AFRL-RI-RS-TR-2022-129

MEDIA FORENSICS INTEGRITY ANALYTICS

PURDUE UNIVERSITY

SEPTEMBER 2022

FINAL TECHNICAL REPORT

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ AIR FORCE MATERIEL COMMAND ■ UNITED STATES AIR FORCE ■ ROME, NY 13441

NOTICE AND SIGNATURE PAGE

AFRL-RI-RS-TR-2022-129 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE CHIEF ENGINEER:

/ S /	**/ S /**
JEFFREY T. CARLO	JAMES S. PERRETTA
Work Unit Manager	Deputy Chief,
	Information Warfare Division
	Information Directorate

REPORT DOCUMENTATION PAGE

1. REPORT DATE	**2. REPORT TYPE**	**3. DATES COVERED**	
		START DATE	**END DATE**
SEPTEMBER 2022	FINAL TECHNICAL REPORT	MAY 2016	APRIL 2022

4. TITLE AND SUBTITLE
MEDIA FORENSICS INTEGRITY ANALYTICS

5a. CONTRACT NUMBER	**5b. GRANT NUMBER** FA8750-16-2-0193	**5c. PROGRAM ELEMENT NUMBER** 62303E
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER** R1ZD

6. AUTHOR(S)
Edward J. Delp (Purdue U); Stefano Tubaro (Polytechnic U of Milan); Mauro Barni (U of Siena); Walter J. Scheirer (U of Notre Dame); C.-C. Jay Kuo (U of Southern California); Nasir Memon (New York U); Luisa A. Verdolvia (U of Naples); Wael Abd-Almageed (U of Southern California)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Purdue University School of Electrical and Computer Engineering 465 Northwestern Avenue West Lafayette IN 47907-2035	**8. PERFORMING ORGANIZATION REPORT NUMBER**

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory/RIGC 525 Brooks Road Rome NY 13441-4505	**10. SPONSOR/MONITOR'S ACRONYM(S)** AFRL/RI	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)** AFRL-RI-RS-TR-2022-129

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The goal of this research was to develop a set of forensics tools to determine the integrity, semantic consistency and evolutionary history of images and videos. We used a data-driven approach. In TA1.1 we designed machine-learning methods trained to analyze the integrity of images and videos. In TA1.2 we used the physical integrity of the scene and the traces of electrical network frequency (ENF) to determine the location and integrity of a video. Our work in TA1.3 generated techniques to correlate the existing objects in the pool in space (spatial coherence) and time (time coherence). We have participated in the NIST evaluations and have delivered our software tools via the APIs for integration with the TA2 efforts.

15. SUBJECT TERMS
Media forensics, machine learning, electrical network frequency. image provenance

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**
a. REPORT	**b. ABSTRACT**	**C. THIS PAGE**		
U	U	U	SAR	212

19a. NAME OF RESPONSIBLE PERSON **JEFFREY T. CARLO**	**19b. PHONE NUMBER** *(Include area code)* **N/A**

TABLE OF CONTENTS

Section **Page**

LIST OF FIGURES

LIST OF TABLES

1.0 SUMMARY

Consumer-grade imaging sensors have become ubiquitous in the past decade. Images and videos, collected from such sensors are used by many entities for public and private communications, including publicity, advocacy, disinformation, and deception. The US Department of Defense (DoD) would like to be able to extract knowledge from and understand this imagery and its provenance. Many images and videos are modified and/or manipulated prior to publication/dissemination. The goal of this research was to develop a set of forensics tools to determine the integrity, semantic consistency and evolutionary history of images and videos.

We have assembled a team of outstanding technical experts from seven universities, with complementary skills and background in computer vision and biometrics, machine learning, digital forensics, as well as signal processing and information theory. We investigated all three subareas of media integrity analysis in Technical Area 1 (TA1). We strongly feel that important synergetic aspects of media integrity are necessary for the successful development of the techniques across the subareas and that the nature of our team positioned us uniquely to broadly attack the research problems in this domain.

In the last decade, a large number of media forensic techniques (mostly model-based) have been developed to assess the integrity of media assets. Unfortunately, the conditions under which media integrity analysis must operate in real life depart significantly from the theoretical settings assumed by such model-based solutions. Rather, we used a data-driven approach. More specifically, in TA1.1 we designed machine-learning methods trained to analyze the integrity of images and videos. In TA1.2 by exploring the physical integrity of the scene, as captured by the images and videos, using the traces of electrical network frequency (ENF) we were able to identify the location where a video was captured and assess its integrity. Complementing the work in TA1.1 and TA1.2, our work in TA1.3 generated techniques to correlate the existing objects in the pool in space (spatial coherence) and time (time coherence). We have participated in the NIST evaluations and have delivered our software tools via the APIs for integration with the TA2 efforts.

The table below is a brief description of the tasks and problems we addressed along with which parts of our team worked on the various parts of TA1.

Task Descriptions

Tasks	Description (Lead Performer)
TA1.1a	Adversary-aware ML methods based on high-order statistics for detection/localization of global color manipulations and multiple compressions (Siena)
TA1.1b	Clone and splicing detection and localization (USC)
TA1.1c	Sensor-wide image attribution (Purdue and Milano)
TA1.1d	Detection of traces left behind by specific editing software suites (Milano)
TA1.1e	Use of application-specific information to secure machine learning forensics tools (Siena)
TA1.2	ENF-based video authentication (NYU)
TA1.3	Develop a comprehensive framework for multimedia phylogeny and manipulation detection (Notre Dame and Milano)

2.0 INTRODUCTION

2.1 Primary Research Goal and Impact

The primary goal of this work was to develop a set of forensics tools to determine the digital, physical and semantic integrity of images and videos. We investigated all three subareas in Technical Area 1 (TA1). The main impact of our work was that instead of relying on traditional model-driven solutions, we developed data-driven solutions, which have several advantages over model-based methods, including not relying on complicated models, the ability to address situations in which the manipulations are only loosely defined, and the ability to exploit possible adversarial settings.

2.2 Innovative Aspects of the Project

Over the last decade, most of the successful forensic approaches developed have relied on custom-tailored models. Unfortunately, the conditions under which media integrity analysis must operate in real-life significantly depart from the theoretical settings assumed in such models. Taking a different path, in this project we focused on designing and developing data-driven solutions for media forensics integrity analytics. Our approach also included adversarial-aware methods which take into consideration how an adversary will operate. In TA1.1, we developed machine-learning methods to distinguish original and manipulated images and videos. In TA1.2, the physical integrity of the scene, as captured by the images and videos, was examined using condition-invariant place recognition techniques and the traces of electrical network frequency (ENF). In TA1.3 we used the principles of media phylogeny to characterize media content, and relationships to support processing of high-level user queries and hypotheses. This includes the construction of representations from media corpora (determination of provenance), content analysis:, semantic-level manipulation detection, and query and hypothesis analysis. We used results in TA1.1 and TA1.2 as input, and described techniques to correlate the existing objects in the pool in space and time.

Our approach for TA1 used recent advances in machine learning, deep learning, computer vision, neuroscience, and phylogenetic representations, and provide a new way of examining forensics integrity. We accomplished our goals because of the depth of experience and the broad areas of expertise that our team possesses. The team produced more than 80 publications, one M.S. thesis and 5 Ph.D. thesis. We have delivered to DARPA more than 30 Docker container software packages. We have also performed well in all the NIST evaluations and in some cases have been ranked as first or second.

3.0 METHODS, ASSUMPTIONS, AND PROCEDURES

In this section we described our methods we used to address several problems in the program. These results are described by each performer sub-team. These results are then linked to our various publications.

3.1 Medifor Project

In this section we describe our efforts on the “main” MediFor project. Later we describe our efforts on the sub-projects: Overhead Forensics and Scientific Integrity.

3.1.1 Purdue University

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Deepfake Detection
Major Technical Approach: Convolution Neural Network and Recurrent Neural Network

The rapid proliferation of a free machine learning-based software tool has made it easy to execute believable face swaps that leave minimal traces of manipulation. These videos are colloquially referred to as “deepfakes”, and could lead to potentially disastrous consequences if used to political or social effect. In this work, we introduce a basic approach to deepfake detection: a convolutional neural network (CNN) to extract frame level features, and then a recurrent neural network (RNN) that uses these features to classify an entire video.

As shown in Figure 1, our method takes image sequence from a video as input and outputs a score indicating if the input video has been manipulated or not. We have empirically chosen a frame sequence length of 16. For entire videos, multiple frame sequences can be extracted to determine the class in a robust manner. The CNN extracts features of importance from the entire frame. These features are extracted for each frame in the sequence, and then input into the RNN, which learns to identify inconsistencies in the features that indicate a manipulation.

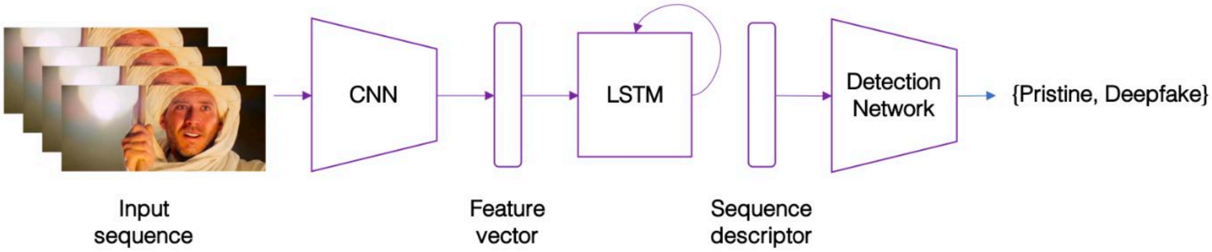


Figure 1. Block Diagram of the Proposed Deepfake Detection Method

Published Paper:
D. Güera, E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks", *IEEE International Conference on Advanced Video and Signal-based Surveillance (AVSS), Auckland, New Zealand, November 2018*. DOI: 10.1109/AVSS.2018.8639163

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Deepfake Detection
Major Technical Approach: Convolution Neural Network and Recurrent Neural Network with Automatic Weighting

Altered and manipulated multimedia is increasingly present and widely distributed via social media platforms. Advanced video manipulation tools enable the generation of highly realistic-looking altered multimedia. While many methods have been presented to detect manipulations, most of them fail when evaluated with data outside of the datasets used in research environments. In this work, we introduce a method based on convolutional neural networks (CNNs) and recurrent neural networks (RNNs) that extracts visual and temporal features from faces present in videos to accurately detect manipulations.

As shown in Figure 2, our method takes image sequence from a video as input and outputs a score indicating if the input video has been manipulated or not. More specifically, for each input frame, we use a face detector to localize the facial region and then pass the cropped facial images into a CNN to extract features. To obtain discriminative features between real and manipulated faces, we also add the additive angular margin loss (also known as ArcFace) on top of the extracted features. Then we predict scores for all frames that indicates if the frames have been altered and weights that indicates the importance of different frames. Since some frames might contain blurry faces where the presence of manipulations might be difficult to detect, the predicted weights can exclude these frames to achieve a robust detection. Lastly, to adaptively merge the prediction from each frame into a single score for the entire video, we combine the features from different frames using a Gated Recurrent Unit (GRU) to obtain the final manipulation score.

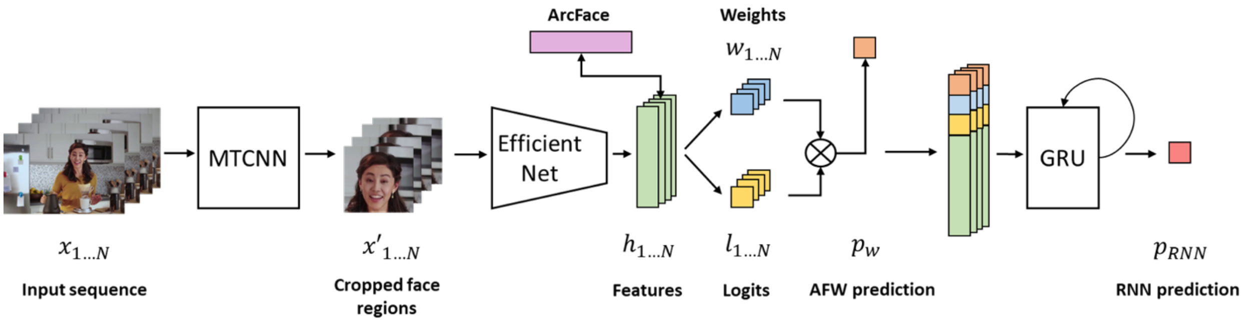


Figure 2. Block Diagram of the Proposed Deepfake Detection Method

[1] D. Güera and E. J. Delp, “Deepfake Video Detection using Recurrent Neural Networks”, *IEEE International Conference on Advanced Video and Signal Based Surveillance*, November 2018, Auckland, New Zealand, https://doi.org/10.1109/AVSS.2018.8639163
[2] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint Face Detection and Alignment using Multitask Cascaded Convolutional Networks”, *IEEE Signal Processing Letters*, vol. 23, April 2016, https://doi.org/10.1109/LSP.2016.2603342
[3] M. Tan and Q. V. Le, “Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks,” *arXiv preprint arXiv:1905.11946*, May 2019, https://arxiv.org/abs/1905.11946
[4] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” *IEEE conference on Computer Vision and Pattern Recognition*, July 2017, Honolulu, HI, https://doi.org/10.1109/CVPR.2017.195
[5] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, “Faceforensics++: Learning to Detect Manipulated Facial Images,” *IEEE International Conference on Computer Vision*, pp. 1–11, October 2019, Seoul, South Korea, https://doi.org/10.1109/ICCV.2019.00009

Published Paper:
D. M. Montserrat, H. Hao, S. K. Yarlagadda, S. Baireddy, R. Shao, J. Horváth, E. Bartusiak, J. Yang, D. Güera, F. Zhu, and E. J. Delp, “Deepfakes Detection with Automatic Face Weighting”, *IEEE Computer Vision and Pattern Recognition Workshops*, June 2020, Seattle, Washington. DOI: 10.1109/CVPRW50498.2020.00342

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Video Manipulation Detection
Major Technical Approach: Metadata-based Machine Learning

Easy-to-use, free or cheap video editing tools are rapidly spreading across the world, greatly expanding the number of manipulated videos being generated. The large majority of detection approaches work in the pixel domain and have proven to be effective, but the adversarial nature of detection vs generation means that these adversaries could quickly adapt their manipulations to avoid detection. By introducing an orthogonal analysis approach, we show that most manipulated videos can be quickly filtered, due to the indicators left in the video metadata by the tools used. These indicators tend to be crucial for playing the videos, so there is a level of stability to this approach.

As shown in Figure 3a, our method takes labeled videos from the MFC dataset to extract all the stream descriptors ("metadata") and create feature vectors representing each video. We then train an ensemble of an SVM and a random forest to distinguish between feature vectors representing pristine and manipulated videos. This trained manipulation detector can then be used directly on any suspect video after similar features have been extracted.

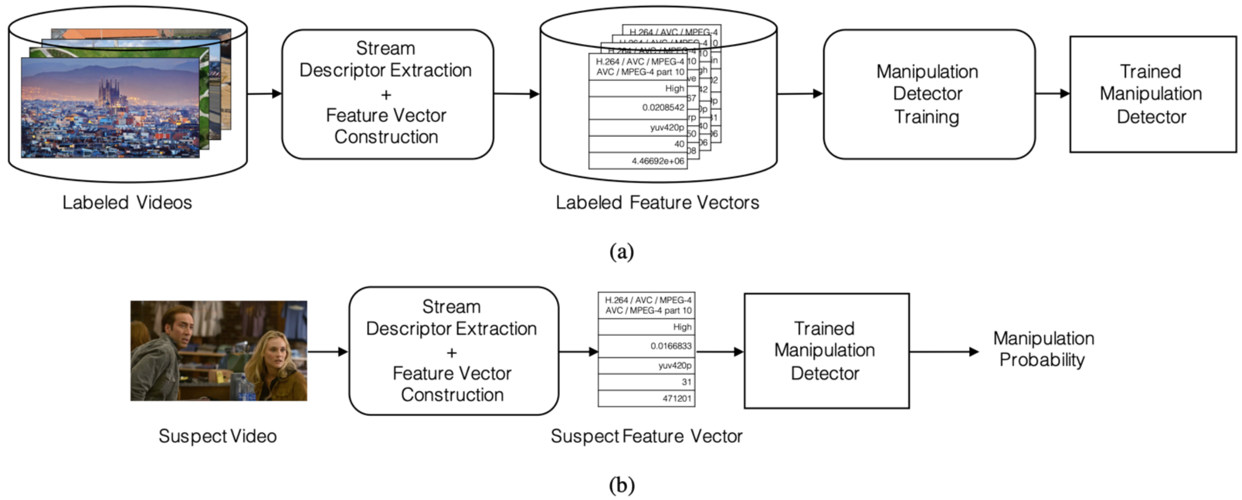

Labeled Videos
Stream Descriptor Extraction + Feature Vector Construction
H.264 / AVC / MPEG-4 AVC / MPEG-4 part 10
High
0.0208542
yuv420p
40
4.46692e+06
Labeled Feature Vectors
Manipulation Detector Training
Trained Manipulation Detector
(a)
Suspect Video
Stream Descriptor Extraction + Feature Vector Construction
H.264 / AVC / MPEG-4 AVC / MPEG-4 part 10
High
0.0166833
yuv420p
31
471201
Suspect Feature Vector
Trained Manipulation Detector
Manipulation Probability
(b)

Figure 3. Block Diagram of the Proposed Video Manipulation Detection Method

Published Paper:
D. Güera, S. Baireddy, P. Bestagini, S. Tubaro, E. J. Delp, "We Need No Pixels: Video Manipulation Detection Using Stream Descriptors", *International Conference on Machine Learning (ICML), Synthetic Realities: Deep Learning for Detecting AudioVisual Fakes Workshop, Long Beach, California, June 2019*. URL: arXiv:1906.08743

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Camera Model Attribution
Major Technical Approach: A Counter-Forensic Method for CNN-Based Camera Model Identification

We propose a counter-forensic method capable of subtly altering images to change their estimated camera model when they are analyzed by any CNN-based camera model detector. Our method can use both the Fast Gradient Sign Method (FGSM) or the Jacobian-based Saliency Map Attack (JSMA) to craft these adversarial images and does not require direct access to the CNN. Figure 4 shows the block diagram of our proposed counter-forensic method.

As shown in Figure 4, the adversarial image generator module takes the set of ***K*** extracted patches as input. When presented with new image patches, our module can work in two different modes. In the first operation mode, the adversarial image generator module uses the fast gradient sign method (FGSM) to do an untargeted image manipulation. The adversarial example is generated as

$$x^* = x + \epsilon sign(\nabla_x J(\Theta, x, y))$$

where ϵ is a parameter that determines the perturbation size.

In the second operation mode, the adversarial image generator module implements the Jacobian-based saliency map attack (JSMA) to do a targeted image manipulation. In this case, we try to perturb the image patches to produce a specific misclassification class ***L'***, different from the true real label ***L*** that is associated with the analyzed image ***I*** and its associated $\boldsymbol{P_k}$ patches.

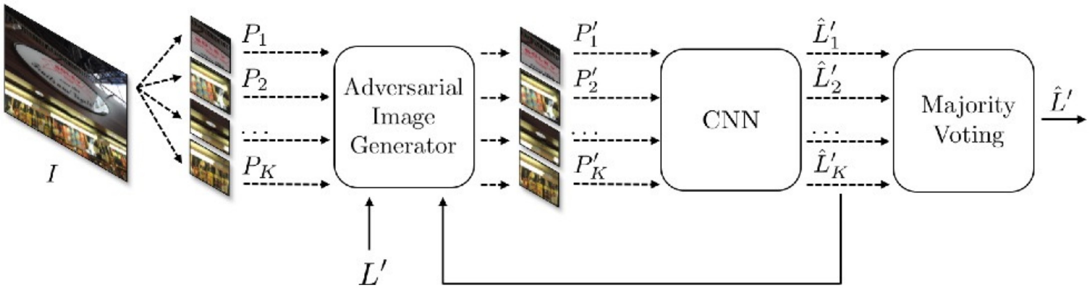


Figure 4. Block Diagram of Our Proposed Method

Published paper:
D. Güera, Y. Wang, L. Bondi, P. Bestagini, S. Tubaro, E. J. Delp, "A Counter-Forensic Method for CNN-Based Camera Model Identification", *IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW*), pp. 1840-1847, Honolulu, Hawaii, July 2017.
DOI: 10.1109/CVPRW.2017.230

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Camera Model Attribution
Major Technical Approach: Supervised approach with CNN

Program Objectives: Reliability Map Estimation for Camera Model Attribution

One important aspect of forensic image analysis is camera model attribution, which is the process of identifying a camera model used to capture an image. We utilize a convolutional neural network (CNN) to analyze an image **I** and identify the camera that acquired it (from a known set of camera models). More specifically, our approach splits an image **I** into patches of uniform size. Then, the CNN analyzes each patch. It estimates the camera model used to capture the patch and the likelihood that the predicted camera model is correct. Next, the likelihood scores are used to construct a reliability map of the same dimensions as the image **I**. The reliability map highlights the regions within an image that the CNN model classifies with the most confidence. It also indicates which regions of the image are more difficult for the CNN to classify. The camera model attributed to the entire image is based on the predictions of the patches which the CNN most confidently classifies. Trained and evaluated on the Dresden Image Database, this approach achieves camera model attribution with high success.

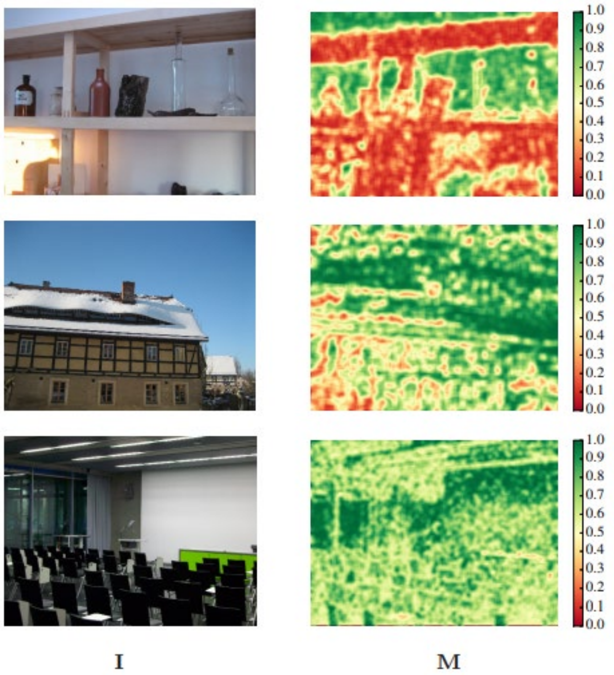


Figure 5. Reliability Map Estimates

The above figure shows the images under analysis and their reliability map estimates. Green indicates high confidence in correct camera attribution, whereas red indicates low confidence in camera attribution.

Published Paper:
D. Güera, S. K. Yarlagadda, P. Bestagini, F. Zhu, S. Tubaro, E. J. Delp, “Reliability Map Estimation for CNN-Based Camera Model Attribution”, *IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2018, Lake Tahoe, NV,
DOI: 10.1109/WACV.2018.00111

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Scanner Model Attribution
Major Technical Approach: Convolution Neural Network

We propose a new robust Convolutional Neural Network (CNN)-based system for scanner model identification. As shown in Figure 1, an input image $\boldsymbol{I}$ is split into sub-images $\boldsymbol{I_s}$ (n × m pixels) in zig-zag form. From each $\boldsymbol{I_s}$, a patch of size 64 × 64 is extracted from a random location. We denote the extracted patch as $\boldsymbol{I_p}$. In the training stage, these extracted patches $\boldsymbol{I_p}$ along with their corresponding labels $\boldsymbol{S}$ are inputs into the network. The proposed system will evaluate two tasks on scanned images: scanner model classification and reliability map generation. In Task 1 (scanner model classification), we assign the predicted scanner labels to both patches $\boldsymbol{I_p}$ and original images $\boldsymbol{I}$. The classification decision for the original image $\boldsymbol{I}$ is obtained by majority voting over the decisions corresponding to its individual sub-images $\boldsymbol{I_s}$. In Task 2, a reliability map is generated based on the majority vote result from Task 1. The pixel values in the reliability map indicate the probability of the corresponding pixel in the original image being correctly classified. The probability of pixel $\boldsymbol{x}$ belonging to scanner $\boldsymbol{s}$ is the average value of the corresponding probabilities for the sub-images that contain pixel $\boldsymbol{x}$.

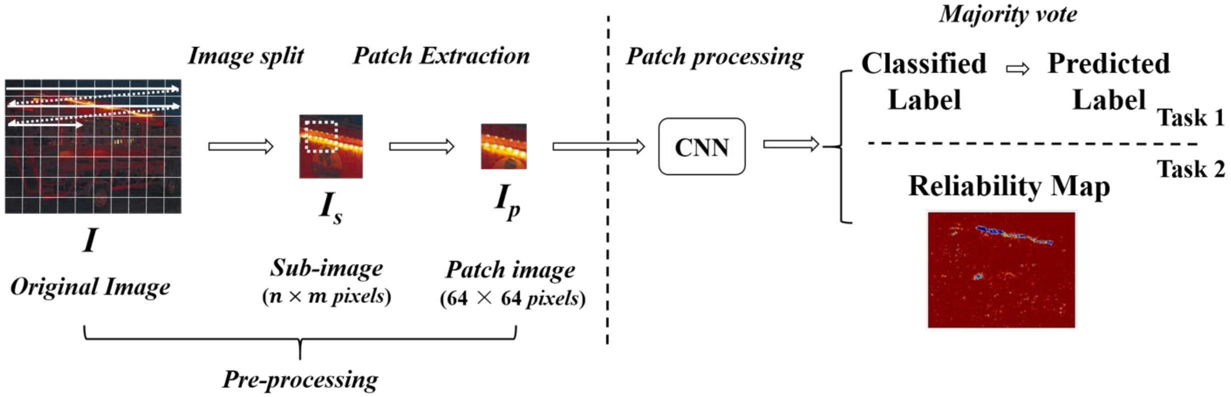


Figure 6. The proposed system for scanner model identification

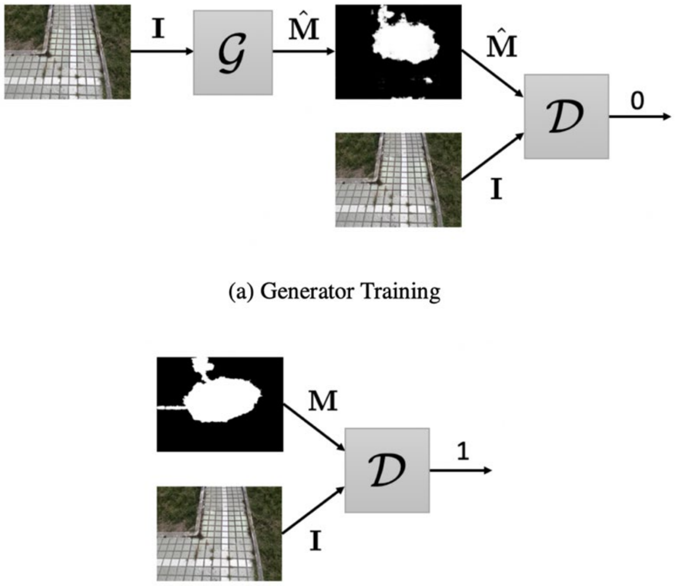
Published paper:
R. Shao, E. J. Delp, "Forensic Scanner Identification Using Machine Learning", *IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI),* Santa Fe, New Mexico, March 2020

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Shadow Removal Detection
Major Technical Approach: GAN-based Detection

Easy-to-use, free or cheap media editing tools are rapidly spreading across the world, greatly expanding the amount of manipulated content being generated. The large majority of traditional detection approaches work by exploiting statistical traces of alterations to prove lack of integrity. Another approach is verifying the physical integrity of the image. With the rise of deep learning, it has become easier to correct miscast shadows automatically. Here, we propose an approach to identify when such automatic techniques have been applied.

As shown in Figure 7, our method utilizes a conditional GAN. Figure 7b shows the discriminator is trained to determine if the mask belongs to the conditional image provided to the GAN (i.e., is it correctly detecting shadow removal), while Figure 7a shows the generator is trained to create accurate shadow removal masks to "fool" the discriminator. After satisfactory training, the discriminator is discarded, and the generator is used to identify manipulated shadow areas.



(a) Generator Training

(b) Discriminator training

Figure 7. Block Diagram of the Proposed Shadow Removal Detection Approach

Published Paper:
S. K. Yarlagadda, D. Güera, D. Mas Montserrat, F. Zhu, P. Bestagini, S. Tubaro, E. J. Delp, "Shadow Removal Detection And Localization For Forensics Analysis", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brighton, UK, May 2019*. DOI: 10.1109/ICASSP.2019.8683695

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Explore various approaches to multimedia analytics
Major Technical Approach: Machine Learning-Based Multimedia Analytics

In this thesis, we have introduced multiple problems in the field of multimedia analytics and solutions based on new machine learning methods.

We stated the problem of logo detection and presented multiple solutions. We first showed that object detection networks such as Faster R-CNN can be successfully adapted to detect logos in the wild. Additionally, we showed that by combining object detection networks with image classification networks such as DenseNet, the detection accuracy can be further improved. In the scenario where the number of training samples is small or non-existent, image synthesis techniques can be applied to create new training samples. We presented two different techniques to create new images. The first technique consists of randomly splicing logo images in a background image. The second technique improves upon the first one and extracts information of the background image in order to splice logo images in a realistic manner. In combination with image synthesis, we showed that bootstrapping techniques can be used to further increase the logo detection accuracy. As logos are largely found on the internet, using weakly-labeled images from the internet is a useful approach for logo detection.

However, there is still a large gap in the accuracy of object detection methods trained with synthetic images and object detection methods trained with real images. Therefore, more research needs to be done in image synthesis techniques, in order to create realistic images with the same statistical properties as the real-world images. It is important to analyze which aspects of the image synthesis process (i.e. location of logos, distortions, and transformations applied to the foreground and background images, statistical variation...) are determining factors when it comes to train networks that generalize well. Methods that find image synthesis pipelines in a data-driven manner instead of ad-hoc hand-crafted pipelines could generate more realistic images that in turn might provide a significant increase in detection accuracy. Furthermore, semi-supervised, self-supervised, and unsupervised techniques (e.g. triplet-loss, Siamese networks...) coupled with few-shot and one-shot learning methods, should be included in the training process in order to fully benefit of a large number of unlabeled images and videos containing logos available on the internet.

We presented the problem of pose estimation. We introduced the MultiView Matching Network (MV-Net) and the Single View Matching Network (SV-Net) to perform pose estimation and tracking. The pair of networks provides an initial estimate of the pose and then it refines the pose in an iterative manner. The same iterative process can be used to track the pose within a video. Additionally, we showed how photorealistic rendering techniques can be used to generate datasets that can be used to train the neural networks, removing the need for manually annotating the 6D pose of images.

While these techniques provide promising results, most AR applications are required to work in real-time and typically in smartphone devices. These devices have memory and computing limitations and require low-weight highly optimized neural networks. Therefore, there is a need for neural networks that are highly accurate while being compact and fast. Neural architecture search (NAS) has proved to be highly successful in image classification and object detection task. Future work includes exploring NAS techniques to design novel neural networks for pose estimation. Furthermore, neural network-based techniques to render images from 3D information could be coupled with 6D pose estimation methods in order to further increase the estimation accuracy.

We presented a new method to detect face manipulations within videos. We showed that combining convolutional and recurrent neural networks achieves high detection accuracies on the DFDC dataset. We described a method to automatically weight different face regions and showed that boosting techniques can be used to obtain more robust predictions. The method processes videos quickly (in less than eight seconds) with a single GPU. Although the results of our experiments are promising, new techniques to generate deepfake manipulations emerge continuously. The modular nature of the proposed approach allows for many improvements, such as using different face detection methods, different backbone architectures, and other techniques to obtain a prediction from features of multiple frames.

Furthermore, this work focuses on face manipulation detection and dismisses any analysis of audio content which could provide a significant improvement of detection accuracy in future works. Incorporating adversarial losses during training and making use of discriminators from GANs could improve the manipulation detection accuracy. Additionally, NAS techniques could be applied to obtain more compact and accurate architectures.

We presented a new method to detect manipulations within satellite images. The wide range of manipulations that can be applied to images and the large diversity of imaging technologies used in satellites makes their detection a challenging problem that still remains unsolved. We introduced an unsupervised splicing detection method. The method consists of an ensemble of generative autoregressive models that estimates the pixel distribution of the image. The method is capable to accurately detect manipulated pixels by selecting the regions of the image where the network predicts a low likelihood value. The presented method is fully unsupervised and doesn't use any prior knowledge from the applied manipulation during training. Our experiments show that the localization accuracy of our method surpasses the previous works and shows that generative models, specially autoregressive-based networks, provide a promising approach to detect pixel-level manipulations.

Published Thesis:
D. Mas Montserrat, "Machine Learning-Based Multimedia Analytics", *PhD dissertation*, Purdue University, West Lafayette, IN, June 2020.
URL: https://engineering.purdue.edu/~ace/thesis/danni/daniel-thesis-final.pdf

3.1.2 University of Siena

Contract Information:
Team: Purdue Team, Siena Sub-Team
POC: Prof. Mauro Barni, University of Siena, barni@diism.unisi.it

Technical Problem: Detection of multiple JPEG compression in the presence of laundering and counter-forensic attacks
Technical Approach: Adversary-aware, SVM-based detection

Our goal is to address the following binary hypotheses testing problem (see Figure 8).

H0): the image has been JPEG compressed once (JPEG camera-native), corresponding to the absence of manipulation. H1): the image has been either compressed twice or compressed, attacked (processed) and then compressed again.

An attack placed in the middle between the two compression stages may correspond to the application of a processing operation or to a counter-forensic (CF) attack, i.e. an attack aimed at erasing single compression traces so as to make the image look like an uncompressed one (single compressed after the final compression stage).

A rich feature is selected to train an adversary-aware SVM detector able to withstand: i) laundering attacks, including resizing, zooming, denoising, median filtering, histogram enhancement, cropping, mirroring, blurring, seam-carving, rotation (up to a certain extent), copy move; ii) the dithering counter-forensic attack in [1] against double JPEG detection:

Specifically, the classification is based on handcrafted (high-order) features selected from both spatial (image pixel) and frequency (DCT) domain to capture the artifacts introduced by double JPEG compression under various attacking conditions. The selected features include well known steganalysis features: the Subtractive Pixel Adjacency Model (SPAM) features for the spatial domain and the CC-PEV features for the DCT domain.

To improve the resilience against various processing and attacks, the SVM is trained with single, double and attacked images for a subset of selected attacks that can be used as a proxy for the most powerful attack (MPA).

Different SVMs are trained for different quality factors of the to-be-inspected image. The quality factor QF of the image is read from the JPEG bit-stream or estimated from image pixels. For double compressed and attacked images, this QF corresponds to the quality factor of the second (last) compression or second quality factor (i.e., QF2).

To train the SVM for a given QF, a large number of training image samples are built under the two hypotheses, starting from uncompressed images, according to the following procedure: for the first class (H0), the uncompressed images are single compressed with quality factor QF; for the second class (H1), the double compressed versions are obtained by compressing the images first with various QF1s and then with QF2 = QF; the attacked images are built by first compressing the images with the QF1s, then attacking them with the selected processing operators (approximating the MPA) and finally re-compressing them with QF2= QF.

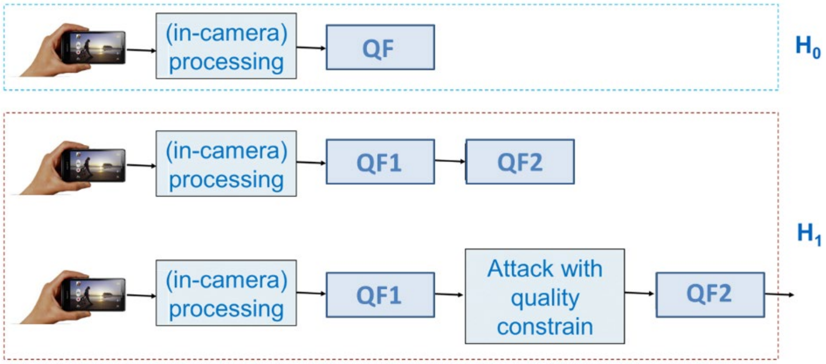


Figure 8. Adversarial Double JPEG Detection Setup

[1] Stamm, Matthew C., et al. "Undetectable Image Tampering through JPEG Compression Anti-forensics." *Image Processing (ICIP), 2010 17th IEEE International Conference on. IEEE,* 2010.

Published Paper:
M. Barni, E. Nowroozi and B. Tondi, "Higher-order, adversary-aware, double JPEG-detection via selected training on attacked samples," *EUSIPCO 2017*

Contract Information:
Team: Purdue Team, Siena Sub-Team
POC: Prof. Mauro Barni, University of Siena, barni@diism.unisi.it

Technical Problem: Detection and localization of contrast adjustment in the presence of JPEG
Technical Approach: Adversary-aware, SVM and CNN-based detection

Development of a contrast adjustment detector based on SVMs
Adjustment of contrast and lighting conditions of image subparts is often performed during forgery creation. In this work, we study the problem of contrast adjustment detection in adversarial setting, with specific focus on the adaptive histogram equalization (AHE), applying contrast enhancement on a local basis. The detection of AHE is more challenging than the detection of global contrast enhancement operators (like for instance gamma correction and histogram stretching), since it does not introduce easily identifiable artifacts in the image histogram.

We first identified a suitable set of high-order features, named CRSPAM, inspired to the rich feature model for color images proposed in [1], that uses the SPAM set as base feature set. We keep the dimensionality to this feature set limited (1372 features in total) so to be able to train an SVM classifier, without needing to resort to multiple classifier approaches (e.g., ensemble classifiers) that would be more difficult to train in the adversary-aware modality.

A JPEG-aware version of the SVM detector is trained to improve the robustness against JPEG compression, which is found to be the most harmful laundering attack against contrast manipulation detectors. Moreover, JPEG compression is a very common post-processing operation in practice. Figure 9 illustrates the detection task considered.

A general SVM-based detector is developed that works directly on the image pixel as described in the following: it first gets an estimate of the Quality Factor (QF) of the JPEG compression by exploiting the idempotency property of the JPEG compression (that is, the fact that JPEG compression with the same QF is an -almost- idempotent operation); then, such an estimate is used to select the SVM to be used for testing among a pool of SVM classifiers - each one trained on a QF in a set - according to the minimum distance criterium. By exploiting the estimation, this approach gives superior performance with respect to using an SVM classifier trained on the mixture of QFs.

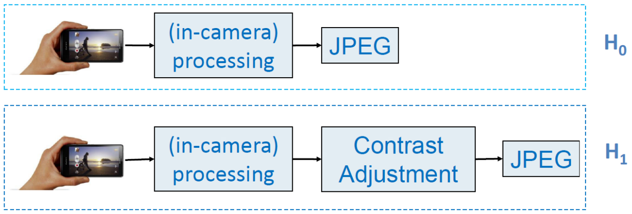


Figure 9. Contrast Manipulation Detection Task With the JPEG Laundering

Development of a CNN for the detection and localization of generic contrast adjustment

By exploiting the superior performance of deep learning-based methods, we design a JPEG-aware, patch-based CNN that can be applied for the detection and the localization of a generic contrast adjustment and localization task. The general setup is the same one illustrated in Figure 9.

The capability of the detector to generalize to unseen contrast manipulations is achieved by considering contrast adjustments of different nature for training. Specifically, the manipulations considered, in equal percentage, are: Adaptive Histogram Equalization (Cl-AHE), Histogram Stretching (HS), and Gamma Correction (γ Corr) - both a compression ($\gamma < 1$) and an expansion ($\gamma > 1$) of the contrast.

The localization capability is achieved by training the network on small image patches (64x64), that makes it possible to apply the method on sliding windows on a JPEG image to get a tampering map of contrast manipulation. For the detection task, a final score is obtained by aggregating the soft CNN outputs for each patch.

With regard to the design of the architecture, we depart from the trend in multimedia forensics of using pretty shallow architectures [2], and consider a pretty deep architecture with a large number of filters with fixed kernel size (3x3), that allows to get superior performance compared to that achievable with the state-of-the-art architecture. Among the main features of this structure, we point out: i) the use of many convolutions before the first pooling layer permits to consider a large receptive field for each neuron, which is good to capture relationships among pixels in large neighborhoods; ii) the stride fixed to 1 permits to retain as much spatial information as possible.

[1] M. Goljan, J. Fridrich, and R. Cogranne, “Rich Model for Steganalysis of Color Images”, *IEEE WIFS 2014*
[2] B. Bayar and M. C. Stamm, “A deep learning approach to universal image manipulation detection using a new convolutional layer,” *ACM IH&MMSEC 2016*

Published Paper:
M. Barni, A.Costanzo, E. Nowroozi and B. Tondi, " CNN-based detection of generic contrast adjustment with JPEG post-processing," *IEEE ICIP 2018*

Contract Information:
Team: Purdue Team, Siena Sub-Team
POC: Prof. Mauro Barni, University of Siena, barni@diism.unisi.it

Technical Problem: Splicing detection and localization in JPEG images based on inconsistencies of double JPEG compression
Technical Approach: CNN-based estimation of primary quantization matrix, clustering and morphological reconstruction.

We design an approach to perform localization of spliced regions in a JPEG image, based on the analysis of recompression traces and inconsistences in the former QF (for images compressed twice), under the assumption that different spliced areas and background have been compressed with a different Quality Factor QF1. By assuming that the forged image is finally saved in JPEG format (hence a final JPEG compression is performed with QF2), the splicing can be revealed by looking at the inconsistencies in the 8x8 quantization matrix of former compression (Q1 matrix).

To address this task, the first step is the development of a method for accurate estimation of the Q1 matrix, that can work well under general operative conditions (both when the quality of the former JPEG compression is lower or higher than that of the second JPEG, both under aligned or non-aligned double JPEG compression) and on small window size or patch size (the patch size determines the resolution capability of the localization task). State-of-the art solutions for this problem are statistical model-based approaches, that work only under specific settings and have poor performance on small patches.

Our method for Q1 matrix estimation based on CNNs, that can work under very general conditions, resorts to a dense CNN architecture (DenseNet), which is modified in order to obtain a structure suitable for the estimation. The output of the network is a 15-dim vector of the first quantization coefficients (the most relevant and discriminative ones) of the Q1 matrix in zig-zag order. The CNN is trained to minimize the logarithm of the hyperbolic cosine of the prediction error (log-cosh loss function). The estimation is done on floating point, then rounding to integer is performed to get the final estimated Q1 integer vector. The network works on 64x64x3 patches.

Tampering localization is performed starting from the result of the Q1 estimation (obtained by applying the CNN estimator on sliding windows), performing clustering and map refinement. Figure 10 illustrates the scheme of the proposed method. The main steps are described in the following. The Spectral Clustering (SC) algorithm is considered to localize and identify the multiple spliced regions in the tampered image based on the Q1 estimation result, that is the Q1 map. The number of clusters k considered for the clustering is estimated by means of a CNN model directly trained on the Q1 maps. Finally, the output tampering map is refined by including spatial information into the process via morphological reconstruction (MR). MR helps to remove noisy (small) clusters and reassign ring clusters along the boundary of big clusters to the internal clusters.

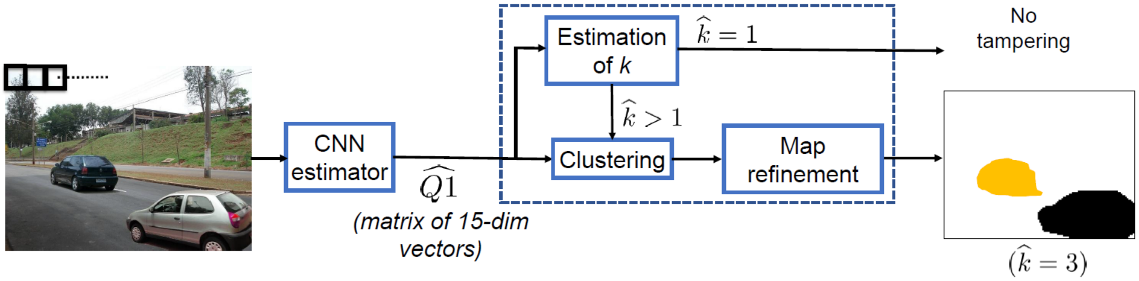


Figure 10. Scheme of the Proposed Method for Image Tampering Localization of JPEG Images

Published Paper:
Y. Niu, B.Tondi, Y.Zhao, M.Barni “Primary Quantization Matrix Estimation from Double JPEG images via Convolutional Neural Network”, *IEEE SPL 2019*
Y. Niu, B. Tondi, Y. Zhao, R. Ni, M. Barni, "Image Splicing Detection, Localization and Attribution via JPEG Primary Quantization Matrix Estimation and Clustering," in *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 5397-5412, November 2021. DOI: 10.1109/TIFS.2021.3129654

Contract Information:
Team: Purdue Team, Siena Sub-Team
POC: Prof. Mauro Barni, University of Siena, barni@diism.unisi.it

Technical Problem: Improving the security of machine-learning based manipulation detectors
Technical Approach: Multiple classification, SVM-based

We propose an architecture based on multiple classifiers that could improve the security against targeted counter-forensic attacks, with respect to standard two class classifiers. Specifically, we adapt to our case the one-and-half class classifier (1.5C) proposed in [1].

The 1.5C classifier consists of 3 parallel intermediate classifiers followed by a combination classifier. The parallel of 3 classifiers consists of: a two-class classifier, a one-class classifier trained on the pristine class (H0) and a one-class classifier trained on manipulated class (H1). The final combination classifier is a one-class classifier (trained on the pristine class) that is fed with the soft outputs of the 3 intermediate classifiers. Figure 11 illustrates the architecture of the 1.5C classifier.

We implemented the classifiers by means of SVMs. With regard to the features used to train each SVM, we considered the 686 features of the well-known SPAM feature set extracted from the luminance (V) channel of the image.

Training the 1C classifiers requires particular care given that the 1C classifiers tend by construction to have poor performance with respect to the alternative class, i.e., the class of samples not used for training. Since we wish to avoid missed detection events (H1 detected as H0), in order to increase the security against integrity violation attacks, we validated the 1C SVMs by weighting differently the two type of error probabilities. Specifically, in order to avoid a missed detection event, that an attacker normally aims at, we weight more the probability of a missed detection and minimize the weighted error probability term. This corresponds to consider a relatively small closed acceptance region for the 1Cs trained on the pristine class (both the intermediate and the final 1C combination classifier) and a relatively large closed region for the 1C trained on the manipulated class.

The performance of the 1.5C classifier are assessed by considering several manipulations of different kind: the adaptive histogram equalization (Cl-AHE), resizing and median filtering. Our experiments aim at showing that the 1.5C classifier can achieve good performance in the absence of attacks (comparable to those achieved by a standard 2C classifier), but at the same time is more difficult to attack in a perfect knowledge (or white box) scenario, thus achieving a better security with respect to the 2C classification case.

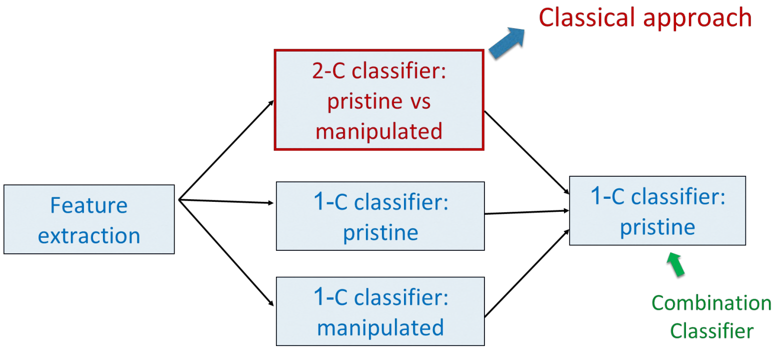


Figure 11. Architecture of the One-and-a-Half Class (1.5C) Classifier

[1] B. Biggio, I Corona, et al, “One-and-a-Half-Class Multiple Classifier Systems for Secure Learning Against Evasion Attacks at Test Time”, *International Workshop on Multiple Classifier Systems,* 2015.

Published Paper:
M. Barni, E. Nowroozi and B. Tondi, " Improving the security of image manipulation detection through one-and-a-half-class multiple classification", *MTAP 2019*.

Contract Information:
Team: Purdue Team, Siena Sub-Team
POC: Prof. Mauro Barni, University of Siena, barni@diism.unisi.it

Technical Problem: Improving the security of machine learning-based manipulation detectors

Technical Approach: Features randomization, SVM and CNN-based detection

Securing data-driven detectors by means of Random Feature Selection (RFS)
We address the problem of data-driven image manipulation detection in the presence of an attacker with limited knowledge about the detector. Specifically, we assume that the attacker knows the architecture of the detector, the training data, and the class of features the detector can rely on. In order to develop a more secure detector, our proposal is to resort to a random feature selection mechanism and build the detector by relying on a random subset of the initial features. The random selection works then as a kind of secret key, and indirectly forces the attack to exit from the white-box scenario. The security model adopted is depicted in Figure 12. As shown in the picture, we assume that the attack is carried out directly in the feature space.
The effectiveness of such approach is demonstrated both in theory and in practice.

Regarding the theory, we were able to characterize, under some simplifying assumptions on the statistical model, the impact that the number of selected features has on the accuracy of the detector, both in absence and in presence of attacks. The theoretical analysis confirms that the detector based on randomization is good for security, that is, thanks to random feature selection, the security of the detector significantly increases at the expense of a negligible loss of performance in the absence of attacks.

The effectiveness of the randomization in real word applications is assessed by considering two specific manipulation detection problems, the adaptive histogram equalization (CLAHE) and the median filtering detection (MF), by considering the case of SVM-based detection. The feature set considered is the SPAM set of dimensionality N = 686. Given its ignorance about the exact feature set, it is assumed that the adversary attacks a version of the detector based on the entire feature set. The gradient-based targeted attack developed in [1] against SVM models is considered for these experiments We also evaluated the security of the randomized feature detector by considering a scenario more favorable to the attacker, where the attacker is aware of the defense mechanism and the number of random features k is publicly available. In this scenario, we consider an attack that targets an expected version of the classifier, obtained by averaging a number of predictions obtained by different classifiers, trained on different random subset of k features.

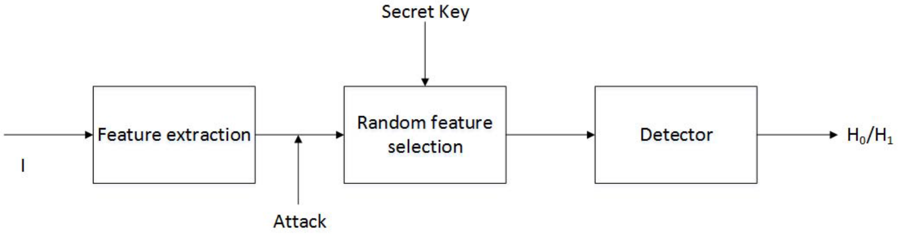


Figure 12. Scheme of a General RFS Detector

Effectiveness of Random Deep Feature Selection (RDFS) for securing CNN-based detectors
We investigate if the random feature selection approach developed for the SVM case to improve the robustness of forensic detectors against targeted attacks, can be extended to detectors based on deep learning features and CNNs in particular. This corresponds to study the transferability of the

so-called adversarial examples targeting an original CNN image manipulation detector to a detector that rely on a random subset of the features extracted from the flatten layer of the original network.

Given a network model, the approach can be extended by working as follows: i) extracting the feature vector at the output of the convolutional part of the network, i.e. the flatten layer, and then, ii) applying the random selection to that set of features, by selecting k out of the total number N of features according to a secret key. The reduced set of features obtained in this way is used to train an SVM or a Neural Network (NN) detector (dense layer). The scheme of the proposed Random Deep Features Selection (RDFS) detector is illustrated in Figure 13. The same scheme is applied during both training and testing, with the same secret key.

The adversarial attack is carried out in the pixel domain, as shown in Figure 13, since this is the case with common adversarial examples against deep learning classifiers. We assume that the attacker does not know the existence of the randomization strategy and then he targets the original CNN classifier (hence implementing a so-called vanilla attack). To implement the attack, we considered several gradient-based, iterative, attacks. Specifically, adversarial examples were built by relying on the following well known algorithms: the original box constrained L-BFGS, the Fast Gradient Sign Method, namely FGSM, and the Projected Gradient Descent (PGD) Attack.

Experiments were carried out on several different image manipulation detection tasks, that is, the resizing, adaptive histogram equalization (CLAHE), the median filtering detection (MF), considering two state-of-the-art CNNs that have been used for multimedia forensic tasks, corresponding to a shallow (BayarNet, [2]) and a deeper (ConstrastNet, [3]) architecture, with a different number of features in the dense layer. These tests reveal that the dangerousness of adversarial examples can indeed be mitigated by the proposed RDFS scheme. However, the degree of effectiveness of RDFS depends on the detection task, the kind of attack and the network. In fact, in several cases, the mismatch between the original classifier targeted by the attack and the one used for classification by itself the attack success rate even without any feature randomization.

As future investigation, the development of adversarial attacks with increased strength would allow a better assessment of the RDFS defense mechanism.

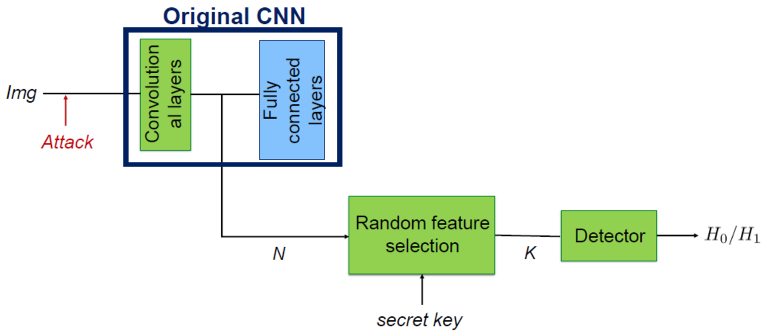


Figure 13. Scheme of the RDFS Detector

[1] Z. Chen, B. Tondi, X. Li, R. Ni, Y. Zhao, and M. Barni." A gradient-based pixel-domain attack against SVM detection of global image manipulations." WIFS 2017
[2] B. Bayar and M. Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer," in ACM IH&MMSEC 2016.
[3] M. Barni, A. Costanzo, E. Nowroozi, and B. Tondi, "CNN-based detection of generic contrast adjustment with JPEG postprocessing," ICIP 2018

Published Paper:
Z. Chen, B. Tondi, X. Li, R. Ni, Y. Zhao and M. Barni, "Secure Detection of Image Manipulation by Means of Random Feature Selection," IEEE T-IFS 2019

M. Barni, E. Nowroozi, B. Tondi and B. Zhang, "Effectiveness of Random Deep Feature Selection for Securing Image Manipulation Detectors Against Adversarial Examples," IEEE ICASSP 2020

3.1.3 University of Southern California

Contract Information:
Team: Purdue Team, USC Sub-Team
POC: Prof. C.-C. Jay Kuo, University of Southern California, cckuo@sipi.usc.edu, 213-740-4658

Major Technical Problem: Image Splicing Localization
Major Technical Approach: Contrastive PCA++

In this work, we propose a new data visualization and clustering technique for discovering discriminative structures in high-dimensional data. This technique, referred to as cPCA++, utilizes the fact that the interesting features of a "target" dataset may be obscured by high variance components during traditional PCA. By analyzing what is referred to as a "background" dataset (i.e., one that exhibits the high variance principal components but not the interesting structures), our technique is capable of efficiently highlighting the structure that is unique to the "target" dataset. Similar to another recently proposed algorithm called "contrastive PCA" (cPCA), the proposed cPCA++ method identifies important dataset-specific patterns that are not detected by traditional PCA in a wide variety of settings. However, the proposed cPCA++ method is significantly more efficient than cPCA, because it does not require the parameter sweep in the latter approach. We applied the cPCA++ method to the problem of image splicing localization. In this application, we utilize authentic edges as the background dataset and the spliced edges as the target dataset. The proposed method is significantly more efficient than state-of-the-art methods, as the former does not require iterative updates of filter weights via stochastic gradient descent and backpropagation, nor the training of a classifier. Furthermore, the cPCA++ method is shown to provide performance scores comparable to the state-of-the-art Multi-task Fully Convolutional Network (MFCN).

In summary, the main contributions are:

(1) Proposed a new dimensionality reduction method which is based on Principle Component Analysis (PCA), which is inspired by contrastive PCA, called cPCA++.

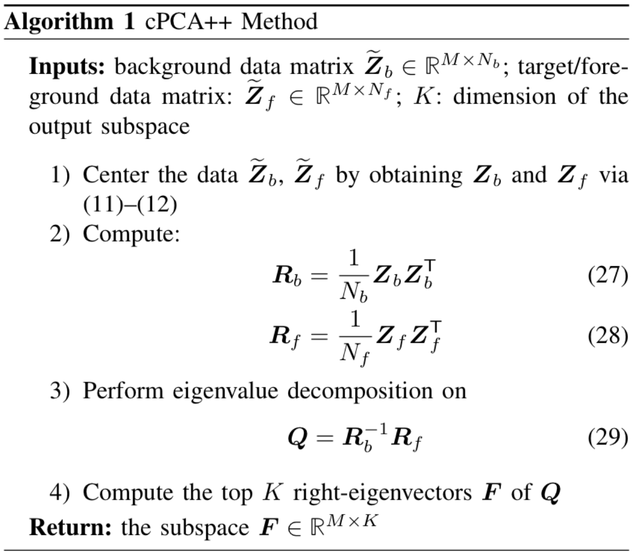
(2) Proposed a new approach for image splicing localization based on cPCA++, which is significantly more efficient than state-of-the-art techniques such as the Multi-task Fully Convolutional Network (MFCN), and still achieves comparable performance scores.

(3) Conducted experiments on different scenarios (datasets) of image splicing localization.

(4) Compared with different dimensionality reduction techniques in terms of separability of foreground and background information.

(5) Compared computational time performance with cPCA and t-SNE. Our proposed method is the most efficient.

(6) Compared image splicing localization using cPCA++ and MFCN. Experimental result shows that our proposed method can achieve comparable or even better performance than MFCN, but with much efficient algorithm.



```
Algorithm 1 cPCA++ Method
  Inputs: background data matrix $\widetilde{\boldsymbol{Z}}_b \in \mathbb{R}^{M \times N_b}$; target/foreground data matrix: $\widetilde{\boldsymbol{Z}}_f \in \mathbb{R}^{M \times N_f}$; $K$: dimension of the output subspace
    1) Center the data $\widetilde{\boldsymbol{Z}}_b$, $\widetilde{\boldsymbol{Z}}_f$ by obtaining $\boldsymbol{Z}_b$ and $\boldsymbol{Z}_f$ via (11)–(12)
    2) Compute:
         $\boldsymbol{R}_b = \frac{1}{N_b}\boldsymbol{Z}_b\boldsymbol{Z}_b^{\mathsf{T}}$   (27)
         $\boldsymbol{R}_f = \frac{1}{N_f}\boldsymbol{Z}_f\boldsymbol{Z}_f^{\mathsf{T}}$   (28)
    3) Perform eigenvalue decomposition on
         $\boldsymbol{Q} = \boldsymbol{R}_b^{-1}\boldsymbol{R}_f$   (29)
    4) Compute the top $K$ right-eigenvectors $\boldsymbol{F}$ of $\boldsymbol{Q}$
  Return: the subspace $\boldsymbol{F} \in \mathbb{R}^{M \times K}$
```

Figure 14. cPCA++ Algorithm

Published paper:
Salloum, R. and Kuo, C.C.J., "Efficient image splicing localization via contrastive feature extraction," 2019. arXiv:1901.07172

Contract Information:
Team: Purdue Team, USC Sub-Team
POC: Prof. C.-C. Jay Kuo, University of Southern California, cckuo@sipi.usc.edu, 213-740-4658

Major Technical Problem: Image Splicing Localization
Major Technical Approach: Interpretable Convolutional Neural Network

The model parameters of convolutional neural networks (CNNs) are determined by backpropagation (BP). In this work, we propose an interpretable feedforward (FF) design without any BP. The FF design adopts a data-centric approach. It derives network parameters of the current layer based on data statistics from the output of the previous layer in a one-pass manner. To construct convolutional layers, we develop a new signal transform, called the Saab (Subspace approximation with adjusted bias) transform. It is a variant of the principal component analysis with an added bias vector to annihilate activation's nonlinearity. Multiple Saab transforms in cascade yield multiple convolutional layers. As to fully connected layers, we construct them using a cascade of multi-stage linear least squared regressors. The classification and robustness performances of BP- and FF-designed CNNs applied to the MNIST and the CIFAR-10 datasets are compared. Finally, we comment on the relationship between BP and FF designs.

An interpretable CNN design based on the FF methodology was proposed in this work. We design FC layers in CNN as a sequence of label-guided linear least-squared regressors. Interpretable CNN design offers an alternative approach to CNN filter weights selection. We conducted extensive comparison between these two design methodologies, in the aspect of robustness, training complexity, classification performance, etc.

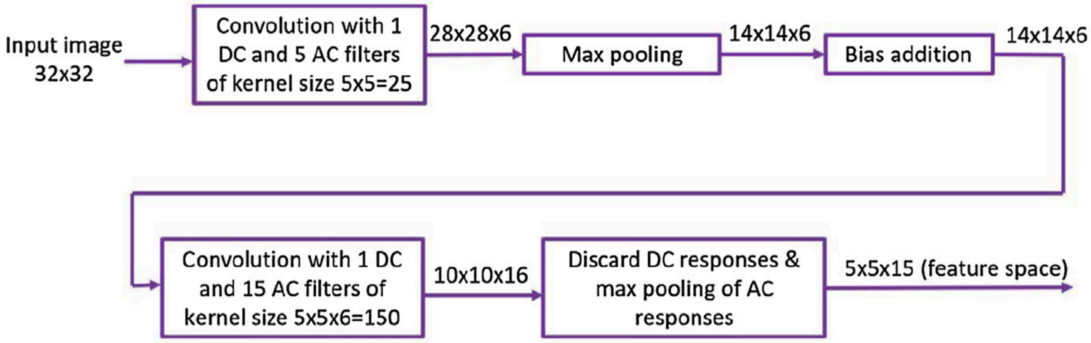

Input image
32x32
Convolution with 1 DC and 5 AC filters of kernel size 5x5=25
28x28x6
Max pooling
14x14x6
Bias addition
14x14x6
Convolution with 1 DC and 15 AC filters of kernel size 5x5x6=150
10x10x16
Discard DC responses & max pooling of AC responses
5x5x15 (feature space)

Figure 15. Summary of the FF Design of the First Two Convolutional Layers of the LeNet-5

Published paper:
Kuo, C.C.J., Zhang, M., Li, S., Duan, J. and Chen, Y., "Interpretable convolutional neural networks via feedforward design," *Journal of Visual Communication and Image Representation*, **60**, Apr 2019, pp.346-359. DOI: 10.1016/J.JVCIR.2019.03.010

Contract Information:
Team: Purdue Team, USC Sub-Team
POC: Prof. C.-C. Jay Kuo, University of Southern California, cckuo@sipi.usc.edu, 213-740-4658

Major Technical Problem: Image Splicing Localization
Major Technical Approach: Multi-task Fully Convolutional Network (MFCN)

In this work, we propose a technique that utilizes a fully convolutional network (FCN) to localize image splicing attacks. We first evaluated a single-task FCN (SFCN) trained only on the surface label. Although the SFCN is shown to provide superior performance over existing methods, it still provides a coarse localization output in certain cases. Therefore, we propose the use of a multi-task FCN (MFCN) that utilizes two output branches for multi-task learning. One branch is used to learn the surface label, while the other branch is used to learn the edge or boundary of the spliced region. We trained the networks using the CASIA v2.0 dataset, and tested the trained models on the CASIA v1.0, Columbia Uncompressed, Carvalho, and the DARPA/NIST Nimble Challenge 2016 Science datasets. Experiments show that the SFCN and MFCN outperform existing splicing localization algorithms, and that the MFCN can achieve finer localization than the SFCN.

We present an effective solution to the splicing localization problem based on a fully convolutional network (FCN). The base network architecture is the FCN VGG-16 architecture with skip connections, but we incorporate several modifications, including batch normalization layers and class weighting. We first evaluated a single-task FCN (SFCN) trained only on the surface label or ground truth mask, which classifies each pixel in a spliced image as spliced or authentic. Although the SFCN is shown to provide superior performance over existing techniques, it still provides a coarse localization output in certain cases. Thus, we next propose the use of a multi-task FCN (MFCN) that utilizes two output branches for multi-task learning. One branch is used to learn the surface label, while the other branch is used to learn the edge or boundary of the spliced region. It is shown that by simultaneously training on the surface and edge labels, we can achieve finer localization of the spliced region, as compared to the SFCN.

In summary, the main contributions are:

(1) Single-task fully convolutional network is utilized. We added batch normalization layers and class weighting.

(2) Proposed multi-task fully convolutional network, utilized two output branches for multi-task learning. One branch is used to learn the surface label, while the other branch is used to learn the edge or boundary of the spliced region.

(3) Evaluated two different inference approaches. The first approach utilizes only the surface output probability map in the inference step. The second approach, which is referred to as the edge-enhanced MFCN, utilizes both the surface and edge output probability maps to achieve finer localization.

(4) Trained the SFCN and MFCN using the CASIA v2.0 dataset and tested the trained networks on the CASIA v1.0, Columbia Uncompressed, Carvalho, and the DARPA/NIST Nimble Challenge 2016 Science datasets.

(5) Showed that after applying various postprocessing operations such as JPEG compression, blurring, and addition of noise to the spliced images, the SFCN and MFCN methods still outperform the existing methods.

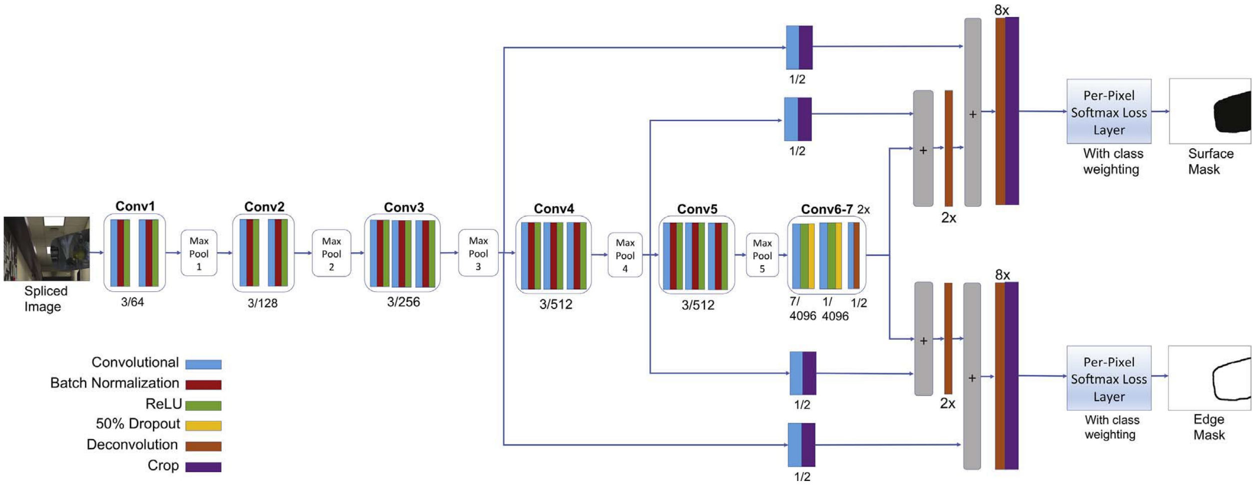
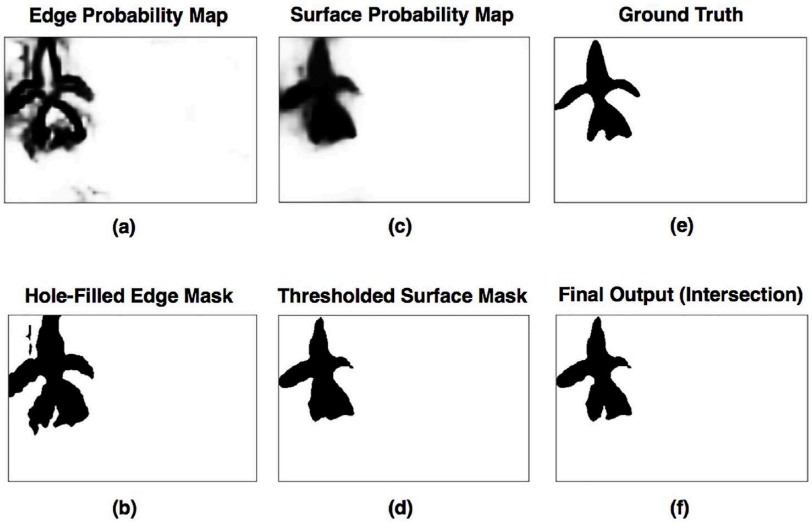


Figure 16. The MFCN Architecture for Image Splicing Localization



Edge Probability Map (a)

Surface Probability Map (c)

Ground Truth (e)

Hole-Filled Edge Mask (b)

Thresholded Surface Mask (d)

Final Output (Intersection) (f)

Figure 17. Illustration of MFCN Inference with Edge Enhancement: (a) Edge Probability Map, (b) Hole-Filled, Thresholded Edge Mask, (c) Surface Probability Map, (d) Thresholded Surface Mask, (e) Ground Truth Mask, and (f) Final System Output Mask

Published paper:
Salloum, R., Ren, Y. and Kuo, C.C.J., “Image splicing localization using a multi-task fully convolutional network (MFCN),” *Journal of Visual Communication and Image Representation*, **51**, Feb 2018, pp.201-209. DOI: 10.1016/J.JVCIR.2018.01.010

Contract Information:
Team: Purdue Team, USC Sub-Team
POC: Prof. C.-C. Jay Kuo, University of Southern California, cckuo@sipi.usc.edu, 213-740-4658

Major Technical Problem: Image Splicing Localization
Major Technical Approach: Theoretical Understanding of Convolutional Neural Networks

There is a resurging interest in developing a neural-network-based solution to the supervised machine-learning problem. The convolutional neural network (CNN) will be studied in this lecture note. We introduce a rectified-correlations on a sphere (RECOS) transform as a basic building block of CNNs. It consists of two main concepts: 1) data clustering on a sphere and 2) rectification. We then interpret a CNN as a network that implements the guided multilayer RECOS transform with two highlights. First, we compare the traditional single-layer and modern multilayer signal-analysis approaches, point out key areas that enable the multilayer approach, and provide a full explanation to the operating principle of CNNs. Second, we discuss how guidance is provided by labels through backpropagation (BP) in the training.

To discuss the effect of label guidance in backpropagation, we first compare two network initialization schemes: 1) the random initialization and 2) the k-means initialization. For the latter, we perform k-means at each layer based on its corresponding input data samples (with zero mean and unit-length normalization), and we repeat this process from the input to the output layer after layer. Today, random initialization is commonly adopted. Based on the previous discussion, we expect the k-means initialization to be a better choice. This is verified by our experiments in the LeNet-5 applied to the MNIST data set.

In summary, the main contributions are:

(1) Neural networks architecture evolution. We provide a survey on the architecture evolution of neural networks, including computational neurons, multilayer perceptron (MLPs), and CNNs.

(2) Signal analysis via multilayer RECOS transform. We point out the differences between the single- and multilayer signal-analysis approaches and explain the working principle of the multilayer RECOS transform.

(3) Network initialization and guided anchor vector update. We carefully examine the CNN initialization scheme since it can be viewed as an unsupervised clustering and the CNN self-organization property can be explained. We compared two network initialization schemes: random initialization and K-means initialization.

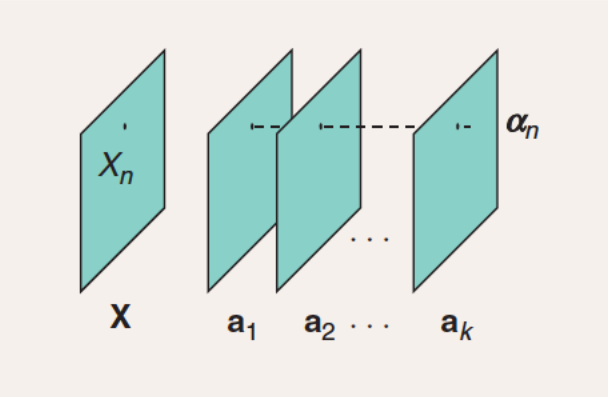


Figure 18. The Visualization of the Anchor-position Vector α_n

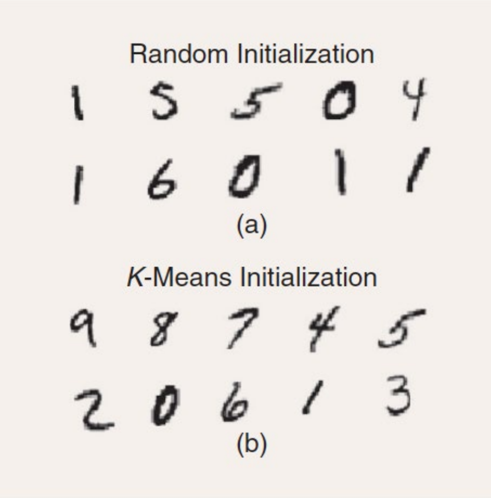


Figure 19. The Comparison of MNIST Unsupervised Classification Results of the LeNet-5 Architecture with the (a) Random and (b) K-Means Initializations, where the Images that are Closest to Centroids of Ten Output Nodes are Shown

Published paper:
Kuo, C.C.J., "The CNN as a guided multi-layer RECOS transform," *IEEE Signal Processing Magazine*, **34**, 3, May 2017, pp. 81-89. DOI: 10.1109/MSP.2017.2671158

Contract Information:
Team: Purdue Team, USC Sub-Team
POC: Prof. C.-C. Jay Kuo, University of Southern California, cckuo@sipi.usc.edu, 213-740-4658

Major Technical Problem: Image Splicing Localization
Major Technical Approach: PixelHop: a successive subspace learning method

A new machine learning methodology, called successive subspace learning (SSL), is introduced in this work. SSL contains four key ingredients: (1) successive near-to-far neighborhood expansion; (2) unsupervised dimension reduction via subspace approximation; (3) supervised dimension reduction via label-assisted regression (LAG); and (4) feature concatenation and decision making. An image-based object classification method, called PixelHop, is proposed to illustrate the SSL design. It is shown by experimental results that the PixelHop method outperforms the classic CNN model of similar model complexity in three benchmarking datasets (MNIST, Fashion MNIST and CIFAR-10). Although SSL and deep learning (DL)have some high-level concept in common, they are fundamentally different in model formulation, the training process and training complexity. Extensive discussion on the comparison of SSL and DL is made to provide further insights into the potential of SSL.

In this work, we have three major contributions. First, we introduce the SSL notion explicitly and make a thorough comparison between SSL and DL. Second, the LAG unit using soft pseudo labels is novel. Third, we use the PixelHop method as an illustrative example for SSL, and conduct extensive experiments to demonstrate its performance. In contrast with traditional subspace methods, SSL examines the near-and far-neighborhoods of a set of selected pixels. It uses the training data to learn three sets of parameters: (1) Saab filters for unsupervised dimension reduction in the PixelHop unit, (2) regression matrices for supervised dimension reduction in the LAG unit, and (3) parameters required by the classifier. Extensive experiments were conducted on MNIST, Fashion MNIST and CIFAR-10 to demonstrate the superior performance of the PixelHop method in terms of classification accuracy and training complexity.

In summary, the main contributions are:

(1) introduced the SSL notion explicitly and make a thorough comparison between SSL and DL
(2) introduced novel idea of LAG unit for feature transformation
(3) PixelHop method as illustrative example for SSL and conducted experiments to demonstrate its performance
(4) Applied PixelHop method on object classification on MNIST, Fashion MNIST and CIFAR-10 dataset, which achieves comparable performance as CNN.

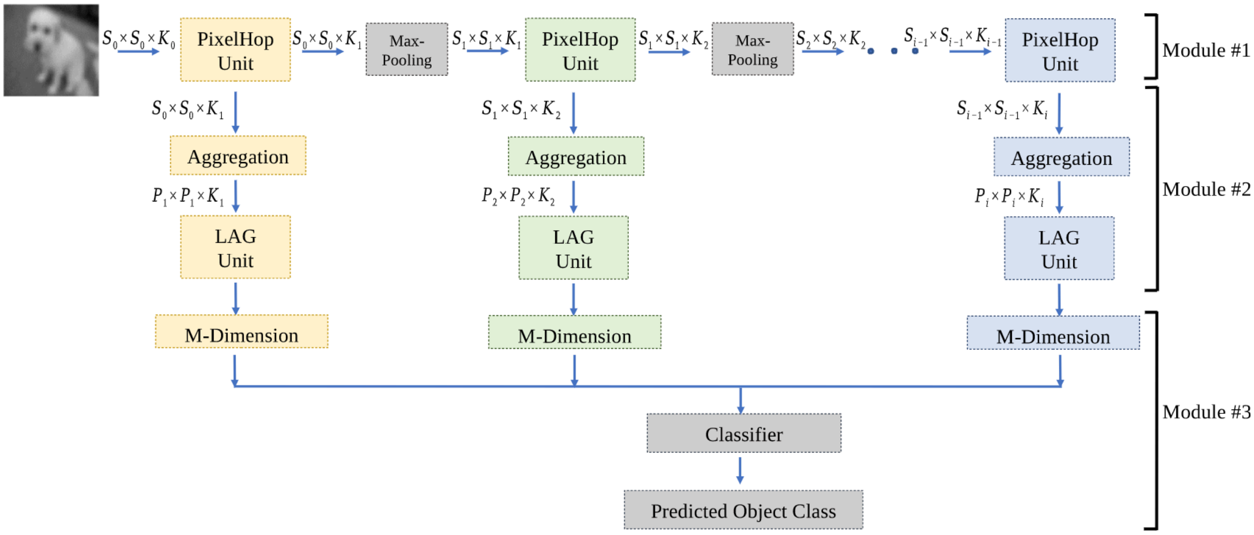

$S_0 \times S_0 \times K_0$
PixelHop Unit
$S_0 \times S_0 \times K_1$
Max-Pooling
$S_1 \times S_1 \times K_1$
PixelHop Unit
$S_1 \times S_1 \times K_2$
Max-Pooling
$S_2 \times S_2 \times K_2$
$S_{i-1} \times S_{i-1} \times K_{i-1}$
PixelHop Unit
Module #1
$S_0 \times S_0 \times K_1$
Aggregation
$P_1 \times P_1 \times K_1$
LAG Unit
$S_1 \times S_1 \times K_2$
Aggregation
$P_2 \times P_2 \times K_2$
LAG Unit
$S_{i-1} \times S_{i-1} \times K_i$
Aggregation
$P_i \times P_i \times K_i$
LAG Unit
Module #2
M-Dimension
M-Dimension
M-Dimension
Classifier
Predicted Object Class
Module #3

Figure 20. Block Diagram of the PixelHop Method

Published paper:
Chen, Y. and Kuo, C.C.J., "PixelHop: A Successive Subspace Learning (SSL) Method for Object Classification," *Journal of Visual Communication and Image Representation*, **70**, Jul 2020, pp. 102749. DOI: 10.1016/J.JVCIR.2019.102749

Contract Information:
Team: Purdue Team, USC Sub-Team
POC: Prof. C.-C. Jay Kuo, University of Southern California, cckuo@sipi.usc.edu, 213-740-4658

Major Technical Problem: Image Splicing Localization
Major Technical Approach: Subspace approximation with augmented kernels (Saak) transform

Being motivated by the multilayer RECOS (REctified-COrrelations on a Sphere) transform, we develop a data-driven Saak (Subspace approximation with augmented kernels) transform in this work. The Saak transform consists of three steps: (1) building the optimal linear subspace approximation with orthonormal bases using the second-order statistics of input vectors, (2) augmenting each transform kernel with its negative, (3) applying the rectified linear unit (ReLU) to the transform output. The Karhunen-Loéve transform (KLT) is used in the first step. The integration of Steps 2 and 3 is powerful since they resolve the sign confusion problem, remove the rectification loss and allow a straightforward implementation of the inverse Saak transform at the same time. Multiple Saak transforms are cascaded to transform images of a larger size. All Saak transform kernels are derived from the second-order statistics of input random vectors in a one-pass feedforward manner. Neither data labels nor backpropagation is used in kernel determination. Multi-stage Saak transforms offer a family of joint spatial-spectral representations between two extremes; namely, the full spatial-domain representation and the full spectral-domain representation. We select Saak coefficients of higher discriminant power to form a feature vector for pattern recognition, and use the MNIST dataset classification problem as an illustrative example.

Data-driven forward and inverse Saak transforms were proposed. By applying multi-stage Saak transforms to a set of images, we can derive multiple representations of these images ranging from the pure spatial domain to the pure spectral domain as the two extremes. There is a family of joint spatial-spectral representations with different spatial-spectral trade-offs between them. The Saak transform offers a new angle to look at the image representation problem and provides powerful spatial-spectral Saak features for pattern recognition. The MNIST dataset was used as an example to demonstrate image reconstruction, feature selection and classification process.

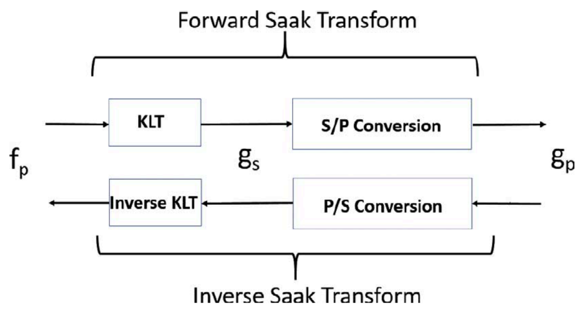


Figure 21. The Block Diagram of Forward and Inverse Saak Transforms

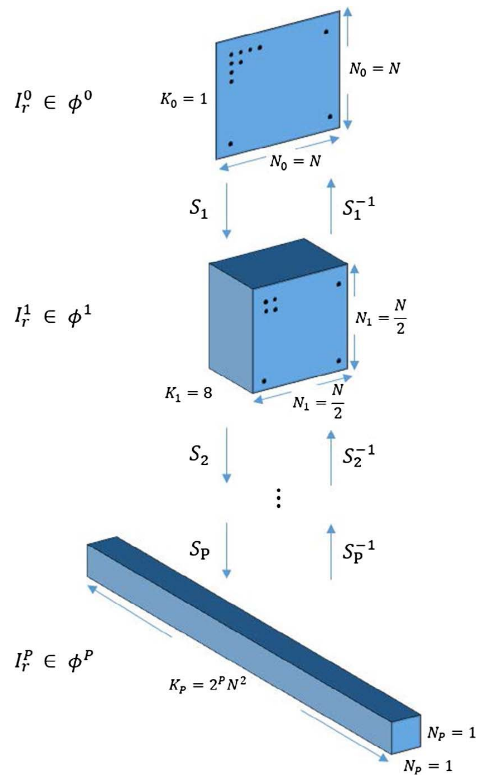


Figure 22. The Conversion Between an Input Image and Its Spatial-Spectral Representations via Forward and Inverse Multi-Stage Saak Transforms

Published paper:
Kuo, C.C.J. and Chen, Y., "On data-driven saak transform," *Journal of Visual Communication and Image Representation*, **50**, 1, Jan 2018, pp. 237-246. DOI: 10.1016/J.JVCIR.2017.11.023

Contract Information:
Team: Purdue Team, USC Sub-Team
POC: Prof. C.-C. Jay Kuo, University of Southern California, cckuo@sipi.usc.edu, 213-740-4658

Major Technical Problem: Image Splicing Localization

Major Technical Approach: Theoretical Understanding of Convolutional Neural Networks

This work attempts to address two fundamental questions about the structure of the convolutional neural networks (CNN): (1) why a nonlinear activation function is essential at the filter output of all intermediate layers? (2) what is the advantage of the two-layer cascade system over the one-layer system? A mathematical model called the "REctified-COrrelations on a Sphere" (RECOS) is proposed to answer these two questions. After the CNN training process, the converged filter weights define a set of anchor vectors in the RECOS model. Anchor vectors represent the frequently occurring patterns (or the spectral components). The necessity of rectification is explained using the RECOS model. Then, the behavior of a two-layer RECOS system is analyzed and compared with its one-layer counterpart. The LeNet-5 and the MNIST dataset are used to illustrate discussion points. Finally, the RECOS model is generalized to a multilayer system with the AlexNet as an example.

In summary, the main contributions are:

(1) Addressed the question of why a nonlinear activation operation is needed at the filter output of all intermediate layers.

(2) Addressed the question of what the advantage of the cascade of two layers in comparison with a single layer is.

(3) A new model ''Rectified COrrelations on a Sphere (RECOS)" is proposed. We view a CNN as a network formed by basic operational units that conducts. It is called the RECOS model. A set of anchor vectors is selected for each RECOS model to capture and represent frequently occurring patterns. For an input vector, we compute its correlation with each anchor vector to measure their similarity. All negative correlations are rectified to zero in the RECOS model, and the necessity of rectification as well as the advantage of multi-layer is explained.

(4) A simple matrix analysis is used to explain the advantage of the two-layer RECOS model over the single-layer RECOS model.

(5) Functions of subnets in LeNet-5 is explained. A CNN can be decomposed into two subnetworks (or subnets): the feature extraction subnet and the decision subnet. For LeNet-5, the decision subnet has the following three roles: (1) converting the spectral-spatial feature map from the output of S4 into one feature vector of dimension 120 in C5; (2) adjusting anchor vectors so that they are aligned with the coordinate-unit vector with correct feature/digit pairing in F6; and (3) making the final digit classification decision in Output.

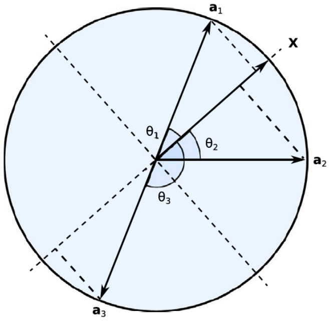


Figure 23. An Example to Illustrate the Need of Correlation Rectification in the Unit Circle

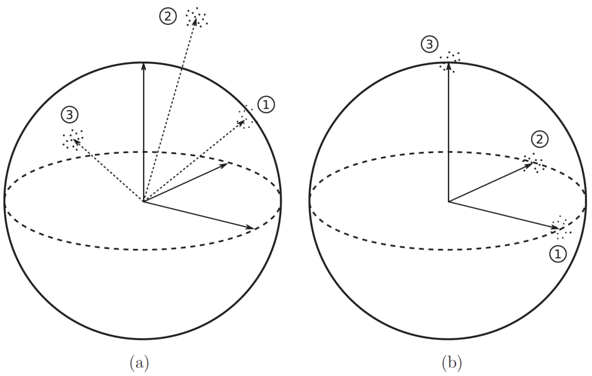


Figure 24. Illustration of Functions of Convolutional Layer C5 and Fully Connected Layer F6 in LeNet-5 Architecture

Published paper:
Kuo, C.C.J., "Understanding convolutional neural networks with a mathematical model,"
Journal of Visual Communication and Image Representation, **41**, 1, Nov 2016, pp. 406-413.
DOI: 10.1016/J.JVCIR.2016.11.003

3.1.4 Politecnico di Milano

Contract Information:
Team: Purdue Team, PoliMi Sub-Team
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: Camera model forensics
Major Technical Approach: CNN-based solutions for camera model attribution

Thanks to the astonishing performance obtained through Convolutional Neural Networks (CNNs), we focused on developing new techniques to perform sensor attribution for visual multimedia data (i.e., images and videos). This is, given an image or a video under analysis, detect the device used for its acquisition (e.g., scanner, camera model, specific camera instance, etc.). Within this context, we specifically focused on detecting camera model attribution traces, i.e., footprints left by all camera devices of the same model. We also focused on how to estimate the reliability of an image in terms of camera model detectability and moved the first steps into the investigation of the open-set scenario (i.e., being able to recognize whether a picture comes from one of the camera under analysis or not).

The proposed camera model attribution method [1, 2] works as follows. Given an image, our algorithm detects which camera model has been used to shoot it within a set of known models. Despite this issue being already topic of research in the literature in the last few years, several state-of-the-art studies tend to show inaccurate results when working on small resolution images. Conversely, our algorithm works in the challenging scenario of small patches (i.e., 64x64 pixels), still achieving state-of-the-art accuracy. This technique is based on machine learning supervised classification paradigm. Specifically, we developed an ad-hoc convolutional neural network (CNN) architecture to extract characteristic camera footprints from each small image patch.

As the proposed CNN exploits very small image patches to take decision, it is paramount that this input data is as reliable as possible. As a matter of fact, not all image patches contain enough information to enable accurate detection of the camera model used to acquire them (i.e., a completely saturated patch is useless as it could have been acquired with any device). Therefore, we developed a methodology that enables to compute patch reliability [3]. This means understanding which portion of an image carries a high amount of camera model information. This technique is based on transfer learning principle applied to a CNN, followed by a dense multi-layer perceptron that acts as regressor.

Through the joint use of the proposed patch reliability estimation technique [3] and camera model detector [1, 2], we also progressed in the field of detection and localization of splicing forgeries obtained through images shot with different cameras [4]. Specifically, we proposed a system that leverage our camera model CNN to extract significant information from image patches. This information in forms of feature vector is fed to a classifier that understands whether the picture is original from a camera, or it is a composition. If a composition is detected, an iterative procedure is used to localize which cluster of patches comes from a different camera model than the rest of the image, i.e., splicing was performed. The proposed solution has proven accurate also in case of spliced images from camera models never used for CNN training, given that no additional disruptive editing operations are performed.

One of the main issues of the great part of camera model detectors in the literature, is that they work in a closed set scenario. This means, they are only able to attribute an image to a camera in a controlled set of known camera models. However, in practical situations, an analyst may be unaware of a set of possible camera candidates. In this case, it is paramount to be able to work in an open set scenario. This means being able to understand whether a picture comes from a camera in the known set, or not. To face this problem, we proposed a set of training strategy explicitly tailored to the open set scenario applied to our camera model CNN [5].

Published Papers:

[1] L. Bondi, L. Baroffio, D. Güera, P. Bestagini, E. J. Delp, S. Tubaro, "First Steps Towards Camera Model Identification with Convolutional Neural Networks", IEEE Signal Processing Letters, vol. 24, no. 3, pp. 259-263, March 2017. DOI: 10.1109/LSP.2016.2641006

[2] L. Bondi, D. Güera, L. Baroffio, P. Bestagini, E. J. Delp, S. Tubaro, "A Preliminary Study on Convolutional Neural Networks for Camera Model Identification", IS&T Electronic Imaging (EI), vol. 2017, no. 7, pp. 67-76, Burlingame, California, January 2017. Preprint: Link DOI: 10.2352/ISSN.2470-1173.2017.7.MWWSF-327

[3] D. Güera, S. K. Yarlagadda, P. Bestagini, F. Zhu, S. Tubaro, E. J. Delp, "Reliability Map Estimation For CNN-Based Camera Model Attribution", IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 964-973, Lake Tahoe, Nevada, February 2018. DOI: 10.1109/WACV.2018.00111

[4] L. Bondi, S. Lameri, D. Güera, P. Bestagini, E. J. Delp, S. Tubaro, "Tampering Detection and Localization through Clustering of Camera-Based CNN Features", IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW), pp. 1855-1864, Honolulu, Hawaii, July 2017. DOI: 10.1109/CVPRW.2017.232

[5] P. R. M. Júnior, L. Bondi, P. Bestagini, S. Tubaro, and A. Rocha, “An In-Depth Study on Open-Set Camera Model Identification,” IEEE Access, 2019.

Contract Information:
Team: Purdue Team, PoliMi Sub-Team
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: Adversarial techniques for camera model attribution
Major Technical Approach: Adversarial CNNs, inpainting-based solutions

Camera model attribution through PRNU-based analysis is a well-known subject in the forensic literature. Indeed, PRNU traces characterize each camera sensor, and prove to be particularly robust to common post processing operations such as resizing and compression. In spite of this, PRNU traces cannot always be trusted to determine whether a picture has been shot with a specific device. Indeed, an informed attacker can manipulate a photograph in order to delete PRNU traces, thus anonymizing the given picture. This process is known as image anonymization.

The importance of studying PRNU-based image anonymization is twofold. On one hand, the ability of anonymizing a photograph is paramount to ensure press freedom in countries rules by regimes and in zones of war. On the other hand, determining at which level PRNU can actually be removed is essential to study the robustness of PRNU-based forensic detectors, and possibly improve them.

In this context, we developed two PRNU-based image anonymization methods. The first one [1] leverages image inpainting as shown in Figure 25. The main idea is to delete some random pixels creating some “holes” in the picture to be anonymized and fill these holes back through inpainting techniques. The method is repeated on multiple holes, and special care is taken around the edges.

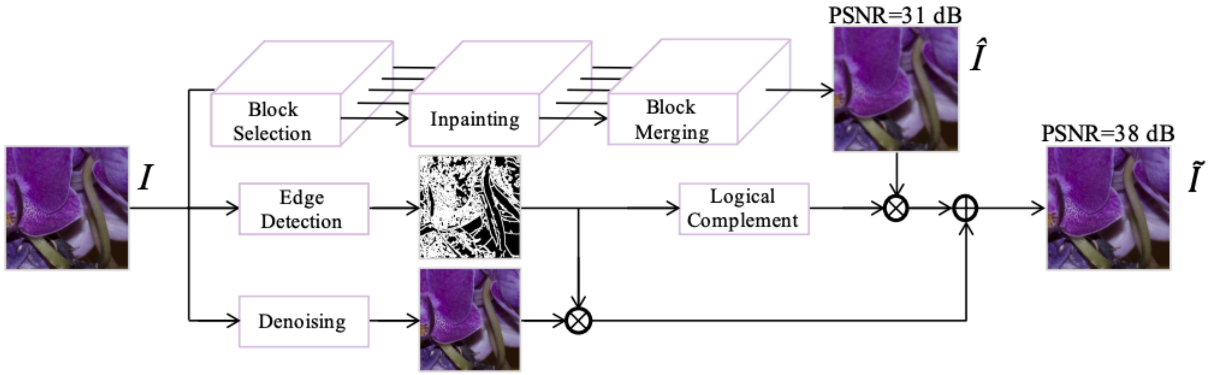


Figure 25. Main Pipeline of the Inpainting-based Method

As an alternative, we also proposed a method leveraging a convolutional neural network (CNN) [2]. In this case, the CNN is trained to remove PRNU traces from an image as a sort of denoiser.

Published Papers:
[1] S. Mandelli, L. Bondi, S. Lameri, V. Lipari, P. Bestagini, S. Tubaro, "Inpainting-Based Camera Anonymization", IEEE International Conference on Image Processing (ICIP), pp. 1522-1526, Beijing, China, September 2017. DOI: 10.1109/ICIP.2017.8296536
[2] N. Bonettini, L. Bondi, D. Güera, S. Mandelli, P. Bestagini, S. Tubaro, E. J. Delp, "Fooling PRNU-Based Detectors Through Convolutional Neural Networks", European Signal Processing Conference (EUSIPCO), Rome, Italy, September 2018. DOI: 10.23919/EUSIPCO.2018.8553596

Contract Information:
Team: Purdue Team, PoliMi Sub-Team
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: PRNU-based forensics
Major Technical Approach: PRNU projections, CNNs for PRNU matching, search of video PRNU parameters

Image source attribution refers to the problem of detecting which device has been used to acquire a specific photograph or video. This means being able to discriminate between different physical cameras of the same brand and model. In the literature, one of the most robust methodologies to solve source attribution problem is based on photo response non uniformity (PRNU). PRNU is a trace left on each image as a multiplicative noise, which is characteristic of the sensor used to shoot the photo. It is therefore possible to attribute an image to a device by comparing a noise footprint extracted from the image under analysis with the PRNU of the suspect device. In this project, we tackled proposed a series of solutions to three major issues related to PRNU forensics: PRNU storage; fast PRNU attribution; video attribution in presence of motion stabilization.

One drawback of large-scale approaches for PRNU-based camera model attribution is the need to store in a central database a huge amount of data. Indeed, PRNU fingerprints need to be extracted at higher resolutions, up to the size of the imaging sensor, to achieve better matching and detection performance and avoid false alarms. In a large-scale retrieval setup, the need of storing several thousands of reference fingerprints at full resolution poses issues regarding the amount of storage space. A second issue arises in terms of computational complexity, when a query fingerprint needs to be matched against many device fingerprints stored in a central camera fingerprints database.

In light of these considerations, we proposed an improved processing chain composed by decimation, gaussian random projections, ternary quantization and coding tailored to increase the compression rate while preserving the highest possible matching accuracy. Exploiting the artifacts generated by JPEG compression on images, we can greatly reduce the required bitrate necessary to store or send a fingerprint [1, 2].

Another main issue of PRNU-based approaches is that running a large-scale test through correlation analysis can be very time consuming. In order to overcome this issue, we developed a method based on convolutional neural networks whose goal is to compare an image noise with a candidate PRNU to detect whether the image comes from a specific device [3]. In particular, we developed two different approaches based on two networks: a deeper one that provides higher attribution accuracy; and a shallow one that is less computational demanding. These approaches have been trained to work on both uncompressed and JPEG-compressed images.

To decide whether a digital video has been captured by a given device, multimedia forensic tools usually exploit characteristic noise traces left by the camera sensor on the acquired frames (i.e., the PRNU). This analysis requires that the noise pattern characterizing the camera and the noise pattern extracted from video frames under analysis are geometrically aligned. However, in many practical scenarios (e.g., if electronic image stabilization is used) this does not occur, thus a re-alignment or synchronization has to be performed. In this context, we proposed two methods to align frame fingerprints with reference PRNU by recovering the scaling, shift and rotation parameters introduced by the electronic stabilization [4, 5]. In [4] we exploit a global optimization technique to recover scale, rotation and cropping parameters. In [5] we overcome the problem of computational complexity by searching for scaling and rotation parameters in the frequency domain thanks to a modified version of the Fourier Mellin transform (FM). The proposed pipeline is based on the following steps: (i) noise extraction; (ii) geometric transformation estimation; (iii) geometric compensation and matching. More specifically, we propose to use a global optimizer to estimate shift parameters, whereas a modified version of the FM transform using both phase and magnitude Fourier information is used to estimate scaling and rotation parameters in closed form.

Published Papers:

[1] L. Bondi, P. Bestagini, F. Pérez-González, S. Tubaro, "Improving PRNU Compression through Preprocessing, Quantization and Coding", IEEE Transactions on Information Forensics and Security, vol. 14, no. 3, pp. 608-620, July 2018. DOI: 10.1109/TIFS.2018.2859587

[2] L. Bondi, F. Pérez-González, P. Bestagini, S. Tubaro, "Design of Projection Matrices for PRNU Compression", IEEE Workshop on Information Forensics and Security (WIFS), Rennes, France, December 2017. DOI: 10.1109/WIFS.2017.8267652

[3] S. Mandelli, D. Cozzolino, P. Bestagini, L. Verdoliva, and S. Tubaro, “CNN-based fast source device identification,” IEEE Signal Processing Letters (SPL), 2020.

[4] S. Mandelli, P. Bestagini, L. Verdoliva, and S. Tubaro, “Facing Device Attribution Problem for Stabilized Video Sequences,” IEEE Transactions on Information Forensics and Security (TIFS), 2019.

[5] S. Mandelli, F. Argenti, P. Bestagini, M. Iuliani, A. Piva, and S. Tubaro, “A Modified Fourier-Mellin Approach for Source Device Identification on Stabilized Videos,” in IEEE International Conference on Image Processing (ICIP), 2020.

Contract Information:
Team: Purdue Team, PoliMi Sub-Team
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: Software identification through JPEG traces
Major Technical Approach: Study of JPEG quantization matrix

The vast majority of images available online are JPEG compressed. Images are compressed directly onboard during acquisition, are compressed after almost any processing operation, and are also compressed whenever uploaded on a social platform or shared through messaging apps. For this reason, studying the history of JPEG traces can reveal important pieces of information related to the past of a specific image and the possible use of editing software suites. Within this context, we focus on two specific JPEG-related problems: detection of multiple JPEG compression; and detection of the used JPEG implementation.

Considering the problem of double JPEG compression, we developed a series of CNNs that consider different kinds of input. We studied the possibility of feeding the CNN with image patches, with processed patches obtained by applying some noise extraction algorithms, and we also investigated CNNs that estimate DCT histograms [1].

Considering the problem of multiple JPEG compressions using editing software suites, we investigated a solution capable of detecting up to four image compressions [2]. This means, given an image, detect how many times (up to four) it has been JPEG compressed. To do so, we leverage perturbations of DCT histograms that capture traces of multiple compressions and train a supervised classifier to discriminate between images compressed different number of times. Multiple-JPEG detection is performed using Task-driven Non-negative Matrix Factorization (TNMF).

Due to the widespread diffusion of JPEG compression standards we also focused on the detection of traces left on images by the use of different JPEG implementations (e.g., characteristic of proprietary software suites). Specifically, we focused on capturing traces left by different JPEG implementations in order to distinguish between images that have been compressed with different software suites, even if the very same quantization matrix has been used [3]. As a matter of fact, the literature has recently shown that it is possible to detect whether an image has been compressed with different JPEG implementations according to the used quantization rule (i.e., flooring, ceiling or rounding). Motivated by this finding, we explored the rationale behind eigen-algorithms to propose a compact descriptor that captures JPEG implementation traces. Specifically, given a JPEG image under analysis, we re-encode it using different controlled implementations of JPEG compression algorithm. We then compare the image under analysis with the re-compressed versions to expose salient differences. These differences are collected into a descriptor, that can be fed to a simple supervised classifier to detect the JPEG implementation used to originally encode the image under analysis.

Published Papers:

[1] M. Barni, L. Bondi, N. Bonettini, P. Bestagini, A. Costanzo, M. Maggini, B. Tondi, S. Tubaro, "Aligned and non-aligned double JPEG detection using convolutional neural networks", Journal of Visual Communication and Image Representation, vol. 49, pp. 153-163, November 2017. DOI: 10.1016/J.JVCIR.2017.09.003

[2] S. Mandelli, N. Bonettini, P. Bestagini, V. Lipari, S. Tubaro, "Multiple JPEG Compression Detection through Task-driven Non-negative Matrix Factorization", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2106-2110, Calgary, Canada, April 2018. DOI: 10.1109/ICASSP.2018.8461904

[3] N. Bonettini, L. Bondi, P. Bestagini, and S. Tubaro, "JPEG Implementation Forensics Based on Eigen-Algorithms," in IEEE International Workshop on Information Forensics and Security (WIFS), 2018.

Contract Information:
Team: Purdue Team, PoliMi Sub-Team
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: Video manipulation forensics
Major Technical Approach: Coding-based traces, PRNU-based techniques, deepfake detection through CNNs.

Considering the problem of video forgery detection, we proposed different methods exploiting different kind of traces: coding traces; PRNU traces; deepfake traces.

One of the reasons behind failure of many video forgery detection methods is that videos are customary stored and distributed in a compressed format, and codec-related traces tends to mask previous processing operations. For this reason, we decided to exploit coding traces left by specific video processing suites as an asset for solving forgery detection and localization problems. Specifically, we propose to capture video codec traces through convolutional neural networks (CNNs) [1]. To do so, we train two CNNs to extract information about the used video codec and coding quality, respectively. Building upon these CNNs, we propose a system to detect and localize forgeries generated by splicing content from different videos, characterized by inconsistent coding schemes and / or parameters (e.g., video compilations from different sources or broadcasting channels). The developed solution for video forgery detection and localization follows this pipeline: (i) each video frame is split into patches; (ii) each patch is fed to two different CNNs (i.e., on extracting quality-based features, and one extracting codec-based features); (iii) features for each patch are concatenated into an 8-element vector; (iv) feature vectors coming from each patch are compared to detect incoherent behavior indicating forgery.

As PRNU is considered to be a very robust forensic trace, we also developed two solutions for video forgery detection based on PRNU analysis. In [2] we propose a method to estimate whether videos from different sources are spliced together. The method tries to estimate a PRNU from a portion of the video and verifies which frames of the videos are compatible with this estimate. In [3] we propose a different pipeline that works in open-set scenario. Rather than performing a costly PRNU estimation and matching search, we cast the problem in terms of PRNU feature classification.

As the recent developments in video tampering have shown the dangerous implications introduced by deepfakes, we also investigated a deepfake video detector [4]. Properly training a deepfake detector is challenging task. As a matter of fact, deepfake traces are particularly subtle. Moreover, it is easy to get tricked and learn actors' faces rather than forensic footprints. For this reason, we developed a training routine that exploit a triplet loss. The network is fed triplets of faces during training. In particular, we show the network two faces of the same actor, and one face of another one, in order to push real and fake faces of the same actor further away. In order to help the CNN focusing on relevant portion of the face, the CNN itself is forced to learn an attention mask that hides non-relevant portion of the input picture.

Published Papers:

[1] S. Verde, L. Bondi, P. Bestagini, S. Milani, G. Calvagno, S. Tubaro, "Video Codec Forensics Based on Convolutional Neural Networks", IEEE International Conference on Image Processing (ICIP), pp. 530-534, Athens, Greece, October 2018. DOI: 10.1109/ICIP.2018.8451143

[2] S. Mandelli, D. Cozzolino, P. Bestagini, L. Verdoliva, S. Tubaro, "Blind Detection and Localization of Video Temporal Splicing Exploiting Sensor-Based Footprints", European Signal Processing Conference (EUSIPCO), Rome, Italy, September 2018. DOI: 10.23919/EUSIPCO.2018.8553511

[3] P. R. M. Júnior, L. Bondi, P. Bestagini, A. Rocha, and S. Tubaro, "A PRNU-Based Method to Expose Video Device Compositions in Open-Set Setups," in IEEE International Conference on Image Processing (ICIP), 2019.

[4] N. Bonettini, E. D. Cannas, S. Mandelli, L. Bondi, P. Bestagini, and S. Tubaro, "Video Face Manipulation Detection Through Ensemble of CNNs," in International Conference on Pattern Recognition (ICPR), 2020.

3.1.5 New York University

Contract Information:
Team: Purdue Team, NYU Sub-Team
POC: Prof. Nasir Memon, New York University, memon@nyu.edu, 732-241-6128

Major Technical Problem: Factors Affecting ENF based Time-of-Recording Verification
Major Technical Approach: Investigation of how the quality of the ENF signal to be estimated from video is affected by different light source illumination, and investigation of how the quality

of the estimated ENF signal is affected by different compression ratios depending on the type of light source.

Ambient conditions in the scene and data properties may have a noticeable impact on the quality of the ENF signal to be estimated from a video. Consequently, the performance of an ENF based forensic application may be affected by these conditions. Accordingly it was first investigated how the performance of ENF based time-of-recording detection and verification is affected by different illumination sources. Next, it was explored how different compression ratios, and different lengths of the reference ENF data (ground-truth) affect the performance depending on the type of light source and depending on the video length.

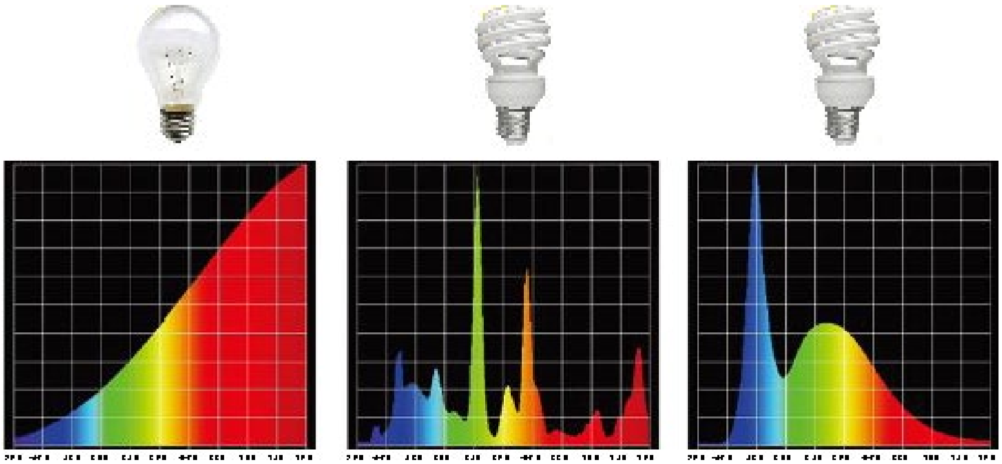


Figure 26. Spectra of Various Light Sources

Figure 26 illustrates emission spectra of various commonly used light sources, namely Incandescent, white CFL (Compact Fluorescent), and white LED (Light-emitting Diode). From the figure, it can be seen that Incandescent tungsten has a lower level of power in blue light, though it has the greatest power in red. CFL has a relatively lower level of power across the spectrum except for some distinguishable spikes, including green and red. LED provides the highest power in blue light; though relatively lower power in the red. Given these differences in the power of wavelengths across the visible spectrum for different light sources, we investigate how the type of light source may affect the quality of the estimated ENF signal, and consequently, the ENF based time-of-recording verification.

In summary, the following contributions were introduced:

1) Exploration of how the quality of the ENF signal to be estimated from video is affected by different light source illumination.

2) Exploration of how the quality of the estimated ENF signal is affected by different compression ratios, and by social media encoding depending on the type of light source and depending on the video length.

3) Exploration of how the length of ground-truth ENF database affects the time of recording verification performance.

Published Paper: S. Vatansever, A. E. Dirik and N. Memon, "Factors Affecting Enf Based Time-of-recording Estimation for Video," *Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing*, Brighton, United Kingdom, (2019)

Contract Information:
Team: Purdue Team, NYU Sub-Team
POC: Prof. Nasir Memon, New York University, memon@nyu.edu, 732-241-6128

Major Technical Problem: The Effect of Rolling Shutter Sampling Mechanism and of Idle Period between Successive Frames on Capture and Estimation of ENF in a Given Video
Major Technical Approach: Rolling Shutter based Luminance Samples Estimation

In the global shutter mechanism, widely used in CCD sensors, an entire frame is exposed at one time instance, i.e., each row of pixels of a single frame is sampled simultaneously. Whereas in rolling shutter mechanism, mostly used in CMOS sensors, each row of pixels in a frame is captured sequentially at different time instances. Figure 27 illustrates the timing for these two distinct procedures for a single frame. ENF estimation techniques are designed under consideration of these sampling mechanisms. In global shutter based ENF estimation approach, one illumination sample is obtained per frame by averaging all steady pixels within the frame. Hence, the sampling frequency in this technique is essentially the video frame rate, which is much lower than twice the nominal ENF frequency. As this does not satisfy the Nyquist criteria, it has to work with alias ENF. However when the video frame rate is a divisor of the nominal ENF, e.g. an 30 fps video captured in the US (60 Hz mains-power), alias ENF is observed at the 0 Hz DC component. So, it is a great challenge for this scheme to work under this condition. Rolling Shutter based ENF estimation method leads to a sampling frequency as high as number of rows × camera frame rate, as each row is treated as a different illumination sample. Consequently, since the number of samples is much higher than twice the ENF frequency (Nyquist rate), alias ENF is not a phenomenon for this scheme. Although the rolling shutter mechanism can lead to an increase of sampling frequency and consequently avoids alias frequency, it brings with it the idle period issue. That is, some illumination samples in each frame are missed. A representation of this phenomenon is depicted in Figure 28. Accordingly, we developed an analytical model to explore how the frequency of mains-powered illumination and so ENF is shifted and is attenuated in relation to idle period length. Based on the model developed, a novel idle period estimation approach is proposed. Our model and analysis also leads to a novel time-of-recording verification technique that performs better than the existing techniques in the literature. A systematic search of possible ENF components emerging as a result of the idle period, followed by idle period assumptions in each component, and interpolation of missing samples for each assumption result in better quality ENF signal estimations. A more accurate ENF signal extraction consequently leads to a better time-of-recording verification performance. In summary, the following contributions were introduced:

1) A model for where frequency of main ENF harmonic is shifted and how the power of the captured ENF is attenuated, depending on idle period length.

2) A novel idle period estimation technique targeted at camera forensics.

3) A novel time-of-recording verification method for videos captured using a rolling shutter mechanism.

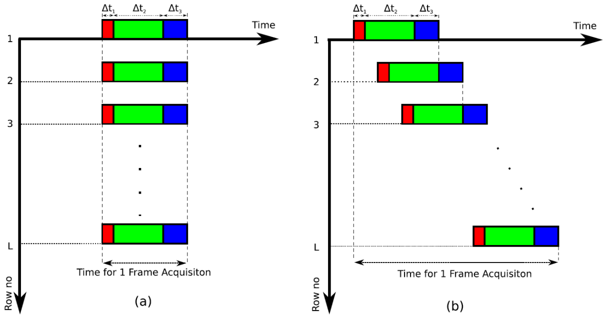


Figure 27. Demonstration of (a) Global Shutter Sampling Mechanism – Each Row of a Frame Is Exposed at the Same Time Instance (B) Rolling Shutter Sampling Mechanism

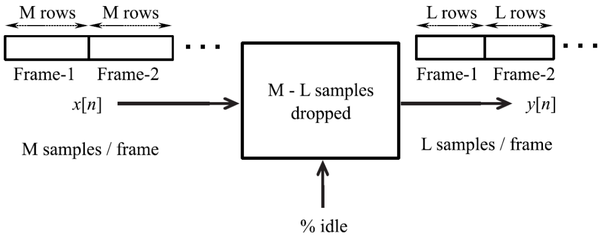


Figure 28. Inherent Reduction in Captured Luminance Samples in a Video Due to the Implementation of Idle Period at the End of Each Frame by Camera

Published Paper: S. Vatansever, A. E. Dirik and N. Memon, "Analysis of Rolling Shutter Effect on ENF-Based Video Forensics," *IEEE Transactions on Information Forensics and Security,* **14**, 9, Sept. 2019, pp. 2262-2275.

Contract Information:
Team: Purdue Team, NYU Sub-Team
POC: Prof. Nasir Memon, New York University, memon@nyu.edu, 732-241-6128

Major Technical Problem: Detecting the Presence of ENF in a Given Video
Major Technical Approach: Super-pixel based ENF estimation

In ENF-based video forensics, it is important to test whether a video contains any traces of ENF before moving on to further analysis. For instance, if a video does not contain any ENF signal it would be useless to search in existing ENF databases for video time-of-recording verification. More importantly, a substantial amount of computational load and time can be saved if a quick test can establish the absence of an ENF signal. For this purpose, we propose a super-pixel-based ENF signal presence detection technique. The proposed method performs multiple “so-called ENF” signal estimations from different steady object regions having very close reflectance properties, i.e., super-pixels. Our motivation to use super-pixels is that each pixel in a super-pixel region is almost uniform in brightness, color, and texture, and hence has uniform reflectance characteristics. Figure 29 provides a sample image with super-pixel regions. Working on a super-pixel region provides the possibility of estimating ENF from videos taken by not only CCD camera but also by CMOS camera, which uses rolling shutter mechanism. In the proposed algorithm, a “so-called ENF signal” is estimated from each steady super-pixel separately. The reason we use the term “so-called ENF” is because the estimated signal is initially unknown to be actually an ENF signal. Depending on the similarity of the estimated signals from each steady super-pixel, it can be decided whether any ENF signal is present in the test video or not. The proposed method does not require any verification against a reference ENF database. In summary, the following contributions were mainly introduced:

1) A novel ENF estimator that exploits only some portion, i.e., a small region of a given video for ENF extraction is introduced.

2) An ENF detector which achieves a very high ENF signal presence detection accuracy for videos captured by any sensor type is proposed.

3) Both the ENF estimator and ENF detector can operate independently of the source camera sensor type, CCD or CMOS is developed.

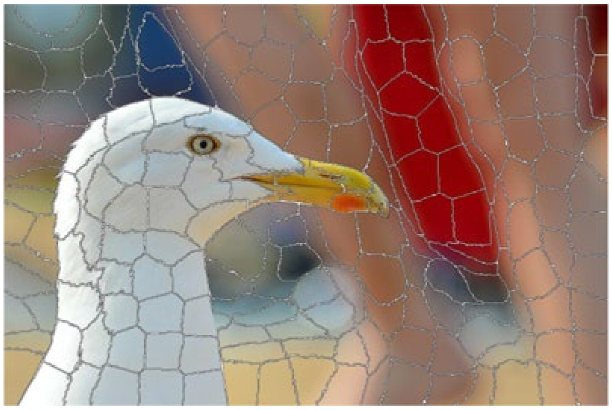


Figure 29. A Sample Image With Super-Pixels

Published Paper: S. Vatansever, A. E. Dirik and N. Memon, "Detecting the Presence of ENF Signal in Digital Videos: A Superpixel-Based Approach," *IEEE Signal Processing Letters*, **24**, 10, Oct. 2017, pp. 1463-1467

Contract Information:
Team: Purdue Team, NYU Sub-team
POC: Prof. Nasir Memon, New York University, memon@nyu.edu, 732-241-6128

Major Technical Problem: The need for ground-truth ENF for time-of-recording verification.
Major Technical Approach: Development of a ground-truth ENF recorder to be connected to any power outlet from where it can directly read the reduced mains power voltage and can estimate ENF at 1 samples/second resolution by using these voltage time series by means of STFT based ENF estimation technique.

Ground truth ENF database is required particularly for time-of-recording verification task. Hence, it is essential to instantaneously acquire the electric network frequency from a power outlet and save it. For this purpose, we developed an ENF recording device which can obtain ground-truth ENF directly from any mains power outlet in real time at 1 samples/second resolution, and which can send the recorded data to a remote server. As can be seen in **Figure *30***, we have 5 active devices that currently record ENF, 1 in NYU, 1 in Partech, 1 in University of Colorado Denver, and 2 in Turkey.

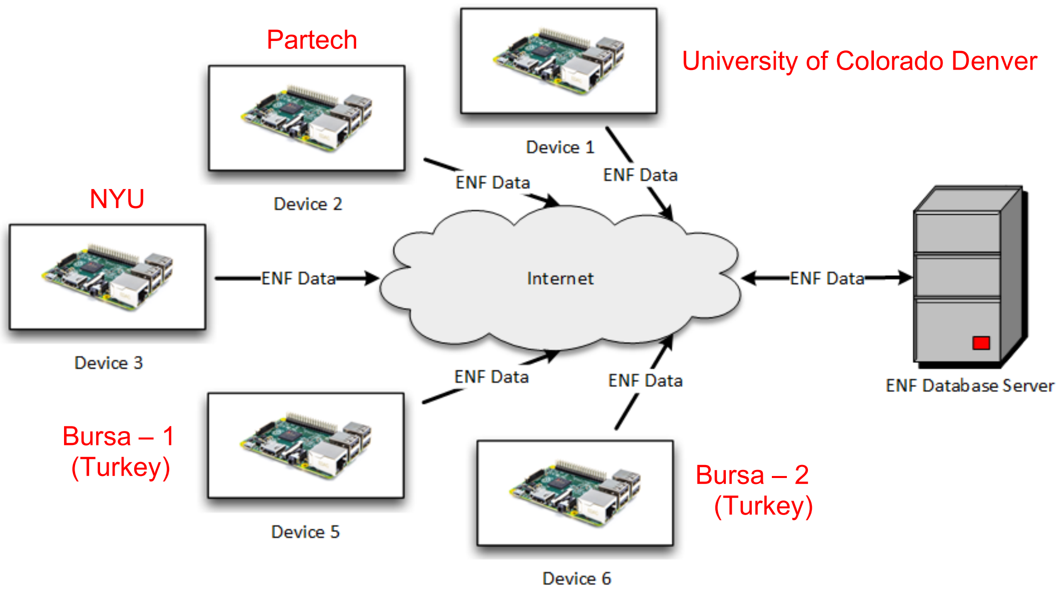


Figure 30. Active ENF Recorders Which Currently Record Ground-Truth ENF Data and Send Them to Remote Server

In summary, the following contributions were mainly introduced:

1) Development of an ENF recording device that can obtain ENF directly from any mains power outlet in real time.

2) Creation of ground-truth ENF databases for eastern network (US), western network (US), and Turkey.

Contract Information:
Team: Purdue Team, NYU Sub-team
POC: Prof. Nasir Memon, New York University, memon@nyu.edu, 732-241-6128

Major Technical Problem: Time-of-Recording Verification
Major Technical Approach: Attenuation of frame rate harmonics

A critical phenomenon with the rolling shutter sampling mechanism is that there occur strong peaks at multiples of video frame rate. Consequently, ENF harmonics overlap with the frame rate harmonics for videos whose frame rate is a divisor of nominal ENF. For such a case, if the power of ENF is not greater than the power of frame rate harmonic due to some reasons including long idle period, type of light source, compression, etc., estimated ENF signal is obtained in a poor quality, or it is completely destroyed. It was explored that luminance intensity on consecutive rows of a frame is in the trend of increase or decrease due to the inverse square law, which causes a gap in frame transitions as shown in Figure 31(a). It consecutively causes strong frame rate harmonics to arise as provided in Figure 31(b). Hence, in order to handle these frame rate harmonics, we needed to bring the up-going or down-going luminance variations into a form such that they will fluctuate around a straight line so that frame transitions become relatively smooth. For this purpose, a number of techniques were tried including derivative based approach and curve fitting to handle this phenomenon. Based on our analysis, we found the best working technique is "curve fitting". We implemented a second degree polynomial curve fitting to the luminance variation data obtained for each frame. After subtracting sample of luminance value of each row of a frame from the corresponding fitted curve point, a considerable attenuation in the frame rate harmonics were yielded. For some videos, curve fitting approach result a small improvement. In these videos, the performance can be further improved by deleting slightly of luminance samples in the end of a frame and in the beginning of the next frame. The main understanding to the deletion approach is that if some data from the end of a frame and from the beginning of the next frame are removed, the end point of the previous frame and the first point of the next frame come closer. However, as sample deletion would increase idle period, and reduces the power of ENF, the proportion of samples to be deleted should be decided very carefully.

In summary, the following contributions were mainly introduced:

1) Exploration of the source of frame rate harmonics.

2) A method for reducing the frame rate effect is proposed.

3) A further improved time-of-recording verification approach.

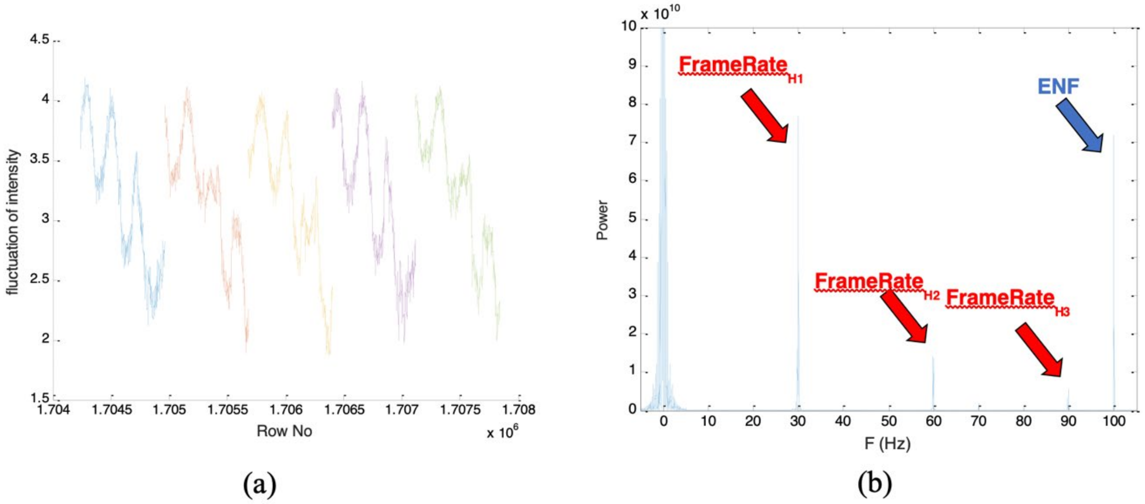


Figure 31. (a) Intensity Fluctuations (b) Frame Rate Harmonics for an Exemplary Video Before Curve Fitting Operation

Published Paper:
We are currently working on a TIFS paper including the above-mentioned concepts!

Contract Information:
Team: Purdue Team, NYU Sub-team
POC: Prof. Nasir Memon, New York University, memon@nyu.edu, 732-241-6128

Major Technical Problem: Time-of-Recording Verification
Major Technical Approach: Classification of quality ENF samples

An estimated ENF signal consisting of relatively high number of outliers requires a special attention to which part of the signal to select and use for time-of-recording verification. Indeed, for some cases particularly when they consist of much greater amount of outliers, elimination of the signal may be required before time-of-recording verification application. Accordingly, we propose a novel method that is based on detection and classification of all quality samples of an estimated ENF signal, and creation of a mask for these samples. For the similarity test, we propose an adapted normalized cross-correlation operation accordingly in a way that it works based on the mask. The new approach is distinguishable for noisy ENF.

Table 1. Classification of Useful ENF Samples and an Adapted Normalized Cross Correlation

Steps	Description
1.	The outliers are detected based on the derivative based approach.
2.	Each outlier clip is expanded by accepting some non-outlier samples at either side as outlier.
3.	All detected outliers are made zero.
4.	A mask for all good quality samples is built.
5.	The normalized cross-correlation operation is adapted in a way that it works based on the mask

The adapted normalized cross correlation operation stated in step 5 in Table 1 is provided as:

$$c(i) = \frac{\sum_{n=1}^{N}\left[F_{est}(n) - \mu_{est}\right]\left[F_{ref_i}(n) - \mu_{ref_i}\right]}{\sqrt{\sum_{n=1}^{N}\left[F_{est}(n) - \mu_{est}\right]^2 \sum_{n=1}^{N}\left[F_{ref_i}(n) - \mu_{ref_i}\right]^2}}$$

N: Length of the estimated ENF signal

$F_{est}(n)$: n^{th} sample of the estimated ENF signal ($n \neq$ index of noisy samples)

μ_{est}: Mean of non-zero samples of the estimated ENF signal

$F_{ref_i}(n)$: n^{th} sample of the i^{th}ground-truth ENF clip of the length N ($n \neq$ index of noisy samples)

μ_{ref_i}: Mean of non-zero samples of the i^{th}ground-truth ENF clip

$c(i)$: normalized correlation coefficient between the Estimated ENF signal and i^{th}ground-truth ENF clip

$i = \{1, 2, \ldots, L_{ref} - N + 1\}$

In summary, the following contributions were mainly introduced:

1) A method for detecting and classifying quality ENF samples of an estimated ENF signal.

2) An adapted normalized cross-correlation operation.

3) A further improved time-of-recording verification approach.

Published Paper:
We are currently working on an SPL paper for the above-mentioned concepts!

3.1.6 University of Notre Dame

Contract Information:
Team: Purdue Team, Notre Dame Sub-Team
POC: Prof. Walter Scheirer, University of Notre Dame, walter.scheirer@nd.edu, 574-631-2436

Major Technical Problem: Image Provenance Analysis
Major Technical Approach: Image Provenance at Scale Pipeline

Image Provenance Filtering and Image Provenance Graph Building motivate the need for any developed techniques to be scalable. Specialized algorithmic components are necessary to solve the problem at hand. First, one needs an accurate and scalable image retrieval algorithm that is able to operate over very large collections of images (realistically, on the order of millions of images) to find related candidates. Such an algorithm also has to address the particularities of the provenance image filtering task: it must perform well at retrieving the near-duplicate host images that are highly related to the query (a well-known problem in the image retrieval literature), but also perform well at retrieving donors (images that potentially donated small portions to the query) and the donors' respective near duplicates (which might not be directly related to the query). Second, the identification of likely image transformations that explain how each retrieved image might have been used to generate the others is required, as it is used to create the ordering of the images in the provenance graph. And third, methods from graph theory are necessary to organize the relationships between images, yielding a directed graph that is human-interpretable. All of these components must be integrated as a coherent and scalable processing pipeline.

This approach introduces, for the first time, a fully automated large-scale end-to-end pipeline that starts with the step of provenance image filtering (over millions of images) and ends up with the provenance graphs. The following new contributions are introduced this work:

1) Distributed interest point selection: a novel interest point selection strategy that aims at spatially diversifying the image regions used for indexing within the provenance image filtering task.

2) Iterative Filtering: a novel querying strategy that iteratively retrieves images that are directly or indirectly related to the query, considering all possible hosts, donors, composites, and their respective near duplicates.

3) Clustered Provenance Graph Construction: a novel graph construction algorithm that clusters images according to their content (joining near duplicates into the same clusters), prior to establishing their intra- and inter-cluster relationship maps.

4) State-of-the-art results on the provenance analysis benchmark released by the American National Institute of Standards and Technology (NIST).

5) A new dataset of real-world scenarios containing composite images from Photoshop battles held on the Reddit website. Experiments performed over this dataset highlight the real-world applicability of the approach.

Published Paper:
[1] Moreira et al. "Image Provenance at Scale," IEEE T-IP, Vol. 27, No. 12, 2018

Contract Information:
Team: Purdue Team, Notre Dame Sub-Team
POC: Prof. Walter Scheirer, University of Notre Dame, walter.scheirer@nd.edu, 574-631-2436

Major Technical Problem: Image Provenance Analysis
Major Technical Approach: Image Provenance Leveraging Metadata

Image processing and computer vision techniques can be employed to detect correspondences between images or other digital art forms. This kind of correspondence can range from object matching in images to comparing the style and semantics of the two. Provenance analysis can be thought of as ordering pair similarities between multiple image pair sets and is therefore a natural extension to pairwise image comparison. These subsequent ordered parings can be modeled as a graph, where each edge denotes a correspondence between a pair, and the end vertices of the edge signify the two respective images. A provenance analysis algorithm could be analyzing multiple very close-looking realistic versions of the same visual object. Complex scenarios like this can make content-based similarity metrics unreliable.

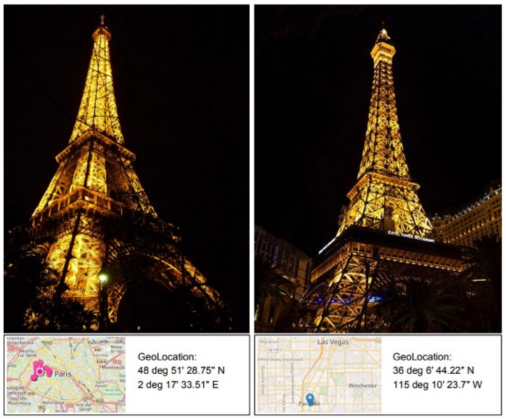


Figure 32. Two Eiffel Towers. Only One Is Authentic, but Both Are Real Objects

Due to the vast range of possible versions of a single original image, the metrics for quantifying the similarity between pairs of images can be noisy. Relying solely upon visual cues to order the different versions into a graph can result in poor provenance reconstructions. Therefore, it becomes pertinent to utilize other sources of data to determine connections. For example, it is difficult to point out a semantic difference between the two images in Figure 32, but the images can be differentiated by inspecting the metadata of the image files. Such a pair of images can be termed semantically similar, as they are related to each other in a semantic way but do not originate from the same source. Matching difficulty can also arise within sets of near-duplicate images, which are generated from a single origin having undergone a series of transformations (e.g., crop→saturate→desaturate). The pixel-level data within these image sets can exhibit ambiguous provenance directionality. Information beyond pixel-level data may be required to detect differences between such images.

To handle scenarios where image content fails to explain image evolution, file metadata can be used to help fill in the gaps. In this work, we explored the use of commonly present file metadata tags to improve image provenance analysis. We compared these results against image content-based methods and highlight the advantages and disadvantages of both.

Published Papers:
Moreira et al. “Image Provenance at Scale,” IEEE T-IP, Vol. 27, No. 12, 2018
Bharati et al. “Beyond Pixels: Image Provenance Analysis Leveraging Metadata,” IEEE WACV, 2019

Contract Information:
Team: Purdue Team, Notre Dame Sub-Team
POC: Prof. Walter Scheirer, University of Notre Dame, walter.scheirer@nd.edu, 574-631-2436

Major Technical Problem: Image Provenance Analysis
Major Technical Approach: Fast Local Spatial Verification for Feature-Agnostic Large-Scale Image Retrieval

Our goal is to adapt image retrieval to different contexts, such as composite images, especially those containing small, spliced objects. For example, a composite image created from many smaller donor images. This type of image poses a significant challenge for existing image retrieval algorithms because if the host (i.e., background) and donor images are matched globally, the latter group would not be highly scored since the content shared with the composite is very small. Nonetheless, tasks like meme analysis and disinformation debunking require retrieving each meaningful piece of content in an image under scrutiny. Because the image in question is a conglomeration from many sources, we can assume that the goal of a retrieval system should be to return instances of all images contributing to it.

To do so, we propose a new spatial verification method that allows object-level instance scoring of retrieved results without the need for costly object detection steps. We devise a feature-agnostic algorithm that utilizes a geometrically consistent voting measure inspired by the Spatially Constrained Similarity Measure (SCSM) technique, with the major difference being that object regions of interest in the query need not be known ahead of time, in order to make the solution more appropriate to the current reality of unspecified retrieval context. As we show through experiments, the proposed method quickly and accurately localizes and ranks rigid objects contained within the query image to objects contained within a large image database. We call this method the Objects in Scene to Objects in Scene (OS2OS) score, and it is optimized for fast matrix operations on CPUs or GPUs.

In summary, the contributions are:

- A new perspective on the problem of image retrieval, which aims to address the deficiencies in existing problem formulations for retrieval in cases of complex composite images and other manipulated images.
- A new method called the OS2OS score for spatial verification of matching objects between images, including tiny objects that are important for understanding memes and other emerging Internet media.
- A series of experiments showing the viability of the approach on the Oxford 5K, Paris 6K, Google-Landmarks, and NIST MFC2018 datasets, as well as meme-style imagery from Reddit.
- A new experimental protocol for the Reddit Photoshop Battles dataset, preparing it to be used for bench-marking potential solutions to the problem of retrieving the donors of composite images.

Our Papers:
Published Paper: Moreira et al. “Image Provenance at Scale,” IEEE T-IP, Vol. 27, No. 12, 2018

J. Brogan *et al*., "Fast Local Spatial Verification for Feature-Agnostic Large-Scale Image Retrieval," in *IEEE Transactions on Image Processing*, vol. 30, pp. 6892-6905, July 2021. DOI: 10.1109/TIP.2021.3097175

Contract Information:
Team: Purdue Team, Notre Dame Sub-Team
POC: Prof. Walter Scheirer, University of Notre Dame, walter.scheirer@nd.edu, 574-631-2436

Major Technical Problem: Image Provenance Analysis
Major Technical Approach: Learning Transform-Aware Embeddings for Image Forensics

Depending on the type of manipulations performed on an image, the provenance graph can be complicated with multiple donor images (images that share partial content), and long chains formed by near-duplicate images (images derived from a single image through a series of transformations). Previous approaches rely heavily on the quality of keypoints (i.e., points of interest for matching). Although keypoint matching is efficient to compare two images, the existing keypoint detection and matching techniques focus on general invariance among local regions tolerant to a small set of transformations. For example, SIFT and SURF are invariant to most changes in scale and orientation. Even though this property is very desirable for provenance analysis, establishing sequences of images that vary slightly based on different transformations requires differentiating between slightly modified versions of the same content. This work focuses on improving the fidelity of reconstructing the chains of globally-related images in the provenance graphs by encoding this awareness in the first stages of graph construction, i.e., in the dissimilarity computation. The subsequent stage of applying greedy graph algorithms treats dissimilarity values as adjacency weights and will then rely heavily on transformation-aware image matching, thus improving the process.

Image matching is a common task in many computer vision problems such as object recognition, 3D reconstruction, scene understanding, and image retrieval. A desired property for representations used to solve these problems is invariance to view changes, compression, and other image transformations. Optimizing this property while learning the mapping between the data X and label Y can easily overlook capturing the subtle differences between image transformations.

For that reason, we propose to learn representations that are aware of different versions of images in the transformation space, and depending on the number of transformations, can encode appropriate distance among near-duplicate images. This approach can be useful for improving dissimilarity computation between different versions of an image. A better understanding of the subtle differences in the variants can lead to improved rank-based output in deducing a sequence of transformations for image forensics, cultural analytics, and other applications. It can also be used to define acceptable standards of edited data for learning algorithms. To our knowledge, this is the first work that focuses on solving the image ordering aspect of provenance analysis. This work also highlights the importance of awareness of transformation-based differences among near-duplicate images while learning representations.

Published Paper:
A. Bharati, D. Moreira, P. J. Flynn, A. de Rezende Rocha, K. W. Bowyer, W. J. Scheirer, "Transformation-Aware Embeddings for Image Provenance," in *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 2493-2507, January 2021. DOI: 10.1109/TIFS.2021.3050061

3.2 Overhead Forensics Project

In this section we describe our efforts on the Overhead Forensics project.

3.2.1 Purdue University

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Unsupervised splicing detection, localization in satellite images
Major Technical Approach: Sat-SVDD (Satellite Support Vector Data Descriptor)

We propose a new robust deep anomaly detection and localization method, targeting splicing manipulations inserted in overhead images. In our work, we assume that we do not have access to forged images for training. Our contributions for this method are:

- The proposed approach significantly outperforms all previously presented satellite manipulation detection methods.
- By using anomaly detection techniques, we do not require access to any manipulated data for training.
- We introduce a new splicing detection function that takes into account the statistical properties of satellite data.

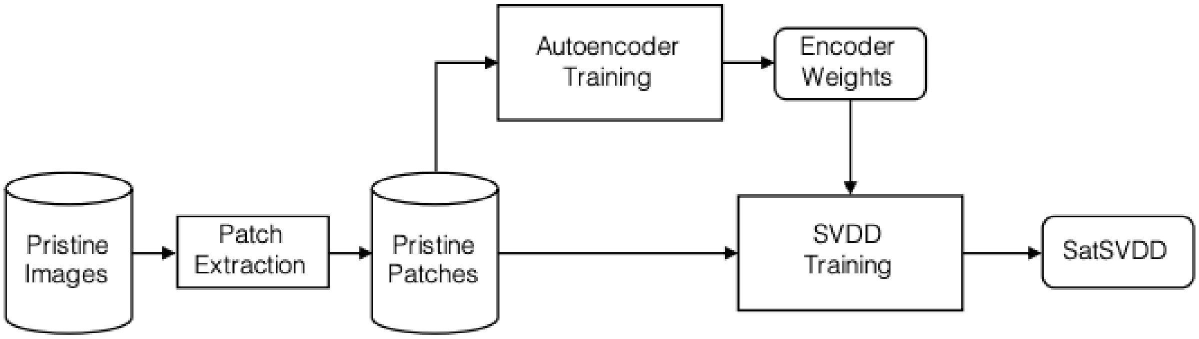


Figure 33. Overview of the Whole Training Process of Sat-SVDD

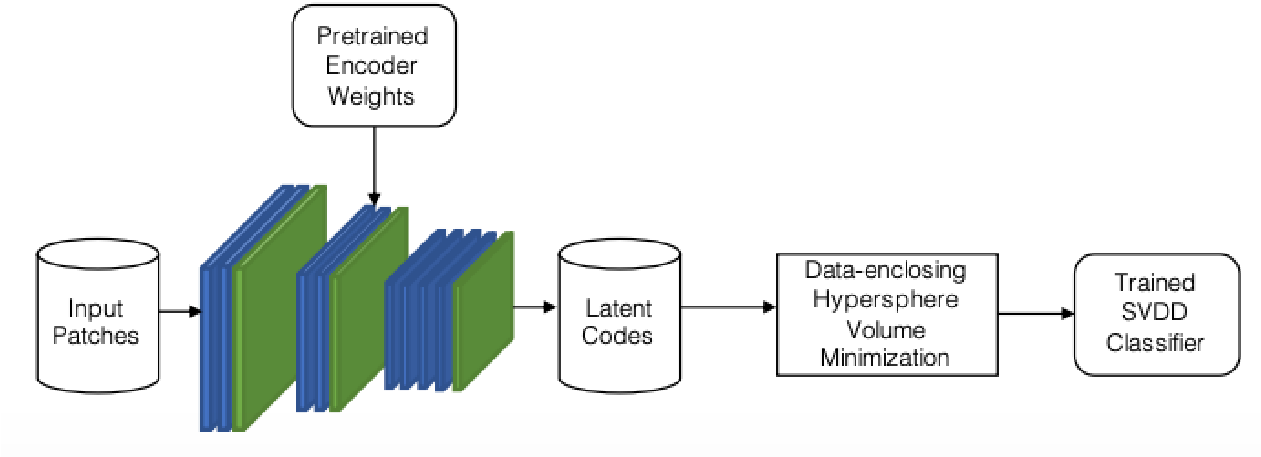


Figure 34. Overview of the Sat-SVDD Training Step

Published Paper:
J. Horváth, D. Güera, S. K. Yarlagadda, P. Bestagini, F. Zhu, S. Tubaro, E. J. Delp, "Anomaly-based Manipulation Detection in Satellite Images", *IEEE Conference on Computer Vision and Pattern Recognition*, Workshop on Media Forensics (CVPRW), Long Beach, California, June 2019.

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Forensic Analysis of Satellite Imagery
Major Technical Approach: Supervised approach with cGAN

Satellite images could be modified in a number of ways, such as inserting objects into an image to hide existing scenes and structures. We utilize a Conditional Generative Adversarial Network (cGAN) to detect the presence of such spliced forgeries in satellite images. Additionally, we identify the forgery shapes and locations. A trained cGAN analyzes a satellite image **I** and estimates a forgery mask **M** of the same resolution that indicates the likelihood that each pixel belongs to a spliced region. Trained on both pristine and falsified images, our supervised method achieves high success on these detection and localization objectives.

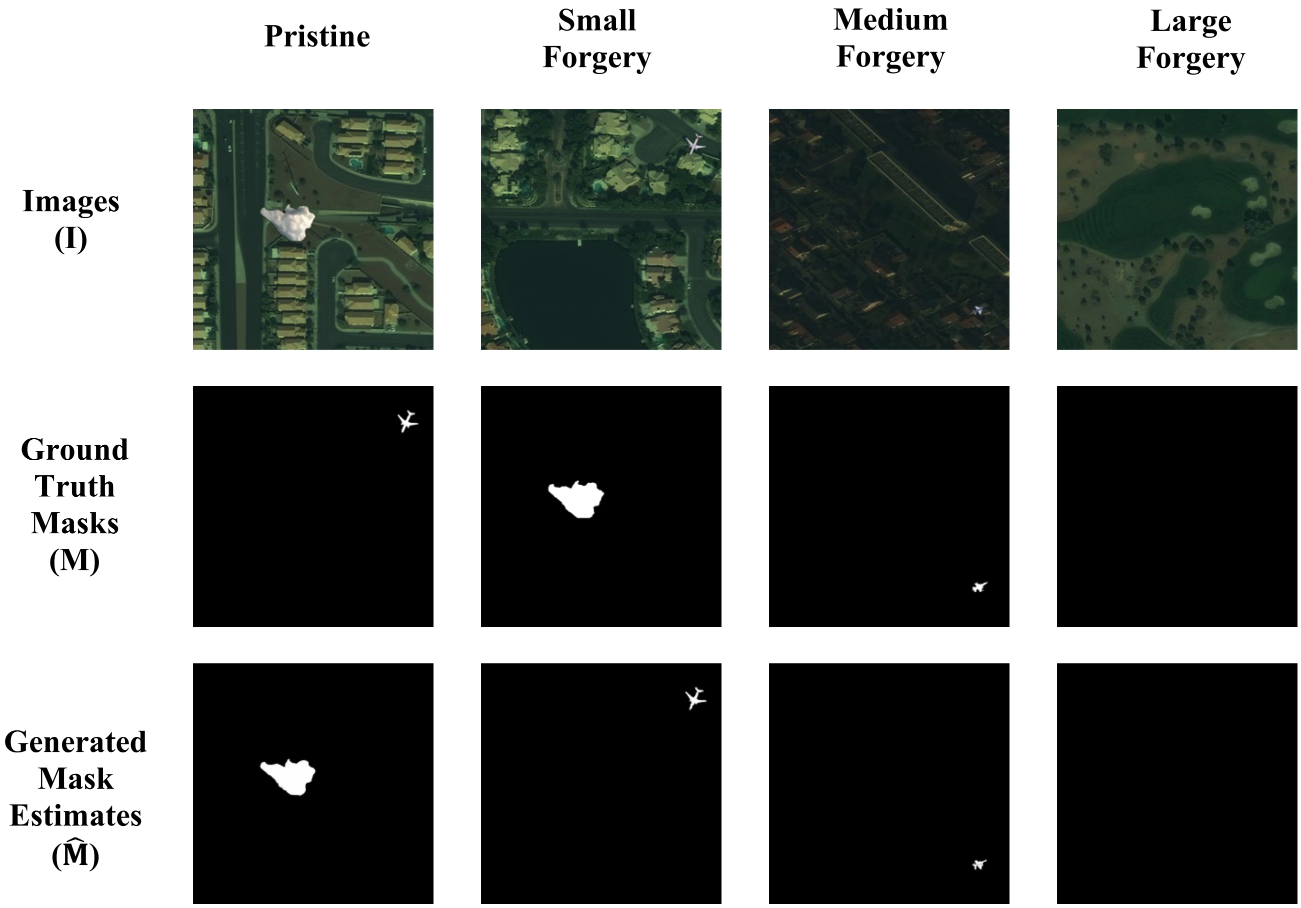


Figure 35. Input Images, Ground Truth Masks, and Generated Mask Estimates

Published Paper:
E. R. Bartusiak, S. K. Yarlagadda, D. Güera, F. M. Zhu, P. Bestagini, S. Tubaro, E. J. Delp, “Splicing Detection and Localization in Satellite Imagery using Conditional GANs”, *IEEE International Conference on Multimedia Information Processing and Retrieval (MIPR),* March 2019, San Jose, CA

Published Thesis:
E. R. Bartusiak, "An Adversarial Approach to Sliced Forgery Detection and Localization in Satellite Imagery", Master's dissertation*, Purdue University*, West Lafayette, IN, May 2019.

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Unsupervised splicing detection, localization in satellite images
Major Technical Approach: Deep Belief Networks with Uniform-Uniform RBMs

Satellite images are more accessible with the increase of commercial satellites being orbited. These images are used in a wide range of applications including agricultural management, meteorological prediction, damage assessment from natural disasters and cartography. Image manipulation tools including both manual editing tools and automated techniques can be easily used to tamper and modify satellite imagery. One type of manipulation that we examine in this paper is the splice attack where a region from one image (or the same image) is inserted ("spliced") into an image. In this paper, we present a one-class detection method based on deep belief networks (DBN) for splicing detection and localization without using any prior knowledge of the manipulations. We evaluate the performance of our approach and show that it provides good detection and localization accuracies in small forgeries compared to other approaches.

The main contributions of the paper are that we present a new technique to generate datasets containing satellite images with splicing manipulations. We also introduce a method for splicing manipulation detection and localization using DBNs that does not require any manipulated data during training. We then evaluate our method with multiple configurations of RBMs

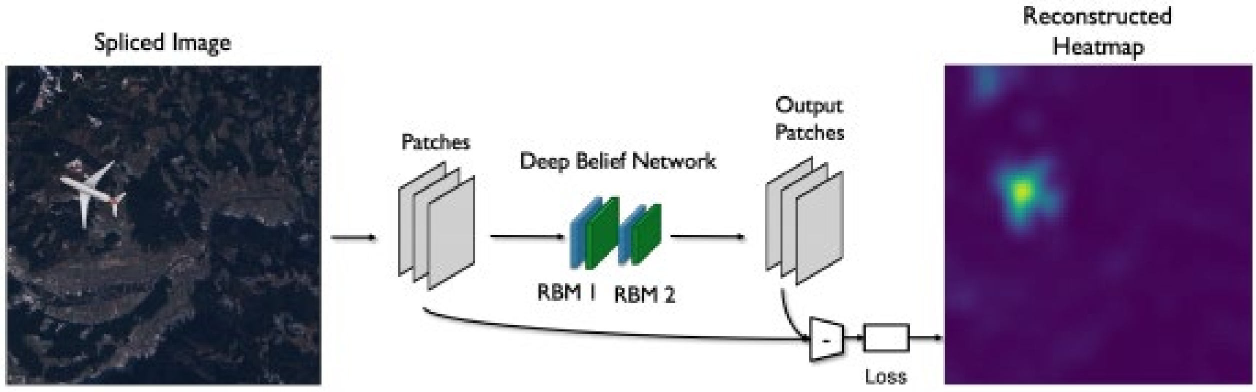


Figure 36. Method Overview for Generating the Heatmap and the Detection Score

Published Paper:
J. Horváth, D. Mas Montserrat, H. Hao, E. Delp, "Manipulation Detection in Satellite Images Using Deep Belief Networks", IEEE Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, Washington, June 2020. https://arxiv.org/abs/2004.12441

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Unsupervised splicing detection, localization in satellite images
Major Technical Approach: Gated PixelCNN

In this work, we show how PixelCNN and Gated PixelCNN, two generative autoregressive models, can be used to detect pixel-level manipulations. These neural networks, commonly used to generate new images, can model the distribution of a pixel given a set of previously seen pixels (neighboring pixels). These neural networks can assign a conditional likelihood value to a given pixel, and in turn, a likelihood value to a complete image. Through sampling from the pixel distribution, new images can be generated in a sequential fashion. Furthermore, manipulated pixels can be detected by selecting the pixels with a low likelihood assigned by the neural network. By averaging the likelihood estimated by an ensemble of multiple networks, the method is able to obtain a more accurate manipulation localization. Figure 37 presents the proposed ensemble where multiple networks process the input image and its flipped and rotated versions. Then, all predictions are averaged in order to obtain a robust prediction.

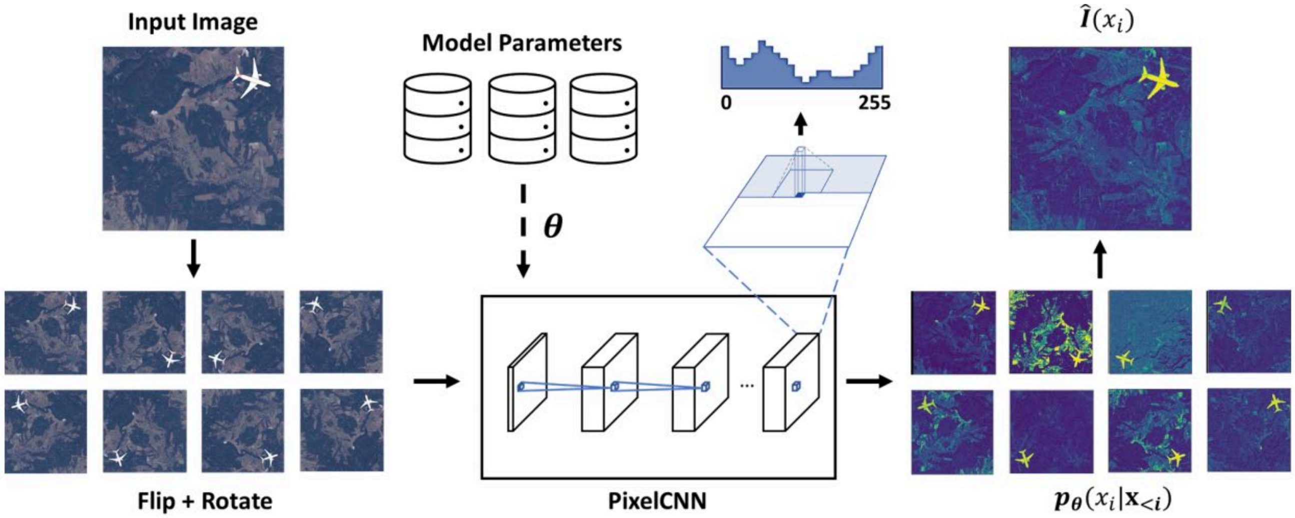


Figure 37. Proposed Ensemble: Multiple Models Process the Flipped and Rotated Input Images. The Prediction of Every Network Is Averaged Obtaining a Final Robust and Accurate Likelihood Estimate for Each Pixel of the Image. The 8 Images of pθ(xi|x<i) are Plotted as − log pθ(xi|x<i) for Visualization Purposes.

Published Paper:
D. M. Montserrat, J. Horváth, S. K. Yarlagadda, F. Zhu, and E. J. Delp, "Generative Autoregressive Ensembles for Satellite Imagery Manipulation Detection", IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1-6, December 2020.

Contract Information:
Team: Purdue Team, Purdue Sub-Team

POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Unsupervised satellite image forgery detection and localization
Major Technical Approach: Generative adversarial network (GAN) and one-class support vector machine (SVM)

Current satellite imaging technology enables shooting high resolution pictures of the ground. As any other kind of digital images, overhead pictures can also be easily forged. However, common image forensic techniques are often developed for consumer camera images, which strongly differ in their nature from satellite ones (e.g., compression schemes, post-processing, sensors, etc.). Therefore, many accurate state-of-the-art forensic algorithms are bound to fail if blindly applied to overhead image analysis. Development of novel forensic tools for satellite images is paramount to assess their authenticity and integrity. In this paper, we propose an algorithm for satellite image forgery detection and localization. Specifically, we consider the scenario in which pixels within a region of a satellite image are replaced to add or remove an object from the scene. Our algorithm works under the assumption that no forged images are available for training. Using a generative adversarial network (GAN), we learn a feature representation of pristine satellite images. A one-class support vector machine (SVM) is trained on these features to determine their distribution. Finally, image forgeries are detected as anomalies. The proposed algorithm is validated against different kinds of satellite images containing forgeries of different size and shape.

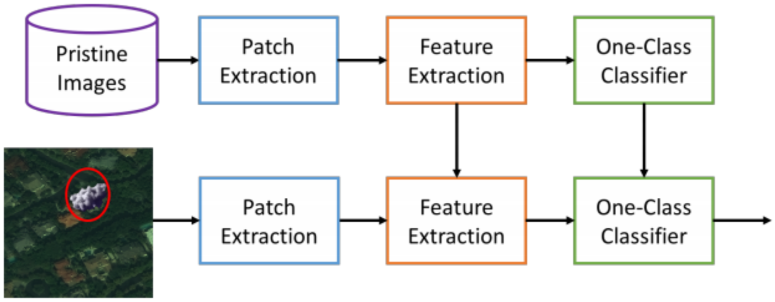


Figure 38. Pipeline of the Proposed Method. At Training Time, the Feature Extractor and One-Class SVM Learn Their Models from Pristine Images Only. At Testing Time, Forged Areas Are Detected as Anomaly With Respect to the Learned Model.

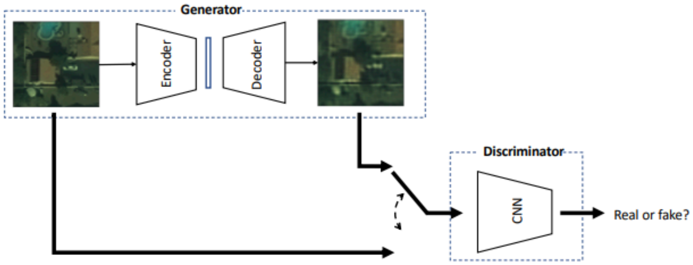


Figure 39. Architecture of the Used GAN

Published Paper:
S. K. Yarlagadda, D. Güera, P. Bestagini, F. Zhu, S. Tubaro, E. J. Delp, "Satellite Image Forgery Detection and Localization Using GAN and One-Class Classifier", *IS&T Electronic Imaging (EI)*, vol. 2018, no. 7, pp. 214-1-214-9, Burlingame, California, January 2018.
DOI: 10.2352/ISSN.2470-1173.2018.07.MWSF-214

3.2.2 University of Siena

Contract Information:
Team: Purdue Team, Siena Subteam
POC: Prof. Mauro Barni, University of Siena, barni@diism.unisi.it

Technical Problem: Copy-move localization with source-target disambiguation
Technical Approach: disambiguation based on multiple-branch CNNs, geometric transformation estimation

There are many Copy-Move (CM) detection and localization algorithms in the literature, determining whether a given image contains cloned regions, or so called nearly duplicate regions. Such algorithms can only detect the copy-move forgery and localize the nearly duplicate areas, providing a binary mask that highlights both the source region and its displaced version, without identifying which of the two regions corresponds to the source area and which to the target one. However, distinguishing between source and target region in a copy-move forgery is of primary importance in order to be able to correctly localize the tampering given that only the target region corresponds to the tampered area.

To address this problem, we propose a new CNN-based architecture. Given the binary localization mask produced by a generic copy-move detector, our method permits to derive the actual tampering mask, by identifying the target and source region. The main idea behind the proposed method is to exploit the non-invertibility of the copy-move transformation, due to the presence of interpolation artifacts and local post-processing traces in the spliced region.
More specifically, the method exploits the non-invertibility of the CM process, caused by the interpolation (and also local post-processing), to determine the *direction* of the transformation (forward direction). The algorithm first estimates the transformation matrix that maps one region into the other (and vice versa); then, the forward direction of the transformation is guessed by comparing each region with the one obtained by applying the transformation.
To do this, we resort to a multi-branch CNN architecture (DisTool) consisting of two main parallel branches, looking for two different kinds of CM-traces. The first branch, named 4-Twins Net, consists of two parallel Siamese networks, trained in such a way to exploit the non-invertibility of the copy-move process caused by the interpolation artifacts often associated to the copy-move operation. The second branch is a Siamese network designed to identify artifacts and inconsistencies present at the boundary of the copy-moved region. The soft outputs of the two branches are fused through a simple fusion module. The fusion strategy weights the outputs of the two networks based on the estimated transformation, that takes into account the reliability of the networks' decision in the various cases (e.g., the 4-Twins does not work well when the CM

transformation is close to rigid, while it is very reliable in the other cases). Based on our tests, fusion can effectively improve the overall performance of the disambiguation.
A remarkable strength of the proposed method is that it works independently of the CM detection algorithm and hence it can be used on top of any such method. The difficulty of training an end-to-end architecture for both localization and disambiguation providing good performance on both tasks, also motivates the use of an independent tool for the disambiguation (that was the attempt of the only prior work addressing the source-target disambiguation problem, BusterNet [1]).

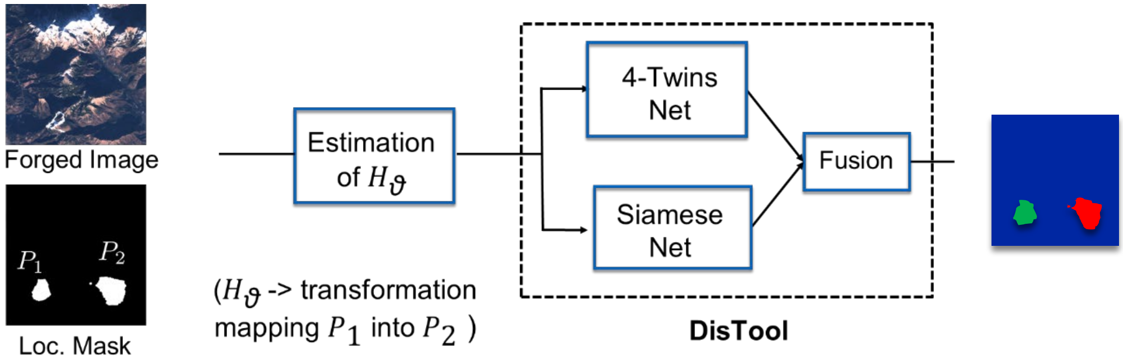

Forged Image
P_1
P_2
Loc. Mask
Estimation of H_ϑ
(H_ϑ -> transformation mapping P_1 into P_2)
4-Twins Net
Siamese Net
Fusion
DisTool

Figure 40. Scheme of the proposed CM disambiguation system (DisTool)

For the Siamese network to work well, it is necessary that large part of the boundary is contained in the network input patch, and/or the presence of such boundary can be easily revealed in the input patch. This is often not the case with satellite images due to the different statistics of these images and the more general uniformity of content. In this case, the Siamese network might loose some effectiveness, especially when the size of the copy-moved area is large.

In order to have large part of the boundary that is captured in the input patch of the Siamese network, we proposed to consider the following pre-processing steps: i) proportionally resize the CM regions to a smaller size, before patch extraction, so that min(height, width) < 64 (64x64 being the network input patch size); ii) test the top-let, top-right, bottom-left, bottom-right crops (instead of considering the central crop), and consider the patch resulting in the *most confident*. This second measure increases further the chances that large part of boundary falls within the 64x64 input patch.
The modification is only applied during testing, without needing to retrain the model. The resizing operation in fact it is not supposed to damage the network behavior (given that the Siamese Net is trained to look at the behavior across the boundary of the regions). This is confirmed in the experiments.

[1] Y. Wu, W. Abd-Almageed, and P. Natarajan, “BusterNet: Detecting copy-move image forgery with source/target localization,” ECCV 2018.

Published Paper:
M. Barni, Q. -T. Phan, B. Tondi, "Copy Move Source-Target Disambiguation Through Multi-Branch CNNs," *IEEE Transactions on Information Forensics and Security (TIFS), vol. 16, pp. 1825-1840, December 2020.* DOI: 10.1109/TIFS.2020.3045903

Contract Information:
Team: Purdue Team, Siena Subteam
POC: Prof. Mauro Barni, University of Siena, barni@diism.unisi.it

Technical Problem: Generation of multispectral images using GANs
Technical Approach: Generation using progressive GAN and style transfer using NICEGAN, CycleGAN and pix2pix.

Images generated using GANs are progressively improving to an extent that they are hard to distinguish even from an expert’s perspective. Their deceiving qualities made them very popular for image generation or image style transfer. However, all of the proposed architectures were targeted towards RGB images. Hence, we trained several GAN architectures to generate multispectral satellite images mainly Sentinel-2 images. The first proposed method was to use progressive GAN to generate sentinel 2 like images from scratch by adapting the architecture to 13 bands images. For this model, we used an already curated dataset of sentinel-2 images of resolution 256x256 for training.

We also applied style transfer for two tasks. The first being land cover transfer where images transfer from vegetation to barren and vice versa while the second is season transfer where images are transferred from winter to summer and vice versa. For the first task, we experimented with two GAN architectures, one being NICEGAN while the other was CycleGAN. NICEGAN is an adaptation of CycleGAN. For these architectures, we built our own land cover dataset of sentinel-2 level-1C that has two domains one of which was vegetation while the second was barren. We used 512x512 image resolution for training and also experimented with training only the 4 bands that are sampled at 10m GSD or training all 13 bands. For the second task that was season transfer, we trained pix2pix models on paired datasets where each image had its representation in both the source domain and the target domain. Subsequently, we created several paired datasets where each pair of image has the same region but one part representing summer while the other representing winter. We trained different models on different regions (China, Scandinavian countries and Alps). In addition, the models were trained either on 4 or 13 bands.

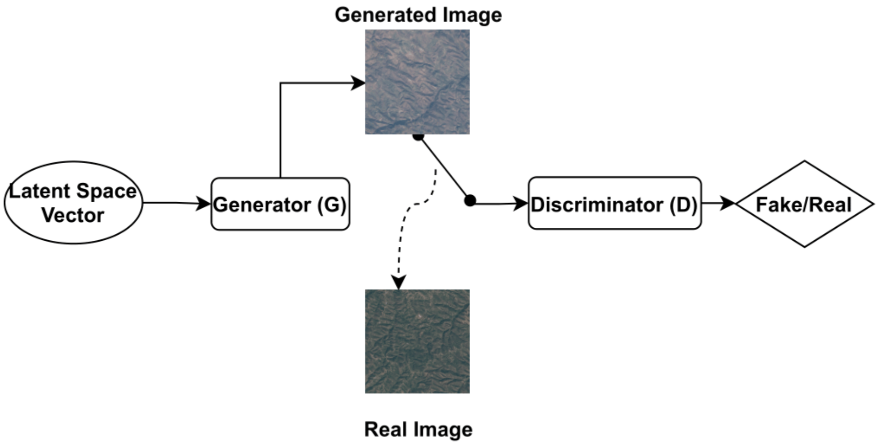


Figure 41. Basic GAN

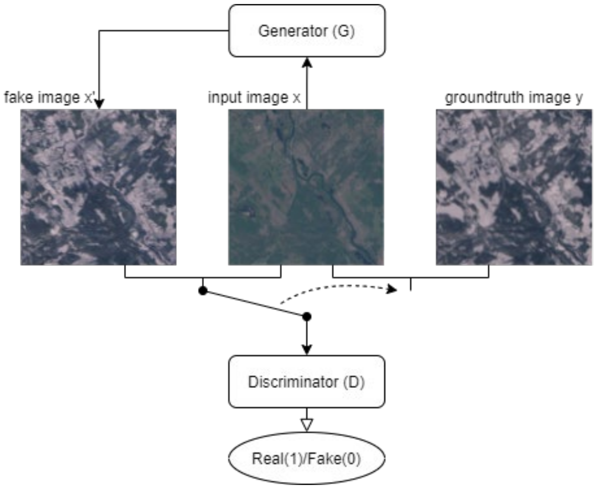


Figure 42. pix2pix for season transfer

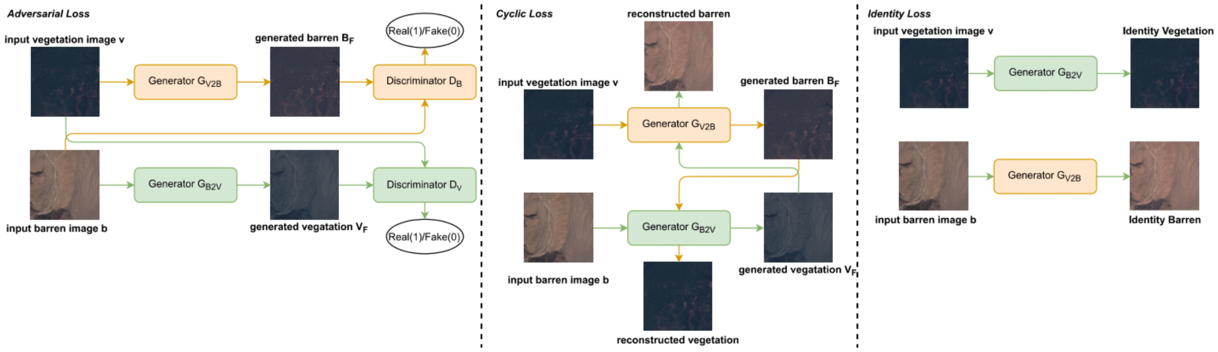


Figure 43. CycleGAN

Published Papers:
L. Abady, M. Barni, A. Garzelli, B. Tondi, "GAN Generation of Synthetic Multispectral Satellite Images," *SPIE Image and Signal Processing for Remote Sensing XXVI, vol. 11533, pp. 122-133, Online, September 2020.* DOI: 10.1117/12.2575765

L. Abady, J. Horváth, B. Tondi, E. J. Delp, M. Barni, "Manipulation and Generation of Synthetic Satellite Images Using Deep Learning Models", *SPIE Journal of Applied Remote Sensing* (submitted)

3.2.3 Politecnico di Milano

Contract Information:
Team: Purdue Team, PoliMi Sub-Team
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: SAR image forgery detection and localization
Major Technical Approach: CNN and clustering-based approaches

Considering the problem of localizing forgeries on SAR images, we focused on the scenario of image copy-paste, also in presence of editing operations. This is, given two SAR pictures (a donor and a host), a region is selected from the donor, it is optionally edited, and it is finally pasted on the host image. Given one of these images, our goal is to detect the use of this manipulation technique and localize the forged region.

To do so, we leverage a pipeline composed by the following steps (as shown in Figure 44): i) we extract a peculiar fingerprint from the SAR image under analysis; ii) we analyze the fingerprint with a binarization technique in order to provide a binary forgery mask.

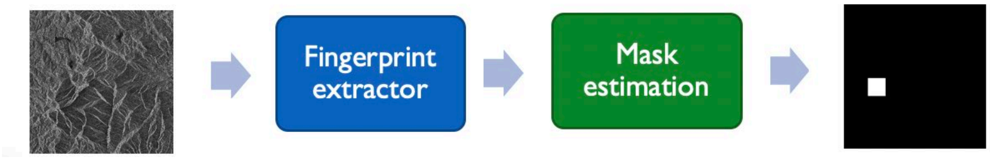


Figure 44. SAR Forensic Pipeline

Concerning the first step, we developed a SAR fingerprint extraction technique. This technique is based on a convolutional neural network that acts as noise extractor but is trained using a siamese procedure. At training stage, two instances of the network are used to extract a noise pattern from two image patches. If the two patches come from the same image, the MSE between the noise patterns is imposed to be low. Conversely, if the two patches come from different images and / or have undergone different processing steps, the MSE is imposed to be high. After training, only one noise extraction network is used.

Given a noise-like fingerprint extracted with the fingerprint extractor, we have three different options. Indeed, we developed three different versions of mask estimation algorithms. The first one is the fastest one and it is based on K-means clustering: the methods try clustering the fingerprint into two regions, the pristine and the forged parts. The second method is based on gaussian mixture models: we try fitting a mixture of gaussian on the fingerprint noise, and check whether different regions belong to different gaussian distributions. This method is more accurate but slower. Finally, the third method exploits a U-net for image segmentation: given the fingerprint, a CNN outputs a binary mask highlighting the possible forgery. This method makes use of a GPU if available, thus providing a significant speed-up.

Published Paper:
E. D. Cannas, N. Bonettini, S. Mandelli, P. Bestagini, S. Tubaro, "Amplitude SAR Imagery Splicing Localization**,"** *IEEE Access, vol. 10, pp. 33882-33899, March 2022.* DOI: 10.1109/ACCESS.2022.3161836

Contract Information:
Team: Purdue Team, PoliMi Subteam
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: Panchromatic image source attribution
Major Technical Approach: CNN and uncertainty-analysis approach

Manipulation of overhead images has become a threat only recently due to the current possibility of gathering satellite data. Therefore, state-of-the-art techniques for their forensic analysis are still underdeveloped, and they mostly focus on forgery detection and localization.

In this work, we consider a different forensic problem that has been broadly studied for natural images but not yet investigated for satellite data to the best of our knowledge: image source attribution. Attributing an image to its source means detecting which acquisition device has been used to capture the image under analysis. For natural images, this problem has been studied at different granularity in the literature: some methods consider the problem of detecting which kind of device was used for the acquisition (e.g., camera vs. scanner); other methods consider the problem of detecting the specific brand or model (e.g., Sony vs. Canon); other methods considered the problem of detecting the specific instance of the device (e.g., this iPhone X vs. that iPhone X).

Our goal is to understand which satellite has been used to acquire a panchromatic image under analysis. More specifically, we consider this attribution problem both in closed-set and open-set. In the first one, we assume that the image may only come from a set of known satellites. In the second one, we assume instead that the image may also come from a satellite that is unknown to the system, yet the system should be able to label the image as such. The proposed method is based on the use of a Convolutional Neural Network (CNN) that acts as classifier. In order to deal with the open-set problem, we exploit ideas from the world of model uncertainty analysis. Specifically, we adapt and compare two different strategies: Deep Ensemble (DE) and Monte Carlo Dropout (MCD). The main idea is that it is possible to run multiple attribution tests on a single image under analysis. If the image comes from an unknown satellite, the trained CNN should exhibit an uncertain behavior across its responses. This behavior can be captured and used to reject the image as coming from an unknown satellite. Otherwise, the image is attributed to the originating satellite.

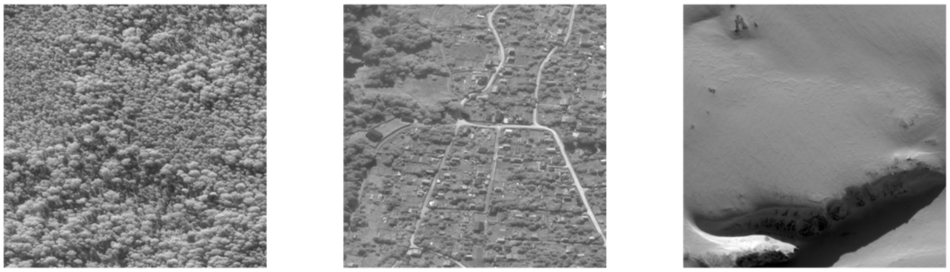


Figure 45. Examples of panchromatic images

The satellite attribution problem we tackle can be formulated as follows. Given a panchromatic image *I* (see Figure 45) and a set of *M* known satellites, we want to understand if *I* has been acquired with one of the *M* known satellites or not (i.e., open-set). If the answer is positive, we also want to determine which satellite's sensor among the *M* considered ones has generated it (i.e., closed-set).

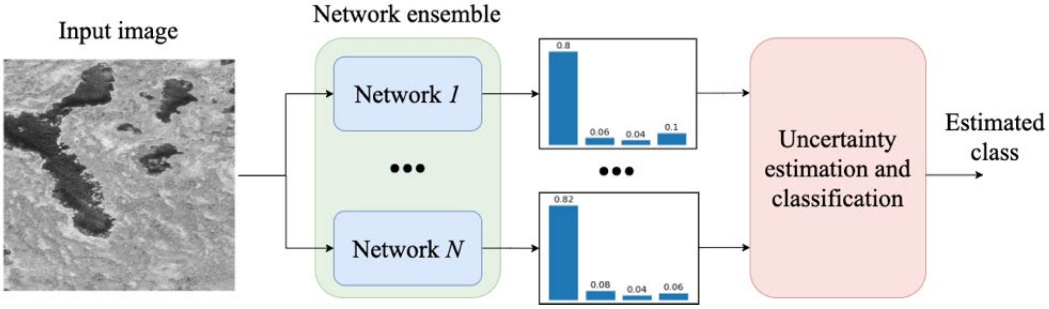


Figure 46. Pipeline of the proposed attribution method

In order to do so, we rely on a CNN ensembling method as depicted in Figure 46. Considering *N* CNNs trained in closed-set over the *M* known classes, we test the image against all of them, obtaining *N* attribution scores, and then we evaluate the classification uncertainty of the ensemble based on these responses. If the evaluation of the uncertainty is high, we conclude that the image does not belong to any of the known satellites. Otherwise, we attribute the image according to the closed-set classification response.

Published Paper:
E.D. Cannas, S. Baireddy, E. R. Bartusiak, S.K. Yarlagadda, D. Mas Montserrat, P. Bestagini, S. Tubaro, E. J. Delp, "Open-Set Source Attribution for Panchromatic Satellite Imagery", *IEEE International Conference on Image Processing, pp. 3038-3042, Anchorage, AK, September 2021.*
DOI: 10.1109/ICIP42928.2021.9506600

3.3 Scientific Integrity Project

In this section we describe our efforts on the Scientific Integrity project.

3.3.1 Purdue University

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: PDF Image Extraction
Major Technical Approach: Image Extraction for Scientific Integrity

We provide a tool to extract images from scientific papers saved as PDF files. The tool is implemented in Python, integrated with PDFMiner, a permissive free python-based tool for extracting information from PDF documents. The tool can extract raster images such as JPG, PNG, BMP and GIF images, but cannot handle vector images, for example EPS and SVG images. We extract vector images as series of sub-images rather than an entire image. It supports most regular and special color spaces, such as gray scale, RGB, CMYK, indexed and separation color spaces.

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Overlaid Text Content on Image
Major Technical Approach: Image Segmentation for Scientific Integrity

We provide a method to reduce the false alarms introduced by text instances by incorporating text detection in the system. The module is able to output a mask indicating text location which can be then ignored by copy-move detection algorithm.

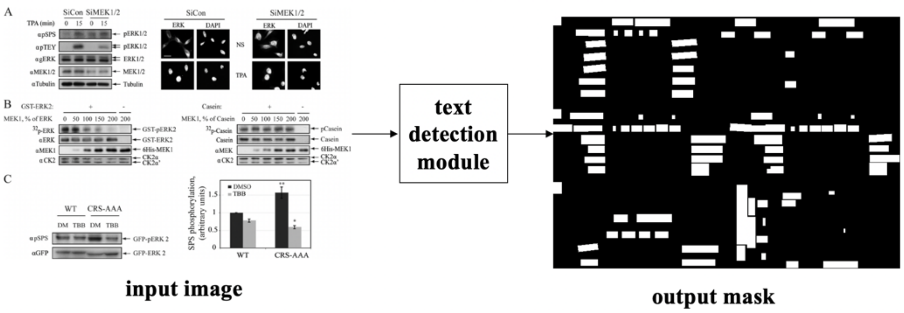


Figure 47. An Illustration of Module Input and Output

We use Character Region Awareness for Text Detection (CRAFT) model as the text detector for biomedical figures. CRAFT is a character level text detector that can effectively localize text instance by exploring each character and affinity between characters. In terms of architecture, it is a two-channel fully convolutional network that has a structure similar to U-Net. It adopts VGG-16 with batch normalization as backbone and also introduces skip connections in decoding. The network outputs both region map and affinity map. In post-processing, we apply thresholding, morphological operations (opening and erosion) followed by connected-components labeling in OpenCV to get the final text segmentation mask.

3.3.2 Politecnico di Milano

Contract Information:
Team: Purdue Team, PoliMi Sub-Team
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: Analysis of scientific images
Major Technical Approach: CNNs for PDF parsing, Fourier Mellin transform for blot copy-move detection, image enhancement operations.

Scientific publications containing forged images undermine the credibility of good science. On one hand, publishing “wrong” images might be due to unwanted mistakes during the writing process. On the other hand, this might be due to author misconduct. In any case, it is of paramount importance to be able to detect forgeries in scientific images. This process is often done manually by operators trained to spot forgeries by visual inspection. Indeed, after a publication is marked as suspect, different experts analyze the published results and pictures searching for possible inconsistencies. To do so, they often edit the images under analysis applying standard image processing operations to enhance possible artifacts. In order to help forensic investigators with their duties, we focused on three main activities: PDF document analysis through CNNs; copy-move detection of western blots; implementation of image enhancement operations.

The first steps toward the verification of scientific integrity of digital images have been done by considering the problem of image and text segmentation. As a matter of fact, being able to collect and extract digital images from digital versions of scientific papers is the first problem to solve to enable automatic assessment of the veracity and authenticity of scientific images. To this purpose, we developed a technique to process scanned versions of scientific papers, and extract images from the text. This is done by considering two different convolutional neural networks trained on purpose. On one hand, these solutions can be helpful in segmenting older papers that results from scanning operations (e.g., for provenance analysis). On the other hand, these tools can be used together with pdf analysis libraries to enhance their capabilities.

One of the recurring forgeries applied to scientific images is western blots duplication / copy-move. This means copying one or many blots, optionally applying some processing operation (e.g., resizing, rotation, etc.), and paste them in a different region of the same image, or in a different image as well. General purpose copy-move detectors are meant to spot these kinds of forgeries.

However, if the corpus of images to be analyzed is conspicuous, execution time might become an issue. In order to alleviate this problem, we focused on the development of a fast blot detection and matching solution. This is based on a series of image processing operations paired with a segmentation algorithm for blot detection and segmentation. The matching step is based on the use of Fourier Mellin Transform, which enables comparing images that have been possibly rotated, scaled and cropped.

Finally, we implemented a series of image processing operations ranging from brightness to contrast adjustment in order to help the analysist visualizing images the way they are used to. This new feature is embedded in the tool developed with the rest of the team and improves the available user interface. Indeed, thanks to this new addition, analysts can avoid opening suspect images with other image processing tools. Conversely, they can perform their analysis within a single software suite.

Contract Information:
Team: Purdue Team, PoliMi Subteam
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: Analysis of synthetic scientific images
Major Technical Approach: Feature extraction and one-class classifier

The widespread diffusion of synthetically generated content is a serious threat that needs urgent countermeasures. As a matter of fact, the generation of synthetic content is not restricted to multimedia data like videos, photographs or audio sequences, but covers a significantly vast area that can include biological images as well, such as western blot and microscopic images.

In this work, we focused on the detection of synthetically generated western blot images. These images are largely explored in the biomedical literature and it has been already shown they can be easily counterfeited with few hopes to spot manipulations by visual inspection or by using standard forensic detectors.

To overcome the absence of publicly available data for this task, the first step has been to create a new dataset comprising more than 14K original western blot images and 24K synthetic western blot images, generated using four different state-of-the-art generation methods. To generate synthetic western blot images, we adopt well-known CNN architectures from the literature of natural images generation. In particular, we used: pix2pix; CycleGAN; StyleGAN2-ADA; Denoising Diffusion Probabilistic Model (DDPM). As show in Figure 48, the first two techniques allow us to generate blots starting from binary masks, thus controlling blots position. Conversely, the last two techniques randomly generate western blots according to different styles.

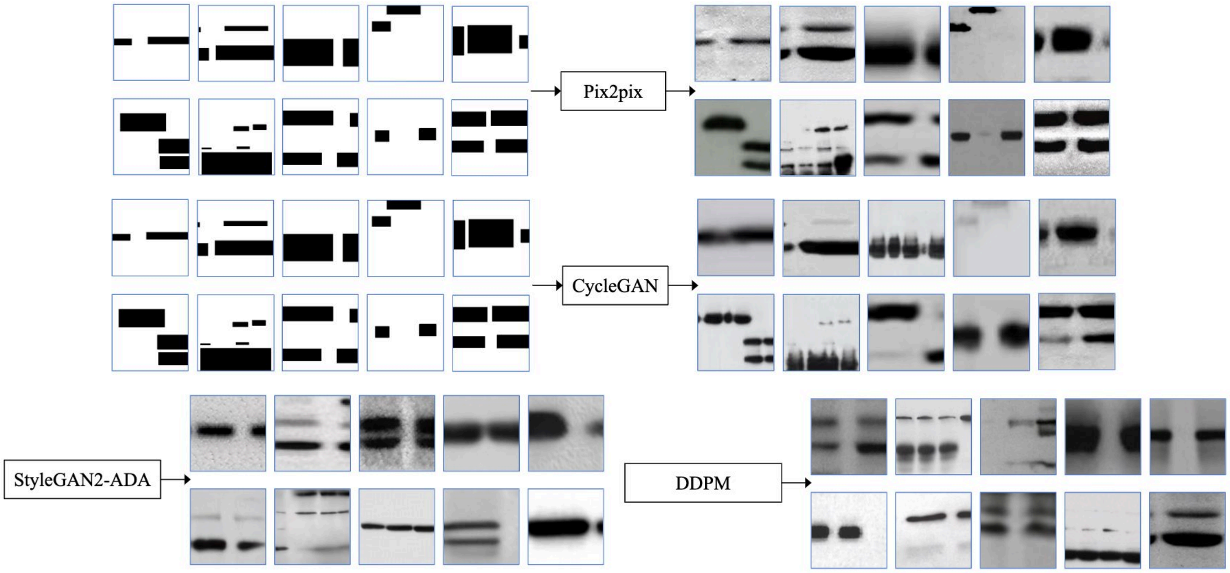


Figure 48 - Blots generation process using the four different techniques

Concerning the synthetic blot detection part, we proposed a method that first pre-process the image under analysis, then extract a series of features, and finally feed these features to a one-class classifier. The pre-processing steps consists of a high-pass filtering operation, as illustrated in Figure 49.

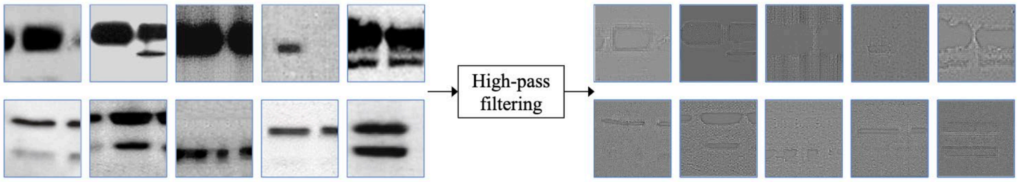


Figure 49 - High-pass filtering process

As features, we made use of co-occurrences matrices computed on top of the high-pass filtered version of each image. Finally, we used a one-class SVM as classifier. This was trained only considering pristine western blots images. Synthetic images are detected as non-belonging to the pristine class distribution.

3.3.3 University of Notre Dame

Contract Information:
Team: Purdue Team, Notre Dame Sub-Team
POC: Prof. Walter Scheirer, University of Notre Dame, walter.scheirer@nd.edu, 574-631-2436

Major Technical Problem: Image Provenance Analysis
Major Technical Approach: Provenance Analysis for Scientific Integrity

Images have been used in scientific publications since the early days of science, to illustrate the proposed processes, to aid in the explanation of theories, or to present the outcome of experiments.

Later on, with the advent and popularization of photography, photographs were added to the scientific repertory, quickly replacing abstract and graphical arts, especially in the stage of documenting experimental results. In some scientific fields, such as Biomedicine, images captured by dedicated apparatuses are even accepted as the results themselves, constituting the elements to be scrutinized while examining a hypothesis.

More recently, with the transition from classical photography to digital imaging, editing software entered the scene, allowing researchers to easily retouch and compose images. On the one hand, some edits are benign and acceptable, such as intensity calibrations for visual enhancement, or compositions that aim at making the comparison of different outcomes easier. On the other hand, some edits comprise either mistakes (when the authors, for instance, inadvertently provide misrepresentations of experimental results) or misconduct (when there is the intent to deceive readers). In this work, regardless of the purpose of the modification, we simply call it image post-processing, since it happens at the end of the digital imaging pipeline, after capture and digitalization. Therefore, when we detect an image post-processing operation, we are not claiming that it is a mistake or misconduct. We leave this judgement to the experts and to the competent audiences.

The usage of post-processed images in science is a relevant problem, since the cases of mistake and misconduct are obviously unwanted by publishers, demanding acts of retraction as soon as possible, in the case of unfortunate publication. Indeed, in addition to the negative impact on a researcher's reputation, it affects science credibility and research outcomes, leads to unfair funding decisions, and causes the development of ungrounded techniques.

Hence, in this work, we propose a workflow and introduce a system that implements this workflow, to uncover image post-processing attempts in scientific publications. The main aspect of the workflow is the systematic provision of a series of state-of-the-art image forensic algorithms to a collection of human experts, whose iterative feedback is fundamental to the final outcome. The implemented system, in turn, presents a rich graphical user interface (GUI) and was developed with the aim of easy extension, to quickly allow the inclusion of novel image forensic tools. Notre Dame provided a solution for the provenance analysis step. The integration of provenance into the system is done in the following way: the analyst selects an image for which they want to see the provenance graph. The provenance module then builds the provenance graph, which is displayed in the system GUI.

This solution [1] has been submitted to the appreciation of the image processing community for formal journal peer reviews.

[1] Paper Submitted to Scientific Reports: Moreira et al. "Provenance Analysis for Scientific Integrity," 2022

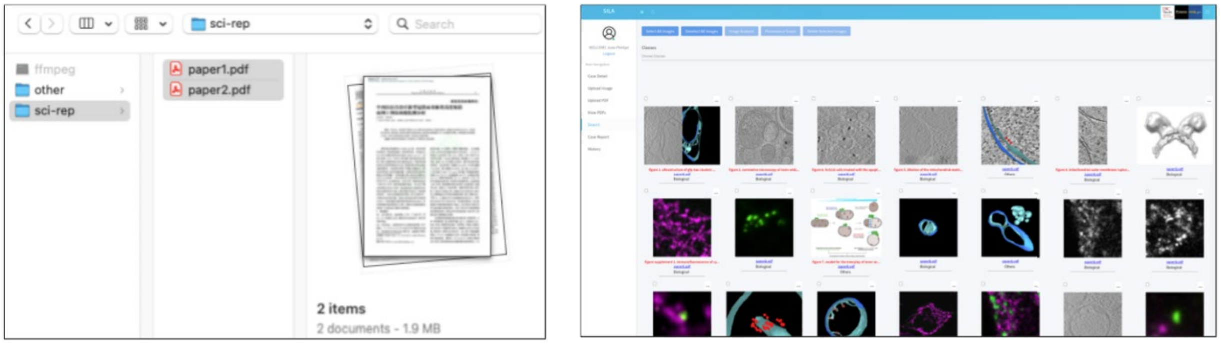
Contract Information:
Team: Purdue Team, TA.1.3, Notre Dame Sub-Team
POC: Prof. Walter Scheirer, University of Notre Dame, walter.scheirer@nd.edu, +1 (574) 631-2436

Major Technical Problem: Scientific Image Analysis
Major Technical Approach: SILA: A System for Scientific Image Analysis

Program Objectives: Find transition partners and transfer to them some of the technologies developed along the MediFor project.

In a period of five years (from 2011 to 2015), nearly 78% of the cases of research misconduct identified by the US Department of Health and Human Services (HHS), through the Office of Research Integrity (ORI), involved image manipulations and improper reuse [1]. As a response to the growing problems of mistakes and misconduct related to images in scientific publications, HHS has demonstrated interest to become a transition partner and absorb some of our forensic image analysis solutions to help them perform paper screenings in a more scalable way.

To address their interest and following their guidance as stakeholders, we developed SILA: a system for Scientific Image Analysis. SILA implements a novel human-in-the-loop computational workflow for scientific integrity verification to uncover image manipulations and reuse in scientific publications. It contains an easy-to-use graphical user interface (GUI) for people outside of computer science (see Figure 50) and was developed with the aim of easy extension to allow the inclusion of novel image analysis tools (see Figure 51).



(a) (b)

Figure 50. Implemented GUI. In (a), the operating-system agnostic interface allows the analyst to upload multiple scientific papers of interest as PDF files. In (b), the web browser-based system GUI is populated with content automatically extracted from the provided files.

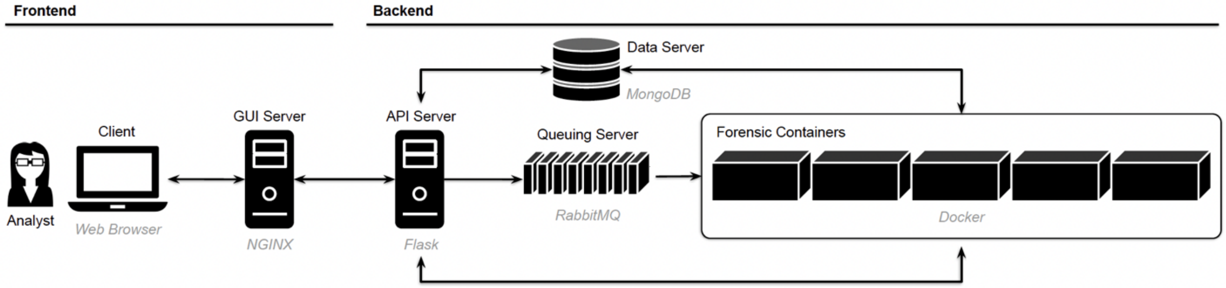


Figure 51. The architecture of SILA. We adopted a client-server model, with frontend components focused on usability, and backend components focused on scalability and extensibility. Other solutions can be added to the system if they are made available as Forensic Containers. Labels in italic on the bottom of the components detail the technology used to implement them. Arrows express the data flow between components.

SILA provides a unique and principled combination of image processing, image forensics, and computer vision tools, such as copy-move detection [2] and provenance analysis [3], all custom-tailored to the case of scientific images. It implements an end-to-end workflow – from Portable Document Format (PDF) files to image provenance graphs – to help experts in (1) spotting image post-processing events and (2) deciding how legitimate they are. While most tasks are automated (such as figure and caption extraction from the binary stream of the PDF files), the ultimate decision-making is left to the human experts, who have the final word based on the provided evidence.

Last but not least, we have also collected and annotated a new dataset containing both retracted (due to the documented presence of problems with figures) and unchallenged scientific papers, which we are making available to the community, in the hope that it becomes a benchmark for algorithm and system development. This dataset has 988 publications from different countries and provides a complete set of annotations for copy-move detection and provenance analysis.

A paper introducing SILA and the proposed dataset has been submitted to Springer Nature Scientific Reports and is currently under the second round of peer review.

[1] Office of Research Integrity. Data Graphs 2006–2015. Available at https://ori.hhs.gov/images/ddblock/ORI%20Data%20Graphs%202006-2015.pdf (2015). Accessed on Jun. 23, 2021.

[2] Cozzolino et al. “Efficient dense-field copy-move forgery detection,” IEEE T-IFS, Vol. 10, 2015.

[3] Moreira et al. “Image Provenance at Scale,” IEEE T-IP, Vol. 27, No. 12, 2018

3.3.4 University of Naples Federico II

Contract Information:
Team: Purdue Team, Unina Subteam
POC: Prof. Luisa Verdoliva, University Federico II of Naples, verdoliv@unina.it, +39 081 7683929

Major Technical Problem: Analysis of manipulated scientific images
Major Technical Approach: Development of a dense-based copy-move detector for scientific images

In the literature there are several cases of research misconduct in scientific publications. In this work we focused on the detection of copy-move manipulations. Copy-move operations comprise those manipulations in which a region of an image is cloned somewhere else within itself, usually with the intent to cover an undesired feature, by replicating either an existing object (duplication) or background pixels (removal). There are many algorithms that have been proposed for copy-move forgery detection in the literature. However, none of them was designed to handle the detection of cloned content within scientific images properly. Scientific images are often composed of multiple sub-components (panels) and text describing the content. To avoid false alarms due to text matching and analyze all the existing sub-panels within an image, our solution uses the panel segmentation step and includes an Optical Character Recognition (OCR) system to localize and remove text. Once panels are cleaned from text and extracted, the copy-move detector examines all the possible pairs of panels inside each figure, looking for visually similar content. Each analyzed pair thus leads to the generation of a pair of cloned-region masks. Based on the original location of the extracted panels, the multiple masks are combined at the end of the process to generate the final image clone mask, as shown in Figure 52.

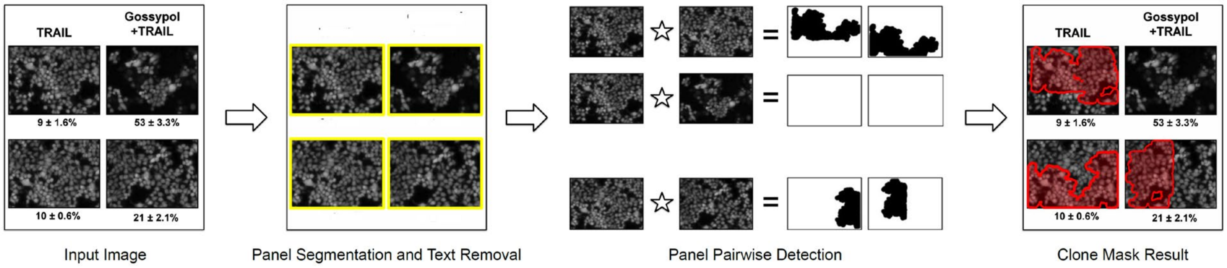


Figure 52. A block diagram depicting the operation of the copy-move detection approach

Concerning the panel pairwise copy-move detection step, we leverage dense-field copy-move approaches. Compared to other algorithms, the dense-field-based ones have the advantage of handling both additive (when entire objects are duplicated) and occlusive copy-move operations (when uniform and background small portions of pixels are duplicated to hide content). More specifically we rely on an efficient dense-field method for finding similar patches within an image. It thus includes a dense feature extraction step that uses Zernike moments to be robust to distortion and rotation operations, a randomized iterative algorithm for computing nearest neighbor fields, and a post-processing procedure based on dense linear fitting and morphological operations. Nonetheless, our approach adds extra steps to adapt to the peculiar nature of the images extracted from scientific papers, which largely differ from natural scenes, especially given the presence of uniform backgrounds, lower resolution, and limited range of pixel intensity values. Figures with western blots, for example, span a small range of pixel values, whose blots are very similar in appearance, even if they are not the same. As a consequence, many false-alarm matches arise through the images, needing mitigation. To reduce these random false matches, we introduce an additional constraint to the matching procedure: matching between two regions must hold both ways. This means that a region A is declared forged and a copy of a region B only if two compatible matches are found: the match from A towards B and vice-versa.

Lastly, besides the Zernike features, we also consider the Red, Green, and Blue (RGB) values of the image pixels. By considering the three-color channels of the images, the solution turns out to be more robust even to small changes in pixel intensity values. The strategy based on RGB data, on the contrary, is successful in the case of the cell images but misses the western blots. To take the best of both worlds, we implement a fusion of them.

3.3.5 USC Information Sciences Institute

Contract Information:
Team: USC Information Sciences Institute
POC: Prof. Wael Abd-Almageed, USC-ISI, wamageed@isi.edu, +1 703 248 6174
Major Technical Problem: Scientific Integrity
Major Technical Approach: A GUI for Scientific Integrity System
Program Objectives: develop a system GUI including multiple image analyzing methods for scientific papers
System Design:

- Design concept (modularity, scalability, deployability, etc)
- Backend
- Front end
- Module description
- Screen shots

Design Concept:
For the design we use a standard model-view-controller (MVC) design pattern, where the graphical user interface (GUI), backend servers, algorithms and data storage are designed to be isolated components that interact with each other to maintain state. The system has a front-end component designed in AngularJS, which is then served with a production grade Nginx server. The backend server is hosted separately and made to communicate with the front-end component using a Representational State Transfer (REST) framework.

The database used for this data storage is MongoDB, which is a NoSQL database and uses a JSON-like documents. MongoDB serves as the persistent storage for all user related data. All the components used in the application like Nginx server, MongoDB, Flask server and the various algorithms are Dockerized as containers for portability and scalability.

Backend:
The backend of the system contains a number of components, as follows:

- The backend server is implemented using Python Flask, with Secure Sockets Layer (SSL) security. For authentication we use JSON Web Token (JWT), which secures the server and verifies requests coming from the frontend system.
- The *Copy-Move* algorithm is implemented in pm-sci-int docker container, and also uses Python Flask and has its own endpoint running on different port.
- We use RabbitMQ message-broker, which interacts with copy move server and manages its requests in queue fashion.
- For database we use MongoDB which is also a docker container.
- Provenance analysis is performed and generated using the *Sciint-simple-image-retrieval* docker container
- Sub-panel extraction is performed using *panel-extraction* docker container

Frontend:
Our front-end system uses AngularJS 7 and is deployed on the hardware server as a docker container using Nginx web server. Each element within the front-end are further abstracted as components and the front-end is designed to be an interaction between various components.

4.0 RESULTS AND DISCUSSION

In this section we describe our results of our work. This is organized by sub-teams.

4.1 Main Medifor Project

In this section we describe our efforts on the “main” MediFor project. Later we describe our efforts on the sub-projects: Overhead Forensics and Scientific Integrity.

4.1.1 Purdue University

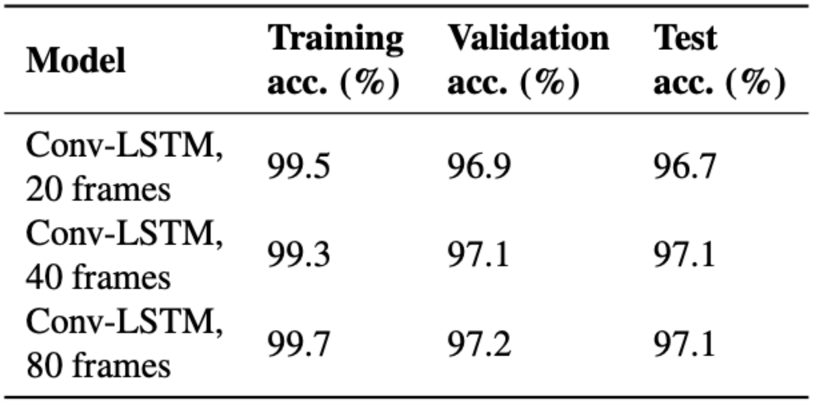
Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Deepfake Detection
Major Technical Approach: Convolution Neural Network and Recurrent Neural Network

Program Objectives: predict a score for each input video that indicates if the video has been manipulated or not.

The following table presents the results of experimentation with different sequence lengths provided to the network. At the time of publication, no state-of-the-art approach to deepfake detection existed, and there was no agreed standard dataset created with “deepfake” technology to evaluate approaches on.

Table 2. The Performance of our Model with Different Input Sequence Lengths



Model	Training acc. (%)	Validation acc. (%)	Test acc. (%)
Conv-LSTM, 20 frames	99.5	96.9	96.7
Conv-LSTM, 40 frames	99.3	97.1	97.1
Conv-LSTM, 80 frames	99.7	97.2	97.1

Published Paper:
D. Güera, E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks", *IEEE International Conference on Advanced Video and Signal-based Surveillance (AVSS), Auckland, New Zealand, November 2018*. DOI: 10.1109/AVSS.2018.8639163

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Deepfake Detection
Major Technical Approach: Convolution Neural Network and Recurrent Neural Network with Automatic Weighting

Program Objectives: predict a score for each input video that indicates if the video has been manipulated or not.

Table 3 presents the results of balanced accuracy. Because it is based on extracting features on the entire video, Conv-LSTM [1] is unable to capture the manipulations that happen within face regions. However, if the method is adapted to process only face regions, the detection accuracy improves considerably. Classification networks such as Xception [4], which provided state-of-the-art results in Faceforensics++ dataset [5], and EfficientNet-b5 [3] show good accuracy results. Our work shows that by including an automatic face weighting layer and a GRU, the accuracy is further improved.

Table 3. Balanced Accuracy of the Presented Method and Previous Work

Method	Accuracy
Conv-LSTM [1]	55.82%
Conv-LSTM [1] + MTCNN [2]	66.05%
EfficientNet-b5 [3]	79.25%
Xception [4]	78.42%
Ours	92.61%

Additionally, we evaluate the accuracy of the predictions at every stage of our method. Table 4 shows the balanced accuracy of the prediction obtained by the averaging the logits predicted by EfficientNet, the prediction of the automatic face weighting layer, and the prediction after the gated recurrent unit. We can observe that every stage increases the detection accuracy, obtaining the highest accuracy with the GRU prediction.

Table 4. Balanced Accuracy of at Different Stages of Our Method.

Method	Accuracy
Ours (Logits)	85.51%
Ours (Weights)	87.90%
Ours (GRU)	92.61%

[1] D. Güera and E. J. Delp, "Deepfake Video Detection using Recurrent Neural Networks", *IEEE International Conference on Advanced Video and Signal Based Surveillance*, November 2018, Auckland, New Zealand, https://doi.org/10.1109/AVSS.2018.8639163

[2] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint Face Detection and Alignment using Multitask Cascaded Convolutional Networks", *IEEE Signal Processing Letters*, vol. 23, April 2016, https://doi.org/10.1109/LSP.2016.2603342

[3] M. Tan and Q. V. Le, "Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks," *arXiv preprint arXiv:1905.11946*, May 2019, https://arxiv.org/abs/1905.11946

[4] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *IEEE conference on Computer Vision and Pattern Recognition*, July 2017, Honolulu, HI, https://doi.org/10.1109/CVPR.2017.195

[5] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics++: Learning to Detect Manipulated Facial Images," *IEEE International Conference on Computer Vision*, pp. 1–11, October 2019, Seoul, South Korea, https://doi.org/10.1109/ICCV.2019.00009

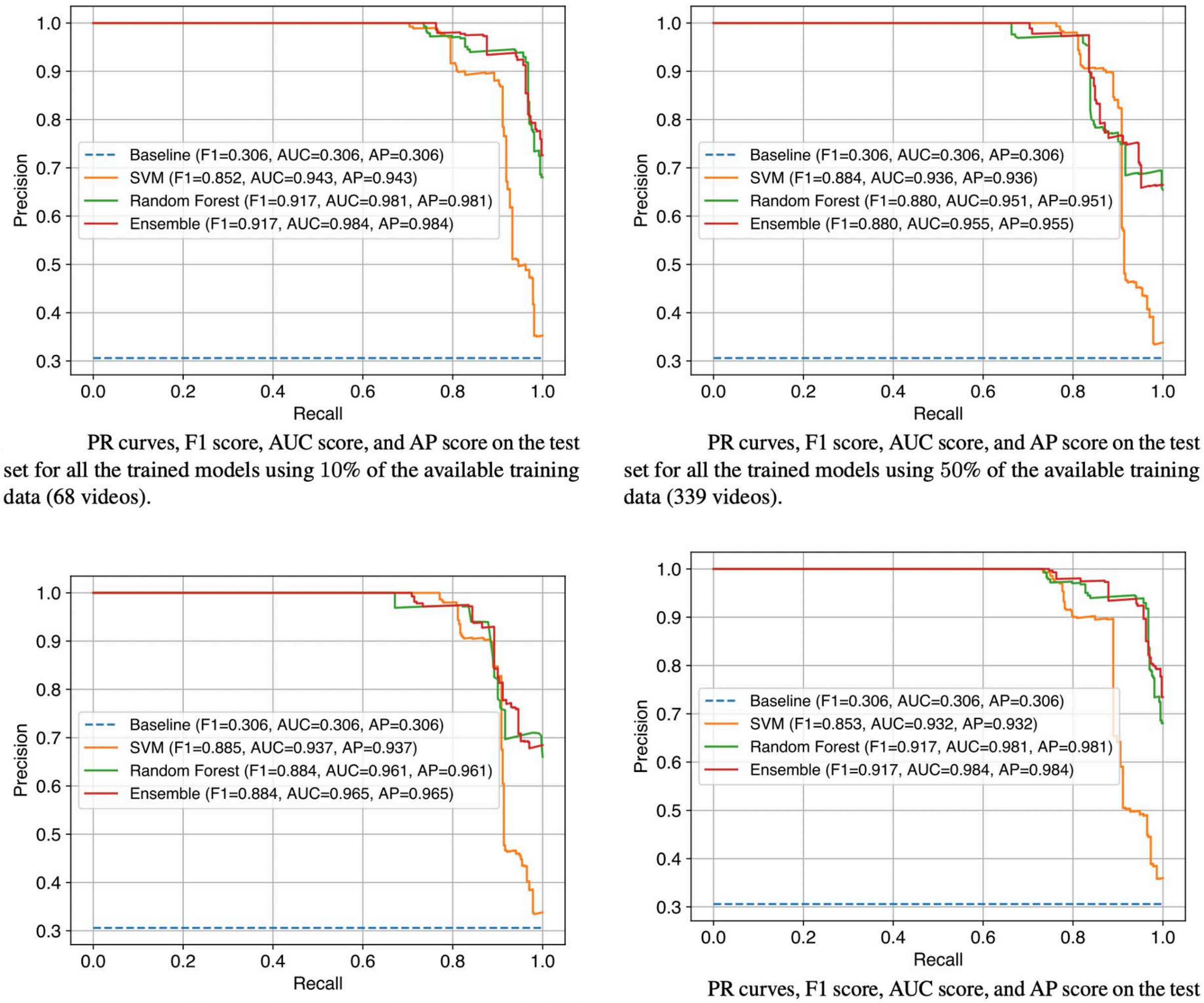
Published Paper:
D. M. Montserrat, H. Hao, S. K. Yarlagadda, S. Baireddy, R. Shao, J. Horváth, E. Bartusiak, J. Yang, D. Güera, F. Zhu, and E. J. Delp, "Deepfakes Detection with Automatic Face Weighting", *IEEE Computer Vision and Pattern Recognition Workshops*, June 2020, Seattle, Washington.
DOI: 10.1109/CVPRW50498.2020.00342

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Video Manipulation Detection
Major Technical Approach: Metadata-based Machine Learning

Program Objectives: predict a score for each input video that indicates if the video has been manipulated or not.

The following figure shows the effectiveness of our trained detector on videos from the MFC19 dataset created by NIST. Even when trained on a small portion of the entire dataset (10%), the detector is remarkably effective at identifying manipulated videos.



PR curves, F1 score, AUC score, and AP score on the test set for all the trained models using 10% of the available training data (68 videos).

PR curves, F1 score, AUC score, and AP score on the test set for all the trained models using 50% of the available training data (339 videos).

PR curves, F1 score, AUC score, and AP score on the test set for all the trained models using 25% of the available training data (169 videos).

PR curves, F1 score, AUC score, and AP score on the test set for all the trained models using 75% of the available training data (508 videos).

Figure 53. PR Curves for Different Amounts of Training Data

Published Paper:
D. Güera, S. Baireddy, P. Bestagini, S. Tubaro, E. J. Delp, "We Need No Pixels: Video Manipulation Detection Using Stream Descriptors", *International Conference on Machine Learning (ICML), Synthetic Realities: Deep Learning for Detecting AudioVisual Fakes Workshop*, Long Beach, California, June 2019. URL: arXiv:1906.08743

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Camera Model Attribution
Major Technical Approach: A Counter-Forensic Method for CNN-Based Camera Model Identification

Program Objectives: given a method to change the estimated result in a CNN-based camera model detector

A DenseNet model with 40 layers is selected as the CNN camera model detector. It is trained on the patch dataset using untargeted attacks with FGSM and targeted attacks with JSMA. We only perturb images from the test split which were correctly classified by the CNN in their original states.

For untargeted attacks with FGSM, Table 5 reports the error rate and the average confidence score on the test split of the patch dataset for different values of ϵ which have been shown to generate high misclassified adversarial images while not producing appreciable visual changes. We find that using $\epsilon = 0.005$ offers the best compromise between error rate and visual changes in the image.

Table 5. Error Rate and Confidence Score Values of Our Trained DenseNet Model after an Untargeted Attack with FGSM to the Test Split with Different Values of ϵ

ϵ value	Error Rate (%)	Confidence Score (%)
0.001	91.4	97.7
0.002	91.7	97.2
0.003	92.2	96.7
0.004	92.7	95.8
0.005	93.1	95.3
0.006	94.1	95.1
0.007	94.5	94.2
0.008	95.3	93.6
0.009	95.9	93.0
0.01	96.2	92.3

For targeted attacks with JSMA, Table 6 reports the error rate and the average confidence score for each possible camera model target class. JSMA allows us to generate image patches that get misclassified into a specific camera model with high error rates and confidence scores, the modifications that it applies to the images can usually be spotted through visual inspection

Table 6. Error Rate and Confidence Score Values of Our Trained DenseNet Model for each Possible Target Camera Model after Applying a Targeted Attack with JSMA to the Test Split

Target Camera Model	Error Rate (%)	Confidence Score (%)
AS-One	99.5	87.7
ES-D5100	99.3	88.6
MK-Powershot	99.3	88.4
MK-s860	99.7	88.5
PAR-1233	99.7	87.9
PAR-1476	99.4	88.1
PAR-1477	99.5	88.2
PAR-A015	99.6	88.4
PAR-A075	99.3	87.8
PAR-A106	99.2	87.9

Published paper:
D. Güera, Y. Wang, L. Bondi, P. Bestagini, S. Tubaro, E. J. Delp, "A Counter-Forensic Method for CNN-Based Camera Model Identification", *IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW), pp. 1840-1847*, Honolulu, Hawaii, July 2017.
DOI: 10.1109/CVPRW.2017.230

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Camera Model Attribution
Major Technical Approach: Supervised approach with CNN

Program Objectives: Reliability Map Estimation for Camera Model Attribution

Table 7. Results of Our Approach

$\boldsymbol{M_{ip}}$	**Strategy**	**Patches**	**Accuracy**	**Acc. Delta**
$\boldsymbol{M_{ip}^2}$	Scratch	55,475	0.9009	0.0342
	Pre-Trained	618,958	0.9478	0.0811
	Transfer	637,135	**0.9513**	**0.0845**
$\boldsymbol{M_{ip}^3}$	Scratch	518,228	0.9041	0.0374
	Pre-Trained	626,767	0.9520	0.0853
	Transfer	641,808	**0.9556**	**0.0888**
$\boldsymbol{M_{ip}^4}$	Scratch	562,897	0.8963	0.0296
	Pre-Trained	649,515	0.9499	0.0832
	Transfer	647,998	**0.9532**	**0.0865**
$\boldsymbol{M_{ip}^5}$	Scratch	511,425	0.9045	0.0378
	Pre-Trained	648,665	0.9529	0.0862
	Transfer	651,508	**0.9530**	**0.0863**
$\boldsymbol{M_{ip}^6}$	Scratch	517,386	0.9035	0.0367
	Pre-Trained	651,405	0.9501	0.0834
	Transfer	652,308	**0.9531**	**0.0864**

This table summarizes the results of our method. $\boldsymbol{M_{ip}}$ refers to the network architecture. *Strategy* indicates the training scheme used. The *patches* column depicts the total number of patches deemed "reliable" by the network. *Accuracy* indicates the average camera attribution result, where each image's camera model is determined based estimates of the reliable patches. Finally, the *accuracy delta* indicates the accuracy gained by selecting reliable patches to estimate the camera model as opposed to randomly sampling patches to estimate the camera model. In addition to the impact on camera model attribution, our method constructs a reliability mask that visualizes easy and difficult regions of an image for a CNN to classify. Thus, our approach increases interpretability of CNN results and understanding of which image details are more important for camera model attribution.

Published Paper:
D. Güera, S. K. Yarlagadda, P. Bestagini, F. Zhu, S. Tubaro, E. J. Delp, "Reliability Map Estimation for CNN-Based Camera Model Attribution", *IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2018, Lake Tahoe, NV,
DOI: 10.1109/WACV.2018.00111

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Scanner Model Attribution
Major Technical Approach: Convolution Neural Network

Program Objectives: predict the scanner label or generate a reliability map to indicate the manipulated regions in a scanned image.

Figure 54 reports the results in terms of the confusion matrices for the 10-scanner dataset. The overall classification accuracy is 93.72% per patch (i.e. without majority vote) and 96.83% per image (i.e. with majority vote). The high accuracies on patch-level and image-level classification tasks indicate our model is very effective on the 10-scanner dataset. The results for both the 10-scanner dataset and the entire dataset are reported in Table 8. On the 10-scanner dataset, our method achieves the highest classification accuracy on both patches and the images. On the entire dataset, our method achieves the highest patch-level accuracy. Our image-level classification accuracy is very close to the highest, the one which achieved by Xception.

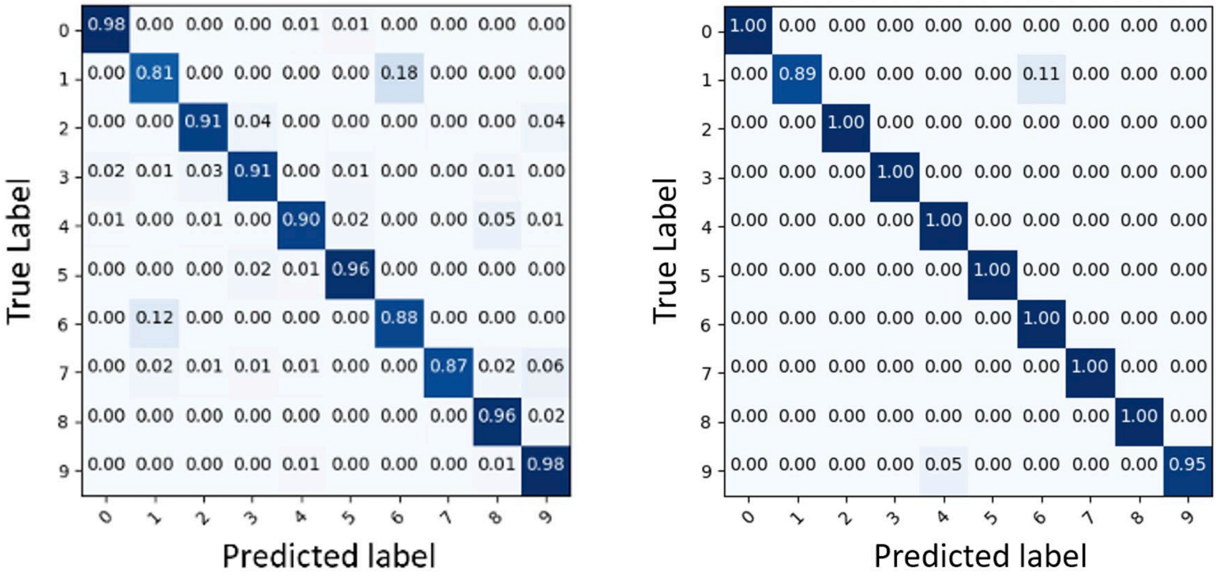


Figure 54. Confusion Predict a Score for Each Input Video That Indicates if the Video Has Been Manipulated or Not (Left: Per Patch; Right: Per Image)

Table 8. The Scanner Model Classification Accuracy: "Per Patch" Rows Indicate the Classification Accuracy on Patches IP; "Per Image" Rows Indicate the Classification Accuracy for Full Size Images I

Network		10 scanners	169 scanners
Ours	per image	96.83%	92.97%
	per patch	93.72%	89.69%
Xception	per image	95.24%	93.24%
	per patch	92.11%	88.85%
Inception3	per image	94.44%	90.37%
	per patch	91.69%	88.62%
Resnet34	per image	96.03%	91.67%
	per patch	91.72%	88.73%

Figure 55 shows an example of the reliability map. In the reliability map, the color of the pixel represents the probability that it is generated by the predicted scanner model. Color "dark red" indicates a probability value equal to 1.0, and color "dark blue" indicates a probability value equal to 0.0. Then we generate 2 manipulated images based on the original image from Figure 55. Figure 56 shows the manipulated images and their corresponding reliability maps with different parameters. The top one is generated by self-image copy-move with translation operations. The bottom one is generated by copy-pasting regions in another image source from different scanner model.

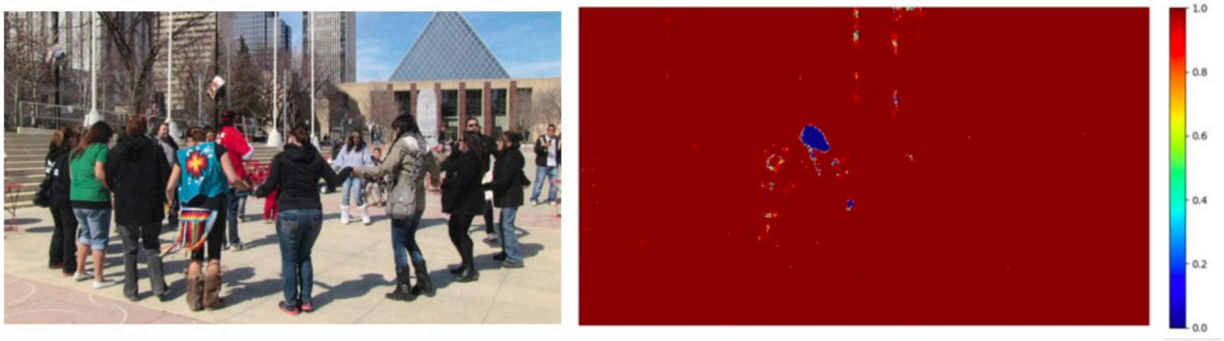


Figure 55. An Original Scanned Image Used for Forged Image Creation and Its Corresponding Reliability Map

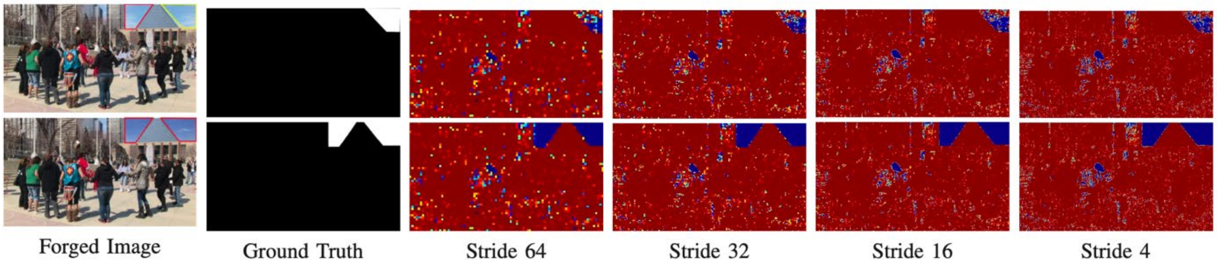


Figure 56. The Forged Scanned Images and Corresponding Reliability Maps with Different Strides

Published paper:
R. Shao, E. J. Delp, "Forensic Scanner Identification Using Machine Learning", *IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI),* Santa Fe, New Mexico, March 2020

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Shadow Removal Detection
Major Technical Approach: GAN-based Detection

Program Objectives: maximize AUC for detection

Figure 57 shows the effectiveness of our approach to shadow removal detection. The left graph shows the detection performance (image classification) while the right graph shows localization performance (accuracy of masks generated). The proposed BCE loss (that replaces the l1 loss used by conditional GANs) improves the detection AUC from 0.751 to 0.788. It has a similar effect for localization, increasing from 0.701 to 0.803. All other approaches traditionally used had an AUC of less than 0.6 for the localization.

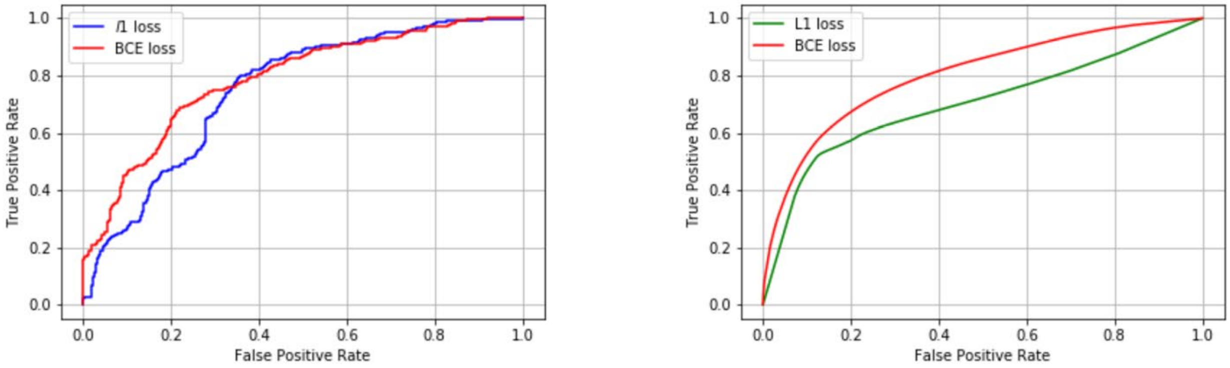


Figure 57 AUC of Standard (l1) vs Proposed (BCE) Loss

Published Paper:
S. K. Yarlagadda, D. Güera, D. Mas Montserrat, F. Zhu, P. Bestagini, S. Tubaro, E. J. Delp, "Shadow Removal Detection and Localization For Forensics Analysis", *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brighton, UK, May 2019*. DOI: 10.1109/ICASSP.2019.8683695

4.1.2 University of Siena

Contract Information:
Team: Purdue Team, Siena Sub-Team
POC: Prof. Mauro Barni, University of Siena, barni@diism.unisi.it

Technical Problem: Detection of multiple JPEG compression in the presence of laundering and counter-forensic attacks
Technical Approach: Adversary-aware, SVM-based detection

Program Objectives: improve the detection performance of double JPEG detection in presence processing operations and CF attacks, on different reference datasets. The Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve is considered as evaluation metrics.

Table 9 shows the performance of the aware detector for QF2 = 85. The resizing, dithering and grid desynchronization attacks are experimentally identified and selected as approximation of the MPA and considered to perform the aware training. Form the table, we see that desynchronization and cropping are the most difficult cases to face with, possibly because they do not introduce specific artifacts in the final manipulated image.

Table 9. Performance (AUC) of the Aware SVM for QF2 = 85. Results Over the RAISE Dataset.

Processing	QF1			
	50:75	**80**	**83**	**90**
D-JPEG only	0.99	0.98	0.95	0.95
Wavelet Denoise	0.90	0.78	0.70	0.74
Median Filtering	0.94	0.93	0.95	0.96
Blurring	0.96	0.96	0.97	0.98
Stamm Dithering	0.95	0.93	0.93	0.93
Adaptive Hist Equalization	0.93	0.95	0.96	0.92
Resize	0.94	0.87	0.86	0.83
Rotation	0.93	0.86	0.84	0.81
Zooming	0.98	0.97	0.97	0.94
Desynchronization	0.75	0.63	0.59	0.49
Cropping (not aligned)	0.83	0.70	0.68	0.57
Seam-Carving	0.90	0.71	0.74	0.50
Mirroring	0.99	0.97	0.94	0.95
Copy-Move	0.97	0.97	0.94	0.94
Hist stretching	0.94	0.87	0.82	0.62

As further results, in the presence of database mismatch between training and testing, the AUC decrease but not significantly so. With regard to the impact of a mismatch between the size/resolution of the images used for training and those used for testing, the performance drop yielding very bad results when the size of the test images is much larger than that of the training images (double size), while the degradation is small when the size of the test images is smaller or larger, but not too much. Finally, the mismatch between the software used to build the images used for training and that used for the test does not have a significant impact on the performance.

[1] Stamm, Matthew C., et al. "Undetectable Image Tampering through JPEG Compression Anti-forensics." Image Processing (ICIP), 2010 17th IEEE International Conference on. IEEE, 2010.

Published Paper:
M. Barni, E. Nowroozi and B. Tondi, "Higher-order, adversary-aware, double JPEG-detection via selected training on attacked samples," EUSIPCO 2017

Contract Information:
Team: Purdue Team, Siena Sub-Team
POC: Prof. Mauro Barni, University of Siena, barni@diism.unisi.it

Technical Problem: Detection and localization of contrast adjustment in the presence of JPEG
Technical Approach: Adversary-aware, SVM and CNN-based detection

Program Objectives: improve the detection performance of contrast adjustment detection in presence of JPEG post-processing. The Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve and the Test Accuracy (Acc) are considered as evaluation metrics.

Development of a contrast adjustment detector based on SVMs

The proposed system provides improved performance in the presence of JPEG compression over a wide range of QFs, while it maintains good performance in the absence of attacks, that is when the AHE is the last step of the manipulation chain. The performance of the system based on the pool of aware SVM classifier when the QF is estimated by means of idempotency are reported in Figure 58. The average error of the QF estimation in terms of L1 distance between real and estimated QFs is reported in

, for the contrast manipulated images (for the pristine images the estimation error is similar).

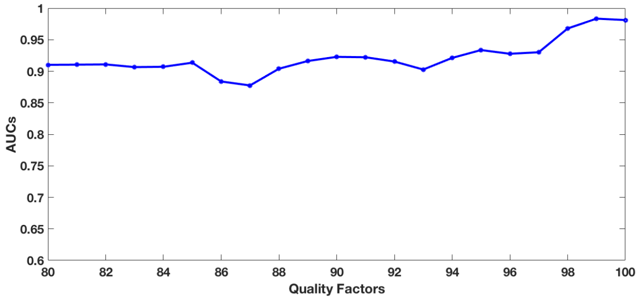


Figure 58. Performance of the Aware SVM-based Detector

Table 10. Average Error of QF Estimation via the Idempotency-Based Approach

QF	**80**	**81**	**82**	**83**
Average Error	0.1017	0.1202	0.0686	0.0721
QF	**84**	**85**	**86**	**87**
Average Error	0.0631	0.1107	0.0325	0.0015
QF	**88**	**89**	**90**	**91**
Average Error	0.0015	0.0010	0.0010	0
QF	**92**	**93**	**94**	**95**
Average Error	0	0.1778	0.1402	0.0451
QF	**96**	**97**	**98**	**99**
Average Error	0.0451	0.0015	0	1
QF	**100**			
Average Error	2			

Figure 59 reports the performance of the system when JPEG compression is carried out with a different software with respect to the one used to generate the training images and to perform the QF estimation: specifically, the test images are compressed with GIMP while the compression of the images used for training and the idempotency-based estimation is implemented in Matlab.

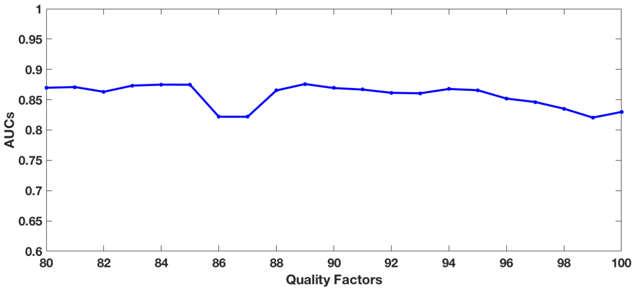


Figure 59. Performance of the Aware Detector Under JPEG Compression Software Mismatch

Published Paper:
M. Barni, E. Nowroozi and B. Tondi, " Detection of Adaptive Histogram Equalization Robust Against JPEG Compression," IAPR/IEEE IWBF 2018

Development of a CNN for the detection and localization of generic contrast adjustment
The average accuracies that can be obtained with the proposed network at the patch level in the range of QFs [80; 100] are: 0.84 for CLAHE, 0.72 for γ Corr and 0.79 for HS. Due to the difficulty of the task, these are remarkable results.
reports the overall performance of the detector on full images in terms of AUC, for both matched and mismatched processing parameters. The results significantly improve those achieved by the approaches from the literature. Good robustness to JPEG compression is achieved (at least for CLAHE and HS) also when the QF is 85 and 80, which are outside the training range (which is limited to the QFs in [90:100]).

Table 11. Performance (AUC) of the Detector Under Matched Processing. (The Matched Parameters Are in Bold)

		no jpeg	100	98	95	90	85	80	75
CLAHE	0.003	100	99.9	99.8	98.9	97.6	97.1	96.8	96
	0.005	100	99.9	99.9	99.4	98.9	98.8	98.5	98
	0.007	100	99.9	100	99.6	99.1	98.9	98.7	98.5
γ Corr	**1.5**	98.8	98.5	94.2	89.2	87	84	81.2	81
	1.7	99.4	98.9	95.7	91.8	90.4	89.7	89.2	88.1
	0.7	99.1	97.1	92.3	87.3	85.6	81	78	69
	0.6	99.7	99.5	97.3	91.6	86.7	83.7	80.1	77.3
HS (%)	3	99.6	98.1	95.8	91.4	87.8	85.7	83.5	83
	5	99.5	98.9	97.6	93.7	92.6	91.5	90.3	89.4
	7	100	99.3	98.3	95.5	94	93.7	93.6	93

Table 12 shows the results under various contrast/brightness adjustment performed with Photoshop (PS). Based on these results, we can argue that the CNN-based detector scales well with respect to the adjustment type maintaining good performance when the tones of the image are adjusted in different ways and, possibly, selectively in different tonal ranges (Curve_S), and when the adjustment operates differently on the color channels (the Auto processing). The AUC is large with respect to all the QFs for some of the processing (AutoTone, Curve_S, HistEq) and, in general, it remains above 90% in most of the cases.

Table 12. Performance (AUC) of the Detector for Different Tonal Adjustments in PS

	no jpeg	100	98	95	90	85	80
HistEq	100	99.9	99.9	99.5	98.3	96.9	94.8
Brightness+	97.5	97.7	95.2	93.6	91.2	87.8	85.6
Contrast+	99.1	100	99.6	97.9	94.7	91.9	87.1
Brightness-	96.7	97.3	93.3	90.1	84.2	78.8	75.6
Contrast-	98.8	99.6	96.4	91.2	87	82	80
Curve_S	99.6	99.8	99.8	99.1	97.7	96	93.6
AutoContrast	95.9	94.7	93	91.9	90.2	89	86.5
AutoColor	98.2	98.6	96.8	95.3	93.7	91.8	89.1
AutoTone	99.5	99.5	99	98.2	97.2	96.1	94.5

[1] M. Goljan, J. Fridrich, and R. Cogranne, “Rich Model for Steganalysis of Color Images”, IEEE WIFS 2014

[2] B. Bayar and M. C. Stamm, “A deep learning approach to universal image manipulation detection using a new convolutional layer,” ACM IH&MMSEC 2016

Published Paper:
M. Barni, A.Costanzo, E. Nowroozi and B. Tondi, " CNN-based detection of generic contrast adjustment with JPEG post-processing," IEEE ICIP 2018

Contract Information:
Team: Purdue Team, Siena Sub-Team
POC: Prof. Mauro Barni, University of Siena, barni@diism.unisi.it

Technical Problem: Splicing detection and localization in JPEG images based on inconsistencies of double JPEG compression
Technical Approach: CNN-based estimation of primary quantization matrix, clustering and morphological reconstruction.

Program Objectives: improving tampering localization based on double JPEG analysis and identifying different sources of tampering (donors). The localization capability is measured in terms of MCC, while the Normalized Mutual Information (NMI) is used to evaluate the quality of the clustering to measure the capability of identifying different sources.

For the experimental validation of the tool, we created ourselves a large-scale tampering dataset consisting of a large number of spliced images, tampered with a different number of tampering sources (at most 3), different tampering sizes, and for several combinations of QF1s. The pristine images, used to build the tampered versions, are taken from the RAISE dataset. Then, the number of clusters is at most 3, i.e., k=4 (the background cluster is included).

The average accuracy that we obtained for the estimation of k is 79%. Somewhat expectedly, the estimation works better for a smaller number of clusters k, since in these cases the minimum difference between the QF1 values considered tends to be smaller on the average (i.e. the QF1 values are closer) when there are multiple tampering regions. The detection accuracy, that is the probability that the estimated value of k is 1 when there is no tampering (k = 1), is 95.5%; and we get a True Positive (TP) rate (correct detection of the tampering) of 96.7% at a False Positive (FP) rate (uncorrect detection of pristine images) of 5%.

Table 13 shows the results obtained for tampering localization are compared to two state of the art methods performing tampering localization of double JPEG images, that is [1] and [2] ([1] provides a separate method for the aligned – Al - and non-aligned - NAl - case). For each method, performance is averaged on the corresponding set of TP images. The proposed method works better in the more challenging non-aligned double JPEG scenario, since the CNN-based Q1 estimator is designed to work particularly for this case (the aligned case is assumed to occur with probability 1/64). The performance in the aligned double JPEG scenario are also good and slightly better than those achieved by the best performing method on the average. It is worth stressing that having a same method that works in a general setting (both in the aligned and non-aligned case, and for every combination of QFs of first and second compression) represents a significant advantage in practice, since such information about the alignment or not of the compression grid is not known a priori.

Table 13. Average Localization Performance (MCC), Averaged on the Set of TP Images

	Mixed	Aligned			Non-Aligned		
		All	QF1 < QF2	QF1 > QF2	All	QF1 < QF2	QF1 > QF2
[1]-Al	0.623	0.623	0.772	0.090	0.000	0.000	0.000
[1]-NAl	0.471	0.000	0.000	0.000	0.471	0.582	0.195
[2]	0.589	0.646	**0.815**	0.024	0.011	0.013	0.009
Our	**0.639**	**0.648**	0.614	**0.685**	**0.631**	**0.588**	**0.676**

With regard to the average clustering performance of our method, we get an average NMI value of 0.475 (0.481 for the aligned case, 0.468 for the aligned case). Notice that, even if the NMI indexes can theoretically reach 1, satisfactory clustering results can be obtained in practice with much lower NMI values (a value of the NMI around 0.5 can be satisfactory, while the NMI is close to zero when either the localization or the clustering result is poor).

[1] T. Bianchi and A. Piva, “Image forgery localization via block-grained analysis of JPEG artifacts,” IEEE T-IFS, 2012.
[2] W. Wang, J. Dong, and T. Tan, “Exploring DCT coefficient quantization effects for local tampering detection,” IEEE T-IFS, 2014.

Published Paper:
Y. Niu, B.Tondi, Y.Zhao, M.Barni “Primary Quantization Matrix Estimation from Double JPEG images via Convolutional Neural Network”, IEEE SPL 2019
Y. Niu, B. Tondi, Y. Zhao, R. Ni, M. Barni, "Image Splicing Detection, Localization and Attribution via JPEG Primary Quantization Matrix Estimation and Clustering," in *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 5397-5412, November 2021. DOI: 10.1109/TIFS.2021.3129654

Contract Information:
Team: Purdue Team, Siena Sub-Team
POC: Prof. Mauro Barni, University of Siena, barni@diism.unisi.it

Technical Problem: Improving the security of machine-learning based manipulation detectors
Technical Approach: Multiple classification, SVM-based

Program Objectives: improve the security of the detector against targeted attacks while maintaining good performance in absence of attacks. The Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve and the Test Accuracy (Acc) are considered to measure the detection performance. The improved robustness against attacks is measures in terms of Mean Square Error (MSE) and percentage (%) of pixels modified by the attack.

Our tests confirm that both the 2C SVM classifier and the 1.5C multiple SVMs classifier have perfect performance in the absence of attacks (the AUC is 100% with the 2C and 99% with the 1.5C). Experiments to assess the robustness of the 1.5C architecture against post-processing were also carried out by considering Gaussian noise addition and JPEG compression. These tests show that, even if the performance of the 1Cs classifiers are significantly impaired by the post-processing, the 1.5C classifier is robust and its performance remain comparable to those of the 2C. The tests are performed on the RAISE dataset.

The tests we carried out under attacks confirm the expectation that, in order to make fail the 1.5C SVM classifier, the attacker has to introduce a much larger distortion into the images, with respect to the 2C case. The gradient-based white-box targeted attack against SVM models developed in [2] is considered for these experiments. Table 14 compares the attack against the 2C and the 1.5C detector in terms of MSE. As shown in the table, the average MSE for the attacked images in the case of attack against the 1.5C SVM classifier is more than double in the case of resizing and contrast enhancement (via adaptive histogram equalization- CLAHE) and almost double in the case of median filtering, with respect to the case of attack to the 2C SVM. The average percentage of pixels modified by the two attacks are reported in
Table ***15***. The table confirms that, in order to be successful against the 1.5C classifier, the attacker has to modify a larger number of pixels.

Table 14 Average MSE of the Attack (on 500 attacked images)

	Resizing	Median Filtering	CL-AHE
Attack against 2C	0.10	0.22	0.27
Attack against 1.5C	0.17	0.60	0.43

Table 15. Average Percentage of Pixels Modified by the Attack (on 500 Attacked Images)

	Resizing	Median Filtering	CL-AHE
Attack against 2C	9.5%	15.1%	12.3%
Attack against 1.5C	13.4%	25.1%	15.1%

[1] B. Biggio, I Corona, et al, “One-and-a-Half-Class Multiple Classifier Systems for Secure Learning Against Evasion Attacks at Test Time”, International Workshop on Multiple Classifier Systems, 2015.
[2] Z. Chen, B. Tondi, X. Li, R. Ni, Y. Zhao, and M. Barni.” A gradient-based pixel-domain attack against SVM detection of global image manipulations.” WIFS 2017.

Published Paper:
M. Barni, E. Nowroozi and B. Tondi, " Improving the security of image manipulation detection through one-and-a-half-class multiple classification", MTAP 2019.

Contract Information:
Team: Purdue Team, Siena Sub-Team
POC: Prof. Mauro Barni, University of Siena, barni@diism.unisi.it

Technical Problem: Improving the security of machine learning-based manipulation detectors
Technical Approach: Features randomization, SVM and CNN-based detection

Program Objectives: improve the security of the detector against targeted attacks while maintaining good performance in absence of attacks. The missed detection error probability (Pd) of the detector with and without attacks is considered as evaluation metric.

Securing data-driven detectors by means of Random Feature Selection
Figure 60 reports the performance of the RFS detector under attacks, for different values of the attack strength (a smaller value of ε corresponding to a larger strength), for the case of AHE. The case of attack carried out in the feature domain (left) and in the pixel domain (right) is considered. From the figures we see that, whereas the targeted attack is able to fool the SVM detector based on the full feature set (missed detection probability 100%), that is, when k = N, when the randomized detector is considered (k < N), the error probability of missed detection decreases. Specifically, for a pretty low number of features, e.g. 1/30 of the initial dimensionality of the feature space, the error probability under attack in the pixel domain case is below 50% for all the attack strengths considered. As an overall trend, reducing the dimension of the feature set does not impact much the performance of the detector (few features are enough to perform correct detection in absence of attacks). For the case of MF results are similar.

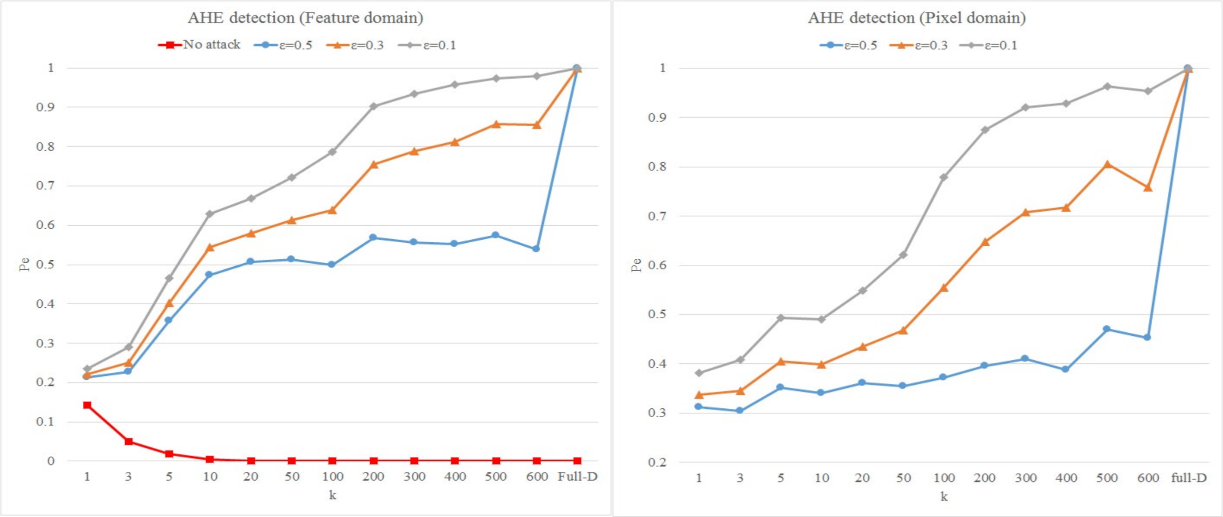


Figure 60. The Performance of the RFS Detector Under Attacks

For the attack carried out against the expected version of the classifier (attack aware of randomized defence), the results we got (by averaging on 50 and 100 classifiers) are comparable to that of the attack carried on the full feature detector, thus confirming the security improvement achieved by the RFS detector, even in the presence of more elaborated attacks.

Published Paper:
Z. Chen, B. Tondi, X. Li, R. Ni, Y. Zhao and M. Barni, "Secure Detection of Image Manipulation by Means of Random Feature Selection," IEEE T-IFS 2019

Effectiveness of Random Deep Feature Selection (RDFS) for securing CNN-based detectors
Table 16 reports the Accuracy (Acc = 1 - Pd) of the RDFS detector under attacks in the case of BayarNet, when the FC detector (same FC layers as the original CNN) is considered to build the detector. By inspecting Table 16 we observe that, when the attack works well against the FC detector with k = N (last line of the table), that is, when the attack targeted to the original model can be successfully transferred to the full features FC detector, the proposed randomization strategy helps and a significant gain in the Accuracy can be achieved, at the expense of a minor Accuracy reduction in the absence of attacks. Specifically, the Accuracy gain is about 20-30% for k= 30 and 30-50% for k = 10, while the accuracy reduction in the absence of attacks is only 2-4%, the exact value depending on the task.

In some cases, however, it happens that the Accuracy is already large also for k = N, i.e., the attack fails against the full feature FC detector, meaning that the attack targeted to the original CNN can not be transferred to the full feature FC detector. In these cases, just re-training the FC network on a different set decreases by itself the attack success rate (confirming that, at least for image forensic applications, adversarial examples tend to be non-transferable, in contrast to what happens in typical pattern recognition applications.

A similar behavior can be observed in Table 17, where the performance of the RDFS detector under attacks in the case of ContrastNet are reported. In this case, for k = N the attack is even less effective than before (being then less transferable). When an SVM is considered for the classification, expectedly, the mismatch in the architecture decreases further the success rate of the attack against the full feature SVM detector (case with k = N), i.e. it increases the Accuracy, without even resorting to randomization.

Table 16. Accuracy (%) of the RDFS Detector Based on FC Network for the Case of BayarNet

	Resize				Median Filtering				CL-AHE			
K	No Att.	PGD	FGSM	BFGS	No Att.	PGD	FGSM	BFGS	No Att.	PGD	FGSM	BFGS
5	91.0	69.9	61.6	65.6	88.7	79.8	51.0	73.0	73.0	87.4	89.2	88.0
10	95.0	68.0	55.7	62.0	93.2	80.6	44.5	67.1	78.0	88.0	89.1	78.6
30	97.0	58.5	43.4	48.8	96.8	79.7	30.8	56.1	80.1	89.5	90.7	64.7
50	97.4	52.0	35.9	40.1	97.7	80.0	24.6	53.5	80.7	90.2	91.3	56.3
200	97.8	31.0	13.7	17.4	98.7	77.6	10.8	44.8	81.5	91.6	94.0	42.8
400	97.7	20.7	7.2	9.1	98.8	76.6	7.5	42.6	81.3	91.8	94.5	41.0
600	97.9	16.4	5.4	7.1	98.9	80.5	6.0	43.0	80.5	91.5	93.8	36.6
N	98.0	31.5	0.6	20.3	99.0	81.9	4.3	39.7	80.5	91.8	93.9	35.1

Table 17. Accuracy (%) of the RDFS Detector Based on FC Network for the Case of ContrastNet

	Resize				Median Filtering				CL-AHE			
K	No Attk	**PGD**	**FGSM**	**BFGS**	No Attk	**PGD**	**FGSM**	**BFGS**	No Attk	**PGD**	**FGSM**	**BFGS**
5	74.4	67.7	58.7	60.7	97.2	83.3	48.3	77.1	87.4	47.2	63.7	47.0
10	78.6	71.9	59.9	63.0	98.8	86.1	44.3	79.2	91.1	55.6	68.8	48.3
30	92.7	81.8	65.5	70.7	99.4	88.5	30.0	79.6	94.3	56.7	76.3	39.8
50	96.8	85.2	66.8	73.0	99.6	87.4	21.9	76.6	95.1	50.6	80.0	35.3
200	99.7	88.0	69.6	77.9	99.6	88.6	17.0	76.2	96.9	48.5	83.0	26.0
400	99.8	89.3	71.8	80.0	99.6	88.1	15.6	75.6	97.1	30.1	83.6	21.0
N	100	89.8	75.2	81.2	99.7	85.2	13.7	71.3	98.2	33.5	34.0	26.2

[1] Z. Chen, B. Tondi, X. Li, R. Ni, Y. Zhao, and M. Barni.” A gradient-based pixel-domain attack against SVM detection of global image manipulations.” WIFS 2017

[2] B. Bayar and M. Stamm, “A deep learning approach to universal image manipulation detection using a new convolutional layer,” in ACM IH&MMSEC 2016.

[3] M. Barni, A. Costanzo, E. Nowroozi, and B. Tondi, “CNN-based detection of generic contrast adjustment with JPEG postprocessing,” ICIP 2018

Published Paper:
M. Barni, E. Nowroozi, B. Tondi and B. Zhang, "Effectiveness of Random Deep Feature Selection for Securing Image Manipulation Detectors Against Adversarial Examples," IEEE ICASSP 2020

4.1.3 University of Southern California

Contract Information:
Team: Purdue Team, USC Sub-Team
POC: Prof. C.-C. Jay Kuo, University of Southern California, cckuo@sipi.usc.edu, 213-740-4658

Major Technical Problem: Image Splicing Localization
Major Technical Approach: Contrastive PCA++

Program Objective: propose a new data visualization and clustering technique for discovering discriminative structures in high-dimensional data. Apply proposed dimensionality reduction method on image splicing localization task.

Figure 61 shows the visualization of dimension-reduced data for synthetic dataset using different dimensionality reduction method. The top row of plots show the performance of the cPCA algorithm for different positive values of the contrast parameter _. Clearly, a contrast factor for $\alpha = 2.7$ is ideal, but must be found by sweeping α. The bottom-left plot shows the performance of traditional PCA (which, as expected, fails to separate the classes). The bottom-center plot shows the performance of the t-SNE algorithm, which again fails to discover the underlying structure in the high-dimensional data. Finally, the bottom right figure shows the output obtained by the cPCA++ method, which obtains the ideal clustering without a parameter sweep.

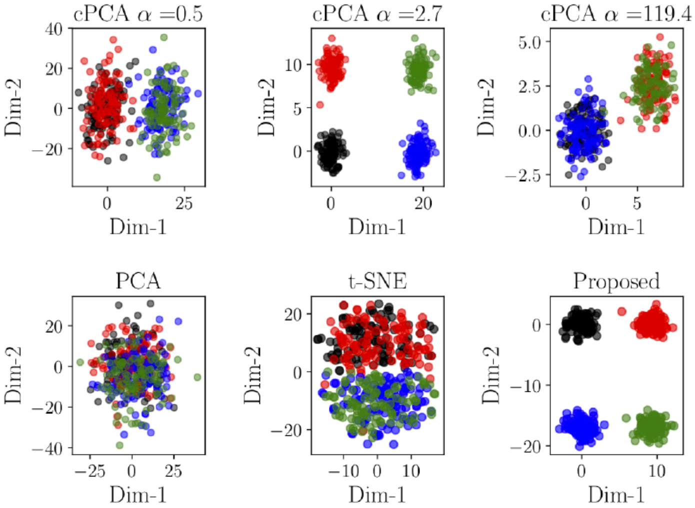


Figure 61. Performance on a Synthetic Dataset, where Different Colors are Used for the Four Different Classes

Table 18. Time Required for the Different Algorithms to Perform the Required Dimensionality Reduction for the Various Datasets Studied in Sec. III-B. All Times Listed in the Table Are in Seconds. Boldface is Used to Indicate Shortest Runtimes and Average cPCA++

Example	cPCA	t-SNE	cPCA++
Synthetic	0.062	6.62	**0.0017**
MNIST over Grass	26	811	**1.00**
Mice Protein Expression	0.13	2.23	**0.0033**
MHealth Measurements	0.091	1388	**0.00065**
RNA-Seq of Leukemia Patient	11.70	2155	**0.86**
Average cPCA++ Speedup	**51x**	**428654x**	**1x**

Table 19. Edge-based F1 Scores (Left) and MCC Scores (Right) on Columbia and Nimble WEB Datasets. Boldface is Used to Emphasize Best Performance

Dataset	cPCA++	MFCN
Columbia	**0.359**	0.312
Nimble WEB	**0.376**	0.273

Dataset	cPCA++	MFCN
Columbia	**0.385**	0.329
Nimble WEB	**0.388**	0.297

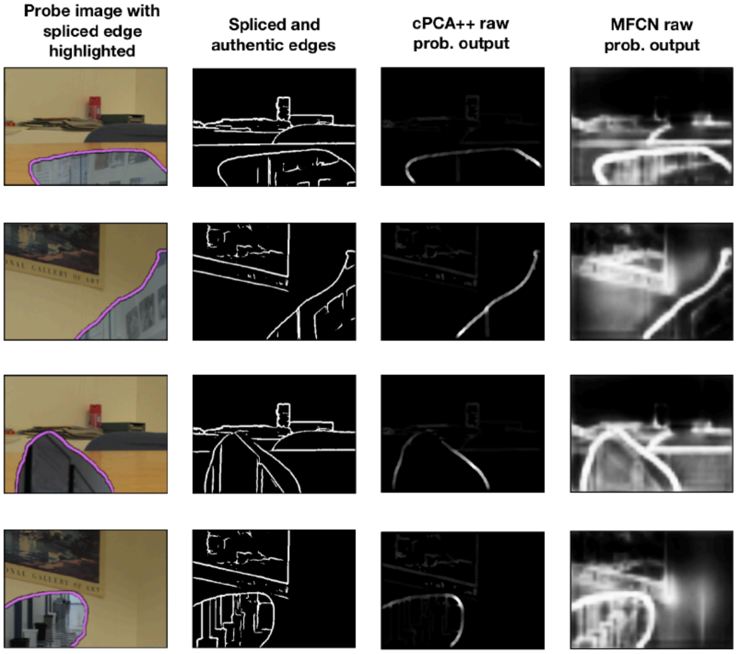

Probe image with spliced edge highlighted
Spliced and authentic edges
cPCA++ raw prob. output
MFCN raw prob. output

Figure 62. Localization Output Examples from Columbia Dataset

Published paper:
Salloum, R. and Kuo, C.C.J., "Efficient image splicing localization via contrastive feature extraction," 2019. arXiv:1901.07172

Contract Information:
Team: Purdue Team, USC Sub-Team
POC: Prof. C.-C. Jay Kuo, University of Southern California, cckuo@sipi.usc.edu, 213-740-4658

Major Technical Problem: Image Splicing Localization
Major Technical Approach: Interpretable Convolutional Neural Network

Program Objective: propose an interpretable CNN design based on the Feedforward methodology. Understand CNN in both theoretical and experimental aspect. Achieve comparable or even better performance than traditional CNN using interpretable CNN design.

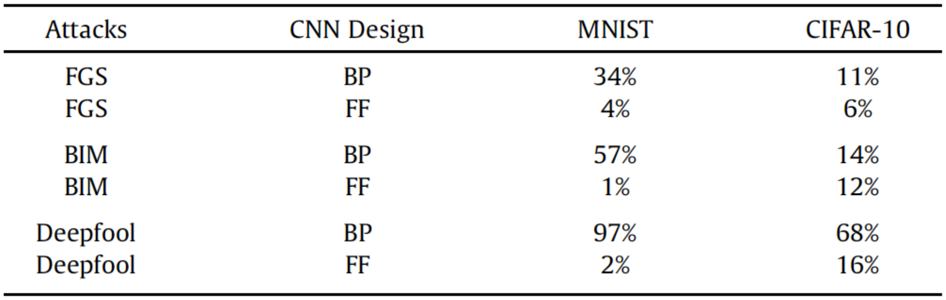
Table 20. Comparison of Testing Accuracies of BP and FF Designs for the MNIST and the CIFAR-10 Dataset

Datasets	MNIST	CIFAR-10
FF	97.2%	62%
Hybrid	98.4%	64%
BP	99.1%	68%

Table 21. Comparison of Testing Accuracies of BP and FF Designs Against FGS, BIM and Deepfool Three Adversarial Attacks Targeting at the BP Design

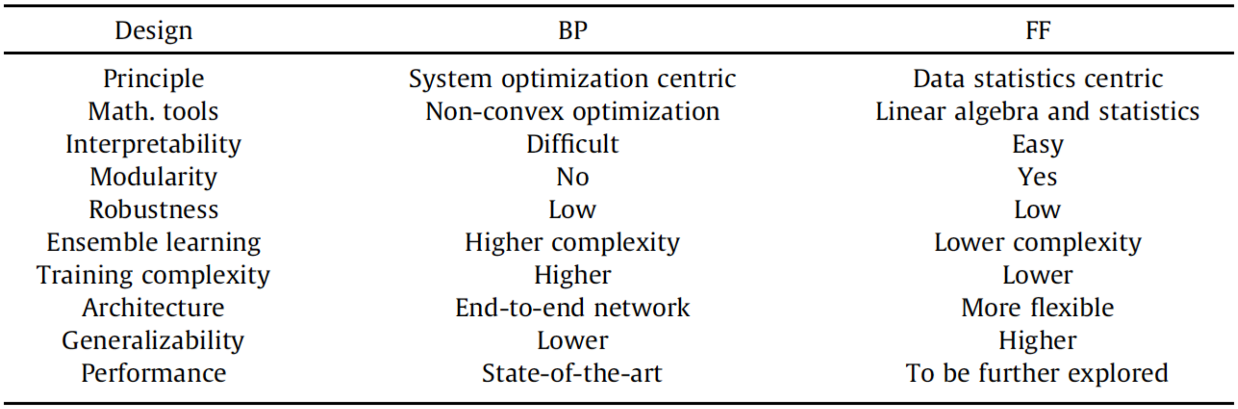
Attacks	CNN Design	MNIST	CIFAR-10
FGS	BP	6%	15%
FGS	FF	56%	21%
BIM	BP	1%	12%
BIM	FF	46%	31%
Deepfool	BP	2%	15%
Deepfool	FF	96%	59%

Table 22. Comparison of Testing Accuracies of BP and FF Designs Against FGS, BIM and Deepfool Three Adversarial Attacks Targeting at the FF Design



Attacks	CNN Design	MNIST	CIFAR-10
FGS	BP	34%	11%
FGS	FF	4%	6%
BIM	BP	57%	14%
BIM	FF	1%	12%
Deepfool	BP	97%	68%
Deepfool	FF	2%	16%

Table 23. Property Comparison of BP and FF Designs



Design	BP	FF
Principle	System optimization centric	Data statistics centric
Math. tools	Non-convex optimization	Linear algebra and statistics
Interpretability	Difficult	Easy
Modularity	No	Yes
Robustness	Low	Low
Ensemble learning	Higher complexity	Lower complexity
Training complexity	Higher	Lower
Architecture	End-to-end network	More flexible
Generalizability	Lower	Higher
Performance	State-of-the-art	To be further explored

Published paper:
Kuo, C.C.J., Zhang, M., Li, S., Duan, J. and Chen, Y., “Interpretable convolutional neural networks via feedforward design,” *Journal of Visual Communication and Image Representation*, **60**, Apr 2019, pp.346-359. DOI: 10.1016/J.JVCIR.2019.03.010

Contract Information:
Team: Purdue Team, USC Sub-Team
POC: Prof. C.-C. Jay Kuo, University of Southern California, cckuo@sipi.usc.edu, 213-740-4658

Major Technical Problem: Image Splicing Localization
Major Technical Approach: Multi-task Fully Convolutional Network (MFCN)

Program Objective: propose a convolutional neural network method for image splicing localization problem, which can perform better than other state-of-the-art algorithms. Discuss the robustness of proposed method on compressed, blurred, noise added images.

Table 24 presents average F1 scores of proposed and existing methods for various datasets. For each dataset, we highlight in bold the top-performing method. As noted in [1], ADQ2, ADQ3, and NADQ require JPEG images as input because they exploit certain JPEG data directly extracted from the compressed files. Therefore, these three algorithms could only be evaluated on the CASIA v1.0 and Nimble 2016 Science datasets, which contain images in JPEG format. For the Columbia and Carvalho datasets (which do not contain images in JPEG format), we put "NA" in the corresponding entries in the table to indicate that these three algorithms could not be evaluated on these two datasets.

Table 24. Average F1 Score of Proposed and Existing Methods for Various Datasets

Method	CASIA v1.0	Columbia	Nimble 2016 SCI	Carvalho
SFCN	0.4770	0.5820	0.4220	0.4411
MFCN	0.5182	0.6040	0.4222	0.4678
Edge-enhanced MFCN	**0.5410**	**0.6117**	**0.5707**	**0.4795**
NOI1	0.2633	0.5740	0.2850	0.3430
DCT	0.3005	0.5199	0.2756	0.3066
CFA2	0.2125	0.5031	0.1587	0.3124
NOI4	0.1761	0.4476	0.1635	0.2693
BLK	0.2312	0.5234	0.3019	0.3069
ELA	0.2136	0.4699	0.2358	0.2756
ADQ1	0.2053	0.4975	0.2202	0.2943
CFA1	0.2073	0.4667	0.1743	0.2932
NOI2	0.2302	0.5318	0.2320	0.3155
ADQ2	0.3359	NA	0.3433	NA
ADQ3	0.2192	NA	0.2622	NA
NADQ	0.1763	NA	0.2524	NA

Table 25 presents Average MCC scores of proposed and existing methods for various datasets. For each dataset, we highlight in bold the top-performing method. As noted in [1], ADQ2, ADQ3, and NADQ require JPEG images as input because they exploit certain JPEG data directly extracted from the compressed files. Therefore, these three algorithms could only be evaluated on the CASIA v1.0 and Nimble 2016 Science datasets, which contain images in JPEG format. For the Columbia and Carvalho datasets (which do not contain images in JPEG format), we put "NA" in the corresponding entries in the table to indicate that these three algorithms could not be evaluated on these two datasets.

Table 25. Average MCC Score of Proposed and Existing Methods for Various Datasets

Method	CASIA v1.0	Columbia	Nimble 2016 SCI	Carvalho
SFCN	0.4531	0.4201	0.4202	0.3676
MFCN	0.4935	0.4645	0.4204	0.3901
Edge-enhanced MFCN	**0.5201**	**0.4792**	**0.5703**	**0.4074**
NOI1	0.2322	0.4112	0.2808	0.2454
DCT	0.2516	0.3256	0.2600	0.1892
CFA2	0.1615	0.3278	0.1235	0.1976
NOI4	0.0891	0.2076	0.1014	0.1080
BLK	0.1769	0.3278	0.2657	0.1768
ELA	0.1337	0.2317	0.1983	0.1111
ADQ1	0.1262	0.2710	0.1880	0.1493
CFA1	0.1521	0.2281	0.1408	0.1614
NOI2	0.1715	0.3473	0.2066	0.1919
ADQ2	0.3000	NA	0.3210	NA
ADQ3	0.1732	NA	0.2512	NA
NADQ	0.0987	NA	0.2310	NA

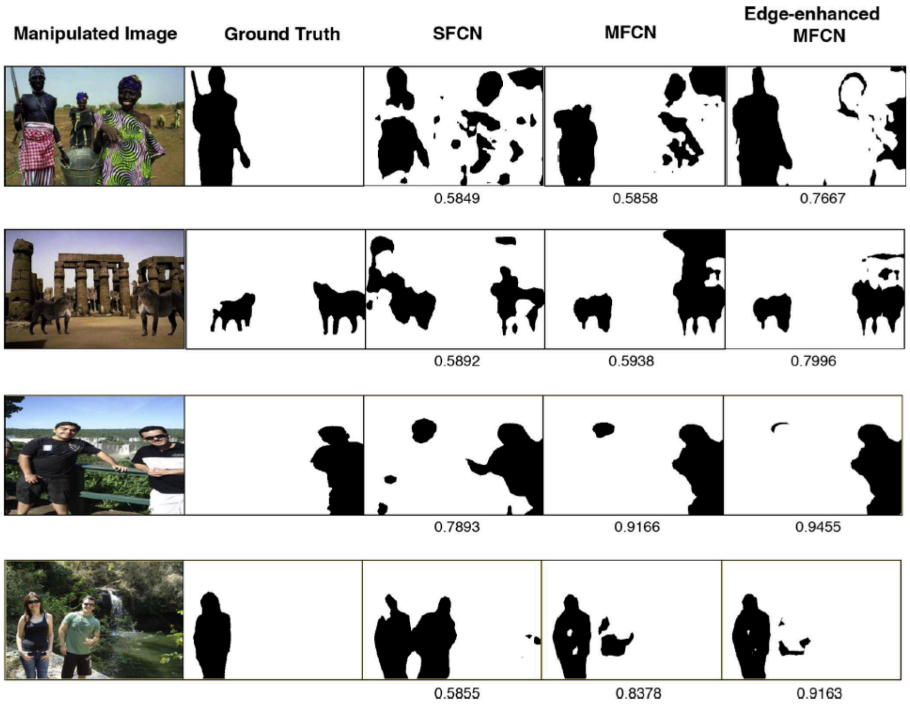


Figure 63. System Output Mask Examples of SFCN, MFCN, and Edge-enhanced MFCN on the CASIA v1.0 and Carvalho Datasets. We Refer to the MFCN without Edge-enhanced Inference Simply as MFCN. Each Row in the Figure Shows a Manipulated or Spliced Image, the Ground Truth Mask, the SFCN Output, the MFCN Output, and The Edge-enhanced MFCN Output. The Number Below Each Output Example is the Corresponding F1 Score

[1] Zampoglou, M., Papadopoulos, S., Kompatsiaris, Y., "Large-scale evaluation of splicing localization algorithms for web images," *Multimedia Tools and Applications*, **76**, 4, Feb 2017, pp. 4801-4834.

Published paper:
Salloum, R., Ren, Y. and Kuo, C.C.J., "Image splicing localization using a multi-task fully convolutional network (MFCN)," *Journal of Visual Communication and Image Representation*, **51**, Feb 2018, pp.201-209. DOI: 10.1016/J.JVCIR.2018.01.010

Contract Information:
Team: Purdue Team, USC Sub-Team
POC: Prof. C.-C. Jay Kuo, University of Southern California, cckuo@sipi.usc.edu, 213-740-4658

Major Technical Problem: Image Splicing Localization
Major Technical Approach: Theoretical Understanding of Convolutional Neural Networks

Program Objective: interpret a CNN as a network that implements the guided multilayer RECOS transform, provide a full explanation to the operating principle of CNNs and discuss how guidance is provided by labels through backpropagation (BP) in the training.

Figure 64 presents the correct classification rate when we train the network until its performance converges for a fixed number of training samples. The two points along the y-axis indicate the correct classification rates without any labeled training sample. The rates are around 32 and 14% for the k-means and random initializations, respectively. Note that the 14% is slightly better than the random guess on the outcome, which is 10%. Then, both performance curves increase as the number of labeled training samples grows. The k-means can reach a correct classification rate of 90% when the number of labeled training samples is around 250, which is only 0.41% of the entire MNIST training data set (i.e., 60,000 samples). This shows the power of the LeNet-5 even under extremely low supervision.

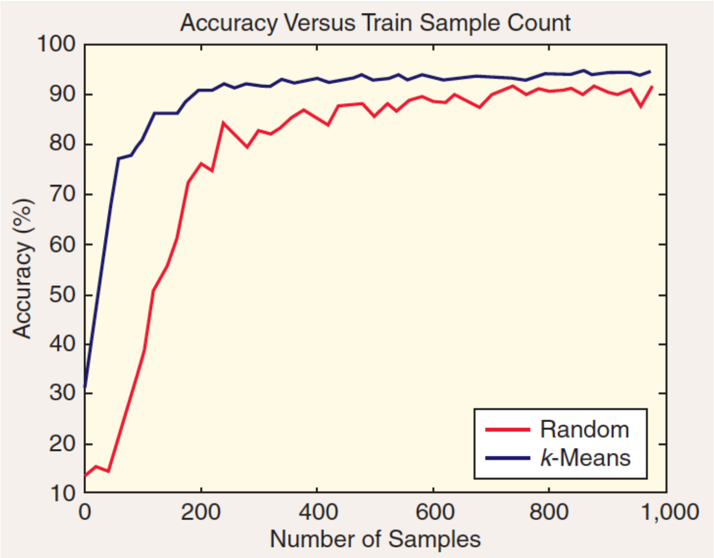
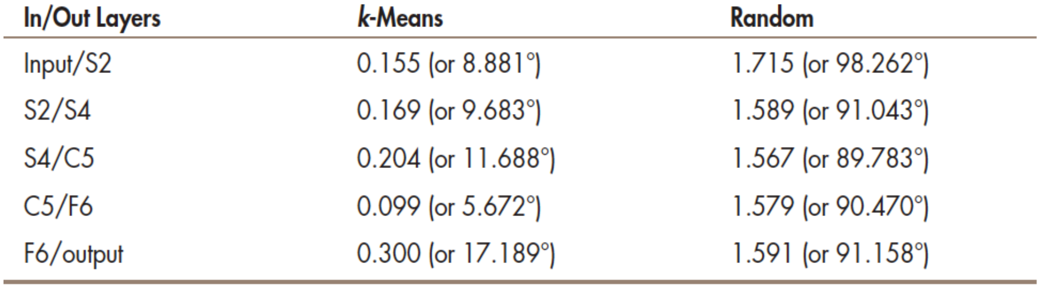


Figure 64. The Comparison of MNIST Weakly Supervised Classification Results of the LeNet-5 Architecture with the Random and K-Means Initializations, where the Correct Classification Rate is Plotted as a Function of Training Sample Numbers

Table 26 presents the averaged orientation changes of anchor vectors in terms of radian (or degree) for the two initialization cases. They are obtained after the convergence of the network with all 60,000 MNIST training samples. This orientation change is the result due to label guidance through the BP. It is clear from the table that a good network initialization (corresponding to unsupervised learning) leads to a faster convergence rate in supervised learning.

Table 26. The Averaged Orientation Changes of Anchor Vectors in Terms of the Radian (or Degree) for the K-Means and the Random Initialization Schemes



In/Out Layers	*k*-Means	Random
Input/S2	0.155 (or 8.881°)	1.715 (or 98.262°)
S2/S4	0.169 (or 9.683°)	1.589 (or 91.043°)
S4/C5	0.204 (or 11.688°)	1.567 (or 89.783°)
C5/F6	0.099 (or 5.672°)	1.579 (or 90.470°)
F6/output	0.300 (or 17.189°)	1.591 (or 91.158°)

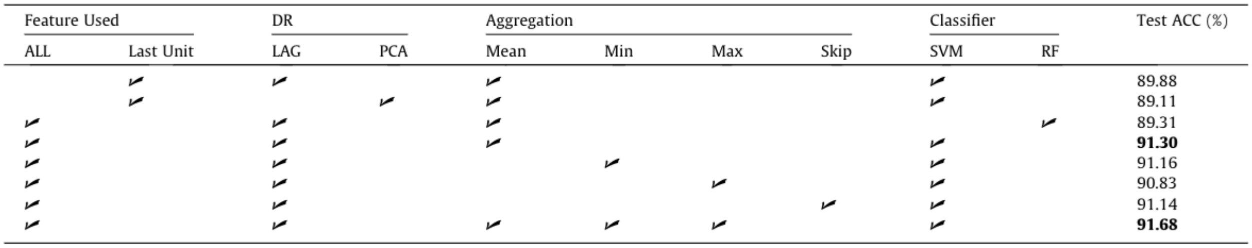
Published paper:
Kuo, C.C.J., “The CNN as a guided multi-layer RECOS transform,” *IEEE Signal Processing Magazine*, **34**, 3, May 2017, pp. 81-89. DOI: 10.1109/MSP.2017.2671158

Contract Information:
Team: Purdue Team, USC Sub-Team
POC: Prof. C.-C. Jay Kuo, University of Southern California, cckuo@sipi.usc.edu, 213-740-4658

Major Technical Problem: Image Splicing Localization
Major Technical Approach: PixelHop: a successive subspace learning method

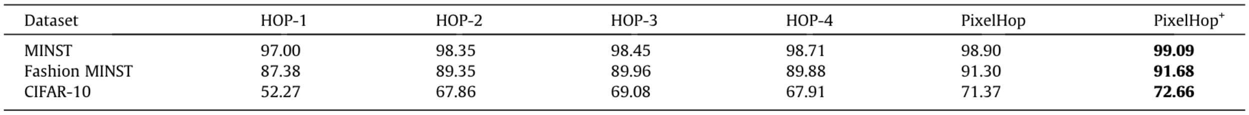
Program Objective: introduce a new machine learning methodology, successive subspace learning, use PixelHop method to illustrate the SSL idea. Prove that PixelHop method can achieve better performance in object classification on three benchmarking datasets.

Table 27. Ablation Study for Fashion MNIST, where the Fourth and the Eighth Rows Are the Settings Adopted by PixelHop and PixelHop+, Respectively



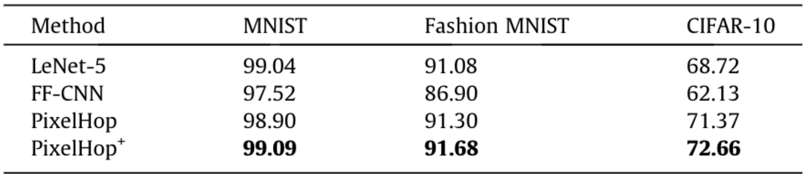
Feature Used		DR		Aggregation				Classifier		Test ACC (%)
ALL	Last Unit	LAG	PCA	Mean	Min	Max	Skip	SVM	RF	
	✓	✓		✓				✓		89.88
	✓		✓	✓				✓		89.11
✓		✓		✓					✓	89.31
✓		✓		✓				✓		**91.30**
✓		✓			✓			✓		91.16
✓		✓				✓		✓		90.83
✓		✓					✓	✓		91.14
✓		✓		✓	✓	✓		✓		**91.68**

Table 28. Comparison of the Classification Accuracy (%) Using Features from an Individual PixelHop Unit, PixelHop and PixelHop+ for Fashion MNIST



Dataset	HOP-1	HOP-2	HOP-3	HOP-4	PixelHop	PixelHop$^+$
MINST	97.00	98.35	98.45	98.71	98.90	**99.09**
Fashion MINST	87.38	89.35	89.96	89.88	91.30	**91.68**
CIFAR-10	52.27	67.86	69.08	67.91	71.37	**72.66**

Table 29. Comparison of Testing Accuracy (%) of LeNet-5, Feedforward-designed CNN (FF-CNN), PixelHop and PixelHop+ for MNIST, Fashion MNIST and CIFAR-1



Method	MNIST	Fashion MNIST	CIFAR-10
LeNet-5	99.04	91.08	68.72
FF-CNN	97.52	86.90	62.13
PixelHop	98.90	91.30	71.37
PixelHop$^+$	**99.09**	**91.68**	**72.66**

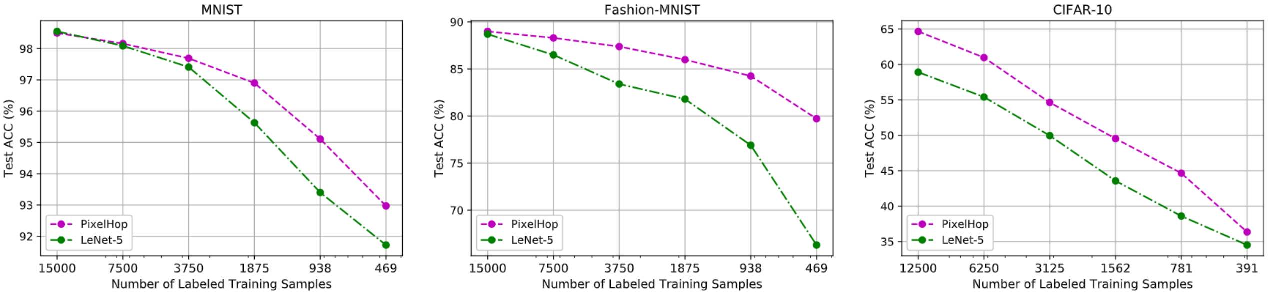


Figure 65. Comparison of Testing Accuracy (%) of LeNet-5 and PixelHop with Different Training Sample Numbers for MNIST, Fashion MNIST and CIFAR-10

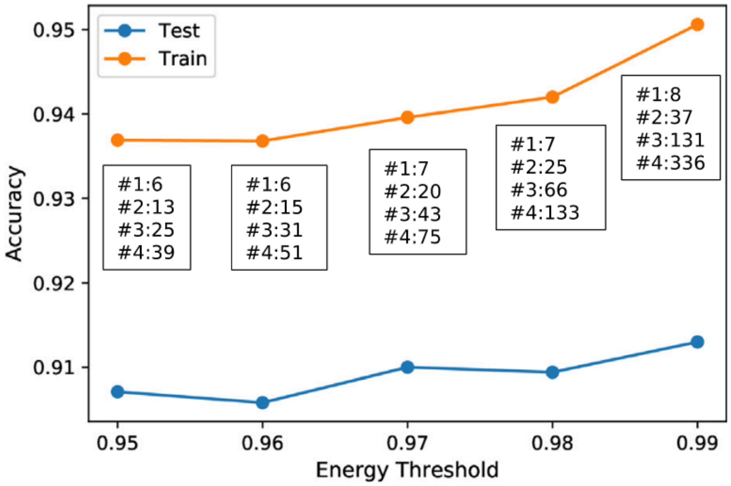
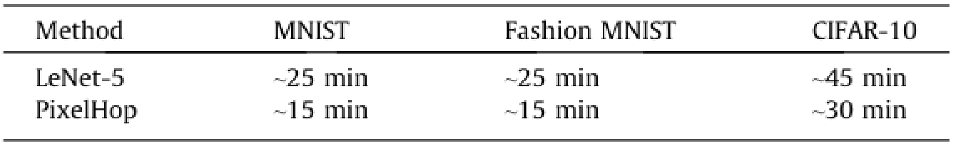


Figure 66. The Classification Accuracy as a Function of the Total Energy Preserved by ACfilters Tested on Fashion MNIST

Table 30. Comparison of Training Time of the LeNet-5 and the PixelHop Method on the MNIST, the Fashion MNIST and the CIFAR-10 Datasets



Method	MNIST	Fashion MNIST	CIFAR-10
LeNet-5	~25 min	~25 min	~45 min
PixelHop	~15 min	~15 min	~30 min

Published paper:
Chen, Y. and Kuo, C.C.J., “PixelHop: A Successive Subspace Learning (SSL) Method for Object Classification,” *Journal of Visual Communication and Image Representation*, **70**, Jul 2020, pp. 102749. DOI: 10.1016/J.JVCIR.2019.102749

Contract Information:
Team: Purdue Team, USC Sub-Team
POC: Prof. C.-C. Jay Kuo, University of Southern California, cckuo@sipi.usc.edu, 213-740-4658

Major Technical Problem: Image Splicing Localization
Major Technical Approach: Subspace approximation with augmented kernels (Saak) transform

Program Objective: propose a data driven Saak transform for feature extraction. Neither data labels nor backpropagation is used. Further understand convolutional neural networks in theoretical aspect.

We use 60,000 image samples (namely, 6000 training images for each digit) to compute the F-test score of all Saak coefficients. Then, we test with the following three settings.
• Setting No. 1: Selecting the leading 2000 last-stage Saak coefficients (without considering their F-test scores).
• Setting No. 2: Selecting 2000 last-stage Saak coefficients with the highest F-test scores.
• Setting No. 3: Selecting 2000 Saak coefficients with the highest F-test scores from all stages.

Furthermore, we conduct the PCA on these 2000 selected Saak coefficients (called the raw feature vector) to reduce the feature dimension from 2000 to 64, 128 or 256 (called the reduced-dimension feature vector). Finally, we apply the SVM classifier to the reduced dimension feature vectors of each test image.

Table 31. The Classification Accuracy Using the SVM Classifier, where Each Column Indicates the Reduced Feature Dimension From 2000 Saak Coefficients. Each Row Indicates a Particular Setting in Selecting the Saak Coefficients

	64	128	256
Setting No. 1	97.22	97.08	96.82
Setting No. 2	97.23	97.10	96.88
Setting No. 3	98.46	98.50	98.31

Table 32. The Classification Accuracy Using the SVM Classifier Under Setting No. 3, where Each Column Indicates the Raw Feature Dimension and Each Row Indicates the Reduced Feature Dimension

	1000	2000	3000	4000	5000
64	98.49	98.46	98.42	98.43	98.41
128	98.46	98.50	98.52	98.51	98.52
256	98.20	98.31	98.27	98.29	98.25

Table 33. The Classification Accuracy Using the KNN Classifier Under Setting No. 3, where Each Column Indicates the Raw Feature Dimension and Each Row Indicates the Reduced Feature Dimension

	1000	2000	3000	4000	5000
64	97.53	97.52	97.50	97.52	97.45
128	97.49	97.45	97.46	97.37	97.39
256	97.50	97.42	97.39	97.41	97.39

Published paper:
Kuo, C.C.J. and Chen, Y., “On data-driven saak transform,” *Journal of Visual Communication and Image Representation*, **50**, 1, Jan 2018, pp. 237-246. DOI: 10.1016/J.JVCIR.2017.11.023

Contract Information:
Team: Purdue Team, USC Sub-Team
POC: Prof. C.-C. Jay Kuo, University of Southern California, cckuo@sipi.usc.edu, 213-740-4658

Major Technical Problem: Image Splicing Localization
Major Technical Approach: Theoretical Understanding of Convolutional Neural Networks

Program Objectives: address two fundamental questions about the structure of the convolutional neural networks (CNN): (1) why a nonlinear activation function is essential at the filter output of all intermediate layers? (2) what is the advantage of the two-layer cascade system over the one-layer system? Two questions are answered by a mathematical model called the “REctified-COrrelations on a Sphere” (RECOS).

The correlation between two anchor vectors can be viewed as a projection from an anchor vector to the input (and vice versa). For positive correlations, the geodesic distance is a monotonically decreasing function of the projection value. The larger the correlation, the shorter the distance.

Table 34 shows the experimental result conducted on MNIST dataset. Negative correlated images and filters gives poor accuracy. It proves that nonlinear activation is essential at the filter output of all intermediate layers.

Table 34. The Experimental Result Conducted on MINST Dataset

Type of Correlation between image and filter	Accuracy
Positive correlated	98.94%
Negative correlated	37.36%

Figure 67 presents the visualization of output images in 6 spectral channels from 1st layer and 16 spectral channels from 2nd layer of LeNet-5. Top 2 rows show the MNIST dataset with 10 different background scenes. The structured background has an impact on the 6 channel responses at the first layer yet their impact on the 16 channel responses at the second layer diminishes. This shed light to the advantage of two-layer cascade system over single-layer system.

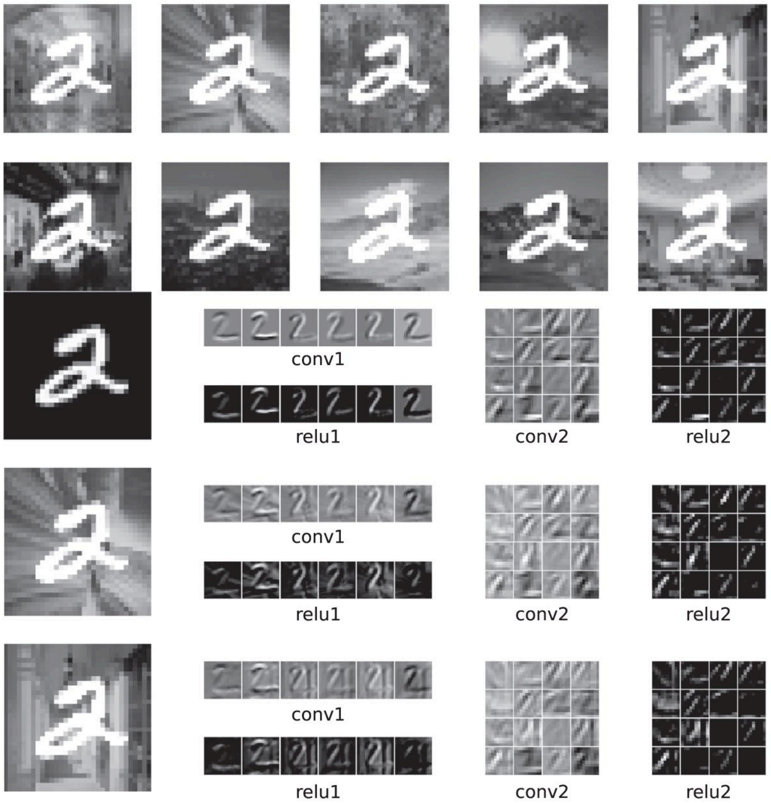


Figure 67. The MNIST Dataset with 10 Different Background Scenes (Top 2 Rows). Filter Response of 1st and 2nd Layer of LeNet-5

Published paper:
Kuo, C.C.J., "Understanding convolutional neural networks with a mathematical model," *Journal of Visual Communication and Image Representation*, **41**, 1, Nov 2016, pp. 406-413.
DOI: 10.1016/J.JVCIR.2016.11.003

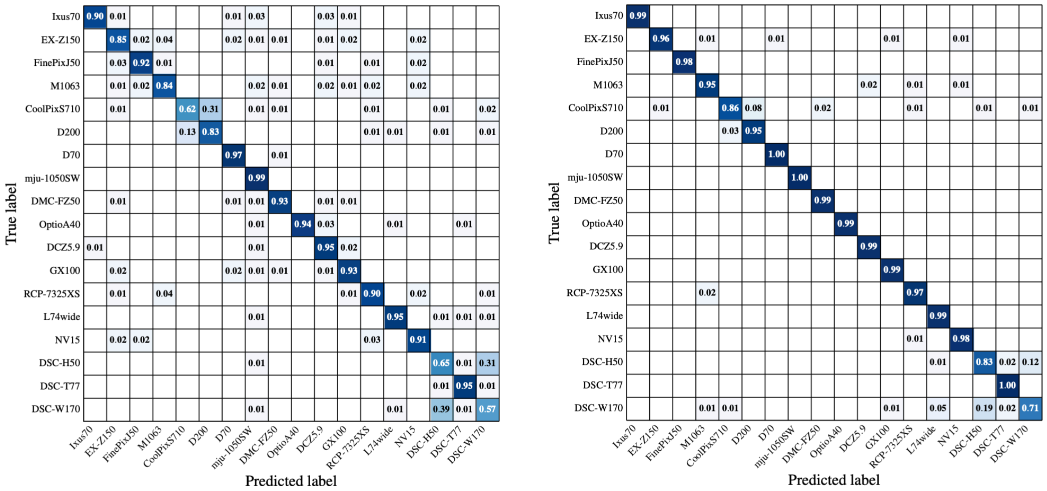
4.1.4 Politecnico di Milano

Contract Information:
Team: Purdue Team, PoliMi Sub-Team
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: Camera model forensics
Major Technical Approach: CNN-based solutions for camera model attribution

Program Objectives: develop a camera model identification system that in closed-set and open-set, enable forgery localization and detection through camera model traces

For camera model identification in close-set [1, 2], we considered the Dresden image database, which consists of 73 devices belonging to 25 different camera models. Images have been clustered depending on the depicted scenery and split into 64x64 pixel patches. Figure 68 shows the results obtained using [1, 2] (left) and the improvement obtained using the strategy in [3] in terms of confusion matrix.



	Ixus70	EX-Z150	FinePixJ50	M1063	CoolPixS710	D200	D70	mju-1050SW	DMC-FZ50	OptioA40	DCZ5.9	GX100	RCP-7325XS	L74wide	NV15	DSC-H50	DSC-T77	DSC-W170
Ixus70	0.90	0.01					0.01	0.03			0.03	0.01						
EX-Z150		0.85	0.02	0.04			0.02	0.01	0.01		0.01	0.02			0.02			
FinePixJ50		0.03	0.92	0.01							0.01		0.01		0.02			
M1063		0.01	0.02	0.84				0.02	0.01		0.02	0.01	0.02		0.02			
CoolPixS710		0.01			0.62	0.31		0.01	0.01				0.01			0.01		0.02
D200					0.13	0.83							0.01	0.01		0.01		0.01
D70							0.97		0.01									
mju-1050SW								0.99										
DMC-FZ50		0.01					0.01	0.01	0.93		0.01	0.01						
OptioA40								0.01		0.94	0.03			0.01			0.01	
DCZ5.9	0.01							0.01			0.95	0.02						
GX100		0.02					0.02	0.01	0.01		0.01	0.93						
RCP-7325XS		0.01		0.04								0.01	0.90		0.02			0.01
L74wide								0.01						0.95		0.01	0.01	0.01
NV15		0.02	0.02										0.03		0.91			
DSC-H50								0.01								0.65	0.01	0.31
DSC-T77																0.01	0.95	0.01
DSC-W170								0.01						0.01		0.39	0.01	0.57

	Ixus70	EX-Z150	FinePixJ50	M1063	CoolPixS710	D200	D70	mju-1050SW	DMC-FZ50	OptioA40	DCZ5.9	GX100	RCP-7325XS	L74wide	NV15	DSC-H50	DSC-T77	DSC-W170
Ixus70	0.99																	
EX-Z150		0.96		0.01			0.01					0.01			0.01			
FinePixJ50			0.98															
M1063				0.95							0.02		0.01		0.01			
CoolPixS710		0.01			0.86	0.08			0.02				0.01			0.01		0.01
D200					0.03	0.95												
D70							1.00											
mju-1050SW								1.00										
DMC-FZ50									0.99									
OptioA40										0.99								
DCZ5.9											0.99							
GX100												0.99						
RCP-7325XS				0.02									0.97					
L74wide														0.99				
NV15													0.01		0.98			
DSC-H50														0.01		0.83	0.02	0.12
DSC-T77																	1.00	
DSC-W170				0.01	0.01							0.01		0.05		0.19	0.02	0.71

Figure 68. Camera Model Identification Results [1, 2] (left) and the Improvement Obtained with [3] (right)

Concerning the problem of tampering detection using camera features, we tested the proposed method [4] on a dataset built by splicing together images coming from different Dresden devices. The method depends on two thresholds that can be tuned differently and provide a detection score and a tampering mask as output. Figure 69 shows the results on tampering detection in terms of Area Under the Curve (AUC) obtained thresholding the tampering score. In terms of tampering localization, the proposed method achieves a pixel-wise accuracy greater than 0.8.

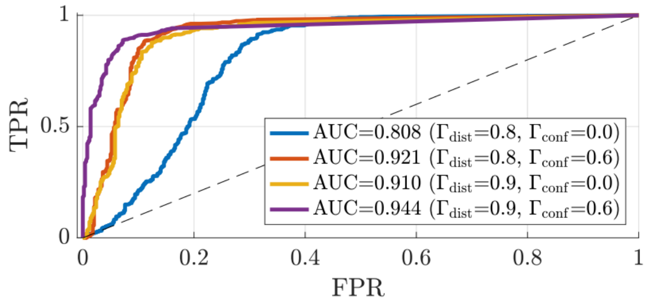


Figure 69. Area Under the Curve (AUC) for Tampering Detection Using Different Parameter Settings

To evaluate the proposed open-set strategy [5], we considered some camera models within our dataset as unknown. Figure 70 shows the obtained confusion matrix. Notice that most of the "unknown" models are correctly classified as such, while known models are also recognized.

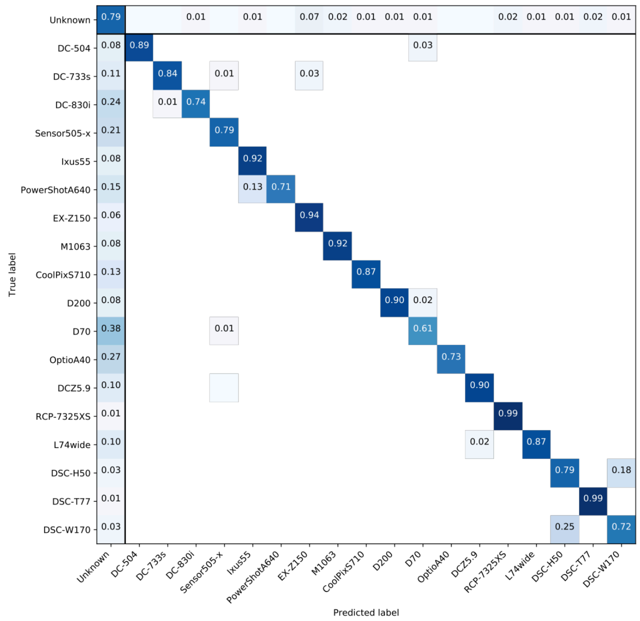


Figure 70. Open-set Camera Model Identification Confusion Matrix

Published Papers:

[1] L. Bondi, L. Baroffio, D. Güera, P. Bestagini, E. J. Delp, S. Tubaro, "First Steps Towards Camera Model Identification with Convolutional Neural Networks", IEEE Signal Processing Letters, vol. 24, no. 3, pp. 259-263, March 2017. DOI: 10.1109/LSP.2016.2641006
[2] L. Bondi, D. Güera, L. Baroffio, P. Bestagini, E. J. Delp, S. Tubaro, "A Preliminary Study on Convolutional Neural Networks for Camera Model Identification", IS&T Electronic Imaging (EI), vol. 2017, no. 7, pp. 67-76, Burlingame, California, January 2017. Preprint: Link DOI: 10.2352/ISSN.2470-1173.2017.7.MWWSF-327
[3] D. Güera, S. K. Yarlagadda, P. Bestagini, F. Zhu, S. Tubaro, E. J. Delp, "Reliability Map Estimation For CNN-Based Camera Model Attribution", IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 964-973, Lake Tahoe, Nevada, February 2018. DOI: 10.1109/WACV.2018.00111

[4] L. Bondi, S. Lameri, D. Güera, P. Bestagini, E. J. Delp, S. Tubaro, "Tampering Detection and Localization through Clustering of Camera-Based CNN Features", IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW), pp. 1855-1864, Honolulu, Hawaii, July 2017. DOI: 10.1109/CVPRW.2017.232
[5] P. R. M. Júnior, L. Bondi, P. Bestagini, S. Tubaro, and A. Rocha, "An In-Depth Study on Open-Set Camera Model Identification," IEEE Access, 2019.

Contract Information:
Team: Purdue Team, PoliMi Sub-Team
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: Adversarial techniques for camera model attribution
Major Technical Approach: Adversarial CNNs, inpainting-based solutions

Program Objectives: Study the robustness of PRNU detectors

To validate the inpainting-based solution [1], as image dataset we randomly selected 600 never-compressed Adobe Lightroom images from the Dresden Image Database coming from six camera instances (Nikon D70, Nikon D70s, Nikon D200, two devices each). All images were synchronized to landscape orientation and cropped to the central portion of size 512 × 512 pixels. The estimation of the clean sensor fingerprint K for each camera was obtained from 25 homogeneously lit flatfield images as typically suggested in the literature. Noise residuals W were computed with the Wavelet-based filter commonly used in PRNU extraction. Results are shown in Figure 71 in terms of ROC curves for different sets of algorithm parameters. Notice that the goal is to obtain a ROC curve as close to the 45-degree diagonal as possible. Indeed, the anonymization test must break the PRNU attribution test.

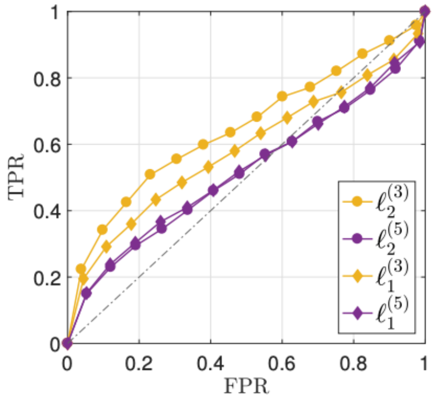


Figure 71. ROC Curve Related to the Inpainting-based Anonymization Method

Concerning the method based on CNNs [2], the same dataset has been used. In this case, Figure 72 shows the ROC curves obtained when a CNN denoiser is used to extract the PRNU (blue) or wavelet denoiser is used (orange). Images are anonymized in both cases. However, if one takes the absolute value of the correlation, the wavelet denoiser case provides the ROC curve in green.

This means that it is possible to slightly adapt the PRNU test to avoid being fooled by the network, if wavelet denoising is used.

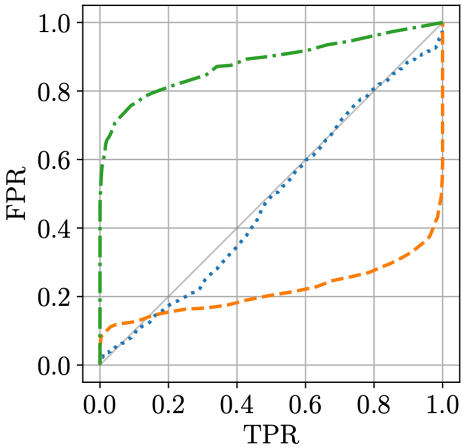


Figure 72. Results Achieved Through the Use of the Anonymization CNN

Published Papers:
[1] S. Mandelli, L. Bondi, S. Lameri, V. Lipari, P. Bestagini, S. Tubaro, "Inpainting-Based Camera Anonymization", *IEEE International Conference on Image Processing (ICIP)*, pp. 1522-1526, Beijing, China, September 2017. DOI: 10.1109/ICIP.2017.8296536
[2] N. Bonettini, L. Bondi, D. Güera, S. Mandelli, P. Bestagini, S. Tubaro, E. J. Delp, "Fooling PRNU-Based Detectors Through Convolutional Neural Networks", *European Signal Processing Conference (EUSIPCO)*, Rome, Italy, September 2018. DOI: 10.23919/EUSIPCO.2018.8553596

Contract Information:
Team: Purdue Team, PoliMi Sub-Team
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: PRNU-based forensics
Major Technical Approach: PRNU projections, CNNs for PRNU matching, search of video PRNU parameters

Program Objectives: PRNU compression based on projections, video PRNU attribution, fast PRNU attribution

Concerning PRNU compression [1, 2], the methods have been tested on a series of datasets built starting from Dresden Image Dataset and simulating different working conditions. Three controlled compression datasets are built upon RAW images coming from 6 camera devices, two for each model of Nikon-D200, Nikon-D70, Nikon-D70s: i) the first dataset is composed of 6 camera fingerprints extracted from flatfield RAW images; ii) the second dataset is composed of 1317 query residual extracted from natural RAW images; iii) the third dataset is composed of 1317 query residual extracted from natural RAW images compressed in JPEG format with different quality factors before the noise extraction process. Two uncontrolled compression datasets are built upon JPEG images from 53 camera models: i) the first dataset is a composed of 53 camera fingerprints extracted from flatfield JPEG images as encoded by cameras' firmware; ii) the second dataset is a composed of 9092 query residual extracted from natural JPEG images as encoded by cameras' firmware.

Choice of the resizing factor for the first step of the pipeline, has been performed evaluating the impact in terms of true positive rate when database fingerprints are extracted from raw images while query residuals are extracted from the other datasets, also considering different JPEG quality factors. For weak JPEG compression (QF>70) we observed a drop in detection performance when decimating with a factor greater than 3, whereas accuracy is preserved almost without loss for decimation factor below 3. For stronger JPEG compression factors (QF<70) decimation with factor 2 results beneficial, as it increases the Signal-to-Noise Ratio between the PRNU (signal) and the PRNU-unrelated noise components remaining after the noise extraction process. It is also interesting to notice that the loss-less behavior of decimation with factor 2 might be related to CFA interpolation, even though we have no experimental evidences to prove it at this time.

Concerning the proposed CNN-based framework for PRNU image attribution [3], Figure 73 shows the achieved performance in terms of Area Under the Curve (AUC) for different networks compared to the standard PCE test. AUC are computed for different patch size P. It is possible to notice that the PCE standard approach provides less accurate results that any of the used networks.

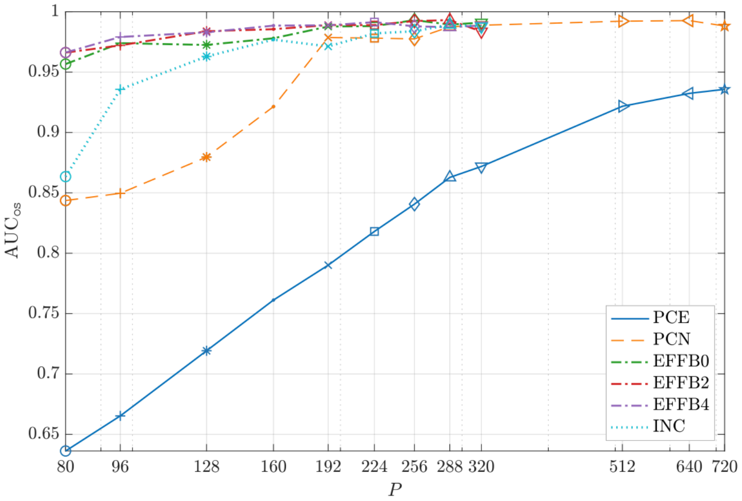


Figure 73. PRNU Attribution AUC vs. Patch Size for Different CNNs and the Standard PCE Test

Concerning video device attribution in presence of stabilization, results are reported in Figure 74. The method [4] (i.e., Mandelli 2019) in blue proves more accurate than method [5] in orange. However [5] provides faster results than [4].

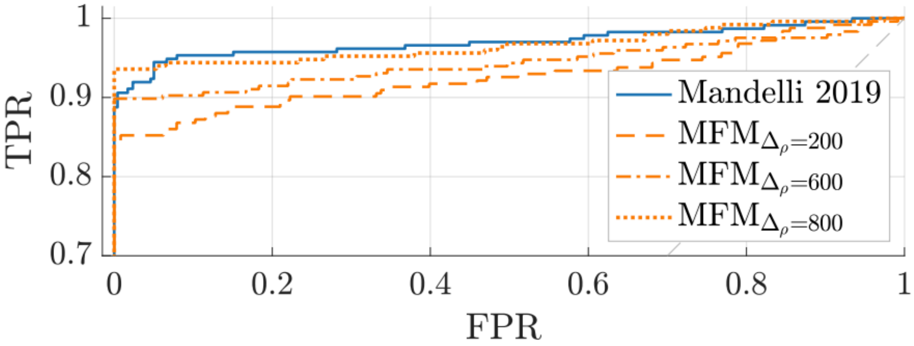


Figure 74. ROC Curves for Video PRNU Attribution

Published Papers:
[1] L. Bondi, P. Bestagini, F. Pérez-González, S. Tubaro, "Improving PRNU Compression through Preprocessing, Quantization and Coding", IEEE Transactions on Information Forensics and Security, vol. 14, no. 3, pp. 608-620, July 2018. DOI: 10.1109/TIFS.2018.2859587
[2] L. Bondi, F. Pérez-González, P. Bestagini, S. Tubaro, "Design of Projection Matrices for PRNU Compression", IEEE Workshop on Information Forensics and Security (WIFS), Rennes, France, December 2017. DOI: 10.1109/WIFS.2017.8267652
[3] S. Mandelli, D. Cozzolino, P. Bestagini, L. Verdoliva, and S. Tubaro, "CNN-based fast source device identification," IEEE Signal Processing Letters (SPL), 2020.
[4] S. Mandelli, P. Bestagini, L. Verdoliva, and S. Tubaro, "Facing Device Attribution Problem for Stabilized Video Sequences," IEEE Transactions on Information Forensics and Security (TIFS), 2019.
[5] S. Mandelli, F. Argenti, P. Bestagini, M. Iuliani, A. Piva, and S. Tubaro, "A Modified Fourier-Mellin Approach for Source Device Identification on Stabilized Videos," in IEEE International Conference on Image Processing (ICIP), 2020.

Contract Information:
Team: Purdue Team, PoliMi Sub-Team
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: Software identification through JPEG traces
Major Technical Approach: Study of JPEG quantization matrix

Program Objectives: Development of software identification detectors based on JPEG traces

Concerning the problem of double JPEG compression [1] Table 35 and Table 36 report the achieved accuracy using different test quality factors (QF), different patch size B, considering the case of training on JPEG images whose second compression QF is 75 and 85, respectively.

Table 35. Double JPEG Detection Accuracy When Training on Images Whose Second Compression is Operated with QF=75

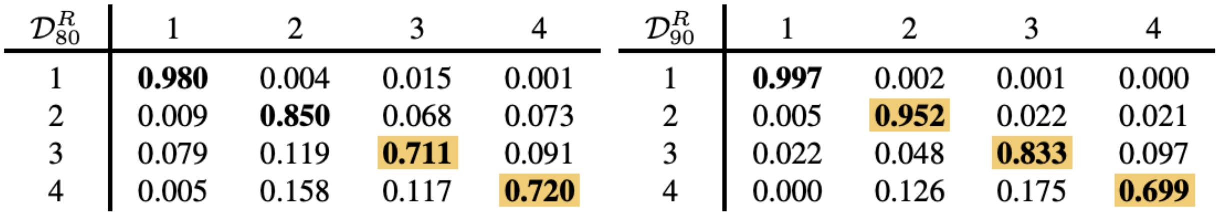
Testing (QF1, QF2)	$B = 64$	$B = 256$
(55, **75**)	0.816	0.876
(65, **75**)	0.805	0.866
(75, **75**)	0.764	0.842
(85, **75**)	0.674	0.776
(**60**, 78)	0.777	0.845
(**70**, 78)	0.765	0.830
(**60**, 80)	0.723	0.794
(**70**, 80)	0.720	0.790

Table 36. Double JPEG Detection Accuracy When Training on Images Whose Second Compression is Operated with QF=85

Testing (QF1, QF2)	$B = 64$	$B = 256$
(55, **85**)	0.897	0.972
(65, **85**)	0.878	0.972
(75, **85**)	0.865	0.961
(85, **85**)	0.793	0.954
(**70**, 88)	0.751	0.786
(**80**, 88)	0.738	0.785
(**70**, 90)	0.650	0.610
(**80**, 90)	0.634	0.600

Concerning the problem of multiple JPEG compression detection [2], results are shown in Table *37*. This picture shows the achieve confusion matrixes considering images whose QF ranges around 80 or 90, respectively.

Table 37. Multiple JPEG Compression Detection Confusion Matrices in Case of Images QF Around 80 and 90



$\mathcal{D}^R_{80}$	1	2	3	4
1	**0.980**	0.004	0.015	0.001
2	0.009	**0.850**	0.068	0.073
3	0.079	0.119	**0.711**	0.091
4	0.005	0.158	0.117	**0.720**

$\mathcal{D}^R_{90}$	1	2	3	4
1	**0.997**	0.002	0.001	0.000
2	0.005	**0.952**	0.022	0.021
3	0.022	0.048	**0.833**	0.097
4	0.000	0.126	0.175	**0.699**

The proposed method for JPEG implementation detection [3] is based on the following pipeline: i) the image under analysis is re-compressed using a custom JPEG implementation using the same quantization matrix of the original image; ii) the re-compressed and the original image are compared in the quantized DCT domain; iii) a feature vector is derived from the mean squared error obtained for each of the 64 JPEG DCT coefficients; iv) feature vectors are fed to a trained classifier.

As all the experiments involved the use of a supervised classifier (i.e., a random forest), we always proceeded in the following way: i) given a dataset, we randomly split its samples into 50% training and 50\% test; ii) the random forest is trained using default parameters\footnote according to scikit-learn implementation on the training set; iii) results are reported on the test set. As no hyper-parameters tuning is performed, there is no need for a validation set of data. This is done on purpose in order to evaluate the proposed feature discrimination capability with a simple classifier, rather than our ability of fully optimizing a machine-learning technique.

The first experiment was run to detect whether a single compressed JPEG image has been saved using Adobe Photoshop CC 2017 (hereinafter Photoshop) or the Python Imaging Library (PIL), considering that the same quantization matrix has been used. To this purpose, we built a dataset starting from the 1338 uncompressed color images at 512x384 pixel resolution of the UCID dataset. Notice that considering low resolution images makes the problem more challenging. Indeed, less pixels lead to less reliable statistics, thus feature vectors. We first JPEG compressed each uncompressed image using Photoshop at different JPEG quality levels, thus obtaining four distinct datasets. Then, we compressed each UCID uncompressed image with PIL, forcing the use of the respective JPEG quantization matrix used by Photoshop, thus obtaining other four datasets. With this setup, we extracted the proposed feature vector from the luminance component of every picture and trained a different classifier for each dataset pair (i.e., Photoshop vs. PIL). This experiment showed that accuracy for JPEG low quality images is higher than 86%, and it drops to 75% when high quality images are considered.

The second experiment we performed is a more challenging version of the first one: to detect whether an image has been originally JPEG compressed using Photoshop or PIL, given that afterward it is re-compressed with the same quantization matrix using PIL. For this experiment, we took all previously built datasets, and re-compressed each image using PIL and the same quantization matrix used for the first compression. Accuracy ranged from 89% for low quality images, to 65% for higher quality pictures. However, given the challenging task, we can conclude that the proposed feature vector is still a viable solution toward capturing JPEG implementation traces.

Published Papers:

[1] M. Barni, L. Bondi, N. Bonettini, P. Bestagini, A. Costanzo, M. Maggini, B. Tondi, S. Tubaro, "Aligned and non-aligned double JPEG detection using convolutional neural networks", Journal of Visual Communication and Image Representation, vol. 49, pp. 153-163, November 2017. DOI: 10.1016/J.JVCIR.2017.09.003

[2] S. Mandelli, N. Bonettini, P. Bestagini, V. Lipari, S. Tubaro, "Multiple JPEG Compression Detection through Task-driven Non-negative Matrix Factorization", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2106-2110, Calgary, Canada, April 2018. DOI: 10.1109/ICASSP.2018.8461904

[3] N. Bonettini, L. Bondi, P. Bestagini, and S. Tubaro, “JPEG Implementation Forensics Based on Eigen-Algorithms,” in IEEE International Workshop on Information Forensics and Security (WIFS), 2018.

Contract Information:
Team: Purdue Team, PoliMi Sub-Team
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: Video manipulation forensics
Major Technical Approach: Coding-based traces, PRNU-based techniques, deepfake detection through CNNs.

Program Objectives: Development of forensic techniques for video manipulation detection

Concerning the codec-based video tampering detection method, the proposed technique has been tested by considering a dataset of videos with 720p resolution, each one consisting of 210 frames. Each original video has been encoded using three different codecs, and 4 different quality values per codec. Video forgeries have been generated by splicing together portions coming from differently encoded versions of the same original video. In doing so, we ensure that the proposed solution does not detect changes in video content, but actual codec and quality changes. Results on a set of 330 tampered videos show the it is possible to detect the presence of a forgery on a single frame analysis achieving AUC equal to 0.825, 0.800 or 0.889 if coding-based, quality-based or both features are used, respectively. In terms of tampering localization, by visual inspection we verified that it is possible to highlight which portion of the frame is incoherent with the rest of it.

In order to validate the proposed PRNU-based detectors, splicing video portions have been collected from Vision dataset, acquired with more than 30 mobile devices. More specifically, we created two distinct datasets for non-stabilized and stabilized compilations. From Vision dataset, we select non-stabilized devices with minimum Video Resolution set to HD-Ready (VR ≥ 720p). Then, 5 videos per device are collected, randomly picking from indoor/outdoor scenarios and considering only move/panrot acquisition modes. This choice comes from the idea of generating plausible results, since combinations of flat or static videos are actually less likely to be found. For each device, we cut the 5 selected videos at frame index 150, 300, 450, 600, 750, respectively, in order to generate splicing portions of different lengths. The splicings are then cropped to common resolution of 720 × 720 pixels and grayscale converted. Since the available non-stabilized devices are 19, we end up with a pool of 95 distinct splices. The final video compilation is obtained as the temporal concatenation of Ns $\in$ [3, 6] splicing portions, randomly extracted from the pool. Following this pipeline we generated two datasets, covering 150 non-stabilized videos and just as many stabilized. Results show that the correct number of clusters can be detected around 90% of the times.

Concerning deepfake detection, the proposed solution has been tested on two well-known datasets: FaceForeniscs++; the DeepFake Detection Challenge (DFDC). On the former, we achieve AUC of 0.94. On the latter, we achieve AUC of 0.88. In both scenarios we outperform XceptionNet used as baseline. Moreover, by aggregating eight differently trained versions of our network in a voting pipeline, we reached the top 2% on the DFDC challenge on Kaggle (i.e., position 41).

Published Papers:

[1] S. Verde, L. Bondi, P. Bestagini, S. Milani, G. Calvagno, S. Tubaro, "Video Codec Forensics Based on Convolutional Neural Networks", IEEE International Conference on Image Processing (ICIP), pp. 530-534, Athens, Greece, October 2018. DOI: 10.1109/ICIP.2018.8451143
[2] S. Mandelli, D. Cozzolino, P. Bestagini, L. Verdoliva, S. Tubaro, "Blind Detection and Localization of Video Temporal Splicing Exploiting Sensor-Based Footprints", European Signal Processing Conference (EUSIPCO), Rome, Italy, September 2018. DOI: 10.23919/EUSIPCO.2018.8553511
[3] P. R. M. Júnior, L. Bondi, P. Bestagini, A. Rocha, and S. Tubaro, "A PRNU-Based Method to Expose Video Device Compositions in Open-Set Setups," in IEEE International Conference on Image Processing (ICIP), 2019.
[4] N. Bonettini, E. D. Cannas, S. Mandelli, L. Bondi, P. Bestagini, and S. Tubaro, "Video Face Manipulation Detection Through Ensemble of CNNs," in International Conference on Pattern Recognition (ICPR), 2020.

4.1.5 New York University

Contract Information:
Team: Purdue Team, NYU Sub-Team
POC: Prof. Nasir Memon, New York University, memon@nyu.edu, 732-241-6128

Major Technical Problem: Factors Affecting ENF based Time-of-Recording Verification
Major Technical Approach: Investigation of how the quality of the ENF signal to be estimated from video is affected by different light source illumination, and investigation of how the quality of the estimated ENF signal is affected by different compression ratios depending on the type of light source.

Program Objective: exploration of how the type of illumination source affects ENF based video forensics, and also exploration of how it is affected by some other factors depending on the type of illumination source and video length.

In order to accomplish the above mentioned objectives, we first investigated light source effect. For this purpose, we tested videos captured under illumination of different type of light sources, namely white CFL, yellow CFL, white LED, yellow LED, and Halogen bulb (yellow color). Table 38 provides the rate (in %) of correct time-of-recording estimation when ENF signal is searched in true reference database, i.e., the task of time-of-recording verification. As can be seen from the table, ENF is best estimated for videos recorded under LED illumination, though the quality of the ENF signal noticeably drops under CFL illumination. Also explored is the performance increases as the video length increases, and at least 2-minute video length is required for a reliable analysis. The experiments for other factors were conducted using the light sources that lead the best and the worst performances, i.e. white LED and CFL only. Table 39 provides the rates of correct time-of-recording estimations (%) for different compression rates for CFL versus LED. As can be observed from the table, up to a rate of 500 Kbps, the performance is 100% for each size of video clips for LED, whereas for CFL, the performance drops for 2-minute videos as the compression ratio increases. For 100 Kbps, and Facebook compression, the performance for LED is also relatively better than that for CFL. Table 40 provides the rates of correct time-of-recording estimations (%) for different lengths of ground-truth ENF data for CFL versus LED. For LED, 100% success rate is obtained for each length of video clip and each length of reference data, whereas for CFL, the performance for 2-minute videos drops as the length increases.

Table 38. The Rates of Correct Time-of-Recording Estimations (%) for Different Light Sources

Clip L. (min.)	Halogen	CFL (white)	CFL (yellow)	LED (white)	LED (yellow)
1	2.25	0.00	0.50	26.37	20.00
2	100	51.75	47.25	100	97.75
5	100	100	100	100	100
10	100	100	100	100	100

Table 39. The Rates of Correct Time-of-Recording Estimations (%) for Different Compression Rates for CFL vs. LED

Bit Rate	2 min.		5 min.		10 min.	
	CFL	LED	CFL	LED	CFL	LED
Original	51.75	100	100	100	100	100
500 Kbps	6.00	100	100	100	100	100
100 Kbps	0	19.75	0	87.50	51.25	100
Facebook	0	12.25	0	56.88	0	100

Table 40. The Rates of Correct Time-of-Recording Estimations (%) for Different Lengths of Ground-Truth ENF Data for CFL vs. LED

Database Length	2 min.		5 min.		10 min.	
	CFL	LED	CFL	LED	CFL	LED
One-day	51.75	100	100	100	100	100
One-week	27.50	100	100	100	100	100
One-month	11.00	100	98.13	100	100	100

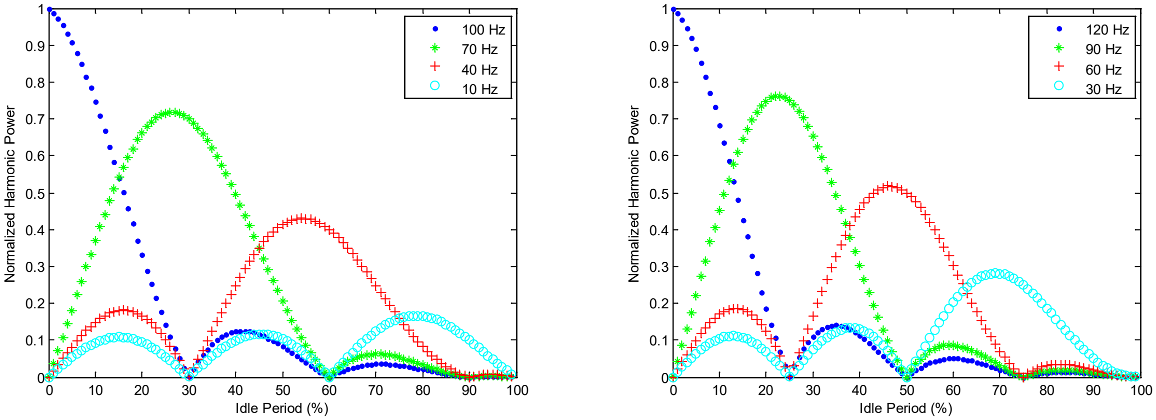
Published Paper:
S. Vatansever, A. E. Dirik and N. Memon, "Factors Affecting Enf Based Time-of-recording Estimation for Video," *Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing*, Brighton, United Kingdom, (2019)

Contract Information:
Team: Purdue Team, NYU Sub-Team
POC: Prof. Nasir Memon, New York University, memon@nyu.edu, 732-241-6128

Major Technical Problem: The Effect of Rolling Shutter Sampling Mechanism and of Idle Period between Successive Frames on Capture and Estimation of ENF in a Given Video
Major Technical Approach: Rolling Shutter based Luminance Samples Estimation

Program Objective: The main objective is development of an ENF-based video authentication method in the sense of time-of-recording verification. This task requires a reliable ENF signal estimation. In order to obtain an accurate ENF signal estimation from a given video, we need to know how the main ENF harmonic is changed depending on the idle period in a rolling shutter based sampling mechanism. In addition to that we need to know which harmonics to exploit and how to use the estimated ENF signals from this harmonics for an improved time-of-recording verification performance.

In order to accomplish the above mentioned objective, we derived an analytical model illustrating how the frequency of the main ENF harmonic is replaced with new ENF components depending on the length of the idle period per frame. By exploiting multiple ENF harmonics introduced by the proposed model, we obtain a better time-of-recording verification performance. This approach also leads us to estimate the idle period as auxiliary information possibly to use in camera forensics. Figure 75 illustrates the nominal illumination harmonic, 100/120 Hz (twice the nominal ENF), is shifted to some other frequency depending on the idle period length. It also reveals that the power of the captured ENF signal is inversely proportional to idle period length. Hence, in the presence of noise, videos with long idle periods may be a great challenge for ENF based forensic analysis. Table 41 provides a comparative analysis for the proposed time-of-recording verification technique. Accordingly, proposed metrics (Metric 3 and Metric 4) outperform those in the literature (Metric 1 and Metric 2) with true decision rates 79.16% and 83.33%, respectively.



(a) (b)

Figure 75. Variation in Frequency of Main ENF Harmonic vs. Idle Period for a 30 FPS Video: (a) Captured in EU (50 Hz Mains Power), (b) Captured in the US (60 Hz Mains Power)

Table 41. Experimental Results with the Videos With Moving Content

Method	**Metric 1**	**Metric 2**	**Metric 3 (Proposed)**	**Metric 4 (Proposed)**
TD (%)	62.50	75.00	79.16	83.33
FD (%)	8.33	8.33	12.50	8.33
ND (%)	29.16	16.66	8.33	8.33

TD = True Decision, FD = False Decision, ND = No Decision

Published Paper:
S. Vatansever, A. E. Dirik and N. Memon, "Analysis of Rolling Shutter Effect on ENF-Based Video Forensics," *IEEE Transactions on Information Forensics and Security,* **14**, 9, Sept. 2019, pp. 2262-2275.

Contract Information:
Team: Purdue Team, NYU Sub-Team
POC: Prof. Nasir Memon, New York University, memon@nyu.edu, 732-241-6128

Major Technical Problem: Detecting the Presence of ENF in a Given Video
Major Technical Approach: Super-pixel based ENF estimation

Program Objective: One objective is development of a multi-region ENF estimator utilizing multiple ENF vectors estimated from separate regions. Another objective is development of an ENF presence detector to decide whether a given video is applicable to an ENF based forensic analysis.

In order to accomplish the above mentioned objectives, we divide the content into small regions in a way that all pixels within a region have uniform reflectance characteristics, i.e., we divide the content super-pixels. We use simple linear iterative clustering (SLIC) segmentation algorithm to compute super-pixel regions in a given video. We mainly estimate ENF from each super-pixel region, and compute a representative ENF by means of element-wise median or mean operation of all estimated vectors. This operation eliminates some regions' negative contribution to the final ENF signal. We then check the similarity of each estimated signal to the representative to decide whether the given video contains ENF. The proposed technique can operate on video clips as short as 2 min and is independent of the camera sensor type, i.e., CCD or CMOS. Table 42 provides the area under the ROC curve (AUC) values obtained for 4 different decision metrics that are calculated based on the utilization of median based representative ENF. According to the table, the detection metric *f1* (mean of estimated correlation coefficients between each regional ENF vector and the representative) outperforms other metrics not only for mixture of sensor types but also for each sensor type independently.

Table 42. ENF Detection Performance (AUC) based on Median based Representative ENF

Sensor Type	# Videos	*f1*	*f2*	*f3*	*f4*
CCD	80	**0.985**	0.947	0.931	0.960
CMOS	80	**0.959**	0.942	0.941	0.944
Any (Mixed)	160	**0.973**	0.939	0.931	0.952

Published Paper:
S. Vatansever, A. E. Dirik and N. Memon, "Detecting the Presence of ENF Signal in Digital Videos: A Superpixel-Based Approach," *IEEE Signal Processing Letters*, **24**, 10, Oct. 2017, pp. 1463-1467

Contract Information:
Team: Purdue Team, NYU Sub-team
POC: Prof. Nasir Memon, New York University, memon@nyu.edu, 732-241-6128

Major Technical Problem: The need for ground-truth ENF for time-of-recording verification.
Major Technical Approach: Development of a ground-truth ENF recorder to be connected to any power outlet from where it can directly read the reduced mains power voltage and can estimate ENF at 1 samples/second resolution by using these voltage time series by means of STFT based ENF estimation technique.

Program Objective: Creation of ground-truth ENF databases for different mains-power networks.

As Rome NY and in NYU is located in same mains power network, i.e., Eastern, ground-truth ENF variations obtained from these 2 different locations in the same day, on 10th July 2020, are the same as can be seen in **Figure *76***. ENF signal obtained for Denver on 10th July 2020 is different than that for Rome and NYU since it is located in Western network.

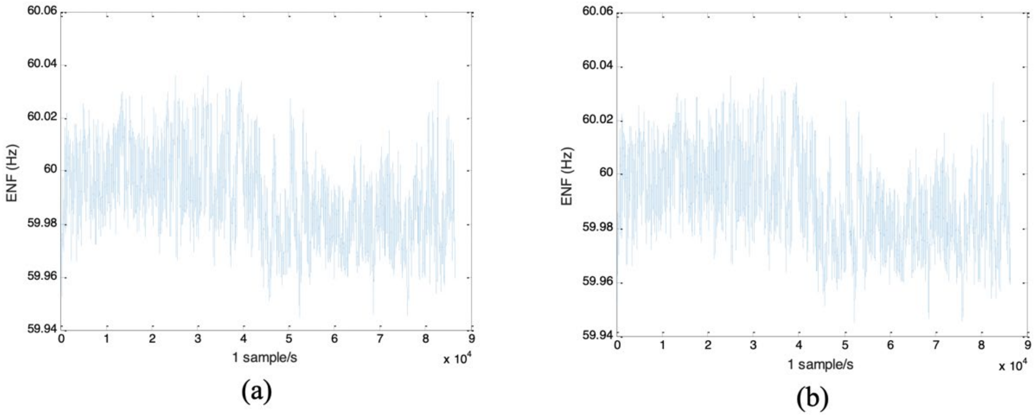


Figure 76. Ground Truth ENF Signal Computed on 10th July 2020 from (a) NYU, (b) Rome NY

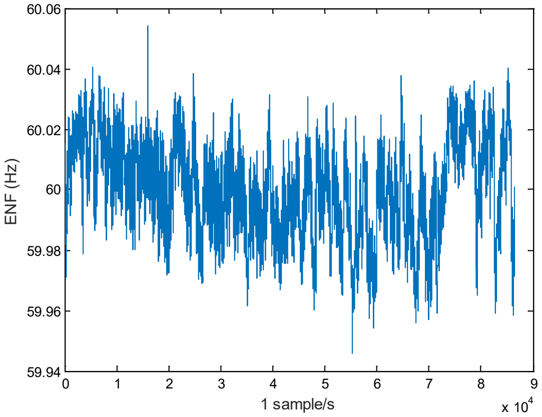


Figure 77. Ground Truth ENF Signal Computed on 10th July 2020 from Denver

Different locations/regions in the same power network in the US, i.e., Eastern, Western, etc., have different clock times, which is a crucial phenomenon to be taken into consideration in time-of-recording verification task. That is, we record the ground truth ENF at one point only, i.e. Denver, in the Western network. Similarly, in the Eastern network we record the ground truth ENF also at one region only, i.e. New York. Nevertheless, there is one hour difference between the east part and west part for each network. Consequently, when we search the estimated ENF signal of a video recorded in the west part with the ground truth ENF recorded in the east part of the same network, the time lag is estimated at the wrong point. For instance, recording time of a video captured in Las Vegas that is located in the Western Network is estimated one hour ahead of the actual recording time, when the estimated ENF signal is compared with the ground-truth ENF obtained in Denver. This is because; the time in Las Vegas is one hour behind the Denver time. Similarly, recording time of a video captured in Chicago that is located in the Eastern Network is estimated one hour ahead of the actual recording time, when the estimated ENF signal is compared with the ground-truth ENF obtained in Rome NY.

Contract Information:
Team: Purdue Team, NYU Sub-team
POC: Prof. Nasir Memon, New York University, memon@nyu.edu, 732-241-6128

Major Technical Problem: Time-of-Recording Verification
Major Technical Approach: Attenuation of frame rate harmonics

Program Objective: The main objective is development of an ENF-based video authentication method. For this purpose, we mainly target the task of video time-of-recording verification which requires a reliable ENF signal estimation.

The proposed algorithm for attenuation of frame rate harmonics best performs with a 2nd degree curve fitting, followed by 5% removal of the data. Figure 78 and Figure 79 respectively provide intensity fluctuations (a) and the frame rate harmonics (b) before and after the algorithm run.

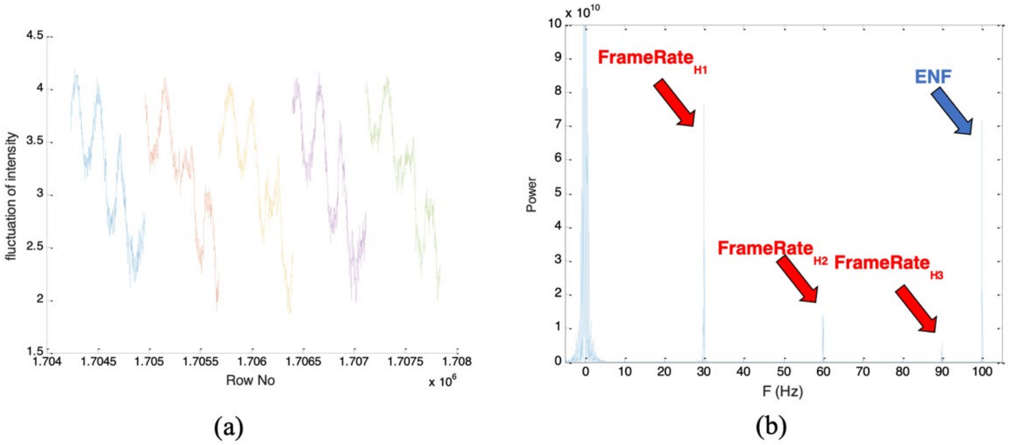


Figure 78. (a) Intensity Fluctuations (b) Frame Rate Harmonics for an Exemplary Video Before Curve Fitting Operation

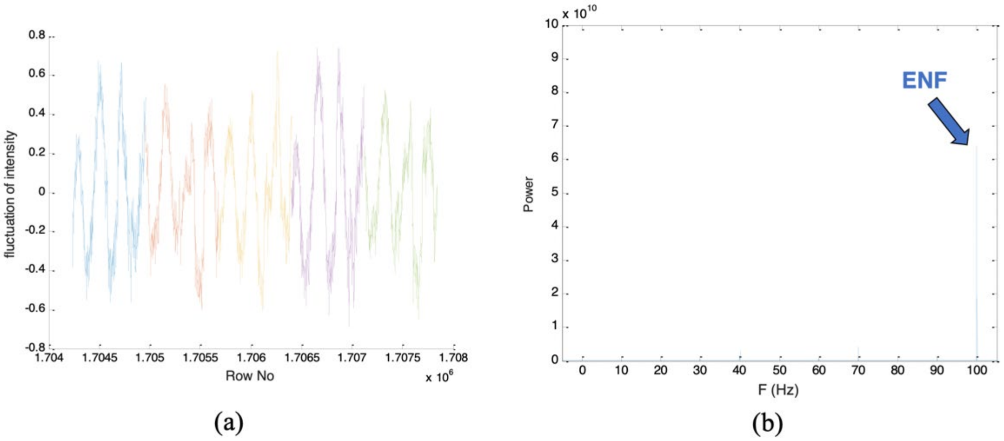


Figure 79. (a) Intensity Fluctuations (b) Frame Rate Harmonics for an Exemplary Video After Curve Fitting Operation

We also tested the effectiveness of the proposed approach on our TIFS paper [1] dataset for time-of-recording verification task. Table 43 provides the comparative results. Accordingly, we outperform the presented method in the TIFS.

Table 43. A Comparative Analysis

Method	TIFS [1]	Proposed Approach
TD (%)	83.33	**91.67**
FD (%)	8.33	0.00
ND (%)	8.33	8.33

TD = True Decision, FD = False Decision, ND = No Decision

[1] S. Vatansever, A. E. Dirik and N. Memon, "Analysis of Rolling Shutter Effect on ENF-Based Video Forensics," in IEEE Transactions on Information Forensics and Security, vol. 14, no. 9, pp. 2262-2275, Sept. 2019, doi: 10.1109/TIFS.2019.2895540.

Published Paper:
We are currently working on a TIFS paper including the above-mentioned concepts!

Contract Information:
Team: Purdue Team, NYU Sub-team
POC: Prof. Nasir Memon, New York University, memon@nyu.edu, 732-241-6128

Major Technical Problem: Time-of-Recording Verification
Major Technical Approach: Detection and classification of quality ENF samples

Program Objective: The main objective is development of an ENF-based video authentication method. For this purpose, we mainly target the task of video time-of-recording verification which requires a proper similarity test between the estimated ENF signal and the ground truth.

The proposed approach for detection and classification of quality ENF samples along with the adapted normalized cross correlation is distinguishable for noisy ENF signals. It even shows an extraordinary performance for some cases as in Figure 80, for which the proposed approach leads a time-of-recording match even for such a noisy ENF signal.

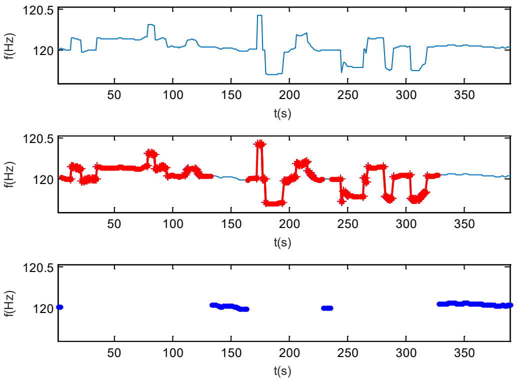


Figure 80. Analysis of an Estimated ENF Signal from an Exemplary Video. Blue Color in the Bottom Figure Represents Quality ENF Samples, Whereas Red Color in the Middle Figure Illustrates Useless ENF Samples. The Proposed Approach Leads a Time-of-Recording Match for Such a Noisy Signal

In order to test the effectiveness of the proposed method, we made a comparative analysis for non-modified and at least 2-minute length Medifor videos at collection 1089 and 974 (35 videos) and live-stream videos (34 videos). Overall, we increase the accuracy from 68.57% to 74.28% for the Medifor videos, and from 72.05% to 80.88% for the stream videos as provided in Table 44.

Table 44. Comparative Results for Medifor Videos and Live-Stream Videos

Approach	ACC	
	Medifor Videos	**Stream videos**
Method 1: Directly use of the estimated ENF signal	68.57	72.05
Method 2: Use of the longest good-quality ENF clip	71.42	75.00
Method 3 (Part of the proposed): Use of the longest good-quality ENF clip with samples removal at either side of the noisy parts	71.42	76.47
Method 4 (Proposed): Use of all good-quality ENF clips with samples removal at either side of the noisy parts	**74.28**	**80.88**

The proposed approach can also be used as a quality ENF signal presence detector. That is, it can also be used to test whether the estimated ENF is appropriate for an ENF based forensic analysis.

Published Paper:
We are currently working on an SPL paper for the above-mentioned concepts!

4.1.6 University of Notre Dame

Contract Information:
Team: Purdue Team, Notre Dame Sub-Team
POC: Prof. Walter Scheirer, University of Notre Dame, walter.scheirer@nd.edu, 574-631-2436

Major Technical Problem: Image Provenance Analysis
Major Technical Approach: Image Provenance at Scale Pipeline

Program Objectives: maximize evaluation metrics for both tasks of (i) provenance image filtering – namely Recall considering the top-50 (R@50), top-100 (R@100), and top-200 (R@200) images of the retrieved image rank – and of (ii) provenance graph construction – namely vertex overlap (VO), edge overlap (EO), and their combination (VEO).

Table 45 summarizes the average results of the proposed approach for the task of image filtering, along the four years of official project evaluation. The NC17 baseline results were obtained with our team's previous solution [31] and are reported here for comparison sake. As one might observe, the proposed approach has improved previous results and was robust across years of challenge, in spite of the changes in dataset size, number of probes, and different types of image transformations.

Table 45. Official Results Along Project Years for Image Filtering

Year	Dataset	Size (# images)	Probes (# images)	R@50	R@100	R@200
2017	NC17 (baseline)	~1 million	151	0.552	0.587	0.635
2018	MFC18	~1 million	~10000	0.846	0.882	0.884
2019	MFC19	~1 million	~1000	0.798	0.804	0.814
2020	MFC20	~2 millions	~6000	0.805	0.831	0.855

Table 46 contains the average results of the proposed approach for the task of graph construction, along the four years of official evaluation. The technique was again consistently robust across evaluations.

Table 46. Official Results Across Project Years for Graph Construction

Year	Dataset	Size (# images)	Probes (# images)	VO	EO	VEO
2017	NC17	~1 million	151	0.613	0.208	0.412
2018	MFC18	~1 million	~10000	0.798	0.298	0.553
2019	MFC19	~1 million	~1000	0.703	0.295	0.519
2020	MFC20	~2 millions	~6000	0.714	0.250	0.490

Finally, Table 47 contains the average results of graph construction over other program’s and third-party real-world more complex datasets, all reported in the published paper.

Table 47. Results of Graph Construction Over Other Datasets

Dataset	Graphs (# cases)	VO	EO	VEO
NC17-Dev1-Beta4	65	0.853	0.353	0.613
Professional	80	0.985	0.218	0.604
Reddit	100	0.924	0.121	0.526

This work is further detailed in our published paper [1].

Published Paper:
[1] Moreira et al. “Image Provenance at Scale,” IEEE T-IP, Vol. 27, No. 12, 2018

Contract Information:
Team: Purdue Team, Notre Dame Sub-Team
POC: Prof. Walter Scheirer, University of Notre Dame, walter.scheirer@nd.edu, 574-631-2436

Major Technical Problem: Image Provenance Analysis
Major Technical Approach: Image Provenance Leveraging Metadata

Program Objectives: improve evaluation metrics for the task of provenance graph construction, namely vertex overlap (VO), edge overlap (EO), and their combination (VEO), through the use of image metadata.

Table 48 presents the effect of applying image metadata to the task of provenance graph construction, for the case of the NC17-Dev1-Beta4 dataset, which has 65 cases. As one might observe, the usage of metadata information only (second data row) does not surpasses the usage of visual information (represented by the baseline row, which depicts our previous results, published in [1]). However, the combination of metadata and visual information leads to a significant improvement of EO (and VEO, consequently), in the cases where metadata is available and was not tampered. Such improvement can be seen in the last row of *Table 48*, attesting the efficacy of the proposed solution.

Table 48. Average Results of Provenance Graph Construction Over the NC17-Dev1-Beta4 Dataset

Solution	VO	EO	VEO
Visual (baseline)	0.853	0.353	0.613
Metadata only	0.249	0.009	0.130
Metadata + Visual	0.853	0.384	0.628

This work is further detailed in our published paper [2].

Published Papers:
[1] Moreira et al. "Image Provenance at Scale," IEEE T-IP, Vol. 27, No. 12, 2018
[2] Bharati et al. "Beyond Pixels: Image Provenance Analysis Leveraging Metadata," IEEE WACV, 2019

Contract Information:
Team: Purdue Team, Notre Dame Sub-Team
POC: Prof. Walter Scheirer, University of Notre Dame, walter.scheirer@nd.edu, 574-631-2436

Major Technical Problem: Image Provenance Analysis

Major Technical Approach: Fast Local Spatial Verification for Feature-Agnostic Large-Scale Image Retrieval

Program Objectives: improve evaluation metrics for the task of provenance image filtering, namely Recall considering the top-50 (R@50), top-100 (R@100), and top-200 (R@200) images of the retrieved image rank, through the use of the proposed OS2OS image matching score.

Figure 81 depicts the performance of four methods of image retrieval over the MFC18 development dataset, which contains nearly 3300 probes. Average R@25, R@50, R@100 and R@200 are plotted for comparison sake. As one might observe, baseline solutions such as DSURF (proposed in [1]) and DELF (proposed in [2]) are improved whenever combined to the proposed OS2OS image matching score approach (see, for instance, DSURF versus DSURF+OS2OS, and DELF versus DELF+OS2OS approaches).

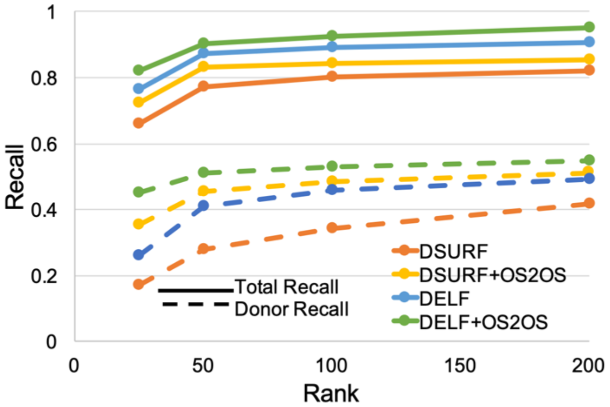


Figure 81. Average Recall Values for the MFC2018 Development Dataset, With Nearly 3,300 Queries

In addition, the original goal of the OS2OS score is to improve the retrieval of “small donors”, i.e., images that donate small objects rather than background to composite images. Thanks to the groundtruth annotation provided to MFC18, we can inspect the recall values of only small donors (“Donor Recall”, represented by dashed lines). In these cases, the application of the OS2OS score significantly improves the recall of small donors, bringing to further steps of provenance analysis images donating small objects that were being missed in previous approaches.
This work is further detailed in our published paper [3].

[2] Noh, H., Araujo, A., Sim, J., Weyand, T., and Han, B., "Large-Scale Image Retrieval with Attentive Deep Local Features," *The IEEE International Conference on Computer Vision*, Venice, Italy, October 2017, pp. 3476-3485, doi: 10.1109/ICCV.2017.374.
Our Papers:
[1] Published Paper: Moreira et al. “Image Provenance at Scale,” IEEE T-IP, Vol. 27, No. 12, 2018
[3] J. Brogan *et al.*, "Fast Local Spatial Verification for Feature-Agnostic Large-Scale Image Retrieval," in *IEEE Transactions on Image Processing*, vol. 30, pp. 6892-6905, July 2021. DOI: 10.1109/TIP.2021.3097175

Contract Information:
Team: Purdue Team, Notre Dame Sub-Team
POC: Prof. Walter Scheirer, University of Notre Dame, walter.scheirer@nd.edu, 574-631-2436

Major Technical Problem: Image Provenance Analysis
Major Technical Approach: Learning Transform-Aware Embeddings for Image Forensics

Program Objectives: improve evaluation metrics for the task of provenance graph construction, namely vertex overlap (VO), edge overlap (EO), and their combination (VEO), through the use of a novel deep-learning-based method of image patch representation.

Table 49 contains various average results of provenance graph construction over the MFC18-Eval-Ver1-Part1-Oracle dataset, which contains 445 cases. This dataset was selected by NIST in 2018 as an official program dataset to evaluate performers on the so-called “oracle” scenario. The oracle scenario comprises the situation where graph construction is executed on top of the results of an ideal provenance graph filtering solution, meaning that all images belonging to the provenance graph are provided, without distractors (i.e., images that do not belong to the provenance graph are not available).

Table 49. Results of Provenance Graph Construction Over the MFC18-Eval-Ver1-Part1-Oracle Dataset

Solution	VO	EO	VEO
Official submission #907	0.84	0.51	0.68
Official submission #973	0.80	0.15	0.50
Official submission #1268	0.80	0.50	0.65
SURF-based approach	0.90	0.55	0.74
DELF-based approach	0.89	0.49	0.70
ResNet-18-based approach	1.00	0.46	0.73
Proposed approach	1.00	0.58	0.79

Due to the nature of the proposed deep-learning-based approach, which does not provide directions to the edges of the provenance graphs, all reported values regard undirected provenance graphs. As one might observe, the first three data rows of Table 49 regard the official submissions of regular performers of the program in 2018, kindly provided by NIST. The following three rows regard exploratory experiments, with variation in the low-level image representation (through either SURF [1], DELF [2], or ResNet-18 [3] descriptors). The last row depicts the performance of the proposed approach, which surpasses all other methods, including official performers.

This work is further detailed in our published paper [4].

[1] Bay, H., Tuytelaars, T., and Van Gool, L., “SURF: Speeded Up Robust Features”, *The European Conference on Computer Vision,* Graz, Austria, May 2006, pp. 404-417, doi: 10.1007/11744023_32
[2] Noh, H., Araujo, A., Sim, J., Weyand, T., and Han, B., "Large-Scale Image Retrieval with Attentive Deep Local Features," *The IEEE International Conference on Computer Vision*, Venice, Italy, October 2017, pp. 3476-3485, doi: 10.1109/ICCV.2017.374.

[3] He, K., Zhang, X., Ren, S., and Sun, J., "Deep Residual Learning for Image Recognition," *The IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, June 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
Our Papers:
[4] A. Bharati, D. Moreira, P. J. Flynn, A. de Rezende Rocha, K. W. Bowyer, W. J. Scheirer, "Transformation-Aware Embeddings for Image Provenance," in *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 2493-2507, January 2021. DOI: 10.1109/TIFS.2021.3050061

4.2 Overhead Forensics Project

In this section we present further discussion of the Overhead Forensics project.

4.2.1 Purdue University

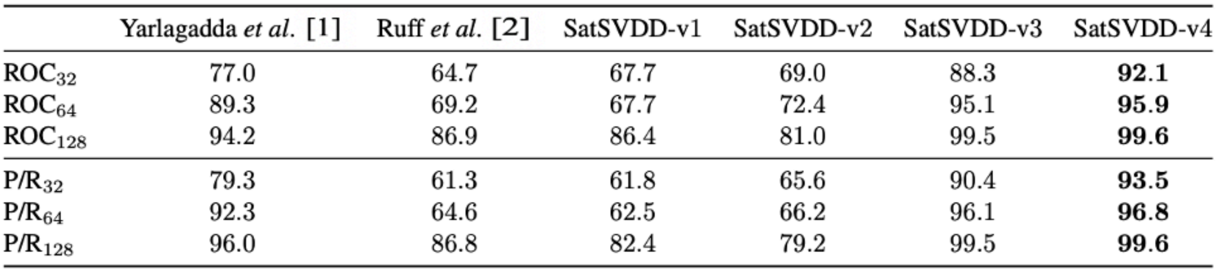
Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Unsupervised splicing detection, localization in satellite images
Major Technical Approach: Sat-SVDD (Satellite Support Vector Data Descriptor)

Program Objectives: Applying Sat-SVDD to detect spliced objects from unknown origin in overhead images.

The combined modifications to the Deep SVDD classifier lead to an increased anomaly detection and localization performance. Applying only one modification did not lead to better performance compare to the unchanged implementation of Deep SVDD, but by applying all of them we can achieve a significant improvement in all the reported metrics. The proposed detection function makes the pristine and forged image scores distributions more distinguishable.

Table 50. AUC Scores (%) for the Detection Task (ROC and P/R Metrics). The Subscript Denotes the Manipulation Size. Results That Surpass All Competing Methods Are Bold. Our Final Proposed Model, SatSVDD-v4, Outperforms All Previous Approaches



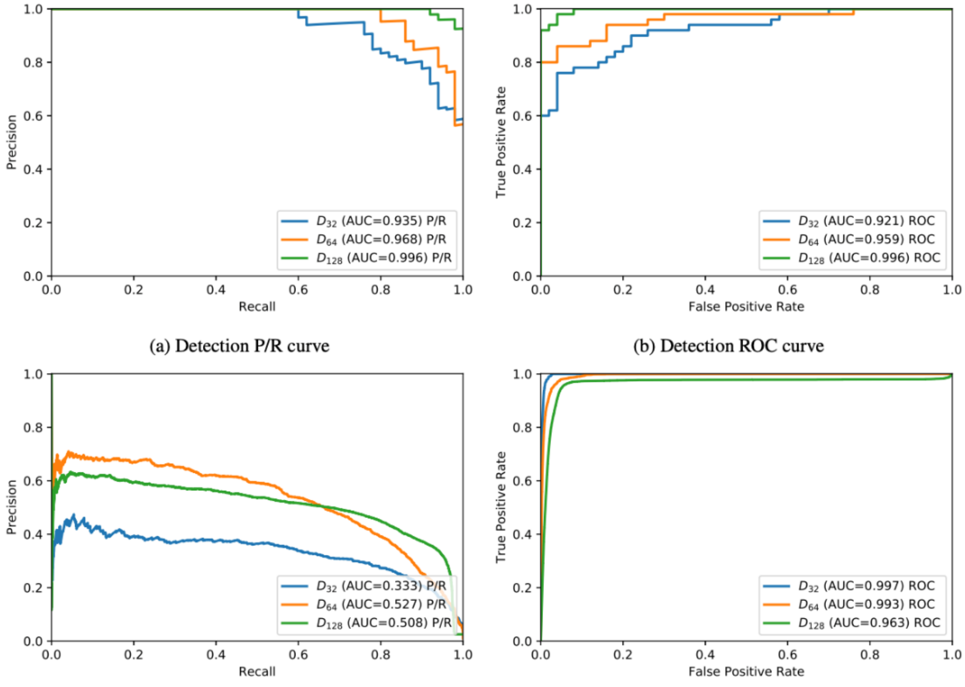
	Yarlagadda *et al.* [1]	Ruff *et al.* [2]	SatSVDD-v1	SatSVDD-v2	SatSVDD-v3	SatSVDD-v4
ROC_{32}	77.0	64.7	67.7	69.0	88.3	**92.1**
ROC_{64}	89.3	69.2	67.7	72.4	95.1	**95.9**
ROC_{128}	94.2	86.9	86.4	81.0	99.5	**99.6**
P/R_{32}	79.3	61.3	61.8	65.6	90.4	**93.5**
P/R_{64}	92.3	64.6	62.5	66.2	96.1	**96.8**
P/R_{128}	96.0	86.8	82.4	79.2	99.5	**99.6**

Table 51. AUC Scores (%) for the Localization Task (ROC and P/R Metrics). The Subscript Denotes the Manipulation Size. Results That Surpass All Competing Methods Are Bold. Our Final Proposed Model, SatSVDD-v4, Outperforms Almost All Previous Approaches

	Yarlagadda *et al.* [1]	Ruff *et al.* [2]	SatSVDD-v1	SatSVDD-v2	SatSVDD-v3	SatSVDD-v4
ROC_{32}	77.0	64.7	67.7	69.0	88.3	**92.1**
ROC_{64}	89.3	69.2	67.7	72.4	95.1	**95.9**
ROC_{128}	94.2	86.9	86.4	81.0	99.5	**99.6**
$\mathrm{P/R}_{32}$	79.3	61.3	61.8	65.6	90.4	**93.5**
$\mathrm{P/R}_{64}$	92.3	64.6	62.5	66.2	96.1	**96.8**
$\mathrm{P/R}_{128}$	96.0	86.8	82.4	79.2	99.5	**99.6**

$$d(\mathbf{M}) = \frac{\max(\mathbf{M}) - \mu_{\mathbf{M}}}{\sqrt{\frac{\sum_{x \in I}(x - \mu_{\mathbf{M}})^2}{\max(|\mathbf{M}|)}}},$$

Equation 1. Detection Function



(a) Detection P/R curve

(b) Detection ROC curve

(c) Localization P/R curve

(d) Localization ROC curve

Figure 82. P/R and ROC Curves for the Anomaly Detection and Localization Tasks

[1] S. Kalyan Yarlagadda, D. Guera, P. Bestagini, F. Zhu, ¨ S. Tubaro, and E. Delp, "Satellite image forgery detection and localization using gan and one-class classifier," *Proceedings of the*

IS&T International Symposium on Electronic Imaging, vol. 2018, pp. 214–1–214–9, Feb. 2018. Burlingame, CA.

[2] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S. A. Siddiqui, A. Binder, E. Muller, and M. Kloft, “Deep one-class ¨ classification,” *Proceedings of the International Conference on Machine Learning*

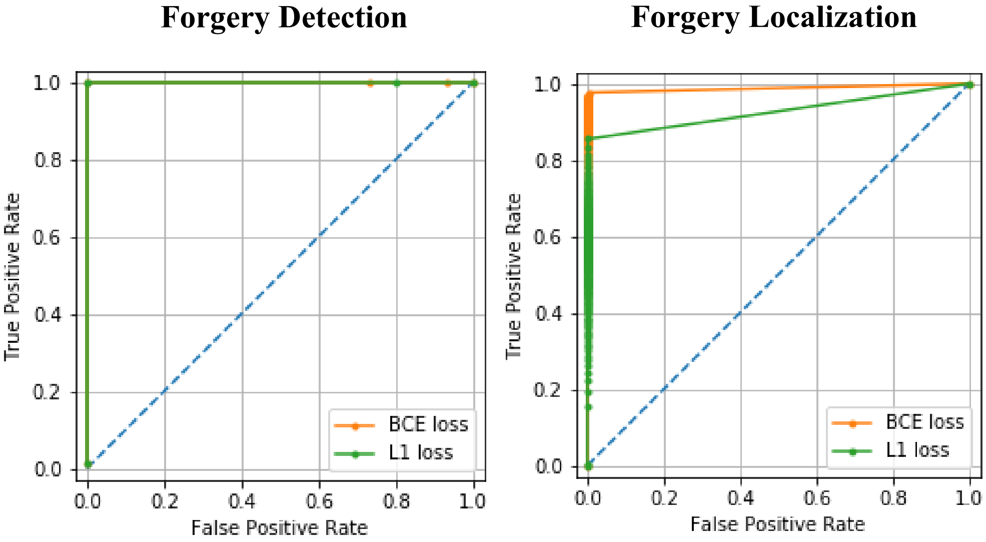
Published Paper:
J. Horváth, D. Güera, S. K. Yarlagadda, P. Bestagini, F. Zhu, S. Tubaro, E. J. Delp, "Anomaly-based Manipulation Detection in Satellite Images", *IEEE Conference on Computer Vision and Pattern Recognition*, Workshop on Media Forensics (CVPRW), Long Beach, California, June 2019.

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Forensic Analysis of Satellite Imagery
Major Technical Approach: Supervised approach with cGAN

Program Objectives: Splicing Detection and Localization in Satellite Imagery

Results show that the developed method accomplishes both tampering detection and localization with incredibly high accuracy on the used dataset. The receiver operating characteristic (ROC) curves reveal the performance of our approach. The area under the curve (AUC) for the detection objective is 1.000, indicating that it is possible to achieve perfect detection accuracy with thresholding. The AUC for the localization objective is 0.988, indicating that localization results are also very good. This forensic image technique exploits a data driven approach, learning to distinguish forged regions from pristine ones directly from the available training data.


Forgery Detection

Forgery Localization

Figure 83. ROC Curves for Detection and Localization of Spliced Forgeries

Table 52. Metrics for Detection and Localization of Spliced Forgeries

Metric	Forgery Detection	Forgery Localization
ROC AUC	1.000	0.988
Average Precision	1.000	0.953

Published Paper:
E. R. Bartusiak, S. K. Yarlagadda, D. Güera, F. M. Zhu, P. Bestagini, S. Tubaro, E. J. Delp, "Splicing Detection and Localization in Satellite Imagery using Conditional GANs", *IEEE International Conference on Multimedia Information Processing and Retrieval (MIPR),* March 2019, San Jose, CA

Published Thesis:
E. R. Bartusiak, "An Adversarial Approach to Sliced Forgery Detection and Localization in Satellite Imagery*", Master's dissertation, Purdue University*, West Lafayette, IN, May 2019.

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Unsupervised splicing detection, localization in satellite images
Major Technical Approach: Deep Belief Networks with Uniform-Uniform RBMs

Program Objectives: We evaluate our method as a splicing detection and localization method with the dataset presented in in the publication found at the end of the document. We also evaluate multiple configurations of our network in a one-class classification framework providing competitive results compared to the common one-class classification methods. We also examine the proposed method of the one-class classification task with the MNIST dataset that contains images of handwritten digits from 0-9.

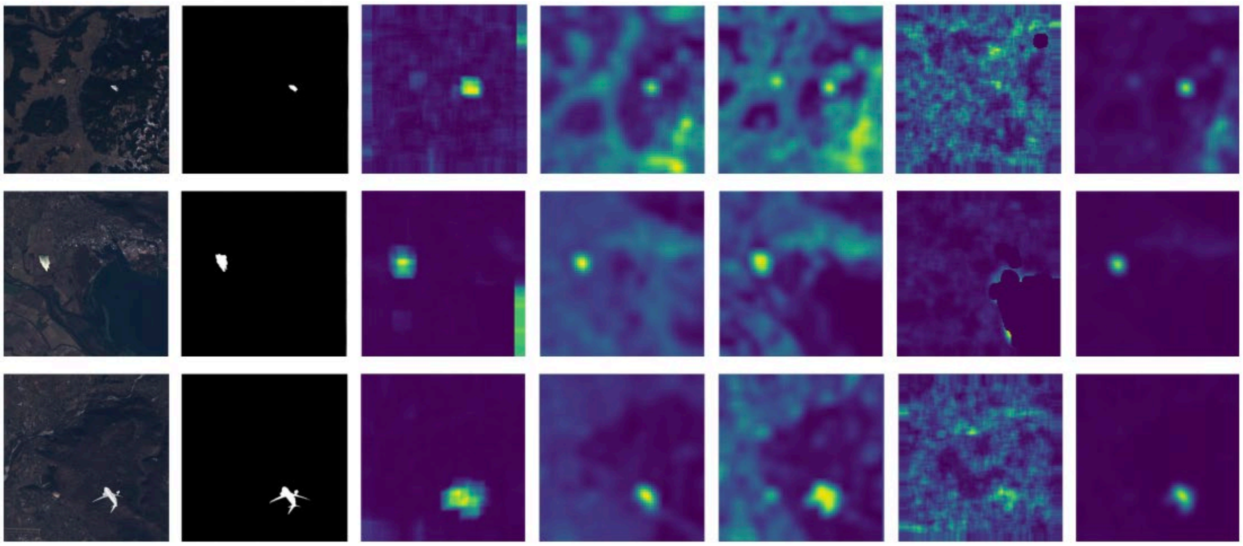
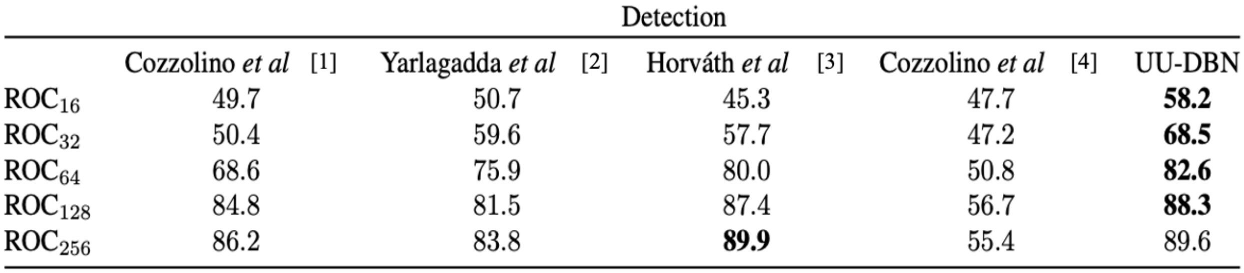


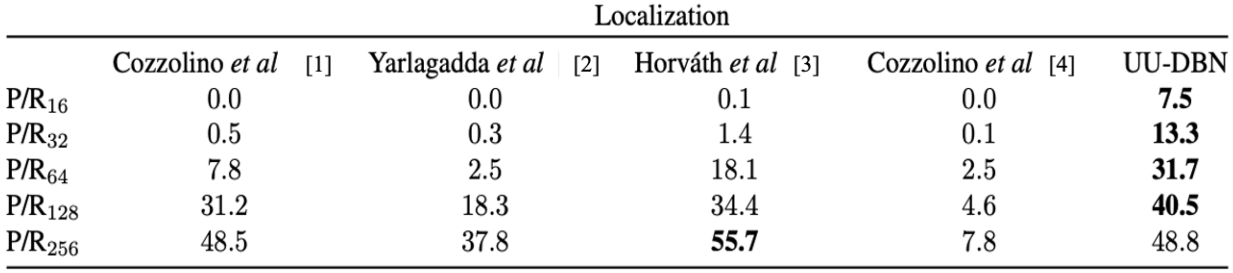
Figure 84. From Rows Left to Right: Input Images, Ground Truth Masks, Cozzolino et al [1], Yarlagadda et al [3], Horvath ‘ et al [4], Cozzolino et al [2] and UU-DBN

Table 53. AUC Scores (%) for the Detection and Localization Task (ROC and P/R Metrics). The Subscript Denotes the Manipulation Size. Best Performing Methods Are Bold



Detection

	Cozzolino *et al* [1]	Yarlagadda *et al* [2]	Horváth *et al* [3]	Cozzolino *et al* [4]	UU-DBN
ROC_{16}	49.7	50.7	45.3	47.7	**58.2**
ROC_{32}	50.4	59.6	57.7	47.2	**68.5**
ROC_{64}	68.6	75.9	80.0	50.8	**82.6**
ROC_{128}	84.8	81.5	87.4	56.7	**88.3**
ROC_{256}	86.2	83.8	**89.9**	55.4	89.6



Localization

	Cozzolino *et al* [1]	Yarlagadda *et al* [2]	Horváth *et al* [3]	Cozzolino *et al* [4]	UU-DBN
P/R_{16}	0.0	0.0	0.1	0.0	**7.5**
P/R_{32}	0.5	0.3	1.4	0.1	**13.3**
P/R_{64}	7.8	2.5	18.1	2.5	**31.7**
P/R_{128}	31.2	18.3	34.4	4.6	**40.5**
P/R_{256}	48.5	37.8	**55.7**	7.8	48.8

[1] D. Cozzolino, G. Poggi, and L. Verdoliva, “Splicebuster: A new blind image splicing detector,” *Proceedings of the IEEE International Workshop on Information Forensics and Security*, pp. 1–6, November 2015, Rome, Italy.

[2] D. Cozzolino and L. Verdoliva, “Noiseprint: A cnn-based camera model fingerprint,” *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 144–159, 2020.

[3] S. Kalyan Yarlagadda, D. Guera, P. Bestagini, F. Zhu, ¨ S. Tubaro, and E. Delp, “Satellite image forgery detection and localization using gan and one-class classifier,” *Proceedings of the IS&T International Symposium on Electronic Imaging*, vol. 2018, no. 7, pp. 214–1–214–9, February 2018, Burlingame, CA.

[4] J. Horvath, D. Guera, S. Kalyan Yarlagadda, P. Bestagini, F. Maggie Zhu, S. Tubaro, and E. J. Delp, “Anomaly-based manipulation detection in satellite images,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 62–71, June 2019, Long Beach, CA.

Published Paper:
J. Horváth, D. Mas Montserrat, H. Hao, E. Delp, "Manipulation Detection in Satellite Images Using Deep Belief Networks", *IEEE Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, Washington, June 2020. https://arxiv.org/abs/2004.12441

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Unsupervised splicing detection, localization in satellite images
Major Technical Approach: Gated PixelCNN

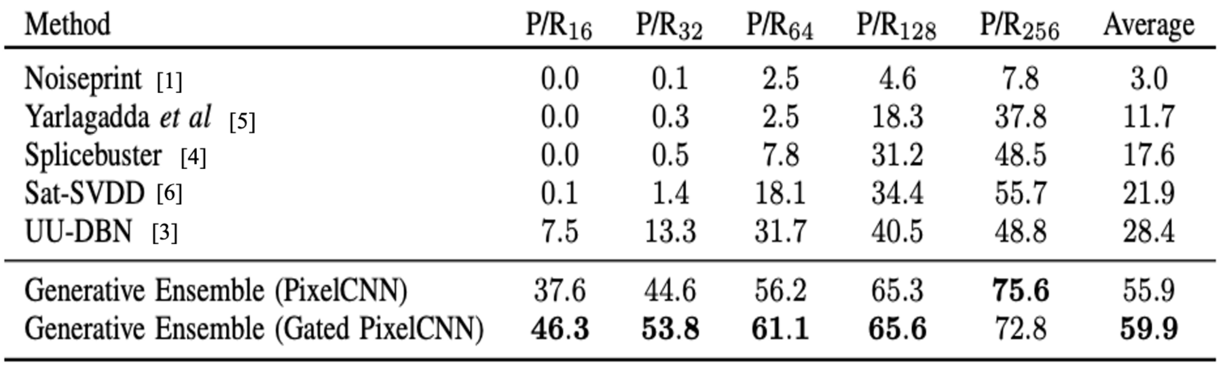
Program Objectives: Applying Gated PixelCNN to detect spliced objects from unknown origin in overhead images.

The ensemble of networks is trained only with original images and no manipulated images are used during the training process. In order to evaluate the localization performance of the presented method, we compute the area under the curve (AUC) of the Precision/Recall (P/R) curves by changing the threshold T applied to the estimated negative loglikelihood I(xi).

Our experimental results show that the generative ensemble of PixelCNNs and Gated PixelCNNs outperform previously presented methods. While most of the methods fail to detect objects smaller than 64 × 64, the presented generative ensembles are able to properly detect small forgeries. While methods such as [5], [6], and [3] produce estimates within patches of the input image, and therefore lacking enough resolution to detect small forgeries, PixelCNN and Gated PixelCNN process the whole image in a fully-convolutional manner and detects manipulations in a pixel-level providing higher detection accuracy.

We can observe that Gated PixelCNN provides more accurate results than the regular PixelCNN network. These results are aligned with previous works [2] which have shown that Gated PixelCNN is able to model the image distribution of the training images more accurately (with a lower negative log-likelihood score) than PixelCNN.

Table 54. AUC Scores (%) of the P/R Curves for the Localization Task. The Subscript (P/Rx) Denotes the Manipulation Size



Method	P/R_{16}	P/R_{32}	P/R_{64}	P/R_{128}	P/R_{256}	Average
Noiseprint [1]	0.0	0.1	2.5	4.6	7.8	3.0
Yarlagadda *et al* [5]	0.0	0.3	2.5	18.3	37.8	11.7
Splicebuster [4]	0.0	0.5	7.8	31.2	48.5	17.6
Sat-SVDD [6]	0.1	1.4	18.1	34.4	55.7	21.9
UU-DBN [3]	7.5	13.3	31.7	40.5	48.8	28.4
Generative Ensemble (PixelCNN)	37.6	44.6	56.2	65.3	**75.6**	55.9
Generative Ensemble (Gated PixelCNN)	**46.3**	**53.8**	**61.1**	**65.6**	72.8	**59.9**

[1] D. Cozzolino and L. Verdoliva, “Noiseprint: A cnn-based camera model fingerprint,” *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 144–159, 2020.

[2] A. Van den Oord, N. Kalchbrenner, L. Espeholt, O. Vinyals, A. Graves, et al., “Conditional image generation with pixelcnn decoders,” *Proceedings of the Advances in Neural Information Processing Systems*, pp. 4790–4798, December 2016, Barcelona, Spain.
[3] J. Horvath, D. M. Montserrat, H. Hao, and E. J. Delp, “Manipulation detection in satellite images using deep belief networks,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, June 2020, Seattle, WA.
[4] D. Cozzolino, G. Poggi, and L. Verdoliva, “Splicebuster: A new blind image splicing detector,” *Proceedings of the IEEE International Workshop on Information Forensics and Security*, pp. 1–6, November 2015, Rome, Italy.
[5] S. Kalyan Yarlagadda, D. Guera, P. Bestagini, F. Zhu, S. Tubaro, and ¨ E. Delp, “Satellite image forgery detection and localization using gan and one-class classifier,” *Proceedings of the IS&T International Symposium on Electronic Imaging*, vol. 2018, no. 7, pp. 214–1–214–9, February 2018, Burlingame, CA.
[6] J. Horvath, D. Guera, S. Kalyan Yarlagadda, P. Bestagini, F. Maggie Zhu, S. Tubaro, and E. J. Delp, “Anomaly-based manipulation detection in satellite images,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 62–71, June 2019, Long Beach, CA.

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Unsupervised satellite image forgery detection and localization
Major Technical Approach: Generative adversarial network (GAN) and one-class support vector machine (SVM)

Program Objectives: Tests on copy-paste attacked images with different forgery size show promising accuracy in both detection and localization.

We proposed a solution for satellite imagery forgery detection and localization. The rationale behind the proposed method is that it is possible to train an autoencoder to obtain a compact representation of image patches coming from pristine satellite pictures. This autoencoder can than be used as a feature extractor for image patches. During testing, a one-class SVM is used to detect whether feature vectors come from pristine images or not, thus representing forgeries. The solution proposed in this work makes use of generative adversarial networks to train the autoencoder for the forgery detection task. Moreover, it is worth noting that the whole system is trained only on pristine data. This means that no prior knowledge on the forgeries is assumed to be available.

Table 55. Detection Results in Terms of AUC for the Different Datasets. AUCs are Reported in Two Different Cases: Autoencoder Trained with or without the GAN. Best Results are Reported in Italics

Forgery Size	AUC (without Gan)	AUC (with GAN)	AUC Difference
Small	0.784	*0.797*	+0.013
Medium	0.904	*0.920*	+0.016
Large	0.950	*0.972*	+0.022

Table 56. Localization Results in Terms of AUC for the Different Datasets. AUCs Are Reported in Two Different Cases: Autoencoder Trained with or without the GAN. Best Results Are Reported in Italics

Forgery Size	AUC (without Gan)	AUC (with GAN)	AUC Difference
Small	*0.913*	0.902	-0.009
Medium	*0.963*	0.961	-0.002
Large	*0.970*	0.974	-0.004

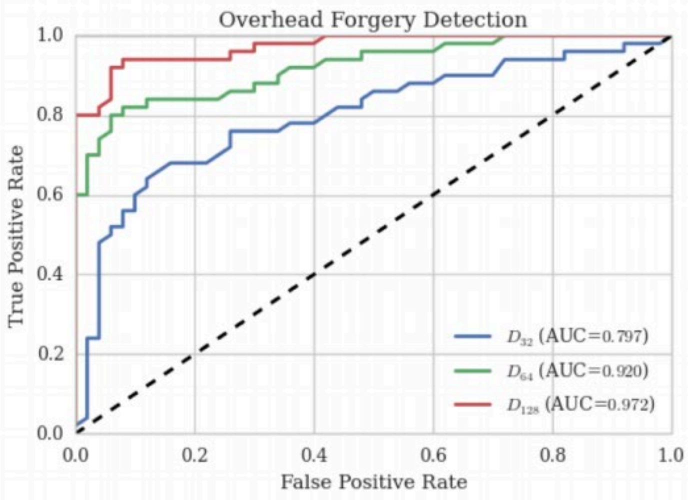


Figure 85. Forgery Detection ROC Curves. Each Curve Represents Results on a Different Dataset According to the Forgery Average Size

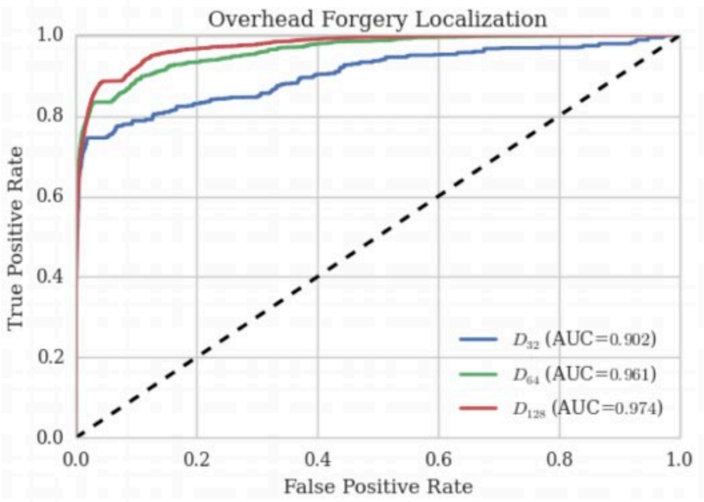


Figure 86. Forgery Localization ROC Curves. Each Curve Represents Results on a Different Dataset According to the Forgery Average Size

Published Paper:
S. K. Yarlagadda, D. Güera, P. Bestagini, F. Zhu, S. Tubaro, E. J. Delp, "Satellite Image Forgery Detection and Localization Using GAN and One-Class Classifier", *IS&T Electronic Imaging (EI),* vol. 2018, no. 7, pp. 214-1-214-9, Burlingame, California, January 2018.
DOI: 10.2352/ISSN.2470-1173.2018.07.MWSF-214

4.2.2 University of Siena

Contract Information:
Team: Purdue Team, Siena Subteam
POC: Prof. Mauro Barni, University of Siena, barni@diism.unisi.it

Technical Problem: Copy-move localization with source-target disambiguation
Technical Approach: Disambiguation based on multiple-branch CNNs, geometric transformation estimation

Objectives: Achieving good disambiguation capabilities of the CM forgeries. The Accuracy of the disambiguation, computed as the ratio of correctly disambiguated copy-moves over the total number of tested images, is considered as evaluation metric.

To address this objective, we propose a multi-branch CNN architecture (DisTool) consisting of two main parallel branches, looking for two different kinds of CM-traces. The first branch, named 4-Twins Net, consists of two parallel Siamese networks, trained in such a way to exploit the non-invertibility of the copy-move process caused by the interpolation artifacts often associated to the copy-move operation. The second branch is a Siamese network designed to identify artifacts and inconsistencies present at the boundary of the copy-moved region. The disambiguation capability of DisTool are assessed by considering the case where both the binary localization mask (ground-truth localization mask) and the transformation are given. Results are shown in Table 57. Performances are reported on SYN-Ts, a synthetic dataset of CM forgeries we built ourselves (starting from images of RAISE, Dresden and Vision), with the same characteristic of the one considered for training, and the USCISI dataset, which is another dataset for which the information on binary localization mask and CM transformation are available.

Table 57. Accuracy (%) of 4-Twins Net, Siamese Net and DisTool (fusion)

Dataset	#Imgs	4-Twins Net	Siamese Net	DisTool
SYN-Ts-Rigid	1000	47.06	97.00	97.00
SYN-Ts-Rot	1000	99.40	93.10	99.50
SYN-Ts-Res	1000	99.30	95.60	99.90
USCISI	9984	94.48	30.51	91.40

Table 58 reports the performance of an end-to-end system performing simultaneously copy-move localization and source-target disambiguation by means of DisTool. For copy-move localization, we considered the patch-based algorithm in [1] (Dense Field Copy-Move Forgery Detection, DF-CMFD), which works reasonably well under general conditions, with the exception of the USCISI dataset, where this method works poorly, and we used the BusterNet-CMFD CNN-based method. The performance of DisTool are tested and compared to the BusterNet disambiguation method on SYN-Ts, USCISI and other publicly available datasets of realistic CMs. The OptIn images are the images for which the two duplicated regions can be correctly identified after the application the CM localization algorithm and the preprocessing, that can be different for the two methods.

Table 58. Accuracy (%) of DisTool for the end-to-end system

Dataset	# Imgs	End-to-end (DF-CMFD / BusterNet + DisTool)			BusterNet	
		CM detector	OptIn	OptIn Accuracy	OptInB	OptInB Accuracy
SYN-Ts-Rigid	1000	DF-CMFD	992	94.86	143	80.99
SYN-Ts-Rot	1000	DF-CMFD	962	98.44	33	75.76
SYN-Ts-Res	1000	DF-CMFD	956	96.75	146	86.99
CASIA	1276	DF-CMFD	482	74.04	688	52.18
Grip	80	DF-CMFD	75	74.67	21	28.57
USCISI	9984	BusterNet-CMFD	5531	78.61	5051	85.57

To assess the performance of DisTool on satellite images, we generated a large scale datasets of synthetic CM forged satellite images starting from Sentinel-2 images from the EOlearn EOPatches Slovenia 2017 database, consisting of 293 images, of size 1000x1000, taken from Sentinel-2, with resolution 10 meters. The CM forgeries were produced by considering different settings, that is, several sizes for the duplicated regions, different shapes of the cloned area, different applied transformations (rotation, resizing, and both) and different post-processing applied to the copied /target area, to mimic a realistic case of CM forgery.

In this case, given the particular nature of satellite images and their textural content, CM detection and localization is performed using the algorithm based on SIFT keypoints matching. In particular, a refined version of the al SIFT-based CM detector in [2] has been considered.

Our tests revealed that, as expected, the 4-Twins works better in the case of general geometric transformation (RST case), while the Siamese network works better in the rigid translation case (T case), since in this case the boundary of the two regions can be easily - almost perfectly -

matched, via translation. However, as argued, the base Siamese network achieves poor performance, with a disambiguation accuracy in the range 58%-62%, when the size of the copy-moved area is large (> 250x250).

By applying the proposed pre-processing to the input regions, the Siamese network works much better both in the RST and T case, with both large and small CM. Specifically, in the T case, with the new SiameseNet branch that includes pre-processing, the disambiguation accuracy of the final fusion module (DisTool) is + 20% higher with large CMs (250x250), 10-15% higher with smaller CMs (150x150).

Table 59. Accuracy (%) of DisTool on Sentinel-2 CM images (with and w/o pre-processing)

	RST case		T case	
	150x150	250x250	150x150	250x250
4-TwinsNet	87,8%	89,5%	51,6%	55,0%
SiameseNet	70,6%	58,2%	75,1%	61,9%
SiameseNet (with pre-processing)	**86,7%**	**84,0%**	**85,9%**	**85,2%**
Fusion	90,6%	83,2%	71,4%	62,5%
Fusion (Siamese with pre–processing)	**91,1%**	**89,1%**	**83,5%**	**83,0%**

The improvement brought by the pre-processing in the fusion of DisTool have also been confirmed on a smaller set of realistic handmade CM forgeries generated starting from different sources, of satellite and overhead images.

[1] D.Cozzolino, G. Poggi, and L. Verdoliva, “Efficient dense-field copy–move forgery detection,” IEEE T-IFS 2015

[2] I Amerini, L Ballan, R Caldelli, A Del Bimbo, L Del Tongo, G Serra, “Copy-move forgery detection and localization by means of robust clustering with J-Linkage”, Signal Processing: Image Communication Vol.28, No.6, 2013

Published Paper:
M. Barni, Q. -T. Phan, B. Tondi, "Copy Move Source-Target Disambiguation Through Multi-Branch CNNs," *IEEE Transactions on Information Forensics and Security (TIFS), vol. 16, pp. 1825-1840, December 2020.* DOI: 10.1109/TIFS.2020.3045903

Contract Information:
Team: Purdue Team, Siena Subteam
POC: Prof. Mauro Barni, University of Siena, barni@diism.unisi.it

Technical Problem: Generation of multispectral images using GANs
Technical Approach: Generation using progressive GAN and style transfer using NICEGAN, CycleGAN and pix2pix.
Objectives: Generate multispectral images with good quality using GAN architectures. The image assessment was done using qualitative measures and also by comparing spectrum views. For land cover translation task, land cover classification was used for assessment.

To assess the images generated from scratch using progressive GAN we relied on visual qualities and also compared the spectral correlation between bands in both pristine and GAN generated images. Figure 87 shows an example of the spectrum view of some pixels for pristine and generated images.

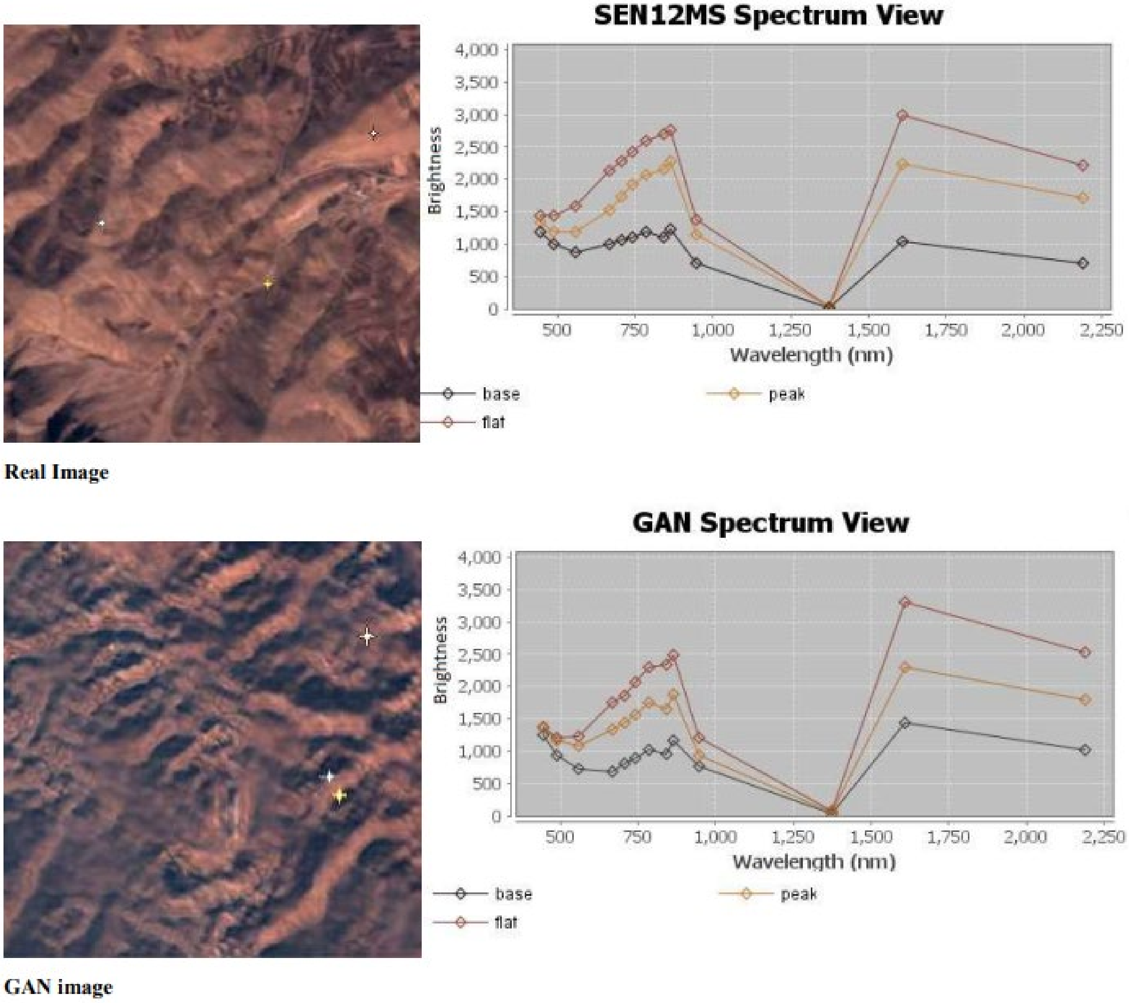


Figure 87. Similar spectral correlation between Pristine and Generated

Whereas for the land cover transfer task using NICEGAN, we relied on land cover classification using LDA as a classifier on 2000 images per each domain of our land cover test dataset. Table 60 shows the results of NDVI, Image classification and pixel classification of pristine and generated images. The results emphasize a similar land cover classification for images from the same domain. RGB representation of some examples of the land cover transfer using NICEGAN are shown in Figure 88.

Table 60. NDVI and classification of land cover domains for pristine images and images generated by NICEGAN

Metric	Real Vegetation	GAN Vegetation	Real Desert	GAN Desert
Average NDVI	0.6816	0.4311	0.1025	0.1058
Average Vegetation Image Classification	75.3877	84.25	0	0.6
Average Vegetation Pixel Classification	72.1239	67.6068	0.0508	5.9074
Average Desert Image Classification	0.9505	0.45	99.95	85.35
Average Desert Pixel Classification	0.9621	0.6203	99.5248	82.1605

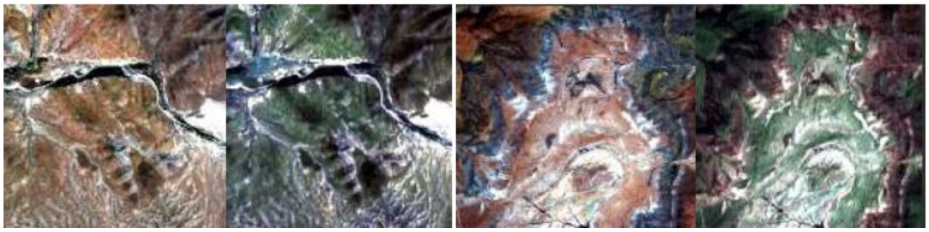


Figure 88. Sample images of land cover transfer from barren to vegetation using NICEGAN

For the land cover transfer task using CycleGAN, we applied classification based on NDVI values specifically, pixels of which the NDVI is lower than -0.1 are classified as water pixels and as barren when NDVI ∈ [−0.1, 0.1], low vegetation when NDVI ∈ [0.1, 0.4], and high vegetation when NDVI is larger than 0.4. The pixel classification is obtained on 2000 image per each class. In Table 61, we report the results whereas in Figure 89 we show RGB representation of the land cover transfer using CycleGAN. For season transfer task, we assessed the images on qualitatively so in Figure 90 we show an example of an RGB representation of season transfer using pix2pix.

Table 61. Classification of land cover for pristine image and images generated by GAN

	High Vegetation	Low Vegetation	Barren	Water
Real Vegetation	77.87%	21.81%	0.29%	0.01%
GAN Vegetation	91.80%	6.70%	1.10%	0.30%
Real Barren	0%	0%	100%	0%
GAN Barren	0%	2%	97.54%	0.46%

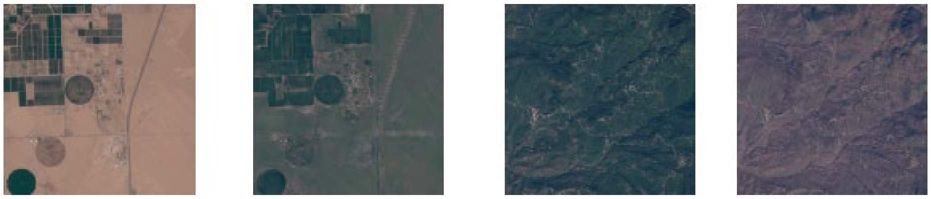


Figure 89. Sample images of land cover transfer using CycleGAN

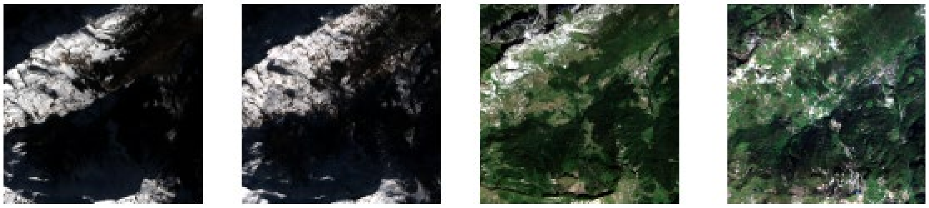


Figure 90. Sample of season transfer using pix2pix

Published Papers:
L. Abady, M. Barni, A. Garzelli, B. Tondi, "GAN Generation of Synthetic Multispectral Satellite Images," *SPIE Image and Signal Processing for Remote Sensing XXVI, vol. 11533, pp. 122-133, Online, September 2020.* DOI: 10.1117/12.2575765

L. Abady, J. Horváth, B. Tondi, E. J. Delp, M. Barni, "Manipulation and Generation of Synthetic Satellite Images Using Deep Learning Models", *SPIE Journal of Applied Remote Sensing* (submitted)

4.2.3 Politecnico di Milano

Contract Information:
Team: Purdue Team, PoliMi Sub-Team
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: SAR image forgery detection and localization
Major Technical Approach: CNN and clustering-based approaches

Program Objectives: Development of SAR forensic techniques for splicing detection and localization

The network for fingerprint extraction has been tested on a set of images under different circumstances: only copy-paste; copy-paste plus resizing; copy-paste plus rotation; copy-paste plus resizing and rotation. Different patch sizes have been used for the forgeries. Results show that copy-paste can be exposed with AUC=0.7 in the most difficult conditions, and AUC>0.95 if resizing or rotation are also applied. These results are obtained considering that the binary mask is extracted through simple thresholding, hence much better performance can be obtained by optimizing the second step of the pipeline. Indeed, Figure 91 shows some examples of masks obtained using the three different proposed solutions.

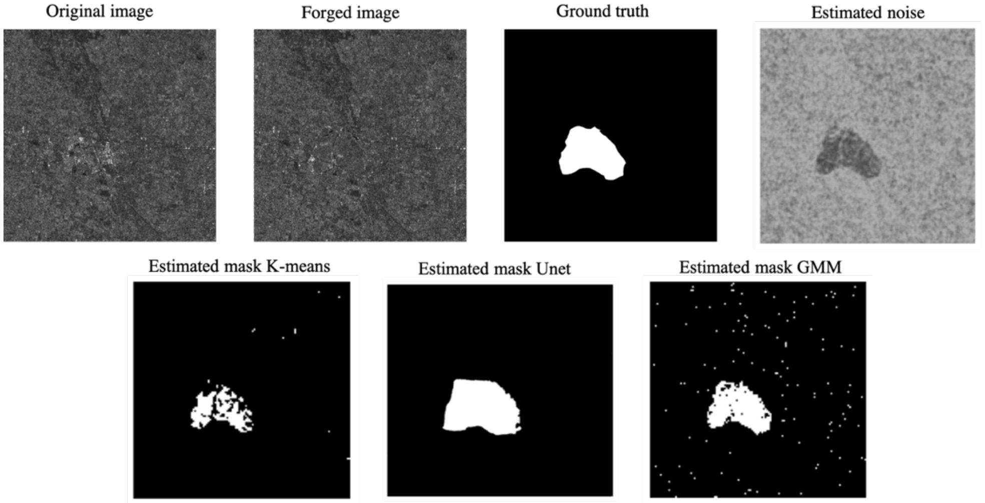


Figure 91. Example of the SAR Forensic Technique

Published Paper:
E. D. Cannas, N. Bonettini, S. Mandelli, P. Bestagini, S. Tubaro, "Amplitude SAR Imagery Splicing Localization**,"** *IEEE Access, vol. 10, pp. 33882-33899, March 2022.* DOI: 10.1109/ACCESS.2022.3161836

Contract Information:
Team: Purdue Team, PoliMi Subteam
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: Panchromatic image source attribution
Major Technical Approach: CNN and uncertainty-analysis approach

Program Objectives: Development of panchromatic image source attribution techniques in closed-set and open-set

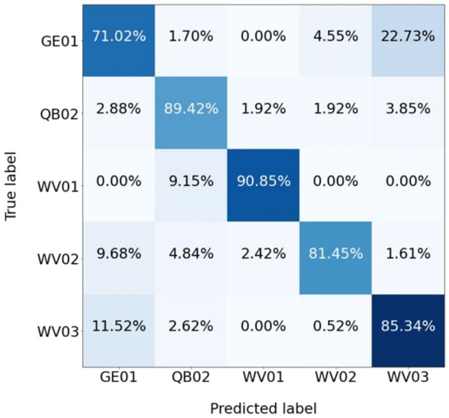
In order to test the effectiveness of the investigated pipeline, we created a dataset of satellite images. We collected 8-bit panchromatic images from the DigitalGlobe portal coming from 5 different satellites: GeoEye (GE01), QuickBird (QB02), WorldView 1, 2 and 3 (WV01-02-03). All images are provided in GeoTIFF format, with no compression, in non-overlapping tiles of 16,384 x 16,384 pixels. All tiles are orthorectified, sensor and radiometrically corrected. To avoid a possible bias due to the semantics of the scene represented in the images, we have selected different geographical regions, so that for each satellite we have samples coming from urban, snowy, barren, forest and field areas. However, as these images were too large to be processed by our baseline network, we performed a patch extraction procedure on each sample. Specifically, we extracted pixel patches of size 1024 x 1024, ending up with a dataset of roughly

3000 images equally balanced in number among all satellite classes and in the geographical areas represented.

The baseline network we considered in our setup is an EfficientNetB0 that we trained as an *M*-class classifier. The network is trained from scratch, as most available pre-trained models work on natural images, therefore having a strong mismatch with respect to the panchromatic imagery we are analyzing. For closed-set experiments, we consider *M*=5 (i.e., all available satellites in the dataset) and we follow a train-validation-test scheme, where we use roughly 50% of samples for training, 25% for validation, and the remaining 25% for testing. For the open-set experiments, we consider *M*=4 (i.e., we leave one satellite out of the training set) and we add to the test set all images coming from the unknown satellite (i.e., the one left out during training). Open-set experiments are repeated five times following a leave-one-out procedure (i.e., each satellites is considered as unknown in one experiment).

For MCD we adopted a single seed for random initializing the model. For DE, we trained 10 networks using 10 different seeds increasing in value computed with a fixed step of 10. Finally, regarding CNN hyperparameters, all networks have been trained using batches of 5 images and a simple cross-entropy loss for 200 epochs, relying on Adam (optimization algorithm) for optimization. We started with a learning rate of 0.001 reduced on a plateau of the validation loss after 10 consecutive epochs, and early stopping of the training if the validation loss did not improve after 50 consecutive epochs or if the learning rate reached a minimum of 10^-8.

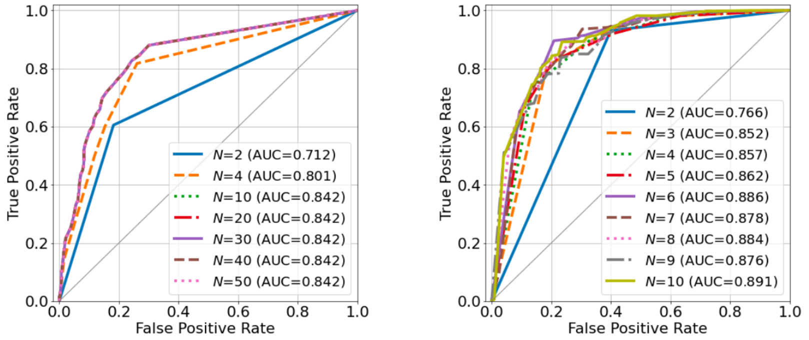
The first experiment is performed to evaluate whether the selected backbone network can solve the source attribution problem in closed-set. To this purpose, we trained the EfficientNetB0 considering all the available satellites (i.e., *M*=5). Figure 92 reports the obtained confusion matrix. These results show that it is possible to exploit EfficientNetB0 for this task, however also highlighting that the problem is more challenging with respect to classic camera attribution.



True label \ Predicted label	GE01	QB02	WV01	WV02	WV03
GE01	71.02%	1.70%	0.00%	4.55%	22.73%
QB02	2.88%	89.42%	1.92%	1.92%	3.85%
WV01	0.00%	9.15%	90.85%	0.00%	0.00%
WV02	9.68%	4.84%	2.42%	81.45%	1.61%
WV03	11.52%	2.62%	0.00%	0.52%	85.34%

Figure 92. Confusion matrix for a baseline EfficientNetB0 trained with all satellite classes available in closed set (accuracy of 83%)

For the open-set classification task, our efforts initially focused in determining an appropriate number of networks for both MCD and DE. To this purpose, Figure 93 shows the ROC curves obtained using MCD and DE with a different N in the case EfficientNetB0 is trained over four satellites and WV03 samples are used as unknown class. It is possible to observe that the AUC increases with N, but has diminishing returns after N=10.

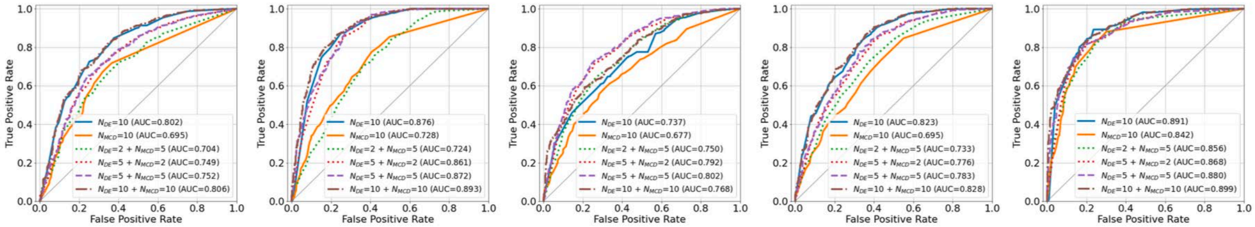


(a) MCD ensembles with $N \in \{2, 4, 10, 20, 40, 50\}$.

(b) DE ensembles with $N \in \{2, \ldots, 10\}$.

Figure 93. ROC curves obtained using MCD or DE ensembles with different N values

The next experiment we performed aimed at evaluating DE, MCD and their combination considering all five leave-one-class out scenarios (i.e., four satellites in the training set, and one left out as unknown). Figure 94 shows these open-set classification performances. In addition to the DE and MCD with N=10, we considered two equivalent combinations of 10 networks obtained with 5 and 2 MCD inferences of a DE of 2 and 5 networks respectively, and an ensemble of 25 and 100 networks considering a combination of 5 and 10 MCD inferences with a DE of 5 and 10 networks respectively. In all these experiments we use N_{DE} and N_{MCD} to distinguish the N value used for DE and MCD.



(a) Training without GE01 samples.

(b) Training without QB02 samples.

(c) Training without WV01 samples.

(d) Training without WV02 samples.

(e) Training without WV03 samples.

Figure 94. Open-set results for the leave-one-class experiments. N_{DE} and N_{MCD} denote N values for DE and MCD whenever they are combined

Published Paper:
E.D. Cannas, S. Baireddy, E. R. Bartusiak, S.K. Yarlagadda, D. Mas Montserrat, P. Bestagini, S. Tubaro, E. J. Delp, “Open-Set Source Attribution for Panchromatic Satellite Imagery”, *IEEE International Conference on Image Processing, pp. 3038-3042, Anchorage, AK, September 2021.*
DOI: 10.1109/ICIP42928.2021.9506600

4.3 Scientific Integrity Project

In this section we present further discussion of the Scientific Integrity project.

4.3.1 Purdue University

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: PDF Image Extraction
Major Technical Approach: Image Extraction for Scientific Integrity

Program Objectives: efficiently and automatically extract images from PDF files

To evaluate the quality of the extracted images, we compare these images with the original images from the publisher's website. This evaluation aims to measure the number of failure cases when extracting the vector images. A similarity score is determined for each pair of extracted and original images. If the similarity score is above a threshold, the extracted image will be labeled as "matched", i.e. the target image is extracted correctly. Otherwise, the extracted image is mismatched, since it is a sub-image of the original image. We use the complex wavelet structural similarity index (CW- SSIM) to evaluate the extracted image quality, since this type of index is less sensitive to the image translation, scaling and rotation:

$$\tilde{S}(c_x, c_y\ ) = \frac{2\left|\sum_{i=1}^{N} c_{x,i}\, c_{y,i}^{*}\right| + K}{\sum_{i=1}^{N}\left|c_{x,i}\right|^2 + \sum_{i=1}^{N}\left|c_{y,i}\right|^2 + K}$$

Equation 2. CW-SSIM Equation

Here c are local coefficients extracted from the complex wavelet transformation of the images, respectively, c^* denotes the complex conjugate of c and K is a small positive constant. The subscripts x and y indicate for two compared images. The purpose of K is mainly to improve the robustness of the CW-SSIM measure when the local signal-to-noise ratios are low.
Based on this experiment, the empirical matching threshold is 0.65. When we test the proposed system on dataset version 4, 2,493 images are extracted correctly with total 3,186 original images. The extraction rate is 78.25%.

Contract Information:
Team: Purdue Team, Purdue Sub-Team
POC: Professor Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740

Major Technical Problem: Overlaid Text Content on Image
Major Technical Approach: Image Segmentation for Scientific Integrity

Program Objectives: provide a mask to indicate the overlaid text region on the image

We manually annotate 563 figure images from our retracted papers dataset. Text areas were removed based on the annotations. We also generate 1000 synthetic images where text instances of different size, color and font were added on top of the clean figures. Then the detection results are evaluated by comparing the detection mask with ground truth mask pixel-wise directly over the widely used mean average precision metric for object detection. The similarity metrics we adapted here are Intersection over Union (IoU), Dice Co-efficient, Cosine Coefficient and Precision/Recall.

Table 62. Evaluation Results for Two Type of Datasets. The Scores Reported are Averaged within the Dataset

Dataset Type	IoU score	Dice score	Cosine score	Precision	Recall
563 scientific figures	0.8883	0.9325	0.9361	0.9574	0.9227
1000 synthetic figures	0.6488	0.7676	0.7787	0.7499	0.8339

Contract Information:
Team: Purdue Team
POC: Prof. Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740
Prof. Walter Scheirer, University of Notre Dame, walter.scheirer@nd.edu, +1 574 631 2436
Prof. Paolo Bestagini, Politecnico di Milano, paolo.bestagini@polimi.it, +39 02 2399 3571
Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647
Prof. Wael Abd-Almageed, USC-ISI, wamageed@isi.edu, +1 703 248 6174

Major Technical Problem: Dataset for Published Scientific Papers
Major Technical Approach: Dataset Collection for Scientific Integrity
Program Objectives: construct a dataset for automatically or semi-automatically analyzing the reliability of the published scientific papers

The dataset contains both retracted/corrected papers and suspected papers which might contain scientific misconduct. All the suspected papers are collected because at least one of the authors have had one or more retracted papers. Most of the papers are in the life sciences area. They were retracted/corrected due to the duplication, manipulation, plagiarism, falsification/fabrication of images. The information for the papers in the database is obtained from Retraction Watch, a blog that reports on retractions of scientific papers.

The latest version of the dataset (version 7) contains 650 papers from 298 authors.

The dataset is organized by authors' name. Each author has two directories, one for the retracted/corrected papers and the other for suspected papers.

Each paper will have a directory which contains:

- Original paper
- Supplemental material
- Figure (sub-directory): Contains the source images of the original published paper and supplemental material. For corrected papers, the corrected images, if available, will also be included. These images are originally archived at the publisher's website. For each

image shown in original paper, here is a .txt file including the corresponding caption and legend.

- Retraction/Correction notice: Only retracted/corrected papers contain this document. The official letter indicates the reason of paper retraction/correction.
- Related paper (sub-directory): Only retracted/corrected papers contain this document. It contains the related plagiarized papers based on the retraction/correction notice.
- Ground truth(sub-directory): Only some retracted/corrected papers contain this directory. It contains the manually labeled ground truth of the manipulated, duplicated, falsified, fabricated images based on the information given by the retraction/correction notice.

The dataset also includes a ReadMe file and a spreadsheet. The ReadMe file describes the structure of the entire dataset and gives a description of each filename. The spreadsheet provides more detail information of the collected papers. In the spreadsheet, for each paper, we include this information: the paper title, name of the authors, publication year, publication source, reason for retraction/correction, and paper DOI. We also indicate the availability of the original paper, source images, supplemental images, corrected images, and plagiarized papers.

Contract Information:
Team: Purdue Team
POC: Prof. Edward J. Delp, Purdue University, ace@ecn.purdue.edu, +1 765 494 1740
Prof. Walter Scheirer, University of Notre Dame, walter.scheirer@nd.edu, +1 574 631 2436
Prof. Paolo Bestagini, Politecnico di Milano, paolo.bestagini@polimi.it, +39 02 2399 3571
Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647
Prof. Wael Abd-Almageed, USC-ISI, wamageed@isi.edu, +1 703 248 6174

Major Technical Problem: Scientific Integrity
Major Technical Approach: A GUI for Scientific Integrity System

Program Objectives: develop a system GUI including multiple image analyzing methods for scientific papers

We have developed an extensible, end-to-end cloud-based scientific integrity system that incorporates various elements needed by an analyst, including content extraction and segmentation, visual content ranking, manipulation detection and provenance analysis. The architecture of the system is shown in Figure 95. The current system can perform five main different tasks under the analyst's suspicions and desire. These tasks are represented as rounded rectangles lined-up from left to right in the middle of the diagram. They may be demanded by the analyst in an iterative way, as long as they obey the order defined by the numbers that precede their labels. Each task receives a particular input from the analyst and produces a result that must be presented to the expert through the system GUI, to help with their decision-making process.

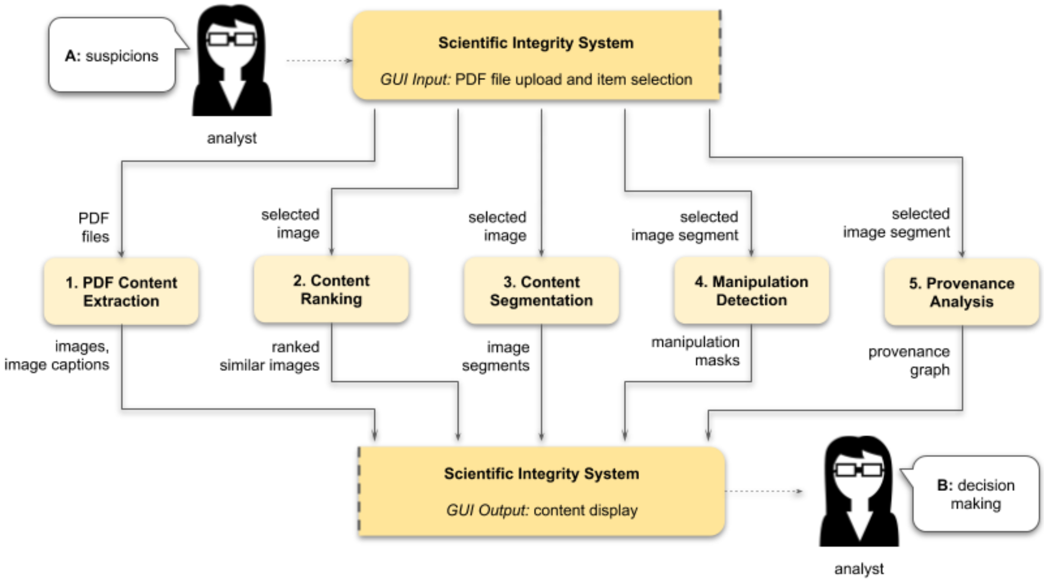
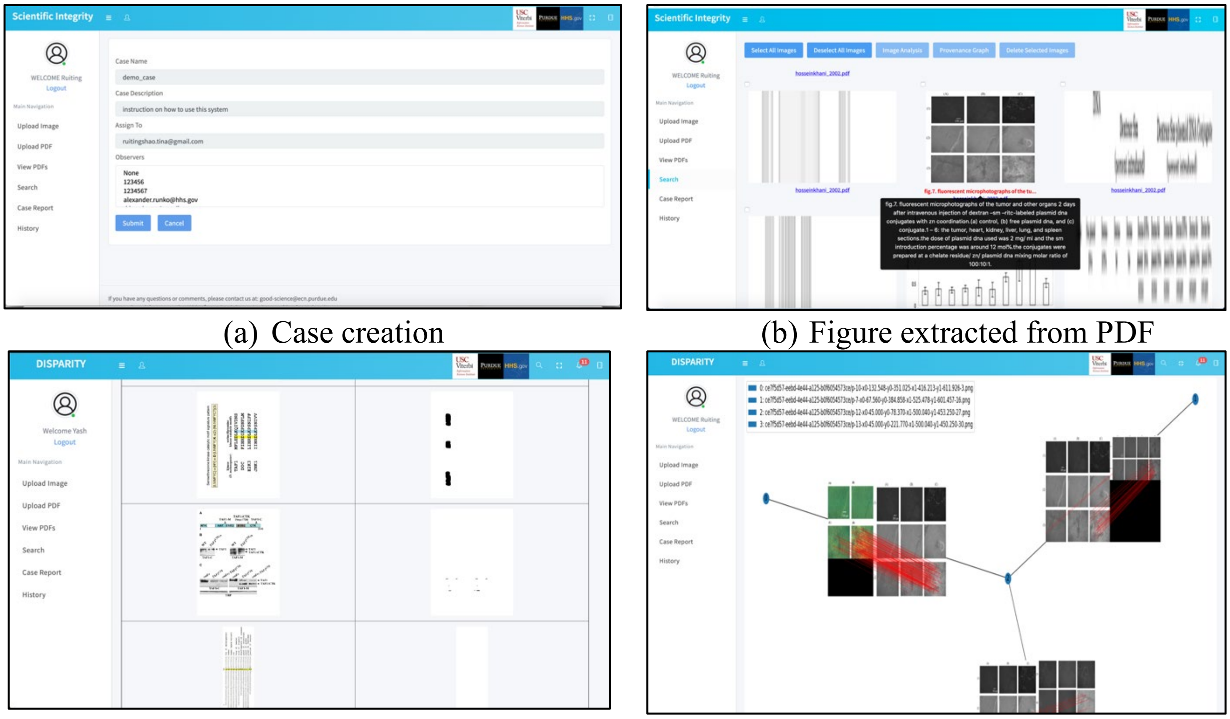


Figure 95. Proposed System Workflow

The system is developed as a web application that uses NGINX webserver to handle the JavaScript user interface. The user interface interacts with the backend via a RESTful application programming interface (API), implemented using Flask. We use RabbitMQ as brokerage middle ware to enable multiple users using the system to handle extensive workloads. And all integrated functionalities are implemented as Docker containers to make the system extensible.



Case Name
demo_case
Case Description
instruction on how to use this system
Assign To
ruitingshao.tina@gmail.com
Observers
None
123456
1234567
alexander.runko@hhs.gov

If you have any questions or comments, please contact us at: good-science@ecn.purdue.edu

fig.7. fluorescent microphotographs of the tumor and other organs 2 days after intravenous injection of dextran –sm –ritc-labeled plasmid dna conjugates with zn coordination.(a) control, (b) free plasmid dna, and (c) conjugate.1 – 6: the tumor, heart, kidney, liver, lung, and spleen sections.the dose of plasmid dna used was 2 mg/ ml and the sm introduction percentage was around 12 mol%.the conjugates were prepared at a chelate residue/ zn/ plasmid dna mixing molar ratio of 100:10:1.

0: ce7f5d57-eebd-4e44-a125-b0f6054573ce/p-10-x0-132.548-y0-351.025-x1-416.213-y1-611.926-3.png
1: ce7f5d57-eebd-4e44-a125-b0f6054573ce/p-7-x0-67.560-y0-384.858-x1-525.478-y1-601.457-16.png
2: ce7f5d57-eebd-4e44-a125-b0f6054573ce/p-12-x0-45.000-y0-78.370-x1-500.040-y1-453.250-27.png
3: ce7f5d57-eebd-4e44-a125-b0f6054573ce/p-13-x0-45.000-y0-221.770-x1-500.040-y1-450.250-30.png

(a) Case creation

(b) Figure extracted from PDF

(c) Example of copy-move detection result

(d) Example of provenance graph

Figure 96. Screenshots of the Scientific Integrity System

Published paper:
Paper is still in progress.

4.3.2 Politecnico di Milano

Contract Information:
Team: Purdue Team, PoliMi Sub-Team
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: Analysis of scientific images
Major Technical Approach: CNNs for PDF parsing, Fourier Mellin transform for blot copy-move detection, image enhancement operations.

Program Objectives: Development of a tool for scientific images manipulation detection

To be able to distinguish images from text starting from scanned versions of articles and papers, we developed a data driven approach. Specifically, we considered a series of different architectures proposed in the literature, namely dhSegment, Fast CNN and Pix-2-pix. Among the inspected architectures, we focused on pix-2-pix in particular. This method makes use of a generative adversarial network (GAN) to translate images from one domain (i.e., scanned documents in our case) to another domain (i.e., labeled pages showing which portions of them are text, and which represent pictures). The network has been adapted and re-trained from scratch using the "newspaper and magazine image segmentation dataset" available in the literature. This consists of pairs of scanned documents / labeled images of documents. All documents belong to different kinds of newspapers and magazines. Despite the training set is strongly different from the test set (i.e., no scientific publications are used in training), results on image/text segmentation of digital versions of scanned scientific papers proved promising.

In order to detect text within scanned documents, we made use of the Efficient and Accurate Scene Text Detector (EAST). This is a multi-channel fully convolutional network showing impressive results even if the considered text is not perfectly aligned with the page (e.g., in case of paper distortion or suboptimal scanning procedures). Once the EAST model is applied, we obtain a first rough estimate of text location within the page. We then apply an enhancement procedure based on morphological operations and connected-components linking in order to estimate the final text location. An example of image and text extraction achieved with the proposed methods is shown in Figure 97.

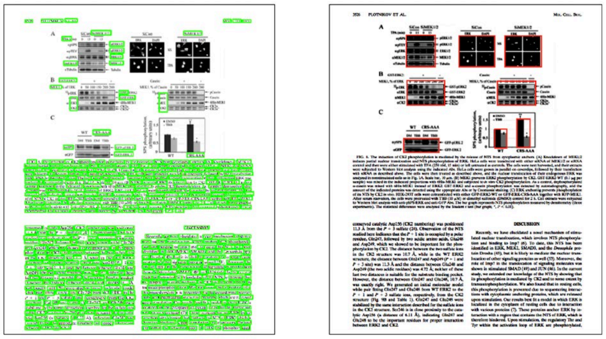


Figure 97. Example of Text and Image Extraction from Scanned PDF Page

The proposed method for blot detection and matching is based on the following series of operations: given a paper, we use the tool developed at Unicamp to extract images containing blots; we then segment all blots within a picture by applying thresholding and watershed segmentation followed by post-processing morphological operations; we finally compare all pairs of blots using Fourier Melling Transform and phase-correlation. To test the proposed solution, we created a dataset of more than 50k blot pairs obtained by manually segmenting scientific images and applying a series of synthetic transformations (i.e., gaussian noise addition, constant multiplication and addition, color inversion, JPEG compression, rotation, gamma correction, and flipping). An example of matching blots under different transformations is reported in Figure 98. Results show that it is possible to correctly detect more than 50% of the matchings, with zero false positives. The overall AUC is of 0.80.

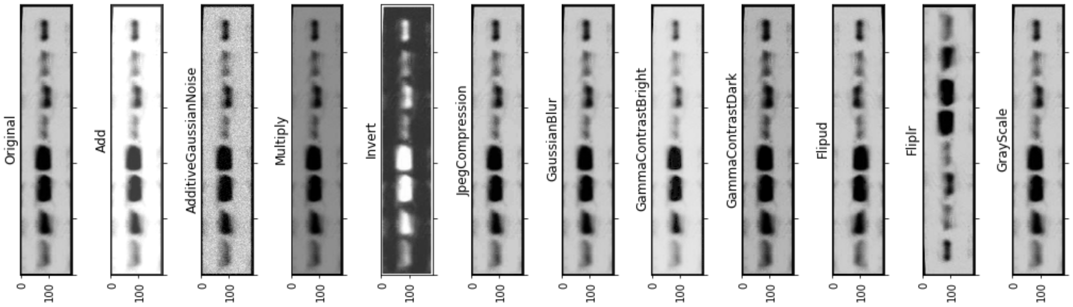


Figure 98. Examples of Blots Undergoing Different Processing Operations

Contract Information:
Team: Purdue Team, PoliMi Subteam
POC: Prof. Stefano Tubaro, Politecnico di Milano, stefano.tubaro@polimi.it, +39 02 2399 3647

Major Technical Problem: Analysis of synthetic scientific images
Major Technical Approach: Feature extraction and one-class classifier

Program Objectives: Development of synthetic western blots detector for the analysis of scientific publications

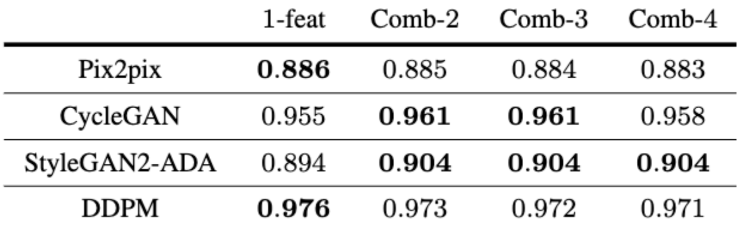
The dataset built for our experiment is composed by

- 14,200 real images with size 256 X 256 pixels
- 6,000 squared images with size 256 X 256 generated by the Pix2pix model, providing as input to the generator the same blot-masks seen in training phase;
- 6,000 squared images with size 256 X 256 generated by the CycleGAN model, providing as input to the generator the same blot-masks seen in training phase;
- 6,000 squared images with size 256 X 256 generated by the StyleGAN2-ADA model, providing as input to the generator different seeds for each new image to be synthesized.
- 6,000 squared images with size 256 X 256 generated by DDPM providing as input to the generator different noisy samples, corresponding to an equal number of new images to be synthesized.

The training step of the one-class classifier was carried out on 50% of the real pristine images only. All the other images have been used for the test set.

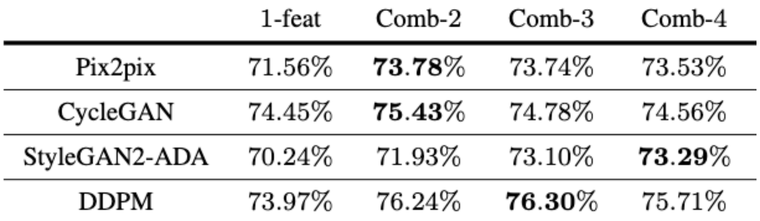
Table 63 shows the AUC obtained on the different synthetic classes when we use 1, 2, 3 or 4 features from the computed co-occurrences matrixes. Table 64 shows the same results in terms of balanced accuracy. These results show promising results, considering that synthetic images have never been used in training.

Table 63 - AUC achieved with different feature combinations (from 1 to 4)



	1-feat	Comb-2	Comb-3	Comb-4
Pix2pix	**0.886**	0.885	0.884	0.883
CycleGAN	0.955	**0.961**	**0.961**	0.958
StyleGAN2-ADA	0.894	**0.904**	**0.904**	**0.904**
DDPM	**0.976**	0.973	0.972	0.971

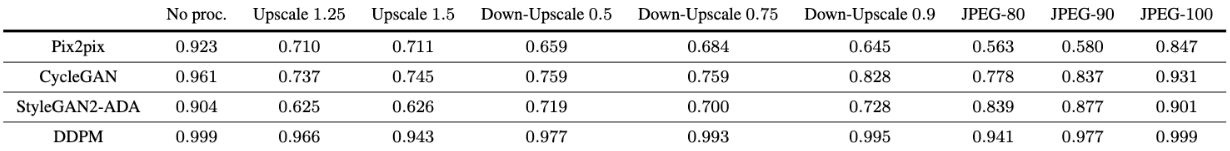
Table 64 - Accuracy achieved with different feature combinations (from 1 to 4)



	1-feat	Comb-2	Comb-3	Comb-4
Pix2pix	71.56%	**73.78**%	73.74%	73.53%
CycleGAN	74.45%	**75.43**%	74.78%	74.56%
StyleGAN2-ADA	70.24%	71.93%	73.10%	**73.29**%
DDPM	73.97%	76.24%	**76.30**%	75.71%

Further results have been obtained by testing the proposed classifier on top of images further processed through laundering operations (i.e., compression and resizing). Table 65 shows these results in terms of AUC.

Table 65 - AUC achieved on test data attacked through laundering operations



	No proc.	Upscale 1.25	Upscale 1.5	Down-Upscale 0.5	Down-Upscale 0.75	Down-Upscale 0.9	JPEG-80	JPEG-90	JPEG-100
Pix2pix	0.923	0.710	0.711	0.659	0.684	0.645	0.563	0.580	0.847
CycleGAN	0.961	0.737	0.745	0.759	0.759	0.828	0.778	0.837	0.931
StyleGAN2-ADA	0.904	0.625	0.626	0.719	0.700	0.728	0.839	0.877	0.901
DDPM	0.999	0.966	0.943	0.977	0.993	0.995	0.941	0.977	0.999

4.3.3 University of Notre Dame

Contract Information:
Team: Purdue Team, TA.1.3, Notre Dame Sub-Team
POC: Prof. Walter Scheirer, University of Notre Dame, walter.scheirer@nd.edu, +1 (574) 631-2436

Major Technical Problem: Scientific Image Analysis
Major Technical Approach: SILA: A System for Scientific Image Analysis

The current version of SILA – a system for scientific image analysis – counts on five main tasks, which are made available to the human analyst:

* Image and Caption Extraction: This is the task responsible for automatically extracting figures and their respective captions from the suspect PDF files that were uploaded to SILA.

* Panel Segmentation: This task allows the analyst to select multi-panel figures (for instance, figures composed of one or two outputs of microscopy, graphs, and bar charts collated together) and automatically segment them into smaller parts, one for each constituent panel.

* Image Ranking: Once images extracted from the PDF files are made available, analysts are allowed to select, as many times as they want, an image of interest and retrieve similar content existing across the various PDF files, such as exact copies, near-duplicates, and semantically similar elements, in a similar fashion to Google reverse image search.

* Copy-Move Detection: This task allows the analyst to perform a single-image analysis over a selected figure, with the aim to inspect the image for cloned regions (a.k.a. copy-move detection [1]).

* Provenance Analysis: In addition to inspecting single images, provenance analysis [2] is made available to the analyst as a tool to detect both content splicing and reuse across different sections of one or more publications.

In addition to these tasks, we also collected and annotated the Scientific Papers (SP) dataset: a set of scientific publications containing samples collected from all over the world with the documented presence of image manipulations and reuse. We are making this dataset available to the community [3], along with task-specific ground-truth annotations (e.g., cloned regions within the manipulated images, provenance graphs detailing reused images across papers).

Table 66 summarizes the results of SILA over the SP dataset. The reported metrics are: IR (image recall, for the task of image extraction), LD (normalized Levenshtein Distance [4], for the task of caption extraction), BS (BertScore [5], also for the task of caption extraction), IoU (intersection over union, for the task of panel segmentation), P@N (precision at top-N images, for the task of image ranking), F1 (F1-score [1], for the task of copy-move detection), VO (vertex overlap [2], for the task of provenance analysis), EO (edge overlap [2], for provenance analysis), and VEO (vertex and edge overlap [2], also for provenance analysis). All values belong to the real interval [0..1]. For all metrics but LD, the larger the value, the better the solution. All values are averaged over the sets of elements in the bottom row (± deviations), except for IR.

Table 66. Quantitative results of the tasks performed by SILA over the SP dataset.

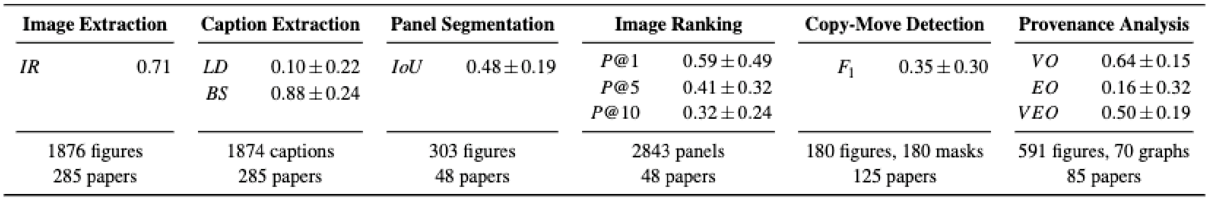


Image Extraction		Caption Extraction		Panel Segmentation		Image Ranking		Copy-Move Detection		Provenance Analysis	
IR	0.71	*LD*	0.10 ± 0.22	*IoU*	0.48 ± 0.19	*P@1*	0.59 ± 0.49	F_1	0.35 ± 0.30	*VO*	0.64 ± 0.15
		BS	0.88 ± 0.24			*P@5*	0.41 ± 0.32			*EO*	0.16 ± 0.32
						P@10	0.32 ± 0.24			*VEO*	0.50 ± 0.19
1876 figures 285 papers		1874 captions 285 papers		303 figures 48 papers		2843 panels 48 papers		180 figures, 180 masks 125 papers		591 figures, 70 graphs 85 papers	

Discussion: in the topic of scientific image analysis, there is a lack of a unified benchmark that allows the principled evaluation of different techniques. Currently, it is not easy for researchers to either compare or reproduce each others' results. There is also no large dataset freely available out there containing diverse, well-documented, and confirmed cases of image manipulations, with rich annotations that are precise enough to allow the calculation of objective metrics. We believe the herein presented results take a fundamental first step in the direction of establishing a baseline and providing such a benchmark. A paper introducing SILA and the proposed dataset has been submitted to Springer Nature Scientific Reports and is currently under the second round of peer review.

[1] Cozzolino et al. "Efficient dense-field copy-move forgery detection," IEEE T-IFS, Vol. 10, 2015.

[2] Moreira et al. "Image Provenance at Scale," IEEE T-IP, Vol. 27, No. 12, 2018.

[3] Moreira et al. "Scientific Integrity Dataset," available at:
https://github.com/danielmoreira/sciint/tree/dataset, 2021.

[4] Levenshtein. "Binary codes capable of correcting deletions, insertions, and reversals," Sov. Phys. Doklady Vol. 10, 1966.

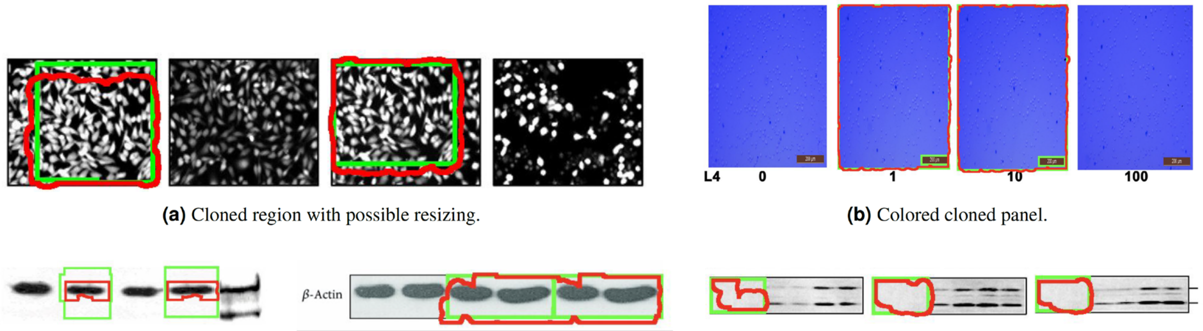
[5] Zhanget al. "BERTscore: Evaluating text generation with BERT", ArXiv
e-prints, available at: https://arxiv.org/abs/1904.09675, 2019.

4.3.4 University of Naples Federico II

Contract Information:
Team: Purdue Team, Unina Subteam
POC: Prof. Luisa Verdoliva, University Federico II of Naples, verdoliv@unina.it, +39 081 7683929

Major Technical Problem: Analysis of manipulated scientific images
Major Technical Approach: Development of a dense-based copy-move detector for scientific images

For the dataset we manually selected 180 figures from 126 retracted scientific papers. We then recruited six volunteers to go through the retraction notices and manually annotate the described cloned regions. We asked three folks to annotate each figure and adopted the intersection of the provided annotations as the final manipulation mask. As a result, we obtained a set of 180 pixel-wise copy-move manipulated region masks. In Figure 99, we show some results that are aligned with the human annotations. Green boxes highlight the ground truth provided by annotators, while the red regions were automatically obtained through the system. As one might observe, we can detect copies coming from different panels of the same figure, even if the cloned region was resized (see Figure 99a). It can also deal with both gray-level and colored images (see Figure 99 a and b), and even with scientific imagery such as western blots (see Figure 99c, d and e). In the latter case, we were able to detect the replication of small areas, including cloning possibly related to content removal (see Figure 99e). Of course, these analyses do not necessarily point out malicious manipulations. Instead, they aim at helping analysts focus on suspiciously similar content.



(a) Cloned region with possible resizing.

(b) Colored cloned panel.

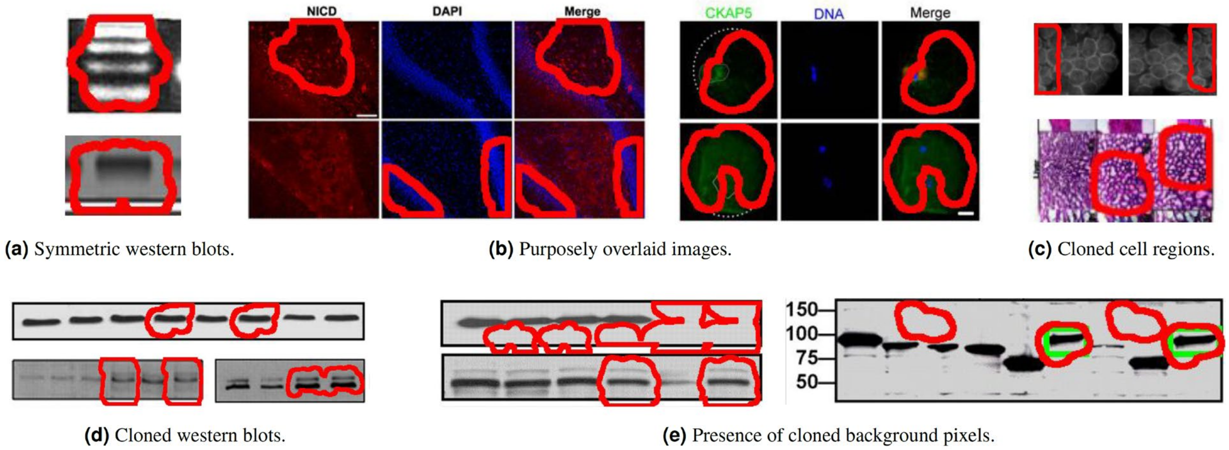
(c) Cloned blot.

(d) Cloned pair of blots.

(e) Cloned uniform pixels possibly to hide elements.

Figure 99. Results of copy-move detection that agree with existing retraction notices. In green boxes, the manual annotations provided as ground truth, while in red the cloned regions identified by our algorithm.

To gain insights into the complexity of the task at hand and the need for human intervention, we show some interesting results in Figure 100. In Figure 100a, for instance, there are bilaterally symmetric western blots that may cause false alarms to our detector. Our approach recognizes two symmetric regions in the image, but this does not mean manipulation was carried out since the algorithm cannot distinguish between fabricated mirrored areas and authentic symmetric patterns. Another common situation is presented within Figure 100b. It often happens that, in a timed experiment, the same substrate is imaged several times and under different conditions, to ultimately be either compared side-by-side or overlaid to generate a richer representation. This is the case of the two rightmost "Merge" panel columns depicted within Figure 100b, which are indeed a combination of their respective previous panels (therefore sharing similar regions). We can detect these replications, but their existence is legitimate and would not be considered evidence of manipulation by a human analyst. Lastly, Figure 100c, d, and e depict examples of suspect cases that were probably overlooked by human analysts. Many of these images were not previously acknowledged as containing duplicated content. Nonetheless, we can spot cloned content within them (see Figure 100c and d), and even probable content removal (see Figure 100e, where uniform background pixels might have been used to make blot lanes clear). It is worth mentioning that all of these figures came from already retracted papers, and only one of them was previously identified by a person as containing a duplication (see the pair of green boxes within the rightmost panel of Figure 100e).



(a) Symmetric western blots.

(b) Purposely overlaid images.

(c) Cloned cell regions.

(d) Cloned western blots.

(e) Presence of cloned background pixels.

Figure 100. Results of copy-move detection that are not mentioned in existing retraction notices. Regions with red borders depict the cloned content found by our method. Many of these regions do not present a respective ground-truth green box because they have never been acknowledged in a retraction notice.

Finally, we report the results in terms of F1-score of copy-move detection over the same scientific images obtained by different approaches from the forensic literature in Table 67. Within this group, there were methods specialized in highlighting noise inconsistencies, methods that tackled the problem of copy-move detection with classic content-matching approaches such as sparse or dense local descriptors, and more recent methods relying on deep-learning-based strategies. As one might observe, all the methods presented performance far from those belonging to our proposed solutions.

Table 67. Copy-move detection results of different forensic techniques.

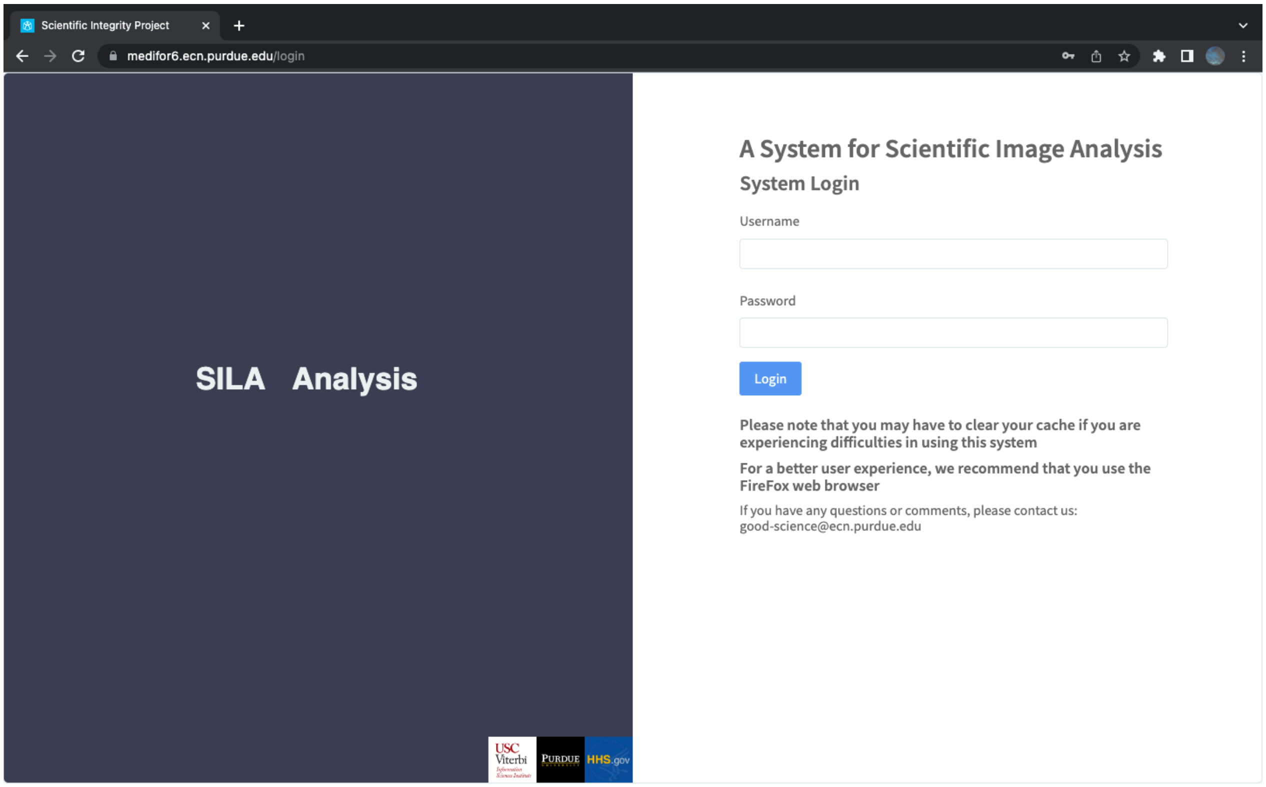
Description	F_1-score	Method
Methods based on noise analyses	0.081	ELA
	0.128	Noise-based
	0.101	Splicebuster
Methods based on content matching	0.164	SIFT-based
	0.202	EDF
	0.149	FE-CMFD-HFPM
Methods based on deep learning	0.092	BusterNet
	0.065	MSD/STRD
Proposed methods	0.296	Basic Zernike (no panel segmentation)
	0.307	Zernike
	0.325	RGB
	0.345	Zernike+RGB

4.3.5 USC Information Sciences Institute

Contract Information:
Team: USC Information Sciences Institute
POC: Prof. Wael Abd-Almageed, USC-ISI, wamageed@isi.edu, +1 703 248 6174
Major Technical Problem: Scientific Integrity
Major Technical Approach: A GUI for Scientific Integrity System
Program Objectives: develop a system GUI including multiple image analyzing methods for scientific papers

Screenshots:
Following are screenshots taken from various components of the system, representing typical flow of using the system:

- User login



A System for Scientific Image Analysis

System Login

Username

Password

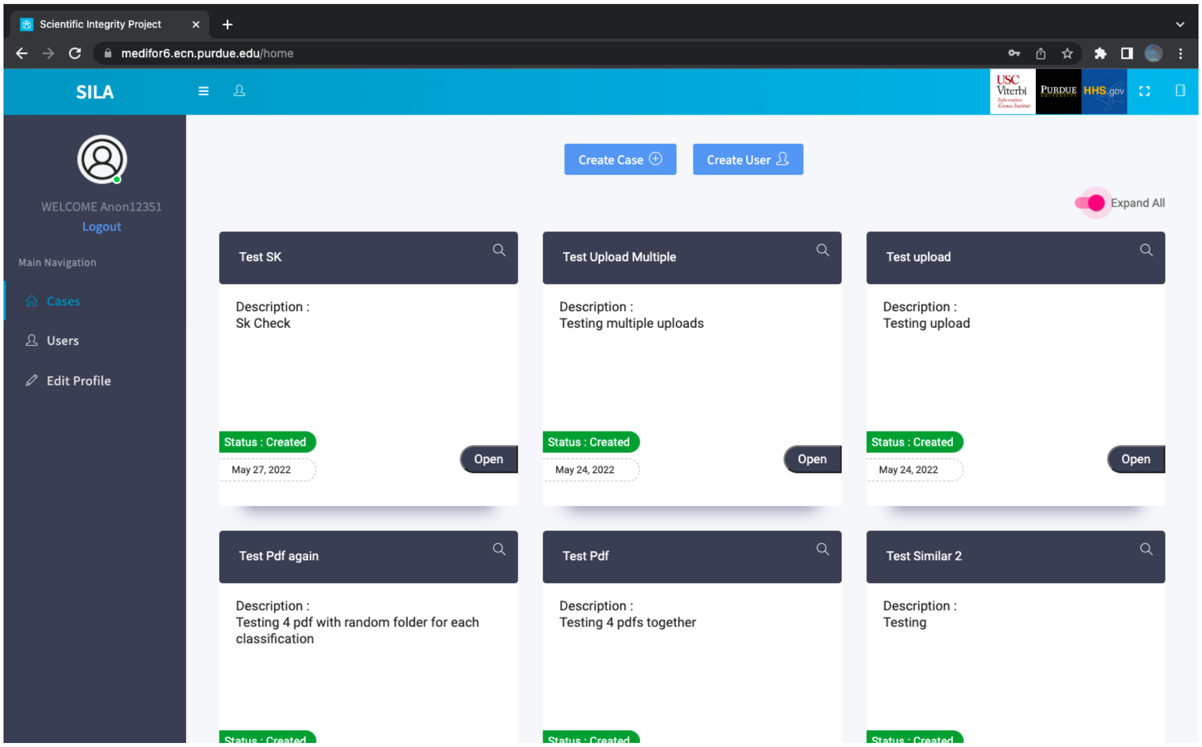
Please note that you may have to clear your cache if you are experiencing difficulties in using this system

For a better user experience, we recommend that you use the FireFox web browser

If you have any questions or comments, please contact us: good-science@ecn.purdue.edu

Figure 101. Login page

- Home page, on which the user can hover on magnifying glass to preview the images inside the case



Test SK

Description :
Sk Check

Status : Created
May 27, 2022

Test Upload Multiple

Description :
Testing multiple uploads

Status : Created
May 24, 2022

Test upload

Description :
Testing upload

Status : Created
May 24, 2022

Test Pdf again

Description :
Testing 4 pdf with random folder for each classification

Status : Created

Test Pdf

Description :
Testing 4 pdfs together

Status : Created

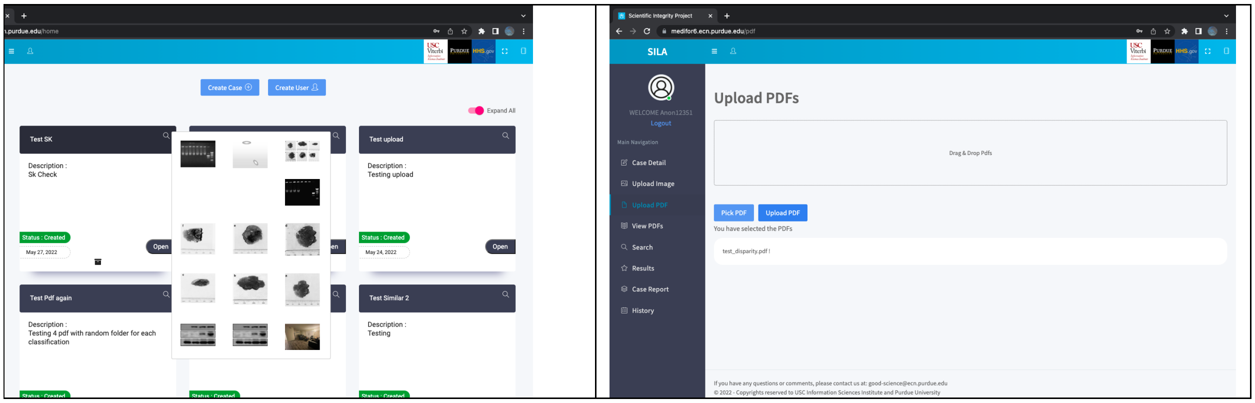
Test Similar 2

Description :
Testing

Status : Created

Figure 102. Once the user logs in, the go to the home page

- The user can drag and drop image files or PDF files that need to be analyzed



Upload PDFs

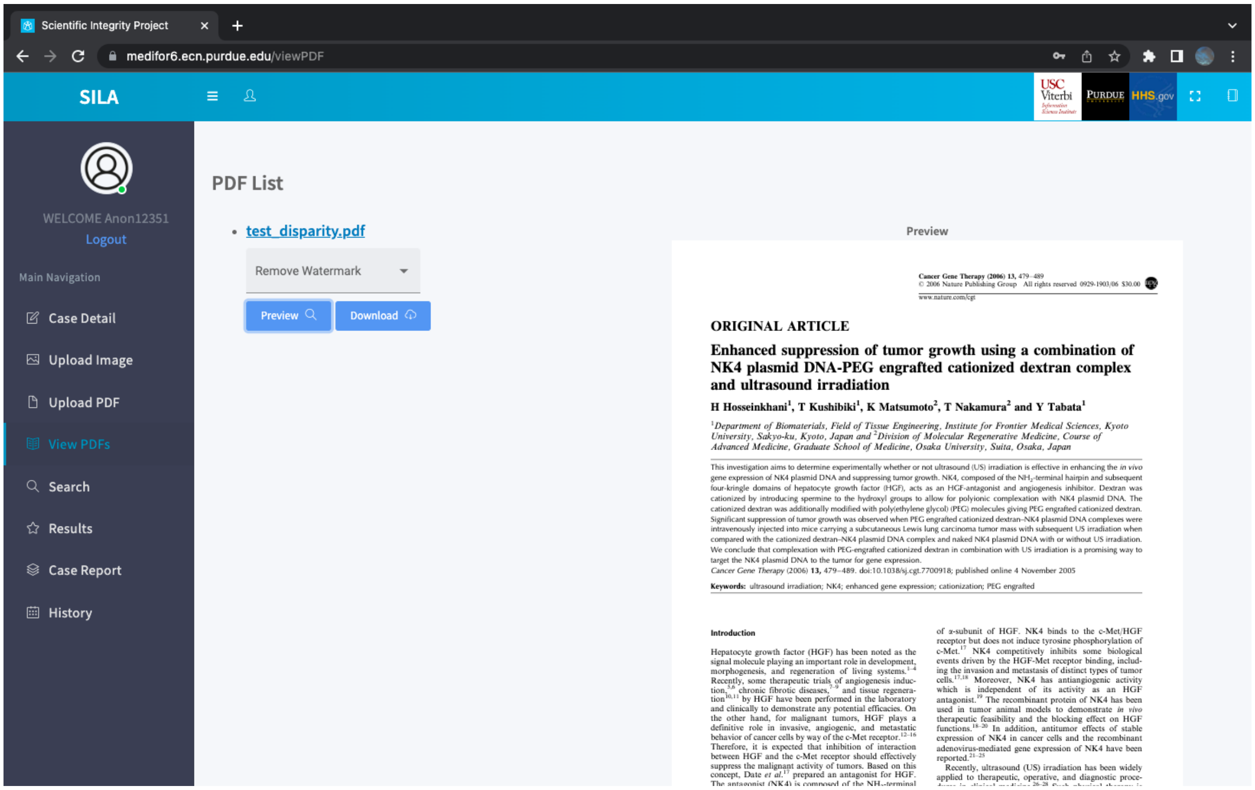
You have selected the PDFs

test_disparity.pdf !

If you have any questions or comments, please contact us at: good-science@ecn.purdue.edu
© 2022 - Copyrights reserved to USC Information Sciences Institute and Purdue University

Figure 103. Users can drop and drop files

- User can preview pdf and remove watermark if any



Cancer Gene Therapy (2006) 13, 479–489
© 2006 Nature Publishing Group All rights reserved 0929-1903/06 $30.00
www.nature.com/cgt

ORIGINAL ARTICLE

Enhanced suppression of tumor growth using a combination of NK4 plasmid DNA-PEG engrafted cationized dextran complex and ultrasound irradiation

H Hosseinkhani[1], T Kushibiki[1], K Matsumoto[2], T Nakamura[2] and Y Tabata[1]

[1]Department of Biomaterials, Field of Tissue Engineering, Institute for Frontier Medical Sciences, Kyoto University, Sakyo-ku, Kyoto, Japan and [2]Division of Molecular Regenerative Medicine, Course of Advanced Medicine, Graduate School of Medicine, Osaka University, Suita, Osaka, Japan

This investigation aims to determine experimentally whether or not ultrasound (US) irradiation is effective in enhancing the *in vivo* gene expression of NK4 plasmid DNA and suppressing tumor growth. NK4, composed of the NH_2-terminal hairpin and subsequent four-kringle domains of hepatocyte growth factor (HGF), acts as an HGF-antagonist and angiogenesis inhibitor. Dextran was cationized by introducing spermine to the hydroxyl groups to allow for polyionic complexation with NK4 plasmid DNA. The cationized dextran was additionally modified with poly(ethylene glycol) (PEG) molecules giving PEG engrafted cationized dextran. Significant suppression of tumor growth was observed when PEG engrafted cationized dextran–NK4 plasmid DNA complexes were intravenously injected into mice carrying a subcutaneous Lewis lung carcinoma tumor mass with subsequent US irradiation when compared with the cationized dextran–NK4 plasmid DNA complex and naked NK4 plasmid DNA with or without US irradiation. We conclude that complexation with PEG-engrafted cationized dextran in combination with US irradiation is a promising way to target the NK4 plasmid DNA to the tumor for gene expression.
Cancer Gene Therapy (2006) **13,** 479–489. doi:10.1038/sj.cgt.7700918; published online 4 November 2005

Keywords: ultrasound irradiation; NK4; enhanced gene expression; cationization; PEG engrafted

Introduction

Hepatocyte growth factor (HGF) has been noted as the signal molecule playing an important role in development, morphogenesis, and regeneration of living systems.[1–4] Recently, some therapeutic trials of angiogenesis induction,[5,6] chronic fibrotic diseases,[7–9] and tissue regeneration[10,11] by HGF have been performed in the laboratory and clinically to demonstrate any potential efficacies. On the other hand, for malignant tumors, HGF plays a definitive role in invasive, angiogenic, and metastatic behavior of cancer cells by way of the c-Met receptor.[12–16] Therefore, it is expected that inhibition of interaction between HGF and the c-Met receptor should effectively suppress the malignant activity of tumors. Based on this concept, Date *et al.*[17] prepared an antagonist for HGF. The antagonist (NK4) is composed of the NH_2-terminal of α-subunit of HGF. NK4 binds to the c-Met/HGF receptor but does not induce tyrosine phosphorylation of c-Met.[17] NK4 competitively inhibits some biological events driven by the HGF-Met receptor binding, including the invasion and metastasis of distinct types of tumor cells.[17,18] Moreover, NK4 has antiangiogenic activity which is independent of its activity as an HGF antagonist.[19] The recombinant protein of NK4 has been used in tumor animal models to demonstrate *in vivo* therapeutic feasibility and the blocking effect on HGF functions.[18–20] In addition, antitumor effects of stable expression of NK4 in cancer cells and the recombinant adenovirus-mediated gene expression of NK4 have been reported.[21–25]

Recently, ultrasound (US) irradiation has been widely applied to therapeutic, operative, and diagnostic proce-

Figure 104. PDF preview and watermark removal

- Users can perform operations, such as Panel Extraction

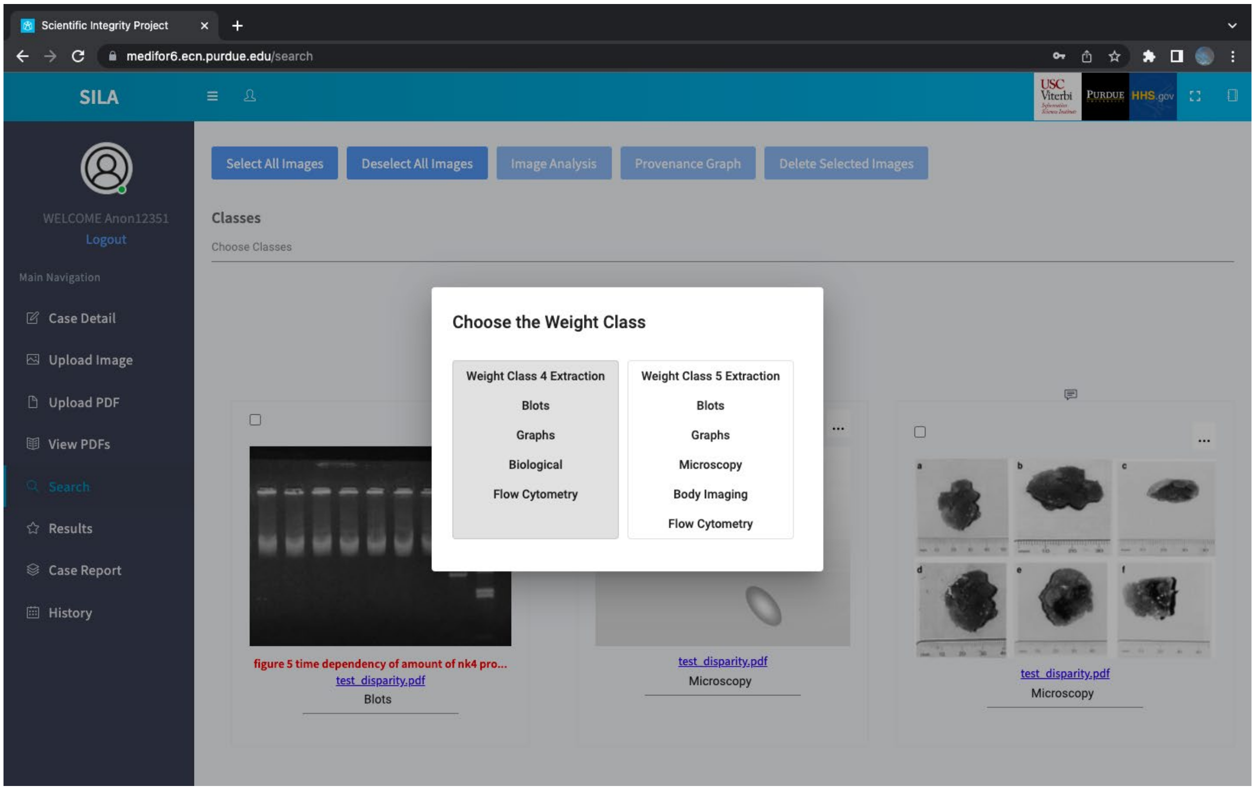


Figure 105. Panel extraction from selected images

- Users can also add custom annotations to select images

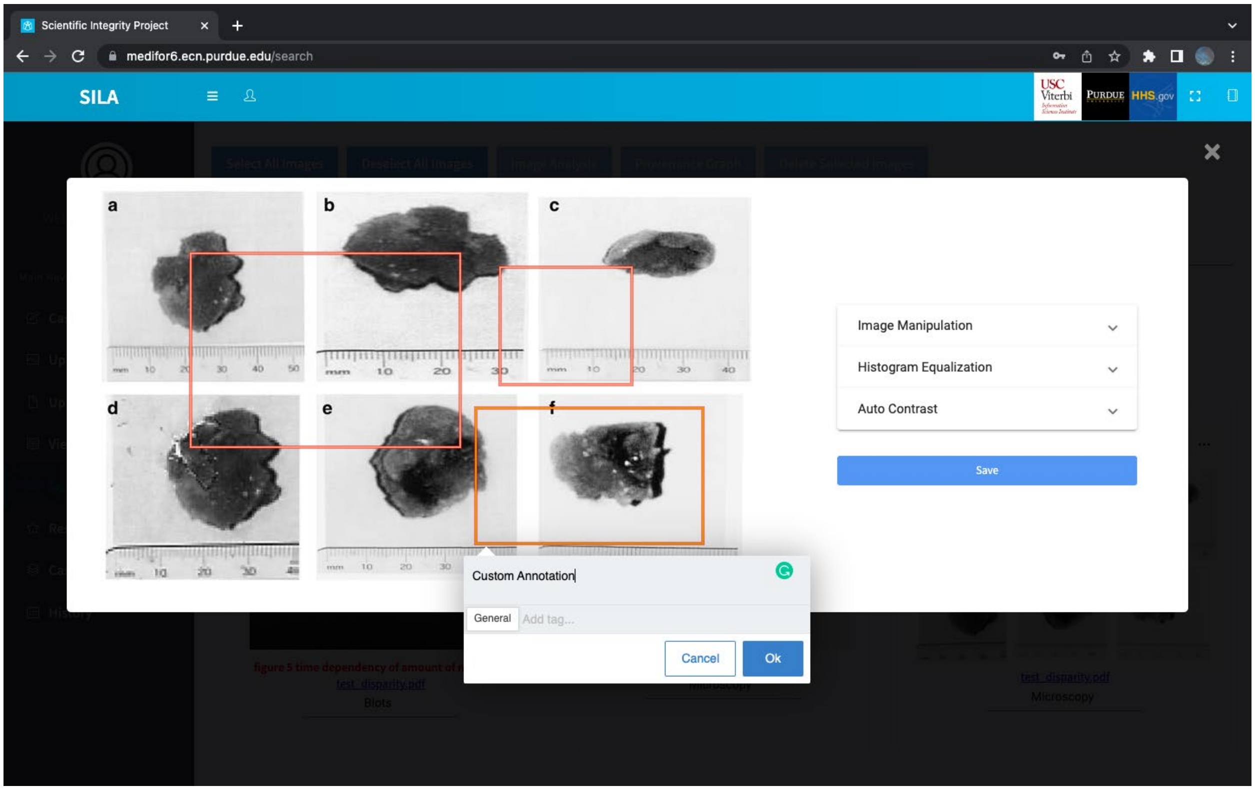


Figure 106. Custom annotations of select images

- User can manually correct automatically generated images labels

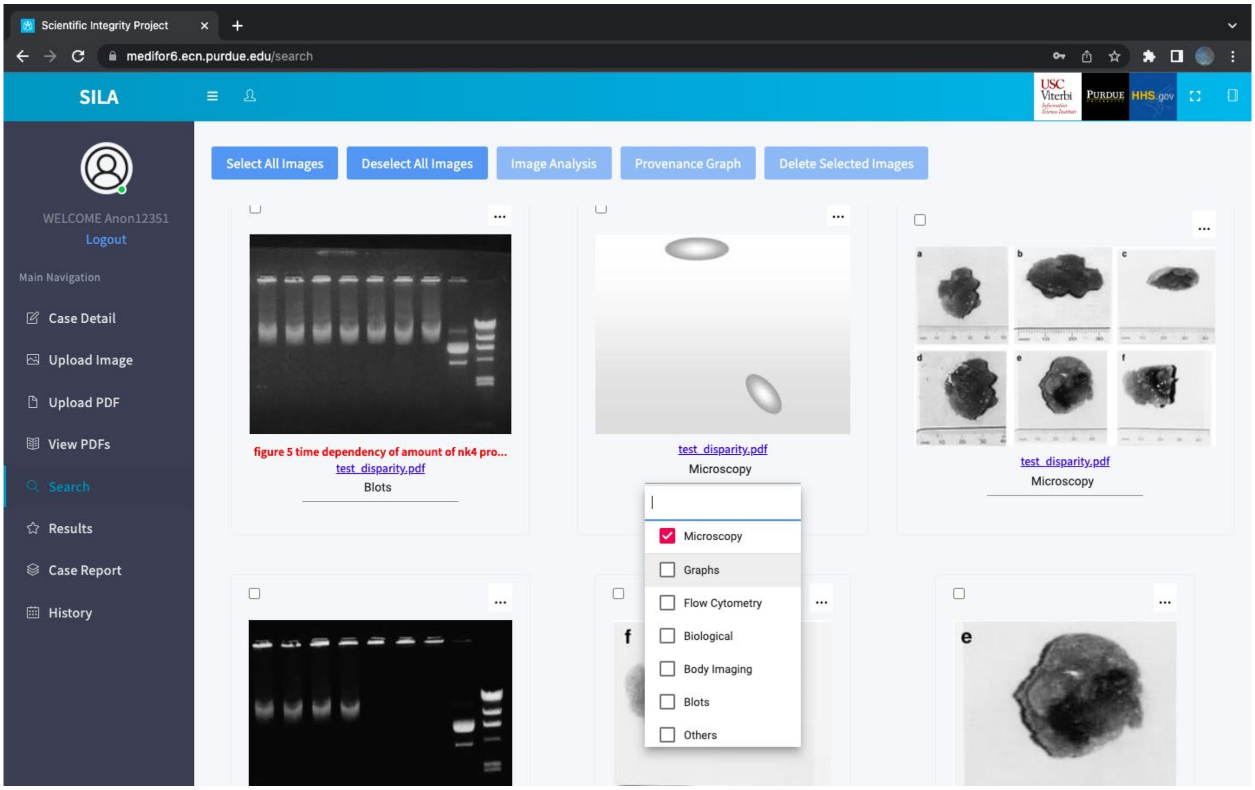
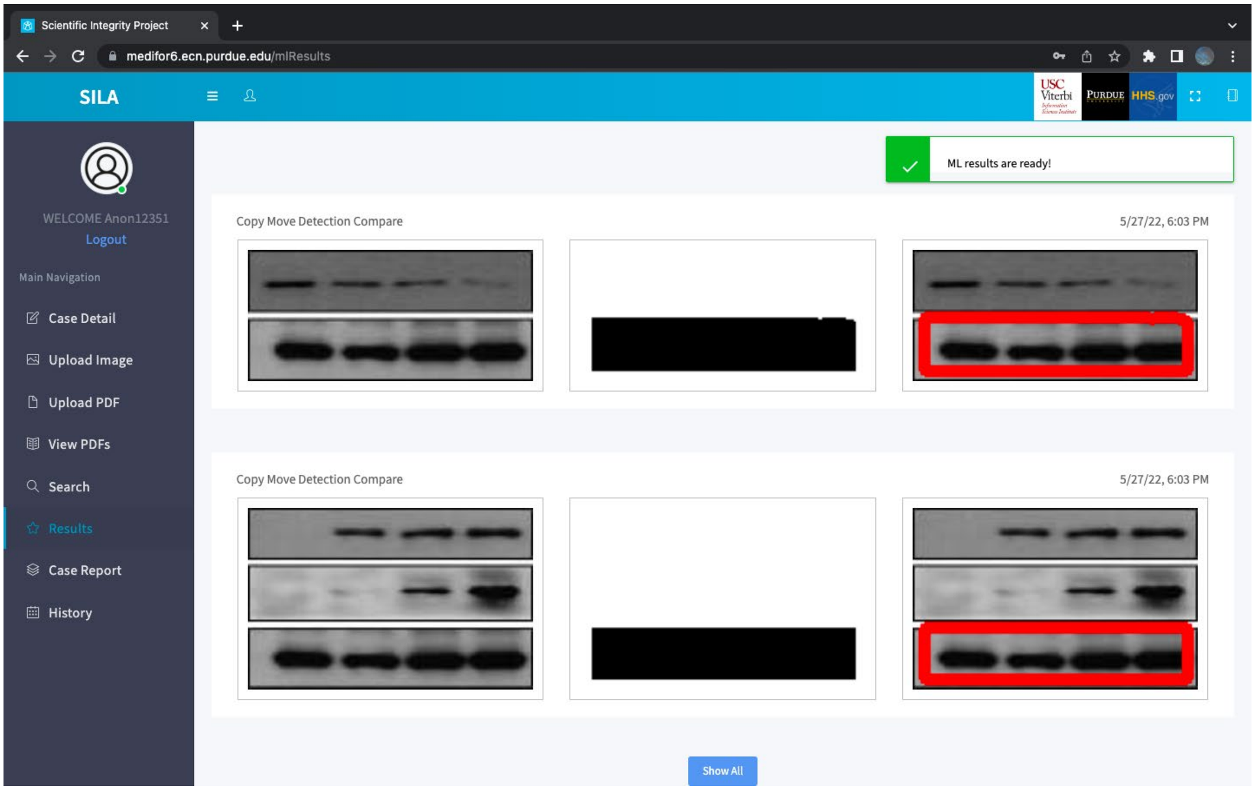


Figure 107. Manual correction of image labels

- User can select images and perform copy-move analysis



ML results are ready!

Copy Move Detection Compare 5/27/22, 6:03 PM

Copy Move Detection Compare 5/27/22, 6:03 PM

Figure 108. Copy-move analysis on select images

- Users can perform provenance analysis of select images to find relationships between images represented as a provenance graph

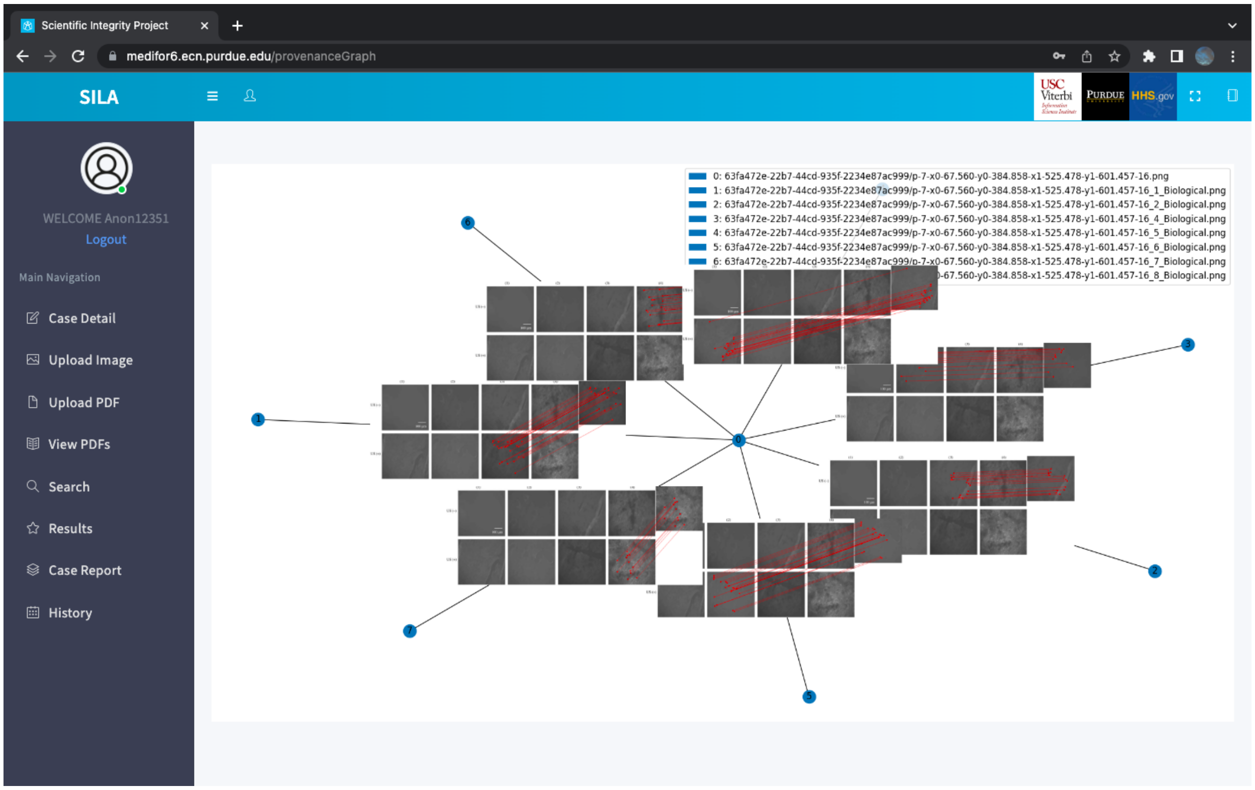
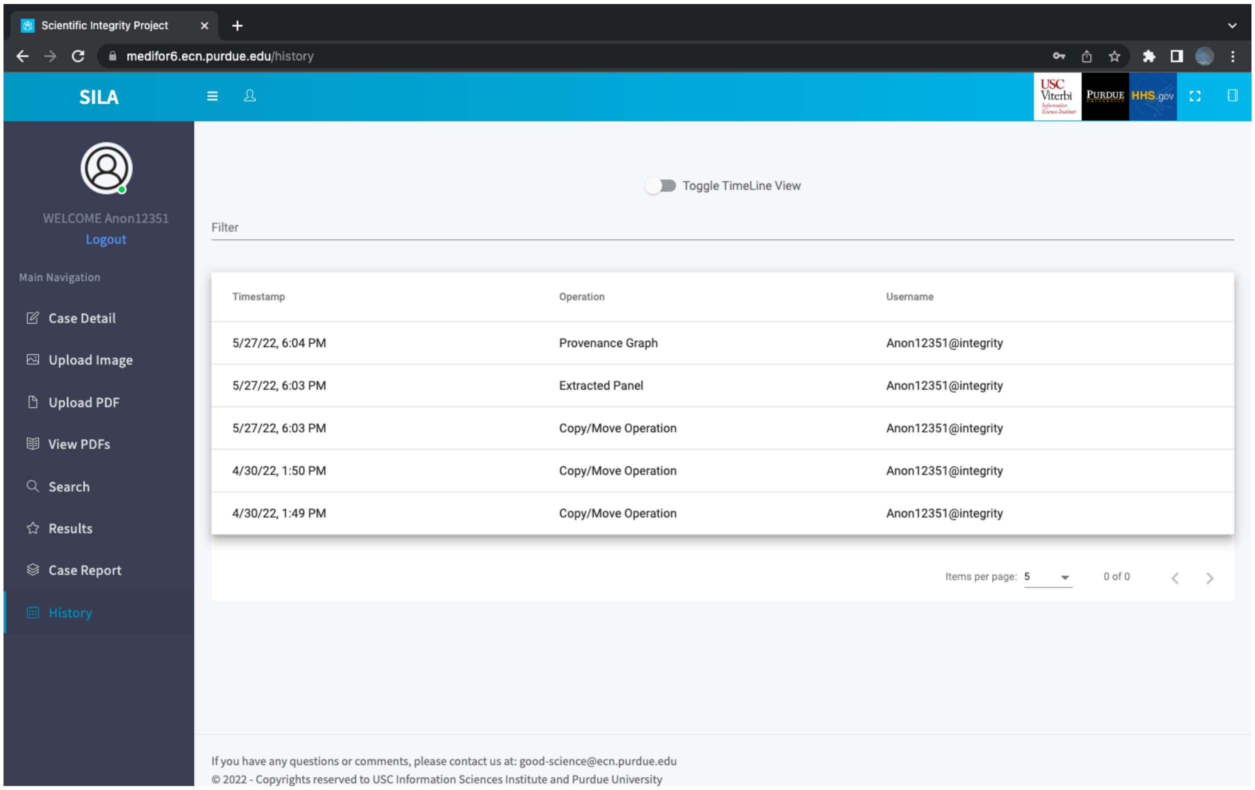

Scientific Integrity Project
mediforб.ecn.purdue.edu/provenanceGraph
SILA
USC Viterbi Information Sciences Institute
PURDUE
HHS.gov
WELCOME Anon12351
Logout
Main Navigation
Case Detail
Upload Image
Upload PDF
View PDFs
Search
Results
Case Report
History
0: 63fa472e-22b7-44cd-935f-2234e87ac999/p-7-x0-67.560-y0-384.858-x1-525.478-y1-601.457-16.png
1: 63fa472e-22b7-44cd-935f-2234e87ac999/p-7-x0-67.560-y0-384.858-x1-525.478-y1-601.457-16_1_Biological.png
2: 63fa472e-22b7-44cd-935f-2234e87ac999/p-7-x0-67.560-y0-384.858-x1-525.478-y1-601.457-16_2_Biological.png
3: 63fa472e-22b7-44cd-935f-2234e87ac999/p-7-x0-67.560-y0-384.858-x1-525.478-y1-601.457-16_4_Biological.png
4: 63fa472e-22b7-44cd-935f-2234e87ac999/p-7-x0-67.560-y0-384.858-x1-525.478-y1-601.457-16_5_Biological.png
5: 63fa472e-22b7-44cd-935f-2234e87ac999/p-7-x0-67.560-y0-384.858-x1-525.478-y1-601.457-16_6_Biological.png
6: 63fa472e-22b7-44cd-935f-2234e87ac999/p-7-x0-67.560-y0-384.858-x1-525.478-y1-601.457-16_7_Biological.png
-67.560-y0-384.858-x1-525.478-y1-601.457-16_8_Biological.png
0
1
2
3
5
6
7

Figure 109. Provenance graph representing relationships between images

- All operations performed by the user are recorded in a detailed audit trail such that can be referenced later



Timestamp	Operation	Username
5/27/22, 6:04 PM	Provenance Graph	Anon12351@integrity
5/27/22, 6:03 PM	Extracted Panel	Anon12351@integrity
5/27/22, 6:03 PM	Copy/Move Operation	Anon12351@integrity
4/30/22, 1:50 PM	Copy/Move Operation	Anon12351@integrity
4/30/22, 1:49 PM	Copy/Move Operation	Anon12351@integrity

If you have any questions or comments, please contact us at: good-science@ecn.purdue.edu
© 2022 - Copyrights reserved to USC Information Sciences Institute and Purdue University

Figure 110. Audit trail of all operations performed by a user

- Users can generate a report summarizing the finding of a given case and export it as a PDF

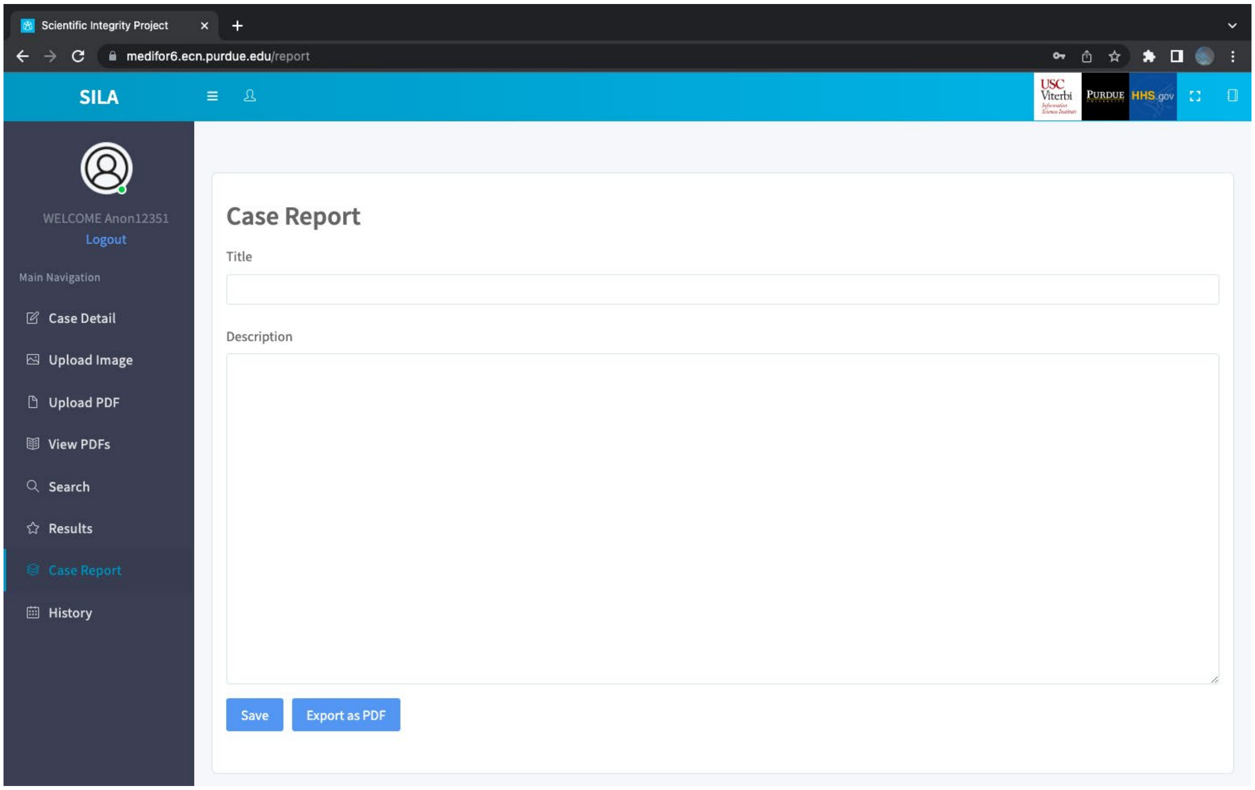


Figure 111. Case report generated by a user and exported as a PDF

5.0 CONCLUSION

Here we briefly review what are our contributions to the MediFor program for each part of TA1. We also list all of the software products we have delivered to DARPA.

Task TA 1.1a: Adversary-Aware ML Methods Based On High-Order Statistics For Detection/Localization Of Global Color Manipulations And Multiple Compressions

We addressed the problems of detection of double JPEG compression and contrast manipulation in the presence of counter-forensic attacks, considering various setups and different approaches for the analysis.

With regard to the detection of double compression, in our first approach, we focused on SVM-based detection and resort to adversary-aware training. To design a detector capable to withstand both laundering attacks (that may be intentional or not) and counter-forensics attacks aimed at making the image look like a single compressed one (erasing single compression traces after the first compression), a suitable rich feature set is identified and used, by exploiting steganalysis features extracted from both spatial (image pixel) and frequency (DCT) domain. For this task, the capability of CNNs is also exploited to design a double JPEG detector that can work on small window sizes, and that can then be applied for localization. Moreover, by resorting to CNNs, we also designed a method for the estimation of the primary quantization matrix in double JPEG images. The estimator shows superior performance compared to the state-of-the-art methods and is capable to work under very general operative conditions (regarding the alignment of the compression grids and the qualities of first and second compression). The method is then integrated in an algorithm for splicing localization based on compression inconsistencies, that resorts to clustering and morphological reconstruction to get a final map of the tampering. Under the assumption that the donor images have a different JPEG quality, the method can identify the different sources of tampering.

For the detection of contrast manipulation, several approaches have been considered. A JPEG-aware version of an SVM detector that exploits color rich feature models is trained to improve the robustness of contrast enhancement detection against JPEG compression (which is shown to be the most harmful laundering attack against contrast manipulation detectors). The method first estimates the JPEG quality from the pixel image by exploiting the idempotency property of the JPEG compression and then choose, among a pool of trained models, the aware SVM trained for the corresponding (or the closest) quality. By exploiting the superior performance of deep learning-based methods, a JPEG-aware, CNN architecture is designed for the detection of a generic contrast adjustment, showing good generalization capabilities to different tonal adjustments performed by various software. In this case, robustness against JPEG is achieved by training the model on a mixture of JPEG qualities. The CNN works patch-based and then can be applied for localization on sliding windows.

Finally, a method for copy-move disambiguation has been developed by exploiting a multi-branch CNN architecture that we explicitly designed for the purpose. The tool allows to identify the source and target region of a copy-move forgery, thus permitting to localize the tampering. The disambiguator shows good results also when global post-processing is applied to the image.

TA1.1b Clone and splicing detection and localization

We conducted fundamental research on interpretable convolutional neural networks and theory of successive subspace learning. We also conducted applied research on image splicing localization and published two papers. Our research work received quite a few recognitions including one best paper award (2018) and two best paper runner-up citations (2019 and 2020).

One of our papers is highly cited paper in the field of image splicing localization. It has been cited 63 times since its publication in 2018. In this work, we proposed a technique that utilizes a fully convolutional network (FCN) to localize image splicing attacks. We first evaluated a single-task FCN (SFCN) trained only on the surface label. Although the SFCN is shown to provide superior performance over existing methods, it still provides a coarse localization output in certain cases. Therefore, we propose the use of a multi-task FCN (MFCN) that utilizes two output branches for multi-task learning. One branch is used to learn the surface label, while the other branch is used to learn the edge or boundary of the spliced region. We trained the networks using the CASIA v2.0 dataset, and tested the trained models on the CASIA v1.0, Columbia Uncompressed, Carvalho, and the DARPA/NIST Nimble Challenge 2016 Science datasets. Experiments show that the SFCN and MFCN outperform existing splicing localization algorithms, and that the MFCN can achieve finer localization than the SFCN.

Task: TA1.1c - Sensor-Wide Image Attribution

The goal of TA1.1c is to develop forensic techniques related to sensor-based image attribution. This means to be able to attribute images or videos to the sensor (or family of sensors) used to acquire them. Within this context, we developed a series of detectors for many different tasks: camera model identification through Convolutional Neural Networks (CNNs); image and video device attribution by means of PRNU-based analysis; scanner attribution.

Camera model identification refers to the problem of understanding which is the camera brand and model used to shoot a photograph. To this purpose, we developed a detector based on CNNs. In particular, given an image, our method detects which camera model has been used to shoot it within a set of known models (closed-set) or if the image has been acquired with an unknown model (open-set). Our algorithm works in the challenging scenario of small patches (i.e., 64x64 pixels), still achieving state-of-the-art accuracy.

As the proposed ad-hoc CNN exploits very small image patches to take decision, it is paramount that this input data is as reliable as possible. As a matter of fact, not all image patches contain enough information to enable accurate detection of the camera model used to acquire them (i.e., a completely saturated patch is useless as it could have been acquired with any device). Therefore, we developed a methodology that enables to compute patch reliability. This means understanding which portion of an image carries a high amount of camera model information.
Through the joint use of the proposed patch reliability estimation technique and camera model detector, we also progressed in the field of detection and localization of splicing forgeries obtained through images shot with different cameras.

In order to detect the specific camera instance used to acquire a video or an image, we also developed a series of techniques based on photo response non uniformity (PRNU). PRNU is a trace left on each image as a multiplicative noise, which is characteristic of the sensor used to shoot the photo. It is therefore possible to attribute an image to a device by comparing a noise footprint extracted from the image under analysis with the PRNU of the suspect device. As one drawback of large-scale approaches for PRNU-based camera model attribution is the need to store in a central database a huge amount of data, we proposed a way of compressing PRNU traces. In particular we proposed an improved processing chain composed by decimation, gaussian random projections, ternary quantization and coding tailored to increase the compression rate while preserving the highest possible matching accuracy.

Moreover, we focused on the challenging problem of video attribution in presence of motion stabilization. In this context, we proposed two methods to align frame fingerprints with reference

PRNU by recovering the scaling, shift and rotation parameters introduced by the electronic stabilization by exploiting a global optimization technique. We also managed to overcome the problem of computational complexity by searching for scaling and rotation parameters in the frequency domain thanks to a modified version of the Fourier Mellin transform (FM).
Finally, in order to address the issue of the computational burden introduced by the PRNU correlation test, we also developed a method based on convolutional neural networks whose goal is to compare an image noise with a candidate PRNU to detect whether the image comes from a specific device.

In additional to consumer camera models, we also considered the problem of scanner attribution, i.e., detect which is the scanner used to acquire a picture. To solve this problem, we proposed a new robust CNN-based system that takes as input a patch extracted from an image split inro sub-images through zig-zag scan. The proposed system evaluates two tasks on scanned images: scanner model classification and reliability map generation. In Task 1 (scanner model classification), we assign the predicted scanner labels to the original images under analysis through majority voting. In Task 2, a reliability map is generated based on the majority vote result from Task 1. The pixel values in the reliability map indicate the probability of the corresponding pixel in the original image being correctly classified.

Task: TA1.1d - Detection Of The Traces Left By The Use Of Specific Editing Software Suites

The goal of TA1.1d is to develop forensic techniques capable of detecting image and video editing traces left by specific software suites. In this task we considered a series of different problems: detection of JPEG-based traces; video forgery detection and localization based on different codecs and qualities; detection of video forgeries operated by means of different deepfake software suites.

The vast majority of images available online are JPEG compressed. Images are compressed directly onboard during acquisition, are compressed after almost any processing operation, and are also compressed whenever uploaded on a social platform or shared through messaging apps. For this reason, studying the history of JPEG traces can reveal important pieces of information related to the past of a specific image and the possible use of editing software suites. Considering the problem of double JPEG compression, we developed a series of CNNs that consider different kinds of input.

Considering the problem of multiple JPEG compressions using editing software suites, we investigated a solution capable of detecting up to four image compressions. This means, given an image, detect how many times (up to four) it has been JPEG compressed. To do so, we leverage perturbations of DCT histograms that capture traces of multiple compressions and train a supervised classifier to discriminate between images compressed different amount of times.
Due to the widespread diffusion of JPEG compression standard we also focused on the detection of traces left on images by the use of different JPEG implementations (e.g., characteristic of proprietary software suites). Specifically, we focused on capturing traces left by different JPEG implementations in order to distinguish between images that have been compressed with different software suites, even if the very same quantization matrix has been used.

Considering the problem of video forgery detection, we proposed different methods exploiting different kind of traces. One of the reasons behind failure of many video forgery detection methods is that videos are customary stored and distributed in a compressed format, and codec-related traces tends to mask previous processing operations. For this reason, we decided to exploit coding traces left by specific video processing suites as an asset for solving forgery detection and localization problems. Specifically, we propose to capture video codec traces through convolutional neural networks (CNNs). To do so, we train two CNNs to extract information about the used video codec and coding quality, respectively. Building upon these CNNs, we propose a system to detect and localize forgeries generated by splicing content from different videos, characterized by inconsistent coding schemes and / or parameters (e.g., video compilations from different sources or broadcasting channels).

As the recent developments in video tampering have shown the dangerous implications introduced by deepfakes, we also investigated a deepfake video detector. Properly training a deepfake detector is a challenging task. As a matter of fact, deepfake traces are particularly subtle. Moreover, it is easy to get tricked and learn actors' faces rather than forensic footprints. For this reason, we developed a training routine that exploits a triplet loss. The network is fed triplets of faces during training. In particular, we show the network two faces of the same actor, and one face of another one, in order to push real and fake faces of the same actor further away.

Task: TA1.1e - Use of private information to secure machine learning forensic tools

The goal of TA1.1e was to develop image forensics techniques that can withstand the efforts of an adversary to make the forensics analysis fail. Several scenarios can be considered based on the knowledge that the adversary has on the system targeted by the attack. Specifically, we can distinguish between white-box and black-box attacks. In the former case, the adversary knows all the details of and has full access to the attacked system. In the latter, the attacker has no knowledge, or only partial knowledge, of the system and hence he works by attacking a surrogate forensics detector and then transfers the attack to the targeted detector. Intermediate scenarios wherein the adversary knows the general architecture of the system under attack but he is unaware of the exact parameters the analyst used to tune it are sometimes referred to as gray-box attacks. Within TA1.1e we have addressed both the more challenging white box scenario and the grey box scenario.
With regard to the white box case, we have developed an image forensics detector which is intrinsically more difficult to attack than conventional architectures. The detector combines the good properties of 2-class classifiers (producing better performance in the absence of attacks) and those of 1-class classifiers (known to provide better performance in the presence of unforeseen attacks). The result is a composite system, named 1.5-class classifier, consisting of one 2-class classifier and two 1-class classifiers run in parallel, followed by a final 1-class classifier in charge of making a final decision by relying on the outputs of the 3 classifiers run in the first stage. The use of a 1-class classifier as a last stage results in a closed detection region for the class of pristine images, thus classifying as manipulated any image whose characteristics do not correspond to those of pristine images. The 1.5-class architecture has been implemented by relying on SVM classifiers and has been applied to detect adaptive histogram equalization, resizing and median filtering. The performance of the detectors in the presence of a white box attack have been

evaluated and the expected advantage in terms of improved security, with respect to the single classifiers the architecture consists of, were confirmed.

With regard to the grey-box setting, we have studied the possibility of securing machine learning forensic tools (noticeably SVMs and CNNs) by keeping part of the information the tools rely on secret. In particular, we considered a case wherein the detector is based on a subset of features chosen from a larger set. The attacker knows the exact architecture of the detector and the overall feature set, but he is unaware of the exact subset of the features the detector relies on. We considered two cases, in the first case the classifier consists of an SVM based on handcrafted features. In the second case, we considered a CNN fed with a random set of deep learning features extracted from the output of the convolutional layers of the CNN classifier when the final stages are trained on the entre feature set. We tested both solutions by applying them to the cases of histogram equalization and median filtering detection. In both cases, randomization contributes to diminish the transferability of the attacks built of the full feature detector, thus increasing the security of the detection in a grey-box setting. In the CNN case, however, a certain degree of transferability between the full feature and the reduced feature detector is observed. For this reason we enriched the defense by modifying the loss function used to train the reduced feature detector so as to include within it a term that forces the gradient of the detector to be orthogonal to that of the full feature CNN. The experiments we carried out confirm that gradient orthogonalization contributes to further decrease the transferability of the attacks, hence improving the security of the proposed system in the presence of grey-box attacks.
A more detailed description of the approaches described above and the results of the experiments we run to test them, can be found in the 1-page technical summary and the 1-page result summary.

Task: TA1.2 - ENF-Based Video Forensics

Our goal as TA1.2 sub-team is to conduct research on video forensics utilizing electrical network frequency (ENF), and develop ENF-based video authentication method. For this purpose, we mainly target the task of video time-of-recording verification. This task requires a reliable ENF signal estimation from video, followed by a proper similarity test with the ground truth.

For creation of ground-truth ENF database, we developed an ENF recording device that can obtain ENF directly from any mains power outlet in real time at 1 samples/second resolution, and can send the recorded data to a remote server. We currently have 5 active devices recording, 1 in NYU, 1 in Partech, 1 in University of Colorado Denver, and 2 in Turkey.

For an accurate ENF signal estimation from a given video, we first proposed a super-pixel based approach. It estimates ENF from different object regions having very close reflectance properties, i.e., super-pixels, and computes a so-called representative ENF by element-wise median operation of all estimated vectors. The reason we use the term “so-called” is because the estimated signal is initially unknown to be actually an ENF signal. Depending on the similarity of each estimated signal to the representative, it is decided whether the estimated signal is ENF. Hence, it consequently leads to an ENF detector, which achieves a very high accuracy. It can operate on video clips as short as 2 minutes and is independent of the camera sensor type.

Although the proposed super-pixel approach can work for both CCD and CMOS sensors, it is designed as one luminance sample per frame to be obtained. Accordingly, the sampling frequency in this technique is essentially the video frame rate, which does not satisfy the Nyquist criteria for ENF, and has to work with the alias. When the video frame rate is a divisor of the nominal ENF, alias ENF is observed at the 0 Hz DC component. So, an alternative approach is needed. A rolling Shutter based ENF estimation method leads to a sampling frequency as high as number of rows × camera frame rate. Hence, alias ENF is not a phenomenon for this scheme. However, rolling-shutter brings with it the idle period issue. That is, some illumination samples in each frame are missed. Accordingly, we developed an analytical model to explore how the frequency of mains-powered illumination and so ENF is changed and is attenuated in relation to idle period length. Based on this model, we proposed an improved time-of-recording verification technique performing better than the existing methods in the literature. Later on, we further improved the time-of-recording approach. We first explored strong frame rate harmonics. Luminance intensity on consecutive rows of a frame is in the trend of increase or decrease due to the inverse square law, which causes a gap in frame transitions. It consecutively causes strong frame rate harmonics to arise. For videos whose frame rate is a divisor of nominal ENF, ENF harmonics overlap with the frame rate harmonics. To handle this effect, we implemented a second degree polynomial curve fitting to the luminance variation data obtained for each frame. After subtracting sample of luminance value of each row of a frame from the corresponding fitted curve point, a considerable attenuation in the frame rate harmonics were yielded. We further improved the time-of-recording verification approach by classifying quality ENF samples of an estimated ENF signal, and building a mask for them. We adapted the normalized cross-correlation operation accordingly in a way that it works based on the mask. The new approach is distinguishable for noisy ENF signals. It even shows an extraordinary performance for some cases.

TA 1.3 – Provenance Analysis

As a consequence of our research on the topic of Provenance Analysis, we have obtained the following results: generation of four conference papers, three journal papers, one dataset, and two Ph.D. dissertations. All publications were accompanied by source code, properly made available to the scientific community.

In summary, our publications describe individual baseline results from the beginning of the project. Publications present varied enhancements to the solution proposed where we have introduced a complete end-to-end pipeline to solve the problem of provenance analysis, for the first time in the literature. It was with this approach that we have participated in the majority of the evaluations yearly proposed by NIST and DARPA, along the existence of the MediFor program. We also investigated a crowd-sourced dataset of provenance graphs that were collected from Reddit, as an effort to test our solution on more realistic scenarios.

Table 68 summarizes our average results for the task of provenance image filtering, along the four years of project evaluation. All results were obtained with the solution proposed during this program, except for the NC17 baseline results, which were obtained with our previous approach. As one might observe, the proposed technique has improved previous results and was robust across years of challenge, in spite of the changes in dataset size, number of probes, and different types of image transformations.

Table 68. Official results along project years for image filtering.

R@N is the official project metric. It means recall at top N images and should be as close to 1.0 as possible.

Year	Dataset	Size (# images)	Probes (# images)	R@50	R@100	R@200
2017	NC17 (baseline)	~1 million	151	0.552	0.587	0.635
2018	MFC18	~1 million	~10000	0.846	0.882	0.884
2019	MFC19	~1 million	~1000	0.798	0.804	0.814
2020	MFC20	~2 millions	~6000	0.805	0.831	0.855

Table 69 contains the average results of the proposed approach for the task of graph construction, along the four years of official evaluation. Again, the technique was consistently robust across evaluations.

Table 69. Official results across project years for graph construction.

VO, EO, and VEO are the official project metrics and should be as close to 1.0 as possible.
VO means vertex ovelap, EO means edge overlap, and VEO is a combination of both (i.e., graph overlap).

Year	Dataset	Size (# images)	Probes (# images)	VO	EO	VEO
2017	NC17	~1 million	151	0.613	0.208	0.412
2018	MFC18	~1 million	~10000	0.798	0.298	0.553
2019	MFC19	~1 million	~1000	0.703	0.295	0.519
2020	MFC20	~2 millions	~6000	0.714	0.250	0.490

As an outcome of such research, we are currently looking for new applications of provenance analysis to apply our solutions, such as verifying the integrity of images from scientific papers (scientific integrity).

Task: OF – Overhead Forensics

The goal of OF is to develop forensic techniques for overhead image analysis. Indeed, satellite images can be modified in a number of ways, such as inserting objects into an image to hide existing scenes and structures. Moreover, differently from images acquired by consumer cameras, satellite images can be of different nature and be acquired with different technologies. Within this context, we developed a series of tools to solve different overhead forensics problems: RGB image splicing detection and localization; SAR forensics; copy-move forensics.

In order to detect and localize forgeries within RGB images, we developed a series of different solutions. The first one is based on the use of autoencoders. The rationale behind the proposed method is that it is possible to train an autoencoder to obtain a compact representation of image patches coming from pristine satellite pictures. This autoencoder can then be used as a feature extractor for image patches. During testing, a one-class SVM is used to detect whether feature vectors come from pristine images or not, thus representing forgeries. Moreover, we investigated additional techniques fully based on CNNs. For instance, we introduced a method for splicing manipulation detection and localization using deep belief networks (DBNs) that does not require any manipulated data during training. Moreover, we propose to use a Conditional Generative Adversarial Network (cGAN) to detect the presence of spliced forgeries in satellite images.

Considering the problem of localizing forgeries on SAR images, we focused on the scenario of image copy-paste, also in presence of editing operations. This is, given two SAR pictures (a donor and a host), a region is selected from the donor, it is optionally edited, and it is finally pasted on the host image. Given one of these images, our goal is to detect the use of this manipulation technique and localize the forged region. To do so, we leverage a pipeline composed by the following steps: i) we extract a peculiar fingerprint from the SAR image under analysis; ii) we analyze the fingerprint with a binarization technique in order to provide a binary forgery mask. Concerning the first step, we developed a SAR fingerprint extraction technique. This technique is based on a convolutional neural network that acts as a noise extractor but is trained using a siamese procedure. Given a noise-like fingerprint extracted with the fingerprint extractor, we have three different options. Indeed, we developed three different versions of mask estimation algorithms based on clustering, gaussian mixture models and CNNS.

Considering the problem of copy-move (CM), many methods are available in the literature. However, most of them simply highlight duplicated regions within an image not being able to disambiguate between the original and the copy. To address this problem, we proposed a new CNN-based method. Given the binary localization mask produced by a generic copy-move detector, our method permits to derive the actual tampering mask, by identifying the target and source region. The main idea behind the proposed method is to exploit the non-invertibility of the copy-move transformation, due to the presence of interpolation artifacts and local post-processing traces in the displaced region. To do so, we resort to a multi-branch CNN architecture (DisTool) consisting of two main parallel branches, looking for two different kinds of CM-traces.

Task: SI -- Scientific Integrity

The goal of the SI project is to develop forensic techniques that enable analysts to identify misconducts in scientific publications. In this task, we propose a workflow that considers several different problems: PDF content extraction, content segmentation, content ranking, manipulation detection, and provenance analysis. We also introduce a system that implements this workflow.

Considering the problem of PDF content extraction, its execution is divided into two sub-tasks, namely image extraction and image caption extraction. For image extraction, we developed a method to parse PDF streams using MuPDF. We also propose mitigations to the four main problems we find: transparency band error, multiple images error, multiple copies error and scanned pages error. For image caption extraction, we first use PDFMiner to extract paragraphs and their respective metadata. Here we only consider the paragraphs that start with keywords selected by the analyst, like “Figure” or “Fig.”. Then, we rely on the matching of image and caption metadata to associate images and captions.

Considering the problem of content segmentation, we chose a recent data-driven approach introduced by Tsutsui and Crandall. This method has potential to be more robust to the diversity of image composition layouts one may find in scientific papers. The prediction process is learned with a convolutional neural network, called YOLOv2 system. By employing this solution, multi-panel images give origin to multiple sub-images, one for each panel, while single-panel images simply generate a copy of itself. And the panels coming from the same compound image share the same caption.

Considering the problem of content ranking, we rely on techniques from the field of Content-Based Image Retrieval (CBIR) to sort the gallery panels (a set of panels extracted from PDF files). Using content-based approaches allows us to apply these techniques across scientific publications belonging to a variety of disciplines. Exact and almost-duplicate copies are ranked first, then semantically similar panels, and then irrelevant content. Here we recommend matching local features such as SIFT and SURF in the lowest levels of the CBIR representation. When a query image is selected, each of its feature vectors is used to retrieve the k-nearest gallery feature vectors. Each vector can point back to its source image panel and receive a weighted vote. Once all the query feature vectors are processed, votes are summed, and gallery panels are sorted from the most to least voted.

Considering the problem of manipulation detection, we leverage on the dense-field copy-move detector which is proposed for natural images. We carry out a series of modifications, such as proper morphological operations and L_2-norm between RGB patches, in order to deal with the spurious matchings and background, panels and graphs or diagram matches. We also developed a copy-move detector specialized in exposing duplicated western blots. It is not generalized to other types of images, but it is robust to additional editing operations applied after copy.

Considering the problem of provenance analysis, we propose a combination of solutions from U-phylogeny and Beyond pixels to build provenance graphs. Given a selected suspect image panel and a set of available panels of interest, a symmetric adjacency matrix is calculated to express the similarities between every pair of image panels. Then we use Kruskal's algorithm to compute the maximum spanning tree from this matrix. Finally, we use the publication dates of each panel's paper as metadata to determine the orientation of the edges of the obtained spanning tree. The graph links older to newer publications, indicating that newer papers are reusing the content from the older ones.

In this work, we also introduce an extensible, end-to-end, cloud-based scientific integrity system to uncover image-post-processing attempts in scientific publications from the five different aspects under the suspicions and desire of analysts. The system is developed as a web application that uses NGINX webserver to handle the JavaScript user interface. The user interface interacts with the backend via a RESTful application programming interface (API), implemented using Flask. We use RabbitMQ as brokerage middle ware to enable multiple users using the system to handle extensive workloads. And all integrated functionalities are implemented as Docker containers to make the system extensible.

5.1 Delivered Software Products Including Docker Containers

Container Name: Purdue_TA11b_MFCN_6_2019Validation
Brief Description: This container performs localization of image splicing attacks using a multi-task fully convolutional network (MFCN)

Container Name: prov-filter-svc
Available at: https://hub.docker.com/r/dmoreira/prov-filter-svc
Brief Description: Provenance image filtering (image retrieval) service following the specs of the DARPA MediFor program.

Container Name: prov-index-svc
Available at: https://hub.docker.com/r/dmoreira/prov-index-svc
Brief Description: Provenance image indexing (index access) service following the specs of the DARPA MediFor program.

Container Name: ndpurdue:2019_container_validation_0.1
Available at: gitlab-registry.mediforprogram.com/ndpurdue/provenance/ndpurdue:2019_container_validation_0.1
Brief Description: Provenance graph building container following our in-house specs (due to missing program specs).

Container name: Purdue-Polimi MFC19 Video Analytic
Brief description: Manipulation detection is carried out exploiting multimedia streams characteristics. Specifically, we process multimedia streams and generate a feature vector which is later fed to a supervised classifier for global manipulation detection.

Container name: Purdue Polimi MFC20 Video
Brief description: Manipulation detection is carried out exploiting multimedia streams characteristics. Specifically, we process multimedia streams and generate a feature vector which is later fed to a supervised classifier for global manipulation detection.

Container name: Purdue Satellite Anomaly Detection
Brief description: Sat-SVDD a deep learning based method for detecting and localizing splicing manipulations in overhead images.

Container name: Purdue Satellite Splice Detection with Deep Belief Network
Brief description: Uniform Deep Belief Network for detecting and localizing splicing manipulations in overhead images. Our method uses recent advances in anomaly detection and does not require any prior knowledge of the type of manipulations that an adversary could insert in the satellite imagery.

Container name: satellite_mipr
Brief description: The docker contains Conditional Generative Adversarial Network (cGAN) to identify the presence of spliced forgeries within satellite images. Additionally, we identify their

locations and shapes. Trained on pristine and falsified images, our method achieves high success on these detection and localization objectives.

Container name: satellite_forgeries
Brief description: This project page describes our tool used in the paper "Satellite Image Forgery Detection and Localization Using GAN and One-Class Classifier." to create forged satellite images. We also include an additional tool that was not used in the previously mentioned paper that also allows one to generate manipulated overhead imagery.

Container name: sat_one_class_gan
Brief description: The docker method contains generative adversarial network (GAN), which learnt a feature representation of pristine satellite images. A one-class support vector machine (SVM) which is trained on these features to determine their distribution and to detected anomalies.

Container name: Scanner Project
Brief description: Source scanner classification using deep learning method is carried out to identifying the original scanner model used to generate the scanned images/documents. The classification results will be used to generate a reliability map to indicate the suspicious forged region.

Container name: verify-recording-time-v2
Location: https://containereval.mediforprogram.com/analytics/d4445875-5e2f-4720-ac31-17a723c1ef79
Brief description: time of recording verifier -This container verifies the recording time of a given video utilizing Electric Network Frequency (ENF) criteria. It estimates ENF from the given video, and compares the quality samples of the estimated ENF signal with the ground-truth. It requires GPU Different from ENF work we have also submitted the following docker containers for video manipulation detection tasks:

Container name: nyu-vid-manipulation-detect-frame-audio-fusion
Location: https://containereval.mediforprogram.com/analytics/f9004c41-5abb-4ef5-8cd5-7e04e6f74c87
Brief description: Video manipulation detector - fusion Video tamper detection using fusion of three different manipulation detectors. The first and the second detectors use I frame features to detect recompression artifacts. To extract theses features, the first method uses optical flow and the second method uses laplacian filtering and edge blurriness. The third method uses audio channel to detect video manipulation. The model runs on a CPU device.

Container name: nyu-vid-manipulation-detect-opticflow-laplacian
Location: https://containereval.mediforprogram.com/analytics/0efa7628-476b-45ec-ad4a-d6f7aa45aa6f

Brief description: Video manipulation detector - optic flow & edge blurriness This algorithm detects video manipulations. The proposed method determines the group of pictures (GOP) value by looking at the traces left from the previous compression of the video using optical flow and edge blurriness variations. If the estimated GOP value and the video's GOP value differ from each other, the video is marked as tampered. The model runs on a CPU device.

Container name: nyu-vid-manipulation-detect-audio
Location: https://containereval.mediforprogram.com/analytics/706e5769-124d-4ca2-a6e5-605ac9d01fff
Brief description: Audio based video manipulation detector This algorithm searches inconsistencies in the audio channel of a given video due to any performed frame drop / copy / duplicate operations. This algorithm especially detects the frame drop positions.

Container Name: API_migration_ta11a_jpeg_grpc
Location: https://gitlab.mediforprogram.com/mbarni/api_migration_ta11a_jpeg_grpc
Brief description: The software allows to test a forensic detector based on machine learning methods (specifically Support Vector Machines) addressing the following binary hypotheses testing problem: Hypothesis 0: the image has been JPEG compressed once (JPEG camera-native), corresponding to the absence of manipulation. Hypothesis 1: the image has been either compressed twice or compressed, attacked and then compressed again. Specifically, given a JPEG input image, the service produces a score that indicates the probability that the input image is double compressed/attacked.

Container Name: API_migration_ta11a_cenh_loc_grpc
Location: https://gitlab.mediforprogram.com/mbarni/api_migration_ta11a_cenh_loc_grpc
Brief description: JPEG-aware contrast manipulation detection and localisation tool -forensic detector based on machine learning methods (specifically Convolutional Neural Networks) addressing the following binary hypotheses testing problem: Hypothesis 0: the image has undergo contrast manipulation, corresponding to the absence of manipulation. Hypothesis 1: the image has undergo contrast manipulation and (eventually) JPEG compression. Specifically, when presented with JPEG/PNG/TIFF/BMP image data stream, the service produces a score that indicates the probability that the input image is contrast manipulated.

Container Name:vpurdue-ta11a-cmfd-loc-grpc
Location: https://gitlab.mediforprogram.com/mbarni/purdue-ta11a-cmfd-loc-grpc
Brief description: CopyMove - SIFT J-Linkage : Purdue TA1.1a copy-move forgery detection and localisation - Copy-Move manipulation detection, localisation and disambiguation. The algorithm works as follows: (1) Keypoint extraction and matching; (2) Clustering of matches (RANSAC); (3) Localisation of regions underlying clusters; (4) Image post-processing to extract tampering maps.

Container Name: purdue-ta11a-cmfd-loc-dis-grpc
Location: https://gitlab.mediforprogram.com/mbarni/purdue-ta11a-cmfd-loc-dis-grpc

Brief description: CopyMove - Twin CNN Detector: Purdue TA1.1a copy-move forgery detection and source/target disambiguation - The material provided here allows to build a Docker Image for the Purdue-TA1.1a-Copy-Move manipulation detection, localisation and disambiguation. The first part of the detection algorithm works as follows: (1) Keypoint extraction and matching; (2) Clustering of matches (RANSAC); (3) Localisation of regions underlying clusters; (4) Image post-processing to extract tampering maps.The second part of the detection method allows to determine which of the cloned region is the source region (hence pristine pixels) and which is the target region (hence manipulated pixels) by means of machine learning and deep neural networks

Container Name: purdue-ta11a-cmfd-loc-dis-cpu
Location: https://gitlab.mediforprogram.com/purdueunisi/purdue-ta11a-cmfd-loc-dis-cpu
Brief description: Twin CNN Detector: Purdue TA1.1a copy-move forgery detection and source/target disambiguation - CPU-ONLY VERSION added upon request by TA2 for container purdue-ta11a-cmfd-loc-dis-grpc

Container Name: purdue-polimi-camid/fast
Location: https://gitlab.mediforprogram.com/pbestagini/mfc19-purdue-polimi-camid
Brief description: PRNU camera model attribution (shifts) - The system performs a basic photo-response non uniformity (PRNU) test to attribute one picture to a given acquisition device. Possible shifts are taken into account.

Container Name: purdue-polimi-camid/multiple
Location: https://gitlab.mediforprogram.com/pbestagini/mfc19-purdue-polimi-camid
Brief description: PRNU camera model attribution (multi PRNU) - The system performs a basic photo-response non uniformity (PRNU) test to attribute one picture to a given acquisition device. Multiple PRNUs are tested against the image. Possible shifts are taken into account.

Container Name: purdue-polimi-camid/single
Location: https://gitlab.mediforprogram.com/pbestagini/mfc19-purdue-polimi-camid
Brief description: PRNU camera model attribution (shift and resize) - The system performs a photo-response non uniformity (PRNU) test to attribute one picture to a given acquisition device. Possible shifts and resizing are taken into account.

Container Name: purdue-polimi-mfc19-video-base
Location: https://gitlab.mediforprogram.com/dguera/purdue-polimi-mfc19-video
Brief description: This analytic analyzes the multimedia stream descriptor of the video and gives a manipulation probability based on the descriptor similarity to the descriptors of previously uncovered manipulations.
Reference
(https://mediforprogram.com/wiki/download/attachments/1245706/Guera2019_Streams.pdf) D. Güera, S. Baireddy, P. Bestagini, S. Tubaro, E. J. Delp, "We Need No Pixels: Video Manipulation Detection Using Stream Descriptors", International Conference on Machine Learning (ICML), Synthetic Realities: Deep Learning for Detecting AudioVisual Fakes Workshop, Long Beach, CA.

Container Name: purdue-polimi-mfc19-video-ganbase
Location: https://gitlab.mediforprogram.com/dguera/purdue-polimi-mfc19-video
Brief description: Multimedia Stream Descriptor Detector (GAN-Aware) - This analytic analyzes the multimedia stream descriptor of the video and gives a manipulation probability based on the descriptor similarity to the descriptors of previously uncovered manipulations, including DeepFakes and GAN-based manipulations.
Reference
(https://mediforprogram.com/wiki/download/attachments/1245706/Guera2019_Streams.pdf) D. Güera, S. Baireddy, P. Bestagini, S. Tubaro, E. J. Delp, "We Need No Pixels: Video Manipulation Detection Using Stream Descriptors", International Conference on Machine Learning (ICML), Synthetic Realities: Deep Learning for Detecting AudioVisual Fakes Workshop, Long Beach, CA.

Container Name: purdue-polimi-video-codec-base
Location: https://gitlab.mediforprogram.com/pbestagini/mfc19-purdue-polimi-video-codec
Brief description: Codec inconsistency detector - This container detects codec inconsistencies in videos.
Reference
(https://mediforprogram.com/wiki/download/attachments/1245706/Verde2018_Video.pdf) S. Verde, L. Bondi, P. Bestagini, S. Milani, G. Calvagno, S. Tubaro, "Video Codec Forensics Based on Convolutional Neural Networks", IEEE International Conference on Image Processing (ICIP), pp. 530-534, Athens, Greece, October 2018. DOI: 10.1109/ICIP.2018.8451143

Container Name: purdue-polimi-video-codec-cpu
Location: https://gitlab.mediforprogram.com/pbestagini/mfc19-purdue-polimi-video-codec-cpu
Brief description: Codec inconsistency detector (cpu) - This container detects codec inconsistencies in videos. It runs on CPU only
Reference
(https://mediforprogram.com/wiki/download/attachments/1245706/Verde2018_Video.pdf) S. Verde, L. Bondi, P. Bestagini, S. Milani, G. Calvagno, S. Tubaro, "Video Codec Forensics Based on Convolutional Neural Networks", IEEE International Conference on Image Processing (ICIP), pp. 530-534, Athens, Greece, October 2018. DOI: 10.1109/ICIP.2018.8451143

Container Name: purdue-polimi_qmatrix
Location: https://gitlab.mediforprogram.com/pbestagini/api-migration-purdue-polimi_qmatrix
Brief description: Multiple JPEG Compression Detection based on Quantization Matrix - This container detects whether an image has undergone a JPEG compression with a specific implementation seen at training time.
Reference
(https://mediforprogram.com/wiki/download/attachments/1245706/Bonettini2018_jpeg.pdf, https://ieeexplore.ieee.org/document/8630765) N. Bonettini, L. Bondi, P. Bestagini, S. Tubaro, "JPEG Implementation Forensics Based on Eigen-Algorithms", IEEE International Workshop on Information Forensics and Security (WIFS), 2018

6.0 APPENDICES

6.1 Appendix A Purdue MediFor Team Contact Information

To contact the entire Purdue MediFor Team, send an email to *medifor-team@ecn.purdue.edu*

The team web site is located at: *https://engineering.purdue.edu/MEDIFOR/*

6.2 Appendix B PI and Co-PIs Contact Information

Listed below is the complete contact information list for the principal investigator and all of the co-principal investigators on the team.

Edward J. Delp
Purdue University
School of Electrical and Computer Engineering
465 Northwestern Avenue
West Lafayette, Indiana 47907-2035
Telephone Number: 765 494 1740
Fax Telephone Number: 765 494 3358
Email: ace@ecn.purdue.edu

Walter Scheirer
University of Notre Dame
Dept. of Computer Science & Engineering
384 Fitzpatrick Hall
Notre Dame, IN 46556
Email: wscheire@nd.edu
Telephone: +1 574 631 2436

Anderson Rocha
Institute of Computing
University of Campinas (Unicamp)
Albert Einstein, 1251
Campinas, SP, Brazil
13.083-852
E-mail: anderson.rocha@ic.unicamp.br
Telephone: +55 19 3521-2979

Kevin W. Bowyer
384 Fitzpatrick Hall
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, IN 46556

Email: kwb@nd.edu
Telephone: +1 574 631-9978

Patrick J. Flynn
Dept of Computer Science & Engineering
384 Fitzpatrick Hall
University of Notre Dame
Notre Dame, IN 46556 USA
Telephone: 574 631-8803
flynn@nd.edu

C.-C. Jay Kuo
3740 McClintock Ave. Room 400
Ming-Hsieh Department of Electrical Engineering
University of Southern California
Los Angeles, CA 90089-2564
Tel: +1-213-740-4658
Fax:+1-213-740-4651
E-mail: cckuo@sipi.usc.edu

Stefano Tubaro
Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB) www.deib.polimi.it
Politecnico di Milano
Piazza Leonardo da Vinci 32
20133 Milano
Italy
stefano.tubaro@polimi.it
Office Tel. Number: +39 02 2399 3647

Paolo Bestagini
Politecnico di Milano
Dipartimento di Elettronica, Informazione e Bioingegneria
Via Ponzio, 34/5
20133, Milano
Italy
Email: paolo.bestagini@polimi.it
Telephone: +39 02 2399 9654

Mauro Barni
University of Siena,
Department of Information Engineering and Mathematics
Via Roma 56,
53100, Siena
ITALY
e-mail: barni@diism.unisi.it
Telephone: +39 0577 234850 (int. 1005)

Nasir Memon
New York University
Department of Computer Science and Engineering
6 Metrotech Center
Brooklyn, NY 11201.
email: memon@nyu.edu
Telephone: +1 718-260-3970

Luisa Annalisa Verdoliva
Department of Industrial Engineering
University Federico II of Naples
Via Claudio, 21
Naples, 80125, Italy
Email: verdoliv@unina.it
Telephone: +39 081 7683929

Wael AbdAlmageed
Viterbi School of Engineering
University of Southern California
Los Angeles, CA 90089-2564
Tel.: 310-448-8332, 703-248-6174
Email: wamageed@isi.edu

7.0 BIBLIOGRAPHY

Journal Papers

1) C.-C. Jay Kuo, **"Understanding convolutional neural networks with a mathematical model"**, *Journal of Visual Communications and Image Representation, vol. 41, pp. 406-413, Best Paper Award, November 2016*. DOI: 10.1016/J.JVCIR.2016.11.003

2) G. Marmerola, M. Oikawa, Z. Dias, S. Goldenstein, A. Rocha, **"On the Reconstruction of Text Phylogeny Trees: Evaluation and Analysis of Textual Relationships"**, *PLOSONE, vol. 11, no. 12, pp. 1–35, December 2016*. DOI: 10.1371/journal.pone.0167822

3) L. Bondi, L. Baroffio, D. Güera, P. Bestagini, E. J. Delp, S. Tubaro, **"First Steps Towards Camera Model Identification with Convolutional Neural Networks"**, *IEEE Signal Processing Letters, vol. 24, no. 3, pp. 259-263, March 2017*. DOI: 10.1109/LSP.2016.2641006

4) C.-C. Jay Kuo, **"The CNN as a guided multi-layer RECOS transform"**, *IEEE Signal Processing Magazine, vol. 34, no. 3, pp. 81-89, May 2017*. DOI: 10.1109/MSP.2017.2671158

5) S. Taspinar, M. Mohanty, N. Memon, **"PRNU-based camera attribution from multiple seam-carved images"**, *IEEE Transactions on Information Forensics and Security, 12(12), pp.3065-3080, May 2017*. DOI: 10.1109/TIFS.2017.2737961

6) S. Vatansever, A. E. Dirik, N. Memon, **"Detecting the Presence of ENF Signal in Digital Videos: A Superpixel-Based Approach"**, *IEEE Signal Processing Letters, vol. 24, no. 10, pp. 1463-1467, October 2017*. DOI: 10.1109/LSP.2017.2741440

7) M. Barni, L. Bondi, N. Bonettini, P. Bestagini, A. Costanzo, M. Maggini, B. Tondi, S. Tubaro, **"Aligned and non-aligned double JPEG detection using convolutional neural networks"**, *Journal of Visual Communication and Image Representation, vol. 49, pp. 153-163, November 2017*. DOI: 10.1016/J.JVCIR.2017.09.003

8) F. Costa, A. Oliveira, P. Ferrara, Z. Dias, S. Goldenstein, A. Rocha, **"New Dissimilarity Measures for Image Phylogeny Reconstruction"**, *Pattern Analysis and Applications, vol. 20, no. 4, pp. 1289-1305, November 2017*. DOI: 10.1007/s10044-017-0616-9

9) R. Salloum, Y. Ren, C.-C. Jay Kuo, **"Image Splicing Localization using a Multi-task Fully Convolutional Network (MFCN)"**, *Journal of Visual Communication and Image Representation, vol. 51, pp. 201-209, January 2018*. DOI: 10.1016/J.JVCIR.2018.01.010

10) C.-C. Jay Kuo, Y. Chen, **"On data-driven Saak transform"**, *Journal of Visual Communication and Image Representation, vol. 50, pp. 237-246, January 2018*. DOI: 10.1016/J.JVCIR.2017.11.023

11) D. Moreira, A. Bharati, J. Brogan, A. Pinto, M. Parowski, K. W. Bowyer, P. J. Flynn, A. Rocha, W. J. Scheirer, **"Image Provenance Analysis at Scale"**, *IEEE Transactions on Image Processing, vol. 27, no. 12, pp. 6109-6123, July 2018*. DOI: 10.1109/TIP.2018.2865674

12) L. Bondi, P. Bestagini, F. Pérez-González, S. Tubaro, **"Improving PRNU Compression through Preprocessing, Quantization and Coding"**, *IEEE Transactions on Information Forensics and Security, vol. 14, no. 3, pp. 608-620, July 2018*.
DOI: 10.1109/TIFS.2018.2859587

13) M. Barni, H. Santoyo-Garcia, B. Tondi, **"An Improved Statistic for the Pooled Triangle Test against PRNU-Copy Attack"**, *IEEE Signal Processing Letters, vol. 25, no. 10, pp. 1435-1439, October 2018*. DOI: 10.1109/LSP.2018.2863045

14) B. Tondi, **"Pixel-domain Adversarial Examples Against CNN-based Manipulation Detectors"**, *Electronics Letters, Vol. 54, No. 21, pp 1220 - 1222, October 2018*.
DOI: 10.1049/el.2018.6469

15) R. Salloum, C.-C. Jay Kuo, **"Efficient Image Splicing Localization via Contrastive Feature Extraction"**, *arXiv:1901.07172, January 2019*.

16) S. Vatansever, A. E. Dirik and N. Memon, **"Analysis of Rolling Shutter Effect on ENF-Based Video Forensics"**, *IEEE Transactions on Information Forensics and Security, vol. 14, no. 9, pp. 2262-2275, January 2019*. DOI: 10.1109/TIFS.2019.2895540

17) Z. Chen, B. Tondi, X. Li, R. Ni, Y. Zhao, M. Barni, **"Secure Detection of Image Manimpulation by means of Random Feature Selection"**, *IEEE Transactions on Information Forensics and Security, vol. 14, no. 9, pp. 2454-2469, February 2019*.
DOI: 10.1109/TIFS.2019.2901826

18) C.-C. Jay Kuo, M. Zhang, S. Li, J. Duan, Y. Chen, **"Interpretable Convolutional Neural Networks via Feedforward Design"**, *Journal of Visual Communication and Image Representation, vol. 60, pp. 346-359, April 2019*. DOI: 10.1016/J.JVCIR.2019.03.010

19) S. Taspinar, M. Mohanty, N. Memon, **"Source Camera Attribution of Multi-Format Devices"**, *arXiv:1904.01533, April 2019*.

20) S. Mandelli, P. Bestagini, L. Verdoliva, S. Tubaro, **"Facing Device Attribution Problem for Stabilized Video Sequences"**, *IEEE Transactions on Information Forensics and Security (TIFS), vol. 15, pp. 14-27, May 2019*. DOI: 10.1109/TIFS.2019.2918644

21) P. R. M. Júnior, L. Bondi, P. Bestagini, S. Tubaro, A. Rocha, **"An In-Depth Study on Open-Set Camera Model Identification"**, *IEEE Access, vol. 7, pp. 180713-180726, June 2019*.
DOI: 10.1109/ACCESS.2019.2921436

22) S. Taspinar, M. Mohanty, N. Memon, **"Camera Fingerprint Extraction via Spatial Domain Averaged Frames"**,*arXiv:1904.04573, September 2019*.

23) Y. Chen, C.-C. Jay Kuo, **"PixelHop: A Successive Subspace Learning (SSL) Method for Object Classification"**,*Journal of Visual Communication and Image Representation, December 2019*. DOI: 10.1016/J.JVCIR.2019.102749

24) Y.Niu, B. Tondi, Y.Zhao, M.Barni, **"Primary Quantization Matrix Estimation of Double Compressed JPEG Images via CNN"**, *IEEE Signal Processing Letters, vol. 27, pp 191 - 195, December 2019*. DOI: 10.1109/LSP.2019.2962997

25) M.Barni, E. Nowroozi, B. Tondi, **"Improving the Security of Image Manipulation Detection through One-and-a-half-class Multiple Classification"**, *Multimedia Tools and Applications, Springer, vol. 79, pp 2383–2408, January 2020*. DOI: 10.1007/s11042-019-08425-z

26) M. Yankoski, T. Weninger, W. Scheirer, **"An AI early warning system to monitor online disinformation, stop violence, and protect elections"**, *Bulletin of the Atomic Scientists, vol. 76, no. 2, pp. 85-90, March 2020*. DOI: 10.1080/00963402.2020.1728976

27) S. Mandelli, D. Cozzolino, P. Bestagini, L. Verdoliva, S. Tubaro, **“CNN-based fast source device identification,”** *IEEE Signal Processing Letters (SPL), vol. 27, pp. 1285-1289, July 2020.* DOI: 10.1109/LSP.2020.3008855

28) M. Barni, Q. -T. Phan, B. Tondi, **"Copy Move Source-Target Disambiguation Through Multi-Branch CNNs,"** *IEEE Transactions on Information Forensics and Security (TIFS), vol. 16, pp. 1825-1840, December 2020.* DOI: 10.1109/TIFS.2020.3045903

29) A. Bharati, D. Moreira, P. J. Flynn, A. de Rezende Rocha, K. W. Bowyer, W. J. Scheirer, **"Transformation-Aware Embeddings for Image Provenance**," in *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 2493-2507, January 2021. DOI: 10.1109/TIFS.2021.3050061

30) J. Brogan *et al*., **"Fast Local Spatial Verification for Feature-Agnostic Large-Scale Image Retrieval**," in *IEEE Transactions on Image Processing*, vol. 30, pp. 6892-6905, July 2021. DOI: 10.1109/TIP.2021.3097175

31) Y. Niu, B. Tondi, Y. Zhao, R. Ni, M. Barni, **"Image Splicing Detection, Localization and Attribution via JPEG Primary Quantization Matrix Estimation and Clustering,"** in *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 5397-5412, November 2021. DOI: 10.1109/TIFS.2021.3129654

32) E. D. Cannas, N. Bonettini, S. Mandelli, P. Bestagini, S. Tubaro, **"Amplitude SAR Imagery Splicing Localization,"** *IEEE Access, vol. 10, pp. 33882-33899, March 2022.* DOI: 10.1109/ACCESS.2022.3161836

33) L. Abady, J. Horváth, B. Tondi, E. J. Delp, M. Barni, "Manipulation and Generation of Synthetic Satellite Images Using Deep Learning Models", *SPIE Journal of Applied Remote Sensing* (submitted)

Conference Papers

1) F. Costa, S. Lameri, P. Bestagini, Z. Dias, S. Tubaro, A. Rocha, **"Hash-Based Frame Selection for Video Phylogeny"**,*IEEE International Workshop on Information Forensics and Security (WIFS), Abu Dhabi, UAE, December 2016*. DOI: 10.1109/WIFS.2016.7823906

2) M. Barni, Z. Chen, B. Tondi, **Adversary-aware, data-driven detection of double JPEG compression: how to make counter-forensics harder"**, *IEEE International Workshop on Information Forensics and Security (WIFS), Abu Dhabi, UAE, December 2016*.
DOI: 10.1109/WIFS.2016.7823902

3) S Taspinar, M Mohanty, N Memon, **"Source camera attribution using stabilized video",** *IEEE International Workshop on Information Forensics and Security (WIFS), Abu Dhabi, UAE, December 2016*. DOI: 10.1109/WIFS.2016.7823918

4) S. Milani, P. Bestagini, S. Tubaro, **"Phylogenetic analysis of near-duplicate and semantically-similar images using viewpoint localization"**, *IEEE International Workshop on Information Forensics and Security (WIFS), Abu Dhabi, UAE, December 2016*.
DOI: 10.1109/WIFS.2016.7823909

5) L. Bondi, D. Güera, L. Baroffio, P. Bestagini, E. J. Delp, S. Tubaro, **"A Preliminary Study on Convolutional Neural Networks for Camera Model Identification"**, *IS&T Electronic Imaging (EI), vol. 2017, no. 7, pp. 67-76, Burlingame, California, January 2017*.
Preprint: Link DOI: 10.2352/ISSN.2470-1173.2017.7.MWWSF-327

6) D. Güera, Y. Wang, L. Bondi, P. Bestagini, S. Tubaro, E. J. Delp, **"A Counter-Forensic Method for CNN-Based Camera Model Identification"**, *IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW), pp. 1840-1847, Honolulu, Hawaii, July 2017*. DOI: 10.1109/CVPRW.2017.230

7) L. Bondi, S. Lameri, D. Güera, P. Bestagini, E. J. Delp, S. Tubaro, **"Tampering Detection and Localization through Clustering of Camera-Based CNN Features"**, *IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW), pp. 1855-1864, Honolulu, Hawaii, July 2017*. DOI: 10.1109/CVPRW.2017.232

8) S. Milani, P. Bestagini, S. Tubaro, **"Video Phylogeny Tree Reconstruction Using Aging Measures"**, *European Signal Processing Conference (EUSIPCO), pp. 2181-2185, Kos, Greece, August 2017*. DOI: 10.23919/EUSIPCO.2017.8081596

9) M. Barni, E. Nowroozi, B. Tondi, **"Higher-Order, Adversary-Aware, Double JPEG-Detection via Selected Training on Attacked Samples"**, *European Signal Processing Conference (EUSIPCO), pp. 281-285, Kos, Greece, August 2017*.
DOI: 10.23919/EUSIPCO.2017.8081213

10) A. Bharati, D. Moreira, A. Pinto, J. Brogan, K. Bowyer, P. Flynn, W. J. Scheirer, A. Rocha, **"U-Phylogeny: Undirected Provenance Graph Construction in the Wild"**, *IEEE International Conference on Image Processing (ICIP), pp. 1517-1521, Beijing, China, September 2017*. DOI: 10.1109/ICIP.2017.8296535

11) S. Taspinar, H.T. Sencar, S. Bayram, N. Memon, **"Fast Camera Fingerprint Matching in Very Large Databases"**,*IEEE International Conference on Image Processing (ICIP), pp. 4088-4092, Beijing, China, September 2017*. DOI: 10.1109/ICIP.2017.8297051

12) A. Pinto, D. Moreira, A. Bharati, J. Brogan, K. Bowyer, P. Flynn, W. J. Scheirer, A. Rocha, **"Provenance Filtering for Multimedia Phylogeny"**, *IEEE International Conference on Image Processing (ICIP), pp. 1502-1506, Beijing, China, September 2017*. DOI: 10.1109/ICIP.2017.8296532

13) J. Brogan, P. Bestagini, A. Bharati, A. Pinto, D. Moreira, K. Bowyer, P. Flynn, A. Rocha, W. Scheirer, **"Spotting the Difference: Context Retrieval and Analysis for Improved Forgery Detection and Localization"**, *IEEE International Conference on Image Processing (ICIP), pp. 4078-4082, Beijing, China, September 2017*. DOI: 10.1109/ICIP.2017.8297049

14) S. Lameri, L. Bondi, P. Bestagini, S. Tubaro, **"Near-Duplicate Video Detection Exploiting Noise Residual Traces"**,*IEEE International Conference on Image Processing (ICIP), pp. 1497-1501, Beijing, China, September 2017*. DOI: 10.1109/ICIP.2017.8296531

15) S. Mandelli, L. Bondi, S. Lameri, V. Lipari, P. Bestagini, S. Tubaro, **"Inpainting-Based Camera Anonymization"**, *IEEE International Conference on Image Processing (ICIP), pp. 1522-1526, Beijing, China, September 2017*. DOI: 10.1109/ICIP.2017.8296536

16) Z. Chen, B. Tondi, X. Li, R. Ni, Y. Zhao, M. Barni, **"A Gradient-Based Pixel-Domain Attack Against SVM Detection of Global Image Manipulation"**, *IEEE Workshop on Information Forensics and Security (WIFS), Rennes, France, December 2017*. DOI: 10.1109/WIFS.2017.8267668

17) L. Bondi, F. Pérez-González, P. Bestagini, S. Tubaro, **"Design of Projection Matrices for PRNU Compression"**, *IEEE Workshop on Information Forensics and Security (WIFS), Rennes, France, December 2017*. DOI: 10.1109/WIFS.2017.8267652

18) S. Verde, S. Milani, P. Bestagini, S. Tubaro, **"Audio Phylogenetic Analysis using Geometric Transforms"**, *IEEE Workshop on Information Forensics and Security (WIFS), Rennes, France, December 2017*. DOI: 10.1109/WIFS.2017.8267650

19) S. K. Yarlagadda, D. Güera, P. Bestagini, F. Zhu, S. Tubaro, E. J. Delp, **"Satellite Image Forgery Detection and Localization Using GAN and One-Class Classifier"**, *IS&T Electronic Imaging (EI), vol. 2018, no. 7, pp. 214-1-214-9, Burlingame, California, January 2018*. DOI: 10.2352/ISSN.2470-1173.2018.07.MWSF-214

20) D. Güera, S. K. Yarlagadda, P. Bestagini, F. Zhu, S. Tubaro, E. J. Delp, **"Reliability Map Estimation For CNN-Based Camera Model Attribution"**, *IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 964-973, Lake Tahoe, Nevada, February 2018*. DOI: 10.1109/WACV.2018.00111

21) S. Mandelli, N. Bonettini, P. Bestagini, V. Lipari, S. Tubaro, **"Multiple JPEG Compression Detection through Task-driven Non-negative Matrix Factorization"**, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 2106-2110, Calgary, Canada, April 2018*. DOI: 10.1109/ICASSP.2018.8461904

22) M. Barni, E. Nowroozi, B. Tondi, **"Detection of Adaptive Histogram Equalization Robust Against JPEG Compression"**, *IAPR/IEEE International Workshop on Biometrics and Forensics (IWBF), pp. 1-8, Sassari, Italy, June 2018*. DOI: 10.1109/IWBF.2018.8401564

23) M. Barni, M. C. Stamm, B. Tondi, **"Adversarial Multimedia Forensics: Overview and Challenges Ahead"**, *European Signal Processing Conference (EUSIPCO), pp. 962-966, Rome, Italy, September 2018*. DOI: 10.23919/EUSIPCO.2018.8553305

24) N. Bonettini, L. Bondi, D. Güera, S. Mandelli, P. Bestagini, S. Tubaro, E. J. Delp, **"Fooling PRNU-Based Detectors Through Convolutional Neural Networks"**, *European Signal Processing Conference (EUSIPCO), Rome, Italy, September 2018*.
DOI: 10.23919/EUSIPCO.2018.8553596

25) S. Mandelli, D. Cozzolino, P. Bestagini, L. Verdoliva, S. Tubaro, **"Blind Detection and Localization of Video Temporal Splicing Exploiting Sensor-Based Footprints"**, *European Signal Processing Conference (EUSIPCO), Rome, Italy, September 2018*.
DOI: 10.23919/EUSIPCO.2018.8553511

26) M. Barni, A. Costanzo, E. Nowroozi, B. Tondi, **"CNN-based Detection of Generic Contrast Adjustment with JPEG Post-processing"**, *IEEE International Conference on Image Processing (ICIP), pp. 3803-3807, Athens, Greece, October 2018*.
DOI: 10.1109/ICIP.2018.8451698

27) W. Yaqub, M. Mohanty, N. Memon, **"Towards camera identification from cropped query images"**, *IEEE International Conference on Image Processing (ICIP), pp. 3798-3802, Athens, Greece, October 2018*. DOI: 10.1109/ICIP.2018.8451749

28) S. Verde, L. Bondi, P. Bestagini, S. Milani, G. Calvagno, S. Tubaro, **"Video Codec Forensics Based on Convolutional Neural Networks"**, *IEEE International Conference on Image Processing (ICIP), pp. 530-534, Athens, Greece, October 2018*.
DOI: 10.1109/ICIP.2018.8451143

29) D. Güera, E. J. Delp, **"Deepfake Video Detection Using Recurrent Neural Networks"**, *IEEE International Conference on Advanced Video and Signal-based Surveillance (AVSS), Auckland, New Zealand, November 2018*. DOI: 10.1109/AVSS.2018.8639163

30) M. Barni, M. Nakano-Miyatake, H.S-Garcıa, B. Tondi, **"Countering the Pooled Triangle Test for PRNU-based camera identification"**, *IEEE Workshop on Information Forensics and Security (WIFS), Hong Kong, December 2018*. DOI: 10.1109/WIFS.2018.8630778

31) N. Bonettini, L. Bondi, P. Bestagini, S. Tubaro, **"JPEG Implementation Forensics Based on Eigen-Algorithms"**, *IEEE Workshop on Information Forensics and Security (WIFS), Hong Kong, December 2018.* DOI: 10.1109/WIFS.2018.8630765

32) A. Bharati, D. Moreira, J. Brogan, P. Hale, K. Bowyer, P. Flynn, A. Rocha, W. J. Scheirer, **"Beyond Pixels: Image Provenance Analysis Leveraging Metadata,"** *IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, Hawaii, January 2019.* DOI: 10.1109/WACV.2019.00185

33) E. R. Bartusiak, S. K. Yarlagadda, D. Güera, F. Zhu, P. Bestagini, S. Tubaro, E. J. Delp, **"Splicing Detection and Localization In Satellite Imagery Using Conditional GANs"**, *IEEE International Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, California, March 2019*. DOI: 10.1109/MIPR.2019.00024

34) S. Vatansever, A. E. Dirik, N. Memon, **"Factors Affecting ENF Based Time-of-recording Estimation for Video"**, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 2497-2501, Brighton, UK, May 2019.* DOI: 10.1109/ICASSP.2019.8682419

35) M. Barni, E. Nowroozi, K. Kallas, B. Tondi, **"On the Transferability of Adversarial Examples Against CNN-Based Image Forensics"**, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brighton, UK, May 2019.* DOI: 10.1109/ICASSP.2019.8683772

36) S. K. Yarlagadda, D. Güera, D. Mas Montserrat, F. Zhu, P. Bestagini, S. Tubaro, E. J. Delp, **"Shadow Removal Detection And Localization For Forensics Analysis"**, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brighton, UK, May 2019*. DOI: 10.1109/ICASSP.2019.8683695

37) A. Lieto, D. Moro, F. Devoti, C. Parera, V. Lipari, P. Bestagini, S. Tubaro, **"“Hello? Who Am I Talking To?” A Shallow CNN Approach for Human vs. Bot Speech Classification"**, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brighton, UK, May 2019.* DOI: 10.1109/ICASSP.2019.8682743

38) D. Güera, S. Baireddy, P. Bestagini, S. Tubaro, E. J. Delp, **"We Need No Pixels: Video Manipulation Detection Using Stream Descriptors"**, *International Conference on Machine Learning (ICML), Synthetic Realities: Deep Learning for Detecting AudioVisual Fakes Workshop, Long Beach, California, June 2019*. URL: arXiv:1906.08743

39) J. Horváth, D. Güera, S. K. Yarlagadda, P. Bestagini, F. Zhu, S. Tubaro, E. J. Delp, **"Anomaly-based Manipulation Detection in Satellite Images"**, *IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Media Forensics (CVPRW), Long*

Beach, California, June 2019.
URL: http://openaccess.thecvf.com/content_CVPRW_2019/html/Media_Forensics/Horvath_Anomaly-Based_Manipulation_Detection_in_Satellite_Images_CVPRW_2019_paper.html

40) N. Bonettini, D. Güera, L. Bondi, P. Bestagini, E. J. Delp, S. Tubaro, **"Image Anonymization Detection With Deep Handcrafted Features"**, *IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, September 2019*.
DOI: 10.1109/ICIP.2019.8804294

41) P. R. M. Júnior, L. Bondi, P. Bestagini, A. Rocha, S. Tubaro, **"A Prnu-Based Method to Expose Video Device Compositions in Open-Set Setups"**, *IEEE International Conference on Image Processing (ICIP)*, *Taipei, Taiwan, September 2019.* DOI: 10.1109/ICIP.2019.8802981

42) G. Pinheiro, M. Cirne, P. Bestagini, S. Tubaro, A. Rocha, **"Detection and Synchronization of Video Sequences for Event Reconstruction"**, *IEEE International Conference on Image Processing (ICIP)*, *Taipei, Taiwan, September 2019.* DOI: 10.1109/ICIP.2019.8803545

43) M. Barni, D. Huang, B. Li and B. Tondi, **"Adversarial CNN Training Under JPEG Laundering Attacks: a Game-Theoretic Approach"**, *IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1-6, Delft, Netherlands, December 2019.*
DOI: 10.1109/WIFS47025.2019.9035110

44) C. Borrelli, P. Bestagini, F. Antonacci, A. Sarti, S. Tubaro, **"Automatic Reliability Estimation for Speech Audio Surveillance Recordings"**, *IEEE Workshop on Information Forensics and Security (WIFS)*, *Delft, Netherlands, December 2019*. DOI: 10.1109/WIFS47025.2019.9034986

45) R. Shao, E. J. Delp, **"Forensic Scanner Identification Using Machine Learning"**, *IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), Santa Fe, New Mexico, March 2020*. https://arxiv.org/abs/2002.02079

46) M. Barni, E. Nowroozi, B. Tondi, B. Zhang, **"Effectiveness of Random Deep Feature Selection for Securing Image Manipulation Detectors Against Adversarial Examples"**, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 2977-2981, Barcelona, Spain, May 2020.*
DOI: 10.1109/ICASSP40776.2020.9053318

47) A. G. Poyraz, A. E. Dirik, A. Karakucuk, N. Memon, **"Fusion of Camera Model and Source Device Specific Forensic Methods for Improved Tamper Detection"**. *arXiv:2002.10123, May 2020*

48) J. Horváth, D. Mas Montserrat, H. Hao, E. Delp, **"Manipulation Detection in Satellite Images Using Deep Belief Networks"**, *IEEE Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, Washington, June 2020.*https://arxiv.org/abs/2004.12441

49) D. Mas Montserrat, H. Hao, S. K. Yarlagadda, S. Baireddy, R. Shao, J. Horváth, E. Bartusiak, J. Yang, D. Güera, F. Zhu, E. J. Delp, **"Deepfakes Detection with Automatic Face Weighting"**, *IEEE Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, Washington, June 2020.* https://arxiv.org/abs/2004.12027

50) L. Abady, M. Barni, A. Garzelli, B. Tondi, **"GAN Generation of Synthetic Multispectral Satellite Images,"** *SPIE Image and Signal Processing for Remote Sensing XXVI, vol. 11533, pp. 122-133, Online, September 2020.* DOI: 10.1117/12.2575765

51) S. Mandelli, F. Argenti, P. Bestagini, M. Iuliani, A. Piva, S. Tubaro, **"A Modified Fourier-Mellin Approach for Source Device Identification on Stabilized Videos,"** *in IEEE International Conference on Image Processing (ICIP), Abu Dhabi, UAE, October 2020.* DOI: 10.1109/ICIP40778.2020.9191001

52) E.D. Cannas, S. Baireddy, E. R. Bartusiak, S.K. Yarlagadda, D. Mas Montserrat, P. Bestagini, S. Tubaro, E. J. Delp, **"Open-Set Source Attribution for Panchromatic Satellite Imagery",** *IEEE International Conference on Image Processing, pp. 3038-3042, Anchorage, AK, September 2021.* DOI: 10.1109/ICIP42928.2021.9506600

Theses

1) E. R. Bartusiak, **"An Adversarial Approach to Sliced Forgery Detection and Localization in Satellite Imagery"**, *Master's dissertation, Purdue University, West Lafayette, IN, May 2019.* URL: https://engineering.purdue.edu/~ace/thesis/emily/emily-thesis-april-2019.pdf

2) Joel Brogan, "**Advancing Biometrics and Image Forensics Through Vision and Learning Systems"**, *Ph.D. Thesis, University of Notre Dame, South Bend, IN, Summer 2019.* https://curate.nd.edu/show/2v23vt17r9s

3) D. Güera, **"Media Forensics Using Machine Learning Approaches"**, *PhD dissertation, Purdue University, West Lafayette, IN, December 2019.*
URL: https://engineering.purdue.edu/~ace/thesis/david-guera/david-thesis-final.pdf

4) E. Nowroozi, **"Machine Learning Techniques for Image Forensics in Adversarial Setting"**, *PhD dissertation, University of Siena, Italy, April 2020.*
URL: http://clem.dii.unisi.it/~vipp/website_resources/theses/Phd_thesis_EhsanNowroozi.pdf

5) D. Mas Montserrat, **"Machine Learning-Based Multimedia Analytics"**, *PhD dissertation, Purdue University, West Lafayette, IN, June 2020.*
URL: https://engineering.purdue.edu/~ace/thesis/danni/daniel-thesis-final.pdf

6) Aparna Bharati, **"Vision and Learning-based Methods for Image Forensics"** *Ph.D. Thesis, University of Notre Dame, South Bend, IN, Summer 2020.*

8.0 LIST OF ABBREVIATIONS AND ACRONYMS

1.5C	One-and-half Class classifier
1C	One Class classifier
2C	Two Class classifier

a.k.a	also known as
ACM	Association for Computing Machinery
AFW	Automatic Face Weighting
AHE	Adaptive Histogram Equalization
Al	Aligned
AP	Average Precision
AC	Alternating Current
API	Application Programming Interface
AR	Augmented Reality
AUC	Area Under the Curve
AVSS	IEEE Conference on Advanced Video and Signal-based Surveillance

BCE	Binary Cross Entropy
BERT	Bidirectional Encoder Representations from Transformers
BMP	Bitmap
BP	Backpropagation
BS	BertScore

CASIA	Chinese Academy of Sciences, Institute of Automation
CBIR	Content-Based Image Retrieval
CC-PEV	Cartesian Calibrated features by PEVny
CCD	Charged-Couple Device
CF	Counter-Forensic
CFA	Color Filter Array
CFL	Compact Fluorescent Light
cGAN	Conditional Generative Adversarial Network
cPCA	Contrastive Principal Component Analysis
CIFAR	Canadian Institute for Advanced Research
Cl-AHE	Contrast Limited Adaptive Histogram Equalization
CM	Copy-Move
CMFD	Copy-Move Forgery Detection
CMOS	Complementary Metal-Oxide Semiconductor
CMYK	Cyan, Magenta, Yellow, and Key (black)
CNN	Convolutional Neural Network
Conv-LSTM	Convolutional Long Short-Term Memory
CPU	Central Processing Unit
CRAFT	Character Region Awareness for Text Detection
CRSPAM	ColoR Subtractive Pixel Adjacency Matrix
CVPRW	IEEE Computer Vision and Pattern Recognition Workshop
CW-SSIM	Complex Wavelet Structural Similarity Index

CycleGAN	Cycle-Consistent Generative Adversarial Network
D-JPEG	Double JPEG
DARPA	Defense Advanced Research Projects Agency
DB	DataBase
DBN	Deep Belief Network
DC	Direct Current
DCT	Discrete Cosine Transform
DDPM	Denoising Diffusion Probabilistic Model
DE	Deep Ensemble
DELF	DEep Local Feature
DF-CMFD	Dense Field Copy-Move Forgery Detection
DFDC	DeepFake Detection Challenge
DL	Deep Learning
DoD	Department of Defense
DOI	Digital Object Identifier
DSURF	Speeded Up Robust Features Detector
EAST	Efficient and Accurate Scene Text Detector
ECCV	European Conference on Computer Vision
EDF	Efficient Dense-Field
EFF	EfficientNet
ELA	Error Level Analysis
ENF	Electric Network Frequency
EO	Edge Overlap
EPS	Encapsulated PostScript
EUSIPCO	European Signal Processing Conference
FC	Fully Connected
FCN	Fully Convolutional Network
FE-CMFD-HFPM	Fast and Effective Image Copy-Move Forgery Detection via Hierarchical Feature Point Matching
FF	Feed Forward
FGS	Fast Gradient Sign
FGSM	Fast Gradient Sign Method
FM	Fourier Mellin transform
FP	False Positive
GAN	Generative Adversarial Network
GIF	Graphics Interchange Format
GIMP	General Image Manipulation Program
GOP	Group of Pictures
GPU	Graphics Processing Unit
GRU	Gated Recurrent Unit
GUI	Graphical User Interface

H0	Pristine Class
H1	Manipulated Class
HHS	US Department of Health and Human Services
HS	Histogram Stretching
IAPR	International Association of Pattern Recognition
ICASSP	IEEE International Conference on Acoustics, Speech and Signal Processing
ICIP	IEEE International Conference on Image Processing
ICML	International Conference on Machine Learning
ICPR	International Conference on Pattern Recognition
IEEE	Institute of Electrical and Electronics Engineers
IH&MMSEC	IEEE Conference on Information Hiding and Multimedia Security
INC	Inception
IoU	Intersection over Union
IR	Image Recall
IS&T	Society for Information Systems & Technology
ISI	Information Sciences Institute
IWBF	IEEE International Workshop on Biometrics and Forensics
JPEG	Joint Photographic Experts Group
JPG	Joint Photographic Experts Group
JS	JavaScript
JSMA	Jacobian-based Saliency Map Attack
JSON	JavaScript Object Notation
JWT	JSON Web Token
KLT	Karhunen-Loéve Transform
KNN	k-Nearest Neighbor
L-BFGS	Limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm
LAG	Label-Assisted Regression
LD	Levenshtein Distance
LDA	Linear Discriminant Analysis
LED	Light Emitting Diode
LSTM	Long Short-Term Memory
MCC	Matthews Correlation Coefficient
MCD	Monte Carlo Dropout
MF	Median Filtering
MFC	Media Forensic Challenge
MFCN	Multi-task Fully Convolutional Network
MNIST	Modified National Institute of Standards and Technology
MFM	Modified Fourier-Mellin transform
MLP	Multilayer Perceptron
MPA	Most Powerful Attack

MQ	Message Queuing
MR	Morphological Reconstruction
MSD/STRD	Copy-Move Similarity Detection and Source/Target Regions Distinguishment Networks
MSE	Mean Square Error
MTAP	Multimedia Tools and Applications
MTCNN	Multi-Task Cascaded Convolutional Neural Networks
MV-Net	MultiView Matching Network
MVC	Model-View-Controller

NAl	Non-Aligned
NAS	Neural Architecture Search
NC	Nimble Challenge
NDVI	Normalized Difference Vegetation Index
NICEGAN	No Independent Component for Encoding Generative Adversarial Network
NIST	National Institute of Standards and Technology
NMI	Normalized Mutual Information
NN	Neural Network
NoSQL	Not only SQL
NYU	New York University

OCR	Optical Character Recognition
ORI	Office of Research Integrity
OS2OS	Objects in Scene to Objects in Scene

P/R	Precision/Recall
PCA	Principal Component Analysis
PCE	Peak to Correlation Energy
PCN	Pair-wise Correlation Net
PDF	Portable Document Format
PGD	Projected Gradient Descent
PIL	Python Imaging Library
PixelCNN	Pixel Convolutional Neural Network
PNG	Portable Network Graphics
PoliMi	Polytechnic University of Milan
PR	Precision-Recall
PRNU	Photo-Response Non-Uniformity
PS	Photoshop
PSNR	Peak Signal-to-Noise Ratio

QF	JPEG Quality Factor
QF1	First JPEG Quality Factor
QF2	Second JPEG Quality Factor

R-CNN	Regions with Convolutional Neural Networks

RAISE	Raw Images Dataset
RANSAC	Random Sample Consensus
RBM	Restricted Boltzmann Machine
RDFS	Random Deep Feature Selection
RECOS	Rectified-Correlations on a Sphere
ReLU	Rectified Linear Unit
REST	Representational State Transfer
RFS	Random Feature Selection
RGB	Red, Green, and Blue
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
RST	Rotation-Scaling-Translation
SAR	Synthetic Aperture Radar
Sat-SVDD	Satellite Support Vector Data Descriptor
SC	Spectral Clustering
SCSM	Spatially Constrained Similarity Measure
SFCN	Single-task Fully Convolutional Network
SIFT	Scale-Invariant Feature Transform
SILA	A System for Scientific Image Analysis
SLIC	Simple Linear Iterative Clustering
SP	Scientific Papers
SPAM	Subtractive Pixel Adjacency Model
SPL	Signal Processing Letters
SSIAI	Southwest Symposium on Image Analysis and Interpretation
SSL	Secure Socket Layer
SSL	Successive Subspace Learning
STFT	Short-Time Fourier Transform
SURF	Speeded Up Robust Features
SV-Net	Single View Matching Network
SVDD	Support Vector Data Descriptor
SVG	Scalable Vector Graphics
SVM	Support Vector Machine
T-IFS	IEEE Transactions on Information Forensics and Security
T-IP	IEEE Transactions on Image Processing
T-MM	IEEE Transactions on Multimedia
t-SNE	t-Distributed Stochastic Neighbor Embedding
TA1	Technical Area 1
TIFF	Tag Image File Format
TIFS	IEEE Transactions on Information Forensics and Security
TNMF	Task-driven Non-negative Matrix Factorization
TP	True Positive
U-Net	Convolutional Networks for Biomedical Image Segmentation
UCID	Uncompressed Colour Image Database

Unicamp	The State University of Campinas
Unina	University of Naples Federico II
USC	University of Southern California
UU-DBN	Uniform-Uniform Deep Belief Network
VEO	Vertex and Edge Overlap
VGG	Visual Geometry Group
VO	Vertex Overlap
WACV	IEEE Winter Conference on Applications of Computer Vision
WIFS	IEEE Workshop on Information Forensics and Security
γ Corr	Gamma Correction