TOWARDS AUTOMATION OF TIPPING AND CUEING BETWEEN SMALL SATELLITES IN A CONSTELLATION

THESIS

Cassandra R. Post, 2nd Lieutenant, USAF

AFIT-ENG-MS-17-M-061

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENG-MS-17-M-061

TOWARDS AUTOMATION OF TIPPING AND CUEING BETWEEN SMALL SATELLITES IN A CONSTELLATION

THESIS

Presented to the Faculty

Department of Electrical Engineering and Computer Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Master of Science in Cyber Operations

Cassandra R. Post, B.S.C.E.

2nd Lieutenant, USAF

March 2017

AFIT-ENG-MS-17-M-061

TOWARDS AUTOMATION OF TIPPING AND CUEING BETWEEN SMALL SATELLITES IN A CONSTELLATION

THESIS

Cassandra R. Post, B.S.C.E.
2nd Lieutenant, USAF

Committee Membership:

Kenneth M. Hopkinson, PhD
Chair

Bryan J. Steward, PhD
Member

Maj Alan C. Lin, PhD
Member

AFIT-ENG-MS-17-M-061

Abstract

The use of low-fidelity sensors of satellites in a constellation for accurate surface target detection has the potential to lower costs while increasing flexibility, replacement time, and fault tolerance. This thesis investigates the possibility of utilizing an array of satellites with a heterogeneous mix of sensor types to optimize the validation process of surface target detection. Generation of synthetic scenes allows identification and extraction of optical features that are useful in remote sensing practices. Features of interest are generated from specified resolutions, representing different sensor types using Rochester Institute of Technology's Digital Imaging Remote Sensing Image Generation platform. Synthetic images are jittered to varying degrees to represent the pointing stability, which is one measure of performance of low-fidelity sensors.

These synthetic images are utilized to train an artificial intelligence platform to automatically detect targets of interest on the earth's surface. The Berkeley Caffe Convolution-Based Deep Learning open source platform is trained and employed to automatically detect features of interest. Berkeley Caffe is fast, powerful, and well-supported with deployed projects by major corporations including Pinterest, Google, and others, which makes it particularly valuable for our experiments. The main contributions for this research effort are four-fold: (1) Generation of a dataset which includes synthetic imagery, (2) Schematic of a Tip and Cue Communication Protocol, (3) A satellite imagery classifier with 92.1% test set accuracy, and (4)Successfully modeled uncorrelated jitter impact on classification performance. Through these experiments, we demonstrate that remote sensing platforms can provide features of interest to an artificial intelligence platform to increase overall feature identification

effectiveness. This is useful for the verification and validation for targets of interest in various applications.

Dedicated to my God.

Acknowledgements

I would like to express my sincere appreciation to my research advisor, Dr. Hopkinson, for his constant enthusiasm, encouragement and guidance throughout the course of this research effort. The insight and wisdom was deeply appreciated. I would, also, like to thank my sponsor Air Force Research Laboratories for making this research possible.

I am also indebted to the students and Department of Electrical Engineering and Computer Engineering faculty who shared their precious time by offering their insights when I found myself presented with a wall. Finally, I would like to express my sincere appreciation to all of the individuals who have also spent their valuable time in mentoring and encouraging me during this research effort.

Cassandra R. Post

Table of Contents

Page

Page

Page

List of Figures

List of Tables

Table Page

TOWARDS AUTOMATION OF TIPPING AND CUEING BETWEEN SMALL SATELLITES IN A CONSTELLATION

I. Introduction

Presented is the investigation of the potential in utilization of Machine Learning in application to the space vehicle domain where a heterogeneous set of varied fidelity sensor types in a constellation of small satellites enhances accurate surface target detection in the case of an Emergency Weather scenario. Remote sensing techniques are exercised in application of the remote sensing platform, DIRSIG[2]. This remote sensing platform produces features of interest from specified degrees of clarity to compose a database used to test and examine the classification accuracy of the artificial intelligence platform, Berkeley Caffe[9]. This artificial intelligence platform has been adapted into a single artificial neural network intended to model a heterogeneous set of varied fidelity sensor types. Variance in fidelity is demonstrated through the collection of satellite imagery with varying ranges of pixel resolution and in the generation of synthetic scenes with varying jitter levels. Data utilized emulates the basic terrain in Emergency Weather scenarios. This chapter covers the background information required to set the stage for this investigation. The Motivation, Problem Statement, Research Objectives, Contributions, Research Focus, Methodology, Assumptions, Limitations, and Implications are described in the following sections.

1.1 Background

Components that belong to the proposed framework stem from various sources originating in diverse fields from differing domains. The details drawn from these fields to provide for the framework makeup is described.

Convolutional Neural Networks.

A Convolutional Neural Network is a variant of the Artificial Neural Network field, where layers inside the platform were intended to follow the design of the mammalian cortex. Currently, there are several Convolutional Neural Networks in use in daily life that power recommendations for various software applications. One such Convolutional Neural Network of interest for the research investigation, Berkeley Caffe[9], developed by the Berkeley Vision and Learning Center is applied as the core of the proposed solution framework, which serves as a barrier to determine the classification of imagery that is sent to it, by giving a probability percentage to measure classification certainty.

Satellites in Constellation.

A single satellite is only able to cover a limited portion of the Earth's surface, for instance a Geo-stationary (GEO) satellite can only cover 30% of the Earth's surface [10]. For more complete coverage you need a number of satellites – a satellite constellation. We can describe a satellite constellation as a number of similar satellites, of a similar type and function, designed to be in similar, complementary, orbits for a shared purpose, under shared control. Satellite constellations have been proposed and implemented for use in communications, including networking [10]. Further, details for the different levels of orbit relevant to this research is provided in the Literature Review Chapter.

Present day satellites in constellation serve multiple purposes. Disaster monitoring, voice and data coverage for satellite phones, remote sensing are a few of the practical uses that satellites in constellation provide. Satellites of varying mass class are utilized in a constellation. The proposed solution framework consists of satellites in the nano-satellites class. Satellites in this mass class range are typically utilized in Low Earth Orbits(LEO), where the proximity to the Earth's atmosphere intensifies the angular velocity of the satellite and allows it to scan the earth's surface for only a brief amount of time. This serves as a point of provision for the proposed framework where, satellites in a tight LEO orbit are tipped by satellites in a farther orbit to a location to scan, and may be cued for the precise moment to scan the surface.

Tipping and Cueing.

The concept tipping and cueing is heavily utilized in Activity-Based Intelligence(ABI), where information gathered is focused on transactions which connect entities and objects over space and time in real time[11]. The digital age has generated a vast amount of data at an exponential rate. According to the Cisco Virtual Networking Index, annual global IP traffic will reach 2.3 Zettabytes by year 2020[12]. The continual growth and influx of data poses a difficult challenge in the efficient discovery of significant events or patterns of life. Tipping and cueing served as an efficient solution to such a problem in [13] through the usage of highlighting key characteristics of a target image in a high-volume database of images which can be utilized to tip off an intelligence analyst or algorithm to circumvent needless searching efforts. Cueing involves the creation of a list of key characteristics to look for in the image database. This idea of persisting data collection of a target is under the umbrella of persistent surveillance or persistent intelligence, surveillance and reconnaissance(ISR)[11] and served as inspiration for developing the proposed solution framework.

Remote Sensing.

Remote Sensing has been around since the early 1940's where application included image classification. Great strides have been made in progressing from the usage of air balloons in the allied European theater for gathering intelligence on targets of interest. The current use for remote sensing techniques is multi-fold from the prediction of agriculture yields[14] to measuring the melting of ice in the polar ice caps[15] and image detection of targets such as airports as demonstrated by the Chinese Navy Department [16].

DIRSIG.

Digital Imaging and Remote Sensing Generation(DIRSIG), is a complex synthetic image generation application which produces simulated imagery in the visible through thermal infrared regions. The original intent of this model is to produce broad-band, multi-spectral and hyper-spectral imagery through the integration of a suite of first principles based radiation propagation sub-models. For the purpose of testing the framework design, the application is utilized to generate a series of synthetic spectral images which simulate the output of a low-fidelity sensor aboard a Geostationary Earth Orbit Search and Rescue (GEOSAR) or Low Earth Orbit Search and Rescue (LEOSAR) Satellite in a constellation formation.

1.2 Motivation

The main thread of this research effort is in the application of Machine Learning to aid understanding in the Search and rescue domain. The motivation of this research effort is threefold: (1) Proliferation of nanosatellites, (2) Tipping and Cueing between small satellites in a constellation, (3) Application of the system to optimize the Search and Rescue process.

Nanosatellite Proliferation.

The development of small satellites such as nanosatellites indicate the potential for tipping and cueing between small satellites in a constellation. Strides have been made in the development of nanosatellites, specifically, nanosatellite of interest to the research effort is the integration of a Google Nexus One Smartphone with a STRAND-1 cubesat [17]. The Surrey Training Research and Nanosatellite Demonstrator (STRAND) has been expanded to a second integration where the internals of a cubesat were integrated with a Rhaspberry Pi [18]. Currently, there is development in creation of a Convolutional Neural Network that can be run on a Rhaspberry Pi [19]. This indicates the potential for a satellite to be equipped with a Convolutional Neural Network in the future.

Tipping and Cueing.

The inspiration for creation of this framework stemmed from the concept of tipping and cueing which is heavily utilized in the Activity Based Intelligence field. This is where information is gathered with a focus on transactions which connect entities and objects over space and time in real time. As we are currently in an era of continual growth of data, a challenge is posed in the determination of what information is most meaningful for analysis.

Optimization of the Search and Rescue Process.

The current search and rescue practices depend heavily on human interactions. Several steps are involved in the process and the proposed solution framework can provide a faster turn-around time in the event of an event of distress. Consider the COSPAS-SARSAT system where users depend completely on distress beacons to emit signals that are sent to Search and rescue satellites. These signals are reflected back

to the surface for processing by 2 data centers. The distress signal is sent to a Local User Terminal on the ground surface and is then sent to a Mission Control Center to receive and process the data. Finally, a rescue Coordination center evaluates the information and in the case of an actual emergency sends for help.

The application of the solution framework to the Search and Rescue domain requires automated image understanding. The main motivators for this research effort integrate with the overall problem statement described in the subsequent section.

1.3 Problem Statement

Defense satellites are traditionally large, powerful, and robust. The possibility of replacing a single large satellite with many smaller and less capable satellites has the potential to lower costs while increasing flexibility, replacement time, and fault-tolerance. This is similar to the way that RAID systems have replaced expensive robust hard disk drives with arrays of inexpensive alternatives. Through the experimental design, results collected and conclusions drawn, Machine Learning demonstrates it's suitability in the space vehicle domain as a source of automation for the tipping and cueing between small satellites in a constellation to increase surface target detection effectiveness. The automation of image understanding towards the tipping and cueing between small satellites in a constellation is investigated in the subsequent section.

1.4 Research Objectives/Questions/Hypotheses

The thesis statement for this research effort is as follows: Machine learning platforms can facilitate the tipping and cueing between heterogeneous constellation members in order to increase its overall feature identification effectiveness. In order to determine a means of support for the given thesis statement an investigation of the

possibility of such a technological advance must be conducted. This thesis investigation concentrates on one specific surface target detection problem. Specifically, the optical detection of targets of interest by an array of satellites with a heterogeneous mix of sensor types. Synthetic data emulating the basic terrain where National Emergency scenarios have been observed to occur for the experiments. This also provides ground truth data. RIT's Digital Imaging and Remote Sensing Image Generation (DIRSIG) will accomplish the need to generate such synthetic data. The system allows for synthetic ground images to be constructed and the application of filters to simulate different resolutions and jitter/clutter/timing effects to emulate different types of standard sensor types. The classification strength of Berkeley Caffe is measured against a collection of images representing a set of images produced by a constellation of heterogeneous satellite sensor types. Data representing each sensor type's output must be collected and given to the classifier in equal amounts to ensure optimal classification accuracy. A large diverse dataset representing heterogeneous sensor types has been produced in an effort to determine the classifier's strength in the proposed framework.

Research Questions.

The Research objectives are framed into three investigative questions to capture the overall research process. First, the conceptual development of a tipping and cueing system is executed to determine if it is possible to leverage the automation of terrain classification for a tipping and cueing system across a constellation of satellites with a mixture of heterogeneous sensor types. Secondly, the development of a Convolutional Neural Network which performs terrain classification is developed. The terrain selected for classification reflect the common areas of a search and rescue event: Urban, Water, Farm and Forest where each area correlates to a specific classi-

fication category. Thirdly, synthetic imagery is incorporated into the Convolutional Neural Network to model satellite sensor capabilities. These investigative questions are answered in the subsequent section.

1.5 Contributions

The main contributions for this research effort are four-fold: (1) Generation of a dataset which includes synthetic imagery, (2) Schematic of a Tip and Cue Communication Protocol, (3) A satellite imagery classifier with 92.1% test set accuracy, and (4)Successfully modeled uncorrelated jitter impact on classification performance. Each contribution described addresses each of the three investigative questions posed in the previous section. Research question 1 is mainly addressed by the development of the schematic of a Tip and Cue Communication Protocol and by the generation of a dataset intended to represent a heterogenous mixture of satellite sensor types. Research question 2 is addressed by the generation of a dataset with high resolution satellite imagery and development of a satellite imagery classifier with 92.1% test set accuracy. Research question 3 is addressed by the addition of synthetic imagery into the dataset generated to answer research question 2 and the successful modeling of uncorrelated jitter impact on classification performance. The subsequent section describes each contribution.

Heterogeneous Sensor Framework Contribution.

A dataset intended to represent a heterogeneous mixture of satellite sensor types is described in the Image Acquisition section of the Methodology chapter. Images collected from the Image Acquisition section are grouped into different Tiers based off of their approximate pixel resolutions, depicted in Figure 1. Imagery of approximately 1 to 0.5 meters pixel resolution are grouped together in Tier 2. While images with

an approximate pixel resolution of 0.305 meters are grouped into Tier 3. Images in Tier 2 and Tier 3 are sent into the HybridCaffe network for image classification. The images from each Tier are randomized for training and testing purposes.

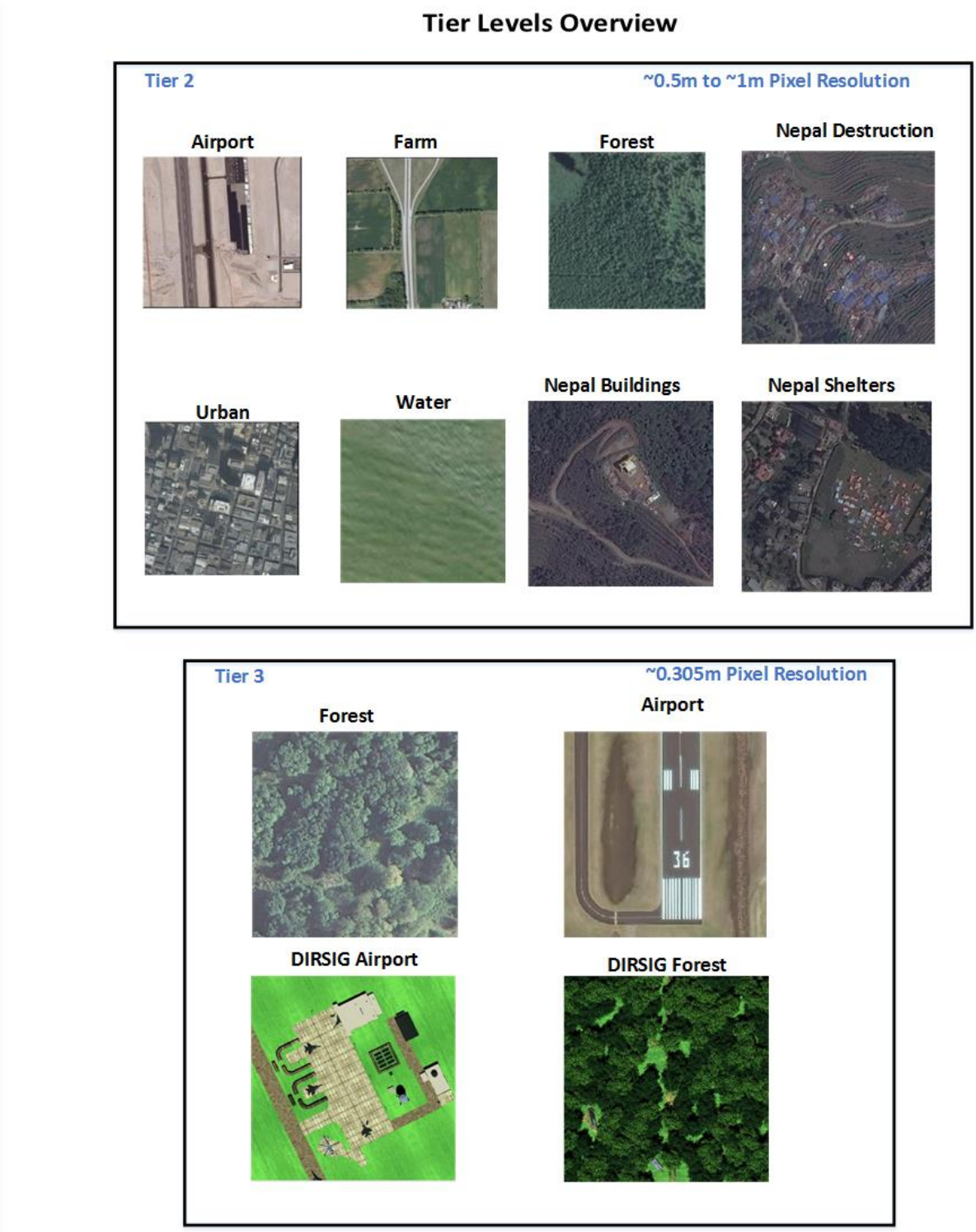


Figure 1. Overview of Tier Levels in Solution Framework, courtesy USGS, TOMNOD, University of California Merced[1], DigitalGlobe and Rochester Institute of Technology[2].

The general solution framework described in Figure 2, depicts images from Tier 2 and Tier 3 sent into HybridCaffe as inputs. It is important to note that the Convolutional Neural Network, named HybridCaffe is treated as a black box where the Tier level organization does not affect the layers during the process of conducting a forward pass. These images are processed in the Convolutional Neural Network layers described in the basic CNN architecture Section. The steps of processing through the layers are in the form of a forward pass. Once, the images are successfully processed they are output as probability percentages. These probability percentages are viewed as network activations.

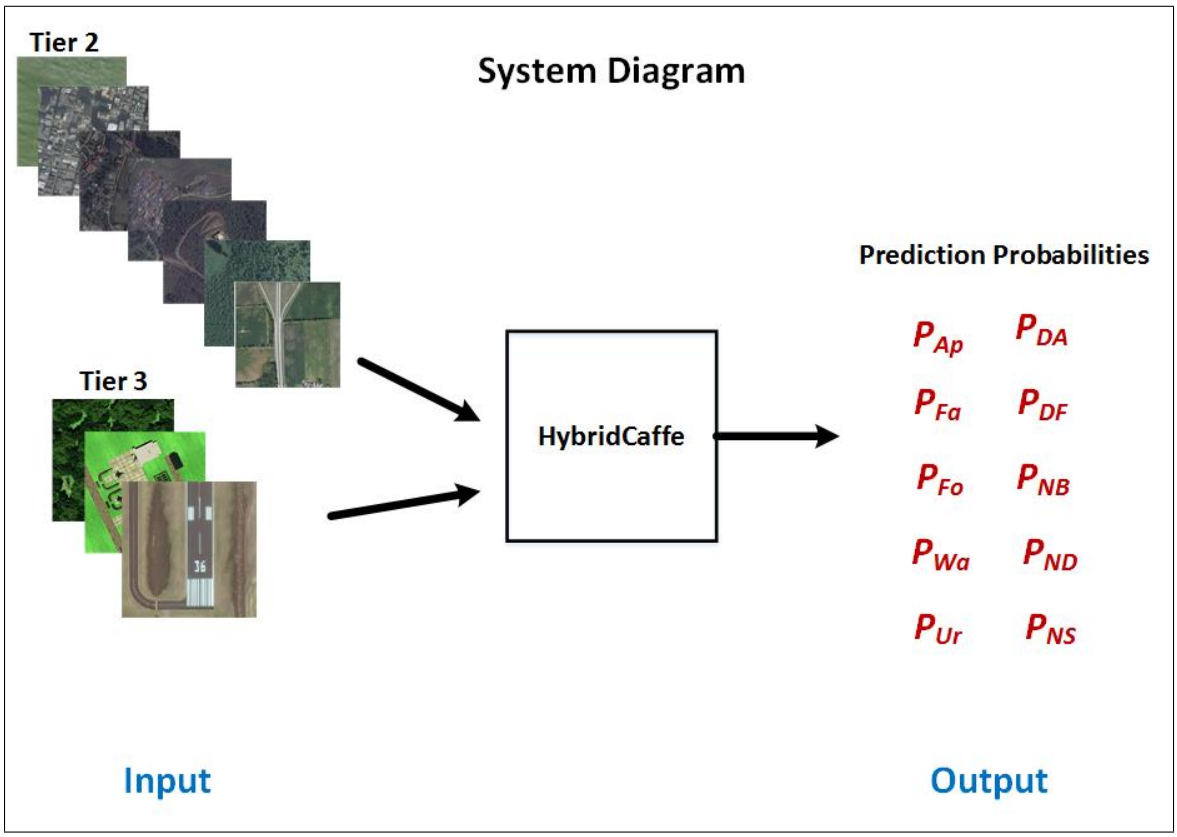


Figure 2. Overview of Solution Framework, where Tier 2 and Tier 3 Level imagery is sent as inputs into HybridCaffe, resulting in prediction probability for each respective class.

Tipping and Cueing Scheme Contribution.

The thesis opens with the the question, that investigates how to leverage the automation of terrain classification for a tipping and cueing system across a constellation of satellites with a mixture of heterogeneous sensor types. Mainly, this question was answered through the contribution of a satellite communication protocol described in Figure 3 and corresponding algorithm depicted in Figure 4.

The tip and cue scheme in Figure 3, depicts a conceptual diagram for how the tipping and cueing communication protocol would work between a selection of satellites in a constellation. Each satellite in the constellation is attributed to a specific orbit altitude. Satellites in a Geostationary Orbit (GEO) have a mean orbit altitude of 35,790 kilometers above the Earth's surface. Satellites in a Low Earth Orbit (LEO) have an orbit altitude from 200 to 1200 kilometers above the Earth's surface. Each of the orbit altitudes described are attributed to Tier levels where the GEO orbit is associated with Outer Tier Level, and the two ranges specified in LEO orbit is associated with Inner and Detail Tier Levels.

The process of the tip and cue communication protocol is initiated by a ground station sending a set of GPS coordinates and a request for a GEO satellite to sense the Earth's surface for identification of a distress event. Requests depicted in Figure 3 are indicated by dashed lines and confirmations are indicated by solid lines. A color specific to each satellite in the Tip and Cue Protocol is depicted in Figure 3 as well. The GEO satellite will send the sensed image and GPS coordinates in the form of a tip message to a LEO satellite in an orbit of 1200 kilometers. The LEO satellite will receive the tip and send a confirmation to the GEO satellite. The LEO satellite has sensors that allow it to have higher precision and spatial resolution as it is closer to the Earth's surface. However, the close proximity to the Earth's surface causes the angular velocity of the LEO satellite to be high enough for persistence to be lost. This

loss of persistence is evident in the limitation a LEO satellite has in only being able to cover a particular region on the ground at a given time. The high angular velocity also limits their field of view. These limitations on the LEO satellite cause a need for appropriate timing to ensure the satellite is able to sense the Earth's surface at the correct time. The GEO satellite sends a cue message indicating the appropriate time for the LEO satellite to sense the Earth's surface. The same process is repeated for the Inner and Detail LEO satellites where the LEO satellite in inner orbit sends tip information to Search and Rescue respondents nearest to the location of the distress event. Further details describing the process is provided in the Future Work Chapter.

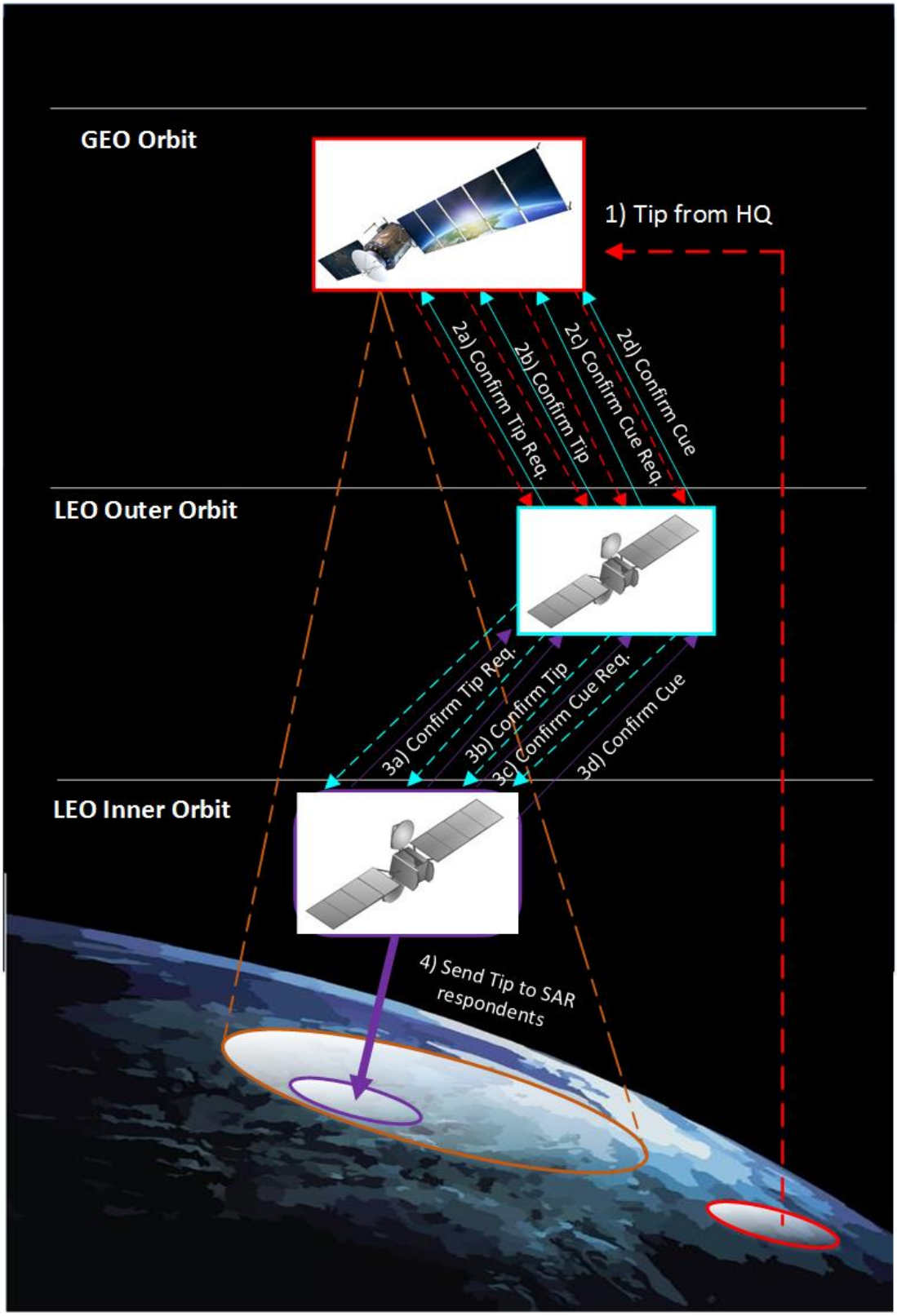


Figure 3. Tip and Cue Concept Diagram

Algorithm Design Contribution.

The Tip and Cue Algorithm described in Figure 4 depicts a conceptual layout of how a tip and cue would be executed between a selection of three satellites in a constellation. The tip and cue exchange are executed according to the steps depicted in Figure 3. Three Convolutional Neural Networks are assigned labels correlated to each Tier Level. For the sake of simplicity, it is assumed that any classification above 0.80 is considered a positive classification. This then triggers the activation of the TipSuccessor function followed by the CueSuccessor function. Each function must complete the subroutines within their scope before progressing to the next function. This convention of order is applied to each Tier Level until the final tier, Detail Tier completes its process.

Algorithm 2 Tip and Cue Algorithm

```
procedure TipCue
    CNN1 ← Outer Tier ← Current Tier
    CNN2 ← Inner Tier ← Successor Tier
    CNN3 ← Detail Tier
    p ← Current CNN Certainty Percentage
    p′ ← Successor CNN Certainty Percentage
top:
    if p >= 0.80 then
        Current Tier ← Confirmation Signal
        TipSuccessor ← Activation Signal
        CueSuccessor ← Activation Signal
        goto loop.
loop:
procedure TipSuccessor
    Send Tip Request to Successor
    Receive Tip Request Confirmation from Successor
    Send Tip information to Successor
    Receive Tip information Confirmation from Successor
    goto bottom.
bottom:
procedure CueSuccessor
    Send Cue Request to Successor
    Receive Cue Request Confirmation from Successor
    Send Cue information to Successor
    Receive Cue information Confirmation from Successor
if p′ >= 0.80 then
    Inner Tier ← Current Tier
    Detail Tier ← Successor Tier
    goto top.
    close;
repeat classification
```

Figure 4. Tipping and Cueing Algorithm

HybridCaffe Satellite Imagery Classifier Contribution.

The Convolutional Neural Network designed for this research effort is called: HybridCaffe. HybridCaffe is a Convolutional Neural Network designed with the intention of classifying high resolution satellite imagery and synthetic satellite imagery with the pixel resolution ranges specified in Figure 1. This CNN is based off of the basic Alexnet [20] architecture provided in Berkeley Caffe Deep Learning Framework [9]. Alexnet is a Convolutional Neural Network designed to classify 1.3 million high resolution images in the LSVRC-2010 ImageNet training set [20]. Table 1, depicts the hierarchy in naming convention where Alexnet, AfCaffe and HybridCaffe are all derivatives of the Berkeley Caffe Deep Learning framework and each CNN builds off of the previous one. For instance HiReSatCaffe was developed from the AfCaffe CNN through fine-tuning practices described in the Fine-Tuning a CNN section of the Methodology.

Table 1. Naming Convention Hierarchy

CNN	Description
Berkeley Caffe	type of CNN designed to be a ready-out-of-the-box Deep Learning Framework for convenience
Alexnet	derivative of Berkeley Caffe which classifies high resolution imagery from the ImageNet database
AfCaffe	derivative of Alexnet which classifies aircraft
HiReSatCaffe	derivative of AfCaffe which classifies high resolution satellite terrain imagery
HybridCaffe	derivative of HybridCaffe which classifies combination of synthetic imagery and high resolution satellite terrain imagery

The reasoning behind the naming convention stems from the fact that synthetic imagery is mixed with real satellite imagery. Figure 4 depicts examples from each class category, including the synthetic images generated by DIRSIG. HybridCaffe was generated through fine-tuning practices applied to AfCaffe [5]. Analysis of the classification capacity of HybridCaffe is described in the Methodology and Results Chapters.

(a) Airport

(b) Farm

(c) Forest

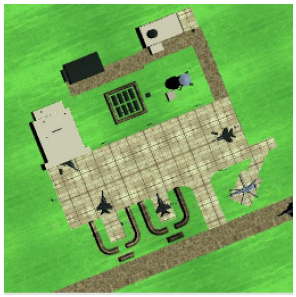
(d) Urban

(e) Water

(f) Nepal Building

(g) Nepal Destruction

(h) Nepal Shelter



(i) DIRSIG Airport [2]

(j) DIRSIG Forest [2]

Figure 5. Heterogeneous Sensor Network Dataset, courtesy USGS, images (a) through (e) and courtesy TOMNOD, images (f) through (h).

Modeling Uncorrelated Jitter Impact Contribution.

The capability of a satellite sensor is assessed by performance measures, such as spatial resolution, noise and jitter. In an effort to obtain a sense of realism in this research effort the performance measure jitter will be analyzed. The rotating sensors on board satellites in LEO orbits commonly encounter vibrations as they are propelled at high angular velocities. The modeling of the impact uncorrelated jitter has on the classification capacity of HybridCaffe indicates the classifier's potential in the space vehicle domain.

1.6 Research Focus

This will involve a machine learning based approach in optimizing the performance of tipping and cueing between low resolution and high resolution sensors aboard constellation satellites. The current system in place for responding to significant sensor events depend heavily on human interaction for sensor cueing. Automation in the communication between low resolution and high resolution sensors would be beneficial.

1.7 Methodology

The Berkeley Caffe Convolution-Based Deep Learning system is utilized as the basis for the machine learning aspect of the project. It is fast, powerful, and well-supported with deployed projects by major corporations including Pinterest, Google, and others [9]. The goal will be to train the system on examples of what are potential targets of interest at different resolutions and use them to train the system. The system will then be fed examples not used in training to see how effectively it can correctly identify targets for further investigation by other satellites (true positives) versus incorrect identifications. Once sweeps from different small satellite sensors are

available, a confusion matrix and a True Positive Curve will be generated as a basis for determining whether there is, in fact, a target present or not. The results can then be compared to results emulating those from a larger more less-capable satellite sensor with jittered images.

Experiments conducted will determine the classification strength of Berkeley Caffe. The classifier will be sent jittered synthetic images, images of typical search and rescue terrain, and high resolution images of rural areas in Nepal after the earthquake of April 2015 occurred.

1.8 Assumptions

For every significant aspect of this thesis, the following assumptions apply:

Remote Sensing.

The pixel resolution for imagery collected in the Image Acquisition Section of the Methodology Chapter is described. Images collected from the United States Geological Survey(USGS) High Resolution Orthoimagery(HRO) and National Agricultural Imagery Program(NAIP) datasets are assumed to have pixel resolutions described in their respective data sheets. Refer to the Image Acquisition Section for exact pixel resolutions. Imagery collected from the Nepal Earthquake dataset is assumed to have a pixel resolution between 1 meter and 0.5 meters. Synthetic imagery generated and collected from DIRSIG has a pixel resolution range of approximately 0.305 meters.

Convolutional Neural Network.

Described is the size of images the Convolutional Neural Network (CNN) is trained on. Every image sent into the CNN to train on is automatically snipped into 227 x 227 snapshots in the Data Layer of the CNN. The Data Layer in a CNN describes, the

origin for where the data is retrieved for sending into the network for training along with any protocol for normalizing the data to be sent into the neural network. The process of snipping an image into a standardized format ensures that the network is trained on data of an equal size, which is critical for optimal training performance. All Berkeley Caffe CNN's are fully functional in this capacity out of the box, where the Data Layer dictates the size by which the images are sent into the neural network, as 227 x 227 pixels.

Heterogeneous Sensor Framework.

Specifications for the inputs and outputs for the Heterogeneous Sensor Framework are described. The USGS HRO, NAIP and Nepal Earthquake images collected represent inputs for the CNN. DIRSIG-generated images come from the DIRSIG domain and represent inputs to the CNN. The DIRSIG domain represents the DIRSIG Simulator Suite described in the Methodology chapter and will not be used as input into the CNN. Classifications conducted by the CNN will be given in probability percentages. Classifications will be made with all classes described in the Heterogeneous sensor dataset.

Tipping and Cueing Scheme.

All of the assumptions described in this subsection apply to the Tipping and Cueing Scheme Section of the Future Work Chapter. The tipping and cueing scheme is purely conceptual, backed with facts stated in the description. A small sample of the satellite constellation is considered, where a single satellite from each tier in the example is described in the tip and cue exchange. The pixel resolution range of each satellite's imagery will be specific to their corresponding Tier Level. For instance, a satellite in Tier Level 3 will produce images with a pixel resolution of 0.305 meters.

For every tip sent between each satellite, the tip will have GPS coordinates. Cues sent between the sender satellite and the receiver satellite will have timing information that specifies the exact moment to scan the Earth's surface. Confirmation messages are sent between satellites in receipt of each message received. Initiation of the Tip and Cue exchange will be given by headquarters with an image and GPS coordinates given.

1.9 Limitations

Experimentation and design is limited by several factors to be described.

Remote Sensing.

Scarcity of high resolution satellite imagery influenced all aspects of this research effort. Limitation on abundance required resourceful approach to generate as many images as possible for the CNN to train on. This involved the design of a python script which snipped large 512 x 512 pixel images into equal sized tiles. This generated a bizarre result, where the sliced images became slightly blurred. Similarly, the DIRSIG software required an equally resourceful approach to generate as many images as possible to satisfy the CNN's training requirements. Two previously generated scenes possessed characteristics in common with scenes collected from the USGS HRO dataset. The capabilities of the DIRSIG software suite were used to generate multiple sets of images. Refer to the Methodology chapter for specific number of images and programmed flight line associated with each synthetic image set generated.

Convolutional Neural Network.

The strategies described in the Remote Sensing subsection resulted in limitations characterized by questionable classifier performance. The main source of questionable

classifier performance stems from test data originating from the same source as that of training data. This indicates a possible false sense of true classification power when applied to the real-world domain. Ground truth for this limitation is evaluated and tested in the Results and Discussion chapter. Additionally, the CNN is limited to training iterations of 10,000 and datasets of 1000 for each class where the images that compose the training set and testing sets are randomly allocated. This limitation is imposed to ensure proper adherence to the research scope. Where, the main research focus is on application of Machine Learning in a small scale.

The CNN is limited to 10 different classes: Airport, Farm, Forest, Water, Urban, DIRSIG Airport, DIRSIG Forest, Nepal damaged Buildings, Nepal total Destruction, and Nepal Shelter. Again, the limitation described is one imposed by the author to maintain a small scale scope. Each class possess limitations which influence the classification capacity of the CNN modeled. Airport images were collected from the USGS HRO and DigitalGlobe datasets. A small percentage of original airport tiles were included with sliced tiles. Farm images were collected from the USGS HRO and NAIP datasets, where a majority of farm images stem from the NAIP dataset. Forest, Water and Urban images were collected from the USGS HRO dataset and both training and testing data are sourced from that dataset. DIRSIG Airport and DIRSIG Forest images were generated from the same source, DIRSIG. Similarly, the testing and training data originate from the same source. The Nepal damaged Buildings, Nepal total Destruction and Nepal Shelter earthquake post-event imagery were collected from the same TOMNOD data repository [3]. Further, explanation of the source of Nepal Earthquake imagery, through the company TOMNOD, is provided in the Literature Review Chapter. Additionally, each image in the three Nepal classes are from the same geographic location which caused slight confusion in the classification

process. Finally, the CNN is limited to the weights utilized in fine-tuning off of the AfCaffe architecture[5] and the classic characteristics of the AlexNet framework[20].

Tipping and Cueing Scheme.

The pixel resolution associated with each Tier Level is limited by the pixel resolutions associated with the dataset created from image collection.

Thesis Organization.

The remainder of the thesis is organized in a traditional thesis structure. Chapter II, the Literature Review, describes related works which served as inspiration for this research effort. Chapter III, the Methodology, describes the experimental design used in the evaluation of this architecture's performance when applied to the search and rescue domain. Chapter IV, the Results and Discussion, details results obtained during the experiments conducted. Chapter V, Conclusion, describes conclusions and reflections on lessons learned. Chapter VI, Future Work, describes future work to further develop the framework's potential for upcoming research efforts.

II. Literature Review

This chapter describes related works which served as inspiration for this research effort. Innovations in small scale satellite development and usage in a constellation are highlighted. A background on the history, strengths and fine-tuning practices of Berkeley Caffe is given. Selected works highlight convolutional neural network practices which applied remote sensing techniques to classification of high resolution satellite imagery. In particular, software products in the search and rescue demonstrate application of neural networks in the classification of imagery. Insights found on all accounts given are described in the proceeding sections of this chapter.

2.1 Satellites in Constellation

Satellites in constellation provide many capabilities, such as distributed observations/measurements and an ability to be launched in large numbers[21]. These capabilities are advantageous for execution of remotely sensing the Earth's surface for real-time coverage in application to the Search and Rescue domain. The distributed nature of the satellite constellation provides a sense of robustness, where the degradation of system performance is safely inhibited. According to author Sukumar Ghosh, "A distributed system provides an excellent opportunity for incorporating fault tolerance and graceful degradation[22]". The satellite constellation as itself a distributed system, presents numerous benefits as a cost efficient and robust observation system which fosters fault tolerance.

A diverse field of novel missions claim to benefit from the capabilities described, such as meteorology, climate science, disaster detection, disaster warning, atmospheric, magnetosphere, and ionospheric measurement/observation[21]. Strides have been taken in the past decade as several successful small class satellite missions, in

particular, nanosatellites have demonstrated the utility of this class of spacecraft in independent operation through engineering/technology (e.g. STRaND-1[17]) and military (e.g. SMDC-One[23]) capabilities. There have been cases of small less capable satellites being used for remote sensing projects as indicated in Surrey Training, Research and Nanosatellite Demonstrator (STRaND-1[17]), which utilized a Google Nexus One smart phone with an android operating system to provide for the computing capabilities of the satellite[17]. Imagery of low resolution quality is generated from the smart phone camera equipped with Google One Nexus. Continued expansion of the android operating system's capabilities offer great potential in the design of a similar nanosatellite with a more powerful and capable sensor. Particularly, such work has been done in adapting the embedded internals of a CubeSat computer with a raspberry pi embedded system[18]. Opportunities of expansion from these satellite project's findings makes development of an automated tipping and cueing satellite constellation framework highly possible.

2.2 Artificial Neural Networks

In the past fifty years, Artificial Neural Networks have gained a significant amount of attention from researchers in the Deep Learning field. Development of this technology in the past decade produced services and products to meet the needs of the fast paced, high resolution, high data driven demands of the 21st century. Artificial Neural Networks are utilized by big-data heavyweights like Google, Pinterest and Netflix to power self-driving cars and recommendations for users. The following section describes the origins of Artificial Neural Networks.

History of ANN's.

The roots of Artificial Neural Networks can be traced back to biological origins and embryonic development challenges overcome by modern breakthroughs. An investigation conducted by Hubel et al, revealed a layer hierarchy in the mammalian cortex, subsequently, this inspired researchers in the computer vision domain to develop similar pattern recognition mechanisms in neural networks[24]. In particular, their discovery was extended through Fukushima, who designed the Neocognitron[25]. However, his work was hindered by modest computational means in the late 1970's as Artificial Neural Networks require impressive computing hardware to produce ground breaking results. Consequently, between the 1940's and 1970's interest in further development of the potential in Artificial Neural Networks was sparse.

Renewed academic interest was spearheaded by CalTech Professor, John Hopfield, who presented a paper which proposed the classic Neural Network architecture. His work formed a basis for modern Neural Network development to draw insights from[26]. Artificial Neural Networks train on data through Stochaistic Gradient Descent(SGD), to seek out approximately optimal neuron weights appropriate for pattern recognition. SGD requires gradients of pattern recognition error with respect to each network parameter. The back propagation chain derivative technique which provides these gradients, was applied in the ANN research community in 1986 when the backpropagation algorithm was proposed by Rumelhart[27]. As the early designs of early neural networks began development in the footsteps of the realizations drawn by these forefront predecessors[24, 26, 25] many challenges were encountered, as they were incipient in design. The vast computing power required for the intricate layer design was not yet available at the end of the 20th century. It was at the dawn of the 21st century when an explosion of research and interest developed sophisticated successors to the original ANN's. In the late 90's a new state of the art classifica-

tion accuracy of 99.2% on MNIST dataset was recorded from the inception of LeNet neural architecture for recognition of handwritten digits and words[28].

Convolutional Neural Networks, a variant of Artificial Neural Networks, best suited for the image classification problem domain have been applied to image recognition tasks in speech, biology, medicine, handwriting and scene classification. The CNN is a variant of the Neocognitron, an early neural network built with a biologically inspired design[25]. There are many Convolutional Neural Networks in existence today, the Convolutional Neural Network of interest, Berkeley Caffe developed by the Berkeley Vision and Learning Center at Berkeley University California has been released to researchers for further development of its classification capacity[9].

Modern Convolutional Neural Networks.

There have been strides made in the utilization of the CNN's classification power in speech, biology[29], medicine[30], handwriting and scene classification[31, 32, 33, 16, 34, 35, 36]. Presently, it is used for large-scale image recognition by industry heavy-weights like Facebook, Google and Baidu. It has since then been expanded to DeCAF[37], an implementation of deep convolution activation features and the annual ImageNet Largescale Visual Recognition Challenge[38], a benchmark for large-scale object recognition. DeCAF demonstrated deep learning features can deliver general improvements in visual recognition and can be fine-tuned to specific tasks[37]. The ImageNet Largescale Visual Recognition Challenge[38] held every year has produced a variety of impressive variants of Caffe, such as VGG[39] and GoogLeNet[40]. Contributions involved evaluation and expansion of the depth and width of traditional Convolutional Neural Network architecture[40, 39].

Fine-Tuning.

The concept of fine-tuning was realized in the work of DeCAF, through an investigation of semi-supervised multitask learning of deep convolutional representations, where representations are learned on a set of related problems but applied to new tasks which have too few training examples to learn a full deep representation[37]. Grand scale deep learning models with the representational capacity of ImageNet[20] have demonstrated superb computational abilities when applied to enormous amounts of training data. However, when the same deep learning model is applied to training data of limited size over-fitting occurs. When the ratio between the model's complexity and the magnitude of the training set is too high this phenomenon occurs.

Justification for choosing Berkeley Caffe.

The benefits of Caffe are clearly defined in Caffe: Convolutional Architecture for Fast Feature Embedding. The modularity of Caffe allows for extension to different data formats, layer modification and loss function expansion which are direct and easy to access. There is a high number of layers and loss functions in use and a generous number of examples demonstrate how the layers and loss functions are ensemble together to create trainable classifiers. The separation of the implementation of Caffe and its representation is made clear. Model definitions are written as configuration files in the Protocol Buffer language. Caffe's network architecture is in the form of arbitrary directed acyclic graphs. Once instantiated Caffe, abstracts from its underlying location in host or GPU and allocates the exact amount of memory required for the network. A single function call serves to switch between a CPU and GPU implementation.

Strengths of Berkeley Caffe.

Strict testing coverage policies are in place to ensure rapid improvement and refactoring of the codebase. Each module in Caffe is tested and no new code is allowed on the project without the appropriate test. Caffe's interface flexibility is optimal for research capacity needs such as graph generation, trouble shooting and diagnostics. Python and MATLAB bindings are made available to provide some flexibility in the development and testing of the network. MATLAB and Python languages can classify inputs and build networks. Pre-trained reference models are provided to serve as ready-out-of-the-box examples for visual classification tasks. This includes the famous "AlexNet" ImageNet model with some slight variations along with the R-CNN detection model. These examples were made available and were intended by the authors to foster quick progress in any research effort requiring a neural network.

Convolutional Neural Networks in Scene Classification.

Contributions have been made in utilizing CNN's for classifying satellite imagery. Some of the most popular, High resolution satellite imagery datasets have been generated like UC Merced[1] and DeepSat[31]. However, a CNN requires a large volume of training images to draw out reliable classifications. These datasets are limited in size and in pixel resolution. It is difficult to find a dataset that possesses high resolution quality characteristics. Another limitation is the scope of scene type. Most other datasets are focused on the sensing of a certain object type in a scene, for instance detecting roads[34] and buildings[41] and detecting airports[16]. These are datasets that are useful but do not possess enough features to allow for a global scope of coverage.

In the case of [31], [36], [41] and [16] images are supplemented by use of straight segments, and channel map patches to help the convolutional neural network label and

detect the scenes of interest. In the case of the proposed framework, HybridCaffe, the main contribution is a direct implementation of application of high resolution satellite images to the CNN. A high percentage of accuracy is achieved. Categories to test the CNN's capacity to label airports, and three classes of images from a Nepal Earthquake scenes have been selected. The set of images chosen represent the different search and rescue objectives specified in training of areas of different terrain: Urban, Farm, Forest, and water.

A supplemental pair of synthetic images have been added: DIRSIG Airport and DIRSIG Forest. Comparisons are drawn to determine the CNN classifier's power. Synthetic imagery is jittered to varying ranges of radian intensity to determine the drop off for when the CNN can no longer classify images. Experiments were performed to determine what classes of scene served as points of confusion in the CNN.

Extraction of buildings and roads from aerial imagery has many applications in a wide range of areas including automated map making, urban planning, change detection for real-estate management, land use analysis, and disaster relief. However, these tasks have been performed by human experts manually, which is a very costly costly and time-consuming process. Because buildings and roads have much variation in their shape and they may be occluded by other objects such as trees, accurate labeling of large aerial imagery is a complex attentional task for human. Hence, automatic extraction of buildings and roads is highly demanded, and many attempts at automatic aerial imagery interpretation have been proposed in remote sensing literature [41].

2.3 Remote Sensing

Remote sensing highly benefits from satellites in constellation. As satellite in constellation provide many capabilities as mentioned in [21]. The capability for dis-

tributed observations and measurements makes it possible to cover the globe in a more real-time fashion. Remote sensing, the conducting of observations from afar, uses remote sensors to collect data and reproduce output with spatial, temporal and spectral sampled characteristics. The fundamentals, calculation of the Ground Sampling Distance (GSD), and DIRSIG are described in the proceeding subsections of this section.

Fundamentals.

Remote sensing is viewed as a multi-disciplinary science that includes a combination of various fields such as optics, spectroscopy, photography, computer, electronics and telecommunication [42]. Remote sensing imagery has applications in mapping land-use and cover, agriculture, soils mapping, forestry, city planning, archaeological investigations, military observations , geomorphological surveying, land cover changes, deforestation, vegetation dynamics, water quality dynamics, urban growth, disaster monitoring, oil spill detection, mineral deposit extraction, air quality monitoring, ocean flow monitoring, ice sheet detection, disease prevention, etc [42]. The process of remotely sensing an Earth-based object involves the collection of information from the Earth's surface without the human in direct contact of the detected surface. Aspects of the Electro-Magnetic spectrum is used in the process of remotely sensing an object on the Earth's surface. The radiant flux or power received by a surface per unit area W/cm^2. Radiance is directional power per unit area $W/sr - cm^2$, where the radiant flux is emitted, reflected, transmitted or received by a surface. Electromagnetic energy is either reflected or emitted by the Earth's surface. Radiation which leaves an object, is called radiant exitance, which includes both emitted and reflected irradiance. While radiation that hits a surface is known as incident irradiance. Radiation which leaves the surface of an object is influenced by the physical

properties of the object. Incident irradiance is dependent on the environment such as solar conditions and the atmosphere but is independent of the object's physical properties.

According to [42] the steps involved in the process of remotely sensing an object are six-fold: (1) Initially, the EMR source will emit energy in the form of the emission of electromagnetic radiation, or EMR (sun/self-emission), (2) Thereafter, energy is transmitted from the source to the surface of the Earth, (3) Consequently, the EMR interacts with the Earth's surface, resulting in reflection and emission, (4) Finally, energy from the surface is transmitted to the remote sensor, (5) Followed by, Sensor data output, (6) Concluding in, data transmission, processing and analysis of the sensor data.

There are two forms of remote sensing: passive remote sensing and active remote sensing. Energy provided by the remote sensing platform is known as active remote sensing. Remote sensing measurements depend on the external energy source is known as passive remote sensing.

EMR from the particles which pass between a target and a remote sensor carry information that bears characteristics about the target's identity and nature, called the signature. In the case of this research particles of interest can be debris from damaged building remains and dust plumes which lead back to sources which created them. Targets possess spectral, spatial, and temporal characteristics which make each class of a target unique. These features are developed and simulated in the DIRSIG simulator suite. The replica that bears all the spatial, spectral, temporal and radiometric characteristics of the target is called the image. Images generated by remote sensors are far from perfect as the collection is conducted with the single remote sensor looking at one direction at one time using only one small slice of the

Electro-Magnetic spectrum. Furthermore, some information about the image is lost before it is input to a remote sensor and the image can be formed.

Data collection with traditional satellites require perseverance through challenges in order to reach a successful collection. The geometry of a satellite's orbit usually cannot be changed to satisfy specific tasking requirements, so the time of day and frequency of revisit of a particular remote sensor to a tasked collection site may preclude success. This difficulty can be alleviated by the Tipping and Cueing framework. In particular, the cueing scheme will enable a satellite to know the time of day and location on the Earth surface to conduct imagery collection.

Ground Sampling Distance.

The Ground Sampling Distance (GSD) is the distance between pixel centers measured on the ground. Typically, calculation of the ground sample distance requires metrics such as the platform altitude, pixel size and focal length.

The Ground Sampling Distance (GSD) is calculated as a similar triangles problem:

$$\frac{GSD}{Altitude} = \frac{PixelSize}{FocalLength} \tag{1}$$

After algebraic manipulation becomes:

$$GSD = \frac{PixelSize * Altitude}{FocalLength} \tag{2}$$

DIRSIG.

Innovation of remote sensing in satellite imagery has been demonstrated by the development of various Image Generation platforms such as DIRSIG, Digital Imaging Remote Sensing Image Generation. This platform has the capability of modeling sensors of aircraft and satellites orbiting the Earth's atmosphere. This is very useful in

propelling the capabilities in development of technology for target detection, scenario simulation and related topics. DIRSIG also possesses the capability of generation of various aircraft models through utilization of CAD software. Development of various aspects of scenario building is essential for the formation of various strategies to be put in place before practical application in the real world. Such a capability as this saves resources such as time and money and demonstrates efficient planning in the long run. The synthetic scenes and objects generated by DIRSIG possess potential for application to an Berkeley Caffe that serves as the main backbone of the image classification aspect for this proposed framework.

The DIRSIG model is a complex synthetic image generation application which produces simulated imagery in the visible through thermal infrared regions [2]. As indicated in Figure 6, the majority of the imagery generated for the test dataset is in the visible spectrum.

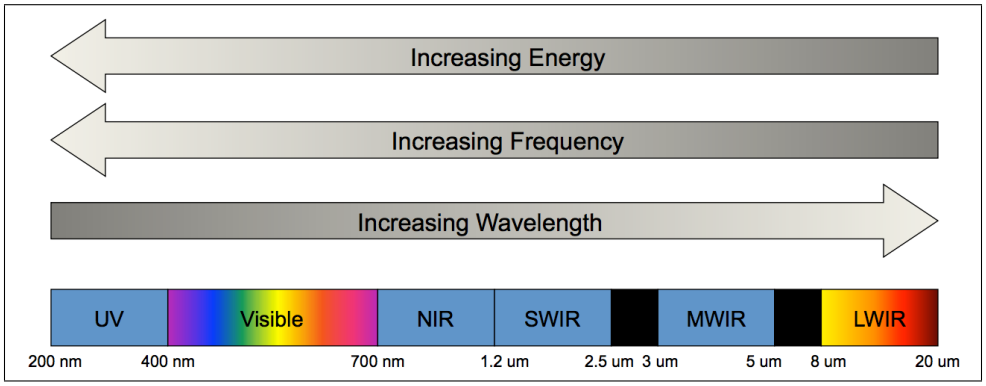


Figure 6. The Electromagnetic Spectrum, DIRSIG imagery ranges from visible through infrared (0.2 - 20.0) microns.[2]

The DIRSIG model address four important aspects of scene modeling: (i) geometric complexity, (ii) spatial-spectral diversity, (iii) directional reflectance and illumination loading and (iv) spectral mixing [43]. Fundamentally, the model is a ray-tracer model, where surfaces along paths are determined within scenes to contribute to radiance fluxes at specific points.

2.4 Search and Rescue

There have been many efforts in utilizing the diverse remote sensing capabilities of satellites in constellation for application to disaster detection. The Disaster Monitoring Constellation provides daily global coverage at moderate resolution, as part of a system to provide global disaster monitoring. Companies focused on using high resolution satellite imagery for serving the public in catastrophic events, Tomnod, Interra and Global Forest Fire Watch have emerged in recent years. A freshly minted data science term, crowdsourcing, has leveraged the browsing appetite of the average Internet user. Disastrous events calling for search and rescue capabilities, such as the missing Malaysian Airline 407 was spearheaded by Digital Globe who partnered with Tomnod to provide high resolution satellite imagery for users to classify. These crowdsourcing efforts extended into a single tipping and cueing framework could pose as a favorable solution in a catastrophic event requiring surface detection of survivors. A background on the crowdsourced intelligence pioneer, Tomnod, is given in the subsequent subsection.

Tomnod.

The crowdsourcing platform, Tomnod, is a project owned by Colorado-based satellite company Digital Globe that uses crowd-sourcing to identify objects and places in satellite images. This company is at the forefront of innovation in the burgeoning field of crowdourcing of Earth observation imagery, combined with advanced algorithms and deep GIS and imagery knowledge. It was originally created as a research project in the University of California. Their work involved the usage of Tomnod in location of the tomb of Genghis Khan [44]. The company's name, Tomnod, is coined from the Mongolian word, "big eye." This is a crowd based solution to the satellite imagery remote sensing challenge of both data volume and search target ambiguity

[44]. Tomnod charged a group of volunteer participants the challenge of finding the tomb of Genghis Khan, an archaeological enigma of unknown characteristics widely believed to be hidden somewhere within the range of our satellite imagery.

Tomnod's Purpose.

The purpose of Tomnod is similar to the vision which motivated this proposed solution framework. In particular, the combination of imagery, geospatial analytics and all-source analysts possess the potential to make a significant difference in evacuation planning, disaster response, recovery, and rebuilding in regions worldwide. Commonly, it is assumed Digital Globe and other satellite providers have tools which automate the process to search through the massive number of images taken each day by satellites circling the Earth, but that's not quite the case yet. Presently, this automation boils down to the analysts to identify the most important targets and events and to provide context and insight.

Crowdsourcing.

In recent years, crowdsourcing has emerged as a rapidly growing field in research and online content creation, facilitated by the development of new technologies, greater incentive for outreach among researchers, a growing public interest in applied science and the desire to have a positive impact on the world [45]. Crowdsourcing displays the powerful impact people serve as an excellent resource in disaster response. The creators of Tomnod believe the power of crowdsourcing lies not only in harnessing parallel networks for scalable analytics, but in forming the collaborative frameworks necessary to cultivate collective reasoning [44] where the system depends on the crowd to process large volumes of information and identify the unexpected features of interest.

Nepal Earthquake.

High Resolution satellite imagery collected from the 7.8-magnitude earthquake which heavily impacted central Nepal on April 25 2015 was utilized as training input data. A brief background is described on the scenario in central Nepal and the important role Tomnod played during this crisis. During this cataclysmic event, thousands of homes and roads were destroyed resulting in a large-scale humanitarian crisis. In response, Digital Globe collected thousands of square kilometers of high-resolution satellite imagery and opened up access to the imagery for labeling through their Cloud Services. Imagery provided by Digital Globe was also made available to web-connected volunteers with Tomnod accounts. It has been estimated that more than 58,000 people identified and tagged more than 21,000 damaged buildings and roads and areas of major destruction in central Nepal.

In an effort, to serve ongoing response to the aftermath of the earthquake, Digital Globe had released satellite imagery to the public domain. Tomnod instructed their users to label images with the following features: damaged buildings, blocked roads, tents/shelters and areas of major destruction. Quality control of input entered by users is conducted by an intelligent algorithm, CrowdRank and experts in quality control. Figure 7, displays a map of central Nepal with relevant feature points entered by Tomnod users.

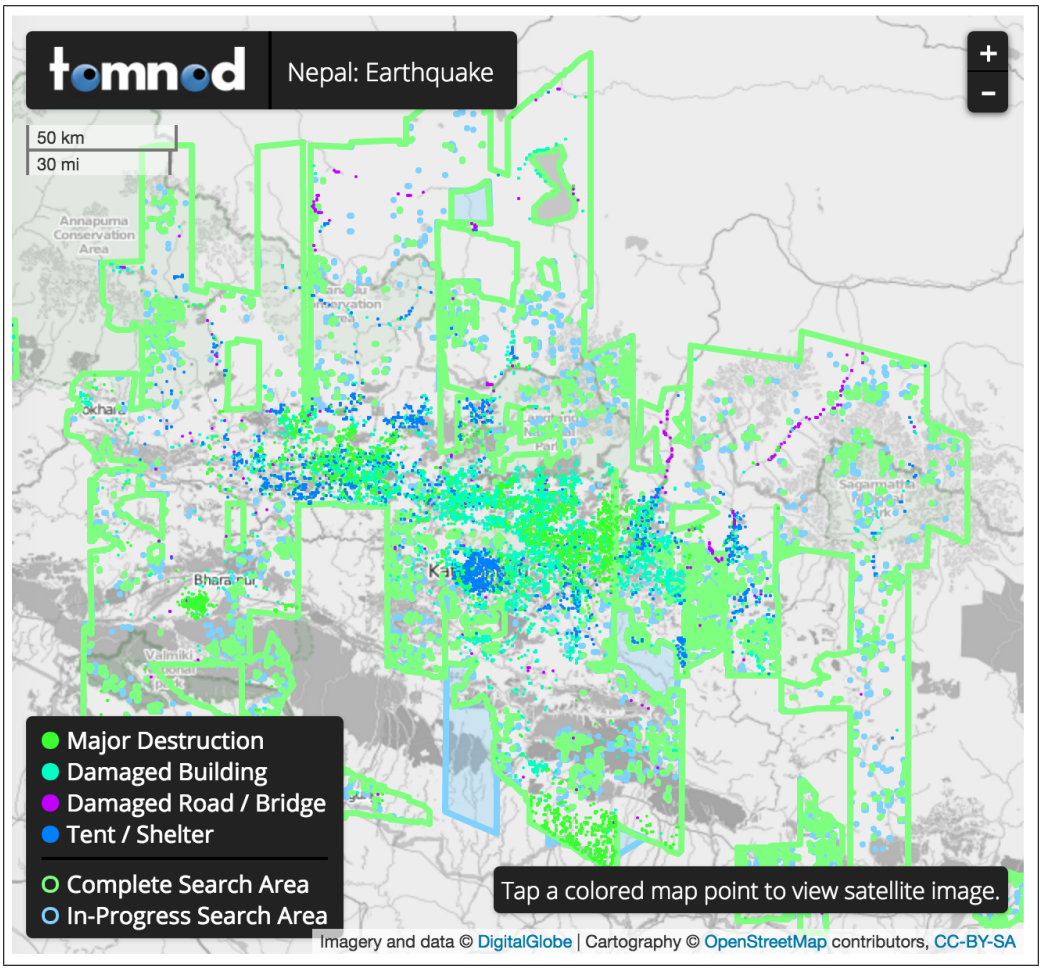


Figure 7. Map of Tomnod user inputs, from the Nepal Earthquake Data Portal.[3]

COSPAS-SARSAT.

The background of the current search and rescue satellite aided tracking system used in present day is discussed. Search and Rescue Satellite Aided Tracking (SARSAT) is a portion of the International COSPAS-SARSAT Program which constitutes of 41 nations and two independent SAR organizations[4]. Where COSPAS stands for Cosmicheskaya Sistema Poiska Avariynyh Sudov, that translates from Russian to English as: "Space System for the Search of Vessels in Distress." This utilizes Geostationary Search and Rescue (GEOSAR) and Low Earth Orbit Search and Rescue (LEOSAR) satellites from the National Oceanic and Atmospheric Administration (NOAA) to detect the location of stranded mariners, hikers and aviators[4]. Since, the system's inception great success has been achieved. From September 1982 to December 2015, the COSPAS-SARSAT System provided assistance in rescuing at least 41,750 persons in 11,788 SAR events [46]. Additionally, from January to December 2015, the COSPAS-SARSAT System provided assistance in rescuing 2,185 persons in 718 SAR events.

User Domain Restriction and Future Application.

The user domain restriction and future applications for the COSPAS-SARSAT system is described. The COSPAS-SARSAT program is applicable only to individuals who possess a distress beacon that operates in the 406.0 - 406.1 MHz frequency band. A total of approximately 1,513,000 beacons are registered in the COSPAS-SARSAT program [46]. This is a high price item that is not readily accessible by individuals from less fortunate demographics. As global weather continues to change in varying extremes[47], it is possible that search and rescue (SAR) efforts may increase throughout the world. In an effort to be best equipped for facing such a challenge it is beneficial to possess the ability to select events of interest to optimize time in

response to a weather event. Such an example is evident in Hurricane Matthew. It is possible to equip first response personnel with training and equipment for mass rescue efforts.

Response to Beacon Activation.

The process of response to beacon activation is described, as indicated in Figure 8. First, an Emergency Position Indicating Radio Beacon (EPIRB) or Personal Locater Beacon (PLB) is activated. This transmits a distress signal in the form of digitally coded short bursts to search and rescue satellites. Location information, beacon identification information, and aircraft/vessel information are included in each burst. Next, search and rescue satellites from the space segment of the COSPAS-SARSAT system receive the signal and relay the signal to the ground segment. Next, The signal is sent to a Local User Terminal (LUT) on land, to analyze the signal and determine the approximate geographic location. At this point the beacon activation site location has been estimated and the appropriate rescue authorities must be routed to the beacon activation site. Next, the LUT transmits all information included in the beacon's distress message to the Mission Control Center, who receives and processes the data. Next, the Rescue Coordination Center (RCC) evaluates the information to determine if the emergency is legitimate. If it is the beacon distress message is sent to a national-government Search-and-Rescue Point of Contact (SPOC). These points of contact are responsible for reacting to distress alerts within their local area. Finally, rescuers are deployed to the location of the EPRIB or PLB.

The process described requires several steps until the deployment of rescuers to the beacon activation site. Usage of satellites with optical sensors poses as a favorable solution. The process would at a minimum involve recording the GPS coordinates for each survivor then sending them to first response personnel closest to the coordinate's

exact location. This system would be highly beneficial in future emergency weather scenarios where sending first responders to an approximated location sacrifices time. The main objective of the proposed solution framework is to optimize time in an effort to direct first response personnel to precise locations in an emergency weather scenario.

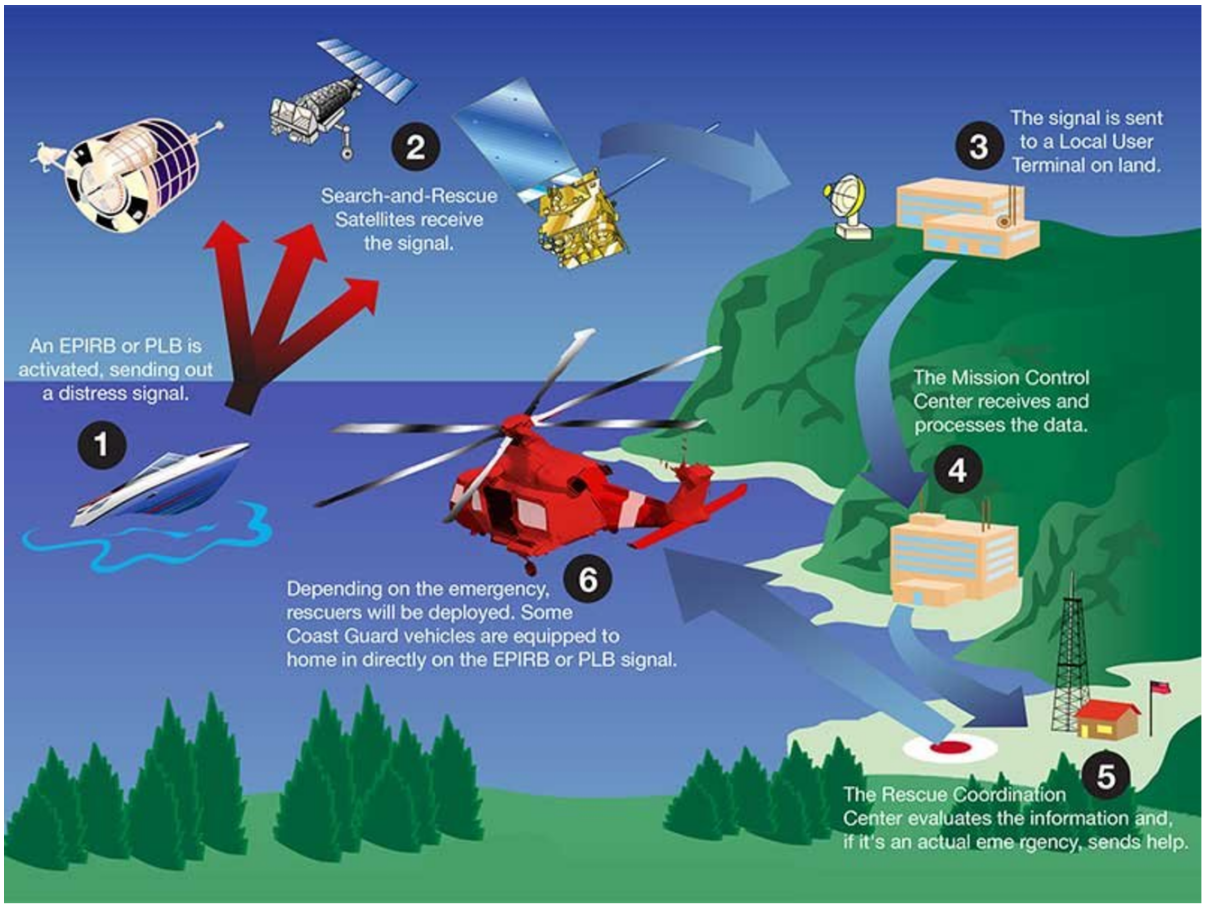


Figure 8. System overview of COSPAS-SARSAT system [4]

III. Methodology

This chapter describes the process of data preparation, Convolutional Neural Network preparation, and the general experimental design. These methods of preparation described investigate the research questions posed in the introduction where the automation of terrain imagery classification is investigated followed by incorporation of synthetic imagery to model satellite sensor capabilities. Before such an investigation is to be conducted in the form of an experiment set, a series of necessary preparations are to be made. The Convolutional Neural Network model, HybridCaffe, required a large dataset to render valuable classifications. The process involved in the creation of an extensive dataset is described in the Image Acquisition section. The basic architecture requirements for the successful completion of experimentation is described in the Convolutional Neural Network Development Section. Additionally, the process of fine-tuning a Convolutional Neural Network in preparation for experimentation is described in this section. Finally, the experiments section describes the general experiment design for the Classification Factor Investigation Experiment Set and the HyridCaffe Classification Evaluation Experiment Set. Insights found on all accounts given are described in subsequent sections of this chapter.

3.1 Image Acquisition

This section describes the methods of data collection from the High Resolution Orthoimagery (HRO) and National Agricultural Imagery Program (NAIP) datasets residing in the United States Geological Survey (USGS). Imagery obtained from these datasets possesses a high pixel resolution quality range of 1 meter to 0.305 meters and 0.5 to 1 meter, respectively. These images were collected for the sole purpose of training the Convolutional Neural Network, Hybrid Caffe.

High Resolution Orthoimagery.

Initial attempts at acquiring HRO were drawn from pools of images organized by feature and geographic location found in the United States Geological Survey (USGS). Specifications set for the features of interest were drawn out by interfacing with the USGS data pool with EarthExplorer. The selected features of interest were Airports, All, Beach, Sea, Lake, Forest, Plain, Woods and Swamp intended to correlate with the five categories of data used to train and test the Convolutional Neural Network model, HybridCaffe. Correlations between the USGS EarthExplorer features, geographic locations and HybridCaffe categories are described in Table 1.

Table 2. HybridCaffe Categories with associated USGS Features, courtesy USGS

HybridCaffe Category	USGS Feature	Geographic Location
Airport	Airports	Arizona, Virginia, Texas
Farm	Plains	Kansas
Forest	Forest, Woods, Swamp	Montana, Oregon, Washington
Water	Sea, Beach, Lake	American Samoa, Louisiana, Florida
Urban	All	Arizona, Louisiana, Alabama

Image Scarcity.

Image scarcity of features with Airport, Farm, Forest and Water characteristics required pulling imagery from multiple feature pools with similar traits. Abundance of imagery with urban characteristics required a search through specification of "All" in the USGS EarthExplorer Feature list. The majority of the images pulled from the feature pools described in Table 1 were from the HRO dataset where a fairly high pixel resolution from approximately 1 meter to 0.305 meters was evident in each of the images. The pixel area of the images acquired from the HRO dataset was approximately 1028 x 1028 pixels. The Caffe model, HybridCaffe is required to train

on an enormous amount of imagery to ensure an effective classification capability is obtained. In particular, the training data given to the Caffe model must be maintained at an optimal pixel range of approximately 227 x 227 pixels as the model is programmed to take input as 227 square images from larger sized images. Images that are under the pixel size described are zoomed in to expand to the tile snippet size. This presents a risk of deteriorating the Caffe model's classification capacity as any further decrease in the pixel dimensions may cause the features of interest to become obstructed. This provided an opportunity for image parsing of the large pixel area images.

Advantages of Differing Spatial Scale and Mitigation.

The Convolutional Neural Network, HybridCaffe is trained on multiple images with different spatial scales. When working with Convolutional Neural Networks the ability to learn at one spatial scale is not useful as training at multiple spatial scales allows for the Convolutional Neural Network to learn how to view a target image from all possible perspectives. Thus, training HybridCaffe on images with different spatial resolutions is ideal.

A form of mitigation against training on images with a single similar spatial scale is through the custom data augmentation where the scale is randomly adjusted for each training image. This random adjustment consists of random zooming in and out of training images. This form of mitigation gives the Convolutional Neural Network more experience with a broader set of ways a target image can look. The more variety of representations of a particular class, the higher the likelihood the Convolutional Neural Network has in obtaining a well-rounded perspective of the target image it is trained on.

(a) Airport (b) Farm (c) Forest

(d) Water (e) Urban

Figure 9. USGS High Resolution Orthoimagery dataset samples, courtesy USGS

Image Parsing.

The method for image parsing of the HRO dataset imagery will be described. A python script was written to slice the large pixel area images into four equal sized tiles of approximately 257 x 257 pixels. The original tiles were removed in an effort to prevent unnecessary parsing of the original 1028 x 1028 pixel sized images. The same python script, modified to slice two equal sized tiles, was applied to each of the four 257 x 257 sliced images which generated two equal sized tiles of approximately 128 x 128 pixels from each of the four 257 x 257 pixel sized images. This process generated eight equal sized tile images of approximately 128 x 128 pixels. The python script utilizes the function image slicer which appeared to magnify images with dimensions that do not render even numbers when divided. For instance the number 257 divided

by 2 renders the value 128.5, however, in an effort to maintain whole numbers in the pixel dimensions the decimal value is dropped resulting in a magnified image. The slight magnification did not cause any noticeable impact in the performance of any of the Convolutional Neural Networks developed during this research effort.

National Agriculture Imagery Program.

A large portion of imagery making up the Farm category were drawn from the National Agricultural Imagery Program where a fairly high pixel resolution of approximately 1 meter was evident in each of the images. The pixel area of the images acquired from the NAIP dataset was approximately 772 x 1028 pixels. As stated previously in the HRO subsection, a Caffe model requires an enormous amount of data in order to obtain an efficient classification capability. In particular, HybridCaffe was given input imagery with a pixel resolution of approximately 96 x 128 pixels. This is intended match with the dimensions established in the parsing of images gathered from the HRO dataset.

Figure 10. Sample from National Agricultural Imagery Program, courtesy USGS

Image Parsing.

The method for image parsing of the NAIP dataset imagery will be described. A python script was written to slice the large pixel area images into four equal sized tiles of approximately 192 x 256 pixels in size. The original tiles were removed in an effort to prevent unnecessary parsing of the original 770 x 1026 pixel sized images. The same python script, modified to slice two equal sized tiles was applied to each of the four 192 x 257 sliced images which generated two equal sized tiles of approximately 96 x 128 pixels from each of the four 192 x 256 pixel sized images. This process generated eight equal sized tile images of approximately 96 x 128 pixels in size. As in the parsing of the HRO imagery dataset, the function utilized in the python script magnifies images with dimensions that do not render even numbers. For instance the number 770 divided by 4 renders the value 192.5, however, in an effort to maintain whole numbers in the pixel dimensions the decimal value is dropped resulting in an image that appears magnified. This process of slight image fuzzing is a result of the magnification which reflects the loss of pixels during the parsing process. Again, as described in the image parsing process applied the HRO dataset, no noticeable impact was observed in the performance of any of the Convolutional Neural Networks developed during this research effort.

Supplementary Data Additions.

Scarcity of certain scene categories called for a search of additional imagery sources. In addition to the data acquired in from the HRO and NAIP datasets, acquisition of supplementary data for the Forest and Airport category is described. Supplementary Forest category imagery obtained from the UC Merced dataset[1] provided very high resolution imagery with a pixel resolution of 12 inches. Forest images possessed ideal dimensions of 256 x 256 pixels. Origins of the data source

stems from the National Map of the High Resolution Orthoimagery dataset from the United States Geological Survey. Supplemental Airport imagery obtained from the DigitalGlobe provided high resolution imagery of Airports with a pixel resolution of 0.305 meters. Airport images snipped with Microsoft Snipping tool generated varying pixel dimensions in size, from 350 x 350 pixels to 128 x 128 pixels.

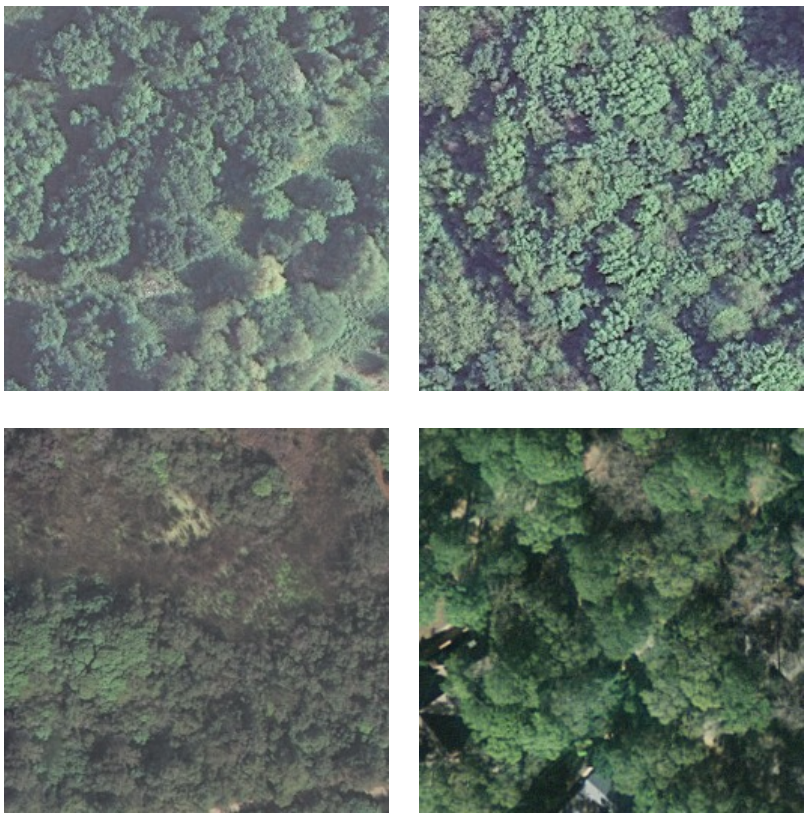


Figure 11. Samples from the Forest Class of the UCMerced Dataset [1], courtesy USGS

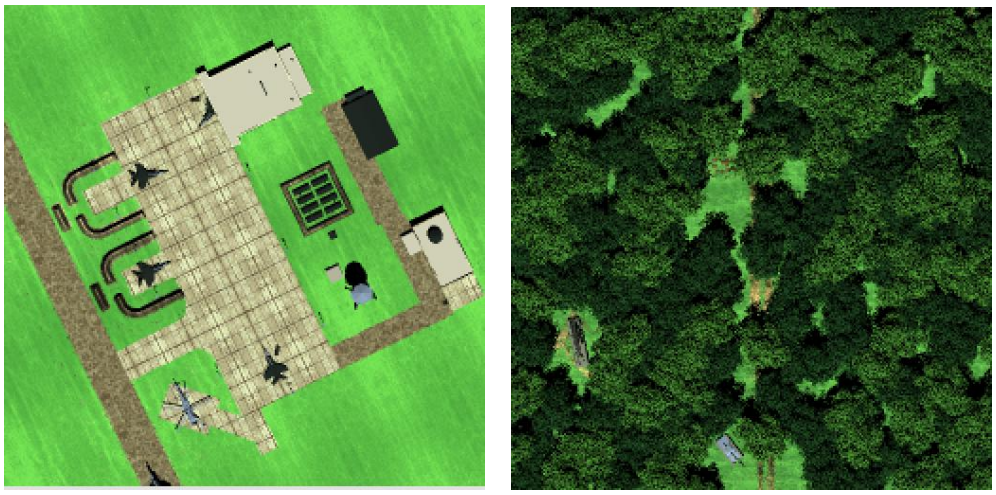
Figure 12. Sample from DigitalGlobe Collection

3.2 Synthetic Image Generation

The process of synthetic image generation utilizing the remote sensing platform, DIRSIG, is described in subsequent section.

DIRSIG Simulator Suite.

The synthetic image generation tool, DIRSIG, was utilized to generate two additional categories, DIRSIG Airport and DIRSIG Forest, to be included in the Hybrid-Caffe training dataset. Methods for generating the set of images will be described. Generation of both categories involved appropriate utilization of the imaging platform, platform motion and data collection settings of the DIRSIG simulator suite.



(a) DIRSIG Airport (b) DIRSIG Forest

Figure 13. Samples from DIRSIG [2]

Imaging Platform.

The imaging platform required specification of the clockrate of the RGB Focal Plane to be set at a low rate of 1 Hertz. Additionally, the focal length was specified at 1000 millimeters.

Platform Motion.

The intention of the platform motion editor is to model four unique passes over each of the DIRSIG Airport and DIRSIG Forest scenes. Each pass generated set of Ground Sample Distance (GSD) values that contributed to the generation of a unique set of synthetic images. For all flight passes a GSD of 1 meter/scan was recorded. This low GSD was recorded to account for the small size of each of the scenes.

Straight Flight Path.

The first pass, an altitude of 5,000 meters, a velocity of 1.00 m/s, time delta of 1.00 and a duration of 250 seconds was specified. These settings in combination with

the Imaging Platform settings previously described allowed for the platform to pass across the scene in a straight flight pattern without any heading variations. Imagery was collected every second during the flight line and stored into a single capture file. This generated a total of 250 unique scenes.

45° Flight Path.

The second pass possessed all of the settings described in the first pass with a change to a heading specified at 45 degrees. These settings, in combination with the Imaging Platform settings previously described, allowed for the platform to pass across the scene in a diagonal flight path where the heading of 45 degrees is specified. Imagery was collected every second during the flight line and stored into a single capture file.

135° Flight Path.

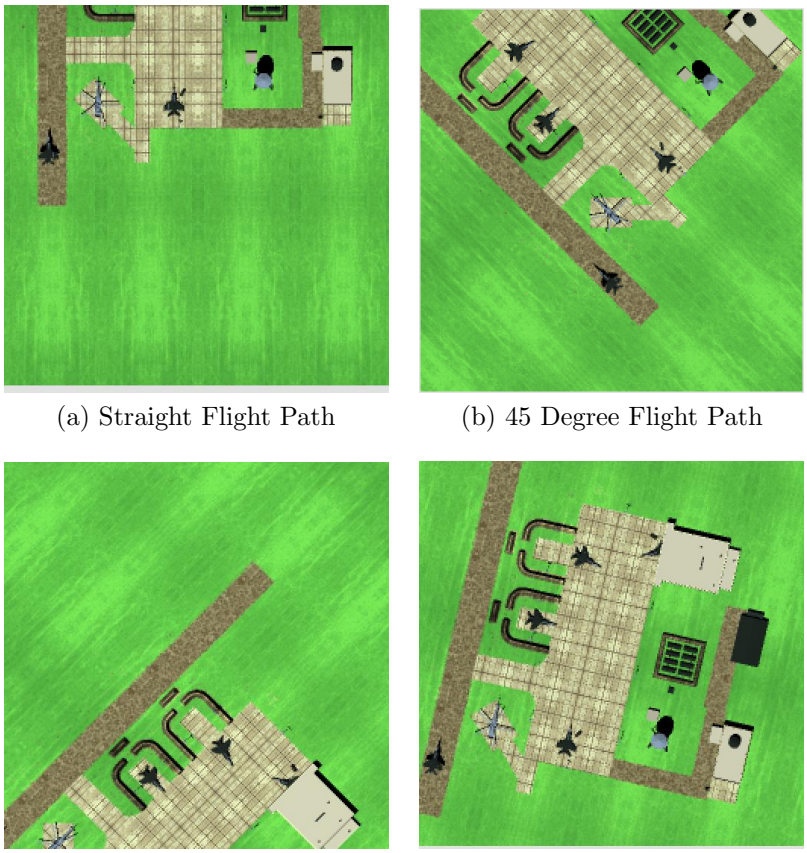
The third pass possessed all of the settings described in the first pass with an addition of a heading specified at 135 degrees. These settings in combination with the Imaging Platform settings previously described, allowed for the platform to pass across the scene in a diagonal flight path where the heading of 135 degrees is specified. Imagery was collected every second during the flight line and stored into a single capture file.

Racetrack Flight Path.

The fourth pass, involved the creation of a racetrack flight path. The specifications for this flight path required an altitude of 5,000 meters, a velocity of 1.00 m/s, time delta of 1.00 and a duration of 250 seconds was specified. These settings in combination with the Imaging Platform settings previously described allowed for the

platform to pass across the scene in a straight flight pattern without any heading variations. Imagery was collected every second during the flight line and stored into a single capture file.

Each of the passes generated a total of 250 unique images, where in summation all of the passes generated a total of 1,000 unique images. In Figure 14, a sample from each of the flight paths described applied to the DIRSIG Airport scene shows the variance in direction of each of the sensor platforms.



(a) Straight Flight Path

(b) 45 Degree Flight Path

(c) 135 Degree Flight Path

(d) Racetrack Flight Path

Figure 14. DIRSIG Flight Path Samples [2]

Data Collection.

A continuous capture window was specified with a capture time of 250 seconds was applied to each of the Straight Flight Path, 45° Flight Path, 135° Flight Path and the Racetrack Flight Path previously described. This in conjunction with the clock pattern specified, ensured an amount of 250 images were collected from each flight path.

Summary.

Each of the Flight Paths described were utilized within a single simulation file. This indicates that there were a total of 4 simulations for the generation of 1000 unique DIRSIG Airport and DIRSIG Forest scenes. This is further described in Table 2, where the number of unique scenes generated are associated with their respective flight path and simulator file.

Table 3. Simulation files applied to DIRSIG Airport and DIRSIG Forest scenes, where each simulation generated 250 images based off of the flight path specifications described.

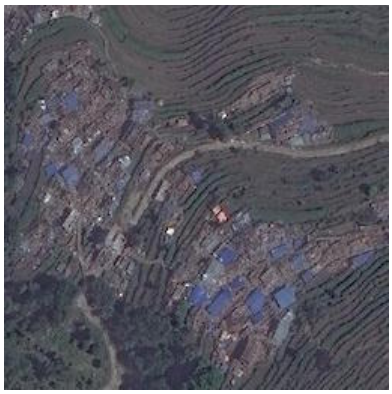
Flight Paths	Generated Image Counts	Simulation Title
Straight	250	Straight.sim
45°	250	45.sim
135°	250	135.sim
Racetrack	250	Racetrack.sim

3.3 Search and Rescue Imagery Acquisition

This section describes the methods of data collection from the Nepal Earthquake Portal provided by the company Tomnod. The company Tomnod, leverages the utilization of crowd-sourcing to identify features of interest in thousands of high res-

olution satellite images. The categories that developed from features extracted by users served as points of confirmation to aid response strategies for blocked roads, tents/shelters, damaged buildings and areas of mass destruction. Imagery selected for training the Caffe model, HybridCaffe were from the categories of damaged buildings, areas of mass destruction and tents/shelters.

(a) Nepal Damaged Buildings



(b) Nepal Total Destruction

(c) Nepal Shelters

Figure 15. Samples from Nepal Earthquake Post-Event, courtesy TOMNOD

3.4 Convolutional Neural Network Development

This section describes the background required for setting the stage of designing a Convolutional Neural Network (CNN) with the purpose of application to experimental design. The CNN designed was made in part with fine-tuning. This is the process

by which the lower layers of a pre-trained CNN are interchanged with a new set of weights similar to the weights that originally trained the pre-trained CNN.

Basic Architecture.

The basic architecture of HybridCaffe is a derivative of the AlexNet network[20]. The basic architecture of a CNN is characterized by alternatively stacked convolutional layers and spatial pooling layers often followed by one or more fully connected layers as in multi-layer perceptron [41]. Similarly, the basic architecture of AfCaffe follows the same architectural pattern. Figure 16, depicts the basic architecture of AfCaffe where the first five main layers of the architecture constitute of convolutional layers and the last three layers comprise of three fully connected layers. The subsequent subsections briefly describe the theory behind the layers involved in each of the main layers which constitute the main architecture of HybridCaffe.

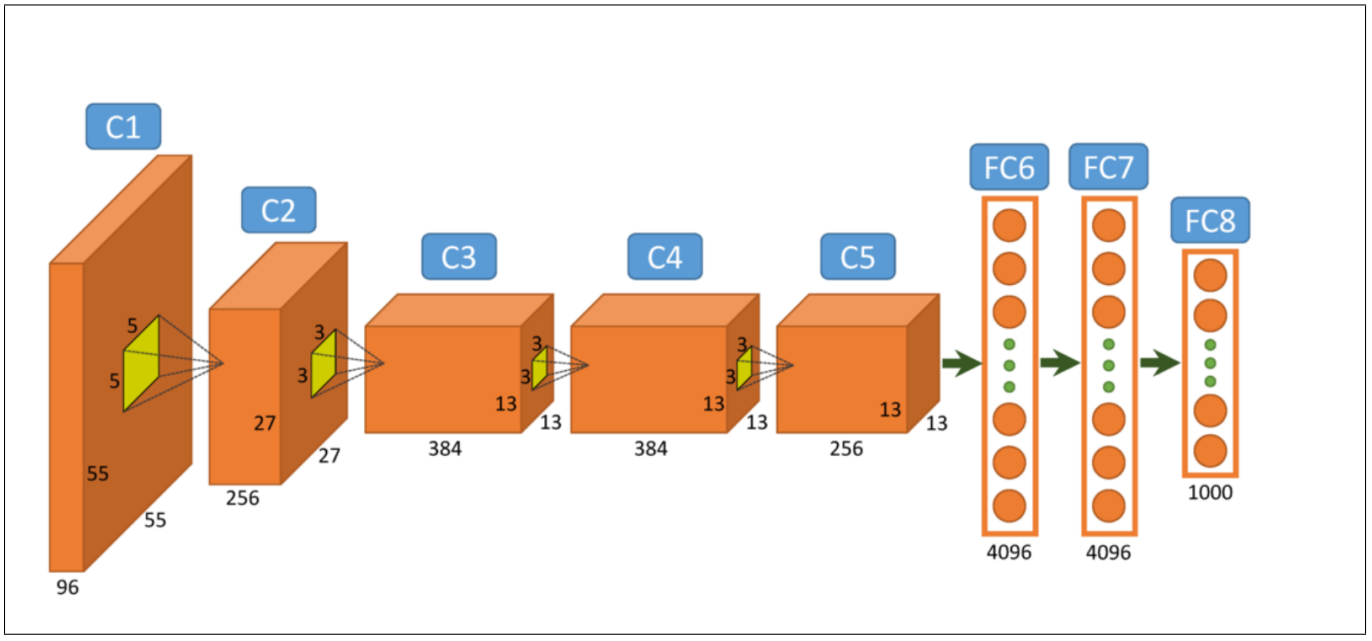


Figure 16. Architecture of HybridCaffe, similar to that of AfCaffe [5]

The Artificial Neuron.

The artificial neuron represents the biological neuron in the human brain. The artificial neuron is also known as the "perceptron." Figure 17 indicates an artificial neuron with it's respective inputs represented as I and connection weights represented by W. This figure is used to describe equations (3) through (5).

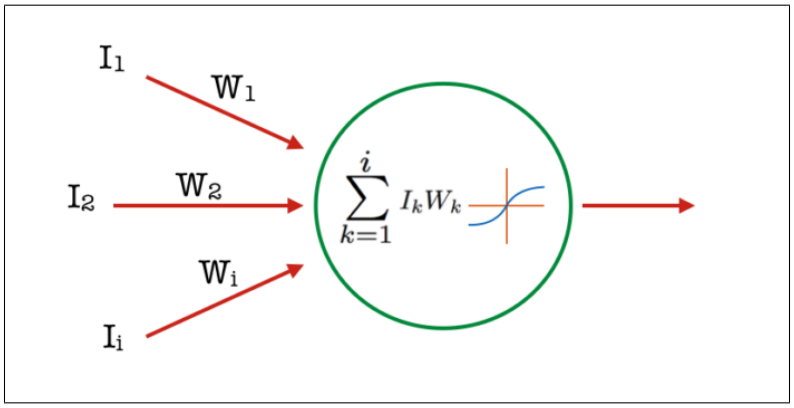


Figure 17. Artificial neuron diagram. Inputs on the left are multiplied by their associated weights (learned values). The sum of these weighted inputs is applied to an activation function which computes the neuron's output[5].

Consider, the input activation function for a given neuron:

$$a(I) = b + \sum_{k=1}^{i} I_k W_k \tag{3}$$

Where b represents the neuron bias, I represents the input and W represents their associated connection weights, respectively. The kth instance of the input and weight vectors are multiplied together inside the summation over all instances within the vector. The connection weights represent the strength of the connection between the input neurons and neighboring neuron.

$$a(I) = b + IW \tag{4}$$

The pre-activation and output activation processes are depicted in Figure 17 where the inputs are multiplied by their respective weights and outputted. Inputs of an artificial neuron are described by the phrase, "pre-activations" and inputs are described by the phrase "output activations". The output activation function is described in Equation (5).

$$h(I) = g(a(I)) = g(b + \sum_{k=1}^{i} I_k W_k) \tag{5}$$

Convolutional Layer.

Initial convolutional layers extract low-level features such as edges, lines and corners within input imagery. Higher-level layers extract higher-level features. Typically, images which are input into the convolutional layer are of dimensions height x width x depth. Where, depth relates to the third dimension of the input volume. Typically, for input images the third dimension is represented by the value 3, indicating the three color channels of the image where the three color channels are represented by Red(R), Green(G), and Blue(B).

Pooling Layer.

Typically, in the architecture of a Convolutional Neural Network, pooling layers are inserted between successive Convolutional layers. The main purpose of this layer is three-fold: (1) reduce the spatial size of the representation, (2) reduce the amount of parameters and computation in a network, (3) control over-fitting [6].

This layer is responsible for filtering sites of the convolutional layer and performs spatial pooling on them. Spatial pooling goes also by the terms sub-sampling and down-sampling. Essentially, this layer performs the reduction of resolution in the features of the images. This makes the features robust against noise and distortion.

There are several types of spatial pooling: max pooling, average pooling, sum pooling, etc. In all cases, the input is divided into non-overlapping two-dimensional spaces. These two dimensional spaces can be viewed as spatial neighborhoods, known as filters which house features of interest. Conceptually, features of interest are represented as numbers as described in Figure 18.

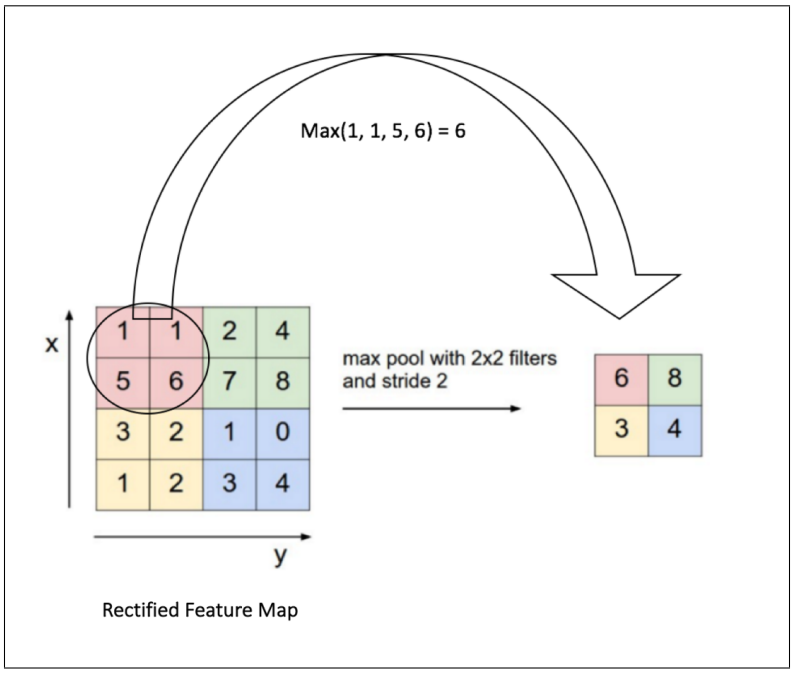


Figure 18. An illustrated example of max pooling with a 2x2 filter and a stride of 2. The maximum values are input into a two dimensional space[6].

Max Pooling.

Max pooling is a type of spatial pooling where the maximum value of each filter is taken. This concept is defined in Figure 18, where the features of interest are represented by numbers where a 2x2 window takes the largest element from the feature map within that window. Max pooling is advantageous as it produces fast convergence during training.

ReLU Layer.

The Rectified Linear Unit (ReLU) Layer is a layer of neurons that applies the non-saturating activation function, indicated in equation 6.

$$f(x) = max(0, x) \tag{6}$$

This increases the nonlinear properties of the decision function and the overall network without affecting the receptive fields of the convolution layer. Deep convolutional neural networks with ReLU's train several times faster than their equivalents with tanh units[20]. Unfortunately, ReLU units tend to be fragile and output the same value for any input during training when the learning rate is set too high. This results in the ReLU unit no longer engaging in its role of discerning the differences between inputs. Typically, a ReLU unit that enters this state is unable to recover into a normal state. In some cases, recovery requires several trial and error attempts to determine the appropriate learning rate for the network.

Fully Connected Layer.

The fully connected layer, is a portion of the architecture that high-level reasoning in the neural network is executed through fully connected layers. All units in a fully-connected layer are connected to all units in the next lower layer. Information that is output from the pooling and convolutional layers hold high-level features of the input image. The main purpose of the fully connected layer is to use the high level features to classify the input image into various classes based off of the training dataset. The sum of the output probabilities from the fully-connected layer is 1. This is indicated in Figure 19. Where the sum of all the Output Predictions equate to 1, with the highest Output Prediction as boat with an output prediction of 0.94.

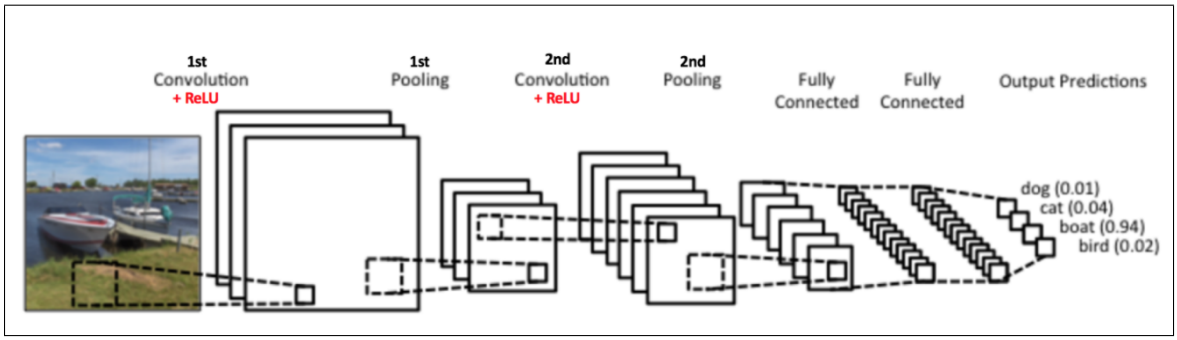


Figure 19. An illustration of the basic Convolutional Nueral Network with the four main operations indicated. [7].

Utilization of the softmax as the activation function in the output layer of the fully-connected layer ensures production of the appropriate Output Predictions. In other words, the convolutional and pooling layers serve to extract features from the input image, while the fully connected layer acts as the classifier.

Fine-Tuning a CNN.

The process of fine-tuning a CNN network with the purpose of rapid and efficient CNN model development is described. Designing a Caffe network from the ground-up can be a rather challenging and time-consuming task as several hundred lines of code will be required to build the network. An appropriate work around to this is the utilization of the weights of a Caffe network that has already been trained. Advantageous results can be obtained from fine-tuning a network as demonstrated by works in [16] and [48] in the Literature Review Section.

Fine-Tuning HybridCaffe.

Fine-tuning will be described from a practical standpoint in terms of the two CNN models developed during this research effort, HybridCaffe and HiReSatCaffe. HybridCaffe was originally HiReSatCaffe where the five categories; Airport, Farm,

Forest, Water and Urban imagery drawn from the HRO and NAIP datasets were used for training purposes. HiReSatCaffe was developed by fine-tuning off of the AfCaffe network [5]. This was done by executing the appropriate edits in the AfCaffe train validation file, namely, replacing the lower layers of the CNN, specifically, fully connected layers seven and eight with weights from the HiReSatCaffe dataset. The inner product layer required specification of the output layers to correlate to the number of categories which the CNN model trained on. For the case of HiReSatCaffe the output layers were specified to be five to correlate with each of the categories: Airport, Farm, Forest, Water, and Urban. HybridCaffe was developed when the two categories of synthetic imagery from DIRSIG was generated and added along with the imagery from Tomnod's open source Nepal Earthquake dataset. These two major additions to the training image categories constituted for a new total of 10 categories: Airport, Farm, Forest, Water, Urban, DIRSIG Airport, DIRSIG Forest, Nepal Buildings, Nepal Destruction and Nepal Shelter. The appropriate adjustments were executed in the inner product layer to ensure the new number of categories correlated to the correct number of output layers.

3.5 Research Question 2 Experiments

Further investigation was realized after the development of the Tipping and Cueing framework consisting of a Tip and Cue Exchange Protocol and Tier Organization based off of the differing pixel resolutions. This development indicated the need for further exploration in the reliability of a Convolutional Neural Network in the space vehicle domain. The investigation conducted centered on automation of terrain classification with Machine Learning.

First CNN Iteration.

The approach taken is an investigation of the possibility to automate terrain classification with a Machine Learning framework. In order to effectively train a Convolutional Neural Network a large dataset must be generated. This approach followed the steps established in the Data Acquisition Section where understanding the Remote Sensing domain, followed by parsing images, and building a dataset is conducted. Once the appropriate dataset is generated, a Convolutional Neural Network must be developed. The process of building the first CNN is described by understanding basic CNN architecture, and by fine-tuning a CNN as described in the fine-tuning section. The first CNN iteration utilized image classes depicted in Figure 9. A total of 3,198 images were used in the network where 10% of the images were randomly allocated to a testing set and 90% of the images were randomly allocated to a training set. The training set consisted of 2,875 images and the validation set consisted of 323 images. All observations and insights found during the training phase is described in the Results Chapter.

3.6 Research Question 3 Experiments

In this section the experimental design of Research Question 3 is described where incorporation of synthetic imagery to model satellite sensor capabilities is investigated. HiReSatCaffe is a CNN trained on terrain specific imagery, a combination of synthetic imagery and Nepal Earthquake Post-Event imagery is incorporated for the creation of a new CNN called: HybridCaffe. A sequential approach is executed in an effort to achieve the general objective for this set of experiments. As stated in the Assumptions section, experiments will be executed in appropriate sequential order where no experiment can be started until the previous one is finished. The overall purpose of the series of experiments is to determine if it is possible to incorporate

synthetic imagery to model satellite sensor capabilities and determine what the most important factors are in influencing the performance of HybridCaffe.

3.7 HybridCaffe Classification Evaluation

The purpose of this experiment set is to determine whether or not the classification accuracy of HybridCaffe is truly genuine when synthetic imagery is incorporated.

The Experiment Design.

The CNN model, HybridCaffe is trained on a large training dataset of 9,000 images for a span of 10,000 iterations. Once the classifier is trained on the appropriate number of iterations a confusion matrix and a True Positive Percentage curve is generated to determine the genuine accuracy of the classifier.

3.8 Assumptions

All experiments are run in sequential order with recorded data reflecting that. Only 1 unique DIRSIG synthetic scene will be applied to each Jitter Class, as indicated in Figure 20. Each Jitter Class will generate a jittered replicate of the original unique image.

Response Variable.

The table below indicates the response variables from the HybridCaffe Classification Evaluation experiment sets.

True Positive Threshold Values.

True Positive values indicate decision points where images sent into HybridCaffe are confirmed as an airport or not based on the probability percentage. Refer to the

Results Chapter for the exact Probability Percentage values utilized in the calculation of the True Positive Percentages. The True Positive Percentage (TP%) is the percent of images classified as airports out of the total number of images that are airports. Threshold values between 0 and 1 are specified to determine the percentage of True Positive (TP) results. For instance, a threshold of 0.80 indicates that any score above 0.80 will be recorded as a True Positive result. In the case of a threshold of 0, a total of 100% will be recorded, as all of the probability percentages recorded are greater than 0. A true positive vs threshold graph illustrating this behavior is provided in the Results Chapter.

3.9 Classification Factor Investigation

The purpose of this experiment set is to determine what the important factors of influence are in HyrbridCaffe's ability to accurately classify an image. In particular, the relationship between stochastic gradient descent and HybridCaffe's classification sensitivity is analyzed. The main factor evaluated is image clarity represented by "Probability Percentage", as evident in the Experiment Test Matrix.

The relationship between stochastic gradient descent and Caffe's classification sensitivity exists for a number of reasons. Initially, it is evident that the classifier is heavily influenced by the inconsistent drop and rise in rendered probability percentages when performing classifications. The sensitivity of the classification capability is drawn out from the certainty and failure thresholds which are calculated from the Probability percentages that are influenced by stochastic gradient descent.

Jitter Definition.

In an effort to set the stage for the description of the experiment design and control variables utilized the sensor performance measure is defined. The definition

of jitter in a synthetic image model depends on what distribution the synthetic model of interest utilizes when it generates the jitter. In the case of the DIRSIG model, jitter is modeled according to a Gaussian distribution where it is assumed that the jitter is characterized by variation in standard deviation. Therefore, in this research domain jitter is defined as a deviation in the sensor's pointing with a Gaussian distribution. Again, this is based off of the distribution modeled by the synthetic tool used to generate the synthetic images.

Jitter Classes and Jitter Variance Ranges.

In the Classification Factor experiment set each experiment utilizes an image generated from DIRSIG that is intended to simulate a jittered satellite image. A specified amount of uncorrelated jitter variance is applied to an image. Each step of increase in uncorrelated jitter variance will constitute its own respective Jitter Class. A series of Jitter Classes will constitute a single Jitter Variance Range. Table 4 summarizes the amount of uncorrelated jitter variance applied to a single image with each respective jitter class label. The single table can be viewed as a single jitter variance range. In this case the controllable noise factor is the Jitter Class, or uncorrelated jitter variance.

Table 4. Visual of each Jitter Class and their respective jitter variances

Jitter Class	Jitter Increase
Jitter Class 1	0.0001
Jitter Class 2	0.0002
Jitter Class 3	0.0003
Jitter Class 4	0.0004
Jitter Class 5	0.0005
Jitter Class 6	0.0006
Jitter Class 7	0.0007
Jitter Class 8	0.0008
Jitter Class 9	0.0009
Jitter Class 10	0.0010

The Experiment Design.

A single set of synthetic DIRSIG images are sent as inputs into HybridCaffe one time for each Jitter Variance Range. The probability percentage for each Jitter Class image is recorded in the Test Matrix. After the final run is executed and the data is recorded, the average for all Probability Percentages recorded from each run is calculated and provided in Table 10.

Probability Percentage Averages.

The table below indicates the averages recorded for the probability percentage averages of Jitter Variance Range A, Jitter Variance Range B, Jitter Variance Range C and Jitter Variance Range D. These averages are correlated with error bars in Figure 29 provided in the Results Chapter.

Table 5. Jitter Matrices Averages

Jitter Matrix	Probability Average
Jitter Matrix A	0.16995
Jitter Matrix B	0.77534
Jitter Matrix C	0.29183
Jitter Matrix D	0.91066

Control Variables.

The control variables utilized in the Classification Factor Investigation Experiment set are described. In an effort to obtain a sense of realism, each experiment utilizes an image generated from DIRSIG which is intended to simulate a jittered satellite image. A step of uncorrelated jitter variance in radians is applied to an image. Each step size of increase in jitter variance will constitute it's own respective Jitter Class. Where for instance Jitter Class 1 has a jitter step of 0.0001 radians, rendering the Jitter Class value of 1. The subsequent Jitter Class will be applied the same jitter step size rendering 0.0002, correlating to Jitter Class 2. Table 3 summarizes the amount jitter variance applied to a single image with each respective jitter class label. The controllable noise factor will be the Jitter Class.

Constant Factors.

For each experiment certain factors are held constant for the sake of simplicity. This ensures the necessary limit of scope in each experiment. Figure 20 depicts the constant factors applied in the Classifier Factor experiment set. For each experiment the same DIRSIG scene will be applied to each Jitter Class to observe effects. Effects are recorded in test matrices included in this section and curves are generated to indicate trends for further analysis in the Results Chapter.

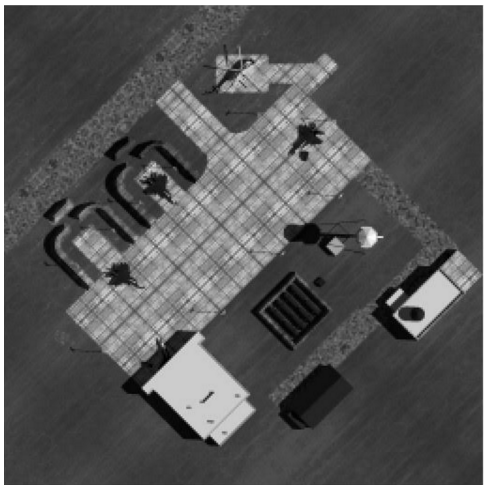


Figure 20. Screen shot of a synthetic DIRSIG scene, generated by a simulated Hyper-Spectral Imaging Pushbroom Sensor. [2]

Modeling the Hyper-Spectral Imaging Sensor.

This subsection of the Constant Factor subsection describes the process of modeling the Hyper-Spectral Imaging (HSI) Sensor using the DIRSIG simulator suite. Generation of the image indicated in Figure 20, involved appropriate utilization of Imaging Platform to appropriately model the pushbroom spectrometer. As added background, this spectrometer type is a 1D linear array that instantaneously captures the across-track dimension and consecutive read-outs are combined to form an image. The radiance image generated contains 30 spectral channels, where for this research the radiance image associated with channel 26 is utilized.

Test Matrix.

Table 6. Jitter Variance Range A: High Uncorrelated Jitter(Step of 0.0001 radians)

Jitter Class	Jitter Step Size [radians]	Jitter Variation [radians]
Jitter Class 1	0.0001	0.0001
Jitter Class 2	0.0001	0.0002
Jitter Class 3	0.0001	0.0003
Jitter Class 4	0.0001	0.0004
Jitter Class 5	0.0001	0.0005
Jitter Class 6	0.0001	0.0006
Jitter Class 7	0.0001	0.0007
Jitter Class 8	0.0001	0.0008
Jitter Class 9	0.0001	0.0009
Jitter Class 10	0.0001	0.0010

Table 7. Jitter Variance Range B: Medium-Low End Uncorrelated Jitter(Step of 0.00001 radians)

Jitter Class	Jitter Step Size [radians]	Jitter Variation [radians]
Jitter Class 1	0.00001	0.00011
Jitter Class 2	0.00001	0.00012
Jitter Class 3	0.00001	0.00013
Jitter Class 4	0.00001	0.00014
Jitter Class 5	0.00001	0.00015
Jitter Class 6	0.00001	0.00016
Jitter Class 7	0.00001	0.00017
Jitter Class 8	0.00001	0.00018
Jitter Class 9	0.00001	0.00019
Jitter Class 10	0.00001	0.00020

Table 8. Jitter Variance Range C: Medium-Higher End Uncorrelated Jitter(Step of 0.00001 radians)

Jitter Class	Jitter Step Size [radians]	Jitter Variation [radians]
Jitter Class 1	0.00001	0.00021
Jitter Class 2	0.00001	0.00022
Jitter Class 3	0.00001	0.00023
Jitter Class 4	0.00001	0.00024
Jitter Class 5	0.00001	0.00025
Jitter Class 6	0.00001	0.00026
Jitter Class 7	0.00001	0.00027
Jitter Class 8	0.00001	0.00028
Jitter Class 9	0.00001	0.00029
Jitter Class 10	0.00001	0.00030

Table 9. Jitter Variance Range D: Low Uncorrelated Jitter(Step of 0.00001 radians)

Jitter Class	Jitter Step Size [radians]	Jitter Variation [radians]
Jitter Class 1	0.00001	0.00001
Jitter Class 2	0.00001	0.00002
Jitter Class 3	0.00001	0.00003
Jitter Class 4	0.00001	0.00004
Jitter Class 5	0.00001	0.00005
Jitter Class 6	0.00001	0.00006
Jitter Class 7	0.00001	0.00007
Jitter Class 8	0.00001	0.00008
Jitter Class 9	0.00001	0.00009
Jitter Class 10	0.00001	0.00010

IV. Results and Discussion

This chapter describes the process of executing a set of experiments to address each of the research questions posed in the introduction. The investigation of the automation of terrain imagery is covered in the appropriate application of training HiReSatCaffe. Followed by the investigation of the incorporation of synthetic imagery which created a new CNN called HybridCaffe. A larger number of experiments were applied in an effort to determine how synthetic imagery influences the CNN's classification capacity where factors that influence classification are investigated. Observations drawn from the Classification Factor Investigation and the HybridCaffe Classification Evaluation Experiment Sets. The results from each experiment in the Classification Factor Investigation experiment set indicate key factors of influence in the classification capacity of HybridCaffe. The HybridCaffe Accuracy Evaluation experiment set results indicate the true strength of HybridCaffe's classification accuracy. Insights found on all accounts given are described in subsequent sections of this chapter.

4.1 Research Question 2 Experiments

The number of experiments are less verbose compared to the set of experiments developed to investigate research question 3. As described in the Methodology Chapter, much of the investigation of research question 2 focused on understanding the problem domain, generating a dataset and developing the appropriate Convolutional Neural Network to apply classifications. A total of 5 classes were utilized in the training of HiReSatCaffe to represent terrain typical for a search and rescue event. As stated in the Methodology chapter, a total of 3,198 images were used in the network where

the training set consisted of 2,875 images and the testing set consisted of 323 images. Observations of the CNN's performance is described in the next subsection.

HiReSatCaffe Performance.

As the purpose of the first CNN iteration involved development of a basic CNN a small dataset was used along with a short training time. HiReSatCaffe was set to train for a total of 1,000 iterations where every 10 iterations the CNN applied images from the testing set to assess HiReSatCaffe's classification performance. During training HiReSatCaffe demonstrated a high network accuracy of 98%, which served as a indicator that over-fitting possibly occurred within the CNN. Creation of the CNN, HybridCaffe mitigated over-fitting by adding five categories to increase the volume of the training dataset. As over-fitting occurs when the ratio between the CNN complexity and the training set is too high. In the case of HiReSatCaffe over-fitting occurred since it was fine-tuned off of Imagenet a vastly complex CNN model and was trained on a small training set, with only 323 images in the testing set.

4.2 Jitter Variance Range Results

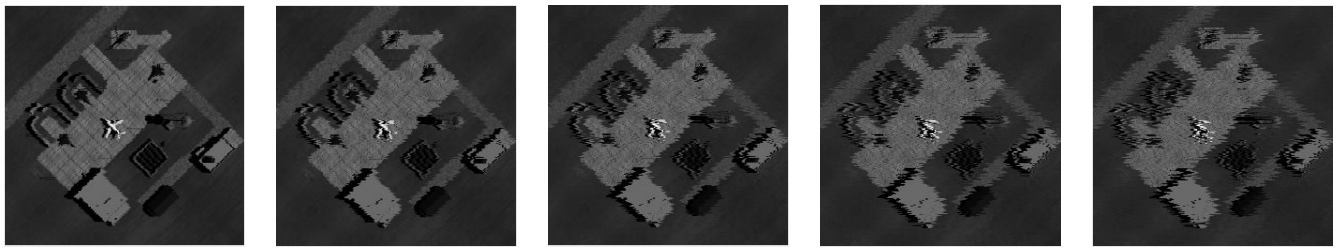
This section describes results rendered from experiments that investigate impacts synthetic imagery has on HybridCaffe's classification capacity. As stated in the Methodology chapter, the purpose of this experiment set is to unveil characteristics and features that influence classifications. Gradient descent, image edge sharpness and image contrast appeared to be the major sources of influences in classification performance. Four sets of 10 synthetic DIRSIG images were applied a percentage of jitter and sent to the classifier. Every image was applied a unique amount of jitter, in radian step size.

Jitter Variance Range A.

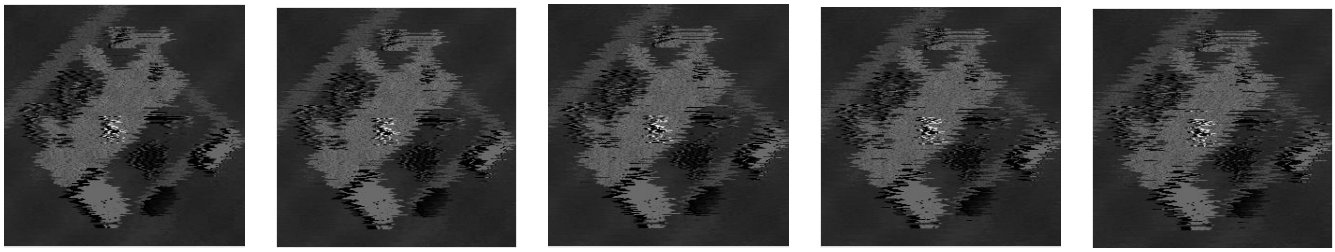
Jitter Variance Range A constitutes of DIRSIG synthetic images jittered to a step of 0.0001 micro radians. The control variable for this experiment is a single synthetic DIRSIG image A sequential step increase of 0.0001 applied to each synthetic image produced a drastic drop in classification ability as indicated in Table 10.

Table 10. Jitter Variance Range A: Percentage Probabilities

Jitter Class	**Probability Percentage**
Jitter Class 1	0.9304
Jitter Class 2	0.7143
Jitter Class 3	0.0516
Jitter Class 4	0.0012
Jitter Class 5	0.0015
Jitter Class 6	0.0003
Jitter Class 7	0.0002
Jitter Class 8	0
Jitter Class 9	0
Jitter Class 10	0



(a) 0.0001 radians (b) 0.0002 radians (c) 0.0003 radians (d) 0.0004 radians (e) 0.0005 radians



(f) 0.0006 radians (g) 0.0007 radians (h) 0.0008 radians (i) 0.0009 radians (j) 0.0010 radians

Figure 21. Jitter Variance Range A: High Uncorrelated Jitter(Step of 0.0001 radians)

Jitter Curve Discussion and Observations.

Observations of Jitter Curve A, indicated in Figure 22, is described. The characteristics of the curve were as expected, a pronounced drop was evident after the third Jitter Class image was applied. This drop in probability percentage indicated that scrambling of the clean edges within the image caused the classifier severe confusion. An interesting parallel is drawn where synthetic images applied jitter above 0.0003 micro radians were classified as images from the Nepal Earthquake Class. Of the three Nepal Earthquake Classes, Nepal Total Destruction denoted by (ND), possessed the highest degree of destruction in comparison to Nepal Damaged Buildings (NB) and Nepal Shelter (NS). This behavior is evident in Figure 22, as the increase of uncorrelated jitter decreases HybridCaffe's sensitivity to airport images.

Confidence Interval Interpretation.

The wide confidence interval indicated in Figure 22, is typical for a small sample set. As an inverse square root relationship exists between the confidence interval and the sample size, the small sample size of 10 values generated a wide confidence interval. A non-parametric regression method, local regression is applied to the data set to draw out key characteristics. The LOESS method in R is utilized to generate the dark blue trend line, to highlight key characteristic behavior indicated in the data points. As the data points indicate there is a substantial drop between Jitter Class 2 and Jitter Class 3 where the synthetic airport images are mistakenly classified as Nepal Destruction Class images. This is an example of a False Negative, where the true condition is not met.

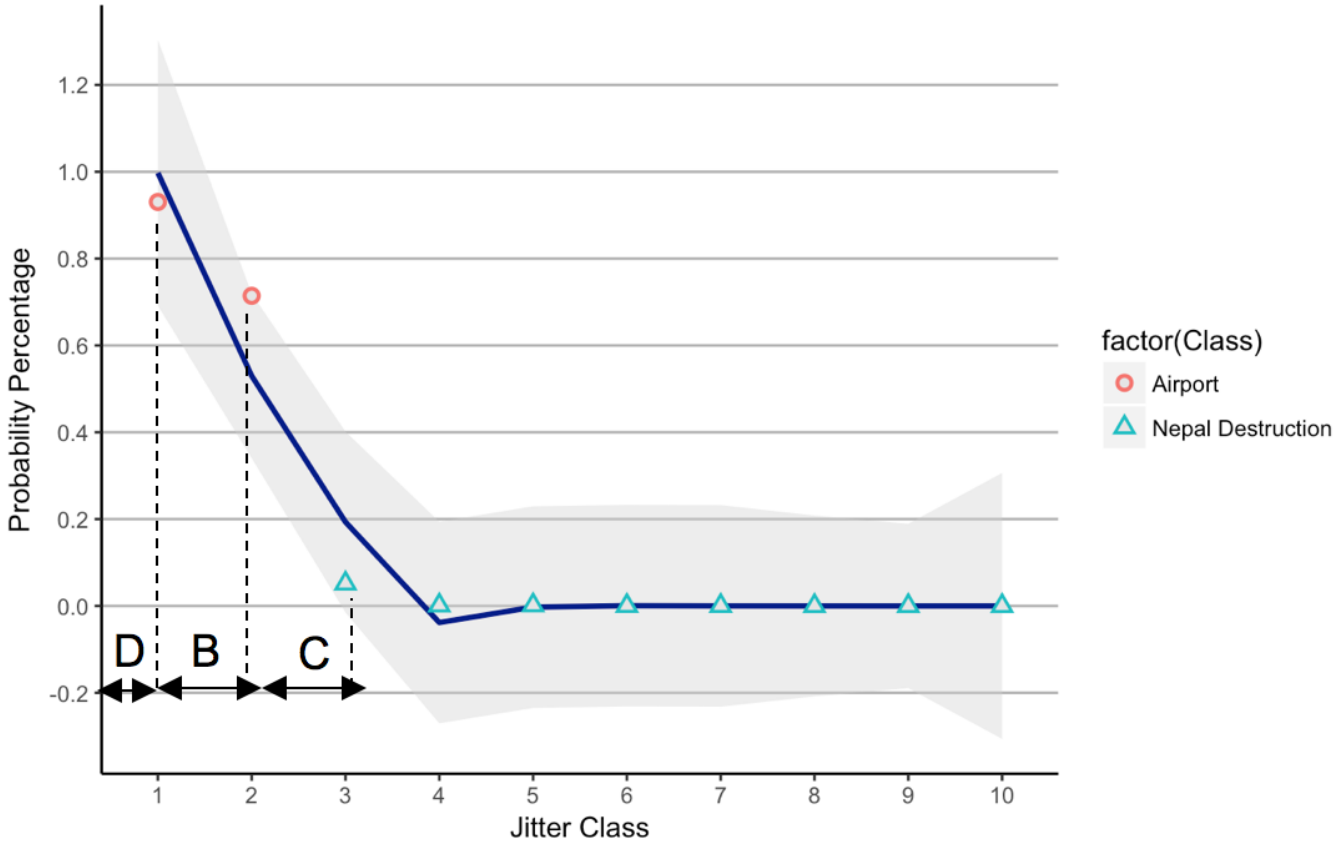


Figure 22. Jitter curve for Jitter Variance Range A: High Uncorrelated Jitter(Step of 0.0001 radians)

Further Exploration.

Based off of the behavior of the curve it was hypothesized that the certainty threshold resided between probability percentages 0.9304 and 0.7143 and the failure threshold resided between probability percentages of 0.7143 and 0.0516. The threshold for certainty will constitute for a positive classification and the failure threshold will constitute for a negative classification. In other words, any classifications drawn with a percentage below the positive classification threshold is not to be counted as a positive classification. The purpose behind generation of Jitter Curves B, C and D is described. Investigation of the exact positive classification threshold is conducted in the analysis of Jitter Variance Range B. Investigation of the exact negative classification threshold is conducted in Jitter Variance Range C. For added measure, determination of a high positive classification threshold is investigated in Jitter Variance Range D.

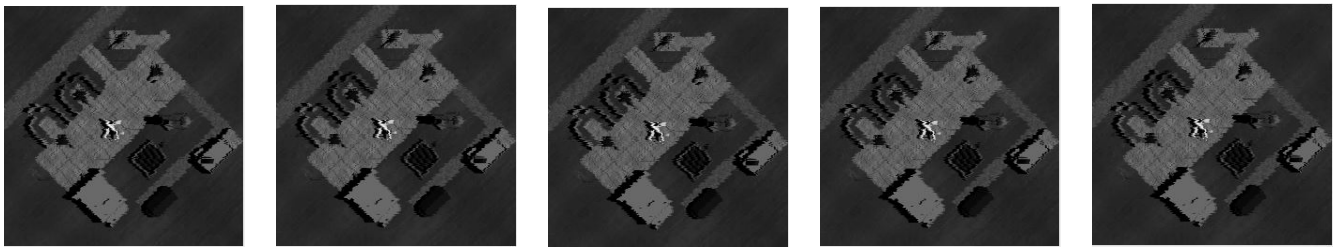
Jitter Jitter Variance Range B.

Jitter Variance Range B constitutes of DIRSIG synthetic images jittered to a step of 0.00001 micro radians. The step of micro radians were chosen to determine a classification threshold. A sequential step increase of 0.00001 applied to each synthetic image produced a curve with consistent True Positive results as indicated in Jitter Curve B. The threshold was calculated by taking the average of the probability percentages in Jitter Variance Range B. The confirmation threshold was calculated to be 0.7753, this indicates that for the purposes of this investigation the confirmation threshold was found to be the probability percentage of 0.7753.

Table 11. Jitter Variance Range B: Percentage Probabilities

Jitter Class	Probability Percentage
Jitter Class 1	0.9304
Jitter Class 2	0.8222
Jitter Class 3	0.7783
Jitter Class 4	0.7578
Jitter Class 5	0.8106
Jitter Class 6	0.8086
Jitter Class 7	0.8352
Jitter Class 8	0.5652
Jitter Class 9	0.7869
Jitter Class 10	0.6582

(a) 0.00011 radians (b) 0.00012 radians (c) 0.00013 radians (d) 0.00014 radians (e) 0.00015 radians



(f) 0.00016 radians (g) 0.00017 radians (h) 0.00018 radians (i) 0.00019 radians (j) 0.00020 radians

Figure 23. Jitter Variance Range B: Medium-Low-End Uncorrelated Jitter(Step of 0.00001 radians)

Jitter Curve Discussion and Observations.

Observations of Jitter Curve B, shown in Figure 24, are described. The characteristics of the curve were as expected, a slow drop was evident after the first Jitter Class image was applied. This slow drop in accuracy indicated that HybridCaffe possessed a form of resiliency against the lighter amount of uncorrelated jitter applied. The characteristics of stochastic gradient descent were evident in the inconsistent trend indicated in the slow drop. This includes a small peak in Jitter Class six and a small drop indicated in Jitter Class eight. This behavior is evident in Figure 24, as the increase of low-end uncorrelated jitter appears to have minimal effect on HybridCaffe's sensitivity to airport images.

Confidence Interval Interpretation.

Jitter Variance B shows a wide confidence interval, similar to Jitter Variance Range A, again due to a small sample size. As the data points indicate there is a slow drop between Jitter Class 1 and Jitter Class 10, where the synthetic airport images are correctly classified as Airport Class images. This is an example of a True Positive, where the true condition is met.

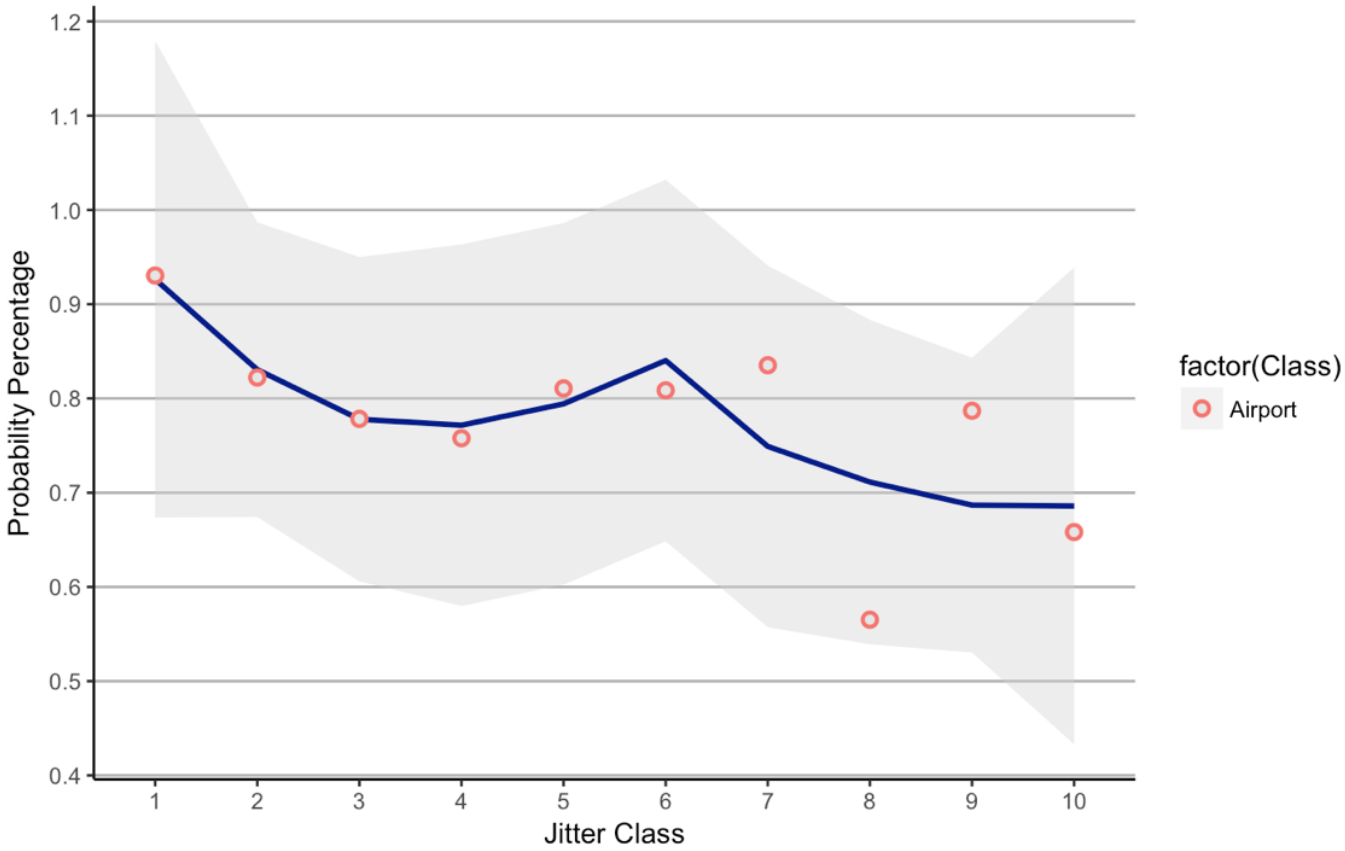


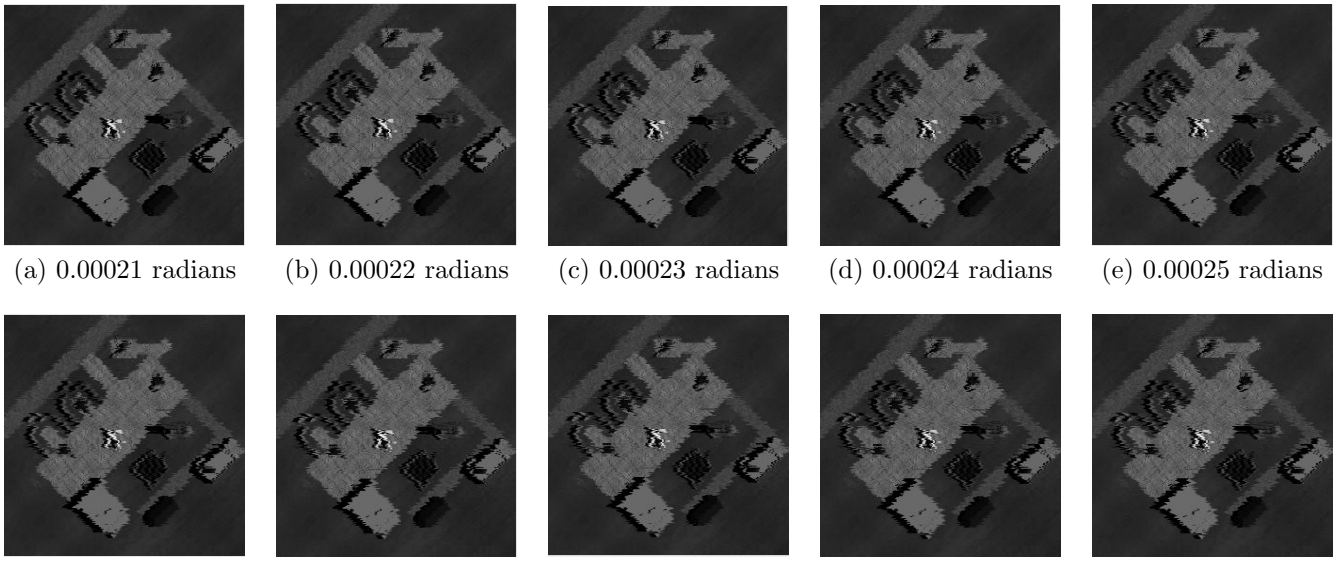
Figure 24. Jitter Curve for Jitter Variance Range B: Medium-Low-End Uncorrelated Jitter(Step of 0.00001 radians)

Jitter Variance Range C.

Jitter Variance Range C constitutes of DIRSIG synthetic images jittered to a step of 0.00001 radians where each step of jitter variation were between 0.00021 radians and 0.00030 radians. The step of micro radians were chosen to determine a classification threshold. A sequential step increase of 0.00001 applied to each synthetic image produced a gradual drop in classification accuracy as indicated in Jitter Curve C. A small peak was indicated in the fifth Jitter Class. As indicated in Figure 26 Jitter Variance Range C, the amount of jitter is very pronounced after Jitter Class six. The threshold was calculated by taking the average of the probability percentages in Jitter Variance Range B. The threshold was calculated to be 0.2918. This indicates that for the purposes of this investigation the failure threshold was found to be the probability percentage of 0.2918.

Table 12. Jitter Variance Range C: Percentage Probabilities

Jitter Class	**Probability Percentage**
Jitter Class 1	0.7143
Jitter Class 2	0.6481
Jitter Class 3	0.4074
Jitter Class 4	0.4037
Jitter Class 5	0.1861
Jitter Class 6	0.8086
Jitter Class 7	0.1089
Jitter Class 8	0.1551
Jitter Class 9	0.0145
Jitter Class 10	0.0516



(a) 0.00021 radians (b) 0.00022 radians (c) 0.00023 radians (d) 0.00024 radians (e) 0.00025 radians

(f) 0.00026 radians (g) 0.00027 radians (h) 0.00028 radians (i) 0.00029 radians (j) 0.00030 radians

Figure 25. Jitter Variance Range C: Medium-High-End Uncorrelated Jitter(Step of 0.00001 radians)

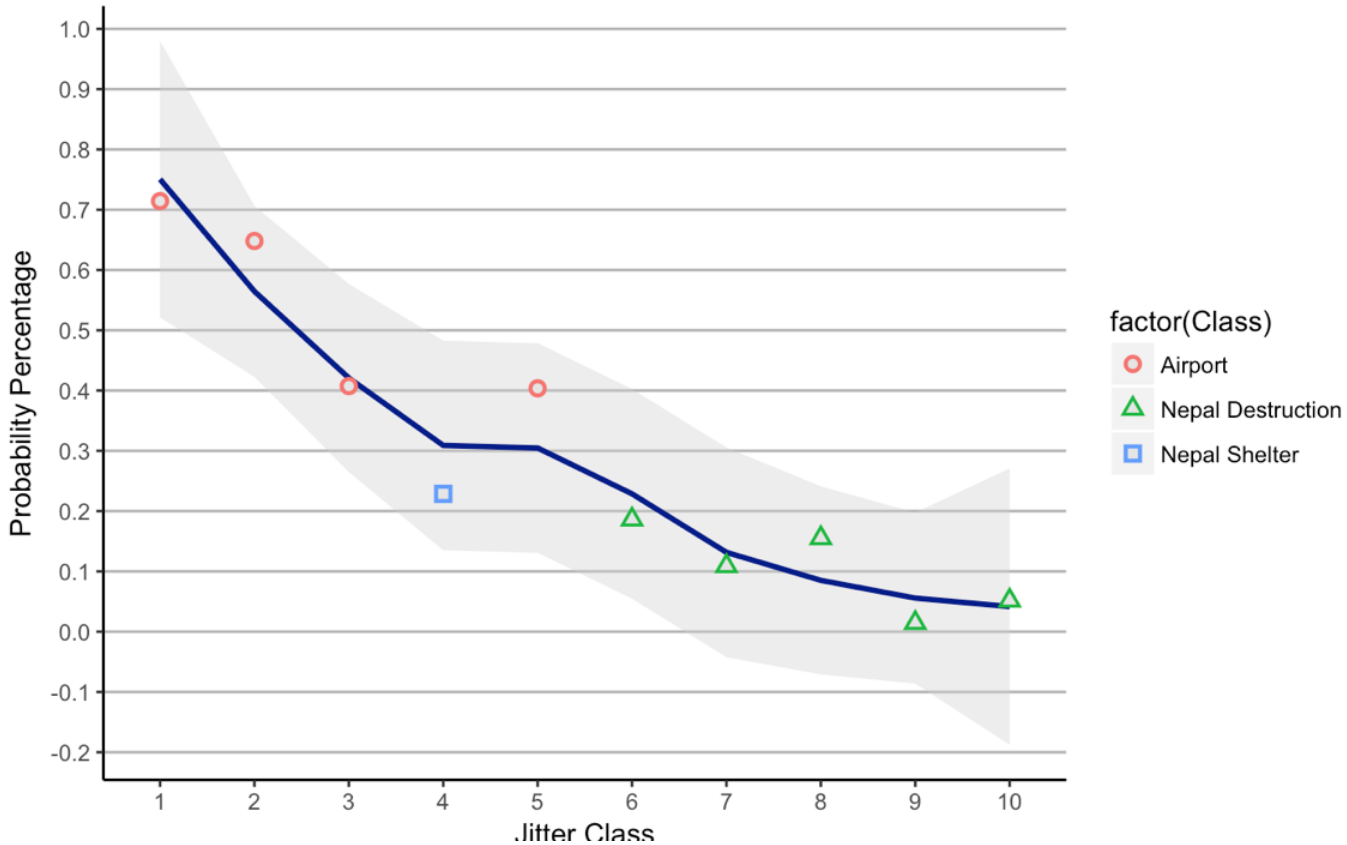


Figure 26. Jitter Curve for Jitter Variance Range C: Medium-High-End Uncorrelated Jitter(Step of 0.00001 radians)

Jitter Curve Discussion and Observations.

Observations of Jitter Curve C, indicated in Figure 26, is described. The characteristics of the curve were as expected, a pronounced drop was evident after the fourth Jitter Class image was applied. This drop in accuracy indicated that scrambling of the clean edges within the image caused the classifier severe confusion. Again, an interesting parallel is drawn where synthetic images applied jitter above 0.00024 micro radians were classified as images from the Nepal Earthquake Class. Of the three Nepal Earthquake Classes, Nepal Total Destruction denoted by (ND), possessed the highest degree of destruction in comparison to Nepal Damaged Buildings (NB) and Nepal Shelter (NS). The increase in uncorrelated jitter decreased features that would make the image recognizable as an airport image. Consequently, a trade-off between uncorrelated jitter and HybridCaffe's sensitivity emerged where increases of uncorrelated jitter increased HybridCaffe's sensitivity to images from the Nepal Earthquake dataset and decreases of uncorrelated jitter increased HybridCaffe's sensitivity to airport images. This trade-off is evident in Figure 26, as the increase in Jitter Class shows evidence of HybridCaffe's sensitivity to Nepal Earthquake images increasing.

Confidence Interval Interpretation.

Again, the wide confidence interval indicated in Figure 26, is typical for a small sample set. As the data points indicate there is a steady drop in sensitivity to airport images between Jitter Class 3 and Jitter Class 10 where the synthetic airport images are mistakenly classified as Nepal Destruction Class images. This is an example of a False Negative, where the true condition is not met even though the image is in fact a positive (i.e. airport).

Jitter Variance Range D.

Jitter Variance Range D constitutes of DIRSIG synthetic images jittered to a step of 0.00001 micro radians between probability percentages 0.9261 and 0.9304 where each step of jitter variation were between 0.00001 radians and 0.00010 radians. A sequential step increase of 0.00001 applied to each synthetic image produced a consistent curve in classification accuracy as indicated in Jitter Curve D. The upper confirmation threshold was calculated by taking the average of the probability percentages in Jitter Variance Range D. The threshold was calculated to be 0.91066.

Table 13. Jitter Variance Range D: Percentage Probabilities

Jitter Class	**Probability Percentage**
Jitter Class 1	0.9261
Jitter Class 2	0.9664
Jitter Class 3	0.9707
Jitter Class 4	0.8498
Jitter Class 5	0.9014
Jitter Class 6	0.9682
Jitter Class 7	0.8873
Jitter Class 8	0.9018
Jitter Class 9	0.8045
Jitter Class 10	0.9304

(a) 0.00001 radians (b) 0.00002 radians (c) 0.00003 radians (d) 0.00004 radians (e) 0.00005 radians

(f) 0.00006 radians (g) 0.00007 radians (h) 0.00008 radians (i) 0.00009 radians (j) 0.00010 radians

Figure 27. Jitter Variance Range D: Low Uncorrelated Jitter(Step of 0.000001 radians)

Jitter Curve Discussion and Observations.

Observations of Jitter Curve D, indicated in Figure 28, is described. The characteristics of the curve were as expected, an inconsistent positive curve was evident between all the Jitter Classes. The positive trend indicates HybridCaffe's robustness against low amounts of uncorrelated jitter. The characteristics of stochastic gradient descent were evident in the inconsistent trend indicated in the slow drop. This includes a small peak in Jitter Class six and a small drop indicated in Jitter Class eight. This behavior is evident in Figure 28, as the increase of low-end uncorrelated jitter appears to have minimal effect on the inhibition of HybridCaffe's sensitivity to airport images.

Confidence Interval Interpretation.

Again, the wide confidence interval indicated in Figure 28, is typical for a small sample set. The data points indicate there is a slow drop between Jitter Class 1 and

Jitter Class 10 where the synthetic airport images are correctly classified as Airport Class images. This is an example of a True Positive, where the true condition is met.

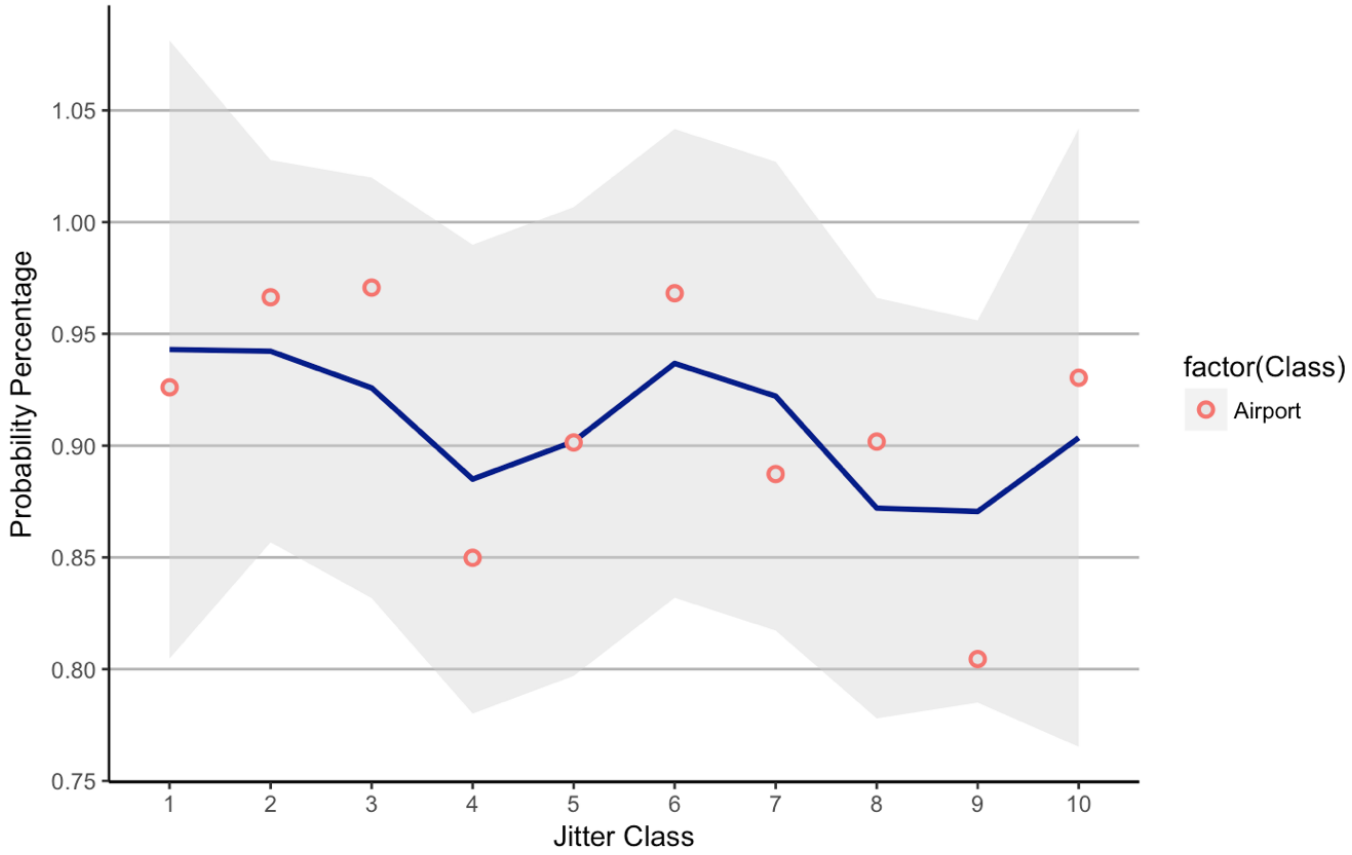


Figure 28. Jitter Curve for Jitter Variance Range D: Low Uncorrelated Jitter(Step of 0.000001 radians)

Jitter Variance Range Averages.

The average of each Jitter Variance Range was calculated and measured for data point variation in Figure 29. Observations drawn from the Jitter Variance Range Averages indicated in Figure 29, are described in the subsequent subsections.

Jitter Variance Range Threshold Interpretation.

The thresholds for each Jitter Variance Range were calculated by calculating the average of each Jitter Variance Range's set of values. Jitter Variance Range B had a confirmation threshold of 0.7753. This calculated value indicates that any classification which renders a probability percentage above 0.7753 is confirmed as a positive classification. Jitter Variance Range C had a failure threshold of 0.2918. This calculated value indicated that any classification which renders a probability percentage below 0.2918, is a failed classification. Jitter Variance Range D had an upper confirmation threshold of 0.9106. This calculated value indicated that any classification which renders a probability percentage above 0.9106 is considered an absolute positive classification.

General Interpretation.

The standard deviation error bars generated are based off of the variation of the spread in numerical value between the data points. Longer standard deviation error bars indicate a pronounced variation between the recorded predicted probability averages. Jitter Variance Range A possessed values with a pronounced range in data point variation where the highest values were between 0.90 and 0.71 and the rest of the values were fairly close to 0. Jitter Variance Range B possessed data values which were more closely linked together. Consequently, a shorter standard deviation error bar was evident. Jitter Variance Range C, possessed values that were closer to 0 but with a greater range of variance. This resulted in a longer standard deviation error bar. Jitter Variance Range D possessed values that were close together with a smaller variation in data values; this is reflected in the smaller error bar. In general, Jitter Variance Range D had the lowest amount of data point variation while Jitter Variance Range A possessed the highest amount of data point variation. This is attributed

to the jitter jitter variance range being higher in Jitter Variance Range A and being lower in Jitter Variance Range D. Conclusively, this confirms the fact that increase in uncorrelated jitter significantly degrades classification performance.

Interpretation of Standard Deviation Error Bar Overlap.

Conclusions drawn from interpretation of the overlap of the standard deviation error bars between each Jitter Variance Range's bar graph is described. It is a common practice in the Data Science field to view overlaps in standard deviation error bars as indicators of non-significance between the means of two populations of data. The standard deviation error bars of Jitter Variance Range A and Jitter Variance Range C appear to overlap. This indicates that the means of the two different populations are not statistically significant. However, the population means of Jitter Variance Ranges A,B and D are all statistically significant as their standard deviation error bars appear to not overlap. Consequently, this lines up with the low p-value of $2.53x10^9$, which indicates statistical significance between the 3 population means of Jitter Variance Ranges A, B and D.

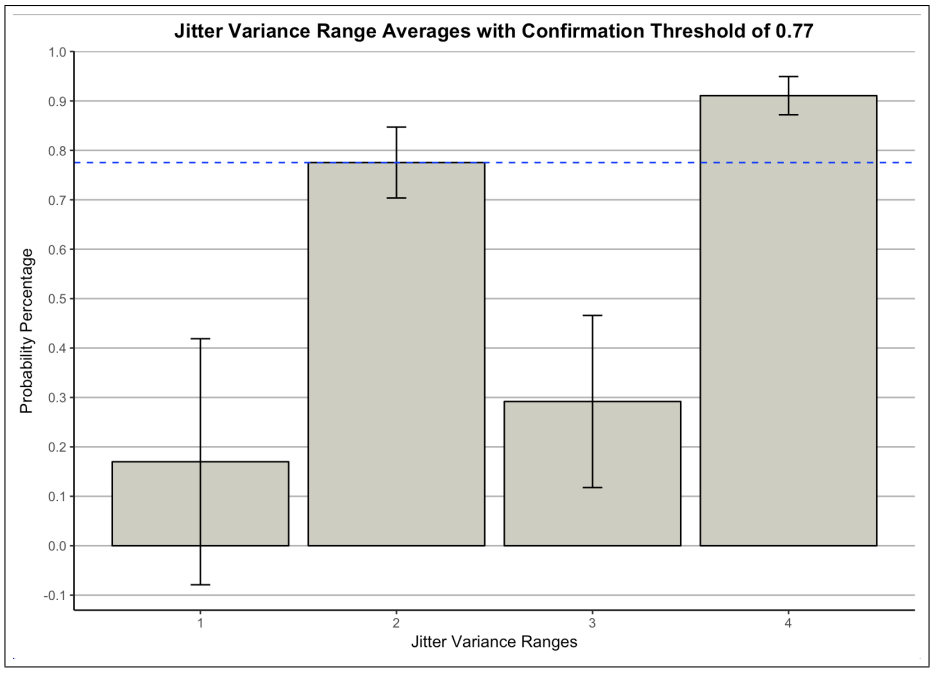


Figure 29. The average for Jitter Variance Range A through D is displayed, including standard deviation error bars indicating the general variation of the data point spread and the positive classification threshold of 0.77.

4.3 Hypothesis Test

A Hypothesis test was generated to check for the general statistical significance of the 4 sets of data generated from Jitter Variance Range A, Jitter Variance Range B, Jitter Variance Range C, and Jitter Variance Range D. The Null Hypothesis states that the confirmation threshold for HybridCaffe to execute a positive classification of the Hyper-spectral Airport image, indicated in Figure 29, is 0.80 percent. The Alternative Hypothesis states that the confirmation threshold for HybridCaffe to execute a positive classification of the Hyper-spectral Airport image, indicated in Figure 29, is 0.70 percent. Measurement of the statistical significance of the population means are drawn out from observations conducted in the previous subsections regarding the standard deviation error bar interpretations and through performance of Analysis of Variance by generating an ANOVA table in the RStudio suite. Conclusively, the results are strongly significant, as a p-value <0.05 indicates the results are statistically

significant. The p-value in Figure 30 also indicates that the Null Hypothesis is to be rejected. Therefore, the Alternative Hypothesis, the confirmation threshold for HybridCaffe to execute a positive classification of the Hyper-spectral Airport image, is accepted. This conclusion lines up with the experimental results recorded in this experiment set in Figure 29.

```
            Df Sum Sq Mean Sq F value   Pr(>F)
ind          3  3.913  1.3042   26.99 2.53e-09 ***
Residuals   36  1.740  0.0483
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 30. Analysis of Variance (ANOVA) Table produced in RStudio Suite

4.4 CNN Training Results Interpretation

Evaluation of Confusion Matrix.

This section describes results rendered from the confusion matrix. Overall, the raw counts for the classification of each class were high with an exception to the Nepal Earthquake Classes, Nepal Building (NB) and Nepal Destruction (ND). This is due to the similarities between the classes as indicated in Figure 31.

(a) Nepal Building Damage Class (b) Nepal Total Destruction Class

Figure 31. Selections from Tomnod Nepal Earthquake Dataset, courtesy TOMNOD

Both of the classes possessed similarities that were strong enough to cause the classifier to mistakenly classify one for the other. There are 23 instances of Nepal Damaged Building Class images classified as Nepal total Destruction. In addition, there are 12 instances of Nepal Damaged Building Class images classified as Nepal total Destruction. Both of these classes have the lowest counts for True Positive results in the confusion matrix. Nepal Earthquake Class Nepal Shelter possessed 4 counts of being classified as Nepal Damaged Buildings. The resultant points of confusion in the confusion matrix described in Figure 32, are as expected. This resulted in the Nepal Building class rendering low precision and recall values of 0.630 and 0.797 respectively.

DIRSIG Airport (DA) and DIRSIG Forest (DF) are the two categories with the highest number of True Positive counts. Both classes were not erroneously classified as other classes. This is presumably due to the synthetic features that makeup the images. When compared with classes in the set of scenes with realistic high resolution

satellite imagery characteristics there are few similarities that may cause confusion. This resulted in both classes rendering high precision and recall values of 1.

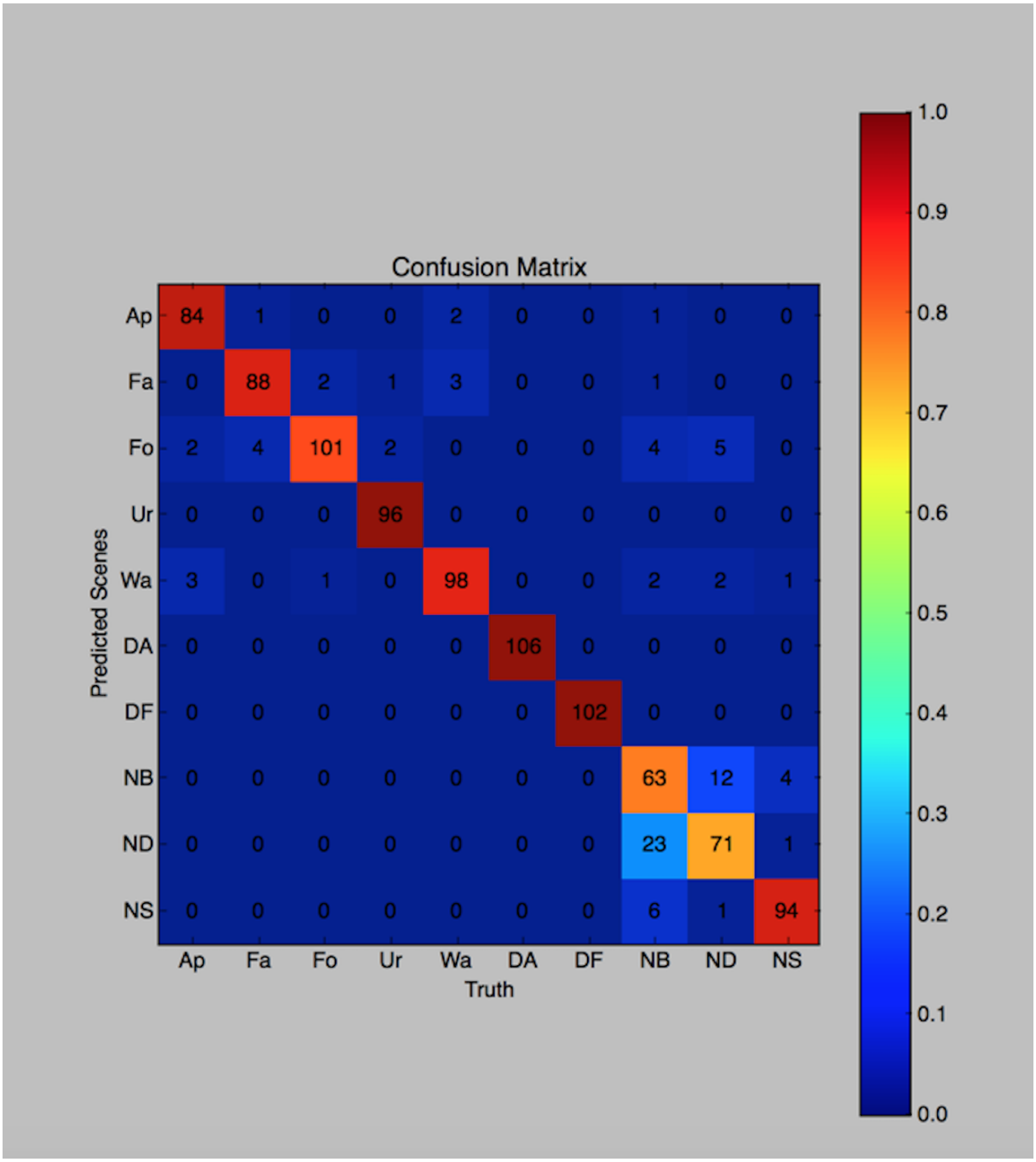


Figure 32. Confusion Matrix, where values on the diagonal line are True Positive (TP) classifications.

True Positive Results.

As described in the Methodology Chapter, the data in Figure 33 consist of Thresholds with the percentage of True Positive values associated with each threshold. The threshold indicates the decision point where images sent into HybridCaffe are confirmed as an airport based on their respective probability percentages. The probability percentages utilized in the calculation of the True Positive Percentage calculations are provided in Table 14. The values provided in Table 14 depict the Percent Probabilities produced by HybridCaffe when a single class of high resolution satellite imagery is applied, specifically, airport images.

Table 14. Raw Values Rendered from Initial Experiment

Instance Number	Probability Percentage	Instance Number	Probability Percentage	Instance Number	Probability Percentage
1	1.000	**17**	1.000	**33**	1.000
2	0.667	**18**	0.999	**34**	1.000
3	0.951	**19**	1.000	**35**	0.999
4	1.000	**20**	1.000	**36**	0.867
5	0.994	**21**	1.000	**37**	1.000
6	1.000	**22**	1.000	**38**	1.000
7	0.978	**23**	1.000	**39**	0.956
8	0.664	**24**	0.965	**40**	1.000
9	0.544	**25**	0.999	**41**	0.825
10	1.000	**26**	0.581	**42**	0.993
11	0.998	**27**	0.984	**43**	0.993
12	1.000	**28**	1.000	**44**	1.000
13	1.000	**29**	0.993	**45**	1.000
14	0.983	**30**	0.998	**46**	0.976
15	1.000	**31**	0.589	**47**	0.979
16	0.999	**32**	0.927	**48**	0.992

Table 14. Continuation

Instance Number	Probability Percentage
49	1.000
50	0.981
51	1.000
52	0.995
53	0.979
54	0.999
55	0.997
56	0.999
57	0.989
58	1.000
59	0.826
60	0.626
61	1.000
62	1.000
63	0.963
64	1.000

Instance Number	Probability Percentage
65	1.000
66	1.000
67	0.997
68	1.000
69	0.995
70	0.997
71	0.725
72	0.993
73	0.658
74	1.000
75	0.999
76	0.999
77	1.000
78	0.991
79	1.000
80	1.000

Instance Number	Probability Percentage
81	0.599
82	0.999
83	0.859
84	0.762
85	1.000
86	1.000
87	0.979
88	1.000

The True Positive Percentage (TP%) is the percent of images successfully classified as airports out of the total number of images that were successfully classified as airports. Threshold values between 0 and 1 are specified to determine the percentage of True Positive (TP) results. In the case of a threshold of 0, a total of 100% will be recorded, as all of the probability percentages recorded are greater than 0. A low threshold is not ideal for practical testing but is valuable for the small scale evaluation of this research effort. A True Positive vs Threshold graph illustrating this behavior, is provided in Figure 30.

Table 15. True Positive Percentages with correlated Thresholds

Threshold	TP Percentage
0.000	0.909
0.050	0.909
0.080	0.909
0.100	0.909
0.150	0.909
0.180	0.909
0.200	0.909
0.210	0.909
0.220	0.909
0.230	0.909
0.240	0.909
0.250	0.909
0.260	0.909
0.270	0.909
0.280	0.909
0.290	0.909

Threshold	TP Percentage
0.300	0.909
0.310	0.909
0.320	0.909
0.330	0.909
0.340	0.909
0.350	0.909
0.360	0.909
0.370	0.909
0.380	0.909
0.390	0.909
0.400	0.909
0.410	0.909
0.420	0.909
0.430	0.909
0.440	0.909
0.450	0.909

Threshold	TP Percentage
0.460	0.909
0.470	0.909
0.480	0.909
0.490	0.909
0.500	0.909
0.510	0.909
0.520	0.909
0.530	0.909
0.540	0.909
0.550	0.898
0.560	0.898
0.570	0.898
0.580	0.898
0.590	0.886
0.600	0.875
0.610	0.875

Table 15. Continuation

Threshold	TP Percentage
0.620	0.875
0.630	0.875
0.640	0.875
0.650	0.875
0.660	0.864
0.670	0.841
0.680	0.841
0.690	0.841
0.700	0.841
0.710	0.841
0.720	0.841
0.730	0.830
0.740	0.830
0.750	0.830
0.760	0.830
0.770	0.818

Threshold	TP Percentage
0.780	0.818
0.790	0.818
0.800	0.818
0.810	0.818
0.820	0.818
0.830	0.807
0.840	0.807
0.850	0.807
0.860	0.795
0.870	0.784
0.880	0.784
0.890	0.784
0.900	0.784
0.910	0.784
0.920	0.784
0.930	0.784

Threshold	TP Percentage
0.940	0.784
0.950	0.784
0.960	0.761
0.970	0.750
0.980	0.705
0.990	0.659
1.000	0.227

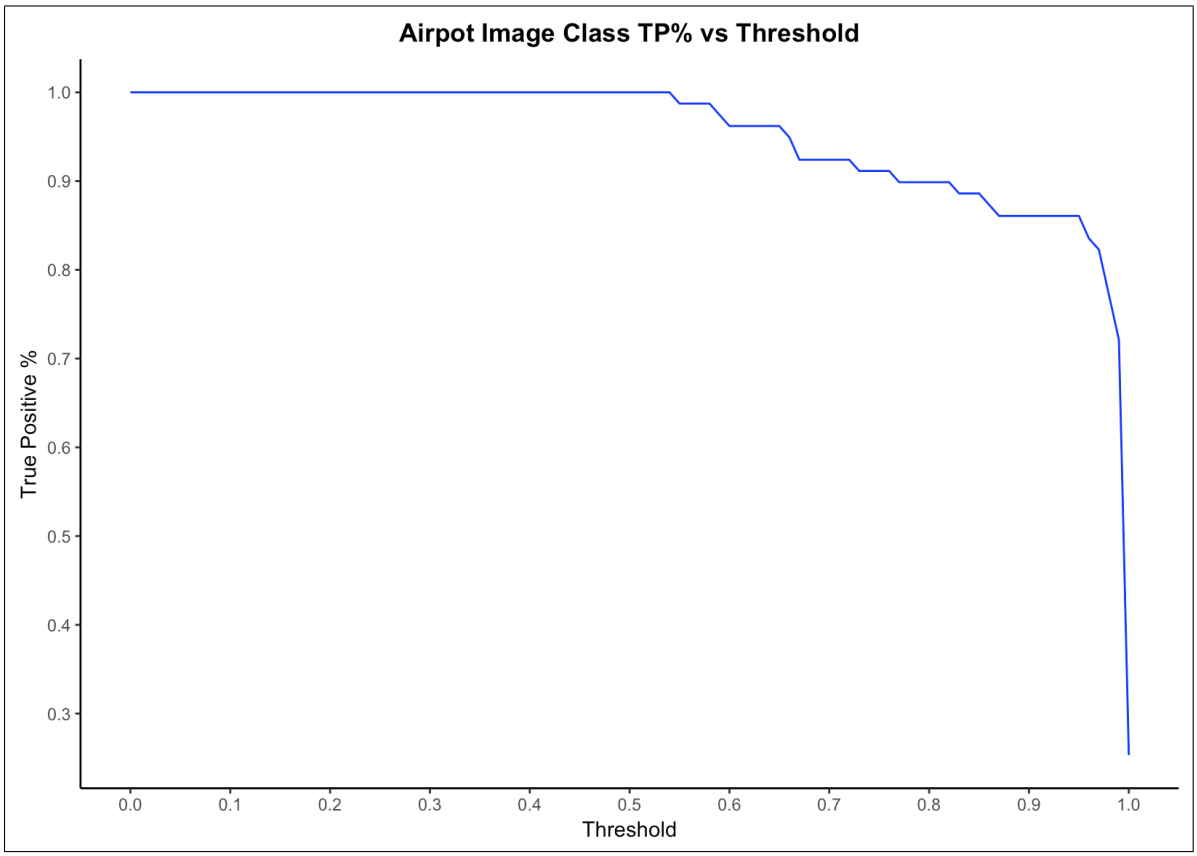


Figure 33. True Positive Percentage vs Threshold

A total of 88 airport images were sent into HybridCaffe for classification. A total of 79 images were correctly classified as airport images. There were 9 images that were mistakenly classified as Farm, Urban, Water Class images. These unsuccessful Airport Class classifications were categorized as False Negatives (FN) where the classifier failed to correctly classify an image according to the established true condition. In the case of the experiment, the true condition for a correct classification was classification of an Airport Class image, as all the images sent into HybridCaffe were Airport Class images.

The True Positive Percentage curve is useful in indicating how well HybridCaffe is in it's ability to sense high resolution satellite imagery of airports with differing spatial scale qualities. Again, the data indicated in Table 15 shows only True Positive values. A consideration for future work would entail calculation of a False Positive rate to determine how sensitive HybridCaffe is to high resolution satellite imagery of non-airport classes with differing spatial scale qualities. The combination of both the True Positive results and False Positive results have the potential of being combined together as a ROC curve as discussed in the Future Work Chapter.

Relationship between Training Accuracy and Loss.

Observations drawn from the Training vs Loss Curve, indicated in Figure 34 are described. The loss represents any misses rendered by the network during the training process. The testing accuracy represents any hits rendered by the network during the training process. An illustration of the inverse relationship between Training Accuracy and Training Loss is evident in Figure 34. Specifically, as the number of training iterations increased the testing accuracy of the network increased. HybridCaffe was programmed to test the network's accuracy every 10 iterations.

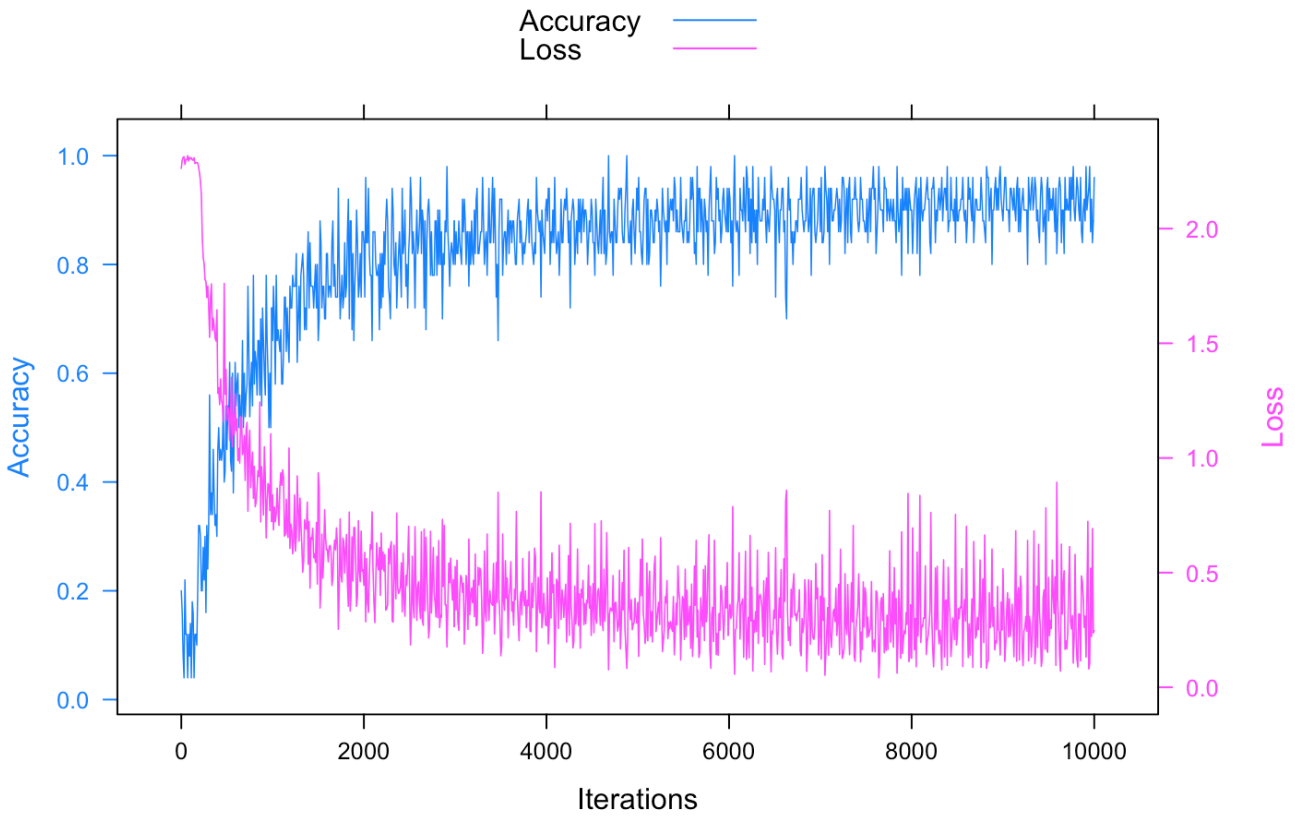


Figure 34. Training vs Loss Curve

4.5 Training Berkeley Caffe

In an effort to maintain a sense of realism, HybridCaffe, was trained on a non-homogeneous spread of training data. Where the number of testing and training images for each category were randomly allocated. Specifically, there was a total of 9,013 images in the training set and a total of 987 images in the testing set. This makes for a ratio of 90% of the total number of images assorted to training and 10% of the total number of images assorted to testing. The uneven numbers are attributed to the random allocation of images into the training and testing folders. Refer to Table 16 for the exact number of testing and training images per category.

Table 16. HybridCaffe Training and Testing Statistics Summary, where images in Train Set accounts for 90% and Test set accounts for 10% of the total amount of imagery.

Category	Train Set Count	Test Set Count
Airport	911	89
Farm	907	93
Forest	896	104
Water	901	99
Urban	897	103
DIRSIG Airport	894	106
DIRSIG Forest	898	102
Nepal Destruction	900	100
Nepal Building	909	91
Nepal Shelter	900	100

V. Future Work

This chapter describes multiple avenues of future extension to this research effort. The concept of the tip and cue process between three satellites in three unique orbits is described in the Tipping and Cueing Scheme section. The continued expansion of small scale satellite development from smart phones[17], raspberry pi[18] and development of Berkeley Caffe on raspberry pi [19] hardware indicates the potential of an extension of Berkeley Caffe into space. Details of this proposition is described in the Expansion to Rasberry Pi section. The expansion of Berkeley Caffe into the International Space Station to accommodate the limited memory capacity on board satellites in space and cyber-security measures are described in the Expansion to ISS section. The processing mechanisms in Convolutional Neural Networks are biologically inspired, the potential in modeling a traumatized neuron is described in the Post Traumatic Stress Disorder Study through Application of Convolutional Neural Networks section. Finally, potential for further evaluation of the true robustness of HybridCaffe is described in the robustness evaluation section.

5.1 Tipping and Cueing Scheme

The framework is divided into sections called "Resolution Tier Levels" intended to correlate with the Geostationary Orbits (GEO) and Low Earth(LEO) of each satellite in the framework. Each Tier Level represents the specific resolution quality of an image that has been sensed by a satellite within that respective orbit. For instance, an image with a pixel resolution of approximately 1 meter would be sensed by a satellite in a LEO orbit. This is due to the fact that LEO satellites reside at a close distance to the Earth's upper atmosphere.

Orbit Specifications.

Each satellite in this domain is at the mercy of the physics of space and their unique spacecraft design. Specifications in the functionality of such satellites will be described. Satellites in Geostationary Orbit have a mean orbit altitude of 35,790 kilometers above the Earth's surface. It is typical for imagery sensed by satellites at this orbit to have a pixel resolution from 1 to 30 kilometers. This produces imagery that is of a grainy quality where finer details required for surface target detection can be abstracted within the immense scale produced. Satellites in Low Earth Orbit have an orbit altitude from 200 to 1200 km above the Earth's surface. It is common for imagery sensed by satellites within this orbit range to have a pixel resolution from 0.5 to 1 meter. This produces imagery of a much cleaner quality where small details such as roads and cars are evident. Nevertheless, a more sensitive sensor would be required to extract finer features from surface targets (i.e. license plate numbers, runway strip names). A LEO satellite with a tighter orbit at 200 km above the Earth's surface is capable of sensing the surface for the tiny details required of a LEO satellite to sense but is otherwise abstracted. Attributing the pixel resolution of a remotely sensed image to the orbit from which the satellite resides served as a basis in the design of the tier framework described in Table 16.

Table 17. Tier Framework Summary, where GEO, LEO-inner, LEO-outer orbits are specified with their respective ranges.

Tier Level	Tier No.	Orbit Level	Resolution Range (in meters)
OUTER	1	Geostationary Orbit	1000
INNER	2	Low Earth Orbit	1
DETAIL	3	Low Earth Orbit	0.305

Inspiration for Tipping and Cueing Algorithm.

The organization scheme described in Table 16 also served as inspiration for the proposed tipping and cueing scheme. The requirements binding the tipping and cueing exchange between selected satellites in a constellation is described, under the use case of the target image being an airport. The description follows a chronological sequence moving from the outermost orbit satellite to the innermost orbit satellite. Exchanging information between satellites of different tiers will be in pairs. Tipping of a satellite from one tier to the other only involves two satellites, completing a one-to-one mapping of one satellite communicating with the subsequent one. Overlaps of the one-to-one mapping is illegal to the bounds described in the requirements outlined for the tipping and cueing scheme described in Tipping and Cueing Algorithm depicted in Figure 4.

Tipping and Cueing Sequence.

The tipping and cueing scheme is described in sequence, starting with Tier 1: Outer Tier, followed by Tier 2: Inner Tier and concluded by Tier 3: Detail Tier. A geostationary satellite sends a tip to a low earth orbit satellite. A geostationary satellite receives a set of coordinates from various airports from the Head Quarters ground station. Once the geostationary satellite is able to successfully sense and collect images with the specific coordinates, it sends them to it's respective Convolutional Neural Network (CNN). Classifications conducted by the CNN is under the assumption that the CNN has been trained on the target image of interest, in this case, an airport. The imagery collected by the GEO satellite possess a pixel resolution of about 30 kilometers. When the classifier reaches a probability percentage greater than or equal to 0.80, a tip request is sent to the successor. A satellite in a LEO orbit of 1200 km receives the tip request and responds to the GEO satellite with a

confirmation of receiving the tip request. Once the confirmation has been received the GEO satellite sends a tip information request to the LEO satellite. The LEO satellite responds to the GEO satellite with a confirmation of receiving the tip info request. Once the confirmation has been received the GEO satellite sends the tip information to the LEO satellite. After the LEO satellite successfully processed the tip information, a tip information confirmation is sent to the GEO satellite.

Advantages of Tip and Cue Scheme.

As stated previously, LEO satellites reside at an orbit fairly close to the Earth's surface. This causes their orbital velocity to be at a high value that creates a noticeable deterioration in it's capability to remotely sense the Earth's surface with absolute precision. The sun is a source of radiance required to successfully remotely sense the Earth's surface to conduct a collection. It is at this point in the sequence of the tipping and cueing scheme that the LEO satellite has information for the exact location to sense. Nevertheless, it requires information detailing when to remotely sense the surface and collect high pixel resolution quality imagery. Otherwise, precious resources may be wasted, as the LEO satellite is at high risk of inadvertently sensing the target area at an inopportune time (i.e the sun is on the opposing side of the earth's surface).

Initiation of Tip and Cue Exchange.

A tip detailing where to sense followed by a cue describing when to sense is highly valuable in this domain. In continuation of the airport use case, the GEO satellite sends a cue request to the LEO satellite. The LEO satellite receives the cue request and responds to the GEO satellite with a confirmation of receiving the cue request. Once the confirmation has been received the GEO satellite sends a cue

information request to the LEO satellite. The LEO satellite responds to the GEO satellite with a confirmation of receiving the cue info request. Once the confirmation has been received the GEO satellite sends the cue information to the LEO satellite. After the LEO satellite successfully processed the cue information, a cue information confirmation is sent to the GEO satellite.

LEO Outer Orbit Data Collection.

Then, the LEO satellite remotely senses and collects the appropriate amount of images for it's CNN to train on. The imagery collected are airport images with a pixel resolution of about 1 kilometer. Once the LEO satellite's CNN reaches a probability percentage greater than or equal to 0.80, it sends a tip to a satellite in a tighter LEO orbit at 200 kilometers above the Earth's surface.

The successor satellite receives the tip request and responds to the LEO satellite with a confirmation of receiving the tip request. Once the confirmation has been received the LEO satellite sends a tip information request to the successor satellite. The successor satellite responds to the LEO satellite with a confirmation of receiving the tip information request. Once the confirmation has been received the LEO satellite sends the tip information to the successor satellite. After the successor satellite successfully processed the tip information, a tip information confirmation is sent to the LEO satellite.

LEO Outer Orbit to LEO Inner Orbit Exchange.

Once the LEO satellite receives this information, it sends a cue request to the successor satellite. The successor satellite receives the cue request and responds to the LEO satellite with a confirmation of receiving the cue request. Once the confirmation has been received the LEO satellite sends a cue information request

to the LEO satellite. The successor satellite responds to the LEO satellite with a confirmation of receiving the cue info request. Once the confirmation has been received the LEO satellite sends the cue information to the successor satellite. After the successor satellite successfully processed the cue information, a cue information confirmation is sent to the LEO satellite. Then, the successor satellite remotely senses and collects the appropriate amount of images for it's CNN to train on. The imagery collected are airport images with a pixel resolution of about 0.305 meters. Once the LEO satellite's CNN reaches a probability percentage greater than or equal 0.80, it sends a tip to Head Quarters.

5.2 Expansion to Raspberry Pi

Description.

Work has been conducted in developing "smart phone satellites", where the guts of the satellite runs off of the main operating system of a smart phone. An example of this is work done on the STRAND-1 satellite, where the satellite exterior houses the internal guts of the android smart phone[17]. Such work has inspired other researchers to conduct similar research on creating the internals of a cubesat out of a Rhaspberry Pi's embedded system[18]

Potential of Berkeley Caffe in Space.

Furthermore, development has been done on creating a version of Berkeley Caffe that is capable of fully operating on a Rhaspberry Pi[19]. Finally, it has been proven that Rhaspberry Pi's are completely operational in the vacuum of space[18]. This presents an interesting extension for the proposed framework of this thesis to be applied in the embedded systems domain with another possible future extension to application in the space domain. This would make a rather interesting thesis topic

as it utilizes multiple domains at once similar to the nature of this thesis only a bit more specific. Specific in regards to the way the materials are orchestrated. The evaluation domain would be confined to the testing of Berkeley Caffe's performance on the Rhaspberry Pi and comparing the embedded classifier's performance to a Caffe model on a laptop with the CPU option specified and to a Caffe model on a GPU computer with the GPU option specified. Further analysis of the results would be fair indicators of how feasible sending a satellite equipped with a classifier into space would be.

5.3 Moving Tiers to Caffe Networks

In the event that more time was made available to the author, multiple Caffe models would be created. Where a single Caffe model would be attributed to each Tier Level in Table 1, creating an ensemble of Caffe models.

Association between Caffe Models and Tier Levels.

The association between each Tier Level and their respective Caffe model will be described. Tier 1: Outer Tier would be represented by a single Caffe model trained only on remotely sensed images with a pixel resolution of approximately 30 km. Tier 2: Inner Tier would be represented by a single Caffe model trained only on remotely sensed images with a pixel resolution of approximately 1 meter. Tier 3: Detail Tier would be represented by a single Caffe model trained only on remotely sensed images with a pixel resolution of approximately 12 centimeters. The evaluation between the three Caffe models would generate conclusions as to how these Caffe models can be interlinked together to "tip and cue" each other based off of imagery with similarities in certain characteristics such as target size or shape. It would be interesting also to see if there is way for a Caffe model to process GPS coordinates and find a way to map

the GPS coordinates to a certain image. This can also serve as another characteristic to perform the tip and cue scheme from.

5.4 Expansion to the International Space Station

The description in this section is based solely off of the assumption that Berkeley Caffe, has been successfully deployed in space. Additionally, this section is an extension of the tipping and cueing framework scheme proposed in the Tipping and Cueing Scheme section. Furthermore, majority of the content in this section is purely speculation.

ISS hosts Virtual Machine Association between Caffe models.

In an effort to accommodate the limited memory capacity on board satellites in space, the satellites in constellation will each possess their own respective Caffe model which will run off of a distributed system, where each node represents a single instance of Caffe. The Caffe distribution will run off of a CPU based on the Earth's surface. The International Space Station(ISS) will house a single laptop that will have a series of Virtual Machines(VM's). Where each VM will represent a single node from the Caffe distribution. Astronauts on mission to conduct experiments on the ISS will be able to view the VM for each Caffe model attributed to each satellite in the constellation. This is ideal as the ISS is in a LEO orbit, and majority of the satellites in the constellation of the framework proposed will be in the LEO orbit. Additionally, this shared orbit ensures quick and efficient imagery processing, as the ISS may serve as a source of computational power in the constellation.

Proposed Cyber-Security Measures.

The VM's can be guarded against cyber-intrusions through appropriate utilization of firewall protocols in the Caffe distribution. It is also expected that each VM will be monitored by a human operator to ensure there are no cyber-intrusions. A future iteration of this effort will have an autonomous mechanism in place where cyber-intrusions will set off an alarm alerting human operators. The autonomous mechanism is intended to fulfill the original intention of this proposed tipping and cueing framework, removing the human from the loop and allowing the machine to take full control and reign of the system.

5.5 Post Traumatic Stress Disorder Study through Application of Convolutional Neural Networks

The description in this section is based solely off of the assumption that ensembles of Berkeley Caffe, have been successfully utilized in modeling clusters of neurons in the human brain. Additionally, this section is an extension of the tipping and cueing framework scheme proposed in the Tipping and Cueing Scheme section. Where, a tip and cue occurs between neurons, in the place of satellites. Furthermore, majority of the content in this section is based purely on speculation.

Modeling a Traumatized Neuron.

A single CNN can represent a traumatized neuron inside the brain of a PTSD survivor. The image content of the traumatized neuron would be filled with screenshots taken from violent action films such as Saving Private Ryan and Enemy at the Gates. Screenshots can be taken using a python script reading in movie film through the SciPy software suite.

Modeling a Non-Traumatized Neuron.

Another single CNN can represent daily images in the household. The researcher can wear a Gopro in and around the house for approximately 1 day. All the motion imagery will be read into the SciPy software that will record screenshots. The researcher will wear the Gopro to a shooting range, or during a hunting excursion with friends and/or family. These action-oriented images will be randomly shuffled with the training images of the daily activities neuron. A third party CNN can have a hybrid dataset of the two neurons and serve as a bridge that will complete the connection between the two neurons.

Speculation for the Cause of Flashbacks of Traumatic Events in a Survivor.

It is hypothesized that when a PTSD survivor experiences a flashback from a traumatic event, the traumatic neuron is triggered by a third party neuron which causes the traumatic neuron to fire to a neuron filled with daily household images. When the traumatic neuron fires to a neuron filled with daily images it fills the household neuron with traumatic images. Conversely, it is possible that the trigger neuron will activate a household neuron with a hybrid mixture of traumatic images and household images, where the household neuron refuses to train on household images and will only train on traumatic images.

5.6 Further Evaluations

Potential for expansion of the True Positive Percentage curve in Figure 33 into a Receiver Operator Curve (ROC) is described. Where the ROC curve would indicate the true classification capacity of HybridCaffe, when faced with the challenge of proper classification of Airport Class images. Further exploration of the robustness

of HybridCaffe can be conducted by sending images of areas impacted by National Weather Scenarios, such as images obtained from NOAA's Emergency Response Imagery National Geodetic Survey. Selections from [8] for robustness evaluation are indicated in the Figure 35.

Figure 35. Images from Tornado Joplin, courtesy NOAA [8]

VI. Conclusions

Conclusions drawn during the investigation of the potential in utilization of Machine Learning in application to the space vehicle domain where a heterogeneous set of varied fidelity sensor types in a constellation of small satellites enhances accurate surface target detection in the case of an Emergency Weather Scenario is presented. Application of the remote sensing platform, DIRSIG demonstrated remote sensing techniques crucial in the investigation conducted in the research effort. Features of interest were generated from specified degrees of clarity to aid in the generation of a database utilized to test and examine the classification accuracy of the artificial intelligence platform Berkeley Caffe. The artificial intelligence platform was adapted into a single artificial neural network called HybridCaffe to model a heterogeneous set of varied fidelity sensor types. Variation in fidelity was demonstrated through the collection of satellite imagery with varying jitter levels. The data utilized emulates the basic terrain in Emergency Weather scenarios.

6.1 General Conclusion

Through the experimental design, results collected and conclusions drawn, Machine Learning demonstrates its suitability in the space vehicle domain as a source of automation for the tipping and cueing between small satellites in a constellation to increase surface target detection effectiveness. In spite of the limitations faced through image paucity the Machine Learning platform, Berkeley Caffe exhibited it's capability through the platform's modular design. The Tier Level design implemented in the Heterogeneous Sensor Framework exhibited the basic concept of three different satellites, in three unique orbits, in a constellation tipping and cueing each other according to the description in the Contribution Section of the Introduction Chapter.

6.2 Conclusions to Research Questions posed in Introduction

The leveraging of the automation of terrain classification for a tipping and cueing system across a constellation of satellites with a mixture of heterogeneous sensor types was realized in the building of the Tip and Cue Exchange Protocol, Tier Organization and Image Classifier described in the Contribution section of the Introduction. Once a Tipping and Cueing system was designed two research questions were investigated in parallel to determine the suitability a Machine Learning platform has in the space vehicle domain. A Convolutional Neural Network was designed with the intent of classifying terrain common in search and rescue events. Automation of terrain classification with Machine Learning was realized in the creation of the Convolutional Neural Network, HiReSatCaffe where terrain imagery was successfully classified. After a Convolutional Neural Network was successfully developed and tested synthetic imagery was incorporated to model and test satellite sensor capabilities. The incorporation of synthetic imagery into a Convolutional Neural Network paired with positive experimental results has successfully modeled satellite sensor capabilities. Overall, through the investigation of the 3 research questions the overall conclusion is realized that Machine Learning platforms can facilitate the tipping and cueing between heterogeneous constellation members in order to increase its overall feature identification effectiveness.

6.3 Research Effort Contributions

As stated in the Introduction Chapter, the main contributions for this research effort are four-fold: (1) Generation of a dataset which includes synthetic imagery, (2) Schematic of a Tip and Cue communication Protocol, (3) A satellite imagery classifier with 92.1% test set accuracy, and (4)Successfully modeled uncorrelated jitter impact on classification performance. The contributions are significant as they provide a

glimpse into the future development of optimizing current Air Force satellite systems with application of a communication protocol. In general, this indicates that a Convolutional Neural Network is able to classify images of different pixel resolutions which represent imagery produced by a constellation of satellites with a heterogeneous mixture of sensors. This proposed framework can be applied to problem domains outside of the search and rescue domain, such as the intelligence field where surface targets of interest can be detected and identified.

6.4 Significance of Methodology

Varying the amount of uncorrelated jitter between different intensity ranges demonstrates Berkeley Caffe's capability to learn a class of imagery and commit to correct classifications. The appropriate failure and confirmation thresholds indicate Berkeley Caffe's capability for strong classification power in spite of increase in light amounts of uncorrelated jitter. Berkeley Caffe has been utilized in many domains for image classification. The continued development of small satellites in constellation demonstrates the capability for satellite's embedded circuitry to host a Berkeley Caffe classifier. The fact that Berkeley Caffe is capable of learning multiple classes of images and is able to classify a high rate of True Positive (TP) results indicates potential for Berkeley Caffe has to be extended into additional research ventures described in the Future Work chapter.

Bibliography

1. Shawn Yang, Yi Newsam, "Bag-of-visual-words and spatial extensions for land-use classification", in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '10*, 2010.

2. J.R Schott, S.D. Brown, R.V. Raqueno, H.N. Gross, and G. Robinson, "An advanced syhtnetic image generation ˜ model and its application to multi/hyperspectral algorithm development", 1999, vol. 15, pp. 99–111.

3. Caitlyn Milton, "Nepal earthquake data portal", `http://blog.tomnod.com/Nepal-Earthquake-Data-Portal`, Accessed: 2016-12-30.

4. NOAA, "NOAA-Search and Rescue Satellite Aided Tracking", `http://www.sarsat.noaa.gov/`, Accessed: 2016-8-30.

5. Robert Mash, Nicholas Becherer, Brian Woolley, and John Pecarina, "Going deeper with convolutions", in *IEEE National Avionics & Electronics Conference*, 2016, pp. 1–9.

6. Andrej Karpathy, "CS231n Convolutional Neural Networks for Visual Recognition", `http://cs231n.github.io/convolutional-networks/`, Accessed: 2017-1-30.

7. Ujjwal Karn, "The Data Science Blog: Machine Learning, Deep Learning, NLP, Data Science", `https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/`, Accessed: 2017-1-30.

8. NOAA, "Joplin, MO Tornado Response Imagery", `http://storms.ngs.noaa.gov/storms/joplin`, Accessed: 2017-1-30.

9. Yangqing Jia, Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick, Sergio Guadarrama, and Trevor Darrell, "Caffe: Convolutional architecture for fast feature embedding", *arXiv preprint arXiv:1408.5093*, 2014.

10. Lloyd Wood, "Satellite constellation networks", 2003.

11. Jeremy Crampton, "Open Geography: GEOINT 2012 reflections on the Intelligence", `https://opengeography.wordpress.com/2012/10/17/geoint-2012-reflections-on-intel/`, Accessed: 2016-12-30.

12. Cisco, "The zettabyte era: Trends and analysis", White Paper, June 2016.

13. Christian M. Lewis, David Messinger, and Michael G. Gartley, "Activity-based intelligence tipping and cueing using polarimetric sensors", in *Proc. SPIE*, 2014, vol. 9099, pp. 90990C–90990C–13.

14. Hadi H. Jaafar and Farah A. Ahmad, "Crop yield prediction from remotely sensed vegetation indices and primary productivity in arid and semi-arid lands", *International Journal of Remote Sensing*, vol. 36, no. 18, pp. 4570–4589, 2015.

15. Praveen Rao Teleti and Alvarinho J. Luis, "Sea ice observations in polar regions: Evolution of technologies in remote sensing", *International Journal of Geosciences*, pp. 1031–1050, 2013.

16. Peng Zhang, Xin Niu, Yong Dou, and Fei Xia, "Airport detection from remote sensing images using transferable convolutional neural networks", in *International Joint Conference on Neural Networks(IJCNN)*, 2016, pp. 2590–2595.

17. Shaun Kenyon, CP Bridges, Doug Liddle, R Dyer, James Parsons, David Feltham, Rupert Taylor, Dale Mellor, Andrew Schofield, and Rosie Linehan, "Strand-1: Use of a $500 smartphone as the central avionics of a nanosatellite", in *Proceedings of the 62nd International Astronautical Congress.*, 2011, pp. 1–19.

18. David Honess, "Astro Pi: Mission Update 6 - Payload Handover - Raspberry Pi", `https://www.raspberrypi.org/blog/astro-pi-mission-update-6-payload-handover/`, Accessed: 2016-12-30.

19. Yufei Chen, "GitHub:Caffe-RPI", `https://github.com/benjibc/caffe-rpi`, Accessed: 2016-12-30.

20. Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk, and Darko Stefanovic, "Imagenet classification with deep convolutional neural networks", *Advances in Neural Information Processing Systems 25 (NIPS2012)*, pp. 1–9, 2012.

21. N. H. Crisp, K. Smith, and P. Hollingsworth, "Launch and deployment of distributed small satellite systems", *Acta Astronautica*, vol. 114, pp. 65–78, 2015.

22. Sukumar Ghosh, *Distributed Systems: An Algorithmic Approach, 2nd ed.*, CRC Press: Taylor and Francis Group, 6000 Broken Sound Parkway NW, Suite 300 Boca Raton, FL, 2008.

23. John R London, Mark E Ray, David J Weeks, and A Brent Marley, "The first us army satellite in fifty years : Smdc-one first flight results", in *25th Annual AIAA/USU Conference on Small Satellites*, 2011, pp. 1–5.

24. D.H. Hubel and T.N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex", *The Journal of Physiology*, vol. 160, pp. 106–154, 1962.

25. Kunihiko Fukushima, "Biological cybernetics neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position", *Biol. Cybernetics*, vol. 36, 1980.

26. John J Hopfield, "Neural networks and physical systems with emergent collective computational abilities.", in *Proceedings of the National Academy of Sciences of the United States of America*, 1982, vol. 79, pp. 2554–2558.

27. Geoffrey E. Rumelhart, David E. Hinton and Ronald J. Williams, "Learning representations by back-propagating errors", *Nature*, vol. 323, pp. 533–536, 1986.

28. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov 1998.

29. Srdjan Sladojevic, Marko Arsenovic, Andras Anderla, Dubravko Culibrk, and Darko Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification", *Computational Intelligence and Neuroscience*, vol. 2016, 2016.

30. Alessandro Ferrari, Stefano Lombardi, and Alberto Signoroni, "Bacterial colony counting with convolutional neural networks in digital microbiology imaging", *Pattern Recognition*, vol. 61, pp. 629–640, 2017.

31. Saikat Basu, Sangram Ganguly, Supratik Mukhopadhyay, Robert DiBiano, Manohar Karki, and Ramakrishna Nemani, "Deepsat - a learning framework for satellite imagery", *ACM SIGSPATIAL GIS*, vol. 32, pp. 1–22, 2015.

32. Yanfei Zhong, Feng Fei, Yanfei Liu, Bei Zhao, and Hongzan Jiao, "Satcnn: satellite image dataset classification using agile convolutional neural networks", *Remote Sensing*, 2016.

33. Marco Castelluccio, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva, "Land use classification in remote sensing images by convolutional neural networks", *DIETI*, 2015.

34. Volodymyr Hinton, Mnih Geoffrey E, Volodymyr Mnih, and Geoffrey E Hinton, "Learning to detect roads in high-resolution aerial images", in *11th European Conference on Computer Vision (ECCV)*, 2010, vol. 6316, pp. 210–223.

35. Saikat Basu, Sangram Ganguly, Ramakrishna R Nemani, Supratik Mukhopadhyay, Gong Zhang, Cristina Milesi, Petr Votava, Ralph Dubayah, Laura Duncanson, Bruce Cook, Yifan Yu, Sassan Saatchi, Robert Dibiano, Manohar Karki, Edward Boyda, Uttam Kumar, and Shuang Li, "A semi-automated probabilistic framework for tree cover delineation from 1-m naip imagery using a high performance computing architecture", *Journal of Latex Class Files*, vol. 11, pp. 1–20, 2012.

36. M Papadomanolaki, M Vakalopoulou, S Zagoruyko, and K Karantzalos, "Benchmarking deep learning frameworks for the classification of very high resolution satellite multispectral data", *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. III-7, pp. 83–88, 2016.

37. Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell, "Decaf: A deep convolutional activation feature for generic visual recognition", in *31st International Conference on Machine Learning*, 2014, vol. 32, p. 647–655.

38. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C Berg, and Li Fei-Fei, "Imagenet large scale visual recognition challenge", *International Journal of Computer Vision*, vol. 115, pp. 211–252, 2015.

39. Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition", in *International Conference on Learning Representations (ICLR)*, 2015, pp. 1–14.

40. C. Szegedy, Wei Liu, Yangqing Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions", in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 1–9.

41. Shunta Saito and Yoshimitsu Aoki, "Building and road detection from large aerial imagery", in *Proceedings of SPIE - The International Society for Optical Engineering*, 2015.

42. Shefali Aggarwal, "Remote sensing for natural or man-made disasters and environmental changes", *Satellite Remote Sensing and GIS Applications in Agricultural Meteorology*, vol. 3221, pp. 23–28, 2004.

43. Emmett J. Ientilucci and Scott D. Brown, "Advances in wide area hyperspectral image simulation", in *Proc: SPIE*, 2003, vol. 5075, pp. 110–121.

44. Albert Yu-Min Lin, Andrew Huynh, Gert Lanckriet, and Luke Barrington, "Crowdsourcing the unknown: The satellite search for genghis khan", *PLOS ONE*, vol. 9, no. 12, pp. 1–17, 12 2015.

45. Andrew Baruch, Avinoam May and Dapeng Yu, "The motivations, enablers and barriers for voluntary participation in an online crowdsourcing platform", *Computers in Human Behavior*, vol. 64, pp. 923 – 931, 2016.

46. Secretariat of the International Cospas-Sarsat Programme, "Cospas-sarsat system data", vol. 42, December 2016.

47. E. M. Fischer and R. Knutti, "Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes", *Nature Clim. Change*, vol. 5, 2015.

48. Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic, "Learning and transferring mid-level image representations using convolutional neural networks", in

Proc: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1717–1724.

Vita

Second Lieutenant Cassandra R. Post graduated from Manassas Park High School in Manassas Park, Virginia. She entered undergraduate studies at the University of Kansas in Lawrence, Kansas where she graduated with a Bachelor of Science degree in Computer Engineering in May 2015. She was commissioned through the Detachment 280 AFROTC at the University of Kansas where she was nominated for a Regular Commission.

Her first assignment was at Wright-Patterson AFB as a graduate student in the Graduate School of Engineering and Management, Air Force Institute of Technology in June 2015. Upon graduation, she will be assigned to Keesler AFB as a student in Undergraduate Cyberspace Training in April 2017.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
23–03–2017	Master's Thesis	June 2015 — Mar 2017

4. TITLE AND SUBTITLE

Towards Automation of Tipping and Cueing Between Small Satellites in a Constellation

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Post, Cassandra, R., 2nd Lieutenant, USAF

5d. PROJECT NUMBER

16G156

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENG-MS-17-M-061

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Naval Postgraduate School, Center for Multi-INT Studies
Director of Research Programs and Operations
Attn: Deborah Shifflett, COMM: 831-656-3758, EMAIL: dsshiffl@nps.edu
833 Dyer Road
Bldg. 232, SP437
Monterey, CA 93943

10. SPONSOR/MONITOR'S ACRONYM(S)

NPS/CMIS

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The use of using low-fidelity sensors of satellites in a constellation for accurate surface target detection has the potential to lower costs while increasing flexibility, replacement time, and fault tolerance. This thesis investigates the possibility of utilizing an array of satellites with a heterogeneous mix of sensor types to optimize the validation process of surface target detection. Generation of synthetic scenes allows identification and extraction of optical features that are useful in remote sensing practices. Features of interest are generated from specified resolutions, representing different sensor types using Rochester Institute of Technology's Digital Imaging Remote Sensing Image Generation platform. Synthetic images are jittered to varying degrees to represent the pointing stability. These synthetic images are utilized to train The Berkeley Caffe Convolution-Based Deep Learning open source platform to automatically detect features of interest. Through these experiments, we demonstrate that remote sensing platforms can provide features of interest to an artificial intelligence platform to increase overall feature identification effectiveness. Through these experiments, we demonstrate that remote sensing platforms can provide features of interest to an artificial intelligence platform to increase overall feature identification effectiveness.

15. SUBJECT TERMS

Satellites in Constellation, Tipping and Cueing, Convolutional Neural Networks, Remote Sensing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Kenneth Hopkinson (ENG)
U	U	U	UU	142	19b. TELEPHONE NUMBER *(include area code)* (937)255-3636,x4579 kenneth.hopkinson@afit.edu