REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
03/27/2023	Conference Briefing	

4. TITLE AND SUBTITLE	**5a. CONTRACT NUMBER**
Investigations Into Image Interpretability For Machine Learning	W56KGU-18-D-0004
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Jared Taylor, Nicholas Noboa, Sofia D. Hirschmann, James Tanis, Paul Brown, John M. Irvine	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
The MITRE Corporation 202 Burlington Road Bedford, MA 01730	PRS-23-0674
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
United States Army (ACC-APG)	ACC-APG
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES

Preprint to Joint Agency Commercial Imagery Evaluation

14. ABSTRACT

Humans and ML algorithms are sensitive to difference image properties; Image chain analysis aids in understanding factors affecting ML performance; Clutter affects the false alarm rate: Developing metrics to quantify clutter; Target occlusion affects detection performance; There is a weak, but significant relationship between clutter and target occlusion; Putting it all together: Understanding clutter will help us understand both false alarms and target detection.

15. SUBJECT TERMS

Image Processing; Artificial Intelligence; image chain analysis; machine learning

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Susan Carpenito
				22	**19b. TELEPHONE NUMBER** *(Include area code)* 781-271-7646

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Investigations Into Image Interpretability For Machine Learning

Jared Taylor, Nicholas Noboa,
Sofia D. Hirschmann, James Tanis, PhD,
Paul Brown, John M. Irvine, PhD

March 28, 2023

This invention was made with Government support under W56KGU-18-D-0004 and associated task orders, awarded by Army Contracting Command, Aberdeen Proving Ground. The Government has certain rights in the invention.

MITRE | SOLVING PROBLEMS FOR A SAFER WORLD®

Data Rights

NOTICE

This software was produced for the U. S. Government under Contract No. FA8702-19-C-0001, and is subject to the Rights in Noncommercial Computer Software and Noncommercial Computer Software Documentation Clause DFARS 252.227-7014 (FEB 2014)

This technical data deliverable was developed using contract funds under Basic Contract No. W56KGU-18-D-0004

Outline

- Need to understand image quality for machine learning
- Image Chain Analysis (ICA) for machine learning
- Example ICA analysis
- Complexity of the background clutter
- Machine Learning (ML) for object detection
- Clutter and target occlusion
- Next steps

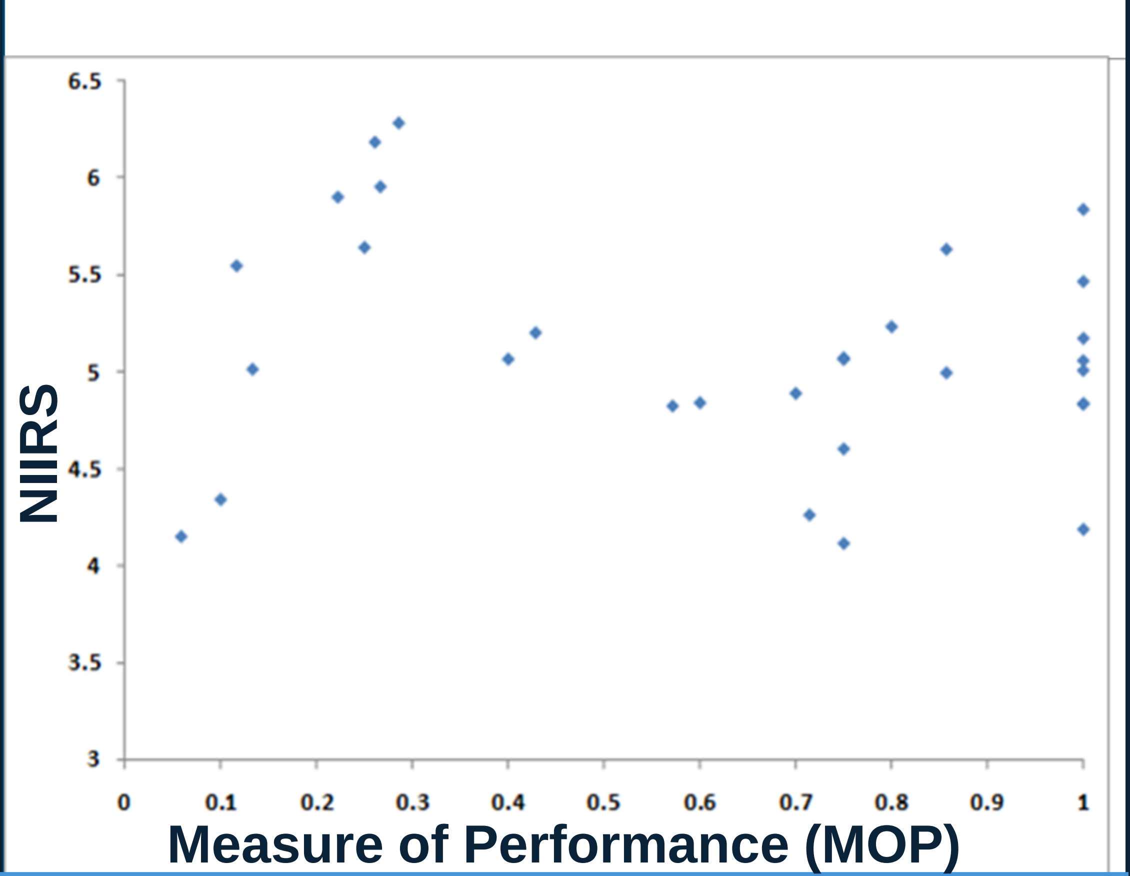
Image Quality for Machine Learning (ML)

A standard measure of image quality for ML is needed to

- Ensure imagery collection satisfies needs
- Enable sensor design studies

The National Imagery Interpretability Rating Scale (NIIRS) has served this need based on human perception of the imagery

NIIRS measures human perception which differs from AI/ML performance



$$MOP = \frac{1 + Detections}{1 + True\ Targets + False\ Alarms}$$

Image Chain Analysis

IMAGERY

TRAIN AND TEST AI/ML MODELS IN CONTROLLED SETTING

AI/ML Models
Multiple Experimental Threads

OBJECTIVE MOP

EMPIRICAL PERFORMANCE

EVALUATE POSSIBLE SYSTEM CHANGES:

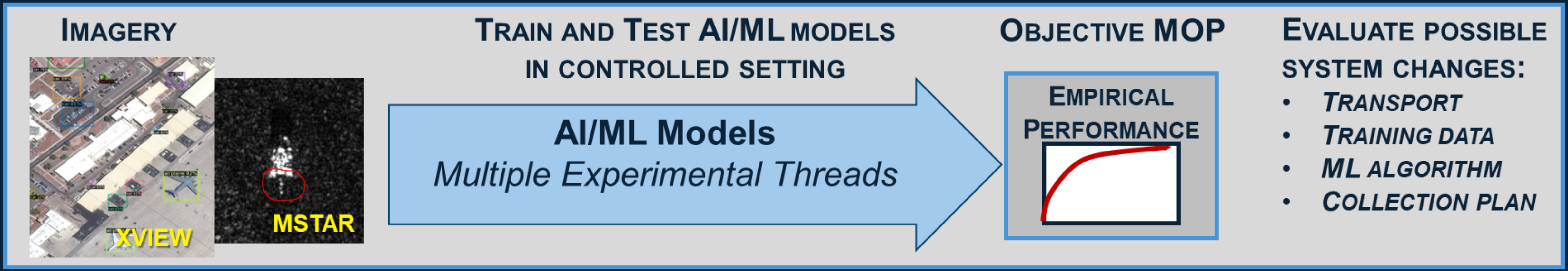
- *TRANSPORT*
- *TRAINING DATA*
- *ML ALGORITHM*
- *COLLECTION PLAN*

Image Chain Analysis (ICA) provides a framework for understanding image quality for machine learning

– Systematic experimentation to quantify image quality effects on ML performance

ICA Framework for Exploring Image Quality

- Leverage unclassified imagery data and deep learning tools to develop and demonstrate the ICA framework



SAR Experiments

- AI/ML tool: TensorFlow
- Imagery: MSTAR from AFRL
- Experiments: Training set, sensor noise emulation, viewing geometry

MITRE

Machine Learning Dataset

Targets Used:

T72 variants

T62

2S1

BRDM

BTR60

D7

ZIL131

ZSU23

Images from MSTAR Public Release available from AFRL

	2S1	A04	A05	A07	A10	A32	A62	A63	A64	BRDM	BTR60	D7	T62	ZIL131	ZSU
2S1	0	346.4	350.2	345.6	341.5	348.8	345.7	342.5	347.1	344	349.1	345.6	344	348.1	347.5
A04	346.4	0	337.3	337.7	341	339.6	341.1	337.4	339.6	343.5	346	344	342	344.6	342.6
A05	350.2	337.3	0	340.6	342.9	336.6	338.4	337.6	340.1	343.6	347.6	342.6	341.3	344.2	338.3
A07	345.6	337.7	340.6	0	339.2	339.6	338	338.3	340.8	344.2	347.6	342.7	341.7	343.2	338.3
A10	341.5	341	342.9	339.2	0	339.6	337.8	336.2	341.6	349.5	347.8	343.2	348.3	339.8	340.4
A32	348.8	339.6	336.6	339.6	339.6	0	339.2	334.6	340.1	348	343.9	344.2	345	347.2	347.2
A62	345.7	341.1	338.4	338	337.8	339.2	0	335.1	335.6	340.5	347.2	341.1	343	341.4	335.3
A63	342.5	337.4	337.6	338.3	336.2	334.6	335.1	0	342.6	344.4	343.1	339.7	342.6	342.1	343.8
A64	347.1	339.6	340.1	340.8	341.6	340.1	335.6	342.6	0	336.6	348.8	345.2	339.3	343.2	336.7
BRDM	344	343.5	343.6	344.2	349.5	348	340.5	344.4	336.6	0	351.2	345.1	339.4	347.1	342.5
BTR60	349.1	346	347.6	347.6	347.8	343.9	347.2	343.1	348.8	351.2	0	344.4	350.6	351.2	353.8
D7	345.6	344	342.6	342.7	343.2	344.2	341.1	339.7	345.2	345.1	344.4	0	346.8	347.9	341.2
T62	344	342	341.3	341.7	348.3	345	343	342.6	339.3	339.4	350.6	346.8	0	346.4	338.4
ZIL131	348.1	344.6	344.2	343.2	339.8	347.2	341.4	342.1	343.2	347.1	351.2	347.9	346.4	0	339.4
ZSU	347.5	342.6	338.3	338.3	340.4	347.2	335.3	343.8	336.7	342.5	353.8	341.2	338.4	339.4	0

Euclidian distances of target chips: differences between T72 variants are small enough to treat them as a single target type when determining correct identifications

Optimized Model With Image Filtering

- Original model trained on full set of targets at 15-degree incident angle
- New test model trained with enhanced images consisting of original image & an inverse processed separately and combined into separate layers
  - Noise reduction filter
  - Dilation & Erosion filter
- New model theoretically uses shapes of shadows to aid in detection
- Slight increase in performance demonstrated with enhanced data set

	Unedited			Filtered	
	Class	Score		Class	Score
1	2S1	0.981988	1	2S1	0.995488
2	A04	0.629363	2	A04	0.710956
3	A05	0.964956	3	A05	0.979554
4	A07	0.954984	4	A07	0.995919
5	A10	0.973192	5	A10	0.99016
6	A32	0.729914	6	A32	0.819216
7	A62	0.922954	7	A62	0.923272
8	A63	0.986053	8	A63	0.987068
9	A64	0.94531	9	A64	0.977449
10	BRDM_2	0.945393	10	BRDM_2	0.988638
11	BTR60	0.857389	11	BTR60	0.997824
12	D7	0.983577	12	D7	0.979696
14	T62	0.990533	14	T62	0.964247
15	ZIL131	0.986856	15	ZIL131	0.951047
16	ZSU23_4	0.982481	16	ZSU23_4	0.992994
	Average:	0.92233		Average:	0.950235

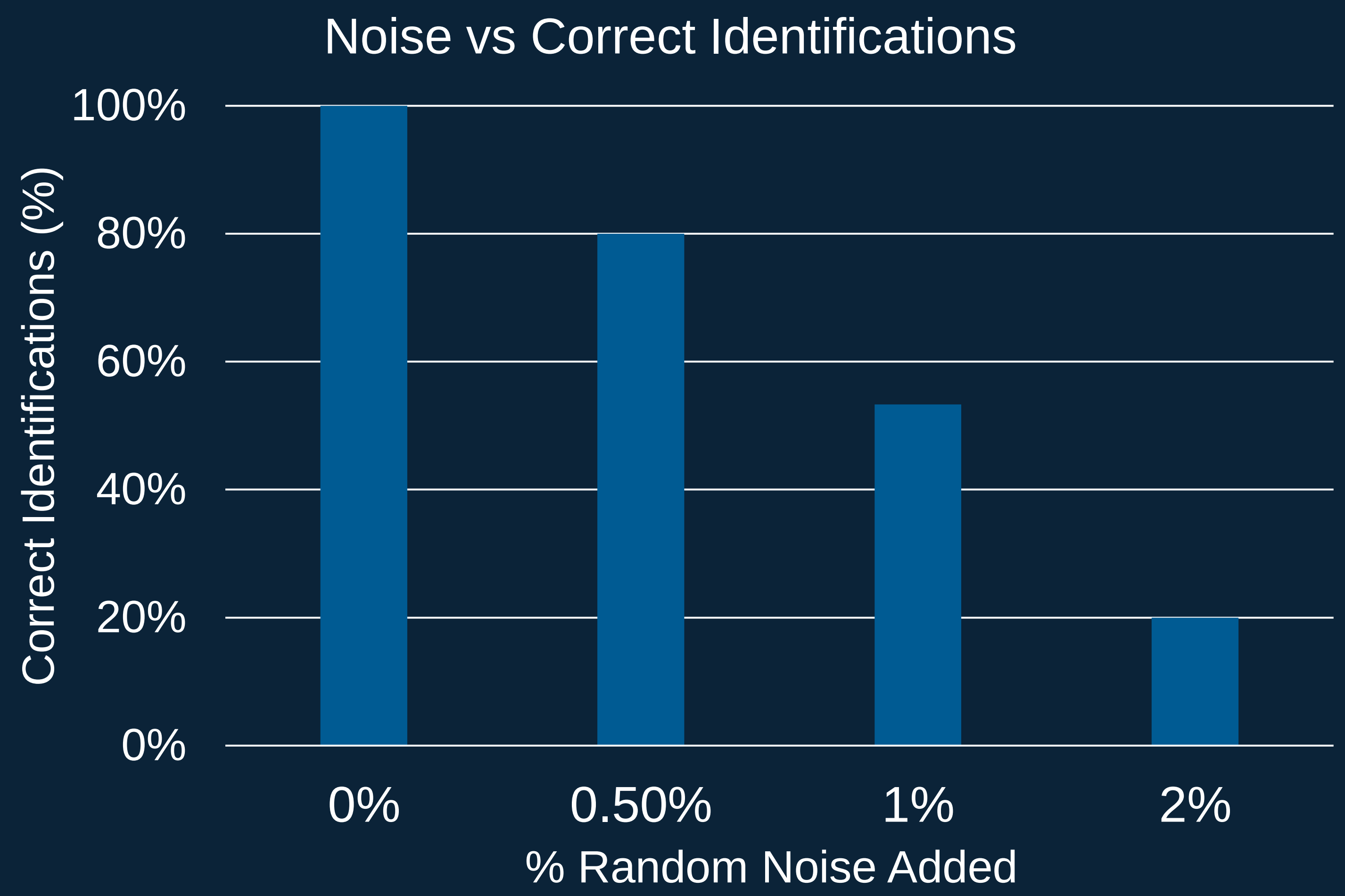
Performance of model with unedited versus filtered images

Noise Added (% Pixels Affected, Condensed T72 Variants)

Noisy 0.5%	2S1	T72	BRD	BTR	D7	T62	ZIL1	ZSU
2S1	1							
T72		6				2		
BRDM	1							
BTR60	1							
D7					1			
T62						1		
ZIL131							1	
ZSU								1

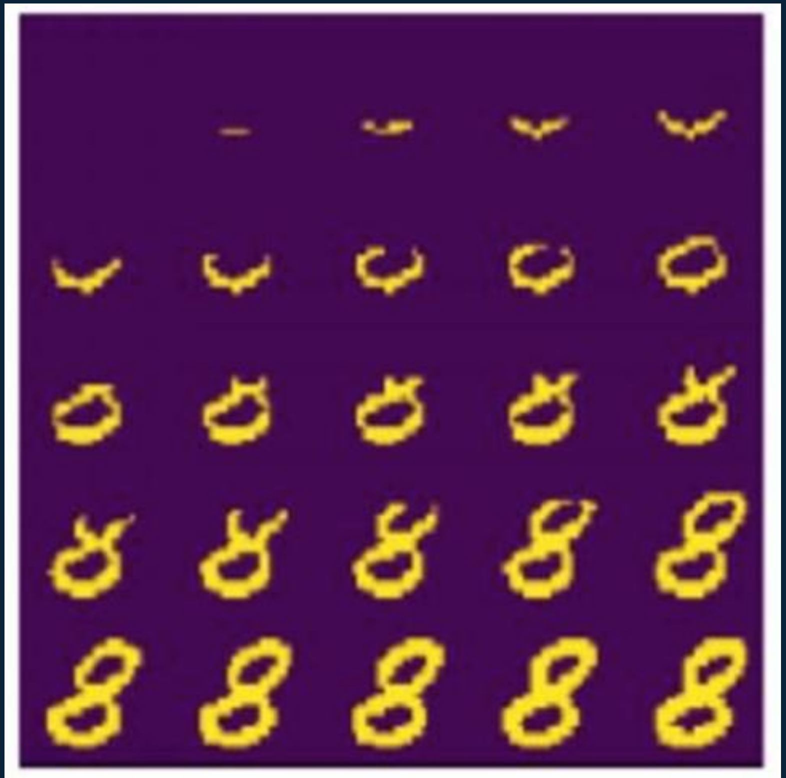
Noisy 1%	2S1	T72	BRDM	BTR60	D7	T62	ZIL131	ZSU
2S1						1		
T72		4				4		
BRDM	1							
BTR60	1							
D7					1			
T62						1		
ZIL131							1	
ZSU								1

Noisy 2%	2S1	T72	BRD	BTR	D7	T62	ZIL1	ZSU
2S1						1		
T72						7	1	
BRDM						1		
BTR60	1							
D7					1			
T62						1		
ZIL131							1	
ZSU					1			



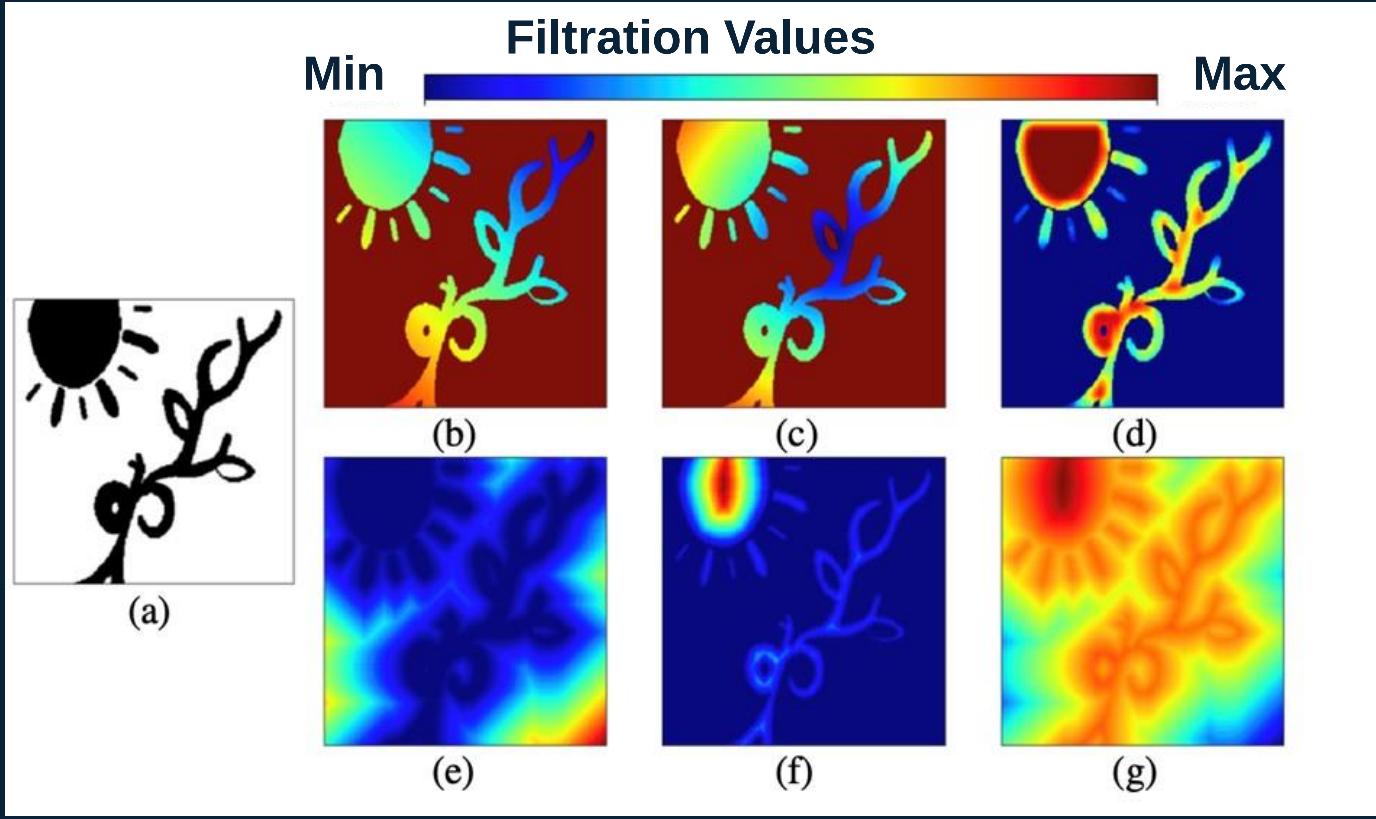
Topological Data Analysis

https://cs230.stanford.edu/projects_spring_2018/reports/8290918.pdf



A height incrementally uncovers the cycles and connected components of the “Figure 8” symbol

There Are a Variety of Filtrations



a) Binary
b) Height
c) Radial
d) Density
e) Dilation
f) Erosion
g) Signed Distribution

Topological Features and Classification of Clutter Level

50 Trials

50 Random splits

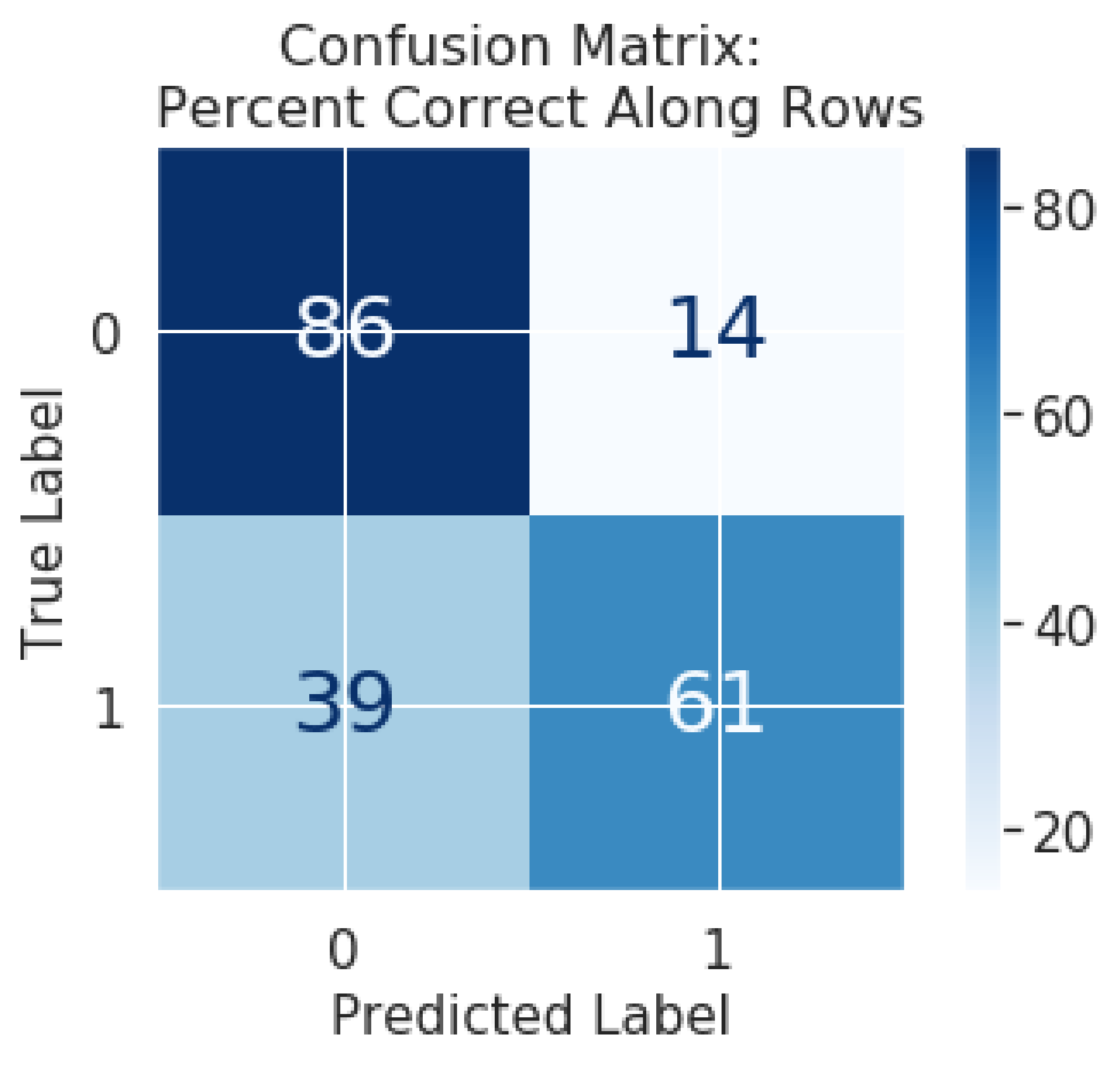
Reported as a percent

Accuracy: 72.5 ± 8.5

- Mean = 73.3
- Std Dev = 4.4
- 96% confidence interval = [64, 81]

F1: 68 ± 10

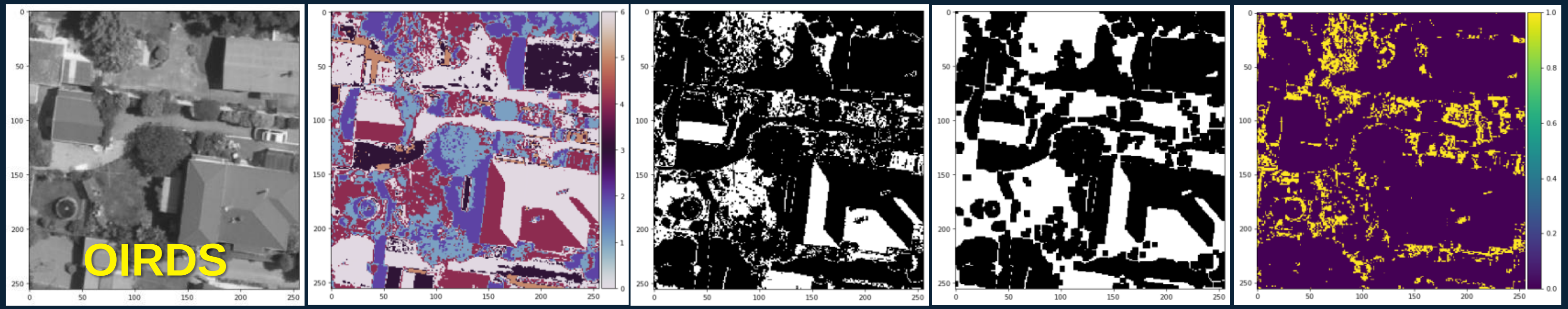
- Mean = 67.9
- Std Dev = 5.9
- 96% confidence interval = [58,78]



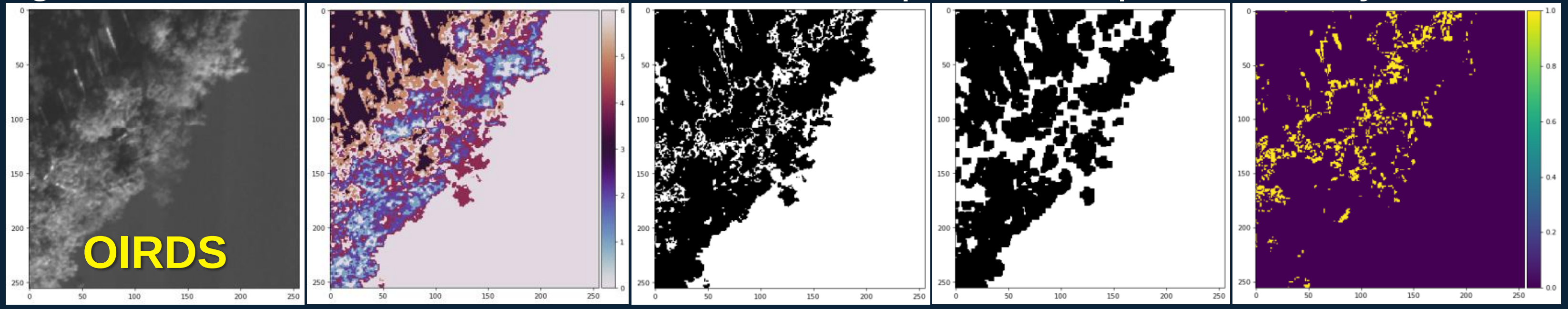
Clutter Analysis Using K-Means & Morphology Features

- Clusters
- Morphological Filtering
- Structural Similarity
- Texture

Image | **K-Means** | **Cluster Mask** | **Filtered** | **Difference**

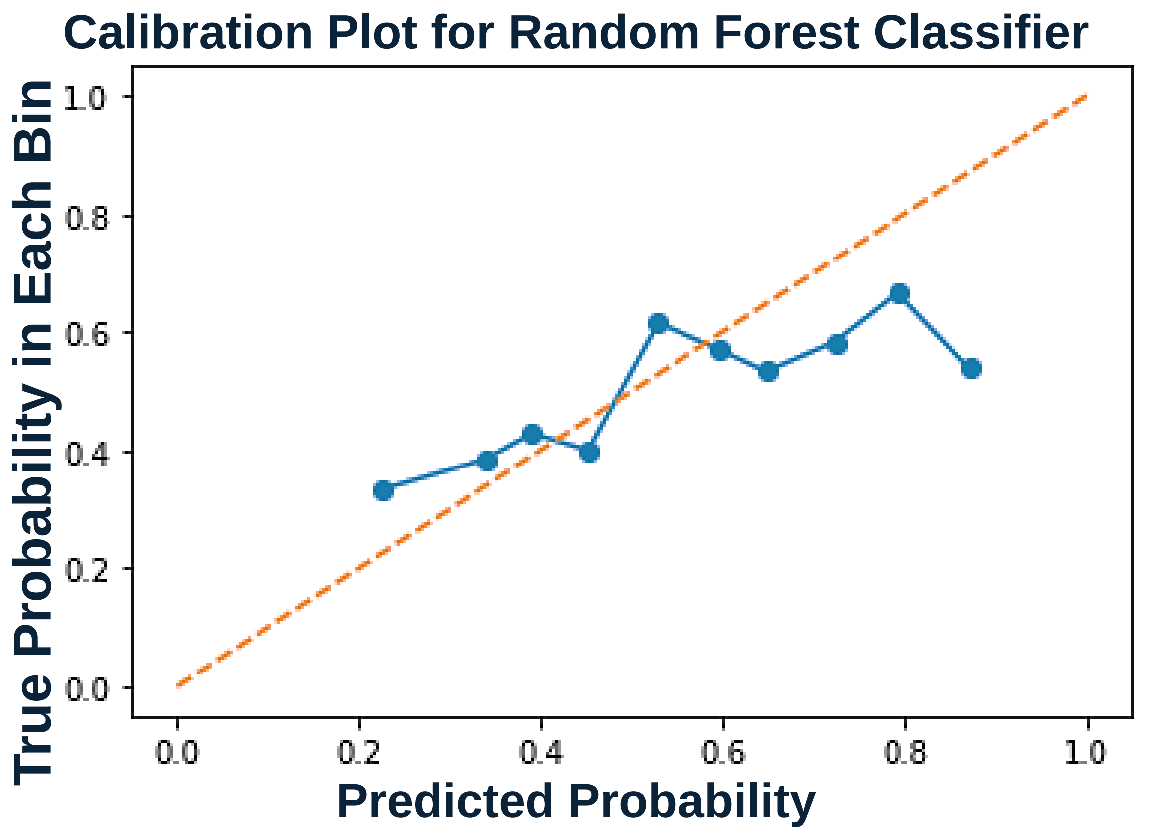


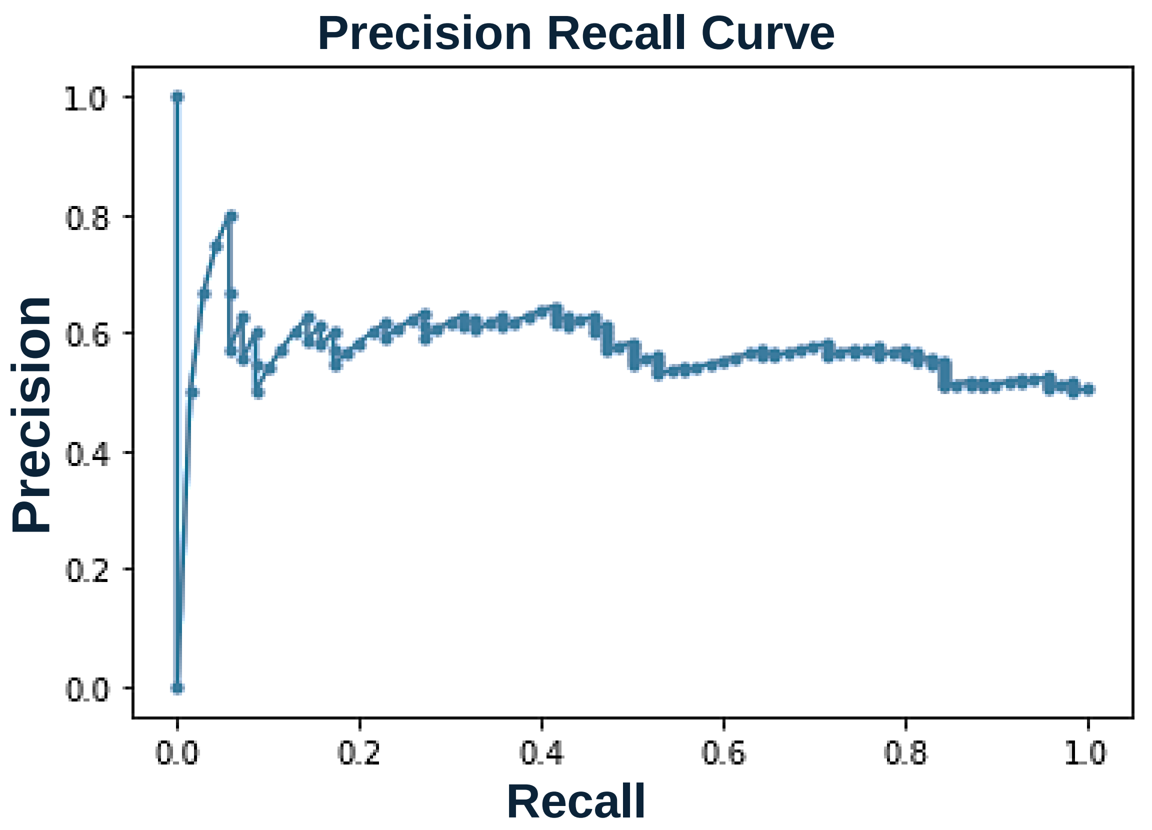
High Clutter: K-means cluster with maximum pixels is represented by **maroon**



Low Clutter: K-means cluster with maximum pixels is represented by **light pink**

Random Forest Performance





New Classifier

Standard Random Forest

True Labels

0	34	35
1	22	48
	0	1

Predicted Labels

F1: 0.627

Accuracy: 0.589

Probabilistic Random Forest

True Labels

0	33	36
1	21	49
	0	1

Predicted Labels

F1: 0.632

Accuracy: 0.589

Training Models for Ensemble Predictor

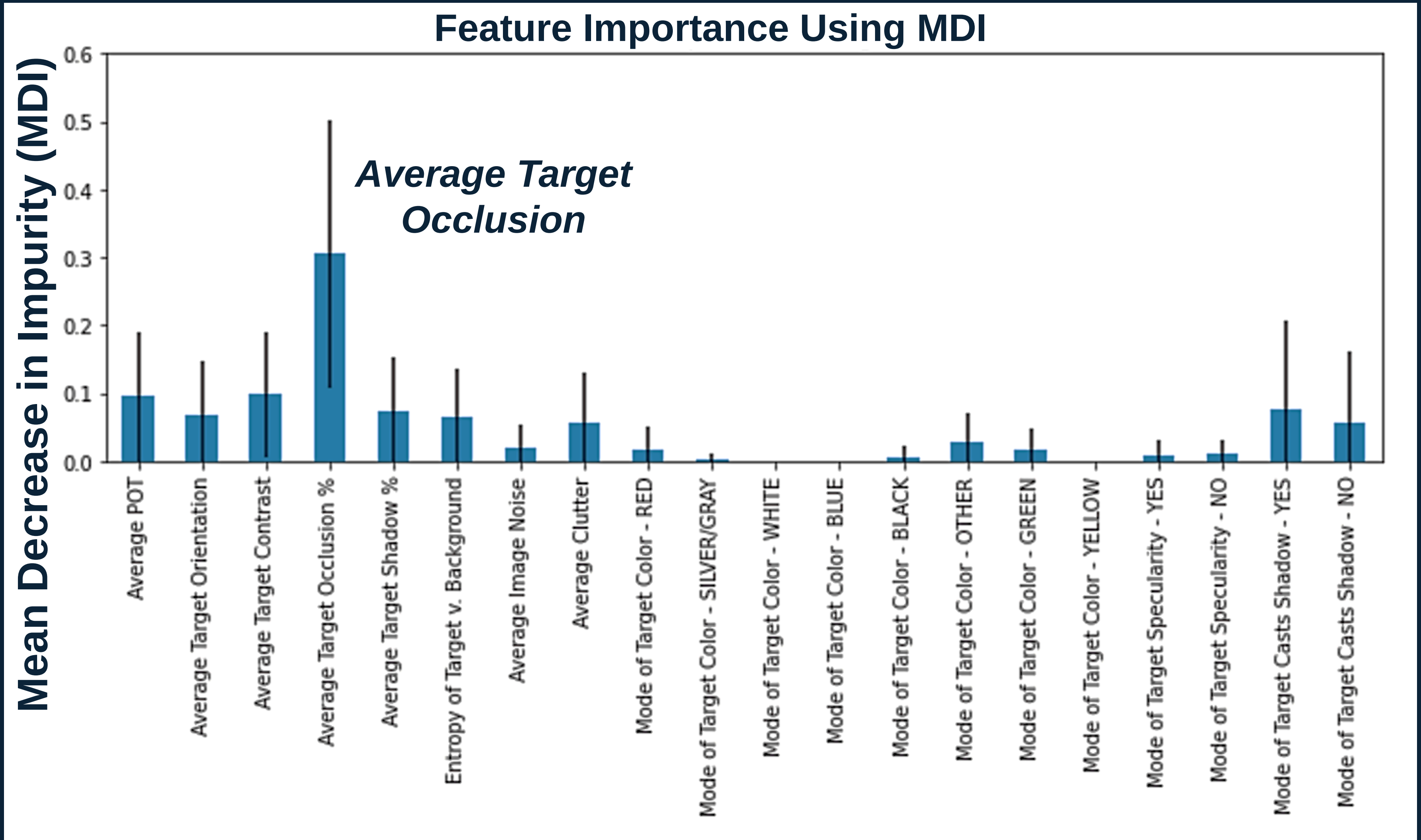
OIRDS/39867435_1537_769_1793_1025.tif

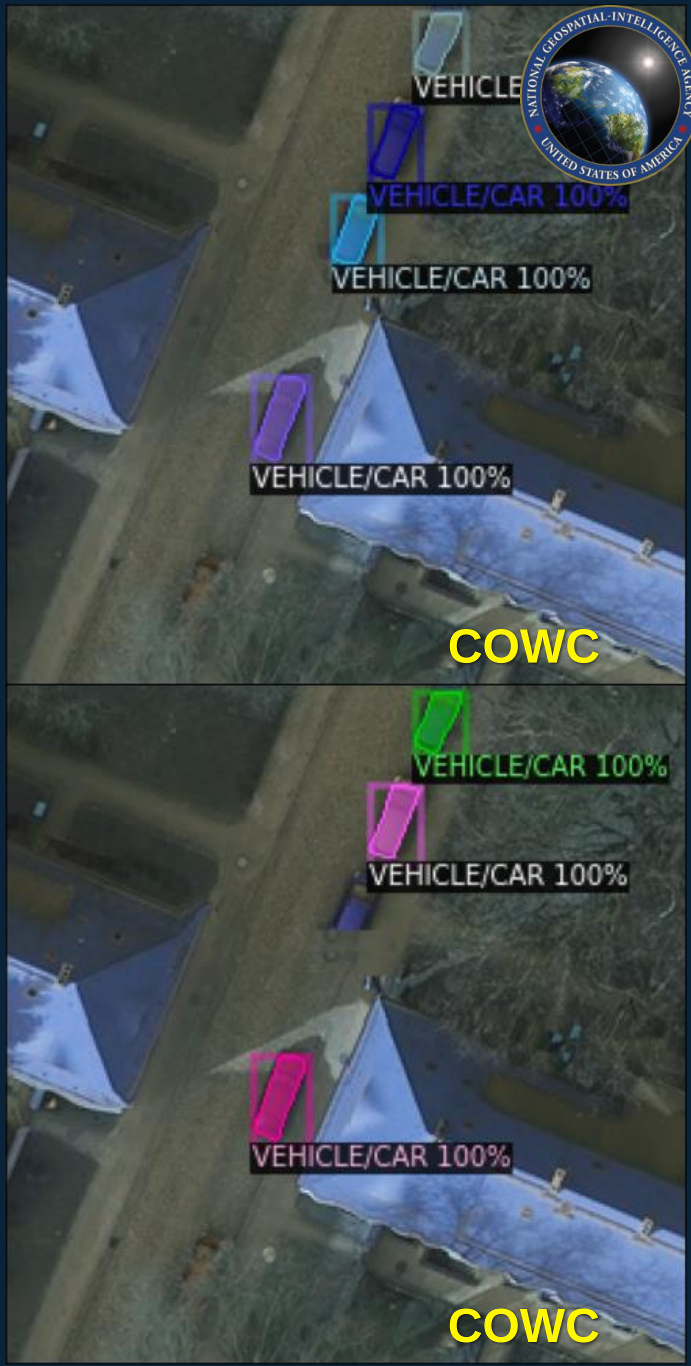
Unmodified

Mixed

200% Gamma

Factors Affecting Performance



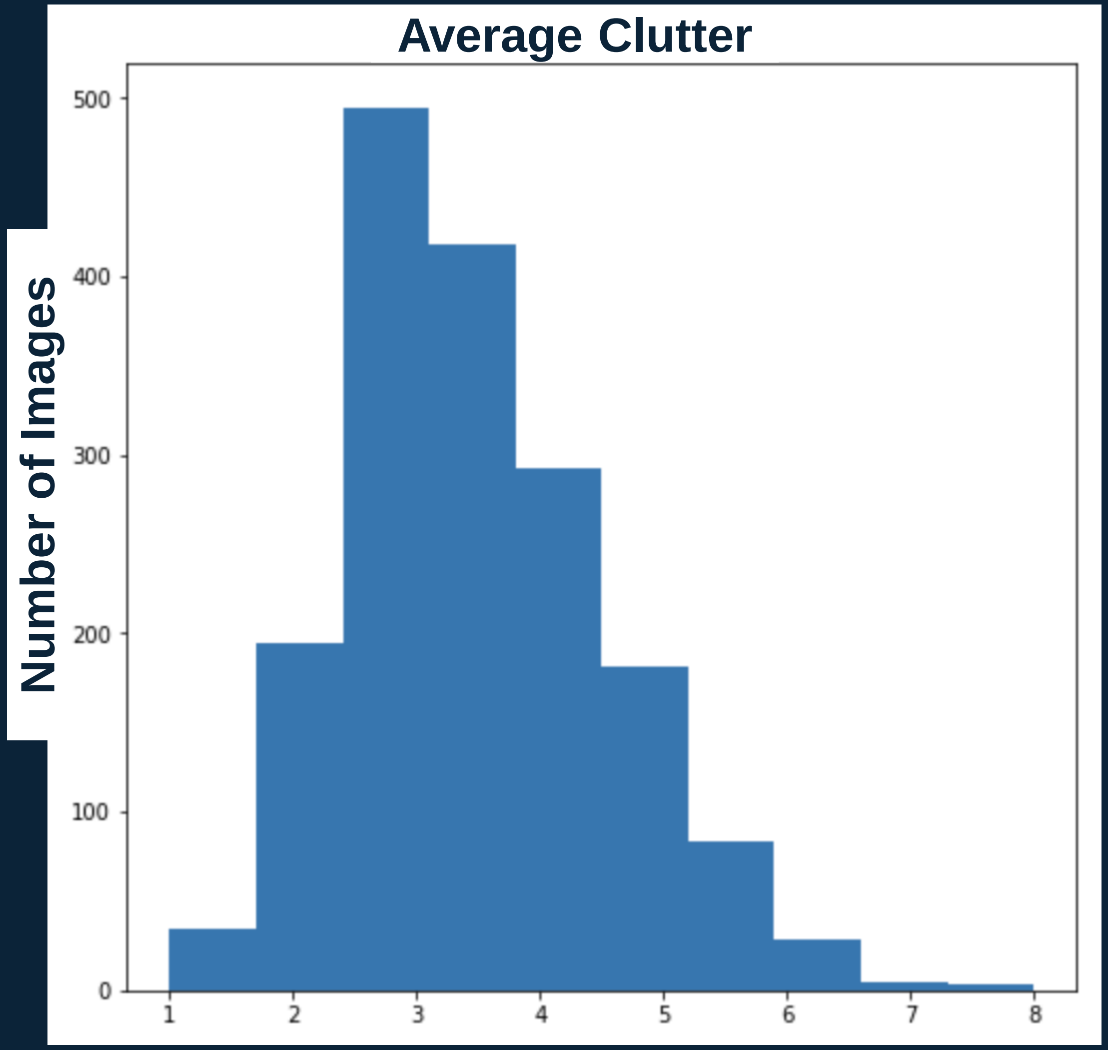


COWC/Potsdam_2_10_RGB.0.0.jpg

Assessment of Clutter and Target Occlusion

The majority of OIRDS imagery maintain an average clutter level somewhere between 2.5 and 4.5

	% Correlation with Clutter
Probability of Target (POT)	-2.564%
Average POT %	-7.330%
Average Target Occlusion %	9.784%



Clutter itself has little correlation with a human’s ability to identify targets; however, there is a small but significant correlation between the clutter level and rate of occlusion, and a negative correlation with the average pixels on target, which will decrease as occlusion increases.

Summary and Next Steps

- Summary:
  - Humans and ML algorithms are sensitive to difference image properties
  - Image chain analysis aids in understanding factors affecting ML performance
  - Clutter affects the false alarm rate: Developing metrics to quantify clutter
  - Target occlusion affects detection performance
  - There is a weak, but significant relationship between clutter and target occlusion
  - Putting it all together: Understanding clutter will help us understand both false alarms and target detection

- Next Steps:
  - Analysis with new data sets – do the findings hold up across a range of conditions?
  - Development of performance models for ML
  - Establish the basis for a standard measure of image quality for ML

QUESTIONS?

Relevant Scoring Definitions

- **True Positives = Number of correct target identifications**
- **False Positives = Number of incorrect object identifications**
- **False Negatives = Number of true targets missed**
- **Precision = True Positives / (True Positives + False Positives)**
  - *Measure of model's ability to correctly identify targets*
- **Recall = True Positives / (True Positives + False Negatives)**
  - *Measure of model's ability to find all true targets*
- **F1 Score = 2 * (Precision * Recall) / (Precision + Recall)**
  - *Measure of model's overall performance*

Ground truth for OIRDS test subset = 167 cars

Ground truth (visible targets) for OIRDS test subset = 145 cars

Appendix – ML Model

- Ensemble model – minimum two models agree on detection
- Precision: 0.964
- Recall: 0.796
- Adjusted Recall: 0.917
  - Adjusted Recall based on targets with probability of target 1 (100% certainty)
- F1 Score: 0.872
- Adjusted F1 Score: 0.940
  - Uses Adjusted Recall

Imagery Sources

- xView: xView (xviewdataset.org)
- MSTAR: SDMS Public Web Site (af.mil)
- OIRDS: Overhead Imagery Research Data Set download | SourceForge.net
- COWC: Cars Overhead with Context (COWC) | Library Digital Collections | UC San Diego Library (ucsd.edu)

- All data sets are publicly available, open source.