Distribution Statement "A" (Approved for Public Release, Distribution Unlimited). This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR001120C0055 and 140D0422C0043. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Defense Advanced Research Projects Agency (DARPA); or its Contracting Agent, the U.S. Department of the Interior, Interior Business Center, Acquisition Services Directorate, Division V.

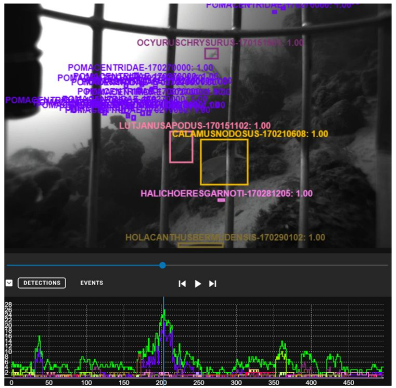
Detecting Novel Objects and Anomalies in Marine Imagery using VIAME

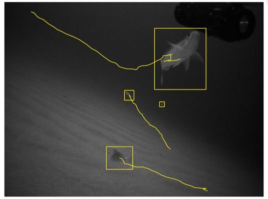
Christopher Funk, Alexander Lynch, Roddy Collins, Sarah Brockman, Bryon Lewis, Mary Salvi, Roni Choudhury, Matt Dawkins, Anthony Hoogs

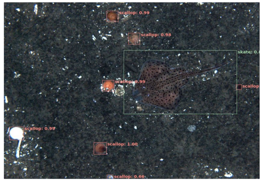
Sponsors:

Video and Imagery Analytics for Marine Environments (VIAME)

- Do-It-Yourself system for end-users to create specialized AI models for a variety of visual analytics
- Open-source with permissive licensing on viametoolkit.org with a public web server at viame.kitware.com
- Used around the world for detection and classification of fish, scallops, and marine mammals from imagery
- Contains web and desktop interfaces with advanced features such as AI-assisted annotation, video tracking, and video search
- Includes specializations to underwater domain such as stereo measurement & image enhancement, but can be applied to other domains and problems





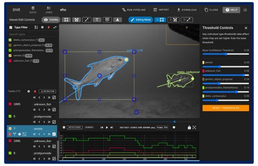


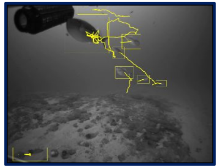
Video/Image Annotation

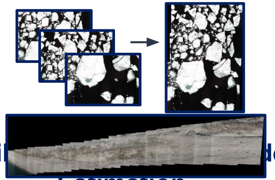
Automatic Object Tracking

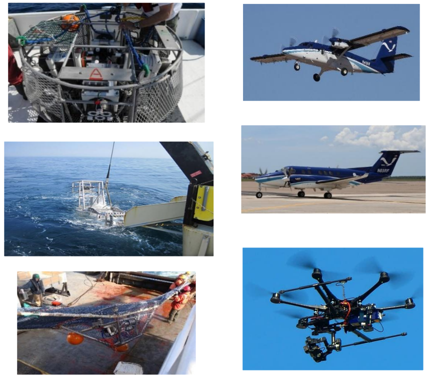
Image Enhancement

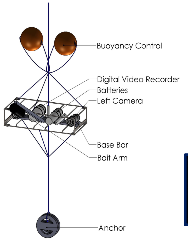
Image Registration and Mosaicing



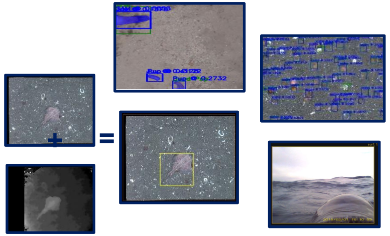




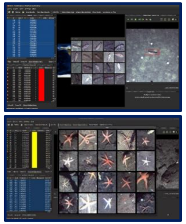




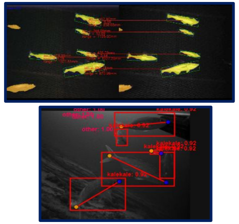
Object Detection and Frame Classification



Video Search and Rapid Model Generation

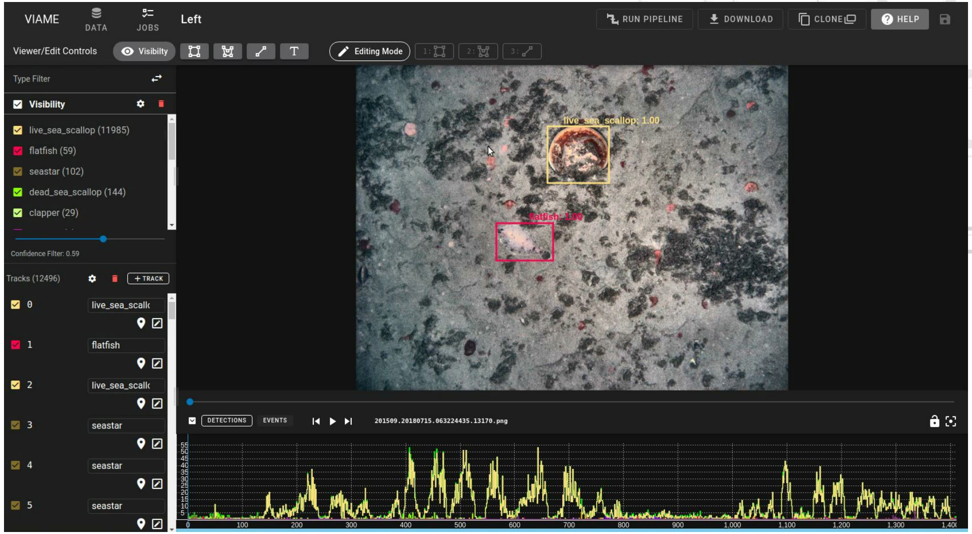


Stereo Measurement

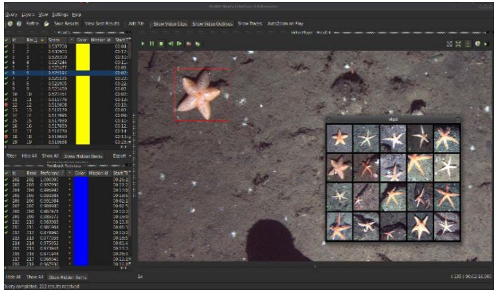


VIAME GUIs

- **VIAME-Web**: viame.kitware.com
- **DIVE**: For standard deep learning annotation and model training
- **SEARCH**: Utilizes video search for rapid model generation
- **SEAL**: Specialized for multi-modal (EO/IR/UV) annotation
- **VIEW**: Original desktop annotator for large resolution imagery
- **Project Folders**: Bulk processing outside of graphical interfaces using sh/bat scripts
- **Command Line Interfaces**



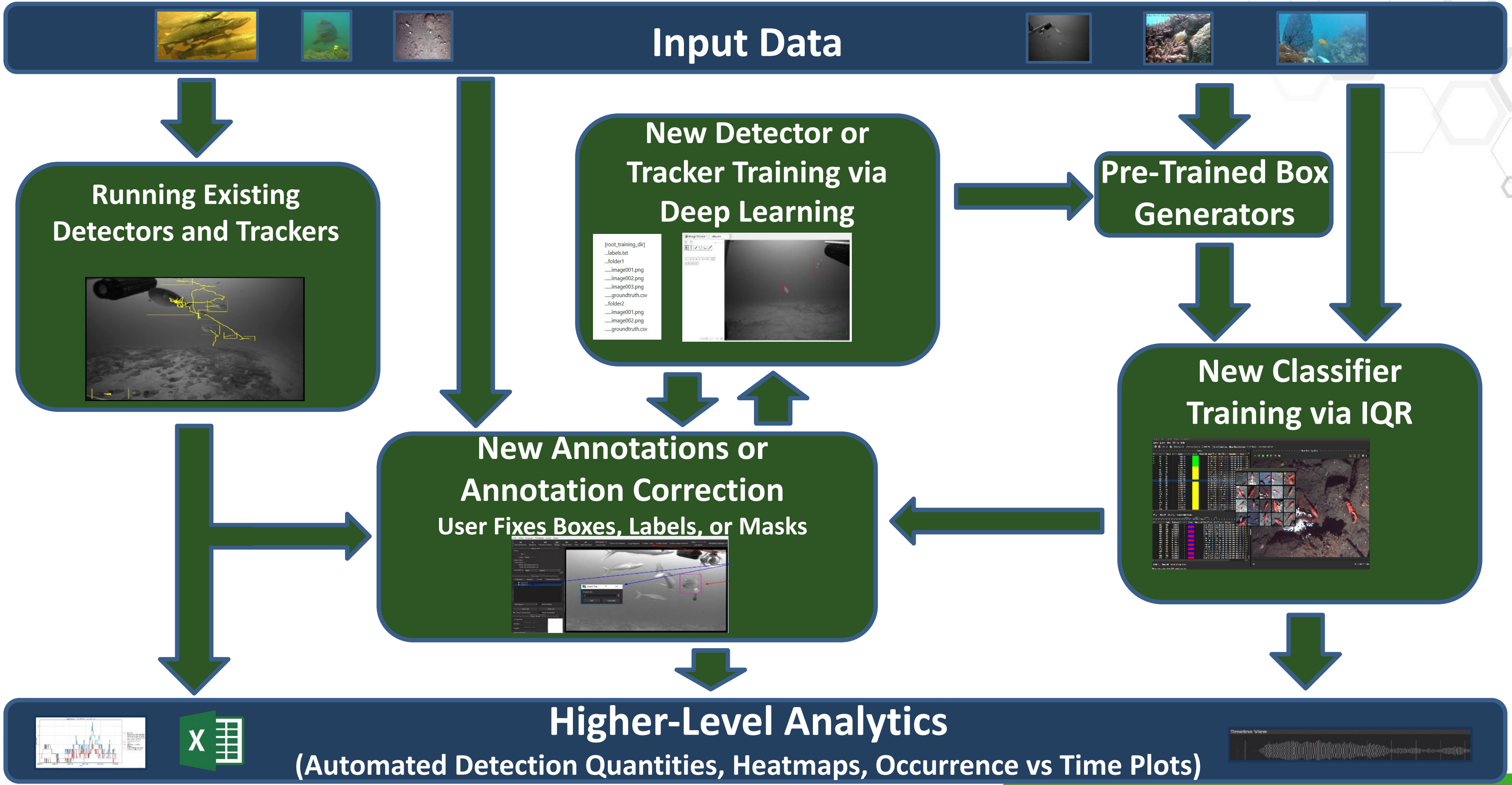
DIVE Interface (Desktop / Web)



SEARCH: Desktop Search Engine

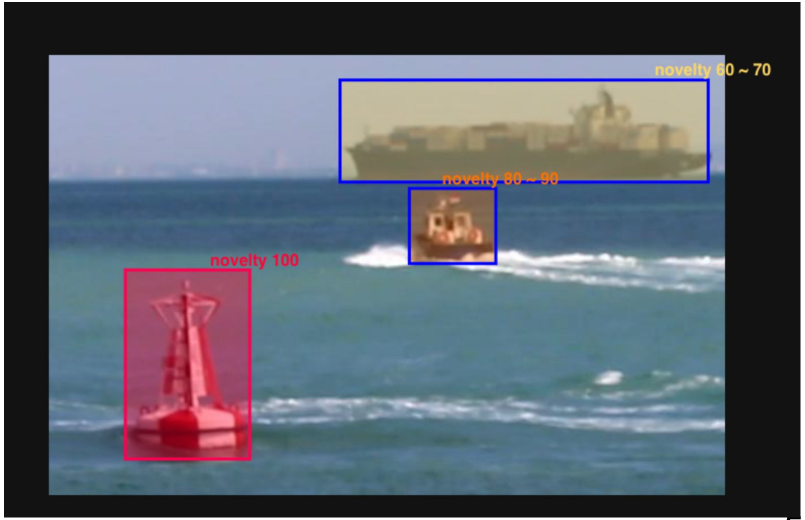
SEAL: Multi-Modal Annotator

VIAME Model Training and Execution



Problem: Anomaly Detection

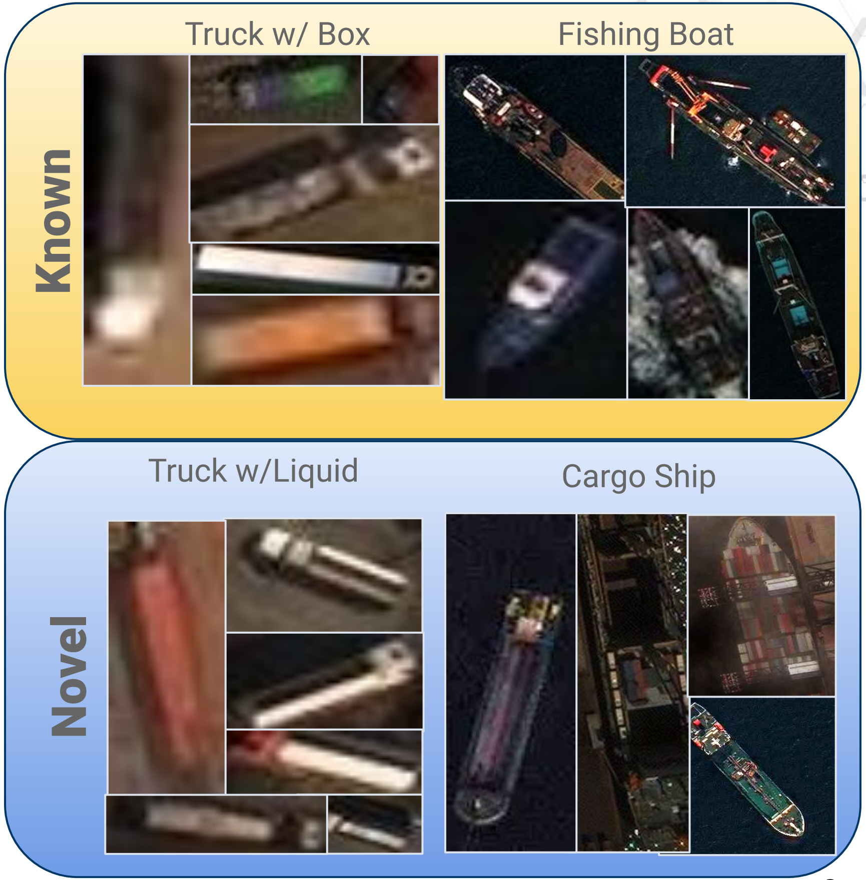
- Detect novel objects of interest – salient unknown unknowns – in marine images
  - Previously not in VIAME
- Novel objects are new types or new variants of known types
- Avoid false alarms from novel scene conditions or incidental object appearance changes
  - Weather, season
  - Viewpoint, lighting



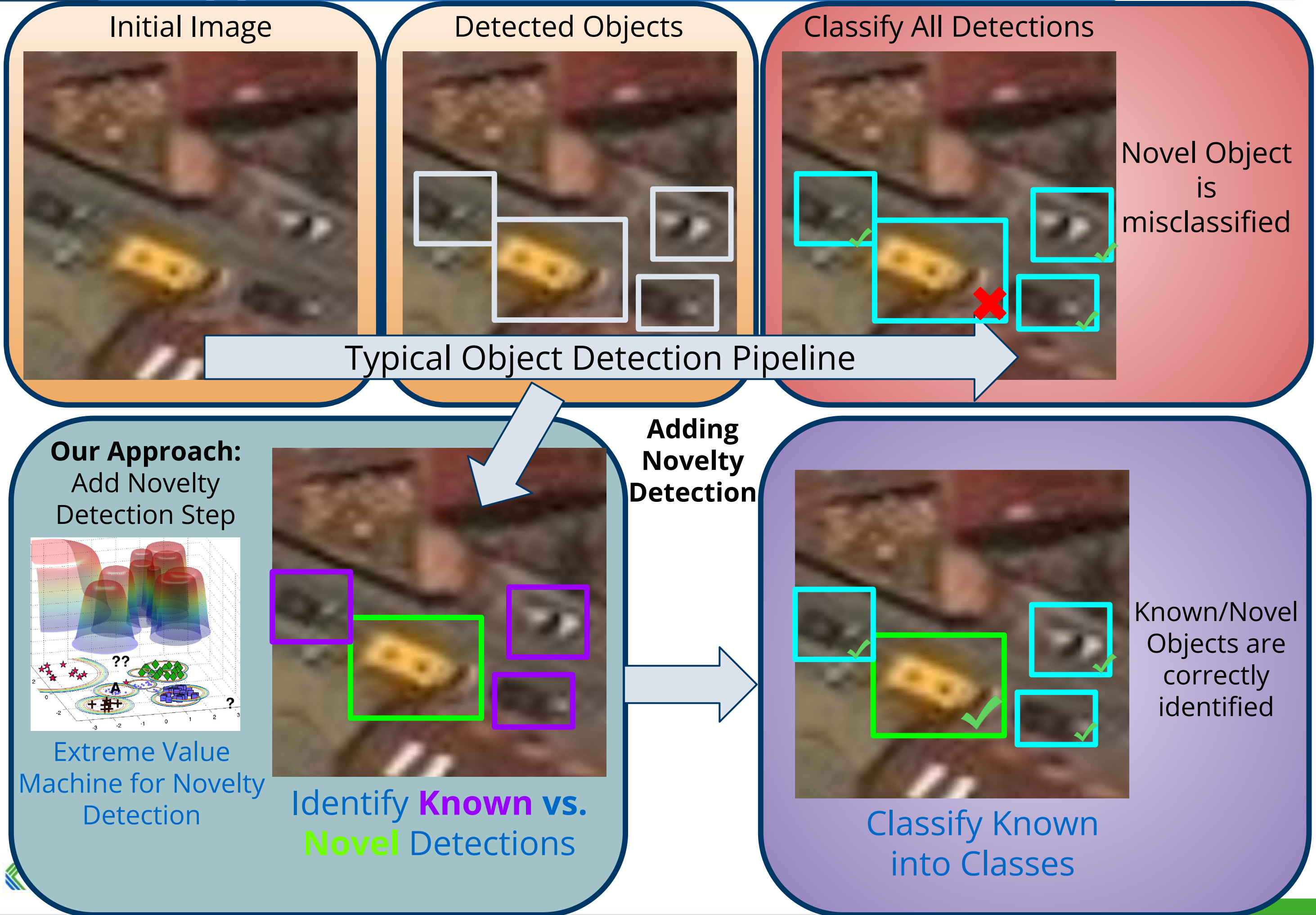
Challenges

- Salient, novel objects of interest can be very similar to known classes
- Some classes have high intra-class variation
- New imaging and scene conditions (viewpoint, lighting) can seem like novelties… but are they meaningful to an analyst?
- Thresholding on class scores can be unreliable for novelty detection
  - Most methods just misclassify novelties as known classes or false negatives

- **End-users need discovery of target novelties AND robustness against nuisance novelties**



Approach



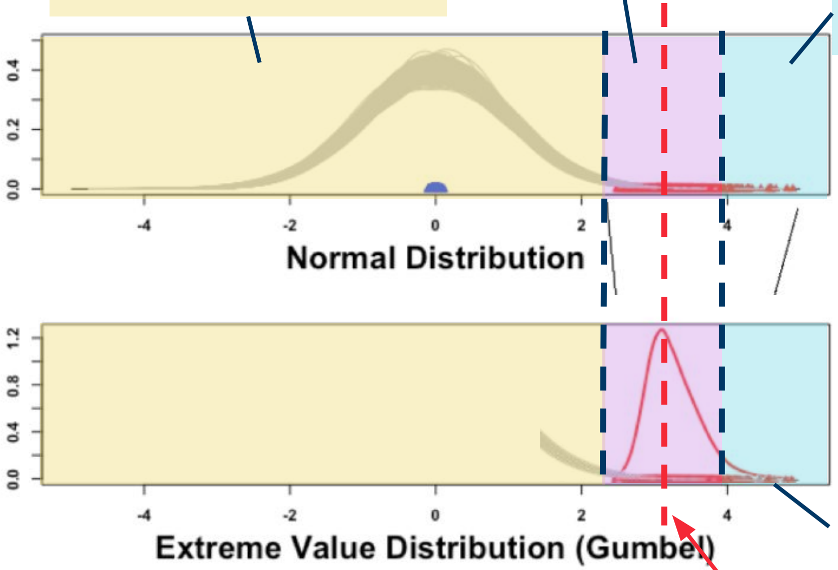
Each detection gets a novelty score

Legend:

Known Detection
Car Detection
Novel/Unknown Detection
Correct: ✓
Incorrect: ✗

Extreme Value Theory (EVT)

- Distribution of observed data is known, but not novelties. How far is “novelty” from the known distribution?
- EVT models only the tail of a distribution. In long-tailed problems, distributions of the tails can be very different for different problems and datasets. In EVT, the distribution of the tail is estimated from data.
- EVT is used to model the probability of very rare events such as extreme floods & tornados, to predict insurance losses.



Normal tail is very flat in the tough zone, lacking an obvious point for the decision boundary

EVT uses either Weibull, Gumbel, or Frechet (Generalized Pareto) depending on the type of tail

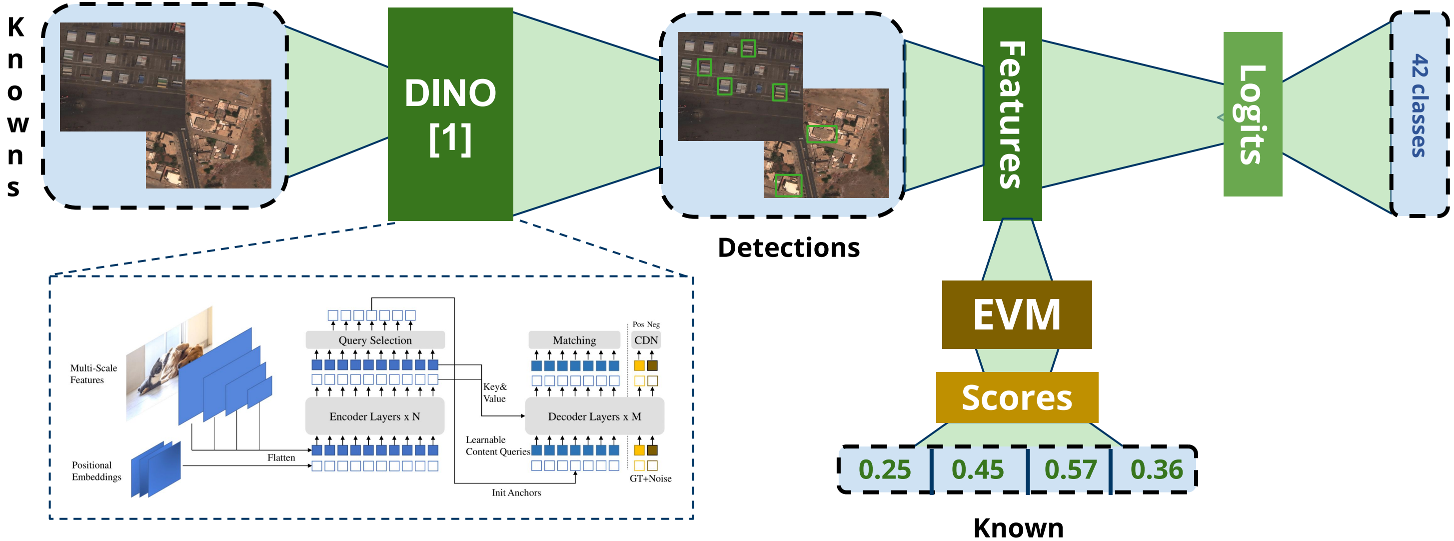
With a distribution fitted only to the tail, a meaningful decision threshold can be set

Source Image https://www.dataanalysisclassroom.com/lesson60/

Detector Model

Our method works with virtually any existing detector model, trained within VIAME or through another platform.

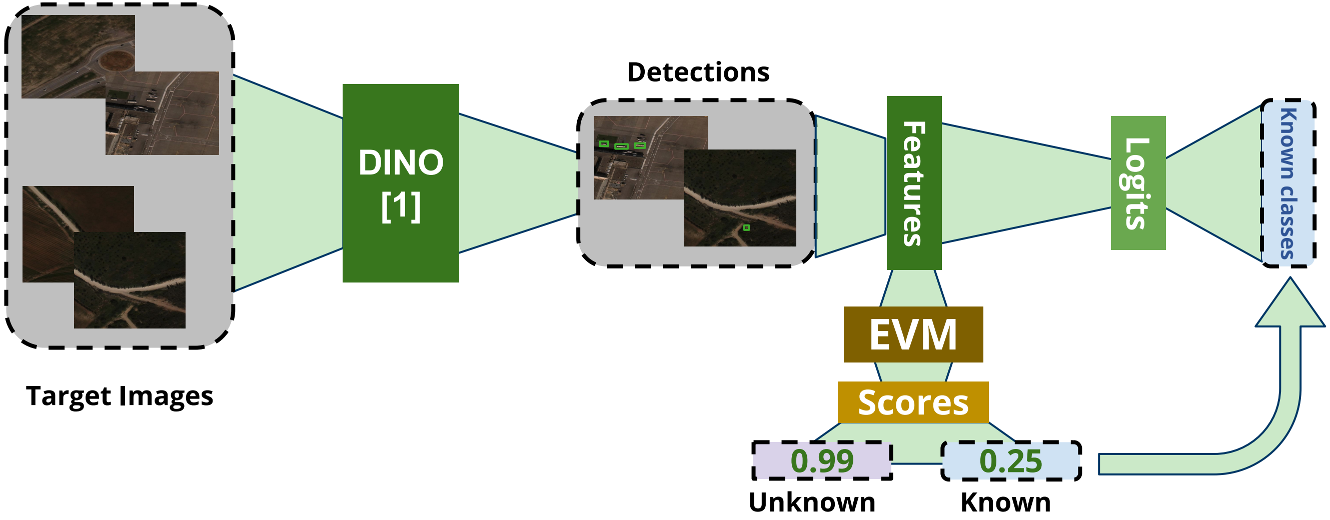
Given a detector, the Extreme Value Machine is trained to distinguish novelties from known classes. **Novelties are not used in training.** This is not deep-learning training.



[1] Zhang et al. DINO: DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection. International Conference on Learning Representations (ICLR) 2023.

Testing

DINO detects generic objects. The EVM model determines if each object instance is known or novel. If known, then it is classified into a known class.



Every detection gets a novelty score.

If an instance is detected as known, classify into a known class

Leveraging a trained detector is why our method detects novelties that are similar to the known classes, but sufficiently different from the background.

kitware

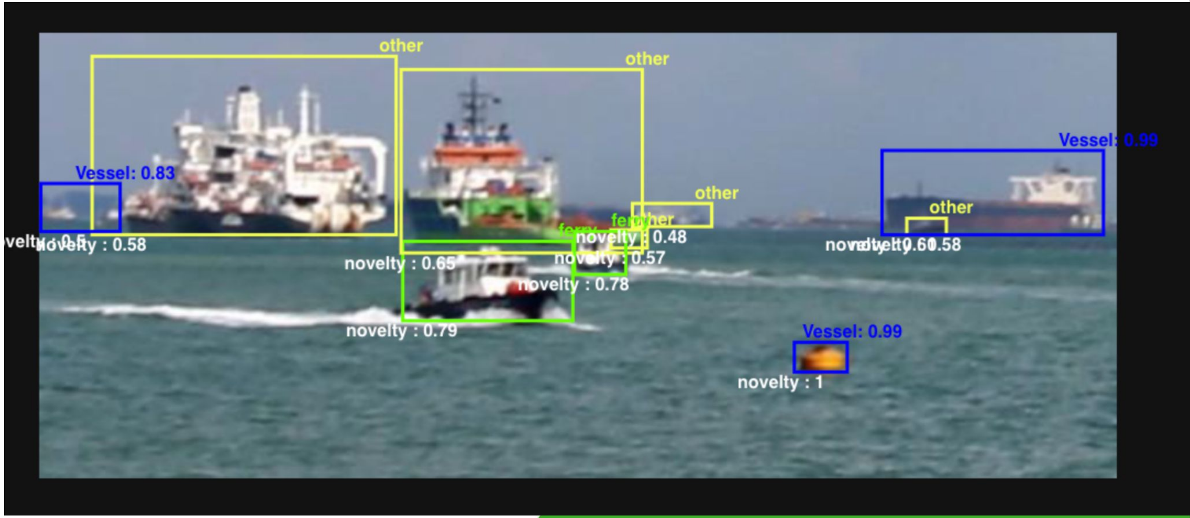
Singapore Maritime Dataset

Dilip K. Prasad, Deepu Rajan, Lily Rachmawati, Eshan Rajabally, and Chai Quek. “Video Processing from Electro-optical Sensors for Object Detection and Tracking in Maritime Environment: A Survey.” IEEE Tranactions on Intelligent Transportation Systems, 2017.

- **On-shore and on-board 1080p videos, both RGB and NIR**
- **81 videos; 20 minutes; 5.4GB**
  - 11 onboard cameras, 40 onshore, 30 NIR
  - Our experiments focused on the onshore RGB data
- **9 categories; 5-200 examples per class; long-tailed (“vessel” has 10X more than any other class)**

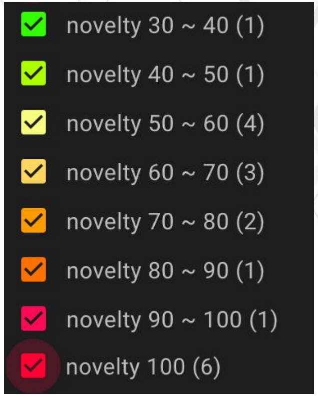
Anomaly Detection Results

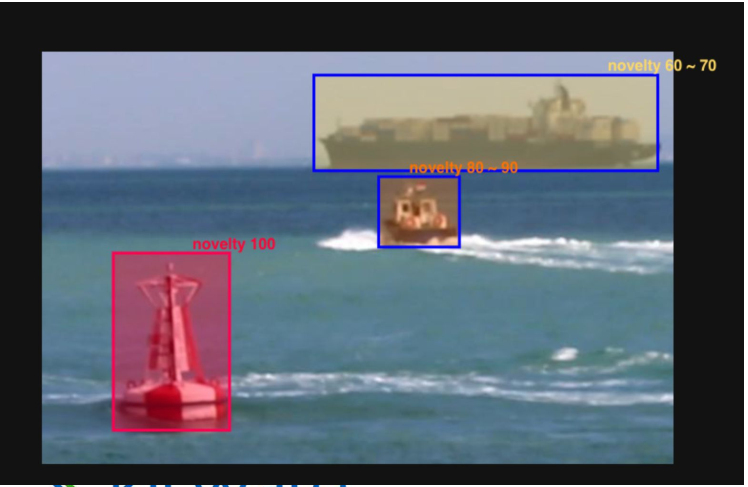
- Train detector and EVM on images with 8 known classes, but **not buoys**
- Test on images with buoys
- Instances of known classes should be detected and classified
- Instances of novel classes should be predicted as novel
  - Buoys and other novelties



Anomaly Detection Results

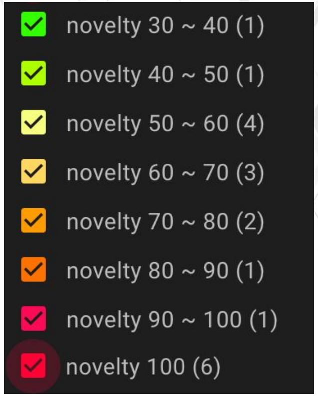
- Color-coded by degree of novelty
- Most detections have some degree of novelty





Anomaly Detection Results

- Problematic image - most objects are given high novelty even when they are known classes



Satellite Imagery Results

Standard detection/classification | **Novelty Detection**

The bus is detected and classified as a Small_Car. Our system correctly assigned a high novelty score, indicating it is a new variant of Small_Car or a novel vehicle type.

Satellite Imagery Quantitative Results

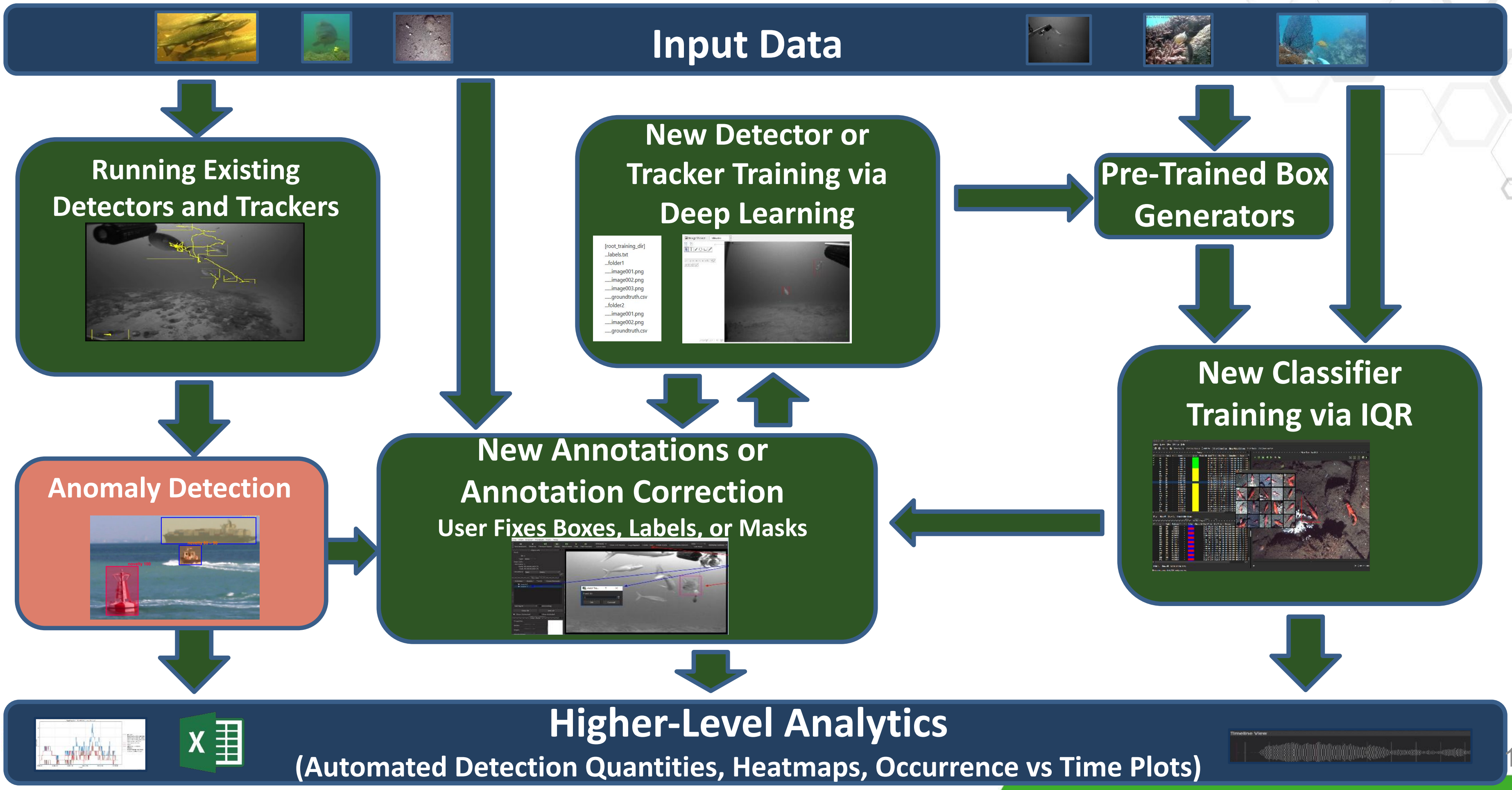
We compare the performance of traditional Softmax thresholding and our EVM on novelty detection on the xView Dataset. 42 known classes, 18 novel classes.

Method	AUC	AP	F1	mean known score	mean unknown score
Softmax	64.62%	6.55%	12.12%	0.534	0.601
EVM	**77.45%**	**12.55%**	**27.75%**	0.930	0.977

Softmax Thresholding: detection is novel if max class probability < T, over all known classes. The AUC is higher than F1 and AP because AUC uses TN in FPR balancing FPs from all known classes. TN is not used by F1 and AP.

~1500 novel instances
~15,000 known instances
~90% of pixels are background (no ground truth label)

Anomaly/Novelty Detection in VIAME



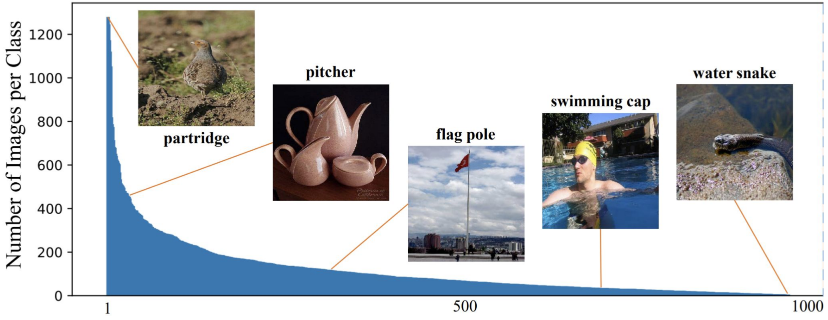
Novelty Detection Publications

- Dawei Du, Christopher Funk, Katarina Doctor, and Anthony Hoogs. “Novel Object Detection in Remote Sensing Imagery.” IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 2023.
- Dawei Du, Christopher Funk, Anthony Hoogs. “Novelty Detection in Remote Sensing Imagery.” IEEE International Geoscience and Remote Sensing Symposium, Kuala Lumpur, Malaysia, 2022.
- Dawei Du, Ameya Shringi, Anthony Hoogs, Christopher Funk. “Reconstructing Humpty Dumpty: Multi-feature Graph Autoencoder for Open Set Action Recognition.” Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2023, pp. 3371-3380.
- Chen Zhao, Dawei Du, Anthony Hoogs, Christopher Funk. “Open Set Action Recognition via Multi-Label Evidential Learning.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Contract No. HR001120C0055 and 140D0422C0043. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Defense Advanced Research Projects Agency (DARPA); or its Contracting Agent, the U.S. Department of the Interior, Interior Business Center, Acquisition Services Directorate, Division V.

Learning Detectors from Very Few Examples

Many problems have long-tailed distributions in the number of examples per class.



How do we build effective detectors for classes in the tail, with very few training examples?

Leverage lots of unlabeled data via unsupervised pre-training

- Same technique used by LLMs
- Developed on DARPA Learning with Less Labeling program
- Poster tomorrow, #1633

Our low-shot methods are now available in VIAME

kitware

Conclusion

- VIAME is an open-source toolkit for scientists to create their own AI solutions to unique problems, with no programming or AI expertise
  - Open source, no license fee, highly permissive license
- Anomaly/novelty detection is now available in VIAME, based on prior detection models for particular domains
  - All functionality is accessed through the web GUI without programming or scripting, including model training and novelty results display
  - Applicable to any domain with an object detection/classification application
  - Find rare species or variants automatically
  - Discover unknown unknowns!

Questions?

- Contact: anthony.hoogs@kitware.com
- For more information and discussions, come to our related poster:

OT44B-1633 Creating Deep Learning Detectors in VIAME for Rare Objects in Marine Imagery

Panel 1633

In session OT44B - Combining Underwater Imaging with Deep Learning for Better Ocean Observations

Backup

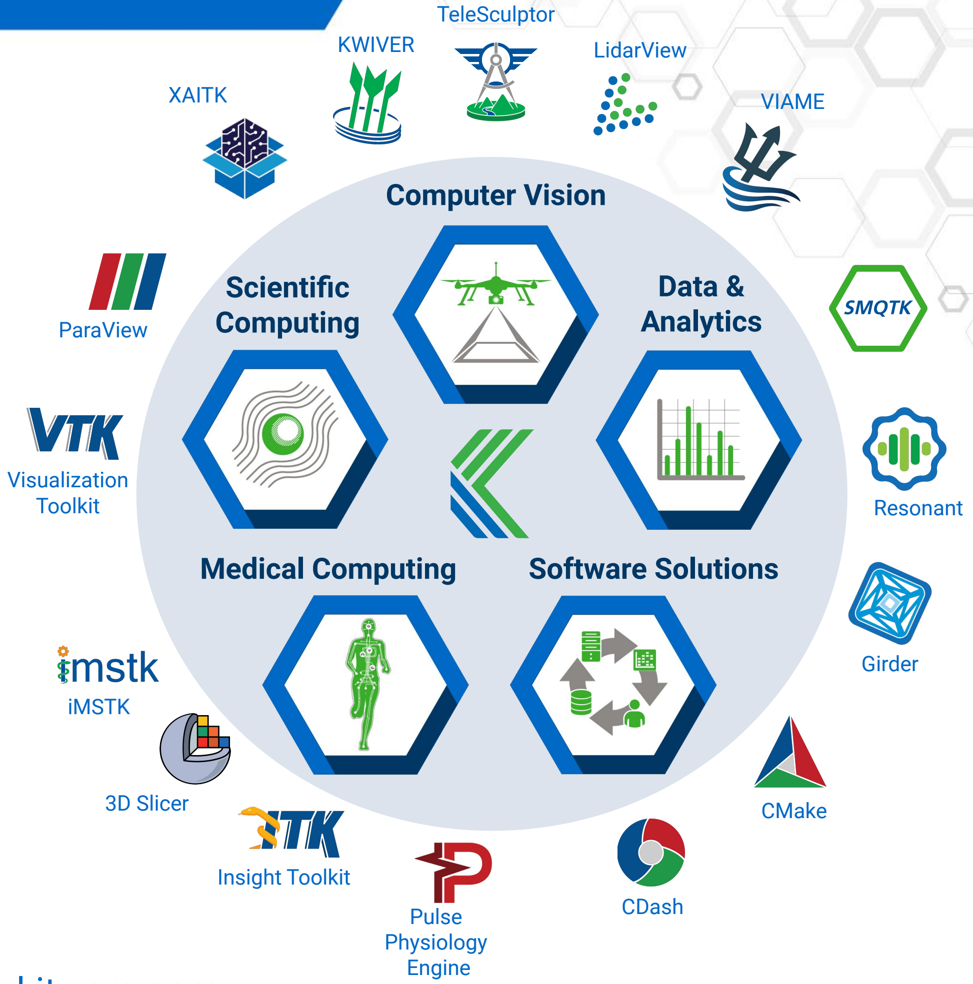
Kitware Company Overview

- Open-source software R&D: algorithms & applications, image & data analysis, training data, integration, & testing
- 200+ employees: ⅓ PhD, ⅓ masters
- Offices: Albany, NY; Chapel Hill, NC; Arlington, VA; Minneapolis, MN; Lyon, France
- Secure facility: Albany, NY; 44+ cleared personnel, TS+ clearances

Commercial and Government R&D Services

- Commercial: 10% of Revenue
  - Multiple purchasing, contracting and licensing mechanisms
  - Commercial business models to suit commercial business needs
- Government: 90% of Revenue
  - Wide range of federal customers including DoD, IC, DOE, NIH, NOAA
  - All government-funded software is provided with **unlimited rights** to the government
  - Software released as **open source** when permitted

100% Employee Owned

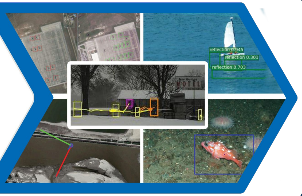


kitware.com

Computer Vision at Kitware

www.kitware.com/expertise/#computer-vision

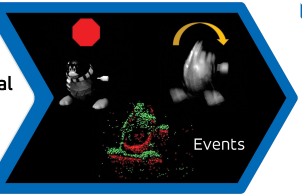
Deep Learning



Dataset Collection and Annotation

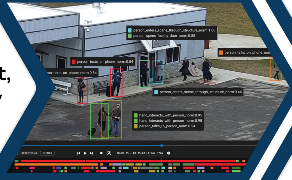
Cyber Physical Systems (CPS)

Computational Imaging

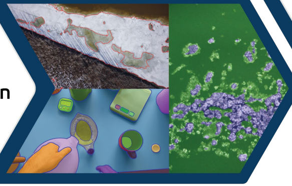


Object Detection and Tracking

Complex Event, Threat, and Activity Detection



Semantic Segmentation



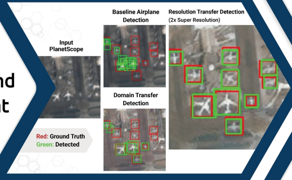
kitware

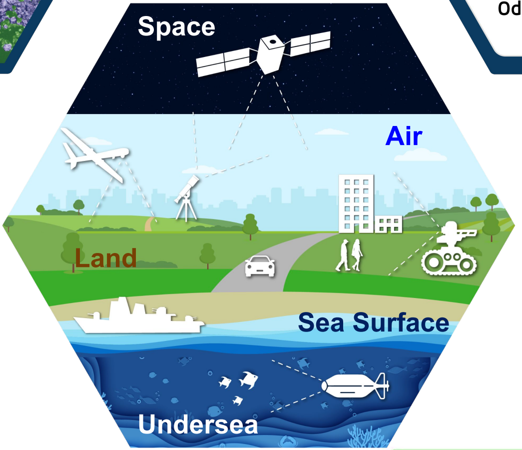
Delivering Innovation

70+ team members
24 PhDs
40+ cleared
40+ active projects
Vision-ary since 2007

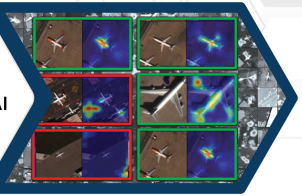
Image and Video Forensics

Super Resolution and Enhancement

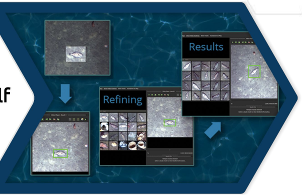




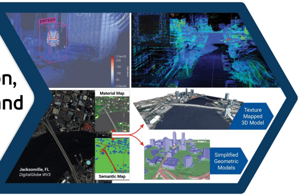
Explainable and Ethical AI



Interactive Do-It-Yourself AI



3D Reconstruction, Point Clouds, and Odometry

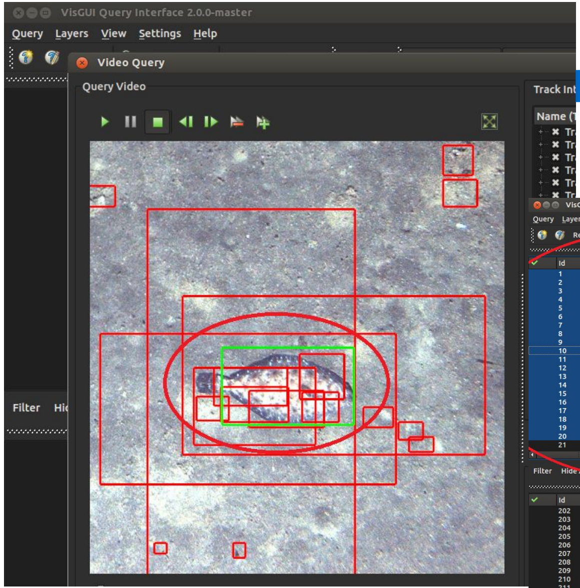


Open Source Toolkits

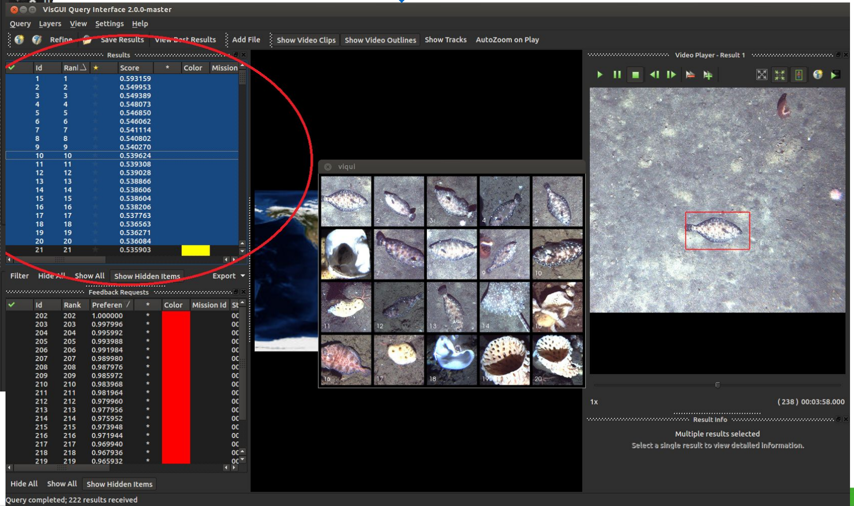
WATCH
VIAME
XAITK
KWIVER
LidarView
TeleSculptor
SMQTK

Iterative Query Refinement for Rapid Detector Generation

- VIAME supports multiple workflows for interactive detector generation
- In Iterative Query Refinement, starting with a single exemplar, the user adjudicates system matches. Feedback is used to create a model instantly, and improved results are adjudicated until the user is satisfied.

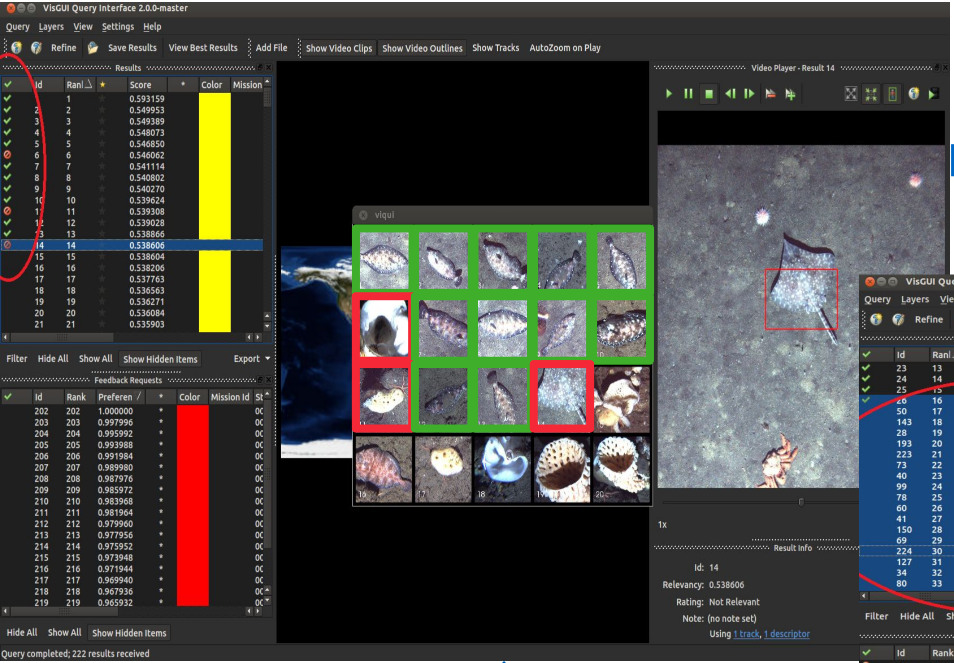


Select target exemplars; generate initial result list

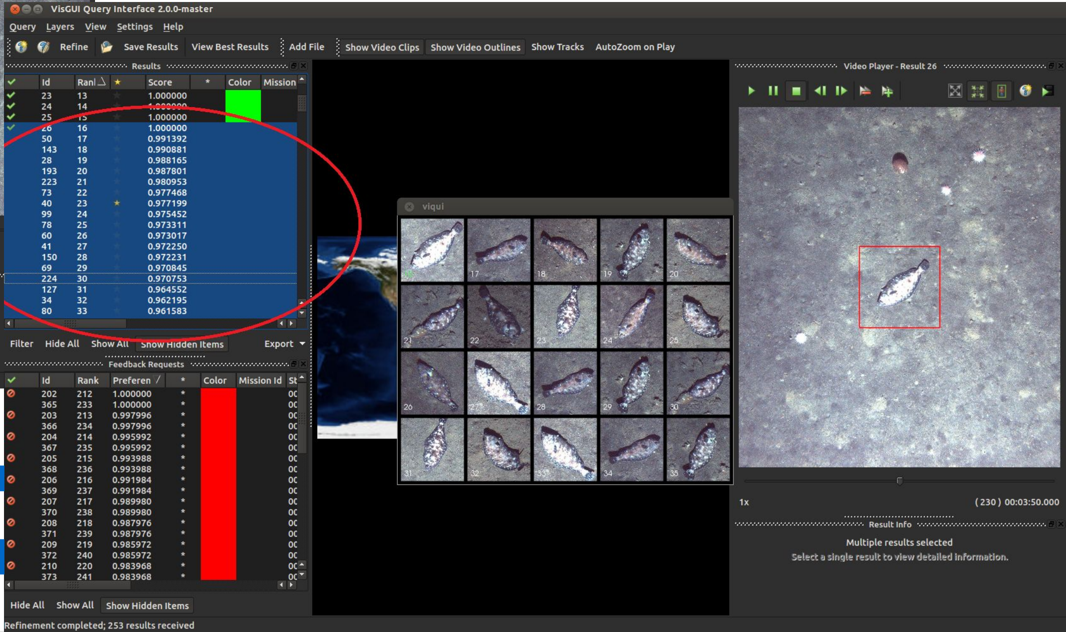


Iterative Query Refinement for Rapid Detector Generation

User adjudicates the result list



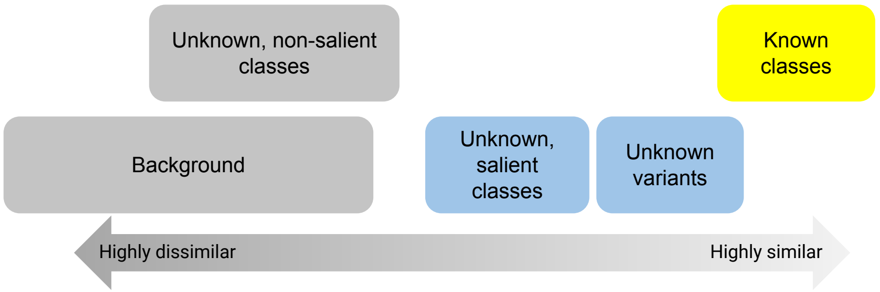
Create/update model in a few seconds; generate improved result list



Target detector

Novel Contributions

- Our approach finds **salient, novel objects** by leveraging an object detector trained on mission-relevant objects. Our novelties are likely to be salient because they are:
  - Similar to known classes
  - Different from the background and non-salient classes
- Our method is theoretically grounded in Extreme Value Theory [1], which provides a statistically-valid dissimilarity score for distinguishing between known classes, novelties and background
  - Other methods simply threshold the class probabilities, or perform logistic regression which requires training on known background data
- We generate a novelty score for each detection, enabling the analyst to set the operating point for novelty detection/filtering
  - Other methods operate on the entire image (image novelty score)

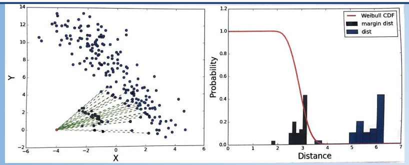


[1] Rudd, Ethan M., Lalit P. Jain, Walter J. Scheirer, and Terrance E. Boult. "The extreme value machine." IEEE transactions on pattern analysis and machine intelligence 40, no. 3 (2017): 762-768.

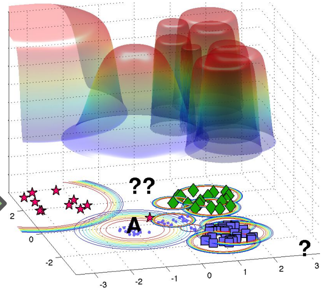
Novel Object Classification

We leverage the Extreme Value Machine (EVM), a model which we derive from statistical extreme value theory (EVT).

EVM Training



Feature Extraction → All Pairwise Dissimilarity → Weibull fit to K nearest "negatives" → **Domain Independent** Set-cover to keep only "Extreme Vectors" →



EVM Testing

Feature Extraction → Dissimilarity to Extreme vectors → Evaluate Weibulls, p = Max probability over all classes → **Domain Independent** Prob Unknown = 1-p → **Novelty Score** K+1 probabilities

[1] Rudd, Ethan M., Lalit P. Jain, Walter J. Scheirer, and Terrance E. Boult. "The extreme value machine." IEEE transactions on pattern analysis and machine intelligence 40, no. 3 (2017): 762-768.

kitware

Dataset and Evaluation

xView [1] is a dataset of 700+ RGB satellite images fully labeled with bounding boxes for 60 classes.

We split xView into a known set (42 classes) and unknown set (18 classes). Unknown classes are distributed across the 8 high-level categories to avoid simplifying the problem.

Instances of unknown classes should be detected as novelties.

***** Dataset statistics *****
- # classes: 42/18
- # train: 28,517
- # test: 16,244

Fixed-Wing Aircraft	Passenger Vehicle	Truck	Railway Vehicle	Maritime Vessel	Engineering Vehicle	Building	None
Small Aircraft	Small Car	Pickup Truck	Passenger Car	Motoboat	Tower Crane	Hut/Tent	Helipad
Cargo Plane	**Bus**	Utility Truck	Cargo Car	Sailboat	Container Crane	Shed	Pylon
		Cargo Truck	Flat Car	Tugboat	Reach Stacker	Aircraft Hangar	Shipping Container
		Truck w/Box	**Tank Car**	Barge	Straddle Carrier	**Damaged Building**	Shipping Container Lot
		Truck Tractor	**Locomotive**	Fishing Vessel	Mobile Crane	**Facility**	Storage Tank
		Trailer		Ferry	Dump Truck		Vehicle Lot
		Truck w/Flatbed		Yacht	Haul Truck		Construction Site
		Truck w/Liquid		**Container Ship**	Scraper/Tractor		**Tower Structure**
				Oil Tanker	Front Loader		**Helicopter**
					Excavator		
					Cement Mixer		
					Ground Grader		
					Crane Truck		

Table 1. The class two-level hierarchy from the xView dataset [2]. The parent classes are listed at the top and child classes are below them. The 'None' parent class is for classes without a parent. The **bold classes** are the ones we consider to be unknown in our experiments. The other classes classes are considered known classes.

[1] Darius Lam, Richard Kuzma, Kevin McGee, Samuel Dooley, Michael Laielli, Matthew Klaric, Yaroslav Bulatov, and Brendan McCord. "xview: Objects in context in overhead imagery." arXiv preprint arXiv:1802.07856 (2018).

Evaluation Metrics

Novel object detection poses unique challenges requiring modifications to standard object detection metrics. It is a 3-class problem: novel objects, known objects, and background objects. Novel objects are similar to the known classes, which differentiates them from background. However, the object detector may find background detections for which we have no ground truth. We consider such detections as (unannotated) background, rather than novelties, and do not include them in our metrics.

True Positives (TP) are novel objects correctly detected as novel.
False Negatives (FN) are novel objects incorrectly detected as a known class.
False Positives (FP) are known objects incorrectly detected as novel.
True Negatives (TN) are known objects correctly detected as a known class.

For the binary novelty detection problem (whether it is a novel or known object), we use standard detection metrics: Area under the ROC Curve (AUC), Average Precision (AP), F1. The ROC curve is a plot of TPR vs. FPR. The F1 score is the Max F1 for all novelty thresholds for each algorithm. We do not include any non-detected, background pixels in our metrics, although these would be included in a spatially-grounded metric such as false alarms per km^2. We do not include misclassification errors (known objects detected as the wrong known class) in these metrics, but standard metrics for these can be computed independently.

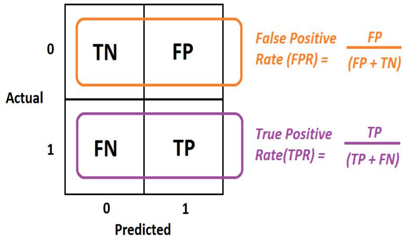
We measure the per-class recall/TPR for novel classes (the unknown set). To measure per-class novelty recall/TPR, i.e. the accuracy of detecting instances of a specific novel class, we do not include correct novel detections from other classes nor incorrect detected known classes. Incorrectly-detected, known classes have no correspondence to any novel class, distorting both False Positives and True Negatives. Without a specific novel class association, the global FP and TN would have to apply to each novel class. This would penalize the novel classes with fewer instances since these have fewer possible True Positives to balance out FP or TN.

$$TPR = \frac{TP}{Actual\ Positive} = \frac{TP}{TP + FN}$$

$$FNR = \frac{FN}{Actual\ Positive} = \frac{FN}{TP + FN}$$

$$TNR = \frac{TN}{Actual\ Negative} = \frac{TN}{TN + FP}$$

$$FPR = \frac{FP}{Actual\ Negative} = \frac{FP}{TN + FP}$$



$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

TP = True positive
TN = True negative
FP = False positive
FN = False negative

Quantitative Results Per Class

Performance of our EVM approach vs. traditional Softmax thresholding per novel class. Ideal score for each measure is 100%. Bold indicates best score.

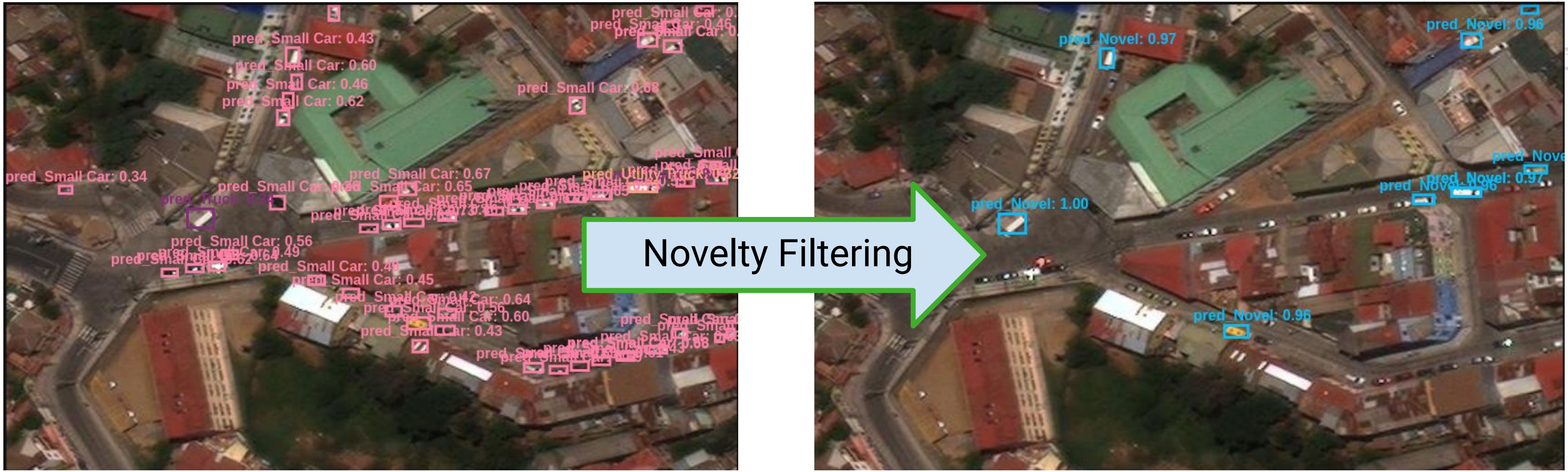
Class	Cargo Plane	Helicopter	Bus	Cargo Truck	Trailer	Truck w/Flatbed	Truck w/Liquid	Tank Car (rail)	Locomotive	Container Ship	Oil Tanker	Excavator	Cement Mixer	Ground Grader	Damaged Building	Facility	Construction Site	Tower
Recall	16.16	0.00	**92.57**	78.98	**74.83**	82.86	100	0.00	14.29	8.89	4.00	**66.67**	**100**	50.00	**100**	3.45	50.00	**100**
	97.98	0.00	86.14	**87.50**	37.09	82.86	100	0.00	**42.86**	**42.22**	**12.00**	8.33	66.67	**75.00**	50.00	**13.79**	50.00	33.33
Count	140	8	260	212	281	58	2	18	10	97	49	78	6	15	13	100	88	19

Only Recall is computed because FP and TN are invalid for individual novel classes.

Two classes, Helicopter and Tank Car (rail), have 0% accuracy because they are not detected, probably because they are far away from any known class.

Without novelty detection, an analyst must examine all detections to find unknown unknowns.

With novelty detection, hypothesized unknown unknowns are easily highlighted.



Novelty detection computes a novelty score for each object detection, enabling the analyst to interactively filter detections based on level of novelty.

Each detected object has a classification probability for some known class, and a novelty score

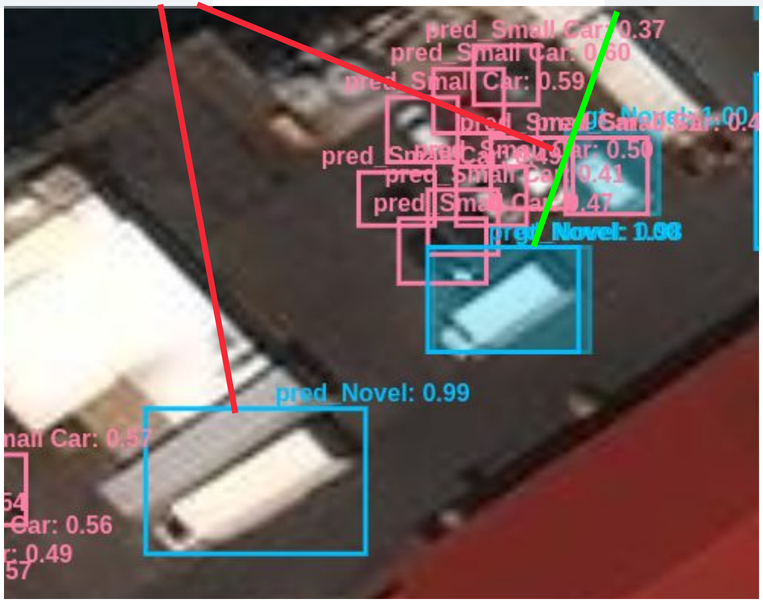
In this case, the bus is detected and classified as a Small_Car. However our system correctly assigned a high novelty score, indicating it is a new variant of Small_Car or a novel vehicle type.

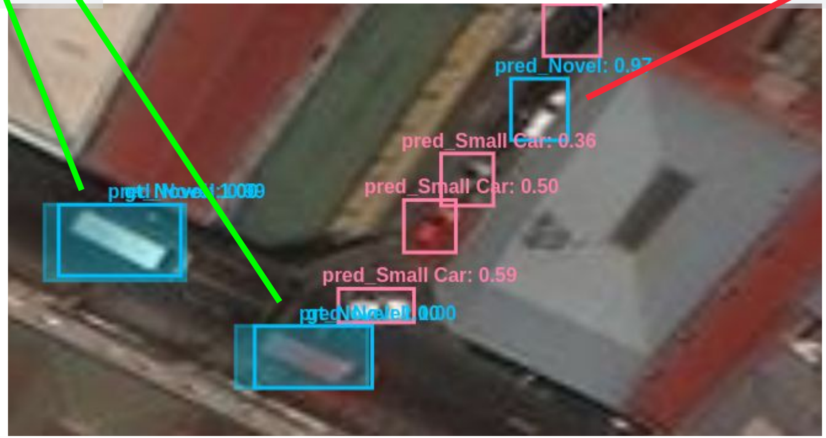
Results Assessment against Ground Truth

Incorrect

Correct

Incorrect





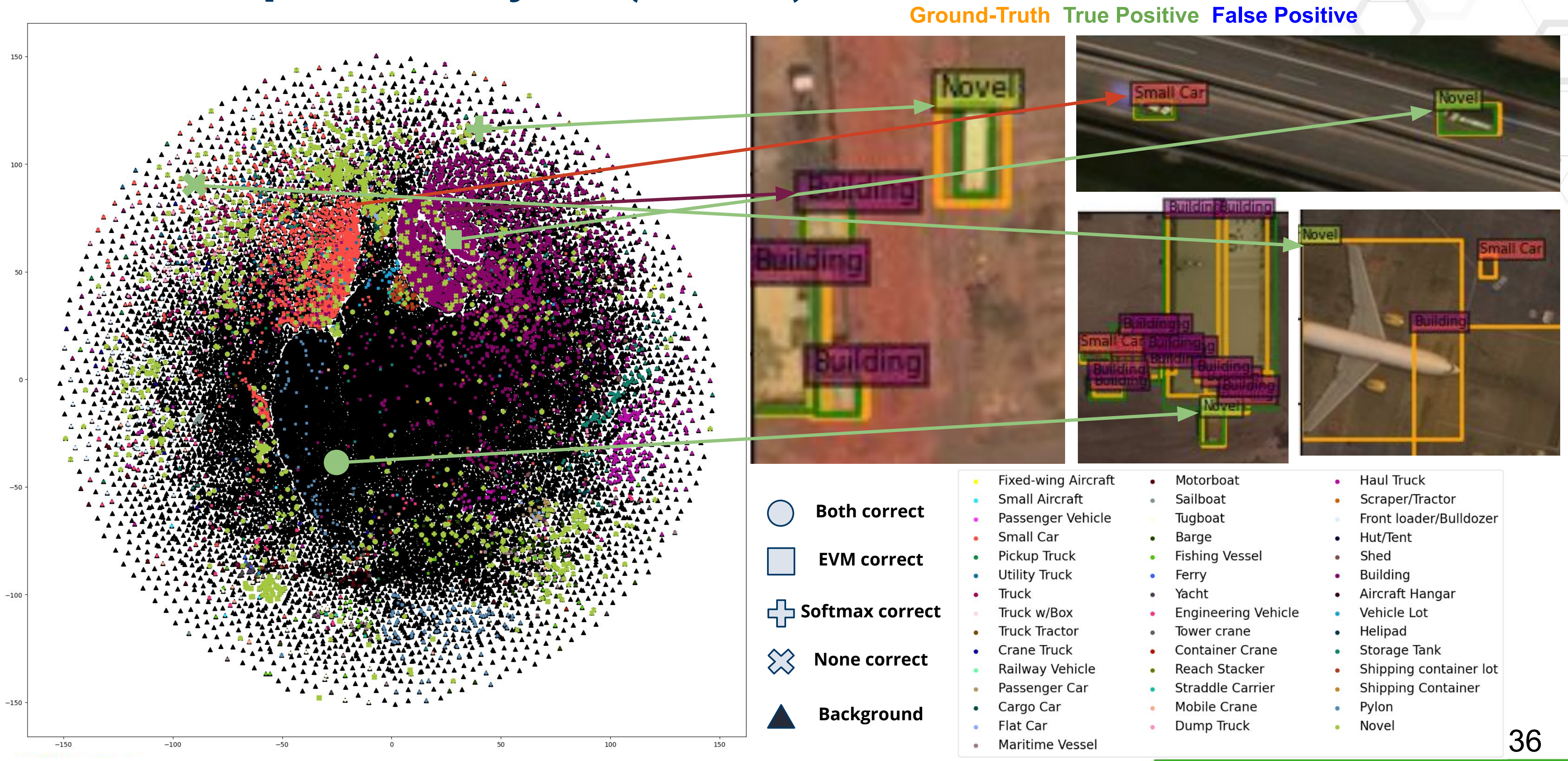
The novelty error should be Truck w/Box. Difficult case because of similarity to correct novelty on the right.

Legend

- Detected Novelty
- Ground Truth Novelty

Our approach detects novelty across a wide range of objects and scenes

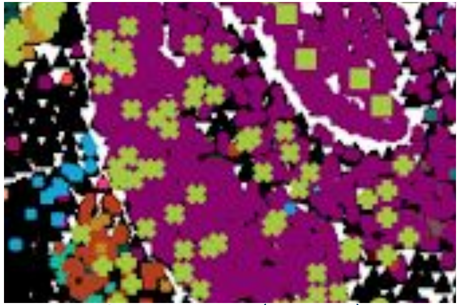
Feature Space Analysis (T-SNE)



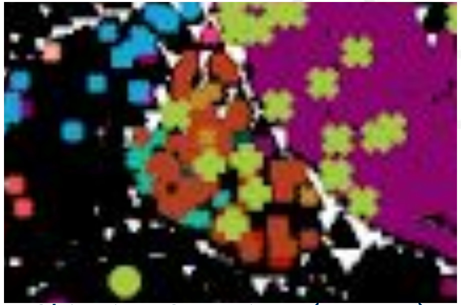
Feature Space Analysis (T-SNE)

- Both correct
- EVM correct
- Softmax correct
- None correct
- Background

Storage container (forest green) near haul trucks (purple) and pink dump trucks missed by both

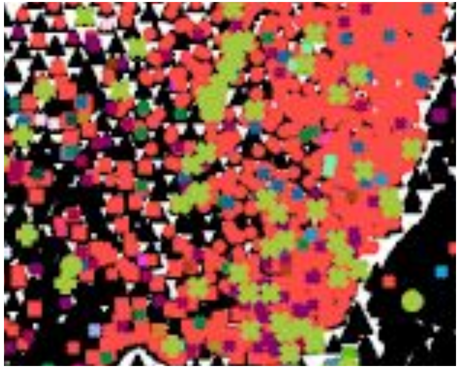


Cluster of buildings (purple) clearly separated from other classes and some overlapping novel objects (green)



Shipping Container (orange) surrounded by background

Mixture of bulldozers (light blue) and novel (green) overlapping background



Cars (orange) and some building both labeled were known

- Fixed-wing Aircraft
- Small Aircraft
- Passenger Vehicle
- Small Car
- Pickup Truck
- Utility Truck
- Truck
- Truck w/Box
- Truck Tractor
- Crane Truck
- Railway Vehicle
- Passenger Car
- Cargo Car
- Flat Car
- Maritime Vessel
- Motorboat
- Sailboat
- Tugboat
- Barge
- Fishing Vessel
- Ferry
- Yacht
- Engineering Vehicle
- Tower crane
- Container Crane
- Reach Stacker
- Straddle Carrier
- Mobile Crane
- Dump Truck
- Haul Truck
- Scraper/Tractor
- Front loader/Bulldozer
- Hut/Tent
- Shed
- Building
- Aircraft Hangar
- Vehicle Lot
- Helipad
- Storage Tank
- Shipping container lot
- Shipping Container
- Pylon
- Novel