RESEARCH REVIEW 2019

Video Summarization and Search

Dr. Rachel Brower-Sinning
Adam Harley
Dr. Jeffery Hansen
Jacob Ratzlaff
Chris Grabowski
Professor Katerina Fragkiadaki
Ed Morris

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Copyright 2019 Carnegie Mellon University.

This material is based upon work funded and supported by the Department of Defense under Contract No. FA8702-15-D-0002 with Carnegie Mellon University for the operation of the Software Engineering Institute, a federally funded research and development center.

The view, opinions, and/or findings contained in this material are those of the author(s) and should not be construed as an official Government position, policy, or decision, unless designated by other documentation.

NO WARRANTY. THIS CARNEGIE MELLON UNIVERSITY AND SOFTWARE ENGINEERING INSTITUTE MATERIAL IS FURNISHED ON AN "AS-IS" BASIS. CARNEGIE MELLON UNIVERSITY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. CARNEGIE MELLON UNIVERSITY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

[DISTRIBUTION STATEMENT A] This material has been approved for public release and unlimited distribution. Please see Copyright notice for non-US Government use and distribution.

This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other use. Requests for permission should be directed to the Software Engineering Institute at permission@sei.cmu.edu.

Carnegie Mellon® is registered in the U.S. Patent and Trademark Office by Carnegie Mellon University.

DM19-1116

The goal of video summarization and search is to assist analysts by

- increasing volume of data that can be analyzed
- providing insights into patterns of life that would be difficult to otherwise identify

© 2019 Carnegie Mellon University

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

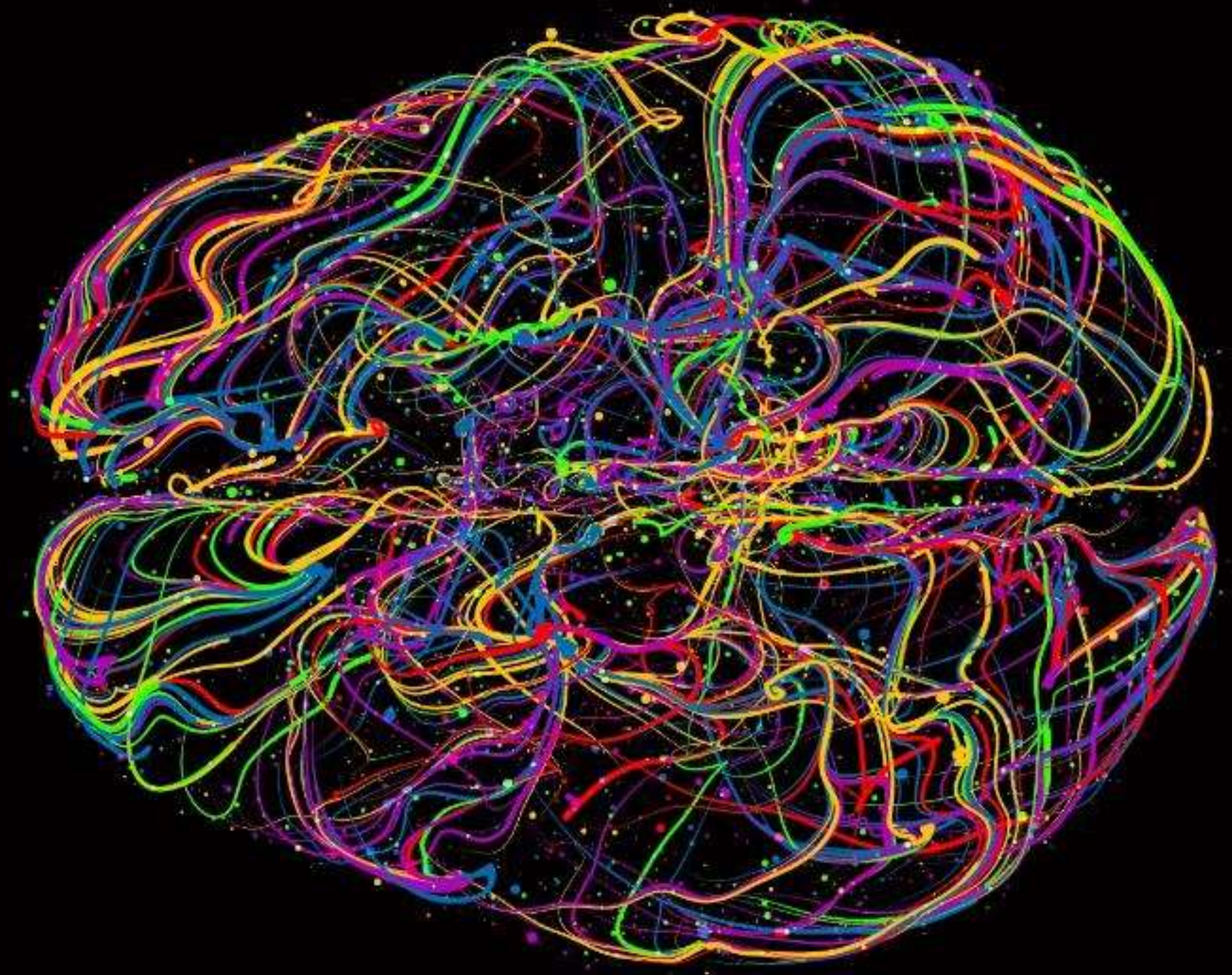
Three core technologies to make video summarization and search possible in the DoD problem space:

- Domain adaptation to address limitations of ***training data*** to improve object identification
- Geometry-aware visual surveillance to improve ***tracking*** of moving objects
- Pattern-of-life analysis to ***characterize*** relationships and behaviors of those objects

© 2019 Carnegie Mellon University
[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Video Summarization and Search:

Domain Adaptation



Jacob Ratzlaff
Chris Grabowski
Dr. Rachel Brower-Sinning

© 2019 Carnegie Mellon University

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Object Identification in the DoD Problem Space

Solving machine learning problems starts with having sufficient data

- DoD datasets have limitations:
  - May not reflect future mission environments
  - Often unlabeled
  - Few examples of the target object
- DoD datasets and public datasets are not good analogs:
  - Semantic differences
    - Perspective
    - Scale
    - Types of objects
    - Density of objects
  - Texture differences
    - Non-standard environments
    - Diverse objects

Image reference: *Vision Meets Drones: A Challenge*. Zhu, Pengfei; Wen, Longyin; Bian, Xiao; Haibin, Ling; and Hu, Qinghua. arXiv. 2018

© 2019 Carnegie Mellon University

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Our Approach: Domain Adaptation

Use existing datasets to train a classifier for our new target domain

- Semantic information in the metadata of images is used to create a matched subset
  - Currently we only match on object density
- To match texture, we use CycleGAN
  - CycleGAN uses machine learning to transform a set of images to appear more similar to another set of images

Train Mask R-CNN classifiers using a combination of

- Original dataset
- Small set of target data
- Adapted dataset

© 2019 Carnegie Mellon University

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Influence of Training Data on Object Identification

Poor Object Identification

- Using pretrained MSCOCO weights
- Fails to classify many objects, likely due to training on ground-level perspective
- Major problems with misclassifying building components as vehicles/personnel

Improved Object Identification

- Using a combination of adapted and target domain datasets for training
- Significantly more objects are correctly classified
- Minor problems with misclassifying building components as vehicles/personnel

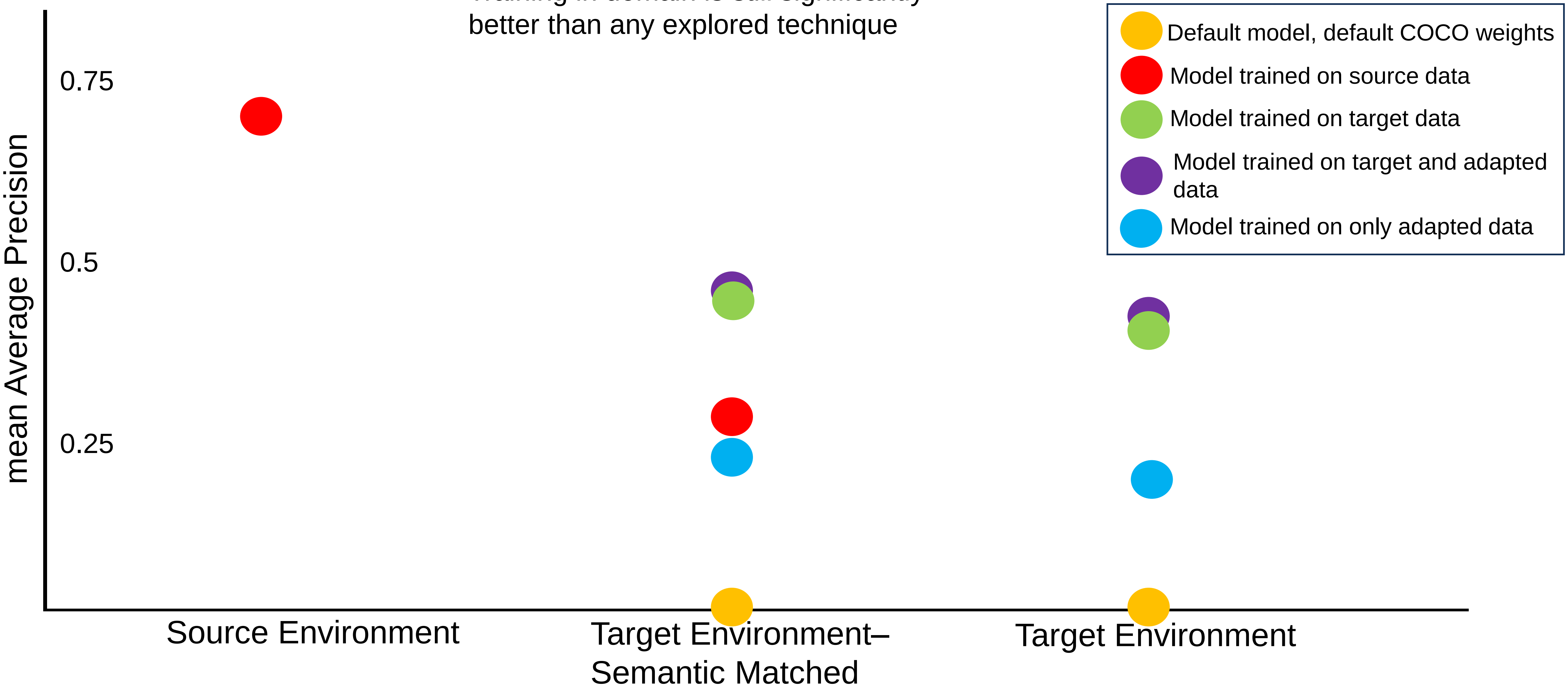
Image reference: *Vision Meets Drones: A Challenge*. Zhu, Pengfei; Wen, Longyin; Bian, Xiao; Haibin, Ling; and Hu, Qinghua. arXiv. 2018

© 2019 Carnegie Mellon University

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Results

- Poor performance with a pretrained model
- Incremental performance increases with adapted data
- Training in domain is still significantly better than any explored technique



© 2019 Carnegie Mellon University

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Future Directions

Current results may be improved by additional matching of target and source image semantics:

- We match only on object density. We will be extending this to include other image features such as perspective and scale.
- After improved semantic matching, can addition of supplemental data increase performance or act as a substitute for target data?

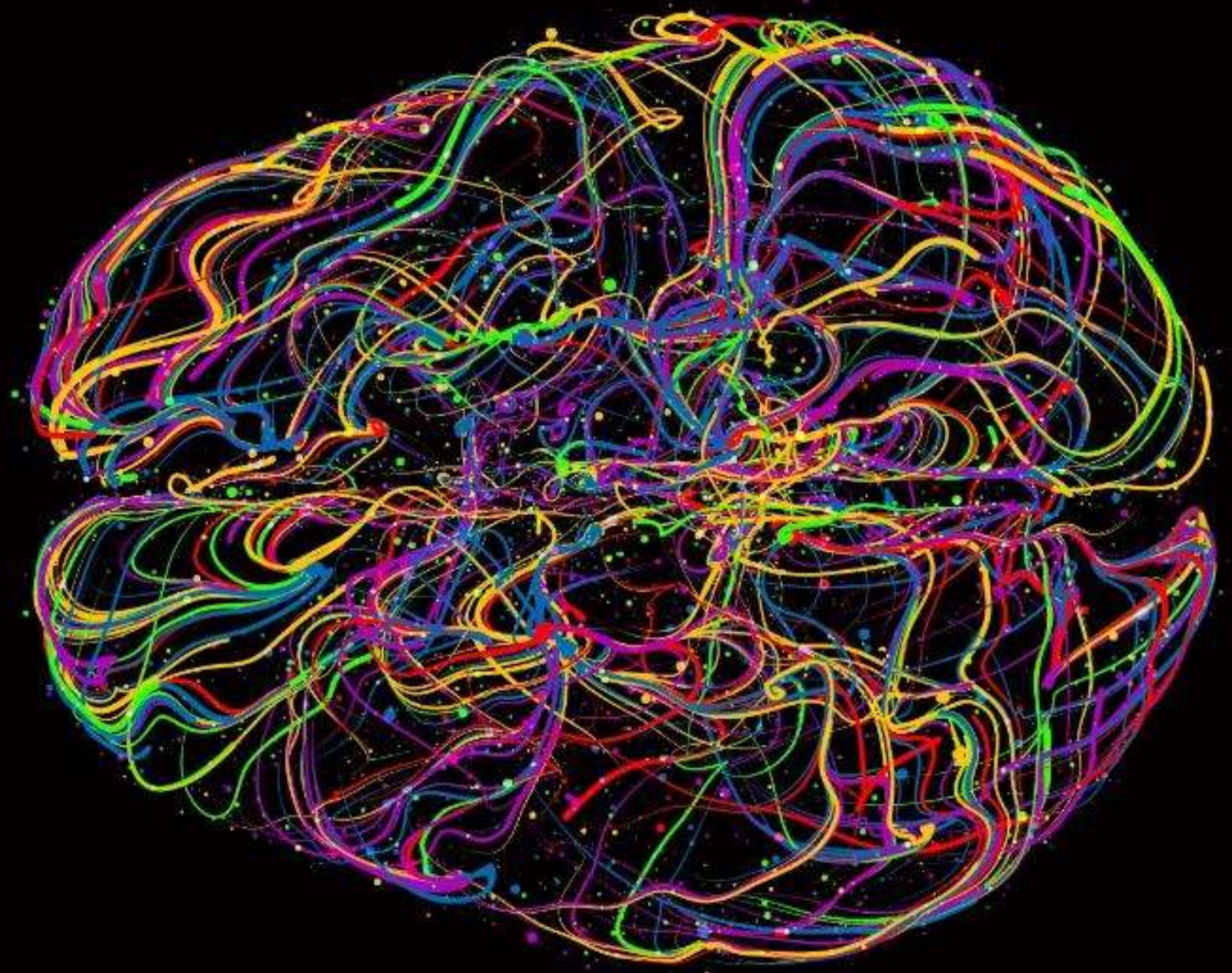
Exploration of feature statistics indicated that automatic determination of object identification quality may be possible.

© 2019 Carnegie Mellon University

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Video Summarization and Search:

Geometry-Aware Visual Surveillance

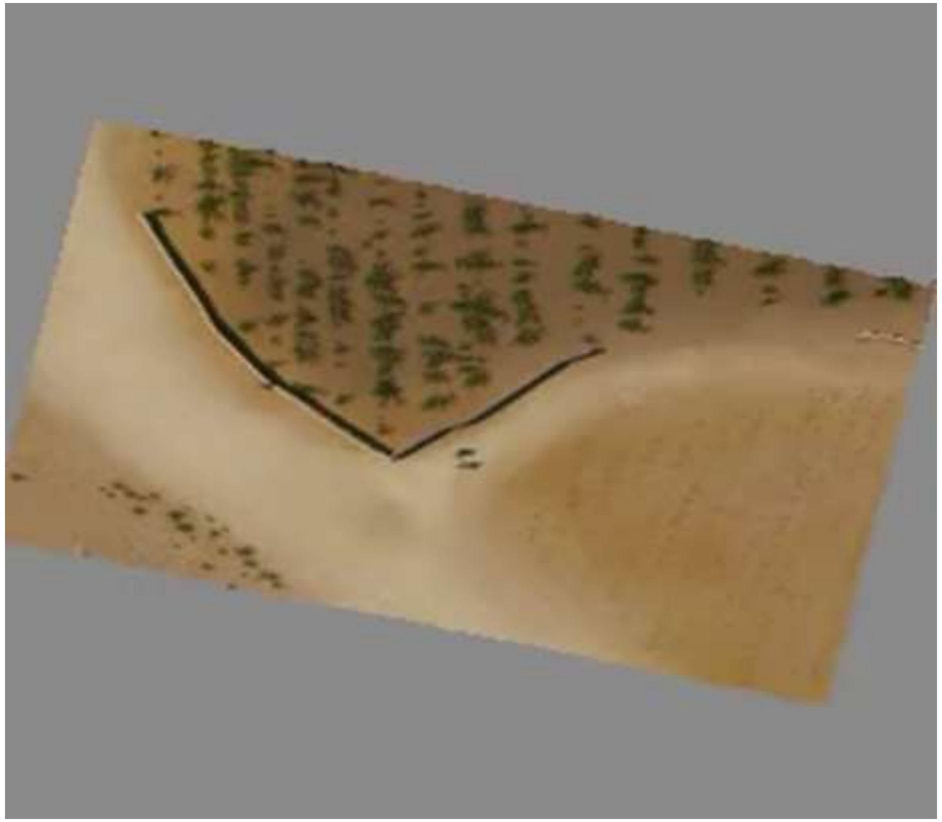


Adam Harley
Professor Katerina Fragkiadaki

© 2019 Carnegie Mellon University

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Egomotion-Stabilized Perception

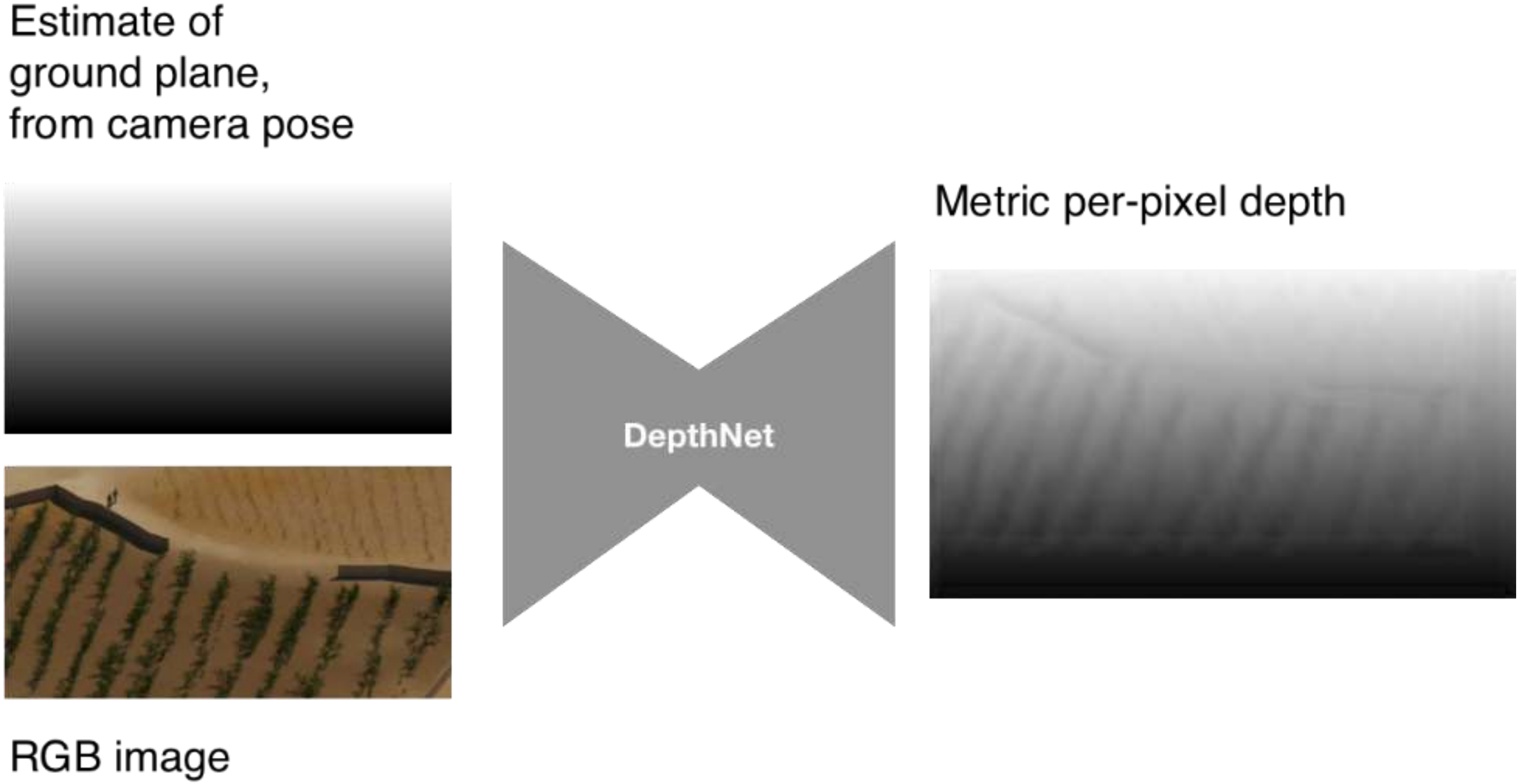


Using the relative camera motion information from the drone

- Learn a depth map of the scene
- Use this depth map to project pixels from a virtual overhead camera

© 2019 Carnegie Mellon University
[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

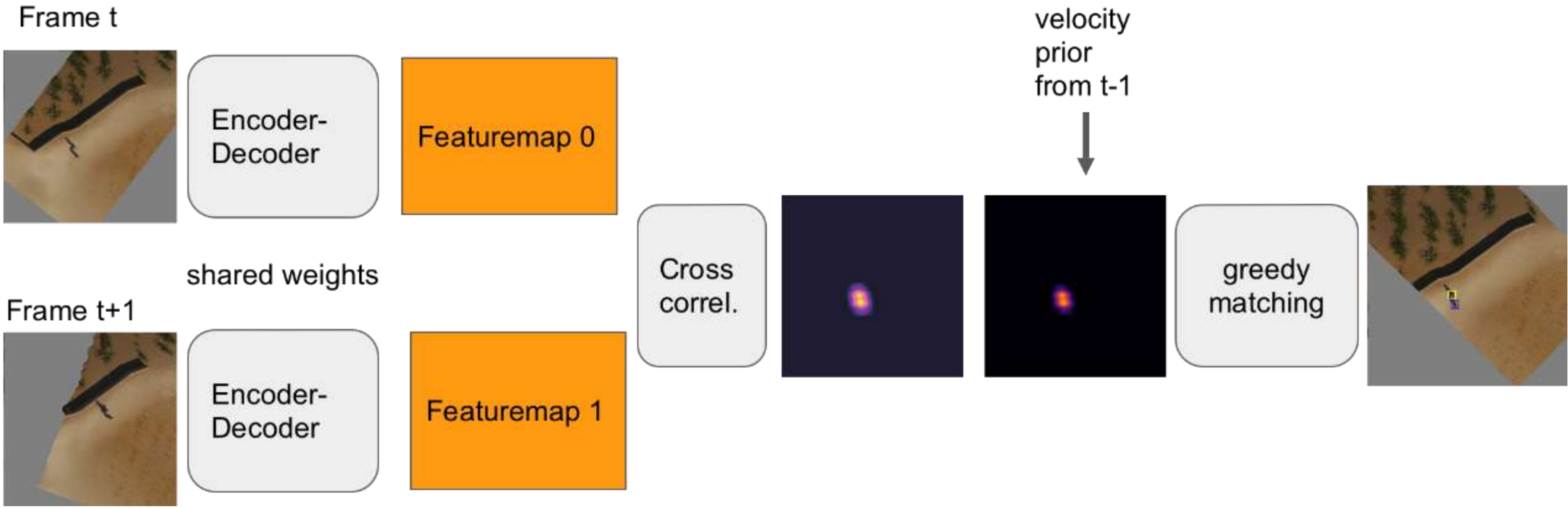
Learning Depth from Telemetry



© 2019 Carnegie Mellon University

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Tracking Model: Template Matching in the Stabilized Map

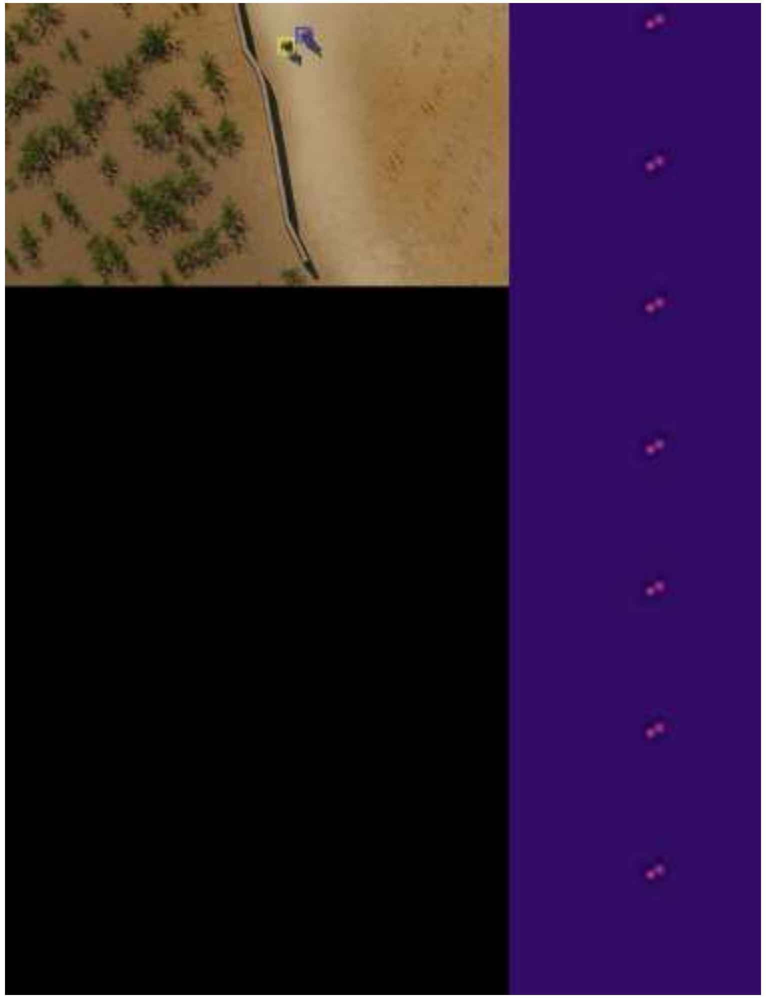


© 2019 Carnegie Mellon University

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Tracking Visualization

Unstabilized



Stabilized

© 2019 Carnegie Mellon University
[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Stabilization Helps

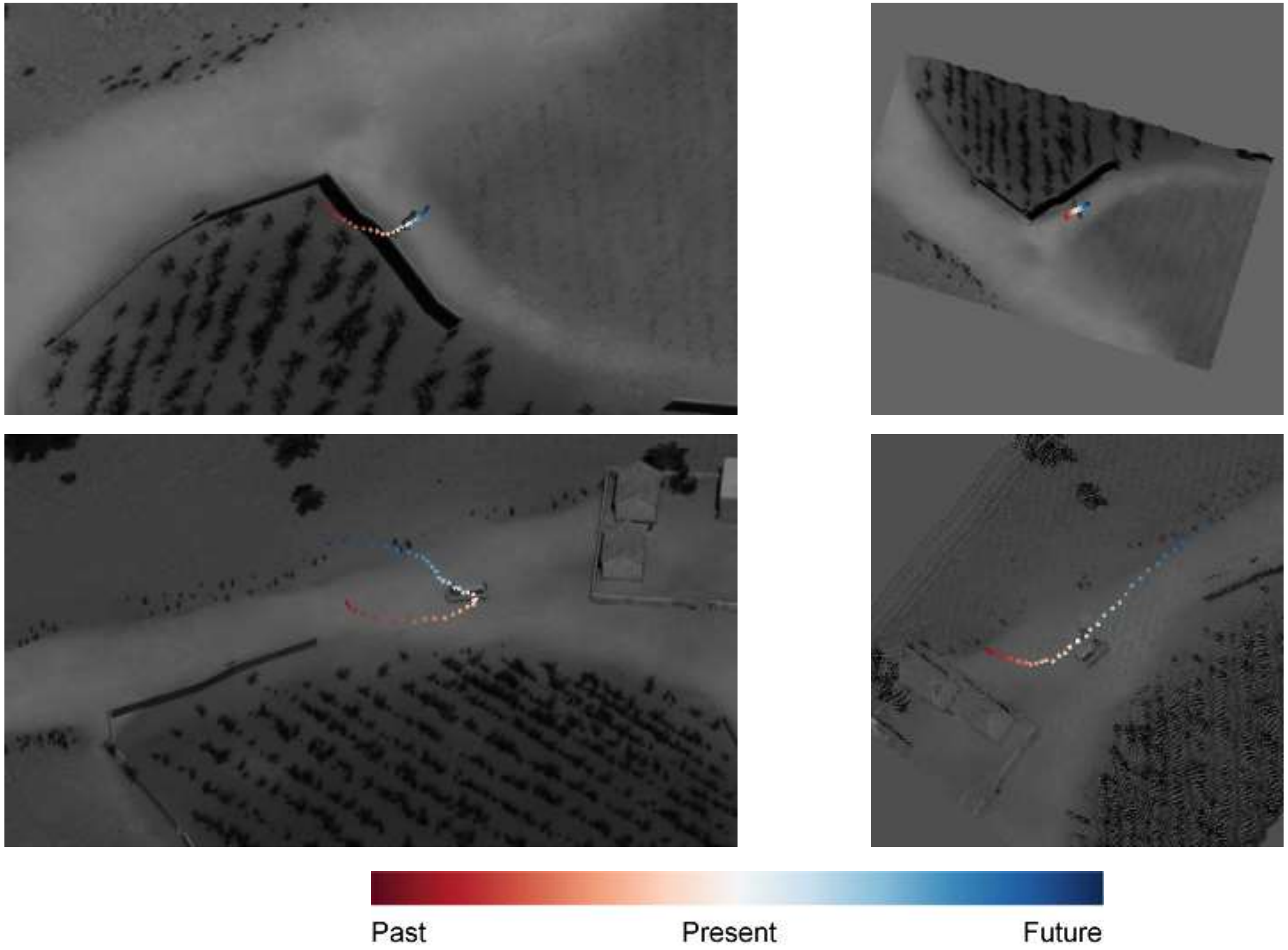
Method	Mean IOU
Unstabilized	0.5185
Unstabilized+DataAug	0.5567
Unstabilized+DataAug+Hungarian	0.6240
Unstabilized+DataAug+Hungarian+VelocityPrior	0.6428
Stabilized+DataAug+Hungarian+VelocityPrior	**0.7958**

Stabilization gives >10% boost

© 2019 Carnegie Mellon University

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

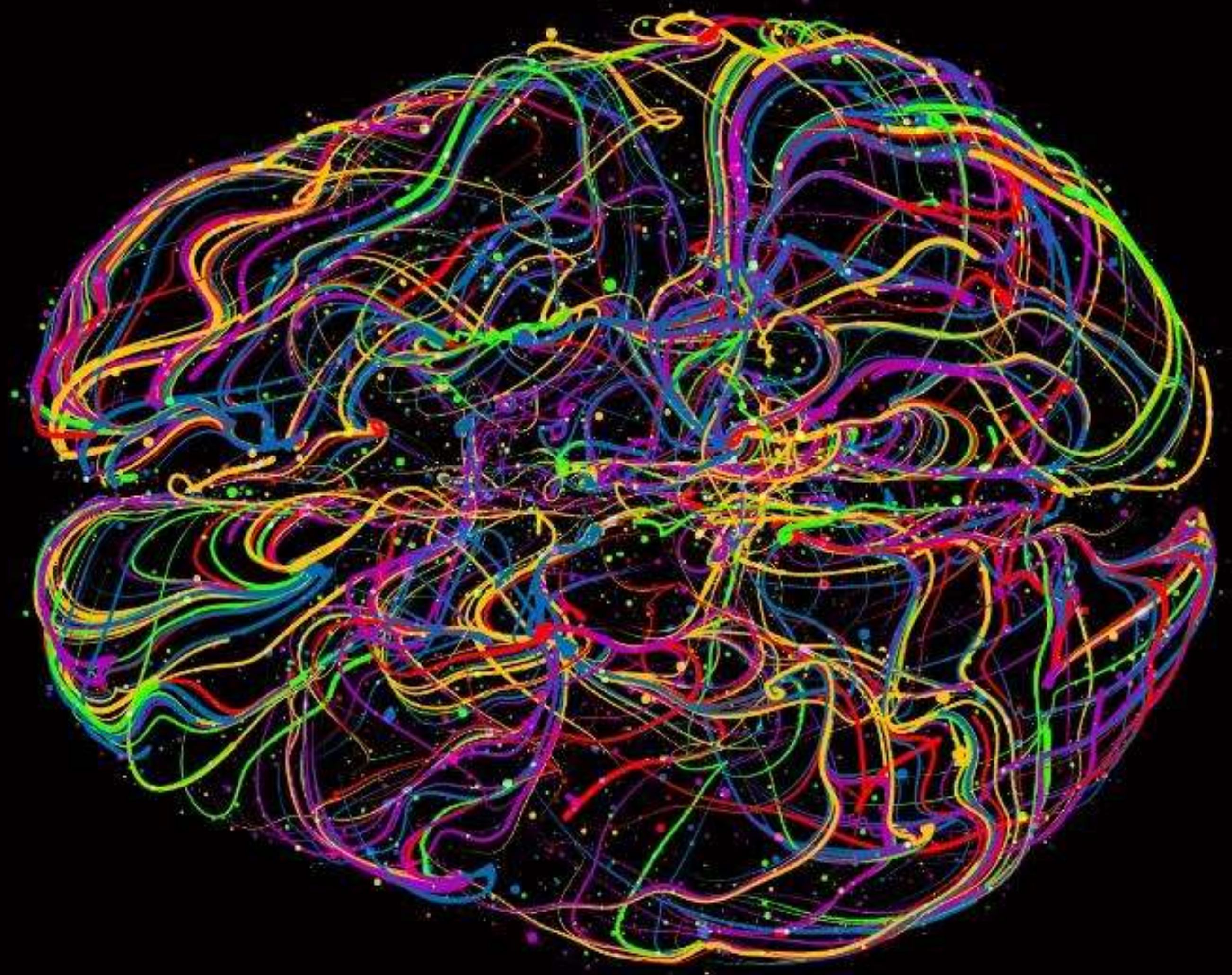
Trajectory Forecasting is Possible Only in Stabilized Space



© 2019 Carnegie Mellon University
[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Video Summarization and Search:

Pattern-of-Life Analysis

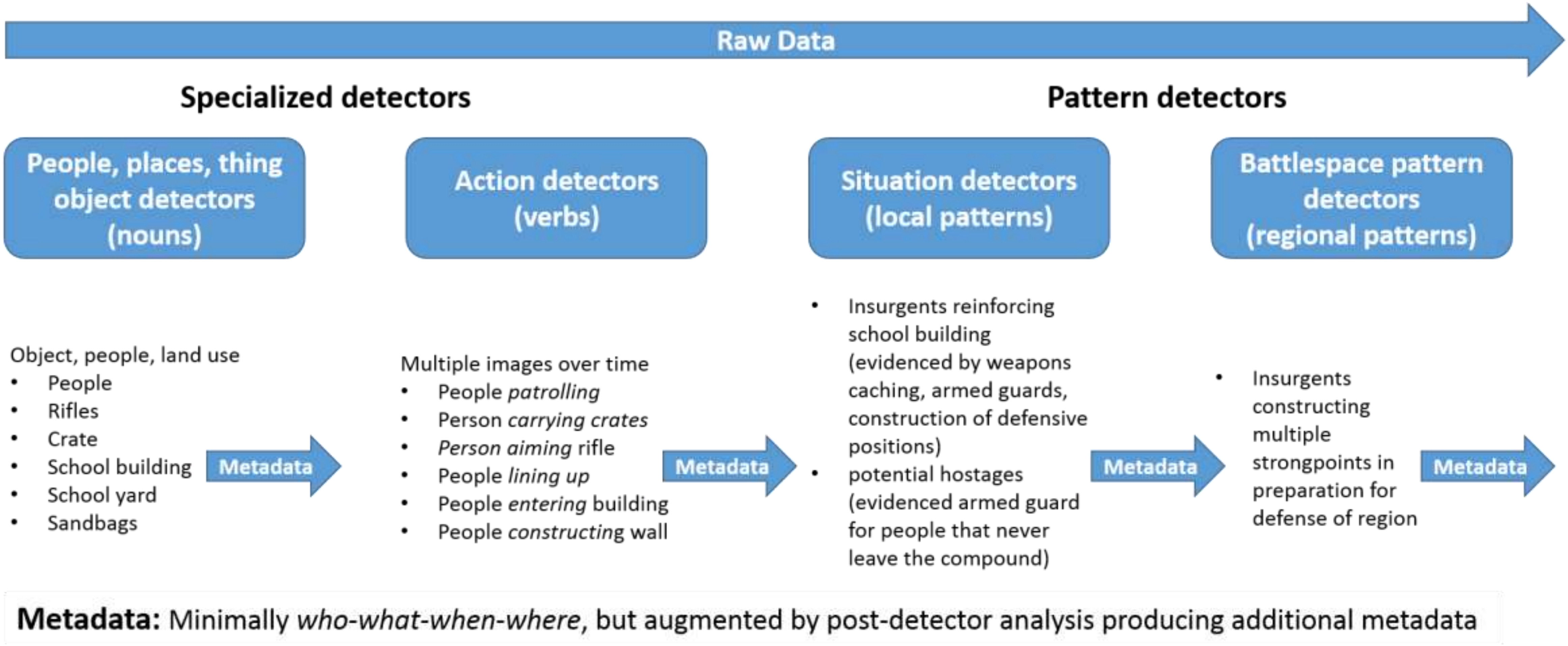


Dr. Jeffery Hansen

© 2019 Carnegie Mellon University

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Pattern-of-Life Analysis: Hierarchy of Detectors

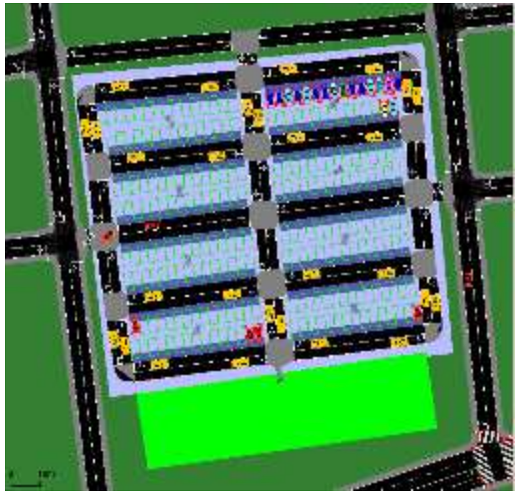


© 2019 Carnegie Mellon University

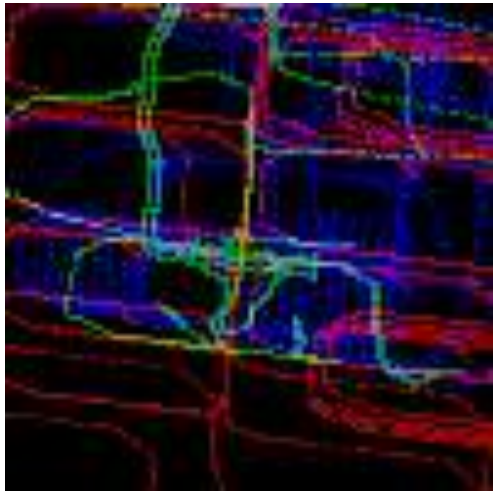
[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Pattern-of-Life Analysis: “Mall Parking Lot” Experiment

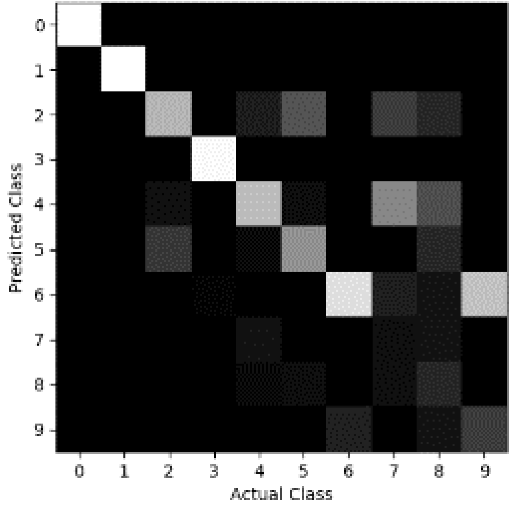
We have conducted initial experiments on a “Mall Parking Lot” example for using machine learning techniques to analyze patterns of life.



SUMO traffic simulator used to simulate vehicles and pedestrians including customers, employees, carpoolers, and “drug dealers.”



For each track, an ego-centric patch is generated characterizing the movement of that vehicle or pedestrian and its relationship to other vehicles, pedestrians, and fixed reference points.

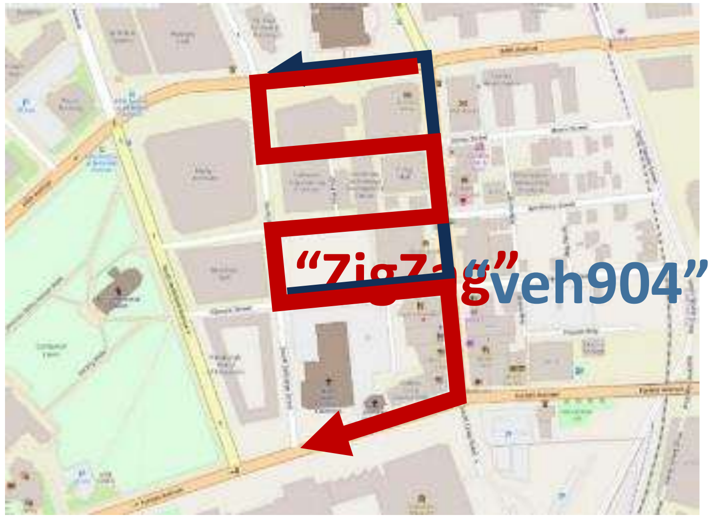


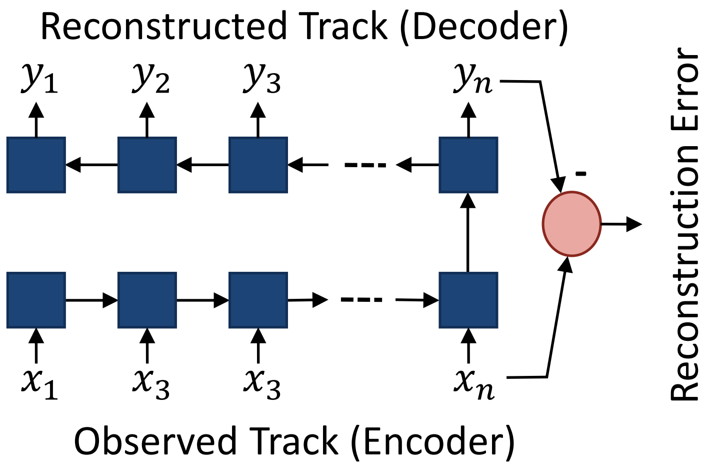
Using a CNN classifier, we correctly identified the track type with 86% accuracy out of 10 possible classes.

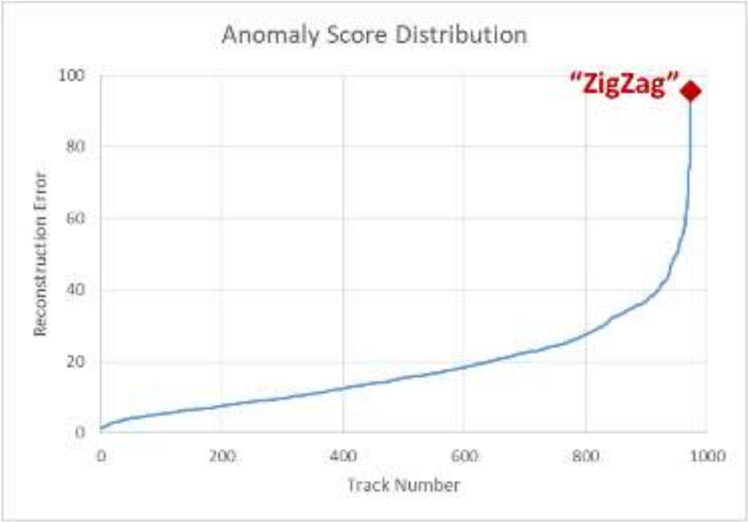
© 2019 Carnegie Mellon University

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Pattern-of-Life Analysis: Anomaly Detection







In another experiment, we simulated 950 “best-path” vehicle trips into which we injected one with an unusual “zig zag” path.

We used an LSTM autoencoder to learn typical track behavior and compute a “reconstruction error” representing the degree to which a track is consider “unusual.”

Of the 950 tracks, our injected “zig-zag” track consistently scored in the top 5 in terms of reconstruction error demonstrating the ability to identify anomalous behaviors.

© 2019 Carnegie Mellon University

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.

Increased training data for improved object identification,

improved object tracking,

and characterization of patterns of life

Enables the analyst to effectively **process more data** and **identify more complex patterns**

© 2019 Carnegie Mellon University

[DISTRIBUTION STATEMENT A] Approved for public release and unlimited distribution.