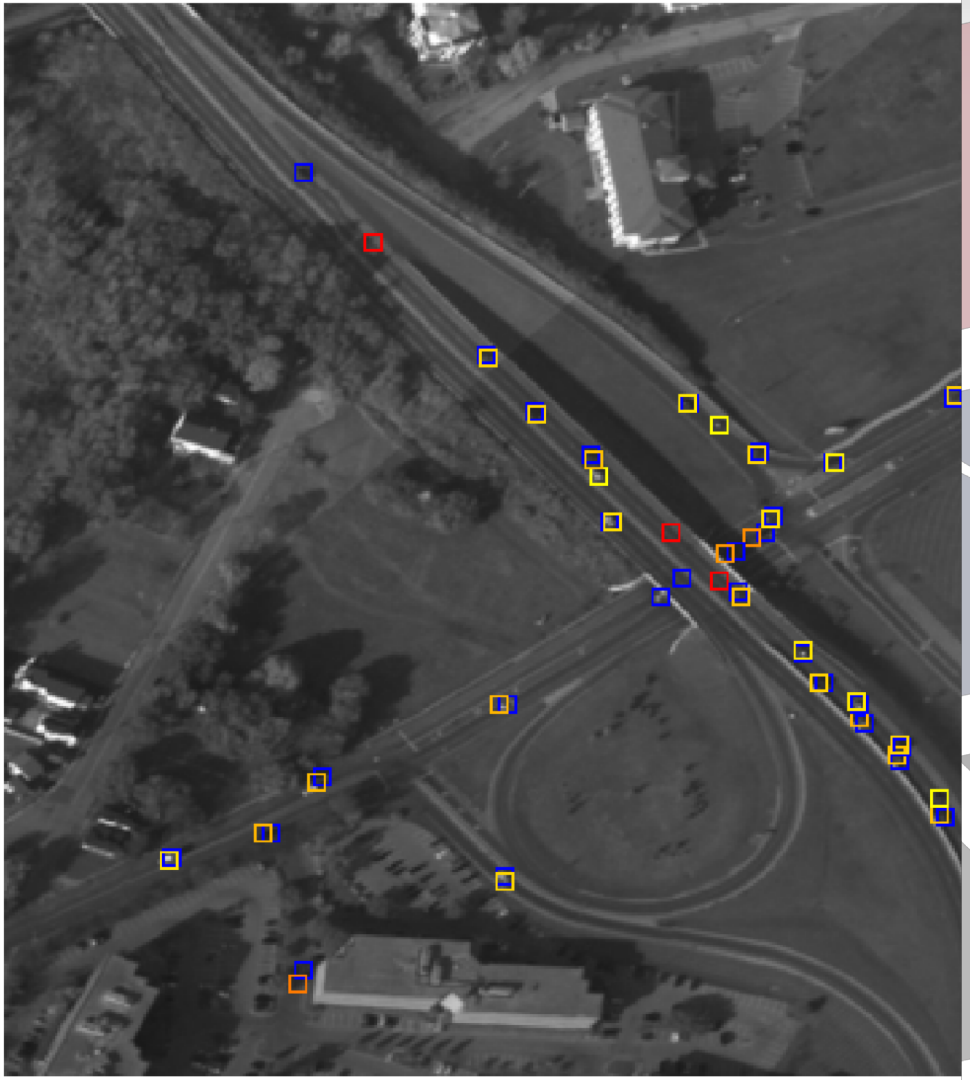


Object Detection and Tracking from Overhead Video with Deep Learning

Joseph Ryan Crawford

Senior Engineer

Expedition Technology

EXPEDITION TECHNOLOGY

Overhead Object Tracking Pipeline

OVERHEAD IMAGERY & DEEP LEARNING

A Very Brief History of Overhead Imagery

- 1858 – First aerial photographs by Nadar from a balloon
- 1880s – Tethered kites (upper left)
- 1897 – First photograph from a rocket
- Early 1900s – "Bavarian Pigeon Fleet" (bottom left)
- 1911 – First airplanes during Italo-Turkish War

Photos from Wikipedia

EXPEDITION TECHNOLOGY

Modern Overhead Imagery

- Modern Platforms
  - Satellites, UAVs, Airplanes
- Why Deep Learning?
  - Abundance of high-resolution imagery sources
  - Advances in deep learning algorithms and hardware

Advances in Imagery Platforms and Deep Learning enable new opportunities

Deep Learning & Image Processing

- Deep Learning for Image Classification & Analysis
  - State of the art since 2012 (AlexNet)
  - Network Improvements
    - Networks are becoming both more efficient and more powerful
    - ResNet (2015)
      - enabled training of very deep networks (hundreds of layers)
    - SqueezeNet (2016), MobileNet v1/2(2017/2018), ShuffleNet(2017)
      - fast architectures that maintain quality results
  - Hardware advances
    - Faster GPUs with Increased Memory (I.e. 32GB V100)

Network and hardware advances enable practical processing of large high-resolution imagery

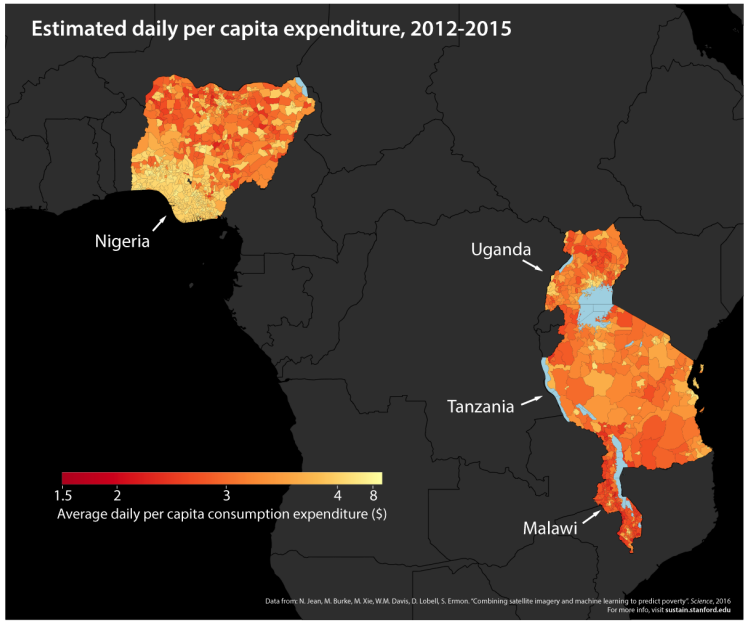
Applications of DL & Overhead Imagery

Kaggle Competition:
Classifying land usage in the Amazon

Example Image Chip Classes:

https://www.kaggle.com/c/planet-understanding-the-amazon-from-space

Estimating Poverty Rates from Nighttime Imagery

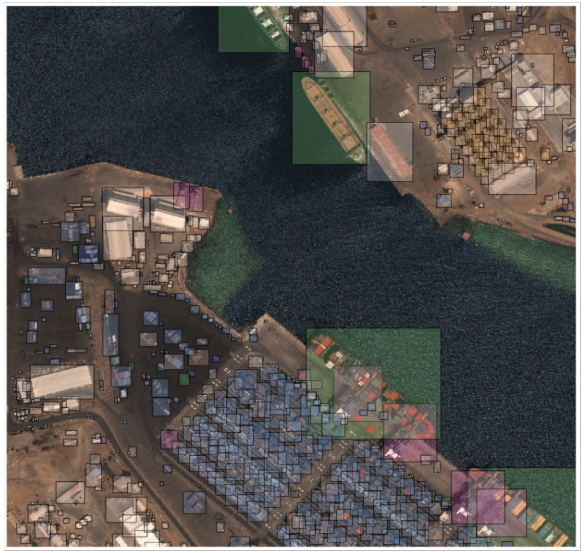


http://sustain.stanford.edu

EXPEDITION TECHNOLOGY

Sponsored Challenges

DIUx xView 2018 Detection Challenge



http://xviewdataset.org
Sponsors: DIUx, NGA

DeepGlobe Satellite Challenge - CVPR18

Road Extraction

Building Detection

Land Cover Classification

http://deepglobe.org
Sponsors: Facebook, DigitalGlobe, Purdue, and MIT

Industry and government are sponsoring challenges to encourage innovation in Deep Learning for Overhead Imagery

EXPEDITION TECHNOLOGY

OVERHEAD/AERIAL IMAGERY SENSORS & PRE-PROCESSING

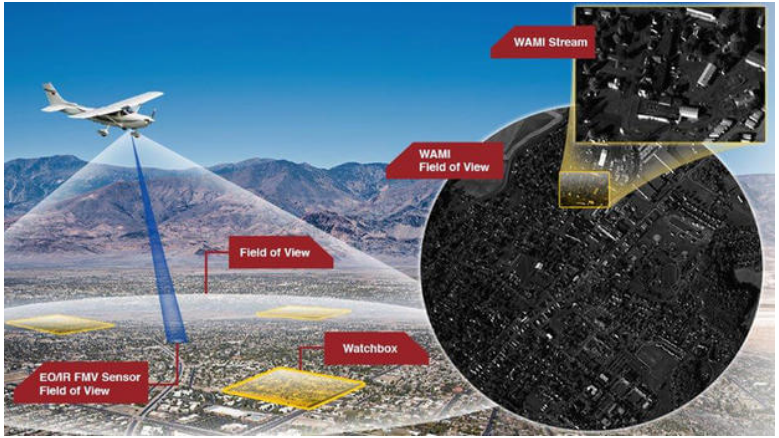
Overhead Aerial Imagery Sensors

- **Satellite Systems**
  - Revisit Rates: Few times per day
  - Sufficient spatial resolution, typically low temporal resolution
  - Source of most competition data
- **Unmanned Aerial Systems/Vehicles (UAS/UAV)**
  - High temporal resolution – 30 to 60 FPS available
  - High spatial resolution – very low ground sample distance (GSD) values per pixel
  - Limited area of regard
- **Wide-Area Motion Imagery (WAMI)**
  - 1 to 2 Hz frame rates
  - Gigapixel images – resulting in GSDs of 0.25-0.5m
  - Wide-Area of Regard – able to collect small city-sized areas simultaneously

Wide-Area Motion Imagery (WAMI)

- **Wide-area motion imagery (WAMI)** sensors
  - Flown on helicopters, balloons, small aircraft, or UAVs
  - Used to image small city-sized areas at approximately **0.25-0.5m/pixel** and about **1-2 frames/s**

- Challenges for Object Detection and Tracking:
  - **Low spatial resolution** – objects of interest are very small
  - **Low temporal resolution** – large object displacement between frames
  - **Difficult Georegistration/stabilization** - can lead to significant motion clutter



https://www.military.com/defensetech/2016/05/25/logos-touts-new-wide-area-surveillance-sensor

WAMI sensors enable city-scale object detection and tracking – despite some challenges

Source: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5562653

Public Dataset:

Wright-Patterson Air Force Base 2009 Dataset (WPAFB2009)

- **Six cameras** with ortho-rectified imagery
- Frame Rate: ~1.25 Hz
- Duration: 14 minutes (1,025 frames) with ~18k ground truth tracks
- WPAFB2009 dataset available at 6 resolution levels
  - Resolution 0 @ ~0.25m GSD
  - **Resolution 1 @ ~0.5m GSD – typical vehicles ~9x9 pixels**
  - Resolution 2 @ ~1m GSD
  - Each remaining level a doubling of GSD from there

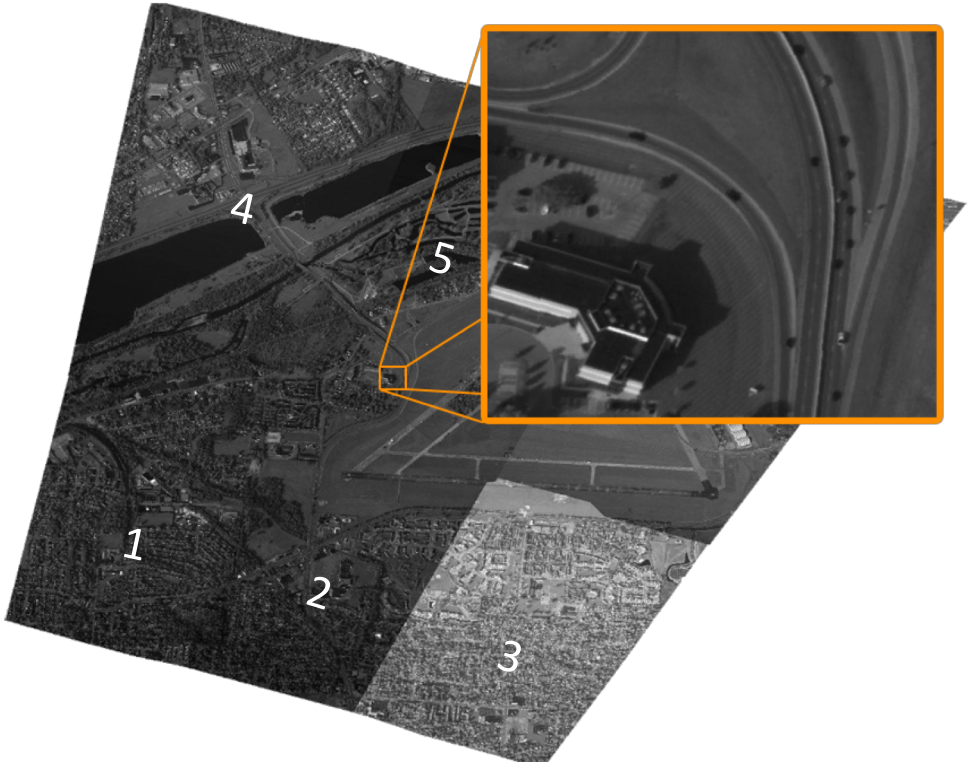
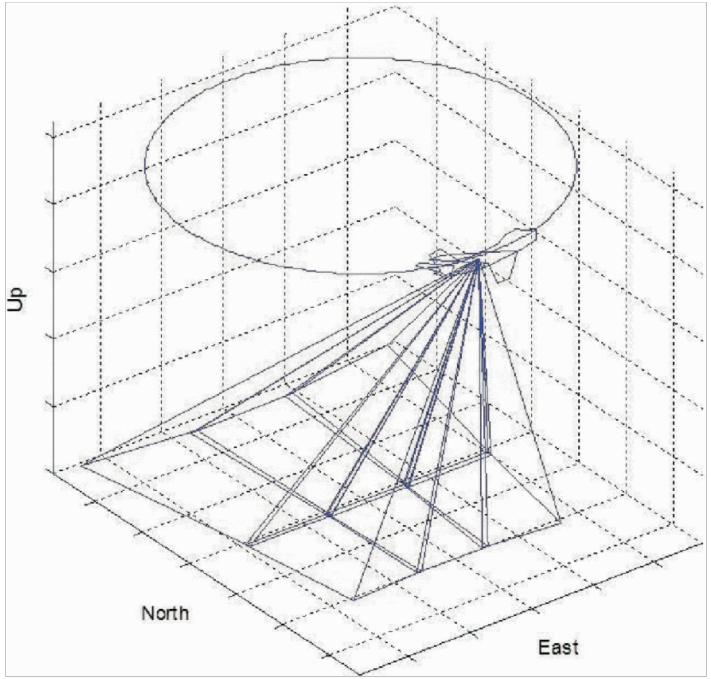


Image: https://www.sdms.afrl.af.mil/index.php?collection=wpafb2009

Public Dataset:

Wright-Patterson Air Force Base 2009 Dataset (WPAFB2009)



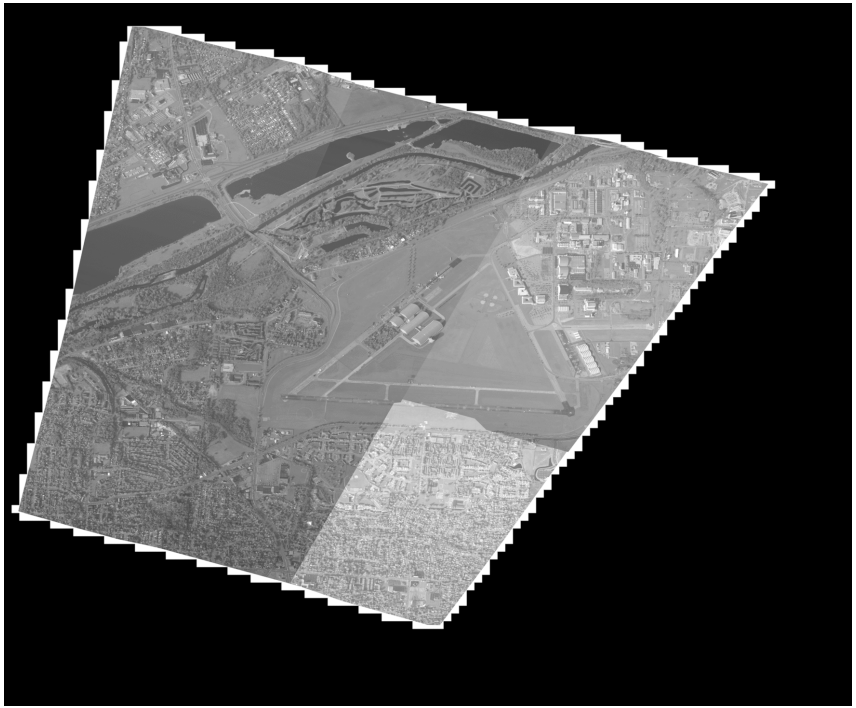
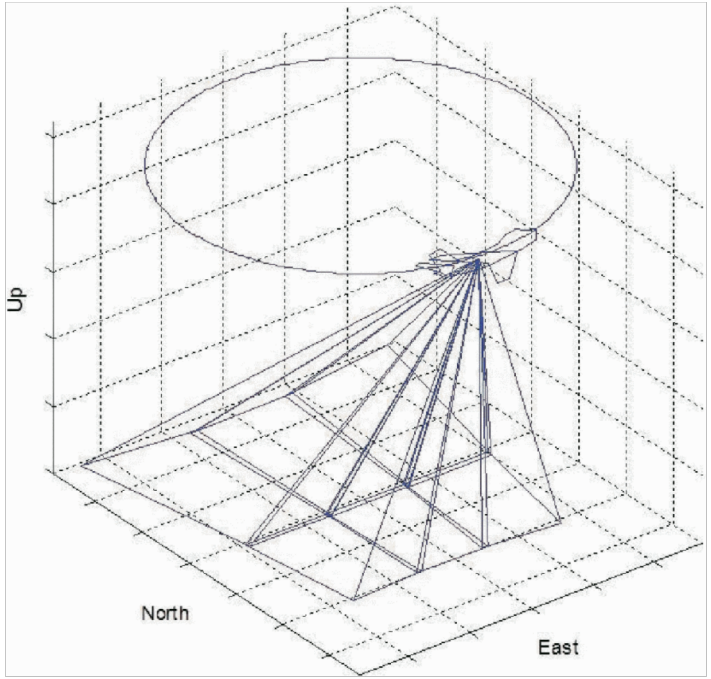


Image: https://ieeexplore.ieee.org/document/7158947/

Public Dataset:

Wright-Patterson Air Force Base 2009 Dataset (WPAFB2009)



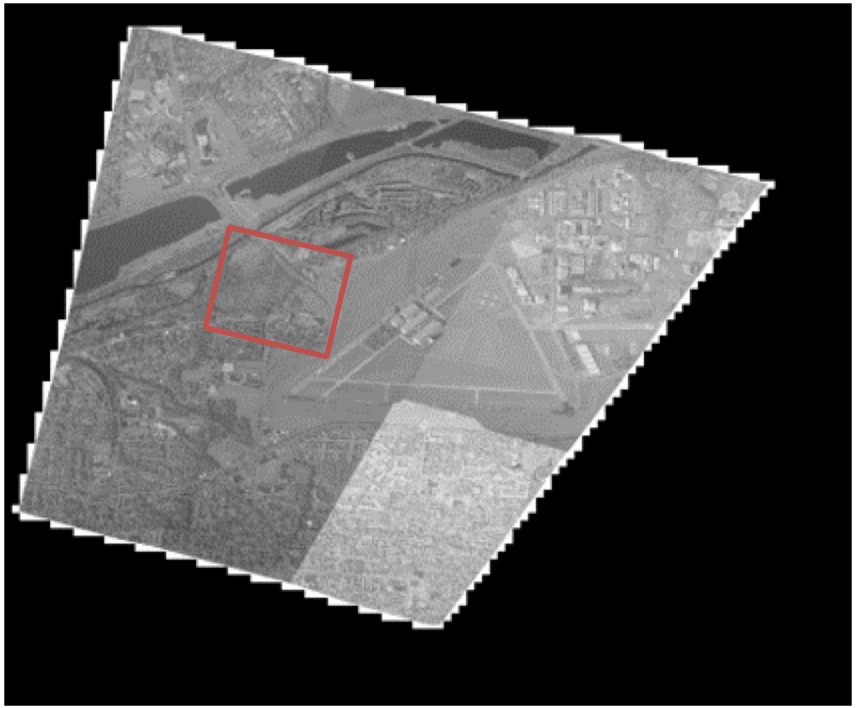


Image: https://ieeexplore.ieee.org/document/7158947/

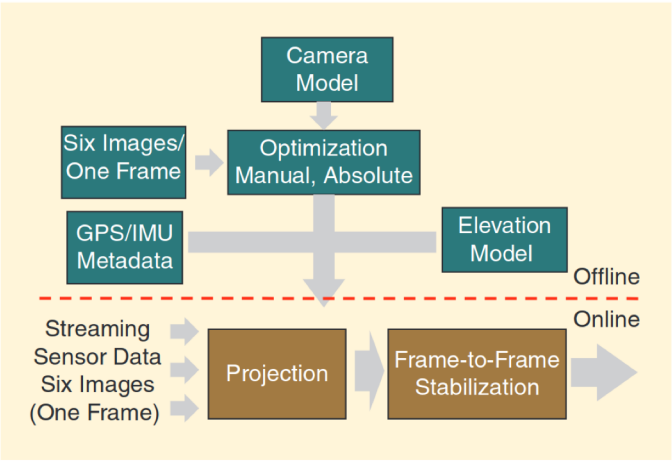
EXPEDITION TECHNOLOGY

Sample WPAFB2009 Clip

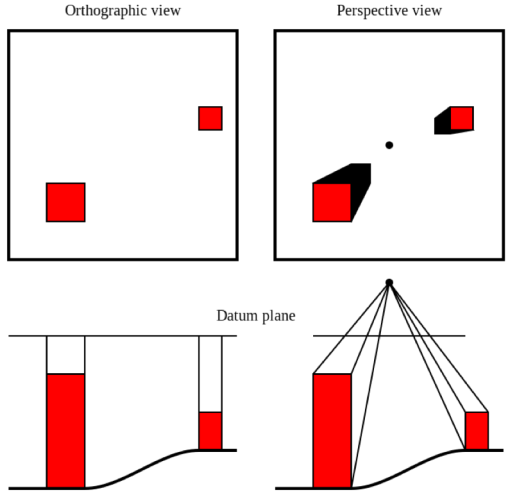
- Example of WAMI collection from WPAFB2009 dataset
- Zoomed to a single ROI
- **Note the jitter and lighting artifacts from the six sensor images stitched together**
  - Lighting and color changes are difficult for background subtraction / frame differencing techniques

Image Georegistration / Orthorectification

- As the plane circles the region of interest, image mosaic must be stitched together and georegistered



[FIG2] A real-time WAMI georegistration and stabilization pipeline.



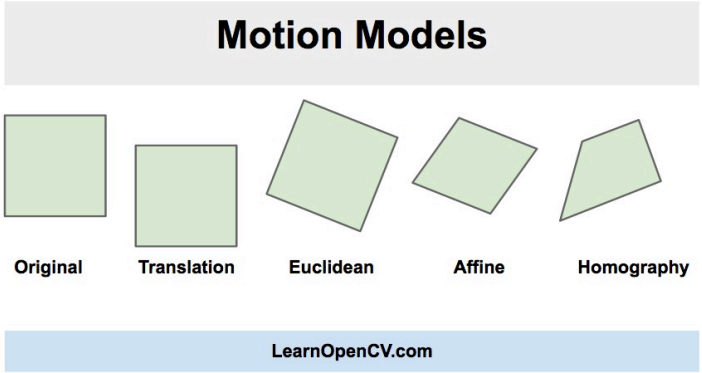
Orthographic views project at a right angle to the data plane. Perspective views project from the surface onto the datum plane from a fixed location.

Images: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=5562653, https://en.wikipedia.org/wiki/Orthophoto

Motion Analysis of Frame Sequence

Methods for frame-to-frame alignment vary in four considerable ways:

1. **Feature Space:** Information useful for matching
   (e.g. key points, edges, Fourier representation, **pixel intensity**)
2. **Search Space:** Class of transformation
   (e.g. translation, Euclidean, affine, **homography**)
3. **Search Strategy:** Method to choose the next test transformation
   (e.g. exhaustive search, relaxation, dynamic programming, **gradient descent**)
4. **Similarity Metric:** Value of a test transformation
   (e.g. sum of squares difference, **Enhanced Correlation Coefficient**[*])

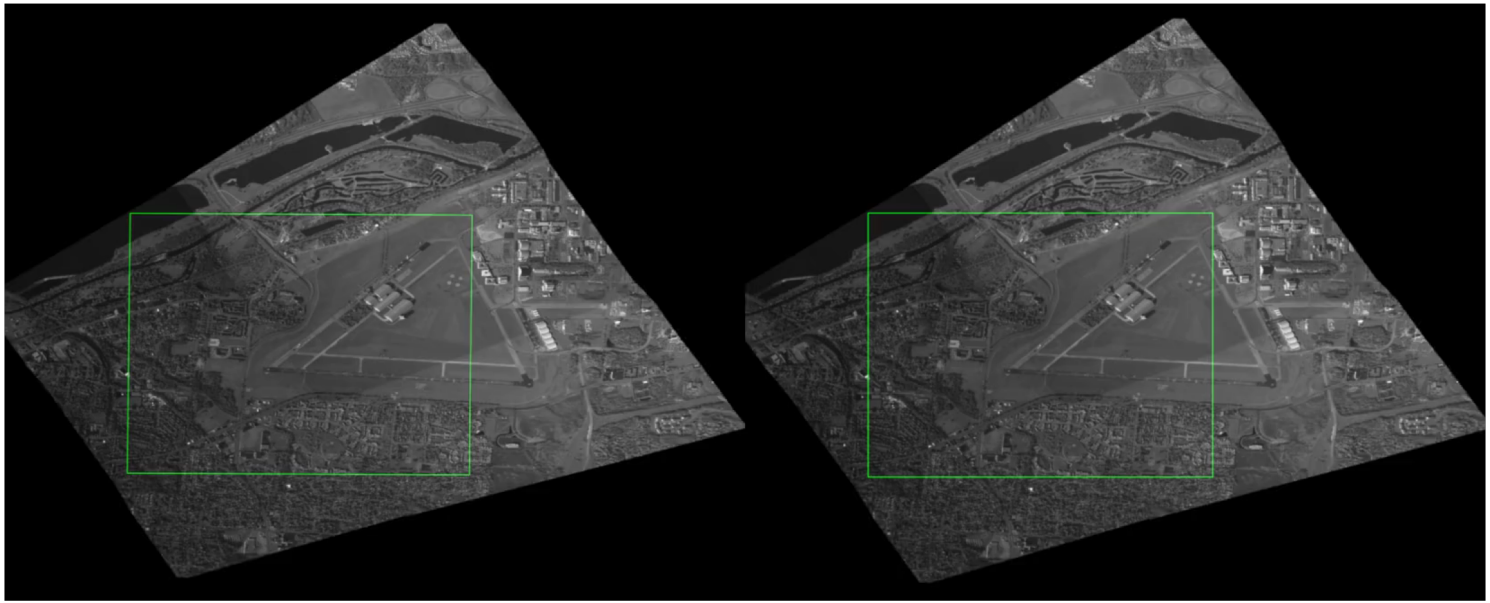


[*] Evangelidis, G.D. and Psarakis E.Z. "Parametric Image Alignment using Enhanced Correlation Coefficient Maximization", IEEE Transactions on PAMI, vol. 32, no. 10, 2008

EXPEDITION TECHNOLOGY

Result of Perspective Transform

Original Frames **Stabilized Frames**



Stabilized images enable increased detection and tracking performance

EXPEDITION TECHNOLOGY

OVERHEAD OBJECT DETECTION

Traditional Approaches

Traditional Image Feature-based Approaches

- Background Subtraction
- Vehicle Extraction
- Shadow Removal
- Feature-based
  - HOG, SIFT, SURF, BRIEF

Shortfalls:

- Not robust to aerial imagery in motion
  - Poor stabilization, georegistration, lighting changes, etc.
- Objects typically large relative to image size or require erosion-dilation

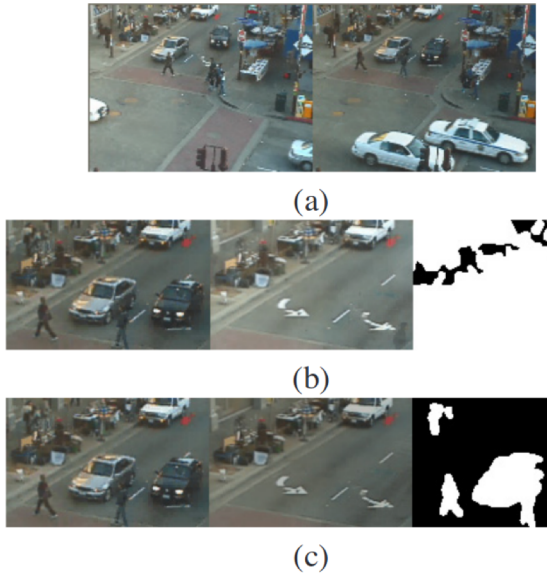
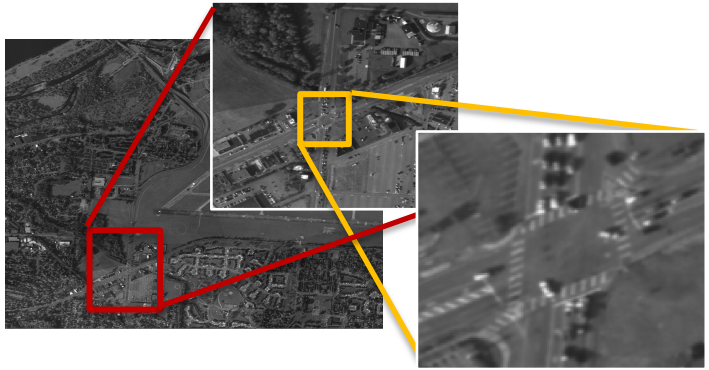


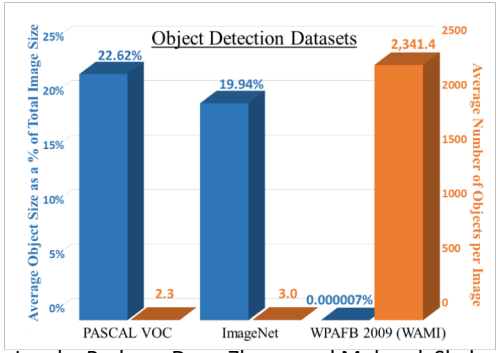
Figure 1. (a) The entire scene suddenly becomes dark by an auto-iris camera as the two white vehicles in the bottom pass by. (b) A disastrous detection result without the illumination correction. (c) An enhanced result with the illumination correction.

Kim, Z. Real time object tracking based on dynamic feature grouping with background subtraction. CVPR 2008.

Small Object Detection with Deep Learning

- Traditional vs Small Object Detection
  - Typical object detection has focused on objects that are large relative to the scene
  - Commonly 20% of the total image size
  - Self-driving car data can have significantly smaller objects than ImageNet on the order of 0.1-1% of pixels
  - Truly "Small Objects" in aerial imagery can be **one millionth of one percent** of the image





LaLonde, Rodney, Dong Zhang, and Mubarak Shah. "Clusternet: Detecting small objects in large scenes by exploiting spatio-temporal information." In *Computer Vision and Pattern Recognition*. 2018.

COCO 2018 Object Detection Task

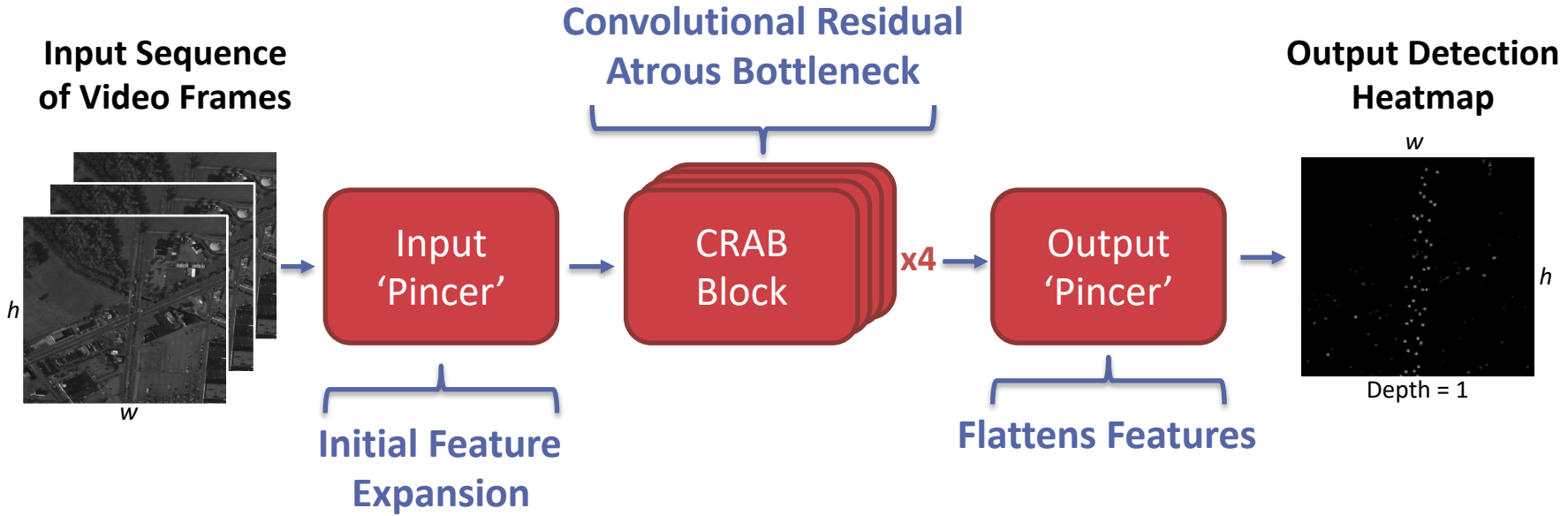
EXPEDITION TECHNOLOGY

Our Vehicle Detection Approach

- **Objective**
  - Discrete detection of vehicles to seed a tracker
  - Architecture than can process large images
  - Favor detections of moving objects
    - Reduces false positives from buildings
    - Less interest in stationary objects

- **Approach**
  - Fully convolutional network
  - Input multiple consecutive frames
    - Allows network to infer object motion

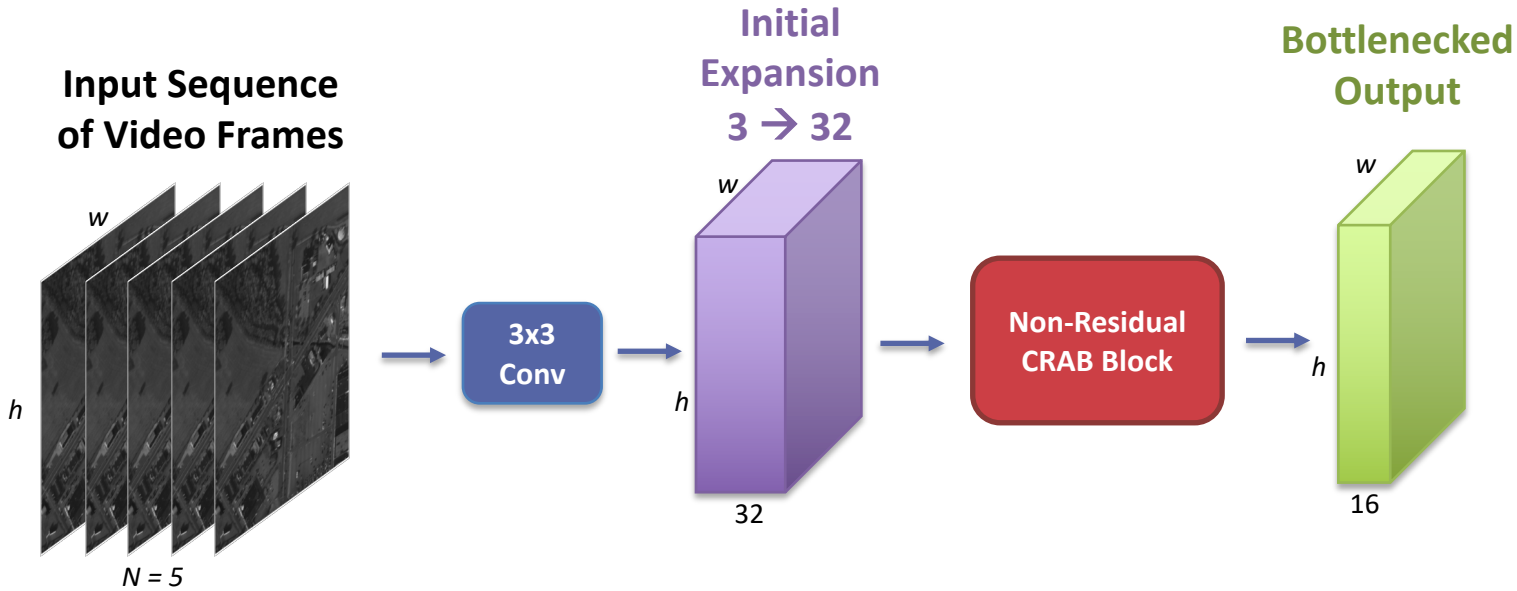
CRABnet – Vehicle Detection Architecture



Inspiration from ClusterNet by LaLonde et al. CVPR 2018.

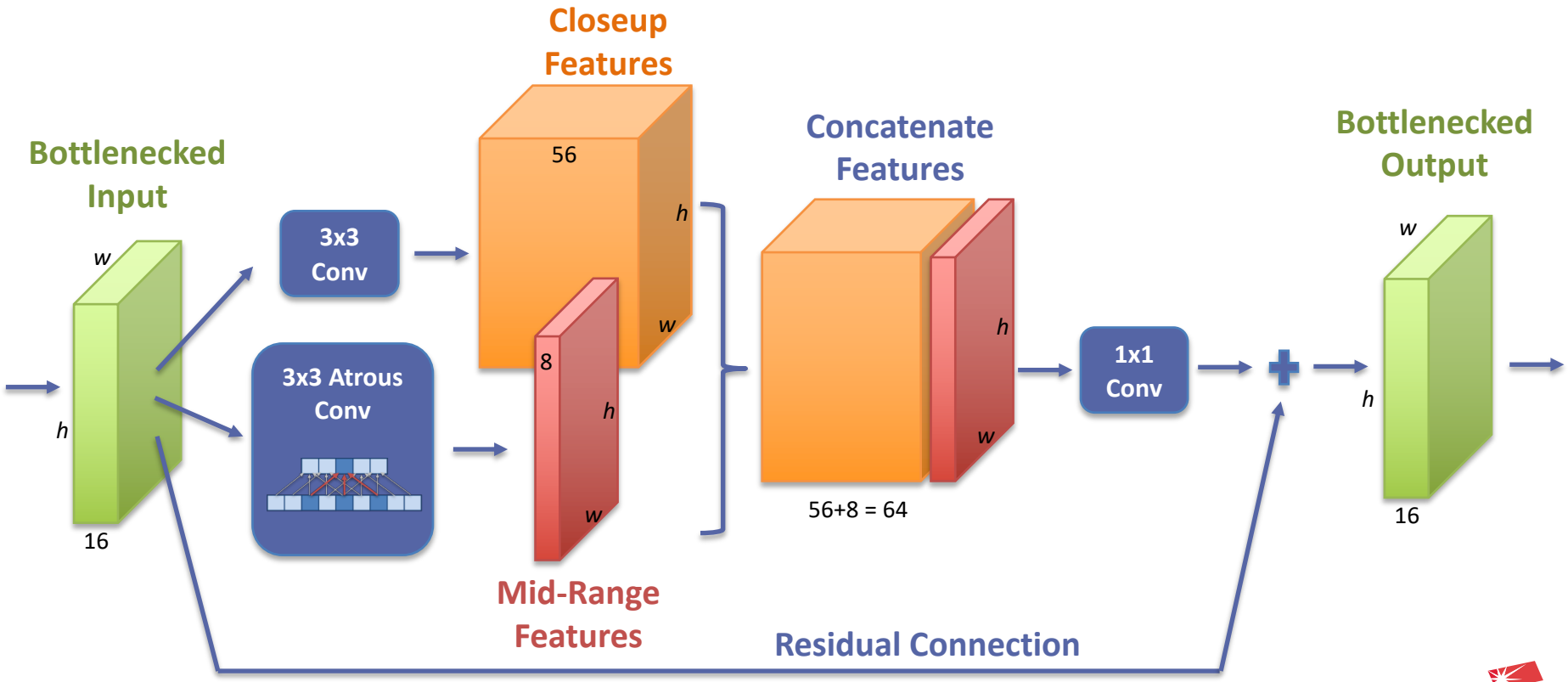
EXPEDITION TECHNOLOGY

CRABnet – Input Pincer Block



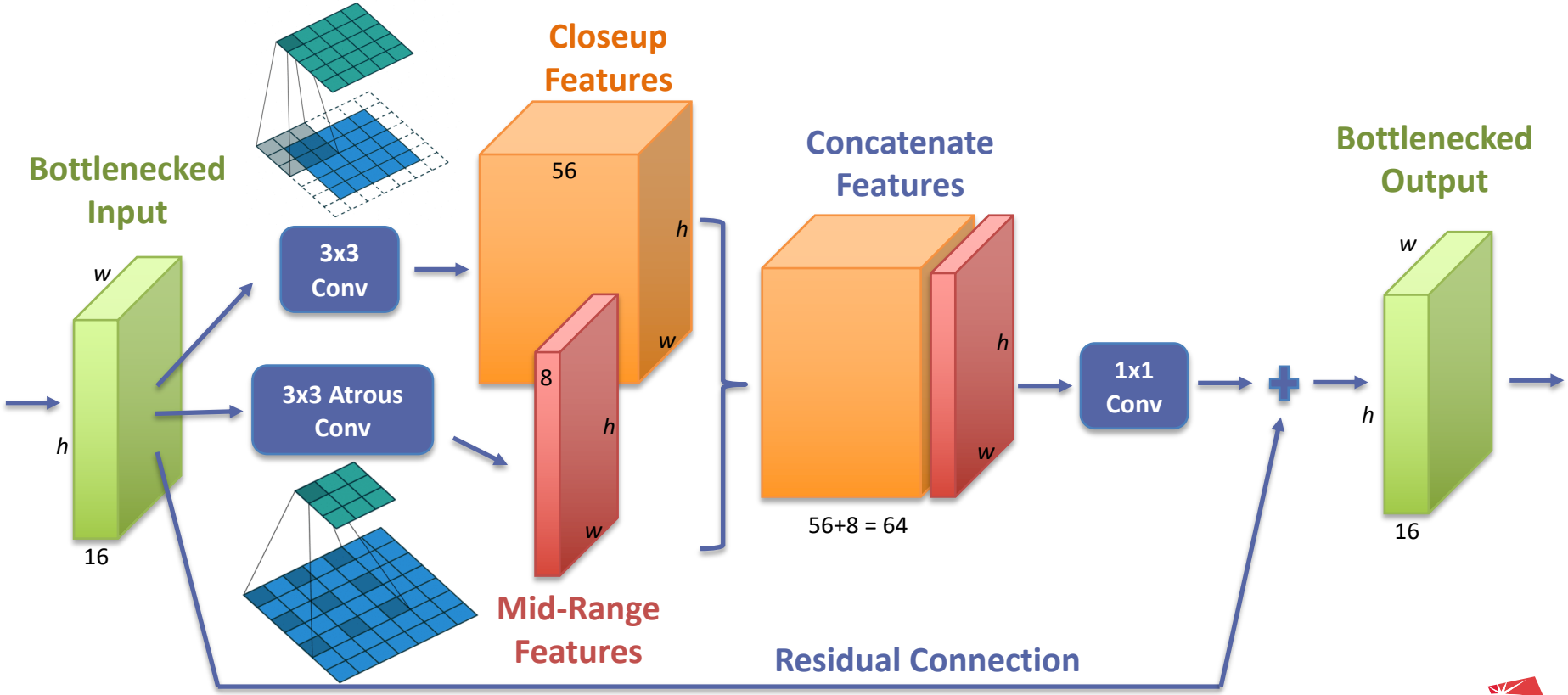
EXPEDITION TECHNOLOGY

CRABnet – CRAB Block



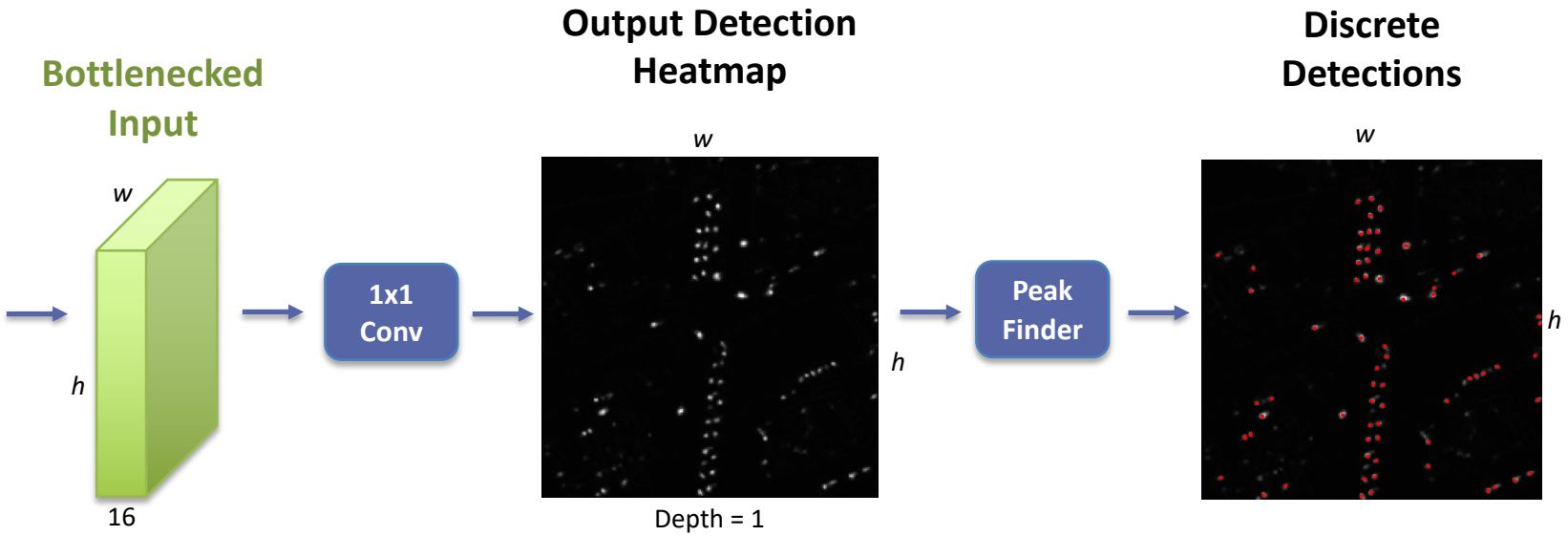
EXPEDITION TECHNOLOGY

CRABnet – CRAB Block



EXPEDITION TECHNOLOGY

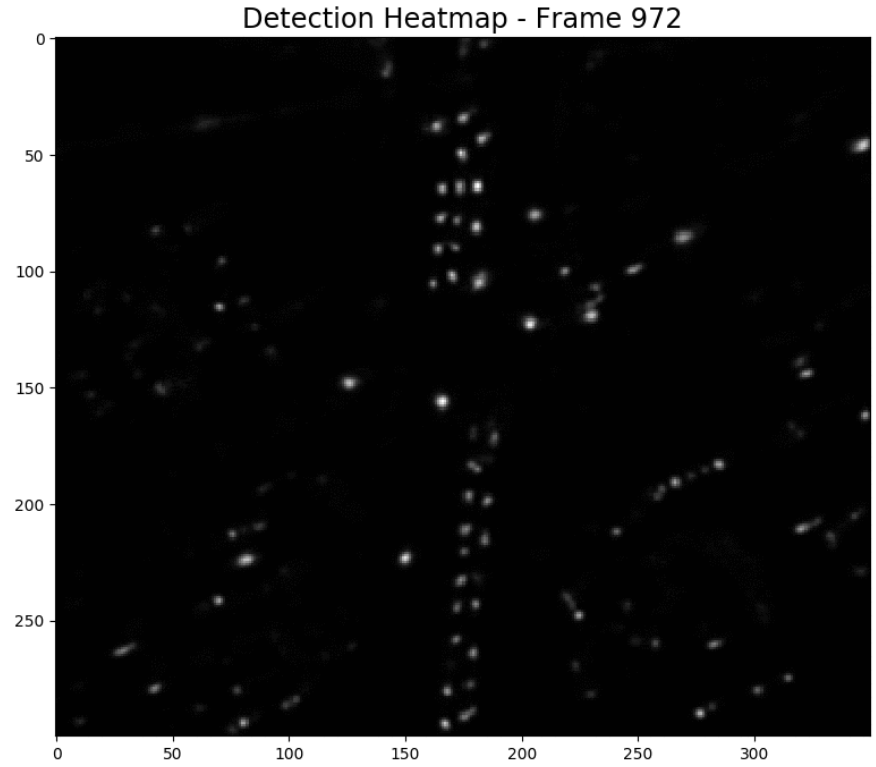
CRABnet – Output Pincer Block

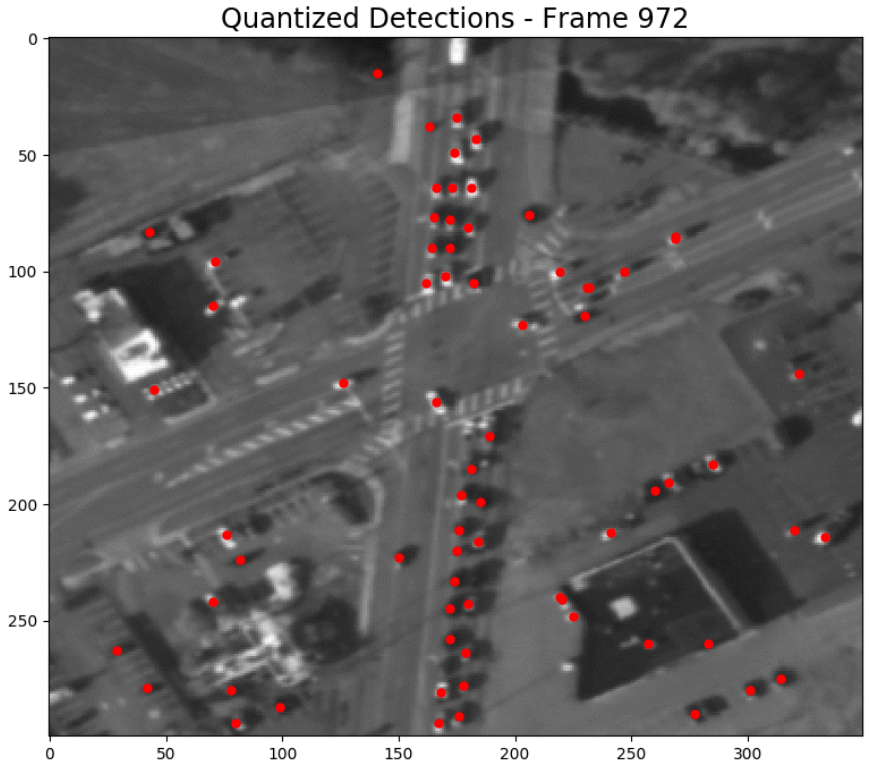


Output Detection Heatmap is used for Discrete Detections and passed to the tracker

EXPEDITION TECHNOLOGY

Sample Detection Results



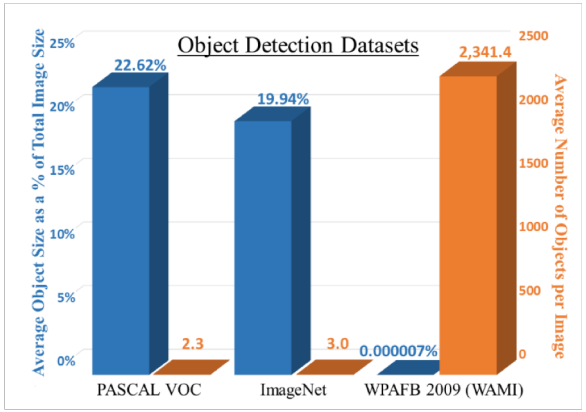


EXPEDITION TECHNOLOGY

OVERHEAD IMAGERY TRACKING

Small Object Tracking with Deep Learning

- Traditional vs Small Object Tracking
  - Typical object tracking has focused on objects that are large relative to the scene
  - Multi Object Tracking (MOT) Benchmark test trackers on crowded scenes of pedestrians
  - **To repeat:** Truly "Small Objects" in aerial imagery can be **one millionth of one percent** of the image

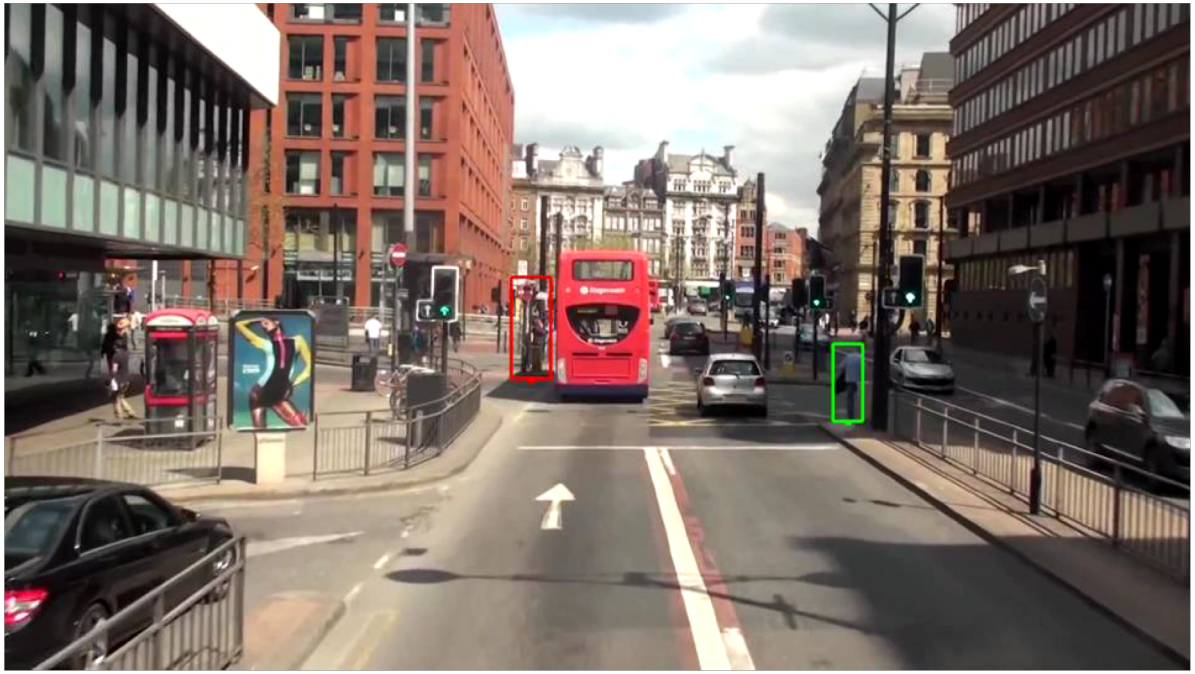


LaLonde, Rodney, Dong Zhang, and Mubarak Shah. "Clusternet: Detecting small objects in large scenes by exploiting spatio-temporal information." In *Computer Vision and Pattern Recognition*. 2018.

EXPEDITION TECHNOLOGY

Object Tracking for Self-Driving Cars

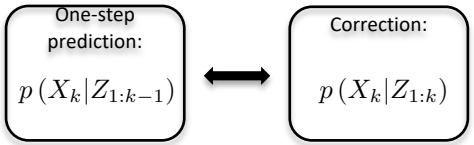
- Current research focuses on objects of a large relative size to the scene
- Significant research driven by self-driving cars
- Example result from the **Multiple Object Tracking Benchmark 2016**

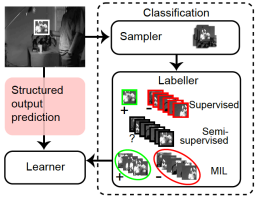


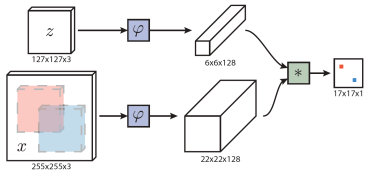
Clip from MOT16 submission: https://motchallenge.net/tracker/LMP
S. Tang, M. Andriluka, B. Andres, and B. Schiele. Multi people tracking with lifted multicut and person re-identification. CVPR 2017.

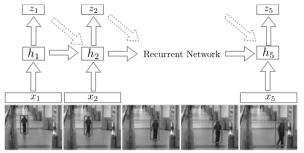
Approaches to Tracking

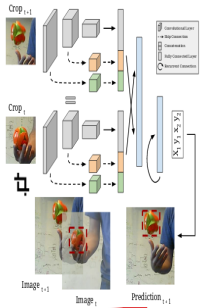
	Track by (particle) filtering	Track by detection	Track by matching	Visual recurrent tracking
Pros	- Usually very accurate - Well understood Bayesian theory - Non-linear dynamics	- Usually very accurate - Can use any detection algorithm - Detect object(s) in every frame	- Usually very accurate - Builds target appearance model either offline or online - Match appearance models from frame to frame	- Joint learning of target dynamics and appearance - No Markovian assumption - Completely trained offline
Cons	- Markovian target motion model - Large number of particles needed - Target appearance and dynamics modeled independently	- Markovian target motion model - Computationally very expensive - Target appearance and dynamics modeled independently	- Markovian target motion model - Computationally expensive - Needs high frame rate - Target appearance and dynamics modeled independently	- Need large amounts of training data - Can be difficult to train depending on the model complexity and optimization objective











References:

[1] P. Perez, C. Hue, J. Vermaak, M. Gangnet, "*Color-based probabilistic tracking*," ECCV, 2002.

[2] S. Hare, A. Saffari, P. H. S. Torr, "*Struck: Structured output tracking with kernels*," ICCV, 2011

[3] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, "Fully-Convolutional Siamese Networks for Object Tracking," ECCV, 2016

[4] Q. Gan, Q. Guo, Z. Zhang, K. Cho, "*First Step toward Model-Free, Anonymous Object Tracking with Recurrent Neural Networks*," https://arxiv.org/abs/1511.06425

[5] Gordon, D., Farhadi, A., Fox, D.: Re3: Real-time recurrent regression networks for visual tracking of generic objects. IEEE Robotics and Automation Letters (2018)

EXPEDITION TECHNOLOGY

Recent Challenge Approaches

Visual Object Tracking (VOT) Challenge

- Focuses on short-term (ST), long-term (ST), and real-time tracking of a target object

Deep Learning Usage

- **2013-2014:** Dominated by traditional CV techniques
  - Background-subtraction, optical flow, key-point, model-based etc.
- **2015:** 3 “Deep” trackers out of 62
  - 1st and 2nd place by large margin
- **2018:** all top trackers use “deep” feature extractors
  - CNNs, ResNets, Siamese networks

VOT2018 challenges Summary

- VOT2018 ST baseline:
  - DCF the dominant methodology, Deep features the dominant features
  - Wider use of deep features trained for localization
- VOT2018 ST realtime:
  - Best performance: fully convolutional deep approaches and DCFs
  - Some of the fastest trackers also rank among best on baseline
- VOT2018 LT:
  - Explicit object detection integrated
  - Nearly all top-performers CNN-based, only one purely DCF
  - Top two performers do not update the detector

Matej Kristan (matej.kristan@fri.uni-lj.si) 32/33

State of the art Object Trackers all leverage Deep Learning

Image: http://www.votchallenge.net/

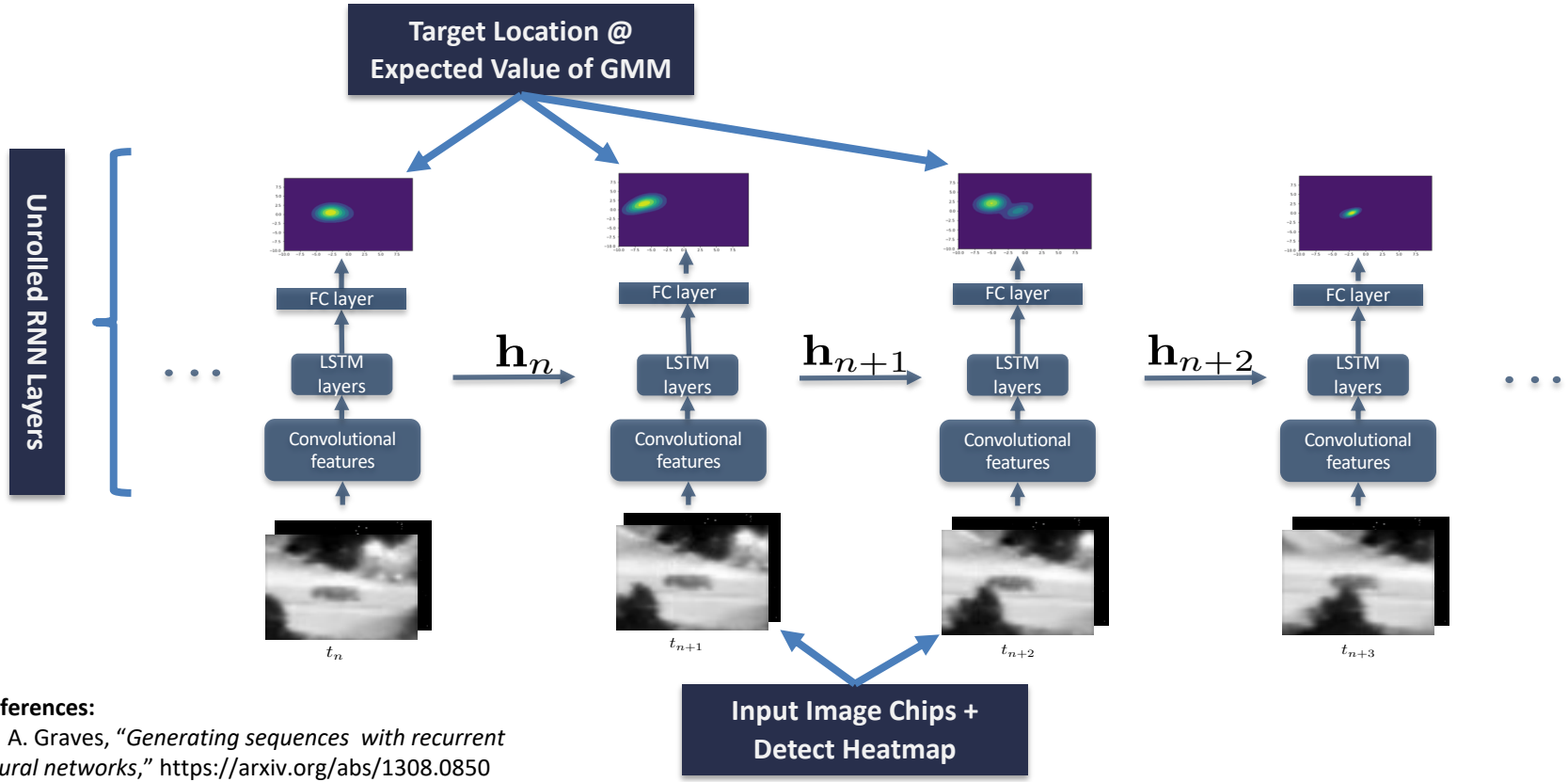
Our Tracker Approach

- Learn frame-to-frame target displacement via Recurrent Neural Network (RNN)
  - Long Short-Term Memory (LSTM) units used in RNN

- At each time step, the target displacement is the expected value under a Gaussian Mixture Model (GMM)
  - Negative log-likelihood of training data given the learned GMM acts as a smooth loss function

At each time step, the output GMM is conditioned on the current input and the history of ALL previous inputs

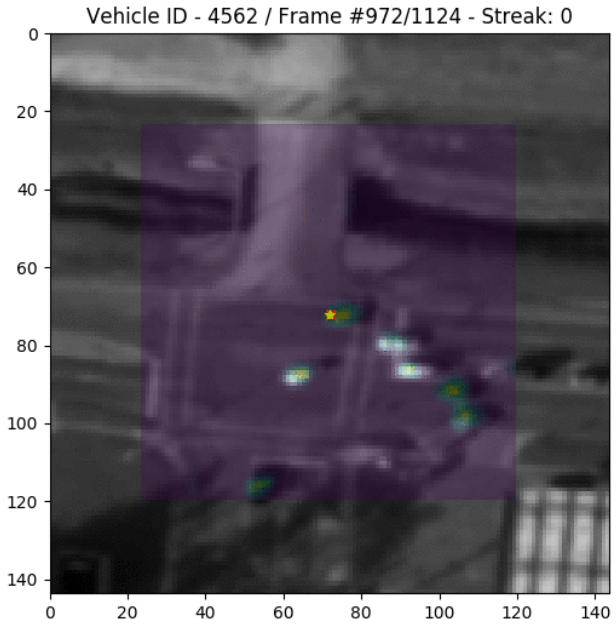
Tracker Approach

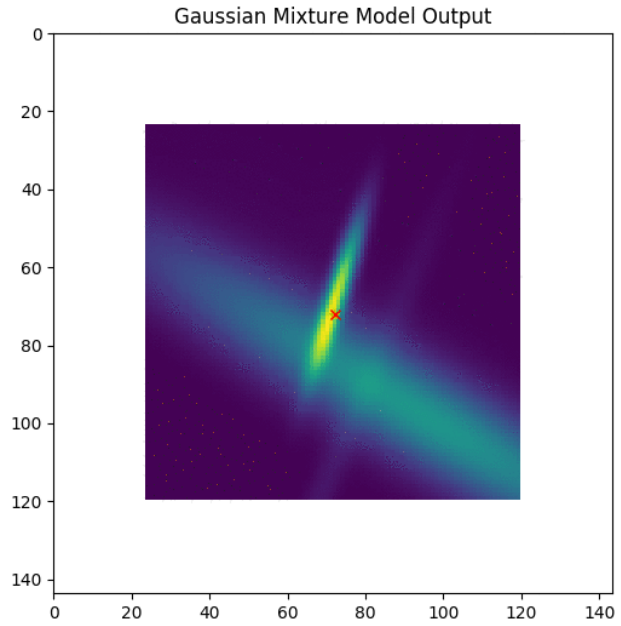


References:
[1] A. Graves, "*Generating sequences with recurrent neural networks*," https://arxiv.org/abs/1308.0850

EXPEDITION TECHNOLOGY

Tracker Samples – Early Training Progress





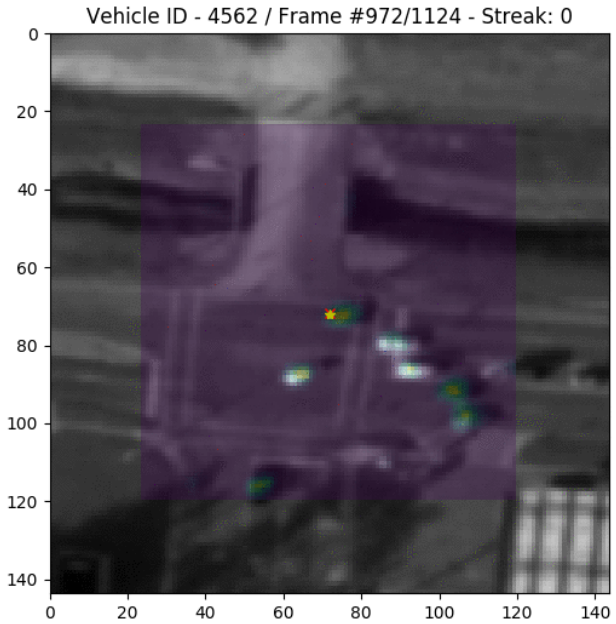
Truth Location

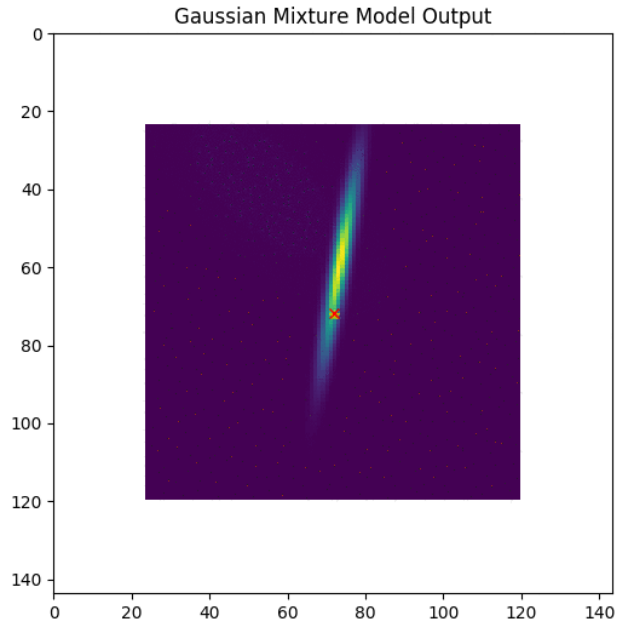
Predicted Location

After 12,000 steps – GMM output can barely locate the correct vehicle and typically loses track immediately

EXPEDITION TECHNOLOGY

Tracker Samples – Further Progress





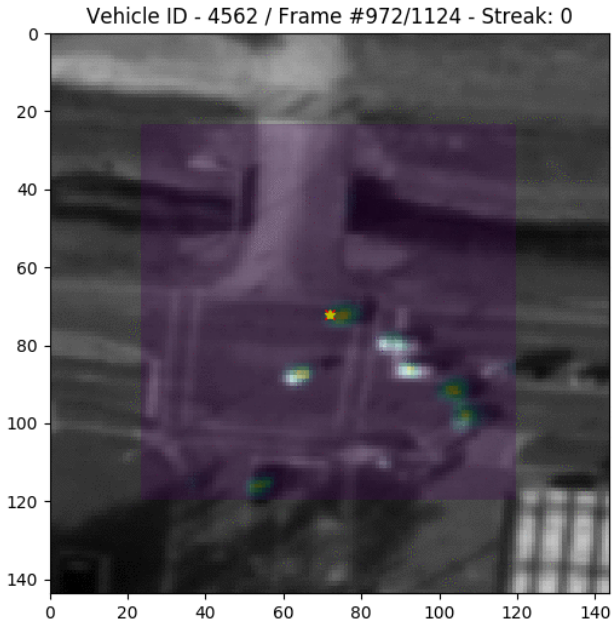
Truth Location

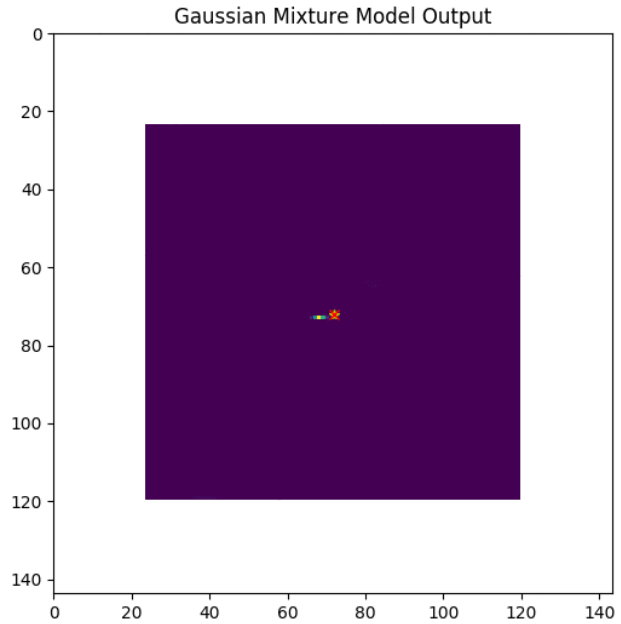
Predicted Location

After 30,000 steps – GMM output able to locate the vehicle, but does not react quickly to changes in vehicle dynamics

EXPEDITION TECHNOLOGY

Tracker Samples – Fully Trained Model





Truth Location

Predicted Location

After 400,000 steps – GMM output able to track the vehicle well once lock is established

EXPEDITION TECHNOLOGY

Results on WPAFB

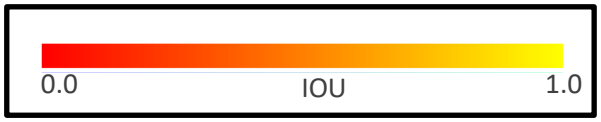
Training Setup	
Resolution	1: ~0.5m GSD
Regions of Interest	Data split into five distinct ROIs
Sequence Length	20 Frames
Training Set	First 850 WPAFB Frames
Target Split	Approximately half moving, half stationary targets

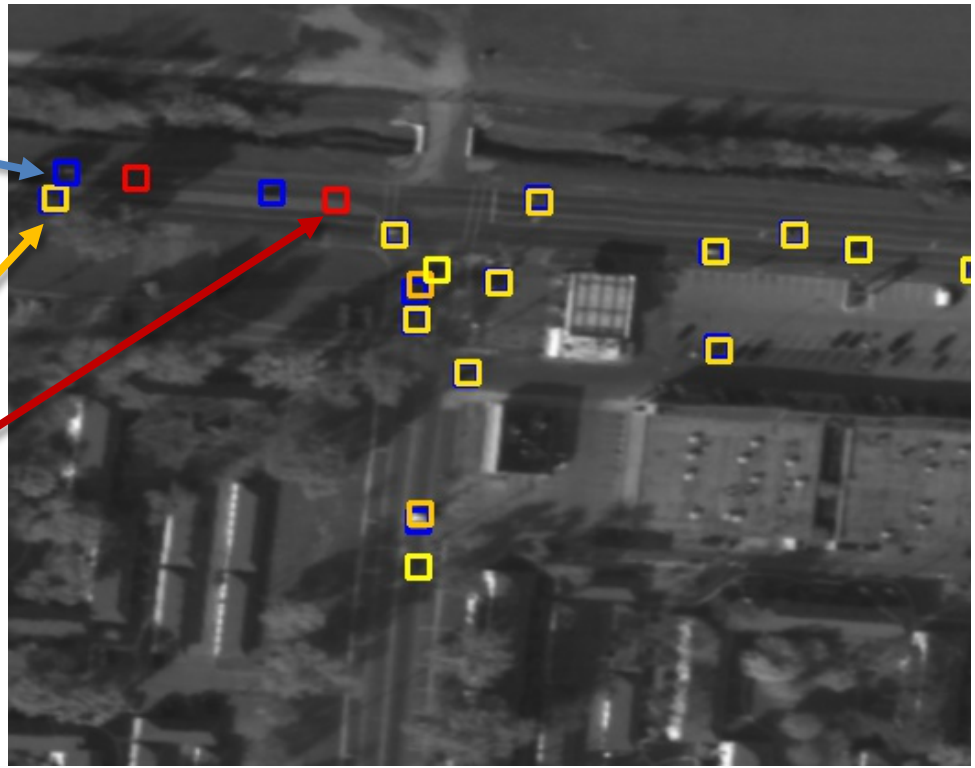
Evaluation Setup	
Evaluation Set	Final 150 WPAFB Frames
Detections	Truth data is used to seed the initial tracker position
Track Death / Rebirth	Tracks are re-seeded to the truth position when a track is lost

Note: Current results are trained on the entire ROI and tested on held out frames. Future results will test extrapolation to new scenes.

Evaluation Results Video

- Ground truth
  - Labeled as **Blue**
- Tracker Prediction
  - Predicted boxes are colormapped against IOU
  - **Yellow** is full overlap
  - **Bright Red** for lost track



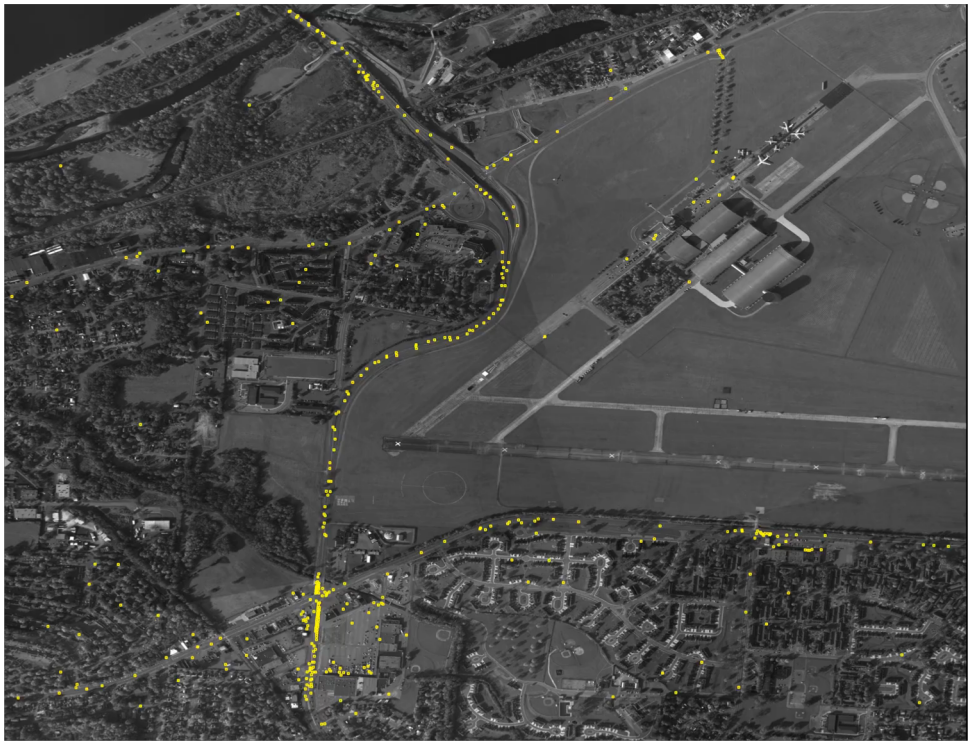


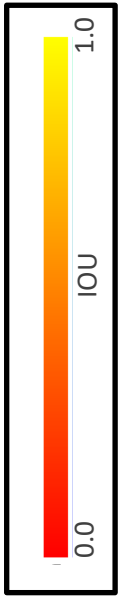
EXPEDITION TECHNOLOGY

Tracker Results Seeded from Ground Truth

- Sample Result on Full Scene

When the tracker loses the target vehicle it is reset to the true vehicle location (a frame of bright red will be shown)





EXPEDITION TECHNOLOGY

Tracker Results Seeded from Ground Truth

- Sample Result on Neighborhood Scene

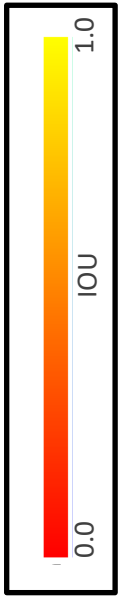
When the tracker loses the target vehicle it is reset to the true vehicle location (a frame of bright red will be shown)

EXPEDITION TECHNOLOGY

Tracker Results Seeded from Ground Truth

- Sample Result on Complex Scene – Overpass/Underpass

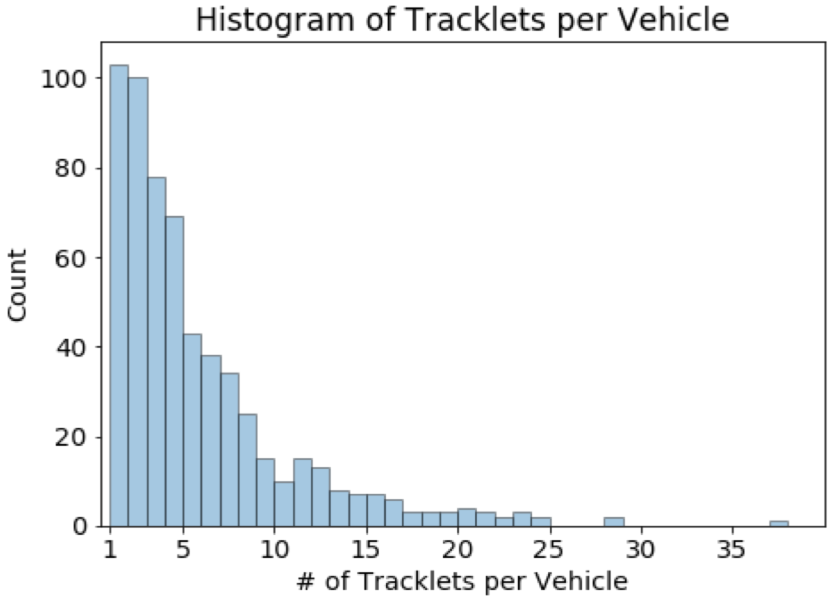
When the tracker loses the target vehicle it is reset to the true vehicle location (and a frame of bright red will be shown)

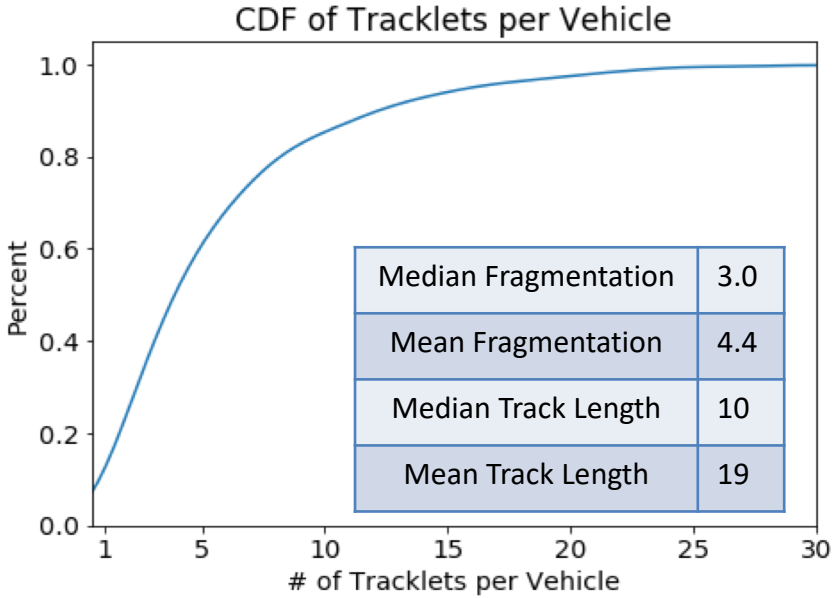


EXPEDITION TECHNOLOGY

Performance Metrics on WPAFB Dataset

Full Scene Single Object Tracker Performance – 153 Evaluation Frames





Median Fragmentation	3.0
Mean Fragmentation	4.4
Median Track Length	10
Mean Track Length	19

End-to-End Track Manager will further reduce fragmentation by re-associating fragmented tracks.

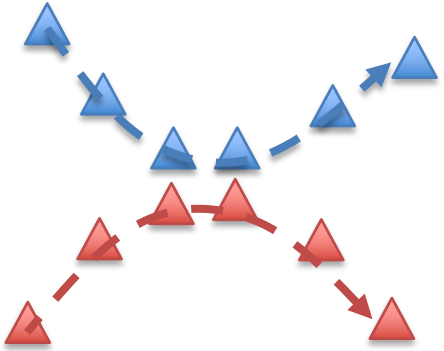
EXPEDITION TECHNOLOGY

NEXT STEPS:
END-TO-END TRACK MANAGEMENT

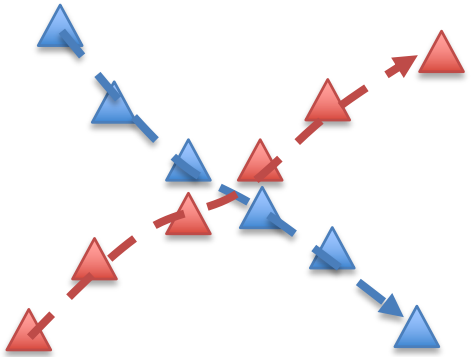
EXPEDITION TECHNOLOGY

End-to-End Tracking

- Given a set of detections and current tracks
  - Are all current detections mapped to an active track?
  - Are all active tracks still tracking a valid object?
- Traditional Association Methods
  - Global Nearest Neighbor
  - Hungarian/Munkres Algorithm / Linear Assignment
  - Multi-Hypothesis Tracking
  - Markov Chain Monte Carlo Methods
- Proposed Solution – Two Staged Approach
  - Run Detector -> Tracker in parallel
  - Run Supervisory RNN network to map detections to tracks to address death/birth/reassociation

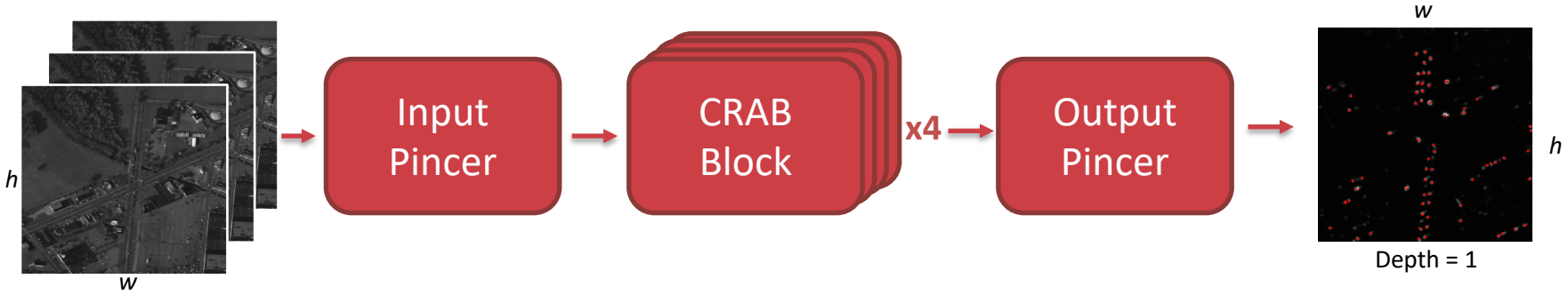


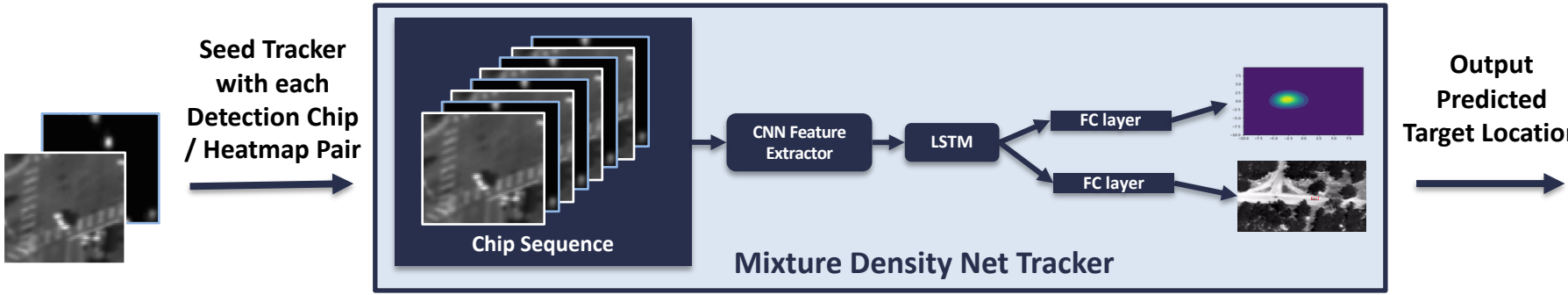
Which track is correct?



Output of Detector Seeds Tracker

Discrete CRABnet Mover Detections Feed the MDN Tracker Network





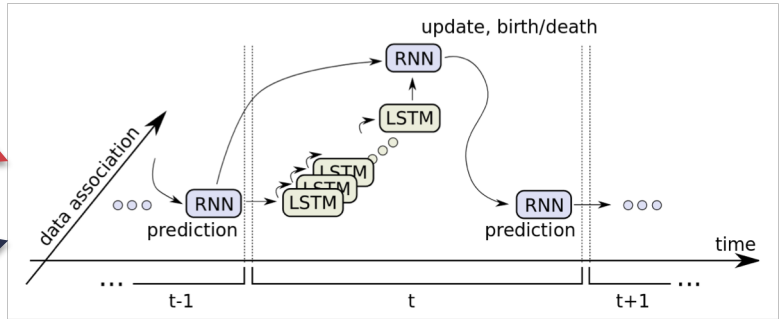
EXPEDITION TECHNOLOGY

Supervisory RNN Monitors Tracks/Detections

- Feed output of the Detector and Tracker into a RNN-based Online Multi-Target Track Association Network
  - Manage Birth, Death and Re-association of Tracks
  - Leverage **visual feature layers** from Tracker to assist re-association

CRABnet Detector

Mixture Density Net Tracker



Online Multi-target Tracking Using Recurrent Neural Networks
A. Milan, S. H. Rezatofighi, A. Dick, I. Reid, K. Schindler. In: AAAI 2017

EXPEDITION TECHNOLOGY

WRAP-UP

Future of Deep Learning and Overhead Imagery

- **Advancements in Imagery**
  - Higher resolution imaging sensors
    - 150+ MP single sensors (Phase One IQ4 150MP) @ 1.4 FPS
    - 35+ MP Video Sensors (RED MONSTRO/HELIUM 8K) @ 60 FPS
  - Advances in satellite constellations
    - Higher revisit rates / Improved resolutions
- **Advances in Deep Learning Hardware & Algorithms**
  - Faster Hardware with Increased Memory
    - i.e. V100 with 32GB enables larger base images
  - More efficient algorithms
    - Faster executing and faster converging architectures (MobileNet v2)

Advances in Imagery and Deep Learning Open the Door to New Opportunities

The Team

Tyler Balsam

Joe Succar

Ryan Crawford

Enrico Mattei

Stephen Johnson

Greg Harrison

EXPEDITION TECHNOLOGY

Contact Us

We're hiring!

Machine Learning, Image Processing,
RF Signal Processing, Navigation & Autonomy

https://www.exptechinc.com/pages/careers/

Ryan Crawford
Expedition Technology, Inc.
E-mail - ryan.crawford(at)exptechinc.com
(571) 429-6141

EXPEDITION TECHNOLOGY