ClusterNet: Detecting Small Objects in Large Scenes by Exploiting Spatio-Temporal Information

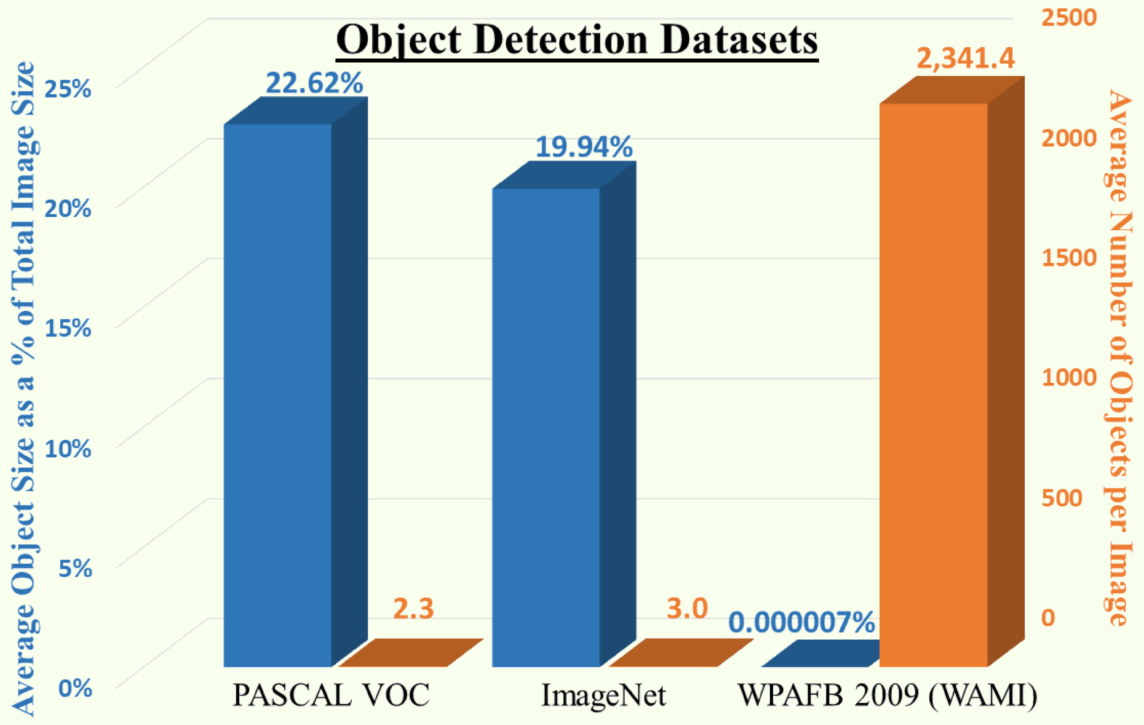
Rodney LaLonde lalonde@knights.ucf.edu
Dong Zhang dromstons@gmail.com
Mubarak Shah shah@crcv.ucf.edu

Center for Research in Computer Vision (CRCV), University of Central Florida

UCF Center for Research in Computer Vision

2018 CVPR Salt Lake City • June 18-22

Problem: Object Detection in Wide Area Motion Imagery



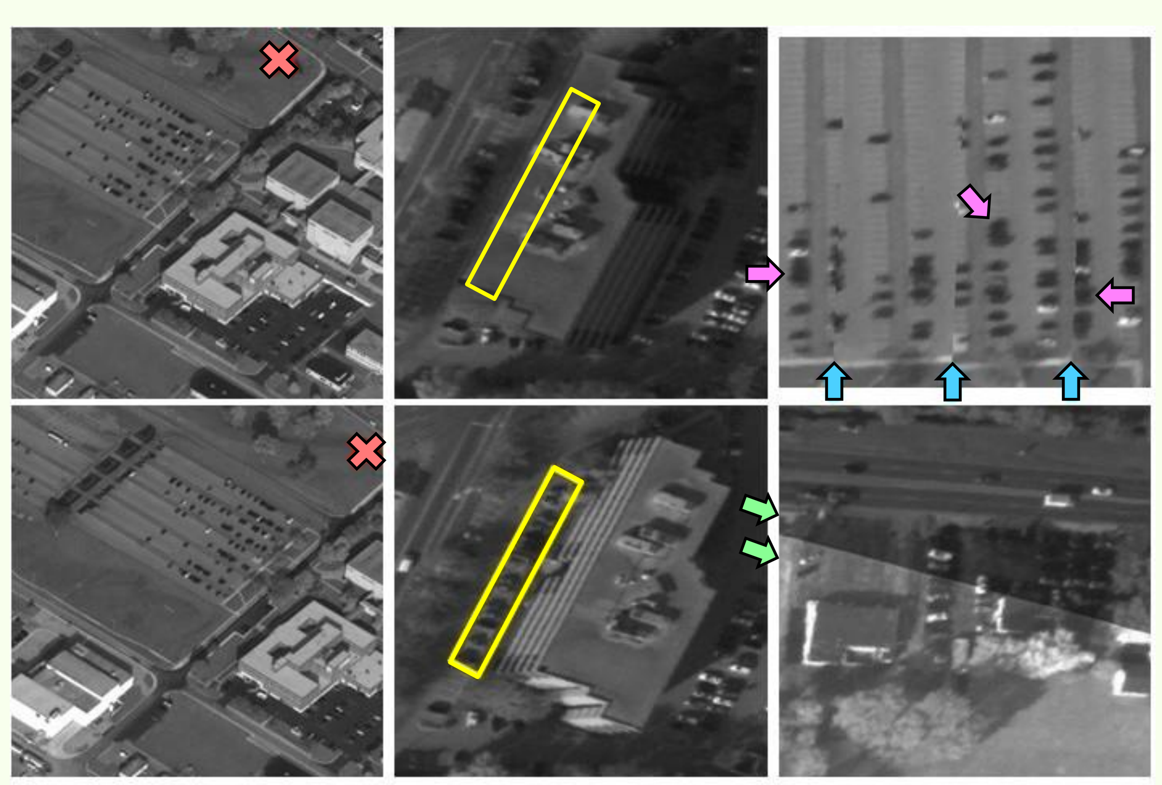
- ❑ WAMI video frames contain thousands of objects compared to 2-3 in "standard" object detection.
- ❑ WAMI objects are extremely small in relation to the search space.

WPAFB 2009 Dataset

- ❑ 6 slightly-overlapping cameras; 19 sq. km.; 1.25 fps; Grayscale;
- ❑ ~315 MP per frame; Vehicle size: ~9x18 pixels; Pixel size ~1/4 m^2
- ❑ ~2.4 million annotations across 1,025 frames of video.

Challenges

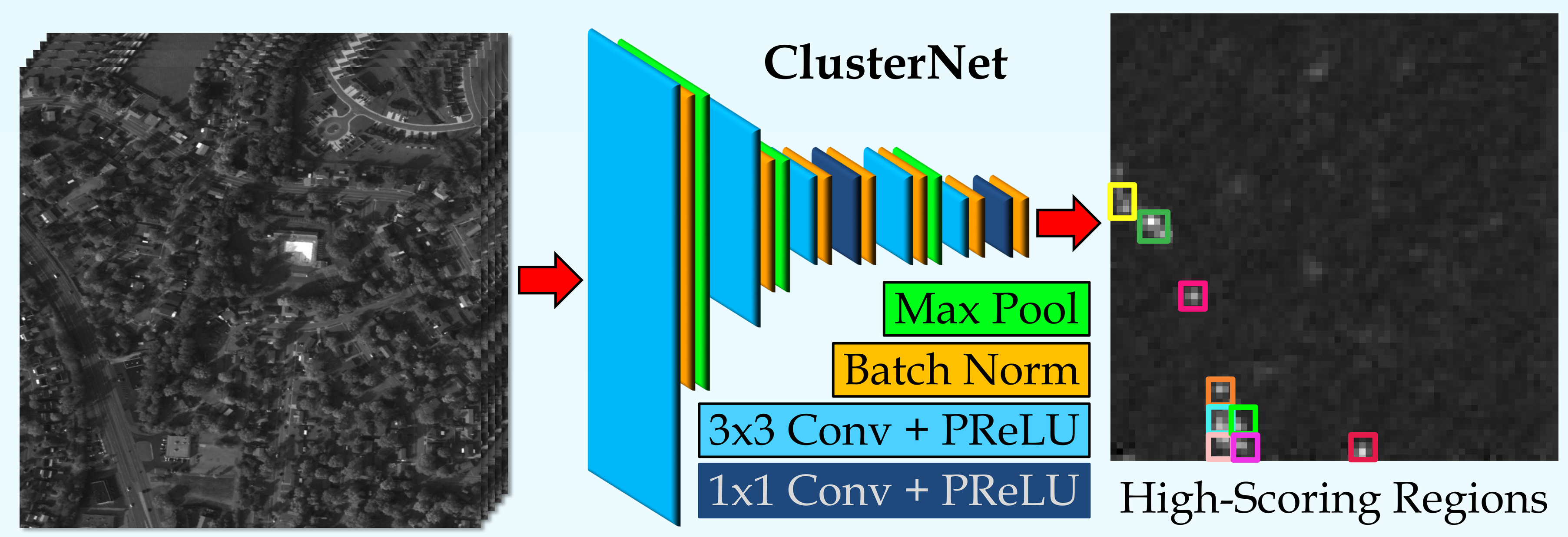
- ❑ Low frame rate
- ❑ Motion parallax
- ❑ Mosaic Seams
- ❑ Blurred/Unclear Object Boundaries
- ❑ Inconsistent Illumination



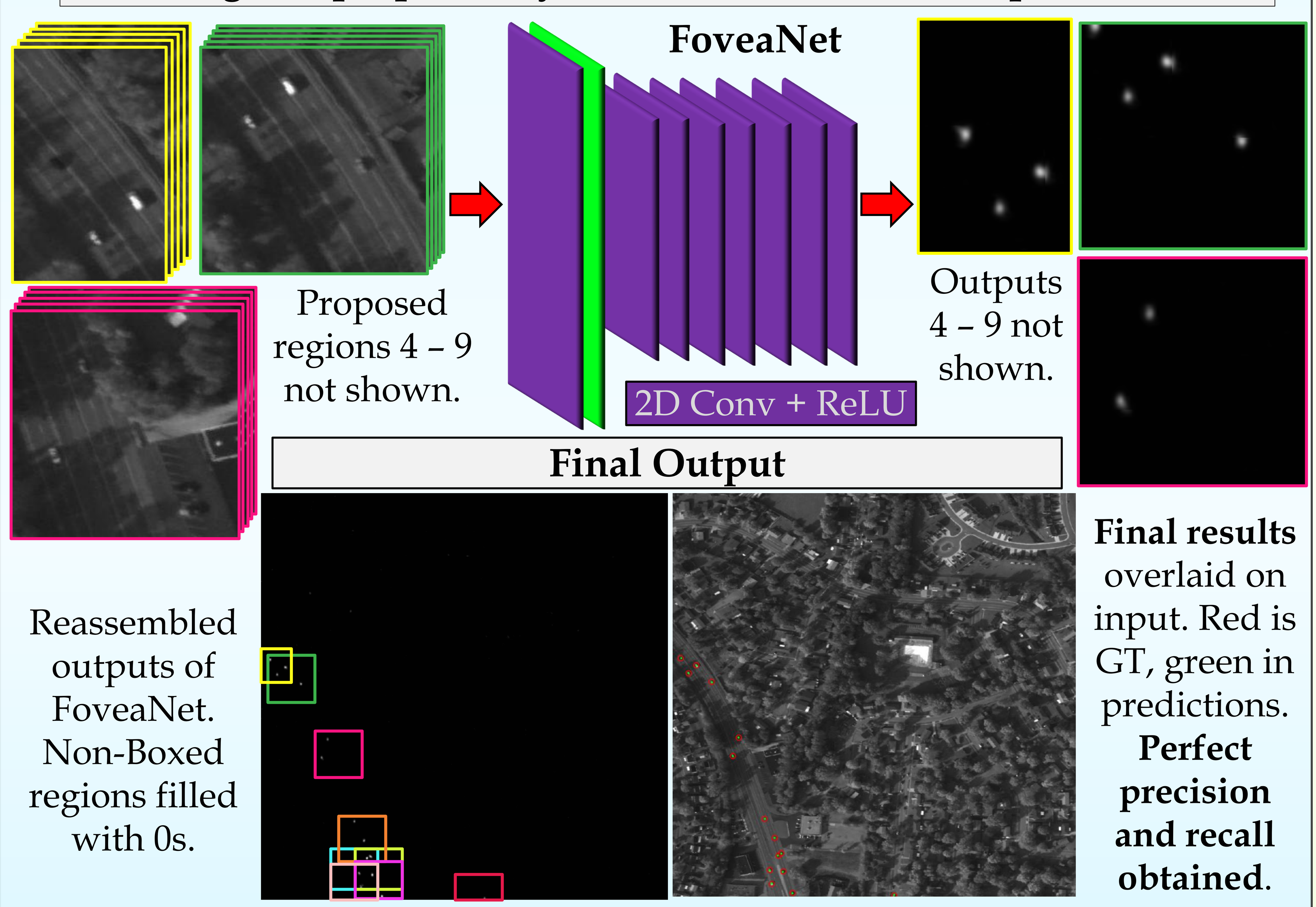
Proposed Method: ClusterNet & FoveaNet

Two-Stage Fully Convolutional Neural Network

- ❑ First Stage: ClusterNet detects Regions of Objects of Interest.
- ❑ ClusterNet effectively reduces the search space to regions containing one to several hundred vehicles per output neuron.
- ❑ Second Stage: FoveaNet extracts the effective receptive field of high-scoring output neurons for fine-grained localization.

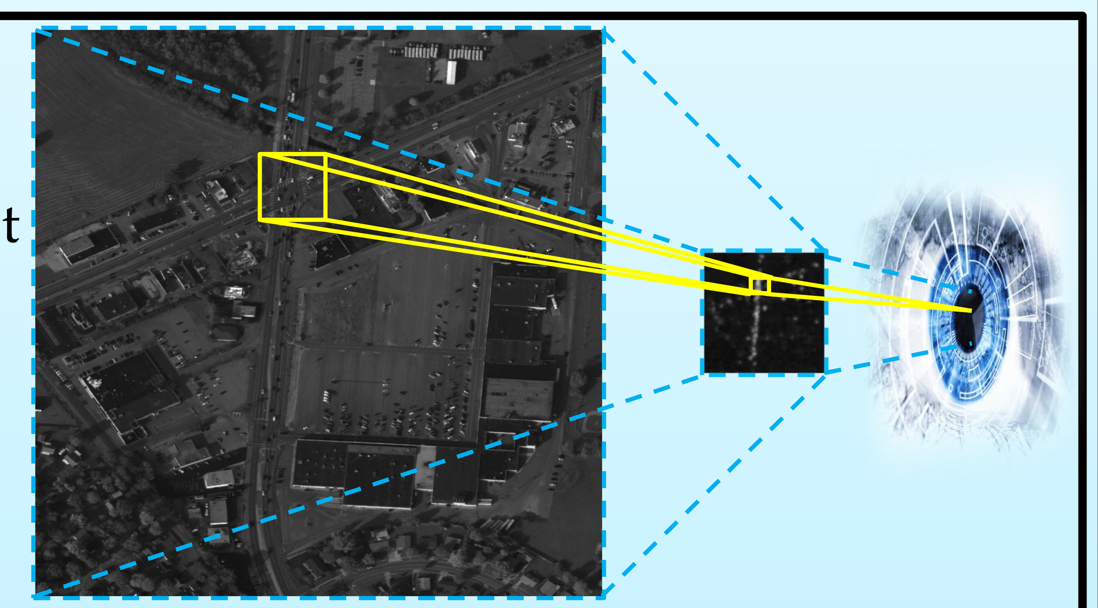


Regions proposed by ClusterNet: 9 out of 324 possible



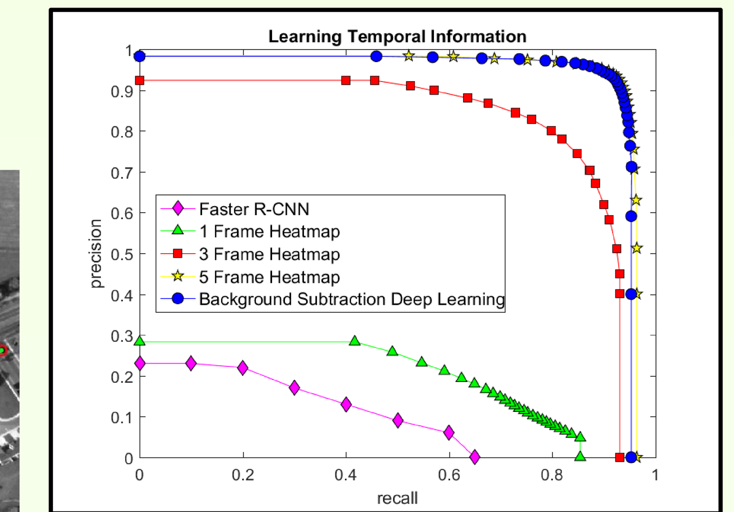
Final results overlaid on input. Red is GT, green in predictions. **Perfect precision and recall obtained.**

SUMMARY: ClusterNet effectively reduces the search space while FoveaNet scans the receptive field of high scoring neurons in ClusterNet's output to produce fine-grained localization of vehicles.



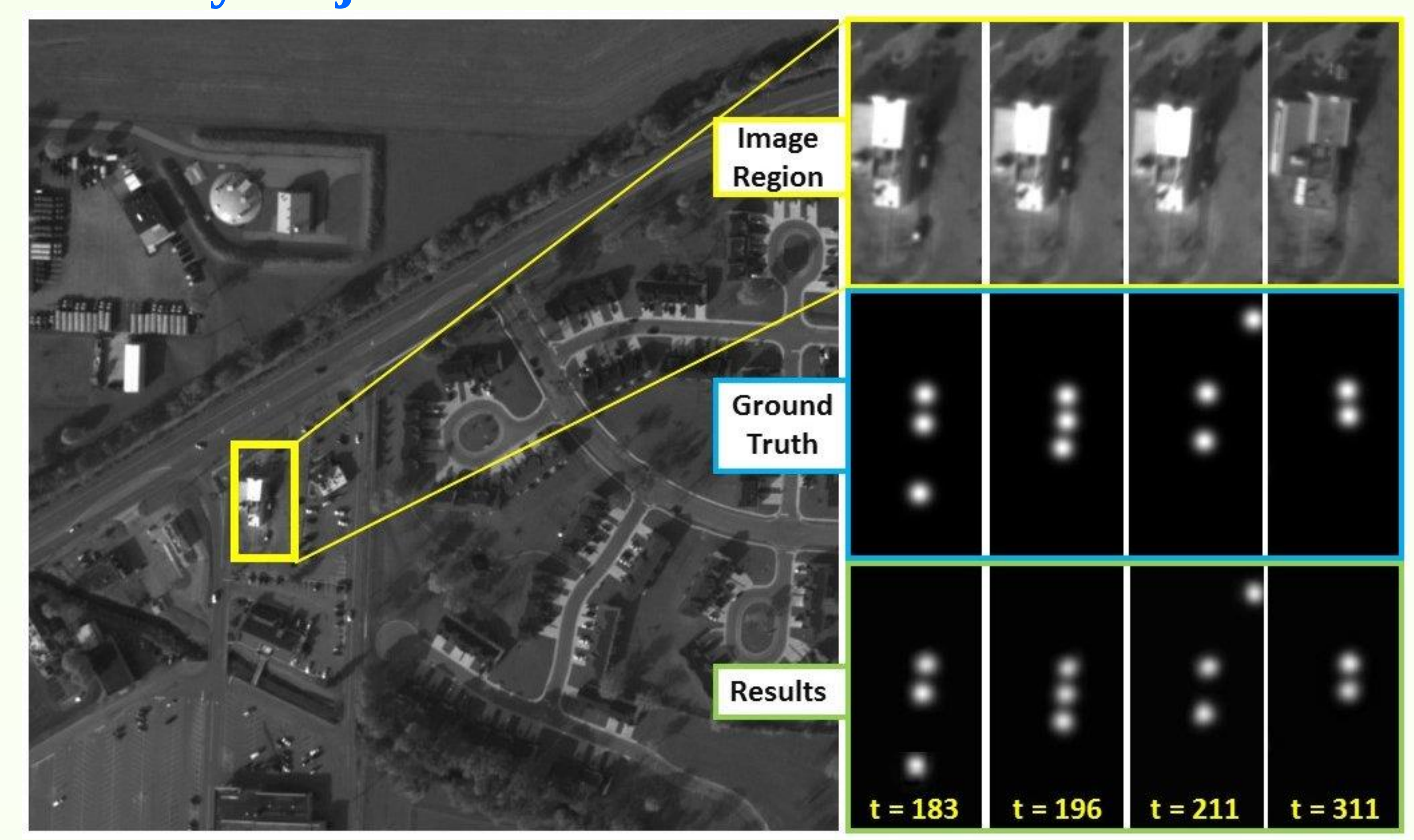
Qualitative Results:

Moving Object Detection Results

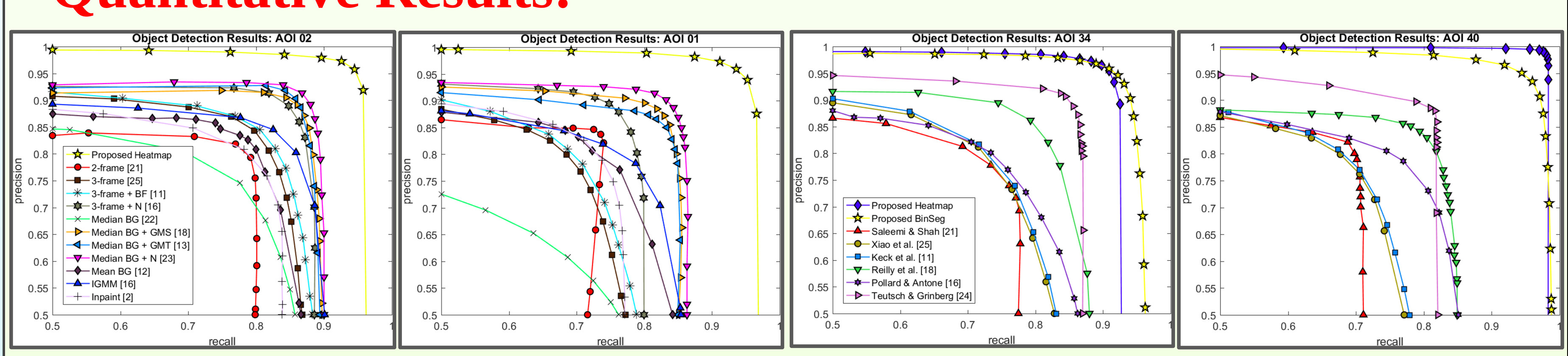


^ Explicit temporal relationship computed (background subtraction) vs. varying amounts of temporal information implicitly modeled by CNN vs. Faster R-CNN.

Stationary Object Detection Results



Quantitative Results:



Comparison of F_1 Scores on Eight Crop and Aligned Sections of the WPAFB 2009 Dataset

Method	01	02	03	04	34	40	41	42
Sommer *et al.* [27]	0.866	0.890	0.900	0.804	x	x	x	x
Shi [25]	0.645	0.760	0.861	0.575	x	x	x	x
Liang *et al.* [12]	0.842	0.880	0.903	0.760	x	x	x	x
Kent *et al.* [10]	0.767	0.807	0.668	0.711	x	x	x	x
Aeschliman *et al.* [2]	0.764	0.795	0.875	0.679	x	x	x	x
Pollard & Antone (3-frame + N) [16]	0.816	0.868	0.892	0.805	x	x	x	x
Saleemi & Shah [24]	0.783	0.793	0.876	0.733	0.755	0.749	0.762	x
Xiao *et al.* [29]	0.738	0.820	0.868	0.687	0.761	0.733	0.700	x
Keck *et al.* [9]	0.743	0.825	0.876	0.695	0.763	0.737	0.708	x
Reilly *et al.* [19]	0.850	0.876	0.889	0.783	0.826	0.817	0.799	x
Pollard & Antone (IGMM) [16]	0.785	0.835	0.776	0.716	0.766	0.778	0.616	x
Teutsch & Grinberg [28]	x	x	x	x	0.874	0.847	0.854	x
Prokaj & Medioni [17]	x	x	x	x	x	x	x	0.631
Proposed Method	**0.947**	**0.951**	**0.942**	**0.887**	**0.933**	**0.983**	**0.928**	**0.927**

Project Pages

http://crcv.ucf.edu/projects/FullyConvolutionalDNN/

https://bit.ly/2sTvhJu