Moving Object Detection with Deep Convolutional Networks and LSTM

Gregoire Robinson
Advisors: Dong Zhang and Dr. Mubarak Shah

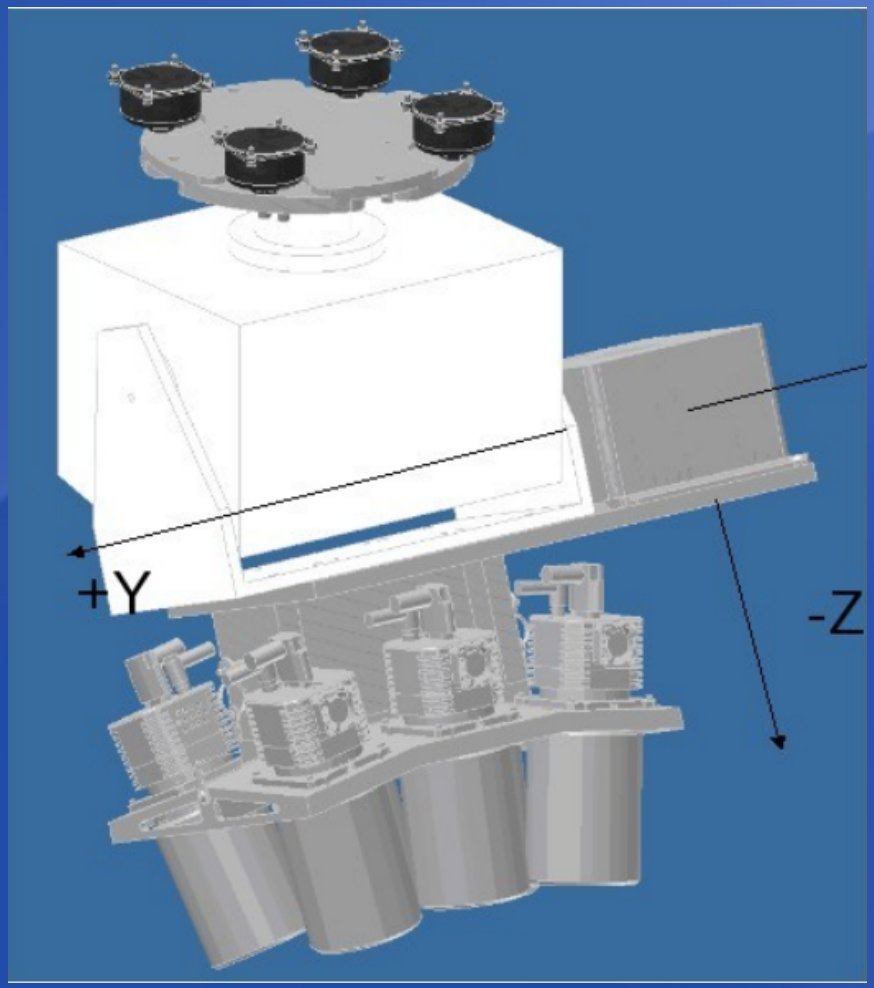
Week 2

- Background
- This week’s results

Background

- Objective is to detect vehicles on roads in wide area surveillance (WAS) video
- Large electro-optic platform flying at 7,000 ft.
- These are very large images (over 20,000 x 20,000 pixels)
- Thousands of objects per image, low resolution, low saliency of objects from background, and low frequency of image capture means this is a challenging problem

Background



Series of 6 cameras on EO platform

Background

- Used the Wright Patterson Air Force Base dataset from 2009
- Many potential applications for this technology
- Security surveillance
- Traffic monitoring/management
- Emergency management

Results

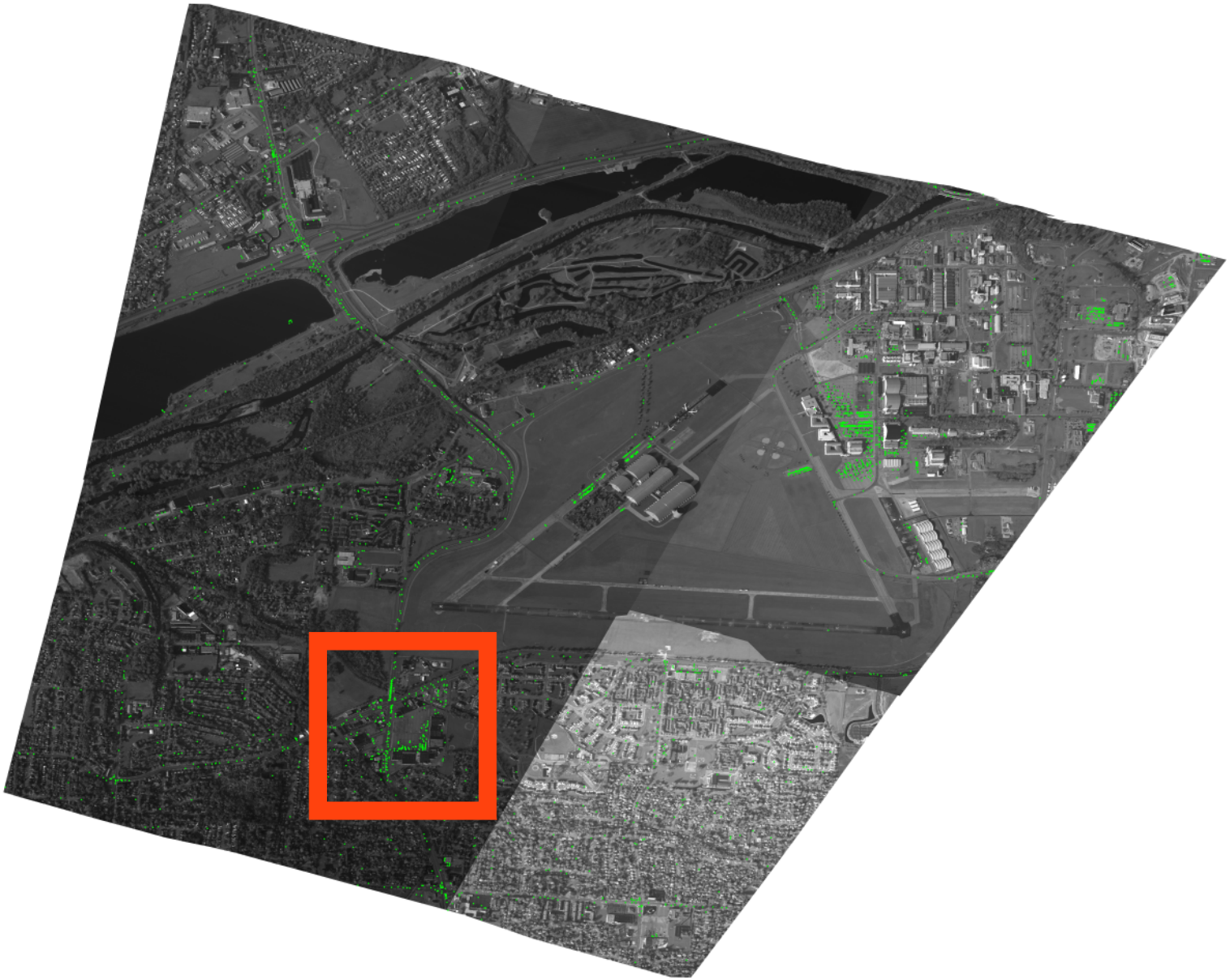
1) Since last week, obtained full WPAFB 2009 dataset

2) Verified ground truth annotations from dataset (see subsequent slides for examples of this)

3) Generated a heat map for a single image

4) Coded a 5 layer CNN: 2DConv-MaxPool-2DConv-MaxPool-2DConv

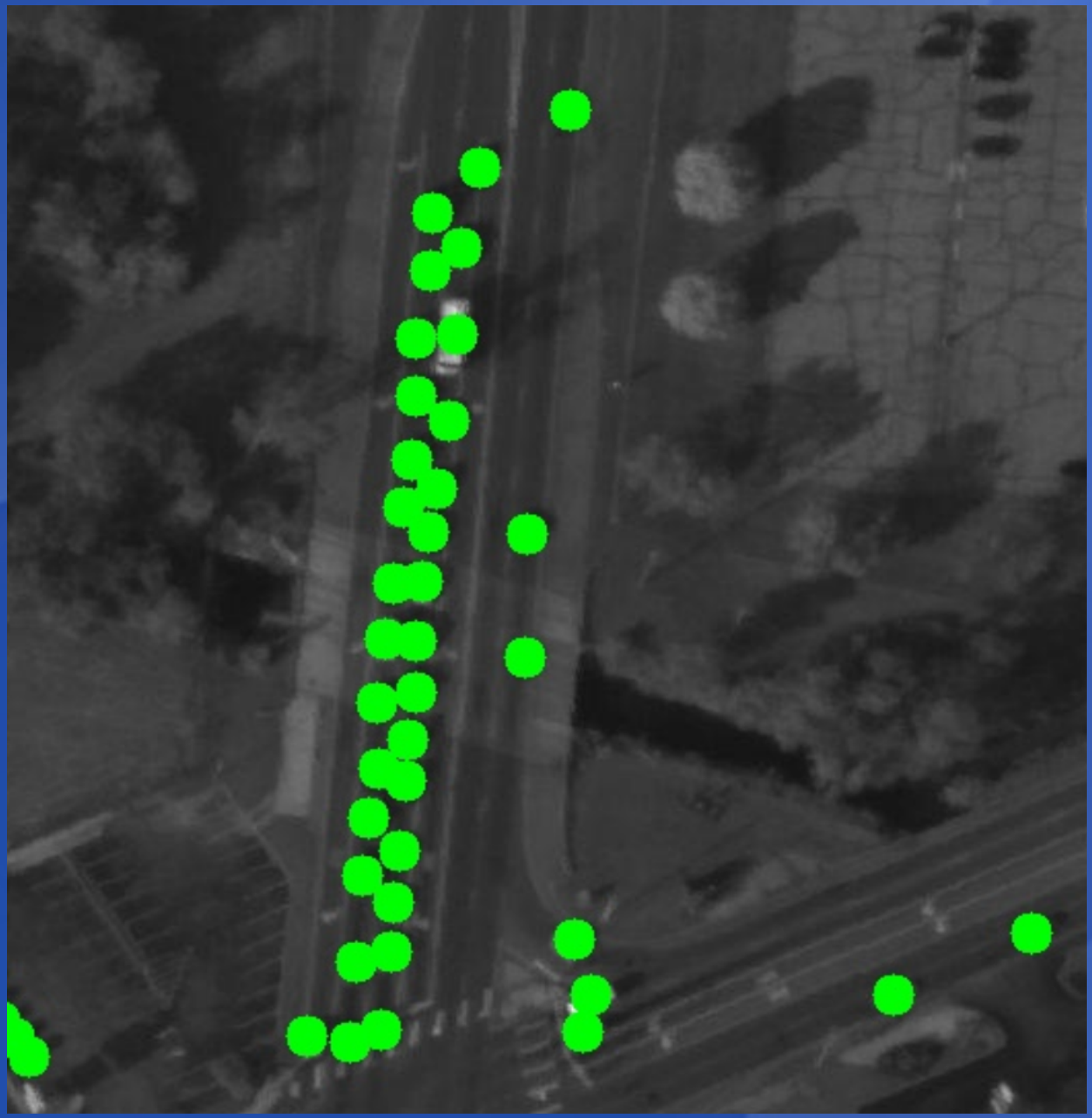
Results: Ground Truth Annotations Verified



Results: Ground Truth Annotations Verified



Results: Ground Truth Annotations Verified



Results

1) Since last week, obtained full WPAFB 2009 dataset

2) Verified ground truth annotations from dataset (see subsequent slides for examples of this)

3) Generated a heat map for a single image

4) Coded a 5 layer CNN: 2DConv-MaxPool-2DConv-MaxPool-2DConv

Results

- **Generated a heatmap for a single image:** Overlaid a 5 x 5 pixel 2D Gaussian over every truth point. Every Gaussian then mostly demarcates a vehicle. This can be used as a training label for the CNN. Each image is over 30,000 x 30,000 pixels, so I was only able to process a single image for the moment. When I have access to a GPU, I will be able to process the entire dataset and incorporate object movement.
- Since the images are of different dimensions and are large, I resized them to have equal dimensions and sliced them into 512 x 512 tiles.
- I used the heatmap tiles as training labels for the CNN, and I used the original image (split into 512 x 512 tiles) as the training set.

Results

- Dong gave me a preliminary 5 layer CNN to code: 2D Convolution Layer - 2D MaxPool Layer - 2D Convolution Layer – 2D MaxPool Layer - 2D Convolution Layer
- I successfully finished coding that, but unfortunately, my laptop is not able to perform the computation, which is too intensive.
- Currently, I’m working with IT to gain access to a GPU cluster. Steve Dick has been very helpful, but there have been some issues installing certain libraries such as OpenCV.

Next Week:

- Continue implementing and training CNN