Vehicle Detection in Wide Area Surveillance Imagery with Deep Convolutional Networks

CENTER FOR RESEARCH IN COMPUTER VISION

Gregoire Robinson
gregoirerobi@umass.edu
University of Massachusetts Amherst

Dong Zhang
dzhang@eecs.ucf.edu
University of Central Florida

I. Problem & Motivation

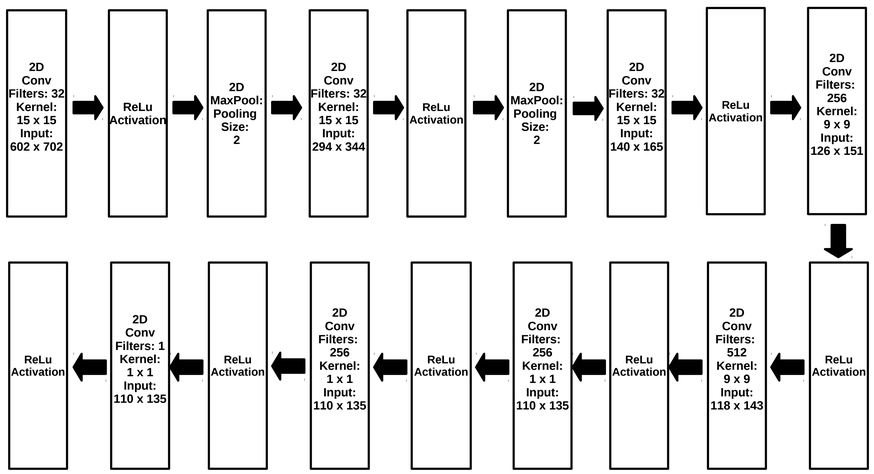
- Detect moving vehicles in Wide Area Surveillance (WAS) imagery
- WAS images are recorded by a UAV electro-optic platform flying at ~7000 feet
- Applications in security, traffic monitoring, and emergency management
- Data set is from Air Force Research Laboratory Columbus Large Image Format (CLIF) 2007 and PV Labs
- Ground Truth Annotations:



II. Challenges & Contribution

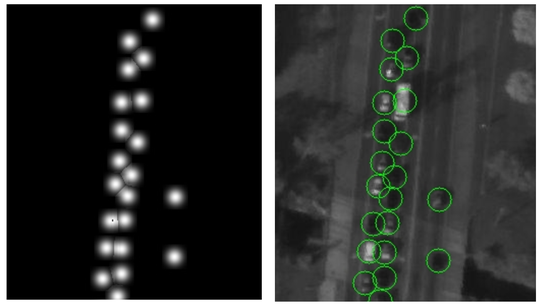
- Thousands of objects per image, low resolution, low object saliency from background, and low frequency of image capture means this is a challenging problem
- Previous attempts have used median background modeling and frame differencing
- Convolutional neural network is trained to recognize both appearance features and frame differences

III. Convolutional Neural Network



III. Heatmaps

- Neural nets generally use words as training labels
- Instead, create a two-dimensional Gaussian matrix at each ground truth annotation point
- Use the resulting heatmap as a "label" to train a convolutional neural network



Heatmap *Input Image*

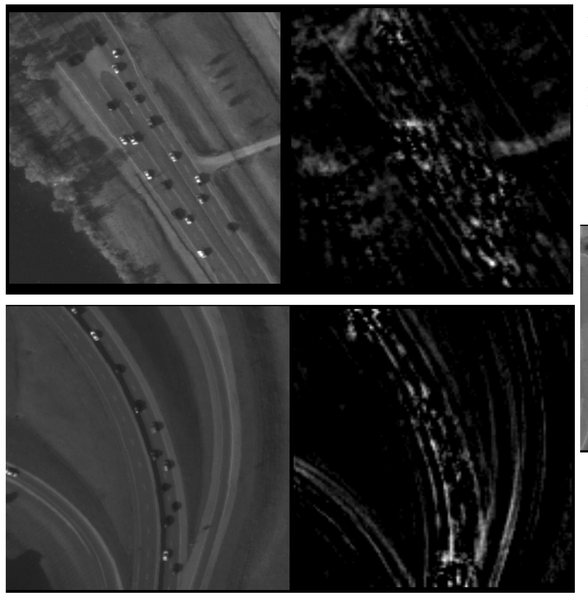
IV. Image Preprocessing

- CLIF and PV Labs raw data comes from a series of multiple cameras recording simultaneously and must first be stitched together
- Resulting stitched images are on the order of ~20K x 20K pixels
- Stitched images are then registered to compensate for motion and sliced into smaller sub-images
- Sub-images are used as training input for convolutional network

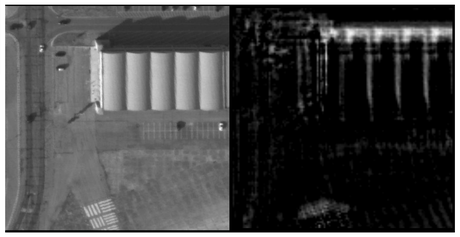


- During training images are progressively convolved and resized from 602 x 702 to 110 x 135 pixels
- Network output is regressed on a single heatmap

V. Results CLIF Dataset



- Due to low object saliency, network primarily detects background
- Some vehicles are detected but results are not sufficiently meaningful



Sharp edges found between roads and buildings as well as road markings interfere with network learning

VI. Results PV Labs Dataset