Object Tracking and Geo-localization from Street Images

Daniel Wilson, Xiaohan Zhang, Kamiku Xue, and Safwan Wshah
Department of Computer Science
University of Vermont

Introduction

Our goal is to construct an automated system which processes a sequence of road images and detects, classifies, and geo-localizes signs visible in the images. We will construct an interactive web widget enabling easy visualization on a satellite map. We have also built a large traffic sign recognition dataset to support future research projects in the field.

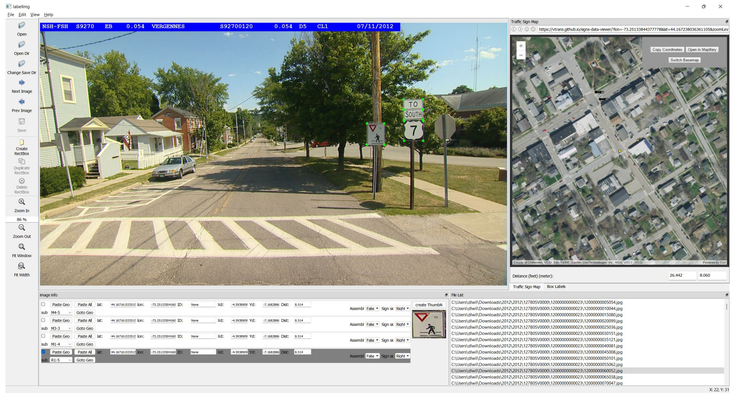

Figure 1. The annotation tool use to label signs during construction of our dataset.

Methods

We have constructed a new object detection model by modifying Cascade RCNN, a deep learning architecture which uses a cascade of networks to generate progressively more accurate predictions, with the capability to geolocate signs. A second neural network, called the similarity network, is trained to receive a pair of detections as input to predict the probability that two detections depict the same sign. Finally, our system employs the Hungarian algorithm to compute the optimal assignments to merge repeated detections from the same signs.

To visualize the detection results, we customized an ArcGIS web application named *VailViewer*. The user can navigate through street view images collected by VTrans in different years. The detected signs are visualized in both the map and street view images.

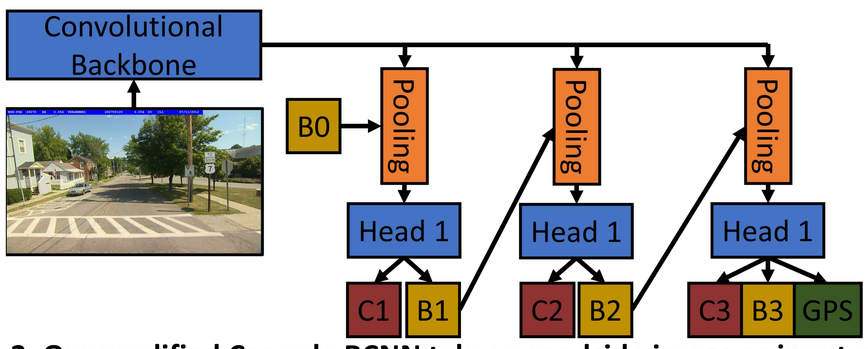

Figure 2. Our modified Cascade RCNN takes a roadside image as input and predicts the class, bounding box, and GPS of each visible sign.

Results

Our full pipeline steps though each image in a sequence, uses our detector to generate sign detections, computes the similarity between sequential detections with our similarity network, and then merges repeated detections with the Hungarian algorithm, resulting in a list of sign predictions as the final output. This end-to-end system currently achieves 92.5% recall and 80% precision.

Our web widget is already in use by VTrans and is capable of visualizing image and sign locations on an interactive map enabling efficient assessment and planning of asset management.

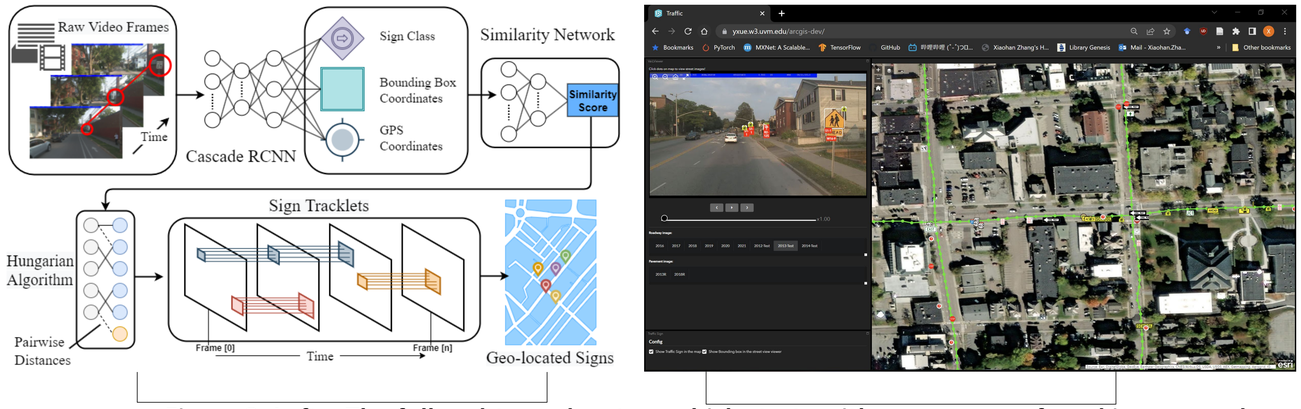

Figure 3. Left – The full end-to-end system which starts with a sequence of road images and ultimately outputs geo-localized sign predictions. Right – A screenshot of *VailViewer*.

Future Work

We have constructed an unlabeled dataset from additional images we have gathered and are implementing self-supervised learning techniques to enhance the performance of our detection model. We are augmenting our tracking algorithm with the capability to merge repeated detections across separate years in which the vehicle drove through the same road segments. We are expanding the capabilities of the VailViewer widget in coordination with VTrans to support various requested features such as fast traveling speed and shortcut keys in full screen mode.

Acknowledgments

We would like to acknowledge *Rick Scott* and *Alex Geller* for supporting this research.

References

Wilson, D.; Alshaabi, T.; Van Oort, C.; Zhang, X.; Nelson, J.; Wshah, S. Object Tracking and Geo-Localization from Street Images. Remote Sens. 2022, 14, 2575. https://doi.org/10.3390/rs14112575

Z. Cai and N. Vasconcelos, "Cascade R-CNN: High Quality Object Detection and Instance Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 5, pp. 1483-1498, 1 May 2021, doi: 10.1109/TPAMI.2019.2956516.