# Object Tracking and Geo-localization from Street Images

Daniel Wilson and Safwan Wshah
Department of Computer Science
University of Vermont

## Introduction

The goal of our research is to develop an automated system which processes a sequence of images, detects and classifies the visible signs, and predicts their respective GPS coordinates. Our project introduces the largest available dataset to serve as a benchmark in the domain of Traffic Sign Recognition (TSR).

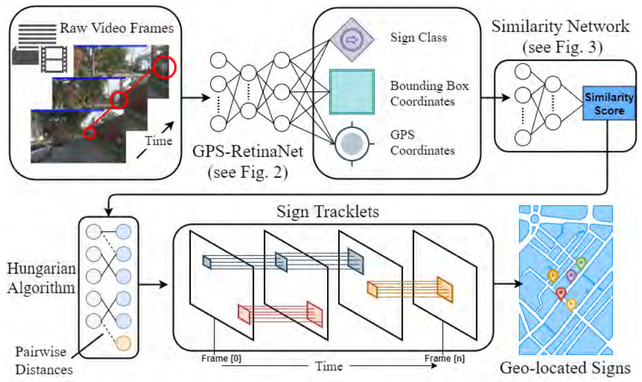


Figure 1. The full end-to-end system takes raw image sequences as input and ultimately outputs geolocalized sign predictions.

## Methodology

We constructed a modified variation of the Deep Learning architecture RetinaNet to detect signs and predict distance offsets which are converted to GPS predictions using a coordinate transform. We train a second a neural network to compute a similarity score between each pair of RetinaNet detections, and then use the Hungarian Algorithm to condense similar detections into individual predictions of each separate sign.

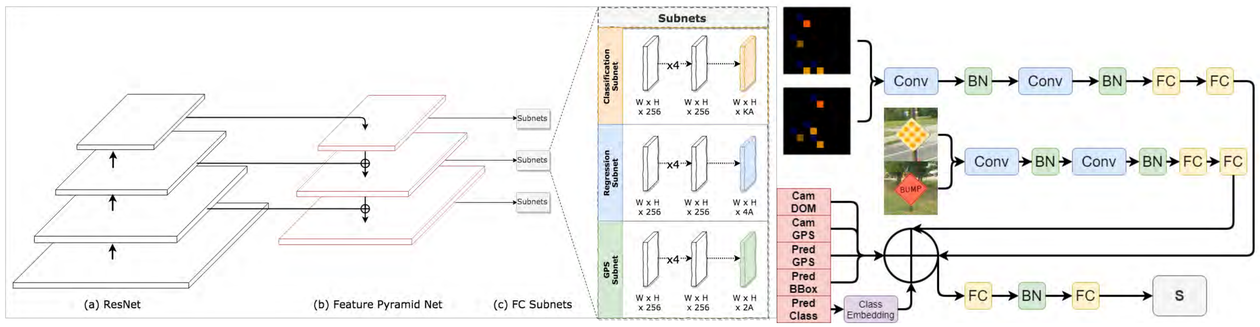


Figure 2. GPS-Retinanet (left) is our modified version of RetinaNet which predicts GPS offsets in addition to classifying objects. Our custom neural network (right) learns to compute similarity scores between detections.

## Results

Our dataset contains 199 classes of signs, 74,326 high resolution images, and 47,589 annotations, making it the largest TSA dataset available. Each annotation specifies a sign's GPS, class, bounding box, road side, assembly attribute, and contains a unique physical identifier. Our object detector achieves a 75th percentile mean average precision of 85%. The system scores a mean of 4.3 meters geospatial margin of error.

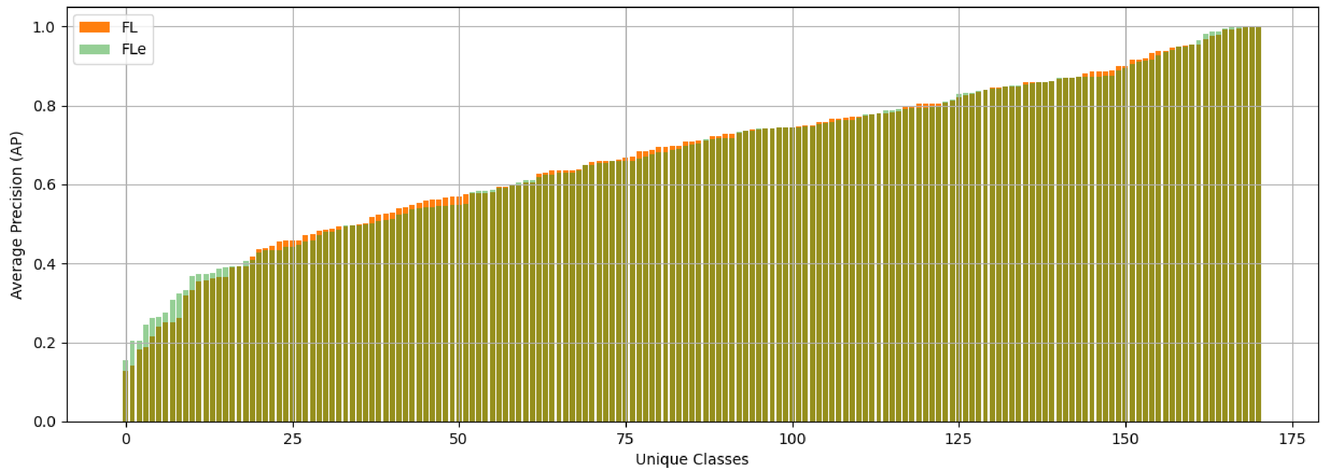


Figure 3. The distribution of Average Precision values (an evaluation metric commonly used in the field of object detection) across each class in the dataset.

## Conclusion

Our system automates the task of locating road assets from road images, which can be used in logistical planning and maintenance assessments. We introduce a large TSR dataset to support future research in this field.

## Acknowledgments

We would like to acknowledge Rick Scott, Ken Valentine and Alex Geller for supporting our project.

## References

Møgelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey." IEEE Trans. Intelligent Transportation Systems,vol. 13, no. 4, pp. 1484–1497, 2012.

F. Almutairy, T. Alshaabi, J. Nelson and S. Wshah, "ARTS: Automotive Repository of Traffic Signs for the United States," in IEEE Transactions on Intelligent Transportation Systems, doi: 10.1109/TITS.2019.2958486.