A Comparison of Deep Learning Methods for Two-Wheeler Detection in Images

Michel Jelsma & Douwe Pieter Reitsma
Supervisors: Klaas Dijkstra & Lucas Ramos

Winter 2022

Introduction

- More than 10000 cyclists are hospitalized with severe injuries annually in the Netherlands [1].
- Traffic safety analysis is currently performed by a trained expert who must either physically stand near a road or sift through hours of video footage manually.
- Deep learning methods could speed up this process significantly.
- External datasets might have missing annotations for certain classes causing models to forget certain features during training.
- We constructed an ensemble of models to prevent this.
- Since ensembles are relatively slow, we also constructed a ''surrogate'' model, trained using annotations generated by the ensemble.

Materials and Methods

Models [2]:
- YOLOv5s & YOLOv5m
- Faster-RCNN
- Ensemble of YOLOv5s
- Surrogate model

Data Augmentations [3]:
- ChannelShuffle
- HorizontalFlip
- Rotate

(External) Datasets:
- Berkley Deep Drive 100k [4]
- Specialized Cyclist Dataset [5]
- MB10000 [6]
- Custom Traffic Intersection Dataset (TID)
  - Used for training and testing.

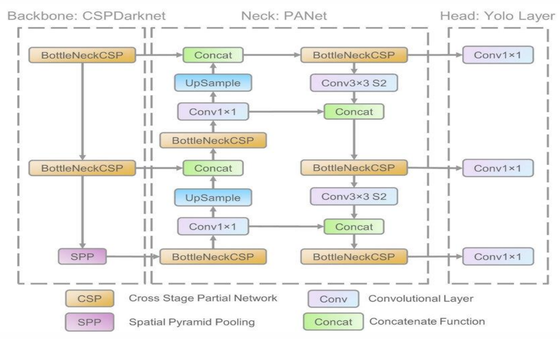

Figure 1. YOLOv5 architecture [7].

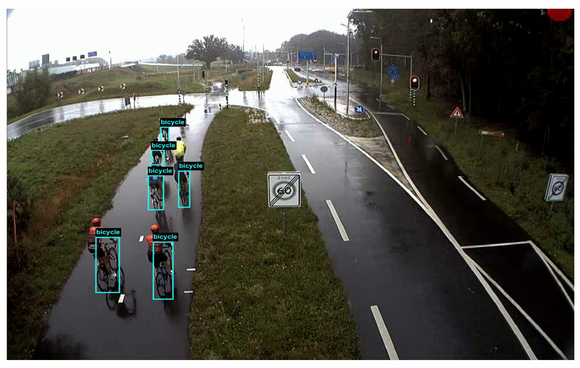

Figure 2. Sample image from our traffic intersection dataset.

Abstract

Manually analyzing traffic is a laborious task. Deep learning techniques could help speed up this process significantly. In this research we compared several detection models to assess which model has the best accuracy for detecting two-wheelers in traffic camera footage. We concluded that Faster-RCNN had the best accuracy with 0.88 mAP & 0.81 F1-score followed by YOLOv5s with 0.73 mAP and 0.78 F1-score.

Materials and Methods

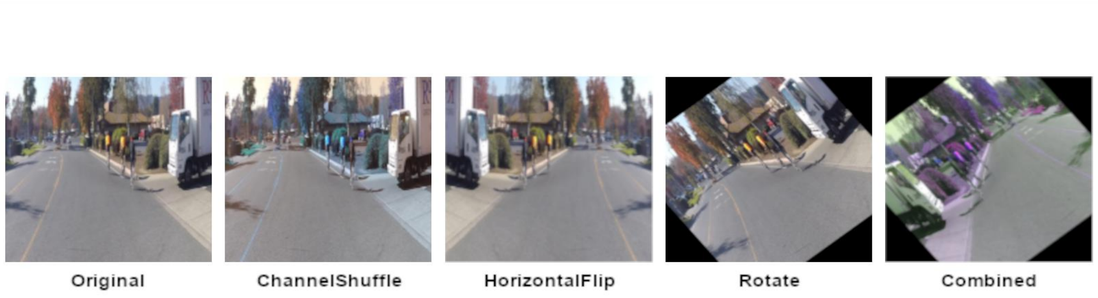

Figure 3. Data augmentations.

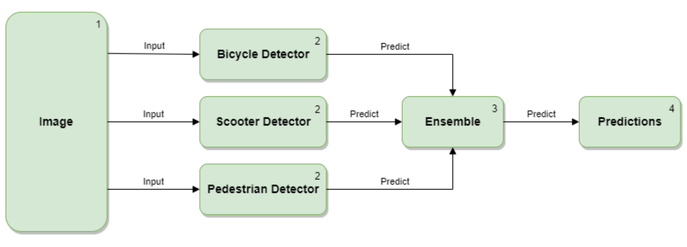

Figure 4. Ensemble pipeline.

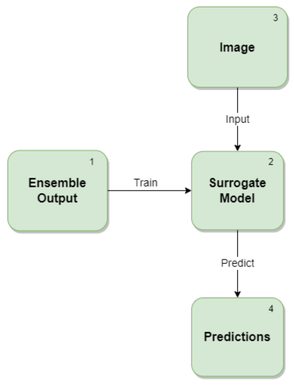

Figure 5. Surrogate pipeline.

Experiments and Results

Comparing object detection models
- We compared YOLOv5s, YOLOv5m and Faster-RCNN to make a substantiated choice on which model to use for creating the ensemble and surrogate models.

Ensemble & surrogate learning
- We compared the base models to:
  - A fine-tuned YOLOv5s model trained on a combination of the specialized external datasets.
  - An ensemble of fine-tuned YOLOv5s models.
  - A fine-tuned surrogate model pre-trained on the combined specialized datasets with annotations generated by the ensemble model.

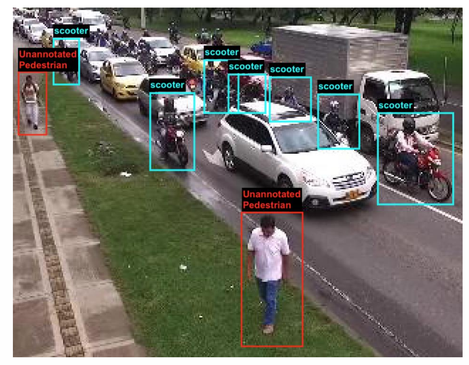

Figure 6. Example of an image with missing annotations.

Experiments and Results

Model	Pre-trained	mAP	F1-score
Faster-RCNN	ImageNet	**0.88**	**0.81**
YOLOv5s	COCO	0.73	0.78
YOLOv5m	COCO	0.72	0.74

Table 1. Comparing object detection models.

Model	Pre-trained	mAP	F1-score
Combined external datasets	COCO, BDD100k, SCD & MB10000	**0.68**	**0.76**
Ensemble YOLOv5s	COCO, BDD100k, SCD & MB10000	0.18	0.51
Surrogate	COCO, BDD100k, SCD & MB10000	0.67	0.73

Table 2. Ensemble & surrogate learning

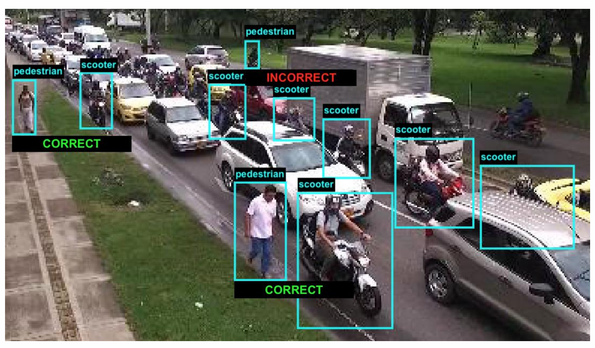

Figure 7. Example of (mostly) correctly generated annotations by the ensemble model for an image in the MB10000 dataset.

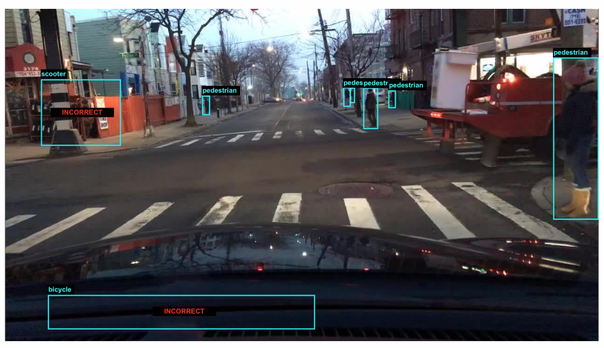

Figure 8. Example of incorrect annotations generated by the ensemble model for an image in the BDD100k dataset.

Conclusions

- Faster-RCNN is better at detecting two-wheelers in the TID than YOLOv5.
- Creating an ensemble led to worse results.
- The surrogate model performed better than the ensemble but worse than the base models.

References

[1] van der Horst, A. R. A., de Goede, M., de Hair-Buijssen, S., & Methorst, R. (2014). Traffic conflicts on bicycle paths: A systematic observation of behaviour from video. Accident Analysis & Prevention, 62, 358–368.
[2] Gandhi, R. (2018, 12). R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms.
[3] Buslaev, A., Iglovikov, V. I., Khvedchenya, E., Parinov, A., Druzhinin, M., & Kalinin, A. A. (2020). Albumentations: Fast and flexible image augmentations. Information, 11(2).
[4] Yu, F., Xian, W., Chen, Y., Liu, F., Liao, M., Madhavan, V., & Darrell, T. (2018). Bdd100k: A diverse driving video database with scalable annotation tooling. *arXiv preprint arXiv:1805.04687*, *2*(5), 6.
[5] Masalov, A., Matrenin, P., Ota, J., Wirth, F., Stiller, C., Corbet, H., & Lee, E. (2019, June). Specialized Cyclist Detection Dataset: Challenging Real-World Computer Vision Dataset for Cyclist Detection Using a Monocular RGB Camera. In *2019 IEEE Intelligent Vehicles Symposium (IV)* (pp. 114-118). IEEE.
[6] Espinosa, J. E., Velastin, S. A., & Branch, J. W. (2018). Motorcycle detection and classification in urban Scenarios using a model based on Faster R-CNN.
[7] A Forest Fire Detection System Based on Ensemble Learning - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/The-network-architecture-of-Yolov5-It-consists-of-three-parts-1-Backbone-CSPDarknet_fig1_349299852 [accessed 13 Jan, 2022]

Acknowledgements

This project is financially supported by Regieorgaan SIA (part of NWO) under the KIEM SI project STATS.

NHL STENDEN computer vision & data science