Turkish Journal of Forecasting vol. 06 no. 1 (2022) pp. 40-45

Content list available at JournalPark

Turkish Journal of Forecasting

Journal Homepage: tjforecast.xyz

YOLOv5-based Vehicle Objects Detection Using UAV Images

Z. N. Duman[1,*], M. B. Çulcu[2], O. Katar[3]

[1]*Firat University, Faculty of Technology, Department of Civil Engineering, Elazig, Turkey*
[2]*Firat University, Faculty of Technology, Department of Energy Systems Engineering, Elazig, Turkey*
[3]*Firat University, Faculty of Technology, Department of Software Engineering, Elazig, Turkey*

A R T I C L E   I N F O

Article history:

Received	19	July	2022
Revision	24	August	2022
Accepted	25	August	2022
Available online	31	August	2022

Keywords:

Object detection
Vehicle
YOLOv5
Deep learning
UAV

R E S E A R C H   A R T I C L E

A B S T R A C T

Traffic is the situation and movement of pedestrians, animals, and vehicles on highways. The regulation of these movements and situations is also a basic problem of traffic engineering. It is necessary to collect data about traffic in order to produce suitable solutions to problems by traffic engineers. Traffic data can be collected with equipment such as cameras and sensors. However, these data need to be analysed in order to transform them into meaningful information. For a difficult task such as calculating and optimizing traffic density, traffic engineers need information on the number of vehicles to be obtained from the image data they have collected. In this process, artificial intelligence-based computer systems can help researchers. This study proposes a deep learning-based system to detect vehicle objects using YOLOv5 model. A public dataset containing 15,474 high-resolution UAV images was used in the training of the model. Dataset samples were cropped to 640×640px sub-images, and sub-images that did not contain vehicle objects were filtered out. The filtered dataset samples were divided into 70% training, 20% validation, and 10% testing. The YOLOv5 model reached 99.66% precision, 99.44% recall, 99.66% mAP@0.5, and 89.35% mAP@0.5-0.95% during the training phase. When the determinations made by the model on the images reserved for the test phase are examined, it is seen that it has achieved quite successful results. By using the proposed approach in daily life, the detection of vehicle objects from high-resolution images can be automated with high success rates.

1. Introduction

One of the most important inventions of humanity to make life easier is the invention of the wheel. Loads that were transported using human and animal power in the past began to be transported by wheeled vehicles later on. Wheeled vehicles have developed with the effect of industrialization and technology and have become an important component of transportation systems [1]. However, with the increasing population, the number of vehicles above the structural limits creates negative consequences such as traffic jams, accidents, and delays [2]. In order to eliminate these negativities, many traffic planning and analysis studies are carried out by researchers. Vehicle count and distribution should be collected in a traffic planning study. Data specified by traditional methods such as fixed cameras, and sensors can be collected, and this data can often produce accurate results [3]. However, traditional methods are not efficient methods for resource management. In addition to traditional methods, collecting traffic data

* Corresponding author.
E-mail addresses: znduman@firat.edu.tr (Zeynep Nur Duman)

https://doi.org/10.34110/forecasting.1145381

using unmanned aerial vehicles is a very popular approach today [4]. Unmanned Aerial Vehicles (UAV) are complex robots that can fly autonomously or with the aid of a ground station. UAVs are very suitable for use in traffic monitoring due to their size, cost, and mobility [5]. By integrating the camera and other necessary hardware parts, UAVs can be given the ability to vision. The UAVs equipped with a vision feature have a great potential for development for studies on topics such as parking lot management, traffic control, and disaster rescue [6]. When parameters such as flight speed, altitude, and viewing angle are taken into account, high-resolution images can be obtained with the help of UAVs. These images must be analysed and converted into meaningful information in order to be used in applications. Images are analysed in two different ways, manually and autonomously. In manual analysis methods, only a real person assigned for this job is needed. This person is required to detect the vehicle objects one by one by examining the images transferred from the UAV on a frame-by-frame basis. Vehicle detection is difficult due to factors such as vehicle objects occupying small areas in high-resolution images, and the presence of images containing fog-like weather conditions [7]. However, manual image review takes a lot of time and effort. Vehicle detection becomes an error-prone process with the emergence of human factors such as fatigue due to the excess time spent. Computers with fast processing capacity and the ability to be unaffected by human factors can help with related tasks. Thanks to the developments in computer hardware technology, the number of areas where autonomous technologies are used has also increased [8]. The main purpose of developing an autonomous system is to minimize the human factor. In this way, the margin of error can be minimized in the task performed. Artificial intelligence-based solutions are actively used for the autonomous system to analyse an image by itself [9]. Although classical image processing methods can produce results using fewer resources, artificial intelligence-based systems are more preferred due to their high success rates. Today, studies involving the use and development of deep learning models for the classification and detection of images have gained attention [10]. Chen et al. [11] proposed a system that automatically detects vehicle objects from high-resolution UAV images. They used an adaptive cropping algorithm to incorporate high-resolution images into the training of deep learning models. A publicly available dataset was used in their study. Researchers who preferred YOLOv5 as a deep learning model shared that after the test phase, the model reached 91.90% precision, 82.50% recall, 89.60% mAP@0.5, and 62.40% mAP@0.5:0.95. Sang et al. [12] proposed a YOLOv2-based system for detecting vehicle objects. In their study, two different publicly available datasets containing a total of eight different vehicle types were preferred. Normalization was applied to the bounding boxes since it is thought that different scales may affect vehicle detection. The model was trained for a maximum of 160 epochs using 8 batch size, 0.001 learning rate, and down-sampling factor 32 values. The trained model reached 94.78% mean Average Precision (mAP) in the testing phase. Song et al. [13] proposed a YOLOv3-based approach for vehicle detection and counting on highways. 57,290 annotated images from 11,129 raw images were prepared for this study. Compared to existing public datasets, the proposed dataset contains small annotated objects. The dataset, which has three different classes car, bus, and truck, is divided into 80% train and 20% test. As a result of the deep model test phase, whose training was completed, it reached 88.00% precision, 89.00% recall, 71.32% average Intersection Over Union (IoU), and 87.88% mAP values. In most similar studies in the literature, You Only Look Once (YOLO) architecture was used as a deep learning model due to its up-to-dateness and high performance. In this study, a deep learning-based approach is proposed for computer-assisted automatic vehicle detection. The YOLOv5 model was preferred for the deep learning approach. Vehicle images obtained from a publicly available dataset were used for model training and performance evaluations. The main contributions of this study are as follows:

- Pixels containing vehicle objects can be detected in an end-to-end structure without the need for any feature extraction
- The proposed deep model can produce results quickly without the need for high-capacity equipment.
- The detection success of small objects has been examined thanks to the image cropping method and filtering.

The rest of this paper is organized as follows. Section 2 contains information about the method proposed for this study, the dataset used, and the deep learning model. The numerical values obtained during the training phase of the deep learning model and the predictions made on the images reserved for the testing phase are given in Section 3. Conclusion part of the study is in Section 4.

2. Materials and Method

In this study, a deep learning approach is proposed to detect vehicle objects from UAV images. Instead of creating a model from scratch, YOLO architecture was preferred. Because this architecture is much better in terms of time and accuracy compared to other popular architectures. In detecting vehicle objects from high-resolution images, the pixel areas where the vehicle objects are located are detected in the single image output given as input to the model.

Returns the coordinates of the bounding box of the vehicle object in YOLO architecture predictions. In this way, the detected areas can be visually expressed. Fig. 1 shows a block representation of the proposed approach employed in this study.

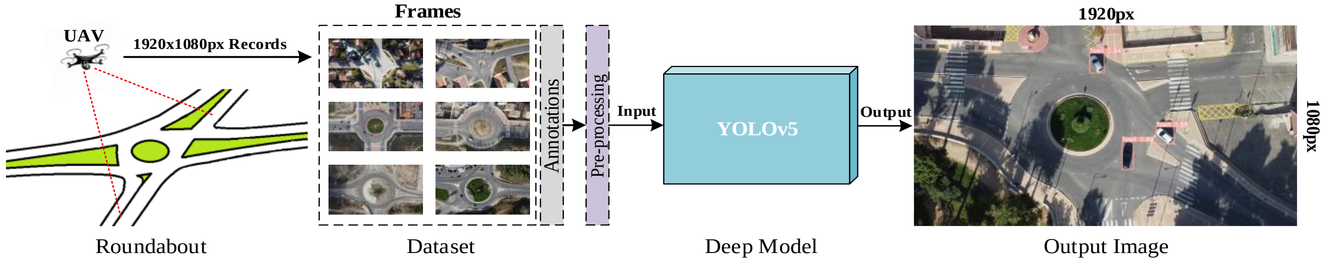


Figure 1. Block representation of the proposed approach

2.1. Dataset

The dataset used in this study was created by imaging the roundabouts in Spain with the help of UAVs [14]. The image frames contain the traffic flow with 1920×1080px resolution. After the completion of image capture, image processing, and labelling processes, the dataset includes 947,400 cars, 19,596 motorcycles, 2262 trucks, and 7008 bus objects. Various data augmentation techniques were used in the dataset with eight different roundabout scenes. In addition to the images, there are description files for each image that contain the information on the pixel locations where the objects, are to be used in detection studies. Fig. 2 shows a sample randomly selected from the dataset, and bounding boxes drawn using the coordinates in the description file.

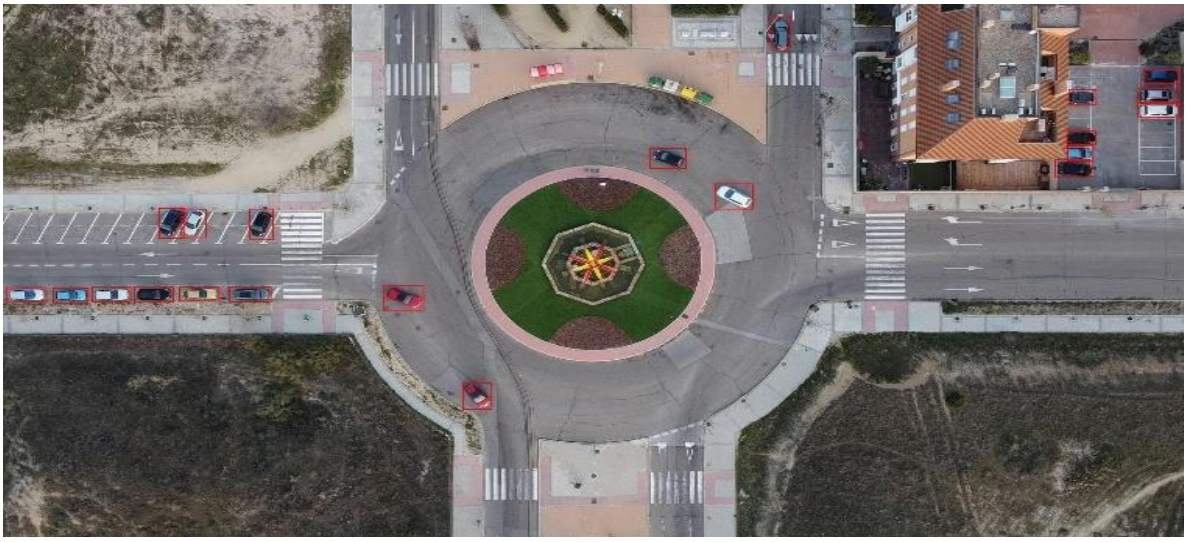


Figure 2. Sample a high resolution UAV image

2.2. Proposed detection method

Considering the difficulties of problem-specific model development, proven models on large datasets provide convenience for object detection studies. In this study, the YOLOv5 model, which is one of the single-stage and popular models, was used. Object detection algorithms generally work for the detection and classification of the desired number of objects in two different steps. These two-stage architectures are insufficient for performance. Redmon et al. [15] proposed an architecture that can combine two different steps into one step. This architecture, called YOLO, performs object detection using convolutional neural networks. As a real-time object detection system, YOLO uses a single neural network. For this reason, it is faster than other object detection algorithms. The YOLO divides the input image into N×N grids. Each grid considers whether there is an object in it. Deciding that the object has a centre point, the grid finds the class, height, and width of that object and draws a bounding box around that object. Multiple grids may think the object is within itself. In this case, unnecessary bounding boxes appear on the screen. The Non-Max Suppression (NMS) algorithm draws the bounding boxes with the highest confidence value on the screen for the objects detected on the image. The NMS algorithm is illustrated in Fig. 3.

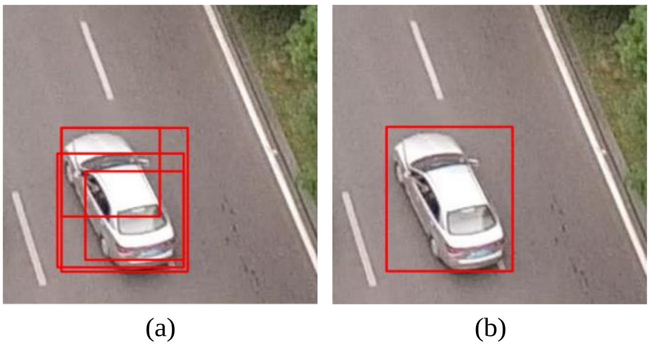


Figure 3. Before (a) and after (b) applying the NMS algorithm

3. Experiments

3.1. Experimental setups

The YOLOv5 model has been included in the training thanks to the PyTorch library created with the Python programming language. The output layers of the model have been revised to be suitable for producing output for the vehicle objects only. The training process was carried out for 100 epochs by keeping the 16 batch size and the learning rate constant at 0.001. All of these processes were carried out in the Google Colab environment. The default input size of the YOLOv5 model is 640×640px. Since the dataset samples have 1920×1080px resolution, it is not possible to carry out the training processes without making any changes to the default parameters of the YOLOv5 architecture. The input size of the model can be equated to the resolution of the dataset samples, but it is not a logical approach given our hardware specifications. For this reason, 1920×1080px images are cropped into 640×640px pieces. Padding was applied for the parts that could not complete the 640×640px size. In this way, six sub-images were obtained from one 1920×1080px image without any loss of information. After sub-images were obtained from all the images in the dataset, the sub-images that did not include the vehicle object were filtered. Thanks to the filtering process, the cost of the training phase has been reduced. Fig. 4 shows the diagram representing the image cropping, padding, and filtering processes. Before starting the model training, the samples were randomly divided to be used 70% in the training, 20% in the validation, and 10% in the testing phases.

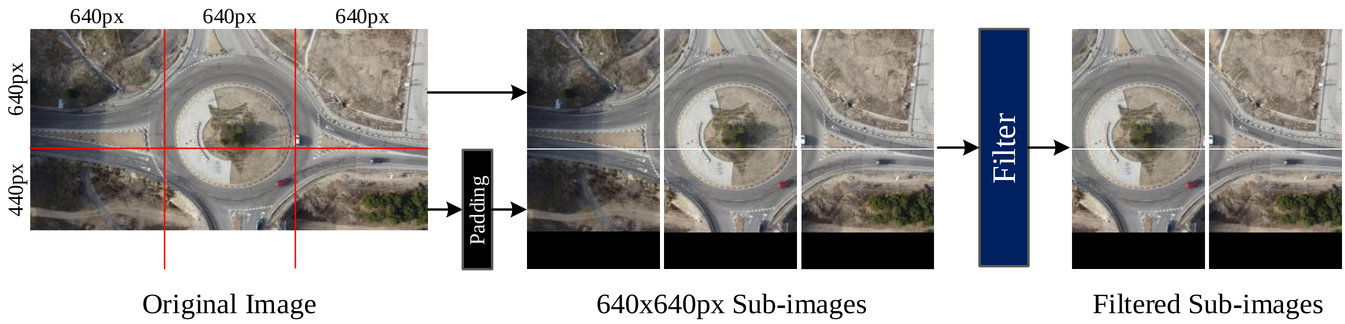


Figure 4. Block representation of the pre-processing stage.

3.2. Performance evaluation metrics

In detection studies, the mAP value is used to evaluate the performance of the deep learning model. The mAP value corresponds to the area under the precision–recall curve. Precision and recall are also computed as described in equations (1) and (2).

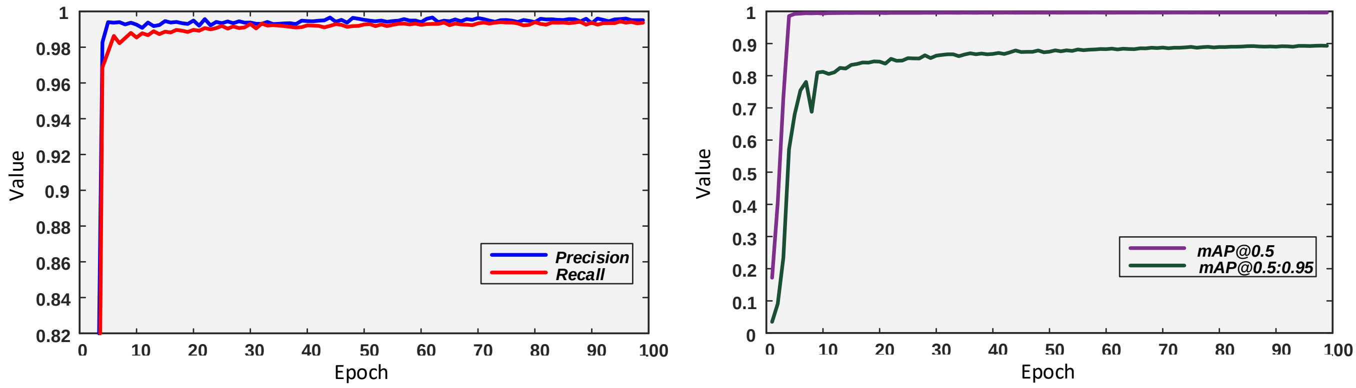
$$Precision = \frac{TP}{TP+FP} \tag{1}$$

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

The precision represents the correct detection rate of all predicted results, while the recall corresponds to the correct detection rate with respect to all ground truths. In addition, TP (true positive) is the number of bounding boxes correctly predicted as containing a vehicle, FP (false positive) is the number of bounding boxes inaccurately predicted as containing a vehicle, and FN (false negative) is number of bounding boxes that are incorrectly predicted as containing a vehicle.

3.3. Experimental results

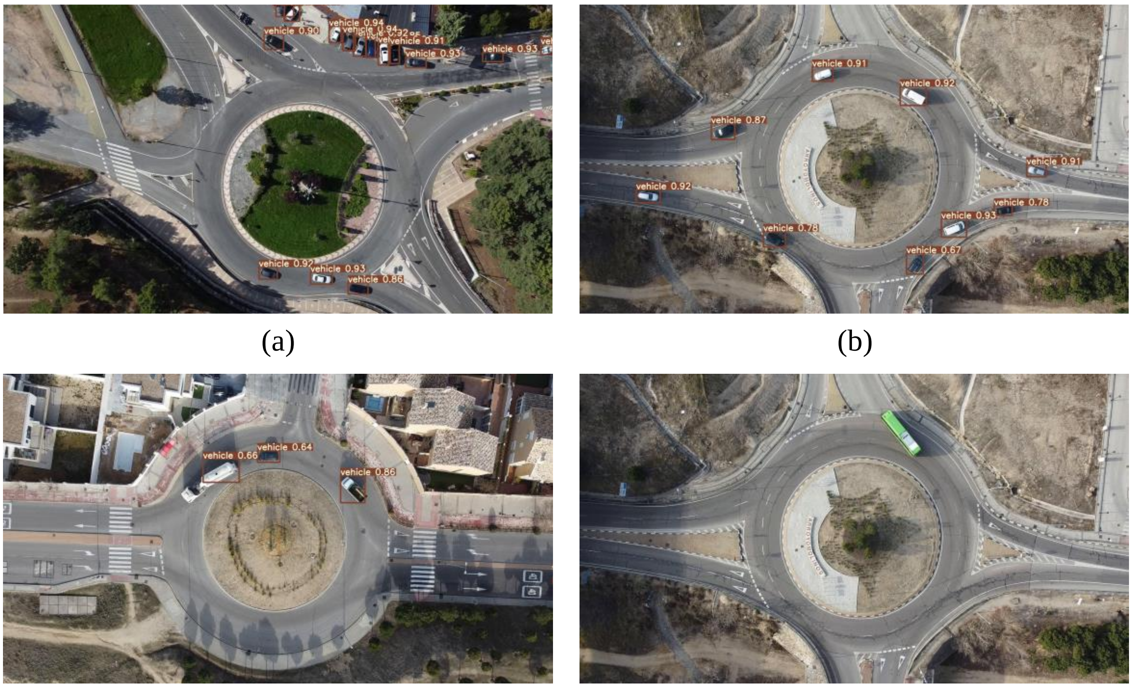
After completing the training of the model determined for automatic vehicle detection, performance results were obtained on the test data. Fig. 5 shows the precision-recall curves and mAP curves during training.



(a) (b)

Figure 5. Precision-recall (a) and mAP (b) curves.

The highest precision achieved by the model during training is 0.9966, and the highest recall is 0.9944. The final mAP@0.5 achieved by the YOLOv5 is 0.9966, and the mAP@0.5-0.95 is 0.8935. In Fig. 6, the outputs produced by the model for the test images are given.



(a) (b) (c) (d)

Figure 6. Test images predicted by the proposed model.

It is seen that the model detects the car objects with the highest number of samples in the dataset quite successfully. The model has difficulty in detecting large vehicles such as buses and trucks.

4. Conclusion

With the developments in the field of computer hardware, the use of machine learning methods in the field of transportation has become widespread. In this way, computer-based automatic detection systems have gained importance in traffic monitoring. High detection success is critical for the active use of the models developed in this field in daily life. In this study, an approach is proposed for automatic detection of vehicle objects from high resolution UAV images. Experimental studies were conducted on a publicly available dataset. YOLOv5 architecture is employed with default parameters. During the training phase, the value of 0.8935 mAP@0.5-0.95 was reached. Considering the determinations made by the model on the test images, there are problems in the detection of objects such as buses and trucks. The main reason for this is the small number of samples containing large vehicle objects in the dataset. However, the model has shown that it can detect areas with high reliability and accuracy for car objects.

References

[1] G. Mattioli, C. Roberts, J.K. Steinberger, A. Brown, The political economy of car dependence: A systems of provision approach, Energy Research and Social Science. 66 (2020). doi:10.1016/j.erss.2020.101486.

[2] C. Buchanan, Traffic in Towns: A study of the long term problems of traffic in urban areas, Routledge. (2015)

[3] Z. Sun, G. Bebis, R. Miller, On-road vehicle detection: A review, IEEE Transactions on Pattern Analysis and Machine Intelligence. 28 (2006) 694–711. doi:10.1109/TPAMI.2006.104.

[4] E.V. Butilă, R.G. Boboc, Urban Traffic Monitoring and Analysis Using Unmanned Aerial Vehicles (UAVs): A Systematic Literature Review, Remote Sensing. 14 (2022). doi:10.3390/rs14030620.

[5] R. Shrestha, R. Bajracharya, S. Kim, 6G Enabled Unmanned Aerial Vehicle Traffic Management: A Perspective, IEEE Access. 9 (2021) 91119–91136. doi:10.1109/ACCESS.2021.3092039.

[6] Y. Akbari, N. Almaadeed, S. Al-maadeed, O. Elharrouss, Applications, databases and open computer vision research from drone videos and images: a survey, Artificial Intelligence Review. 54 (2021) 3887–3938. doi:10.1007/s10462-020-09943-1.

[7] X. Li, F. Men, S. Lv, X. Jiang, M. Pan, Q. Ma, H. Yu, Vehicle detection in very-high-resolution remote sensing images based on an anchor-free detection model with a more precise foveal area, ISPRS International Journal of Geo-Information. 10 (2021). doi:10.3390/ijgi10080549

[8] G. Fragapane, R. de Koster, F. Sgarbossa, J.O. Strandhagen, Planning and control of autonomous mobile robots for intralogistics: Literature review and research agenda, European Journal of Operational Research. 294 (2021) 405–426. doi:10.1016/j.ejor.2021.01.019.

[9] B. Liu, C. Han, X. Liu, W. Li, Vehicle Artificial Intelligence System Based on Intelligent Image Analysis and 5G Network, International Journal of Wireless Information Networks. (2021). doi:10.1007/s10776-021-00535-6.

[10] S. Ghaffarian, J. Valente, M. van der Voort, B. Tekinerdogan, Effect of attention mechanism in deep learning-based remote sensing image processing: A systematic literature review, Remote Sensing. 13 (2021). doi:10.3390/rs131529

[11] Z. Chen, L. Cao, Q. Wang, YOLOv5-Based Vehicle Detection Method for High-Resolution UAV Images, Mobile Information Systems. 2022 (2022). doi:10.1155/2022/1828848.

[12] J. Sang, Z. Wu, P. Guo, H. Hu, H. Xiang, Q. Zhang, B. Cai, An improved YOLOv2 for vehicle detection, Sensors (Switzerland). 18 (2018). doi:10.3390/s18124272.

[13] H. Song, H. Liang, H. Li, Z. Dai, X. Yun, Vision-based vehicle detection and counting system using deep learning in highway scenes, European Transport Research Review. 11 (2019). doi:10.1186/s12544-019-0390-4.

[14] E. Puertas, G. De-Las-heras, J. Fernández-Andrés, J. Sánchez-Soriano, Dataset: Roundabout Aerial Images for Vehicle Detection, Data (Basel). 7 (2022). doi:10.3390/data7040047.

[15] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2016: pp. 779–788. doi:10.1109/CVPR.2016.91.