2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA) | 978-1-7281-8470-8/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICMLA51294.2020.00155

Deep Learning Based Vehicle Detection From Aerial Images

Hüseyin Seçkin DIKBAYIR
Turkish Aerospace, Gazi University
Ankara, Turkey
seckindikbayir@gmail.com

Halil İbrahim BÜLBÜL
Computer Education and Instructional Technologies,
Gazi University
Ankara, TURKEY
bhalil@gazi.edu.tr

***Abstract*— With the help of today's developing technology; many applications such as traffic monitoring, target detection, empty parking lot assignment etc. have become applicable with machine learning and deep learning methods. These applications are aimed to determine the desired object from the photographs. In this study, it is aimed to develop an approach that able to identify the vehicles through aerial images by using the YOLO (You Only Look Once) algorithm, feeding it with a trained convolutional neural network structure. As a result of the study, an application that can detect vehicles was developed, increased the performance rate of the YOLO algorithm by 3.2%.**

Keywords- vehicle detection, deep learning, machine learning, image processing.

I. INTRODUCTION

Vehicle detection is an issue that is needed in many areas in military and civilian fields. Unmanned aerial vehicles have been actively used to provide inputs for such applications with the help of the high-resolution photos they provide.

In recent years, vehicle detection applications can be developed faster and more accurately by taking advantage of machine learning and deep learning methods. When such applications are examined in general, it is seen that the structures are based on convolutional network or image processing algorithms.

While convolutional based approaches are examined; it is seen that a set of features is revealed from the input images at the first stage. Classifiers are then used to describe these objects. Depending on the method chosen, the detection process is performed by scanning the whole image in a sliding window structure or working on the selected regions on the image. Although the accuracy rate produces high results, the processing of the input and the generation of the result takes longer than other deep learning structures.

In structures such as YOLO (You Only Look Once) and SSD (Single Shot Detector), on the other hand, it is important to be able to detect in real time. So, timing is more important than performance rate in these approaches. It is observed that the successful detection rate of such algorithms decreases as the object size gets smaller.

When these two different structures (Faster R-CNN vs YOLO) are compared; It has been observed that Faster R-CNN has the highest performance rate, but the lowest speed. YOLO, on the other hand, has a high-performance rate for large objects and decreases as the objects get smaller. (Lin & Girshick 2017).

Since the small dimensions of the vehicles in the drone images affect the performance rate of the YOLO algorithm, using it alone does not give the desired result. In the study, the Faster R-CNN algorithm was used to increase the performance rate of the YOLOv3 algorithm, and it was aimed to improve vehicle detection. Munich Vehicle Data Set was used as the data set because it includes images of different vehicle types over 100m, and high resolution images obtained from the "Google Earth" application and DJI drone images were also included in the set to expand the data set.

The rest of the paper is planned as follows: Section II provides references to previous research on vehicle detection and data sets used generally in these applications. Section III introduces the methodology and aim of the work. Section IV gives our work result and compares it with literature. Section V summarize the study and marks the future studies.

II. RELATED WORKS

In the literature, there are many vehicle detection and classification applications on photographs obtained from systems such as drones and satellites. Within the scope of this study, relevant articles were investigated by using the keywords of vehicle detection, image processing, parking area detection, machine learning, and deep learning. The outstanding studies in this context are briefly summarized below.

1. Hybrid deep artificial neural network (HDNN), which is a new model and approach that differs from existing algorithms, has been introduced by Chen (Chen, Xiang, Liu, & Pan, 2014). This model includes sliding window and deep CNN (convolutional neural network) method. The main idea is to detect different sized vehicles by adjusting the convolutional layers to different sizes. In this study, a special data set, collected from San Francisco with the help of Google Earth, was used. Although the study shows a different method from other ones, it takes up to 7 seconds, even on the most powerful processors in terms of operation. It is stated that this method is not a highly preferred, because of its time consuming.
2. A two-step method for vehicle detection has been introduced by Ammour et al. (Ammour et al., 2017). In the first stage, the picture is divided into parts with the help of region extraction and average-shift algorithm. In the second

978-1-7281-8470-8/20/$31.00 ©2020 IEEE
DOI 10.1109/ICMLA51294.2020.00155

stage, feature extraction was performed using the VGG16 model for vehicle detection. With the help of SVM (support vector machine features are determined as vehicle or not. As a data set, specially obtained images from Trento University, Faculty of Science Campus was used. Although it is ahead of many algorithms in terms of accuracy, it is considered unsuitable for real-time use due to its detection time and misleading approach in small objects.

3. In the study conducted by Deng et al. (Deng, Sun, Zhou, Zhao, & Zou, 2017), a method that aims to produce an R-CNN (region based) feature map is presented. It was aimed to prevent the slowness offered by the CNN structure. Paired network structure (combination of two different CNN methods) is used in the study. "Munich Vehicle Data Set" was preferred as the data set. Although it gives better results than similar studies in terms of accuracy and speed, it produces erroneous results in small objects. It has been stated that it has a slower detection mechanism for real-time applications because of CNN structure.

4. In another study (Ren, He, Girshick, & Sun, 2017), it was aimed to shorten the operating time of the CNN network structure. VGG16 model was used for feature extraction, and the network structure was expanded by feeding with this model. As a result of the study, the CNN structure was accelerated. However, erroneous detection occurring in small objects could not be prevented. Lateral images were used instead of drone images in the study, and performance values were measured in the range of 5-17 fps.

5. Fast R-CNN and Faster R-CNN methods were used in another study by Yu et al. (Yu, Westfechtel, Hamada, Ohno, & Tadokoro, 2017). The application for these network structures has been revealed, and an optimized approach for small objects is desired. Vedai and Munich Vehicle Data Set were used. Other algorithms with proven success were not included in the study.

6. In a different study (Kyrkou, Timotheou, Kolios, Theocharides, & Panayiotou, 2018), focusing on traffic monitoring, using CNN algorithm, road and vehicle extraction was performed. Masking process is used for path extraction and CNN algorithm is preferred for decision making. It was stated that drone and satellite photographs taken in Cyprus were used in the study. It is observed that the study has a high success rate, combining masking technique and different methods.

7. In the study conducted by Liu et al. (Liu, Yang, & Li, 2018), vehicle detection was performed by referring to infrared images. "NPU_CS_UAV_IR_DATA" is used as the data set. A CNN-based detection application was developed, infrared photographs were taken in real time and sent to the system. A CNN-based detection process was carried out by running the software, which had previously tested and verified steps on the system. It was reported that high success rates were made on black and white photographs.

8. In a study conducted by Yang et al. (Yang, Liao, Li, Cao, & Rosenhahn, 2019), a network structure called Double Focal Lose CNN was developed. ITCVD data set was preferred as the data set. Basically, it is aimed to reduce the detection error with the help of the network structure using CNN. It is seen that a suitable structure for real-time applications cannot be provided due to the long processing time.

As observed from the studies in the literature, the generally applied approach aims that increase the accuracy and shortening the detection time. In this sense, it has been observed that there is a need for an approach with high accuracy and short detection time.

III. METHODOLOGY

In this study, the CNN-based Faster R-CNN deep learning algorithm, which stands out in the literature with its performance and working speed, and the v3 version of YOLO algorithm (v1, v2, v3, v4 (Mahto, Garg, Seth, & Panda, 2020)) are used.

Faster R-CNN is a fast regionally based convolutional neural network and has a structure that operates by combining similar regions. The aim of Faster R-CNN (region based convolutional neural network) is to create a certain number of regions with a selective search method and search through these regions instead of searching through the whole picture and find the right object.

The YOLO algorithm, on the other hand, aims to offer a structure suitable for real-time processing by taking the picture completely convolutional rather than a regional-based approach. Depending on the size of the picture, it is divided into a grid of m x m and the distinction is made according to their similarities.

Within the scope of this study, it is aimed to make vehicle detection faster and to increase the detection accuracy rate. Accordingly, YOLO algorithm is preferred for detection, while a faster-regional convolutional based network is presented as an aid in the learning phase of the YOLO algorithm to increase the accuracy. The Munich Vehicle Data Set was used as the data set, as it contains images of different vehicle types from the data sets in “Table I”, with high resolution and over 100m. In additional to this, high-resolution images obtained from the “Google Earth” application and DJI drone images were included in this set to expand the data.

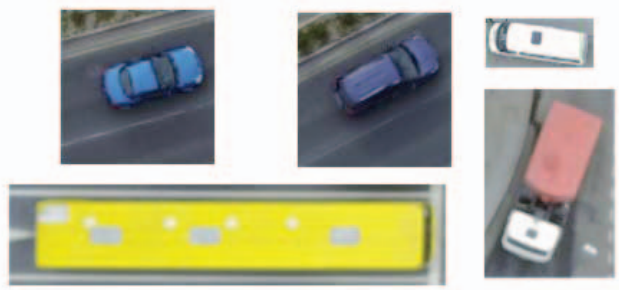

FIGURE 1- VEHICLE EXAMPLES USED IN DATASET

Firstly, to demonstrate the methodology, definition of fast regional convolutional neural network was made. Convolutional neural networks have a convolutional operator inside that extracts the properties of the input image. This operator preserves the relationship between pixels by learning the image properties of the input data. Basic structure consists of four layers; the convolutional operator, ReLU (Rectified Linear Units), subsampling and fully linked

layer. After each convolutional processing, areas with negative values for each pixel are replaced with a zero-value using the ReLU activation function. The subsampling layer, on the other hand, reduces the size of each feature map and transfers the important information to the other layer. The fully linked layer classifies the input image based on the training data set by linking the input image with the extracted images.

While defining the convolutional neural network, it is aimed to feed the algorithm that will detect the actual object by selecting small vehicle objects from the available data sets. In addition, the description is planned to be a reference to compare the algorithm performance. The data set is labeled in accordance with the algorithm. They were randomly divided into testing and learning using the 5-fold cross validation method, using the desktop application developed with the C # programming language.

Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross-validation, where k used as 5.

With the help of code developed with Python programming language, the Faster R-CNN network was trained and prepared to feed YOLO.

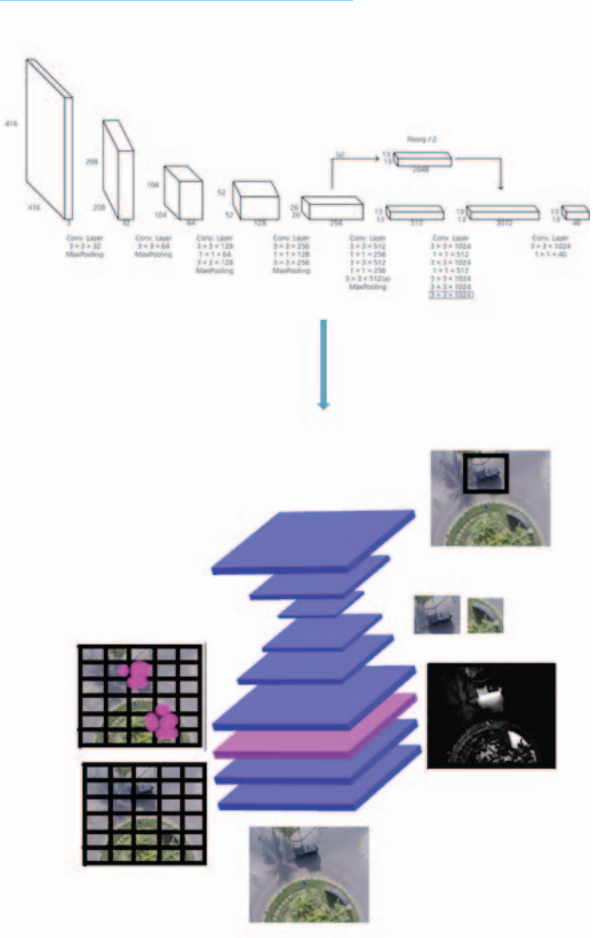

FIGURE 2- NETWORK STRUCTURE OF FASTER R-CNN FED YOLO

The Faster R-CNN Fed YOLO and YOLO algorithms are both run on a GeForce 1060 GPU Video card and a 16GB Ram computer. To compare algorithms, precision, recall, F1-Score, and quality metrics are used.

Precision metric (1): It is used to show how many of the values we guess positively which are positive.

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

Sensitivity metric (2): It is used to show how much of the transactions we should have predicted positively.

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

F1-Score metric (3); shows us the harmonic mean of precision and precision values. Harmonic averaging is preferred because it is for processing extreme cases.

$$F1\ Score = \frac{2 * Precision * Sensitivity}{Precision + Sensitivity} \quad (3)$$

Quality metric (4): It expresses the probable success rate that the algorithm will show on a new input.

$$Quality = \frac{TP}{TP + FP + FN} \quad (4)$$

TP: True Positive (Objects that are positive, but the model predicts positive)

FP: False Positive (Objects that are not actually positive, but the model predicts positive)

FN: False Negative (Objects that are positive, but the model predicts negatively)

To measure the success of the algorithm, the available data are divided into three different classes (A, B, C) according to the size and complexity of the structure it contains.

Class A: Photographs obtained from a height of 30m and below, includes clear vehicle images.

Class B: Photographs including medium sized vehicle images taken between 50 m - 75 m and different objects such as houses, pools etc.

Class C: Photographs including small sized vehicle images taken from a height above 75 m and objects that will mislead the algorithm.

IV. EVALUATION

YOLO is an algorithm that can respond faster in terms of performance than convolutional neural networks and their derivatives. Since it is desired to obtain a result with more successful detection rate by using the speed of the YOLO algorithm in the scope of the study, the convolutional approach has been used. However, it has been observed that the rate of successful detection of the algorithm decreases, especially in areas where the shadows are dropped, and when the color of the terrain is close to the vehicle color.

Within the scope of the study, labeling of the tool sets was done separately for the fast regional convolutional and YOLO algorithm.

YOLO and Faster R-CNN Fed YOLO is tested by using data classes (A, B, C). Results are shared on "Table I – Comparison of YOLO and Faster R-CNN Fed YOLO Algorithms".

TABLE I - COMPARISON OF YOLO AND FASTER R-CNN FED YOLO ALGORITHMS

Class	Algorithm	Precision	Sensitivity	F1 Score	Quality
A	YOLO	0,93	0,84	0,89	0,8
	Faster R-CNN Fed YOLO	0,92	0,9	0,91	0,83
B	YOLO	0,88	0,86	0,87	0,77
	Faster R-CNN Fed YOLO	0,9	0,88	0,89	0,8
C	YOLO	0,89	0,74	0,81	0,68
	Faster R-CNN Fed YOLO	0,87	0,88	0,87	0,78

As a result of the study, it was observed that the structure offered as an aid to YOLO increased the detection rate of the YOLO algorithm by 3.2%. However, it has been observed that the algorithm fails to detect dark colored vehicles in shadowy areas. The comparison of the sample detection results of the algorithm is given in the figure below.

FIGURE 3 - DETECTION RESULT OF YOLO ALGORITHM

FIGURE 4- DETECTION RESULTS OF R-CNN FED YOLO ALGORITHM

V. CONCLUSION

In this paper, it is aimed to develop a vehicle detection application with higher accuracy and close to the speed of the YOLO algorithm. In this context, firstly; Munich Vehicle Data Set was used, because it includes images of different vehicle types over 100m, and high resolution images obtained from the "Google Earth" application and DJI drone images were also included in the set to expand the data set. The data set is divided into test and learning sets with the help of 5-fold cross validation method.

Secondly, the studies in the literature are analyzed and the prominent deep learning algorithms were determined as Faster R-CNN and YOLO. First, the available data set was labeled and adapted to the Faster R-CNN algorithm. The YOLO algorithm is fed with the help of this trained network structure, aiming to increase the accuracy of vehicle detection. Classification data has been presented to YOLO as input. Then, for comparison, the YOLO algorithm has been trained with the same data sets and made suitable for detection. These algorithms were compared with the help of different classes created over the data set. In the tests performed with inputs with a maximum size of 400x400, it was observed that the created structure increased the performance rate of the algorithm by 3.2%.

With the accuracy rates revealed by the study, it is thought that it can be used in many areas such as traffic monitoring, target determination, vehicle counting, parking applications.

In this study, some studies that could not be done due to both the time and the obtained data sets because the conditions were not yet established; It can be said to reduce the error rate of the algorithm, increase the variety of tools to enable the algorithm to recognize more tools, and evaluate the performance of different network structures in small objects in order to reduce the error rate.

REFERENCES

[1] Ammour, N., Alhichri, H., Bazi, Y., Benjdira, B., Alajlan, N., & Zuair, M. (2017). Deep learning approach for car detection in UAV imagery. Remote Sensing. https://doi.org/10.3390/rs9040312.

[2] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2016). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv preprint arXiv:1606.00915

[3] Chen, X., Xiang, S., Liu, C. L., & Pan, C. H. (2014). Vehicle detection in satellite images by hybrid deep convolutional neural networks. IEEE Geoscience and Remote Sensing Letters. https://doi.org/10.1109/LGRS.2014.2309695

[4] Deng, Z., Sun, H., Zhou, S., Zhao, J., & Zou, H. (2017). Toward Fast and Accurate Vehicle Detection in Aerial Images Using Coupled Region-Based Convolutional Neural Networks. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 10(8), 3652–3664. https://doi.org/10.1109/JSTARS.2017.2694890

[5] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[6] Ioffe, S., & Szegedy, C. (2015, June). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International conference on machine learning (pp. 448-456).

[7] Kyrkou, C., Timotheou, S., Kolios, P., Theocharides, T., & Panayiotou, C. G. (2018). Optimized vision-directed deployment of UAVs for rapid traffic monitoring. In 2018 IEEE International Conference on Consumer Electronics, ICCE 2018 (Vol. 2018-Janua, pp. 1–6). https://doi.org/10.1109/ICCE.2018.8326145

[8] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[9] Liu, K.; Mattyus, G. DLR 3k Munich Vehicle Aerial Image Dataset. Available online: http://pba-freesoftware.eoc.dlr.de/3K_VehicleDetection_dataset.zip (accessed on 31 December 2015)

[10] Liu, X., Yang, T., & Li, J. (2018). Real-time ground vehicle detection in aerial infrared imagery based on convolutional neural network. Electronics (Switzerland), 7(6). https://doi.org/10.3390/electronics7060078

[11] Mahto, P., Garg, P., Seth, P., & Panda, J. (2020). Refining Yolov4 for vehicle detection. International Journal of Advanced Research in Engineering and Technology. https://doi.org/10.34218/IJARET.11.5.2020.043

[12] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2016.91

[13] Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards Real-Time Object Detection with. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

[14] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[15] Sommer, L. W., Schuchert, T., & Beyerer, J. (2017). Fast deep vehicle detection in aerial images. In Proceedings - 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017 (pp. 311–319). https://doi.org/10.1109/WACV.2017.41

[16] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar. ´Focal loss for dense object detection. arXiv preprint arXiv:1708.02002, 2017. 1, 3, 4.

[17] Yang, M. Y., Liao, W., Li, X., Cao, Y., & Rosenhahn, B. (2019). Vehicle detection in aerial images. Photogrammetric Engineering and Remote Sensing, 85(4), 297–304. https://doi.org/10.14358/PERS.85.4.297

[18] Zhong, J., Lei, T., & Yao, G. (2017). Robust vehicle detection in aerial images based on cascaded convolutional neural networks. Sensors (Switzerland). https://doi.org/10.3390/s17122720.