# LANE TRACKING WITH DEEP LEARNING: MASK RCNN AND FASTER RCNN

Fatma Nur ORTATAŞ
*Computer Engineering*
*Yozgat Bozok University*

Yozgat, TURKEY
f.nur.ortatas@yobu.edu.tr

Emrah ÇETİN
*Electrical and Electronics Engineering*
*Yozgat Bozok University*

Yozgat, TURKEY
emrah.cetin@yobu.edu.tr

2022 Innovations in Intelligent Systems and Applications Conference (ASYU) | 978-1-6654-8894-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ASYU56188.2022.9925296

***Abstract*—Traffic accidents are caused by driver faults and non-compliance with traffic rules mostly. It is aimed to reduce the accident rates by preventing driver-related accidents was the developments in electric and autonomous vehicle technologies by the studies carried out in recent years. Lane tracking systems for autonomous vehicles are critical to reducing these kinds of risks in accidents. Within the scope of this study, a test environment was created in the Gazebo environment for lane detection and tracking. In addition, segmentation process and object detection algorithms were used for lane detection by utilizing Mask RCNN and Faster RCNN. The software developed within the scope of the study has been successfully tested in the created simulation environment.**

***Keywords— lane detection, mask RCNN, faster RCNN, autonomous vehicle***

## I. INTRODUCTION

With the developing technology, unmanned systems have started to appear in all areas of our lives to facilitate our daily life and to provide time efficiency. The rapid development unmanned systems technology, combined with electric vehicle technology. The growing experience and production of this technology have become widespread in recent years. Thus, it pushes autonomous vehicle technology. There are many automotive and technology companies working in this field all over the world. The main reason for this rapidly developing technology is to increase safe driving rates and thus to reduce traffic accident rates. Looking at the 2019 data of the World Health Organization, it has been reported that more than 1.35 million people die every year due to various traffic accidents [1]. %94 of the accidents are caused by risky driving and mistakes made by drivers at the wheel [2]. Considering the 2019 data of the Turkish Statistical Institute, it was revealed that %91.04 of the traffic accidents that occurred were caused by drivers [3]. The main purpose of all studies is to reduce the current rates by preventing driver-related accidents with the developments in electric and autonomous vehicle technology.

With today's technology, we come across an improved driver assistance system (ADAS), which increases driving quality in terms of driver and pedestrian safety and is used to prevent possible accidents. ADAS systems come with many different features. Features such as distance sensors, adaptive speed control, collision warning system, and lane departure warning systems are common in today' s vehicle [4,6]. One of the important features of ADAS systems that have come toa the fore in recent years is computer vision assisted systems in which cameras are used [7]. Lane detection and tracking system, which is one of them, has a critical importance in ADAS systems. Lane tracking systems for autonomous vehicle are critical to reduce the risk of accidents. Violation of the road border lane on the roadside or incorrect crossings account for this high rate. For this reason, road boundary lanes have a very important place for safe driving.

The development of lane tracking systems has a critical place in minimizing traffic accident rates, especially due to lane violations and incorrect overtaking. When the literature is examined for lane tracking, it is seen that there are different studies that use image processing techniques and machine learning algorithms. Image processing methods are frequently used for lane detection operations. Within this method, the first thing to do is to provide the color space conversion of the image from incoming camera image and define the Region Of Interest (ROI). ROI process is the process of selecting the area to be studies on the image by entering the pixel values and removing the remining part from the image. After this process, it is seen that basic image processing techniques such as the application of different edge detection filters on the area to be processed and the threshold value are used [8]. In addition, it has been seen that the Hough method is frequently used together with machine learning algorithms in the studies. This method is mostly used to detect known and obvious geometric shapes such as ellipses, circles or lines in the image [9].

Another working example shows a 3-step road lane detection process. It is seen that these stages are the separation of road and sky with the horizon line detection, then the analysis of lane region by marking color space transformations, and finally the determination by the Hough method [10]. In another study, for lane detection, first of all, the world is considered as a flat area, the image coming from the camera is converted into a bird's eye view, and then Canny edge detection algorithms are used. Thus, a clean image is obtained by getting rid of unnecessary information and pollution that may occur in the image. However, this method is insufficient in cases where the vehicle turns or where the road is inclined such as intersections. In addition, it was observed that power spectra and Fast Fourier methods

978-1-6654-8894-5/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: Bozok Universitesi. Downloaded on November 12,2022 at 12:22:06 UTC from IEEE Xplore. Restrictions apply.

were applied to the images [11]. most of the current lane detection studies show that features are extracted manually. This means that the extracted manually. This mean that the extracted features are weak. These weak features, which are manually extracted, cannot be successful in lane detection in the face of possible complex scenarios that the vehicle will encounter [12,13]. Machine learning algorithm approaches put forward the necessity of dataset. In different studies, datasets were created with different methods. In the studies, cameras placed on the vehicle were used to create the dataset [14]. Huval et al. revealed a large dataset of highway images. Afterwards, he introduced the use of deep learning method in vehicle and lane detection [15]. In another study, it was seen that lane detection was achieved by combining different types of convolutional neural networks [16]. In some studies, a dataset was created by directly taking human behaviors to use imitation learning methods [17]. There is segmentation process as the method used in the literature [18]. In another study, it was seen that certain colors or a part of the image were used to segment an image. This study paves the way to work only with the information of the segmented area [19]. The success of computer vision technology is basically directly proportional to the success of image processing and pattern recognition. Because in this area, it is expected that the driver's vision, perception and correct decision-making mechanisms will be realized by image processing and pattern recognition techniques. However, since the basic image processing techniques are directly affected by the basic physical conditions such as bad weather conditions or the difference in the lighting condition of the environment, it cannot be fully efficient [20]. Liu et al. another study he made used the Mask R-CNN model to embody the lane line [21]. Yet another study showed that aerial road images were segmented with Mask R-CNN [22]. In another study carried out in China, it was reported that some traffic markers and the Mask R-CNN model on road lane lines were applied to a ready-made dataset, but it could not be fully met in a real-life traffic scenario [23]. In another study, it is seen that the convolutional neural network model was used for lane detection [24].

In this study, a 3-stage method was applied for lane detection and tracking. In order to eliminate the lane loss problem in complex scenarios, studies have been carried out on the detection of road detection with Mask-RCNN, mainly due to the fixed ROI region encountered in image processing techniques used in lane detection. Thus, it is aimed to eliminate the ROI related restriction in path detection. As a second step, studies have been carried out on the detection of lane lines with Faster RCNN in order to eliminate the possibility of being affected by changes that will occur depending on the environment in which the vehicle is located, or the precipitation conditions. After all these determinations were made, image processing techniques were used to realize the lane tracking as the third step.

This paper was created as follows. In section II, details about simulation and dataset, Mask RCNN, Faster RCNN, Lane Detection and Tracking are given together with each simulation outcome. Afterwards, in section III, the results are discussed in the conclusion.

## II. METHOD

A simulation environment has been developed for the implementation and testing of the performed work. In addition, within the scope of the study, Mask RCNN, Faster RCNN algorithms as well as image processing techniques were used to perform lane detection and tracking. The workflow diagram of the processes carried out throughout the study is given in Fig. 1.

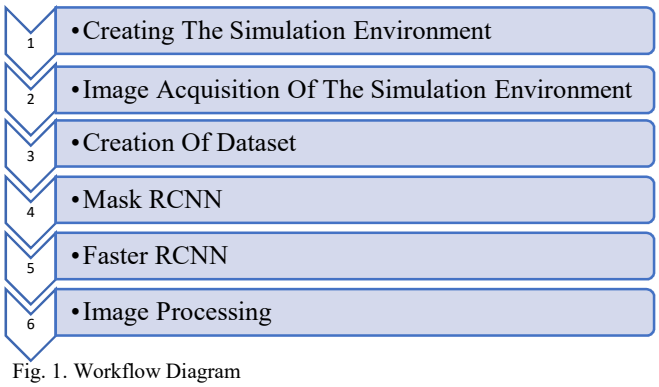



Fig. 1. Workflow Diagram

### *A. Simulation Environment and Dataset*

To carry out the model training and test processes of the developed software, a track design was made in the open-source Gazebo environment, where a two-lane vehicle can move. The visual of the track developed in the gazebo environment is given in Fig. 2.

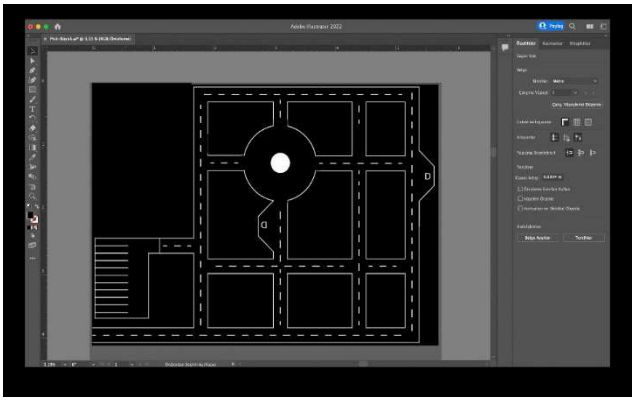


Fig. 2. Created Track Image

A dataset was created from 900 images obtained from the created environment with the help of the camera placed on the vehicle. An example of the image taken from the camera used in the simulation environment is given in Fig. 3. After the dataset was created, it is polygonally labeled for Mask RCNN. The main reason for labeling it as polygonal is that independent and as many points as necessary are needed for the selection of the desired road inside the image. For this reason, the polygonal selection is more appropriate for the labeling to be made from the boundary line of the road. For Faster RCNN, its labeling as a rectangle. The main reason for using rectangular selection is due to that road lane images are stable and rectangular selection is sufficient for their selection. The road lanes are labeled as small zones. With the small region selection, it is aimed to minimize the unnecessary information on the label. One of the labeling samples is given in Fig. 4.

978-1-6654-8894-5/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: Bozok Universitesi. Downloaded on November 12,2022 at 12:22:06 UTC from IEEE Xplore. Restrictions apply.

The collected dataset images are labeled with polygonal road sections for Mask R-CNN and rectangular lane lines on the road for Faster R-CNN. In this study, image dimensions are used as 640*480. During the labeling process, the labeling made for Mask R-CNN was also made polygonal labeling, generally covering 210*160 and 400*300 dimensions. Three different classes were used for labeling for Faster R-CNN. These were created separately for 3 different lane lines on the road. Since the right and left lane lines consist of a whole line, they were divided into small parts and labeled. In general, labeling processes ranging from 10*10 to 20*20 were used for left lane labeling. For the right lane, labeling was done in sizes ranging from 15*15 to 30*30. For the middle lane, labeling was done in sizes ranging from 40*40 to 60*60.

The obtained data were separated as training and test data, with %80-%20 as the optimum value.

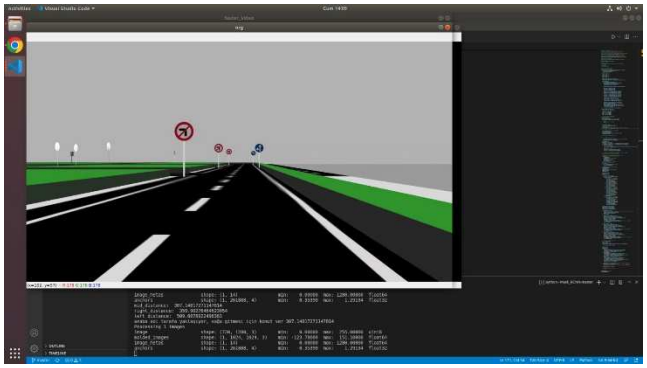
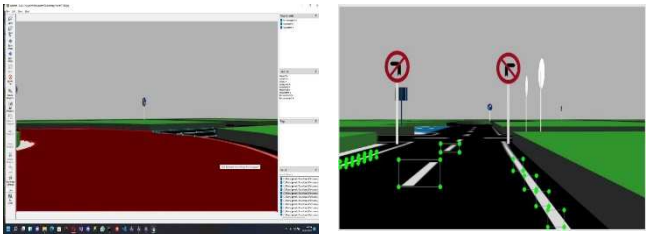


Fig. 3. An Example of The Image Taken from Camera



a) Polygonal b) Rectangle

Fig. 4. An Example of Labelling

## B. Mask R-CNN

Mask RCNN, in the simplest sense, is formed by adding mask output to Faster RCNN outputs. Faster RCNN has two outputs as in all object recognition algorithms. The first of these is label information, and the second is the bounding box location information of the object [25]. By adding the mask property with Mask RCNN, the pixels of the object within the bounding box are defined. Thus, mask information takes place. In the Mask RCNN algorithm, feature mapping is provided from the images that will be used in model training structurally. After this process, using this feature map, the predicted objects in the image are surrounded by the Region Proposal Network. Then the ROI pooling layer is applied to the bounding boxes in the image to make all predictions of the same size. After this process, the fully connected layer is applied to classify the last bounding boxes in the image. In the last case, the intersections of the bounding boxes that are too many on the image are grouped and the highest ones are selected by looking at the prediction ratios. Mask operation is performed on the identified object in the bounding boxes with the highest predictive value selected [25,26]. The structure information of Mask RCNN and Faster RCNN is given in Fig. 5. Faster RCNN structure is given in the frame shown in orange in the figure. The Mask RCNN structure, on the other hand, is surrounded by green color. The necessary procedure for all these operations has been completed and the configuration settings have been arranged. Polygonal labeling was done on the created dataset and model training was carried out in 1000 iterations.

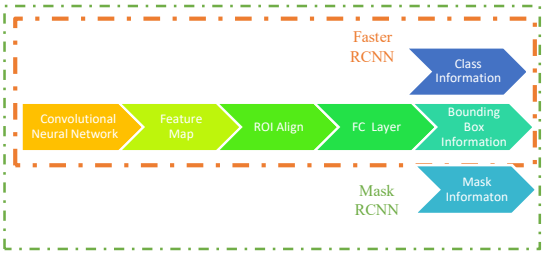



Fig. 5. Mask RCNN and Faster RCNN Structures

Within the scope of the study, Mask RCNN was applied to the image taken from the simulation environment and the detection process was carried out successfully. An example image of this study is given in Fig. 6.

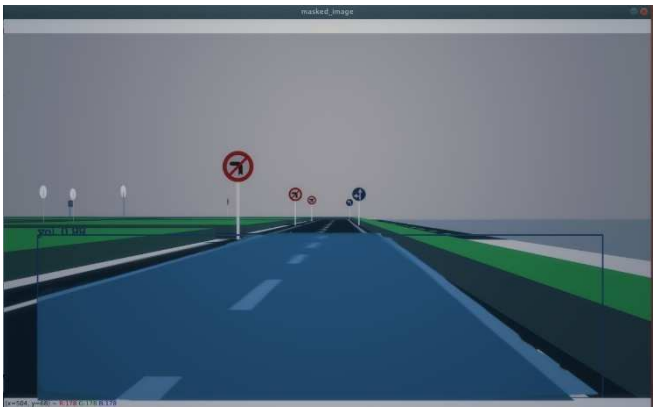


Fig. 6. Mask RCNN Applied Image

## C. Faster RCNN

Faster RCNN, the most widely used and advanced version of the RCNN family, is a deep convolutional network for object detection introduced in 2015 that appears as an end-to-end, unified network. The network can accurately and quickly predict the positions of different objects. In the RCNN family, first RCNN is followed by Fast RCNN and then Faster RCNN.

R-CNN first divides the input image into about 2000 regions. The convolutional neural network is applied to the separated independent regions, respectively. At this stage, fixed-length feature vector extraction of each region is performed. The features of each extracted region are evaluated for classification by applying a linear Support Vector Machine (SVM). After all the processes, the classification process is carried out according to the results of the evaluation by looking at the regions [27,28].

978-1-6654-8894-5/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: Bozok Universitesi. Downloaded on November 12,2022 at 12:22:06 UTC from IEEE Xplore. Restrictions apply.

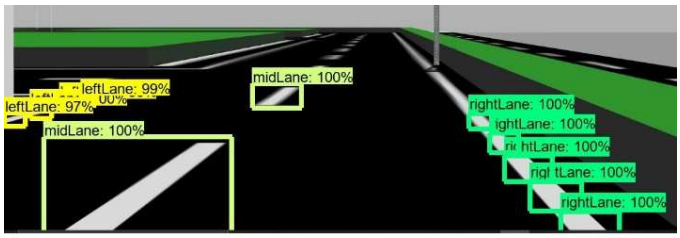



Fig. 7. Faster RCNN Applied Image

While RCNN and Fast RCNN use selective search algorithm, Faster RCNN uses the Region Proposal Network algorithm [29].

The selective search algorithm is used to divide the image into 2000 regions. The testing speed of the R-CNN model is slow due to the application of the CNN algorithm, which will be performed separately and sequentially for each region. However, the region proposal network used in Faster R-CNN caused a great improvement in the overall detection speed of the model since it was applied directly after feature extraction. Thus, the number of boxes suggested before the bounding box that will be formed in the final image has been reduced from approximately 2000 to 300 and has led to an improvement in the quality of these boxes. Therefore, the pattern detection rate is higher [30].

Since Mask RCNN is derived from Faster RCNN, the model structures are very similar, the only difference is that Mask RCNN also gives extra mask information as output. Within the scope of the project, it is planned to use this model for the detection of lanes on the road. For this reason, 3 lane lines on the two-lane road image on the created dataset are named and labeled separately. Then, after the necessary configuration setting was made, the model training was carried out. Within the scope of the study, Faster RCNN was applied to the image taken from the simulation environment and the detection process was carried out successfully. An example image of this study is given in Fig. 7.

### *D. Lane Detection and Tracking*

The model was trained with the Mask RCNN algorithm and when it was applied to the image taken from the simulation environment, the path detection process was carried out successfully. After this process, it is aimed to obtain a clean road image by removing the road detected in the image from the entire image. For this purpose, the road image masked by color filtering processes is separated from the image coming from the camera. The visual expression of the completed work is given in Fig. 8.



Fig. 8. Filtered Image Result

After these processes, the Faster RCNN algorithm, which was modeled on the filtered image output, was applied. With this process, it is aimed to determine the lane lines on the road image separated from the image. The visual of the process is given in Fig. 9.

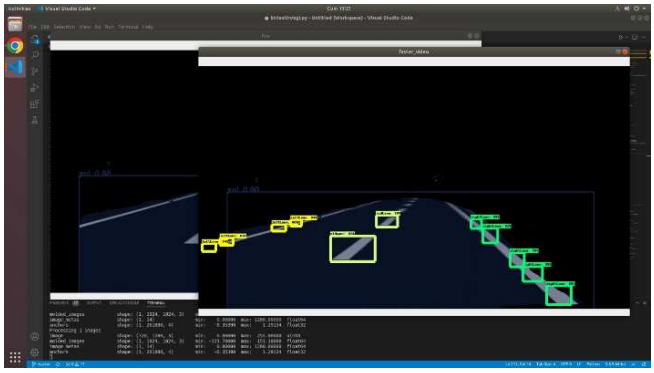


Fig. 9. Image with Mask RCNN and Faster RCNN Applied

Finally, in an image taken with the camera, the road and the lanes on the road were determined. However, since it would not be efficient for lane tracking in this way, after this step, position information of the bounding box on each final image was obtained by using image processing techniques. According to received location information, the averages of each bounding box location were determined. Based on this information, straight lines were drawn on the road lanes in the image. By changing the pixel distances of the vehicle' s position to these road lane lines, the approach to the road lanes is determined. Based on these pixel's distance information received, vehicle steering control is provided. The visual of this process is given in Fig.10.

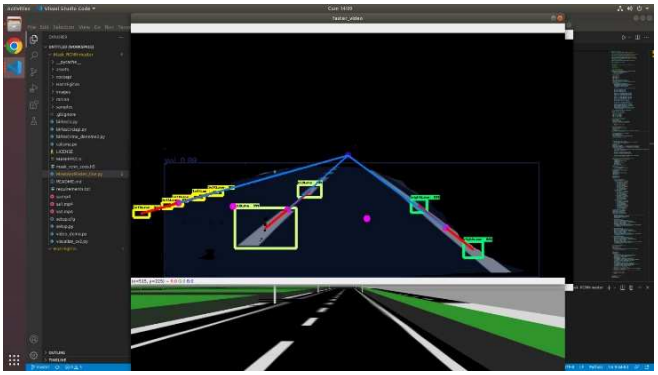


Fig. 10. Result Image

## III. CONCLUSION

This study, it is carried out to solution two basic problems in the realization of lane tracking with image processing methods. The first of this problem is realized due to the problem of vehicle access lost during the masking process based on the secure four-point call received in the ROI process. The Mask R-CNN algorithm applied within the scope of the study detects %99 correctly in the simulation environment and realizes the solution of this problem. The second problem is the color conversions and method used in the road lane. The reason for this problem is that the values of the color filters change due to the change in the basic physical environment conditions such as lighting, shading, clouding or reflection, such as changing weather conditions or different sun angles in ordinary life. The difficulty of detecting different filter values at different times of the day by using Faster RCNN, direct detection of lane lines in the simulation environment has been achieved with %99 success.

978-1-6654-8894-5/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: Bozok Universitesi. Downloaded on November 12,2022 at 12:22:06 UTC from IEEE Xplore. Restrictions apply.

If the vehicle approaches the right or left lane lines, the approach amounts were determined by using the image processing techniques used after this stage and the steering control was carried out successfully. Simulation results have promising result to apply these techniques for further studies such as an autonomous vehicle in real-time.

## REFERENCES


[1] World Health Organization. Road traffic injuries. https://www.who.int/news-room/fact-sheets/%20detail/road-traffic-injuries/. Last Access Date: 15.04.2022

[2] National highway traffic safety administration. Risky driving. https://www.nhtsa.gov/risky-driving . Last Access Date: 15.04.2022

[3] Police Academy. https://www.pa.edu.tr/upload/editor/files/trafik_kaza_ve_denetim_istatistikleri.pdf , Last Access Date: 15.04.2022

[4] J. Piao & M. McDonald (2008) Advanced Driver Assistance Systems from Autonomous to Cooperative Approach, Transport Reviews, 28:5, 659-684, DOI:10.1080/01441640801987825

[5] Venhovens, P., Naab, K. and Adiprasito, B. (2000) Stop and go cruise control, FISITA World Automotive Congress, Seoul, 12–15 June.

[6] D. Gerónimo, A. M. López, A. D. Sappa and T. Graf, "Survey of Pedestrian Detection for Advanced Driver Assistance Systems," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1239-1258, July 2010, doi: 10.1109/TPAMI.2009.122.

[7] V. K. Kukkala, J. Tunnell, S. Pasricha and T. Bradley, "Advanced Driver-Assistance Systems: A Path Toward Autonomous Vehicles," in *IEEE Consumer Electronics Magazine*, vol. 7, no. 5, pp. 18-25, Sept. 2018, doi: 10.1109/MCE.2018.2828440.

[8] Y. Yeniaydin and K. W. Schmidt, "A lane detection algorithm based on reliable lane markings," 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1-4.

[9] Satti, S., Suganya Devi, K., Dhar, P., Srinivasan, P." A machine learning approach for detecting and tracking road boundary lanes", ICT Express, Volume 7, Issue 1,2021, Pages 99-103.

[10] K. H. Lim, K. P. Seng, L. -M. Ang and S. W. Chin, "Lane Detection and Kalman-Based Linear-Parabolic Lane Tracking," *2009 International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2009, pp. 351-354, doi: 10.1109/IHMSC.2009.211.

[11] José María Armingol, Arturo de la Escalera, Cristina Hilario, Juan M. Collado, Juan Pablo Carrasco, Marco Javier Flores, José Manuel Pastor, F José Rodríguez, IVVI: Intelligent vehicle based on visual information, Robotics and Autonomous Systems, Volume 55, Issue 12, 2007, Pages 904-916, ISSN 0921-8890, https://doi.org/10.1016/j.robot.2007.09.004.

[12] Sun TY, Tsai S, Chan V (2006) HSI color model-based lanemarking detection. In: IEEE intelligent transportation systems conference, pp 1168–1172.

[13] Nedevschi S, Schmid R, Graf T, Danescu R (2004) 3D lane detection system based on stereovision. In: IEEE conference on intelligent transportation systems, pp 161–166.

[14] Formosa, N., Quddus, M., Ison S., Abdel-Aty M., Yuan, J. "Predicting real-time traffic conflicts using deep learning," Accid. Anal. Prev., vol. 136, no. December 2019, 2020.

[15] Huval B, Wang T, Tandon S (2015) An empirical evaluation of deep learning on highway driving, arXiv preprint. arXiv: 1504.01716.

[16] Li X, Mei X, Prokhorov D (2016) Deep neural network for structural prediction and lane detection in trafc scene. IEEE Trans Neural Netw Learn Syst 28:690–703.

[17] Kebria, P. M., Khosravi, A., Salaken S. M. and Nahavandi, S. "Deep imitation learning for autonomous vehicles based on convolutional neural networks," in IEEE/CAA Journal of Automatica Sinica, vol. 7, no. 1, pp. 82-95.

[18] Neven, D., Brabandere, B. D., Georgoulis, S., Proesmans M. and Gool, L. V. "Towards End-to-End Lane Detection: an Instance Segmentation Approach," *2018 IEEE Intelligent Vehicles Symposium (IV)*, Changshu, China, 2018, pp. 286-291.

[19] Muthalagu, R., Bolimera, A. & Kalaichelvi, V. Vehicle Lane markings segmentation and keypoint determination using deep convolutional neural networks. Multimed Tools Appl 80, 11201–11215 (2021). https://doi.org/10.1007/s11042-020-10248-2.

[20] Zhang, W., & Mahale, T. (2018). End to End Video Segmentation for Driving: Lane Detection For Autonomous Car. *ArXiv, abs/1812.05914*.

[21] B. Liu, H. Liu, J. Yuan, "Lane Line Detection based ON Mask R-CNN", Proceedings of the 3rd International Conference on Mechatronics Engineering and Information Technology (ICMEIT),2019,pp. 696-699,doi: 10.2991/icmeit-19.2019.111.

[22] Z. Sun, H. Geng and B. Jia, "Road Segmentation Algorithm Based on Simplified Mask R-CNN for GF-3 SAR Images," *2021 SAR in Big Data Era (BIGSARDATA)*, 2021, pp. 1-4, doi: 10.1109/BIGSARDATA53212.2021.9574293.

[23] J. Tian, J. Yuan and H. Liu, "Road Marking Detection Based on Mask R-CNN Instance Segmentation Model," *2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL)*, 2020, pp. 246-249, doi: 10.1109/CVIDL51233.2020.00-92.

[24] Liang, D., Guo, YC., Zhang, SK. *et al.* Lane Detection: A Survey with New Results. *J. Comput. Sci. Technol.* **35,** 493–505 (2020). https://doi.org/10.1007/s11390-020-0476-4

[25] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980-2988, doi: 10.1109/ICCV.2017.322.

[26] A. Nur Ömeroğlu, N. Kumbasar, E. Argun Oral and I. Y. Ozbek, "Mask R-CNN Algoritması ile Hangar Tespiti Hangar Detection with Mask R-CNN Algorithm," 2019 27th Signal Processing and Communications Applications Conference (SIU), 2019, pp. 1-4, doi: 10.1109/SIU.2019.8806552.

[27] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142-158, 1 Jan. 2016, doi: 10.1109/TPAMI.2015.2437384.

[28] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580-587, doi: 10.1109/CVPR.2014.81.

[29] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

[30] B. Liu, W. Zhao and Q. Sun, "Study of object detection based on Faster R-CNN," *2017 Chinese Automation Congress (CAC)*, 2017, pp. 6233-6236, doi: 10.1109/CAC.2017.8243900.


978-1-6654-8894-5/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: Bozok Universitesi. Downloaded on November 12,2022 at 12:22:06 UTC from IEEE Xplore. Restrictions apply.