International Journal of Engineering Sciences & Emerging Technologies, Dec. 2024
ISSN: 22316604 Volume 11, Issue 4, pp: 924-931 ©IJESET

VEHICLE DETECTION WITH DEEP LEARNING BASED APPROACH: A REVIEW

Muskan Jain [1] Dr. Ranjana Sharma[2]
Research Scholar[1] Student, Teerthanker Mahaveer University, Moradabad (U.P.), India
Associate Professor[2], Teerthanker Mahaveer University, Moradabad (U.P.), India

Email: muskanjain.scholar@tmu.ac.in[1], Ranjana.computers@tmu.ac.in[2]

ABSTRACT

The essential role of vehicle detection and tracking technology in the field of intelligent transportation management (ITMS), This essay suggests an innovative approach tailored to enhance accuracy, particularly for small target vehicles. The focal point of this methodology is the introduction of the vehicle detection model YOLOv5-NAM, an improved a variant of theYOLOv5s model achieved through the integration of The attention module based on normalization (NAM). Notably, this augmentation yields a notable 1.6% improvement in mean Average Precision compared to the original YOLOv5s model [1].
Additionally, the paper introduces a real-time tracking technique for tiny target automobiles, labeled as JDE-YN, leveraging using the YOLOv5-NAM model as the car sensor. This tracking method incorporates feature extraction into the joint training prediction head, resulting in a significant 0.9% enhancement in Multiple Object Tracking Accuracy (MOTA) compared to the original JDE technique.
The study employs algorithms for primary-stage target detection, specifically utilizing the SSD and YOLOv3 algorithms. Utilizing image data from an road vehicle dataset available for public use, the models are trained and subsequently compared for their effectiveness in vehicle detection. This analysis extend beyond vehicle detection, impacting areas such as semantic segmentation and target tracking, and applications in driverless operation.

KEYWORDS: Vehicle Detection, Accuracy, YOLOv3, YOLOv5, MOTA, MOTP.

1. INTRODUCTION

In the era among intelligent transportation networks, the proliferation of communication services for vehicles has surged, leading to a strain for vehicle network connectivity using radio frequency resources. Cognitive radio-equipped automobile networks emerge as a solution to this challenge by offering additional spectrum resources, which could transform a number of industries, including as aerospace, railroad, and road transportation, and military domains. This paper delves into the critical role of tracking and detection of vehicles technology in the context.
The development of vehicle detection and tracking technology is essential for various applications, including intelligent transportation systems and driving. This paper explores different approaches to vehicle detection **[1], [3]**. focusing on traditional algorithms and deep learning-based methods, specifically the YOLOv5-NAM and JDE-YN algorithms. The contributions of this paper lie in enhancing The precision and instantaneous execution of small aim vehicle discovery and multi-vehicle tracking through attention mechanisms and advanced tracking strategies.
Technology for tracking and detecting vehicles is crucial to intelligent transportation management and control systems. **[21]**In order to achieve high recognition and tracking accuracy for small target vehicles, this research introduces a unique attention-based vehicle detection and tracking strategy. First, we extend the traditional YOLOv5s model with the normalization-based attention module (NAM) to create a new vehicle identification model called YOLOv5-NAM.
.

2. REVIEW OF THE LITERATURE

Vehicle detection can be done in a number of ways, both conventional and modern. While supervised learning makes use of known and labelled input data, unsupervised classification makes use of unknown data and labels. The following three strategies-

2.1 STRATEGIES

2.1.1 Traditional Vehicle Detection Algorithms:

The paper begins by categorizing vehicle detection algorithms into three main types: those based on prior knowledge, shallow machine learning, and deep learning. Traditional methods relying on prior knowledge often struggle with changing lighting conditions and presence of non-vehicle objects with similar shadows. Shallow machine learning methods**[4],[6]** while improving accuracy, face challenges in feature selection and model scalability. These methods, including frame difference, streamer, and background modeling**[7],[8]**.and affected by environmental factors and struggle with real-time requirements.

2.1.2 YOLOv5-NAM: Enhancing Small Target Vehicle Detection:

The paper introduces YOLOv5-NAM as an improvement over the YOLOv5 model, incorporating normalization-based attention modules (NAM). This enhancement aims to boost the detection accuracy of small target vehicles. By focusing on spatial and channel attention mechanisms, YOLOv5-NAM addresses limitations in traditional models and aims to provide better performance, particularly in scenarios where small vehicles might be challenging to detect accurately.

2.1.3 Algorithm for Vehicle Detection Utilising Deep Learning:

In this paper then delves into the realm Various techniques for vehicle detection based on deep learning **[9] ,[11]** focusing on both one- and two-stage detection methods**[14].** Two-stage methods like R-CNN, Fast R-CNN, and Faster R-CNN**[11-12]** generate candidate regions before processing them through convolutional neural networks (CNN)**[15].** While achieving improved accuracy, they face various Challenges related to redundant calculations and increased computational costs. In contrast, one-stage methods such as YOLOv3[16] offer real-time detection by directly predicting bounding boxes and classifications, and making them suitable for applications like intelligent transportation system.

3. SYSTEM MODEL

3.1.1 Algorithm Selection Analysis:

The subsection delves into the evaluation of common target detection algorithms, emphasizing the key factors of detection rate and accuracy and analysis. The analysis justifies the choice of YOLOv3[16], considering its superior performance in identifying tiny items and real-time capabilities. A comparative study of algorithms, including Faster R-CNN, provides insights into the decision-making process.

3.1.2 Detection Model:

The model incorporates the YOLOv5 algorithm, known for its iterative improvements from YOLOv1 to YOLOv5, enhancing both detection accuracy and speed. Shown in Figure 1. The YOLOv5 network model consume of distinct components: Prediction Head, Neck, Backbone, and Input. The Contribution phase involves data preprocessing using Mosaic data enhancement and an adaptive anchor box strategy. The Backbone, based on Darknet53, employs Conv, C3, and SPPF modules to extract image**[17]** features as shown in figure 2. The Neck fuses these features using the FPN+PAN structure, leading to the Prediction Head, which performs detection and classification. The YOLOv5 algorithm makes use of GIoU Training Loss and non-maximum suppression to refine the detection results.

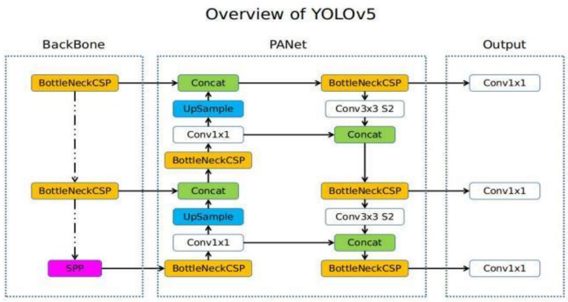

Figure 1. Overview of YOLOv5 (Backbone, PANet, Output)

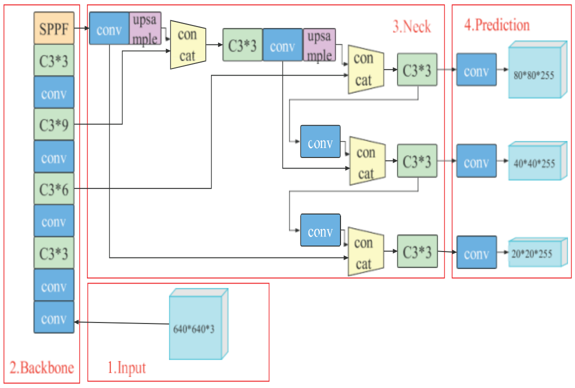

Figure 2. Network model of YOLOv5 [17].

- **Input**

The input section supplies the data needed to train the model and preprocess the experimental dataset. The

algorithm focuses on effective data preprocessing to enhance model training using Mosaic data augmentation and an adaptive anchor box strategy. In the Input phase, which serves as the foundation for model training, the algorithm leverages Mosaic data augmentation inspired by CutMix. The adaptive anchor box strategy is another crucial component implemented during the Input phase.

- **Backbone**

The Backbone serves as a crucial component within the vehicle detection network model, responsible for extracting key features from the input image through diverse (CNN) operations. Over the years, researchers have developed several noteworthy Backbones, including VGG, Darknet53, ResNet, and MobileNet. In this contex YOLO series algorithms, the Backbone is rooted in Darknet53, which has undergone extensive exploration and adaptation.

- **Neck**

The Neck serves as a critical intermediary component situated between the Prediction Head and the Backbone within the Network model for vehicle detection. It's primary function is to amalgamate the characteristics of the image that the Backbone, facilitating a seamless transition for subsequent processing by the Prediction Head. In this context of YOLOv5 algorithm.

The FPN (Feature Pyramid Network) within the Neck plays a pivotal role by transmitting high-level semantic information downward and merging it with the features from the Backbone[20]. This integration ensures a comprehensive representation of the image features as shown in figure 4., combining both high and low-level information. Concurrently, the PAN (Path Aggregation Network) structure within the Neck is responsible for conveying information on low-level features upward and merging it with the FPN network. This hierarchical fusion process enables the Neck to create a cohesive and enriched feature map.

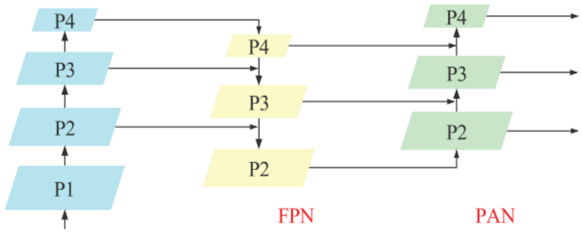

Figure 3. The FAN+PAN Structure

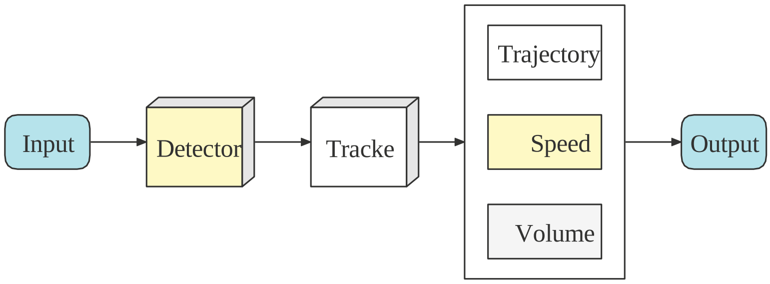

Figure 4. The model of vehicle tracking based on detection

4. MODEL DESIGN

4.1.1 Detection Module of Network:

This subsection explores the architecture of the detection and image detection module within the network, focusing on the utilization of Using DarkNet-53 as the primary network. It addresses the challenges associated with deep-level networks, emphasizing the improvements brought about by residual connections. Additionally, it provides an in-depth examination of the loss function, highlighting its role in training the model effectively.

DarkNet-53, a pivotal component of the architecture, serves as a robust foundation for the detection. The utilization of this backbone network is motivated by its proven effectiveness in handling complex visual tasks. The paper highlights the network's capability to overcome challenges associated with deep-level architectures, thereby ensuring improved feature extract and representation.

5. Dataset and Analysis

5.1.1 Dataset Selection: This dataset contains images of road vehicles captured by driving vehicles, encompassing various vehicle types,, and other objects. scenarios, featuring varying vehicle sizes, backgrounds, and lighting conditions shown in figure 5. The detection and tracking system was implemented and tested on this dataset to evaluate its performance in accurately identifying and tracking vehicles.

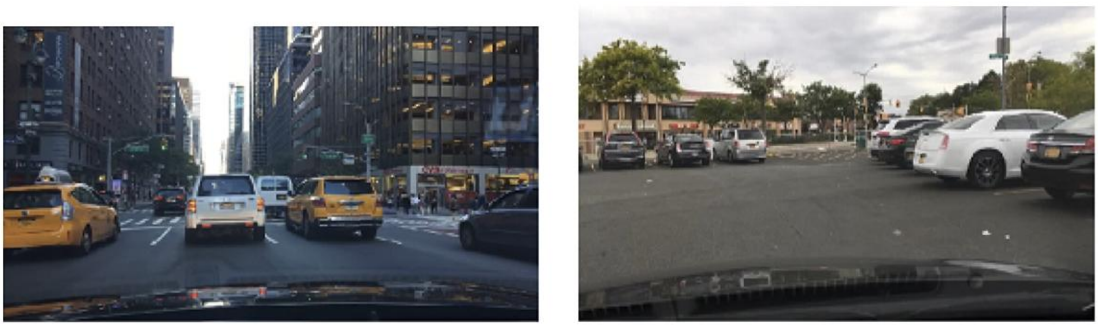

Figure 5. BDD100k data image.

5.1.2 Model Detection: The YOLOv3 algorithm was employed and trained for model training, focusing on the detection is divided in three characteristic layers within the primary network. Key parametersas displayed in Figure 2, were set to facilitate effective training. The learning rate was adjusted using a breakpoint continuation method, initially training with a larger learning rate and then optimizing with a smaller learning rate. The training processes spanned 50 epochs, with additional retraining options.

5.1.3 Detection Accuracy: The primary metric assessed was the detection accuracy of this. The YOLOv5 detection algorithm, known for its iterative improvements and real-time capabilities, demonstrated commendable accuracy in identifying vehicles within frames. The use of an adaptive anchor box approach and mosaic data improvement in the Input phase significantly contributed of this robustness of the detection model.

(a) (b)

Figure 6. Data association failure

5.1.4 Tracking Precision: The Kalman filter-based tracking model complemented the detection system by improving multi tracking precision**[18].** By continuously estimating the state of the system and adjusting estimates based on observations shown in figure 6, the Kalman filter compensated for the limitations of vehicle detector.

This was particularly evident in scenarios involving occlusion or abrupt changes in vehicle movement, where the tracker maintained accurate trajectory information.

this tracking precision was quantified through the correlation of automobiles in nearby frames to create a full trajectories. This Kalman filter's ability to continuously observe and estimate the state of the system ensured that the tracked vehicles closely aligned with their actual positions, enhancing the overall reliability of the tracking model..

5.1.5 Real-Time Performance: An essential aspect of the evaluation was the performance in real time of the integrated system. This algorithm of YOLOv5, renowned for its efficiency, combined with the Kalman filter's computational optimization, demonstrated the system's suitability for real-time applications. The timely processing of each frame, coupled with accurate detection and tracking, validated the system's practicality for dynamic, real-world environments.

5.1.6 Map Value Analysis: mAP value, a crucial metric, was calculated to measure the model's overall detection effectiveness. The model achieved a mAP value of 72.8. Although numerically slightly lower, the detection performance was notable. The analysis revealed outstanding performance in detecting cars, with buses and motorcycles showing slightly lower detection scores.

Despite being able to be recognised as a bus category, the figure on only shows three buses. The left, andonly the first bus has a higher score. Another side, The right graphic illustrates how the bus category's box is not entirely full, leaving some of the vehicle's body outside of the detection box.

International Journal of Engineering Sciences & Emerging Technologies, Dec. 2024
ISSN: 22316604 Volume 11, Issue 4, pp: 924-931 ©IJESET

Figure 7: Bus category detection impact.

5.1.7 Detection Effect on Bus Category

The overall accuracy of the model is commendable, especially in capturing instances belonging to the "car" and "person" categories as shown in figure 7.

The overwhelming majority of the training data consists of instances from the "car" category, reaching millions. Conversely the "bus" classification has just over 10,000 instances as well as the "motor " categories even more limited, with only over 4,000 instances. Consequently, the model predominantly encounters and trains on features related to the "car" category during the training process.

The detection effect varied across different categories as shown in figure 7. For cars, the model demonstrated excellent detection, with high scores and accurate bounding boxes. Trucks exhibited a respectable detection effect, while buses showed a comparatively lower score, indicating room for improvement. Motorcycles and bikes.

6. Methods for Various Approaches

This review study requires us to examine different type of methodology for Vehicle Detection.

6.1.1 UA-DETRAC

The (Unmanned Aerial Vehicle benchmark - DETRAC) **[19]** is a collection of is designed for vehicle classification in unconstrained aerial environments. While the DETRAC dataset is primarily used for vehicle detection in road surveillance. Its dataset as shown in figure 8. designed for the evaluation of object detection algorithms, particularly for vehicle classification in aerial imagery. If you are specifically working with the vehicle detection dataset UA-DETRAC using machine learning.

6.1.2 COCO (Common Objects in Context):

For tasks involving object detection, segmentation, and captioning, COCO is a popular large-scale dataset. It is appropriate for training and assessing different computer vision techniques since it has a rich annotation format, a broad collection of images, and complicated situations. models.

The evaluation of four methods on the UA-DETRAC dataset reveals challenges associated with diverse weather conditions and varying scales of vehicle targets, impacting the performance of vehicle target detectors. The MOTA (Accuracy of Multiple Object Tracking) and MOTP (Precision of Multiple Object Tracking) values across the methods consistently range between 0.3 and 0.4, indicating the difficulty of accurately detecting and tracking vehicles in this dataset. These low values can be attributed to the dataset's encompassing nature, presenting obstacles for vehicle detectors.

The evaluation of four methods on the UA-DETRAC dataset reveals challenges associated with diverse weather conditions and varying scales of vehicle targets, impacting the performance of vehicle target detectors.

6.1.3 MOTA

The Multiple Object Tracking Accuracy (MOTA) and Precision (MOTP) measurements values across the methods consistently range between 0.3 and 0.4, indicating the difficulty of accurately detecting and tracking vehicles[23] in this dataset as shown in table 8. These low values can be attributed to the dataset's encompassing nature, presenting obstacles for vehicle detectors.

7. Comparison Table of Different Approaches

Table 8: The dataset's detection findings.

Model	NMS	CAR	BUES	VAN	mAP
YOLOv5s	IoU-NMS	0.697	0.433	0.743	0.503
YOLOv5-NAM	IoU-NMS	0.7	0.46	0.731	0.517
YOLOv5s	SD-NMS	0.712	0.424	0.735	0.504
YOLOv5s	SD-NMS	0.72	0.446	0.726	0.519

8. RESULT

The evaluation of vehicle detection methods on the UA-DETRAC dataset, tailored for unconstrained aerial environments, underscores the challenges posed by diverse weather conditions and varying vehicle scales. The MOTA and MOTP metrics consistently fall within the 0.3 to 0.4 range, indicating the arduous task of accurate detection and tracking in this comprehensive dataset. Notably, the YOLOv5-NAM model, enriched with the normalization-based attention module, exhibits superior (mAP) values, especially in the CAR and VAN categories, compared to the baseline YOLOv5s. The comparison table [8] reveals nuanced insights the impact of different Non-Maximum Suppression (NMS) techniques on detection performance. The findings emphasize the importance of model selection, particularly considering specific challenges posed by datasets like UA-DETRAC. Finally, this study provides valuable guidance for optimizing vehicle detection models in environment.

9. CONCLUSION

The YOLOv5-NAM, an evolved version of the classical YOLOv5s model, demonstrates notable improvements in vehicle detection. Moreover, the devised real-time JDE-YN tracking technique for small target vehicles, showcases its efficacy in tracking multiple vehicles simultaneously.

The experimental evaluations conducted regarding the COCO and UA-DETRAC datasets yield promising results. In comparison Comparing the YOLOv5-NAM model to the YOLOv5s model achieves a commendable 1.6% increase in mAP value, signifying enhanced detection accuracy. The JDE-YN method[21], applied for vehicle tracking, demonstrates a 0.9% improvement in MOTA value over the conventional JDE algorithm. Additionally, the proposed tracking method remarkably reduces the instances of identity switching by 15%. These outcomes collectively affirm the effectiveness of the devised approach in real- timing of occlusion vehicle tracking and tiny target vehicle detection.

The implications of the research extend beyond the immediate improvements in detection and tracking metrics. The methods introduced pave the way for advancing research in the more general area of monitoring and identifying automobiles. The successful application of YOLOv5-NAM[22] and JDE-YN on real-world datasets suggests their crucial for integration into ITS and contributes to the ongoing progress in the domain of artificial intelligence for vehicular applications.

Acknowledgement

I would like to thanks and I am extremely grateful to my supervisor, Dr. Ranjana Sharma for guidance and all the useful discussions and contributed to this research paper or supported me. Her deep insights helped me at various stages of my research.

References

[1] H. Song, J. Zhu, and Y. Jiang, "Two-stage merging network for describing traffic scenes in intelligent vehicle driving system,"*IEEE Transactions on Intelligent Transportation Systems,*no. 99, pp. 1–12, 2021.

[2] P. Barcellos, V. Gomes, and J. Scharcanski, "Shadow detection in camera-based vehicle detection: survey and analysis," *Journal of Electronic Imaging*, vol. 25, no. 5, p. 051205, Article ID 051205, 2016.

[3] H. Thomas and H. Michael, "Camera-based method for distance determination in a stationary vehicle, corresponding computer program and corresponding parking assistance system with camera," 2020, https://patents.google.com/ patent/WO2012076400A1/en.

[4] Y. Chen, W. Li, C. Sakaridis, and D. Dai, "Domain adaptive faster R-CNN for object detection in the wild," 2018 IEEE, https://arxiv.org/abs/1803.03243.

[5] I. Papakis, A. Sarkar, A. Svetovidov, and J. S. Hickman, "Convolutional neural network-based In-vehicle occupant detection and classification method using second strategic highway research program cabin images," *TransportationResearch Record Journal of the Transportation Research Board,*2021.

[6] G. Goedert, D. Jungen, J. Beck et al., Weight -responsive Vehicle Seat Occupant Detection and Classification Method and System, US20180244172A1[P], 2018.

[7] M. A. A. Al-qaness, A. A. Abbasi, H. Fan, R. A. S. H. Ibrahim, and A. Hawbani, "An improved YOLO-based road traffic monitoring system," *Computing*, vol. 103, no. 2, pp. 211–230,2021.

[8] H. Yin, Bo Chen, C. Yi, and Y. D. Liu, "Overview of target detection and tracking based on vision," *Acta Automatica Sinica*, vol. 42, no. 10, pp. 1466–1489, 2016.

[9] Chai, W.; Lu, Y.; Velipasalar, S. Weighted average precision: adversarial example detection for visual perception of au- tonomous vehicles. In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021; Volume 2021, pp. 804–808.

[10] Hong, H.; He, K.; Gan, Z.; Jiang, G. Stability Precision Error Correction of Photoelectric Detection by Unmanned Aerial Vehicle.

J. Sens. **2021**, *2021*, 5564448.

[11] Xu, W.; Huang, H.; Pan, S. Using feature alignment can improve clean average precision and adversarial robustness in object detection. In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021; pp. 2184–2188.

[12] Xu, W.; Huang, H.; Pan, S. Using feature alignment can improve clean average precision and adversarial robustness in object detection. In Proceedings of the 2021 IEEE International Conference on Image Processing (ICIP), Anchorage, AK, USA, 19–22 September 2021; pp. 2184–2188.

[13] Hong, H.; He, K.; Gan, Z.; Jiang, G. Stability Precision Error Correction of Photoelectric Detection by Unmanned Aerial Vehicle.

J. Sens. **2021**, *2021*, 5564448.

[14] Hoffmann, J.E.; Tosso, H.G.; Santos, M.M.D.; Justo, J.F.; Malik, A.W.; Rahman, A.U. Real-Time Adaptive Object Detection and Tracking for Autonomous Vehicles. *IEEE Trans. Intell. Veh.* **2021**, *6*, 450–459.

[15] He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2980–2988.

[16] Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.

[17] Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. .

[18] Wen, L.; Du, D.; Cai, Z.; Lei, Z.; Chang, M.C.; Qi, H.; Lim, J.; Yang, M.H.; Lyu, S. UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking. *Comput. Vis. Image Underst.* **2015**, *193*, 102907.

[19] Wen, L.; Du, D.; Cai, Z.; Lei, Z.; Chang, M.C.; Qi, H.; Lim, J.; Yang, M.H.; Lyu, S. UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking. *Comput. Vis. Image Underst.* **2015**, *193*, 102907.

[20] Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully Convolutional One-Stage Object Detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9626–9635.

[21] P. Maduranga, N. Pethiyagoda, M. Kulasekara, and T. Weerawardane, "Deep Learning-Based Vehicle Type Detection and Classification," *Int. J. Comput. Appl. Math. Comput. Sci.*, vol. 3, Jun. 2023, doi: 10.37394/232028.2023.3.3.

[22] Q. Zhao, W. Ma, C. Zheng, and L. Li, "Exploration of Vehicle Target Detection Method Based on Lightweight YOLOv5 Fusion Background Modeling," *Appl. Sci.*, vol. 13, p. 4088, Mar. 2023, doi: 10.3390/app13074088.

[23] S. Luo, "Comparison and summary of Faster R-CNN, YOLOv3 and YOLOv5 applied in vehicle detection," *J. Phys. Conf. Ser.*, vol. 2649, p. 012015, Nov. 2023, doi: 10.1088/1742-6596/2649/1/012015.