Object Detection in Road Scene Images for Autonomous Driving

A Dissertation
Submitted in Partial Fulfillment of the Requirements
for the Award of the Degree of

Bachelor of Technology
in
Computer Science & Engineering

Submitted by

Rishita Singh(1906022)

Madhu Sah(1906027)

under the Supervision of

Dr. Rajib Ghosh
(Assistant Professor Grade I)

Department of Computer Science & Engineering
National Institute of Technology Patna

Jan-April 2023

राष्ट्रीय प्रौद्योगिकी संस्थान, पटना

NATIONAL INSTITUTE OF TECHNOLOGY PATNA

Certificate

This is to certify that ***Rishita Singh with Roll No.1906022***, ***Madhu Sah with Roll No.1906027***, *have carried out Major Project entitled* ***"Object Detection in Road Scene Images for Autonomous Driving"*** *during their 8th semester under the supervision of* ***Dr.Rajib Ghosh***, *Assistant Professor Grade-I, CSE Department, in partial fulfillment of the requirements for the award of Bachelor of Technology degree in the Department of Computer Science Engineering, National Institute of Technology Patna.*

Dr. Rajib Ghosh
Assistant Professor Grade-I
Department of Computer Science & Engineering
National Institute of Technology Patna
May 2023

राष्ट्रीय प्रौद्योगिकी संस्थान, पटना

NATIONAL INSTITUTE OF TECHNOLOGY PATNA

Declaration & Copyright

We, the students of the 8th semester, hereby declare that we have completed the Major Project entitled ***"Object Detection in Road Scene Images for Autonomous Driving"*** *has been carried out by us in the Department of Computer Science and Engineering of the National Institute of Technology Patna under the guidance of* ***Dr.Rajib Ghosh****, (Assistant Professor Grade-I) of Computer Science and Engineering, NIT Patna. No part of this project has been submitted for the award of the degree or diploma to any other Institute.*

Rishita Singh(1906022): _____

Madhu Sah(1906027): _____

Acknowledgements

We take this opportunity to express our profound gratitude and deep regards to Prof. ***Dr.Rajib Ghosh****, for his exemplary guidance, monitoring, and constant encouragement throughout the course of this project. We also extend our heartfelt thanks to the entire CSE department of NIT Patna, for providing the necessary technical facilities and the technical environment which supported us to perform to the very best of our potential. The motivation that we gained for this project, shall get along a long way in learning and leveraging these technologies further, and incorporating them into our project work in the future. We are extremely humbled and gratified for the opportunity and exposure provided to us in Deep Learning through this project.*

Rishita Singh (Roll. No. **1906022**)
B. Tech. (C.S.E.) - 8th Semester
Madhu Sah (Roll. No. **1906027**)
B. Tech. (C.S.E.) - 8th Semester

Abstract

Object detection is a fundamental task in computer vision, with numerous applications such as autonomous driving, surveillance, and image retrieval.In object detection, the goal is to locate and classify objects of interest within an image or video. The detection of vehicles, pedestrians, and other objects on the road is crucial for ensuring the safety of passengers and pedestrians.In this project, we aim to perform object detection in the road scene images for the purpose of autonomous driving.

In this project, the Faster R-CNN algorithm with ResNet-50 as the backbone and Feature Pyramid Network (FPN) was utilized for object detection in road scene images. Additionally, the YOLOv6 algorithm was employed for the same purpose. Both models were trained using the training samples available in the dataset and fine-tuned using the validation samples.The Faster R-CNN algorithm is a two-stage object detection framework that first generates region proposals using a Region Proposal Network (RPN) and then classifies and refines these proposals.On the other hand, YOLOv6 is a one-stage object detection algorithm known for its efficiency and real-time performance.

The performance of the proposed system has been evaluated on the test samples present in the KITTI dataset.Output images showing the bounding boxes and scores for each object detected in the test images were also generated by us.It was observed from our evaluation results that YOLOv6 outperformed Faster R-CNN in terms of mean Average Precision (mAP@0.75). An impressive mAP@0.75 of 0.78 was achieved by YOLOv6, while Faster R-CNN achieved an mAP@0.75 of 0.62 on the test set. We conducted a comprehensive analysis of the model's performance on different object categories and gained valuable insights for future improvements.

Contents

Chapter 1

Introduction

Object detection plays a crucial role in various computer vision applications, including autonomous driving, surveillance systems, and image understanding. The ability to accurately and efficiently detect objects in images is essential for enabling advanced functionalities such as real-time decision-making, tracking, and scene understanding.

In this project, we focus on object detection specifically tailored for road scene images, with the goal of supporting autonomous driving systems. The accurate detection of objects like vehicles, pedestrians, and cyclists is of utmost importance in ensuring the safety and efficiency of autonomous vehicles on the road.

The advancements in deep learning and convolutional neural networks (CNNs) have revolutionized the field of object detection. These techniques have demonstrated remarkable performance in terms of accuracy and speed, paving the way for more sophisticated and reliable object detection models.

The primary objective of our project is to explore and implement state-of-the-art object detection algorithms and evaluate their performance on the popular KITTI dataset. The KITTI dataset is widely used for benchmarking object detection models in the context of autonomous driving. It contains a diverse collection of road scene images, annotated with ground truth object labels and bounding box coordinates.

To achieve our objective, we have chosen two well-established object detection models: Faster R-CNN with ResNet-50 as the backbone and Feature Pyramid Network (FPN),

and YOLOv6. These models have demonstrated excellent performance in previous studies and are known for their accuracy and efficiency in object detection tasks.

In this project report, we present a detailed analysis of the object detection results obtained using both Faster R-CNN and YOLOv6 models. We evaluate the performance of these models in terms of Average Precision (AP), precision, recall, and other relevant evaluation metrics. Additionally, we provide insights into the strengths and limitations of each model, highlighting potential areas for further improvement.

By conducting this project, we aim to contribute to the growing body of knowledge in the field of object detection for autonomous driving. The results and findings of this project can provide valuable insights and guidance for researchers and practitioners working on similar tasks, ultimately advancing the development of reliable and efficient autonomous driving systems.

Through this project, we hope to showcase the significance of accurate object detection in the context of autonomous driving and highlight the potential of state-of-the-art models in achieving this objective.

Chapter 2

Literature Survey

In, recent years, object detection accuracy has been remarkably improved by Deep Learning based frameworks. Recently, convolutional neural networks (CNN) has been propesed that shows high image clasiification accuracy, CNN can also be used for object detection such as R-CNN [9], which shows a great improvement on object detection accuracy compared with the conventional feature-based detectors.

In [9], region proposals and CNNs are combined which is called RCNN that regions with CNN features. In R-CNN, the possible objects are extracted by selective search, which proposes 2000 object regions, the extracted image content is aligned to the same size (227x227). Finally, the CNN with SVM classifiers assigns what type of objects the image content of region belongs to. However, R-CNN is slow because it performs a ConvNet forward pass for each object proposal.In order to speed up RCNN Spatial pyramid pooling networks (SPPnets) are proposed. Though, some drawbacks are (1)multistage pipeline training that requires disk storage which is very time consuming, (2) fixed convolution layers limits the accuracy of very deep networks.

Fast R-CNN [10], processes the whole image with several convolutional and max pooling layers to produce feature map. For each object proposal, a region of interest (ROI) pooling layer extract a fixed length feature vector from the feature map. Fast R-CNN appends a ROI max pooling layer and two full connection layers with after CNN layers. One of the full connection layers is designed for object category recognition and the other can fine-tune ROI position as a regression method. Then, Faster RCNN [11] proposed Regional Proposal Networks (RPNs) that improves the system performance since, it

shares convolutional layers with object detection networks. It merges a ROI proposal layer which gave k-possible region proposals and decides whether each region proposal contains an object.

Recently, Deep Learning based detection frameworks like Faster R-CNN [12], outperformed methods [13,15,16] that used sliding window approach. The Deep Learninig detection frameworks uses the output of last convolutional layer as feature map, which is used for localisation and classification [12] of objects. To accurately locate tiny vehicles (10X20 pixels), only small networks or shallow layers of standard modals like VGG16 [14] are applicable to provide a sufficiently high feature map resolution [13,16].

In [3], Faster R-CNN is extended by an additional Search Area Reduction modulewhich divides the input image into regions and predicts a confidence score of how likely a region contains at least one object. In Faster R-CNN multiple object proposals are predicted thus, the inference time for generating object proposals increases. In this paper they experiment is performed on aerial images that contain randomly located objets whose size is in the range of a only few pixels, this charecteristic helped in order to reduce inference time.

Two-stage approaches, e.g. Fast R-CNN [17], Faster R-CNN [12], and R-FCN [18], are comprised of an initial object proposal stage followed by a classification stage. Single stage approaches, e.g. SSD [19], YOLO [20], and YOLO9000 [21], directly predict object classes and locations in one step. In [1], a one-stage object detection framework is proposed for improving the detection accuracy while supporting a true real time operation based on YOLOv4. It achieves the best trade of between accuracy and speed for autonomous driving by using the deformable convolution to optimise the backbone network.

Chapter 3

Problem Statement

The aim of this work is to evaluate different object detection techniques that can accurately detect objects within images and draw bounding boxes around them during autonomous driving and evaluating its performance using relevant evaluation metrics such as average precision (AP) and mean average precision (mAP).

Chapter 4

Proposed Methodology

In this project, we explored two different object detection models, Faster R-CNN with ResNet-50 as the backbone and Feature Pyramid Network (FPN), as well as YOLOv6, for object detection on the KITTI dataset.

For Faster R-CNN:

Faster R-CNN with Feature Pyramid Network (FPN) is a powerful object detection model that combines two essential components: the Region Proposal Network (RPN) and the FPN itself. The RPN operates on the feature map generated by the backbone network and efficiently generates region proposals by predicting potential bounding boxes using anchor boxes of different scales and aspect ratios. On the other hand, FPN is a top-down architecture that produces feature maps at multiple scales, allowing the model to capture both fine-grained and high-level contextual information. This multi-scale feature representation enables Faster R-CNN with FPN to effectively detect objects of various sizes and handle complex scenes with occlusions and small objects. By leveraging the RPN and FPN together, Faster R-CNN achieves accurate object localization and classification, making it a popular choice for object detection tasks in computer vision.

We first trained the Faster R-CNN-FPN model on the training samples and fine-tuned it on the validation samples. The input image is passed through a backbone network, which is typically a convolutional neural network (CNN), here ResNet-50, that extracts features from the image. Then the output feature map from the backbone network

is passed through a region proposal network (RPN), which generates object proposals based on anchor boxes.The proposed regions are then aligned to a fixed size and passed through a feature pyramid network (FPN), which generates a set of feature maps at different scales.These feature maps are used to predict the object classes and bounding boxes using a RoI (region of interest) pooling layer and a set of fully connected layers.Finally, the classification and regression outputs are combined to generate the final detection results.

To quantitatively evaluate the performance of the Faster R-CNN model, we calculated the mean average precision (mAP) metric. The mAP was computed for each object class and for different image scales, including small, medium, and large. We utilized an Intersection over Union (IoU) threshold to match predicted bounding boxes with the ground truth annotations. Additional post-processing steps, such as non-maximum suppression, has been applied to refine the detection results.

For YOLOv6:

YOLOv6, another popular object detection model is also used for comparison. YOLOv6 is trained on training samples using a similar approach. YOLOv6 is known for its real-time performance and single-shot detection capability, making it suitable for scenarios where inference speed is crucial.

YOLOv6 incorporates several technical advancements that contribute to its superior performance in object detection. One key improvement is the adoption of the CSP-Darknet backbone network, which utilizes Cross Stage Partial connections for enhanced feature reuse and improved information flow. This architecture allows YOLOv6 to capture and leverage high-level and low-level features effectively, leading to better object detection accuracy. Additionally, YOLOv6 introduces various optimizations, such as the utilization of anchor boxes with multiple aspect ratios, to handle objects of different sizes and scales more effectively. The model also employs advanced data augmentation techniques, such as mosaic augmentation and random perspective transformation, to enhance the diversity and robustness of the training data. These technical advancements collectively contribute to YOLOv6's impressive performance in object detection tasks.

Similar to the evaluation process for Faster R-CNN,the output of the YOLOv6 is visually inspected. The mAP metric is used to evaluate the model's performance on the test

samples.

The results of our experiments, including mAP values are presented in the result section. Through this comparative analysis of Faster R-CNN with ResNet-50 backbone and YOLOv6, we aim to gain insights into the strengths and weaknesses of each model and identify their suitability for object detection in road scene images.

By considering both Faster R-CNN and YOLOv6 in our methodology, we explore the capabilities of two different approaches to object detection and enable a comprehensive evaluation of their performance on the KITTI dataset.

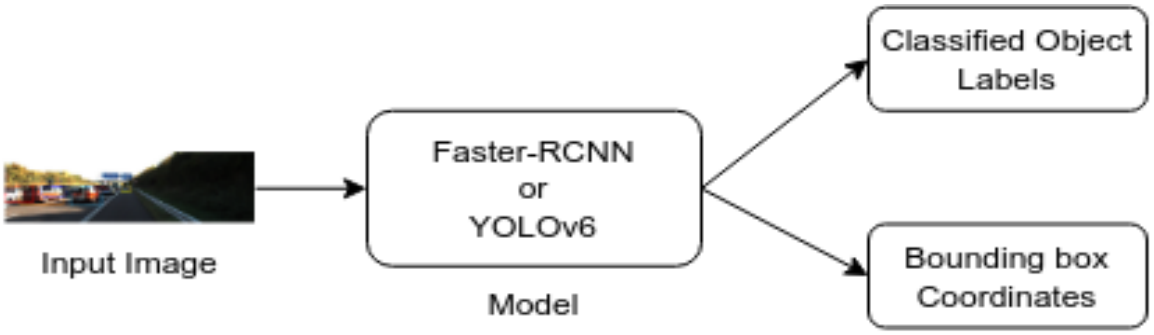

FIGURE 4.1: Overview of Proposed Object Detection System

The overview of Proposed Object Detection System is shown in Figure 4.1.

Chapter 5

Data Description and Result Analysis

The dataset used in this project is the KITTI dataset, which is widely used for object detection tasks. It consists of 7481 images, each labeled with one or more of the following 9 object classes: car, pedestrian, cyclist, truck, van, dontcare, misc, personsitting. The bounding box annotations for each object are provided in the format [xmin, ymin, xmax, ymax], where(xmin,ymin) represents the lower left coordinate of object and (xmax,ymax) represents the upper right coordinate of object. The images in the dataset are around size 1240x370 pixels. This dataset provides a diverse range of real-world images with varying lighting and weather conditions, making it a challenging dataset for object detection.

5.1 Data Description

The dataset used in this project is taken from "https://www.cvlibs.net/datasets/kitti/"

The dataset comprises of two main folders, containing images and corresponding labels, respectively.

1. image2 (image files in .png format)
2. label2 (text files containing annotations in .txt format)

FIGURE 5.1: Example images from the KITTI dataset

The images in the KITTI dataset are captured from a moving vehicle in real-world scenarios. The images are of high resolution and sizes, also contain different lighting conditions, weather conditions, and occlusions, making the dataset challenging for object detection. Overall, the images in your dataset represent a diverse set of scenarios that are relevant to real-world object detection applications.

FIGURE 5.2: Annotated images from the KITTI dataset

Annotated images are provided for illustrative purposes only. The bounding box coordinates and object labels were obtained from the KITTI dataset.

5.2 Quantitative Result Analysis

To evaluate the performance of the the model, a set of quantitative metrics comprising of precision, recall, average precision(AP) and mean average precision(mAP) have been used. They show the highest values of the quantitative metrics obtained until the corresponding epoch number.

Evaluation Metrics: For evaluating the proposed model we used precision, recall, average precision(AP) and mean average precision (mAP) have been used.

Precision: Precision measures the percentage of correct detections among all predicted bounding boxes,. It is computed as given in the equation below.

$$Precision = \frac{TP}{TP + FP} \tag{5.1}$$

Recall: Recall measures the percentage of correct detections among all ground truth annotations.. It is computed as is given in the below equation.

$$Recall = \frac{TP}{TP + FN} \tag{5.2}$$

Average Precision (AP): The area under the precision-recall curve, which captures the trade-off between precision and recall. A higher AP indicates better performance of the model.

Mean Average Precision (mAP): The average of the AP scores across all object categories.

Intersection over Union (IoU): IoU measures the overlap between the predicted bounding box and the ground truth bounding box. It is used to determine whether a detection is considered correct or not.

In this section, we present the results of the evaluation of the Faster R-CNN with ResNet-50 and YOLOv6 models for object detection on the KITTI dataset. We assessed the models' performance using Average Precision (AP), Average Recall (AR), and mean Average Precision (mAP).

AP-Car, AP-Pedestrian, AP-Cyclist, AP-Truck, AP-Van: These metrics represent the Average Precision for each object class achieved by the Faster R-CNN model. They indicate the model's ability to detect and classify objects accurately.

AP@0.50:0.95, AR@0.50:0.95, Small, Medium, Large: These metrics correspond to the Average Precision (AP) and Average Recall (AR) at different IoU thresholds for the YOLOv6 model. The Small, Medium, and Large categories refer to different object sizes. These metrics provide insights into the model's performance across different scales and the trade-off between precision and recall.

Faster R-CNN Results

The performance of the Faster R-CNN model on the KITTI dataset is summarized in Table 5.1.

AP-Car	0.62
AP-Pedestrian	0.37
AP-Cyclist	0.45
AP-Truck	0.66
AP-Van	0.47

APs	0.45
APm	0.50
APl	0.63

TABLE 5.1: Results of Faster R-CNN on the KITTI dataset

The Faster R-CNN model achieved respectable results for car detection (AP-Car) with a score of 0.62. However, the performance for pedestrian, cyclist, truck, and van detection was comparatively lower, with AP scores ranging from 0.37 to 0.66.

The Faster R-CNN model demonstrated its object detection performance on the KITTI dataset using two different Intersection over Union (IoU) thresholds: 0.50 and 0.75.

mAP@0.50=0.75

mAP@0.75=0.62

YOLOv6 Results

The performance of the YOLOv6 model on the KITTI dataset is presented in Table 2.

	AP@0.50:0.95	**AR@0.50:0.95**
Small	0.565	0.641
Medium	0.687	0.744
Large	0.777	0.845

TABLE 5.2: Results of YOLOv6 on the KITTI dataset

The table 5.2 showcases the performance results of YOLOv6. The evaluation metrics, namely Average Precision (AP) and Average Recall (AR), are reported for different object sizes.

Interestingly, YOLOv6 exhibits superior performance in terms of low-level feature extraction compared to Faster R-CNN. This is evidenced by the higher AP values achieved by YOLOv6 across all object sizes: small, medium, and large. YOLOv6 achieves an AP of 0.565 for small objects, 0.687 for medium-sized objects, and an impressive 0.777 for large objects.

Moreover, YOLOv6 demonstrates better performance in accurately detecting objects with complex backgrounds. This is indicated by the high AR values achieved by YOLOv6 across all object sizes: 0.641 for small objects, 0.744 for medium-sized objects, and an impressive 0.845 for large objects.

The YOLOv6 model demonstrated its object detection performance on the KITTI dataset using two different Intersection over Union (IoU) thresholds: 0.50 and 0.75.

mAP@0.50=0.91

mAP@0.75=0.78

Comparison

To compare the two models, we present a performance comparison in Table 5.3.

	AP@0.50	AP@0.75
Faster R-CNN	0.75	0.62
YOLOv6	**0.91**	**0.78**

TABLE 5.3: Performance Comparison of Faster R-CNN and YOLOv6 on the KITTI dataset

The comparison between Faster R-CNN and YOLOv6 reveals notable differences in their object detection performance. YOLOv6 outperformed Faster R-CNN with significantly higher AP scores at both IoU thresholds of 0.50 and 0.75. YOLOv6 achieved an impressive AP of 0.91 at an IoU threshold of 0.50, showcasing its strong detection capabilities. Similarly, at an IoU threshold of 0.75, YOLOv6 achieved an AP of 0.78, indicating its ability to accurately localize objects with higher precision.

These results suggest that YOLOv6 exhibits superior performance compared to Faster R-CNN in terms of detection accuracy and localization precision on the KITTI dataset.

The comparative evaluation of Faster R-CNN and YOLOv6 demonstrates the strengths of each model. While Faster R-CNN provides acceptable performance for car detection, YOLOv6 excels in detecting objects of various sizes and demonstrates higher overall detection accuracy. However, it is important to consider the specific requirements of the application and the trade-offs between accuracy and computational efficiency when selecting the appropriate model for a given scenario.

Overall, these results provide insights into the strengths and limitations of Faster R-CNN and YOLOv6 in the context of object detection on road scene images, contributing to the knowledge and understanding of object detection methodologies for autonomous driving and related applications.

5.3 Qualitative Result Analysis

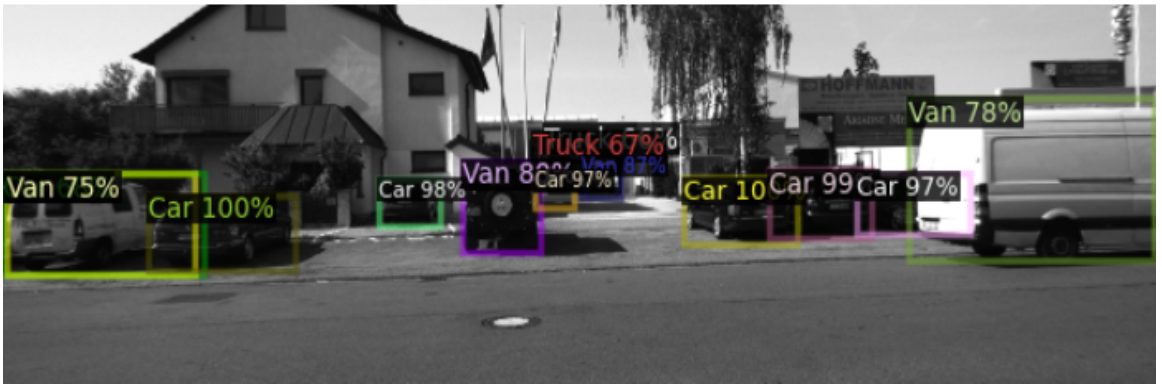

FIGURE 5.3: Example of model's detection on test dataset

Figure 5.3 illustrates the qualitative analysis of the results obtained in this study. The output of the model includes both classification scores and regression values for each detected object. The classification scores indicate the probability that the detected object belongs to each of the predefined classes, while the regression values are used to generate the bounding boxes around the objects. These bounding boxes are represented as rectangles drawn around the objects in the output images. The output images also include labels indicating the class of each detected object and the corresponding classification score.

Chapter 6

Conclusion and Future Scope

6.1 Conclusion

In conclusion, we have presented two object detection models, Faster R-CNN and YOLOv6, for vehicle detection on the KITTI dataset.

The Faster R-CNN model, with a ResNet-50 backbone and FPN feature extraction, achieved a mAP@0.75 of 0.62 on the test sets. Its performance varied across different object categories, with higher accuracy for cars and trucks compared to pedestrians and cyclists.

On the other hand, the YOLOv6 model exhibited superior performance in terms of overall detection accuracy. It achieved an impressive mAP of 0.91 at an IoU threshold of 0.50, showcasing its robustness in detecting objects of different sizes. The YOLOv6 model also demonstrated high recall rates (AR) across different image scales, indicating its ability to accurately localize objects within the images.

Both models have their strengths and limitations. The Faster R-CNN model excelled in detecting cars and trucks, making it suitable for applications that prioritize these object categories. In contrast, YOLOv6 exhibited superior performance across multiple object categories and image scales, making it a more versatile choice for object detection tasks.

In the future, further improvements can be made to both models by incorporating additional training data, optimizing hyperparameters, and exploring advanced techniques such as model ensembling or incorporating attention mechanisms. These advancements

can lead to more accurate and efficient vehicle detection systems, which are crucial for ensuring the safety and reliability of autonomous driving and other computer vision applications.

Overall, our work contributes to the advancement of object detection methodologies and serves as a foundation for future research in computer vision and autonomous driving systems.

6.2 Future Scope

The future work includes working on optimizing the model to achieve real-time object detection performance, which could have many practical applications.

We can explore the possibility of using the object detection model to track objects in a video stream, which could be useful in surveillance and security applications.

References

[1] Cai, Y., Luan, T., Gao, H., Wang, H., Chen, L., Li, Y., Sotelo, M.A. and Li, Z., 2021. YOLOv4-5D: An effective and efficient object detector for autonomous driving. IEEE Transactions on Instrumentation and Measurement, 70, pp.1-13.

[2] Hsu, S.C., Huang, C.L. and Chuang, C.H., 2018, January. Vehicle detection using simplified fast R-CNN. In 2018 International Workshop on Advanced Image Technology (IWAIT) (pp. 1-3). IEEE.

[3] Sommer, L., Schmidt, N., Schumann, A. and Beyerer, J., 2018, October. Search area reduction fast-RCNN for fast vehicle detection in large aerial imagery. In 2018 25th IEEE international conference on image processing (ICIP) (pp. 3054-3058). IEEE.

[4] Kim, H., Lee, Y., Yim, B., Park, E. and Kim, H., 2016, October. On-road object detection using deep neural network. In 2016 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia) (pp. 1-4). IEEE.

[5] Wang, R., Wang, Z., Xu, Z., Wang, C., Li, Q., Zhang, Y. and Li, H., 2021. A real-time object detector for autonomous vehicles based on YOLOv4. Computational Intelligence and Neuroscience, 2021.

[6] Cai, W., Li, J., Xie, Z., Zhao, T. and Kang, L.U., 2018, July. Street object detection based on faster R-CNN. In 2018 37th Chinese Control Conference (CCC) (pp. 9500-9503). IEEE.

[7] Wu, T.H., Wang, T.W. and Liu, Y.Q., 2021, June. Real-time vehicle and distance detection based on improved yolo v5 network. In 2021 3rd World Symposium on Artificial Intelligence (WSAI) (pp. 24-28). IEEE.

[8] Ojha, A., Sahu, S.P. and Dewangan, D.K., 2021, May. Vehicle detection through instance segmentation using mask R-CNN for intelligent vehicle system. In 2021 5th international conference on intelligent computing and control systems (ICICCS) (pp. 954-959). IEEE.

[9] J. Donahue, R. Girshick, T. Darrell and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," IEEE Int. Conf. on Computer Vision and Pattern Recognition, Columbus, 2014.

[10] R. Girshick, "Fast R-CNN," IEEE International Conference on Computer Vision, Santiago, 2015.

[11] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Int. Conf. on Computer Vision and Pattern Recognition, Las Vegas, 2016.

[12] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in neural information processing systems, pages 91–99, 2015.

[13] W. Sakla, G. Konjevod, and T. N. Mundhenk. Deep multi-modal vehicle detection in aerial isr imagery. In Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on, pages 916–923. IEEE, 2017.

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

[15] L. Sommer, T. Schuchert, and J. Beyerer. Deep learning based multi-category object detection in aerial images. In Automatic Target Recognition XXVII, volume 10202, page 1020209. International Society for Optics and Photonics, 2017.

[16] L. Sommer, T. Schuchert, and J. Beyerer. Fast deep vehicle detection in aerial images. In Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on,pages 311–319. IEEE, 2017.

[17] R. Girshick. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, pages 1440–1448, 2015.

[18] J. Dai, Y. Li, K. He, and J. Sun. R-fcn: Object detection via region-based fully convolutional networks. In Advances in neural information processing systems, pages 379–387, 2016.

[19] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. In European conference on computer vision, pages 21–37. Springer, 2016.

[20] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In CVPR, pages 779–788, 2016.

[21] J. Redmon and A. Farhadi. Yolo9000: better, faster, stronger. arXiv preprint arXiv:1612.08242, 2016.

References