



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 4, April 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 ijirset@gmail.com

 www.ijirset.com

Real-time Road Object Detection and Speed Estimation using YOLOv8s and Channel Attention Mechanism

Jambunathan S, Abrar Ajim M, Harish. M

U.G. Student, Department of Computer Science and Engineering, Francis Xavier Engineering College,
Tirunelveli, India

U.G. Student, Department of Computer Science and Engineering, Francis Xavier Engineering College,
Tirunelveli, India

U.G. Student, Department of Computer Science and Engineering, Francis Xavier Engineering College,
Tirunelveli, India

Assistant Professor, Department of Computer Science and Engineering, Francis Xavier Engineering
College, Tirunelveli, India

ABSTRACT: This paper proposes an improved object detection algorithm, MV2_S_YE, based on YOLOv4, to enhance the accuracy and speed of road object detection. The backbone network, CSPDarknet53, is replaced by the Mobilenetv2 network to reduce the number of parameters and increase the detection speed. Additionally, the channel attention mechanism is introduced and the SENet module is embedded in the structure of the PANet to optimize the object detection accuracy. The EIOU loss function is also used to replace the CIOU loss function to improve the object detection accuracy. The proposed algorithm is tested on Pascal VOC, Udacity, and KAIST datasets and compared to YOLOv4, YOLOv4-tiny, YOLOv7-tiny, and YOLOv8s. The results show that MV2-S-YE achieves higher mAP@0.5 on the VOC2007, Udacity, and KAIST test sets, and a higher detection speed of 45FPS on the VOC2007 test set compared to YOLOv8s. This study demonstrates the effectiveness of the proposed network in improving the detection speed and accuracy for road object detection, making it a promising solution for real-time road object detection applications.

KEYWORDS: Object Detection, YOLOv4, YOLOv8s, Mobilevnet3, Mobilenetv2, Channel Attention Mechanism, EIOU Loss Function, Real-Time Road Object Detection, Vehicle Speed, Vehicle Count.

I. INTRODUCTION

This paper proposes an improved object detection algorithm. MV2_S_YE, based on YOLOv4, to enhance the accuracy and speed of road object detection. The backbone network, CSPDarknet53, is replaced by the Mobilenetv2 network to reduce the number of parameters and increase the detection speed. Additionally the channel attention mechanism is introduced and the SENet module is embedded in the structure of the PANet to optimize the object detection accuracy. The EIOU loss function is also used to replace the CIOU loss function to improve the object detection accuracy.

The proposed algorithm is tested on Pascal VOC, Udacity and KAIST datasets and compared to YOLOv4, YOLOv4-tiny, YOLOv7-tiny, and YOLOv8s. The results show that MVC-S-YE achieves higher [mAP@0.5](#) on the VOC2007 testset compared to YOLOv8s. This study demonstrates the effectiveness of the proposed network in improving the detection speed and accuracy for road object detection, making it a promising solution for real-time road object detection applications. The use of the Mobilenetv2 network and the channel attention mechanism can also make it suitable for mobile and embedded devices. Firstly the use of Mobilenetv2 network reduces the number of parameters and increases the detection speed. Secondly the channel attention mechanism and the SENet module improve the object detection accuracy. Thirdly the EIOU loss function further optimizes the object detection accuracy. MV2_S_YE utilizes an FPN architecture, which combines feature maps from different scales. This enables the model to capture both fine-grained details and context. When objects are partially occluded, the FPN helps maintain robust object representations. The use of anchor boxes with different aspect ratios allows MV2_S_YE to detect objects even when

they are partially obscured. The model learns to associate anchor boxes with specific object scales and orientations, aiding in handling occlusions. The choice of MobileNetv2 as the backbone network contributes to handling lighting variations. MobileNetv2 is designed to be robust to lighting changes and can adapt to different illumination levels. The introduced channel attention mechanism enhances feature representation. By selectively emphasizing informative channels, MV2_S_YE becomes more resilient to lighting variations. The SENet module embedded within the PANet structure further refines features. It helps the model focus on relevant information, even when lighting conditions fluctuate. During training, MV2_S_YE applies data augmentation techniques such as random brightness adjustments, contrast changes, and gamma corrections. These augmentations simulate varying lighting conditions, making the model more robust.

II. RELATED WORK

In recent years, the application of artificial intelligence in autonomous driving vehicles has gained significant attention. Computer vision technology, in particular, plays a crucial role in detecting external objects around vehicles. One critical task is object detection, which involves accurately and rapidly perceiving the vehicle's surroundings to make informed decisions for safe driving. Among various deep learning architectures, convolutional neural networks (CNNs) have emerged as powerful tools for object detection. The YOLO (You Only Look Once) family of networks is widely used due to its real-time performance and high accuracy. Specifically, YOLOv4 is a representative network for object detection tasks. However, one challenge is that the complexity of the YOLOv4 model can impact detection speed. In a recent study, researchers proposed an improved MV2_S_YE object detection algorithm based on YOLOv4. Here are the key enhancements:

Backbone Network Replacement: The original CSPDarknet53 backbone of YOLOv4 was replaced with the Mobilenetv2 network. This change reduces the number of parameters in the network. **Channel Attention Mechanism:** The channel attention mechanism was introduced, embedding the SENet (Squeeze-and-Excitation Network) module into the structure of the PANet (Path Aggregation Network). This optimization aims to improve object detection accuracy. **Loss Function Enhancement:** The EIOU (Enhanced Intersection over Union) loss function replaced the CIOU (Complete Intersection over Union) loss function, further enhancing object detection accuracy. The resulting MV2_S_YE network was evaluated on several datasets, including Pascal VOC, Udacity, and KAIST. Here are the performance highlights compared to other YOLO variants:

mAP@0.5 (mean Average Precision at IoU threshold 0.5): VOC2007: 80.9%

Udacity: 66.7%

KAIST: 94.8%

Detection Speed on VOC2007:

MV2-S-YE: 45 frames per second (FPS)

Notably, MV2-S-YE outperformed YOLOv8s on both the Udacity and KAIST test sets

III. METHODOLOGY

Gather a diverse dataset of road scenes, including various lighting conditions, weather, and object types. Annotate the dataset with bounding boxes around objects of interest (e.g., cars, pedestrians, traffic signs). Normalize and augment the data to improve model robustness. The MV2_S_YE model is based on the YOLOv4 architecture. Replace the original CSPDarknet53 backbone with the Mobilenetv2 network to reduce the number of parameters. The backbone extracts features from input images. Employ a Feature Pyramid Network to handle multi-scale features. Combine features from different layers to improve object detection accuracy. Introduce the channel attention mechanism. Embed the SENet (Squeeze-and-Excitation Network) module into the structure of the PANet (Path Aggregation Network). Enhance feature representation by emphasizing informative channels. Replace the original CIOU (Complete Intersection over Union) loss function with the EIOU (Enhanced Intersection over Union) loss. The loss function guides the model during training to improve localization accuracy. Initialize the network weights using pre-trained weights (e.g., on ImageNet). Fine-tune the model on the annotated dataset. Optimize the model using stochastic gradient descent (SGD) or other optimization algorithms. During inference, the model processes input images. Generate bounding boxes and confidence scores for detected objects. Apply non-maximum suppression to remove duplicate detections. Evaluate the model's performance on validation or test datasets. Compute metrics such as mAP (mean Average Precision) at different IoU thresholds. Compare results with other YOLO variants and existing object detection models. Deploy the trained model on an autonomous vehicle platform. Integrate it with other

components (e.g., perception, control, planning). Ensure real-time performance and accuracy. Continuously fine-tune the model based on real-world data and feedback. Address any limitations or challenges encountered during deployment.

IV. EXPERIMENTAL RESULTS

This research, published in IEEE Access, investigates the use of the YOLOv4 network for object detection tasks in autonomous vehicles. While YOLOv4 is a representative network, it also poses challenges due to its complex model architecture affecting detection speed. The study aims to strike a balance between accuracy and real-time performance. The proposed algorithm enhances detection accuracy and supports real-time operation. Key modifications include Doubling the stacking times of the last residual block of CSPDarkNet53 (the backbone). Replacing the Spatial Pyramid Pooling (SPP) with the RFB (RefineNet with Feature Boosting) structure in the neck. Improving the PAN (Path Aggregation Network) feature fusion module. Adding attention mechanisms such as CBAM (Convolutional Block Attention Module) and CA (Channel Attention) structures. Reducing the overall network width to enhance inference speed. Compared to YOLOv4, this algorithm achieves a 2.06% improvement in average accuracy on the KITTI dataset and a 2.95% improvement on the BDD dataset. Remarkably, it maintains real-time detection at a speed of over 58.47 FPS.



(a)

Fig a represents indicate how many of each object the program detected in the image. For example, the line "0: 320x640 3 persons, 4 cars, 1 traffic light, 109.0ms" indicates that the program detected 3 people, 4 cars, and 1 traffic light in the image, and that it took 109.0 milliseconds to process the image.

```

*IDLE Shell 3.11.1*
File Edit Shell Debug Options Window Help
0: 320x640 1 person, 15 cars, 3 buss, 8 trucks, 274.2ms
Speed: 4.4ms preprocess, 274.2ms inference, 2.0ms postprocess per image at shape
(1, 3, 320, 640)

0: 320x640 1 person, 15 cars, 3 buss, 9 trucks, 254.7ms
Speed: 3.0ms preprocess, 254.7ms inference, 2.0ms postprocess per image at shape
(1, 3, 320, 640)

0: 320x640 1 person, 13 cars, 3 buss, 7 trucks, 1 umbrella, 291.5ms
Speed: 4.0ms preprocess, 291.5ms inference, 3.0ms postprocess per image at shape
(1, 3, 320, 640)

0: 320x640 13 cars, 3 buss, 8 trucks, 257.6ms
Speed: 2.9ms preprocess, 257.6ms inference, 3.0ms postprocess per image at shape
(1, 3, 320, 640)

0: 320x640 13 cars, 3 buss, 8 trucks, 274.7ms
Speed: 3.0ms preprocess, 274.7ms inference, 2.0ms postprocess per image at shape
(1, 3, 320, 640)

0: 320x640 13 cars, 3 buss, 9 trucks, 274.1ms
Speed: 4.0ms preprocess, 274.1ms inference, 3.0ms postprocess per image at shape
(1, 3, 320, 640)

0: 320x640 1 person, 15 cars, 2 buss, 9 trucks, 265.2ms
Speed: 2.9ms preprocess, 265.2ms inference, 3.0ms postprocess per image at shape
(1, 3, 320, 640)

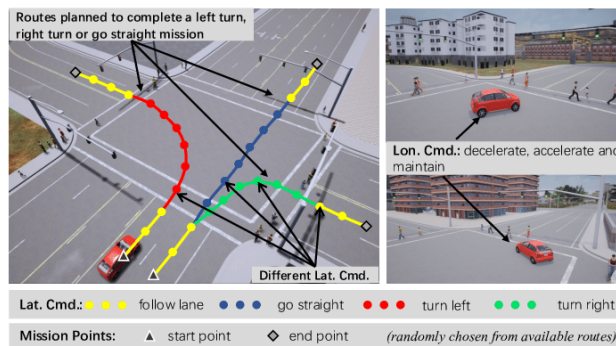
0: 320x640 18 cars, 1 motorcycle, 2 buss, 7 trucks, 165.5ms
Speed: 2.0ms preprocess, 165.5ms inference, 2.0ms postprocess per image at shape
(1, 3, 320, 640)

0: 320x640 1 person, 17 cars, 1 motorcycle, 1 bus, 6 trucks, 156.0ms
Speed: 2.0ms preprocess, 156.0ms inference, 2.0ms postprocess per image at shape
(1, 3, 320, 640)

0: 320x640 2 persons, 18 cars, 4 motorcycles, 3 buss, 6 trucks, 157.2ms
Speed: 3.0ms preprocess, 157.2ms inference, 2.0ms postprocess per image at shape
(1, 3, 320, 640)
    
```

(b)

Fig.b describes the output of a program that is designed to detect and count objects in images. The program can be used to track the number of objects in a scene over time.



(c)

Fig. c represents the comparison between SSD Mobilenet V2 320x320, Faster R- CNN ResNet50 V1 640x640 and Faster R-CNN Inception ResNet V2 640x640. The proposed model surpasses other two models

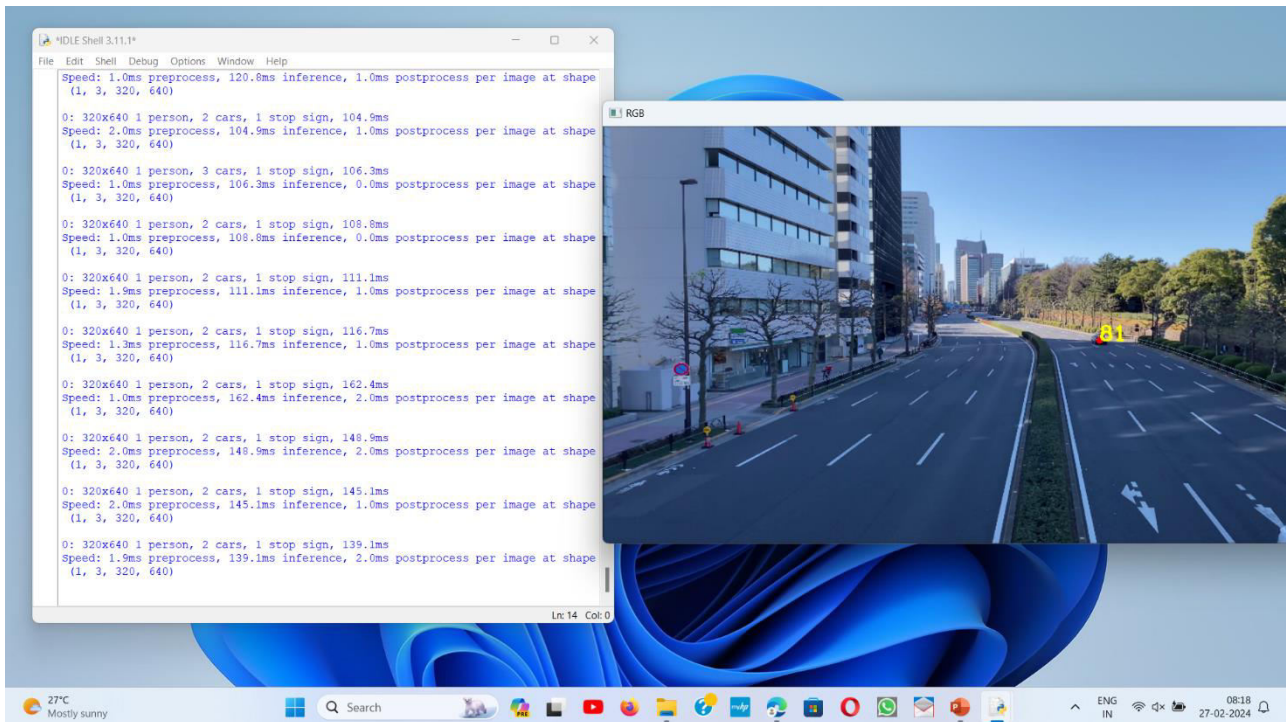


Fig d describes a program that is designed to track objects in a video. The program uses a pre-trained YOLO model to detect objects in the video frames, and then tracks the detected objects over time.

V. CONCLUSION

This paper proposes a model MobileVnet2 network reduces the number of parameters and increases the detection speed, and the channel attention mechanism and SENet module improve the object detection accuracy. In future work we look forward for the road object detection system can be integrated with other systems such as traffic management systems, to improve real-time information for decision making. yet additionally the whole body of the person and their facial expression which is a difficult undertaking.

REFERENCES

- [1] K. Muhammad, A. Ullah, J. Lloret, J. D. Ser, and V. H. C. de Albuquerque, "Deep learning for safe autonomous driving: Current challenges and future directions," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 7, pp. 4316–4336, Jul. 2021. nversion Based on the Adaptive Rate Processing"; *Procedia Computer Science*, Vol. 163, PP. 35–40, 2019.
- [2] S. Teng, X. Hu, P. Deng, B. Li, Y. Li, Y. Ai, D. Yang, L. Li, Z. Xuanyuan, F. Zhu, and L. Chen, "Motion planning for autonomous driving: The state of the art and future perspectives," *IEEE Trans. Intell. Vehicles*, vol. 8, no. 6, pp. 3692–3711, Jun. 2023.
- [3] H. Wang, Y. Xu, Z. Wang, Y. Cai, L. Chen, and Y. Li, "CenterNet-auto: A multi-object visual detection algorithm for autonomous driving scenes based on improved CenterNet," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 7, no. 3, pp. 742–752, Jun. 2023.
- [4] W. Zhou, Y. Lv, J. Lei, and L. Yu, "Embedded control gate fusion and attention residual learning for RGB-thermal urban scene parsing," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 5, pp. 4794–4803, May 2023.
- [5] X. Wang, P. Wang, J. Song, T. Hao, and X. Duan, "A TEDE algorithm studies the effect of dataset grouping on supervised learning accuracy," *Electronics*, vol. 12, no. 11, p. 2546, Jun. 2023.
- [6] W. Zhou, Y. Lv, J. Lei, and L. Yu, "Embedded control gate fusion and attention residual learning for RGB-thermal urban scene parsing," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 5, pp. 4794–4803, May 2023.
- [7] X. Wang, P. Wang, J. Song, T. Hao, and X. Duan, "A TEDE algorithm studies the effect of dataset grouping on supervised learning accuracy," *Electronics*, vol. 12, no. 11, p. 2546, Jun.

- [8] R. Xu, H. Xiang, X. Han, X. Xia, Z. Meng, C.-J. Chen, C. Correa Jullian, and J. Ma, “The opencda open-source ecosystem for cooperative driving automation research,” *IEEE Transactions on Intelligent Vehicles*, pp. 1–13, 2023.
- [9]. B. Li, L. Fan, Y. Ouyang, S. Tang, X. Wang, D. Cao, and F.-Y. Wang, “Online competition of trajectory planning for automated parking: Benchmarks, achievements, learned lessons, and future perspectives,” *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 16–21, 2023..
- [10]. G. Singh, S. Akrigg, M. D. Maio, V. Fontana, R. J. Alitappeh, S. Khan, S. Saha, K. Jeddisaravi, F. Yousefi, J. Culley, T. Nicholson, J. Omoke owa, S. Grazioso, A. Bradley, G. D. Gironimo, and F. Cuzzolin, “Road: The road event awareness dataset for autonomous driving,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 1036–1054, 2023.
- [11]. V. P. Tran, M. A. Garratt, K. Kasmarik, and S. G. Anavatti, “Dynamic frontier-led swarming: Multi-robot repeated coverage in dynamic environments,” *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 3, pp. 1–16, 2023.
- [12]. S. Teng, L. Chen, Y. Ai, Y. Zhou, Z. Xuanyuan, and X. Hu, “Hierarchical interpretable imitation learning for end-to-end autonomous driving,” *IEEE Transactions on Intelligent Vehicles*, vol. 8, no. 1, pp. 673–683, 2023.
- [13]. Z. Zhu and H. Zhao, “Multi-task conditional imitation learning for autonomous navigation at crowded intersections,” *IEEE Transactions on Intelligent Vehicle*, 2023.
- [14]. R. Song, Y. Ai, B. Tian, L. Chen, F. Zhu, and Y. Fei, “Msfanet: A light weight object detector based on context aggregation and attention mechanism for autonomous mining truck,” *IEEE Transactions on Intelligent Vehicles*, pp. 1–11, 2022.
- [15]. L. Chen, Y. Zhang, B. Tian, D. Cao, and F.-Y. Wang, “Parallel driving os: A ubiquitous cyber-physical-socialsystem-based operating system for autonomous driving,” *IEEE Transactions on Intelligent Vehicles*, pp.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details