Enhancing Precision Object Detection and Identification for Autonomous Vehicles through YOLOv5 Refinement with YOLO-ALPHA

Guandong Li[1], Yanzhe Xie[1], Yuhao Lu[1], Zongyan Wen[1], Jingzhen Fan[1], Yuankui Huang[1], Qinghong Ma[1], Wei Hong Lim[2], Chin Hong Wong[1*]

[1]Maynooth International Engineering College, Fuzhou University, Fujian, China
[2]Faculty of Engineering, Technology and Built Environment, UCSI University, 1, Jalan Puncak Menara Gading, UCSI Heights, 56000 Cheras, Kuala Lumpur, Malaysia.
E-mail: guandong.li.2022@mumail.ie, yanzhe.xie.2022@mumail.ie, yuhao.lu.2022@mumail.ie, zongyan.wen.2022@mumail.ie, jingzhen.fan.2022@mumail.ie, yuankui.huang.2022@mumail.ie, qinghong.2021@mumail.ie,, limwh@ucsiuniversity.edu.my, chinhong.wong@mu.ie

Abstract

Advancing swiftly in contemporary society, the rapid growth of autonomous driving technology suggests its potential adoption across continents. The realization of fully autonomous driving relies on proficiently detecting, classifying, and tracking road objects such as pedestrians and vehicles. This research employs the YOLOv5 neural network, enhancing it with YOLO-ALPHA. Modifications, encompassing freeze and attention mechanisms, serve to refine accuracy and expedite training. Furthermore, adjustments to the activation function aim to stabilize precision and recall. The integration of an FCN based on semantic segmentation theory contributes to improved accuracy in detecting road conditions during autonomous driving. Consequently, this enables the successful and highly accurate functionality of automatic identification.

Keywords: Object detection, Autonomous vehicle, YOLOv5, FCN, Attention mechanism, Freeze mechanism, Activation function

1. Introduction

In recent years, the rapid advancement of technology has led to a surge in the popularity of autonomous driving systems. Real-time image processing is a pivotal technology in autonomous driving. Despite reaching an advanced stage, autonomous driving technology faces persistent challenges in low recognition accuracy and sluggish real-time image processing, giving rise to safety concerns. These issues give rise to safety concerns. Although automated driving technology has reached a certain level of maturity, it encounters limitations in adverse weather conditions, complex environments, and nighttime operations. automatic driving sensor recognition accuracy tends to decline, compromising autonomous vehicle safety. Consequently, there is a pressing need to enhance automatic driving technology to bolster autonomous vehicle reliability and safety.

This paper introduces an attention mechanism into the object detection algorithm, fine-tunes activation functions, adds a freezing mechanism and conducts a comparative analysis with the FCN (Fully Convolutional Networks) algorithm. This algorithm offers an enhanced detection speed and accuracy. The implications of this research are highly relevant to the enhancement of autonomous driving safety.

Utilizing the PASCAL VOC (The PASCAL Visual Object Classes) dataset, this study systematically alters various modules and introduces distinct mechanisms, this paper changes a different module and adds a disparate mechanism each time. Then, by evaluating precision, recall, mAP, and mAP0.5 indices of the adapted models, it identifies the optimal placement for incorporating an attention mechanism, freezing layers to prolong training duration, and selecting the activation function that demonstrated superior performance on the PASCAL VOC dataset.

This paper presents a novel detection technique named YOLO-ALPHA. The approach involves incorporating an attention mechanism at various layers within the YOLO network. Specifically, inserting the Squeeze-and-Excitation (SE) attention mechanism (Self-Attention Mechanism) after the Concat module on the 16th layer of the network could enhance both the precision and recall of the model. Additionally, adding a freeze mechanism in the YOLO backbone could accelerate curve fitting. Moreover, replacing the combined structure of SiLU (Sigmoid-weighted Linear Units) and LeakyReLU with the SiLU activation function contributes to a smoother precision and recall curve, thereby, improving the stability of the model by 0.169%.

2. Literature Review

Convolutional neural networks (CNNs) often suffer from information loss, resulting in gradient disappearance and detection inaccuracies during image processing. Researchers continuously strive to mitigate semantic information loss and enhance overall network performance. Lan et al. [1] conducted research to improve the YOLO network, focusing on addressing losses in pedestrian information during image processing. They introduced additional passthrough layers, including a Route layer and Reorg layer, or reconstructed the primitive YOLOv2 network. The resulting YOLO-R network specializes in pedestrian detection, effectively improving accuracy and reducing false detection rates. Wu et al. [2] presented a vehicle detection system in CARLA, introducing YOLOv5-Ghost with adjusted layer structures for reduced computational complexity. The study achieved improved detection accuracy and speed. Lu et al. [3] enhanced a YOLOv5-based model for crack and vehicle detection datasets, addressing sample imbalance and small object presence. Dodia and Kumar [4] compared modern object detectors with YOLO for vehicle detection, suggesting enhancements to improve vehicle prediction capacity. Kaloev et al. [5] investigated activation functions in deep learning, emphasizing their role in introducing nonlinearity. Xiao et al. [6] proposed intelligent layer freezing during training to accelerate training on various networks. Zhao et al. [7] improved an infrared detection model based on YOLOv5s, adding an SE-Net module. Kaymak et al. [8] experimented with FCN architectures in the context of autonomous driving. The study concludes that applying YOLO to vehicle detection faces challenges, leading researchers to improve the algorithm through various modifications.

3. Design Methodology

Fig. 1 shows the flow chart to tackle imminent challenges related to object identification, classification, and tracking in the domain of autonomous driving. Given the effectiveness of YOLOv5 in object identification tasks, as evidenced by Xiao [9] and Kaymak [8], and recognizing the utility of FCN in autonomous driving scenarios, this research chooses to employ both methods for real-time image processing in autonomous driving contexts. To ensure a fair comparison of results across different networks using the control variate method, a consistent dataset, PASCAL VOC 2012, was used to train and validate both YOLO and FCN networks. PASCAL VOC 2012 encompasses 20 classes commonly encountered in autonomous driving scenarios, such as bicycle, bus, car, motorbike, and person.

Following multiple rounds of training and prediction in YOLO, this work collects and scrutinizes detection and classification outcomes using TensorBoard, relying on index curves such as precision, recall, mAP, and mAP0.5 curvature. Higher precision values denote the network's proficiency in accurately classifying items, while elevated recall signifies the network's efficacy in rectifying previous classification errors. The mAP and mAP0.5 metrics evaluate the detection accuracy across all classes present in the dataset. A lower curvature in the curve indicates enhanced stability and a more consistently changing trend, signifying ideal model training stability and convergence.

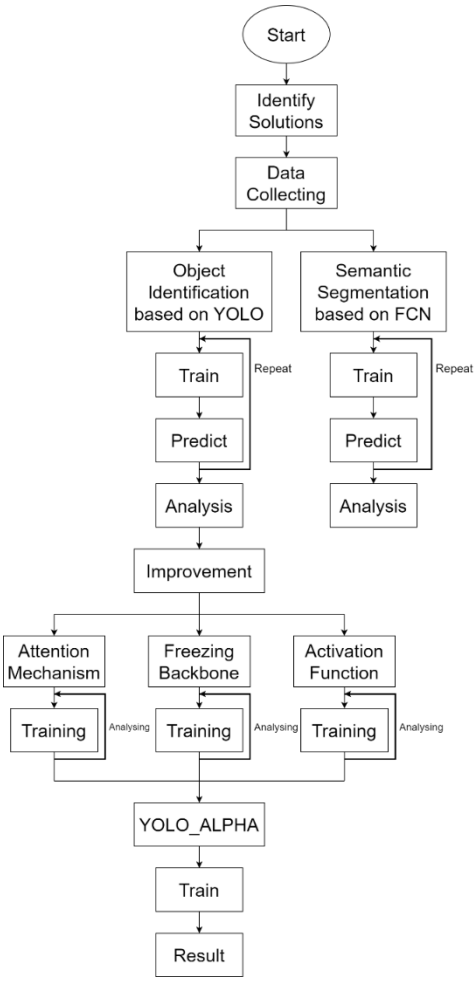

Fig. 1 Flow chart

Then, the project aims to enhance the YOLO algorithm by adding an attention mechanism, freezing backbone, and activation function.

Incorporating an attention mechanism enables the network to focus on crucial input features while disregarding inconsequential ones. Through the integration of SE attention mechanisms in different layers of the YOLO structure and subsequent training, the research evaluates the performance enhancement trend by scrutinizing index curves and determining the optimal location for implementing the SE attention mechanism.

Freezing the backbone is a training strategy aimed at expediting training speed by halting parameters in the backbone upgrade. By comparing the result indexes of modules employing freezing backbone, freezing all layers, and not freezing layers separately, the project seeks to identify the optimal freezing method.

Activation functions amplify the nonlinearity of convolutional neural networks. By replacing the original activation function with SiLU, Mish, LeakyReLU, and ReLU in the modified YOLOv5s, the project utilizes the

average curvature of the precision curve to compare and assess the smoothness and stability of the network.

Ultimately, the study amalgamates the proven optimal strategies from the preceding analysis to formulate YOLO-ALPHA, an enhanced version of YOLOv5. The performance of YOLO-ALPHA is then juxtaposed with that of YOLOv5 to affirm its improvement.

4. Results and Discussions

The object detection output images from YOLOv5s depict vehicles on the road using bounding boxes. However, these bounding boxes lack information regarding the distance between vehicles, a crucial consideration in congested road scenarios, particularly during peak commuting hours, a vital aspect of autonomous driving. To precisely discern the outlines of nearby vehicles and pedestrians, pixel-wise object detection and classification with FCN are employed. Fig. 2 illustrates the contrast in detection results between YOLOv5s and FCN. The depiction from YOLOv5s shows vehicles with pink bounding boxes, outlining the edges of cars as straight lines that do not accurately match the actual boundaries. On the other hand, FCN, with pixel-wise detection, unveils the precise boundaries of cars, enabling autonomous vehicles to effectively avoid collisions.

Fig. 2 Detection result between YOLOv5s and FCN

Fig. 3 shows minimal differences observed between the frozen backbone curve and the original model curve in terms of mAP, precision, and recall at the highest training rounds. Despite a reduction in training time by 2 hours, image accuracy slightly decreases from 0.7166 to 0.6716. The key conclusion drawn is that freezing trained parameters enhances training speed without compromising accuracy. Freezing all layers completely leads to virtually no accuracy, emphasizing the importance of trained parameters. Freezing the backbone during training enhances training efficiency, as the unchanging feature extraction network consumes less memory, leading to faster training.

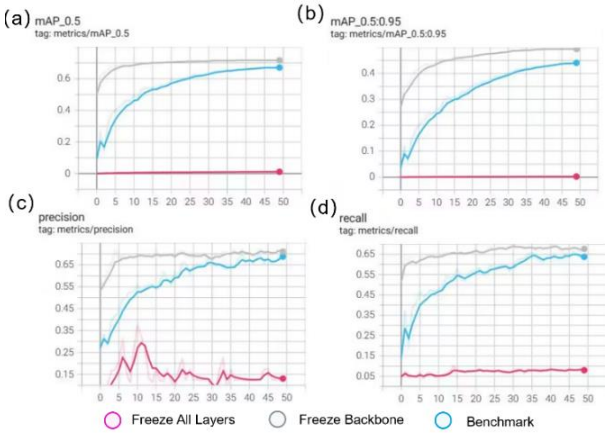

Fig. 2 Freezing Mechanism graphs for (a) mAP_0.5, (b) mAP_0.5:0.95, (c) Precision and (d) Recall

Experimental results suggested that adding the SE attention mechanism to the bottom or middle layers of the network yields optimal outcomes. Fig. 4 illustrates that adding the attention mechanism between layers 16 and 17 achieves higher accuracy than the original version, validating the hypothesis. Fig. 5 further confirms that placing the SE attention mechanism between the middle and bottom layers produces the best results.

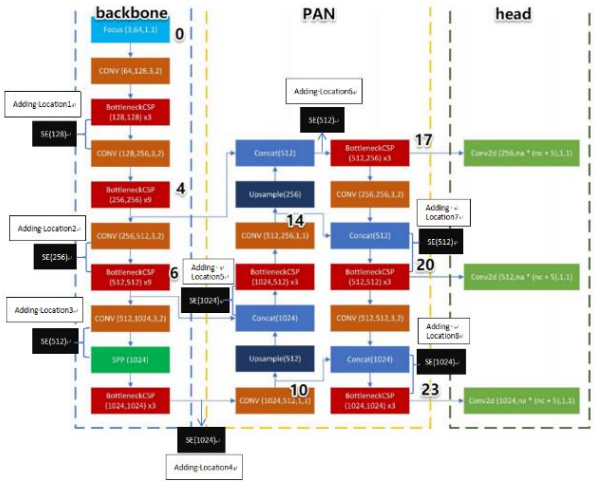

Fig. 3 SE adding locations in the original YOLOv5 network

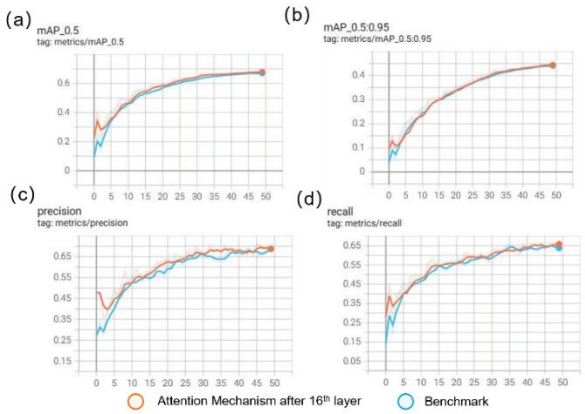

Fig. 4 Attention Mechanism graphs for (a) mAP_0.5, (b) mAP_0.5:0.95, (c) Precision and (d) Recall

Five activation functions (ReLU, Leaky ReLU, Mish, SiLU, and the original activation function) undergo comparative analysis. Fig. 6 displays the precision, recall, mAP, and mAP0.5 indexes at epoch 50 for networks employing various activation functions. The SiLU activation function is selected for the PASCAL VOC dataset due to its smoother curve and increased stability, as tabulated in Table 1.

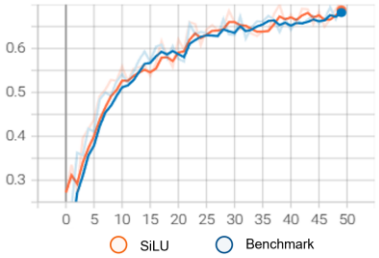

Fig. 5 Activation Function

Table 1 Activation Function Data

Epoch = 50	mAP	mAP0.5	Precision	Recall
benchmark	0.6716	0.4426	0.7064	0.6286
LeakyReLU	0.6437	0.4114	0.6531	0.6238
Mish	0.6649	0.4397	0.6783	0.6363
ReLU	0.6353	0.4058	0.7007	0.5739
SiLU	0.6694	0.4426	0.6961	0.6270

Fig. 7 depicts a less satisfactory outcome when combining the optimal freezing mechanism, attention mechanism, and activation function. The precision initially starts high but continuously decreases, indicating performance degradation. Removing the attention mechanism results in Fig. 8 that surpass those of the original YOLOv5 network, affirming the performance enhancement of the upgraded YOLO-ALPHA network.

5. Conclusion

In this investigation, YOLOv5s is employed to carry out object detection, classification, and tracking, with the FCN (Semantic Segmentation) technique serving as a supplementary method to precisely delineate object boundaries. Enhancements, including the incorporation of attention mechanisms, freezing the network's backbone, and modifying the activation function, are evaluated based on precision, recall, mAP, and mAP0.5 curves. These identified mechanisms and modules, known to augment the performance of YOLOv5s, are amalgamated into an upgraded version referred to as YOLO-ALPHA. However, the collective optimal modules identified in previous research resulted in a degradation of performance. Following meticulous analysis, the freezing backbone mechanism is pinpointed as the cause. Subsequent refinement by removing the freezing backbone mechanism reveals that YOLO-ALPHA showcases performance improvements in comparison to the original YOLOv5s.

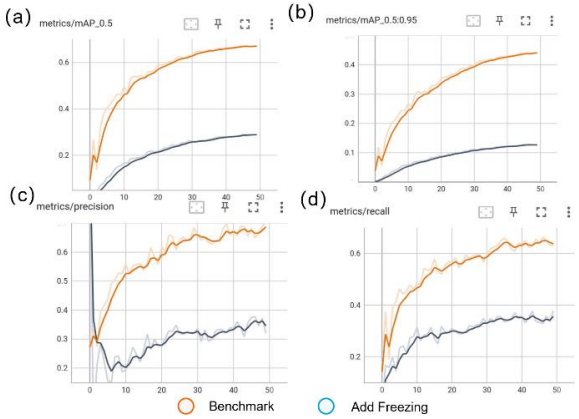

Fig. 6 Unsuccessful Result for (a) mAP_0.5, (b) mAP_0.5:0.95, (c) Precision and (d) Recall

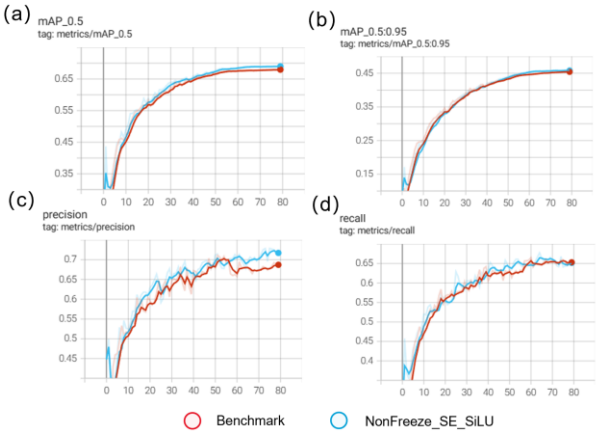

Fig. 7 Result of Applying SE and SiLU excluding Freezing Mechanism graph for (a) mAP_0.5, (b) mAP_0.5:0.95, (c) Precision and (d) Recall

References

1. W. Lan, J. Dang, Y. Wang, and S. Wang, "Pedestrian detection based on yolo network model," in 2018 IEEE International Conference on Mechatronics and Automation (ICMA), 2018, pp. 1547–1551.
2. T.-H. Wu, T.-W. Wang, and Y.-Q. Liu, "Real-time vehicle and distance detection based on improved

yolo v5 network," in 2021 3rd World Symposium on Artificial Intelligence (WSAI), 2021, pp. 24–28.
3. Z. Lu, L. Ding, Z. Wang, L. Dong, and Z. Guo, "Road condition detection based on deep learning yolov5 network," in 2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI), 2023, pp. 497–501.
4. A. Dodia and S. Kumar, "A comparison of yolo based vehicle detection algorithms," in 2023 International Conference on Artificial Intelligence and Applications (ICAIA) Alliance Technology Conference (ATCON-1), 2023, pp. 1–6.
5. M. Kaloev and G. Krastev, "Comparative analysis of activation functions used in the hidden layers of deep neural networks," in 2021 3rd Inter- national Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), 2021, pp. 1–5.
6. X. Xiao, T. Bamunu Mudiyanselage, C. Ji, J. Hu, and Y. Pan, "Fast deep learning training through intelligently freezing layers," in 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), 2019, pp. 1225–1232.
7. H. Zhao, Z. Liang, D. Cai, and Y. Wang, "An improved method for infrared vehicle and pedestrian detection based on yolov5s," in 2022 International Conference on Machine Learning, Cloud Computing and Intelligent Mining (MLCCIM), 2022, pp. 377–383.
8. Kaymak and A. Uc¸ar, "Semantic image segmentation for autonomous driving using fully convolutional networks," in 2019 International Ar- tificial Intelligence and Data Processing Symposium (IDAP), 2019, pp. 1–8.
9. B. Xiao, J. Guo, and Z. He, "Real-time object detection algorithm of autonomous vehicles based on improved yolov5s," in 2021 5th CAA International Conference on Vehicular Control and Intelligence (CVCI), 2021, pp. 1–6.

Authors Introduction

Mr. Guandong Li

He is currently pursuing Bachelor of Robotics and Intelligent Devices as 3rd year student in the Department of Maynooth International Engineering College, Fuzhou University, China. His research interests are embedded systems and machine learning.

Mr. Yanzhe Xie

He is currently pursuing Bachelor of Robotics and Intelligent Devices as 3rd year student in the Department of Maynooth International Engineering College, Fuzhou University, China. His research interest is robotics and machine learning

Mr. Yuhao Lu

He is currently pursuing Bachelor of Robotics and Intelligent Devices as 3rd year student in the Department of Maynooth International Engineering College, Fuzhou University, China. His research interests are data science and embedded system.

Mr. Zongyan Wen

He is currently pursuing Bachelor of Robotics and Intelligent Devices as 3rd year student in the Department of Maynooth International Engineering College, Fuzhou University, China. His research interests are embedded systems and machine learning.

Mr. Jingzhen Fan

He is currently pursuing Bachelor of Robotics and Intelligent Devices as 3rd year student in the Department of Maynooth International Engineering College, Fuzhou University China. His research interests are software development and system learning.

Mr. Yuankui Huang

He is currently pursuing Bachelor of Robotics and Intelligent Devices as 3rd year student in the Department of Maynooth International Engineering College, Fuzhou University, China. His research interests are electronic information and machine learning.

Mr. Qinghong Ma

He is currently pursuing Bachelor of Robotics and Intelligent Devices as 3rd year student in the Department of Maynooth International Engineering College, Fuzhou University, China. His research interests are electronic information and machine learning.

Dr. Wei Hong Lim

He is an Associate Professor in Faculty of Engineering at UCSI University in Malaysia. He received his PhD in Computational Intelligence from Universiti Sains Malaysia in 2014. His research interests are optimization and artificial intelligence.

Dr. Chin Hong Wong

He is a Lecturer at Maynooth International Engineering College at Fuzhou University, China. He received his PhD in Electrical and Electronic Engineering from Universiti Sains Malaysia in 2017. His research interests are Energy harvesting, signals and systems, and machine learning.