Advanced Night time Object Detection in Driver-Assistance Systems using Thermal Vision and YOLOv5

Hoang-Tu Vo, Luyl-Da Quach
Software Engineering Department
FPT University, Cantho city, Vietnam

***Abstract*—Driver-assistance systems have become an indispensable component of modern vehicles, serving as a crucial element in enhancing safety for both drivers and passengers. Among the fundamental aspects of these systems, object detection stands out, posing significant challenges in low-light scenarios, particularly during nighttime. In this research paper, we propose an innovative and advanced approach for detecting objects during nighttime in driver-assistance systems. Our proposed method leverages thermal vision and incorporates You Only Look Once version 5 (YOLOv5), which demonstrates promising results. The primary objective of this study is to comprehensively evaluate the performance of our model, which utilizes a combination of stochastic gradient descent (SGD) and Adam optimizer. Moreover, we explore the impact of different activation functions, including SiLU, ReLU, Tanh, LeakyReLU, and Hardswish, on the efficiency of nighttime object detection within a driver assistance system that utilizes thermal imaging. To assess the effectiveness of our model, we employ standard evaluation metrics including precision, recall, and mean average precision (mAP), commonly used in object detection systems.**

***Index Terms*—Driver-assistance systems; object detection; nighttime object detection; thermal vision; YOLOv5**

I. Introduction

The rise of self-driving cars represents a significant milestone in the automotive industry, promising a paradigm shift in transportation as we know it. With the introduction of autonomous vehicles, there is a pressing need to address the alarming number of fatalities that occur in traffic accidents each year. Road traffic injuries pose a significant threat to the lives of children and young adults aged 5-29 years, making it the leading cause of death within this age group. It is worth noting that a staggering 93% of these fatalities occur in low- and middle-income countries [1]. These tragic incidents have prompted researchers and engineers to explore innovative solutions to improve road safety using machine learning (ML) and deep learning (DL) algorithms [2], [3], [4], [5]. The safety of autonomous vehicles relies on the ability to detect and classify objects correctly. Object detection algorithms need to be robust enough to differentiate between pedestrians, bicycles, cars, and other relevant entities on the road. This distinction is crucial for autonomous vehicles to assess potential risks and determine appropriate responses, such as slowing down, changing lanes, or stopping altogether. Detecting objects on the road is a crucial task for autonomous vehicles to ensure the safety of both passengers and other road users. However, the challenge becomes even more pronounced when it comes to detecting objects at night or in low light conditions. Reduced visibility conditions make it difficult for sensors, such as cameras to capture clear and detailed information about the surrounding environment. Traditional object detection systems heavily rely on visual cues, which can be compromised in low light conditions. This poses a considerable challenge for autonomous vehicles navigating roads at night or in poorly lit environments. To address these challenges [6], [7], [8], researchers have turned to Convolutional Neural Networks (CNNs), a powerful deep learning technique that has revolutionized various fields, including computer vision. CNNs have shown great promise for object detection, providing a robust framework for training models that can learn and extract meaningful features from image data. By incorporating CNN modeling in Driver-Assistance Systems, autonomous vehicles can navigate complex environments more effectively, reducing the risk of accidents and ultimately saving lives.

The objective of this article is to provide an innovative strategy for nighttime object detection in driver-assistance systems using thermal vision and incorporating the YOLOv5 model. The primary objective is to comprehensively evaluate the model's performance by investigating the influence of different activation functions and optimizers. The findings demonstrate the efficiency of the proposed method in enhancing nighttime object detection. The results contribute to the understanding of the role of optimizers and activation functions in training the YOLOv5 model for object detection tasks. The insights gained from this research can guide future endeavors aimed at improving the efficiency and accuracy of driver-assistance systems, ultimately enhancing safety for both drivers and passengers.

The structure of this paper is outlined in the following manner: Section II provides an extensive review of the relevant literature. In Section III, we elaborate on the methodology utilized for Advanced Nighttime Object Detection, covering aspects such as Dataset and Data Preparation, Data annotations/labeling, Activation Functions, and Model Evaluation Metrics. The experimental system and results, accompanied by a comprehensive discussion, are presented in Section IV. Finally, Section V provides concluding remarks to wrap up the paper.

II. Related Works

Ramesh Simhambhatla et al. (2019) [9] undertook a practical examination of three up-to-date meta-architectures, namely SSD, R-CNN, and R-FCN. The aim was to gauge their efficiency and precision in recognizing road objects, including vehicles, pedestrians, and traffic lights, across varying driving scenarios: daytime, nighttime, rainy, and snowy conditions. This research paper was carried out by Ruturaj Kulkarni et al. (2018) [10] introduces a robust deep neural network model that employs transfer learning for the accurate detection and recognition of traffic lights. To facilitate object detection in self-driving cars using deep learning, P Prajwal et al. (2021) [11] have selected the SSD model in conjunction with the MobileNet neural network as the foundational architecture due to its ability to produce results rapidly while maintaining a moderate level of accuracy. VD Nguyen et al. (2018) [12] presents a comprehensive framework that combines deep learning techniques, multiple local patterns, and depth information to identify, classify, and monitor vehicles and walkers on the road. Utilizing a deep CNN, H Yu et al. (2013) [13] employ a sophisticated architecture that effectively detects obstacles in complex scenes by leveraging rich and powerful learned features. P Salavati et al. (2018) [14] presents a novel approach that utilizes Deep Neural Networks (DNN) to detect obstacles using a single camera, employing unsupervised DNNs for extracting global image features and extracting local image features. P Aswathy et al. (2018) [15] explores the influence of deep convolutional layer features within an object tracking framework, showcasing the novel utilization of GoogLeNet CNN architecture's deep layer features for effective object tracking. The primary emphasis of this paper [16] is on the application of a CNN algorithm for computer vision-based object detection. The paper [17] presents a novel real-time approach for object detection in images captured by self-driving vehicles, using a unified neural network that models object detection as a regression problem on predicted bounding boxes and class probabilities, enabling simultaneous prediction of bounding boxes and class probabilities for the entire image. AA Cervera-Uribe et al. (2022) [18] introduces U19-Net, a deep encoder-decoder model designed for the detection of vehicles and pedestrians. This paper [19] introduces a novel and efficient deep learning-based detecting technique called DW-YOLO, which addresses the challenge of detecting objects in images with limited visual cues. G Rjoub et at. (2021) [20] presents a novel object detection system for autonomous vehicles, utilizing the You Only Look Once (YOLO) convolutional neural network (CNN) approach and a Federated Learning (FL) framework to enhance real-time detection accuracy, particularly in challenging weather conditions. This paper [21] demonstrates the utilization of the YOLOv5 model for real-time identification of cars, traffic lights, and pedestrians under different weather conditions, showcasing its effectiveness in typical vehicular environments. The purpose of the paper [22] is to develop a DL model, trained on the YOLOv5s and YOLOv7 architectures, to correctly classify and identify traffic signs in diverse adverse environments. VD Nguyen el at. (2023) [23] introduces an effective feature-based approach that utilizes a sigmoid function based on a triangle pattern to encode and establish strong features of neighboring pixels in local regions, which is then integrated into advanced object detection methods to evaluate its performance.

The purpose of this article is to present an innovative and advanced approach for nighttime object detection in driver-assistance systems. The study focuses on leveraging thermal vision and incorporating YOLOv5 as the proposed method. The primary objective is to comprehensively evaluate the performance of the model, which combines SGD and Adam optimizer. Additionally, the research investigates the impact of different activation functions, such as SiLU, ReLU, Tanh, LeakyReLU, and Hardswish, on the efficiency of nighttime object detection using thermal imaging within a driver assistance system. Standard evaluation metrics, including precision, recall, and mean average precision (mAP), are employed to assess the effectiveness of the model.

III. Methodology

A. Dataset and Data Preparation

The FLIR Thermal Images Dataset consists of a collection of 10,228 thermal images, each hand-labeled with precise bounding boxes. The images have a resolution of 640x512 pixels. Within the dataset, there are a total of 10,228 images, and these images contain a comprehensive set of 79,297 annotated bounding boxes. The dataset focuses on three main categories, namely Person, Bicycle, and Car. In the training set, which includes 8,862 images, there are 67,618 hand-labeled bounding boxes. Specifically, the Person category has 22,372 annotated bounding boxes, the Bicycle category has 3,986 annotated bounding boxes, and the Car category has 41,260 annotated bounding boxes. In the validation set, which consists of 1,366 images, there are 11,679 hand-labeled bounding boxes. The Person category has 5,778 annotated bounding boxes, the Bicycle category has 470 annotated bounding boxes, and the Car category has 5,431 annotated bounding boxes. Details of the distribution of the data set can be seen in Fig. 1 to Fig. 4.

B. Data Annotations/Labeling

Annotation of your training images To ensure the effective training of our object detector, it is imperative to provide supervision during the training process by employing bounding box annotations. The procedure entails outlining a box around each specific object that we intend the detector to detect, and subsequently assigning a corresponding object class label to each box, indicating the desired prediction for the detector. This crucial step allows us to train the object detector accurately. Additional details can be incorporated to provide a comprehensive understanding of the topic. The YOLO labeling format Fig. 6. utilizes a unique approach where a .txt file is generated for every image file in the directory, sharing the same name. These .txt files serve as containers for annotations

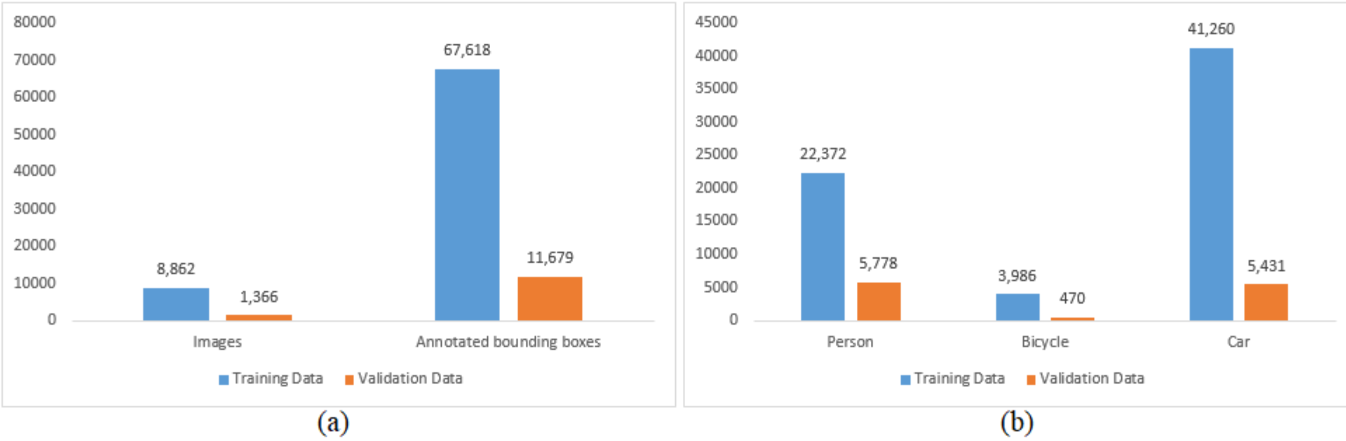


Fig. 1. (a) Total number of images and annotated bounding boxes (b) The annotations are distributed across the three main categories.

Fig. 2. Example of thermal image (left) and bounding boxes manually labeled with class person (right).

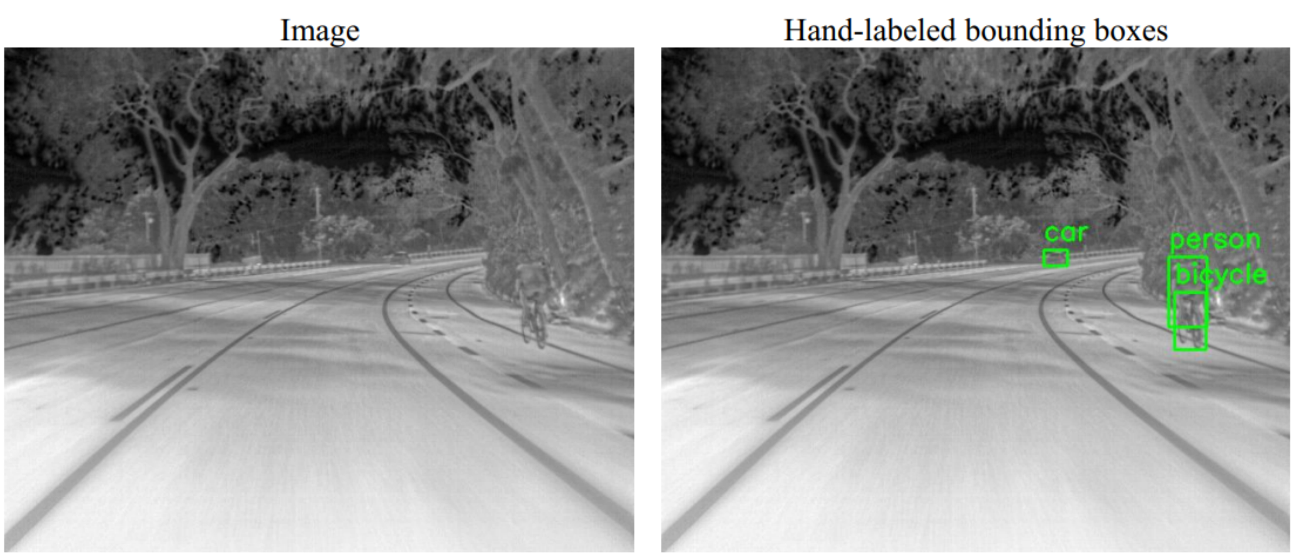


Fig. 3. Example of thermal image (left) and bounding boxes manually labeled with class Person, bicycle and car (right).

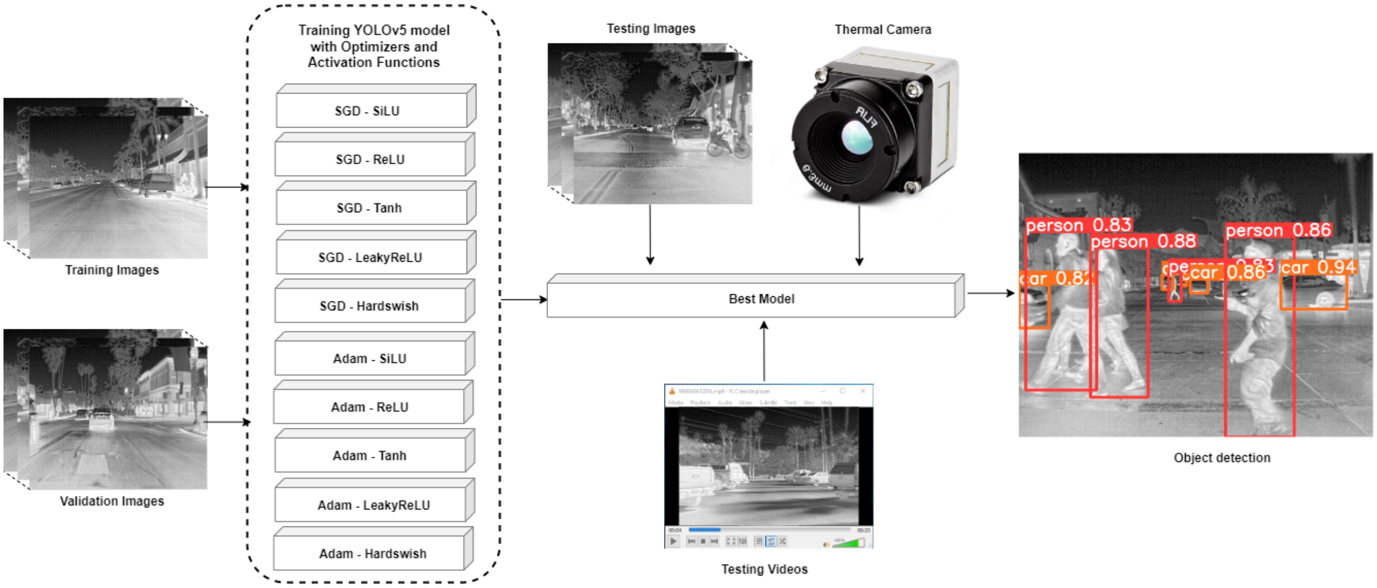
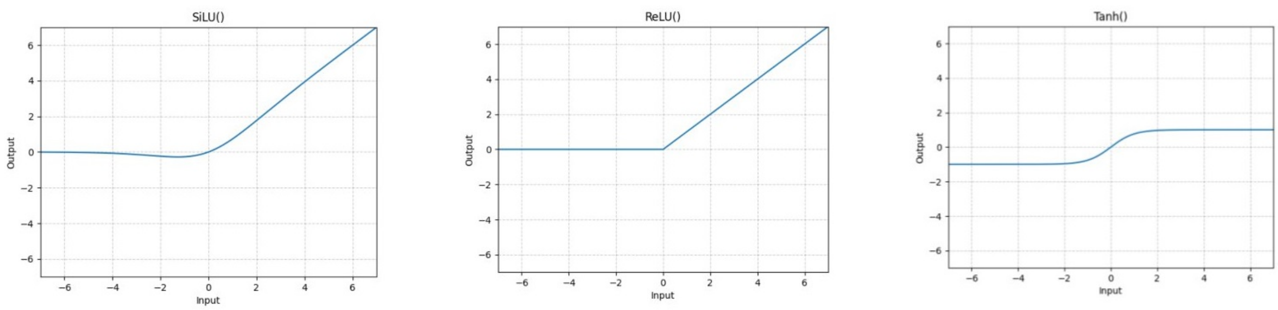


Fig. 4. Flowchart for the overall experiment conducted to train the YOLOv5 model with optimizers and activation functions.

$$\mathrm{SiLU}(x) * \sigma(x)$$
where $\sigma(x)$ is the logistic sigmoid.

$$\mathrm{ReLU}(x) = (x)^{+} = \max(0, x)$$

$$Tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$$



$$\mathrm{LeakyReLU(x)} = \begin{cases} x, & if\ x \geq 0 \\ negative_{slope} * x, & otherwise \end{cases}$$

$$Hardswish(x) = \begin{cases} 0 & if x \leq -3 \\ x & if x \geq +3 \\ x * (x+3)/6 & otherwise \end{cases}$$

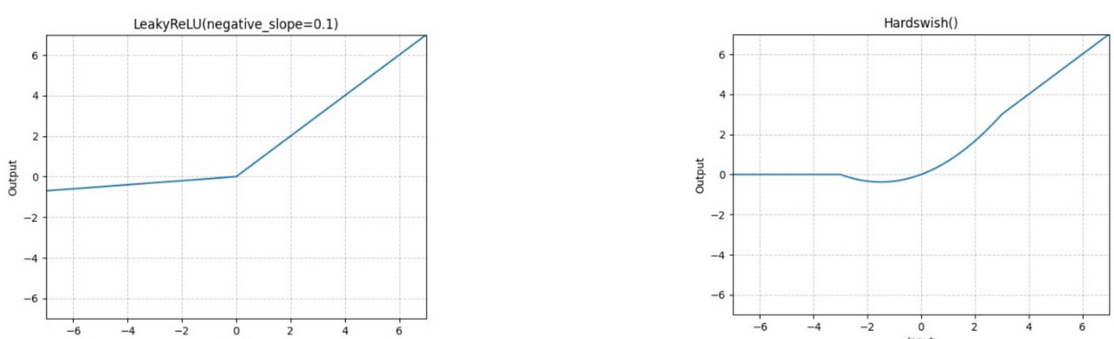


Fig. 5. Non-linear activations.

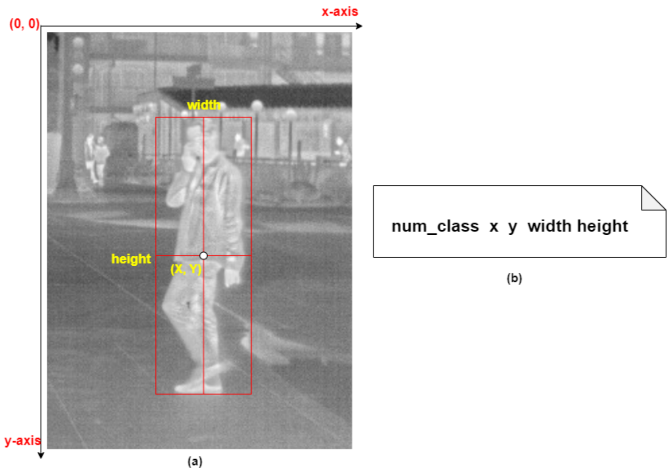


Fig. 6. (a) An example of a bounding box and (b) YOLO annotation format.

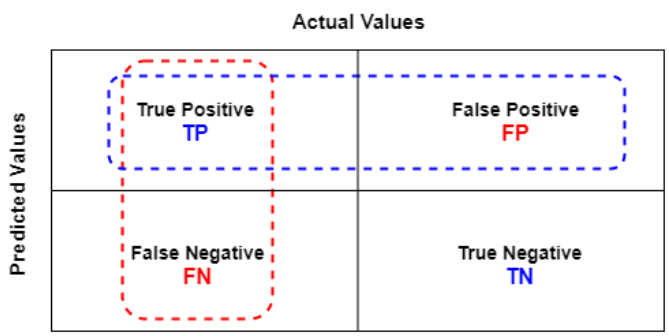


Fig. 7. Precision and recall.

related to the corresponding image file, encompassing object class, object coordinates, height, and width information.

C. Activation Functions

To conduct a comprehensive assessment of the accuracy of the transfer learning network models mentioned earlier, we employ five widely recognized and extensively used activation functions: SiLU (Sigmoid Linear Unit), ReLU (Rectified Linear Unit), Tanh (Hyperbolic Tangent), LeakyReLU (Leaky Rectified Linear Unit), and Hardswish [24]. These activation functions play a crucial role in deep learning methodologies. Each function's corresponding mathematical representation is presented Fig. 5, providing a complete understanding of their functional behavior. The selection of an appropriate activation function depends on a variety of factors, such as the specific requirements of the task at hand and the desired performance outcomes. Each activation function possesses unique characteristics that can influence the learning capabilities and overall performance of the transfer learning network models. By comparing the results obtained from employing these activation functions, we will be able to draw meaningful insights and make informed decisions regarding their suitability for the given task. The findings of this comparative analysis will be shared in detail in the subsequent section, offering a comprehensive evaluation of their effectiveness.

D. Model Evaluation Metrics

This study examined the efficiency of DL models using a range of metrics, including Precision, Recall and mAP in Fig. 7 and in equations (1), (2), and (3). Precision measured the ratio of accurate positive outcomes to all positive predictions, while recall measured the proportion of correctly predicted to all instances of positive outcomes in the dataset. mAP measures the similarity between the ground-truth bounding box and the detected box, resulting in a numerical score. This score serves as an indicator of the model's accuracy in detecting objects. A higher score signifies greater accuracy in the model's detections. By employing multiple evaluation metrics, we gained a comprehensive understanding of the model's performance and made well-informed judgments regarding its effectiveness.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$mAP = \frac{1}{n}\sum_{k=1}^{n} AP_k \tag{3}$$

In which, TP: True Positive, TN: True Negative, FP: False Positive, FN: False Negative, n: the number of classes, AP_k: the average precision of class k.

IV. Results

This section describes the training and validation results obtained for the YOLOv5 model using the SGD and Adam optimizers, along with various activation functions. The experiments were conducted using a learning rate of 0.01 and a momentum value of 0.937. Fig. 8 and Fig. 9 present the results of training the YOLOv5 model using two different optimizers, namely SGD and Adam, along with various activation functions. The performance of the model was evaluated using three key metrics: Precision, Recall, and mAP@0.5 (mean Average Precision at an IoU threshold of 0.5). Precision measures the accuracy of the model in correctly identifying positive instances, while Recall indicates the model's ability to find all positive instances. The mAP@0.5 calculates the average precision across different IoU thresholds.

For the SGD optimizer, the activation functions evaluated were SiLU, ReLU, Tanh, LeakyReLU, and Hardswish. Among these, the SiLU activation function achieved the highest Precision of 0.85247, Recall of 0.73373, and mAP@0.5 of 0.79985. However, other activation functions such as ReLU, LeakyReLU, and Hardswish also demonstrated competitive performance, with Precision ranging from 0.82512 to 0.83494 and mAP@0.5 ranging from 0.79074 to 0.79363. Similarly, for the Adam optimizer, the model was trained with the same set of activation functions. The SiLU activation function yielded the highest Precision of 0.80156, Recall of 0.70725, and mAP@0.5 of 0.77294. The performance of the other activation functions, including ReLU, LeakyReLU, and

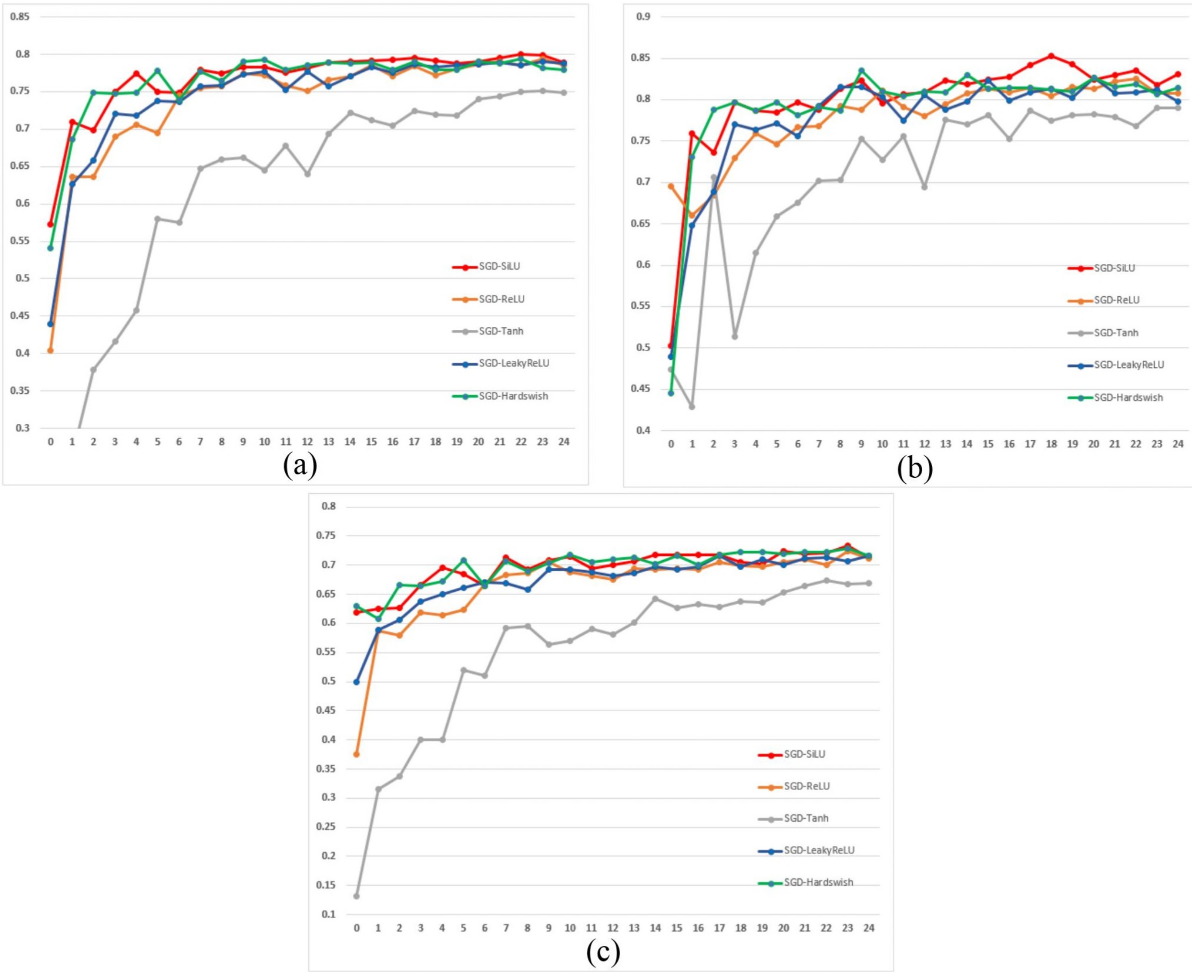


Fig. 8. Training results of YOLOv5 with SGD optimizer and various activation functions (a) The mAP@0.5 scores (b) Training precision, and (c) Training recall.

Hardswish, ranged from Precision values of 0.78057 to 0.7942 and mAP@0.5 values of 0.75974 to 0.77761.

Validation results of YOLOv5 with SGD and Adam optimizer and various activation functions are presented in Table 1. This table shows the evaluation metrics for precision, recall, and mAP@0.5. Each row corresponds to a specific combination of optimizer and activation functions. For the SGD optimizer, the SiLU activation function achieved the highest precision of 0.835, followed by Hardswish with a precision of 0.826. The highest recall was obtained with ReLU at 0.724, closely followed by SiLU at 0.722. The highest mAP@0.5 was achieved with SiLU at 0.800. For the Adam optimizer, the SiLU activation function again obtained the highest precision of 0.801, while ReLU achieved a precision of 0.782. The highest recall was obtained with ReLU at 0.703, closely followed by SiLU at 0.693. The highest mAP@0.5 was achieved with Hardswish at 0.777. Precision-Recall Curve of yolov5 model with SGD optimizer and SiLU activation function is presented in Fig. 10. These results demonstrate the performance of the YOLOv5 model with different combinations of optimizers and activation functions (see Fig. 11). These outcomes indicate the impact of different optimizers and activation functions on the YOLOv5 model's performance. The SiLU activation function consistently exhibited strong performance across both optimizers, while ReLU, LeakyReLU, and Hardswish also showed competitive results. These findings can guide researchers and practitioners in selecting the most effective configuration for training the YOLOv5 model in object detection tasks.

V. Conclusion

In conclusion, this research paper presented an innovative approach for object detection during nighttime in driver-assistance systems, utilizing thermal vision and incorporating the YOLOv5 model. The primary objective was to comprehensively evaluate the performance of the model by exploring the impact of different activation functions and optimizers. The outcomes demonstrated the efficiency of the proposed method in enhancing nighttime object detection. The experiments involved training the YOLOv5 model using two optimizers, SGD and Adam, along with various activation functions, namely SiLU, ReLU, Tanh, LeakyReLU, and Hardswish.

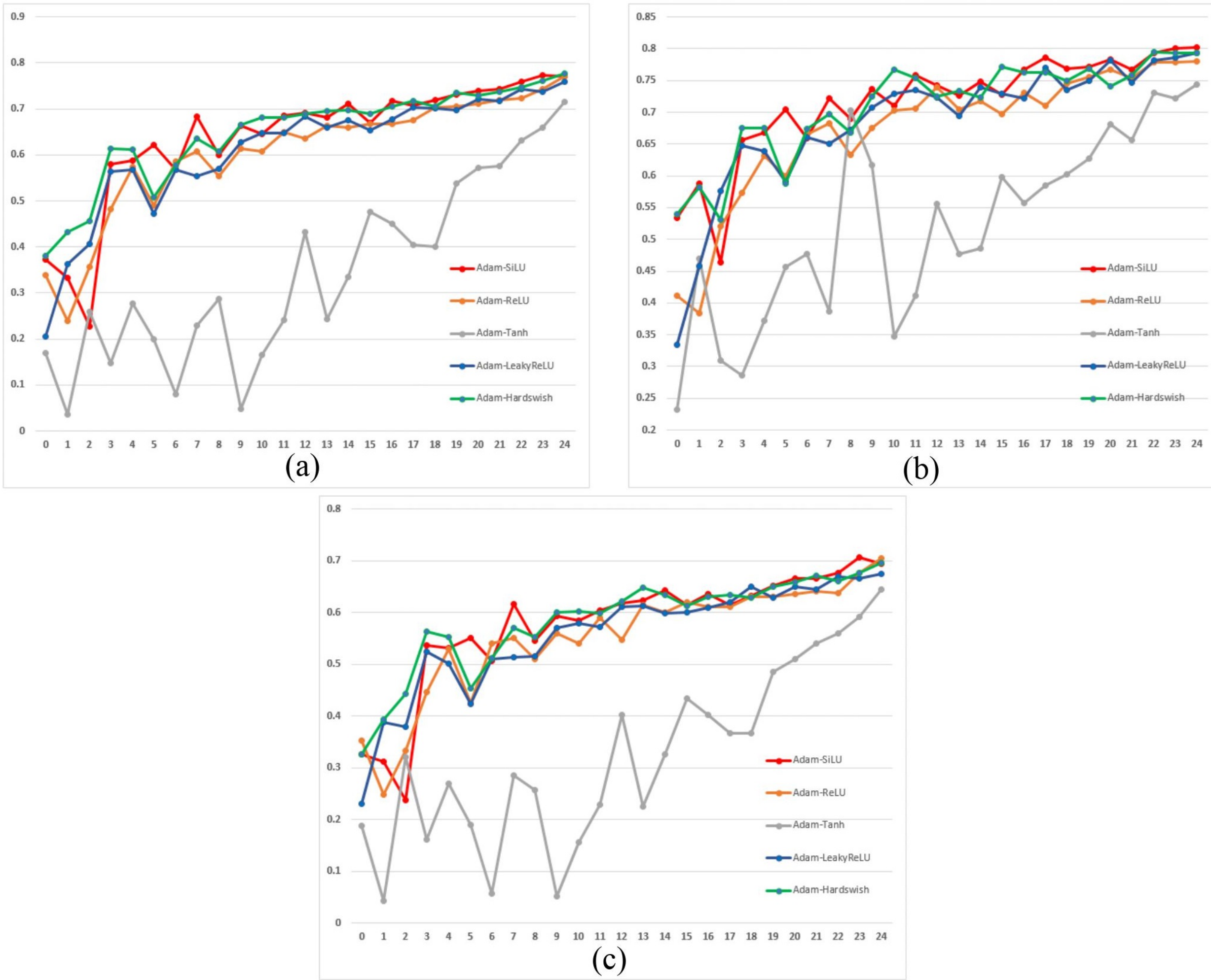


Fig. 9. Training results of YOLOv5 with adam optimizer and various activation functions (a) The mAP@0.5 scores (b) Training precision, and (c) Training recall.

TABLE I. VALIDATION RESULTS OF YOLOV5 WITH SGD AND ADAM OPTIMIZER AND VARIOUS ACTIVATION FUNCTIONS USE LEARNING RATE = 0.01 AND MOMEMTUM = 0.937

Optimizer	Activation Function	Precision	Recall	mAP@0.5
SGD	SiLU	0.835	0.722	0.800
SGD	ReLU	0.807	0.724	0.793
SGD	Tanh	0.787	0.667	0.751
SGD	LeakyReLU	0.813	0.706	0.790
SGD	Hardswish	0.826	0.719	0.790
Adam	SiLU	0.801	0.693	0.771
Adam	ReLU	0.782	0.703	0.771
Adam	Tanh	0.745	0.642	0.715
Adam	LeakyReLU	0.794	0.675	0.759
Adam	Hardswish	0.793	0.696	0.777

The evaluation metrics used, including Precision, Recall, and mAP@0.5, provided insights into the accuracy, coverage, and overall performance of the model. For the SGD optimizer, the SiLU activation function achieved the highest Precision and mAP@0.5 values, indicating its effectiveness in accurately identifying positive instances. However, ReLU, LeakyReLU, and Hardswish also demonstrated competitive performance in terms of Precision and mAP@0.5. Similarly, with the Adam optimizer, the SiLU activation function consistently yielded the highest Precision, while ReLU, LeakyReLU, and Hardswish also performed well. These results highlight the impact of different activation functions on the model's performance. Overall, the findings suggest that the YOLOv5 model, coupled with the SiLU activation function, is a promising configuration for nighttime object detection in driver-assistance systems. However, researchers and practitioners can also consider other activation functions such as ReLU, LeakyReLU, and Hardswish, which showed competitive performance in this study. These results contribute to the understanding of the role of optimizers and activation functions in training the YOLOv5 model for object detection tasks. The insights gained from this research can guide future endeavors in improving the efficiency and accuracy of driver-assistance systems, ultimately enhancing safety for both drivers and passengers.

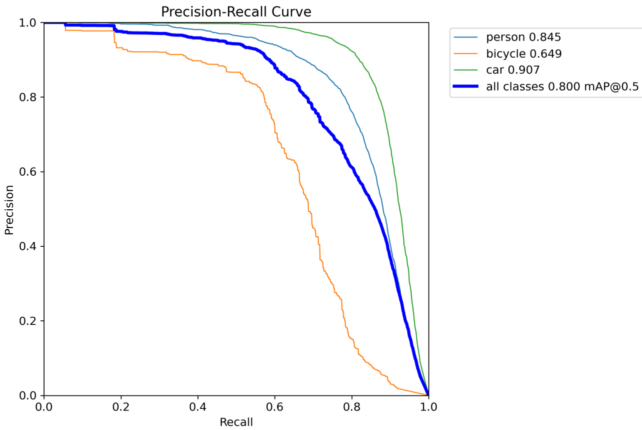


Fig. 10. Precision-Recall curve of yolov5 model with SGD optimizer and SiLU activation function.

REFERENCES

[1] Road traffic injuries, available online: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries. Accessed: 20 June 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

[2] T. N. Hoang and L.-D. Quach, "Adaptive lane keeping assist for an autonomous vehicle based on steering fuzzy-pid control in ros," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, 2022.

[3] P. B. Silva, M. Andrade, and S. Ferreira, "Machine learning applied to road safety modeling: A systematic literature review," *Journal of traffic and transportation engineering (English edition)*, vol. 7, no. 6, pp. 775–790, 2020.

[4] G. Li, H. Xie, W. Yan, Y. Chang, and X. Qu, "Detection of road objects with small appearance in images for autonomous driving in various traffic situations using a deep learning based approach," *IEEE Access*, vol. 8, pp. 211 164–211 172, 2020.

[5] A. Najjar, S. Kaneko, and Y. Miyanaga, "Combining satellite imagery and open data to map road safety," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.

[6] H. Lin, J. D. Deng, D. Albers, and F. W. Siebert, "Helmet use detection of tracked motorcycles using cnn-based multi-task learning," *IEEE Access*, vol. 8, pp. 162 073–162 084, 2020.

[7] W. Wang, B. Wu, S. Yang, and Z. Wang, "Road damage detection and classification with faster r-cnn," in *2018 IEEE international conference on big data (Big data)*. IEEE, 2018, pp. 5220–5223.

[8] H.-T. Vo, H. T. Ngoc, and L.-D. Quach, "An approach to hyperparameter tuning in transfer learning for driver drowsiness detection based on bayesian optimization and random search," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 4, 2023. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2023.0140492

[9] R. Simhambhatla, K. Okiah, S. Kuchkula, and R. Slater, "Self-driving cars: Evaluation of deep learning techniques for object detection in different driving conditions," *SMU Data Science Review*, vol. 2, no. 1, p. 23, 2019.

[10] R. Kulkarni, S. Dhavalikar, and S. Bangar, "Traffic light detection and recognition for self driving cars using deep learning," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*. IEEE, 2018, pp. 1–4.

[11] P. Prajwal, D. Prajwal, D. Harish, R. Gajanana, B. Jayasri, and S. Lokesh, "Object detection in self driving cars using deep learning," in *2021 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*. IEEE, 2021, pp. 1–7.

[12] V. D. Nguyen, H. Van Nguyen, D. T. Tran, S. J. Lee, and J. W. Jeon, "Learning framework for robust obstacle detection, recognition, and tracking," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1633–1646, 2016.

[13] H. Yu, R. Hong, X. Huang, and Z. Wang, "Obstacle detection with deep convolutional neural network," in *2013 Sixth International Symposium on Computational Intelligence and Design*, vol. 1. IEEE, 2013, pp. 265–268.

[14] P. Salavati and H. M. Mohammadi, "Obstacle detection using googlenet," in *2018 8th international conference on computer and knowledge engineering (ICCKE)*. IEEE, 2018, pp. 326–332.

[15] P. Aswathy, D. Mishra *et al.*, "Deep googlenet features for visual object tracking," in *2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS)*. IEEE, 2018, pp. 60–66.

[16] M. Saranya, N. Archana, J. Reshma, S. Sangeetha, and M. Varalakshmi, "Object detection and lane changing for self driving car using cnn," in *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)*. IEEE, 2022, pp. 1–7.

[17] S. H. Naghavi, C. Avaznia, and H. Talebi, "Integrated real-time object detection for self-driving vehicles," in *2017 10th Iranian Conference on Machine Vision and Image Processing (MVIP)*. IEEE, 2017, pp. 154–158.

[18] A. A. Cervera-Uribe and P. E. Mendez-Monroy, "U19-net: a deep learning approach for obstacle detection in self-driving cars," *Soft Computing*, vol. 26, no. 11, pp. 5195–5207, 2022.

[19] Y. Chen, W. Zheng, Y. Zhao, T. H. Song, and H. Shin, "Dw-yolo: an efficient object detector for drones and self-driving vehicles," *Arabian Journal for Science and Engineering*, vol. 48, no. 2, pp. 1427–1436, 2023.

[20] G. Rjoub, O. A. Wahab, J. Bentahar, and A. S. Bataineh, "Improving autonomous vehicles safety in snow weather using federated yolo cnn learning," in *Mobile Web and Intelligent Information Systems: 17th International Conference, MobiWIS 2021, Virtual Event, August 23–25, 2021, Proceedings*. Springer, 2021, pp. 121–134.

[21] T. Sharma, B. Debaque, N. Duclos, A. Chehri, B. Kinder, and P. Fortier, "Deep learning-based object detection and scene perception under bad weather conditions," *Electronics*, vol. 11, no. 4, p. 563, 2022.

[22] T. P. Dang, N. T. Tran, V. H. To, and M. K. Tran Thi, "Improved yolov5 for real-time traffic signs recognition in bad weather conditions," *The Journal of Supercomputing*, pp. 1–19, 2023.

[23] V. D. Nguyen, T. D. Trinh, and H. N. Tran, "A robust triangular sigmoid pattern-based obstacle detection algorithm in resource-limited devices," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2023.

[24] Non-linear activations (weighted sum, nonlinearity), available online: https://pytorch.org/docs/stable/nn.htmlnon-linear-activations-weighted-sum-nonlinearity. [Online]. Available: https://pytorch.org/docs/stable/nn.html#non-linear-activations-weighted-sum-nonlinearity

Fig. 11. Prediction results of the YOLOv5 model trained with SGD optimizer and SiLU activation function.