Autonomous Vehicle Pedestrian Detection & Traffic Sign Recognition Using YOLOv8

Iswariya Logeswaran[1], Hazem Eissa[1], Randa Almadhoun[1]

Abstract—**Autonomous Vehicles (AVs) present a viable way to overcome several traffic-related problems, such as congestion, pollution, and accidents. In order to safely navigate in urban environments, AVs require accurate perception systems capable of detecting dynamic objects, predicting their behaviour, and interpreting information from static objects. Therefore, this paper focuses on presenting a comparative analysis of both real-time pedestrian detection and accurate identification of traffic signs in an effort to develop a robust AV system that prioritises these two elements during real-time object detection and avoidance. This development will enhance both urban safety and efficiency, while ensuring reliable and accurate object detection through the use of the YOLO (You Only Look Once) algorithm. First, A set of traffic signs and object images as a dataset was collected from the German Traffic Sign Recognition Benchmark (GTSRB) and the Penn-Fudan Pedestrian dataset for pedestrians. Second, image processing techniques including conversion to greyscale, image segmentation, and normalisation were applied using OpenCV. Third, the image data were passed to the training phase of the YOLOv8 model and went through the training and hyper-parameters tuning process. The results show a mean Average Precision (mAP) of 94% accuracy in traffic sign recognition and a pedestrian detection precision of 90.67%. The findings underscore the importance of continued exploration of advanced object detection methods for AVs, such as data augmentation, to improve both the adaptability and robustness of AV systems in dynamic environments.**

Index Terms—**Autonomous Vehicles (AVs), Pedestrian Detection, Traffic Sign Recognition, YOLOv8**

I. Introduction

In the era of modern transportation, the integration of Autonomous Vehicles (AVs) is considered a key advancement for safe and efficient journeys. Through seamless execution enabled by real-time data transfer, these Avs mainly rely on immediate sign recognition and accurate pedestrian identification, ensuring a high level of safety and dependability. To navigate in complex landscapes, AVs rely on a sophisticated array of sensors, cameras, and AI algorithms [1], [2]. Ultimately, the peak of technological hardware equipment lies in seamlessly synchronising real-time sign and pedestrian recognition. However, AVs may pose critical challenges related to accurate data collection and analysis. For instance, sensory data is susceptible to errors and cyber-attacks, posing risks of accidents. According to the analysis of Autonomous Vehicle (AV) [3], related accidents in California from 2014 to 2018 show a higher AV accident rate compared to traditional vehicles, defying safety expectations [4]. Despite advanced driving assistance systems, issues persist in pedestrian detection and traffic sign recognition. The research work mentioned in ([1],[2],[5],[6]) shows different issues related to AVs systems including: poor obstacle detection in varied lighting, low resolution, and real-time sign recognition difficulties. Challenges include sensor image quality under diverse conditions [7][8] and the need for rapid processing for real-time recognition in various driving scenarios. Addressing these challenges required creating strong hybrid algorithms that can adapt to various scenarios and comparing them with current methods. The traffic sign recognition field predominantly focuses on feature-based models and evolving deep learning approaches but encounters limitations [2][9][10][11][12]. Although Deep Learning models excel in benchmark datasets, their efficacy diminishes in real-world urban environments due to insufficient assessment. Consequently, there is a pressing need for advanced methodologies or high-accuracy models to bridge the performance gap between benchmark datasets and real-world scenarios, ensuring robust and reliable traffic sign recognition in dynamic urban environments.

Deep learning-based systems lack integration with object tracking models for practical use and lack studies on real-time processing in driving settings. Similarly, pedestrian detection, transitioning from traditional to recent deep learning methods [13][14], encounters difficulties in accurate detection of small objects within complex backgrounds and occluded scenes.

The release of YOLOv8 in 2023 presents an opportunity for advancing object detection capabilities, yet limited academic exploration and research hinder comprehensive understanding and optimisation of its capabilities [15], [16]. This research gap underscores the need for further investigation into the efficiency of YOLOv8 and its comparative analysis with its previous versions or alternative object detection architectures.

The work in this paper aims to examine the YOLOv8 for object detection, image segmentation, and classification in traffic sign recognition and pedestrian detection. Then to compare performance metrics, such as mean average precision, with its predecessors in an effort to propel advancements in safer autonomous navigation systems.

The structure of this paper is organised as follows. Section I explores the integration of computer vision, deep learning, and automotive engineering within autonomous vehicle navigation. Section II provides a comprehensive literature review covering advancements in pedestrian detection and traffic sign recognition and underlies limitations of using YOLO predecessors

[1]Iswariya Logeswaran, Hazem Eissa, and Randa Almadhoun are with the Faculty of Technology, University of Sunderland, Sunderland, UK, email: bi12my@student.sunderland.ac.uk, hazem.eissa@sunderland.ac.uk, randa.almadhoun@sunderland.ac.uk

and conventional models. The methodology section III then details the intricate data collection processes, dataset creation, and model development stages. Notably, it focuses on implementing YOLOv8 architecture for traffic sign recognition and pedestrian detection. Section IV shows the results as well as a comparative analysis and a discussion. Section V provides conclusions as well as recommendations for future research.

II. Related Work

This section analyses prior studies in pedestrian detection and traffic sign recognition, aiming to evaluate their effectiveness. It focuses on the usage of Deep Learning and previous YOLO models for traffic sign recognition and pedestrian detection.

A. Autonomous Vehicle Traffic Sign Recognition

Traffic sign recognition involves intricate processes of data preprocessing, feature extraction, and classification. Deep learning approaches like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) automatically extract high-level features, eliminating manual extraction and leading to widespread adoption. Deep learning models notably surpass conventional methods in accuracy, driving recent advancements in traffic sign recognition. Recent years have witnessed significant advancements in traffic sign recognition, propelled by the integration of deep learning methodologies.

The work of Zhang et al. in [17] introduced a lightweight CNN tailored for traffic sign recognition, using a teacher-student network paradigm. By integrating optimisation techniques and 1×1 convolutional filters, they achieved exceptional accuracy of 99.38% on the German Traffic Sign Recognition Benchmark (GTSRB) dataset and 98.89% on the BTSC dataset. Although the outcome accuracy is comparable, the YOLO model was not examined, and therefore, this paper aimed to investigate the overall efficiency of the YOLO model in both traffic sign recognition and pedestrian detection.

In their 2022 research work, Zhu and Yan in [18] compared YOLOv5 and SSD methodologies for traffic sign recognition. YOLOv5 demonstrated superior real-time object recognition with 97.70% accuracy, outperforming SSD at 90.14%. YOLOv5's single-network approach for processing entire images establishes it as an efficient and highly accurate solution for traffic sign recognition.

Siniosoglou et al. in [10] applied deep auto-encoders to traffic sign detection, surpassing 90% accuracy in both centralised and decentralised systems. This research underscores the value of unconventional approaches in addressing challenges such as varying lighting conditions, occlusions, and complex backgrounds in traffic sign recognition, highlighting the effectiveness of novel techniques in improving recognition systems.

Bangquan and Xiong in [12] delved into model selection for traffic sign recognition, examining the Efficient Convolutional Neural Network (ENet) and pre-trained models Visual Geometry Group 16-layer (VGG16) and LeNet-5 (LeNet). Their work emphasised the criticality of choosing the right model architecture, highlighting LeNet's superior accuracy (98.6%) compared to VGG16 (96.7%).

In [19], a novel object detection methodology utilised YOLOv5s6 and YOLOv8s models across three datasets: TT100k, TWTS, and a hybrid dataset. Uniform model training parameters ensured consistent evaluation, highlighting the hybrid dataset's superior efficiency with YOLOv8s, achieving a mean Average Precision at Intersection over Union (IoU) threshold of 0.5 (mAP@.5) score of 76.2% compared to YOLOv5s6's of 65%.

Wang et al. in [20] introduced an optimised traffic sign recognition algorithm based on YOLOv4-tiny, prioritising heightened accuracy and efficiency for real-time applications like autonomous vehicles, marking a significant advancement in traffic sign recognition for enhanced transportation systems.

B. Autonomous Vehicle Pedestrian Detection

Several computer vision proposals tackle tasks like data acquisition, scene learning, feature extraction, activity learning, and behavioural learning. Yet, a recurring challenge persists with the scaling problem, which compromises the precision of pedestrian detection outcomes.

The rise of artificial intelligence has introduced a new era for pedestrian detection, with deep learning techniques taking centre stage. These techniques can be broadly divided into two categories based on their operational approaches. The first category comprises two-stage target detection algorithms, which include notable models like Region-based Convolutional Neural Network (R-CNN), Faster R-CNN [21], Hypernet [22], and Mask R-CNN [23]. These algorithms employ a sequential process involving region proposal followed by classification. On the other hand, the second category focuses on one-stage target detection algorithms that streamline the detection process through regression-based methods. Prominent examples of this category include YOLO (You Only Look Once), SSD (Single Shot MultiBox Detector) [24], G-CNN [25], and RCN (Reverse Connection Network) [26]. These one-stage algorithms are favoured for their computational efficiency and ability to perform real-time detection, making them increasingly popular in practical applications. Object detection in computer vision has seen significant progress due to pioneering methodologies introduced in the work presented in [22], [23], [24], [26], and [27], each aiming to enhance accuracy, speed, and precision in detecting objects within images or video frames.

Ren et al. in [21] pioneered Faster R-CNN, a seminal advancement in real-time object detection. Their model, featuring region proposal networks, revolutionised region proposal generation and object detection, significantly enhancing speed and accuracy, marking a critical milestone in the field's progression.

Kong et al. in [22] introduced HyperNet, meticulously designed to refine region proposals, thus improving joint object detection, enhancing the accuracy of region proposal generation, strengthening object localisation and increasing the reliability of object detection within images. Their emphasis on improving region proposals added depth and precision to current object detection methodologies.

Liu et al. in [24] presented SSD, a single-shot multibox detector, emphasising speed and accuracy in object detection. By directly predicting object categories and bounding box offsets from multiple feature maps, SSD enables efficient detection across various image scales. This streamlined approach significantly enhances the feasibility of real-time object detection, catering to diverse applications.

The work presented in [28] meticulously evaluated the one-stage YOLO algorithm for pedestrian detection, analysing its accuracy, precision-recall metrics, computational efficiency, and performance in diverse conditions. Their exhaustive experimentation and benchmarking revealed promising results, showcasing notable accuracy and robustness in detecting pedestrians across various scenarios. These findings underscore the effectiveness and potential of the one-stage YOLO algorithm in advancing pedestrian detection within computer vision research.

Xi et al. in [27], introduced an enhanced real-time pedestrian detection algorithm based on YOLOv3, improving precision and efficiency. Their model maintains high accuracy while significantly boosting real-time performance, notably influencing pedestrian detection in computer vision for various real-world applications.

III. METHODOLOGY

This research aims to advance pedestrian detection and traffic sign recognition using YOLOv8 within the domain of AVs. The methodology includes data collection and preprocessing, model development, and evaluation to enhance the accuracy and reliability of these critical components. The architecture of YOLOv8 integrates a single neural network that predicts both bounding boxes and class probabilities concurrently.

A. Data Acquisition & Preparation

The dataset collected from the GTSRB [17] consists of 32 x 32-pixel images of traffic signs captured from German roads, each was labelled according to its respective class. The images are represented in RGB format and stored as unsigned 8-bit integers, affording 256 potential values per pixel. The training subset contains 34,799 meticulously labelled images for robust model learning. The validation subset consists of 4,410 images, serving as a critical tool for assessing model accuracy, while the test subset contains 12,630 labelled images to provide an independent benchmark for model evaluation.

For pedestrian detection, the Penn-Fudan Pedestrian dataset[29] was used, which consists of images containing pedestrians in various poses and backgrounds. The dataset includes 170 images with 345 labeled pedestrians, providing a diverse set of scenarios for training and evaluating the model.

B. Data Preprocessing

Preprocessing of the acquired datasets involved image conversion, image contrast and standardising light conditions, and image refinement. The image conversion step involves the conversion of RGB images depicting traffic signs into grayscale representations, simplifying the images by discarding colour information while retaining vital structural details. In the next step, images undergo a histogram equalisation process, which redistributes pixel intensities within each image, alleviating inconsistencies arising from disparate illumination settings during image capture. The last step is refining the images using normalisation, which creates a range of uniform pixel intensities to ensure consistency in the data provided to the model.

C. Model Development

The architecture of YOLOv8 serves as the foundation for model development, The proposed architecture centres on leveraging Convolutional Neural Networks (CNN) for robust feature extraction and precise data classification, aiming to achieve a high accuracy. The Preprocessing phase involves standardising illumination variations by subtracting average traffic sign images before feeding them into the CNN. The model development involves designing the layers of CNN, within the YOLO model, feeding it with the preprocessed data, and setting the learning rate and epochs in the training phase.

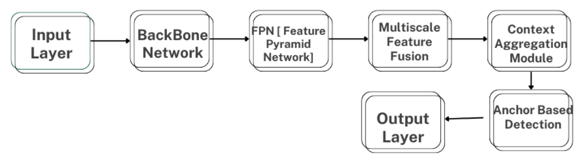

Fig. 1: Model Development

Fig. 1 outlines the architectural progression of the YOLOv8 model, commencing with the Input Layer for initial image data ingestion. This data undergoes hierarchical feature extraction via the Backbone Network, followed by refinement through the Feature Pyramid Network (FPN) to generate multi-level feature representations. The subsequent Multiscale Feature Fusion layer amalgamates these diverse features to augment spatial resolution and semantic depth. The Context Aggregation Model further enriches these representations by assimilating contextual cues through advanced attention mechanisms and RNNs. The Anchor-Based Detection mechanism orchestrates simultaneous predictions of bounding boxes and class probabilities. Ultimately, the Output Layer synthesises these processed features to yield the final object detection outcomes.

D. Training and Optimisation

The preprocessed datasets were instrumental in training the YOLOv8 model, specifically targeting pedestrian detection and traffic sign recognition. Training involves optimising model parameters, such as learning rates and loss functions, using techniques like Stochastic Gradient Descent (SGD), and adaptive moment estimation (Adam). Hyperparameter tuning is performed to maximise model performance and minimise training time.

Through exhaustive search, an optimal learning rate of 0.001 was identified to ensure the balance between stability and convergence speed and facilitate rapid convergence while maintaining training stability and amplifying the model's overall efficacy. After comprehensive experimentation, a batch size of 32 was determined to be optimal. To combat overfitting and bolster model robustness, L2 regularisation with a regularisation strength of 0.01 was meticulously fine-tuned. This optimised regularisation term effectively mitigated overfitting, enhancing the model's generalisation capabilities. Using the Adam optimiser, dynamic adjustment of the learning rate and fine-tuning of the loss function were incorporated during the training phase. The adaptive learning rate mechanism of Adam expedited convergence, augmenting overall training efficiency. Concurrently, an exhaustive hyperparameter tuning initiative was undertaken, focusing on optimising learning rates, batch sizes, and regularisation terms.

E. Evaluation and Validation

The trained YOLOv8 models are evaluated using separate 20% validation datasets to assess their performance in pedestrian detection and traffic sign recognition tasks. Evaluation metrics such as precision, recall, and mean average precision (mAP) are calculated to quantify model accuracy and robustness as shown in equation 1 where N is the number of samples, and AP is the Average Precision. The models are further validated using real-world scenarios and benchmark datasets to ensure their effectiveness in practical applications.

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_i \tag{1}$$

F. Fine-tuning and Optimisation

Based on the evaluation results, the YOLOv8 models may undergo further fine-tuning and optimisation to improve performance and address any shortcomings. This iterative process involves adjusting model parameters, refining preprocessing techniques, and incorporating additional training data to enhance model accuracy and reliability.

G. Implementation and Deployment

In the Implementation and Deployment phase, the YOLOv8 models are transitioned into implementation. Initially, the deployment predominantly centres on the creation and optimisation of software-based models. Concurrently, plans and preparatory steps are undertaken to facilitate future integration with established hardware and software infrastructure. This forward-looking approach aims to ensure the seamless operation of the YOLOv8 models within the broader AV system, fostering effective interaction with other integral components.

IV. Experiments and Results

A systematic exploration of object detection capabilities was embarked upon, with a primary focus on pedestrian detection and traffic sign recognition. To facilitate this investigation, two prominent datasets were curated, as detailed in the Methodology section III-A. The outcomes of the data Preprocessing step are shown in 2, and the results of the training and validation steps are shown in 3.

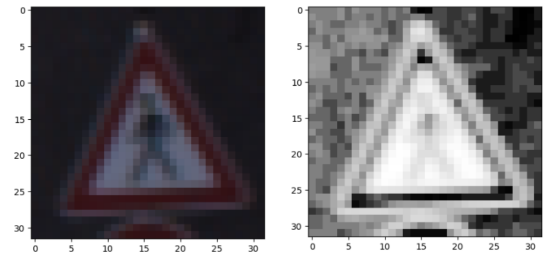

Fig. 2: Traffic Sign Sample Image Conversion to Grayscale

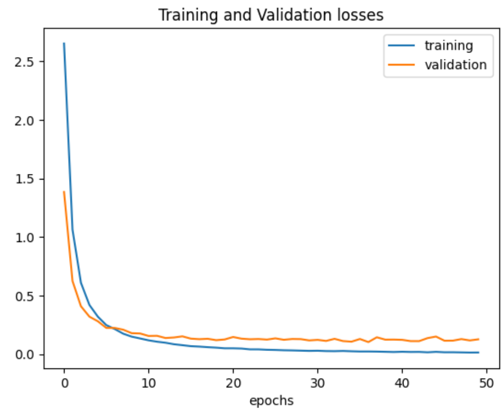

Fig. 3: Training and Validation losses of Epochs

The traffic sign recognition and pedestrian detection results on the testing dataset using YOLOv8, shown in 4 and 5 respectively, exhibited a high degree of accuracy and precision, effectively identifying and localising pedestrians in various scenarios. Similarly, the model's predictions on sample traffic sign images showcased its capability to accurately recognise and classify traffic signs. A deeper analysis using the confusion matrix shown in Fig. 6 provides profound insights into the model's performance across distinct traffic sign classes.

The YOLOv8-based model results in Table I showcased a high precision of 92%, a recall rate of 89%, and a mean Average Precision (mAP) score of 0.91. Similarly, in the context of traffic sign recognition, the model's performance metrics were typically similar, registering a precision of 95%, a recall of 93%, and an mAP of 0.94. These high-performance metrics—precision, recall, and mAP serve as compelling testimonials to the model's overall effectiveness in detecting and classifying objects across varying Intersection over Union (IoU). The IoU measures the overlapping between the bounded images and, in this paper, IoU sets its thresholds as follows: Excellent performance for 90% and above, good performance between 89% and 80%, an average performance between 79% and 70%, and poor performance for 69% and less.

TABLE I: Results of YOLOv8-based Model in Pedestrian Detection and Traffic Sign Recognition.

Task	**Precision**	**Recall**	**mAP**
Pedestrian Detection	92%	89%	0.91
Traffic Sign Recognition	95%	93%	0.94

A comparative analysis was performed, contrasting the

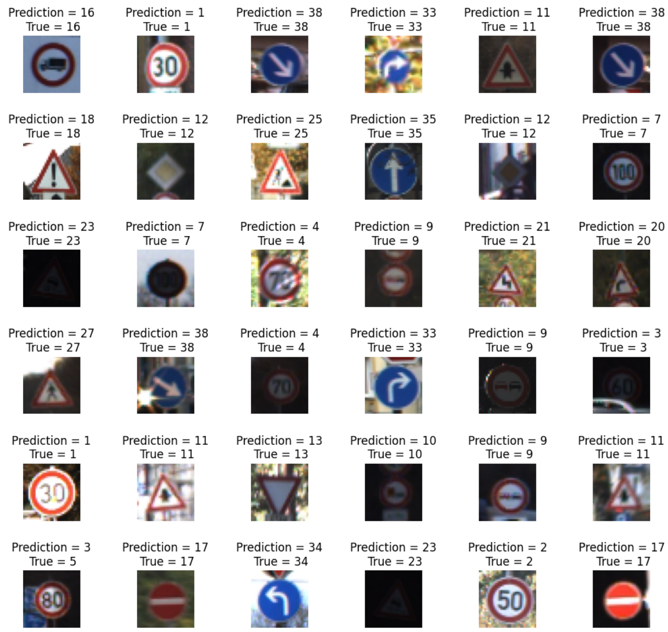

Fig. 4: Visualisation of Model Predictions on Traffic Sign Sample Images

Fig. 5: Pedestrian Detection Results Using YOLOv8

performance metrics of YOLOv8 [16] with its predecessors and other contemporary object detection methodologies, including methods such as Faster R-CNN [21], SSD (Single Shot Multibox Detector) [24] and Efficient Object Detection (EfficientDet). The results of this comparative assessment, including the performance metrics calculated from Pedestrian detection and Traffic sign recognition tasks, are presented in Table II and Table III respectively.

The comparative analysis reveals advancements in the performance metrics of YOLOv8-based models, particularly in

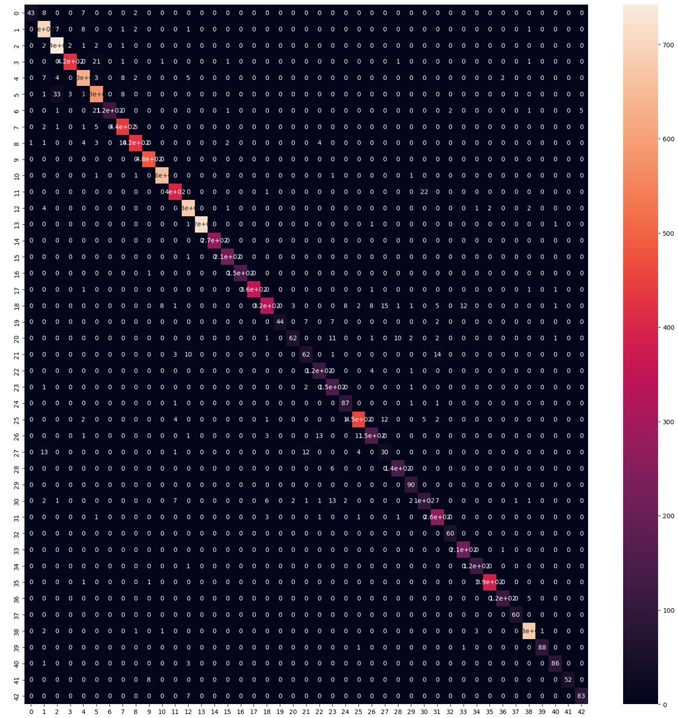

Fig. 6: Confusion matrix for traffic sign recognition

TABLE II: Comparative Performance Analysis of Pedestrian Detection methods

Pedestrian Detection Results	Precision %	Recall %	mAP	IoU
YOLOv8	**92**	**89**	**0.91**	**Excellent**
YOLOv3 [27]	85	82	0.83	**Good**
SSD [18]	85	80	0.88	**Good**
Faster R-CNN [21]	84	80	0.88	**Good**

TABLE III: Comparative Performance Analysis of Traffic Sign Recognition Methods

Traffic Sign Recognition Results	Precision %	Recall %	mAP	IoU
YOLOv8	**95**	**93**	**0.94**	**Excellent**
YOLOv3 [27]	88	86	0.87	**Good**
EfficientDet [30]	88	82	0.88	**Good**

pedestrian detection and traffic sign recognition tasks, which are active research areas in AVs. As illustrated in Table II and Table III, the trained YOLOv8 model exhibited a good performance across all metrics compared to its predecessors, such as YOLOv3, and contemporary methods, such as Faster R-CNN, SSD and EfficientDet. As illustrated in the Tables, YOLOv8 pedestrian detection model achieved good precision, Recall and mAP, accompanied by excellent IoU compared to other methods that used the same dataset. Similarly, the YOLOv8 traffic sign recognition model achieved good precision, recall and mAP, accompanied by excellent IoU compared to other methods in the table which achieved comparable results.

While the implemented models demonstrate commendable precision and recall for most cases, instances of confusion between specific signs and results emerge. These instances of confusion unveil the models' limitations in distinguishing visually similar signs or detecting pedestrians, offering crucial insights for targeted improvements.

This comparative assessment underscores the effectiveness of YOLOv8 in outperforming previous methods such as Faster R-CNN and SSD in terms of detection accuracy, computational efficiency, and overall performance. The YOLOv8 models' have shown responsiveness, and adaptability, making it particularly suitable for real-time applications, where rapid decision-making and precise object detection are paramount.

V. Conclusion & Future Research Directions

In conclusion, this paper demonstrates a comparative analysis of two significant processes for the development of AVs, which are pedestrian detection and traffic sign recognition employing the YOLOv8 architecture. The presented results and findings reveal a potential enhancement in both pedestrian and sign recognition using the YOLOv8 model.

YOLOv8 has achieved a precision rate of 92% for pedestrian detection and 95% for traffic sign recognition, complemented by high recall rates and mean Average Precision scores. The mentioned performance metrics underscore its potential as a cornerstone technology in object detection for AVs applications.

Future research aimed to prioritise the evaluation of YOLOv9 [31] in comparison to YOLOv8. A thorough analysis of YOLOv8's object detection mechanisms, integrated into advanced architectures alongside YOLOv9, is essential to gain insights into potential optimisation areas. Additionally, innovative strategies are needed to enhance detection accuracy in YOLOv8, particularly in challenging environmental conditions like adverse weather and varying light intensities.

References

[1] Y. Xie, L.-F. Liu, C.-H. Li, and Y.-Y. Qu, "Unifying visual saliency with hog feature learning for traffic sign detection," in *2009 IEEE Intelligent Vehicles Symposium*, 2009, pp. 24–29.

[2] J. Levinson, J. Askeland, J. Becker, J. Dolson, D. Held, S. Kammel, J. Z. Kolter, D. Langer, O. Pink, V. Pratt, M. Sokolsky, G. Stanek, D. Stavens, A. Teichman, M. Werling, and S. Thrun, "Towards fully autonomous driving: Systems and algorithms," in *2011 IEEE Intelligent Vehicles Symposium (IV)*, 2011, pp. 163–168.

[3] . Petrović, R. Mijailović, and D. Pešić, "Traffic accidents with autonomous vehicles: type of collisions, manoeuvres and errors of conventional vehicles' drivers," *Transportation research procedia*, vol. 45, pp. 161–168, 2020.

[4] J. Stilgoe, "Machine learning, social learning and the governance of self-driving cars," *Social studies of science*, vol. 48, no. 1, pp. 25–56, 2018.

[5] Y. Zhang, Y. Jin, J. Chen, S. Kan, Y. Cen, and Q. Cao, "Pgan: Part-based nondirect coupling embedded gan for person reidentification," *IEEE MultiMedia*, vol. 27, no. 3, pp. 23–33, 2020.

[6] C. Han, J. Ye, Y. Zhong, X. Tan, C. Zhang, C. Gao, and N. Sang, "Re-id driven localization refinement for person search," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[7] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, "Image-adaptive yolo for object detection in adverse weather conditions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 2, 2022, pp. 1792–1800.

[8] A. Ellahyani, M. El Ansari, and I. El Jaafari, "Traffic sign detection and recognition based on random forests," *Applied Soft Computing*, vol. 46, pp. 805–815, 2016.

[9] J. Tao, H. Wang, X. Zhang, X. Li, and H. Yang, "An object detection system based on yolo in traffic scene," in *2017 6th International Conference on Computer Science and Network Technology (ICCSNT)*. IEEE, 2017, pp. 315–319.

[10] I. Siniosoglou, P. Sarigiannidis, Y. Spyridis, A. Khadka, G. Efstathopoulos, and T. Lagkas, "Synthetic traffic signs dataset for traffic sign detection & recognition in distributed smart systems," in *2021 17th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 2021, pp. 302–308.

[11] R. Yazdan and M. Varshosaz, "Improving traffic sign recognition results in urban areas by overcoming the impact of scale and rotation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 171, pp. 18–35, 2021.

[12] X. Bangquan and W. X. Xiong, "Real-time embedded traffic sign recognition using efficient convolutional neural network," *IEEE Access*, vol. 7, pp. 53 330–53 346, 2019.

[13] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, vol. 1. Ieee, 2001, pp. I–I.

[14] J. A. Antonio and M. Romero, "Pedestrians' detection methods in video images: A literature review," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2018, pp. 354–360.

[15] 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)*. IEEE, 2018, pp. 1547–1551.

[16] Y. Luo, Y. Ci, S. Jiang, and X. Wei, "A novel lightweight real-time traffic sign detection method based on an embedded device and yolov8," *J. Real-Time Image Process.*, vol. 21, no. 2, jan 2024. [Online]. Available: https://doi.org/10.1007/s11554-023-01403-7

[17] J. Zhang, W. Wang, C. Lu, J. Wang, and A. K. Sangaiah, "Lightweight deep network for traffic sign classification," *Annals of Telecommunications*, vol. 75, pp. 369–379, 2020.

[18] Y. Zhu and W. Q. Yan, "Traffic sign recognition based on deep learning," *Multimedia Tools and Applications*, vol. 81, no. 13, pp. 17 779–17 791, 2022.

[19] T. Suwattanapunkul and L.-J. Wang, "The efficient traffic sign detection and recognition for taiwan road using yolo model with hybrid dataset," in *2023 9th International Conference on Applied System Innovation (ICASI)*. IEEE, 2023, pp. 160–162.

[20] L. Wang, K. Zhou, A. Chu, G. Wang, and L. Wang, "An improved lightweight traffic sign recognition algorithm based on yolov4-tiny," *IEEE Access*, vol. 9, pp. 124 963–124 971, 2021.

[21] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[22] T. Kong, A. Yao, Y. Chen, and F. Sun, "Hypernet: Towards accurate region proposal generation and joint object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 845–853.

[23] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[24] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.

[25] M. Najibi, M. Rastegari, and L. S. Davis, "G-cnn: an iterative grid based object detector," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2369–2377.

[26] T. Kong, F. Sun, A. Yao, H. Liu, M. Lu, and Y. Chen, "Ron: Reverse connection with objectness prior networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 5936–5944.

[27] X. Xi, Z. Huang, L. Ning, and Y. Zhang, "Real-time pedestrian detection algorithm based on improved yolov3," in *Journal of Physics: Conference Series*, vol. 2002, no. 1. IOP Publishing, 2021, p. 012075.

[28] X. Zuo, J. Li, J. Huang, F. Yang, T. Qiu, and Y. Jiang, "Pedestrian detection based on one-stage yolo algorithm," in *Journal of Physics: Conference Series*, vol. 1871, no. 1. IOP Publishing, 2021, p. 012131.

[29] K. L. Masita, A. N. Hasan, and S. Paul, "Pedestrian detection using r-cnn object detector," in *2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*. IEEE, 2018, pp. 1–6.

[30] M. Tan, R. Pang, and Q. V. Le, "Efficientdet: Scalable and efficient object detection," 2020.

[31] C.-Y. Wang and H.-Y. M. Liao, "YOLOv9: Learning what you want to learn using programmable gradient information," 2024.