JOIV : Int. J. Inform. Visualization, 9(1) - January 2025 248-257

INTERNATIONAL JOURNAL ON INFORMATICS VISUALIZATION

journal homepage : www.joiv.org/index.php/joiv

Improving YoloPX using YoloP and Yolov8 for Panoptic Driving Perception

Xie Yumeng [a], Noridayu Binti Manshor [a,*], Nor Azura Husin [a], Liu Chengzhi [a]

[a] *Faculty of Computer Science, Universiti Putra Malaysia, Serdang, Selangor, Malaysia*

Corresponding author: [*]ayu@upm.edu.my

***Abstract*—Autonomous driving technology (ADS) has seen significant advancements over the past decade, with car manufacturers investing heavily in its development to meet the growing demand for safer, more efficient, and eco-friendly transportation solutions. The panoptic driving perception system is central to ADS, essential for accurately interpreting the driving environment. This system requires high precision, lightweight design, and real-time responsiveness to detect surrounding vehicles, lane lines, and drivable areas effectively. This study introduces an enhanced YOLOPX model that combines YOLOP and YOLOv8 to create an adaptive multi-task learning network capable of traffic object detection, drivable area segmentation, and lane detection. The model integrates YOLOP's detection head with YOLOPX's anchor-free detection head to improve generalization, incorporates YOLOv8's advanced backbone structure to enhance feature extraction accuracy, and retains YOLOP's three-neck architecture to optimize multi-task processing. The improved model employs a mode loss function for segmentation tasks, enhancing generalization and improving lane detection accuracy. Experiments conducted using the BDD100k dataset demonstrated the model's effectiveness: achieving 98.8% accuracy and 27.6% IoU for lane line detection, 90.4% mIoU for drivable area segmentation, and 85.9% recall and 76.9% mAP50 for traffic object detection. This model represents a significant advancement in ADS, enhancing both the safety and reliability of autonomous vehicles.**

***Keywords*—ADS; YOLOPX; YOLOv8; panoptic driving perception.**

Manuscript received 5 Aug. 2024; revised 19 Nov. 2024; accepted 14 Dec. 2024. Date of publication 31 Jan. 2025.

I. Introduction

Deep learning has made significant advances in computer vision, integrating it into applications such as facial recognition, workpiece gap detection, and medical imaging. One notable application is Autonomous Driving Systems (ADS), which transform driving by increasing comfort, reducing human error, avoiding traffic hazards, maximizing efficiency, and improving safety. The panoramic driving awareness system is essential to ADS, utilizing a multi-task approach to interpret complex traffic scenarios and provide real-time environmental data. The system divides drivable areas, identifies objects, and detects lanes to ensure efficient and safe driving in complex conditions. Despite its advantages, there are still challenges in improving detection capabilities. Ongoing research is essential to improve these systems' flexibility and reliability and ensure ADS's safety and efficiency.

Panoptic driving perception systems [1] form the foundation of advanced driver-assistance system (ADAS) technology, and there are two primary development approaches: systems based on deep learning and computer vision [2] and those based on multi-sensor models [3].

The first approach employs cameras and deep learning-based computer vision algorithms to detect the surrounding environment. This method is cost-effective, easy to integrate, and capable of achieving high real-time performance (over 30 FPS), which is crucial for timely decision-making to ensure driving safety. Accuracy and speed are critical in making decisions to safeguard driving.

The second approach integrates advanced sensors, such as LiDAR [4], [5] to obtain more comprehensive environmental data. However, the complexity of integration presents challenges in achieving real-time performance and managing higher costs. Some neural network models, such as YOLOP, HybridNet, and Faster R-CNN [6], [7], [8], [9], [10], [11] have demonstrated promising results. YOLOP stands out for its accuracy and real-time processing capability, making it an ideal choice for ADS, although it struggles with the precision of small object detection. HybridNet performs well but incurs high computational costs, while Faster R-CNN, despite its high accuracy, is too slow for real-time applications. Considering the real-time requirements of panoptic driving

perception systems, the YOLOP series algorithms are preferred.

In Autonomous Driving Systems (ADS), panoptic driving perception models are vital for autonomous vehicles' safe and efficient navigation. These models integrate multiple tasks, including object detection, drivable area segmentation, and lane detection, into a unified system to provide comprehensive situational awareness. Although models such as YOLOP, YOLOPv2, U-PDP, and HybridNet have progressed, achieving high precision in real-time multi-task processing remains a significant challenge.

In [12], the authors employed a model based on the YOLOPX algorithm. While this model achieved certain optimizations by integrating lane detection, object recognition, and drivable area segmentation, it still faces challenges. Specifically, the model utilizes a Path Aggregation Network (PAN) as the backbone of the encoder. Although PAN can handle features at different scales and adapt to complex scenarios, its multi-scale feature extraction and complex feature aggregation mechanisms significantly increase computational complexity and expand the model's parameter size.

As the latest version in the YOLO series, YOLOv8 incorporates several improvements, including a more efficient network architecture, optimized training processes, and better inference performance. Integrating it into YOLOPX is expected to enhance the model's accuracy further. Combining YOLOP's multi-task capabilities can augment the model's panoptic perception abilities in complex driving scenarios, improving the reliability and safety of ADS. Moreover, the loss functions from publicly available code, including classification, object, and regression loss, were integrated and assigned fixed hyperparameter weights in the experimental setup. This approach may limit the model's applicability across different scenarios and datasets.

This study aims to enhance the panoptic driving perception capabilities of the YOLOPX model in Autonomous Driving Systems (ADS) by optimizing its structure and training strategies. Specific goals include optimizing the encoder model structure to improve the accuracy of model outputs. This study also refines the loss functions in the decoder to achieve a modular design, allowing for the independent adjustment of weights for classification loss, object loss, and regression loss, ensuring accurate detection and effectiveness of the model under various conditions.

This study reviews extensive literature and explores various panoptic driving perception technologies to deepen understanding. It begins with traditional methods and details the development of tasks and integrated processing techniques within panoptic perception technology. The focus is on the model design of the panoptic driving perception system, highlighting contributions and limitations of models like TwinLiteNet, HybridNets, YOLOP, YOLOPv2, and YOLOPX. These studies discuss innovations in neural network architectures, performance metrics, and challenges related to computational demands. The comparison of different algorithms is shown in Table I.

TABLE I
COMPARE DIFFERENT ALGORITHMS OF THE PANOPTIC DRIVING PERCEPTION MODEL.

Ref.	Description and Contribution	Future Research
[8]	This study presents an innovative object detection approach by integrating a Region Proposal Network (RPN) with Fast R-CNN, making detection near real-time. Key contributions include: 1) a cost-effective RPN sharing features with the detection network, 2) integration of RPN and Fast R-CNN into a single efficient network, and 3) achieving state-of-the-art accuracy and high processing speed on multiple datasets.	1). Performance depends on input data quality. 2). Training and fine-tuning complexities for specific applications. 3). Optimization needed for speed-accuracy balance in diverse real-world scenarios.
[12]	This study presents an anchor-free multi-task learning network for panoptic driving perception, integrating object detection, drivable area segmentation, and lane detection. Key contributions include: 1) simplifying training with an anchor-free approach, 2) a novel lane detection head using multi-scale features, and 3) achieving state-of-the-art performance on the BDD100K dataset with higher recall, mAP50, and mIoU.	1). The model's increased complexity and parameter count. 2). The need for further validation of the model's generalizability. 3). The potential for high computational demand, which may limit its deployment on resource-constrained systems.
[13]	This study introduces a novel panoptic driving perception network combining traffic object detection, drivable area segmentation, and lane detection into a single model. Key contributions include: 1) a unified architecture for multiple tasks, 2) high performance on the BDD100K dataset, 3) real-time processing on embedded devices, and 4) ablative studies validating the multi-task approach's effectiveness.	1) Increased complexity and computational demands. 2) Limited exploration of model adaptability to diverse datasets. 3) Balancing accuracy across different tasks.
[14]	This study modifies YOLOv3 for better scene understanding in autonomous driving by integrating object detection with semantic and instance segmentation. Key contributions are: 1) adapting YOLOv3 for panoptic segmentation, 2) implementing dual segmentation heads for comprehensive scene analysis, and 3) achieving real-time performance, crucial for autonomous driving.	1). Increased complexity from the enhanced model structure. 2). Higher computational demands may affect deployment in resource-limited settings. 3). Need to balance accuracy and speed in challenging environments.
[15]	"HybridNets" presents an end-to-end perception network for autonomous driving, efficiently handling multiple tasks like traffic object detection, drivable area segmentation, and lane detection using a weighted bidirectional feature network. The study's main contributions include: 1). An innovative architecture that effectively fuses features for multi-tasking. 2). Customized anchor boxes for improved object detection accuracy. 3). A balanced training strategy and efficient loss function for optimizing network performance.	1) Multitask network complexity impacts computational resources and real-time application. 2) Limited effectiveness across diverse datasets or real-world scenarios. 3) Balancing accuracy and efficiency across tasks is challenging.

Ref.	Description and Contribution	Future Research
	4). High performance on the BDD100K dataset, demonstrating both accuracy and efficiency.	
[16]	"YOLOPv2" enhances the YOLOP network for improved panoptic driving perception, focusing on traffic object detection, drivable area segmentation, and lane detection with better accuracy and speed. Key contributions include a refined multi-task model, advanced data preprocessing, a novel hybrid loss function, increased FPS, and strong generalization capabilities for various scenarios.	1) Increased complexity from enhanced model structure. 2) Higher computational demands, affecting deployment in resource-limited settings. 3) Need to balance accuracy and speed in challenging environments.
[17]	Introducing a neural network for simultaneous detection of drivable areas, lane lines, and traffic objects, optimizing multi-task learning in autonomous vehicles. Contributions include: 1) An integrated framework for concurrent detection of key navigation elements. 2) A context tensor for information sharing between decoders. 3) Improved accuracy and efficiency on benchmark datasets.	1). Occasional misclassification issues, such as incorrectly identifying non-drivable areas. 2). Challenges with detecting intermittent lane markings, leading to potential false negatives.
[18]	Introducing a lightweight neural network for real-time drivable area and lane segmentation, optimized for embedded systems in self-driving cars. Key contributions: 1) Efficient dual-decoder network structure. 2) Integration of dual attention modules for enhanced feature fusion. 3) Achieves high-speed segmentation with a mIoU of 91.3%.	1). Slightly lower accuracy compared to the very latest models. 2). Lane detection performance that, while competitive, does not top all benchmarks. 3). Limited validation across different datasets or diverse real-world driving conditions.
[19]	Introduces a unified multi-task framework for panoptic driving perception, handling vehicle detection, lane detection, and drivable area segmentation. Key contributions: 1) Integrates multiple perception tasks. 2) Utilizes dynamic convolution kernels for efficient feature processing. 3) Achieves high accuracy on BDD100K dataset. 4) Suitable for real-time autonomous driving with fast processing.	1. Complexity in implementing a unified multi-task framework. 2. Performance validation mainly on BDD100K, limited exploration on other datasets. 3. Need for further investigation into model's generalization and scalability.
[20]	Enhance object detection by optimizing the architecture for accuracy and efficiency. Key contributions: 1) Optimized feature extraction with deeper networks for precision. 2) Enhanced data augmentation and preprocessing. 3) Improved multi-scale object detection capabilities.	1) Single-object detection limits multi-task applicability. 2) High computational needs restrict deployment of resource-limited devices. 3) Integrating multi-task models enhances comprehensive driving perception systems.

YOLOv8, the latest model in the YOLO series [8], [21], [22], [23] achieves more efficient detection through its advanced architecture and optimized training processes [20]. However, despite its impressive performance in single-object detection, YOLOv8 may not fully meet the complex needs of panoptic driving perception systems, which require simultaneous multi-task processing. This limitation is particularly evident in scenarios where real-time detection and processing of multiple tasks—such as object detection, lane detection, and drivable area segmentation—are critical. Thus, while YOLOv8 excels in accuracy, its focus on single-object tasks limits its direct applicability in comprehensive ADS environments.

YOLOP, designed to handle multiple sensing tasks simultaneously, attempts to increase computational efficiency and real-time capability by executing multiple tasks in a single forward pass. This capability allows YOLOP to streamline the multi-task process and balance efficiency and performance [13]. However, YOLOP struggles with the accuracy of small object detection and maintaining high precision in dynamic or complex driving scenarios. The updated YOLOPv2 enhances YOLOP's accuracy and speed by incorporating deeper feature extraction networks, improved training methods, and successful multi-task learning algorithms [16]. Despite these improvements, YOLOPv2 continues to face challenges in balancing the complexity of its multi-task architecture with the need for real-time performance. The increased network depth and the associated computational demands can hinder its application in scenarios requiring rapid decision-making.

U-PDP introduces a novel approach by integrating vehicle detection with future path prediction, allowing ADS to better adapt to dynamic environments [19]. While this integration of detection and predictive capabilities offers significant advantages in decision-making and path planning, it also introduces substantial computational challenges. The complexity of combining real-time detection with behavior prediction makes U-PDP less feasible for systems with stringent real-time performance requirements.

HybridNet employs a weighted bi-directional feature network to process multiple tasks, such as traffic object detection, drivable area segmentation, and lane detection [15], [16]. Although HybridNet's architecture effectively combines these tasks, its complexity can impact real-time applications due to the high processing demands required to maintain performance across all tasks.

Given these challenges, the specific limitations of YOLOP, YOLOPv2, U-PDP, and HybridNet underline the necessity of a more streamlined solution that balances multi-task processing with real-time performance. YOLOPX offers such a solution by utilizing an anchor-free multi-task learning strategy that combines object detection, drivable area segmentation, and lane detection within a single model. By employing a shared CNN backbone for initial feature extraction, YOLOPX reduces training time and enhances overall performance. This design effectively manages computational resources, overcoming many limitations in other models. Despite these advantages, YOLOPX [12] is not without its challenges. The model's complexity requires substantial processing resources, which may limit its deployment on resource-constrained devices. Additionally, the training process for multi-task learning is intricate, necessitating extensive tuning and debugging to achieve optimal performance across various tasks. Furthermore, YOLOPX's generalizability can be constrained by the diversity and quality of the training data, leading to

potential misclassifications or detection errors in complex or crowded environments. Small object detection, in particular, remains a challenge when image resolution is low, or object features are minimal.

To address these issues, this paper proposes an enhanced version of YOLOPX that integrates YOLOP's multi-task capabilities with YOLOv8's accuracy. This improved model will leverage anchor-free detection techniques, enabling simultaneous execution of tasks like object detection, drivable area segmentation, and lane detection while maintaining a consistent loss function to minimize inference time and boost real-time processing. The proposed enhancements aim to increase YOLOPX's accuracy, making it more suitable for panoptic driving perception and improving the safety and efficiency of ADS.

II. Materials and Method

This study utilized the BDD100K dataset, which is publicly available at the University of California, Berkeley. This dataset comprises 100,000 images encompassing six weather conditions: clear, cloudy, overcast, rainy, snowy, and foggy. It also includes six scene types: residential, highway, urban streets, parking lot, gas station, and tunnel, as well as three periods: dawn/dusk, daytime, and nighttime.

Inspired by YOLOP, YOLOPX, and YOLOv8, this study proposes an improved YOLOPX model that integrates YOLOP's multi-task detection, YOLOv8's CNN backbone, and YOLOPX's anchorless detection. The model uses an encoder-decoder network with a shared encoder and three task-specific decoders for traffic object detection, drivable area segmentation, and lane detection. A modular loss function handles segmentation tasks, enhancing flexibility, adaptability, and generalization while reducing inference time. This approach significantly increases processing capacity and accuracy, making the model highly suitable for panoptic driving perception systems. Figure 1 illustrates the model structure.

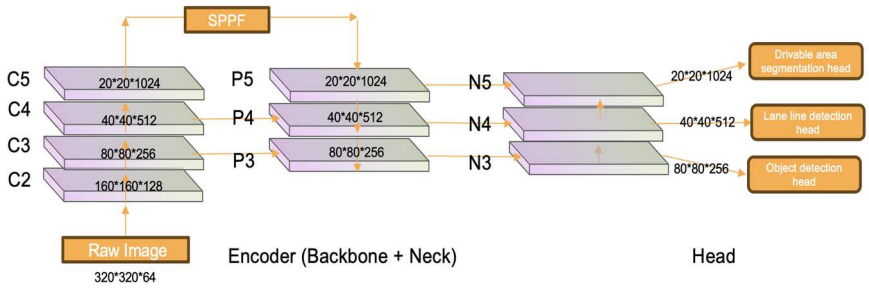

Fig. 1 Architecture of Improved YOLOPX Model

Figure 1 displays the architecture of the improved YOLOPX network for panoptic driving perception. Initially, pre-processed data is input through the encoder-decoder structure of the improved model. The data first reaches the encoder's backbone network, where it undergoes grouped convolution and SPPF (Spatial Pyramid Pooling in Fast R-CNN) layers for initial feature extraction from the input images. These features are then processed and refined in the neck using a Feature Pyramid Network (FPN). Subsequently, the data is passed to the decoder, which is divided into two types of detection heads: an object detection head and a drivable area segmentation head. The heads constitute the final part of the model, responsible for producing final outputs such as class probabilities, object bounding boxes, and confidence. The heads employ the same loss methodology to compute results.

Compared to YOLOPX, the encoder structure in this study essentially retains its backbone and neck configurations but incorporates certain modifications. Specifically, the backbone has been enhanced by replacing the SPP module with the SPPF module. The C3 structure has been substituted in the neck with the C2F structure. This study has preserved the detection head structure utilized in YOLOPX regarding the decoder module. However, this study has optimized the loss function in this research to achieve modularization, allowing the hyperparameter weights for classification loss, object loss, and regression loss to be adjusted independently. This ensures that the model can be broadly applied to various scenarios.

In summary, the encoder in the YOLOPX design is a complex and crucial component, primarily utilizing a shared backbone network and a neck with both bottom-up and top-down structures, while the decoder utilizes three separate detection heads to perform three distinct tasks. The encoder-decoder structure provides a foundation for the network's performance across various tasks. By combining advanced convolutional techniques, architectural developments, and efficiency optimizations, the improved model offers real-time, accurate environmental perception for automated systems.

A. Encoder

The encoder primarily consists of the backbone and neck. The backbone first extracts multi-scale features from raw images, capturing low-level and high-level information. Subsequently, the neck, typically implemented as a feature pyramid or path aggregation network, refines and integrates these features, enhancing the model's capability to detect objects and segment areas across varying resolutions. This interaction ensures the coordination of features at different scales, which is crucial for lane detection tasks that require fine details and broader contextual understanding. The model balances detection accuracy by optimizing the information flow from the backbone to the neck and task-specific heads while reducing the required parameters and computational cost, making it suitable for ADS systems. In our improved model backbone, the SPPF instead of the SPP module was used in YOLOPX. In the neck, use C2F to change the C3 module. The SPPF part can reduce computational cost to improve efficiency

and reduce redundancy and be implemented with fewer layers and operations to improve speed.

B. Decoder

The role of the decoder is to translate the features extracted by the encoder into actual detection outputs, such as bounding boxes, class labels, and performance metric scores. This process is crucial for mapping deep network features to specific detection tasks. The decoder typically resides at the end of the YOLO architecture, following the encoder for feature extraction and possibly the feature fusion part (such as a feature pyramid network). The decoder expected to be employed in this model is responsible for interpreting the high-dimensional feature maps extracted by the encoder and converting them into outputs usable for the specific task. The decoder consists of two detection heads: the detection head and the segmentation head. This paper retains the structure of the YOLOPX head. The details of the architecture of this study model are shown in Figure 2.

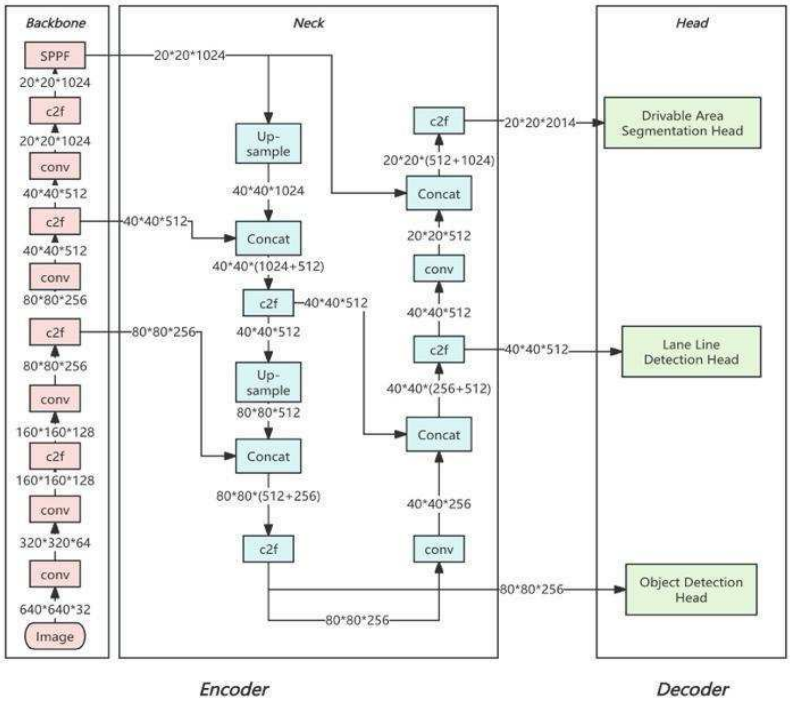

Fig. 2 Encoder-Decoder Architecture.

C. Loss Function

The loss function is a key component in fine-tuning the training process. It quantifies network performance by calculating the discrepancy between the network's predictions and the actual label data. In developing panoptic driving perception models, managing multiple tasks simultaneously, each with its own success metrics, is essential. The loss function plays a crucial role in model training, providing a metric that demonstrates how closely the model's predictions align with actual data. In panoptic driving perception, the model must reliably interpret visual inputs, and the loss function is critical in ensuring that the network learns to generate reliable and accurate predictions that can be used for ADS navigation and obstacle avoidance. This study will utilize the same loss function across all segmentation tasks to improve the model's generalizability.

The loss function strategy adopted by YOLOPX. The loss function strategy in this study wants to address and enhance the weaknesses of the YOLOPX loss function techniques. We propose a new optimization approach:

1) Code Organization and Modularization: We have redesigned the codebase with clearly defined different loss functions, each class focusing on specific functionality, such as handling class imbalance, calculating the IoU of bounding boxes, or computing key point losses. This modular approach allows each component to be used and tested independently, making the code easier to maintain and read.

2) Documentation Clarity and Usability: Each loss function class is accompanied by detailed documentation comments and examples explaining how to use them. This method enables other developers to understand and deploy these loss algorithms more quickly.

3) Applicability: The loss functions to cater to various simple and advanced applications. This approach is instrumental in determining the most suitable loss function for various tasks.

4) Flexibility and Scalability: Each loss function is designed as an independent module, which can be easily added to or modified within the existing architecture. This enhances flexibility and scalability, allowing for future adjustments or enhancements.

These improvements aim to provide more robust and adaptable solutions by addressing the deficiencies of existing loss functions. Thus, they enhance the overall effectiveness and practicality of the loss function strategy and make it suitable for a variety of tasks and scenarios. The composite

loss function consists of three parts: the object detection loss L_{det}, the drivable area segmentation loss L_{seg}, and the lane detection loss L_{detll}. Equation 1 shows the specific formulas:

$$L_{mtl} = L_{det} + L_{seg} + L_{detll} \quad (1)$$

The formula for object detection loss is as shown in Equation 2:

$$L_{det} = \lambda_{BCE} L_{BCE} + \lambda_{DFL} L_{DFL} + \lambda_{IoU} L_{IoU} \quad (2)$$

The category loss isL_{BCE}, corresponding to the 'FocalLossV1' module. A standard loss function calculates the difference between the model's predicted classification probabilities and the actual binary labels. In object detection tasks, it is used to determine whether an anchor contains a target, as shown in Equation 3:

$$L_{BCE} = -[y_n log x_n + (1 - y_n)\, log\, log\, (1 - x_n)\,] \quad (3)$$

The bounding box regression loss L_{DFL}, is used to measure the deviation between the predicted bounding box and the actual bounding box, as shown in Equation 4:

$$L_{DFL}(S_i, S_{i+1}) = -((y_{i+1} - y)\, log\, log\, (S_i) + (y - y_i)\, log\, log\, (S_{i+1})) \quad (4)$$

Among them, S_i, $S_{i+1}$are as shown in Equations 5 and 6:

$$S_i = \frac{y_{i+1} - y}{y_{i+1} - y_i} \quad (5)$$

$$S_{i+1} = \frac{y_i - y}{y_i - y_{i+1}} \quad 6)$$

In this, y represents the actual coordinates of the bounding box, y_i and y_{i+1} represent the two adjacent coordinate values (upper and lower limits) in the predicted bounding box coordinate distribution. This distributed representation method allows the model to predict a continuous distribution of bounding box positions, rather than a fixed point. S_i and S_{i+1} represent the actual bounding box coordinates. Finding the relative position in the prediction distribution can be imagined as looking for two prediction points in the prediction distribution that are closest to the actual coordinate y and calculating the relative position of y based on these two points so that the loss function can optimize this relative position. This method helps to capture the uncertainty of the bounding box.

The bounding box IoU loss L_{IoU}, considers the overlap, distance, aspect ratio, and centroid deviation between the predicted and actual boxes, as shown in Equations 7-10:

$$L_{IoU} = 1 - CIoU \quad (7)$$

$$CIoU = IoU - \frac{\rho^2\ (b, b^{gt})}{c^2} - \alpha v \quad (8)$$

$$v = \frac{4}{\pi^2}\left(\frac{arctanarctan\ \omega^{gt}}{h^{gt}} - \frac{arctanarctan\ \omega}{h}\right)^2 \quad (9)$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (10)$$

CIoU stands for Complete Intersection over Union, which not only considers the IoU but also the distance between the centers of the boxes and the aspect ratio. Here, b represents the center of the predicted bounding box, and b^{gt} represents the center of the actual bounding box. $P^2\ \ (b, b^{gt})$ denotes the Euclidean distance between the two centers. IoU is the square of the length of the diagonal of the minimum bounding rectangle that encloses both the predicted and the actual boxes. ω and h represent the width and height, respectively. α is used to balance the loss due to aspect ratio and v adjusts the consistency of the aspect ratios between the predicted and the actual bounding boxes. If the aspect ratios are the same, v will be zero; as the inconsistency increases, v will also increase. α ensures that when the IoU is already high, v contributes minimally to the overall loss, but increases its influence when IoU is low, thereby encouraging the model to predict more accurate aspect ratios. The design intention is to improve the quality of the model's predicted bounding boxes, making them more aligned with the actual boxes in both position and shape.

The loss function used for object detection will be applied to object and lane detection, offering multiple advantages. It integrates various aspects, such as classification, localization, and shape matching, to enhance detection performance. The loss function used for the segmentation of drivable areas is L_{seg}, and its specific formula is shown from Equations 11 to 13.

$$L_{seg} = \lambda_{FL} L_{FL} + \lambda_{TL} L_{TL} \quad (11)$$

$$L_{FL} = -\alpha_t (1 - p_t)^{\gamma}\, log\, log\, (p_{ct}) \quad (12)$$

$$L_{TL} = 1 - \frac{TP}{TP + \alpha FN + \beta FP} \quad (13)$$

Translation: The λ_{FL} is the weight for the Cross-Entropy Loss, and λ_{TL} is the weight for the Tversky Loss. L_{FL} is the cross-entropy loss used to reduce class imbalance, measuring the difference between the probability distribution predicted by the model and the actual label distribution. L_{TL} is the Tversky Loss portion, which can enhance the model's ability to recognize smaller target categories. α_t is the sample weight, which can be adjusted according to class imbalance. p_t is the probability predicted by the model, which predicts the probability that a given pixel belongs to a certain class. γ represents the true label, usually a binary value where 1 indicates the pixel belongs to a specific category and 0 indicates it does not. TP is the number of true positives, representing the number of pixels correctly predicted to belong to a specific category. FN is the number of false negatives, representing the number of pixels mistakenly predicted to other categories despite belonging to a specific category. FP is the number of false positives, representing the number of pixels incorrectly predicted to belong to a category when they do not. α and β are used to adjust the weights of false negatives and false positives in the Tversky Loss, reflecting the emphasis on these errors. This combination of segmentation losses can help the model better handle class imbalances in segmentation tasks and optimize pixel-level prediction performance. By adjusting the weights λ_{FL} and λ_{TL}, the contributions of cross-entropy loss and Tversky Loss to the total loss can be tailored according to the specific needs of the task. Such a design makes the model more accurate when predicting difficult or uncommon categories.

III. Result and Discussion

This section will describe the experimental process and test results for the panoptic driving perception model improved based on YOLOv8 and YOLOP. Additionally, this chapter will discuss the advantages demonstrated by the model and how the

experimental results meet the research objectives of this study.

A. Datasets

In this experiment, the dataset used is BDD100K, a large and diverse driving dataset created by researchers at the University of California, Berkeley [24]. This dataset comprises 100,000 video clips captured from various urban, suburban, and rural areas, encompassing a wide range of weather conditions and both daytime and nighttime scenes. The primary objective of the BDD100K dataset is to provide abundant training and testing resources for developing Autonomous Driving Systems (ADS) technologies.

In our experiments, the 100,000 images will be divided into a training set, a validation set, and a test set of 70,000, 20,000, and 10,000 photos, respectively. To evaluate the model's performance, this study will use recall and mAP50 for traffic object detection, mIoU for drivable area segmentation, pixel accuracy, and IoU to assess lane detection performance.

B. Parameter setup

Concerning the YOLOPX model experiment, the improved YOLOPX model was implemented on the Pytorch framework. The model was trained for 100 epochs using the SGD optimizer, with an initial learning rate of 0.01. Initially, the model underwent a 3-epoch warm-up training phase. During this warm-up phase, the momentum of the SGD optimizer was set to 0.8, and the learning rate for biases was 0.1. During the training process, a linear learning rate annealing strategy was adopted [25]. This strategy helps ensure that the model learns quickly in the early stages of training and converges more stably later. Additionally, the original image dimensions were resized from 1280x720 to 640x640, and training was conducted on an RTX 3090. A 3-cycle warm-up strategy was applied to the network to ensure stable training. Cosine annealing adjusted the learning rate, and the momentum was set to 0.937. Notably, the model did not use a pre-trained model for fine-tuning. The confidence threshold was set to 0.25 during prediction, and the NMS (Non-Maximum Suppression) threshold was set to 0.45. The confidence and NMS settings followed those of YOLOP.

C. Results

In this experiment, the results are primarily based on two parts: the first part is the final results obtained after the training of the model; the second part compares the results of the YOLOP model, the YOLOPX model, and our model, which trained on the same dataset in the different weather conditions. The training phase is shown in Figure 3.

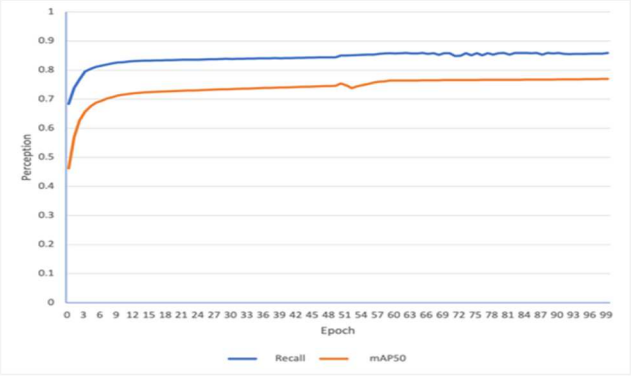

Fig. 3 Experiment process

The consistent performance in the later stages of training indicates that the model has achieved stable convergence, validating the effectiveness of the learning rate annealing and warm-up strategies employed during training. The result of this experiment is shown in Table II:

TABLE II
COMPARISON OF EXPERIMENTAL RESULTS ON LANE LINE DETECTION.

Method	Pixel Accuracy (%)	IoU (%)
YOLOP	70.5	26.2
YOLOPX	88.6	27.2
Ours	98.8	27.6

This data shows that our model surpasses the other two methods in pixel accuracy, nearly reaching 99%, meaning that it closely matches the annotations at the pixel level. YOLOPX has much higher pixel accuracy than YOLOP, and YOLOP shows the weakest performance on both metrics. This indicates that the method used in our model may be more accurate in distinguishing between lane and non-lane pixels, and its IoU metric is higher than the other two methods, thus effectively identifying lane areas and boundaries.

However, when applying these results to real-world scenarios, it's crucial to recognize the model's potential limitations in different driving contexts. Variations in driving conditions, such as lighting changes, worn road markings, adverse weather, and obstructions, may negatively affect the model's accuracy. While the model performs well on the BDD100K dataset, its reliability in unpredictable real-world situations needs further validation to ensure consistent performance in practical applications.

D. Comparison Analysis

This study conducted comparative experiments in different weather conditions on three lane line detection models: YOLOP, YOLOPX, and our model. The experiment aimed to evaluate each model's performance and robustness under four different environmental conditions: straightforward, night, rainy, and snowy.

1) Comparison of model performance during clear weather: In Figure 4, our model excels in object detection, particularly in complex scenes and for distant targets. Compared to YOLOP and YOLOPX, our model provides more accurate bounding box positioning and sizing, significantly reducing missed detections and false positives. In lane line detection, our model performs exceptionally well, offering more accurate and continuous lane line recognition. Our model delivers more stable detection results in complex road environments without interruptions or recognition errors, surpassing YOLOP and YOLOPX.

Our model also stands out in drivable area segmentation, with highly accurate segmentation areas and clear boundaries. Even in complex scenarios, the segmentation remains stable, with no blurred boundaries or incorrect segmentation, clearly outperforming YOLOP and YOLOPX. Our model shows significant advantages in object detection, lane line detection, and drivable area segmentation in clear weather conditions, providing more accurate and stable detection results.

Fig. 4 Clear-day comparison of algorithms.

2) Comparison of model performance during the nights: In Figure 5, nighttime weather, our model demonstrates significant advantages in object detection, lane line detection, and drivable area segmentation. Compared to YOLOP and YOLOPX, our model can more accurately detect vehicles and pedestrians under low light conditions, with more precise bounding box positioning and sizing, reducing missed detections and false positives. Our model offers more accurate and continuous recognition in lane line detection, providing stable detection results even in complex nighttime road environments without interruptions or recognition errors.

Fig. 5 Night Day comparison of algorithm

In drivable area segmentation, our model delivers accurate segmentation areas with clear boundaries, maintaining stability even under poor lighting conditions and outperforming other models. These advantages make our model more suitable for practical applications in panoramic driving perception systems, significantly enhancing the safety and reliability of nighttime autonomous driving.

3) Comparison of model performance during rains: Figure 6 shows that in rainy weather, our model shows significant advantages in object detection, lane line detection, and drivable area segmentation. Compared to YOLOP and YOLOPX, our model can still accurately detect vehicles and pedestrians under poor lighting and obstructed visibility, with more precise bounding box positioning and sizing, reducing missed detections and false positives.

Our model offers more accurate and continuous recognition in lane line detection, providing stable detection results even on wet and reflective roads without interruptions or recognition errors. In drivable area segmentation, our model delivers accurate segmentation areas with clear boundaries, maintaining stability even under blurred visibility due to rain, with no blurred boundaries or incorrect segmentation. These advantages make our model more suitable for practical applications in panoramic driving perception systems, significantly enhancing the safety and reliability of autonomous driving in rainy weather.

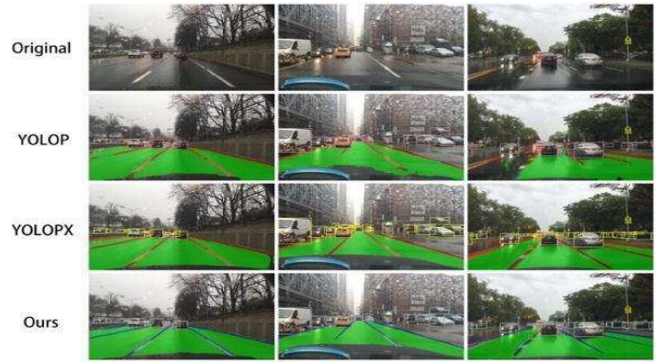

Fig. 6 Rain Day comparison of algorithms.

4) Comparison of model performance during the snow: Figure 7 shows that in snowy weather, our model shows significant advantages in object detection, lane line detection, and drivable area segmentation. Compared to YOLOP and YOLOPX, our model can more accurately detect vehicles and pedestrians, especially under obstructed visibility and strong light reflections, with more precise bounding box positioning and sizing, reducing missed detections and false positives.

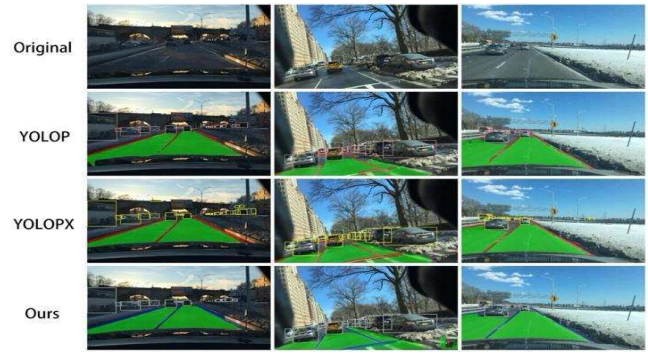

Fig. 7 Snow Day comparison of algorithms

Our model offers more accurate and continuous recognition in lane line detection, providing stable detection results even on snow-covered roads without interruptions or recognition errors. In drivable area segmentation, our model delivers accurate segmentation areas with clear boundaries, maintaining stability even on snow-covered roads. These advantages make our model more suitable for practical applications in panoramic driving perception systems, significantly enhancing the safety and reliability of autonomous driving in snowy weather.

In conclusion, our model performs well and remains robust in various weather situations, particularly regarding lane border precision and area segmentation. This performance indicates that our model employs robust feature extraction and image processing algorithms capable of handling complex and variable environmental conditions. Our model's strong noise suppression and image enhancement abilities ensure stable lane detection capabilities in low light and adverse weather conditions.

IV. Conclusion

This study integrates advancements from YOLOv8 and YOLOP to enhance the YOLOPX model for autonomous driving systems (ADS), addressing key challenges in panoramic driving perception. This study contributed to the improved backbone and neck structure by replacing the SPP module with the more efficient SPPF module and the C3 module with the C2F module to improve multi-scale feature extraction and reduce computational load while retaining the anchor-free detection head of YOLOPX for accurate, efficient detection. The modular loss function as a new modular loss function allows independent adjustment of hyperparameters for classification, object, and regression losses, improving adaptability, generalization, and inference time for different tasks and environments. The extensive experimental validation has contributed to the experiments on the BDD100K dataset, showing significant improvements, with a lane line detection accuracy of 98.8%, a segmentation of the drivable area (mIoU) of 90.4%, and a recognition of traffic objects of 85.9% and mAP50 of 76.9%.

This study combines the latest advancements from YOLOv8 and YOLOP to enhance the YOLOPX model comprehensively. The resulting model exhibits exceptional performance, flexibility, and real-time processing capabilities, contributing valuable advancements to developing robust and efficient panoramic driving perception systems for autonomous driving. Based on the achievements of the improved YOLOPX model in deep learning, future work can focus on multi-sensor fusion, which integrates data from radar, LiDAR, and cameras to enhance perception capabilities, thereby improving the model's accuracy and reliability in complex environments.

Besides, adaptive learning mechanisms should be developed to enable automatic parameter adjustments, allowing systems to respond effectively to changing road conditions, such as fluctuating weather and varying traffic density. Real-time map updates, powered by the improved YOLOPX model, can provide dynamic navigation assistance and timely road obstacle warnings during transportation. Cross-domain validation is essential to test the model across diverse driving scenarios, ensuring its generalizability and practical application. Robustness and safety are critical priorities, necessitating rigorous testing and emergency handling procedures to safeguard the system against sensor failures. Furthermore, improving human-machine interaction by designing intuitive interfaces can help users better understand the vehicle's perception system, thereby fostering trust and confidence in autonomous technologies. Future work in these areas will advance the development of panoptic perception models for driving and thus increase the driving safety of vehicles equipped with this system.

References

[1] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollar, "Panoptic Segmentation," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2019, doi:10.1109/cvpr.2019.00963.

[2] A. Petrovai and S. Nedevschi, "Multi-task Network for Panoptic Segmentation in Automated Driving," 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp. 2394–2401, Oct. 2019, doi: 10.1109/itsc.2019.8917422.

[3] W. K. Fong et al., "Panoptic Nuscenes: A Large-Scale Benchmark for LiDAR Panoptic Segmentation and Tracking," IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 3795–3802, Apr. 2022, doi:10.1109/lra.2022.3148457.

[4] A. Milioto, J. Behley, C. McCool, and C. Stachniss, "LiDAR Panoptic Segmentation for Autonomous Driving," 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 8505–8512, Oct. 2020, doi: 10.1109/iros45743.2020.9340837.

[5] G. Xian et al., "Location-Guided LiDAR-Based Panoptic Segmentation for Autonomous Driving," IEEE Transactions on Intelligent Vehicles, vol. 8, no. 2, pp. 1473–1483, Feb. 2023, doi:10.1109/tiv.2022.3195426.

[6] S. D. Yashwanth, S. V. Rao, Rakshit, Y. P. Meharwade, and R. Kivade, "Autonomous Driving Using YOLOP," 2022 IEEE North Karnataka Subsection Flagship International Conference (NKCon), pp. 1–6, Nov. 2022, doi: 10.1109/nkcon56289.2022.10127088.

[7] X. Dai, "HybridNet: A fast vehicle detection system for autonomous driving," Signal Processing: Image Communication, vol. 70, pp. 79–88, Feb. 2019, doi: 10.1016/j.image.2018.09.002.

[8] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/tpami.2016.2577031.

[9] F. N. Ortatas and E. Cetin, "Lane Tracking with Deep Learning: Mask RCNN and Faster RCNN," 2022 Innovations in Intelligent Systems and Applications Conference (ASYU), pp. 1–5, Sep. 2022, doi:10.1109/asyu56188.2022.9925296.

[10] X. Kong, X. Li, X. Zhu, Z. Guo, and L. Zeng, "Detection model based on improved faster-RCNN in apple orchard environment," Intelligent Systems with Applications, vol. 21, p. 200325, Mar. 2024, doi:10.1016/j.iswa.2024.200325.

[11] D. Jiang, G. Li, C. Tan, L. Huang, Y. Sun, and J. Kong, "Semantic segmentation for multiscale target based on object recognition using the improved Faster-RCNN model," Future Generation Computer Systems, vol. 123, pp. 94–104, Oct. 2021, doi:10.1016/j.future.2021.04.019.

[12] J. Zhan, Y. Luo, C. Guo, Y. Wu, J. Meng, and J. Liu, "YOLOPX: Anchor-free multi-task learning network for panoptic driving perception," Pattern Recognition, vol. 148, p. 110152, Apr. 2024, doi:10.1016/j.patcog.2023.110152.

[13] D. Wu et al., "YOLOP: You Only Look Once for Panoptic Driving Perception," Machine Intelligence Research, vol. 19, no. 6, pp. 550–562, Nov. 2022, doi: 10.1007/s11633-022-1339-y.

[14] M. Diaz-Zapata, Ö. Erkent, and C. Laugier, "YOLO-based Panoptic Segmentation Network," 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC), pp. 1230–1234, Jul. 2021, doi: 10.1109/compsac51774.2021.00170.

[15] V.T Dat, Nvh Bao, and P.D Hung. Hybridnets: End-to-end perception network. arXiv:2203.09035

[16] C. Han, Q. Zhao, S. Zhang, Y. Chen, Z. Zhang, and J. Yuan, "YOLOPv2: Better, faster, stronger for panoptic driving perception," Aug. 2022.

[17] Y. Qian, J. M. Dolan, and M. Yang, "DLT-Net: Joint Detection of Drivable Areas, Lane Lines, and Traffic Objects," IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 11, pp. 4670–4679, Nov. 2020, doi: 10.1109/tits.2019.2943777.

[18] Q.-H. Che, D.-P. Nguyen, M.-Q. Pham, and D.-K. Lam, "TwinLiteNet: An Efficient and Lightweight Model for Driveable Area and Lane Segmentation in Self-Driving Cars," 2023 International Conference on Multimedia Analysis and Pattern Recognition (MAPR), pp. 1–6, Oct. 2023, doi:10.1109/mapr59823.2023.10288646.

[19] H. Wang, M. Qiu, Y. Cai, L. Chen, and Y. Li, "Sparse U-PDP: A Unified Multi-Task Framework for Panoptic Driving Perception," IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 10, pp. 11308–11320, Oct. 2023, doi: 10.1109/tits.2023.3273286.

[20] M. Sohan, T. Sai Ram, and Ch. V. Rami Reddy, "A Review on YOLOv8 and Its Advancements," Data Intelligence and Cognitive Informatics, pp. 529–545, 2024, doi: 10.1007/978-981-99-7962-2_39.

[21] S. Ganapathy and D. Ajmera, "An Intelligent Video Surveillance System for Detecting the Vehicles on Road Using Refined YOLOV4," Computers and Electrical Engineering, vol. 113, p. 109036, Jan. 2024, doi: 10.1016/j.compeleceng.2023.109036.

[22] C. Liu, Y. Tao, J. Liang, K. Li, and Y. Chen, "Object Detection Based on YOLO Network," 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), Dec. 2018, doi:10.1109/itoec.2018.8740604.

[23] D. Balakrishnan, S. Manideep Kumar Reddy, R. Lakshmi Venkatesh, K. Aadith, R. Jebi Nalatharaj, and M. Arshath, “Object Detection on Traffic Data Using Yolo,” 2023 International Conference on Data Science and Network Security (ICDSNS), pp. 1–5, Jul. 2023, doi:10.1109/icdsns58469.2023.10245691.

[24] F. Yu et al., “BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning,” 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2633–2642, Jun. 2020, doi: 10.1109/cvpr42600.2020.00271

[25] J. Ruan, H. Cui, Y. Huang, T. Li, C. Wu, and K. Zhang, “A review of occluded objects detection in real complex scenarios for autonomous driving,” Green Energy and Intelligent Transportation, vol. 2, no. 3, p. 100092, Jun. 2023, doi: 10.1016/j.geits.2023.100092.