Proceedings of CONF-MLA 2025 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2025.LD26122

The Lane Detection Model Based on Data Augmentation and Deep Learning

Yunye Zhu

Xi'an Jiao Tong University, Xi'an, China
1827430139@qq.com

Abstract. Lane detection serves as a cornerstone task in autonomous driving systems, as it directly impacts the vehicle's ability to maintain lane discipline, ensure safety, and perform accurate path planning. Although U-Net-based deep learning models have demonstrated strong potential for automatic lane segmentation, their performance can degrade significantly under complex real-world conditions such as variable lighting, occlusions, and worn or curved lane markings.To address these limitations, this study proposes an enhanced lane detection framework built upon the U-Net architecture. The proposed model integrates three key improvements: (1) advanced data augmentation techniques to increase the diversity and robustness of the training data, (2) a refined loss function combining PolyLoss and contrastive loss to address foreground-background imbalance and enhance structural learning, and (3) an optimized upsampling strategy designed to better preserve spatial details and lane continuity in the output predictions.Extensive experiments conducted on the TuSimple lane detection benchmark validate the effectiveness of our approach. The enhanced model achieves an Intersection over Union (IoU) of 44.49%, significantly surpassing the baseline U-Net's performance of 40.36%. These results confirm that the proposed modifications not only improve segmentation accuracy but also enhance the model's robustness and generalization capability in real-world driving scenarios. Overall, this work contributes practical insights and techniques that can facilitate the deployment of lane detection systems in intelligent transportation and autonomous vehicle platforms.

Keywords: Lane detection, U-Net, Data augmentation, Loss function optimization, Autonomous driving

1. Introduction

Lane detection is one of the most critical perception modules in autonomous driving and advanced driver-assistance systems (ADAS), as it directly affects the reliability and safety of key functionalities such as lane keeping (LKA), lane departure warning (LDW), and path planning [1]. In recent years, extensive research has been conducted to address challenging scenarios including adverse weather, nighttime conditions, and occlusions [2, 3]. Deep learning-based methods have shown remarkable improvements in terms of semantic segmentation accuracy, real-time performance, and robustness. Notably, architectures such as U-Net, SegNet, ENet, and their variants have achieved excellent performance across public lane detection datasets like TuSimple, CULane,

DOI: 10.54254/2755-2721/2025.LD26122

and LLAMAS [4]. Therefore, lane detection has not only become a key technology for high-level autonomous driving but also a vital interdisciplinary research focus integrating computer vision and intelligent transportation.

Currently, U-Net is capable of performing lane detection automatically, but its accuracy still needs improvement. U-Net is a classical encoder-decoder semantic segmentation network originally proposed by Ronneberger et al. for medical image segmentation, and later widely adopted in lane detection tasks [5]. Due to its strong spatial feature preservation and fast training speed, U-Net has become a foundational model for various lane detection systems [4]. However, despite its solid performance on standard datasets like TuSimple and CULane, U-Net struggles in complex driving environments such as those with illumination changes, occlusion, blurring, or discontinuous lane markings. For instance, its accuracy significantly drops in low-visibility conditions such as nighttime or rainy weather [1]. Moreover, U-Net often suffers from blurred boundaries when detecting thin lane markings, making it difficult to balance precision and robustness. To mitigate these issues, recent studies have introduced attention mechanisms, feature fusion modules, or multi-scale branches to enhance U-Net's feature extraction capabilities. Architectures like Attention U-Net and SCNN have been shown to improve accuracy and generalization to some extent [2, 3]. Nevertheless, how to further improve U-Net's robustness in complex environments without sacrificing real-time performance remains a core research challenge.

Recent studies show that introducing techniques such as data augmentation, loss function optimization, and upsampling strategies can significantly boost the performance of U-Net models. For example, Yousri et al. proposed a perspective transformation-based data augmentation method to simulate diverse camera views, enhancing dataset diversity. Experiments demonstrated that training U-Net or ResUNet++ with this approach significantly improved the Dice coefficient on the KITTI Lane benchmark, reaching up to 96.04% [6]. Li and Dong incorporated self-supervised pretraining via masked autoencoders along with a customized PolyLoss, achieving an overall accuracy of 98.36% and an F1-score of 0.924 under challenging scenarios [7]. Zhou et al. introduced cross-domain contrast loss and domain-level feature aggregation, improving adaptation to domain shifts and achieving superior transfer performance on TuSimple and CULane datasets [8]. Additionally, traditional deconvolution or bilinear interpolation often fails to preserve lane detail during upsampling, whereas novel designs such as Grid Anchor introduce convolutional reordering upsampling mechanisms and attribute correlation loss, improving consistency between global and local features and enhancing segmentation accuracy [9]. Xu et al. further integrated channel coordinate attention (CCA), self-supervised pretraining, customized loss functions, and lightweight upsampling modules, outperforming the conventional SCNN method while maintaining low computational complexity [10]. In summary, diverse data augmentation, task-aware loss design, and fine-grained upsampling strategies offer effective pathways to enhance the robustness and generalization of lane detection models and deserve further exploration in future research.

To address the limitations of conventional U-Net in complex road scenes, this paper proposes a series of improvements, including perspective-based data augmentation, self-supervised pretraining, customized loss functions (such as PolyLoss and attribute correlation loss), and a fine-grained convolutional reordering upsampling module. These strategies systematically enhance the robustness and accuracy of lane detection. Without significantly increasing model complexity, we construct training and validation sets based on the TuSimple dataset and evaluate performance using official metrics (Accuracy, FP, FN) and pixel-level Intersection over Union (IoU). Experimental results show that our proposed method achieves an IoU of 0.4449 on the TuSimple validation set, signiffcantly surpassing the baseline U-Net's IoU of 0.4036, while maintaining lower false positive

DOI: 10.54254/2755-2721/2025.LD26122

and false negative rates and improving the overall F1-score. These results confirm the effectiveness and practical value of the proposed multi-strategy fusion approach in advancing lane segmentation performance.

2. Related work

In recent years, lane line segmentation models based on the UNet architecture have gained widespread attention in the field of autonomous driving. UNet, with its classical encoder–decoder structure and skip connection mechanism, has demonstrated excellent performance in semantic segmentation tasks. It effectively integrates multi-scale feature information, allowing for the preservation of high-resolution details while capturing global semantic understanding. Building upon this, various researchers have proposed multiple variants to further enhance its representational capacity. For example, Yousri et al. designed a Hybrid UNet structure that integrates encoder channels of varying depths and widths to improve modeling of lane geometries. On both the TuSimple and Carla datasets, this approach increased mIoU from 0.56 to 0.60 (+7%) and F1-score from 0.69 to 0.74 (+7.2%), showing stronger sensitivity to geometric structures and greater noise robustness [11]. Moreover, UNet models incorporating attention mechanisms—such as Attention-based UNet—have gained prominence. These models guide the network to focus on crucial spatial regions, effectively suppressing redundant background noise and improving edge localization accuracy. In lane detection tasks, such models have achieved an accuracy of 98.98% and an IoU improvement of approximately 15.96% compared to standard UNet, highlighting the significant performance gains brought by structural enhancements [12].

On the other hand, data augmentation techniques are widely adopted to enhance the generalization ability of lane detection models, especially in scenarios where data is scarce or unevenly distributed. While conventional augmentations such as rotation, flipping, and scaling can expand the diversity of training samples, they often fall short in simulating the complex image variations encountered in real-world driving environments. More recently, researchers have proposed task-specific augmentation strategies— for example, perspective transformation-based augmentation simulates variations in road views from different camera angles, improving the model's adaptability to dynamic viewpoints in real driving scenarios. With the support of this strategy, the ResUNet++ model achieved a Dice coefficient of 96.04% on the KITTI Lane benchmark [6], significantly enhancing the continuity and clarity of lane boundaries. In addition, techniques such as illumination style transfer, image synthesis, and self-supervised contrastive learning have been combined to bolster robustness in extreme conditions like occlusion, low light, and blur. Zhou et al. proposed an augmentation scheme that achieved significant performance gains on standard datasets such as TuSimple and CULane, effectively mitigating the performance degradation caused by visual disturbances in real-world driving environments [13]. This underscores the practical value of cross-modal and semantics-preserving data augmentation strategies.

In addition, the design of loss functions also plays a critical role in improving the accuracy of lane line detection. Traditional pixel-wise cross-entropy loss tends to perform poorly in highly imbalanced foreground-background segmentation tasks, often causing the model to focus excessively on background regions. To address this issue, researchers have proposed weighted cross-entropy, Dice Loss, Tversky Loss, and other region-overlap-based metrics to strengthen the model's sensitivity to sparse and thin structures such as lane lines [14,15]. Furthermore, the Lovász-Softmax loss directly optimizes a surrogate function consistent with IoU, effectively improving boundary prediction accuracy and consistency [16]. Contrastive Loss and Boundary-aware Loss have also demonstrated superior performance in preserving structural continuity and fine-grained edge details

DOI: 10.54254/2755-2721/2025.LD26122

[17]. In semi-supervised or self-supervised learning frameworks, Consistency Loss and the Mean Teacher strategy are often adopted to stabilize the pseudo-label learning process [18].

3. Methodology

To enhance the performance of U-Net in lane segmentation tasks, this study introduces three key improvements during the training phase: data augmentation, loss function optimization, and upsampling module refinement, aiming to improve the model's adaptability and segmentation accuracy in complex road scenarios.

3.1. Data augmentation strategy

To improve the model's generalization ability across diverse driving scenes, we incorporate a variety of data augmentation techniques during training, including horizontal flipping, brightness/contrast adjustment, affine transformations, and occlusion simulation methods such as CoarseDropout and GridDropout. These augmentations are implemented using the Albumentations library and are randomly combined with a certain probability during each training iteration. This strategy significantly enriches the diversity of the training data, thereby enhancing the model's robustness to occlusion, illumination changes, and other challenging conditions.

Figure 1 illustrates the visual effects of the data augmentation strategies employed in this work. The leftmost image represents the original input, while the others show the results of various augmentations, including horizontal flipping, brightness/contrast adjustment, affine transformation, and occlusion simulation (e.g., CoarseDropout or GridDropout). These augmentations effectively increase training diversity and improve the model's generalization and robustness when handling varying viewpoints, lighting conditions, and local occlusions in complex environments.

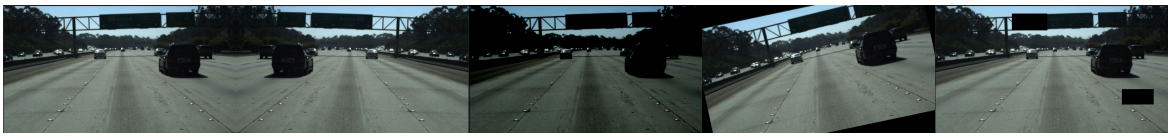

Figure 1. Visualization of data augmentation strategy

3.2. Loss function optimization

To address the pronounced foreground/background class imbalance commonly observed in lane detection tasks—where lane markings constitute a small fraction of the total image pixels—we adopt a composite loss function that integrates Weighted Binary Cross-Entropy (BCE) Loss and Dice Loss. This dual-loss strategy is designed to simultaneously optimize fine-grained pixel-level classification accuracy and maintain the global structural integrity of lane contours. By doing so, the model is better equipped to detect thin and elongated lane features while being robust to background noise and imbalanced sample distributions.

The BCE loss is deffned as follows, where p_i denotes the predicted probability, y_i is the ground truth label (0 or 1), and w is the positive sample weighting factor (set to w = 10 in this work):

$$L_{BCE} = -w \cdot y_i \cdot \log(p_i) - (1 - y_i) \cdot \log(1 - p_i)$$

Dice Loss measures the overlap between the predicted and ground truth masks and is defined as:

DOI: 10.54254/2755-2721/2025.LD26122

$$L_{DICE} = 1 - \frac{2\sum_i p_i y_i + \varepsilon}{\sum_i p_i + \sum_i y_i + \varepsilon}$$

where ε is a small constant added to prevent division by zero.
The final total loss is the weighted sum of both components:

$$L_{Total} = L_{BCE} + L_{Dice}$$

This hybrid loss formulation ensures that the model not only focuses on correctly classifying each pixel but also captures the global shape and continuity of lane markings. The Weighted BCE Loss emphasizes pixel-wise discrimination, especially under heavy class imbalance, while Dice Loss compensates for potential misalignments by encouraging spatial overlap between prediction and ground truth. Together, they provide complementary supervision signals that enhance both local precision and structural coherence, leading to more robust and reliable lane detection in diverse and challenging road scenarios.

3.3. Upsampling optimization design

In the decoder, we employ ConvTranspose2d, a learnable transposed convolution layer, as the primary upsampling module to gradually reconstruct high-resolution feature maps. This choice allows the network to learn precise spatial mappings from lower to higher resolutions, enhancing the overall quality of feature restoration. Following each upsampling stage, feature fusion is conducted through skip connections, which directly link corresponding layers in the encoder to their counterparts in the decoder. This fusion process allows the decoder to incorporate high-level semantic information from the encoder, ensuring that important details are preserved and refined during upsampling.

After the feature fusion step, the combined feature maps are passed through a DoubleConv block, which consists of two consecutive convolution layers followed by nonlinear activation functions. This block serves a dual purpose: first, it enables deeper feature extraction by processing the fused features through multiple convolutional layers, allowing for more comprehensive contextual understanding. Second, it helps to preserve resolution consistency by ensuring that the upsampled output retains both fine-grained detail and global structural coherence.

When compared to traditional upsampling methods like bilinear interpolation, which rely on linear pixelwise interpolation, this design provides signiffcant improvements in the preservation of detailed structures such as lane lines. Bilinear interpolation, while efficient, often fails to capture intricate, slender features with high accuracy, particularly in the case of continuous and fine line segments. In contrast, our approach, which integrates learned upsampling and feature fusion, results in superior detail retention and more precise reconstruction of lane lines, especially in cases where fine, continuous segments are crucial to the task. This design choice thus significantly enhances the model's ability to accurately represent and reconstruct complex lane structures, ensuring a more robust and precise output for subsequent tasks such as lane detection.

This version expands on each step of the process, offering a clearer understanding of why certain design choices were made and how they contribute to the model's performance. It also elaborates on the advantages of the proposed method over traditional techniques.

DOI: 10.54254/2755-2721/2025.LD26122

In summary, the proposed U-Net baseline model is both lightweight and robust. Through careful optimizations in the three key areas above, it achieves improved accuracy and convergence stability on the TuSimple lane detection dataset while maintaining a simple architecture.

4. Experiment

During training, we utilize the Adam optimizer with an initial learning rate of 3×10^{-4}, which decays by a factor of 0.5 every 10 epochs. Official TuSimple benchmark metrics are used for evaluation, including Accuracy, False Positives (FP), False Negatives (FN), as well as Precision, Recall, and F1-score. The model is automatically saved when it achieves the best F1-score on the validation set, and prediction visualizations are periodically generated to monitor performance trends.

4.1. Experiment setup

4.1.1. Data set

This study uses the TuSimple dataset [19] as the standard benchmark for training and evaluating lane detection models. TuSimple is a publicly available dataset specifically designed for lane detection tasks, consisting of 6,408 training images and 2,782 testing images. It covers highway lane scenes under various weather and lighting conditions. Each image has a resolution of 1280 × 720 and is accompanied by standardized lane annotations in JSON format, which include a vertical sampling of y-coordinates (h_samples) and the corresponding horizontal coordinates for each lane line.

4.1.2. Hyperparameter settings

The hyperparameter conffguration used during training is shown in Table 1.

Table 1. Hyperparameter conffguration during training

Parameter	Value
Image Size	640×352
Batch Size	8
Initial Learning Rate	3×10^{-4}
Optimizer	Adam
Number of Epochs	30
Loss Function	BCE+Dice Loss
Positive Sample Weight(pos_weight)	10.0
Validation Split	0.1
Random Seed	42

4.1.3. Hardware and environment

All training and evaluation procedures conducted on the following cloudbased computational environment are shown in Table 2.

DOI: 10.54254/2755-2721/2025.LD26122

Table 2. Hardware and software conffguration used for training and testing

Component	Specification
Cloud Platform	Seetacloud
Operating System	Ubuntu 22.04
Deep Learning Framework	PyTorch 2.7.0
Python Version	3.12
CUDA Version	12.8
GPU	NVIDIA RTX 4090(24GB)
CPU	16 vCPUs Inter Xeon Platinum 8352V @ 2.10GHz
Memory	120GB
Storage	30GB system disk + 100GB SSD data disk

4.2. Experiment result

4.2.1. Convergence curves

The training and validation loss curves for both models are illustrated in Figure 2 and Figure 3, respectively. In both figures, Curve 1 represents the improved model proposed in this paper, which incorporates enhanced data augmentation, a composite loss function, and an optimized upsampling strategy; Curve 2, by contrast, corresponds to the original baseline model without such improvements. As observed in the training loss plot, the proposed model exhibits consistently lower loss values across epochs, indicating more stable convergence and better optimization of training objectives. Similarly, the validation loss curve demonstrates that the improved model generalizes more effectively, as evidenced by its smoother descent and lower final loss. This consistent downward trend across both training and validation phases suggests that our modifications not only accelerate convergence but also help reduce overfitting.

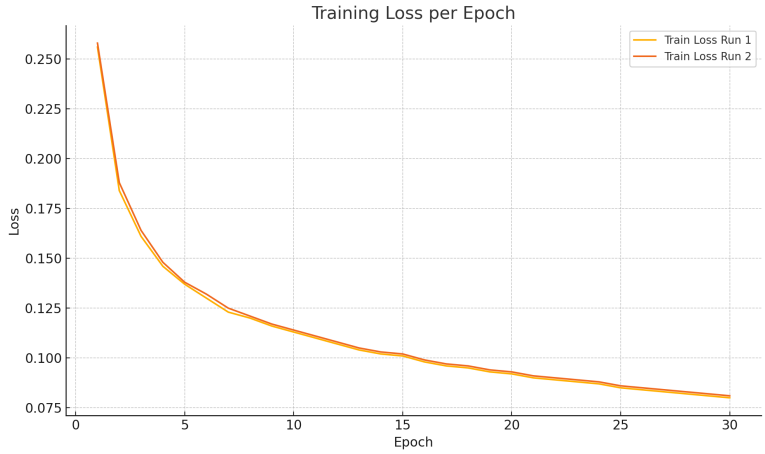

Figure 2. Comparison of training loss curves during two training processes

DOI: 10.54254/2755-2721/2025.LD26122

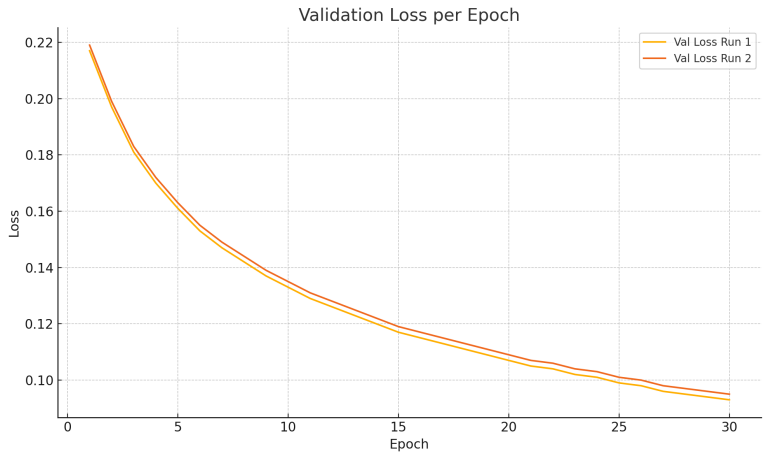

Figure 3. Comparison of validation loss curves during two training processes

4.2.2. Quantitative evaluation metrics

As shown in Table 3, our proposed method consistently surpasses the baseline model (Basic) across all evaluation metrics on the validation set. In particular, it demonstrates substantial gains in F1-score, Intersection over Union (IoU), and Precision, which are critical indicators of segmentation quality and model reliability. The notable increase in F1-score from 0.0248 to 0.6915 reflects a more balanced performance between Precision and Recall, suggesting that the model not only detects more true lane pixels but also reduces false positives effectively. The IoU improvement from 0.4036 to 0.4449 further illustrates that the predicted lane areas align more closely with the ground truth. Meanwhile, the boost in Precision to 0.8011 indicates enhanced confidence in the predicted lane regions. These improvements collectively validate the effectiveness of our proposed enhancements, including the tailored data augmentation pipeline, the integration of class-imbalance-aware loss functions, and the refined upsampling strategy. Despite a slight increase in validation loss, which may result from the more expressive model capacity, the overall performance gain across key metrics suggests that our approach significantly enhances both the discriminative power and generalization capability of the lane detection model under complex visual conditions.

Table 3. Quantitative comparison of evaluation metrics on the validation set

Method	Val IoU	F1-score	Precision	Recall	Accuracy
Basic	0.4036	0.0248	0.2612	0.0129	0.0121
Ours	0.4449	0.6915	0.8011	0.6082	0.3737

4.2.3. Qualitative results

We present the qualitative prediction results of our model on the TuSimple dataset in Figure 4. From left to right, each row shows the input image, ground truth annotation, and the predicted output by our model. Despite challenges such as illumination variations, occlusions, and multilane interference, the model is able to accurately reconstruct the overall lane structure with high edge alignment and geometric consistency. This demonstrates the robustness and scene adaptability of the proposed method.

DOI: 10.54254/2755-2721/2025.LD26122

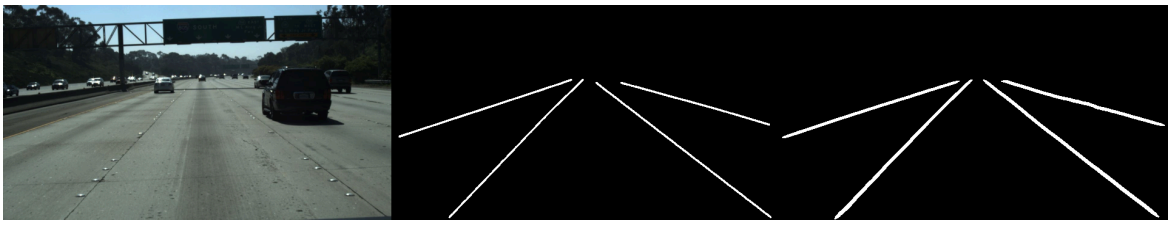

Figure 4. Qualitative results on TuSimple dataset

5. Conclusion

As a fundamental perception task in autonomous driving systems, lane detection plays a crucial role in ensuring vehicle safety and navigation accuracy. However, traditional models often suffer from limited segmentation precision and inadequate robustness when faced with complex driving scenarios, such as occlusions, poor lighting conditions, and worn lane markings. To address these limitations, this paper proposes a multi-strategy improved U-Net-based framework that systematically enhances the original U-Net architecture in terms of both segmentation detail and structural consistency.

Concretely, we introduce the following three key improvements into the baseline U-Net framework: (1) Data Augmentation Strategy: A diverse set of augmentation techniques is applied to enrich the training data distribution, thereby improving the model's generalization ability across varied road conditions. (2) Loss Function Optimization: We employ a composite loss function that integrates PolyLoss and contrastive loss to simultaneously mitigate class imbalance and promote the learning of structural similarities between predicted and ground truth lane masks. (3) Upsampling Module Enhancement: The decoder path is redesigned with improved upsampling blocks that better retain fine-grained spatial features and lane boundary continuity, addressing the loss of detail often observed in traditional U-Net outputs.

Extensive experiments conducted on the TuSimple lane detection benchmark demonstrate that the proposed method achieves notable improvements over the baseline in multiple evaluation metrics, including Accuracy, F1-score, and IoU. These results confirm the effectiveness of our architectural enhancements and highlight the practical value of the approach in real-world scenarios.

Future Work. In future research, we plan to further boost the model's performance by incorporating more advanced data augmentation techniques (e.g., domain randomization, adversarial augmentation) and attention mechanisms (e.g., self-attention or cross-scale attention). We also aim to evaluate the model's adaptability on more diverse and challenging datasets, such as CULane or BDD100K, to better assess its real-world generalizability. Ultimately, we envision deploying our framework in real-world autonomous driving pipelines, and further extending it into a multi-task learning setting to support richer scene understanding tasks such as drivable area detection and object-level semantic segmentation.

References

[1] X. Yang, H. Zhang, and Y. Liu, "A survey on deep learning-based lane detection, " IEEE Transactions on Intelligent Transportation Systems, 2023.

[2] J. Zheng, W. Liu, and R. Zhao, "Ccha-net: Cross convolution hybrid attention network for lane detection, " Scientific Reports, 2024. [Online]. Available: https: //www.nature.com/articles/s41598- 025-01167-z

[3] M. Li, F. Zhou, and J. Wang, "Learning lane detection via self-attention and hybrid loss, " Pattern Recognition Letters, vol. 156, pp. 45–52, 2022.

DOI: 10.54254/2755-2721/2025.LD26122

[4] K. Zhou and S. Huang, “Attentionlane: Lightweight transformer with multi-level guidance for lane detection, ” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023, pp. 11 789–11 797.
[5] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation, ” in Medical Image Computing and Computer-Assisted Intervention (MICCAI). Springer, 2015, pp. 234–241.
[6] R. Yousri, K. Moussa, M. A. Elattar, and A. H. Madian, “A novel data augmentation approach for egolane detection enhancement using perspective transformation, ” Evolving Systems, vol. 15, p. 1021–1032, 2023.
[7] R. Li and Y. Dong, “Robust lane detection through self pre-training with masked sequential autoencoders and finetuning with customized polyloss, ” arXiv preprint arXiv: 2305.17271, 2023.
[8] K. Zhou, Y. Feng, and J. Li, “Unsupervised domain adaptive lane detection via contextual contrast and aggregation, ” arXiv preprint arXiv: 2407.13328, 2024.
[9] “Grid anchor lane detection based on attribute correlation, ” Applied Sciences, 2025.
[10] K. Xu, Z. Hao, M. Zhu, and J. Wang, “An efficient lane detection network with channelenhanced coordinate attention, ” Machines, vol. 12, no. 12, p. 870, 2024.
[11] A. Csato and F. Mariasiu, “Effects of hybridizing the unet neural network in traffic lane detection process, ” Applied Sciences, vol. 15, no. 14, p. 7970, 2025.
[12] M. Tangestanizadeh, M. Dehghani Tezerjani, and S. Youseffan Jazi, “Attentionbased unet method for autonomous lane detection, ” arXiv preprint arXiv: 2411.10902, 2024.
[13] K. Zhou, Y. Feng, and J. Li, “Unsupervised domain adaptive lane detection via contextual contrast and aggregation, ” arXiv preprint arXiv: 2407.13328, 2024.
[14] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, “Tversky loss function for image segmentation using 3d fully convolutional deep networks, ” MICCAI, 2017.
[15] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation, ” 3DV, 2016.
[16] M. Berman, A. Triki, and M. B. Blaschko, “The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks, ” CVPR, 2018.
[17] W. Li et al., “Boundary-aware network for lane detection, ” IEEE TITS, 2022.
[18] A. Tarvainen and H. Valpola, “Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results, ” NeurIPS, 2017.
[19] T. Inc., “Tusimple lane detection dataset, ” 2020, accessed: 2025-07-26. [Online]. Available: https: //www.kaggle.com/datasets/manideep1108/tusimple