Association for the Advancement of Artificial Intelligence

Retraction Note to:

GEONet: Global Enhancement and Optimization Network for Lane Detection

The authors of this paper have formally requested its retraction. This request has been reviewed and accepted by AAAI Press and the Program and General Chairs of AAAI 2025.

The paper was originally submitted as a methodological extension of the authors' prior work published at AAAI 2024 [1]. During submission, the authors prepared an appendix that provided detailed comparisons with their earlier publication, highlighting new architectural elements and improvements in inference speed and accuracy.

However, during the camera-ready preparation the appendix was not included in the final version of the paper. As a result, essential contextual and comparative information was omitted, which led to misunderstandings among readers regarding the originality and ethical standing of the work.

After internal discussion, the authors concluded that, without this comparative documentation, the published paper did not meet their intended standards of academic transparency and integrity. To address these concerns and to uphold the integrity of the scholarly record, the authors have voluntarily requested retraction.

AAAI Press and the Program and General Chairs of AAAI 2025 support this decision. The authors apologize for any confusion or concern this publication may have caused in the research community.

July 7th, 2025

References

[1] Yuan, Zhenlong, Jinguo Luo, Fei Shen, Zhaoxin Li, Cong Liu, Tianlu Mao, and Zhaoqi Wang. 2025. DVP-MVS: Synergize Depth-Edge and Visibility Prior for Multi-View Stereo. In *Proceedings of the AAAI Conference on Artificial Intelligence* 39 (9):9743-52. https://doi.org/10.1609/aaai.v39i9.33056.

GEONet: Global Enhancement and Optimization Network for Lane Detection

Suyang Xi[1*], Yunhao Liu[2*], Hong Ding[1*], Mingshuo Wang[4], Zhenghan Chen[3†], Xiaoxuan Liang[5]

[1]School of Electrical Engineering and Artificial Intelligence, Xiamen University
[2]School of Information Science and Engineering, Fudan University
[3]Microsoft
[4]College of Software, Xinjiang University
[5]School of Electrical and Computer Engineering, University of Massachusetts Amherst

Abstract

Lane detection plays a crucial role in autonomous driving systems, enabling vehicles to navigate safely and efficiently in complex environment. Despite significant advancements in recent years, accurate lane detection remains a challenging task, particularly in scenarios with occlusions, ambiguous lane markings, and diverse lighting conditions. In this paper, we propose the Global Enhancement and Optimization Network (GEONet) for lane detection, which is designed to refine both feature extraction and global feature transmission. Traditional approaches typically depend on deep convolutional layer stacks for global feature extraction, a process that often compromises inference speed and the precision of global feature representation. In contrast, GEONet introduces a novel and more effective methodology. We present the Global Feature Extraction Module (GFEM), which is specifically engineered to capture comprehensive global features with higher accuracy. Additionally, we introduce the Top-Tier Supplementary Module (TTSM), which enhances these features through a bottom-up approach, improving overall lane detection accuracy. To further bolster our framework, we incorporate Whitening Batch Normalization (WBN) and Whitening Contrastive Learning (WCL), which enhance feature robustness and ensure better generalization. In addition to our novel network design, we propose two new loss functions to enhance lane detection accuracy. The Generalized Rectangular Intersection over Union (GRIoU) Loss extends the predicted points into rectangles, optimizing overlap and smoothness of lane predictions.The Angle Loss accounts for angular differenses between predicted and ground truth lanes, improving alignment and continuity. Experimental results demonstrate that our proposed method significantly outperforms current state-of-the-art lane detection techniques.

Introduction

Lane detection is the cornerstone of autonomous driving, playing a vital role in vehicle localization and safe navigation. While deep neural networks (Glorot, Bordes, and Bengio 2011) have advanced the field, lane detection remains challenging, particularly in complex scenarios.

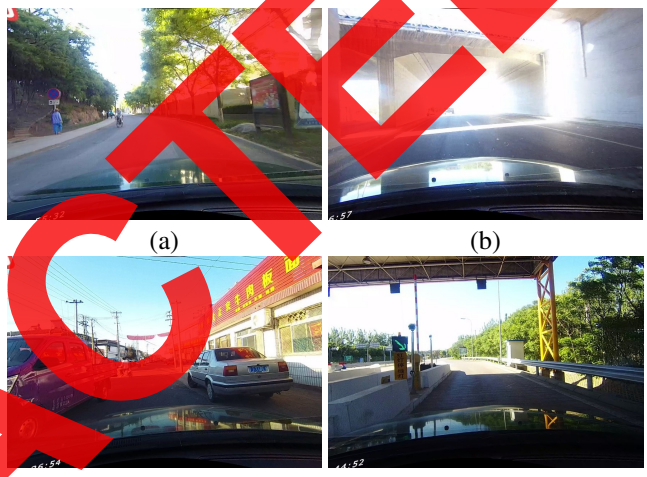

Figure 1: The following are examples of complex scenarios in lane detection: (a) Curved lanes, where the application of our Angle Loss method greatly improves detection accuracy. (b) Lanes affected by strong glare, where leveraging global semantic information is critical for effective detection. (c) Lanes obscured by surrounding vehicles, presenting significant challenges for accurate lane identification. (d) Situations where lane markings are not present.

Earlier methods (Canny 1986; Hough 1962; Sobel, Feldman et al. 1968) depended heavily on manual parameter tuning and edge-based techniques, which often resulted in variability and reduced performance under different conditions. Modern approaches leveraging CNNs have improved feature representation, yet accurately detecting lanes in complex scenarios remains challenging. Recent studies (Pan et al. 2018; Zheng et al. 2021, 2022) introduced advanced methods such as lane coherence and cross-to-fine mechanisms to enhance detection. However, these methods often struggle to fully integrate global semantics with local features, leading to inconsistencies in lane prediction. As illustrated in Fig 1, various complex scenarios further complicate accurate lane detection, emphasizing the need for a robust approach that fuses global semantics with local details to adapt to diverse

*Equal contribution.
†Corresponding author(e-mail: pandaarych@gmail.com).

road conditions.

In complex scenarios, another prevalent issue in lane detection is the model's excessive focus on lightweight backbones to increase inference speed. Early neural network models for lane detection employed instance segmentation and anchor-based object detection techniques, divided into line anchor-based and row anchor-based methods. Line anchor-based methods, such as Line-CNN (Li et al. 2019) and LaneATT (Tabelini et al. 2021a), utilized predefined anchors to accurately regress lane positions, demonstrating high efficiency. In contrast, row anchor-based methods like UFLD (Qin, Wang, and Li 2020) and CondLaneNet (Liu et al. 2021a) predicted potential cells for each predefined row, aiming for rapid inference. However, these methods often underperform in complex environments, lagging behind more advanced and robust approaches

This paper introduces the GEONet framework, developed with the understanding that successful lane detection in difficult scenarios heavily relies on a comprehensive grasp of global context. To address this, we propose a novel method for extracting global features, consisting of the Global Feature Extraction Module (GFEM) and the Top-Tier Supplementary Module (TTSM). The GFEM processes feature maps from the network's backbone to capture precise and extensive global information, which is then employed in subsequent layers and seamlessly integrated into the classification and regression modules via the TTSM, offering additional enhancement. Moreover, we present the Angle Loss, which aims to harmonize the shapes of predicted and ground truth (GT) lanes by taking into account their angular variations. The Generalized Rectangular Intersection over Union (GRIoU) Loss further refines this approach by transforming predicted points into rectangles to improve model accuracy and ensure more consistent lane predictions compared to the traditional Line IoU Loss. Furthermore, we initially implemented Whitening Batch Normalization (WBN) as a replacement for the standard Batch Normalization (BN) in our backbone network. Subsequently, we leveraged Whitening Contrastive Learning (WCL) during the pre-training phase to further boost the discriminative ability and robustness of our features. Benchmark results clearly demonstrate the superiority of our approach in both accuracy and resilience. The main contributions of our work are outlined as follows:

- We have significantly advanced lane detection capabilities by introducing GEONet.
- The GFEM module captures semantic features while the TTSM module optimizes computations, enhancing network performance in complex scenarios.
- We integrate Whitening Contrastive Learning (WCL) and Whitening Batch Normalization (WBN) to enhance feature robustness and generalization.
- We propose the GRIoU Loss to optimize lane prediction performance by improving overlap and smoothness, and introduce Angle Loss to align predicted and ground truth lanes.
- Our approach has been thoroughly validated on multiple benchmark datasets, demonstrating its state-of-the-art performance and robustness.

Related Work

Lane detection techniques utilizing CNNs can be broadly categorized based on their lane representation approach: those that rely on segmentation, anchor-based frameworks, and methods centered on polynomial regression.

Segmentation-Based Methods

Recent advancements in lane detection have embraced pixel-wise prediction strategies, recasting the problem as a semantic segmentation task. While techniques like LaneNet (Neven et al. 2018) have leveraged multi-task networks for comprehensive learning, they often struggle with balancing accuracy and real-time performance. In contrast, methods incorporating self-attention mechanisms (Hou et al. 2019) have demonstrated improved efficiency, allowing for lightweight network designs. However, these approaches still rely on exhaustive pixel-by-pixel processing, ignoring the holistic nature of lanes. Neural architecture search (NAS) (Xu et al. 2020) has been explored as a means to optimize network structures for better accuracy, but the computational overhead remains prohibitive for widespread adoption. Consequently, the quest for an efficient yet highly accurate lane detection system that holistically perceives lanes continues to be a research frontier.

Anchor-Based Methods

Anchor-based techniques enhance efficiency but face difficulties in complex lane geometries or low visibility. Studies like UFLD (Qin, Wang, and Li 2020) have advanced irregular lane detection speed and accuracy, but limitations persist. A comprehensive approach merging global semantics and local features is crucial for a breakthrough. Key milestone studies, including LineCNN (Li et al. 2019) and LaneATT (Tabelini et al. 2021a), have laid the foundation for this evolution, while CLRNet (Zheng et al. 2022) has further refined the methodology through its innovative anchor partitioning system, striking a careful balance between detection precision and computational efficiency. However, the reliance on static anchors poses challenges in adaptability to dynamic environments, highlighting the need for more innovative solutions, such as our enhanced methodology that significantly extends the capabilities of the SOTA CLRNet (Zheng et al. 2022).

Polynomial-Regression-Based Methods

Departing from conventional CNN tactics, parameter curve-based methods model lane lines through curve parameters and detect them by focusing on the regression of curve parameters and related metrics such as confidence scores. This category includes innovative solutions such as PolyLaneNet (Tabelini et al. 2021b) and LSTR(Liu et al. 2021b), which have significantly influenced the field. Additionally, Bezier-LaneNet (Feng et al. 2022) uses Bezier curves to fit lane lines and optimizes the computational bottleneck of LSTR, making breakthroughs in both efficiency and accuracy.

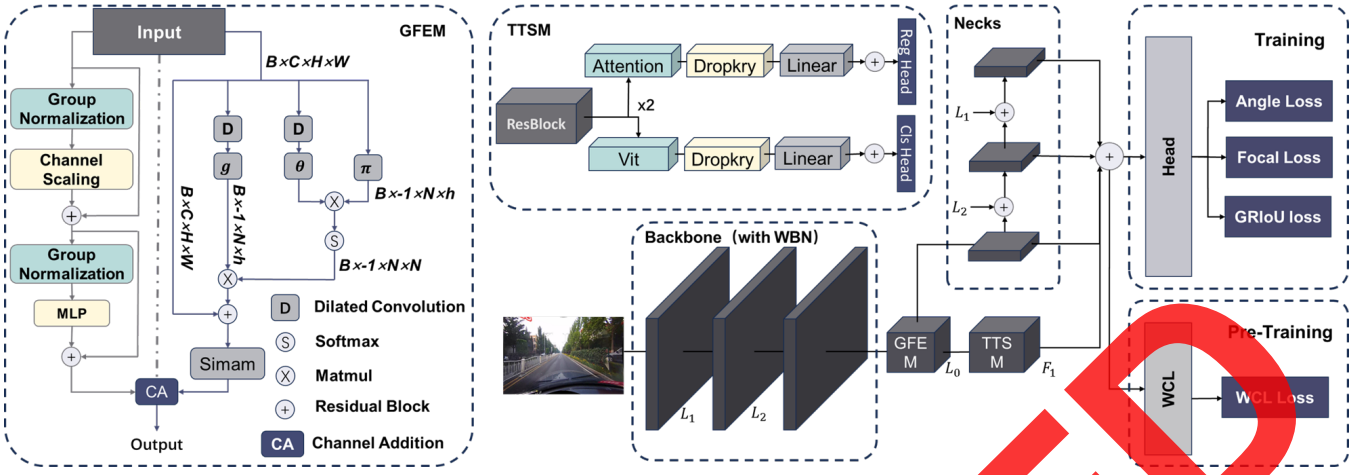

Figure 2: Overview of the proposed GEONet. Image data undergoes preliminary feature extraction through a wbn-processed backbone network. Subsequently, the Feature Pyramid Network (Lin et al. 2017) and the Global Feature Extraction Module (GFEM) generate informative feature maps. The GFEM employs a unique dual-branch structure to capture and integrate global semantic information, crucial for enhancing lane detection accuracy. These feature maps are used for lane classification and regression, while also being fused with global semantic information from the Top-Tier Supplementary Module (TTSM). Additionally, the introduction of Angle Loss and GRIoU Loss significantly improves model performance. A pre-training strategy based on WCL is also implemented to further boost the model's effectiveness.

Method

Global Feature Extraction Module (GFEM)

Motivation. Integrating global semantic information into traditional lane detection models which are based on CNNs (Zheng et al. 2021, 2022; Qin, Wang, and Li 2020; Liu et al. 2021a; Qin, Zhang, and Li 2022) remains challenging, particularly in occluded or low-light conditions. These scenarios demand sophisticated global information synthesis for accurate detection. Conventional deep convolutional methods often struggle with such complexities. To address this, we propose the Global Enhancement Module (GFEM), a novel architecture that enhances the integration of global semantics for improved feature synthesis.

GFEM Structure. The Global Enhancement Module (GFEM) is designed with two parallel branches that work together to extract refined global features. The left branch processes the input feature maps through an MLP-mixer network (Rumelhart, Hinton, and Williams 1986), which includes channel scaling and normalization. This branch establishes a foundational global feature framework but primarily captures coarse features, lacking fine-grained detail. The right branch utilizes a dilated convolution process to expand the receptive field, allowing for broader contextual analysis. After dilation, the feature map is divided into P sub-blocks, distributed across h heads, where pixel similarity is computed using a weighted sum approach. This method enhances fine-grained features by focusing on smaller segments while maintaining an overarching view of long-distance spatial relationships, crucial for handling complex scenarios. Additionally, we integrate the SimAm block (Yang et al. 2021) to improve the depth and quality of global semantic representation. By combining the outputs from both branches, the module synthesizes their strengths, resulting in a detailed and robust global feature set that significantly enhances lane detection accuracy across various driving conditions.

Top-Tier Supplementary Module (TTSM)

Motivation. The Top-Tier Supplementary Module (TTSM) is intended to leverage the extensive global semantic information contained in the highest-level feature maps of the neural network architecture. Inspired by the transformative potential of attention mechanisms in network depth and complexity management (Vaswani et al. 2017), and their successful application in visual tasks (Carion et al. 2020; Dosovitskiy et al. 2020), TTSM aims to optimize the feature distillation process between the Global Feature Extraction Module (GFEM) and the network's decision-making layers.

TTSM Structure. The structural essence of TTSM lies in its dual approach to processing semantic information through self-attention mechanisms (Vaswani et al. 2017). Initially, the feature map L_0, extracted from the network's backbone, undergoes transformation via a basic residual network ϕ to enhance its semantic richness. This processed feature map, denoted as F_{top}, is subjected to two distinct self-attention operations, $Auxihead_1$ and $Auxihead_2$, aimed at optimizing feature representations for both classification and regression tasks. Furthermore, the Dropkey operation on the attention output serves as a regularization technique, preventing overfitting and improving generalization:

$$F_{top} = \phi(\phi(L_0)), \tag{1}$$

$$S_1, S_2 = Auxihead_1(F_{top}), Auxihead_2(F_{top}), \tag{2}$$

$$S_1', S_2' = g(S_1), g(S_2), \tag{3}$$

where ϕ represents a simple residual network, and g represents Dropkey operation. $Auxihead_1$ and $Auxihead_2$ represent multi-head self-attention mechanism(Vaswani et al. 2017) and Vision Transformer (ViT) .

The TTSM module leverages the power of self-attention to capture long-range dependencies and global contextual information from the top-level feature maps. By applying self-attention separately for classification and regression tasks, TTSM enables the network to learn task-specific global representations, enhancing its ability to handle complex lane detection scenarios.

Whitening Batch Normalization (WBN)

Batch Normalization. For a Batch Normalization (BN) layer that processes a batch of D-dimensional vectors $X = x_1, \cdots, x_B \in \mathbb{R}^{D\times B}$, its output is a batch of normalized vectors $Y = y_1, \cdots, y_B \in \mathbb{R}^{D\times B}$, computed as follows:

$$y_{i,k} = \frac{x_{i,k} - \mu_k}{\sqrt{\sigma_k^2 + \epsilon}} \cdot \gamma_k + \beta_k \tag{4}$$

for all $i \in \{1, \cdots, B\}$ and $k \in \{1, \cdots, D\}$, where γ, β are learnable affine parameters, ϵ is a small constant originally proposed for numerical stability. In training time, μ_k, σ_k^2 are mean and variance computed over the k-th row of the input batch X, and in inference time, running estimations from training time is used.

Domain-specific Whitening Transform. As previously mentioned, BN involves standardizing each sample $\mathbf{x}_i \in B$ based on its dimensions. However, this means that even after normalization, the features of batch samples may still be correlated. Essentially, BW involves transforming the data to have zero means and unit variances, while also decorrelating the features via principal component analysis.

$$BW(x_{i,k}; \Omega) = \gamma_k \hat{x}_{i,k} + \beta_k, \tag{5}$$

$$\hat{\mathbf{x}}_i = W_B(\mathbf{x}_i - \boldsymbol{\mu}_B). \tag{6}$$

In our network, we define the WBN to comprehensively replace the BN layers. Using Eq. (5)-(6) we can now define our WBN. Let x^s denote the inputs to the WBN layer from the target domain. Our WBN is defined as follows

$$WBN(x^t; \Omega^t) = BW(x^t, \Omega^t). \tag{7}$$

Whitening Contrastive Learning (WCL)

Motivation. To ensure optimal performance, we need to guarantee that (1) embeddings of images do not come from a degenerate distribution that might lead all representations to collapse into a single feature distribution point; (2) the pairs of positive images (x_i, x_j) with similar semantics should be grouped closely together.

Definition. We donate the number of positive samples per image denoted as d. Let N represent the number of original images, and consider a batch of samples $B = x_1, ..., x_k$, where $K = Nd$. We obtain the corresponding batch of features, denoted by $V = \{\mathbf{v}_1, ...\mathbf{v}_k\}$, using the aforementioned approach. In our proposed W-MSE loss, we calculate the mean squared error (MSE) across all $Nd(d-1)/2$ positive pairs, satisfying the constraint through reparameterization of the $\mathbf{v}$ variables with the whitened variables $\mathbf{z}$.We have formulated this problem as follows:

$$min_\theta \, \mathbb{E}[dist(\mathbf{z}_i, \mathbf{z}_j)] \tag{8}$$

$$\mathcal{L}_{WCL}(V) = \frac{2}{Nd(d-1)} \sum dist(\mathbf{z}_i, \mathbf{z}_j) \tag{9}$$

Here, the sum is over all $(\mathbf{v}_i, \mathbf{v}_j) \in V$, with $pos(i, j) = true$, and $\mathbf{z} = Whitening(\mathbf{v})$ representing the whitened variables. Specifically, we have

$$Whitening(\mathbf{v}) = W_V(\mathbf{v} - \boldsymbol{\mu}_V) \tag{10}$$

Eq. 10 defines $\boldsymbol{\mu}_V$ as the mean of the elements in V: $\boldsymbol{\mu}_V = \frac{1}{K}\sum_k \mathbf{v}_k$. The matrix W_V is determined such that $W_V^\top W_V = \Sigma_V^{-1}$, where Σ_V denotes the covariance matrix of V.

$$\Sigma_V = \frac{1}{K-1} \sum_k (\mathbf{v}_k - \boldsymbol{\mu}_V)(\mathbf{v}_k - \boldsymbol{\mu}_V)^T \tag{11}$$

The proposed loss function in Eq. 9 operates on the premise that it penalizes distant positive pairs, thereby encouraging $g(E(\cdot))$ to reduce their inter-positive distances. Before formal training, we pre-train our model using WCL, which enables the image embeddings to have a favorable distribution, laying a solid foundation for subsequent training.

Generalized Rectangular Intersection over Union Loss (GRIoU Loss)

Motivation. The GRIoU loss aims to optimize the overlap between predicted and ground truth lane points while introducing a geometric penalty term to ensure smoother lane predictions. By utilizing rectangular regions, the GRIoU loss is designed to better capture the spatial relationships between lane markings, especially in complex scenarios where lanes can exhibit various orientations and curvatures. Penalizing the distance between the predicted and ground truth rectangular regions ensures that the model produces more accurate and consistent lane predictions, even in challenging driving environments.

Definition. Let $\mathcal{P} = (x_i^P, y_i^P)_{i=1}^N$ and $\mathcal{G} = \{(x_i^G, y_i^G)\}_{i=1}^N$ denote the sets of predicted and ground truth lane points, respectively, where N is the number of points. For each pair of consecutive points (x_i^P, y_i^P) and (x_{i-1}^P, y_{i-1}^P) in $\mathcal{P}$, we define a rectangular region R_i^P with a length of $\sqrt{(x_i^P - x_{i-1}^P)^2 + (y_i^P - y_{i-1}^P)^2}$ and a width of $2e$, where e is a predefined constant. Similarly, we define rectangular regions R_i^G for each pair of consecutive points in $\mathcal{G}$. The GRIoU Loss $\mathcal{L}_{GRIoU}$ is defined as:

$$\mathcal{L}_{GRIoU} = 1 - \frac{\sum_{i=1}^N d_i^{GI}}{\sum_{i=1}^N d_i^{GU}} \tag{12}$$

where d_i^{GI} and d_i^{GU} are calculated as:

$$d_i^{GI} = \begin{cases} IoU(R_i^P, R_i^G) - \frac{X - \min(X,Y)}{X}, & \text{if } 2 \le i \le N \\ IoU(R_i^P, R_i^G), & \text{if } i = 1 \end{cases} \tag{13}$$

$$d_i^{GU} = \begin{cases} 1, & \text{if } 2 \le i \le N \\ IoU(R_i^P, R_i^G), & \text{if } i = 1 \end{cases} \tag{14}$$

where X and Y represent the areas of the bounding boxes for the predicted and ground truth lanes, respectively, $IoU(R_i^P, R_i^G)$ is the intersection over union of the predicted and ground truth rectangular regions, and $S_{bound}(R_i^P, R_i^G)$ is the area of the smallest bounding box that encloses these regions.

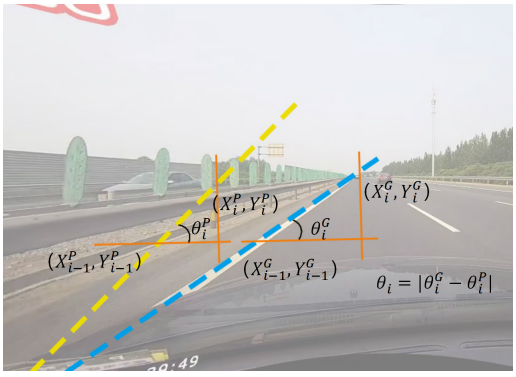

Figure 3: Visualization of Angle Loss

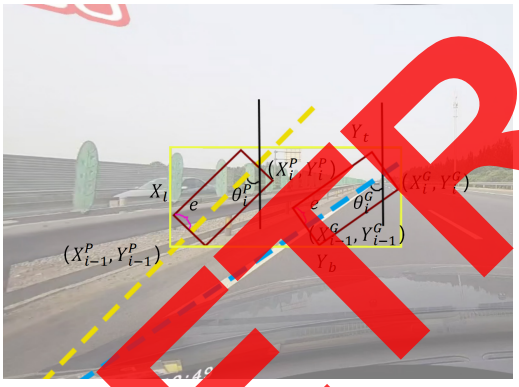

Figure 4: Visualization of GRIoU Loss

Angle Loss

Motivation. Angle Loss is designed to align the predicted lane with the ground truth by accounting for the angular differences between them. This strategy is particularly effective in ensuring the smoothness and continuity of lane lines, especially in complex driving environments.

Definition. Angle Loss functions to reduce the angular deviation between predicted and actual lanes. This deviation is quantified by the angles formed between consecutive points on the predicted and ground truth lanes.

Firstly, the coordinates of a predicted lane point are defined as (x_i^P, y_i^P), and the corresponding GT point is (x_i^G, y_i^G). The total angle θ between the predicted lane and the ground truth (GT) lane is defined as the average of the angles between the individual short lines. It is calculated as follows:

$$\theta = \frac{1}{N-1} \sum_{i=2}^{N} \theta_i \tag{15}$$

where the angle θ_i between them is expressed as follows:

$$\theta_i = \arctan\left(\frac{\left(\frac{y_i^P - y_{i-1}^P}{x_i^P - x_{i-1}^P}\right) - \left(\frac{y_i^G - y_{i-1}^G}{x_i^G - x_{i-1}^G}\right)}{1 + \left(\frac{y_i^P - y_{i-1}^P}{x_i^P - x_{i-1}^P}\right)\left(\frac{y_i^G - y_{i-1}^G}{x_i^G - x_{i-1}^G}\right)} \right) \tag{16}$$

The final angle loss is quantified using the cosine function as follows:

$$\mathcal{L}_{Angle} = 1 - \cos\theta \tag{17}$$

Experiment

Datasets Utilized

Our experimental investigations leverage two prominent and extensively benchmarked datasets in the realm of lane detection: CULane (Pan et al. 2018) and Tusimple [1].

Experimental Details

We apply ResNet (He et al. 2016) and DLA (Yu et al. 2018) as our pre-trained backbones. Input images are resized to 800×320 for all datasets. Data augmentation methods include random affine transformations, such as translation, rotation, scaling, and horizontal flips. The AdamW optimizer with a cosine decay learning rate strategy is employed for optimization. For the CULane and TuSimple datasets, we first perform WCL pre-training for 15 epochs with a learning rate of 4e-4 and a batch size of 40, followed by formal training for 15 epochs with a learning rate of 6e-4 and a batch size of 24, and then 70 epochs with a learning rate of 1e-3 and a batch size of 40. The weight for the angle loss is set to 15 across all datasets, and the interplay between the GRIoU Loss and Angle Loss is finely tuned through a hyperparameter α, which governs their relative contributions to the overall loss function:

$$L_{\text{comb}} = \alpha \times L_{\text{GRIoU}} + (1 - \alpha) \times L_{\text{Angle}}. \tag{18}$$

Training was performed on a GeForce RTX 4090 GPU. This configuration ensures a balanced optimization of prediction accuracy, feature robustness, and lane shape alignment, enhancing lane detection performance in complex scenarios.

Contrasting Our Results with SOTA Approaches

CULane Dataset. We conduct a rigorous analysis of our method's performance on the CULane dataset, situating it within the context of other advanced techniques. By leveraging the DLA34 architecture (Yu et al. 2018), we attained a notable F1@50 score of 82.18, underscoring the efficacy

[1] https://github.com/TuSimple/tusimple-benchmark/

Method	Backbone	F1@50	mF1	F1@75	Arrow	Cross	Crowded	Dazzle	Shadow	Night	No line	Normal
FOLOLane	ERFNet	78.80	-	-	89.00	1569	77.80	75.20	79.30	74.50	52.10	92.70
SGNet	ResNet34	77.67	-	-	87.97	1373	75.41	67.75	74.31	72.69	50.90	92.07
UFLDv2	ResNet34	76.0	-	-	88.80	1910	74.80	65.50	75.50	70.80	49.20	92.50
CANet	ResNet101	79.86	-	-	90.18	1196	78.74	70.07	79.35	74.91	52.88	93.60
CondLane	ResNet34	78.74	53.11	59.39	89.89	1387	77.14	71.17	79.93	73.92	51.85	93.38
LaneATT	ResNet122	77.02	51.48	57.50	86.29	1264	76.16	69.47	76.31	70.81	50.46	91.74
LaneATT	ResNet34	76.68	49.57	54.34	88.38	1330	75.03	66.47	78.15	70.72	49.39	92.14
UFLD	ResNet34	72.30	-	-	85.70	2037	70.20	59.50	69.30	66.70	44.40	90.70
UFLD	ResNet18	68.40	38.94	40.01	81.00	1743	66.00	58.40	62.80	62.10	40.20	87.70
CLRNet	ResNet101	80.13	55.55	62.96	89.79	1262	78.78	72.49	82.33	75.51	54.50	93.85
CANet	ResNet34	79.16	-	-	90.09	1176	77.88	73.11	75.06	73.92	51.68	93.58
CLRNet	DLA34	80.47	55.64	62.78	90.62	1155	79.59	75.30	82.51	75.37	54.58	93.73
CondLane	ResNet101	79.48	54.83	61.23	90.16	1201	77.44	70.93	80.91	74.80	54.13	93.47
LaneAF	DLA34	77.41	50.42	56.79	86.88	1360	75.61	71.78	79.12	73.03	51.38	91.80
E2E	ERFNet	74.00	-	-	85.80	2022	73.10	64.50	74.10	67.90	46.60	91.00
SCNN	VGG16	71.60	38.84	39.84	84.10	1990	69.70	58.50	66.90	66.10	43.40	90.60
RESA	ResNet50	75.30	47.86	53.39	88.30	1503	73.10	69.20	72.80	69.90	47.70	92.10
CLRNet	ResNet34	79.73	55.14	62.11	90.59	1216	78.06	74.57	79.92	75.02	54.01	93.49
PINet	Hourglass	74.40	46.81	51.33	83.70	1427	72.30	66.30	68.40	67.70	49.80	90.30
CLRNet	ResNet18	79.58	55.23	62.21	90.25	1321	78.33	73.31	79.66	75.11	53.14	93.30
GEONet(ours)	ResNet18	81.65	57.25	65.06	91.91	**1061**	80.80	75.63	83.38	76.67	55.41	94.88
GEONet(ours)	ResNet34	**83.07**	57.27	64.93	92.31	1097	81.15	77.13	83.52	77.54	55.89	95.04
GEONet(ours)	ResNet101	82.13	**58.23**	65.06	**93.10**	1184	81.25	76.34	83.38	77.60	56.51	95.13
GEONet(ours)	DLA34	82.18	57.85	**66.02**	92.80	1184	**81.90**	**77.52**	**84.33**	**78.12**	**57.21**	**96.01**

Table 1: Leading Performfmance Achievements on the CULane Dataset.

Method	Backbone	Acc(%)	FP(%)	F1(%)
FOLOLane	ERFNet	**96.92**	4.47	96.59
UFLDv2	ResNet34	95.56	3.18	96.22
RESA	ResNet34	96.82	3.63	96.93
LaneATT	ResNet34	95.63	3.53	96.77
LaneATT	ResNet122	96.10	5.64	96.06
UFLD	ResNet34	95.86	18.91	88.02
CANet	ResNet34	96.66	2.32	97.44
CANet	ResNet101	96.76	1.92	97.77
CondLaneNet	ResNet34	95.37	2.20	96.98
CondLaneNet	ResNet101	96.54	2.01	97.24
CLRNet	ResNet18	96.84	2.28	97.89
CLRNet	ResNet34	96.87	2.27	97.82
CLRNet	ResNet101	96.83	2.37	97.62
GEONet(ours)	ResNet18	96.90	**1.83**	**98.21**
GEONet(ours)	ResNet34	96.91	1.98	98.17
GEONet(ours)	ResNet101	96.88	2.01	98.06

Table 2: Unmatched Performance on the TuSimple Benchmark: Our approach sets a new standard, with F1 scores calculated using the official source code.

of our approach. When utilizing ResNet18 (He et al. 2016) as the backbone, our method achieved an F1@50 score of 81.65, surpassing CLRNet (Zheng et al. 2022) (ResNet18) by 2.07 points. Furthermore, with ResNet34, our approach yielded an F1@50 score of 83.07, further affirming the effectiveness of our global semantic strategy in enhancing lane detection accuracy.

TuSimple Dataset. The differences in performance between various methods are relatively slight, indicating that further improvements on this dataset may be reaching a plateau.

Angle Loss	GRIoU Loss	WCL	WBN	TTSM	GFEM	F1@75	F1@50	mF1
✓	✓	✓	✓	✓	✓	64.81	81.81	56.93
✓	✓	✓	✓	✓		64.58	81.64	56.75
✓	✓	✓	✓			64.14	81.25	56.41
✓	✓	✓				63.71	80.96	56.25
✓	✓					63.23	80.57	53.20
✓						62.90	80.47	55.86

Table 3: Evaluation of Each Technique Through Ablation Study: Results derived using a ResNet18 backbone on the CULane dataset.

Nonetheless, our method achieves a remarkable F1@50 score of 98.21, surpassing the current leading approach by 0.32 points.

Ablation Study

We conducted ablation experiments on GEONet to evaluate key components like angle loss, residual blocks, the Whitening module, and GRIoU loss. Each component improved model accuracy and robustness, particularly in complex scenarios, validating GEONet's design and guiding further optimization.

Comprehensive Ablation Analysis. We first evaluated the impact of introducing Angle Loss and GRIoU Loss into our baseline model, which improved the F1@50 score from 80.12 to 81.28. Subsequent ablation studies focused on the contributions of other key modules. The inclusion of WCL and WBN modules led to further improvements, with F1@75 increasing to 64.81, F1@50 to 81.81, and mF1 to 56.93, proving that WCL contributed to a more robust and discriminative embedding space, while WBN helped in normalizing the feature maps to a more uniform distribution. When the GFEM module was excluded, F1@50

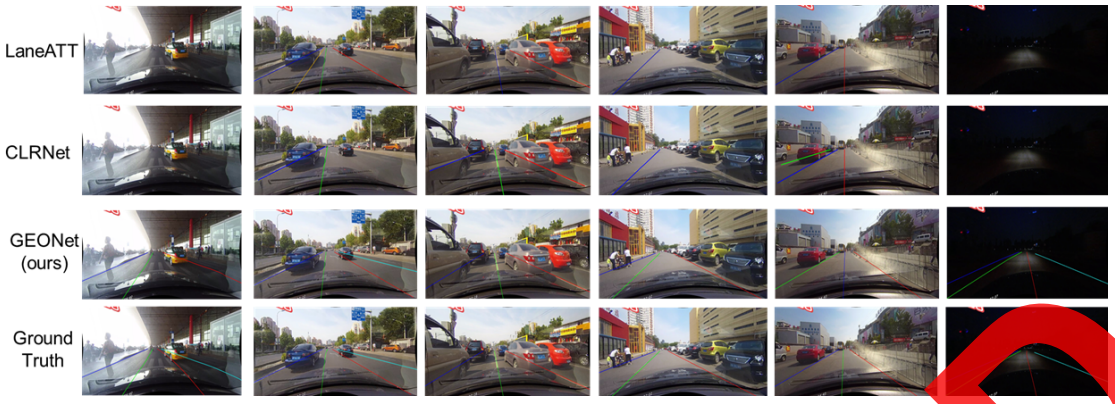

Figure 5: Performance Visualization of Lane Detection Methods on CULane: LaneATT, CLRNet, GEONet (Ours), and Ground Truth.

Residual blocks	F1@75	F1@50	mF1	No line	Shadow
No TTSM	62.88	80.44	55.82	53.71	80.52
4 × blocks	62.94	80.49	55.74	54.52	81.39
3 × blocks	63.19	80.45	55.77	53.79	**82.70**
2 × blocks	**63.34**	**80.69**	55.88	**55.03**	82.38
1 × blocks	62.98	80.46	**56.01**	54.46	81.97
0 × blocks	62.90	80.41	55.81	53.97	82.09

Table 4: Investigation into the Impact of Residual Block Quantity in TTSM.

slightly decreased to 81.64, and F1@75 to 64.58, indicating GFEM's critical role in enhancing performance. Removing both TTSM and GFEM led to a further decline in F1@50 and F1@75, demonstrating that these modules contribute significantly to multi-scale feature extraction and overall accuracy. Finally, when only Angle Loss was used, the model's performance dropped to the lowest, with F1@75 at 62.90, F1@50 at 80.47, and mF1 at 55.86, highlighting the importance of integrating all these components for optimal lane detection performance.

Ablation Study on TTSM's Number of Residual Blocks. To rigorously assess the impact of residual block quantity within TTSM, we performed an ablation study as depicted in Table 4. This study investigates the number of residual blocks (He et al. 2016) in TTSM.The result demonstrates the impact of the number of residual blocks in the Top-Tier Supplementary Module (TTSM) on the performance of the lane detection model. The study indicates that using 0 or 1 residual block results in insufficient global semantic enhancement, leading to performance below the baseline. Conversely, using 3 or 4 blocks causes a decline in performance due to the loss of fine details. The optimal configuration is 2 residual blocks, which significantly improves F1@75, F1@50, and excels in challenging scenarios like "Shadow," confirming the importance of balancing the number of residual blocks to enhance overall model performance.

Ablation Study on Loss. To evaluate the effectiveness of GRIoU loss compared to other commonly used loss functions (IoU loss), we conducted ablation studies on the CU-Lane dataset. In addition, we focused on optimizing the weight parameter of the Angle Loss to elucidate its influence on model accuracy, by incorporating GRIoU Loss and IoU Loss with Angle Loss under different weights. Using ResNet18 as the baseline model, all models were trained for 15 epochs with a learning rate of $1e-3$ and batch size of 24. Table 5 shows the performance of different loss functions. The results indicate that GRIoU loss provides more precise and consistent alignment with ground truth lanes, thus improving overall lane detection accuracy. Furthermore, a judiciously calibrated weight for the Angle Loss can help maximize model performance, particularly under complex detection conditions.

Loss Function	AL' Weight	F1@75	F1@50	mF1
IoU Loss	0	61.23	79.58	54.21
	10	61.35	79.75	54.62
	15	61.61	**79.81**	**54.76**
	20	61.12	79.65	54.42
	25	61.15	79.54	53.95
GRIoU Loss	0	62.89	80.45	55.82
	10	63.30	80.62	55.95
	15	**63.44**	**80.65**	**56.22**
	20	63.08	80.49	55.71
	25	62.61	80.39	55.74

Table 5: Evaluation of Angle Loss Weight Variations in Ablation Studies.

Conclusion

In this work, we introduce GEONet, an innovative lane detection framework tailored for complex scenarios. Key advancements include GFEM and TTSM modules for enhanced global feature extraction and optimization. GRIoU and Angle Loss further refine predictions and lane shape alignment. Notably, GEONet integrates Whitening Batch Normalization (WBN) into its architecture and employs Whitening Contrastive Learning (WCL) during pre-training to bolster feature robustness and generalization. Experimental results on CULane and TuSimple datasets demonstrate GEONet's state-of-the-art performance and robustness, validating the effectiveness of its components.

References

Canny, J. 1986. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6): 679–698.

Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; and Zagoruyko, S. 2020. End-to-end object detection with transformers. In *European conference on computer vision*, 213–229. Springer.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Feng, Z.; Guo, S.; Tan, X.; Xu, K.; Wang, M.; and Ma, L. 2022. Rethinking efficient lane detection via curve modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17062–17070.

Glorot, X.; Bordes, A.; and Bengio, Y. 2011. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, 315–323. JMLR Workshop and Conference Proceedings.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Hou, Y.; Ma, Z.; Liu, C.; and Loy, C. C. 2019. Learning lightweight lane detection cnns by self attention distillation. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1013–1021.

Hough, P. V. 1962. Method and means for recognizing complex patterns. US Patent 3,069,654.

Li, X.; Li, J.; Hu, X.; and Yang, J. 2019. Line-cnn: End-to-end traffic line detection with line proposal unit. *IEEE Transactions on Intelligent Transportation Systems*, 21(1): 248–258.

Lin, T.-Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; and Belongie, S. 2017. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2117–2125.

Liu, L.; Chen, X.; Zhu, S.; and Tan, P. 2021a. Condlanenet: a top-to-down lane detection framework based on conditional convolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3773–3782.

Liu, R.; Yuan, Z.; Liu, T.; and Xiong, Z. 2021b. End-to-end lane shape prediction with transformers. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 3694–3702.

Neven, D.; De Brabandere, B.; Georgoulis, S.; Proesmans, M.; and Van Gool, L. 2018. Towards end-to-end lane detection: an instance segmentation approach. In *2018 IEEE intelligent vehicles symposium (IV)*, 286–291. IEEE.

Pan, X.; Shi, J.; Luo, P.; Wang, X.; and Tang, X. 2018. Spatial as deep: Spatial cnn for traffic scene understanding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Qin, Z.; Wang, H.; and Li, X. 2020. Ultra fast structure-aware deep lane detection. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIV 16*, 276–291. Springer.

Qin, Z.; Zhang, P.; and Li, X. 2022. Ultra fast deep lane detection with hybrid anchor driven ordinal classification. *IEEE transactions on pattern analysis and machine intelligence*.

Rumelhart, D. E.; Hinton, G. E.; and Williams, R. J. 1986. Learning representations by back-propagating errors. *nature*, 323(6088): 533–536.

Sobel, I.; Feldman, G.; et al. 1968. A 3x3 isotropic gradient operator for image processing. *a talk at the Stanford Artificial Project in*, 271–272.

Tabelini, L.; Berriel, R.; Paixao, T. M.; Badue, C.; De Souza, A. F.; and Oliveira-Santos, T. 2021a. Keep your eyes on the lane: Real-time attention-guided lane detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 294–302.

Tabelini, L.; Berriel, R.; Paixao, T. M.; Badue, C.; De Souza, A. F.; and Oliveira-Santos, T. 2021b. Polylanenet: Lane estimation via deep polynomial regression. In *2020 25th International Conference on Pattern Recognition (ICPR)*, 6150–6156. IEEE.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Xu, H.; Wang, S.; Cai, X.; Zhang, W.; Liang, X.; and Li, Z. 2020. Curvelane-nas: Unifying lane-sensitive architecture search and adaptive point blending. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV 16*, 689–704. Springer.

Yang, L.; Zhang, R.-Y.; Li, L.; and Xie, X. 2021. Simam: A simple, parameter-free attention module for convolutional neural networks. In *International conference on machine learning*, 11863–11874. PMLR.

Yu, F.; Wang, D.; Shelhamer, E.; and Darrell, T. 2018. Deep layer aggregation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2403–2412.

Zheng, T.; Fang, H.; Zhang, Y.; Tang, W.; Yang, Z.; Liu, H.; and Cai, D. 2021. Resa: Recurrent feature-shift aggregator for lane detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 3547–3554.

Zheng, T.; Huang, Y.; Liu, Y.; Tang, W.; Yang, Z.; Cai, D.; and He, X. 2022. Clrnet: Cross layer refinement network for lane detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 898–907.