ISSN: 2584-0495 Vol. 4, Issue 2, pp. 1851-1858

International Journal of Microsystems and IoT

ISSN: (Online) Journal homepage: https://www.ijmit.org

Lane Detection for Autonomous Ground Vehicle using Residual U-Net with SAVE-Net [Safe Autonomous vehicle Network]

Suprabhat Maity, Debashish Chakravarty, Maaitrayo Das

Cite as: Maity, S., Chakravarty, D.& Das, M. (2026). Lane Detection for Autonomous Ground Vehicle using Residual U-Net with SAVE-Net [Safe Autonomous vehicle Network]. International Journal of Microsystems and IoT, 4(2), 1851–1858. https://doi.org/10.5281/zenodo.21625814

Published online: 20 February 2026

https://doi.org/10.5281/zenodo.21625814

International Journal of Microsystems and IoT
Vol. 4, Issue 2, pp.1851-1858; DOI: https://doi.org/10.5281/zenodo.21625814

Lane Detection for Autonomous Ground Vehicle using Residual U-Net with SAVE-Net [Safe Autonomous vehicle Network]

Suprabhat Maity[1], Debashish Chakravarty[1], Maaitrayo Das[2]

[1]Department of Mining Engineering, Indian Institute of Technology, Kharagpur, India
[2]Founder, Perceptech Solution, Kolkata.

ABSTRACT
This —The extraction of lane in a road from imagery has garnered significant attention in autonomous ground vehicle. In this study, we introduce semantic segmentation based ANN that leverages the combined residual learning strength and the UNet architecture for effective lane of driving area extraction for Autonomous Ground Vehicle [AGV]. This network is builds and constructed by residual modules, following a same structure to UNet. This model offers two primary advantages: first, the inclusion of residual units simplifies the training of DNN; second, the dense skip like connections enhance information flow, enabling a more compact architecture with improved performances efficiently. We evaluate our network using a publicly available road dataset [CULane] and benchmark it against UNet and two additional latest deep learning methods for road lane extraction. The proposed method outperforms almost all compared approaches, highlighting its superiority over recent advancements in the field. The result of Residual U-Net combined with our Save Net [1] for the Safe Zone Driving Space to represent the final result.

KEYWORDS
Lane Detection;
Autonomous Ground Vehicle (AGV);
Deep Neural Network (DNN);
DNN based Residual UNet;
SAVE-Net

1. INTRODUCTION

Lane detection & extraction is a foundational task in Autonomous Ground Vehicle, with applications spanning automatic road navigation, autonomous vehicles, obstacle avoidance, and safe autonomous driving complying with proper lane, among others. Despite significant research interest over the past decade, accurately extracting the lane from high resolution images remains challenging due to issues such as noises, occlusions, and the complex background present in raw imagery.

Recently, many models have been developed and deployed for lane detection [2] [3] from lane images, which generally fall into 2 categories: Extraction of lane area and center lane extraction. Lane area extraction techniques produce pixel-level lane labeling, whereas center lane extraction methods focus on identifying lane skeletons. Some approaches address both lane area [4] and center lane extraction simultaneously. Since center lane can be derived from lane areas using morphological thinning techniques, this study concentrates on lane extraction from high-resolution images. lane area extraction is essentially a segmentation or pixel level classification task. Traditional methods, such as those us ing shape index features and support vector machines (SVM), have been effective, but recent enhance in deep learning [5] [6] have pretty good amount of improved performance across various computer vision tasks, including remote sensing.

In the field of lane detection [7], early applications of deep learning, like those by Mnih and Hinton, used restricted Boltzmann machines (RBMs) to detect lane areas, employing pre- and post-processing to enhance results. Building on this,

Saito et al. later used convolutional neural networks (CNNs) to directly extract lanes from raw images, achieving better results on datasets like the Massachusetts roads dataset.

Deep neural networks [8] have shown that increased depth often yields better accuracy, but training deep architectures can be challenging due to issues like vanishing gradients. This was addressed by the deep learning framework in combination with residual module, which utilizes identity mappings to simplifying the ease of training. Similarly, UNet coined skip like connections to combine low- and high-level information to achieve strong segmentation performance.

Building on these insights, we propose the DNN with resid ual UNet (ResUnet), which combines residual learning with the UNet architecture. Unlike the original U-Net, ResUnet replaces plain neural units with residual units to enhance training and segmentation accuracy.

2. METHODOLOGY

(A) Deep ResUnet:

1. U-Net: In semantic approached segmentation, achieving fast & fine results requires combining low-level info with high level semantic details [9], [10]. However, efficiently training DNN for this purpose is challenging due to limited training data. One approach to address this is to fine-tune a pre-trained network on the target dataset [9], while another approach, as used in U-Net, involves extensive data augmentation [10]. Be yond augmentation, U-Net's architecture itself aids training by creating paths for information propagation. Copying low-level features to higher levels enables more efficient signal flow be tween

layers, which not only improves back-propagation while training but also integrates fine details with semantic features.

This approach aligns with concepts in residual networks [11]. Here, we demonstrate that substituting UNet's plain units with residual units further enhances its performance.

2. Residual Unit: Increasing the depth of a multi-layer ANN can enhance performance, which may be also complicate training and yields degradation issues [11]. To address this, He et al. [11] introduced the residual neural network, which uses residual units to ease training and mitigate degradation. Each residual unit includes a shortcut connection that adds the input to the learned residual function. This can be expressed as:

$$y_l = h(x_l) + F(x_l, W_l),\ x_{l+1} = f(y_l) \quad \text{----------(1)}$$

here x_l and x_l+1 towards IN & OUT of the l-th residual module, $F(\cdot)$ is the function for residual module, $f(y_l)$ is the activation module, and $h(x_l)$ represents an identity mapping, typically $h(x_l) = x_l$. Fig. 1 shows the difference between a plain & residual unit. Residual units combine layers of convo lution, batch normalization and ReLU activation in different configurations. He et al. [12] recommend a full pre-activation setup, which we adopt for constructing our deep residual UNet to enhance training stability and performance.

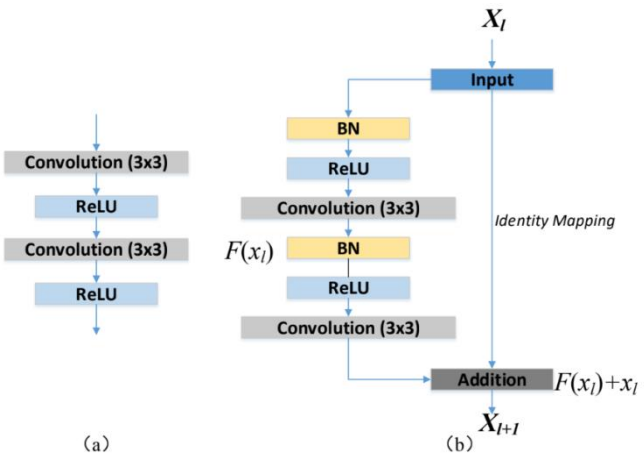

Fig.1 - ANN structure building. (a) Neural modue used in UNet and (b) proposed ResUnet having residual module incorporated with identity mapping.

3. Deep ResUnet: In this work, we present Deep ResUNet, a semantic segmentation architecture that integrates the advantages of UNet with residual-NN. This integration provides two main benefits: first, residual units facilitate easier and more stable network training; second, skip like connections both into the individual residual units and across low- and high level network stages ensure efficient information flow without degradation. As a result, the model maintains a compact architecture with fewer parameters while delivering comparable or improved performance in semantic segmentation tasks.

In this study, we adopt a 7-level Deep ResUNet architecture for extraction of lane, as depicted in Fig. 2. The network is composed of three primary modules: an encoder, a bridge, and a decoder. The encoder progressively converts the input image into compact feature representations, while the decoder restores these features into a pixel-level semantic segmentation map. The bridge serves as a connection between the encoding and decoding paths. All modules are built using residual units, each containing two 3×3 convolutional blocks along with an identity mapping. Each convolutional module consists of batch normalization, ReLU and a convolutional layer, with the identity mapping directly connecting the in-out of the residual unit.

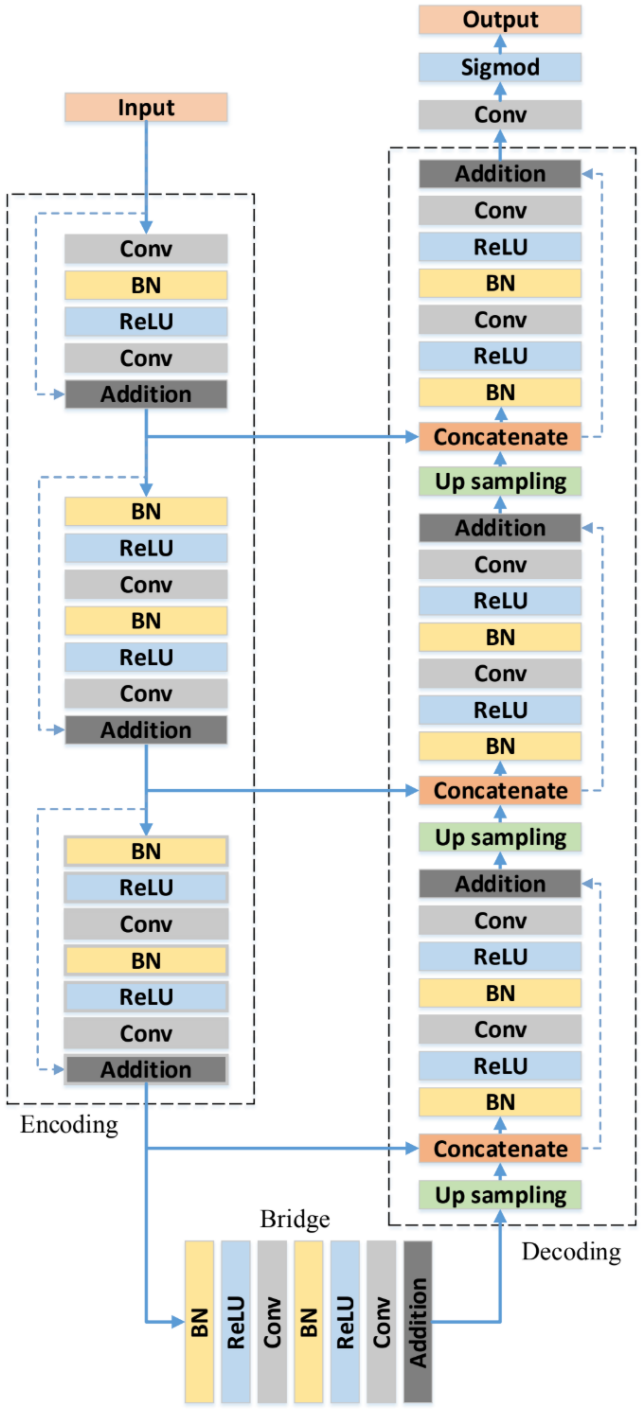

Fig.2 - The architecture of the proposed deep ResUnet.

The encoding path of the Deep ResUNet comprises three residual units. Rather than employing pooling operations for downsampling, feature map resolution is reduced by half by

incorrating in the first convolutional block with stride of 2 for each residual unit. The decoding path likewise construct by 3 residual module. Prior to every residual module in the decoder, feature maps in the lower level are upsampled and concatenated with the corresponding feature maps mapping to encoding path. At the final decoding stage, a 1 × 1 convolution followed by a sigmoid activation function transforms the multi-channel feature maps into the final segmentation output.

Our network includes 15 convolutional layer, compared to U-Net's 23 layers. Additionally, the cropping required in UNet is unnecessary in our design. The parameters and output dimensions at each layer are detailed in Table I.

TABLE I
RESUNET NETWORK LAYERS

Layer	Convolution	Filter	Stride	Output in Size
Input	-	-	-	224 × 224 × 3
Encoding				
Layer 1	Convo 1	3 × 3 / 64	1	224 × 224 × 64
	Convo 2	3 × 3 / 64	1	224 × 224 × 64
Layer 2	Convo 3	3 × 3 / 128	2	112 × 112 × 128
	Convo 4	3 × 3 / 128	1	112 × 112 × 128
Layer 3	Convo 5	3 × 3 / 256	2	56 × 56 × 256
	Convo 6	3 × 3 / 256	1	56 × 56 × 256
Bridge				
Layer 4	Convo 7	3 × 3 / 512	2	28 × 28 × 512
	Convo 8	3 × 3 / 512	1	28 × 28 × 512
Decoding				
Layer 5	Convo 9	3 × 3 / 256	1	56 × 56 × 256
	Convo 10	3 × 3 / 256	1	56 × 56 × 256
Layer 6	Convo 11	3 × 3 / 128	1	112 × 112 × 128
	Convo 12	3 × 3 / 128	1	112 × 112 × 128
Layer 7	Convo 13	3 × 3 / 64	1	224 × 224 × 64
	Convo 14	3 × 3 / 64	1	224 × 224 × 64
Output	Convo 15	1 × 1	1	224 × 224 × 1

(B) Loss Function

Given a module of images $\{I_i, s_i\}$ for training and their corresponding level-0 segmentations, our objective is to learn the network parameters W that enable accurate and robust road area segmentation. We achieve this by minimizing the loss difference the network's predicted segment Net(Ii;W) and the level-0 si. Here, we adopt Mean Squared Error (MSE) as the loss module, defined as:

$$L(W) = \frac{1}{N}\sum_{i=1}^{N} \|\text{Net}(I_i; W) - s_i\|^2 \quad \text{---------(2)}$$

where N represents the total sample training data. The network is trained using stochastic gradient descent (SGD). Notably, other differentiable loss functions can also be applied for network optimization; for example, U-Net employs pixel wise cross-entropy as its loss function to achieve model optimization.

3. DATASET USED

We adopt, the Culane dataset being used for experimental purposes [13]. The dataset contains 306 directories of each image (590 x 1640 x 3) is a RGB image. This data set only includes images and labels for one driver due to size limitations. It has 54k .Mp4 files.

4. EXPERIMENTAL RESULT

(A) Experimental Setup

The proposed methodology was evaluated on a system equipped with a 4-core CPU and 16 GB of DDR4 mem ory. Training of the CNN model was performed using an GPU(nvdia-rtx-2060) with GDDR6 vram (6 gb). During train ing, the initial learning rate was empirically set to 0.001 and subsequently reduced using a learning rate downing strategy called staircase, where rate of learning is decrease in every 10 epochs by factor of 10. The batch size for the proposed algorithm was experimentally chosen as 8. Overall procedure of the proposed algorithm is outlined in Algorithm 2.

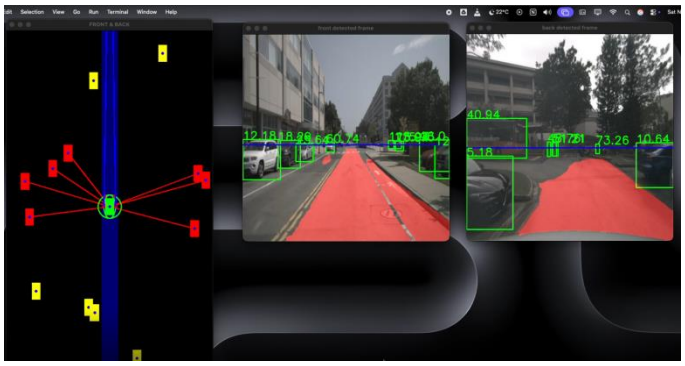

Fig. 3. The result of the proposed deep ResUnet.

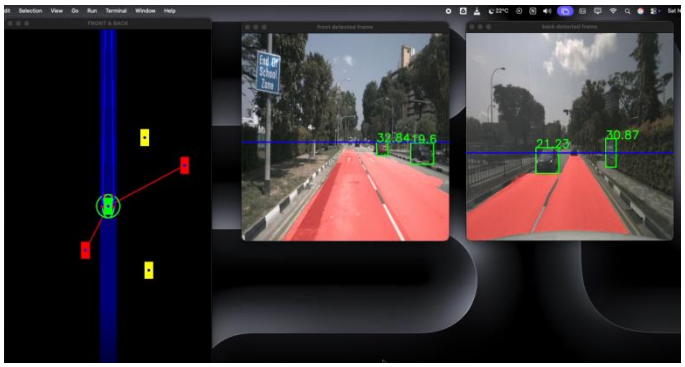

Fig. 4. The result of the proposed deep ResUnet.

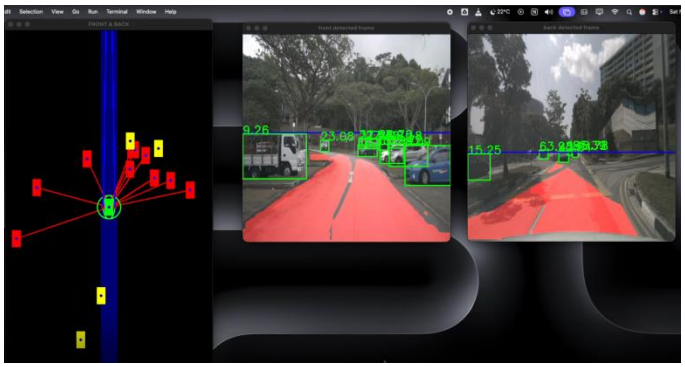

Fig. 5. The result of the proposed deep ResUnet.

Fig. 6. The result of the proposed deep ResUnet.

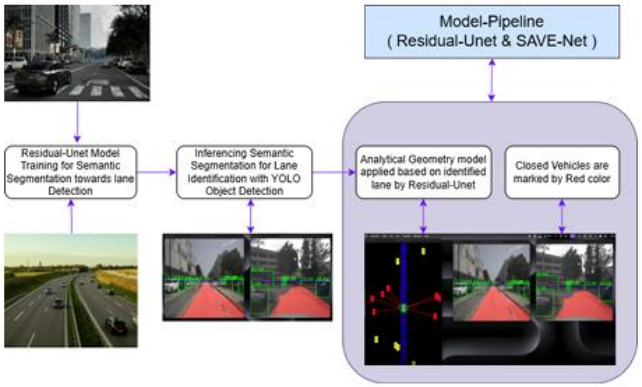

Fig. 7. The Model-Pipeline of the proposed Deep-ResUnet with SaveNet [1]

(B) Results Analysis

The experiment is divided into two sets: First, it is trying to capture the lane from semantic segmentation imposed by

Algorithm 1: Proposed algorithm for training

- Modify Unet with Residual unit.
- Evaluate the total loss employing the binary cross-entropy criterion.
- Adam Optimizer is used for cost function optimization.
- learning rate decay applied by staircase.
- Employ an early stopping strategy in which training is halted when the training loss shows no improvement over 10 successive epochs.
- Best model persisted based on properly optimized training loss.

Algorithm 2: Proposed Model Pipeline

- Inference with saved ResUnet trained model to identify lanes in road.
- Semantically segmented lane in a road attached with source image.
- Placed the lane lines based on the lane segmentation.
- Apply Analytical Geometry [1] model on top the ego-vehicle. • Used YOLO for detecting other vehicle and obstacle in the road.
- Displaying the closed vehicle or obstacle in Red color.

Residual-Unet, and in second part we apply the analytic geometry model [1] to ensure the safe zone for vehicle drive space. The experimental dataset was partitioned into training (11,739 samples), validation (1,677 samples), and testing (3,354 samples) subsets. The test set was treated as a blind dataset, meaning that its samples were not used during either the training or validation phases. This blind testing strategy enhances the robustness of the proposed system. The experimental results are summarized in Table II, showing training, validation, and blind test accuracies of 98.8%, 98.6%, and 98.5%, respectively. The performance of the model was further validated through the evaluation of precision, recall, and F1-score. Additionally, the ROC curve and confusion matrix were analyzed to illustrate the extent of correctly classified samples across each class.

$$\text{Sensitivity or Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$\text{Specificity} = \frac{TN}{FP + TN} \tag{4}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\text{F1-Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

Lane detection is a semantic segmentation task in which each pixel of an image is classified as belonging to a lane or background. The choice of segmentation model significantly affects accuracy, inference speed, and robustness under challenging driving conditions.

1) Comparison of Semantic Segmentation Models:

2) Comparative Insights:

3) Accuracy vs. Speed: DeepLabv3+, SCNN, and UltraFast-Lane-Detection provide the highest accuracy but differ in computational cost. ENet and ERFNet are ideal for real-time edge deployment, while UltraFast achieves state-of-the-art real-time performance.

4) Robustness in Challenging Conditions: SCNN performs better under occlusions, lane curvature, and shadows due to its spatial propagation mechanism. DeepLabv3+ and PSPNet handle variable illumination effectively because of multi-scale context aggregation.

TABLE II
COMPARISON STUDY

Model	Blind Test Accuracy
VGG19 [14]	96.7%
InceptionV3 [15]	97.2%
Mobilenet [16]	97.7%
Resnet50 [17]	98.3%
DenseNet121 [18]	98.7%
Unet [10]	90.5%
Proposed ResUnet	*98.8%*

5) Ease of Training and Dataset Dependence: UNet and SegNet are easier to train on smaller datasets, whereas DeepLabv3+ requires more data and computational power.

6) Recommended Models by Scenario:

7) Benchmark Results on the CULane Dataset: For re search and maximum accuracy, SCNN or DeepLabv3+ are strong choices. For production-level real-time systems, UltraFast-Lane-Detection or ERFNet-SAD are preferred. For resource-constrained embedded platforms, ENet or UNet are efficient and reliable solutions.

TABLE III
COMPARISON OF SEMANTIC SEGMENTATION MODELS FOR LANE DETECTION.

Model	Architecture Type	Accuracy (mIoU/F1)	FPS
ENet	Lightweight encoder-decoder	65–70%	60–80
SegNet	Encoder-decoder (VGG-based)	70–75%	15–25
UNet	Symmetric encoder-decoder	75–80%	25–40
DeepLabv3+	ASPP with encoder-decoder	80–85%	15–25
PSPNet	Pyramid pooling module	78–83%	10–20
ERFNet	Efficient residual CNN	72–78%	40–60
LaneNet	Dual-branch segmentation + embedding	75–80%	25–40
SCNN	Spatial CNN with message passing	88% (CULane)	20
ENet-SAD / ERFNet-SAD	Attention distillation on lightweight nets	85%	50–60
UltraFast-Lane-Detection	Anchor-based + ResNet encoder	90% (CULane)	160
Proposed ResUnet	Symmetric encoder-decoder	90–95%	160

TABLE IV
SEMANTIC SEGMENTATION MODELS WITH USE CASES.

Model	Strengths	Weaknesses	Use Case
ENet	Fast, real-time performance	Lower accuracy on complex lanes	Embedded, edge devices
SegNet	Good generalization	High computational cost	Research, offline analysis
UNet	Good for small datasets	Not optimized for speed	Moderate real-time tasks
DeepLabv3+	Excellent accuracy, contextual understanding	Slow inference	High-end GPU deployment
PSPNet	Strong contextual reasoning	Computationally expensive	Offline processing
ERFNet	Good trade-off between speed and accuracy	Moderate robustness	Real-time systems
LaneNet	Detects individual lanes	Sensitive to shadows/curves	Benchmarking
SCNN	Handles occlusion and curvature	Complex architecture	Advanced ADAS
ENet-SAD / ERFNet-SAD	Efficient and accurate	Harder training	Real-time driving
UltraFast-Lane-Detection	Extremely fast, accurate	Limited lane topology flexibility	Real-time self-driving
Proposed ResUnet	Good for generalization	Optimized for speed	Moderate real-time tasks

TABLE V
RECOMMENDED MODELS FOR DIFFERENT DEPLOYMENT SCENARIOS.

Scenario	Recommended Model(s)	Rationale
Real-time edge deployment	ERFNet-SAD / ENet-SAD / UltraFast-Lane-Detection	Balance between speed and accuracy
High-accuracy offline processing	DeepLabv3+ / SCNN / Residual-Unet	Strong semantic and spatial reasoning
Low-resource embedded systems	ENet / UNet	Lightweight and efficient
Research & benchmarking	LaneNet / SCNN	Interpretable and high-performing

TABLE VI
PERFORMANCE COMPARISON ON THE CULANE DATASET.

Model	F1 (%)	FPS	Remarks
SCNN	88.4	20	High accuracy
ENet-SAD	85.4	60	Efficient, accurate
ERFNet-SAD	86.5	50	Balanced performance
UltraFast-Lane-Detection	90.3	160	State-of-the-art real-time
DeepLabv3+	84.2	18	Very accurate but slow
Proposed ResUnet	91.8	160	Accurate, State-of-the-art

(C) Comparison

Residual-UNet [19] [20] [21] [22] builds upon the strengths of the classic UNet model by incorporating residual blocks within its encoder and decoder. These residual connections allow to flow gradients more easily inside the network, pre venting vanishing-gradient issues and enabling deeper feature extraction. This helps the model better distinguish lane mark ings from the background, even in complex environments.

Key advantages of Residual-UNet include:

- Improved feature propagation through skip and residual connections
- Better learning of fine-grained details, essential for thin structures like lane lines
- Robustness in challenging scenarios (night, rain, shad ows, occlusions)
- Higher segmentation accuracy compared to standard UNet

Standard UNet [23] [24] [25] [26] is widely used for seman tic segmentation because of the encoder–decoder framework combined with skip connections. While it provides reason able performance for lane detection, it lacks the ability to efficiently learn deep hierarchical representations. Residual UNet surpasses standard UNet by:

- Providing improved convergence
- Reducing loss more quickly
- Preserving lane continuity with fewer broken predictions

DeepLabV3+ [27] [28] [29] [30] utilizes atrous spatial pyramid pooling (ASPP) to capture features at multiple scales, making it strong at handling large-scale context. While it performs well on complex backgrounds, it may struggle with extremely thin lane markings without additional refinement. Residual-UNet typically offers:

- Sharper outputs for thin structures
- Lower computational cost
- However, DeepLabV3+ may outperform Residual-UNet in scenes with heavy background clutter due to its larger receptive field.

ENet / Fast-SCNN [31] [32] [33] (Lightweight Models) Lightweight segmentation networks offer high inference speed for real-time applications, especially on embedded hardware. However, this efficiency comes at the cost of reduced accuracy. Compared to these models, Residual-UNet provides:

- Higher accuracy and robustness
- Better detection of subtle lane boundaries Though slower, it offers superior reliability for safety-critical systems.

Attention-UNet [34] [35] [36] mechanisms enhance focus on relevant spatial regions, improving segmentation when targets are small. Attention-UNet may outperform Residual UNet in certain noisy environments, but at a computational expense. Residual-UNet remains more efficient while achiev ing competitive accuracy.

Overall Findings- Across different model comparisons, Residual-UNet maintains an optimal trade-off among accu racy, stability, and computational performance. Its ability to capture detailed lane structures, while maintaining robustness under varied conditions, makes it an excellent choice for lane detection tasks. Although some advanced architectures offer better multi-scale awareness or faster inference, Residual-UNet remains highly effective in scenarios where precision and structural consistency are prioritized.

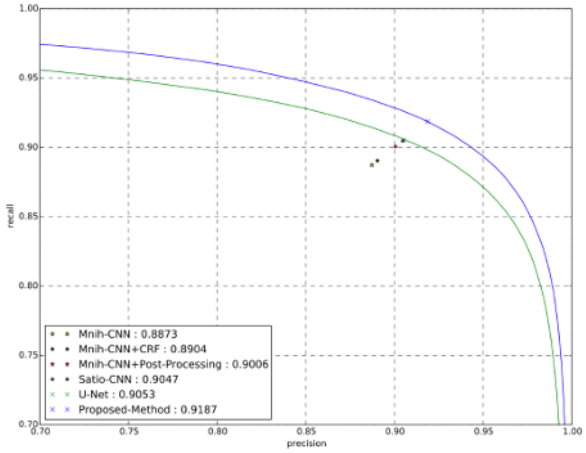

Fig. 8. The relaxed precision-recall curves of U-Net and the proposed method on Culane dataset. The marks '?' and '×' are break-even points of different methods.

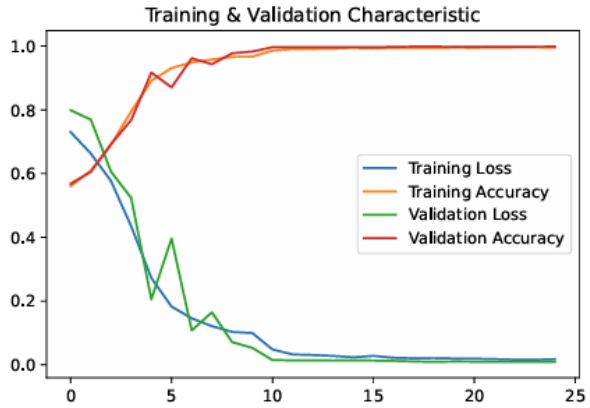

Fig. 9. Training & Validation Characteristics.

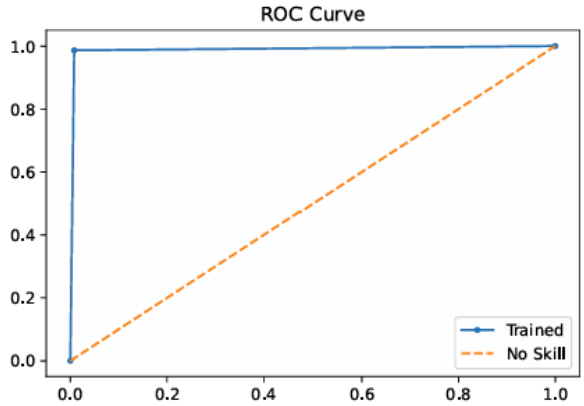

Fig. 10. ROC Curve.

Comparative evaluations with three leading deep learning–based lane extraction methods were conducted on the CULane test dataset. The break-even points of the proposed method and the competing approaches are reported in Table II. Figure 3 illustrates the relaxed precision–recall curves of U-Net and the proposed network, along with their respective break-even points, as well as those of the comparison methods. The results show that the proposed approach outperforms all three competing methods in terms of relaxed precision and recall. Notably, despite having only one-quarter of the parameters of U-Net (7.8M versus 30.6M), the proposed network achieves significant performance improvements in lane extraction.

5. CONCLUSION

In this study, lane detection is performed on road video signals, with the training data augmented by incorporating video noise to achieve more robust performance. The im plementation of a Residual-UNet model for lane detection demonstrates strong capability in accurately segmenting lane markings under diverse driving conditions. By integrating residual connections into the traditional U-Net architecture, the model effectively mitigates vanishing-gradient issues, improves feature propagation, and enhances the network's ability to learn complex spatial patterns. As a result, the Residual UNet achieves higher robustness and precision compared to standard segmentation models, particularly in challenging scenarios such as shadows, low illumination, occlusions, and worn-out lane lines.

Overall, the model proves to be a reliable and efficient solution for real-time lane detection, making it effective for integration into advanced driver-assistance systems (ADAS) and autonomous driving frameworks. Future improvements may include incorporating temporal information, lightweight model optimization for embedded deployment, and fusion with additional sensor data to further enhance accuracy and stability. The proposed model demonstrates strong performance in terms of blind test accuracy. Thus, the model can be used as an advisor for lane detection for Autonomous Ground Vehicle [AGV]. In addition, ensembling road video signals with other conditions such as night and rain videos can lead to further improvements in detection efficiency.

6. REFERENCES

[1] S. Maity, D. Chakravarty, and M. Das, "A novel approach for safe zone drive space of autonomous ground vehicle (self-driving car), international journal of microsystems and iot, 1(4), 231–240. https://doi.org/10.5281/zenodo.10006837," 2023.

[2] W. Liu, Y. Liu, W. Meng, G. He, and J. Li, "D3l: Curvature-constrained denoising diffusion model for 3d lane detection," in Proceedings of the 33rd ACM International Conference on Multimedia, 2025, pp. 4923 4931.

[3] M. Pittner, J. Janai, M. Faigle, and A. P. Condurache, "Sparselanestp: Leveraging spatio-temporal priors with sparse transformers for 3d lane detection," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2025, pp. 29099–29109.

[4] B.-L. Huang, Z.-X. Ni, F.-K. Huang, H.-H. Shuai, and W.-H. Cheng, "When anchors meet cold diffusion: A multi-stage approach to lane detection," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2025, pp. 27917–27926.

[5] Y. Zhao, L. Han, J. Chen, F. Liang, H. Tang, J. Wu, and Z. Zhang, "Real time lane detection based on dual attention mechanism transformer for adas/ad," IEEE Transactions on Vehicular Technology, 2025.

[6] Y. Deng, M. Wang, M. Zhang, and M. Zhou, "A hybrid network of cnn and transformer for 3d lane detection," IEEE Access, 2025.

[7] Q. Ning, J. Zhang, Y. Xie, K. Liu, K. Gao, B. Chen, G. Chen, Q. Fan, H. Liu, and R. Du, "Multi-resolution context augmentation and dual channel attention for 3d lane detection," IEEE Internet of Things Journal, 2025.

[8] Z. Hou, Y. Cao, N. Chen, C. Ren, C. Lin, C. Lang, and Y. Li, "Ada3dlane: Adaptive 3d lane detection from monocular images," IEEE Transactions on Intelligent Transportation Systems, 2025.

[9] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in

Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international con ference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer, 2015, pp. 234–241.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[12] ——, "Identity mappings in deep residual networks," in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. Springer, 2016, pp. 630–645.

[13] "Culane," https://www.kaggle.com/datasets/greatgamedota/culane.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015. 5 10 15 ROC Curve 0.2 0.4 0.6 Fig. 10. ROC Curve.

[15] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," 2015. Training Loss Training Accuracy Validation Loss Validation Accuracy 20 Trained No Skill 0.8 1.0

[16] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. 25

[18] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2018.

[19] Y. Lan and X. Zhang, "Real-time ultrasound image despeckling using mixed-attention mechanism based residual unet," IEEE Access, vol. 8, pp. 195327–195340, 2020.

[20] L. Li, C. Wang, H. Zhang, and B. Zhang, "Residual unet for urban building change detection with sentinel-1 sar data," in IGARSS 2019 2019 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2019, pp. 1498–1501.

[21] K. Niu, Z. Guo, X. Peng, and S. Pei, "P-resunet: Segmentation of brain tissue with purified residual unet," Computers in Biology and Medicine, vol. 151, p. 106294, 2022.

[22] S.-T. Tran, M.-H. Nguyen, H.-P. Dang, and T.-T. Nguyen, "Automatic polyp segmentation using modified recurrent residual unet network," IEEE Access, vol. 10, pp. 65951–65961, 2022.

[23] U. Javaid, D. Dasnoy, and J. A. Lee, "Multi-organ segmentation of chest ct images in radiation oncology: comparison of standard and dilated unet," in International conference on advanced concepts for intelligent vision systems. Springer, 2018, pp. 188–199.

[24] S. Liu, Z. Huang, Z. Xu, F. Zhao, D. Xiong, S. Peng, and J. Huang, "High-throughput measurement method for rice seedling based on improved unet model," Computers and Electronics in Agriculture, vol. 219, p. 108770, 2024.

[25] W. Zhu, X. Chen, P. Qiu, M. Farazi, A. Sotiras, A. Razi, and Y. Wang, "Selfreg-unet: Self-regularized unet for medical image segmentation," in International Conference on Medical Image Computing and Computer Assisted Intervention. Springer, 2024, pp. 601–611.

[26] S. Guan, A. A. Khan, S. Sikdar, and P. V. Chitnis, "Fully dense unet for 2-d sparse photoacoustic tomography artifact removal," IEEE journal of biomedical and health informatics, vol. 24, no. 2, pp. 568–576, 2019.

[27] H. Peng, C. Xue, Y. Shao, K. Chen, J. Xiong, Z. Xie, and L. Zhang, "Semantic segmentation of litchi branches using deeplabv3+ model," Ieee Access, vol. 8, pp. 164546–164555, 2020.

[28] L. Yu, Z. Zeng, A. Liu, X. Xie, H. Wang, F. Xu, and W. Hong, "A lightweight complex-valued deeplabv3+ for semantic segmentation of polsar image," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 930–943, 2022.

[29] C. Wang, P. Du, H. Wu, J. Li, C. Zhao, and H. Zhu, "A cucumber leaf disease severity classification method based on the fusion of deeplabv3+ and u-net," Computers and electronics in agriculture, vol. 189, p. 106373, 2021.

[30] Y. Wang, L. Yang, X. Liu, and P. Yan, "An improved semantic seg mentation algorithm for high-resolution remote sensing images based on deeplabv3+," Scientific reports, vol. 14, no. 1, p. 9716, 2024.

[31] R. P. Poudel, S. Liwicki, and R. Cipolla, "Fast-scnn: Fast semantic segmentation network," arXiv preprint arXiv:1902.04502, 2019.

[32] Y. Luo, Z. Rao, and R. Wu, "Fd-slam: a semantic slam based on enhanced fast-scnn dynamic region detection and deepfillv2-driven back ground inpainting," IEEE Access, vol. 11, pp. 110615–110626, 2023.

[33] S. Zhou, Y. Chen, X. Li, and A. Sanyal, "Deep scnn-based real-time object detection for self-driving vehicles using lidar temporal data," Ieee Access, vol. 8, pp. 76903–76912, 2020.

[34] K. Trebing, T. Staczyk, and S. Mehrkanoon, "Smaat-unet: Precipitation nowcasting using a small attention-unet architecture," Pattern Recogni tion Letters, vol. 145, pp. 178–186, 2021.

[35] N. Das and S. Das, "Attention-unet architectures with

pretrained back bones for multi-class cardiac mr image segmentation," Current Problems in Cardiology, vol. 49, no. 1, p. 102129, 2024.

AUTHORS

Suprabhat Maity received his B.Tech degree in Computer Science and engineering from University Institute of Technology, Burdwan University India in 2004 and MTech degree in Computer Science from Jadavpur University, Kolkata, India in 2009. He is currently pursuing PhD in the Department of Mining Engineering, Indian Institute Technology Kharagpur, India. His areas of interest are Computer Vision, artificial intelligence, machine learning and human computer interaction, deep learning, NLP.

E-mail: suprabhatmaity@gmail.com

Debashish Chakravarty received his B.Tech degree in Mining engineering from Indian Institute of Technology Kharagpur, India in 1993 and MTech degree in Mining Engineering from Indian Institute of Technology, Kharagpur, India in 1995. He did his PhD at the Department of Mining Engineering, Indian Institute of Technology, Kharagpur, India in Digital Image Processing and AI in Rock Mechanics. His areas of interest are Mine Automation, AI, Geo-Resouce Data Analytics for Productivity & Safety, SLAM, IIoT, Locational Surveillance for Environment and Energy Optimization, Rescue Robotics and Robotic Mapping (Aerial UAV and Terrestrial), Autonomy and Intelligence (Autonomous Systems & Autonomous Coordination).

E-mail: profdciitkgp@gmail.com

Maaitrayo Das received his B.Tech degree in Computer Science & Engineering from B. P. Poddar Institute of Management & Technology, Kolkata India in 2022. His areas of interest are Embedded system, artificial intelligence, machine learning and human computer interaction.

E-mail: dasmaaitrayo@gmail.com

,