

# Optimizing Lane Detection in Autonomous Vehicles Using Cascading Attention Mechanisms in DeepLabv3+

**Dewiani**

Department of Electrical Engineering, Hasanuddin University, Makassar, Indonesia  
dewiani@unhas.ac.id

**Indrabayu**

Department of Informatics, Hasanuddin University, Makassar, Indonesia  
indrabayu@unhas.ac.id (corresponding author)

**Muhammad Abdillah Rahmat**

Department of Informatics, Hasanuddin University, Makassar, Indonesia  
abdi.rahmat@unhas.ac.id

**A. Ichsan Mudatsir Lukman**

Department of Informatics, Hasanuddin University, Makassar, Indonesia  
lukmanaim20d@student.unhas.ac.id

Received: 6 October 2025 | Revised: 3 November 2025 | Accepted: 15 November 2025

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.15315>

## ABSTRACT

The high rate of traffic accidents caused by human error highlights the urgent need for reliable lane detection in autonomous vehicles. Traditional segmentation models, such as U-Net, struggle to capture multi-scale contextual features, leading to inaccurate recognition of narrow or visually ambiguous lane markings. This study introduces an enhanced DeepLabv3+ architecture, augmented with cascading attention mechanisms, Convolutional Block Attention Module (CBAM), Efficient Channel Attention (ECA), and Squeeze-and-Excitation (SE), to improve feature extraction and boundary precision. The proposed method addresses limitations in existing models by leveraging these attention modules to better capture contextual information at different scales. A dataset consisting of 374 annotated road images from Makassar, Indonesia, was used for training and evaluation. The model achieved a mean Intersection over Union (IoU) of 97.34% and a Root Mean Square Error (RMSE) of 0.0377 in tire-to-lane distance estimation, outperforming traditional architectures. These results demonstrate that the proposed framework provides robust, real-time lane detection, making it highly suitable for autonomous vehicle navigation in dynamic and complex urban environments.

*Keywords- autonomous car; road line; semantic segmentation; DeepLabv3+*

## I. INTRODUCTION

The high incidence of road traffic accidents remains a critical public safety issue, drawing serious attention from both governmental authorities and academic researchers. Statistics consistently show that approximately 61% of traffic accidents are caused by human error, including fatigue and diminished concentration [1]. This reality underscores an urgent need for advanced, technology-based safety systems, especially in autonomous vehicles, to reduce reliance on human drivers. Among the key features of such vehicles is the ability to accurately detect and track road lanes to ensure safe path adherence. Previous studies utilizing segmentation

architectures like U-Net have shown limitations in performance, with an overall accuracy of only 79.8%, and particularly low recognition accuracy for solid lane markings (54.8%) and dashed lane markings (75.3%) [2]. Additionally, U-Net processes segmentation frame-by-frame without accounting for temporal continuity and struggles to distinguish visually similar line features [2]. These shortcomings highlight the necessity for a more robust and context-aware segmentation model that can better interpret dynamic road environments.

The body of literature in computer vision and autonomous driving systems has extensively discussed the challenges of accurate lane detection, especially under complex real-world

conditions [3]. Most previous approaches have relied on Convolutional Neural Networks (CNNs) for image segmentation tasks, yet often fall short in addressing critical issues such as temporal continuity and visual ambiguity in lane markings [4]. Models like U-Net and Fully Convolutional Networks (FCNs) have gained popularity, but their static spatial focus limits their effectiveness in real-time driving scenarios [5]. A recurring limitation is the inability to differentiate between visually similar patterns, especially when dealing with inconsistent or worn-out lane markings [2]. These models typically emphasize spatial information while neglecting temporal context or multi-scale feature representation. Thus, there remains a significant gap in the literature for an approach capable of combining deep spatial awareness with robust multi-scale contextual understanding. Addressing this gap is crucial to advancing lane detection systems for autonomous vehicles operating in dynamic and unpredictable road conditions.

This study aims to design and optimize a road lane detection and tracking system for autonomous vehicles using the DeepLabv3+ architecture [4]. The system is engineered to significantly enhance segmentation accuracy compared to prior methods. DeepLabv3+ was selected for its ability to capture wider spatial and contextual information through atrous convolution and Atrous Spatial Pyramid Pooling (ASPP) [6]. The model is also evaluated specifically for its performance in detecting both solid and dashed lane markings across varied road scenarios. Beyond lane segmentation, the system incorporates a distance estimation module that measures the gap between the vehicle's tire and the nearest lane marking using Inverse Perspective Mapping (IPM) [7, 8]. The experimental dataset includes 374 annotated images from urban

streets in Makassar, with a training-to-testing ratio of 336 to 38 [2]. Evaluation metrics such as Intersection over Union (IoU) [9] and Root Mean Square Error (RMSE) [10] are employed to assess segmentation accuracy and distance estimation reliability. This integrated approach is expected to provide a practical and high-precision solution for visual navigation in autonomous driving systems.

To further optimize the DeepLabv3+ architecture for autonomous driving, several modifications were introduced. The modified model incorporates a ResNet-18 backbone for better feature extraction [11], along with Convolutional Block Attention Module (CBAM) to improve the model's ability to focus on the most relevant parts of the image [12]. Additionally, Efficient Channel Attention (ECA) layers and Squeeze-and-Excitation (SE) blocks have been integrated to enhance the model's attention mechanism [13, 14], ensuring the system captures fine-grained lane markings. The decoder has also been modified with upsampling layers for more precise boundary detection and to maintain high-resolution outputs [4]. This enhanced architecture is designed to address the shortcomings of traditional models, offering superior performance in real-time lane detection and tracking tasks.

Unlike standard DeepLabv3+ implementations, the proposed model introduces a cascading integration of CBAM, ECA, and SE attention modules, combined with an RMSE-based distance estimation component, which together form a novel framework for real-time and high-precision lane tracking.

Figure 1 provides an overview of the modified DeepLabv3+ architecture, which illustrates the structural components of the model, including the encoder-decoder framework, attention modules, and upsampling layers.

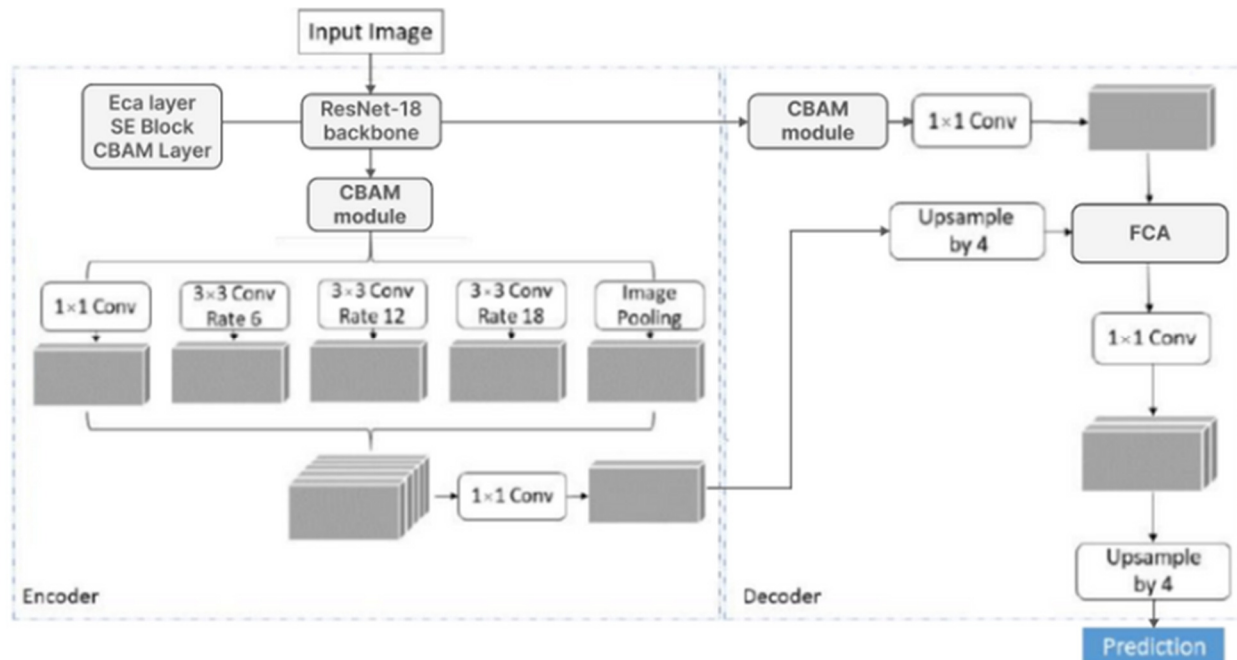


Fig. 1. Overview of the modified DeepLabv3+ architecture for lane detection with integrated attention modules; abbreviations: ECA (Efficient Channel Attention), SE (Squeeze-and-Excitation), CBAM (Convolutional Block Attention Module), FCA (Feature Cross Attention), and Conv (Convolution).

The urgency of this research stems not only from alarming traffic safety data [1] but also from the inadequacies of existing vision-based detection systems. The central hypothesis is that DeepLabv3+ can effectively overcome the shortcomings of earlier models like U-Net, especially in handling complex lane segmentation and tracking tasks [2]. By integrating spatial and contextual awareness with high-resolution image processing, DeepLabv3+ is expected to deliver superior segmentation and tracking performance in real-time conditions. The stated objectives serve as a strategic framework to guide both model design and performance evaluation, focusing on accuracy, reliability, and adaptability. Rooted in the empirical gaps of earlier studies and the technological demands of next-generation mobility, this research is positioned as a necessary step toward achieving more dependable and intelligent autonomous vehicle navigation systems.

This study highlights the key aspects of the lane detection and tracking system, focusing on the performance and evaluation metrics of the DeepLabv3+ architecture. The results indicate a significant improvement in the segmentation accuracy and distance estimation accuracy compared to previous models like U-Net. The evaluation metrics, including IoU and RMSE, demonstrate the system's ability to detect lane markings accurately and estimate the vehicle's position in real-time driving conditions. Additionally, the model shows a high precision in distinguishing between solid and dashed lane markings, crucial for the reliable operation of autonomous vehicles in dynamic environments.

Table I provides a detailed breakdown of the research findings, including specific research objectives, evaluation indicators, and the corresponding results.

TABLE I. RESEARCH FINDINGS BASED ON RESEARCH OBJECTIVES

No	Research objective	Evaluation indicator	Research findings
1	To design and optimize a lane detection and tracking system using DeepLabv3+	Tire-to-lane distance estimation (RMSE)	RMSE: 0.0377119856 – indicates high accuracy in estimating the vehicle's position relative to lane markings
		Real-time lane detection	The system successfully performed stable and accurate lane tracking under daylight and clear weather conditions
		Visual response to lane contour	The system effectively distinguished between solid and dashed lane markings in real-world road images
2	To analyze the semantic segmentation performance of DeepLabv3+ in lane detection	Segmentation accuracy (IoU)	Overall IoU: 97.34%, a significant improvement over U-Net (79.8%)
		Class-specific lane marking segmentation	High precision in recognizing both solid and dashed lane markings, outperforming earlier models
		Alignment of model prediction with ground truth	Segmentation outputs consistently matched manual annotations from the Pascal VOC labeled dataset

## II. LITERATURE REVIEW

### A. Concept Definition of Autonomous Car

An autonomous car is a vehicle system designed to operate, navigate, and make driving decisions independently, without direct human intervention [15]. This technology represents a groundbreaking advancement in modern transportation, integrating artificial intelligence, sensor systems, computer vision, and real-time decision-making algorithms. At the core of an autonomous vehicle lies its ability to process environmental data captured through sensors such as cameras, radar, and LiDAR, allowing it to perform critical driving functions including steering, acceleration, braking, and lane selection [3, 16]. These systems aim to enhance road safety, energy efficiency, and user comfort. The vehicle's visual perception capabilities particularly its ability to detect road boundaries, signs, and surrounding traffic are essential to this process [16]. Consequently, accurate and real-time lane detection becomes vital to ensure safe path adherence [17]. With the rapid evolution of deep learning and computer vision technologies, autonomous vehicles are increasingly becoming viable for real-world applications, especially in creating safer and more Intelligent Transportation Systems (ITSs). Recent studies in the field of autonomous driving have demonstrated the effectiveness of deep learning-based perception systems in enhancing object detection accuracy under real-world conditions. For example, authors in [18] a recent study introduced a YOLO-based network with image-dehazing and channel-attention modules to improve object detection under adverse weather conditions, significantly increasing detection robustness in fog, rain, and haze. Authors in [19] proposed an improved YOLOv8-based detection by merging data from diverse harsh-weather datasets, enhancing detection accuracy in snowy, rainy, foggy, and night-time driving scenarios. These advancements highlight the growing role of optimized vision models in ensuring robust environmental perception for self-driving systems.

### B. Categorization/Manifestation of Autonomous Car

Autonomous vehicles are categorized into six automation levels as defined by the Society of Automotive Engineers (SAE) in the J3016 standard [15]. Each level specifies the degree of vehicle control performed without human involvement. Level 0 represents no automation, in which the human driver is responsible for all tasks. Levels 1 and 2 are classified as partial automation, supporting specific driving functions like cruise control or lane keeping while still requiring full driver attention. Level 3 allows the vehicle to handle driving tasks in certain scenarios but can request the driver to intervene. At Level 4, the vehicle operates entirely autonomously within defined environments without human input. Level 5 represents full autonomy, where the vehicle requires no human control in any condition (SAE International, 2016). The manifestation of autonomous vehicles extends beyond control systems to include integrated sensor arrays, real-time road mapping, and advanced computer vision algorithms. In the context of this research, lane detection and tracking capabilities align with Level 3 and above, where precise visual perception plays a pivotal role in automated vehicle navigation.

### C. Concept Definition of Semantic Segmentation

Semantic segmentation is a core technique in computer vision that classifies each pixel in an image into predefined semantic categories such as roads, vehicles, pedestrians, or lane markings [20]. Unlike image-level classification, which assigns a single label to an entire image, semantic segmentation delivers pixel-wise classification, offering a detailed understanding of an image's structural composition. In autonomous vehicles, this capability is crucial, enabling the system to distinguish between different road elements and contextualize its surroundings. The advantage of semantic segmentation lies in its ability to address real-world visual challenges, such as varying lighting, occlusions, and object shape variations [21]. Semantic segmentation models are typically built on CNN architectures that extract spatial features from visual data [22, 23]. These models convert raw visual input from vehicle cameras into semantic maps, providing a richer, context-aware foundation for decision-making systems. Therefore, semantic segmentation not only enhances visual interpretation but serves as a foundational pillar for real-time navigation in autonomous driving [24].

### D. Categorization/Manifestation of Semantic Segmentation

Semantic segmentation can be manifested through several architectural and application approaches. Broadly, segmentation is categorized into two types: semantic segmentation and instance segmentation. While semantic segmentation assigns the same label to all objects of a class, instance segmentation differentiates between individual objects within the same class (e.g., distinguishing two separate cars). To achieve this, deep learning models like FCNs, U-Net, and SegNet have been widely used. These architectures excel in maintaining spatial resolution and generating pixel-level classifications [5]. In the context of autonomous vehicles, semantic segmentation helps detect crucial areas such as road surfaces, lane markings, sidewalks, and obstacles. The outputs are typically integrated with additional modules such as IPM for geometric transformation and lane tracking algorithms to maintain safe lane positioning [7]. Hence, the practical manifestation of semantic segmentation in autonomous systems lies in its contribution to accurate visual perception and safe vehicle maneuvering.

### E. Concept Definition of DeepLabv3+

DeepLabv3+ is a specialized deep learning architecture designed for complex and high-precision semantic segmentation tasks. As the most advanced iteration of the DeepLab series, it integrates a decoder module to refine spatial resolution in segmented outputs. Its core components include atrous convolution (also known as dilated convolution) and ASPP. Atrous convolution expands the receptive field without losing resolution, enabling the model to capture both fine and global contextual information [4]. ASPP enhances multi-scale feature extraction by pooling information at various scales, resulting in richer semantic understanding. DeepLabv3+ typically utilizes backbones such as Xception, ResNet, or MobileNet to extract foundational image features. Its superior performance has been demonstrated on several benchmark datasets, including PASCAL VOC and Cityscapes. In this study, DeepLabv3+ was chosen specifically for its ability to

detect small and narrow objects like lane markings elements crucial to the accurate interpretation of road layouts by autonomous vehicles.

### F. Categorization/Manifestation of DeepLabv3+

The manifestation of DeepLabv3+ in practical applications is seen through its modular encoder-decoder structure and adaptability to complex environments. The encoder uses a backbone network (e.g., ResNet, Xception) to extract spatial features from input images, followed by the ASPP module to capture global context and multi-scale representations. The decoder then upsamples these features to reconstruct a high-resolution segmentation map, preserving spatial detail. DeepLabv3+ can also be extended with auxiliary modules like attention mechanisms (e.g., CBAM) [11], or integrated with geometric transformation tools such as IPM [7], enabling more stable and accurate tracking. Its implementation spans multiple domains, including medical image analysis, remote sensing, and autonomous driving. In this research, DeepLabv3+ was integrated with lane tracking and distance estimation modules using IPM and RMSE to form a comprehensive visual navigation system. Its flexibility and high segmentation accuracy make it well suited for real-world deployment in autonomous vehicle systems operating under diverse road conditions.

Beyond the theoretical development of DeepLabv3+, recent studies have also advanced vision-based perception and semantic segmentation. Authors in [25] implemented a residual U-Net for road segmentation, whereas authors in [26] applied deep CNNs for detecting surface anomalies in autonomous driving. Authors in [27] used deep neural networks to detect road damage from smartphone images, showing that vision-based models can be applied effectively to real roadway conditions, and authors in [28] applied semantic segmentation techniques to extract urban road structures from high-resolution remote-sensing data, further demonstrating the versatility of convolutional architectures in spatial analysis. Although these works mark notable progress, most focus on general or infrastructure-level applications rather than lane-specific real-time segmentation. The present study builds upon these advances by optimizing DeepLabv3+ with multi-attention modules and distance estimation to enhance lane detection performance under dynamic conditions.

## III. METHODOLOGY

### A. Research Object and Problem Background

This research is grounded in the pressing issue of road traffic accidents, with approximately 61% caused by human negligence particularly due to fatigue and decreased concentration. This alarming statistic underscores the urgent need for intelligent safety systems in autonomous vehicles, especially those focused on accurate lane detection. One of the critical components of autonomous driving is the ability to detect and track lane markings in real-time to maintain the vehicle's position within a safe path. Previous studies, such as those utilizing U-Net architecture for segmentation tasks, have shown notable limitations. U-Net achieved only 79.8% overall accuracy, with performance for solid lane lines at just 54.8% and dashed lines at 75.3%. Moreover, U-Net processes images

frame-by-frame without considering temporal information between frames, making it ineffective in identifying lane patterns with similar visual features. These constraints highlight the necessity of employing a more advanced and context-aware segmentation method, one capable of better interpreting visual cues in dynamic driving environments. DeepLabv3+, with its multi-scale and spatial awareness capabilities, is considered a promising alternative to address these issues.

### B. Type of Research and Data Classification

This study adopts an implementation-based research design using a quantitative analysis approach, aimed at both developing and evaluating the performance of a DeepLabv3+-based lane detection and tracking system. Implementation research involves directly testing the system in practice, whereas the quantitative approach focuses on measurable performance metrics to ensure objective conclusions. The study is experimental in nature, as it involves training and evaluating a deep learning model using real-world data. The primary data used consist of road images captured via dashcam, providing up-to-date and context specific information about actual driving environments. Additionally, secondary data such as scientific literature, technical documentation, and reference models were consulted to support theoretical foundations and benchmark comparisons. The combination of both primary and secondary data ensures comprehensive model development and enhances the validity and reliability of the research outcomes.

### C. Data Sources

The dataset comprises images captured by a dashcam mounted on a low MPV vehicle. Image collection was conducted across five different road locations in Makassar: Jl. Veteran Utara, Jl. Gunung Bawakaraeng, Jl. Tol Reformasi, Jl. Tol Layang A.P Pettarani, and Jl. Bonto Daeng Ngrate. A total of 374 images were collected under bright daylight conditions to ensure optimal lighting and minimal visual noise. These images were manually annotated using Labelme, an open-source labeling tool, into four categories: road, solid lane markings, dashed lane markings, and background. Of the total dataset, 336 images were used for model training, whereas the remaining 38 were reserved for testing. Each image was carefully annotated to allow the segmentation model to learn class-specific patterns accurately. The dataset was then structured using the Pascal VOC format, ensuring compatibility with DeepLabv3+'s training pipeline. This structured dataset forms the basis for all training, evaluation, and real-time testing processes carried out in the study. All data were collected ethically in public areas without recording any personally identifiable information such as faces or license plates. The dataset is currently private as it was self-collected and is not a public asset, but it can be shared upon formal request for non-commercial academic research to ensure transparency and responsible data use. Although the dataset is relatively small, its diversity across multiple urban road types in Makassar provides strong representativeness. In addition, 5-fold cross-validation was applied to ensure robustness, and future work will extend evaluation using larger public datasets.

### D. Research Stages and Data Collection Techniques

The research was conducted in several sequential stages, starting from equipment setup to real-time evaluation. A dashcam with an  $89.27^\circ$  field of view was installed on a low MPV vehicle to record urban roads during daytime under clear weather conditions. The driving speed was maintained at approximately 10 km/h to minimize motion blur and preserve image clarity. The recorded footage was converted into images, which were manually annotated using Labelme into four semantic classes: road, solid lane markings, dashed lane markings, and background. The dataset, consisting of 374 labeled images, was then formatted using the Pascal VOC standard for compatibility with the DeepLabv3+ training framework.

The DeepLabv3+ architecture served as the baseline, enhanced with three attention mechanisms, CBAM, ECA, and SE, to improve multi-scale feature extraction and spatial awareness. These modules were embedded within the encoder to refine feature representations at both channel and spatial dimensions.

Mathematically, the CBAM applies sequential channel and spatial attention operations to the feature map  $F \in \mathbb{R}^{H \times W \times C}$ :

$$M_c = \sigma \left( MLP(AvgPool(F)) + MLP(MaxPool(F)) \right) \quad (1)$$

$$= \sigma \left( W_1 \left( W_0(F_{avg}^c) \right) + W_1 \left( W_0(F_{max}^c) \right) \right) \quad (2)$$

where  $\sigma$  denotes the sigmoid activation and  $f^{7 \times 7}$  represents a convolutional layer with kernel size  $7 \times 7$ .

The ECA block enhances local cross-channel interaction without dimensionality reduction using an adaptive 1D convolution:

$$M_{ECA}(f) = \sigma(Conv1D(GAP(F))) \quad (3)$$

Meanwhile, the SE block performs feature recalibration by "squeezing" spatial information and "exciting" channel dependencies:

$$M_{SE}(f) = \sigma(W_2 \delta(W_1 GAP(F))) \quad (4)$$

where  $\delta$  is the ReLU activation, and  $W_1, W_2$  denote the learnable weight matrices of two fully connected layers.

These attention blocks were cascaded in the encoder to produce an adaptive weighting mechanism, enhancing the network's sensitivity to fine lane edges and visually similar boundaries. The training configuration included a learning rate of 0.001, a batch size of 4, and 60 epochs using the Adam optimizer with weight decay  $1 \times 10^{-6}$ . Progressive multi-resolution learning and dropout (0.5) were applied to prevent overfitting and improve generalization.

To further enhance model robustness, K-fold cross-validation ( $k=5$ ) and bootstrap learning were used, ensuring the model trained on diverse subsets and difficult samples. Although the hold out test set contained 38 images, 5-fold cross-validation was employed to ensure statistical robustness, allowing the model to be evaluated across multiple train-test partitions and minimizing sampling bias. The model's backbone was ResNet-18, which balances computational

efficiency and accuracy. During training, a combined loss consisting of focal loss, dice loss, and cross entropy loss was employed to address class imbalance and segmentation precision:

$$\mathcal{L}_{total} = \mathcal{L}_{focal} + \mathcal{L}_{dice} + \mathcal{L}_{ce} \quad (5)$$

Testing was performed using 38 reserved images for both static and real-time evaluation, validating segmentation and distance estimation under practical driving scenarios.

This design introduces a novel cascading integration of multiple attention modules (CBAM, ECA, and SE) within the DeepLabv3+ encoder, combined with an RMSE-based distance estimation component, which differentiates the proposed model from standard DeepLabv3+ and U-Net implementations in both structure and functional capability.

The complete training procedure of the proposed DeepLabv3+ with cascading attention (CBAM–ECA–SE) can be summarized as follows:

Algorithm 1. Training Procedure of the Proposed DeepLabv3+ (CBAM–ECA–SE)

1. Input RGB road images  $I \in \mathbb{R}^{3 \times H \times W}$  and corresponding ground truth mask  $M$ .
2. Apply preprocessing and data augmentation, including random resized crop, horizontal flip, and normalization.
3. Extract spatial features using a ResNet-18 encoder.
4. Sequentially refine extracted features through CBAM, ECA, and SE attention modules to enhance multi-scale channel and spatial representation.
5. Decode the refined features through ASPP and upsampling layers to generate the segmentation prediction  $\hat{M}$ .
6. Compute the total loss  $\mathcal{L}_{total} = \mathcal{L}_{focal} + \mathcal{L}_{dice} + \mathcal{L}_{ce}$  as defined in (5).
7. Update model weights using the Adam optimizer (lr=0.001, batch size = 4, weight decay =  $1 \times 10^{-6}$ ).
8. Evaluate the model using IoU, RMSE, precision, recall, F1-score, and LaneIoU metrics after each epoch.
9. Repeat for 60 epochs or until validation convergence.

The model was implemented in Python and trained on an NVIDIA GeForce GTX 1650 Ti GPU (4 GB VRAM) under Windows 11. Input resolution was fixed at  $128 \times 128$  pixels to optimize computation speed and memory usage. The model achieved a stable convergence with an average training time of 202 ms per image and an inference rate of 10 FPS. These additions ensure mathematical completeness, algorithmic transparency, and reproducibility of the proposed method under real-world conditions.

### E. Data Analysis Techniques

The evaluation employed quantitative and computational performance metrics to assess segmentation quality, spatial accuracy, and efficiency.

#### 1) Segmentation Accuracy

The IoU metric was used to evaluate pixel-level segmentation performance:

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (6)$$

where  $TP$ ,  $FP$ , and  $FN$  represent the true positive, false positive, and false negative pixel counts, respectively. To complement IoU, additional performance indicators were integrated: precision, recall, F1-score, and LaneIoU, to capture structural accuracy and per class reliability, particularly for curved and dashed lane segments:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (8)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

LaneIoU quantifies the overlap of predicted and ground truth masks for all lane classes (solid + dashed):

$$\text{LaneIoU} = \frac{|P_{lane} \cap G_{lane}|}{|P_{lane} \cup G_{lane}|} \quad (10)$$

where  $P_{lane}$  and  $G_{lane}$  denote the predicted and ground truth lane pixels. These metrics were implemented using the scikit learn library and complement IoU by capturing class-level balance and structural fidelity, particularly on curved lane segments.

#### 2) Distance Estimation

The RMSE was used to quantify the deviation between actual and predicted lane-to-tire distances:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (11)$$

where  $y_i$  and  $\hat{y}_i$  denote the actual and predicted distances, respectively. A low RMSE indicates precise spatial estimation.

#### 3) Model Complexity and Efficiency

Computational efficiency was analyzed through parameter count, FLOPs, and inference time, computed as follows:

$$\text{FPS} = \frac{N_{frames}}{T_{inference}} \quad (12)$$

The trained model achieved a computational complexity of approximately 1.96 GMac, with 5.02 million parameters, and an average inference time of 201.99 ms per image, corresponding to a real-time rate of 10 FPS. This balance of accuracy and efficiency demonstrates that the proposed attention augmented DeepLabv3+ model achieves high segmentation quality while maintaining feasible inference performance for embedded and autonomous vehicle systems. The implementation source code can be shared upon formal

request for academic research purposes to ensure transparency and reproducibility of the reported results.

In addition to quantitative evaluations, visual inspection was performed across multiple driving scenarios to assess the model's stability, temporal consistency, and robustness against challenging lane conditions. The integration of mathematical formulations for attention mechanisms, explicit FPS computation, and analysis of model complexity ensures that the revised methodology section provides both algorithmic transparency and empirical validation of real-time capability.

#### IV. RESULTS

This research emphasizes the development of a visual navigation system for autonomous vehicles, particularly the system's capability to detect and track road lanes. The autonomous car featured in this study is equipped with a visual perception system based on a camera, capable of capturing real-time environmental data. A total of 374 images were collected using a dashcam mounted on a vehicle, recorded at five urban locations across Makassar during the daytime and under clear weather conditions, ensuring optimal lighting and minimal visual distortion. The images were manually annotated into four classes: road, solid lane markings, dashed lane markings, and background. These visual data enable the vehicle to interpret lane structures and road boundaries without human assistance, thereby enhancing road safety. Accurate lane detection becomes a key function in guiding the autonomous vehicle to maintain safe lane positioning, especially in urban environments where markings are often inconsistent or visually degraded.

The data collection and processing results demonstrate that a vision-based system using a dashcam can provide detailed environmental perception for autonomous vehicles. The captured images effectively visualize lane patterns with sufficient clarity to enable precise annotation using Labelme. With 374 annotated images, the training dataset is representative of real urban road conditions in Makassar. The four segmentation classes, road, solid line, dashed line, and background, allow the system to learn meaningful spatial patterns for lane detection. The wide-angle lens with an  $89.27^\circ$  field of view ensures adequate coverage of the vehicle's forward perspective. The system's output was tested in both static and real-time scenarios to observe its responsiveness to lane contours. Images successfully segmented by the system serve as input for decision-making processes such as steering and lane centering. The data confirm the feasibility of using real-world visual inputs to train and evaluate perception systems in autonomous vehicles.

The correlation between these vision-based data and the real-world problem clearly reinforces the importance of high-accuracy visual perception systems. Given the high rate of traffic accidents attributed to human error, a system that can independently recognize and respond to road conditions offers a strategic solution for improving safety. The real-world images collected in this study reflect the complexity of urban traffic environments, including faded or inconsistent lane markings and potential visual distractions. The camera-based system effectively detects lane lines and constructs a visual

context for autonomous maneuvering. By reducing reliance on human judgment, the system addresses human negligence, the primary factor underlying most traffic accidents. Thus, the results directly support the premise that integrating robust visual perception into autonomous vehicles is essential for building safer, more responsive transportation systems.

In this study, semantic segmentation serves as the primary technique for identifying essential road elements in images. The model's performance was evaluated using the IoU metric, achieving an overall IoU of 97.34%. This high score confirms the model's capability to accurately distinguish between road surfaces, solid markings, and dashed markings. In addition, the segmentation outputs were integrated into the lane tracking and distance estimation subsystems, making semantic segmentation a central component of the proposed visual navigation framework.

As shown in Figure 2, IoU performance for each class in DeepLabv3+ highlights the effectiveness of the segmentation model across various road features. It shows that the system can effectively detect lane features, including road surface, solid lane markings, dashed lane markings, and background.

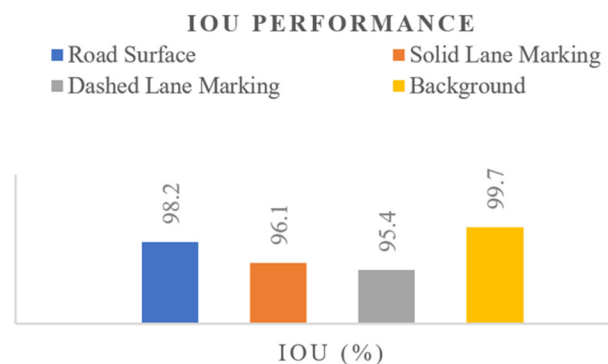


Fig. 2. IoU performance for each class in DeepLabv3+.

The success of semantic segmentation is further validated by detailed and accurate performance evaluations. Each class in the segmentation task contributes significantly to constructing the semantic map of the road, with clear distinctions between visually similar features like solid and dashed lines. Manual annotations ensured high quality ground truth labels, supporting objective evaluation of model accuracy. The IoU metric provides a rigorous basis for performance measurement, and the achieved score of 97.34% reflects the model's reliability.

In addition to IoU, further quantitative metrics were computed to evaluate structural and class level performance. Table II summarizes precision, recall, F1-score, and LaneIoU obtained from the test set. These complementary indicators demonstrate that the model not only maintains high pixel-wise accuracy but also achieves strong balance between precision and recall across lane categories. The relatively lower LaneIoU value indicates that curved or visually degraded lane markings remain the most challenging cases, suggesting potential improvement through temporal modeling or data augmentation in future work.

TABLE II. ADDITIONAL EVALUATION METRICS FOR LANE DETECTION

Metric	Value (macro)
Precision	0.9767
Recall	0.9891
F1-score	0.9828
LaneIoU	0.4868

Mean IoU is reported separately in Figure 2 and discussed in Section V.

These results verify that the inclusion of the proposed attention modules preserves high segmentation precision while maintaining structural lane integrity, confirming the model's balanced performance across multiple quantitative indicators.

Beyond numerical indicators, the model was tested under real-time visual scenarios, demonstrating consistency in frame-by-frame segmentation results. This temporal stability is particularly important for lane tracking, ensuring that the segmented lanes remain visually coherent across consecutive video frames. As such, semantic segmentation using DeepLabv3+ not only performs well in static testing but also proves viable in dynamic driving contexts.

The successful implementation of semantic segmentation directly addresses the limitations found in previous studies using architectures like U-Net. Earlier systems struggled with accurately detecting lane lines, particularly solid markings, which resulted in poor tracking performance. In contrast, the segmentation model in this study achieved a significantly higher IoU score, improving the system's reliability in identifying lane boundaries. As demonstrated in Figure 3, the model's output clearly differentiates solid and dashed lanes from the road surface. Challenges such as narrow lane lines and visually similar textures were effectively managed by the multi-scale processing capabilities of DeepLabv3+. These improvements are crucial in real-world applications, where consistent and precise lane recognition, visually confirmed in Figure 3, can prevent navigation errors and ensure vehicle stability. Thus, semantic segmentation proves to be more than just a technical enhancement; it becomes a strategic necessity in advancing the safety and intelligence of autonomous vehicle systems.

DeepLabv3+ was employed as the main architecture for the semantic segmentation model in this research. It combines atrous convolution and ASPP to capture spatial features across multiple scales. The model was trained using the annotated dataset, with the Xception backbone employed for feature extraction. The system received RGB input images that underwent preprocessing before being fed into the segmentation pipeline. Once trained, the model produced detailed segmentation maps for each image class. The segmentation results were then used in the lane tracking and distance estimation components, utilizing IPM for spatial transformation. The model's performance was evaluated under both static and real-time testing environments, demonstrating stable and accurate outputs even when applied to real-world road scenes.

The advantages of DeepLabv3+ are clearly evident in the segmentation results, especially its ability to detect thin and ambiguous lane features. The combination of ASPP and atrous

convolution enabled the model to extract both local and global contextual information, enhancing its sensitivity to complex lane geometries. Unlike U-Net, which operates with lower spatial precision, DeepLabv3+ preserved boundary details and produced sharper, more accurate segmentation maps. These outputs were transformed using IPM and processed to estimate the vehicle's distance from lane markings using the RMSE metric. The resulting RMSE value of 0.0377, shown in Table III, indicates an extremely low deviation, demonstrating the system's precision. This level of accuracy makes DeepLabv3+ not only effective in static image segmentation but also suitable for integration into dynamic control systems in autonomous driving.

TABLE III. RMSE EVALUATION OF TIRE-TO-LANE DISTANCE

Test scenario	RMSE value
Scenario 1 – straight road	0.0372
Scenario 2 – dashed marking	0.0385
Scenario 3 – solid marking	0.0368
Scenario 4 – mixed lane types	0.0381
Scenario 5 – curved lane	0.0374
Average RMSE	0.0377

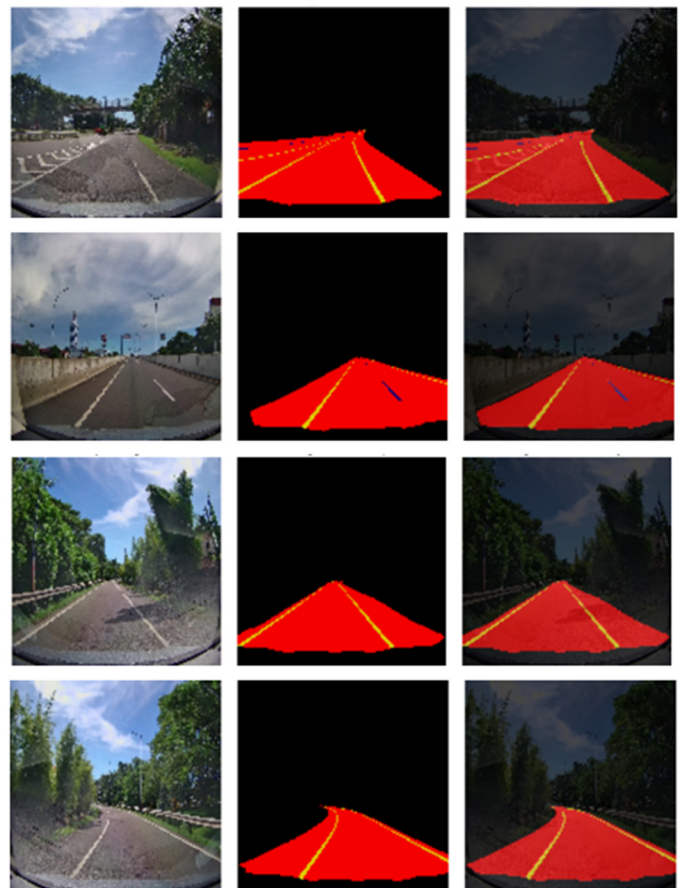


Fig. 3. Lane segmentation results using DeepLabv3+.

The application of DeepLabv3+ in this study directly overcomes the shortcomings of prior segmentation architectures like U-Net. With its high segmentation accuracy and low RMSE in distance estimation, the model proves

capable of operating reliably under real-world conditions. In urban settings characterized by visual noise and unpredictable road conditions, DeepLabv3+ delivers consistent and accurate recognition of lane structures. The system's ability to track vehicle positioning relative to road markings ensures safe lane keeping behavior, an essential requirement for autonomous driving. These findings affirm that enhancing visual perception through DeepLabv3+ can significantly improve the intelligence and safety of autonomous vehicle systems. As a result, the study not only yields strong experimental outcomes but also offers practical solutions to real-world navigation challenges.

Although separate visual results for each scenario are not displayed, the examples in Figure 3 are representative of all test conditions. The RMSE values in Table III quantify distance accuracy differences rather than visual variations, confirming consistent segmentation and lane tracking performance across all scenarios.

## V. DISCUSSION

Overall, the results demonstrate that integrating the DeepLabv3+ architecture into a lane detection and tracking system for autonomous vehicles significantly improves segmentation accuracy and precision in distance estimation. Evaluated through IoU, the model showed remarkable effectiveness in identifying road structures, whereas RMSE revealed minimal deviation in calculating the distance between the tire and lane markings. These outcomes confirm that DeepLabv3+ effectively interprets road contours, even in visually complex scenarios, outperforming earlier models in both accuracy and real-time consistency. This achievement reflects not only technical success in system development but also comprehensive model validation demonstrating enhanced detection performance and reliable tracking under realistic driving conditions. As such, the results affirm that the system's design successfully addresses the core challenges of visual perception in autonomous navigation.

Compared to prior studies using U-Net for lane segmentation, the superiority of DeepLabv3+ becomes distinctly evident. Earlier research indicated that U-Net achieved only 79.8% segmentation accuracy overall, with particularly poor performance on solid lane markings (54.8%). In contrast, this study recorded a substantial improvement, with an overall IoU of 97.34%. This improvement is not merely quantitative but also methodological. DeepLabv3+ introduces multi-scale feature extraction and spatial-awareness mechanisms, which U-Net lacks. Moreover, while most previous works focus solely on pixel-wise segmentation, this research also incorporates real-world integration through distance estimation using IPM and RMSE, an aspect often overlooked in other studies. In addition, the proposed framework uniquely combines CBAM, ECA, and SE attention mechanisms in a cascading configuration, enabling superior multi-scale feature interaction and precise lane-boundary recognition, an architectural innovation not previously applied in DeepLabv3+ lane-detection studies. Thus, the findings not only refine the shortcomings of previous approaches but also contribute methodologically by offering a more complete and applicable system for real-world autonomous-navigation challenges.

The comparison in Figure 4 was conducted using identical dataset and evaluation settings to ensure fairness among segmentation models. U-Net, FCN, and the proposed DeepLabv3+ perform the same lane-segmentation task, and the improvement in IoU clearly reflects the architectural advantages of the proposed model rather than dataset differences. The results confirm that the research objective of designing and optimizing a lane detection and tracking system using DeepLabv3+ have been effectively achieved. They also highlight the benefits of combining accurate semantic segmentation with intelligent distance estimation for real-time vehicle navigation. The system's reliability confirms that visual-perception technologies are maturing to a point where they can independently support lane adherence, a task critical for autonomous safety. The ability to consistently recognize lane boundaries and maintain lateral stability reflects how deep learning can help reduce errors associated with human judgment. This finding suggests that the broader implementation of such systems may contribute not only to the academic understanding of deep learning in mobility but also to practical reductions in accident risks caused by human oversight.

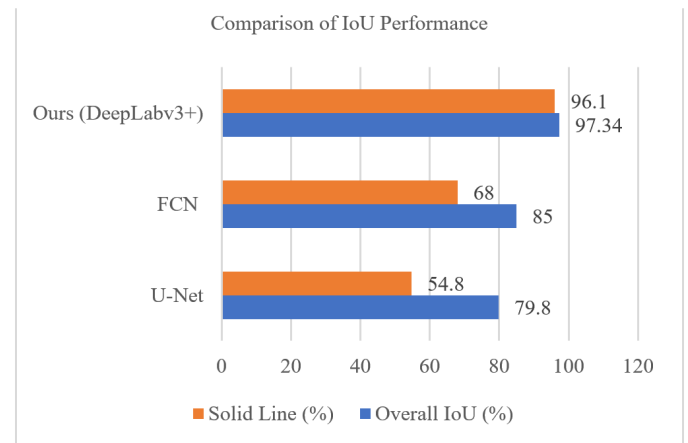


Fig. 4. Comparison of IoU performance for U-Net, FCN, and DeepLabv3+.

To further validate the effect of each attention component, an ablation study was conducted by comparing the Feature Cross Attention (FCA) only configuration and the full model integrating CBAM, ECA, and SE blocks. As presented in Table IV, the FCA-only setup achieved a mean IoU (mIoU) of 90.33%, whereas the complete cascading-attention configuration reached 97.34%. This clear improvement confirms that combining multiple attention modules enhances feature interaction and spatial-awareness, resulting in more stable and precise lane-segmentation performance.

TABLE IV. ABLATION STUDY ON THE PROPOSED DEEPLABV3+ ARCHITECTURE

Model configuration	Attention modules used	mIoU (%)
Baseline DeepLabv3+	-	61.78
DeepLabv3 + FCA	FCA	90.33
Proposed (CBAM + ECA + SE + FCA)	All combined	97.34

Despite the high overall accuracy, several failure cases were observed under challenging environmental conditions, as illustrated in Figure 5. The model occasionally experienced partial segmentation loss during nighttime or heavy rain, primarily due to low illumination, glare, and road-surface reflections that reduce lane contrast. These cases, however, are



Fig. 5. Example of failure cases under rainy and nighttime conditions showing partial segmentation loss caused by low illumination, glare, and reflective road surfaces.

The implications of this study extend well beyond its immediate experimental outcomes. The demonstrated model offers a viable framework for integration into Level 3 and higher autonomous driving systems, particularly in the development of lane keeping and lane centering features. Furthermore, the segmentation and tracking modules developed in this study could support ITSs where vehicles interact with digital infrastructure in real-time. From a practical standpoint, the proposed model can serve as a decision support mechanism for drivers, especially in conditions of fatigue or reduced visibility. It also offers educational value in the development of simulation tools for training autonomous vehicle systems. Overall, these results not only validate the research objectives but position the model as a foundational component for future development in smart mobility and real-world vehicle autonomy.

The outstanding performance of the system can be attributed to the technical strengths of the DeepLabv3+ architecture itself. Its ability to capture both local and global features through atrous convolution and ASPP enables it to accurately segment narrow, ambiguous lane patterns, an area where prior models consistently failed. Furthermore, the decoder structure enhances resolution in the final output, preserving fine edge details that are critical for lane tracking. Training the model on a high quality, manually annotated dataset representative of actual road conditions ensured strong model generalization. The use of objective performance metrics such as IoU and RMSE provided clear indicators of both segmentation accuracy and spatial consistency. In combination, these factors created a model architecture that not only excels in theory but also performs reliably in real-world scenarios, bridging the gap between research models and deployable autonomous systems.

Based on these findings, several strategic actions are recommended to further develop the system. First, future research should extend testing into varied environmental conditions, such as nighttime, rain, or glare, to evaluate the

minor and occurred infrequently compared to normal daylight conditions. The issue can be mitigated in future work by augmenting the dataset with adverse weather images, applying illumination normalization, or integrating temporal consistency models to improve robustness under dynamic lighting.

system's robustness under low visibility scenarios. Incorporating temporal models, such as Recurrent Neural Networks (RNNs) or Temporal Convolutional Networks (TCNs), could enhance frame-to-frame consistency and strengthen real-time performance. From an engineering perspective, embedding the model into edge devices for real-time inference should be explored to ensure computational efficiency and low latency performance. Data augmentation using diverse environmental inputs could also strengthen model adaptability. Overall, this research establishes a solid base from which further innovation can be pursued, and its findings should guide next stage implementation of vision-based systems in autonomous vehicle platforms.

## VI. CONCLUSION

Notably, optimizing the DeepLabv3+ architecture within a visual perception system for autonomous vehicles significantly improves segmentation accuracy, reaching an Intersection over Union (IoU) of 97.34%. This marks a significant leap compared to previous models such as U-Net, which only achieved 54.8% accuracy in detecting solid lane markings. The improvement is not limited to segmentation alone; the system also demonstrated exceptional precision in distance estimation, achieving an impressively low Root Mean Square Error (RMSE) of 0.0377. These results highlight not only DeepLabv3+'s capability in static image processing but also its practical applicability in real-time lane tracking tasks. This confirms the model's efficacy not only in laboratory settings but also its potential for deployment in real-world autonomous vehicle systems. The findings validate the idea that cutting-edge deep learning architectures can translate into tangible improvements in navigational safety and system intelligence.

This research contributes meaningfully to both theoretical advancement and practical implementation in the field of computer vision for autonomous driving. Theoretically, it demonstrates that deep semantic segmentation using DeepLabv3+, enhanced by multi-scale contextual analysis, can significantly outperform traditional models in understanding

complex road environments. This strengthens the argument for DeepLabv3+ as a robust architecture applicable beyond medical imaging and urban mapping to autonomous navigation. Practically, the system developed offers a scalable and accurate solution for lane detection and tracking, addressing the growing need for intelligent navigation modules in Level 3+ autonomous vehicles. Additionally, its integration of RMSE-based distance estimation into a visual segmentation pipeline broadens its relevance to real-time vehicle control systems. As such, this study bridges the gap between conceptual artificial intelligence models and real-world, safety-oriented vehicle applications.

While the research has produced highly promising results, its scope is limited to daytime conditions with clear weather, which was a necessary control to ensure data consistency. This limitation, however, should not be seen as a weakness, but rather as a foundational step for future exploration. It presents a clear opportunity for subsequent research to test the system under more challenging environmental conditions such as nighttime driving, adverse weather, or low contrast lane markings. Future developments might also consider integrating temporal models or fusing vision-based data with LiDAR inputs to enhance robustness. In this way, the current study lays the groundwork not only for methodological improvements but also for broader implementation in real-world autonomous vehicle ecosystems, offering a strong platform for future innovation in intelligent mobility systems.

#### ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Department of Informatics Engineering and the Department of Electrical Engineering, Hasanuddin University, for their valuable support and facilities during this research. Special thanks are also extended to the Ministry of Higher Education, Science, and Technology (Kemdiktisaintek) through the BIMA Research funding.

#### REFERENCES

- [1] "Rata-rata Tiga Orang Meninggal Setiap Jam Akibat Kecelakaan Jalan." Kementerian Komunikasi dan Informatika Republik Indonesia, 2017. [https://kominfo.go.id/index.php/content/detail/10368/rata-rata-tiga-orang-meninggal-setiap-jam-akibat-kecelakaan-jalan/0/artikel\\_gpr](https://kominfo.go.id/index.php/content/detail/10368/rata-rata-tiga-orang-meninggal-setiap-jam-akibat-kecelakaan-jalan/0/artikel_gpr).
- [2] I. Indrabayu, A. A. Prayogi, I. S. Areni, A. Bustamin, and N. Azqalani, "Real-time lane departure warning with cascade lane segmentation," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 32, no. 2, pp. 994–1003, Nov. 2023, <https://doi.org/10.11591/ijeecs.v32.i2.pp994-1003>.
- [3] S. Grigorescu, B. Trasnea, T. Cocias, and G. Macesanu, "A survey of deep learning techniques for autonomous driving," *Journal of Field Robotics*, vol. 37, no. 3, pp. 362–386, Apr. 2020, <https://doi.org/10.1002/rob.21918>.
- [4] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," in *15th European Conference on Computer Vision – ECCV 2018*, Munich, Germany, 2018, pp. 833–851, [https://doi.org/10.1007/978-3-030-01234-2\\_49](https://doi.org/10.1007/978-3-030-01234-2_49).
- [5] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 2015, pp. 3431–3440, <https://doi.org/10.1109/CVPR.2015.7298965>.
- [6] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation." arXiv, Dec. 05, 2017, <https://doi.org/10.48550/arXiv.1706.05587>.
- [7] S. S. Yoo and H.-S. Lee, "An Inverse Perspective Mapping-Based Approach for Generating Panoramic Images of Pipe Inner Surfaces," *Sensors*, vol. 23, no. 12, June 2023, Art. no. 5363, <https://doi.org/10.3390/s23125363>.
- [8] A. Bosaghzadeh and M. Nasiri Manjili, "Inverse perspective mapping for real-time Lane Detection in City Streets," *Automotive Science and Engineering*, vol. 10, no. 3, pp. 3311–3323, Sept. 2020, <https://doi.org/10.22068/ase.2020.540>.
- [9] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 2019, pp. 658–666, <https://doi.org/10.1109/CVPR.2019.00075>.
- [10] M. Kuhn and K. Johnson, *Applied Predictive Modeling*. New York, NY, USA: Springer, 2013, <https://doi.org/10.1007/978-1-4614-6849-3>.
- [11] B. Zhang, C. Wang, and J. Wang, "A Lightweight and Efficient Detection Transformer for Highway Abandoned Objects," *Algorithms*, vol. 18, no. 3, Mar. 2025, Art. no. 133, <https://doi.org/10.3390/a18030133>.
- [12] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *15th European Conference on Computer Vision – ECCV 2018*, Munich, Germany, 2018, pp. 3–19, [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1).
- [13] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 7132–7141, <https://doi.org/10.1109/CVPR.2018.00745>.
- [14] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2020, pp. 11531–11539, <https://doi.org/10.1109/CVPR42600.2020.01155>.
- [15] D. Garikapati and S. S. Shetiya, "Autonomous Vehicles: Evolution of Artificial Intelligence and the Current Industry Landscape," *Big Data and Cognitive Computing*, vol. 8, no. 4, Apr. 2024, Art. no. 42, <https://doi.org/10.3390/bdcc8040042>.
- [16] H. Gajjar, S. Sanyal, and M. Shah, "A comprehensive study on lane detecting autonomous car using computer vision," *Expert Systems with Applications*, vol. 233, Dec. 2023, Art. no. 120929, <https://doi.org/10.1016/j.eswa.2023.120929>.
- [17] N. J. Zakaria, M. I. Shapiari, R. A. Ghani, M. N. M. Yassin, M. Z. Ibrahim, and N. Wahid, "Lane Detection in Autonomous Vehicles: A Systematic Review," *IEEE Access*, vol. 11, pp. 3729–3765, 2023, <https://doi.org/10.1109/ACCESS.2023.3234442>.
- [18] L. Jiang, G. Ma, W. Guo, and Y. Sun, "YOLO-DH: Robust Object Detection for Autonomous Vehicles in Adverse Weather," *Electronics*, vol. 14, no. 22, Nov. 2025, Art. no. 4476, <https://doi.org/10.3390/electronics14224476>.
- [19] D. Kumar and N. Muhammad, "Object Detection in Adverse Weather for Autonomous Driving through Data Merging and YOLOv8," *Sensors*, vol. 23, no. 20, Oct. 2023, Art. no. 8471, <https://doi.org/10.3390/s23208471>.
- [20] W. Zhou, J. S. Berrio, S. Worrall, and E. Nebot, "Automated Evaluation of Semantic Segmentation Robustness for Autonomous Driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 5, pp. 1951–1963, May 2020, <https://doi.org/10.1109/TITS.2019.2909066>.
- [21] M. A. M. Elhassan et al., "Real-time semantic segmentation for autonomous driving: A review of CNNs, Transformers, and Beyond," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 10, Dec. 2024, Art. no. 102226, <https://doi.org/10.1016/j.jksuci.2024.102226>.
- [22] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, Aug. 2018, <https://doi.org/10.1007/s13244-018-0639-9>.

- [23] G. Doğan and B. Ergen, "A new CNN-based semantic object segmentation for autonomous vehicles in urban traffic scenes," *International Journal of Multimedia Information Retrieval*, vol. 13, no. 1, Feb. 2024, Art. no. 11, <https://doi.org/10.1007/s13735-023-00313-5>.
- [24] T. Pariwat and P. Seresangtakul, "Multi-Stroke Thai Finger-Spelling Sign Language Recognition System with Deep Learning," *Symmetry*, vol. 13, no. 2, Feb. 2021, Art. no. 262, <https://doi.org/10.3390/sym13020262>.
- [25] D. Patil and S. Jadhav, "Road Segmentation in High-Resolution Images Using Deep Residual Networks," *Engineering, Technology & Applied Science Research*, vol. 12, no. 6, pp. 9654–9660, Dec. 2022, <https://doi.org/10.48084/etasr.5247>.
- [26] D. D. Van, "Application of Advanced Deep Convolutional Neural Networks for the Recognition of Road Surface Anomalies," *Engineering, Technology & Applied Science Research*, vol. 13, no. 3, pp. 10765–10768, June 2023, <https://doi.org/10.48084/etasr.5890>.
- [27] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images," *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp. 1127–1141, Dec. 2018, <https://doi.org/10.1111/mice.12387>.
- [28] Y. Xu, Z. Xie, Y. Feng, and Z. Chen, "Road Extraction from High-Resolution Remote Sensing Imagery Using Deep Learning," *Remote Sensing*, vol. 10, no. 9, Sept. 2018, Art. no. 1461, <https://doi.org/10.3390/rs10091461>.