JINAV: Journal of Information and Visualization *ISSN: 2746-1440 (online)*
Vol. 7 No. 2 (2026)
https://doi.org/10.35877/454RI.jinav4919

Traffic Sign Image Segmentation using U-Net Architecture Based on Convolutional Neural Network

Mutaqin Akbar[a,*], Budi Sulistiyo Jati[a], Indah Susilawati[a], Moh Ahsan[b]

[a]*Universitas Mercu Buana Yogyakarta, Yogyakarta, Indonesia*
[b]*Universitas PGRI Kanjuruhan Malang, Malang, Indonesia*

Abstract

Accurate pixel-level segmentation of traffic signs in natural scenes is a critical precursor to reliable sign recognition in intelligent transportation systems. This study investigates the effectiveness of the U-Net architecture for semantic segmentation of traffic signs using a dataset comprising 1,750 training subset and 300 testing subset. The model was trained for 20 epochs with the Adam optimizer (learning rate 0.0002, batch size 2), exhibiting stable convergence: training loss decreased from 0.137729 to 0.05989, while the dice coefficient improved from 0.899644 to 0.956839, and binary IoU rose from 0.894176 to 0.94945. Evaluation on a held-out test set of 300 images demonstrated strong generalization, yielding a dice coefficient of 0.930935, binary IoU of 0.907808, precision of 0.90499, recall of 0.981679, and accuracy of 0.953572. The recall–precision asymmetry indicates a mild tendency toward over-segmentation. The consistent covariation between the dice coefficient and binary IoU across both training and testing phases affirms the internal reliability of the evaluation metrics employed in this study. Qualitative analysis revealed high-fidelity segmentation of polygonal signs, with the model faithfully reproducing the sharp vertices and overall silhouette of both the octagonal and the rhomboid sign. In contrast, the circular sign exhibited a localized boundary failure, characterized by conspicuous concave flattening along its upper-left contour that caused the predicted rim to appear clipped relative to the smooth ground-truth circumference. These findings establish U-Net as a robust segmentation foundation for downstream traffic sign recognition, while motivating future work on architectural refinements to improve boundary-level precision for curvilinear signs and the adoption of recall loss functions to further regulate the recall–precision trade-off and mitigate over-segmentation tendencies.

Keywords: convolutional neural network; deep learning; segmentation; traffic sign; U-Net.

Received: 17 June 2026 Revised: 30 July 2026 Published: 31 August 2026

1. Introduction

Traffic signs constitute a fundamental component of road transportation infrastructure, functioning as a visual communication channel between traffic authorities and road users. Their presence is indispensable for regulating traffic flow, alerting drivers to potential hazards, and conveying essential information that collectively supports safety, order, and efficiency on public roads. Traffic signs in Indonesia are regulated by the Ministry of Transportation under Regulation Number 13 of 2014. However, transportation technology keeps evolving, especially with the rapid development of Intelligent Transportation Systems (ITS) and autonomous vehicle systems (Günthner & Proff, 2021; Narkhede & Chopade, 2022; Suriya Prakash et al., 2021). The ability of a system to automatically detect and interpret traffic signs has become an increasingly critical requirement rather than a mere convenience (Brar et al., 2025; Chen et al., 2024). Automated recognition of traffic signs from digital imagery, however, is far from trivial (Lim et al., 2023). Systems must contend with substantial variation in sign shape and color, fluctuating illumination conditions, differing camera viewpoints, and cluttered or visually complex backgrounds (Seraj et al., 2021). Among the various stages involved in building a robust recognition pipeline, image segmentation stands out as a foundational step (S. Li et al., 2023; Sun et al., 2021). Its purpose is to isolate the sign object from the surrounding background, thereby narrowing the region of interest before subsequent analysis or classification takes place (He et al., 2021).

* Corresponding author.
E-mail address: mutaqin@mercubuana-yogya.ac.id

Approaches to image segmentation can broadly be categorized into two major paradigms: conventional techniques and deep learning-based methods. Conventional approaches, including color-based thresholding, edge detection, and clustering algorithms, operate on the basis of manually engineered features and rely heavily on assumptions regarding pixel similarity or discontinuity within an image (Akbar et al., 2019; X. Li et al., 2026; Yao et al., 2025). Color-based thresholding, for instance, typically segments objects by defining fixed boundaries within a particular color space (such as RGB or HSV), while edge detection techniques identify object boundaries through abrupt changes in pixel intensity (Giuliani, 2022; H. Wang et al., 2025; T. Wang & Wong, 2023; Yang & Li, 2026). These methods share a common characteristic: they depend on handcrafted rules and heuristics designed by human experts, rather than on features learned directly from data. While computationally efficient and straightforward to implement, these methods often struggle under dynamic, real-world conditions, including shadows, adverse weather, or backgrounds that closely resemble the color palette of traffic signs themselves (G. Huang et al., 2025).

Advances in deep learning, most notably Convolutional Neural Network (CNN), have considerably altered the trajectory of computer vision, fundamentally transforming how image segmentation tasks are approached and solved. In contrast to conventional approaches that rely on manually designed rules, CNN can automatically extract multi-level feature representations straight from raw images (Akbar et al., 2025). This hierarchical learning process enables far greater adaptability to complex visual variation, including changes in illumination, occlusion, and background clutter (Benfaress et al., 2025; Fredj et al., 2023). Within the domain of semantic segmentation, CNN-based architectures such as Fully Convolutional Networks (FCN) (Hu et al., 2021; S.-Y. Huang et al., 2022; Ren et al., 2023), U-Net (Ben-Abbou et al., 2026; Boly & Akbar, 2024; Ramyashree et al., 2026), SegNet (Priya et al., 2022; Zhang et al., 2024), and DeepLab (Aulia & Akbar, 2026; Memon et al., 2022; J. Wang et al., 2021), demonstrate strong performance in pixel-wise classification tasks, leveraging strategies like encoder-decoder designs, skip connections, and multi-scale feature extraction to retain fine-grained spatial detail alongside deep semantic understanding. As a result, these architectures have repeatedly surpassed traditional segmentation approaches, positioning deep learning as the dominant paradigm for tackling complex, real-world segmentation problems, including traffic sign recognition (Chen et al., 2024).

Motivated by this background, this study focuses on traffic sign image segmentation using a CNN-based approach, specifically employing the U-Net architecture, which are well-suited to preserving spatial detail during pixel-level segmentation. Hyperparameter tuning will be conducted to optimize the model's performance, with a specific focus on the number of convolutional filters, given that this parameter plays a critical role in determining how effectively the network extracts relevant features across different stages of the architecture.

2. Methods

This study's methodology, depicted in Figure 1, is organized into four main phases: collecting and preprocessing the dataset, designing the U-Net architecture, training the model, and finally testing and evaluating its performance.

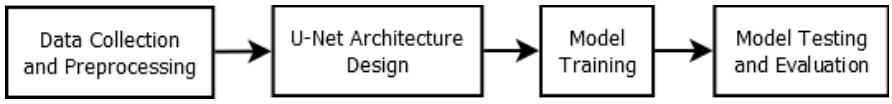


Fig. 1. Research methodology.

2.1. Dataset Collection and Preprocessing

The first phase of this research involves gathering traffic sign image data, which forms the core dataset for training and testing the model. This dataset was sourced from a prior study and consists of 1,750 images allocated for training purposes and 300 images set aside for testing (Akbar, 2021; Akbar et al., 2025). This dataset spans a range of traffic sign categories captured under varying illumination—including both daytime and nighttime conditions—as well as varying camera angles, object scales, and background compositions, providing sufficient visual diversity for the segmentation task. It comprises 10 types of traffic signs encompassing rhombus, circular, and octagonal shapes, spanning several regulatory categories—advisory, prohibition, warning, command, and guidance signs. All images are stored at a resolution of 128×128 pixels. Ground truth annotations for each image are generated using the LabelMe application, an image annotation tool that enables the manual delineation of traffic sign regions to produce pixel-level segmentation masks, which serve as the supervisory signal required for model training (Aulia & Akbar, 2026). During the preprocessing stage, the pixel intensity values are normalized so that they fall within the [0, 1] range.

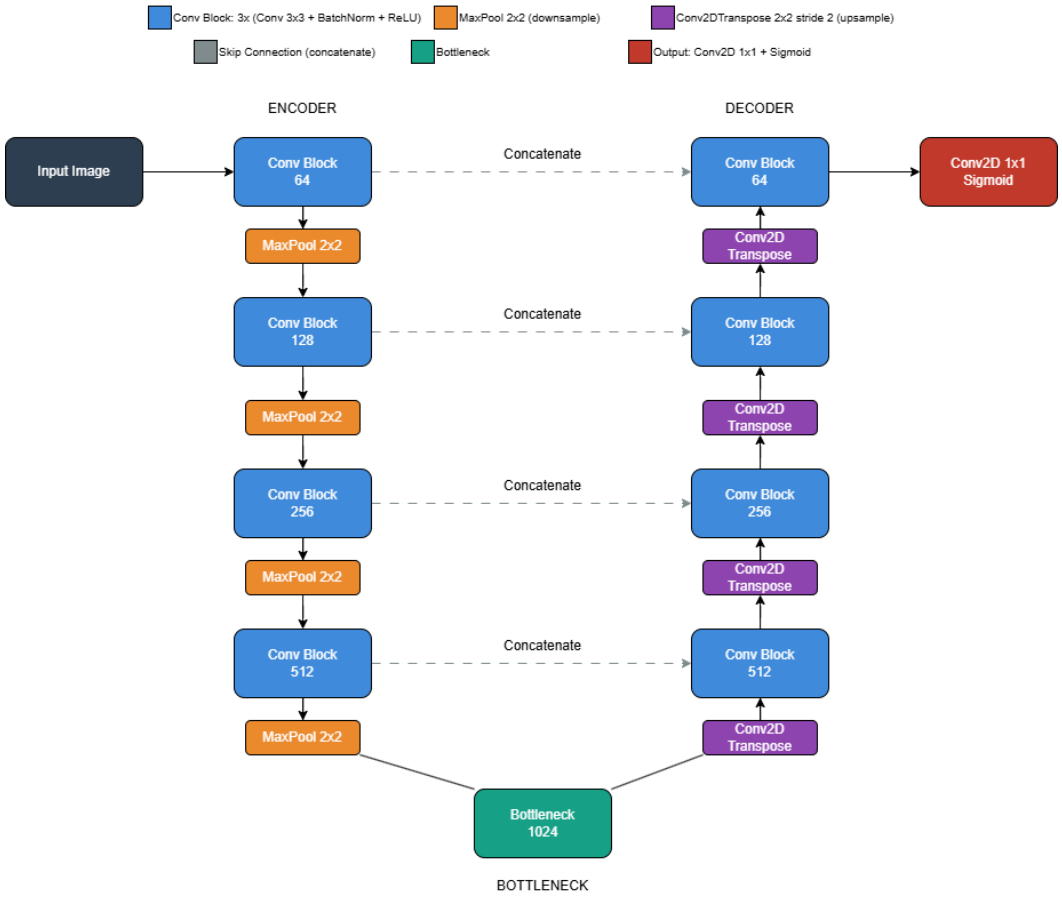


Fig. 2. U-Net architecture.

2.2. U-Net Architecture Design

As depicted in Figure 2, the architecture selected for this research is U-Net, a CNN variant developed specifically to handle image segmentation (Ronneberger et al., 2015). The network is structured around two principal components: an encoder and a decoder, connected through a bottleneck layer at the network's deepest level. The encoder path consists of four encoder blocks, each made up of a convolutional block followed by a max-pooling layer. Each convolutional block passes the input through three successive convolutional layers, with each layer followed by batch normalization and ReLU activation (Dimitrovski et al., 2024; Rajamani et al., 2023). Filter counts of 64, 128, 256, and 512 are applied across the four encoder blocks respectively, doubling progressively from one stage to the next (Ali et al., 2025; Neha et al., 2025). To downsample between encoder blocks, a 2×2 max-pooling operation is used, shrinking the spatial size of the feature maps (Hutchison et al., 2010). At the base of the network, a bottleneck layer employing 1,024 filters connects the encoder and decoder paths, serving as the point of maximum feature abstraction before the reconstruction process begins (Channa et al., 2026). The U-Net architecture is notably characterized by skip connections that join encoder and decoder layers operating at matching spatial resolutions (Manzi, 2026; Rani et al., 2024).

The decoder path is structured as a symmetrical counterpart to the encoder, made up of four decoder blocks containing 512, 256, 128, and 64 filters in turn. Every decoder block starts by applying a Conv2DTranspose layer, configured with a 2×2 kernel and stride of 2, which incrementally upsamples the feature maps to rebuild the original spatial dimensions (Park et al., 2025). The upsampled feature maps are then merged with their corresponding encoder skip connection at the same resolution, allowing spatial information lost during downsampling to be recovered (Zhan et al., 2024). This concatenated feature map is then fed through a convolutional block, mirroring the same three-layer convolution, batch normalization, and ReLU activation pattern used in the encoder, in order to refine the combined features before moving on to the next decoder stage (Javed et al., 2024). The output layer concludes the network with

a 1×1 convolution using one filter and a sigmoid activation, yielding a binary segmentation map where every pixel is assigned to either the traffic sign class or the background (Nwankpa et al., 2018).

2.3. Model Training

For training purposes, the preprocessed dataset is divided into training and testing subsets, with the U-Net model trained on the 1,750-image training subset. Optimization is carried out using the Adam algorithm, applying a learning rate of 0.0002, a batch size of 2, and running for 20 epochs (Akbar et al., 2025; Dogo et al., 2022). Binary cross-entropy functions as the primary loss function directing the optimization process, quantifying the pixel-level difference between the predicted segmentation output and the ground-truth mask. To gain a more thorough understanding of model performance during training, several supplementary metrics are tracked alongside the loss function, including accuracy, dice loss, dice coefficient, binary Intersection over Union (binary IoU), recall, and precision (Everingham et al., 2010; Milletari et al., 2016; Sathyanarayanan, 2024; Yeung et al., 2022). Dice-based and IoU-based metrics are particularly well-suited to segmentation tasks, offering a more meaningful gauge of segmentation quality than accuracy alone, especially given the binary classification of traffic sign versus background pixels (Seghier, 2024).

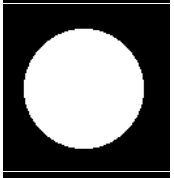
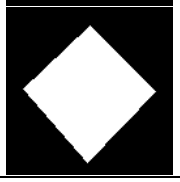
2.4. Model Testing and Evaluation

Upon completion of the training process, the resulting model is evaluated using a separate test dataset that was not involved in either the training stages, which includes 300 images. This testing phase is intended to assess the model's generalization capability in segmenting traffic sign images under previously unseen conditions. Model performance on the test dataset is evaluated quantitatively using the same set of metrics used during training: binary cross-entropy loss, accuracy, Dice loss, Dice coefficient, binary IoU, recall, and precision. These evaluation metrics allow for a direct comparison between the model's performance during training and its performance on unseen data. In addition to this quantitative analysis, visual inspection is conducted by comparing the model's segmentation outputs with the corresponding ground-truth masks across a range of test samples.

3. Results and Discussions

Table 1 presents a representative excerpt from the dataset employed to train and evaluate the proposed U-Net architecture for traffic sign segmentation, illustrating the fundamental input-output pairing that underpins the network's learning process.

Table 1. Example of dataset.

No	Original Image	Mask Image
1		
2		
3		

Each entry comprises an original image at a resolution of 128×128 pixels, capturing traffic signs in situ against heterogeneous natural backgrounds, such as foliage, sky, and supporting infrastructure, that introduce the visual complexity typical of real-world deployment conditions. Complementing each photograph is a corresponding mask image of identical spatial dimensions, manually annotated using the LabelMe application to delineate the precise

boundary of each traffic sign (Aulia & Akbar, 2026). The resulting annotations were subsequently converted into a binary matrix representation, in which pixels are assigned a value of 1 to denote the sign's occupied region and 0 to denote background, resulting in a stark white silhouette against a black canvas. The three exemplars depicted in Figure 1 are an octagonal STOP sign, a circular 40 km/h speed-limit indicator, and a rhomboid crossroad warning sign. They are deliberately curated to encompass the geometric diversity inherent to standardized traffic signage, thereby furnishing the model with sufficient shape variability.

The U-Net model undergoes training using a subset of 1,750 images, optimized through the Adam algorithm with a learning rate of 0.0002, a batch size of 2, and 20 epochs of training. The corresponding training outcomes—covering loss, dice loss, dice coefficient, binary IoU, precision, recall, and accuracy—are summarized in Table 2.

Table 2. U-Net model training results.

Epoch	Loss	Dice loss	Dice Coefficient	Binary IoU	Precision	Recall	Accuracy
1	0.137729	0.100357	0.899644	0.894176	0.937346	0.932262	0.945581
2	0.093903	0.066809	0.933192	0.926738	0.958762	0.952509	0.962985
3	0.088598	0.062852	0.937148	0.930292	0.960923	0.954816	0.964846
4	0.084653	0.0596	0.9404	0.932977	0.961968	0.957166	0.966242
5	0.081388	0.0574	0.9426	0.934889	0.96317	0.958351	0.967239
6	0.077959	0.055027	0.944973	0.93767	0.964285	0.960731	0.968679
7	0.076851	0.054222	0.945779	0.93805	0.965467	0.959975	0.968884
8	0.074109	0.052705	0.947295	0.939659	0.966376	0.961068	0.969717
9	0.075612	0.05331	0.946689	0.938985	0.965589	0.961035	0.969365
10	0.070896	0.05029	0.94971	0.942074	0.967404	0.963054	0.970962
11	0.070583	0.050443	0.949558	0.942049	0.967179	0.963257	0.970947
12	0.068332	0.048683	0.951317	0.943487	0.968226	0.963982	0.971692
13	0.06808	0.048326	0.951674	0.94399	0.968257	0.964583	0.971948
14	0.066792	0.047678	0.952322	0.94468	0.968268	0.965444	0.972301
15	0.065025	0.04657	0.953431	0.945873	0.969389	0.965777	0.972918
16	0.064093	0.04604	0.953959	0.946037	0.969249	0.966128	0.973001
17	0.062506	0.044886	0.955115	0.94754	0.969815	0.967428	0.97377
18	0.061052	0.043833	0.956167	0.948518	0.970442	0.968001	0.974273
19	0.061201	0.0441	0.9559	0.94853	0.970121	0.96835	0.974277
20	0.05989	0.043161	0.956839	0.94945	0.970726	0.968873	0.97475

The training trajectory over the twenty epochs, as shown in Table 2, reveals a model that exhibits consistent, well-behaved convergence. The loss function demonstrates a steady monotonic decline from 0.137729 at the inaugural epoch to 0.05989 by the twentieth. In contrast, the dice loss follows a parallel trajectory, diminishing from 0.100357 to 0.043161 over the same interval. This concomitant reduction in both loss metrics signals that the optimization process successfully navigated the loss landscape without succumbing to instability, oscillation, or premature stagnation. Correspondingly, the model's discriminative capacity, as reflected in the dice coefficient, binary IoU, precision, recall, and accuracy metrics, exhibits a commensurate upward trajectory that corroborates the diminishing loss values. The dice coefficient increases from 0.899644 to 0.956839, indicating a more precise overlap between predicted and ground-truth segmentation masks, while binary IoU shows a comparable improvement, rising from 0.894176 to 0.94945. This consistent co-variation between the dice coefficient and binary IoU aligns with their inherent mathematical relationship, as both metrics reflect the extent of spatial overlap between predicted and actual sign regions; their simultaneous improvement across epochs further strengthens confidence in the reliability of the segmentation outcomes. Precision and recall likewise climb from 0.937346 and 0.932262 to 0.970726 and 0.968873, respectively. This indicates that the model strikes a good balance between reducing false positives and false negatives. Additionally, the Accuracy metric, which reaches 0.97475 by the final epoch, further confirms the model's strong pixel-wise classification capability across the training subset.

Following training, the U-Net model is assessed using a separate testing subset comprising 300 images not included in the training set. Testing evaluation relies on the same set of metrics used throughout the training phase.

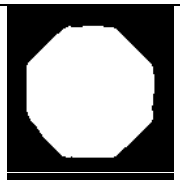
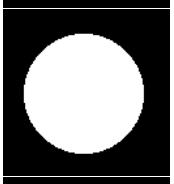
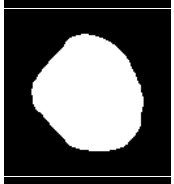
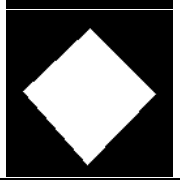
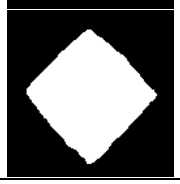
The evaluation of the trained U-Net model on the held-out testing subset, as shown in Table 3, yields performance metrics that largely corroborate the trends established during training, thereby lending credence to the model's capacity for generalization beyond the data on which it was optimized. The testing loss of 0.132471 and dice loss of 0.069065 remain commensurate with values observed in the early-to-mid training epochs, while the dice coefficient

of 0.930935 and accuracy of 0.953572 affirm that the network retains a substantial degree of pixel-wise segmentation fidelity when confronted with previously unseen traffic sign imagery. The binary IoU value of 0.907808 further corroborates this finding, remaining consistent with the dice coefficient and reinforcing the model's capacity to preserve spatial overlap between predicted and ground-truth sign regions even on unseen data. Particularly noteworthy is the disparity between precision (0.90499) and recall (0.981679), a pattern suggesting that the model exhibits a pronounced sensitivity in correctly identifying the vast majority of true traffic sign pixels, albeit at the modest expense of occasionally misclassifying a small proportion of background pixels as belonging to the traffic sign region. This precision-recall asymmetry implies that the model is disposed toward slight over-segmentation rather than under-segmentation (Tian et al., 2021).

Table 3. U-Net model testing results.

Loss	Dice loss	Dice Coefficient	Binary IoU	Precision	Recall	Accuracy
0.132471	0.069065	0.930935	0.907808	0.90499	0.981679	0.953572

Table 4. Example of predicted mask image and segmented original image.

No	Original Image	Mask Image	Predicted Mask	Segmented Original Image
1				
2				
3				

The qualitative results depicted in this figure furnish an instructive visual complement to the quantitative metrics discussed previously, offering a granular, case-by-case appraisal of the model's segmentation fidelity across the three representative traffic sign geometries. For the octagonal STOP sign and the rhomboid crossroad warning sign, the predicted masks exhibit a remarkably congruent correspondence with their respective ground-truth annotations, faithfully reproducing the angular vertices and overall silhouette with only negligible deviation along the peripheral contours. This fidelity underscores the model's competence in handling polygonal shapes characterized by well-defined edges and sharp geometric transitions. The circular speed-limit sign, however, reveals a discernible limitation in the model's predictive capacity: the predicted mask exhibits a conspicuous concave flattening along its upper-left boundary, deviating markedly from the smooth, uninterrupted circumference of the ground-truth mask. This localized segmentation failure propagates directly into the segmented output, where the sign's rim appears slightly clipped and irregular along that same edge rather than following its true round contour. Collectively, these visual findings suggest that while the model demonstrates robust generalization across signs with rectilinear or angular geometry, curvilinear boundaries, such as those inherent to circular signs, may pose a comparatively greater challenge, warranting further architectural adjustments to bolster boundary-level precision in future iterations of the model. Furthermore, the adoption of recall loss may offer an additional avenue for addressing this limitation, as such formulations are designed to penalize the omission of true sign pixels during training, potentially reducing the likelihood of similar boundary-level exclusions in curvilinear regions (Tian et al., 2021).

4. Conclusion

In this research, a CNN-based U-Net model is proposed for traffic sign image segmentation. Using 1,750 annotated images for training, the model was optimized with the Adam algorithm (0.0002 learning rate, batch size of 2) over 20

epochs. Results show smooth convergence behavior, with training loss falling from 0.082138 to 0.058175, dice coefficient climbing from 0.942267 to 0.958337, and binary IoU rising correspondingly from 0.934637 to 0.950863. By the final epoch, precision, recall, and accuracy reached 0.971208, 0.970139, and 0.975471, respectively. To assess generalization, the model was tested on 300 unseen images, producing a dice coefficient of 0.922835, binary IoU of 0.897754, precision of 0.893236, recall of 0.981682, and accuracy of 0.948113. The higher recall relative to precision indicates a slight tendency toward over-segmentation. Qualitative analysis reinforced these quantitative findings: the model segmented polygonal signs with sharp, well-defined edges, such as the octagonal STOP sign and the rhomboid crossroad warning sign, whereas the circular speed-limit sign exhibited a localized boundary failure in the form of a concave indentation. One direction for future work is architectural refinements to improve boundary-level precision on circular signs. In addition, the adoption of recall loss warrants further investigation as a means of explicitly regulating the trade-off between recall and precision during training, potentially curbing the model's tendency toward over-segmentation.

References

Akbar, M. (2021). *Indonesian Traffic Sign Dataset* [Graphic]. Zenodo. https://doi.org/10.5281/ZENODO.21194057

Akbar, M., Susilawati, I., Jati, B. S., & Alamsyah, N. (2025). Multi-Task Learning for Traffic Sign Recognition using Multi-Scale Convolutional Neural Networks. *International Journal of Advances in Data and Information Systems*, *6*(2), 391–402. https://doi.org/10.59395/ijadis.v6i2.1406

Akbar, M., Witanti, A., & Susilawati, I. (2019). GPU Accelerated Fuzzy C-Means (FCM) Color Image Segmentation. *Compiler*, *8*(2). https://doi.org/10.28989/compiler.v8i2.455

Ali, S. H., Ahmad, A., Ali, M., Khan, A., & Shaukat, N. (2025). *Automated MRI Tumor Segmentation using hybrid U-Net with Transformer and Efficient Attention* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2506.15562

Aulia, I., & Akbar, M. (2026). Segmentasi Citra Rambu Lalu Lintas Menggunakan Arsitektur DeepLabV3+ Berbasis Convolutional Neural Network. *Jurnal Sistem Informasi, Teknik Informatika dan Teknologi Pendidikan*, *6*(1), 64–70. https://doi.org/10.55338/justikpen.v6i1.896

Ben-Abbou, T., El Omrani, H., El Fazazy, K., Mahraz, M. A., Tairi, H., & Riffi, J. (2026). R2KAN-U-Net: A Novel Architecture Integrating Kolmogorov–Arnold Networks with Residual U-Net for Robust Traffic Sign Segmentation. *Sensors*, *26*(12), 3797. https://doi.org/10.3390/s26123797

Benfaress, I., Bouhoute, A., & Zinedine, A. (2025). Advancing Traffic Sign Recognition: Explainable Deep CNN for Enhanced Robustness in Adverse Environments. *Computers*, *14*(3), 88. https://doi.org/10.3390/computers14030088

Boly, S. B., & Akbar, M. (2024). Segmentasi Citra Sel Darah Menggunakan Convolutional Neural Network. *Jurnal RESTIKOM : Riset Teknik Informatika Dan Komputer*, *6*(2), 390–398. https://doi.org/10.52005/restikom.v6i2.336

Brar, K. K., Goyal, B., Dogra, A., Mustafa, M. A., Majumdar, R., Alkhayyat, A., & Kukreja, V. (2025). Image segmentation review: Theoretical background and recent advances. *Information Fusion*, *114*, 102608. https://doi.org/10.1016/j.inffus.2024.102608

Channa, A., Khan, A., Chandio, A. A., Akbar, A., Memon, S., Hussain, A., & Hamza, A. (2026). *PEFT-MedSAM: Efficient Fine-Tuning of Medical Foundation Models for Explainable Skin Lesion Segmentation* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2606.18707

Chen, H., Ali, M. A. H., Nukman, Y., Razak, B. A., Turaev, S., Chen, Y., Zhang, S., Huang, Z., Wang, Z., & Abdulghafor, R. (2024). Computational methods for automatic traffic signs recognition in autonomous driving on road: A systematic review. *Results in Engineering*, *24*, 103553. https://doi.org/10.1016/j.rineng.2024.103553

Dimitrovski, I., Spasev, V., Loshkovska, S., & Kitanovski, I. (2024). U-Net Ensemble for Enhanced Semantic Segmentation in Remote Sensing Imagery. *Remote Sensing*, *16*(12), 2077. https://doi.org/10.3390/rs16122077

Dogo, E. M., Afolabi, O. J., & Twala, B. (2022). On the Relative Impact of Optimizers on Convolutional Neural Networks with Varying Depth and Width for Image Classification. *Applied Sciences*, *12*(23), 11976. https://doi.org/10.3390/app122311976

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2010). The Pascal Visual Object Classes (VOC) Challenge. *International Journal of Computer Vision*, *88*(2), 303–338. https://doi.org/10.1007/s11263-009-0275-4

Fredj, H. B., Chabbah, A., Baili, J., Faiedh, H., & Souani, C. (2023). An efficient implementation of traffic signs recognition system using CNN. *Microprocessors and Microsystems*, *98*, 104791. https://doi.org/10.1016/j.micpro.2023.104791

Giuliani, D. (2022). Metaheuristic Algorithms Applied to Color Image Segmentation on HSV Space. *Journal of Imaging*, *8*(1), 6. https://doi.org/10.3390/jimaging8010006

Günthner, T., & Proff, H. (2021). On the way to autonomous driving: How age influences the acceptance of driver assistance systems. *Transportation Research Part F: Traffic Psychology and Behaviour*, *81*, 586–607. https://doi.org/10.1016/j.trf.2021.07.006

He, S., Chen, L., Zhang, S., Guo, Z., Sun, P., Liu, H., & Liu, H. (2021). Automatic Recognition of Traffic Signs Based on Visual Inspection. *IEEE Access*, *9*, 43253–43261. https://doi.org/10.1109/ACCESS.2021.3059052

Hu, T., Deng, Y., Deng, Y., & Ge, A. (2021). Fully Convolutional Network Variations and Method on Small Dataset. *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, 40–46. https://doi.org/10.1109/ICCECE51280.2021.9342059

Huang, G., Cao, H., Sun, J., Chen, Z., & Zhou, Z. (2025). Improved road traffic sign recognition from feature reconstruction. *Scientific Reports*, *15*(1), 42656. https://doi.org/10.1038/s41598-025-26757-9

Huang, S.-Y., Hsu, W.-L., Hsu, R.-J., & Liu, D.-W. (2022). Fully Convolutional Network for the Semantic Segmentation of Medical Images: A Survey. *Diagnostics*, *12*(11), 2765. https://doi.org/10.3390/diagnostics12112765

Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Scherer, D., Müller, A., & Behnke, S. (2010). Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition. In K. Diamantaras, W. Duch, & L. S. Iliadis (Eds.), *Artificial Neural Networks – ICANN 2010* (Vol. 6354, pp. 92–101). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-15825-4_10

Javed, S., Khan, T. M., Qayyum, A., Sowmya, A., & Razzak, I. (2024). *Region Guided Attention Network for Retinal Vessel Segmentation* (Version 3). arXiv. https://doi.org/10.48550/ARXIV.2407.18970

Li, S., Wang, S., & Wang, P. (2023). A Small Object Detection Algorithm for Traffic Signs Based on Improved YOLOv7. *Sensors*, *23*(16), 7145. https://doi.org/10.3390/s23167145

Li, X., Ma, Z., Wang, R., Sun, Z., Dai, M., Wang, Y., Liu, Z., & Ye, H. (2026). A Survey on Image Segmentation and Super-resolution Reconstruction in Visual Sensor Networks. *ACM Computing Surveys*, *58*(2), 1–29. https://doi.org/10.1145/3757730

Lim, X. R., Lee, C. P., Lim, K. M., Ong, T. S., Alqahtani, A., & Ali, M. (2023). Recent Advances in Traffic Sign Recognition: Approaches and Datasets. *Sensors*, *23*(10), 4674. https://doi.org/10.3390/s23104674

Manzi, P. R. (2026). *Volatility Surface Reconstruction using Deep Learning under No-Arbitrage Constraints* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2605.24031

Memon, M. M., Hashmani, M. A., Junejo, A. Z., Rizvi, S. S., & Raza, K. (2022). Unified DeepLabV3+ for Semi-Dark Image Semantic Segmentation. *Sensors*, *22*(14), 5312. https://doi.org/10.3390/s22145312

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *2016 Fourth International Conference on 3D Vision (3DV)*, 565–571. https://doi.org/10.1109/3DV.2016.79

Narkhede, M. M., & Chopade, N. B. (2022). Review of Advanced Driver Assistance Systems and Their Applications for Collision Avoidance in Urban Driving Scenario. In R. Misra, R. K. Shyamasundar, A. Chaturvedi, & R. Omer (Eds.), *Machine Learning and Big Data Analytics (Proceedings of International Conference on Machine Learning and Big Data Analytics (ICMLBDA) 2021)* (Vol. 256, pp. 253–267). Springer International Publishing. https://doi.org/10.1007/978-3-030-82469-3_23

Neha, F., Bhati, D., Shukla, D. K., Dalvi, S. M., Mantzou, N., & Shubbar, S. (2025). An analytics-driven review of U-Net for medical image segmentation. *Healthcare Analytics*, *8*, 100416. https://doi.org/10.1016/j.health.2025.100416

Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). *Activation Functions: Comparison of trends in Practice and Research for Deep Learning* (arXiv:1811.03378). arXiv. http://arxiv.org/abs/1811.03378

Park, M., Oh, S., Park, J., Jeong, T., & Yu, S. (2025). ES-UNet: Efficient 3D medical image segmentation with enhanced skip connections in 3D UNet. *BMC Medical Imaging*, *25*(1), 327. https://doi.org/10.1186/s12880-025-01857-0

Priya, B. L., Jayalakshmy, S., Idayachandran, G., & Kumaran, S. (2022). Performance Analysis of Semantic Segmentation using Optimized CNN based SegNet. *2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*, 1–5. https://doi.org/10.1109/ICSTSN53084.2022.9761293

Rajamani, K. T., Rani, P., Siebert, H., ElagiriRamalingam, R., & Heinrich, M. P. (2023). Attention-augmented U-Net (AA-U-Net) for semantic segmentation. *Signal, Image and Video Processing*, *17*(4), 981–989. https://doi.org/10.1007/s11760-022-02302-3

Ramyashree, Utsavi, S. R., Raghavendra, S., Anoop, B. N., & Venugopala, P. S. (2026). Comparative performance analysis of U-Net and DeepLabV3+ for semantic segmentation in traffic environments. *Scientific Reports*, *16*(1), 15614. https://doi.org/10.1038/s41598-026-46740-2

Rani, A. R., Anusha, Y., Cherishama, S. K., & Laxmi, S. V. (2024). Traffic sign detection and recognition using deep learning-based approach with haze removal for autonomous vehicle navigation. *E-Prime - Advances in Electrical Engineering, Electronics and Energy*, *7*, 100442. https://doi.org/10.1016/j.prime.2024.100442

Ren, X., Deng, Z., Ye, J., He, J., & Yang, D. (2023). *FCN+: Global Receptive Convolution Makes FCN Great Again* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2303.04589

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Vol. 9351, pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28

Sathyanarayanan, S. (2024). Confusion Matrix-Based Performance Evaluation Metrics. *African Journal of Biomedical Research*, 4023–4031. https://doi.org/10.53555/AJBR.v27i4S.4345

Seghier, M. L. (2024). Image Segmentation Evaluation With the Dice Index: Methodological Issues. *International Journal of Imaging Systems and Technology*, *34*(6), e23203. https://doi.org/10.1002/ima.23203

Seraj, M., Rosales-Castellanos, A., Shalkamy, A., El-Basyouny, K., & Qiu, T. Z. (2021). The Implications of Weather and Reflectivity Variations on Automatic Traffic Sign Recognition Performance. *Journal of Advanced Transportation*, *2021*, 1–15. https://doi.org/10.1155/2021/5513552

Sun, C., Wen, M., Zhang, K., Meng, P., & Cui, R. (2021). Traffic sign detection algorithm based on feature expression enhancement. *Multimedia Tools and Applications*, *80*(25), 33593–33614. https://doi.org/10.1007/s11042-021-11413-x

Suriya Prakash, A., Vigneshwaran, D., Seenivasaga Ayyalu, R., & Jayanthi Sree, S. (2021). Traffic Sign Recognition using Deeplearning for Autonomous Driverless Vehicles. *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, 1569–1572. https://doi.org/10.1109/ICCMC51019.2021.9418437

Tian, J., Mithun, N., Seymour, Z., Chiu, H.-P., & Kira, Z. (2021). *Striking the Right Balance: Recall Loss for Semantic Segmentation*. https://doi.org/10.48550/ARXIV.2106.14917

Wang, H., Liu, P., Dou, Q., Song, Y., Luo, M., Han, R., & Zhang, B. (2025). Enhanced edge detection via Dual-branch attention fusion with Canny-assisted supervision. *The Visual Computer*, *41*(12), 9765–9780. https://doi.org/10.1007/s00371-025-03998-3

Wang, J., Wan, X., Li, L., & Wang, J. (2021). An Improved DeepLab Model for Clothing Image Segmentation. *2021 IEEE 4th International Conference on Electronics and Communication Engineering (ICECE)*, 49–54. https://doi.org/10.1109/ICECE54449.2021.9674326

Wang, T., & Wong, H. S. (2023). A Two-Stage Color Image Segmentation Method Based on Saturation-Value Total Variation. *Advances in Applied Mathematics and Mechanics*, *15*(1), 94–117. https://doi.org/10.4208/aamm.OA-2021-0314

Yang, X., & Li, C. (2026). A new edge detection method for noisy image based on discrete fractional wavelet transform and improved Canny algorithm. *Expert Systems with Applications*, *298*, 129668. https://doi.org/10.1016/j.eswa.2025.129668

Yao, Y., Zhang, Y., Liu, Z., & Yuan, H. (2025). A Bridge Crack Segmentation Algorithm Based on Fuzzy C-Means Clustering and Feature Fusion. *Sensors*, *25*(14), 4399. https://doi.org/10.3390/s25144399

Yeung, M., Sala, E., Schönlieb, C.-B., & Rundo, L. (2022). Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics*, *95*, 102026. https://doi.org/10.1016/j.compmedimag.2021.102026

Zhan, S., Yuan, Q., Lei, X., Huang, R., Guo, L., Liu, K., & Chen, R. (2024). BFNet: A full-encoder skip connect way for medical image segmentation. *Frontiers in Physiology*, *15*, 1412985. https://doi.org/10.3389/fphys.2024.1412985

Zhang, C., Lu, W., Wu, J., Ni, C., & Wang, H. (2024). SegNet Network Architecture for Deep Learning Image Segmentation and Its Integrated Applications and Prospects. *Academic Journal of Science and Technology*, *9*(2), 224–229. https://doi.org/10.54097/rfa5x119