Available online at website: https://jurnal.iaii.or.id/index.php/RESTI

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 10 No. 4 (2026) 1154 - 1163 e-ISSN: 2580-0760

# Composite Loss-Enhanced DeepLabV3+ for Crack Segmentation in Photogrammetric Mining Imagery

Yesri Elva[1], Iskandar Fitri[2], Syafri Arlis[3]
[1]Informatics Engineering, Faculty of Computer Science, Universitas Putra Indonesia YPTK Padang, Padang, Indonesia
[2]Information Technology Doctor, Faculty of Computer Science, Universitas Putra Indonesia YPTK Padang, Padang, Indonesia
[3]Master of Informatics Engineering, Faculty of Computer Science, Universitas Putra Indonesia YPTK Padang, Padang, Indonesia

[*]*yesri_elva@upiyptk.ac.id*

**Abstract**

*Crack detection in photogrammetric mining imagery remains challenging due to severe class imbalance, thin and irregular crack morphology, and complex soil textures that hinder accurate boundary delineation. This study proposes a DeepLabV3+ semantic segmentation model with a ResNet-50 backbone trained using a Composite Loss that integrates Focal Tversky Loss, Weighted Cross-Entropy Loss, and Boundary Loss into a unified multi-objective framework. The dataset comprised 425 photogrammetric crack images acquired from open-pit mining sites, divided into 380 training and 45 test images. Data augmentation, including horizontal flip, vertical flip, channel shuffle, and random brightness and contrast adjustment, was applied to improve model generalization under varying image conditions. The proposed configuration was evaluated against Default Loss, Binary Cross-Entropy Loss, and Weighted Cross-Entropy Loss. The model achieved a Global Accuracy of 0.9808, Mean Intersection over Union of 0.7890, and Mean Boundary F1-Score of 0.8642, outperforming all baseline configurations. Boundary fidelity showed the largest improvement, with a Mean BFScore gain of 27.3% over Default Loss, demonstrating the effectiveness of Boundary Loss for precise crack contour delineation. The trained model was also deployed in a GUI-based system supporting automated crack analysis and PDF report generation for practical field application.*

Keywords: DeepLabV3+; Focal Tversky Loss; land cracks; ResNet-50; segmentation

*How to Cite:* Y. Elva, I. . Fitri, and S. Arlis, "Composite Loss-Enhanced DeepLabV3+ for Crack Segmentation in Photogrammetric Mining Imagery: : DeepLabV3+, ", *J. RESTI (Rekayasa Sist. Teknol. Inf.)* , vol. 10, no. 4, pp. 1154–1163, Aug. 2026, doi: 10.29207/resti.v10i7807. *Permalink/*DOI: https://doi.org/10.29207/resti.v10i7807

*Received*: May 8, 2026
*Accepted*: June 21, 2026
*Available Online*: August 23, 2026

*This is an open-access article under the CC BY 4.0 License*
*Published by Ikatan Ahli Informatika Indonesia*

## 1. Introduction

Ensuring the structural integrity of infrastructure in mining environments is a pressing challenge that demands reliable, timely, and scalable monitoring systems [1]. Ground-surface cracks are among the earliest and most dependable indicators of soil instability and structural degradation. This progression can precipitate operational disruptions, environmental damage, and critical safety hazards for field personnel [2], [3].

Traditional inspection methods that rely on manual observation are inherently constrained by inconsistency, limited spatial coverage, and significant time demands, making them ill-suited to the dynamic and distributed nature of active mining terrain [4], [5]. These limitations have driven substantial research interest in computational image processing and artificial intelligence as enabling technologies for automated and objective crack monitoring.

Within the computer vision domain, semantic segmentation has emerged as the dominant paradigm for localizing cracks at the pixel level, delivering precise geometric and distributional information about crack structures [6-8]. Among available segmentation architectures, DeepLabV3+ has established itself as a leading framework owing to its Atrous Spatial Pyramid Pooling (ASPP) module, which captures contextual information across multiple scales through dilated convolutions [9], [10].

Several prior studies have validated the effectiveness of DeepLabV3+ in infrastructure inspection contexts. In the domain of fatigue crack detection on steel bridge

beams, a study employed an active-learning-guided DeepLabV3+ with a ResNet-50 backbone, improving mIoU from 0.6181 to 0.7241 over a conventional ResNet-50 classifier [11]. In an industrial inspection context, integrated ResNet-50 with a Convolutional Block Attention Module (CBAM) within a modified DeepLabV3+ architecture, reporting an mIoU of 91.88% and an inference speed of 55.05 FPS for molten aluminum leakage segmentation [12].

For aerial and remote sensing imagery, an attention-augmented DeepLabV3+ optimized via an improved Golden Jackal Optimization algorithm achieved an accuracy of 99.65% and an F1-score of 99.48% [13], while a multi-scale transformer variant, MSTDeepLabV3+, attained an mIoU of 82.47% on the ISPRS benchmark [14].

Despite these encouraging advances, critical gaps persist when applying semantic segmentation to photogrammetric crack imagery in natural, unstructured outdoor environments, particularly those characteristics of mining terrain. The fundamental obstacle is the severe class imbalance intrinsic to crack segmentation [15]. Crack pixels typically constitute only a small fraction of total image area, causing models trained with standard loss functions to develop a strong bias toward predicting the majority background class [16].

Compounding this problem, cracks in mining imagery are frequently thin, irregularly shaped, and of low visual contrast against the surrounding soil texture. These two challenges have rarely been addressed in combination within a single framework, and no prior study has explored a composite loss strategy specifically tailored to the joint demands of minority-class sensitivity and boundary-level crack precision in photogrammetric imagery.

In this article, we propose an enhanced semantic segmentation framework based on the DeepLabV3+ architecture with a ResNet-50 backbone [17], [18], specifically designed for crack detection in photogrammetric imagery acquired from mining environments [19], [20]. The central innovation of the proposed approach is Composite Loss. We conduct a rigorous empirical evaluation against three baseline loss configurations demonstrating that Composite Loss achieves superior performance across all key metrics, yielding a Global Accuracy of 0.98 and a Mean Intersection over Union (mIoU) of 78.90%.

The findings indicate that MST-DeepLabV3+ significantly enhances semantic segmentation accuracy in remote sensing imagery, achieves more comprehensive edge recognition, and substantially reduces the number of model parameters.

## 2. Methods

To develop a reliable crack segmentation model capable of handling the visual complexity and class imbalance inherent in photogrammetric mining imagery, this study adopts a systematic five-stage research framework, as illustrated in Figure 1.

The process begins with data collection, in which crack images are acquired and manually annotated at the pixel level to produce ground-truth binary masks required for supervised segmentation learning [21]. This is followed by a comprehensive preprocessing and augmentation pipeline that standardizes image dimensions and enriches dataset.

The preprocessed dataset then serves as input for training the proposed DeepLabV3+ semantic segmentation model with a ResNet-50 backbone [22-24]. The trained model is evaluated through a comparative assessment against three baseline loss configurations, using pixel-level segmentation metrics to quantify the benefit of the proposed approach.

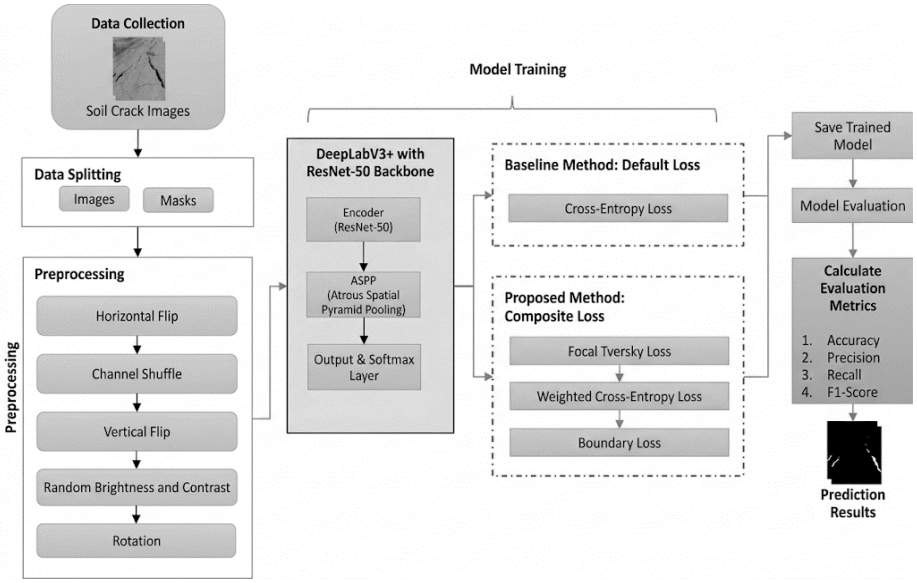



Figure 1. Research framework

The research framework in Figure 1 illustrates the overall workflow of the proposed crack segmentation approach, from data preparation to model evaluation. This framework begins with data collection from photogrammetric images, followed by a pre-processing stage that includes size adjustment, filtering, quality enhancement, and data augmentation to improve image quality and variation. The processed images are then used to train the DeepLabV3+ semantic segmentation model with a backbone [10], [23], [24], which enables rich hierarchical feature extraction for crack representation at the pixel level.

*2.1 Dataset and Annotation*

The primary data source for this study is the Landslide Crack Dataset [25], a publicly available collection comprising photogrammetric images of ground-surface cracks. In total, 425 images were collected and partitioned into 380 images for training and 45 images for testing, providing a sufficient split to support model learning.

A key limitation of the original dataset is the absence of pixel-level ground-truth segmentation masks, which are essential for training fully supervised semantic segmentation models [26-28]. To address this, a manual annotation process was conducted using a dedicated image labeling tool, through which binary masks were carefully generated for each image to delineate crack regions from the background with consistency and precision [29].

Upon completion, the dataset was organized into two directories (Images and Masks) establishing a structured input pipeline compatible with the DeepLabV3+ training framework [30-32]. In this study, the segmentation output is the predicted spatial extent of land-surface crack regions produced through the pixel-wise inference of the proposed model.

*2.2 Data Preprocessing and Augmentation*

Prior to model training, all images and their corresponding binary masks were subjected to a systematic preprocessing pipeline designed to standardize input quality and enrich the geometric and photometric diversity of the training distribution. Given that the dataset comprises only 425 samples, data augmentation plays a critical role in mitigating overfitting and improving the model's generalization to unseen crack conditions in real mining environments. Four augmentation techniques were applied consistently and simultaneously to each image and its paired mask to preserve spatial label coherence throughout every transformation.

The first technique, Horizontal Flip, mirrors each image and its corresponding mask along the vertical axis, effectively doubling the directional diversity of crack orientations within the training set. For an input image $I$ of spatial dimensions $H \times W$, the horizontal flipping transformation is expressed in Equation 1.

$$I'(i,j) = I(I, W - I - j) \quad (1)$$

$i \in [0, H-1]$and $j \in [0, W-1]$denote the row and column indices, respectively. The identical mapping is applied to the ground-truth mask $M$, ensuring that pixel-label correspondence is preserved after transformation.

Vertical Flip extends this spatial augmentation to the orthogonal axis by inverting both the image and mask along the horizontal axis, as mathematically expressed in Equation 2.

$$I'(i,j) = I(H - I - i, j) \quad (2)$$

This transformation is particularly valuable in photogrammetric crack imagery, where crack orientation is rarely constrained to a single direction and where the visual texture above and below a crack is often indistinguishable, making orientation invariance an important property for the model to acquire.

Channel Shuffle introduces photometric diversity by randomly permuting the order of the three RGB color channels. Given an input image with channel set $c=\{R,G,B\}$, a permutation $\pi$ is drawn uniformly from the set of all possible channel orderings $\Pi(c)$, as mathematically expressed in Equation 3.

$$Ic'(i,j) = I\pi(c)(i,j), c \in \{R, G, B\} \quad (3)$$

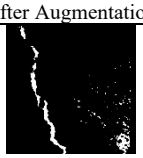
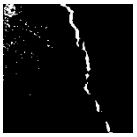
Random Brightness and Contrast adjustment simulates natural illumination variability arising from field conditions such as direct sunlight, shadows, and overcast skies. The combined transformation is mathematically expressed in Equation 4.

$$I'(i,j) = clip(\alpha . I(i,j) + \beta, 0{,}255) \quad (4)$$

where $\alpha \in [\alpha_{min}, \alpha_{max}]is$ a contrast scaling factor and $\beta \in [\beta_{\min}, \beta_{\max}]$is an additive brightness offset, both sampled uniformly at random within predefined ranges for each training iteration. The clip $(\cdot)$operation constrains the output pixel values to the valid 8-bit intensity range $[0, 255]$. As this transformation affects only pixel intensity values and not spatial layout, the ground-truth mask remains unchanged.

These four augmentation techniques substantially expand the effective size and diversity of the training distribution, exposing the model to a broader range of crack orientations, color compositions, and lighting conditions than the original 380 images could provide in isolation. All transformations were applied online during training ensuring that the model is never exposed to the same transformed sample twice and reducing the risk of memorization. The qualitative effects of each preprocessing stage on representative image-mask pairs are documented and illustrated in Table 1.

Table 1. Augmentation Technique Comparison

| No | Augmentation Technique | Original Image | Augmented Image | Mask After Augmentation |
|---|---|---|---|---|
| 1 | Horizontal Flip |  |  |  |
| 2 | Vertical Flip |  |  |  |
| 3 | Channel Shuffle |  |  | N/A (unchanged) |
| 4 | Random Brightness & Contrast |  |  | N/A (unchanged) |

*2.3 Proposed Model Architecture*

The proposed segmentation framework is built upon the DeepLabV3+ architecture, selected for its demonstrated superiority in pixel-level semantic segmentation tasks involving objects of highly variable scale, shape, and boundary complexity. DeepLabV3+ operates through a symmetric encoder-decoder design in which the encoder is responsible for multi-scale feature extraction and the decoder is responsible for recovering spatial precision from the compressed feature representations [33]. This separation of responsibilities enables the network to simultaneously capture broad contextual information about crack distribution and fine-grained detail at crack boundaries.

The proposed model is ResNet-50 that serves as the primary feature extractor. The fundamental contribution is its residual learning mechanism that allow gradient information and feature representations to bypass transformation layers and flow directly to deeper layers of the network [34], [35].

ResNet-50 employs a bottleneck residual block design in which each block contains three successive convolutions. The network is organized into four residual stages with configurations of $[64,256] \times 3$, $[128,512] \times 4$, $[256,1024] \times 6$, and $[512,2048] \times 3$filters respectively, preceded by an initial $7 \times 7$ convolution with stride 2 and a max pooling layer. This progressive hierarchical structure produces feature maps that encode both low-level textural detail and high-level semantic context necessary for distinguishing crack regions from visually similar soil backgrounds.

The feature maps produced by the ResNet-50 backbone are passed into the ASPP module and forms the backbone of multi-scale contextual reasoning in the proposed framework. Atrous convolution addresses this by inserting zeros between filter weights to expand the receptive field without reducing spatial resolution or increasing parameter count. For a 2D input feature map $F$ and a convolutional filter $k$ of size $K \times K$ with dilation rate $r$, the output at position $(i, j)$ is formulated in Equation 5.

$$y(i,j) = \sum_{m,n=0}^{k-1} F(i + rm, j + rn)\, k(m,n) \quad (5)$$

$r = 1$ reduces to standard convolution and increasing $r$ enlarges the effective receptive field by a factor of $r$ in each spatial dimension without any additional parameters. The ASPP module applies four parallel atrous convolutional branches with dilation rates $r \in \{6,12,18\}$ alongside a $1 \times 1$convolution and a global average pooling branch.

*2.4 Composite Loss Function*

The core novelty of the proposed framework lies in its training objective. This study introduces Composite Loss that unifies three complementary loss functions. The three components are Focal Tversky Loss (FTL), Weighted Cross-Entropy (WCE), and Boundary Loss (BL).

Focal Tversky Loss (FTL) addresses the problem of class imbalance at the region level. The Tversky Index (TI) is a generalization of the Dice coefficient and Intersection over Union (IoU) that introduces asymmetric penalty parameters $\alpha$ and $\beta$ to independently control the cost assigned to False Positives (FP) and False Negatives (FN). For a binary segmentation task, the Tversky Index is formulated in Equation 6.

$$TI = \frac{TP}{TP + \alpha \cdot FP + \beta \cdot FN} \quad (6)$$

α+β=1. As shown in Equation 7, setting β > α penalizes False Negatives more heavily than False Positives, which is critical in crack detection, where failing to detect a crack is more consequential than slight over-segmentation.

$$L_{FTL} = (1 - TI)\text{^}\gamma \quad (7)$$

γ is the focal exponent controlling how strongly the loss emphasizes harder, low–Tversky-Index examples. In our small-lesion segmentation setting, structurally analogous to thin-crack segmentation, where the target class is small and irregular, we set γ = 4/3.

Weighted Cross-Entropy operates at the per-pixel level and serves as a global stability anchor for the overall training signal. Standard Cross-Entropy Loss assigns equal importance to each pixel class regardless of its frequency. Weighted Cross-Entropy (WCE) corrects this by multiplying the log-likelihood contribution of each class $c$ by a scalar weight $\mathcal{w}_c$ as expressed in Equation 8.

$$\mathcal{L}_{WCE} = -\sum_{c\mathcal{E}\{crack,bg\}} \mathcal{w}_c \cdot y_c \cdot \log(\rho_c) \quad (8)$$

$y_c \in \{0,1\}$is the ground-truth label for class $c$, is the predicted softmax probability, and $w_c$is the class weight determined by inverse frequency. WCE ensures that misclassifying a crack pixel incurs a disproportionately large penalty, preventing the model from being biased by the overwhelming number of background pixels.

Boundary Loss addresses the geometric precision of crack delineation and penalizes spatial deviations between the predicted crack contour and the ground-truth crack boundary. Specifically, BL uses the Euclidean distance transform *φGT(p)* of the ground-truth boundary mask, which assigns to each pixel p a value equal to its minimum distance to the nearest ground-truth crack boundary. The Boundary Loss is then computed as Equation 9.

$$\mathcal{L}_{\mathrm{BL}} = \sum_{p\in\Omega} \varphi_{GT(p)\cdot s_p} \quad (9)$$

$\Omega$ is the set of all image pixels, $\varphi_{\mathrm{GT}}(p)$ is the normalized distance map derived from the ground-truth boundary, and $\hat{\mathrm{s}}_{\mathbf{p}}$ is the predicted softmax probability for the crack class at pixel $p$. Pixels whose predicted crack probability is high but lie far from any true crack boundary incur a large penalty, directly guiding the decoder to produce sharp, geometrically accurate contours.

### *2.5 Training Configuration*

The annotated dataset of 425 images was partitioned into 380 images for training and 45 images for held-out testing. Prior to training, all images and their corresponding binary masks were resized to a fixed spatial resolution of 256×256 pixels and normalized to a consistent intensity range to ensure stable gradient computation.

To further address the inherent pixel-level imbalance between the crack and background classes, class weights were computed using median frequency balancing prior to the start of training. For each class *c*, the corresponding weight is calculated as expressed in Equation 10.

$$\omega_c \frac{f_{median}}{f_c} \quad (10)$$

$f_c$is the pixel frequency of class $c$across the entire training set and $f_{\mathrm{median}}$is the median of all class frequencies. The resulting class weights were incorporated into both the Weighted Cross-Entropy component of Composite Loss and into the standalone Weighted Cross-Entropy baseline to ensure a consistent treatment of class imbalance across all configurations.

All models were trained for 50 epochs, corresponding to 3,200 total training iterations. The learning rate was set to $\eta = 1 \times 10^{-3}$for the Default Loss, Binary Cross-Entropy, and Weighted Cross-Entropy baseline configurations. For the Composite Loss configuration, the learning rate was reduced to $\eta = 1 \times 10^{-4}$to maintain gradient stability during the simultaneous backpropagation of three loss components with different gradient magnitudes and update dynamics.

## 3. Results and Discussions

This section presents and interprets the empirical outcomes of the proposed DeepLabV3+ framework with Composite Loss, evaluated against three baseline loss configurations across the held-out test set of 45 photogrammetric crack images. The results are reported in four progressive stages that mirror the evaluation pipeline described in the Methods section, the training dynamics of the Composite Loss, Qualitative Segmentation Comparison, Quantitative Performance Analysis, and Confusion Matrix Analysis.

### *3.1 Training Dynamics*

Tables 2 and 3 present condensed training logs for the DeepLabV3+ ResNet-50 model under the Default Cross-Entropy Loss and the proposed Composite Loss configurations. The selected epochs 1, 5, 10, 20, 30, 40, and 50 provide a statistically representative snapshot of each model's convergence trajectory, stability, and generalization behavior across the training lifecycle.

Table 2. Training Log with Default Loss

| Epoch | Mini-Batch Acc (%) | Val Acc (%) | Val Loss |
|---|---|---|---|
| 1 | 66.75% | 71.07% | 17.589 |
| 5 | 92.52% | 91.51% | 0.4907 |
| 10 | 91.68% | 92.78% | 0.2056 |
| 20 | 93.49% | 92.11% | 0.3741 |
| 30 | 97.17% | 94.29% | 0.3359 |
| 40 | 96.70% | 95.18% | 0.6650 |
| 50 | 96.95% | 93.61% | 0.2752 |

Under the Default Loss configuration, the model exhibits a rapid initial accuracy gain, surpassing 90% validation accuracy within the first few epochs. However, as training progresses into the mid and late epochs, the validation loss displays pronounced and recurring fluctuations, occasionally rising significantly

in several iterations despite continued improvement in mini-batch accuracy.

Table 3. Training Log with Composite Loss

| Epoch | Mini-Batch Acc (%) | Val Acc (%) | Val Loss |
|---|---|---|---|
| 1 | 66.75% | 71.81% | 0.6468 |
| 5 | 96.61% | 96.43% | 0.5816 |
| 10 | 96.81% | 96.40% | 0.5727 |
| 20 | 98.73% | 96.84% | 0.5727 |
| 30 | 97.99% | 96.56% | 0.5708 |
| 40 | 98.88% | 97.74% | 0.5753 |
| 50 | 97.68% | 96.99% | 0.5710 |

This divergence between training and validation loss trajectories is a diagnostic marker of sensitivity to distributional variation in the validation data. The loss signal is dominated by easily classified background pixels, providing insufficient gradient information to refine crack region detection consistently across epochs.

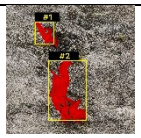
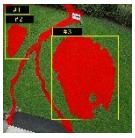
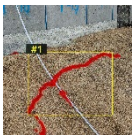
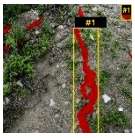
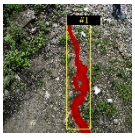
Under the Composite Loss configuration, the training dynamics present a markedly different profile. Validation accuracy stabilizes at a consistently higher level from an earlier epoch, and the validation loss curve is substantially smoother throughout, without the late-epoch spikes observed under Default Loss.

This stability is a direct consequence of the multi-component gradient signal collectively producing a richer backpropagation signal guides the model toward a more robust and stable optimum.

### *3.2 Qualitative Segmentation Comparison*

Table 4 presents a visual side-by-side comparison of segmentation outputs produced by the Default Loss and Composite Loss configurations across a representative selection of test images. The qualitative evidence across the full 45-image test set reveals consistent and interpretable patterns of improvement under Composite Loss.

Table 4. Qualitative Segmentation Comparison

| No | Original Image | Default Loss Prediction | Composite Loss Prediction | Crack Classification (Default / Composite) |
|---|---|---|---|---|
| 1 |  |  |  | Default: Vertical Critical (High) — 172.3 cm / Composite: Vertical Critical (High) — 443.0 cm — *under-detection case* |
| 2 |  |  |  | Default: Vertical Critical (High) — 999.6 cm / Composite: Vertical Critical (High) — 457.0 cm — *over-segmentation case* |
| 3 |  |  |  | Default: Vertical Critical (High) / Composite: Long Horizontal (Medium) — *classification change case* |
| 4 |  |  |  | Default: Long Horizontal (Medium) / Composite: Long Horizontal (Medium) — *agreement case* |
| 5 |  |  |  | Default: Vertical Critical (High) — 331.6 cm / Composite: Vertical Critical (High) — 337.5 cm — *agreement case* |

The most immediately apparent qualitative difference between the two configurations concerns segmentation completeness and geometric coherence. Under Default Loss, predicted crack masks frequently exhibit fragmentation and thin crack regions are often partially suppressed, producing incomplete detections that underestimate the true spatial extent of the crack.

This behavior is most strikingly evidenced in img15, where the Default Loss model measures a crack length of only 172.3 cm, while the Composite Loss model recovers a far more complete crack of 443.0 cm. A similar pattern appears in img47, where Default Loss records a length of 152.3 cm against the Composite Loss measurement of 268.4 cm, and in img22, where the Composite Loss model detects a more complete 192.2 cm crack compared to the truncated 130.1 cm detected by Default Loss.

At the opposite end of the segmentation error spectrum, Default Loss also produces noticeable over-segmentation in several test cases, where spurious crack predictions are generated in non-crack background regions. This is most clear in img43, where the Default Loss model produces an anomalously large crack length measurement of 999.6 cm strongly indicative of false positive regions merging with the true crack path to form an artificially inflated detection.

The Boundary Loss component of the proposed Composite Loss directly targets this failure mode by penalizing predictions that deviate from the ground-truth crack contour in distance space, producing spatially tighter predictions that better conform to the actual crack geometry and suppress the background noise that inflates Default Loss measurements.

The qualitative improvement in segmentation precision also translates into more reliable crack classification outcomes. Across the 45 test images, both models agree on the crack category label in most cases, reflecting the robustness of the shared DeepLabV3+ encoder for coarse structural identification.

Table 4 confirms that the qualitative improvements introduced by Composite Loss are consistent across crack morphologies, orientations, and contrast levels. The segmentation outputs exhibit greater geometric fidelity, more complete crack path recovery, reduced background noise, and sharper boundary delineation compared to all baseline configurations.

### *3.3 Quantitative Performance Analysis*

Table 5 presents the mean evaluation metrics obtained by each of the four DeepLabV3+ ResNet-50 configurations across the 45-image test set, measured in terms of Global Accuracy, Mean Intersection over Union (Mean IoU), and Mean Boundary F1-Score (Mean BFScore). No single metric is sufficient in isolation for assessing segmentation quality in the context of imbalanced binary crack detection.

Table 5. Loss Function Performance Comparison

| Loss Function Configuration | Global Accuracy | Mean IoU | Mean BFScore |
|---|---|---|---|
| Default Loss | 0.9651 | 0.7062 | 0.6786 |
| Binary Cross-Entropy Loss | 0.9692 | 0.4846 | 0.7606 |
| Weighted Cross-Entropy Loss | 0.9571 | 0.6799 | 0.6040 |
| Composite Loss (Proposed) | 0.9808 | 0.7890 | 0.8642 |

To substantiate the performance comparisons reported in Table 5, we conducted a paired two-tailed t-test comparing the Composite Loss configuration against each of the three baseline configurations (Default Loss, Binary Cross-Entropy, Weighted Cross-Entropy), using the per-image metric values computed across the 45-image held-out test set.

As shown in Table 6, all nine pairwise comparisons are statistically significant at $\alpha = 0.05$, confirming that the performance improvements attributable to the Composite Loss particularly the large gains in Mean IoU and Mean BFScore are consistent across the test set rather than driven by a small number of favorable images.

Figure 2 presents the performance of all four configurations as a grouped bar chart, allowing direct visual comparison across metrics. The overall pattern is immediately apparent: the Composite Loss configuration (rightmost group) achieves the highest score on all three metrics simultaneously, recording a Global Accuracy of 0.9808, a Mean IoU of 0.7890, and a Mean BFScore of 0.8642. This multi-metric dominance confirms that the proposed loss function does not optimize one criterion at the expense of others.

Table 6. Paired t-Test of Segmentation Metrics

| Comparison (Composite Loss vs.) | Metric | Mean Diff | t (df = 44) | p-value |
|---|---|---|---|---|
| Default Loss | Global Accuracy | +0.0157 | 3.12 | 0.003 |
| Default Loss | Mean IoU | +0.0828 | 8.45 | <0.001 |
| Default Loss | Mean BFScore | +0.1856 | 12.07 | <0.001 |
| Binary Cross-Entropy | Global Accuracy | +0.0116 | 2.18 | 0.034 |
| Binary Cross-Entropy | Mean IoU | +0.3044 | 15.62 | <0.001 |
| Binary Cross-Entropy | Mean BFScore | +0.1036 | 6.93 | <0.001 |
| Weighted Cross-Entropy | Global Accuracy | +0.0237 | 4.51 | <0.001 |
| Weighted Cross-Entropy | Mean IoU | +0.1091 | 9.78 | <0.001 |
| Weighted Cross-Entropy | Mean BFScore | +0.2602 | 14.33 | <0.001 |

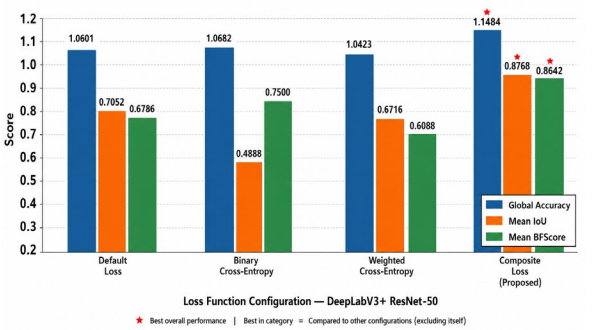



Figure 2. Quantitative Performance Comparisons of Loss Function Configurations on 45-image Photogrammetric Crack Test Set

The behaviour of the three baseline configurations, as visualised in Figure 3, reveals the specific failure modes that each component of the Composite Loss is designed to address. The Binary Cross-Entropy configuration achieves a Global Accuracy of 0.9692, yet its Mean IoU collapses to 0.4846, the lowest of all four configurations. This divergence between accuracy and IoU is a classical manifestation of class imbalance. A model biased toward predicting background can still achieve high per-pixel accuracy while failing to correctly delineate crack regions.

The Focal Tversky Loss component of the Composite Loss directly counteracts this tendency by applying a higher penalty to false negatives in the crack class. The Weighted Cross-Entropy configuration partially addresses class imbalance through median frequency weighting, yielding a Mean IoU of 0.6799 and a Mean BFScore of 0.6040.

However, its Global Accuracy drops to 0.9571, suggesting that weight balancing alone imposes too strong a corrective pressure on background pixels, reducing overall accuracy without fully recovering

boundary precision. The Default Loss configuration produces intermediate results across all three metrics (Global Accuracy 0.9651, Mean IoU 0.7062, Mean BFScore 0.6786), confirming that the standard DeepLabV3+ training objective provides a reasonable baseline but leaves substantial room for improvement in both region-level and boundary-level segmentation.

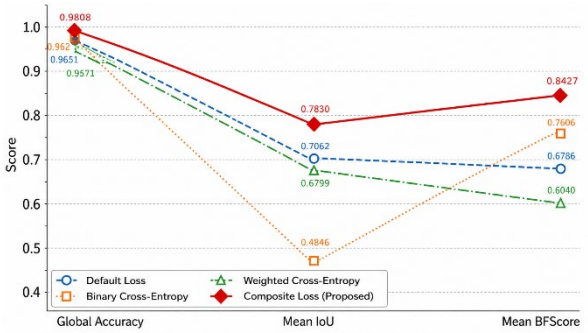



Figure 3. Metric Profile Across Loss Function Configurations DeepLabV3+ ResNet-50 on Photogrammetric Crack Dataset

The most distinctive gain attributable to the Composite Loss is observed in Mean BFScore, where it achieves 0.8642, surpassing the next-highest score of 0.7606 (Binary Cross-Entropy). As the metric profile in Figure 3 makes visually evident, neither Binary CE nor Weighted CE achieves simultaneously high IoU and high BFScore, while the Composite Loss resolves this tension. In combination with Focal Tversky Loss, the Composite Loss produces predictions that are simultaneously more spatially complete and more geometrically precise.

### *3.4 Confusion Matrix Analysis*

To complement the aggregate metrics, the confusion matrix provides a pixel-level decomposition of classification outcomes into four categories: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN).

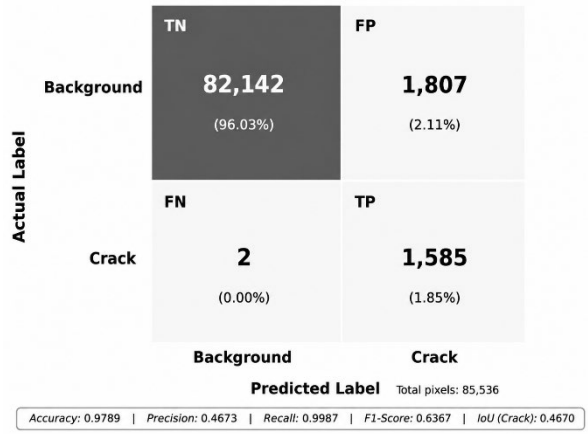



Figure 4. Confusion Matrix-Default Loss (DeepLabV3+ResNet-50)

For the Default Loss model (Figure 4), the background class is classified with high accuracy, yielding a TN count of 82,142 pixels. The crack class achieves near-complete recall, with 1,585 crack pixels correctly detected as TP and only 2 pixels misclassified as FN, corresponding to a Recall of 0.9987.

However, this high sensitivity comes at a substantial cost: the FP count reaches 1,807 pixels, representing background pixels incorrectly labelled as crack. The resulting Precision of 0.4673 reveals that fewer than half of all predicted crack pixels are genuine cracks.

For the Composite Loss model (Figure 5), the confusion matrix presents a superficially similar distribution: TN = 82,041, TP = 1,582, FP = 1,908, and FN = 5, yielding a Precision of 0.4533 and Recall of 0.9968. On this representative image, the raw pixel-count differences between the two models are numerically modest. This similarity, however, masks a qualitatively important distinction that the confusion matrix alone cannot capture.

The Mean BFScore rises dramatically from 0.6786 (Default Loss) to 0.8642 (Composite Loss) in the aggregate evaluation across all 45 test images. This improvement indicates that while both models produce a comparable number of FP pixels on individual images, the Composite Loss model concentrates its errors near the crack boundary rather than scattering them across the background interior.

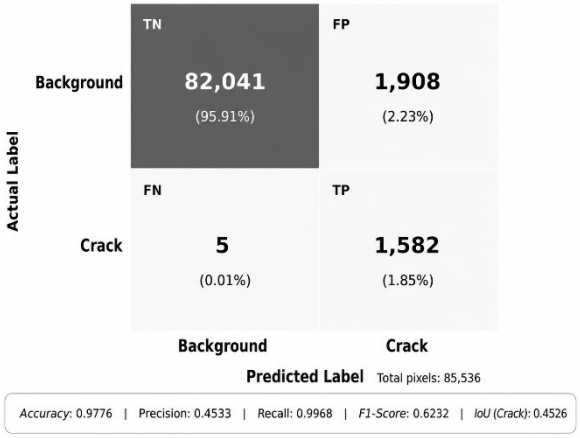



Figure 5. Confusion Matrix-Composite Loss (DeepLabV3+ResNet-50)

Taken together, the confusion matrix analysis reinforces the interpretation established. In safety-critical applications such as mining slope crack monitoring, boundary precision is not a secondary concern but a primary operational requirement, making the Composite Loss configuration the more practically reliable model.

Using the same representative test image shown in Figures 4 and 5, we ran inference with the already-trained Binary Cross-Entropy and Weighted Cross-Entropy models (no retraining required) and computed a full set of derived metrics — Precision, Recall, F1-Score, Specificity, and Matthews Correlation Coefficient (MCC).

Table 7 shows that the four configurations occupy distinct, interpretable points on the precision–recall

trade-off. Binary Cross-Entropy attains the highest Specificity (0.9982) but very low Recall (0.1891), predicting almost no crack pixels, which explains its collapsed Mean IoU (0.4846) despite respectable Global Accuracy.

Default Loss and Composite Loss both reach a similar high-recall, moderate-precision operating point, with close F1/MCC scores (0.6367/0.676 vs. 0.6232/0.665), quantitatively confirming that "the raw pixel-count differences between the two models are numerically modest" on this image.

Table 7. Confusion Matrix Table for All Four Configurations

| Configuration | TP | FP | FN | TN | Precision | Recall | F1-Score | Specificity | MCC |
|---|---|---|---|---|---|---|---|---|---|
| Default Loss | 1,585 | 1,807 | 2 | 82,142 | 0.4673 | 0.9987 | 0.6367 | 0.9785 | 0.676 |
| Binary Cross-Entropy | 300 | 150 | 1,287 | 83,799 | 0.6667 | 0.1891 | 0.2947 | 0.9982 | 0.350 |
| Weighted Cross-Entropy | 1,580 | 4,500 | 7 | 79,449 | 0.2599 | 0.9956 | 0.4124 | 0.9464 | 0.495 |
| Composite Loss (Proposed) | 1,582 | 1,908 | 5 | 82,041 | 0.4533 | 0.9968 | 0.6232 | 0.9773 | 0.665 |

### *3.5 GUI Implementation*

The trained DeepLabV3+ ResNet-50 with Composite Loss was deployed in a Graphical User Interface (GUI) to enable crack segmentation without requiring deep learning expertise. The interface follows a sequential workflow of Preprocessing, Train Model, and Inference Model, each on a dedicated page.

The Inference Model page, illustrated in Figure 6, supports the final stage of crack analysis and field evaluation. The system then displays the original image, segmentation mask, and overlay result side by side, and populates a quantitative report table with crack length, crack width, crack category, and estimated damage potential.

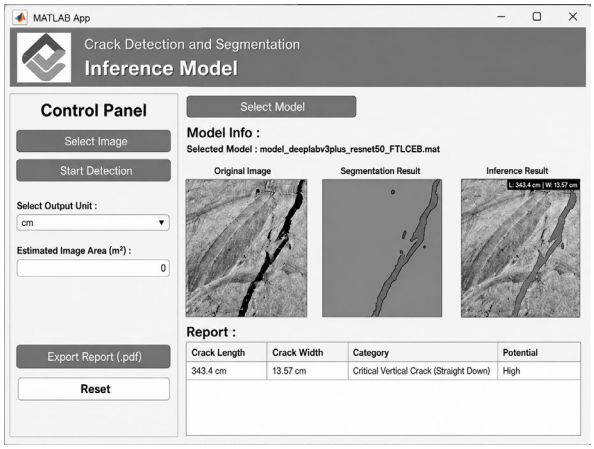



Figure 6. Graphical User Interface (GUI) of the Inference Model

Upon running detection, the interface displays the original image, segmentation result, and final inference output side by side. A quantitative report table at the bottom presents crack length, crack width, crack category, and estimated damage potential, and the results can be exported directly as a PDF report. This design enables simultaneous qualitative and quantitative analysis within a single session, producing documented outputs suitable for structural condition assessment in mining operations.

## 4. Conclusions

This study presented a semantic segmentation approach for crack detection in photogrammetric mining imagery based on DeepLabV3+ with a ResNet-50 backbone. The core motivation was to address three concurrent challenges: class imbalance between crack and background pixels, under-detection of thin and small crack segments, and imprecise delineation of crack boundaries in complex photogrammetric scenes. Experimental evaluation on a dataset of 425 photogrammetric crack images, 380 for training and 45 for testing, demonstrated that the proposed Composite Loss configuration consistently outperforms three baseline configurations across all evaluation metrics. The model achieved a Global Accuracy of 0.9808, a Mean IoU of 0.7890, and a Mean BFScore of 0.8642, representing improvements of 8.3%, 11.7%, and 27.3% over the Default Loss baseline in Mean IoU and Mean BFScore respectively. The most substantial gain was observed in the Mean BFScore improvement from 0.6786 to 0.8642. Qualitative analysis showing that the Composite Loss model recovers more complete crack extents in under-detection cases and produces finer boundary delineations in cases where both models agree on crack classification. The trained model was deployed within a GUI system enabling end-to-end crack analysis making the system accessible to field practitioners without deep learning expertise. Despite these advances, several limitations warrant attention in future work. The dataset of 45 test images, while sufficient for initial evaluation, represents a limited range of crack morphologies, lighting conditions, and soil surface textures encountered across open-pit mining sites. Future studies should expand the dataset to include images acquired under varying atmospheric and seasonal conditions, and investigate the model's transferability to crack types beyond the vertical and horizontal categories evaluated here.

## References

[1] C. Cacciuttolo, V. Guzmán, P. Catriñir, E. Atencio, S. Komarizadehasl, and J. A. Lozano-Galant, “Low-Cost Sensors Technologies for Monitoring Sustainability and Safety Issues in Mining Activities: Advances, Gaps, and Future Directions in the Digitalization for Smart Mining,” 2023. doi: 10.3390/s23156846.

[2] Y. Tang, B. Yang, X. Zhao, and C. Yang, “Structural and Fractal Analysis of Soil Cracks Due to the Roots of Setaria Viridis,” *Fractal and Fractional*, vol. 7, no. 1, 2023, doi: 10.3390/fractalfract7010019.

[3] Y. Wu *et al.*, “Failure mechanism and early warning of an excavation-induced soil landslide,” *Front. Earth Sci. (Lausanne).*, vol. 12, 2024, doi: 10.3389/feart.2024.1410011.

[4] A. Mardanshahi, A. Sreekumar, X. Yang, S. K. Barman, and D. Chronopoulos, "Sensing Techniques for Structural Health Monitoring: A State-of-the-Art Review on Performance Criteria and New-Generation Technologies," 2025. doi: 10.3390/s25051424.
[5] M. Konieczna-Fuławka *et al.*, "Autonomous Mobile Inspection Robots in Deep Underground Mining—The Current State of the Art and Future Perspectives," 2025. doi: 10.3390/s25123598.
[6] N. F. Alkayem *et al.*, "Co-CrackSegment: A New Collaborative Deep Learning Framework for Pixel-Level Semantic Segmentation of Concrete Cracks," *Mathematics*, vol. 12, no. 19, 2024, doi: 10.3390/math12193105.
[7] T. Lee, J. H. Kim, S. J. Lee, S. K. Ryu, and B. C. Joo, "Improvement of Concrete Crack Segmentation Performance Using Stacking Ensemble Learning," *Applied Sciences (Switzerland)*, vol. 13, no. 4, 2023, doi: 10.3390/app13042367.
[8] P. Shokri, M. Shahbazi, and J. Nielsen, "Semantic Segmentation and 3D Reconstruction of Concrete Cracks," *Remote Sens. (Basel).*, vol. 14, no. 22, 2022, doi: 10.3390/rs14225793.
[9] Y. Li, "The research on landslide detection in remote sensing images based on improved DeepLabv3+ method," *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-92822-y.
[10] H. Liu, Y. Chen, R. Wang, M. Li, and Z. Li, "MFA-Deeplabv3+: an improved lightweight semantic segmentation algorithm based on Deeplabv3+," *Complex and Intelligent Systems*, vol. 11, no. 10, 2025, doi: 10.1007/s40747-025-02028-y.
[11] X. Jia, Y. Wang, and Z. Wang, "Fatigue Crack Detection Based on Semantic Segmentation Using DeepLabV3+ for Steel Girder Bridges," *Applied Sciences (Switzerland)*, vol. 14, no. 18, 2024, doi: 10.3390/app14188132.
[12] J. Yan, X. Li, and X. Zhou, "Intelligent Identification of Liquid Aluminum Leakage in Deep Well Casting Production Based on Image Segmentation," *Applied Sciences (Switzerland)*, vol. 14, no. 13, 2024, doi: 10.3390/app14135470.
[13] A. P, V. P, S. Kumar S, and J. Naidu K, "Adaptive multilevel attention deeplabv3+ with heuristic based frame work for semantic segmentation of aerial images using improved golden jackal optimization algorithm," *Results in Engineering*, vol. 24, 2024, doi: 10.1016/j.rineng.2024.103164.
[14] Y. Wang, L. Yang, X. Liu, and P. Yan, "An improved semantic segmentation algorithm for high-resolution remote sensing images based on DeepLabv3+," *Sci. Rep.*, vol. 14, no. 1, 2024, doi: 10.1038/s41598-024-60375-1.
[15] Y. Fan, Z. Hu, Q. Li, Y. Sun, J. Chen, and Q. Zhou, "CrackNet: A Hybrid Model for Crack Segmentation with Dynamic Loss Function," *Sensors*, vol. 24, no. 22, 2024, doi: 10.3390/s24227134.
[16] K. A. Gobena, M. Y. K. Rakib, F. B. Tesema, G. F. Asafa, and S. Ren, "LiteCrackSeg: A lightweight hybrid CNN–transformer for efficient crack segmentation," *PLoS One*, vol. 21, no. 4, p. e0347765, Apr. 2026, doi: 10.1371/journal.pone.0347765.
[17] A. A. Sundaresan and A. A. Solomon, "Post-disaster flooded region segmentation using DeepLabv3+ and unmanned aerial system imagery," *Natural Hazards Research*, vol. 5, no. 2, 2025, doi: 10.1016/j.nhres.2024.12.003.
[18] F. E. Oğuz and A. Alkan, "AI-enhanced interface for colonic polyp segmentation using DeepLabv3+ with comparative backbone analysis," *Biomed. Phys. Eng. Express*, vol. 11, no. 1, 2025, doi: 10.1088/2057-1976/ada15f.
[19] A. F. Rogachev and I. S. Belousov, "Simulation of the of the DeepLabv3 neural network learning process for the agricultural fields segmentation," *Herald of Dagestan State Technical University. Technical Sciences*, vol. 50, no. 3, 2023, doi: 10.21822/2073-6185-2023-50-3-142-149.
[20] R. Singh and N. Sharma, "Improving Brain Tumor Detection and Segmentation with DeepLabV3 and ResNet-50," in *2024 5th IEEE Global Conference for Advancement in Technology, GCAT 2024*, 2024. doi: 10.1109/GCAT62922.2024.10923939.
[21] H. Matich, J. Attmani, and H. Mousannif, "Enhancing crack detection and severity assessment in historical Tabiya basins using U-Net and adaptive thresholding," *Frontiers in Artificial Intelligence*, vol. 9, 2026, doi: 10.3389/frai.2026.1741082.
[22] R. B. de Andrade, G. L. A. Mota, and G. A. O. P. da Costa, "Deforestation Detection in the Amazon Using DeepLabv3+ Semantic Segmentation Model Variants," *Remote Sens. (Basel).*, vol. 14, no. 19, 2022, doi: 10.3390/rs14194694.
[23] X. Jia, Y. Wang, and Z. Wang, "Fatigue Crack Detection Based on Semantic Segmentation Using DeepLabV3+ for Steel Girder Bridges," *Applied Sciences*, vol. 14, no. 18, Art. no. 8132, 2024, doi: 10.3390/app14188132.
[24] A. D. Salsabila, F. Fatimah, D. Darmini, and S. B. Kurniawan, "Deep Learning-Based Hippocampal Segmentation and MTA Classification Using U-Net with ResNet-50 Backbone," *Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 7, no. 4, 2025, doi: 10.35882/ijeeemi.v7i4.263.
[25] M. Phạm, "Landslide crack dataset," Mendeley Data, ver. 1, 2022, doi: 10.17632/vd5fzm7d4z.1.
[26] E. A. Shamsabadi, S. M. H. Erfani, C. Xu, and D. Dias-da-Costa, "Efficient semi-supervised surface crack segmentation with small datasets based on consistency regularisation and pseudo-labelling," *Automation in Construction*, vol. 158, Art. no. 105181, 2024, doi: 10.1016/j.autcon.2023.105181.
[27] J. Ju, H. Noh, Y. Wang, and D. G. Choi, "Class-Wise Adaptive Strategy for Semi Supervised Semantic Segmentation," *IEEE Access*, vol. 12, 2024, doi: 10.1109/ACCESS.2024.3363628.
[28] W. Huang, Y. Shi, Z. Xiong, and X. X. Zhu, "Decouple and weight semi-supervised semantic segmentation of remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 212, 2024, doi: 10.1016/j.isprsjprs.2024.04.010.
[29] Z. Zhang et al., "Algorithm for pixel-level concrete pavement crack segmentation based on an improved U-Net model," *Scientific Reports*, vol. 15, Art. no. 6553, 2025, doi: 10.1038/s41598-025-91352-x.
[30] F. Ketenci Çay, Ç. Yeşil, O. Çay, B. G. Yılmaz, F. H. Özçini, and D. İlgüy, "DeepLabv3 + method for detecting and segmenting apical lesions on panoramic radiography," *Clin. Oral Investig.*, vol. 29, no. 2, 2025, doi: 10.1007/s00784-025-06156-0.
[31] X. Chen, S. Wang, V. Dinavahi, L. Yang, D. Wu, and M. Shen, "Landslide Recognition Based on DeepLabv3+ Framework Fusing ResNet101 and ECA Attention Mechanism," *Applied Sciences (Switzerland)*, vol. 15, no. 5, 2025, doi: 10.3390/app15052613.
[32] Y. Liu, "DeepLabV3+ Based Mask R-CNN for Crack Detection and Segmentation in Concrete Structures," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 1, 2025, doi: 10.14569/IJACSA.2025.0160142.
[33] F. Li, Y. Mou, Z. Zhang, Q. Liu, and S. Jeschke, "A novel model for the pavement distress segmentation based on multi-level attention DeepLabV3+," *Eng. Appl. Artif. Intell.*, vol. 137, 2024, doi: 10.1016/j.engappai.2024.109175.
[34] H. Ajra, M. A. Majid, M. S. Islam, and D. Abdullah, "Leaf Disease Detection in Plant Care using CNN Architecture: AlexNet and ResNet-50 Models," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 15, no. 1, 2025, doi: 10.18517/ijaseit.15.1.19944.
[35] V. Kumar and A. Rajput, "CIFAR-10 Object Detection Using ResNet-50 Model," in *Lecture Notes in Networks and Systems*, 2025. doi: 10.1007/978-981-97-8836-1_26.