TELKOMNIKA Telecommunication, Computing, Electronics and Control
Vol. 23, No. 6, December 2025, pp. 1611∼1625
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v23i6.26829

Mixed attention mechanism on ResNet-DeepLabV3+ for paddy field segmentation

Alya Khairunnisa Rizkita[1], Masagus Muhammad Luthfi Ramadhan[1], Yohanes Fridolin Hestrio[1,2], Muhammad Hannan Hunafa[1], Danang Surya Candra[2], Wisnu Jatmiko[1]
[1]Department of Computer Science, Faculty of Computer Science, University of Indonesia, Depok, Indonesia
[2]Research Center for Geoinformatics, Electronics and Informatics Research Organization, National Research and Innovation Agency, Bogor, Indonesia

Article Info

Article history:

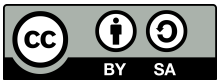
Received Dec 7, 2024
Revised Aug 27, 2025
Accepted Sep 10, 2025

Keywords:

Attention mechanism
DeepLabV3+
Remote sensing
Residual network
Semantic segmentation

ABSTRACT

Rice cultivation monitoring is crucial for Indonesia, where paddy field areas declined by 2.45% according to the Central Bureau of Statistics due to land function changes and shifting crop preferences. Regular monitoring of paddy field distribution is essential for understanding agricultural land utilization by farmers and landowners. Satellite imagery has become increasingly common for agricultural land observation, but traditional neural networks alone provide insufficient segmentation accuracy. This study proposes an enhanced deep learning architecture combining residual network (ResNet)-DeepLabV3+ with coordinate attention (CA) and spatial group-wise enhancement (SGE) modules. The attention mechanisms establish direct connections between context vectors and inputs, enabling the model to prioritize relevant spatial and spectral features for precise paddy field identification. The CA module enhances spectral feature discrimination, whereas the SGE improves spatial characteristic representation. The experimental results demonstrate superior performance over the baseline methods, achieving intersection over union (IoU) of 0.85, dice coefficient of 0.89, and accuracy of 0.95. The proposed mixed attention mechanism significantly improves the accuracy and efficiency of automatic crop area identification from satellite imagery.



Corresponding Author:

Alya Khairunnisa Rizkita
Department of Computer Science, Faculty of Computer Science, University of Indonesia
Depok, 16424, Indonesia
Email: alya.khairunnisa21@ui.ac.id

1. INTRODUCTION

Rice is a prominent agricultural product globally, and it thrives in expansive, unobstructed regions. In Indonesia, rice is the predominant commodity. Based on statistics from the Central Bureau of Statistics in Indonesia, there was a 2.45% decline in rice cultivation from 2022 to 2023 [1], [2]. This reduction in land area is due to changes in land functions and preferences for planted crop commodities. The decreased price of rice relative to other crops and the less-than-ideal paddy field conditions for rice growth are the main causes of the change in commodities and land usage [3], [4]. Therefore, it is important for farmers and owners of paddy fields to know the usefulness of the distribution of existing paddy fields by conducting regular monitoring.

Remote sensing imagery provides a precise and geographically accurate depiction of the land that the device has captured [5]. Remote sensing imagery typically requires geographic information about agricultural

land to accurately monitor crop conditions at specific locations and utilize time efficiently. Additionally, this information aids in monitoring the use of agricultural land, which is rapidly changing its function. Only large and open land crops, such as rice, corn, onions, and oil palm, are suitable for remote sensing monitoring on agricultural land. Consequently, it is possible to map paddy fields using remote sensing technology to monitor crop growth, manage paddy field use, and estimate crop yields [6]. Extracting information about agricultural land can employ methods such as segmentation or object detection to further process remote sensing images [7]. These approaches have been the focus of recent research. In line with the development of remote sensing image processing techniques, image segmentation methods have become one of the techniques relied upon to produce more detailed information from an image. Image segmentation is a technique that classifies each pixel in an image into semantic classes [8]. However, automatically identifying rice fields from satellite images is still very difficult. Current deep learning methods have major problems, such as difficulty in accurately tracing complex and irregular field boundaries, difficulty in distinguishing rice fields from other similar-looking features, such as water bodies, wetlands, or other crops, satellites do not perform well when fields have different sizes, shapes, and orientations in the same image, and satellites do not work reliably under different weather conditions, seasons, or rice growth stages. These problems lead to poor accuracy in identifying rice fields, resulting in unreliable farm monitoring and incorrect crop yield predictions.

In recent years, the development of segmentation methods has continued to grow. Agriculture widely applies neural network approaches, such as crop type segmentation [9]-[12], crop disease detection [13], [14], and farmland mapping [15]. Shelhamer *et al.* [16] first proposed the segmentation neural network architecture, fully convolutional networks (FCN). U-shaped convolutional neural network (U-Net) [17] is a highly improved FCN version that integrates up-sampling, down-sampling, skip connections, and fully convolutional applications. However, U-Net is more symmetric than FCN. Any down-sampling and up-sampling stage of U-Net is endowed with a skip connection at every level. The U-Net architecture is highly applied in applications, and segmentation problems are being upgraded as studies emerge.

Research conducted on agricultural land mapping with segmentation methods was conducted by Zhang *et al.* [18] used refined pyramid scene parsing network (PSPNet) on polarimetric synthetic aperture radar (PolSAR) data. This study proposes a modified PSPNet with the addition of the polarimetric CA module. This module captures polarimetric features from PolSAR data, allowing PSPNet to distinguish existing plant species. The results of this study prove that refined PSPNet provides excellent segmentation results on PolSAR data, especially in terms of accuracy in mapping sharp contours and separating agricultural areas.

Li *et al.* [19] developed U-Net with the addition of a linear attention mechanism (LAM) application to segment remote sensing images. The U-Net structure places LAM on each skip connection. This research led to the development of MAResU-Net, which generally outperforms U-Net in general. Mahmud *et al.* [20] developed DeepLabV3+ by adding a spatial attention (SA) mechanism for road segmentation. The findings of this study demonstrate that adding a SA mechanism increases segmentation accuracy by 1.56% compared to the usual model. Previous research has demonstrated that the integration of attention modules can enhance segmentation accuracy.

Despite these advances, existing approaches exhibit several critical limitations: traditional U-Net architectures struggle with multi-scale feature extraction required for varying paddy field sizes, PSPNet-based methods suffer from high computational complexity limiting their practical deployment, FCN-based approaches lack sufficient contextual information for complex agricultural landscapes, and current attention mechanisms often fail to simultaneously capture both spatial relationships and channel dependencies crucial for distinguishing paddy fields from spectrally similar features. Most existing methods are evaluated on general segmentation datasets rather than domain-specific agricultural scenarios, limiting their applicability to real-world paddy field monitoring tasks.

Previous research has demonstrated that the integration of attention modules can enhance segmentation accuracy. Attention modules can distinguish between important and unimportant image elements. The object boundaries are more precise, and internal pixels are more uniform with the integration of attention modules [18]-[20]. The convolutional block attention module (CBAM) is one of the attention modules designed to enhance the representational power of CNN feature maps. CBAM applies channel attention (CA) and SA. CA focuses on refining the importance of feature channels, allowing the models to know what they must see in the feature. Meanwhile, the SA module focuses on highlighting the important regions within the feature maps. The spatial group-wise enhance (SGE) attention module is another attention module that improves CNN feature maps. Unlike CBAM, which has two modules, SGE focuses on spatial information using a group-wise

strategy [21], [22].

Previous research by Li *et al.* [23] proposes an improved U-Net model combining a multi-axis vision transformer (MaxViT) encoder with a CBAM decoder for automated crop-weed segmentation in precision agriculture. This study addresses the limitations of uniform pesticide spraying and existing segmentation models that struggle with agricultural images' complex backgrounds and multi-scale targets. Using MaxViT's dual attention mechanisms for local and global feature capture alongside CBAM's channel and SA for boundary enhancement, the model achieved 84.28% mean intersection over union (mIoU) and 88.59% mean pixel accuracy (mPA) on sugar beet datasets, improving 3.08% and 3.15% over baseline U-Net while reducing parameters from 43.93 M to 22.08 M and maintaining 0.0559 s inference time.

Developing effective rice field mapping methods faces several major challenges, including the need for computationally efficient solutions that work in real-time for farm monitoring, high data variability from seasonal changes and different growth stages across diverse geographical locations, limited availability of high-quality annotated datasets, requirements for methods that work across different sensor types and imaging conditions, and the need for practical deployment in existing agricultural systems with minimal infrastructure. To address these challenges, this research aims to develop an enhanced deep learning architecture that combines a residual network (ResNet) encoder and DeepLabV3+ decoder to leverage both detailed features and multi-scale contextual information, investigate hybrid attention mechanisms (CBAM and SGE) for capturing spatial and channel dependencies that are crucial for accurate boundary detection, and systematically evaluate performance against existing state-of-the-art methods using comprehensive metrics, such as IoU, dice coefficient, and accuracy. Finally, this study provides a practical, computationally efficient solution for automated rice field monitoring from remote sensing imagery that can be readily deployed in operational agricultural systems.

Therefore, this research focuses on mapping agricultural land use, especially paddy fields, using the ResNet-DeepLabV3+ neural network segmentation method. The segmentation method will also incorporate CBAM and SGE attention modules to enhance ResNet-DeepLabV3+'s ability to segment paddy fields. In summary, this article offers the following contributions:

- We proposed an enhanced deep learning model, ResNet DeepLabV3+ with attention, for segmenting paddy fields from remote sensing images. The proposed model comprises ResNet-34 as the encoder, augmented with an attention module to improve the model's segmentation capabilities, and DeepLabV3+ as the decoder.
- By incorporating the SGE technique, we enhance the CBAM by replacing its SA mechanism. This modification guarantees a more accurate depiction of the spatial characteristics throughout the image.
- We are comparing the performance of the proposed method is compared with other segmentation methods.

This paper is organized into several sections, the first of which covers the introduction to the research. The second section discusses related works. The third section describes the experimental setup. The fourth section discusses the proposed method. The results are discussed in the fifth section, and the last section discusses the research conclusions.

2. RELATED WORK

2.1. Residual network

ResNet was first introduced in 2016 by He *et al.* [24] as an innovation that introduced the concept of residual learning on deep neural networks that are more than 100 layers deep. He *et al.* [24] proposed this concept to address optimization issues in deep neural networks, which frequently face performance degradation difficulties. The basic idea of residual learning is that the network learns the residuals of an existing function. If the function to be learned is denoted as H(x) and the learning network is $F(x) = H(x) - x$, then the formula becomes $F(x) = H(x) + x$.

ResNet uses a residual block with many layers, each of which learns from the residuals of a previous one. A skip connection is used in the network to connect the input and output. There are also skip connections that perform identity mapping, where the outputs of layers get summed at a later point to reduce network complexity and help in optimizing it. ResNet development has shown that residual learning can help to successfully optimize and improve the accuracy of deep neural networks [24].

2.2. DeepLabV3

Chen *et al.* [25] re-implemented their prior model, DeepLab, to the newer version: DeepLabV3. The atrous convolution, used in the DeepLabV3 model, enables a network to globally upscale its field of

view without adding new learned parameters. Atrous convolution to highlight spatial information at a large receptive field and maintain crisp resolution an atrous convolution at multiple parallel levels forms an atrous spatial pyramid pooling (ASPP). ASPP means that the method can effectively extract information from small parameter spaces and has multiple scales using only images. DeepLabV3 combines many new ideas, such as the use of ASPP instead to perform multiscale feature learning globally that will help the deep network be aware of more context and make it capture richer features from multiple scales of detail; add global image-level features directly against the last convolution over grouping (ICOG) convolutional layer for segmenting boundaries; replace SoftMax with sigmoid binary cross-entropy loss while employing the dice coefficient as evaluation metrics; and, importantly, employ batch norm in the ASP module, which has proven to enable a greatly high stability training effectiveness [25].

2.3. Attention mechanism

The attention mechanism creates a direct link between the context vector and the input, allowing the model to focus on the most important parts by assigning scores to each input element based on its relevance. This is especially helpful when working with large datasets where not all information is equally useful, such as in language translation or image recognition where only specific parts matter for accurate results. By using attention, models can learn more efficiently by concentrating on key features and ignoring less important details, which improves accuracy while saving computational resources [21], [22].

2.3.1. Convolutional block attention module

CBAM is an attention module that combines CA and SA to pay more profound attention to relevant features at the channel and spatial levels. CBAM starts by applying CA first to adjust the weight value of each channel, followed by SA, which adjusts the weight value spatially [21]. The combination of CA and SA is shown in (1) and (2). In (1), the initial feature map (F) is performed element-wise multiplication $\otimes$ with the one dimension of the feature map generated from CA $(M_c(F))$. The resulting feature map $(F^{'})$ is also subjected to element-wise multiplication with the two-dimensional feature map generated by SA$(M_s(F))$.

$$F^{'} = M_c(F) \otimes F \tag{1}$$

$$F^{''} = M_s(F) \otimes F^{'} \tag{2}$$

CA is an attention mechanism that assigns varying weights to each channel according to their relative importance. The goal of assigning this weight is to determine the most pertinent channel [21]. The formulation of CA using (3) and (4). Feature maps are assigned average pooling $(AvgP)$ and maximum pooling $(MaxP)$ at each level of the multilayer perceptron (MLP). The featured result $AvgP$ is symbolized as F^c_{avg} and $MaxP$ is symbolized as F^c_{max}. Both feature results are added and given a sigmoid function σ. The sigmoid function determines the most relevant part of the feature based on the $AvgP$ and $MaxP$. There are 2 MLP layers used, so the weight of each layer is denoted as W_1 and W_0.

$$M_c(F) = \sigma(MLP(AvgP(F)) + MLP(MaxP(F))) \tag{3}$$

$$M_c(F) = \sigma(W_1(W_0(F^c_{avg})) + W_1(W_0(F^c_{max}))) \tag{4}$$

SA, in contrast to CA, looks at each channel's relevance. SA is an attention mechanism that focuses on giving weight to the spatial location of features based on how important the location is. This weighting is intended to identify the feature's most relevant location [21]. The formulation of SA using (5) and (6) Feature maps (F) are assigned average pooling $(AvgP)$ and maximum pooling $(MaxP)$ and the resulting two-dimensional feature maps are F^s_{avg} and F^s_{max}. Both results are concatenated and convoluted with the convolution operation (f) with a filter size of 7×7. To provide a feature map that represents the most relevant spatial values, the result of the convolution is given a sigmoid function.

$$M_s(F) = \sigma\left(f^{7\times7}\left(\begin{bmatrix} AvgP(F) \\ MaxP(F) \end{bmatrix} \right) \right) \tag{5}$$

$$M_s(F) = \sigma\left(f^{7\times7}\left(\begin{bmatrix} F^s_{avg} \\ F^s_{max} \end{bmatrix} \right) \right) \tag{6}$$

2.3.2. Spatial group-wise enhanced attention

SGE is an attention module that performs feature grouping at the spatial and channel levels and applies an attention mechanism to determine the relevance level of sub-features based on their spatial location in each formed group. SGE generates a coefficient value (c_i) for each feature based on the similarity of global (g) and local features (x_i), allowing it to focus on the right semantic area. The division of features into several groups allows the processing of features on a smaller scale to obtain the desired details [22]. We compute the formulation of SGE using (7).

$$c_i = g \cdot x_i \tag{7}$$

3. METHODOLOGY

3.1. Data collection

This research utilized remote sensing image data collected to support the field survey implementation of the National Research and Innovation Agency (BRIN) in developing a method to harmonize high-resolution satellite images using artificial intelligence on monitoring and paddy field mapping. Remote mapping activities occurred from 21 to 25 August 2023 in Yogyakarta Province. Thirteen image capture points coincided with rice fields in Sleman Regency, Bantul Regency, and Kulon Progo Regency. The remote sensing tool used to take images of rice fields is the Da-Jiang Innovations (DJI) Phantom 4 multispectral camera, specifically designed for agriculture and environmental monitoring. The camera model used in the drone is FC6360 with five color ranges, including blue, green, red, red edge, and near-infrared (NIR). The average height of the drone in taking images at all capture points is 100 m, resulting in images with a high resolution of 1600 pixels, a width of 1300 pixels, and a ground resolution of 5 cm/pixel.

3.2. Data creation

Before data from BRIN field survey activities can serve as a foundation for a machine learning model, it must undergo several stages of data processing. The stages carried out to prepare the image data of paddy fields include digitization or annotation of paddy field images and the formation of paddy field image datasets. Land designation resulted in five types of paddy field groupings, namely: background, first-phase paddy plants (vegetative), second-phase paddy plants (reproductive), third-phase paddy plants (ready to harvest), and non-paddy fields. Direct identification of the differences in each paddy field's land use is possible. However, some plants have characteristics similar to those of paddy plants, and the paddy phase also sometimes has the same characteristics. The output of this digitization process includes an image file and a shape file, which serve as the initial mask form for segmentation. All existing drone images are digitized to facilitate easier image selection. An example of digitized data is shown in Figure 1. After digitization, each digitized shape file is divided into parts based on the color range from 0 to 255, forming a mask file. Segmentation activities use the mask file as the ground truth of an image to map the class into a pixel-by-pixel image. Once the entire image, the next step was processing the image and shape file were processed to create a label mask and uniformize it to a size of 512×512 pixels. In summary, the dataset used in this study consisted of paddy field images with a mask label of 512×512 pixels. This study used a total of 546 paddy field images, grouped into 5 classes.

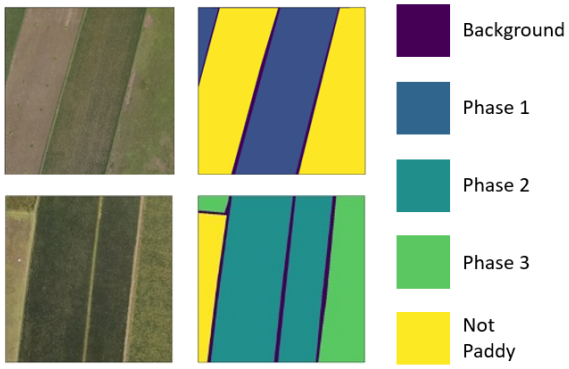


Figure 1. Example of the data

3.3. Data preparation

The research experiment used a dataset of 546 images and their labels. The size of each image and label is 512×512 pixels. The dataset labels contain five classes: background, first-phase rice plant, second-phase rice plant, third-phase rice plant, and nonrice paddy field. We split the dataset into 3 subsets: 70% for training data, 20% for validation data, and 10% for test data. Training data are used to model to learn the data, validation data are used to aid in building the model fast by recognizing at each training where your validation value increases improved or not, and test data are used to evaluate how our model performs segmentation.

3.4. Proposed architecture

In this study, we proposed a deep learning development model, ResNet-DeepLabV3+ with attention, for segmenting paddy fields from remote sensing images. The basic idea behind this model development is to create a segmentation model by combining ResNet-34 as an encoder and DeepLabV3+ as a decoder and placing an attention module on the encoder to improve the encoder's ability to extract features. Figure 2 shows a perspective of the proposed method.

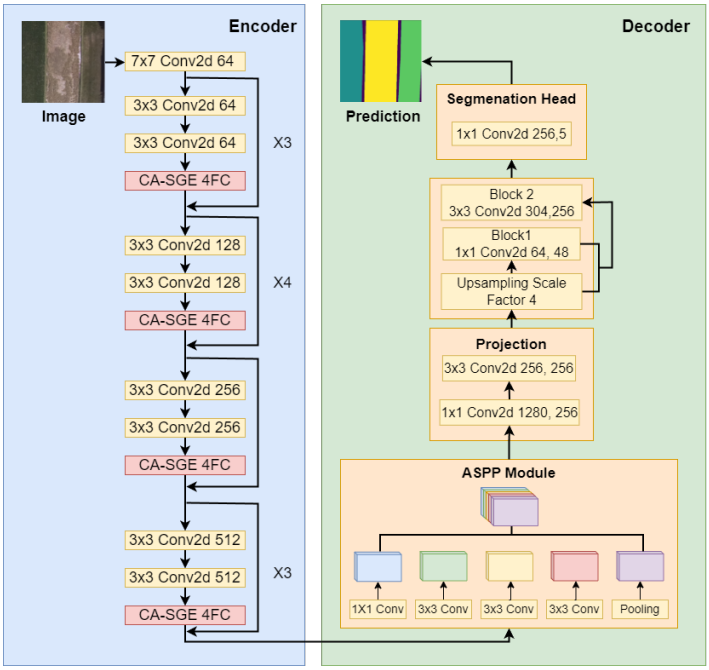


Figure 2. Proposed method

3.4.1. Encoder: ResNet-34 with attention

The encoder of ResNet-34 begins with a 7×7 kernel and converts an incoming image input (3 channels) to be processed in the conv-layer, mapping it into a space with 64 channels. The first layer also performs batch normalization, rectified linear unit (ReLU), and max pooling to reduce the spatial dimension. Each ResNet layer comprises two configurations of convolutional layers with 3×3 kernels, followed by batch normalization and attention addition. The first ResNet layer has an output of 64 channels; the second layer reduces the spatial dimension by increasing the channel size to 128, and the third layer reduces the dimension to 256 channels. The last layer provides an output of 512 channels. Each ResNet layer receives attention from the CA and SGE after the convolution layer has finished. Figure 2 illustrates the combination of CA and SGE. Initially, we combined the application of CA with SA to form CBAM, which increased attention at the channel and spatial levels. We also integrated CA with SGE. In addition, we added a MLP to the depth novel contribution.

CA-SGE attention integration: our key innovation lies in strategically combining CA [21] with SGE [22] mechanisms. Unlike the standard CBAM [21] which uses CA and SA sequentially, we replace the SA component with SGE to create a more effective CA-SGE hybrid. CA-SGE attention integration illustrated in Figure 3. CA enhancement: the standard CA mechanism was modified by incorporating a 4-layer MLP (compared to the default 2-layer) to enable more sophisticated channel relationship modeling. The CA process

uses both adaptive average pooling (AAP) and adaptive max pooling (AMP) followed by the enhanced MLP and sigmoid activation, as illustrated in Figure 4. SGE integration: the SGE mechanism [22] receives the CA-enhanced features and applies group-wise spatial enhancement through: 1) feature grouping, 2) global average pooling (GAP), 3) normalization, 4) position-wise dot product (PWP), and 5) sigmoid activation, as shown in Figure 5. We anticipated the addition of MLP to enable the encoder to extract features from complex data with greater accuracy in CA; we focused on retrieving the values from each AAP and AMP to reduce the channel size to a single value based on the average and maximum channel values. Next, we used sigmoid activation to group the feature values into binary form (0, 1). Figure 4 illustrates the CA process.

For the SGE experiment, we multiplied the CA result by the initial feature before entering the SGE to produce a more focused feature value based on the most relevant channel. SGE receives feature input from CA and splits it into several groups, with the same processing for each group. Following a series of processing steps on each group, all groups were merged back into the resulting feature, which provides a more accurate representation of the feature value. The processing in SGE includes GAP, normalization, PWP, and sigmoid activation. The GAP function takes all feature values and converts them into a single value. The normalization function makes the input range independent. The PWP function finds the dot product between the existing feature vectors. The sigmoid activation function converts feature values into binary form. Figure 5 illustrates the SGE process.

Our key innovation lies in strategically combining CA [21] with SGE [22] mechanisms. Unlike the standard CBAM [21] which uses CA and SA sequentially, we replace the SA component with SGE to create a more effective CA-SGE hybrid. We modified the standard CA mechanism by incorporating a 4-layer MLP (compared with the default 2-layer) to enable more sophisticated channel relationship modeling. The CA process uses both AAP and AMP followed by the enhanced MLP and sigmoid activation, as illustrated in Figure 4. The SGE mechanism [22] receives the CA-enhanced features and applies group-wise spatial enhancement through: 1) feature grouping, 2) GAP, 3) normalization, 4) PWP, and 5) sigmoid activation, as shown in Figure 5.

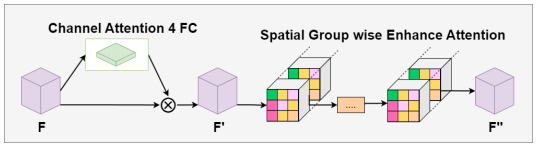


Figure 3. Channel-SGE attention

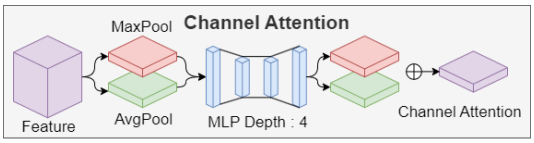


Figure 4. CA

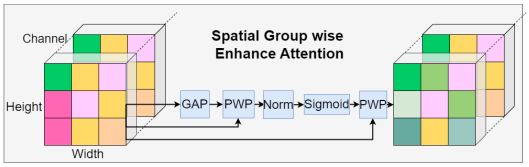


Figure 5. SGE attention

3.4.2. Decoder: DeepLabV3+

In this study, the decoder reconstructs the feature representation extracted by the encoder. The first part of DeepLabV3+ is the ASPP module, which has different dilation levels in parallel. This study uses the ASPP module, which consists of five parallel layers: a convolution layer without dilation, a convolution layer with 12, 24, and 36 dilations, and an AAP layer. In addition, it is good for pooling information from multiscale images. Then, the output from 5 layers will be concatenated, followed by their parallel projection, such that a tone feature representation of 256 channels. Then, the projection output will go through the feature concatenation in a layer block before the segmentation head process. The previous dense prediction layer consisted of a CNN and SoftMax to provide the final result for segmentation image forecast.

3.5. Training configuration

Data augmentation: We applied several augmentation techniques to the training dataset to enhance training data diversity and improve model robustness: 1) random horizontal and vertical flipping with probabil-

ity p=0.5, 2) brightness adjustment with linear scaling factor ranging from 0% to 10% of the original intensity, 3) contrast enhancement using nonlinear adjustment within 0% to 10% range, and 4) gamma correction with random gamma values between 0.8 and 1.2 (80% to 120% of original image gamma value).

Hardware and software setup: The proposed model was implemented using Python 3.8 with the PyTorch 2.0.1 framework and CUDA 11.8 for GPU acceleration. All experiments were conducted on the NVIDIA DGX-1 V100 GPU system at the Tokopedia-UI AI Center of Excellence, which provided sufficient computational resources for training the deep learning model training.

Training parameters: The network was optimized using Adam optimizer [26] with learning rate of 0.00008, which was chosen based on preliminary experiments for stable convergence. Training was conducted for 100 epochs with a batch size of 4 for training data, limited by GPU memory constraints. For the validation and testing phases, the batch size was set to 1 to ensure consistent evaluation across all samples. Regularization strategy: to mitigate overfitting and enhance model generalization, L2 regularization with weight decay coefficient of 0.0001 was applied to all trainable parameters during the optimization process.

3.6. Evaluation

We used IoU, the dice coefficient, accuracy, and the F1 score as metrics to evaluate our segmentation model. IoU is a popular evaluation metric for semantic segmentation is IoU. IoU calculates the overlapping area between predicted segmentation (A) and the ground truth (B). The formula for calculating IoU is shown in (8).

$$IoU = \frac{A \cap B}{A \cup B} \tag{8}$$

The accuracy (ACC) and F1 score are derived from the confusion matrix. A confusion matrix is a table to assess a classification algorithm's performance [27]. The confusion matrix provides an in-depth analysis of the actual vs. predicted classification. Several terms are in the confusion matrix. True positive (TP) is when the model successfully predicts that the actual positive class becomes a positive class. True negative (TN) is when the model predicts the actual negative class as a negative class. If the class is negative and predicted as positive, it is called a false positive (FP). Otherwise, if the class is positive and predicted as negative, it is called a false negative (FN).

Accuracy is the number of correct predictions divided by the total data. The formula for calculating accuracy can be seen in (9). The F1 score is the harmonic mean of precision and recall. In the semantic segmentation task, the F1 score is also called the dice coefficient score (DSC). Precision is the proportion of positive predictions, and recall is the proportion of actual positives that are correctly classified. F1 score is a better metric than accuracy when the testing data are imbalanced. F1 score, dice coefficient, and loss formula are shown in (10), (11), and (12).

$$Accuracy = \frac{TP + TN}{FP + FN + TP + TN} \tag{9}$$

$$F1Score = \left(\frac{TP}{TP + \left(\frac{(FP+FN)}{2} \right)} \right) \tag{10}$$

$$DiceCoefficient = \frac{2|A \cap B|}{|A| + |B|} \tag{11}$$

$$DiceLoss = 1 - DiceCoefficient \tag{12}$$

4. RESULT AND DISCUSSION

4.1. Overview of the key findings

The proposed ResNet-DeepLabV3+ with the CA-SGE attention mechanism achieved superior performance for paddy field segmentation, demonstrating significant improvements across all evaluation metrics. The key findings reveal that: 1) the CA-SGE combination provides synergistic benefits with IoU of 0.85, DSC of 0.91, and accuracy of 0.96, representing improvements of 3.7%, 4.6%, and 2.1% respectively over the baseline, 2) CA mechanisms are more effective than SA for paddy field segmentation tasks, and 3) the proposed method exhibits better training stability with reduced overfitting compared to the baseline model. From the training process of both models, it can be concluded that the proposed ResNet-DeepLabV3+ + CA-SGE is better because it shows more stable performance and lower potential for overfitting.

4.2. Analysis of training performance

The baseline ResNet-DeepLabV3+ model exhibited rapid performance improvement, with training and validation IoU scores reaching approximately 0.9 within the initial 40 epochs. However, subsequent training phases showed gradual improvement accompanied by increasing divergence between training and validation curves, suggesting that overfitting tendencies were emerging despite the gap remaining within acceptable limits. The proposed ResNet-DeepLabV3+ with CA-SGE attention demonstrated distinctly different training characteristics. The model initially exhibited underfitting behavior, where the validation IoU exceeded the training IoU during the early epochs. This pattern was reversed around epoch 30, after which the training IoU began to surpass the validation IoU while gradually improving. Notably, the training-validation gap remained consistently smaller than that of the baseline model throughout the training process. The reduced training-validation divergence observed in our CA-SGE model indicates superior generalization capability and enhanced training stability. This behavior suggests that the integrated attention mechanisms function as effective regularizers, mitigating overfitting risks while preserving strong model performance. The comparative analysis demonstrates that the proposed architecture achieves more robust training dynamics with improved convergence stability compared to the baseline approach. The experimental results demonstrate that the baseline method exhibits a large training-validation gap after epoch 40, indicating overfitting behavior. In contrast, the proposed CA-SGE method shows a smaller training-validation gap with better stability, effectively reducing overfitting issues. A summary of the training is shown in Figures 6 and 7.

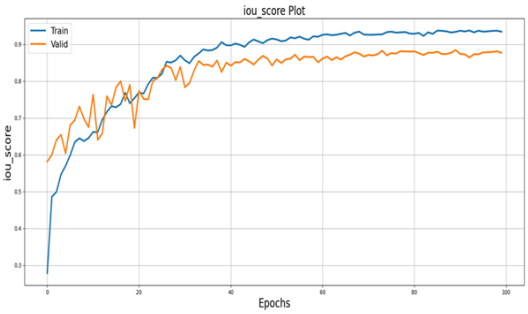


Figure 6. ResNet-DeepLabV3+

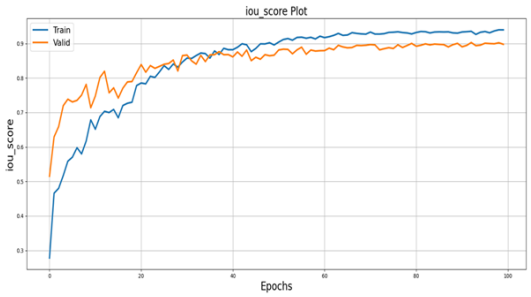


Figure 7. ResNet-DeepLabV3+ CA-SGE 4FC

4.3. Quantitative evaluation of performance

4.3.1. Baseline method comparison

We conducted a comprehensive comparative evaluation against established segmentation architectures to establish the effectiveness of our proposed approach. The baseline methods include ResNet-UNet, ResNet-FPN, and ResNet-DeepLabV3+, all implemented without attention mechanisms to ensure fair comparison with our attention-enhanced variant. Table 1 presents the quantitative performance comparison across three standard evaluation metrics: IoU, DSC, and pixel-wise accuracy. The proposed ResNet-DeepLabV3+ with CA-SGE attention consistently outperformed all baseline methods across every evaluation metric. The results reveal substantial and consistent improvements in performance across all evaluation dimensions. Specifically, the integration of CA and SGE attention mechanisms enhanced the baseline ResNet-DeepLabV3+ performance from IoU 0.82 to 0.85, DSC from 0.87 to 0.91, and accuracy from 0.94 to 0.96. These improvements demonstrate the effectiveness of our hybrid attention approach in capturing both channel-wise feature importance and spatial relationships, which are crucial for accurate paddy field segmentation. The superior IoU and DSC scores indicate enhanced boundary delineation accuracy and improved region overlap precision, which are critical factors for practical agricultural monitoring applications. Consistent performance gains across different architectural baselines validate the proposed attention mechanism's robustness and generalizability.

Table 1. Comparison of testing segmentation methods

Method	IoU	DSC	Acc
ResNet-UNet	0.79	0.86	0.94
ResNet-FPN	0.80	0.87	0.93
ResNet-DeepLabV3+	0.82	0.87	0.94
ResNet-DeepLabV3+CA SGE	0.85	0.91	0.96

4.4. Ablation study

4.4.1. Individual attention mechanism performance

An ablation study was conducted in the attention mechanism part. We compared the performance of CA, SA, CBAM, and SGE on the ResNet-DeepLabV3+ model. We also compared the depth of the MLP layers in the CA attention module. The addition of CA on ResNet-DeepLabV3+ slightly increases the performance. As shown in Table 2, the addition of CA increased all metrics by 0.01. Adding more MLP layers to 4 in CA did not have a significant impact, only increasing the DSC by 0.001, whereas the IoU and accuracy did not increase. The ablation results reveal several critical insights into the effectiveness of the attention mechanism in paddy field segmentation. CA demonstrated consistent performance improvements across all evaluation metrics, with IoU increasing from 0.82 to 0.83 when individually integrated. This improvement validates the effectiveness of channel-wise feature recalibration for agricultural segmentation tasks. Conversely, SA showed counterproductive effects, decreasing IoU performance from 0.82 to 0.81 while maintaining similar DSC scores and reducing accuracy. This degradation suggests that SA mechanisms may introduce unwanted noise rather than beneficial spatial focusing for paddy field boundary detection.

The CBAM evaluation, which combines both CA and SA components, yielded performance identical to CA alone across all metrics. This finding reinforces our observation that SA provides no additional benefit and potentially interferes with CA effectiveness, confirming the limited utility of SA in this specific application domain. SGE demonstrated a unique performance characteristic, achieving exceptionally high pixel-wise accuracy (0.98) while simultaneously showing decreased performance in IoU (0.81) and DSC (0.86) metrics. This pattern indicates that SGE excels at overall pixel classification accuracy but struggles with precise boundary delineation and region overlap precision, which are critical for accurate segmentation. The optimal performance was achieved through our proposed CA-SGE combination with 4-layer MLP depth, demonstrating superior balance across all evaluation metrics. This configuration achieved the highest IoU (0.85) and DSC (0.91) scores while maintaining competitive accuracy (0.96), representing improvements compared to the baseline.

Table 2. Comparison of attention modules

Module	MLP depth	IoU	DSC	Acc
-	-	0.82	0.87	0.94
SA	-	0.81	0.87	0.94
CA	2	0.83	0.88	0.95
CBAM	2	0.83	0.88	0.95
CA+SGE	2	0.83	0.89	0.95
SGE	-	0.81	0.86	0.98
CA	4	0.83	0.89	0.95
CBAM	4	0.82	0.88	0.95
CA+SGE	4	0.85	0.91	0.96

4.4.2. Multilayer perceptron depth analysis

Investigation of MLP depth configurations within the CA module revealed that the 4-layer architecture provides optimal performance without introducing significant computational overhead. Comparison between 2-layer and 4-layer MLP configurations showed that deeper networks yielded marginal improvements (+0.01 DSC) while maintaining computational efficiency. Further exploration of deeper MLP architectures (>4 layers) demonstrated diminishing returns with negligible performance gains, confirming the appropriateness of our 4-layer architectural choice for practical deployment scenarios.

4.4.3. Data augmentation impact

Table 3 displays the results of testing each data augmentation on the neural network segmentation method. The results demonstrate substantial and consistent performance improvements across all segmentation architectures when data augmentation is applied. ResNet-FPN showed the most dramatic improvement, with IoU increasing from 0.55 to 0.80, indicating that this architecture particularly benefits from enhanced data diversity. ResNet-UNet also exhibited significant improvement, with IoU rising from 0.64 to 0.79. The ResNet-DeepLabV3+ baseline showed more modest but meaningful improvements, increasing from 0.76 to 0.82. This smaller relative improvement suggests that the more sophisticated DeepLabV3+ architecture with its atrous convolutions and multi-scale processing already possesses some inherent robustness to variations in input data.

Table 3. Comparison of data augmentation

Method	IoU	DSC	Acc	IoU Aug	DSC Aug	Acc Aug
ResNet-Unet	0.64	0.74	0.89	0.79	0.86	0.94
ResNet-FPN	0.55	0.62	0.79	0.80	0.87	0.93
ResNet-DeepLabV3+	0.76	0.84	0.93	0.82	0.87	0.94

4.5. Visual segmentation results

To complement our quantitative evaluation, we examine the visual segmentation performance across diverse test scenarios. Figure 8 presents representative segmentation results from five carefully selected test samples, showcasing comparative performance between baseline methods and our proposed CA-SGE approach across various challenging conditions, including different lighting conditions, crop growth stages, and field complexities. Our CA-SGE method showed a strong overall performance across all test samples, although there are some areas for improvement. In Sample 1, we observed minor misclassifications at field boundaries where rice fields meet other land types. Sample 2 demonstrated exceptional results with near-perfect accuracy, whereas baseline methods, such as ResNet-UNet, ResNet-FPN, and ResNet-DeepLabV3+, missed important rice field areas. Sample 3 presented the most challenging scenario, requiring distinction between different rice growth stages within the same image - only our CA-SGE method successfully identified young versus mature crops, while other methods struggled with this detailed classification. Samples 4 and 5 confirmed our method's consistent superior performance with accurate boundary detection, whereas the baseline methods made various errors, including false identification of non-agricultural areas as rice fields and missed actual rice fields, particularly in areas with varying lighting conditions.

The proposed ResNet-DeepLabV3+with CA-SGE demonstrated the most reliable and consistent segmentation performance across all test samples. The method excelled in boundary sharpness preservation, misclassification reduction, and fine-grained feature discrimination, making it the most suitable approach for practical paddy field monitoring applications. ResNet-DeepLabV3+(Baseline) showed good overall segmentation results but lacked the precision and consistency of the attention-enhanced variant. The method occasionally struggled with complex boundary conditions and fine-grained classification tasks. ResNet-UNet and ResNet-FPN demonstrated higher tendencies toward misclassification and produced less sharp boundary delineations. They frequently failed to capture fine-grained details and exhibited reduced performance under challenging lighting conditions.

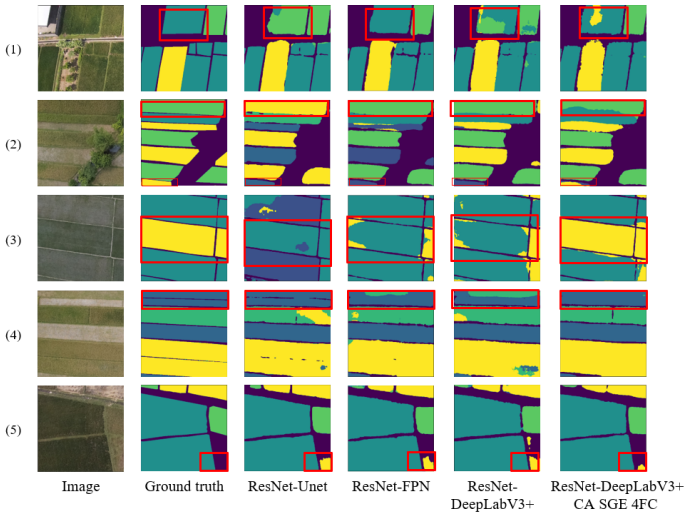


Figure 8. Segmentation result

4.6. Discussion

The experimental results directly address our primary research objective of developing an enhanced attention mechanism for improved paddy field segmentation accuracy. Our proposed CA-SGE method achieved significant improvements across all evaluation metrics compared with the baseline ResNet–DeepLabV3+, thereby successfully fulfilling the objective of enhanced segmentation performance. The secondary objective

of investigating the individual contributions of different attention mechanisms was comprehensively addressed through our ablation study. The findings revealed that CA provides consistent benefits, while SA surprisingly degraded performance, contradicting our initial hypothesis that both mechanisms would contribute positively to segmentation accuracy. Our initial hypothesis that combining CA and SGE Enhancement would create synergistic effects for paddy field segmentation was strongly supported by the experimental evidence. The CA-SGE combination outperformed individual attention mechanisms, validating our theoretical framework that CA focuses on "what" features to attend to while SGE enhances "where" to attend spatially. However, our hypothesis regarding SA effectiveness was not supported. Contrary to expectations based on previous computer vision studies, SA demonstrated counterproductive effects in our agricultural segmentation context, suggesting that SA mechanisms may be less suitable for paddy field boundary detection tasks.

5. CONCLUSION

This study developed a novel semantic segmentation method that combined CA-SGE attention, ResNet encoder, and, DeepLabV3+ decoder with 4 MLP depths, achieving superior performance for paddy field segmentation from drone imagery with IoU of 0.84, DSC of 0.89, and accuracy of 0.95. These improvements have significant implications for precision agriculture, enabling automated crop monitoring, irrigation management, and yield forecasting to support food security initiatives. The method's effectiveness with high-resolution drone imagery makes it particularly valuable for practical applications, including crop insurance assessment, variable-rate fertilizer application, and farm management systems. Future work should focus on evaluating the method's robustness across different geographical regions and crop types, optimizing for real-time edge deployment, and incorporating temporal information from multidate imagery to enhance dynamic crop monitoring capabilities.

ACKNOWLEDGEMENT

The authors would like to thank the BRIN Research Center for Geoinformatics for providing research data, the Tokopedia-UI AI Center, and the Faculty of Computer Science at University of Indonesia for allowing the use of the NVIDIA DGX-1 computer system for the research.

FUNDING INFORMATION

This work was supported in part by *"The Riset dan Inovasi untuk Indonesia Maju (RIIM) Indonesia"* endowment fund for Education Agency (LPDP) and National Research and Innovation Agency (BRIN) under Grant 82/II.7/HK/2022, and in part by the LPDP scholarship funding with Grant LOG-4244/LPDP.4/2022.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Alya Khairunnisa Rizkita	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	✓	
Masagus Muhammad Luthfi Ramadhan	✓	✓		✓	✓					✓		✓		
Yohanes Fridolin Hestrio						✓	✓	✓		✓		✓		✓
Muhammad Hannan Hunafa				✓	✓					✓		✓		
Danang Surya Candra						✓	✓	✓				✓		✓
Wisnu Jatmiko	✓	✓	✓	✓	✓					✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal Analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject Administration
Fu : **Fu**nding Acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are unpublished assets from the National Research and Innovation Agency (BRIN), Indonesia. Restrictions apply to the availability of these data, which were used under a limited-use license for academic purposes. The data are not publicly available and cannot be shared by the authors due to licensing restrictions.

REFERENCES

[1] A. Ansari, Y. P. Lin, and H. S. Lur, "Evaluating and adapting climate change impacts on rice production in indonesia: A case study of the keduang subwatershed, Central Java," *Environments - MDPI*, vol. 8, no. 11, 2021, doi: 10.3390/environments8110117.

[2] A. Massagony, T. Tam Ho, and K. Shimada, "Climate change impact and adaptation policy effectiveness on rice production in Indonesia," *International Journal of Environmental Studies*, vol. 80, no. 5, pp. 1373–1390, 2023, doi: 10.1080/00207233.2022.2099110.

[3] Nasikh, M. Kamaludin, B. S. Narmaditya, A. Wibowo, and I. Febrianto, "Agricultural land resource allocation to develop food crop commodities: lesson from Indonesia," *Heliyon*, vol. 7, no. 7, Jul. 2021, doi: 10.1016/j.heliyon.2021.e07520.

[4] M. Sun *et al.,* "Conversion of land use from upland to paddy field changes soil bacterial community structure in mollisols of Northeast China," *Microbial Ecology*, vol. 81, no. 4, pp. 1018–1028, May 2021, doi: 10.1007/s00248-020-01632-4.

[5] D. Coleman, "Geoinformation: Remote Sensing, Photogrammetry and Geographic Information Systems, Second Edition," *Geomatica*, vol. 68, p. 366+, Oct. 2014.

[6] J. Dong and X. Xiao, "Evolution of regional to global paddy rice mapping methods: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 119, pp. 214–227, Sep. 2016, doi: 10.1016/j.isprsjprs.2016.05.010.

[7] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 117, pp. 11–28, Jul. 2016, doi: 10.1016/j.isprsjprs.2016.03.014.

[8] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 7, pp. 3523–3542, 2022, doi: 10.1109/TPAMI.2021.3059968.

[9] R. Rustowicz, R. Cheong, L. Wang, S. Ermon, M. Burke, and D. Lobell, "Semantic segmentation of crop type in Africa: a novel dataset and analysis of deep learning methods," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2019–June, pp. 75–82, 2019.

[10] Ayushi and P. K. Buttar, "Satellite imagery analysis for crop type segmentation using U-Net architecture," *Procedia Computer Science*, vol. 235, pp. 3418–3427, 2024, doi: 10.1016/j.procs.2024.04.322.

[11] Y. H. Tian, K. Zhang, X. Hu, and Y. Lu, "Crop type recognition of VGI road-side images via hierarchy structure based on semantic segmentation model Deeplabv3+," *Displays*, vol. 81, p. 102574, Jan. 2024, doi: 10.1016/j.displa.2023.102574.

[12] P. Bosilj, E. Aptoula, T. Duckett, and G. Cielniak, "Transfer learning between crop types for semantic segmentation of crops versus weeds in precision agriculture," *Journal of Field Robotics*, vol. 37, no. 1, pp. 7–19, Jan. 2020, doi: 10.1002/rob.21869.

[13] M. A. Khan *et al.*, "CCDF: Automatic system for segmentation and recognition of fruit crops diseases based on correlation coefficient and deep CNN features," *Computers and Electronics in Agriculture*, vol. 155, pp. 220–236, Dec. 2018, doi: 10.1016/j.compag.2018.10.013.

[14] M. Kerkech, A. Hafiane, and R. Canals, "Vine disease detection in UAV multispectral images using optimized image registration and deep learning segmentation approach," *Computers and Electronics in Agriculture*, vol. 174, p. 105446, Jul. 2020, doi: 10.1016/j.compag.2020.105446.

[15] H. Yin, A. V. Prishchepov, T. Kuemmerle, B. Bleyhl, J. Buchner, and V. C. Radeloff, "Mapping agricultural land abandonment from spatial and temporal segmentation of Landsat time series," *Remote Sensing of Environment*, vol. 210, pp. 12–24, Jun. 2018, doi: 10.1016/j.rse.2018.02.050.

[16] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017, doi: 10.1109/TPAMI.2016.2572683.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," (*in Lecture Notes in Computer Science including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics*), 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4_28.

[18] R. Zhang, J. Chen, L. Feng, S. Li, W. Yang, and D. Guo, "A refined pyramid scene parsing network for polarimetric SAR image semantic segmentation in agricultural areas," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3086117.

[19] R. Li, S. Zheng, C. Duan, J. Su, and C. Zhang, "Multistage attention ResU-Net for semantic segmentation of fine-resolution remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2021.3063381.

[20] M. N. Mahmud, M. K. Osman, A. P. Ismail, F. Ahmad, K. A. Ahmad, and A. Ibrahim, "Road image segmentation using unmanned aerial vehicle images and deepLab V3+ semantic segmentation model," in *Proceedings - 2021 11th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2021, IEEE*, Aug. 2021, pp. 176–181, doi: 10.1109/ICCSCE52189.2021.9530950.

[21] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: convolutional block attention module," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 3–19, 2018, doi: 10.1007/978-3-030-01234-2_1.

[22] X. Li, X. Hu, and J. Yang, "Spatial group-wise enhance: improving semantic feature learning in convolutional networks," 2019. [Online]. Available: http://arxiv.org/abs/1905.09646.
[23] Y. Li *et al.*, "An improved U-net and attention mechanism-based model for sugar beet and weed segmentation," *Frontiers in Plant Science*, vol. 15, Jan. 2024, doi: 10.3389/fpls.2024.1449514.
[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE*, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[25] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," Dec. 2017. [Online]. Available: http://arxiv.org/abs/1706.05587
[26] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Jan. 2017. [Online]. Available: http://arxiv.org/abs/1412.6980.
[27] S. Müller, A.C., Guido, "Introduction to machine learning with python: a guide for data scientists," *O'reilly Media*, Inc, 2016.

BIOGRAPHIES OF AUTHORS

Alya Khairunnisa Rizkita received the Bachelor's degree from the Informatics Engineering Study Program, Institut Teknologi Sumatera, and the Master's degree from the Faculty of Computer Science, University of Indonesia. Her research interests include deep learning, face recognition, remote sensing, and computer vision. She can be contacted at email: alya.khairunnisa21@ui.ac.id.

Masagus Muhammad Luthfi Ramadhan received the Bachelor's degree from the Faculty of Computer Science, University of Sriwijaya, and the Master's degree from the Faculty of Computer Science, University of Indonesia. He is currently a research assistant with the Faculty of Computer Science, University of Indonesia. His research interests include deep learning, pattern recognition, and computer vision. He can be contacted at email: mgs.m01@ui.ac.id.

Yohanes Fridolin Hestrio obtained his Bachelor's degree from the Faculty of Mathematics and Natural Sciences at the University of Indonesia. He is currently a researcher at the National Research and Innovation Agency of Indonesia and is pursuing a Master's degree at the Faculty of Computer Science of the University of Indonesia. His research interests include remote sensing, image processing, data fusion, computer vision, machine learning, deep learning, automation, and artificial intelligence. His research interests include rice fields, palm oil plantations, oil and gas, forest fires, slum areas, cloud cover, floods, and drones. He can be contacted at email: yoha012@brin.go.id.

Muhammad Hannan Hunafa received a Bachelor's degree from the Faculty of Informatics, Telkom Technology Institute of Purwokerto. He is currently a Master's degree student at the Faculty of Computer Science, University of Indonesia. His research interests include machine learning, deep learning, and computer vision. He can be contacted at email: muhammad.hannan21@ui.ac.id.

Danang Surya Candra serves as a senior researcher at the National Research and Innovation Agency, Indonesia (BRIN). He completed his Ph.D. in 2019 at the University of Queensland, Australia, specializing in remote sensing imaging. He received his Bachelor of Science from Sebelas Maret University, Solo, Indonesia, in 2002, and his Master of Science from Beihang University, Beijing, China, in 2010. His research interests include remote sensing image processing, remote sensing applications, geographic information system (GIS), machine learning, deep learning, and artificial intelligence. He can be contacted at email: dana004@brin.go.id.

Wisnu Jatmiko is one of the academic staff at the Faculty of Computer Science, head of the Intelligent Robotics and System (IRoS) Laboratory, and head of Artificial Intelligence Cluster Research at the University of Indonesia. He obtained his Bachelor of Engineering and Master of Computer Science degrees from the University of Indonesia in 1997 and 2000, respectively. In 2007, he received his Dr. Eng. degree in Micro-Nano System Engineering at Nagoya University, Japan. In September 2017, he became a professor at the Faculty of Computer Science at the University of Indonesia. He has served as Chair of The Institute of Electrical and Electronics Engineers (IEEE) Indonesia Section for the 2019 and 2020 periods. His current research interests include autonomous robots, biomedical, and computer vision. He can be contacted at email: wisnuj@cs.ui.ac.id.