Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 7, No. 3, June 2026, Page. 2246-2258
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5523

Enhancing Flood Area Segmentation in Remote Sensing Images Using Hybrid Attention Mechanism on DeepLabV3+ with ResNet-50 Backbone

Annisa Syifaul Ummah[1], Esti Suryani*[2], Herdito Ibnu Dewangkoro[3]

[1,3]Department of Informatics, Universitas Sebelas Maret, Indonesia
[2]Department of Data Science, Universitas Sebelas Maret, Indonesia

Email: [1]annisasyifaul09@student.uns.ac.id

Received : Dec 4, 2025; Revised : Dec 30, 2025; Accepted : Jan 19, 2026; Published : Jun 15, 2026

Abstract

Flooding is caused by climate change and urbanization, so rapid and accurate monitoring is essential in supporting emergency response. However, flood segmentation still faces challenges in dense vegetation. This study aims to improve and analyze the performance of the Hybrid Attention Mechanism in the form of Point-wise spatial attention (PSA) and Squeeze-and-Excitation Block (SE Block) in the DeepLabV3+ architecture with the ResNet-50 backbone. The methods used include collecting a dataset of 600 training and 63 validation, data augmentation, model development and Hybrid Attention Mechanism design, hyperparameter optimization, ablation study, and performance evaluation. The ablation results obtained show the best performance with accuracy of 0.9624, F1-score of 0.9618, IoU (Non-Flood) of 0.9323, IoU (Flood) of 0.9208, and mIoU of 0.9265, surpassing previous studies that used Modified U-Net in detecting floods in dense vegetation. This research contributes to the development of a flood segmentation model based on a hybrid attention mechanism, which is more effective in detecting flooded areas in densely vegetated regions.

Keywords: *DeepLabV3+, Flood Segmentation, Hybrid Attention Mechanism, Remote Sensing, ResNet-50, Semantic Segmentation*

1. INTRODUCTION

Natural disasters caused economic losses of more than $300 million and affected 2.3 million people [1]. To reduce the risks caused by flooding, flood monitoring is essential in improving disaster emergency management. In addition, climate change and increased urbanization have exacerbated rainfall intensity, resulting in more frequent flooding [2].

Climate change can trigger flash floods, such as the event that occurred in Hill Country, Texas, on July 4-7, 2025, which killed 120 people and left 170 others missing [3]. Flood damage can cause total destruction to buildings, bridges, and factories [4]. Therefore, mapping flood-prone areas is necessary to highlight the risks and losses caused by disasters [5].

In recent decades, remote sensing has been used as a tool for mapping large-scale flooding [6]. The development of modern deep learning-based image processing methods combined with remote sensing has accelerated flood identification and improved the accuracy of image segmentation from complex data [7]. This technology enables the monitoring of flooded areas on a large scale and in a short time [8].

In addition, Unmanned Aerial Vehicles (UAV) have emerged as unmanned aircraft that collect images quickly and accurately [9]. Unmanned Aerial Vehicles (UAV) are capable of detecting hard-to-reach flooded areas [10]. However, the main challenge in flood segmentation is the difficulty of detecting flooded areas in complex environments, such as areas with dense vegetation and flooded areas filled with various other objects around them [11].

Previous studies have investigated this issue, one of which modified U-Net by adding LeakyReLU, Batch Normalization, and Dropout [11]. The results showed IoU value of 0.9203 and an overall accuracy of 0.9587. Although the accuracy was high, some small details in dense vegetation were not detected.

In an effort to improve results in this segmentation, there is research that uses the DeeplabV3+ method by comparing various backbones, such as ResNet-18, ResNet-50, MobileNetV2, and Xception, for mapping flood areas in nine classes, including Background, BuildingFlooded, BuildingNonFlooded, and others [12]. The results of the study show that DeepLabV3+ with the ResNet-50 backbone provides the best performance with a total of 23.5 million parameters, a global accuracy of 0.92, and mIoU of 0.71. This model is capable of producing detailed and accurate segmentation, as well as clearly recognizing flooded roads and buildings.

In addition, multi-class segmentation research also occurs in scene parsing. In this research, the FCN method with a ResNet-50 backbone was used and supplemented with Point-wise Spatial Attention (PSA) to become Point-wise Spatial Attention Network (PSANet) [13]. This addition produced smoother segmentation outputs, more complete objects, and high performance in the three datasets, namely ADE20K, PASCAL VOC 2012, and Cityscapes.

The problem of detecting floods is similar to challenges in the medical field, particularly in detecting lung tumors that are difficult to distinguish from their surroundings, such as when tumors attach to blood vessels or chest walls. The related research uses the Hybrid Attention Mechanism method combined with DenseNet-12. The use of the Hybrid Attention Mechanism produces clearer tumor edge details, achieving accuracy of 94.61%, which is higher than the CNN, RNN, and U-Net methods [14].

In addition, the Multi-scale strategy-Segmentation Squeeze-and-Excitation Blocks-Conditional Random Field (M-SegSEUNet-CRF) method adds SE Blocks to improve segmentation capabilities compared to U-Net [15]. The SE Block is used to highlight important inter-channel features related to tumor areas, so that the model focuses more on the parts of the image that contain tumor tissue and ignores surrounding tissue such as normal lungs or blood vessels.

Based on previous studies, flood segmentation still faces significant challenges in complex environments, particularly in densely vegetated areas, where existing models often fail to capture fine-grained boundaries and small flooded regions. This limitation indicates a clear research gap in developing flood segmentation models that can simultaneously exploit spatial and channel-wise information. To address this gap, this study proposes the DeeplabV3+ method with a ResNet-50 backbone combined with a Hybrid Attention Mechanism [12]. The use of DeepLabV3+ with a ResNet-50 backbone is expected to provide better segmentation details between flooded and non-flooded areas. The Hybrid Attention Mechanism is a combination of spatial attention and channel attention [14]. Spatial attention uses Point-wise Spatial Attention (PSA) to produce smoother segmentation outputs and high performance [13]. Channel attention uses the Squeeze-and-Excitation (SE) Block to emphasize important features between channels [15].

This study aims to implement and comprehensively analyze the performance of a Hybrid Attention Mechanism combining Point-wise Spatial Attention (PSA) and the Squeeze-and-Excitation (SE) Block within the DeeplabV3+ architecture using a ResNet-50 backbone. Through this integration, the research seeks to improve the model's capability to identify and segment flood-affected areas in a more effective and reliable manner. The findings are expected to serve as a valuable reference for disaster mitigation agencies and SAR teams in enhancing the process of flood area mapping. Furthermore, the results of this study are anticipated to contribute to more efficient mitigation planning, evacuation coordination, and the distribution of emergency assistance during flood events.

Vol. 7, No. 3, June 2026, Page. 2246-2258
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5523

2. METHOD

This study uses the DeeplabV3+ architecture with a ResNet-50 backbone and integrates two types of attention mechanisms, namely Point-wise Spatial Attention (PSA) and Squeeze-and-Excitation (SE) Block. The detailed workflow and processing steps of the proposed method can be seen in Figure 1.

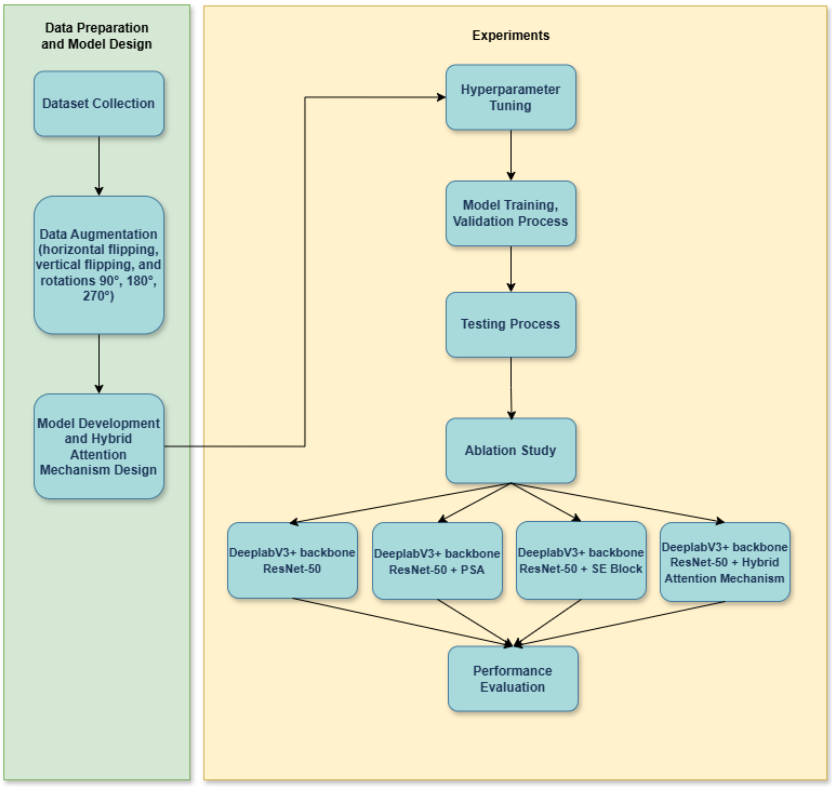


Figure 1. Proposed method

2.1. Data Collection

This study is a continuation of previous research therefore, the dataset used follows the same reference as that study [11]. The dataset was obtained from the Kaggle platform and contained remote sensing images of areas affected and unaffected by flooding. The initial dataset consisted of 600 training data and 63 validation data.

2.2. Data Augmentation

Data augmentation is used to add variety to training datasets through various image transformations. This process makes models more robust, less prone to overfitting, and better able to understand patterns and object shapes from a variety of visual conditions.

2.3. Model Development and Hybrid Attention Mechanism Design

This study uses the DeeplabV3+ with ResNet-50 backbone as the base model. Figure 2 shows the architecture development of DeepLabV3+ with the ResNet-50 backbone, which is supplemented with Hybrid Attention in the encoder [16]. The process begins by inputting images into the ResNet-50 backbone, as shown in Figure 3. ResNet-50 consists of 50 layers [17]. In ResNet-50, the 3×H×W image is processed through a Convolution 7×7, ReLU, and MaxPool so that it has a resolution of 1/4 [18]. After that, the features pass through four cascaded layers, namely Layer1 (256 channels, 1/4) as low-level features, Layer2 (512 channels, 1/8), Layer3 (1024 channels, 1/16), and Layer4 (2048 channels, 1/32).

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2246-2258
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5523

After the features are extracted by the backbone, the second path of the architecture sends the backbone output features to the Hybrid Attention Mechanism as shown in Figure 4. This mechanism consists of two paths, namely Point-wise Spatial Attention (PSA) and SE Block. In the PSA path, the features are simplified and processed through the collect and distribute stages to produce a spatial representation z that highlights important areas in the image. The PSA equation can be seen in eq (1).

$$z_i = \frac{1}{N}\sum_{\forall j} a^c_{i,j} x_j + \frac{1}{N}\sum_{\forall j} a^d_{i,j} x_j \quad (1)$$

Where N is the total number of positions, and $\forall j$ represents the contribution of each point. The values $a^c_{i,j}$and $a^d_{i,j}$correspond to the attention from the point-wise maps $A^c$and A^d. These attention outputs then produce wide-context feature representations $Z^c$and Z^d.

On the SE Block path, features undergo global pooling and are processed through two fully connected layers with sigmoid activation to generate attention weights per channel, thereby amplifying relevant channels and suppressing less important ones [19]. The convolution equation can be seen in eq (2).

$$u_c = \sum_{s=1}^{C'} v^s_c * x^s \quad (2)$$

Where c'denotes the total number of input channels, x^s is the s-th input channel, v^s_c is the 2D convolution kernel, and $*$ represents convolution. This formula shows that the output feature u_c is obtained by convolving all input channels with their corresponding kernels.

The features enriched by Hybrid Attention are then combined with the ASPP output through a concatenate operation. The ASPP module captures object context at various scales through several branches, namely image pooling, 1×1 conv, and 3×3 conv with dilation rates of 6, 12, and 18. After the concatenation process, a 1×1 convolution is performed on the encoder before the features are passed to the decoder stage [20].

In the decoder, the features from the encoder are upsampled four times and combined with features from the attention path. Next, a 3×3 convolution with dilation is performed to smooth the representation, followed by another upsampling until the image resolution reaches its original size [21]. This stage produces the final segmentation map in the output section.

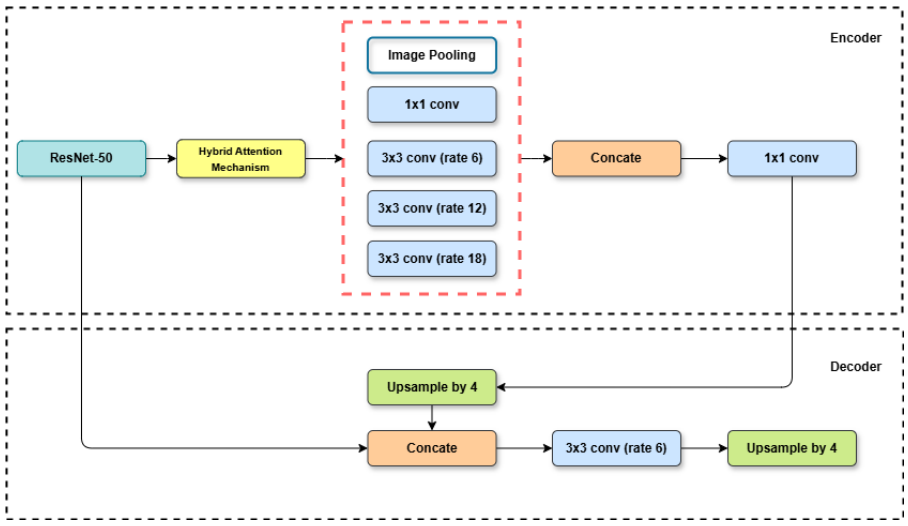


Figure 2. DeeplabV3+ Architecture with ResNet-50 backbone And Hybrid Attention Mechanism

Vol. 7, No. 3, June 2026, Page. 2246-2258
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5523

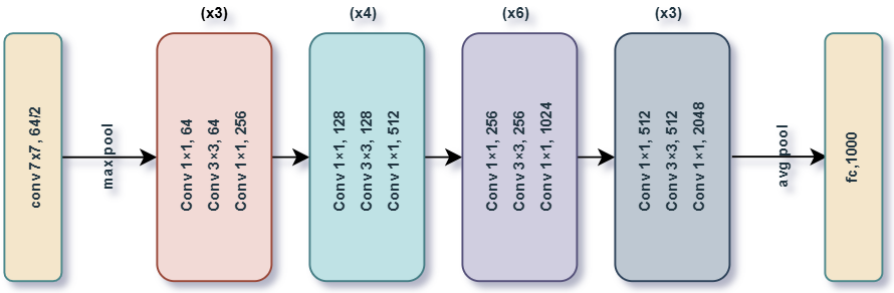


Figure 3. ResNet-50 Architecture

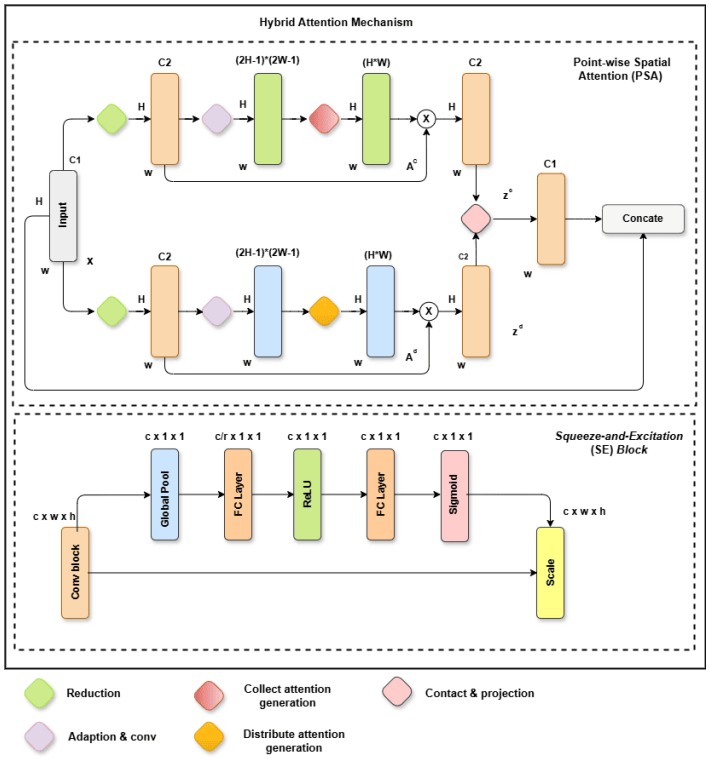


Figure 4. Hybrid Attention Mechanism Architecture

2.4. Model Train, Validation, and Testing Process

The model hyperparameters were determined based on the best outcomes obtained from three experimental trials designed to evaluate training stability and overall model performance. These experiments were conducted to identify the configuration that enabled the model to learn effectively while maintaining consistent convergence throughout the training process. The final selection represents the most stable and optimal setting among the tested variations. The configuration is shown in Table 1.

Table 1. The best hyperparameter

Hyperparameters	Values
Learning Rate	7e-6
Optimizer	Adam
Epochs	100
Loss Function	Cross Entropy Loss
Activation Function	ReLU

Vol. 7, No. 3, June 2026, Page. 2246-2258
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5523

2.5. Performance Evaluation

The evaluation metrics used include:

1. Accuracy, which is the overall correctness of the model's predictions [22].

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3)$$

2. Intersection over Union (IoU), which measures the overlap between predicted and ground truth regions for each class [23].

$$IoU = \frac{TP}{TP + FP + FN} \quad (4)$$

3. Mean Intersection over Union (mIoU), which measures the accuracy of the segmentation model in predicting each pixel class as a whole [24].

$$mIoU = \frac{1}{K}\sum_{c=1}^{K} IoU_c \quad (5)$$

4. F1-Score, which assesses the weighted average between precision and recall with a value range of [0,1] [25].

$$F1 - Score = 2\,x\,\frac{Precision\ x\ Recall}{Precision + Recall} \quad (6)$$

5. Params and FLOPs, which are used to assess the number of parameters and computational complexity of the model [26].
6. Inference Time (ms/image), which represents the average time required by the model to process a single image and generate the corresponding prediction during the inference phase.

Where *TP* is true positive and *TN* is true negative, *FP* is false positive and *FN* is false negative [27]. *K* is the number of classes, *c* denotes the class index**,** mIoU is the IoU value for each class, and precision and recall indicate the accuracy and success of positive detection, respectively.

3. RESULT

3.1. Data Collection

In this study, the dataset was obtained from Kaggle. A sample of the dataset is shown in

Table 2. The data was then split 70:30 into 420 training samples and 180 validation samples. Meanwhile, the original 63 validation images provided in the dataset were used as the test set due to dataset limitations, ensuring that model evaluation could still be conducted using unseen data. Then, the images resized to 128x128 pixels. The dataset can be access in here: https://www.kaggle.com/datasets/lihuayang111265/flood-semantic-segmentation-dataset.

Table 2. Sample of dataset

Data 1	Ground Truth 1	Data 2	Ground Truth 2
			

Vol. 7, No. 3, June 2026, Page. 2246-2258
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5523

3.2. Data Augmentation

Data augmentation is used to increase the diversity of the training dataset through various image transformations. This process makes the model more robust, reduces the risk of overfitting, and enables it to better understand object patterns and shapes under diverse visual conditions. The augmentation techniques applied in this study include horizontal flipping, vertical flipping, and rotations at angles of 90°, 180°, and 270°.

3.3. Hyperparameters Tuning

The results of the established hyperparameter tuning process are presented in Figure 5 and Figure 6. The training and validation loss decreased consistently, reaching 0.046 and 0.140, respectively, at the final epoch. The validation loss began to stabilize within the range of epochs 40–100, indicating that the model maintained controlled performance without significant overfitting. Meanwhile, the final model accuracy reached 0.982 for the training set and 0.958 for the validation set, with a small difference of 0.024.

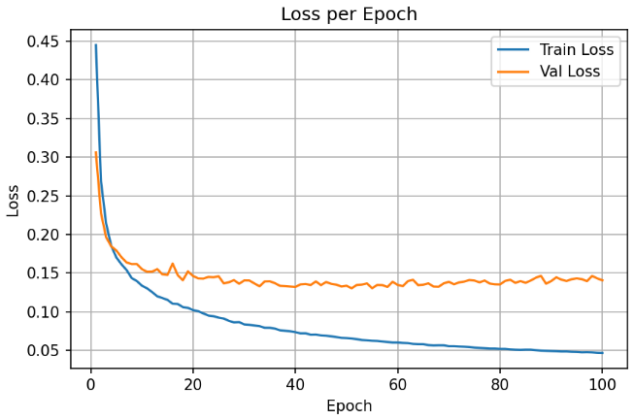


Figure 5. Loss graph based on selected hyperparameters

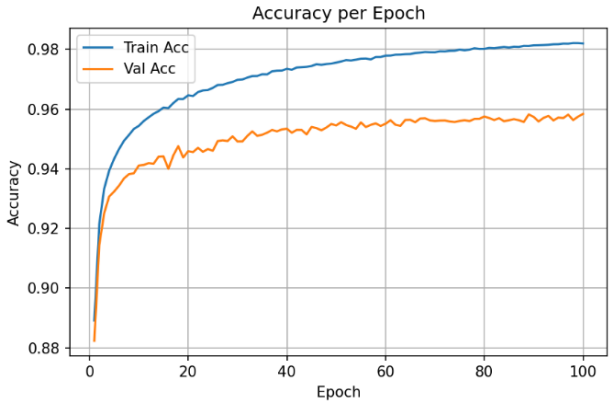


Figure 6. Accuracy graph based on selected hyperparameters

3.4. Ablation Study

Ablation was performed on DeepLabV3+ with ResNet-50 by gradually adding PSA, SE Block, and Hybrid Attention Mechanism to assess the contribution of each module to segmentation performance. The results of the experiment can be seen in Table 3, which is obtained from the average of three experiments. The model with the Hybrid Attention Mechanism achieved an accuracy of 0.9624, an F1-score of 0.9618, IoU (Non-Flood) of 0.9323, IoU (Flood) of 0.9208 and an mIoU of 0.9265.

Vol. 7, No. 3, June 2026, Page. 2246-2258
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5523

Table 3. Performance test results for each ablation model

Models	Accuracy	F1-Score	IoU (Non-Flood)	IoU (Flood)	mIoU
DeeplabV3+ backbone ResNet-50	0.9570	0.9565	0.9232	0.9101	0.9167
DeeplabV3+ backbone ResNet-50 + PSA Attention	0.9605	0.9601	0.9291	0.9176	0.9233
DeeplabV3+ backbone ResNet-50 + SE Block	0.9619	0.9614	0.9313	0.9201	0.9258
DeeplabV3+ backbone ResNet-50 + Hybrid Attention (Proposed method)	**0.9624**	**0.9618**	**0.9323**	**0.9208**	**0.9265**

The computational complexity of each ablation model was analyzed, as summarized in Table 4. The results indicate that the proposed Hybrid Attention model has the highest computational cost, with the largest number of parameters, increased FLOPs, and the longest inference time of 3.36 ms per image. All experiments were conducted on the Kaggle platform using an NVIDIA Tesla P100 GPU with 16 GB memory, providing a consistent and standardized hardware environment for fair performance evaluation.

Table 4. Computational complexity for each ablation model

Models	Number of Parameters	FLOPs	Inference Time (ms/image)
DeeplabV3+ backbone ResNet-50	40,347,298	2,919,211,520	1.51
DeeplabV3+ backbone ResNet-50 + PSA	49,798,818	3,058,955,776	3.21
DeeplabV3+ backbone ResNet-50 + SE Block	43,132,578	2,922,488,320	2.25
DeeplabV3+ backbone ResNet-50 + Hybrid Attention (Proposed method)	58,162,338	3,128,260,096	3.36

In addition, the training and validation loss and accuracy curves shown in Table 5 indicate that all models converge rapidly within the first 20–30 epochs. The proposed Hybrid Attention model demonstrates the most stable convergence behavior, characterized by smoother validation loss and a smaller training validation accuracy gap, indicating improved generalization capability. A comparison of the segmentation quality of each model can be seen in Table 6. The table contains several descriptions, namely (a) Original image Image 1 and 2, (b) Ground truth 1 and 2, (c) DeeplabV3+ backbone ResNet-50, (d) DeeplabV3+ backbone ResNet-50 with PSA, (e) DeeplabV3+ backbone ResNet-50 with SE Block, (f) DeeplabV3+ backbone ResNet-50 with Hybrid Attention.

3.4.1. DeeplabV3+ with ResNet-50 backbone

DeepLabV3+ with ResNet-50 backbone produced accuracy of 0.9570, F1-score of 0.9565, IoU (Non-Flood) of 0.9232, IoU (Flood) of 0.9101, and mIoU of 0.9167. The segmentation results for Image 1 and Image 2 are presented in Table 6. The results show that the model is capable of detecting flooded areas in dense vegetation in Image 2, although its coverage is still limited or under-segmentation, while in Image 1 there are still errors in classifying non-flooded areas as flooded or over-segmentation. This model has the lowest complexity with approximately 40 million parameters and 2.9 billion FLOPs, making it the most efficient but with the lowest accuracy.

3.4.2. DeeplabV3+ with ResNet-50 backbone and PSA

DeepLabV3+ with ResNet-50 and PSA produced accuracy of 0.9605, F1-score of 0.9601, IoU (Non-Flood) of 0.9291, IoU (Flood) of 0.9176, and mIoU of 0.9233. The segmentation results for Image 1 and Image 2 are presented in Table 6.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 7, No. 3, June 2026, Page. 2246-2258
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5523

Table 5. Training and validation loss and accuracy curves for each ablation model

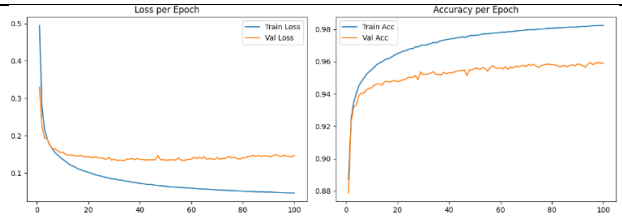
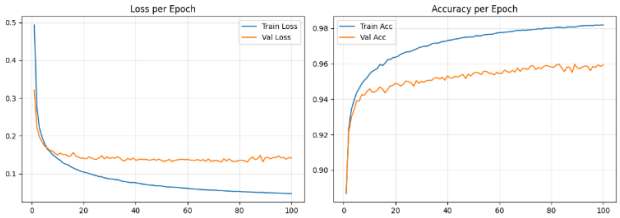
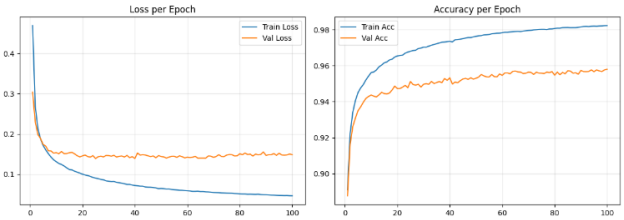
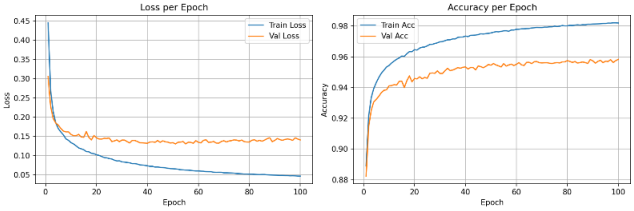
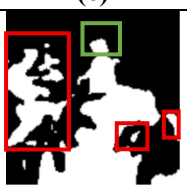
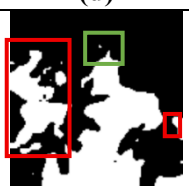
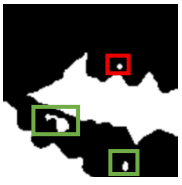
Models	Graph
DeeplabV3+ backbone ResNet-50	
DeeplabV3+ backbone ResNet-50 + PSA	
DeeplabV3+ backbone ResNet-50 + SE Block	
DeeplabV3+ backbone ResNet-50 + Hybrid Attention (Proposed method)	

Table 6. Comparison of segmentation results for each ablation model

Image	(a)	(b)	(c)	(d)	(e)	(f)
1						
	Acc/F1-Score /IoU (Non-Flood)/ IoU (Flood)	0.8673/0.8748 /0.8268/0.6381	0.8646/0.8720 /0.8245/0.6279	0.9053/0.9091 /0.8773/0.7072	0.9344/0.9360 /0.9150/0.7765	
2						
	Acc/F1-Score /IoU (Non-Flood)/ IoU (Flood)	0.9605/0.9604 /0.9425/0.8882	0.9623/0.9621 /0.9441/0.8878	0.9648/0.9647 /0.9455/0.8930	0.9667/0.9666 /0.9464/0.8940	

Visually, the contours of objects and segmentation boundaries in Image 1 still appear jagged and less smooth, but they are better than the baseline model. In Image 2, the object boundaries are clearer and over-segmentation is reduced, although there are still areas of under-segmentation. PSA helps detect

Vol. 7, No. 3, June 2026, Page. 2246-2258
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5523

details of objects that have branched shapes. However, this model has a fairly high complexity with approximately 49 million parameters and 3 billion FLOPs.

3.4.3. DeeplabV3+ with backbone ResNet-50 and SE Block

DeepLabV3+ with ResNet-50 backbone and SE Block produced accuracy of 0.9619, F1-score of 0.9614, IoU (Non-Flood) of 0.9313, IoU (Flood) of 0.9201, and mIoU of 0.9258. With 43 million parameters and 2.9 billion FLOPs, this model is quite efficient and shows improved segmentation. The segmentation results for Image 1 and Image 2 are presented in Table 6. In Images 1 and 2, over-segmentation and noise are reduced, although some flooded areas still do not fully resemble the ground truth, especially in branched and edge areas, even though the image edges are smoother and more stable than the baseline with PSA.

3.4.4. DeeplabV3+ with backbone ResNet-50 and Hybrid Attention Mechanism

DeepLabV3+ with ResNet-50 backbone with Hybrid Attention Mechanism produced accuracy of 0.9624, F1-score of 0.9618, IoU (Non-Flood) of 0.9323, IoU (Flood) of 0.9208, and mIoU of 0.9265, but with the longest computation time. The segmentation results for Image 1 and Image 2 are presented in Table 6. In Image 1, the flooded area is detected more completely with minimal noise and more precise contours. In Image 2, object boundaries and the consistency of flooded regions show improvement, with shapes more closely resembling the ground truth.

3.4.5. Comparison of Baseline and Hybrid Attention Mechanism

The segmentation results for Image 1 and Image 2 are presented in Table 6. The addition of a hybrid attention mechanism improves model performance compared to the baseline. The combination of SE Block and PSA helps the model better understand the position of important parts of the image and the information in each channel, enabling the model to more effectively highlight important features and produce more accurate flood segmentation. In Image 1, the baseline model shows strong over-segmentation, while the hybrid model significantly reduces these errors and yields more precise flood boundaries. In Image 2, the hybrid model provides segmentation closer to the ground truth, especially in dense vegetation, though both models still show under-segmentation in these areas.

4. DISCUSSIONS

To see the contribution of this research, a comparison was made with previous researcher, which used a similar dataset but a different method, in order to highlight the differences, advantages, weaknesses, and performance improvements of the proposed model. The performance comparison can be seen in Table 7.

Table 7. Comparison of performance metrics with previous studies

Methods	Accuracy	IoU (Flood)
Modification of U-Net [11]	0.9587	0.9203
DeeplabV3+ ResNet-50 backbone + Hybrid Attention (Proposed method)	**0.9624**	**0.9208**

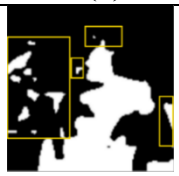
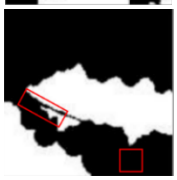
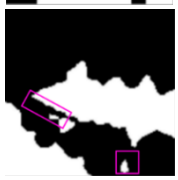
The proposed method shows improved performance compared to the Modified U-Net method. The model achieves average accuracy of 0.9624 and average IoU (Flood) of 0.9208, surpassing the previous method which obtained an accuracy of 0.9587 and an IoU (Flood) of 0.9203.

Additionally, the improvement in segmentation results can be seen in Table 8. In the table, (a) shows the Original image 1 and 2, (b) shows the Ground truth 1 and 2, (c) shows the Modified U-Net [11], and (d) shows the DeeplabV3+ backbone ResNet-50 with Hybrid Attention (the proposed method).

Vol. 7, No. 3, June 2026, Page. 2246-2258
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5523

The proposed method is capable of detecting flood regions in densely vegetated areas, such as in Image 2, although some under-segmentation remains, where previous method fails to identify the flooded region. The DeepLabV3+ ResNet-50 method with Hybrid Attention Mechanism successfully addresses this limitation by producing more precise segmentation results, particularly in areas with dense vegetation that were previously undetected in earlier studies. However, in Image 1, over-segmentation is still observed, mainly due to land areas with color characteristics similar to floodwater, which causes the model to misclassify certain regions.

Table 8. Comparison of the segmentation results of the proposed model with previous studies

Image	(a)	(b)	(c)	(d)
1				
2				

This study has high relevance to the development of computer science applications in Indonesia, particularly in utilizing remote sensing data such as Sentinel and UAV imagery for flood monitoring. The integration of DeeplabV3+ backbone ResNet-50 with Hybrid Attention offers strong potential for automated and accurate flood mapping. Nevertheless, the performance gains achieved by the proposed method come at the cost of increased model complexity in terms of parameter count and computational requirements (FLOPs), highlighting a trade-off between segmentation accuracy and computational efficiency, especially for deployment in resource-constrained environments. Furthermore, the evaluation conducted in this study is limited to a single dataset, which may not fully reflect the model's generalization capability. Therefore, future work should include validation on additional datasets, such as the MediaEval Flood dataset or flood datasets from different regions in Indonesia, to ensure robustness and strengthen the applicability of the proposed method under diverse real-world conditions.

5. CONCLUSION

This study integrates a Hybrid Attention Mechanism into the DeepLabV3+ architecture with a ResNet-50 backbone for flood segmentation. PSA improves boundary clarity by capturing positional relationships, while the SE Block reduces noise and over-segmentation. Their combination achieves the best performance, with an accuracy of 0.9624, an F1-score of 0.9618, IoU (Non-Flood) of 0.9323, IoU (Flood) of 0.9208, and an mIoU of 0.9265, outperforming previous approaches, particularly in dense vegetation areas. These findings contribute to semantic segmentation research in computer vision, especially for remote sensing and disaster informatics, and support automated flood mapping in disaster-prone regions. However, challenges remain in handling color similarity and narrow vegetation structures. Future work should explore more adaptive mechanisms to address color similarity, investigate lighter backbone architectures to improve computational efficiency, and consider integrating multi-modal data sources such as optical and SAR imagery.

ACKNOWLEDGEMENT

This research was supported by the Research Strengthening Program for Research Groups (Penguatan Kapasitas Grup Riset / PKGR) Scheme B, Universitas Sebelas Maret, Fiscal Year 2025, as

stipulated in the research contract No. 371/UN27.22/PT.01.03/2025. The authors would like to express their sincere gratitude and highest appreciation for the financial support, which made this research possible.

REFERENCES

[1] P. Akiva, M. Purri, K. Dana, B. Tellman, and T. Anderson, "H2O-Net: Self-Supervised Flood Segmentation via Adversarial Domain Adaptation and Label Refinement," 2021.

[2] N. A. Muhadi, A. F. Abdullah, S. K. Bejo, M. R. Mahadi, and A. Mijic, "Image Segmentation Methods for Flood Monitoring System," *Water 2020, Vol. 12, Page 1825*, vol. 12, no. 6, p. 1825, Jun. 2020, doi: 10.3390/W12061825.

[3] Mawardi Isal, "121 Orang Tewas Akibat Banjir Bandang di Texas, 170 Lainnya Hilang," Detik News, Jakarta, 10 Juli 2025.

[4] I. A. Hadimlioglu and S. A. King, "Geo-Information Visualization of Flooding Using Adaptive Spatial Resolution", doi: 10.3390/ijgi8050204.

[5] G. Antzoulatos *et al.*, "Flood Hazard and Risk Mapping by Applying an Explainable Machine Learning Framework Using Satellite Imagery and GIS Data," *Sustainability 2022, Vol. 14, Page 3251*, vol. 14, no. 6, p. 3251, Mar. 2022, doi: 10.3390/SU14063251.

[6] L. Hashemi-Beni and A. A. Gebrehiwot, "Flood Extent Mapping: An Integrated Method Using Deep Learning and Region Growing Using UAV Optical Data," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 14, pp. 2127–2135, 2021, doi: 10.1109/JSTARS.2021.3051873.

[7] W. Li *et al.*, "High-Performance Segmentation for Flood Mapping of HISEA-1 SAR Remote Sensing Images," *Remote Sensing 2022, Vol. 14, Page 5504*, vol. 14, no. 21, p. 5504, Nov. 2022, doi: 10.3390/RS14215504.

[8] Y. Li, "The research on landslide detection in remote sensing images based on improved DeepLabv3+ method," *Scientific Reports 2025 15:1*, vol. 15, no. 1, pp. 7957-, Mar. 2025, doi: 10.1038/s41598-025-92822-y.

[9] E. T. Wasehun *et al.*, "UAV and satellite remote sensing for inland water quality assessments: a literature review," *Environmental Monitoring and Assessment 2024 196:3*, vol. 196, no. 3, pp. 1–31, Feb. 2024, doi: 10.1007/S10661-024-12342-6.

[10] M. Fawakherji and L. Hashemi-Beni, "Flood detection and mapping through multi-resolution sensor fusion: integrating UAV optical imagery and satellite SAR data," *Geomatics, Natural Hazards and Risk*, vol. 16, no. 1, Dec. 2025, doi: 10.1080/19475705.2025.2493225;WGROUP:STRING:PUBLICATION.

[11] R. Arya, J. Choudhary, M. Pandey, and D. P. Singh, "Advanced Semantic Segmentation of Flooded Regions in UAV Imagery Using a Modified U-Net Model," *2025 IEEE International Students' Conference on Electrical, Electronics and Computer Science, SCEECS 2025*, 2025, doi: 10.1109/SCEECS64059.2025.10941437.

[12] 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, pp. 363–371, Jun. 2025, doi: 10.1016/J.NHRES.2024.12.003.

[13] H. Zhao *et al.*, "PSANet: Point-wise Spatial Attention Network for Scene Parsing," 2018. Accessed: Aug. 22, 2025. [Online]. Available: https://github.com/hszhao/PSANet

[14] H. Hu, Q. Li, Y. Zhao, and Y. Zhang, "Parallel Deep Learning Algorithms with Hybrid Attention Mechanism for Image Segmentation of Lung Tumors," *IEEE Trans Industr Inform*, vol. 17, no. 4, pp. 2880–2889, Apr. 2021, doi: 10.1109/TII.2020.3022912.

[15] B. Zhang *et al.*, "Multi-scale segmentation squeeze-and-excitation UNet with conditional random field for segmenting lung tumor from CT images," *Comput Methods Programs Biomed*, vol. 222, p. 106946, Jul. 2022, doi: 10.1016/J.CMPB.2022.106946.

[16] Z. Wang, J. Wang, K. Yang, L. Wang, F. Su, and X. Chen, "Semantic segmentation of high-resolution remote sensing images based on a class feature attention mechanism fused with Deeplabv3+," *Comput Geosci*, vol. 158, p. 104969, Jan. 2022, doi: 10.1016/J.CAGEO.2021.104969.

Vol. 7, No. 3, June 2026, Page. 2246-2258
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.2026.7.3.5523

[17] A. Amelio *et al.*, "Representation and compression of Residual Neural Networks through a multilayer network based approach," *Expert Syst Appl*, vol. 215, Apr. 2023, doi: 10.1016/j.eswa.2022.119391.

[18] S. E. Abdallah *et al.*, "Deep Learning Model Based on ResNet-50 for Beef Quality Classification," *Information Sciences Letters*, vol. 12, no. 1, p. 289, 2023, doi: 10.18576/isl/120124.

[19] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," 2018. Accessed: Jun. 30, 2025. [Online]. Available: http://image-net.org/challenges/LSVRC/2017/results

[20] C. Y. Hsu, R. Hu, Y. Xiang, X. Long, and Z. Li, "Improving the Deeplabv3+ Model with Attention Mechanisms Applied to Eye Detection and Segmentation," *Mathematics 2022, Vol. 10, Page 2597*, vol. 10, no. 15, p. 2597, Jul. 2022, doi: 10.3390/MATH10152597.

[21] S. O. Atik, M. E. Atik, and C. Ipbuker, "Comparative research on different backbone architectures of DeepLabV3+ for building segmentation," *https://doi.org/10.1117/1.JRS.16.024510*, vol. 16, no. 2, p. 024510, May 2022, doi: 10.1117/1.JRS.16.024510.

[22] M. Fadhil and R. A. Saputra, "Klasifikasi dan evaluasi performa model random forest untuk prediksi," *Jurnal Teknik*, vol. 12, no. 2, Oct. 2023, doi: 10.31000/JT.V12I2.9099.

[23] D. Wang *et al.*, "SAMRS: Scaling-up Remote Sensing Segmentation Dataset with Segment Anything Model", Accessed: Dec. 22, 2025. [Online]. Available: https://segment-anything.com/demo

[24] X. Hao, L. Yin, X. Li, L. Zhang, and R. Yang, "A Multi-Objective Semantic Segmentation Algorithm Based on Improved U-Net Networks," *Remote Sensing 2023, Vol. 15, Page 1838*, vol. 15, no. 7, p. 1838, Mar. 2023, doi: 10.3390/RS15071838.

[25] S. Cai, Y. Tian, H. Lui, H. Zeng, Y. Wu, and G. Chen, "Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network Cai et al. Dense-UNet for MPM image segmentation," *Quant Imaging Med Surg*, vol. 10, no. 6, 2020, doi: 10.21037/qims-19-1090.

[26] G. Wang, J. Yu, W. Xu, A. Muhammad, and D. Li, "Automated fish counting system based on instance segmentation in aquaculture," *Expert Syst Appl*, vol. 259, p. 125318, Jan. 2025, doi: 10.1016/J.ESWA.2024.125318.

[27] T. Yang, S. Zhou, A. Xu, J. Ye, and J. Yin, "An Approach for Plant Leaf Image Segmentation Based on YOLOV8 and the Improved DEEPLABV3+," *Plants 2023, Vol. 12, Page 3438*, vol. 12, no. 19, p. 3438, Sep. 2023, doi: 10.3390/PLANTS12193438.