Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.4, August 2026, pp. 3132~3139
e-ISSN: 2548-6861

Comparison of U-Net and Attention U-Net for Binary Flood Segmentation from UAV Imagery

Fariida Qurrota 'Aini [1]*, Muhamad Akrom [2]*, Gustina Alfa Trisnapradika [3]*
* Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang, Indonesia
frdaini22@gmail.com[1], m.akrom@dsn.dinus.ac.id [2], gustina.alfa@dsn.dinus.ac.id [3]

Article Info

Article history:

Received 2026-06-09
Revised 2026-07-02
Accepted 2026-06-20

Keyword:

Attention U-Net,
Deep Learning,
Flood Segmentation,
Image Segmentation,
UAV Imagery

ABSTRACT

Flood disasters consistently cause massive damage every year, making rapid mapping of affected areas crucial for coordinating emergency aid. The use of unmanned aerial vehicles (UAVs) offers a practical solution to obtain high-resolution aerial imagery, but manually identifying flood areas from hundreds of images remains time-consuming. This study analyzes and compares two deep learning segmentation architectures, U-Net and Attention U-Net, for automatic flood area detection from UAV RGB images. Both models were trained using 290 image-mask pairs from a public dataset, with a split of 70% for training, 10% for validation, and 20% for testing. Images were processed at a resolution of 256×256 pixels, normalized to the range [0,1], and augmented with horizontal flipping, brightness adjustment, and affine transformations. Attention U-Net enhances the standard U-Net structure by adding attention gates to all skip connections in the decoder to suppress irrelevant background features. Both models were evaluated across five independent training runs using different random seeds to assess result robustness. Across these runs, Attention U-Net achieved a marginally higher mean IoU (77.11% ± 0.76) and Dice/F1 (87.07% ± 0.49) compared to the U-Net baseline (IoU: 76.97% ± 0.54; Dice/F1: 86.98% ± 0.34), but a paired t-test revealed that these differences were not statistically significant (IoU: $p = 0.77$; Dice/F1: $p = 0.77$). These results suggest that, on this dataset, attention gates do not provide a measurable advantage over the standard U-Net architecture, establishing both as comparable practical baselines for future flood mapping research.

I. INTRODUCTION

Floods are one of the most frequent and destructive natural disasters worldwide, destroying infrastructure and threatening the lives of communities in various parts of the world. Rapid and accurate identification of flood-prone areas is essential for emergency response teams so that rescue operations, aid distribution, and damage assessment can be carried out effectively. Conventional mapping methods tend to be slow and heavily reliant on manual intervention, which becomes a major obstacle when time is of the essence.

Automating flood area detection from UAV imagery presents unique technical challenges that distinguish it from general-purpose segmentation tasks. Water surfaces under flood conditions exhibit highly variable visual appearances due to specular reflections, turbidity, and the presence of floating debris, making it difficult to define a consistent pixel-level representation of flood regions. Submerged vegetation and waterlogged terrain can produce spectral signatures that closely resemble dry land, particularly when observed from varying flight altitudes and angles. Building shadows and other dark surfaces further complicate segmentation by introducing false negatives, as shadowed pixels may visually resemble water accumulation. Additionally, UAV imagery is inherently subject to illumination variability caused by changing weather conditions and time of capture, which affects the consistency of color and texture features across images. These compounding factors motivate the use of attention-based mechanisms that can selectively focus on

flood-relevant spatial features while suppressing visually ambiguous background regions.

UAV-based air imagery has now evolved into a reliable tool for flood monitoring. Drones can quickly reach affected areas, capture high-resolution images, and operate in conditions where satellite access is limited [1]. However, the volume of visual data produced by UAVs still requires automation because manual processing is highly inefficient. Various approaches have been developed to address this challenge, ranging from the combination of region growing with deep learning [1] to the implementation of lightweight networks on edge computing platforms [2].

Deep learning-based semantic segmentation addresses this issue by automatically classifying each pixel as either a flood area or not. Lee et al. demonstrated that lightweight deep learning models can perform semantic segmentation of aerial images for real-time flood and wildfire monitoring [2]. U-Net, introduced by Ronneberger et al., has become the most widely used architecture for this task due to its encoder-decoder structure and skip connections that preserve spatial information [3]. Oktay et al. then extended this architecture through Attention U-Net, which adds attention gates to the decoder's skip connections to sharpen focus on target regions and suppress background interference [4].

Several previous studies have confirmed the effectiveness of the encoder-decoder model for flood segmentation from aerial images. Rahnemoonfar et al. introduced FloodNet as a high-resolution UAV dataset for post-flood scene understanding, as well as serving as a benchmark for evaluating various segmentation models [5]. Sinha et al. compared U-Net and DeepLabV3 on UAV flood images and found that both achieved competitive performance, with differences depending on the metrics used [6]. Roohi et al. applied U-Net to optical and radar images from several flood events, reporting an accuracy of 88%, which significantly exceeded the baseline fully convolutional network [7]. Melavanki et al. further demonstrated that U-Net can be adapted for end-to-end flood segmentation from aerial images with high overall accuracy [8].

From the attention mechanism perspective, Li et al. proposed a variant of U-Net that integrates a convolutional block attention module and demonstrated improved flood boundary extraction from SAR images compared to standard U-Net [9]. Sunil et al. applied an attention-based segmentation model called SegNet-ATT on aerial images of flood-affected areas and achieved an accuracy of 89.46%, validating the value of combining spatial and channel attention for flood segmentation [10]. Safavi and Rahnemoonfar compared various segmentation architectures on the FloodNet dataset and found that model selection significantly influences performance, especially in the trade-off between accuracy and computational efficiency [11]. Simantiris and Panagiotakis further showed that unsupervised approaches can produce competitive flood segmentation on UAV images, reporting an F1-score of 80.9%, while the two models in this study achieved higher Dice/F1 scores, at 86.74% and 86.88%, respectively, in a supervised setting.

Although research in this field continues to develop, direct controlled comparisons between U-Net and Attention U-Net under identical experimental conditions for binary flood segmentation from UAV images are still rare. This study aims to fill that gap by systematically evaluating whether the addition of attention gates results in a measurable performance improvement under fully equivalent training conditions.

II. METHODS

A. Dataset

This study uses a publicly available dataset of aerial RGB images collected by UAV, originally published on Kaggle under the CC0 Public Domain license [12]. The dataset consists of 290 image-mask pairs capturing flood-affected areas across visually diverse scene types, including urban settlements, roadways, buildings, vegetation, and open water bodies. Each image represents a distinct aerial viewpoint of flood conditions, providing varied spatial contexts for segmentation learning. An example image-mask pair from the dataset is shown in Fig. 1.

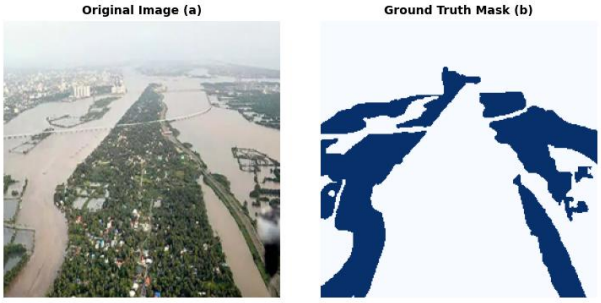


Figure 1. Example of UAV aerial imagery and its corresponding ground truth mask: (a) UAV image, (b) binary flood mask

Ground truth binary masks were annotated at the pixel level using Label Studio, where each pixel is labeled as flood (1) or non-flood (0). The original images were resized to 256×256 pixels and normalized to the range [0,1]. Binary masks were generated by thresholding at pixel value 127. The dataset was split into 203 training images (70%), 29 validation images (10%), and 58 test images (20%) using stratified random sampling with a fixed seed (SEED=42) to maintain a consistent class proportion across all subsets and to ensure reproducibility.

It is acknowledged that the dataset originates from a single publicly available source with no explicit geographic metadata, which limits the diversity of flood conditions represented. Variations in water reflectance, vegetation density, building shadows, and lighting conditions that are typical of real-world UAV deployments are not fully captured within this dataset. This constraint is discussed further in the Limitations section.

B. Data Augmentation

Augmentation is only applied to the training set using the Albumentations library. Techniques used include horizontal flipping (p=0.5), random brightness and contrast adjustments (p=0.2), and affine transformations that include shifts up to 5%, scaling up to 5%, and rotations up to 15 degrees. The validation and test sets are left in their original condition to ensure objective and unbiased evaluation.

C. Model Architecture

Both models adopt an encoder-decoder structure. The U-Net Baseline uses three encoder levels with 64, 128, and 256 filters, a bottleneck with 512 filters, and a symmetric decoder with skip connections. Each level consists of two Conv2D-ReLU blocks, resulting in a total of 7,782,913 trainable parameters. This architecture is illustrated in Fig. 2.

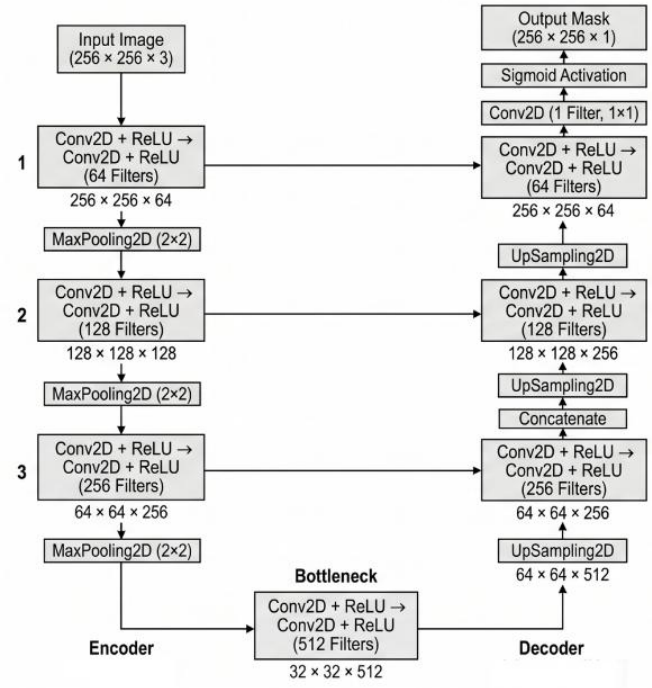


Figure 2. U-Net Baseline architecture

Attention U-Net shares the same backbone structure, with the addition of attention gates at the three decoder skip connections, increasing the number of parameters to 7.912.612. Each attention gate receives skip features and decoder features that have been upsampled as inputs, calculates sigmoid-based spatial coefficients, and then multiplies them back to the skip features before concatenation. This mechanism allows the model to assign more weight to spatial regions relevant to flooding, as illustrated in Fig. 3. Seong and Choi demonstrated that such attention gate mechanisms effectively enhance spatial feature selection in remote sensing image segmentation by weighting based on channel relevance and pixel position relative to the target class [13].

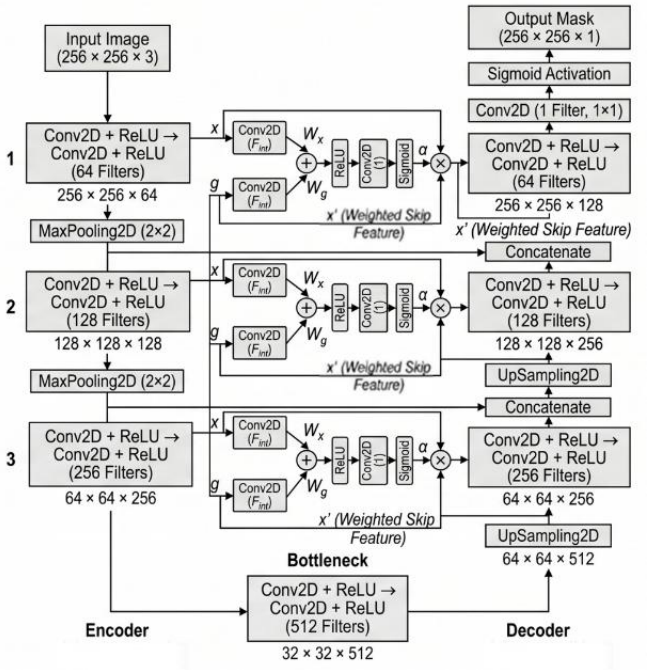


Figure 3. Attention gate mechanism in Attention U-Net

D. Training Configuration

Both models are trained using the Adam optimizer with a learning rate of 1×10^{-4} and a combined loss function of Binary Cross-Entropy and Dice loss. Yeung et al. showed that this combination works well for binary segmentation tasks with class imbalance, which is common in flood datasets where non-flood pixels are usually much more numerous than flood pixels [14]. Training is conducted for a maximum of 50 epochs with a batch size of 4 in float32 precision. Three callbacks are applied: EarlyStopping with a patience of 15 epochs and automatic best weight restoration, ReduceLROnPlateau which halves the learning rate after 5 epochs without improvement down to a minimum of 1×10^{-6}, and ModelCheckpoint to save the weights with the best validation loss. All experiments were conducted on the Kaggle platform using TensorFlow 2.19.0 and Keras, accelerated by dual NVIDIA Tesla T4 GPUs (2×16 GB VRAM). The total number of trainable parameters is 7,782,913 for U-Net Baseline and 7,912,612 for Attention U-Net.

To mitigate the risk of overfitting given the limited dataset size, the regularization strategies above were complemented by data augmentation including random horizontal flipping, brightness and contrast adjustment, and shift-scale-rotate transformations, applied during training to artificially increase sample diversity.

E. Evaluation Metrics

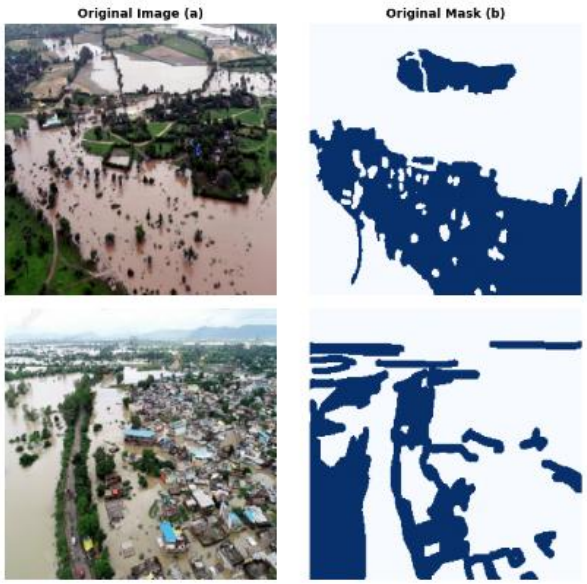
Evaluation is conducted on the test set with a prediction threshold of 0.5 using five pixel-level metrics: Intersection over Union (IoU), Dice/F1 Score, Precision, Recall

(Sensitivity), and Specificity. IoU measures the overlap between the predicted flood mask and the ground truth. Dice/F1 is the harmonic mean of Precision and Recall, providing a balanced measure of segmentation quality. Precision measures the proportion of predicted flood pixels that are correctly identified, reflecting the model's tendency toward false positives. Recall (Sensitivity) measures the proportion of actual flood pixels correctly identified by the model. Specificity measures the proportion of actual non-flood pixels correctly identified as non-flood. To assess robustness beyond a single training run, each model was trained five times using different random seeds (42–46), and a paired t-test was conducted on the IoU and Dice/F1 scores across runs to determine whether differences between architectures were statistically significant.

III. Results And Discussion

A. Data Augmentation Results

Before training, an augmentation pipeline was applied to 203 training images to enhance the diversity of the data distribution. Figure 4 shows examples of the augmentation results on three UAV image samples, along with their ground truth masks. The techniques used include horizontal flipping, random adjustments of brightness and contrast, and affine transformations that encompass shifting, scaling, and rotation within certain limits. As shown in Fig. 4, the augmentation produces variations in orientation and lighting conditions that differ from the original images, while the ground truth masks remain consistent and follow the transformations applied to the images. This approach aims to simulate the variations in UAV image capture conditions in the field, such as different flight angles and light intensities, so that the model does not rely on specific visual patterns during training.



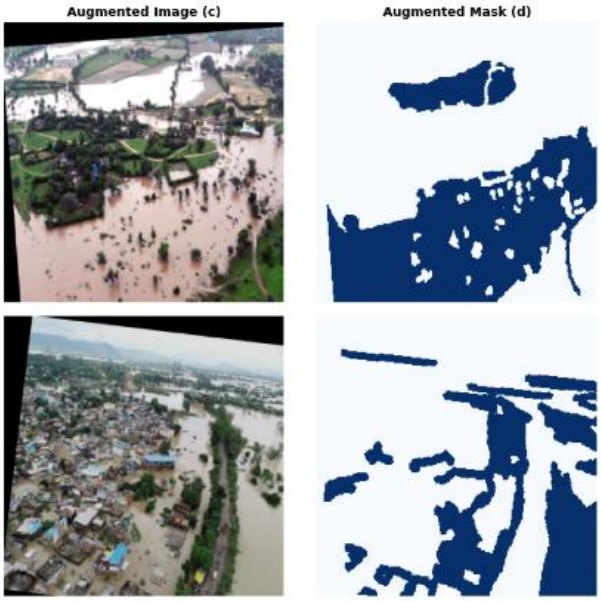


Figure 4. Examples of augmentation results applied to training images: (a) augmented UAV image, (b) corresponding transformed ground truth mask

B. Performance Model

This section presents the quantitative and qualitative evaluation results of both models on the held-out test set of 58 images. To ensure robustness, each model was trained independently five times using different random seeds (42–46); performance is reported as mean ± standard deviation across these five runs, with predictions generated at a threshold of 0.5. Table I summarizes the segmentation performance of the two models on the test set.

TABLE I
SEGMENTATION PERFORMANCE ON THE TEST SET

Metric	U-Net Baseline	Attention U-Net
IoU (%)	76.97 ± 0.54	77.11 ± 0.76
Dice/F1 (%)	86.98 ± 0.34	87.07 ± 0.49
Precision (%)	88.88 ± 0.85	89.25 ± 1.16
Recall (%)	85.19 ± 1.39	85.04 ± 1.89
Specificity (%)	92.34 ± 0.76	92.63 ± 1.03

Across the five runs, Attention U-Net achieved a marginally higher mean IoU and Dice/F1 than U-Net Baseline, along with slightly higher Precision and Specificity, where all values are reported as mean ± standard deviation. However, Attention U-Net also exhibited a marginally lower mean Recall (85.04% ± 1.89) than U-Net Baseline (85.19% ± 1.39), consistent with the precision-recall trade-off typically introduced by attention gating. To determine whether the observed mean differences reflect a genuine architectural effect or random variation, a paired t-test was conducted across the five seeds. The results showed no statistically significant difference between the two architectures (IoU: $t = -0.32$, $p = 0.77$; Dice/F1: $t = -0.31$, $p = 0.77$; $\alpha = 0.05$). This indicates that, despite Attention U-Net's marginally higher

average scores, the addition of attention gates does not provide a measurable performance improvement over the standard U-Net on this dataset.

Fig. 5 displays the training and validation loss curves as well as the Dice scores for the U-Net Baseline over 50 epochs. The model converged smoothly, with the training and validation curves remaining close throughout the process, indicating stable learning without overfitting.

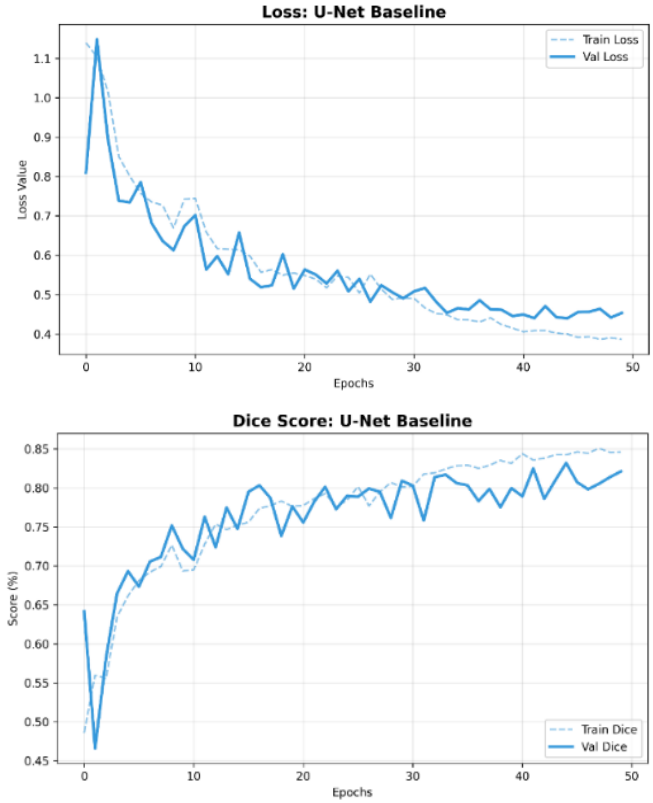
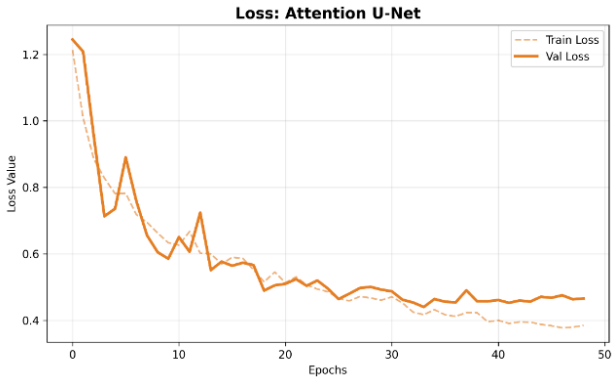


Figure 5. Training and validation curves for U-Net Baseline (representative run, seed 42).

The Attention U-Net shows a similar convergence pattern, as seen in Fig. 6. The loss and Dice score curves for training and validation remain parallel throughout the epochs, confirming that adding attention gates does not disrupt the stability of the learning process.



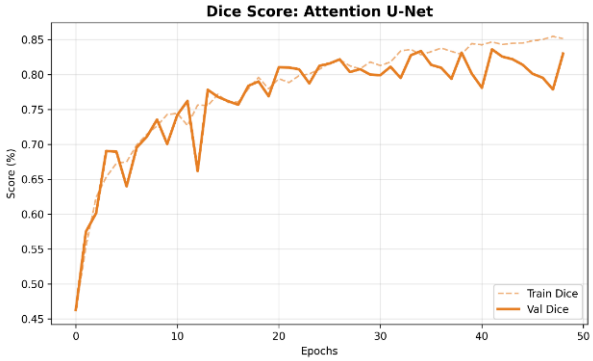


Figure 6. Training and validation curves for Attention U-Net (representative run, seed 42).

Fig. 7 shows the qualitative segmentation results on three representative test samples, comparing the predictions of both models with the ground truth masks. The samples were selected to illustrate varying levels of segmentation difficulty encountered across the test set.In the first sample, both models successfully delineate the main flood extent, producing predictions that closely match the ground truth mask with well-defined boundaries. The attention gates in Attention U-Net appear to suppress background regions more cleanly in this case, resulting in slightly fewer false positive pixels along the flood boundary.In the second sample, both models demonstrate partial success in detecting flood regions but struggle with areas where inundated vegetation produces spectral signatures similar to dry land. In these ambiguous regions, both models tend to under-segment, missing flood pixels at the periphery of the water body. The difference between the two models is minimal in this case, suggesting that attention gates alone are insufficient to resolve this type of ambiguity without additional contextual information.In the third sample, both models correctly identify the dominant flood region but produce fragmented predictions in areas containing building shadows. Shadow regions are visually similar to standing water, leading to false positive detections in both models. This error pattern is consistent across both architectures, indicating that it reflects a fundamental limitation of RGB-based flood segmentation rather than an architectural deficiency specific to either model.

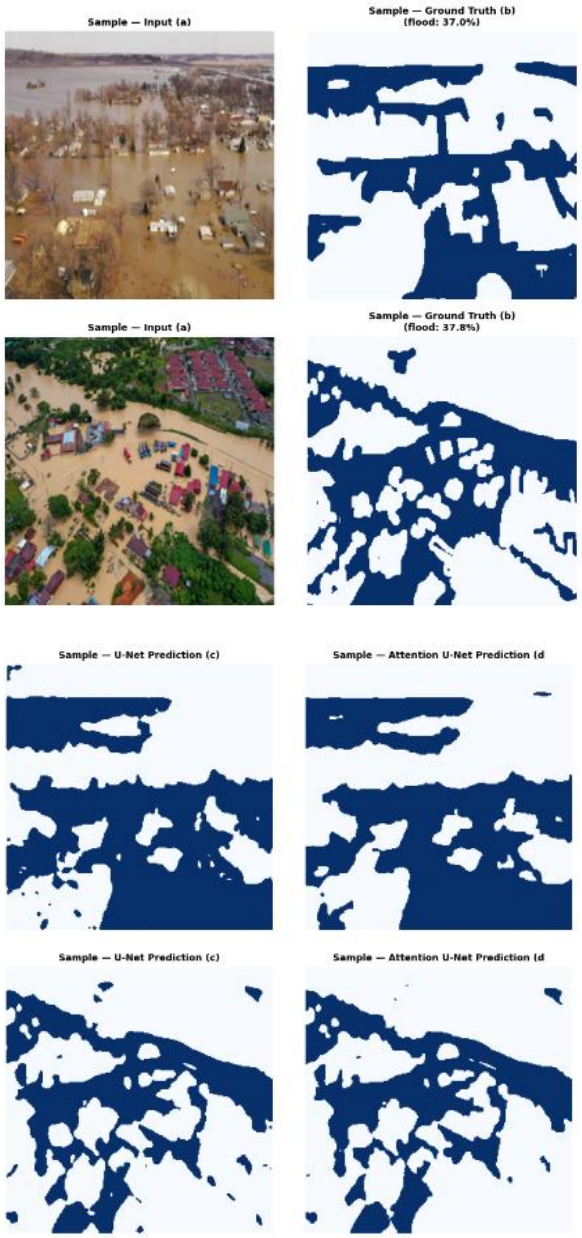


Figure 7. Qualitative results: (a) input, (b) ground truth, (c) U-Net prediction, (d) Attention U-Net prediction.

Fig. 8 presents a qualitative comparison of best-case and worst-case segmentation results from the test set, selected based on per-image IoU scores. In the best-case sample (IoU = 0.95), both models successfully delineate the dominant flood extent in an urban scene with large open water surfaces. The predicted masks closely match the ground truth boundary, with only minor false positive detections in vegetated regions near the flood edge. In the worst-case sample (IoU = 0.10), both models fail to accurately segment a rural scene where flood water covers terrain with complex spatial patterns including roads, vegetation patches, and agricultural land. Both models exhibit significant over-segmentation, classifying large portions of non-flooded land as flood area. This failure is consistent across both architectures and is likely attributable to the visual similarity between turbid floodwater and surrounding soil surfaces, as well as the high spatial complexity of the scene. These error patterns suggest that the primary limitation of both models lies in handling scenes with ambiguous spectral signatures rather than in architectural differences between the two approaches.

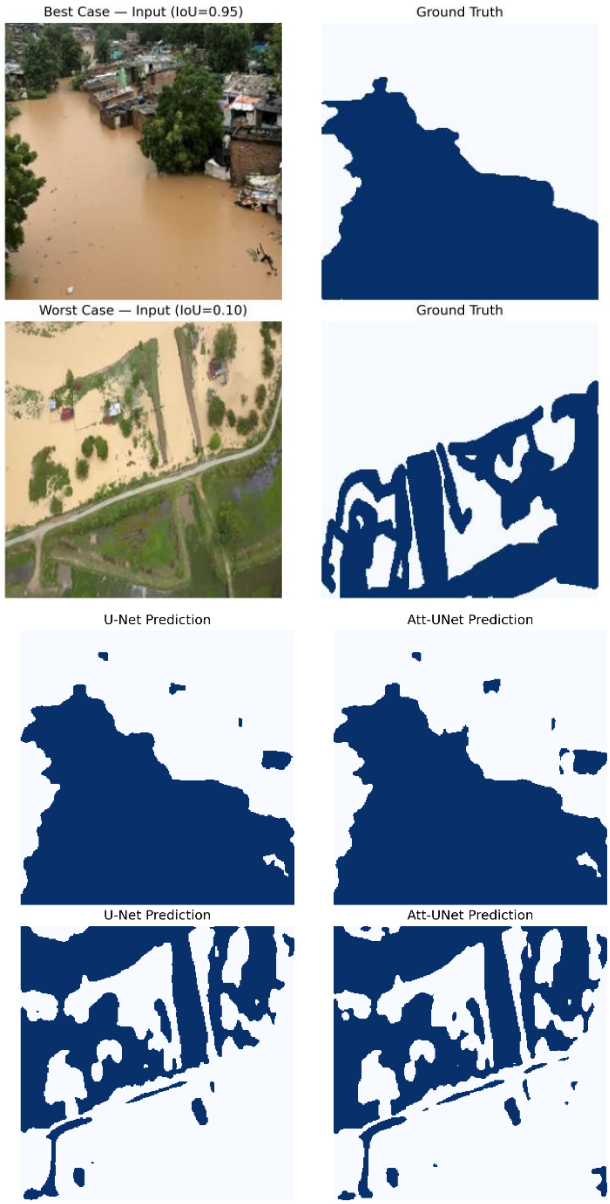


Figure 8. Best-case (top, IoU = 0.95) and worst-case (bottom, IoU = 0.10) segmentation results: (a) input, (b) ground truth, (c) U-Net prediction, (d) Attention U-Net prediction.

These results indicate that, although Attention U-Net consistently achieved marginally higher mean IoU and Dice/F1 scores across five independent runs, the paired t-test confirms that this difference is not statistically significant.

Rather than reflecting a genuine architectural advantage, the small gap between the two models is more plausibly attributable to random variation introduced by weight initialization and stochastic training dynamics. This finding underscores the importance of repeated-run evaluation when comparing closely related architectures with narrow performance margins.

The attention mechanism works by generating a spatial weight map that highlights areas likely to contain water accumulation. This allows the model to suppress areas that visually resemble water but are not relevant, such as building rooftops, roadways, and other bodies of water that are not actual flood boundaries, resulting in a more precise delineation of boundaries. These findings are consistent with Li et al., who showed that the attention mechanism in U-Net improves flood boundary extraction from remote sensing images [9], as well as with Sunil et al., who reported increased accuracy of attention-based segmentation models on flood aerial images [10].

The marginally lower mean Recall of Attention U-Net (85.04% ± 1.89) compared to U-Net Baseline (85.19% ± 1.39) reflects an inherent trade-off between precision and recall. When the model becomes more selective in labeling pixels as flooded, some ambiguous boundary pixels are missed. In the context of real disaster response applications, this trade-off can be acceptable because reducing false positive predictions is operationally more valuable, as limited rescue resources must be directed to truly affected locations [2].

Both models showed stable training curves without signs of overfitting over 50 epochs, which is quite encouraging considering the training set only consists of 203 images. Data augmentation contributed to this stability by increasing the effective diversity of training samples. This aligns with the findings of Hashemi-Beni and Gebrehiwot, who demonstrated that data augmentation significantly improves the accuracy of deep learning models on small UAV flood datasets [1].

Compared to related research, the results of both models are within a competitive range. Sinha et al. reported an IoU of 74.14% for DeepLabV3 on a similar UAV flood segmentation task [6], while Simantiris and Panagiotakis reported an F1-score of 80.9% using an unsupervised approach on UAV flood images [15], whereas both models here achieved higher Dice/F1 scores, at 86.98% and 87.07%, respectively, in a supervised setting. Safavi and Rahnemoonfar also noted that among various architectures tested on the FloodNet benchmark, performance differences between models tend to be close, which aligns with the statistically non-significant difference found between U-Net and Attention U-Net in this study [11]. Şener et al. proposed FASegNet, a lightweight CNN with hybrid attentional atrous convolution, and achieved mIoU of 84.3% on the same Kaggle flood dataset without data augmentation [16], suggesting that task-specific architectural designs may yield higher performance than general-purpose encoder-decoder models on this benchmark. Overall, both models serve as valid baselines for binary flood segmentation from UAV images.

A comprehensive review by Bentivoglio et al. identified encoder-decoder architectures, particularly U-Net variants, as the most widely adopted deep learning approach for flood mapping tasks [17]. More recent studies have begun exploring advanced alternatives. Hernández et al. demonstrated real-time flood detection from UAV images using lightweight segmentation models on edge-computing platforms [18]. Ghosh et al. applied a nested UNet++ architecture to SAR-based flood mapping and reported improved performance over standard UNet [19]. Roy et al. proposed a hybrid CNN-transformer model and demonstrated superior generalization across diverse geographic conditions [20], while Sundaresan and Solomon applied DeepLabV3+ with various backbone architectures to post-disaster UAV imagery from the FloodNet dataset, with ResNet-50 achieving the best balance between segmentation accuracy and computational efficiency [21]. These developments suggest that future work comparing U-Net and Attention U-Net against transformer-based and nested architectures would provide a more comprehensive benchmark to determine whether the advantage of attention gates persists as dataset size and scene diversity increase.

IV. LIMITATIONS

Several limitations of this study should be acknowledged. First, the dataset includes only 290 image-mask pairs from a single publicly available repository with no specific geographic metadata. This relatively small dataset size raises the risk of overfitting, which was addressed through data augmentation and early stopping. However, the effectiveness of these methods on unseen geographic regions has not been verified.

Second, since all images come from one source, the variation in flood conditions, terrain types, and environmental settings in the evaluation is limited. The ability of both models to apply to flood events in different geographic areas, climates, or UAV sensor setups has not yet been evaluated.

Third, this study does not perform cross-validation due to the computational cost associated with repeated full training runs. While five repeated training runs with different random seeds were conducted to assess result stability, k-fold cross-validation would provide a more rigorous estimate of model generalization on a dataset of this size.

Fourth, the comparison is limited to U-Net and Attention U-Net. More recent architectures such as UNet++, DeepLabV3+, SegFormer, and Swin Transformer-based models were not evaluated, and therefore the relative positioning of both models within the broader landscape of modern segmentation methods remains unclear.

Fifth, computational complexity metrics including training time, inference time, GPU memory consumption, and parameter efficiency were not systematically benchmarked, which limits the ability to assess the practical deployment feasibility of both models in real-time UAV monitoring scenarios.

V. CONCLUSION

This study compares the U-Net Baseline and Attention U-Net for binary flood area segmentation from UAV RGB images under identical training conditions, evaluated across five independent runs with different random seeds to assess result robustness. Attention U-Net achieved marginally higher mean IoU (77.11% ± 0.76 vs. 76.97% ± 0.54) and Dice/F1 (87.07% ± 0.49 vs. 86.98% ± 0.34) than the U-Net Baseline, but a paired t-test showed that these differences were not statistically significant ($p = 0.77$ for both metrics). These results indicate that, on this dataset, the addition of attention gates does not yield a measurable performance advantage over the standard U-Net architecture, and that both models perform comparably as practical baselines for automatic flood mapping from UAV images. Both models converged stably without overfitting across all runs. Future research should test these architectures on larger and more geographically diverse datasets, ideally with repeated-run statistical evaluation, to determine whether attention mechanisms provide a measurable benefit as dataset scale and diversity increase.

REFERENCES

[1] 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.

[2] Y. J. Lee, H. G. Jung, and J. K. Suhr, "Semantic Segmentation Network Slimming and Edge Deployment for Real-Time Forest Fire or Flood Monitoring Systems Using Unmanned Aerial Vehicles," *Electronics*, vol. 12, no. 23, p. 4795, Nov. 2023, doi: 10.3390/electronics12234795.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., in Lecture Notes in Computer Science, vol. 9351. , Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[4] O. Oktay *et al.*, "Attention U-Net: Learning Where to Look for the Pancreas," May 20, 2018, *arXiv*: arXiv:1804.03999. doi: 10.48550/arXiv.1804.03999.

[5] M. Rahnemoonfar, T. Chowdhury, A. Sarkar, D. Varshney, M. Yari, and R. R. Murphy, "FloodNet: A High Resolution Aerial Imagery Dataset for Post Flood Scene Understanding," *IEEE Access*, vol. 9, pp. 89644–89654, 2021, doi: 10.1109/ACCESS.2021.3090981.

[6] R. K. Sinha, M. Kushwaha, J. Choudhary, D. P. Singh, and M. Pandey, "Flood image segmentation of uav aerial images using deep learning," in *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, IEEE, 2024, pp. 1–6.

[7] M. Roohi, H. R. Ghafouri, and S. M. Ashrafi, "Advancing flood disaster management: leveraging deep learning and remote sensing technologies," *Acta Geophys.*, vol. 73, no. 1, pp. 557–575, 2025.

[8] N. Melavanki, S. Olekar, S. Biradar, and R. Ganiger, "End-to-End Flood Segmentation in Aerial Images Using U-Net Architecture," in *2025 IEEE 4th International Conference for Advancement in Technology (ICONAT)*, IEEE, 2025, pp. 1–6.

[9] W. Li, J. Wu, H. Chen, Y. Wang, Y. Jia, and G. Gui, "Unet combined with attention mechanism method for extracting flood submerged range," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 15, pp. 6588–6597, 2022.

[10] P. Sunil, S. Sinha, and R. Roy, "SegNet-ATT: cross-channel and spatial attention-enhanced u-net for semantic segmentation of flood affected areas," in *International Conference on Pattern Recognition*, Springer, 2024, pp. 287–300.

[11] F. Safavi and M. Rahnemoonfar, "Comparative study of real-time semantic segmentation networks in aerial images during flooding events," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 16, pp. 4–20, 2022.

[12] F. Karim, "Flood Area Segmentation." Kaggle, 2022. [Online]. Available: https://www.kaggle.com/datasets/faizalkarim/flood-area-segmentation

[13] S. Seong and J. Choi, "Semantic segmentation of urban buildings using a high-resolution network (HRNet) with channel and spatial attention gates," *Remote Sens.*, vol. 13, no. 16, p. 3087, 2021.

[14] M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, "Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation," *Comput. Med. Imaging Graph.*, vol. 95, p. 102026, 2022.

[15] G. Simantiris and C. Panagiotakis, "Unsupervised color-based flood segmentation in UAV imagery," *Remote Sens.*, vol. 16, no. 12, p. 2126, 2024.

[16] A. Şener, G. Doğan, and B. Ergen, "A novel convolutional neural network model with hybrid attentional atrous convolution module for detecting the areas affected by the flood," *Earth Sci. Inform.*, vol. 17, no. 1, pp. 193–209, Feb. 2024, doi: 10.1007/s12145-023-01155-9.

[17] R. Bentivoglio, E. Isufi, S. N. Jonkman, and R. Taormina, "Deep learning methods for flood mapping: a review of existing applications and future research directions," *Hydrol. Earth Syst. Sci.*, vol. 26, no. 16, pp. 4345–4378, Aug. 2022, doi: 10.5194/hess-26-4345-2022.

[18] D. Hernández, J. M. Cecilia, J.-C. Cano, and C. T. Calafate, "Flood Detection Using Real-Time Image Segmentation from Unmanned Aerial Vehicles on Edge-Computing Platform," *Remote Sens.*, vol. 14, no. 1, p. 223, Jan. 2022, doi: 10.3390/rs14010223.

[19] B. Ghosh, S. Garg, M. Motagh, and S. Martinis, "Automatic Flood Detection from Sentinel-1 Data Using a Nested UNet Model and a NASA Benchmark Dataset," *PFG – J. Photogramm. Remote Sens. Geoinformation Sci.*, vol. 92, no. 1, pp. 1–18, Mar. 2024, doi: 10.1007/s41064-024-00275-1.

[20] A. M. R, R. Roy, S. S. Kulkarni, V. Soni, and A. Chittora, "Transformer-based Flood Scene Segmentation for Developing Countries," Oct. 09, 2022, *arXiv*: arXiv:2210.04218. doi: 10.48550/arXiv.2210.04218.

[21] A. A. Sundaresan and A. A. Solomon, "Post-disaster flooded region segmentation using DeepLabv3+ and unmanned aerial system imagery," *Nat. Hazards Res.*, vol. 5, no. 2, pp. 363–371, Jun. 2025, doi: 10.1016/j.nhres.2024.12.003.