International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 642-650, May-2026

https://doi.org/10.47001/IRJIET/2026.105087

Modified U-Net with Dilated Convolution for Flood Water Segmentation

[1]Nitesh Singh, [2]Prakash Chandra Prasad, [3]Anku Jaiswal

[1]Department of Electronics and Computer Engineering, Pulchowk Campus (IOE,Tribhuvan University), Kathmandu, Nepal
[2,3]Assistant Professor, Department of Electronics and Computer Engineering, Pulchowk Campus (IOE, Tribhuvan University), Kathmandu, Nepal

***Abstract* - Flood is one of the natural calamities that presents major threats to human civilization, infrastructure and environment, which requires timely and accurate detection and monitoring to mitigate its impact. Recent development in deep learning has significantly improved semantic segmentation performance. However, conventional convolutional neural networks often struggle to capture large contextual information while retaining smaller spatial details, which are necessary for precise flood water segmentation. This study proposes a Modified U-Net architecture containing dilated convolutional layers to improve flood water segmentation performance. The proposed model includes dilation rates of 2, 4, and 6 within the encoder and bottleneck of U-Net to increase the receptive field without increasing the number of parameters or losing spatial resolution. The proposed models get trained on the training dataset using a combined loss of Binary cross entropy and Dice loss function with AdamW optimizer, validated on a validation dataset and get tested on an unseen separated test dataset using multiple quantitative metrics. Thus, obtained experimental results shows that the Modified U-Net with a dilation rate 2 achieved the improved results overall, with a test Dice coefficient of 0.8839, mean IoU of 0.8331, precision of 0.9013, recall of 0.8709, and accuracy of 0.9367. This improvement shows that an optimal dilation rate balances global context extraction and boundary localization effectively, which is critical for accurate flood water segmentation. The findings of this study highlight the effectiveness of dilated convolutions in improving semantic segmentation performance for flood monitoring applications.**

Keywords: Flood segmentation, U-Net, dilated convolution, semantic segmentation, deep learning.

I. INTRODUCTION

Floods are one of the most destructive natural disasters that affects billions of people and cause huge environmental and socioeconomic harm worldwide annually. They account for roughly half of all weather-related disasters [1]. Flooding is a recurrent natural calamity that is the result of unmanaged urbanization, poor land-use practices, and climate change [2]. Climate resilience, resource allocation, and disaster response all depend on the prompt and accurate segmentation of flooded areas [1]. In order to map flooded areas and enable prompt reaction and mitigation measures, flood water segmentation, the technique of identifying water bodies in photos is essential for disaster management [3]. Conventional flood detection techniques frequently lack accuracy and scalability, especially in complicated situations with different backdrops and water textures [4].

Deep learning models, mainly Convolutional Neural Networks (CNNs), have been found effective in semantic segmentation tasks which have enabled automatic mapping of flood from images [5]. U-Net is an encoder-decoder based model with skip connection placed between encoder and decoder layer, is a popular choice in flood mapping due to its simplicity and better performance on small datasets [6][7]. Though, the traditional U-Net has limited spatial awareness due to its smaller receptive field, hindering performance in complex scenes with variable flood patterns [7].

To address these limitations, researchers have explored advanced semantic segmentation architectures with improved contextual feature extraction capabilities. Models such as DeepLabv3 incorporate atrous or dilated convolutions to enlarge the receptive field without reducing spatial resolution or significantly increasing the number of parameters [10]. These architectures have demonstrated strong performance in various segmentation tasks because of their ability to capture multi-scale contextual information more effectively [11]. However, these architectures may not always be suitable for medium-sized datasets or resource constrained environments as they involve higher computational complexity [12].Dilated convolution also known as atrous convolution, has evolved as an effective technique for improving contextual feature extraction in convolutional neural networks. Dilated convolution inserts spaces between kernel elements unlike the standard convolution, allowing the network to cover a larger receptive field while preserving feature map resolution [13][11]. This property makes dilated convolution particularly

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 642-650, May-2026
https://doi.org/10.47001/IRJIET/2026.105087

suitable for flood segmentation tasks, where both global contextual understanding and accurate boundary localization are essential for reliable segmentation performance [3][14].

Therefore, this study proposes a modified U-Net architecture incorporating dilated convolutional layers within the encoder and bottleneck blocks to improve segmentation of contiguous flood regions while maintaining computational efficiency. Different dilation rates are investigated to analyze their impact on segmentation accuracy and contextual feature learning [14]. The proposed approach aims to generate more accurate and robust flood masks that may contribute to rapid disaster response systems and automated flood monitoring applications.

1.1 Statement of the problem

1. Standard U-Net architectures face difficulty in segmenting large or subtle water regions, particularly in high-resolution images or noisy environments because of the following:
   (a) A narrow receptive field in U-Net.
   (b) Loss of spatial context due to pooling operations.
2. Existing deep learning models like DeepLabv3 shows promising results but introduces higher computational complexity and larger parameter counts.

1.2 Research objectives

1. To design and implement a modified U-Net architecture enhanced with dilated convolution layers.
2. To train and evaluate the model on a dataset of flood imagery with annotated masks.
3. To compare the proposed model's performance with Standard U-Net
4. To analyze the impact of dilated convolutions on capturing multi-scale flood features.
5. To use metrics like IoU, mIoU, Dice, and F1-score to quantify improvements.

II. THEORETICAL BACKGROUND

2.1 Architectural Evolution and the U-Net Model

The field of semantic segmentation advanced significantly with the introduction of Fully Convolutional Networks (FCNs). Unlike earlier models that used fixed-size inputs and fully connected layers, FCNs replaced these with convolutional layers to produce pixel-level prediction maps from images of any size [8]. However, FCNs often resulted in imprecise boundaries due to the loss of spatial information from repeated down-sampling and pooling operations. The FCN's method of integrating features by adding pixel values also proved less effective than other methods for retaining positional information [9]. The U-Net architecture was developed to address these shortcomings [9]. U-Net, introduced by Ronneberger et al. (2015) [6], revolutionized semantic segmentation with its encoder-decoder structure and skip connections, enabling precise boundary detection even with limited training data. The encoder captures high-level semantic features, while the decoder restores the image resolution. The defining feature of U-Net is its "skip connections" that directly link high-resolution feature maps from the encoder to the corresponding layers in the decoder. This mechanism, which uses concatenation rather than addition to integrate features, allows the network to recover and leverage spatial information, leading to highly accurate, pixel-level segmentations with sharp boundaries [7]. U-Net's efficiency with limited training data has made it a popular choice for specialized segmentation tasks. However, standard U-Net architectures can still struggle with the multi-scale nature of flood scenes, where both large bodies of water and small submerged objects must be accurately identified. The fundamental problem lies in the trade-off between increasing the receptive field for broad context and preserving resolution for fine-grained details [7].

2.2 Advanced Architectural Modifications

Dilated convolutions, or atrous convolution, proposed by Yu and Koltun (2016) [13], provides a solution to the trade-off between receptive field size and feature map resolution. This technique is widely employed in computer vision tasks like semantic segmentation and object detection [13]. By introducing a "dilation rate" that defines the spacing between kernel values, a dilated convolution can expand the receptive field exponentially without increasing the number of parameters or losing resolution which allows the network to aggregate multi scale contextual information more effectively [11]. A key challenge with dilated convolutions is the "gridding effect" that occurs with consecutive layers using the same dilation rate, which can lead to a loss of information by causing the receptive field to sample the input at regular, non-continuous intervals. The Hybrid Dilated Convolution (HDC) principle mitigates this by using a "jagged" pattern of dilation rates (e.g., 1, 2, 3) to ensure a continuous receptive field, enabling the capture of both near and far information simultaneously [11]. Chen et al. (2018) [10], integrated dilated convolutions into DeepLabv3+ with Atrous Spatial Pyramid Pooling (ASPP) module, which probes convolutional features at multiple scales to capture global context. The superior performance of DeepLabv3 over standard U-Net on flood segmentation tasks is a strong empirical testament to the power of these dedicated multi-scale modules [9]. Modified U-Nets incorporating dilated layers have shown superior performance. Piao and Liu (2019) [15], enhanced U-Net with dilated convolutions for satellite image segmentation, achieving a mean Intersection over Union (mIoU) of 63.72%

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 642-650, May-2026
https://doi.org/10.47001/IRJIET/2026.105087

on the DeepGlobe dataset by replacing pooling operations to preserve resolution. Yang et al. (2024) [16], applied a similar approach to seismic facies interpretation, reaching 96% accuracy on F3 block data through dilated modules and spatial pyramid pooling. Similarly, Bahrami et al. (2024) reported that models with larger effective receptive fields (e.g. SegNet) had slightly higher precision than U-Net on urban flood data [7]. These studies highlight that dilated convolutional modules can help segment broader contiguous water regions.

2.3 Comparative Analysis of Models for Water Segmentation

In recent years, deep learning has revolutionized the domain of semantic segmentation, especially in the context of natural disaster monitoring. Several studies have explored the effectiveness of various models for water segmentation in flood scenarios. The study by Mou et al. (2025) [9] conducted a comparative evaluation of U-Net, ResNet, and DeepLabv3 on a flood image dataset. It was found that DeepLabv3, which incorporates atrous (dilated) con volution and Atrous Spatial Pyramid Pooling (ASPP), significantly outperformed U-Net in terms of accuracy (90.57% vs. 87.12%). This highlighted U-Net's limited ability to capture large contextual information, particularly in complex flood environments. Further, Bahrami et al. (2024) [7] evaluated SegNet, U-Net, and FCN32 for flood detection and found that SegNet achieved slightly higher precision (88%) than U-Net, attributing U Net's underperformance to its restricted receptive field and sensitivity to noisy or occluded data.

In another domain, Yang et al. (2024) [16] implemented dilated convolution within a U-Net framework for seismic facies segmentation and achieved an impressive 96% classification accuracy. This result emphasized the effectiveness of dilated convolutions in improving spatial awareness while maintaining computational efficiency. Complementing these findings, Piao and Liu (2019) [15] proposed a Deep Dilated U-Net enhanced with a Parallel Dilated Convolution Module designed specifically for high-resolution satellite image segmentation. Their model, evaluated on the DeepGlobe road extraction dataset, showed a significant performance boost improving the mean Intersection-over-Union (mIoU) from 58.29% in standard U-Net to 63.72% using the parallel dilation block. This work demonstrated that inserting a multi-scale receptive block at the bottleneck of U-Net significantly enhances its segmentation performance by aggregating rich semantic and spatial features. The ultimate value of a successful segmentation model lies in its ability to enable quantitative, actionable analysis. A study by Pally and Samadi [17], demonstrates a pipeline that uses segmentation masks as a key enabling technology for flood depth estimation. Their approach uses a model like Mask R-CNN to generate a segmentation mask for the water body, which is then refined with classical computer vision techniques like Canny Edge Detection and aspect ratio to classify the flood water level. This methodology elevates the importance of the core segmentation task, as the quality of the final flood depth estimate is directly dependent on the precision of the initial segmentation mask.

Collectively, these studies suggest that incorporating dilated convolutions into U-Net can improve its ability to generalize in complex, high-resolution, and spatially diverse environ ments such as flood-prone regions. Therefore, leveraging these design enhancements in flood water segmentation presents a promising direction for this research.

2.4 Loss Functions

In this study, combined loss functions incorporating Binary Cross Entropy (BCE) and Dice Loss in the ratio of 0.3:0.7 were used to train the segmentation model. These losses were selected due to their effectiveness in handling the class imbalance, which is common in image segmentation.

Dice Loss

Dice Loss is derived from the Dice coefficient, a metric that measures the overlap between predicted and ground truth segmentation masks. It is especially effective in scenarios with class imbalance, where foreground regions (e.g., water) occupy a small portion of the image.

The Dice coefficient is defined as:

$$Dice\ Loss\ =\ 1 - \frac{2 * |\ Prediction \cap\ Ground\ Truth\ |}{|Prediction| + |Ground\ Truth|}$$

This formulation emphasizes spatial overlap and is less sensitive to class imbalance than pixel-wise losses.

Binary Cross Entropy (BCE)

BCE Loss measures the dissimilarity between predicted probabilities and true binary labels on a pixel-wise basis. It is based on the Bernoulli distribution and works well when class distributions are balanced.

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}[t_i log(p_i) + (1 - t_i) log(1 - p_i)$$

where p denotes predicted mask and t denotes ground truth mask.

The total combined loss function is then computed as:

Combined Loss = 0.3*BCE Loss + 0.7* Dice Loss

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 642-650, May-2026

https://doi.org/10.47001/IRJIET/2026.105087

This combined loss helps the model learn both global shape structures through Dice and fine-grained pixel accuracy through BCE, providing a more robust optimization strategy.

2.5 Evaluation Metrics

To quantitatively assess the performance of the image segmentation model, several evaluation metrics were used. These metrics help evaluate different aspects such as pixel-wise accuracy, overlap, and boundary precision.

1. Intersection-over-Union (IoU)

Intersection over Union evaluates the overlap between predicted segmentation and the ground truth. It is calculated as;

$$IoU = \frac{(Prediction \cap GroundTruth)}{(Prediction \cup GroundTruth)}$$

2. Mean IoU (mIoU)

Mean IoU represents the average IoU score across all classes and is defined as;

$$mIoU = \frac{1}{N}\sum_{i=1}^{N} IoU_i$$

where N is the number of classes.

mIoU is a widely used evaluation metric in semantic segmentation.

3. Accuracy

Accuracy measures the proportion of correctly classified pixels out of the total number of pixels

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

4. F1 Score

The F1 score is calculated by taking the harmonic mean of recall and precision. Both false positives and false negatives are considered in this fair metric.

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

5. Precision

Precision indicates how many of the pixels predicted as positive (e.g., water regions) are correctly classified

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

6. Recall

Recall, also known as sensitivity or true positive rate, The ratio of true positive pre dictions to the total of true positive and false negative predictions. It measures the ability of the model to find all relevant instances.

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

7. Validation Loss

This metric indicates how well the model performs on the validation dataset. Lower values are better and suggest that the model is not overfitting and is generalizing well.

III. METHODOLOGY

3.1 Dataset Collection

This study utilizes three different publicly available datasets containing the actual flood area images and their corresponding mask images collected from the online source Kaggle. The first dataset compromises of 290 images of areas affected by flooding and their corresponding masks. The second datasets consist of 663 flood images and its corresponding masks. Similarly, the third datasets consist of 585 images containing both flooded and non-flooded regions along with its corresponding masks. Each mask image in the first and second datasets is a binary representation of the actual flooded and non flooded area of the image. The Pixel representing water bodies are assigned the value 1 (white) and the region without water is assigned the value 0 (black). Similarly, the third datasets have color coded masks of the actual image. These combination of datasets ensured the diversity in data distribution.

3.2 Dataset Preprocessing

Aconsistent data preprocessing pipeline was applied across all datasets to ensure consistency in datasets and to enhance generalization capability. The preprocessing steps are described below;

a) Mask Conversion

Initially, the third datasets containing the colored mask is converted to binary by initially converting it form BGR to RGB format, and matching the pixels against a predefined list of target flood colors within a tolerance range. All matched pixels are assigned the foreground class, while all other pixels are treated as background. The resulting binary mask is saved in grayscale format. This conversion step is crucial because the segmentation model is trained as a binary classifier at the pixel level.

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 10, Issue 5, pp 642-650, May-2026
https://doi.org/10.47001/IRJIET/2026.105087

b) Image and Mask Resizing

All input images and their corresponding ground truth segmentation masks were resized to a fixed resolution of 256 x 256 pixels. This standardization ensured uniform input size to the model and preserved the structural integrity of water regions across the datasets.

c) Normalization

All the images are first converted from BGR to RGB format. Then, the RGB image pixel values were normalized to the range [0, 1] by dividing each value by 255. This normalization accelerated model convergence and improved numerical stability during training.

d) Dataset Splitting

The whole dataset is split into training, validation, and test subsets using a randomized splitting technique in order to assess model performance and avoid overfitting. The 70% of the data are split as training data which are used for training to learn model parameters, 15% of the data is split as validation data which are used for validation to monitor performance and prevent overfitting, and the remaining 15% of the data split astesting data are reserved as a test set for final unbiased evaluation, ensuring that the model does not encounter test samples during training or hyperparameter tuning. The process was repeated with a fixed random seed to ensure reproducibility.

e) Data Augmentation

We utilized various data augmentation techniques, like rotation, horizontal and vertical flipping, brightness and contrast adjustments on the training datasets to enhance the variability of the dataset. These augmentations improve the diversity of training data, enabling our deep learning model to perform more effectively and adapt to a wider range of scenarios and enhance model robustness and generalization. Geometric transformations were applied jointly to images and masks to preserve spatial correspondence. Augmentation significantly increases the diversity of training samples and reduces overfitting.

3.3 Model Architecture

All the four models for this study contain the symmetric encoder decoder U-Net architecture with skip connections. Each model contains four encoder blocks, a bottleneck block, and four decoder blocks. The first encoder block contains the base filter of size 32 which doubles after each block and reaches the filter size of 512 in the bottleneck block. In the decoder block, the filter size reverses back to 32 in the final decoder block, and the decoder block restores the spatial resolution of the images back to its shapes.

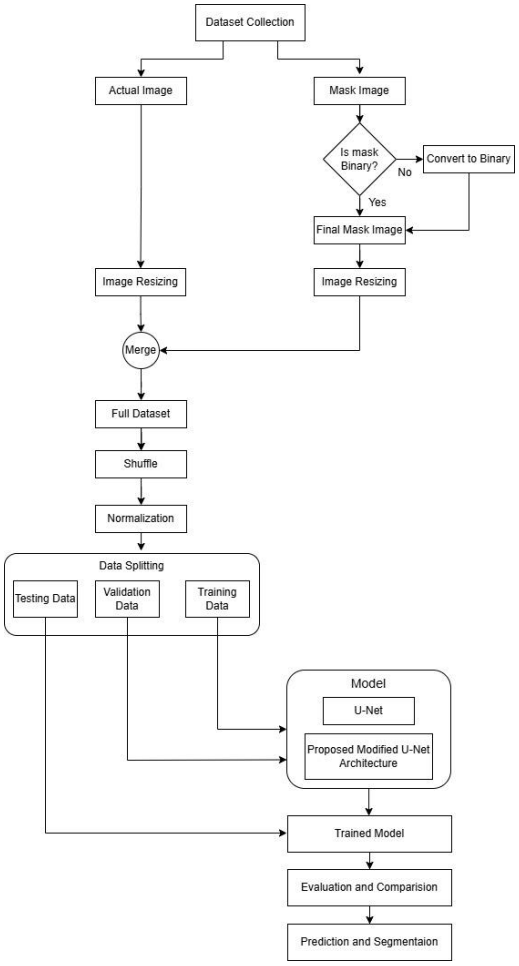

Figure 1: Flowchart of the Model

The skip connections help to preserve fine grained spatial details that would otherwise be lost during downsampling. All the convolutional layers use the ReLu activation function, and in the output layer the sigmoid activation function has been used.

The four models differ only in the dilation rate applied to the convolutional layers within each encoder block and the bottleneck. We increased the dilation rates in each encoder and bottleneck block of each model with the view to increase the effective receptive field of each filter without increasing the number of parameters, enabling the model to capture contextual information at different spatial scales.

The four different models one standard U-Net and three other modified U-Net with different dilation rates in encoder and bottleneck block were trained independently under identical conditions to ensure fair comparison between all the models. All the models use the same training and validation datasets for training, and training was conducted for the maximum upto 200 epochs with batch size of 12. All the

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 642-650, May-2026
https://doi.org/10.47001/IRJIET/2026.105087

images and its corresponding masks were resized to of 256 x 256 pixels. The loss function with a combination of Binary Cross Entropy and Dice Coefficient with a weightage of 0.3 BCE to 0.7 Dice coefficient was adopted during training. This combined BCE-Dice Loss leverages the strengths of both pixel level and region level optimization. BCE loss focuses on pixel wise similarity, while the Dice loss focuses on region overlap and is less sensitive to class imbalance.

AdamW optimizer was used with a learning rate of 3e-4 and weight decay of 1e-4. Three key regularization techniques early stopping, model checkpointing and ReduceLROnPlateau were employed in the all model. Early stopping helps to stop the training process and restores the best weights when the validation loss stops improving. The model checkpointing helps to save the best performing model weights based on the validation loss. Similarly, ReduceLROnPlateau decreases the learning rate when the validation loss stagnates, which helps refine learning in later epochs. These strategies prevent overfitting and ensure that the final models represent the best learned parameters even if the later epochs become unstable. Similarly, Data augmentation techniques like rotation, horizontal and vertical flip, brightness and contrast adjustments was applied only to the training datasets to enhance the variability of the dataset.

IV. RESULTS AND DISCUSSIONS

4.1 Training Performance

The proposed models having different dilation rates in encoder and bottleneck block were trained separately under identical conditions to analyze the impact of dilated convolutions on flood water segmentation. All the four model were compared: one standard U-Net as a baseline model and three dilated U-Net variants with dilation rates of 2, 4, and 6. All the four models used the same preprocessing pipeline, data split, optimizer, loss function, and evaluation metrics to ensure a fair comparison.

The main evaluation criteria were Dice coefficient and Mean IoU, because these metrics better reflects segmentation quality than simple pixel accuracy. Validation loss was also monitored during training to determine the best checkpoint.

The results of each metrics for all the four models on training datasets are shown in the Table 1.

Table 1: Model Comparison on different metrics for different dilation rates on training datasets

Model	Standard U-Net	U-Net Dilation 2	U-Net Dilation 4	U-Net Dilation 6
Best Val Loss	0.159926	0.151001	0.171032	0.188527
Loss Epoch	89	146	73	35
Best Val Dice	0.877215	0.884963	0.869853	0.856261
Dice Epoch	89	134	74	35
Best Val IoU	0.821977	0.828712	0.809284	0.795956
IoU Epoch	89	115	78	36
Precision	0.916675	0.90248	0.933537	0.914396
F1 Score	0.877215	0.884963	0.869853	0.856261
Recall	0.999994	0.999997	1	0.999994
Accuracy	0.928865	0.932409	0.924675	0.916895

The experimental results show that the proposed model achieved high precision and high recall values, with recall values being nearly perfect. This behavior indicates that the model successfully identified most flood pixels while maintaining relatively few false positive predictions. The higher recall suggests that the model tends to prioritize detecting flood regions more aggressively, which is beneficial in flood monitoring applications where missing flooded areas can have serious consequences.

The perfect recall metrics values during training (≈ 1.0) across all the four models are due to two main factors. Firstly, the combined loss function used during training assigns a weight of 0.7 to the Dice loss, which optimizes for foreground overlap and penalizes missed flood pixels false negatives more heavily than false positives. This biases the model to detect every flood pixel during training. Secondly, the data augmentation applied only to the training set, means the model was exposed to a richer and more varied representation of flood patterns during training than during testing which contributes to higher recall values.

4.2 Test Performance

The trained proposed standard U-Net model along with dilated U-Net model was evaluated on the kept aside test datasets. The evaluation considered all the same metrics used during training to provide the comprehensive assessment of the model's effectiveness.

Table 2 summarizes the performance across metrics for all the four model. U-Net model having dilation rate 2 achieved the best results for each metrics. The testing loss of 0.1478 was lowest among all the models, rest all metrics results, Dice coefficient of 0.8839, Mean IoU of 0.8331, Precision of 0.9013, F1 Score of 0.8839, Recall of 0.8709 and Accuracy of 0.9367 was highest among all the four tested model. These best test results for U-Net model with dilation rate 2 even on the test datasets confirms that the best model obtained during training was not due to overfitting, rather it defines the generality of model.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 10, Issue 5, pp 642-650, May-2026

https://doi.org/10.47001/IRJIET/2026.105087

Standard U-Net despite having simpler design and having no dilated convolution achieved competitive results as compared to dilated models and is placed in second place across metrics among the four models. Model having higher dilation rate of 4 and 6 performed worse than the other two model due to the loss of local spatial context. As the dilation rate increases, the receptive field grows substantially, which makes the model to focus on global patterns while ignoring the fine flood boundaries which are essential for accurate segmentation. Overall, the modified U-Net having dilation rate 2 demonstrated the best balance between spatial context and local details and performed better on both datasets and provide best results across metrics.

Table 2: Model comparison on test datasets

Model	Standard U-Net	U-Net Dilation 2	U-Net Dilation 4	U-Net Dilation 6
Loss	0.161435	0.14778	0.179687	0.195917
Dice Coefficient	0.874185	0.8839	0.858492	0.843607
Mean IoU	0.822306	0.833137	0.803009	0.779127
IoU	0.778597	0.793943	0.755587	0.732957
Precision	0.896165	0.90128	0.878312	0.869408
F1 Score	0.874185	0.8839	0.858492	0.843607
Recall	0.858067	0.870859	0.84517	0.827102
Accuracy	0.931483	0.936692	0.922279	0.91531

The drop in recall values from approximately 1.0 on the training set to 0.87 on the test set is as expected and reflects the natural generalization gap rather than data leakage while training of the model. The test recall of 0.8709 achieved by the model having dilation rate 2 remains competitive with the other models and proves that the model generalizes well to unseen data.

4.3 Comparison with SOTA Models

A direct numerical comparison with state of the art methods as mentioned in the literature review section such as DeepLabv3, SegNet, Standard U-Net, ResNet is constrained by different dataset size. This study mainly focuses on the effect of dilation rate under similar experimental conditions for varying dilation rate in the standard U-Net Model. Though, in order to compare the proposed model results against the results of the similar works done in other studies, Table 3 shows a qualitative comparison of reported metrics from closely related works on flood water segmentation tasks to validate our results.

Table 3: Comparison of different test metrics against published benchmarks with our proposed model metrics

Model	Dataset	Dice / F1	mIoU	Precision	Recall	Accuracy
Std. U-Net (Mou et al.)	290 images	0.8361	—	0.82	0.8414	0.8712
DeepLabv3 (Mou et al.)	290 images	0.8749	—	0.884	0.8707	0.9057
ResNet (Mou et al.)	290 images	0.851	—	0.867	0.846	0.887
SegNet (Bahrami et al.)	290 images	0.84	—	0.88	0.8	—
Proposed (Dil-2 U-Net)	1538 images	0.8839	0.8831	0.9013	0.8709	0.9367

While dataset sizes and compositions differ from the other studies, the proposed model achieves superior Dice, precision, Recall and Accuracy metrics results even without pretrained encoders, suggesting the effectiveness of the dilation 2 modification as evaluated in this study.

V. CONCLUSION

This work investigates the impact of dilation rates on the standard U-Net model for flood water segmentation. The study shows that the proposed dilated convolution U-Net model having dilation rate 2 provided the improved segmentation results across all metrics than the standard U-Net model and models having higher dilation rate on both training and test datasets. The model with dilation rate 2 achieved the Dice Coefficient of 0.8839, a Mean IoU of 0.8331, Precision of 0.9013, a Recall of 0.8709 and Accuracy of 0.9367. This result confirms that moderate dilation rates helped to expand the receptive fields without losing structural details in the images and thus improving the segmentation results. Also, the study highlights the importance of balancing contextual information and spatial detail in segmentation tasks.

Model with dilation rate 4, 6 converged faster during training but provided lower results compared to dilation rate 2 model because though higher dilation rate expanded the receptive field by enabling the model to capture multiscale features it came at the cost of losing fine grained spatial information that are necessary for the detection of edges of the flood pixels, which is why only model with moderate dilation improved the segmentation results as compared to the higher dilation rates. This study helped to conclude that only increasing the receptive field does not necessarily improve the results.

In summary, the dilated U-Net model having dilation rate 2 provides an accurate, efficient, and generalizable solution for automated flood water segmentation from multi source imagery.

VI. FUTURE RECOMMENDATIONS

While this study demonstrates the improved results than the standard U-Net model after the use of moderate dilation, there are several key areas for future improvement and expansion. First, the models were evaluated on only 1538 images and its corresponding mask, it seemed not enough for the class imbalanced flood images to evaluate the dilated models. Future works should focus on datasets expansion through adding flood images and making more diverse datasets to improve the dilated models ability to generalize in case of complex scenes.

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 642-650, May-2026
https://doi.org/10.47001/IRJIET/2026.105087

Another important area for future development is the incorporation of hybrid dilation rate in the same model such as lower dilation rates in the initial encoder block to preserve the local features at initial stages and higher dilation rates in the deeper encoder and bottleneck block to capture the broader contextual feature extraction. Also, the use of hybrid parallel dilation rate in the bottleneck block can be explored.

Additionally, advanced techniques like attention mechanism can be implemented in the model which might helps to improve the metrics results. Attention gates in the skip connection between encoder and decoder block could possibly improve flood segmentation.

REFERENCES

[1] Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., ... & Slayback, D. A. (2021). Satellite imaging reveals increased proportion of population exposed to floods. *Nature*, *596*(7870), 80-86.

[2] Alfieri, L., Bisselink, B., Dottori, F., Naumann, G., de Roo, A., Salamon, P., ... & Feyen, L. (2017). Global projections of river flood risk in a warmer world. *Earth's Future*, *5*(2), 171-182.

[3] Derrick Bonafilia, Beth Tellman, Tyler Anderson, and Erica Issenberg. Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 210–211, 2020.

[4] Marco Chini, Renaud Hostache, Laura Giustarini, and Patrick Matgen. A hierarchical split-based approach for parametric thresholding of sar images: Flood inundation as a test case. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12):6975–6988, 2017.

[5] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[6] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *In International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. *Springer*, 2015.

[7] Behrokh Bahrami and Homayoun Arbabkhah. Enhanced flood detection through precise water segmentation using advanced deep learning models. *J. Civ. Eng. Res*, 6(1):1–8, 2024.

[8] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[9] Sanjida Afrin Mou, Tasfia Noor Chowdhury, Adib Ibn Mannan, Sadia Nourin Mim, Lubana Tarannum, Tasrin Noman, and Jamal Uddin Ahamed. Ai driven water segmentation with deep learning models for enhanced flood monitoring. *arXiv preprint arXiv:*2501.08266, 2025.

[10] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. *In Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[11] Panqu Wang, Pengfei Chen, Ye Yuan, Ding Liu, Zehua Huang, Xiaodi Hou, and Garrison Cottrell. Understanding convolution for semantic segmentation. *In 2018 IEEE winter conference on applications of computer vision (WACV)*, pages 1451–1460. *IEEE*, 2018.

[12] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890, 2017.

[13] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv*:1511.07122, 2015.

[14] Furkan Isikdogan, Alan C Bovik, and Paola Passalacqua. Surface water mapping by deep learning. *IEEE journal of selected topics in applied earth observations and remote sensing*, 10(11):4909–4918, 2017.

[15] Shengyuan Piao and Jiaming Liu. Accuracy improvement of unet based on dilated convolution. *In Journal of Physics: Conference Series*, volume 1345, page 052066. *IOP Publishing*, 2019.

[16] Na-Xia Yang, Guo-Fa Li, Ting-Hui Li, Dong-Feng Zhao, and Wei-Wei Gu. An improved deep dilated convolutional neural network for seismic facies interpretation. *Petroleum Science*, 21(3):1569–1583, 2024.

[17] RJ Pally and S Samadi. Application of image processing and convolutional neural networks for flood image classification and semantic segmentation. *Environmental modelling & software*, 148:105285, 2022.

ISSN (online): 2581-3048
Volume 10, Issue 5, pp 642-650, May-2026
https://doi.org/10.47001/IRJIET/2026.105087

AUTHORS BIOGRAPHY

Nitesh Singh, Pursuing Msc in Computer System and Knowledge Engineering at Pulchowk Campus, Kathmandu, Nepal.

Asst. Prof. Prakash Chandra Prasad, at Department of Electronics and Computer Engineering, Pulchowk Campus, Kathmandu, Nepal.

Asst. Prof. Anku Jaiswal, at Department of Electronics and Computer Engineering, Pulchowk Campus, Kathmandu, Nepal.

Citation of this Article:

Nitesh Singh, Prakash Chandra Prasad, & Anku Jaiswal. (2026). Modified U-Net with Dilated Convolution for Flood Water Segmentation. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 10(5), 642-650. Article DOI https://doi.org/10.47001/IRJIET/2026.105087

****

