



OPEN Flood change detection model based on an improved U-net network and multi-head attention mechanism

Fajing Wang & Xu Feng

This work aims to improve the accuracy and efficiency of flood disaster monitoring, including monitoring before, during, and after the flood, to achieve accurate extraction of flood disaster change information. A modified U-Net network model, incorporating the Transformer multi-head attention mechanism (TM), is developed specifically for the characteristics of Synthetic Aperture Radar (SAR) images. By integrating the TM, the model effectively prioritizes image regions relevant to flood disasters. The model is trained on a substantial volume of annotated SAR image data, and its performance is assessed using metrics such as loss function, accuracy, and precision. Experimental findings demonstrate significant improvements in loss value, accuracy, and precision compared to existing models. Specifically, the accuracy of the model algorithm in this work reaches 95.52%, marking a 3.46% improvement over the baseline U-Net network. Additionally, the developed model achieves an accuracy of 90.11% while maintaining a loss value of approximately 0.59, whereas other model algorithms exceed a loss value of 0.74. Thus, this work not only introduces a novel technical approach for flood disaster monitoring but also has the potential to enhance disaster response procedures and provide scientific evidence for disaster management and risk assessment processes.

Keywords SAR imagery, Artificial intelligence, U-Net network, Flood disaster identification, Attention mechanism

Research background and motivations

Flood disasters are among the most prevalent and devastating natural calamities globally. Beyond directly threatening human lives and property, they can cause infrastructure paralysis, agricultural production losses, and long-term ecological damage^{1–3}. With the intensification of global climate change, the increasing frequency and severity of flood disasters have become undeniable realities⁴. However, existing methods face challenges in real-time data processing accuracy and feature extraction. Traditional image processing techniques often rely on manual intervention, making them not only inefficient but also prone to subjective judgment. While deep learning techniques, particularly the U-Net network, have significantly advanced image processing, there remains a gap in optimizing the characteristics of Synthetic Aperture Radar (SAR) images and improving the model's generalization capabilities across different geographic environments. Existing research has predominantly focused on improving network architectures, with relatively little systematic exploration of SAR image processing and model adaptability. This gap limits the effectiveness of current flood monitoring systems in diverse geographic settings.

In the field of flood disaster monitoring, SAR imagery has emerged as a pivotal tool due to its ability to acquire surface information under all-weather and all-day conditions^{5,6}. Compared to optical remote sensing, SAR imagery offers the advantage of being unaffected by weather and lighting conditions, allowing for the acquisition of high-quality surface information under extreme conditions. However, interpreting SAR images is complex, especially when extracting flood-related change information, which requires more advanced image processing techniques. For instance, Wang et al.⁷ employed SAR technology to monitor flood conditions, facilitating a comprehensive assessment of the extent and impact of flood disasters. SAR imagery has the unique ability to penetrate cloud cover and rain, providing high-resolution surface coverage and terrain information, thereby enabling the swift identification of flood extents and disaster assessments⁸. For example, Mastro et al.⁹

School of Transportation and Geomatics Engineering, Yangling Vocational & Technical College, Yangling 712100, Shaanxi, China. email: 18811360968@163.com

utilized SAR imagery for change detection, offering new data and methods for flood disaster monitoring and management.

Aziz et al.¹⁰ demonstrated that SAR imagery offered high accuracy and reliability in delineating flood areas, making it an effective tool for flood risk assessment. Zhang et al.¹¹ utilized gated recurrent unit neural networks to extract flood information from SAR time series, thereby enhancing flood monitoring capabilities. Surampudi and Kumar¹² employed SAR imagery and digital elevation models to estimate agricultural flood depths, achieving precise flood depth estimations. Rao et al.¹³ found that combining SAR imagery with machine learning technology effectively identified earthquake-induced building damage, providing new methods for disaster emergency response. Additionally, Zhang et al.¹⁴ introduced the DaliWS dataset for high-resolution water body segmentation in SAR imagery, offering valuable data for training and evaluating water body segmentation algorithms. Patil et al.¹⁵ used an ensemble modeling approach for flood mapping and damage assessment, demonstrating that this method accurately delineated flood areas and assesses flood damage, thereby providing crucial support for disaster response and recovery efforts.

The U-Net network has garnered significant attention due to its exceptional performance in medical image segmentation tasks. Numerous scholars have extended its application to various domains, including remote sensing image interpretation, land cover classification, and object detection. Chen et al.¹⁶ introduced a multi-level attention-guided U-Net network combined with transformers for medical image segmentation. Experimental results demonstrated that this method achieved high accuracy and robustness in medical image segmentation tasks. Yousef et al.¹⁷ developed a U-Net-based model for MR brain image segmentation, showing that the model could effectively achieve precise segmentation of MR brain images, providing a reliable tool for medical image analysis. Zhang et al.¹⁸ proposed an improved Attention U-Net network for interactive medical image annotation. Experimental results indicated that this method could significantly enhance the efficiency and accuracy of medical image annotation, offering a novel approach to medical image processing.

Chen et al.¹⁹ proposed a U-Net network-based method for geological disaster extraction, demonstrating effective landslide extraction in Sentinel-2 A remote sensing images with high performance and robustness. Zhang et al.²⁰ developed an improved U-Net network for land cover classification of multi-source remote sensing data, significantly enhancing classification accuracy and robustness. Nie et al.²¹ applied an improved U-Net network for sand road extraction in remote sensing images, effectively identifying sand roads. Clark et al.²² utilized an optimized U-Net network for land use and land cover classification of aerial photography images, achieving commendable performance. Quan et al.²³ introduced a method for learning SAR-optical cross-modal features for land cover classification, attaining high classification accuracy. Jian et al.²⁴ proposed an improved nested U-Net network for fluorescence in situ hybridization cell image segmentation, achieving precise and accurate cell image segmentation in experimental results.

In the field of flood disaster monitoring, numerous scholars have explored the application of the U-Net network. Lee et al.²⁵ proposed a method for semantic segmentation network slimming and edge deployment, using it in real-time forest fire and flood monitoring systems based on unmanned aerial vehicles (UAVs). Their experimental results demonstrated that this method could achieve rapid and accurate monitoring of forest fires and floods, providing an effective tool for disaster management. Li et al.²⁶ effectively assessed agricultural disasters in the Poyang Lake area using Sentinel-1 SAR data, offering crucial support for agricultural production. Pech-May et al.²⁷ developed a deep learning model based on the U-Net network for flood area recognition, which proved effective in identifying flood extents. Xu et al.²⁸ introduced an information expansion network based on U-Net networks, achieving high accuracy and robustness in water body extraction tasks. These studies underscore the versatility and efficacy of the U-Net network in flood disaster monitoring, highlighting its potential to significantly enhance real-time disaster response and management.

Concurrently, the rapid advancement of artificial intelligence, particularly deep learning technology, has brought about revolutionary transformations in image analysis and processing. In remote sensing image processing, the U-Net network extracts contextual information from images through its contracting path and achieves precise localization through its expansive path. This facilitates accurate classification and segmentation of objects in remote sensing images^{29–31}. For instance, Ren et al.³² introduced a transfer perception map U-Net network, which enhances the semantic segmentation performance of PolSAR images. Furthermore, the end-to-end training approach of U-Net streamlines the model training process, thereby improving processing efficiency³³.

In summary, floods represent a severe natural disaster that poses threats to ecosystems, economic development, and human safety. According to hydrological studies, the causes of flooding are typically closely related to various factors, including rainfall, topography, and land use. Accurate monitoring and early warning of floods can significantly reduce the losses incurred. Although traditional flood monitoring methods are effective, they still face challenges when dealing with complex terrains and rapidly changing flood situations. Flood disaster monitoring refers to the comprehensive observation and assessment of the entire process of flooding, including pre-event predictions, real-time monitoring during the event, and damage assessments after the flood recedes. This work aims to enhance the extraction of flood change information from SAR images through an improved U-Net network combined with a multi-head attention mechanism, providing a scientific basis for disaster management and emergency response. The primary variable of this work is the change information of flooded areas, specifically identifying flood areas by comparing SAR images taken before and after flooding. This objective is to improve the accuracy of monitoring flood disaster change information to facilitate more effective disaster response and risk assessment. While this work does not directly observe hydrological parameters such as flood depth, peak, intensity, or duration, identifying changes in flood areas is crucial for understanding the impact and distribution of floods.

Research objectives

To address the challenges identified in the current research, this study proposes an advanced methodology that integrates SAR images with deep learning techniques to extract flood disaster change information. The innovation centers on the development of a model based on an enhanced U-Net network, which incorporates the Transformer multi-head attention mechanism (TM) to improve the recognition of flood-related features in SAR images. This optimization enables accurate segmentation and identification of flood disaster areas, thus minimizing reliance on manual processing. The effectiveness and applicability of the proposed method are validated through a series of tests across various flood disaster scenarios.

This work tackles key limitations in existing research as follows:

- (1) **Lack of Time Series Analysis:** To address this gap, the study employs multi-temporal SAR image data and introduces a time series analysis module. This enhancement allows for a more comprehensive capture of dynamic flood changes, leading to improved monitoring of flood evolution and more accurate disaster warnings.
- (2) **Insufficient Generalization Ability Verification:** Extensive experimental validation of the model across multiple flood scenarios—encompassing diverse geographical regions and disaster types—demonstrates its broad applicability and adaptability to various complex environments. This verification strengthens the model's reliability for practical applications.
- (3) **Limitations in Real-Time Monitoring Capabilities:** To overcome the limitations in real-time monitoring, the model is optimized to efficiently process large volumes of SAR image data and generate disaster change information under near-real-time conditions. Experimental outcomes reveal that the model exhibits robust stability and efficiency in handling complex disaster scenarios, providing timely and accurate decision support for disaster management agencies.

These advancements not only enhance the efficiency of SAR image data processing but also reduce the need for manual intervention through automated deep learning models. Additionally, the integration of the attention mechanism further refines the ability to identify critical disaster features. Extensive experimental validation offers disaster management agencies a powerful tool for improved risk assessment and emergency response. These contributions advance flood disaster monitoring technology and facilitate the integration of natural disaster monitoring into broader environmental monitoring systems.

Research model

Analysis of U-net and its improvements

In this work, changes in flood inundation areas are selected as a key variable because this information is crucial for the rapid response to flood disasters and for assessing their impacts. While hydrological parameters such as flood depth and peak are essential for understanding the physical characteristics of floods, they typically require field measurements or complex hydrological models for acquisition. In contrast, information on changes in flood areas identified through SAR imagery can be obtained more quickly, providing immediate geospatial data support for disaster management. Moreover, these change indicators are correlated with hydrological parameters, making them a valuable proxy for assessing flood impacts.

Remote sensing technology allows for the observation of the Earth's surface from the air or space, facilitating the collection of detailed information on surface cover, topography, and environmental changes. SAR images, in particular, are well-suited for monitoring flood hazards during adverse weather conditions due to their ability to penetrate clouds and rain. The Geographic Information System (GIS) platform provides a comprehensive framework for storing, analyzing, and visualizing these remote sensing data. GIS's analytical capabilities enable researchers to consider critical factors such as topography, land use, and water distribution, which are essential for understanding and predicting the occurrence and impact of flood hazards. Integrating deep learning technology with remote sensing and GIS platforms can enable real-time monitoring and early warning of flood disasters. Deep learning models can process and analyze large volumes of data from remote sensing sensors to swiftly identify potential flood areas. The GIS platform provides the necessary geographical context, allowing prediction results to be combined with the actual geographical environment. This integration offers robust decision support for disaster response and resource allocation.

U-Net, a deep learning architecture specifically designed for image segmentation tasks, excels at combining low-level detailed features with high-level semantic information³⁴. The core advantage of U-Net lies in its architecture, which consists of a contracting path and a symmetric expanding path. This design connects feature maps from the contracting path to corresponding layers in the expanding path through skip connections, facilitating gradient backpropagation and preserving more spatial information^{35–37}. Figure 1 illustrates the structure of U-Net as applied to image feature extraction.

Figure 1 shows that when processing SAR images, the original U-Net network may encounter several challenges, such as sensitivity to complex backgrounds and noise, as well as limitations in extracting fine features. SAR images often exhibit significant background noise and intricate environmental features that complicate the image segmentation process in the original U-Net network, affecting the accuracy of the results. In flood disaster scenarios, SAR images frequently contain mixed signals from various water bodies, terrains, and structures, which challenge U-Net's feature extraction capabilities. This issue is supported by research such as Guo et al.³⁸, which demonstrates that background noise in SAR images significantly impacts the segmentation performance of deep learning models. Additionally, U-Net may struggle to capture subtle features, such as the boundaries of water bodies and fine flood traces, when processing high-resolution SAR images, which are crucial for precise disaster identification. For example, Cao et al.³⁹ emphasize that U-Net often encounters difficulties with detail loss in high-resolution remote sensing images.

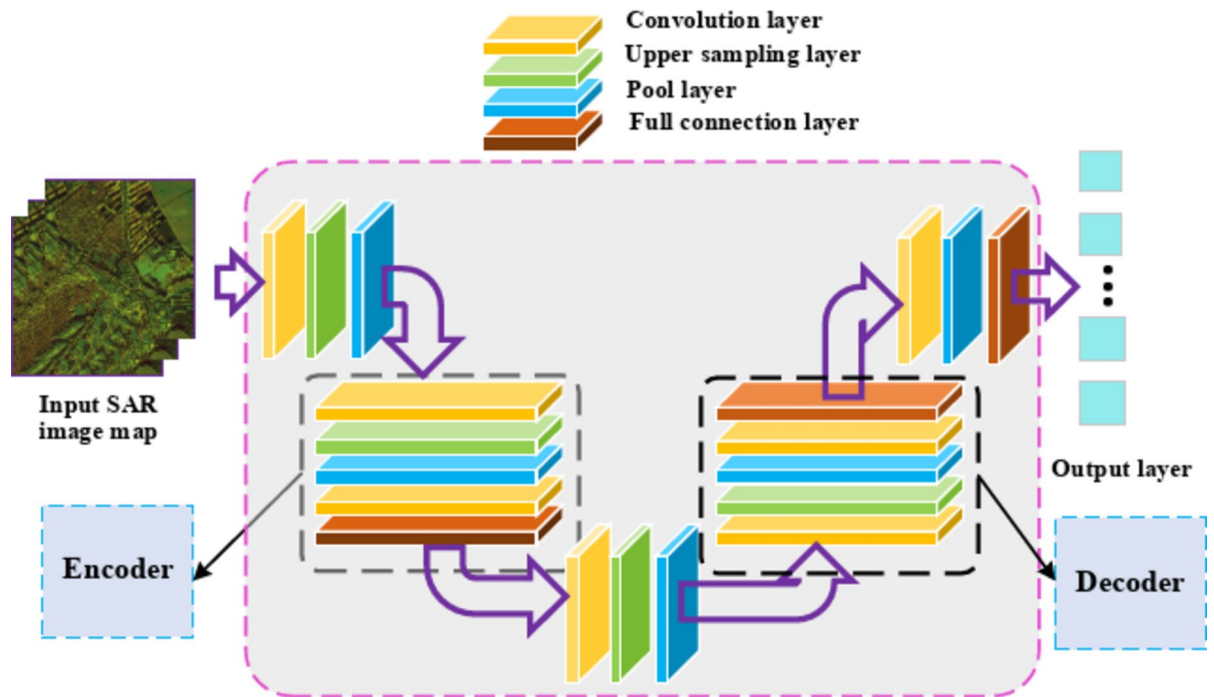


Fig. 1. Diagram illustrating the structure of the U-Net network for image feature extraction.

These challenges are not unique to U-Net networks but are also encountered in other deep Convolutional Neural Network (CNN) models. Many deep learning models face difficulties with background noise and inadequate feature extraction when processing SAR images. While data preprocessing techniques, such as image denoising and pixel normalization, can mitigate the impact of noise, they often fail to fully address the complex backgrounds and high-frequency noise inherent in SAR images. Additionally, preprocessing methods may result in the loss of critical image information, affecting the extraction of fine details. Although increasing the depth or width of a network can improve performance, such modifications often lead to significant increases in computational complexity and may not necessarily resolve the issues associated with complex backgrounds in SAR images.

This work addresses the specific characteristics of flood disaster monitoring by enhancing the U-Net network. Specifically, a multi-head attention mechanism is introduced to improve the model's ability to identify flood-related features. Additionally, the robustness of the model to complex background and noise in SAR images is improved by depth-wise separable convolutions and batch normalization layers^{40–42}. Depth-wise separable convolution breaks down standard convolution into two stages: depth-wise convolution and point-wise convolution. This decomposition minimizes computational overhead and contributes to improved model efficiency and responsiveness. These improvements enable the U-Net network to more accurately identify flood areas, providing a scientific basis for rapid response and effective management of flood disasters.

During the data preprocessing stage, SAR images undergo a series of processing steps to enhance the accuracy of subsequent feature extraction. First, a median filter is applied to remove the commonly encountered salt-and-pepper noise in SAR images. Specifically, a 3×3 median filter is employed, which effectively replaces pixel values with the median from their neighboring values, thereby reducing the influence of extreme values. Additionally, normalization is performed on the images to ensure that the image data maintains a stable distribution during network training. Finally, all images are resized to a uniform dimension of 512×512 pixels to meet the input requirements of the network. Data normalization is employed to enhance model performance and generalization capabilities. This process aims to maintain a stable distribution of the data as it progresses through each layer of the network, thereby mitigating issues such as gradient vanishing or explosion during training. Consequently, this approach accelerates model convergence and improves training accuracy. Following data normalization, the model further adjusts the distribution of input features through “scaling and shifting” steps. In this process, the mean and variance of each data batch are computed, as detailed in Eqs. (1) and (2), and these values are used for data normalization, as shown in Eq. (3). The objective of these steps is to ensure that the input features at each layer have a consistent scale, thereby reducing the risks of gradient vanishing or explosion and expediting network convergence.

$$\mu_{\beta} = \frac{1}{m} \sum_{i=1}^m x_i \quad (1)$$

In Eq. (1), m represents the number of batches of image data. Next, the variance δ_β^2 of the data distribution within each batch is given by Eq. (2):

$$\delta_\beta^2 = \frac{1}{m} \sum_{i=1}^m (x_i - \mu_\beta)^2 \quad (2)$$

Normalization of the data \hat{x}_i within the current batch is then performed, as shown in Eq. (3):

$$\hat{x}_i = \frac{x_i - \mu_\beta}{\sqrt{\delta_\beta^2 + \varepsilon}} \quad (3)$$

In Eq. (3), ε is a small constant, set to 10^{-5} here, to ensure the denominator never becomes zero, thereby maintaining stability. This normalization ensures that the mean of the features is 0 and the variance is 1, consistent with the distribution of the input data. Finally, BN includes a “scale and shift” step, as shown in Eq. (4):

$$y_i = \gamma \hat{x}_i + \beta = BN_{\gamma, \beta}(x_i) \quad (4)$$

In Eq. (4), γ and β represent the scaling and shifting parameters in the BN layer. These parameters allow the network to learn the appropriate scaling and shifting of data, thereby enhancing the network's performance and generalization capability. While Eq. (4) might seem contradictory to the previous normalization operations, it can actually restore the original input, provided the parameters γ and β satisfy Eq. (5):

$$\begin{cases} \gamma^{(k)} = \sqrt{Var(x^{(k)})} \\ \beta^{(k)} = E(x^{(k)}) \end{cases} \quad (5)$$

By implementing these steps, the model effectively manages variations across different data batches while ensuring consistency in the distribution of input data during training. This approach not only accelerates model convergence but also mitigates numerical instabilities such as gradient vanishing or explosion. As a result, the overall performance and generalization capability of the model are significantly enhanced.

Enhanced SAR image-based flood disaster feature extraction model with integrated U-net network and TM

The preceding discussion examines how the integration of remote sensing technologies, SAR images, GIS platforms, and deep learning techniques offers new possibilities for flood disaster monitoring. This technological convergence provides a robust foundation for real-time monitoring and early warning systems. Despite the considerable potential of these methods, several challenges persist that require further attention, including enhancing the precision of complex disaster scenario recognition and improving real-time response capabilities. A thorough analysis of the U-Net network's fundamental structure, as well as its advantages and limitations in image segmentation tasks, is conducted. This analysis, along with a review of existing improvement strategies, highlights several key areas that require further optimization, particularly in the context of managing complex flood disaster scenarios. To enhance U-Net's performance in extracting flood disaster features from SAR images, this work incorporates TM^{43–46}. This mechanism enables the network to focus more effectively on flood disaster-related features by learning different representation subspaces of images in parallel. When integrated with the enhanced U-Net network, it forms an end-to-end flood disaster feature extraction model. Figure 2 illustrates the architecture of the SAR image-based flood disaster feature extraction model, which integrates TM with U-Net.

As illustrated in Fig. 2, the model workflow is as follows: Initially, SAR image data undergoes preprocessing, which includes denoising and normalization operations to enhance data quality and ensure the accuracy of subsequent feature extraction. The enhanced U-Net network is then employed as a feature extractor to derive features from the preprocessed SAR images. This network design effectively captures both contextual and precise positional information.

Following this, the encoder component of the U-Net network extracts high-level feature representations of the SAR images. During this phase, the features are compressed into a more compact representation while preserving essential information. Rather than being passed to the decoder component of the U-Net network, the encoded data is directly forwarded to a subsequent Transformer module for further processing. This approach is intended to prevent potential over-refinement of features by the U-Net decoder, thereby retaining deeper information that can be more effectively processed and utilized by the Transformer module.

In the Transformer module, a multi-head attention mechanism is employed to process these high-level features. This mechanism allows the model to simultaneously handle image features across different subspaces, enhancing its ability to focus on information relevant to flood disasters.

The input SAR image data is first embedded into vectors and then mapped to query (Q), key (K), and value (V) spaces through linear transformations. By computing the similarity between queries and keys, the model generates attention scores that reflect the relevance of different SAR images to each other.

These attention scores are subsequently normalized using the softmax function for each attention head, determining the attention weight distribution of each SAR image relative to others. This normalization results in the final feature representation for each SAR image. The outputs from all attention heads are then concatenated to form the ultimate multi-head attention representation. This final representation integrates information from various subspaces, providing a comprehensive understanding of the SAR image features.

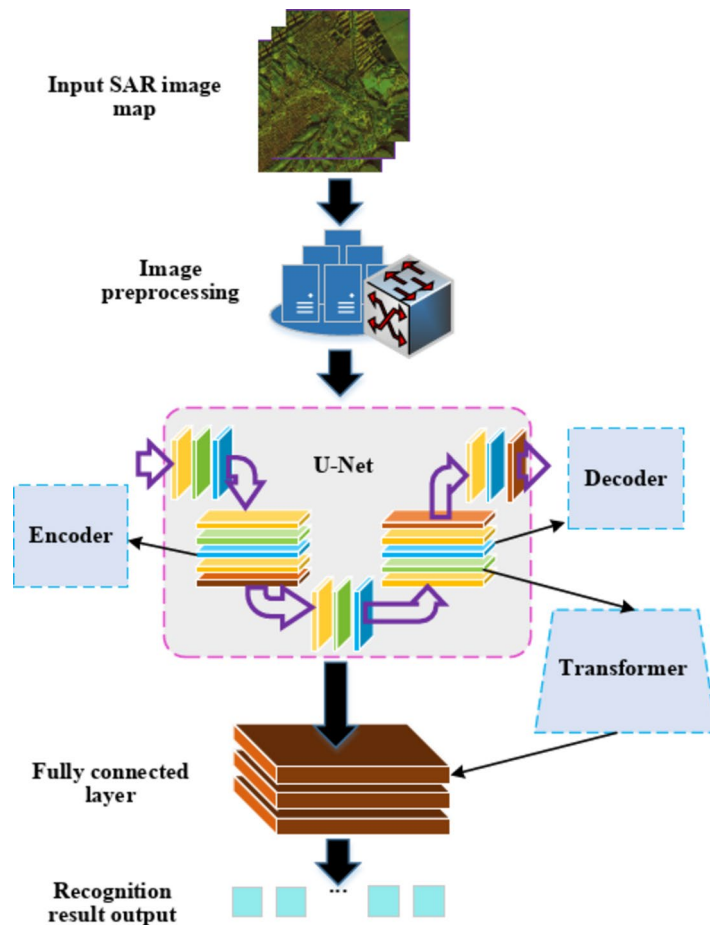


Fig. 2. Architecture of the SAR image-based flood disaster feature extraction model, incorporating the fusion of TM and U-Net.

In the U-Net network, the size of the feature maps is significantly reduced at the bottleneck after the encoder. To make these feature maps suitable for processing by the Transformer, a linear transformation is applied to embed the features into a lower-dimensional representation. However, since the features processed by the Transformer need to be restored to the original resolution for pixel-level analysis, a bilinear interpolation technique is employed to upsample the features from the multi-head attention mechanism to match the size of the original image.

At the final stage of the model, a fully connected layer is introduced to convert the final feature representation into a series of values. These values are not directly used as the final output but are instead remapped back to the image pixel space through an inverse transformation to produce the final output image. This approach ensures that the output is presented in raster image form, maintaining the same size and resolution as the input image.

Consequently, pixel-based accuracy metrics can be used to evaluate model performance. Metrics such as accuracy and precision are calculated by comparing the correspondence between the output image and the ground truth image at each pixel. Thus, even after processing by the Transformer and fully connected layers, effective pixel-level performance evaluation remains feasible.

The calculation for the attention mechanism, $Attention(Q, K, V)$, is represented as Eq. (6):

$$Att(Q, K, V) = \text{softmax} \left(\frac{Q \cdot K^T}{\sqrt{d_k}} \right) \cdot V \quad (6)$$

In Eq. (6), Q represents the matrix of “query” vectors, K refers to the matrix of “key” vectors, and V signifies the matrix of “value” vectors.

The Transformer multi-head attention mechanism involves performing N linear mappings on the matrices Q , K , and V , with the computational expressions depicted in Eq. (7) through (9):

$$MultiHeadAtt_i = Att(QP_i^Q, KP_i^K, VP_i^V) \quad (7)$$

$$MultiHeadAtt = \text{concat} (Att_1, Att_2, \dots, Att_N) \quad (8)$$

$$Y = MultiHeadAtt_i \cdot P^O \quad (9)$$

Here, P_i^Q , P_i^K , and P_i^V represent the representation of the i -th head. Subsequently, all heads are concatenated to derive the final representation $MultiHeadAtt$. Then, a linear transformation with parameters P^O is applied, yielding Y . P^O refers to the linear mapping matrix that integrates these N representations. The equation for Att_i is expressed as Eq. (10):

$$Att_i = \frac{\left(\frac{Q_i \cdot K_i^T}{\sqrt{d_k}} \right)}{\sum_i \left(\frac{Q_i \cdot K_i^T}{\sqrt{d_k}} \right)} \cdot V_i \quad (10)$$

In Eq. (10), $Q_i \cdot K_i$, V_i represents the mapping matrix from the original input to the i -th subspace. Subsequently, feature reconstruction is performed on the SAR image using the fused feature representation. The introduction of fully connected layers facilitates the subsequent mapping, integration, and transformation of features following the TM fusion. This enhancement improves the model's expressive capacity and generalization ability. As a result, the model's effectiveness in SAR image-based flood disaster feature extraction is increased, leading to a more accurate depiction of flood disaster information. Figure 3 provides the pseudocode illustrating the model flow.

Experimental design and performance evaluation

Datasets collection

The acquisition and processing of SAR imagery data involve a certain latency, which is critical for the rapid response in flood disaster monitoring. This work utilizes C-band SAR imagery data provided by the Sentinel-1 satellite, which has a standard revisit cycle of 6 to 12 days, making it suitable for periodic flood monitoring. In emergencies such as floods, data acquisition latency can be shortened through rapid task programming, allowing for the quick retrieval of relevant data after a disaster occurs. The typical latency for this data ranges from a few hours to several days. Coupled with the proposed deep learning model, flood change information can be processed and generated shortly after data acquisition, supporting near-real-time disaster response. The Sentinel-1 data is publicly available and free of charge, featuring both single-polarization and dual-polarization modes. The resolution is 10 m and the temporal resolution is 12 days, making it suitable for large-scale flood detection and monitoring tasks. This data can be accessed via the European Space Agency's Copernicus Open Access Hub, covering a global range with good temporal resolution, meeting the flood detection needs across

```

1 Start
2 Input: SAR image data
3 Output: Identification results of flood disaster information
4 # Preprocessing SAR image data
5 # Feature extractor: Improved U-Net network
6 def U_Net_feature_extractor(SAR_image):
7     # Extract high-level feature representations of SAR images in the encoder
    part
8     encoded_features = U_Net_encoder(SAR_image)
9     return encoded_features
10 # Multi-head attention mechanism
11 def multihead_attention(input_data):
12     # Map to multiple query, key, and value spaces using linear transformation
13     Q = linear_transform(input_data)
14     K = linear_transform(input_data)
15     V = linear_transform(input_data)
16     # Calculate attention scores
17     attention_scores = calculate_attention_scores(Q, K)
18     # Normalize attention scores
19     normalized_attention = softmax(attention_scores)
20     # Calculate weighted sum
21     weighted_sum = weighted_sum(V, normalized_attention)
22     # Multi-head concatenation
23     multihead_concatenation = concatenate(weighted_sum)
24     return multihead_concatenation
25 # Feature reconstruction
26 # Fully connected layer
27 def fully_connected_layer(features):
28     # Perform operations of fully connected layer to enhance feature expression
    capability
29     and generalization capability
30     output = fully_connected(features)
31     return output
32 End

```

Fig. 3. Pseudocode flowchart of TM fusion with U-Net algorithm applied to SAR image-based flood disaster feature recognition.

various geographical locations and climatic conditions. The choice of Sentinel-1 data is based on its global accessibility and openness, ensuring the reproducibility and transparency of this work. The data utilized is Sentinel-1 Level 1 Ground Range Detected (GRD) data. This is a processed data level that includes deblurring and terrain correction, making it appropriate for applications such as flood monitoring and change detection. In terms of polarization, the SAR data of vertical polarization and horizontal polarization are mainly used. These two types of polarization provide different angles of surface information, aiding in the more accurate identification and monitoring of flood areas. This work employs 476 pairs of SAR images, encompassing both pre- and post-flooding events. The dataset primarily focuses on regions in Australia, specifically areas affected by the 2022 floods, covering southeastern Australia and other severely impacted locations. For analysis and validation, 400 image pairs are designated for training and validation, while the remaining 76 pairs are reserved for testing. The images are processed to eliminate noise and normalize them, with all images resized to 512×512 pixels. In addition to Australian regions, the dataset includes images from Southeast Asia, South America, and Europe to represent a range of geographic and climatic conditions. This diverse dataset facilitates the evaluation of the model's adaptability and robustness across different geographic environments, with a particular focus on performance within the Australian context.

During the data collection and processing phase, the focus is on identifying and labeling flood inundation areas. Each pair of SAR images includes pre- and post-flood images, and through image analysis, flood areas are labeled as positive (1) and non-flood areas as negative (0). These annotations generate classification labels for training and evaluating the model. The goal is to classify the SAR imagery using the improved U-Net model, categorizing pixels as 'flood area' or 'non-flood area.' By integrating the multi-head attention mechanism, the model can more accurately identify features related to flood disasters, thereby enhancing classification accuracy.

Additionally, due to flood-related weather conditions characterized by significant cloud cover, some SAR images are affected by cloud obfuscation. Cloud cover can significantly impact SAR image quality, potentially degrading the model's performance in image processing and flood detection. To address this challenge, careful consideration is given to image quality and environmental factors during the selection of the training set.

To effectively address cloud cover issues in the images, this work employs various specific techniques and tools during data preprocessing, including cloud masking, image inpainting, noise removal, and image normalization. First, a threshold-based cloud masking technique is applied, using the radiometric characteristics of SAR imagery to mask or remove areas potentially affected by cloud interference, ensuring that the analysis region remains unaffected by clouds. For areas obscured by clouds, a partial differential equation image inpainting method is utilized to interpolate the spatial features of neighboring pixels. By interpolating the spatial features of adjacent pixels, the image details under the cloud cover can be recovered as much as possible to mitigate the impact of information loss on the analysis. Additionally, to reduce speckle noise in SAR imagery, a combination of multi-level median filtering and wavelet transform techniques is employed to significantly enhance image quality, enabling the model to extract feature information more precisely during training and prediction phases. Finally, OpenCV and Python's image processing libraries are used for image normalization, ensuring consistency in contrast and brightness across images from diverse data sources and regions, thus reducing the impact of radiometric differences on the model.

To evaluate the model's classification performance and generalization capability while maximizing the use of the limited dataset, a 5-fold cross-validation approach is employed. The dataset is initially divided into five subsets, maintaining the original data distribution characteristics. The model training and validation process is then repeated five times, with each iteration using four subsets for training and the remaining subset for testing. Performance metrics are recorded and averaged across iterations to provide a comprehensive assessment of the model's performance. This approach optimizes the use of available data resources and effectively evaluates the model's generalization ability and stability, while minimizing the risk of overfitting.

Figure 4 presents two pairs of SAR images captured before and after flood events. It is important to note that these images are not standard SAR images but have undergone pseudo-color processing. This processing enhances contrast and makes changes in flood-affected areas more visually apparent. SAR images are single-band and lack the color information typical of conventional RGB images. To generate pseudo-color images for a more intuitive display of changes before and after flooding, the amplitude and phase information from the SAR images are utilized. Specifically, the amplitude information is mapped to the red channel, the phase information to the green channel, and the phase change information from the time series analysis to the blue channel. This mapping method aids in visually presenting the changes before and after flooding, particularly highlighting the boundaries and extent of the flood-affected areas. This technique involves specific image processing and color synthesis of SAR data, emphasizing the changes in flood-affected areas. This technique involves specific image processing and color synthesis of SAR data, highlighting changes in flood-affected areas. The visualization method facilitates a clearer understanding of surface changes for non-experts, supporting further analysis and assessment of flood impacts. Additionally, the "cloud layer" phenomenon observed in the images is attributed to auxiliary data used in post-processing and does not represent actual cloud cover in the SAR data.

Experimental environment

The proposed model considers different time points in flood monitoring during the training and validation phases, ensuring that it can effectively identify changes in flood characteristics, including pre-disaster monitoring to detect potential risks in advance. Furthermore, the subsequent monitoring and assessment provide support for post-disaster response. To validate the developed algorithm, the experiments are primarily conducted on the Windows 10 Professional operating system. Table 1 presents detailed information regarding the environment configuration.

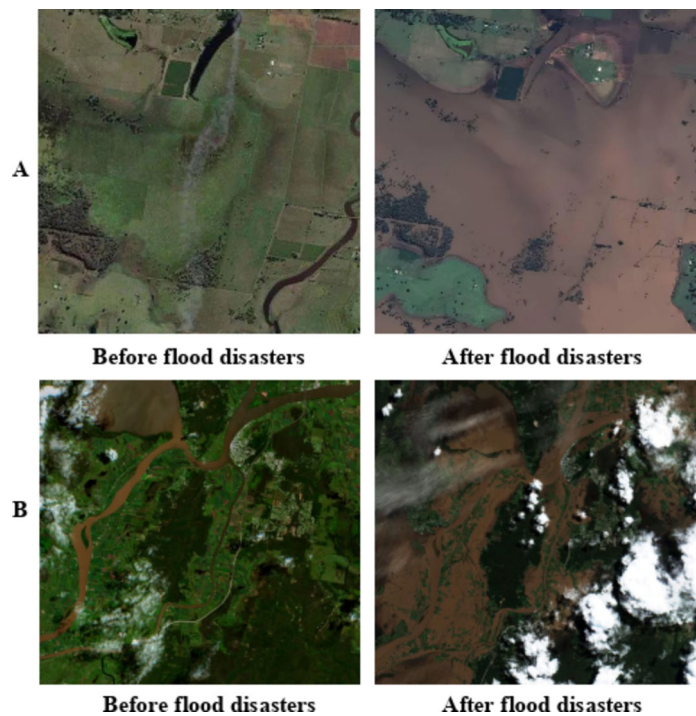


Fig. 4. SAR image data before and after flood disasters ((A) Northern New South Wales region, 2022; (B) East Coast of Australia, 2022).

Name	Information	Name	Information
Operating system	Windows 10 Professional	Deep learning language	Python 3.9
GPU	NVIDIA RTX2070-8GB	Deep learning framework	Tensorflow 2.11.0
CPU	Intel i7-9750 H	Development environment	PyCharm 2022.3
Memory	16GB	Visualization tool	Tensorboard 2.13.0
Hard Disk	512GB		

Table 1. Experimental environment configuration.

Parameters setting

This work references the optimization strategies of Parikh et al.⁴⁷ and Melgar-García et al.⁴⁸ for noise suppression in SAR images and disaster monitoring models in hyperparameter selection, ensuring the stability and accuracy of the models in complex scenarios. The choice of the Adam optimizer is based on theoretical support from Barakat et al.⁴⁹ regarding adaptive optimizers; meanwhile, it combines with validation results from Parikh et al.'s SAR image classification experiments, enhancing the model's capacity to handle complex SAR data. The settings for batch size and training epochs are informed by empirical studies conducted by Liu et al.⁵⁰ and Zeng et al.⁵¹, achieving greater training efficiency and generalization performance. These foundational works provide theoretical and experimental justification for the used hyperparameter settings. The hyperparameters are configured as follows: the optimizer is Adam with an initial learning rate of 0.00001, the number of epochs is set to 100, and the loss function used is Binary Crossentropy.

Performance evaluation

This work focuses on two representative flood events: one is a riverine flood, and the other is a rainstorm-induced flood. These events exhibit significant differences in flood causation, affected areas, and watershed characteristics, thus effectively representing different flood scenarios. Riverine floods occur due to increased inflow from upstream, resulting in rising river levels and subsequent flooding in riparian areas. Such events typically have a large inundation area and a prolonged duration. Conversely, rainstorm-induced floods are characterized by sudden heavy rainfall leading to urban flooding and localized flash floods, exhibiting abruptness, smaller inundation areas, and rapid recession. These events are suitable for testing the model's accuracy and performance in responding to short-duration, high-intensity flood incidents. The reference flood map is generated through data fusion from multiple sources, including post-disaster field measurements, satellite remote sensing imagery, and flood reports published by relevant government agencies. This map is used for the validation and quantitative assessment of model predictions, providing a reliable benchmark for the accuracy

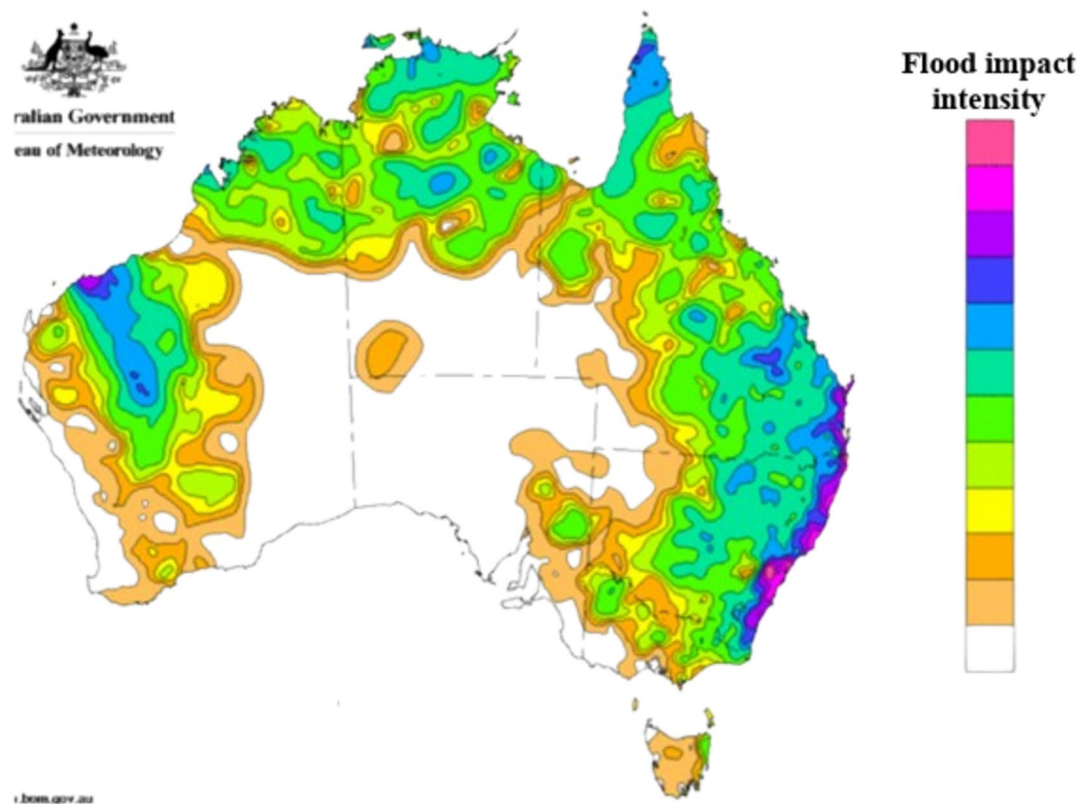


Fig. 5. Reference flood hazard map (Australia 2022).

Region	Annotated flood area (m ²)	Model-predicted area (m ²)	Matching accuracy (%)	Error (m ²)
Region A	5120	4900	95.7	220
Region B	6230	6050	97.1	180
Region C	4670	4600	98.5	70
Region D	5390	5300	98.3	90

Table 2. Comparison of flood area data for various regions.

of the flood monitoring model. The process involves comparing ground-truth flood data with SAR images, employing image processing and data fusion techniques to produce high-precision flood area annotations. These annotations enabled accurate comparisons during the model evaluation phase. For ease of comparison, visual examples of the reference flood maps are provided in the following illustration, highlighting the contours of the actual flood-affected areas. Figure 5 presents an example of the reference flood map for the 2022 Australian region.

Figure 5 displays the annotated areas on the reference flood maps. The shaded areas on the map represent the flood-affected areas, with varying colors indicating the relative intensity of the flood impact. The reference flood map integrates field measurement data, satellite imagery, and official flood reports, accurately reflecting the actual extent of flood impacts and serving as a dependable basis for validating model predictions.

Table 2 provides flood area data for different regions, including reference flood annotations, model-predicted areas, matching rates, and errors. This data enables an analysis of the model's prediction accuracy and its alignment with the actual flood extents. The estimation of flood area is based on the pixel classification results from SAR images. First, the improved U-Net model is used to binarize the SAR images, marking pixels in the flood area as 1 and those in non-flood areas as 0. In the binarized image, each pixel represents a fixed area of actual ground. The number of pixels labeled as flood areas is then counted, and combined with the spatial resolution of the SAR imagery, the total flood impact area can be derived. To enhance accuracy, the area estimation is averaged over multiple experiments, with preprocessing and error correction applied to regions potentially affected by cloud cover and noise.

- Region A: Located in northern New South Wales, encompassing major cities and surrounding rural areas.
- Region B: Covers a portion of the flood-affected area along the east coast of Australia, including major cities and coastal regions.

- Region C: Focused on a specific heavily impacted area along the east coast of Australia.
- Region D: Another major city and its surrounding area in northern New South Wales.

Table 2 illustrates the model's performance across various regions, demonstrating strong predictive capability. Matching accuracies exceed 95% in all regions. Region B achieves the highest matching accuracy at 97.1% with the smallest error of 180 m², indicating particularly precise predictions. Regions A and C also show high matching accuracies of 95.7% and 98.5%, respectively, though the predicted areas are slightly smaller than the reference data, resulting in some error. Region D displays a matching accuracy of 98.3% with an error of 90 m². Overall, while the model exhibits robust performance in predicting flood areas, further optimization is required to minimize prediction errors in certain regions.

To assess the performance of the proposed algorithm, it is compared with ResU-Net⁵², U-Net⁵³, AttU-Net⁵⁴, and models proposed by researchers such as Ren et al. (2024) in the relevant field. The comparison includes various metrics such as loss value, recognition performance, accuracy, and precision. The loss value measures the difference between the model's predictions and the actual labels. Binary Cross-Entropy is used as the loss function, calculated as shown in Eq. (11):

$$Loss = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - \hat{y}_i) \log(1 - \hat{y}_i)] \quad (11)$$

In Eq. (11), N represents the number of samples; \hat{y}_i and y_i are the predicted probability and actual label of the i -th sample. Accuracy measures the proportion of correct predictions among the model's prediction results, expressed in Eq. (12):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

In Eq. (12), TP and TN denote the number of True Positives and True Negatives, respectively, while FP and FN refer to the number of False Positives and False Negatives. Precision measures the proportion of true positives among the samples predicted to be positive, calculated as shown in Eq. (13):

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

Initially, the loss values of the proposed algorithm and other models are analyzed, as depicted in Fig. 6.

In Fig. 6, the performance of each model during the cross-validation process is illustrated, highlighting the variation in loss values with the training cycle. The figure indicates that the proposed algorithm exhibits a higher loss value at the initial stage of training. However, as training progresses, the loss value decreases rapidly. By the 10th cycle, the loss value of the proposed algorithm is significantly lower than that of the other models. In the subsequent training cycles, the loss value of the proposed algorithm gradually stabilizes and remains at a

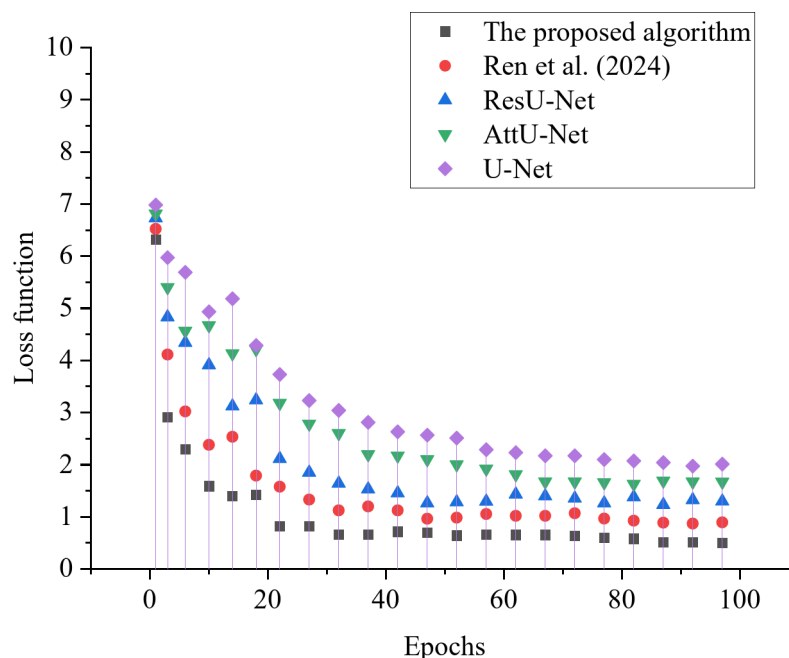


Fig. 6. Loss value results for each algorithm.

low level. Notably, after the 47th cycle, the loss value is markedly lower than that of other models, such as those proposed by Ren et al. (2024), ResU-Net, AttU-Net, and U-Net. This demonstrates that the proposed algorithm has notable advantages in terms of convergence speed and stability.

Figure 7 illustrates the recognition performance of the proposed algorithm compared to other models in identifying flood disasters.

Figure 7 illustrates the effectiveness of various algorithms in flood monitoring and detection. The figure presents the results of different algorithms in identifying flood-affected areas within SAR images. The proposed algorithm is shown to accurately detect submerged areas post-flooding, while other algorithms exhibit notably lower accuracy in flood area identification. In Fig. 7, red boxes highlight key flood-affected regions identified by the model when processing SAR images. These red boxes represent areas of focus determined by the multi-head attention mechanism of the Transformer, indicating regions deemed most relevant to flood characteristics. This visualization method clearly depicts the areas prioritized by the model in SAR images.

Furthermore, the red boxes in Fig. 7 serve as both a visualization of the model's predictions and a reference for subsequent post-disaster analysis. Researchers can use these highlighted regions to validate whether they are indeed critical areas impacted by the flood. Quantitative analysis confirms that the proposed algorithm outperforms other comparative algorithms in accuracy and precision for flood detection. This underscores the model's capability to handle complex backgrounds and noise and validates the effectiveness of the multi-head attention mechanism in enhancing flood area identification.

Figures 8 and 9 depict the results.

Figure 8 reveals the accuracy variations of each model in flood disaster identification during the cross-validation process. The proposed algorithm achieves an accuracy of 0.1423 in the first cycle, which significantly surpasses that of other models. As training progresses, the proposed algorithm's accuracy improves rapidly, reaching a maximum of 0.8571 after the 50th cycle and maintaining a high level thereafter. These results demonstrate that the proposed algorithm exhibits superior recognition capability for flood disaster identification tasks. Additionally, it achieves higher identification accuracy within a short training period and shows better recognition performance and stability compared to other models.

Figure 9 presents the accuracy variations of each model in flood disaster identification, based on the worst-performing results in the cross-validation process. The proposed algorithm's accuracy in the first cycle is 0.1023, notably outperforming other models. Throughout subsequent training, the accuracy of the proposed algorithm continues to rise, reaching 0.6634 after the 50th cycle and stabilizing between 0.8193 and 0.8254 from the 80th to the 100th cycle. This indicates that the proposed algorithm significantly outperforms other models in accuracy

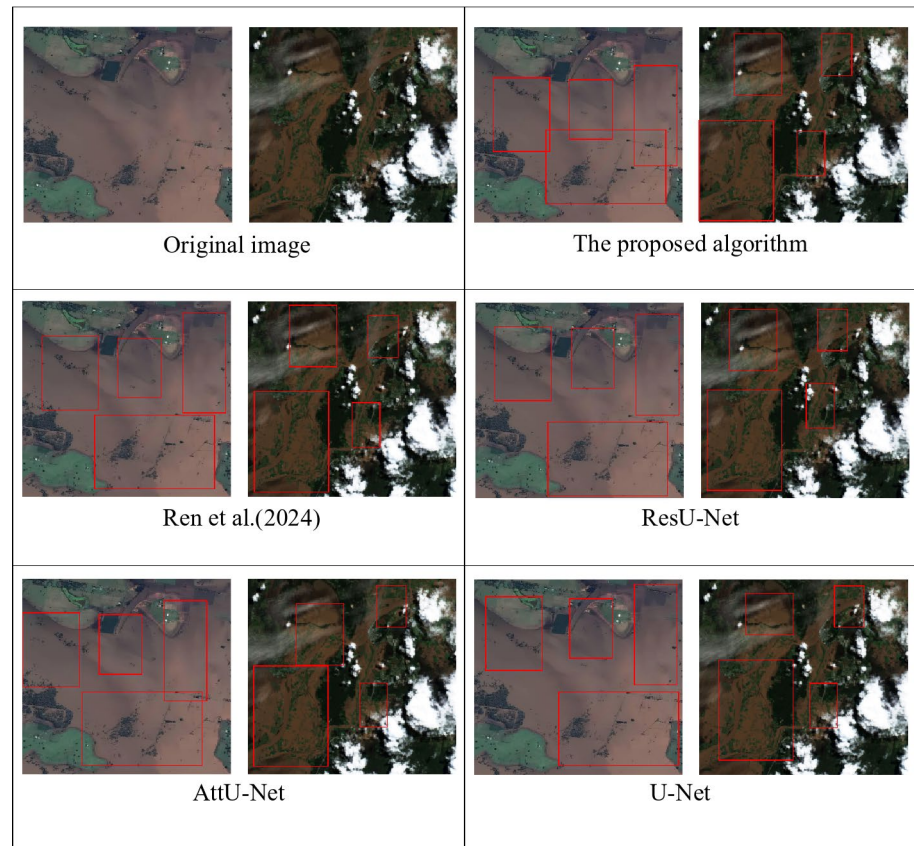


Fig. 7. Recognition performance of each algorithm for flood disaster.

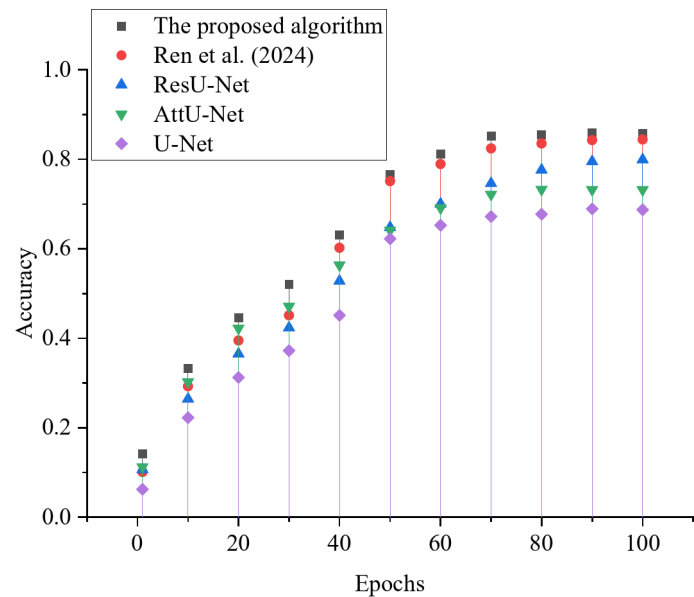


Fig. 8. Accuracy results of flood disaster identification as the iteration cycle progresses for each algorithm.

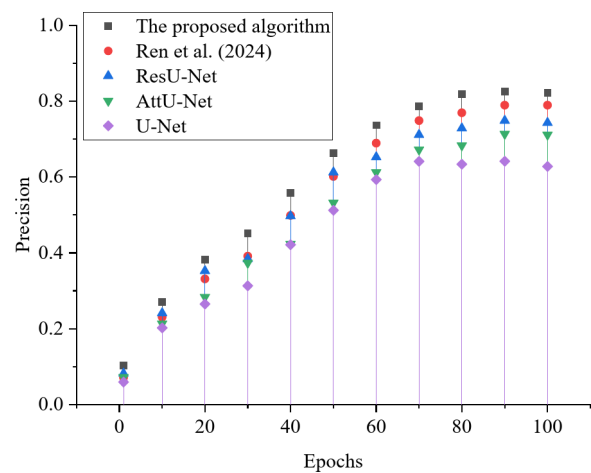


Fig. 9. Precision results of flood disaster identification as the iteration cycle progresses for each algorithm.

Model	Convergence time (average epochs)	Convergence speed (time per epoch, seconds)	Total training time (hours)
U-Net	100	30	50
ResU-Net	90	32	48
AttU-Net	95	31	49
Ren et al. (2024)	105	29	52
The proposed model	85	28	45

Table 3. Comparison of convergence speed among different models.

for flood disaster identification and provides more precise results, demonstrating higher stability and reliability throughout the training process.

To further illustrate the convergence speed of various algorithms, Table 3 presents a comparison of the training times across different models. This table details the average time required per epoch as well as the total training time for each model during the training process.

Table 3 compares the convergence speed of various models, including U-Net, ResU-Net, AttU-Net, Ren et al. (2024), and the proposed improved model. The data indicates that the proposed model outperforms others

in terms of convergence time, time per epoch, and total training time. Specifically, the proposed model achieves convergence in 85 epochs, significantly better than other models, with each epoch taking only 28 s—the shortest among all models—and a total training time of 45 h, reflecting higher training efficiency. In contrast, ResU-Net and AttU-Net require longer convergence times, with 90 and 95 epochs respectively, and slightly longer times per epoch. Although the Ren et al. (2024) model has a shorter time per epoch (29 s), it has the longest total training time of 52 h. Overall, the proposed model demonstrates superior training efficiency and faster convergence speed, providing a notable advantage for handling large-scale data in practical applications.

Discussion

The proposed SAR image-based flood disaster feature extraction model, which integrates TM with U-Net, significantly outperforms other models, including ResU-Net, U-Net, AttU-Net, and the models proposed by Ren et al. Comparative analysis reveals that this model exhibits superior performance in terms of loss value, with rapid and stable convergence, and achieves accuracy and precision rates of 95.52% and 90.11%, respectively, far exceeding those of other models. This aligns with the findings of Qin et al.⁵⁵, who reported significant improvements in hyperspectral image classification using a U-shaped convolution-assisted transformer network. Additionally, Zhang et al.⁵⁶ improved the efficiency and accuracy of change detection in remote sensing images by introducing deep and multi-scale feature representation alongside multi-attention mechanisms. Zhang et al.⁵⁷ also demonstrated that combining the feature extraction capabilities of CNNs with the multi-scale information fusion technology of transformers leads to efficient water body extraction tasks. The superior results in this work can be attributed to the effective integration of the multi-head attention mechanism, which enhances the model's ability to focus on key features, along with improvements to U-Net, including depthwise separable convolutions and customized loss functions. These findings are consistent with the perspectives of Zhang et al.

The achieved accuracy of 95.82% and precision of 90% validate the effectiveness of enhancing U-Net structure through depthwise separable convolutions and customized loss functions. However, it is important to note that the model's performance may be influenced by factors such as SAR image quality, resolution, and acquisition conditions. While the model demonstrates proficiency in flood hazard monitoring within a specific region, its ability to generalize to diverse geographic and climatic conditions has yet to be fully validated. Future research could focus on testing and optimizing the model across a broader range of conditions to strengthen its generalization capability. Additionally, incorporating multi-source data and multi-modal information, such as optical images and meteorological data, may further enhance the accuracy and efficiency of flood hazard monitoring.

Conclusion

This work successfully developed a model that integrates TM with U-Net for extracting flood features from SAR images. The model exhibits exceptional performance across key metrics, including loss value, detection effectiveness, accuracy, and precision, all exceeding 90.11%. These results highlight the model's significant advantage in accurately identifying flood-affected areas in SAR images, offering a robust technical foundation for timely flood monitoring and management. The development of this model significantly advances the automation of SAR data processing, enhancing flood monitoring technology. By integrating the enhanced U-Net with the TM layer, this work improves the model's robustness in handling complex backgrounds and noise, as well as its ability to extract detailed features. This approach not only advances remote sensing image analysis techniques but also paves the way for further innovations in natural disaster monitoring technologies. The improved model proposed here not only enhances the accuracy of flood disaster monitoring but also provides timely response information during different phases of flooding (pre-, during, and post-event), thereby offering a scientific basis for disaster management and risk assessment processes.

This work, however, has certain limitations. Although this work primarily tests and validates the model for riverine and rainstorm-induced floods, the proposed method demonstrates a certain degree of adaptability that can be extended to flood events in other regions around the globe. Given that the climatic, topographical, and hydrological characteristics of different areas may influence flood characteristics, the model's performance may be influenced by the quality, resolution, and acquisition conditions of SAR images. Although the model demonstrates strong results in flood monitoring within specific regions, its adaptability across various geographic and climatic conditions warrants further validation. Future global applications may require model parameters to be adjusted or retrained according to local characteristics to improve applicability and robustness. Such adjustments can enhance the method's performance in addressing various types of floods and geographical conditions, further validating its potential for global flood monitoring applications.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author Xu Feng on reasonable request via e-mail 18811360968@163.com.

Received: 4 September 2024; Accepted: 22 January 2025

Published online: 26 January 2025

References

1. Wang, X., Chen, W., Yin, J., Wang, L. & Guo, H. Risk assessment of flood disasters in the Poyang lake area. *Int. J. Disaster Risk Reduct.* **100**, 104208 (2024).
2. Balgah, R. A., Ngwa, K. A., Buchenrieder, G. R. & Kimengsi, J. N. Impacts of floods on agriculture-dependent livelihoods in Sub-Saharan Africa: an assessment from multiple geo-ecological zones. *Land* **12** (2), 334 (2023).
3. Liu, Z. et al. Research hotspots and frontiers of mountain flood disaster: bibliometric and visual analysis. *Water* **15** (4), 673 (2023).

4. Chowdhuri, I. et al. Evaluating the impact of climate change and geo-environmental factors on flood hazards in India: an integrated framework. *Geol. J.* **58** (9), 3515–3543 (2023).
5. Ambrosio-Ambrosio, J. P. & González-Camacho, J. M. Machine learning algorithms for classifying flood areas from synthetic aperture radar images. *Tecnol. Y Ciencias Del. Agua* **14** (4), 107–153 (2023).
6. Sanderson, J., Tengtrairat, N., Woo, W. L., Mao, H. & Al-Nima, R. R. XFIMNet: an explainable deep learning architecture for versatile flood inundation mapping with synthetic aperture radar and multi-spectral optical images. *Int. J. Remote Sens.* **44** (24), 7755–7789 (2023).
7. Wang, J. et al. Disaster mapping and assessment of Pakistan's 2022 mega-flood based on multi-source data-driven approach. *Nat. Hazards* **120** (4), 3447–3466 (2024).
8. Yu, M. & Gao, Q. Prolonged coastal inundation detected with synthetic aperture radar significantly retarded functional recovery of mangroves after major hurricanes. *Landsc. Ecol.* **38** (1), 169–183 (2023).
9. Mastro, P., Masiello, G., Serio, C. & Pepe, A. Change detection techniques with synthetic aperture radar images: experiments with random forests and Sentinel-1 observations. *Remote Sens.* **14** (14), 3323 (2022).
10. Aziz, M. A. et al. Delineating flood zones upon employing synthetic aperture data for the 2020 flood in Bangladesh. *Earth Syst. Environ.* **6** (3), 733–743 (2022).
11. Zhang, M. et al. Application of gated recurrent unit neural network for flood extraction from synthetic aperture radar time series. *Water* **15** (21), 3779 (2023).
12. Surampudi, S. & Kumar, V. Flood depth estimation in agricultural lands from L and C-band synthetic aperture radar images and digital elevation model. *IEEE Access* **11**, 3241–3256 (2023).
13. Rao, A., Jung, J., Silva, V., Molinaro, G. & Yun, S. H. Earthquake building damage detection based on synthetic-aperture-radar imagery and machine learning. *Nat. Hazards Earth Syst. Sci.* **23** (2), 789–807 (2023).
14. Zhang, S. et al. DaliWS: a high-resolution dataset with precise annotations for water segmentation in synthetic aperture radar images. *Remote Sens.* **16** (4), 720 (2024).
15. Patil, V., Khadke, Y., Joshi, A. & Sawant, S. Flood mapping and damage assessment using ensemble model approach. *Sens. Imaging.* **25** (1), 15 (2024).
16. Chen, B., Liu, Y., Zhang, Z., Lu, G. & Kong, A. W. K. Transattunet: multi-level attention-guided u-net with transformer for medical image segmentation. *IEEE Trans. Emerg. Top. Comput. Intell.* **8** (1), 55–68 (2023).
17. Yousef, R. et al. U-net-based models towards optimal MR brain image segmentation. *Diagnostics* **13** (9), 1624 (2023).
18. Zhang, Y. et al. Interactive medical image annotation using improved attention U-net with compound geodesic distance. *Expert Syst. Appl.* **237**, 121282 (2024).
19. Chen, H. et al. A landslide extraction method of channel attention mechanism U-Net network based on Sentinel-2A remote sensing images. *Int. J. Digit. Earth.* **16** (1), 552–577 (2023).
20. Zhang, G., Roslan, S. N. A. B., Wang, C. & Quan, L. Research on land cover classification of multi-source remote sensing data based on improved U-net network. *Sci. Rep.* **13** (1), 16275 (2023).
21. Nie, Y. et al. An improved U-net network for sandy road extraction from remote sensing imagery. *Remote Sens.* **15** (20), 4899 (2023).
22. Clark, A., Phinn, S. & Scarth, P. Optimised U-Net for land use–land cover classification using aerial photography. *PFG–J. Photogramm. Remote Sens. Geoinf. Sci.* **91** (2), 125–147 (2023).
23. Quan, Y. et al. Learning SAR-optical cross modal features for land cover classification. *Remote Sens.* **16** (2), 431 (2024).
24. Jian, Z. et al. An improved nested U-net network for fluorescence in situ hybridization cell image segmentation. *Sensors* **24** (3), 928 (2024).
25. Lee, Y. J., Jung, H. G. & Suhr, J. K. Semantic segmentation network slimming and edge deployment for real-time forest fire or flood monitoring systems using unmanned aerial vehicles. *Electronics* **12** (23), 4795 (2023).
26. Li, H., Xu, Z., Zhou, Y., He, X. & He, M. Flood monitoring using Sentinel-1 SAR for agricultural disaster assessment in Poyang lake region. *Remote Sens.* **15** (21), 5247 (2023).
27. Pech-May, F., Sanchez-Hernández, J. V., López-Gómez, L. A., Magaña-Govea, J. & Mil-Chontal, E. M. Flooded areas detection through SAR images and U-NET deep learning model. *Computación Y Sistemas.* **27** (2), 449–458 (2023).
28. Xu, X., Zhang, T., Liu, H., Guo, W. & Zhang, Z. An information-expanding network for water body extraction based on U-net. *IEEE Geosci. Remote Sens. Lett.* **21**, 1502205 (2024).
29. Wang, X. et al. Double U-net (W-Net): a change detection network with two heads for remote sensing imagery. *Int. J. Appl. Earth Obs. Geoinf.* **122**, 103456 (2023).
30. Yin, M., Wang, P., Hao, W. & Ni, C. Cloud detection of high-resolution remote sensing image based on improved U-Net. *Multimed. Tools Appl.* **82** (16), 25271–25288 (2023).
31. Wang, Q. et al. High-resolution remote sensing image change detection method based on improved siamese U-net. *Remote Sens.* **15** (14), 3517 (2023).
32. Ren, S., Zhou, F. & Bruzzone, L. Transfer-aware graph U-Net with cross-level interactions for PolSAR image semantic segmentation. *Remote Sens.* **16** (8), 1428 (2024).
33. Amir, Z. et al. Rural settlements segmentation based on deep learning U-net using remote sensing images. *Int. Arch. Photogramm. Remote Sens. Spat. Inform. Sci.* **48**, 1–5 (2023).
34. Hou, J., Hou, B., Zhu, M., Zhou, J. & Tian, Q. Sensitivity analysis of parameters of U-Net model for semantic segmentation of silt storage dams from remote sensing images. *Can. J. Remote Sens.* **49** (1), 2178834 (2023).
35. Maurya, A., Mittal, P. & Kumar, R. A modified u-net-based architecture for segmentation of satellite images on a novel dataset. *Ecol. Inf.* **75**, 102078 (2023).
36. Yu, C., Hong, L., Pan, T., Li, Y. & Li, T. ESTUGAN: enhanced swin transformer with U-net discriminator for remote sensing image super-resolution. *Electronics* **12** (20), 4235 (2023).
37. Yang, P., Wang, M., Yuan, H., He, C. & Cong, L. Using contour loss constraining residual attention U-net on optical remote sensing interpretation. *Visual Comput.* **39** (9), 4279–4291 (2023).
38. Guo, Z. et al. Water-body segmentation for SAR images: past, current, and future. *Remote Sens.* **14** (7), 1752 (2022).
39. Cao, H., Tian, Y., Liu, Y. & Wang, R. Water body extraction from high spatial resolution remote sensing images based on enhanced U-Net and multi-scale information fusion. *Sci. Rep.* **14** (1), 16132 (2024).
40. Talat, N. et al. A novel enhanced normalization technique for a mandible bones segmentation using deep learning: batch normalization with the dropout. *Multimed. Tools Appl.* **82** (4), 6147–6166 (2023).
41. Rudiansyah, R., Kesuma, L. I. & Anggara, M. I. Implementation of image quality improvement methods and lung segmentation on chest X-Ray images using U-Net architectural modifications. *Comput. Eng. Appl. J.* **12** (2), 71–78 (2023).
42. Chang, H. H., Yeh, S. J., Chiang, M. C. & Hsieh, S. T. RU-Net: skull stripping in rat brain MR images after ischemic stroke with rat U-Net. *BMC Med. Imaging* **23** (1), 44 (2023).
43. Dan, Y., Jin, W., Yue, X. & Wang, Z. Enhancing medical image segmentation with a multi-transformer U-Net. *PeerJ* **12**, e17005. (2024).
44. Yecchuri, S. & Vanambathina, S. D. Sub-convolutional U-Net with transformer attention network for end-to-end single-channel speech enhancement. *EURASIP J. Audio Speech Music Process.* **2024**(1), 8 (2024).
45. Chen, J. et al. ETU-Net: efficient transformer and convolutional U-style connected attention segmentation network applied to endoscopic image of epistaxis. *Front. Med.* **10**, 1198054 (2023).

46. Wu, M., Qian, Y., Liao, X., Wang, Q. & Heng, P. A. Hepatic vessel segmentation based on 3D swin-transformer with inductive biased multi-head self-attention. *BMC Med. Imaging*. **23** (1), 91 (2023).
47. Parikh, H., Patel, S. & Patel, V. Classification of SAR and PolSAR images using deep learning: a review. *Int. J. Image Data Fusion*. **11** (1), 1–32 (2020).
48. Melgar-García, L., Martínez-Álvarez, F., Bui, T., Troncoso, A. & D., & A novel semantic segmentation approach based on U-Net, WU-Net, and U-Net++ deep learning for predicting areas sensitive to pluvial flood at tropical area. *Int. J. Digit. Earth*. **16** (1), 3661–3679 (2023).
49. Barakat, A. & Bianchi, P. Convergence and dynamical behavior of the ADAM algorithm for nonconvex stochastic optimization. *SIAM J. Optim.* **31** (1), 244–274 (2021).
50. Liu, Z., Su, B. & Lv, F. Intelligent identification method of crop species using improved U-Net network in UAV remote sensing image. *Sci. Program*. **2022**(1), 9717843 (2022).
51. Zeng, K. & Wang, Y. A deep convolutional neural network for oil spill detection from spaceborne SAR images. *Remote Sens.* **12** (6), 1015 (2020).
52. Ghorbanzadeh, O., Gholamnia, K. & Ghamisi, P. The application of ResU-net and OBIA for landslide detection from multi-temporal sentinel-2 images. *Big Earth Data*. **7** (4), 961–985 (2023).
53. Peng, F., Peng, Q., Chen, D., Lu, J. & Song, Y. Extraction of Terraces in Hilly areas from Remote sensing images using DEM and Improved U-Net. *Photogramm. Eng. Remote Sens.* **90** (3), 181–188 (2024).
54. Atiampo, A. K. & Diédié, G. H. F. New fusion approach of spatial and channel attention for semantic segmentation of very high spatial resolution remote sensing images. *Open J. Appl. Sci.* **14** (02), 288–319 (2024).
55. Qin, R., Wang, C., Wu, Y., Du, H. & Lv, M. A U-shaped convolution-aided transformer with double attention for hyperspectral image classification. *Remote Sens.* **16** (2), 288 (2024).
56. Zhang, M., Liu, Z., Feng, J., Liu, L. & Jiao, L. Remote sensing image change detection based on deep multi-scale multi-attention siamese transformer network. *Remote Sens.* **15** (3), 842 (2023).
57. Zhang, Q., Hu, X. & Xiao, Y. A novel hybrid model based on cnn and multi-scale transformer for extracting water bodies from high resolution remote sensing images. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inform. Sci.* **10**, 889–894 (2023).

Acknowledgements

This work was supported by “Key Research and Development Program of Shaanxi (Program No.2023-YB-NY-042, 2023-YBNY-223, 2024NC-YBXM-207, and 2024NC-YBXM-219)”.

Author contributions

Fajing Wang: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation Xu Feng: writing—review and editing, visualization, supervision, project administration, funding acquisition.

Declarations

Competing interests

The authors declare no competing interests.

Ethics statement

The studies involving human participants were reviewed and approved by School of Transportation and Geomatics Engineering, Yangling Vocational & Technical College Ethics Committee (Approval Number: 2022.865423). The participants provided their written informed consent to participate in this study. All methods were performed in accordance with relevant guidelines and regulations.

Additional information

Correspondence and requests for materials should be addressed to X.F.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025