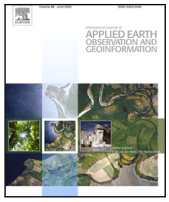
International Journal of Applied Earth Observation and Geoinformation 127 (2024) 103662

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Residual wave vision U-Net for flood mapping using dual polarization Sentinel-1 SAR imagery☆

Ali Jamali [a,*], Swalpa Kumar Roy [b], Leila Hashemi Beni [c], Biswajeet Pradhan [d], Jonathan Li [e], Pedram Ghamisi [f,g]

[a] Department of Geography, Simon Fraser University, 8888 University Dr, Burnaby, BC V5A 1S6, Canada
[b] Department of Computer Science and Engineering, Alipurduar Government Engineering and Management College, West Bengal 736206, India
[c] Geomatics Program, Department of Built Environment, North Carolina A&T State University, Greensboro, NC 27411, USA
[d] The Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), Faculty of Engineering and Information Technology, University of Technology Sydney, Ultimo, NSW 2007, Australia
[e] Department of Geography and Environmental Management, University of Waterloo, Ontario, ON N2L 3G1, Canada
[f] Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Helmholtz Institute Freiberg for Resource Technology, 09599 Freiberg, Germany
[g] Institute of Advanced Research in Artificial Intelligence (IARAI), 1030 Vienna, Austria

ARTICLE INFO

Keywords:
Flood mapping
Deep learning
Semantic segmentation
Vision MLP
ResU-Net

ABSTRACT

The increasing severity, duration, and frequency of destructive floods can be attributed to shifts in climate, infrastructure, land use, and population demographics. Obtaining precise and timely data about the extent of floodwaters is crucial for effective emergency preparedness and mitigation efforts. Deep convolutional neural networks (CNNs) have shown astonishing effectiveness in various remote sensing applications, including flood mapping. One of the key limitations of CNNs is that they can only predict whether a desired feature will appear in an image, not where it can be recognized. To address this limitation, the incorporation of self-attention mechanisms deployed in vision transformers (ViTs) can be particularly effective. However, the self-attention modules in the ViTs are complex and computationally expensive, and they require a wealth of ground data to attain their full capability in image classification/segmentation. Thus, in this paper, we develop the Residual Wave Vision U-Net (WVResU-Net), a deep learning segmentation architecture that utilizes advanced Vision Multi-Layer Perceptrons (MLPs) and ResU-Net for accurate and reliable flood mapping using Sentinel-1 SAR's dual polarization data. Results showed the significant superiority of the developed WVResU-Net algorithms over several well-known CNN and ViT deep learning models, including Swin U-Net, U-Net+++, Attention U-Net, R2U-Net, ResU-Net, TransU-Net and TransU-Net++. For example, the segmentation accuracy of TransU-Net++, SwinU-Net, ResU-Net, R2U-Net, Attention U-Net, TransU-Net, and U-Net+++, was significantly improved by approximately 5, 12, 13, 13, 16, 19, and 23 percentage points, respectively in terms of recall obtained by the WVResU-Net with a recall value of about 69.67%. The code will be made publicly available at https://github.com/aj1365/RWVUNet.

1. Introduction

Floods are among the most common natural disasters, and their frequency has been alarmingly increasing over the past few decades (Schumann et al., 2009). The global economic and social impact of flooding is expected to worsen due to the effects of global warming and growing populations (Intergovernmental Panel on Climate Change (IPCC), 2014). From 2000 to 2019, the estimated annual global cost of flood damage was 65 billion dollars (Tellman et al., 2021). In the years ahead, both the associated costs and the number of people affected by this natural disaster are projected to rise. According to climate change predictions, the global population affected by floods is set to double by 2030, increasing from 72 million to 147 million (Islam and Meng, 2022). To efficiently manage and evaluate this growing flood risk, tools such as Earth Observation (EO) are essential for accurately mapping flooded areas (McCormack et al., 2022).

☆ This research was partially funded by NOAA, United States and NASA, United States.
* Corresponding author.
E-mail addresses: alij@sfu.ca (A. Jamali), swalpa@agemc.ac.in (S.K. Roy), lhashemibeni@ncat.edu (L. Hashemi Beni), biswajeet.Pradhan@uts.edu.au (B. Pradhan), junli@uwaterloo.ca (J. Li), p.ghamisi@gmail.com (P. Ghamisi).

https://doi.org/10.1016/j.jag.2024.103662
Received 12 October 2023; Received in revised form 1 January 2024; Accepted 15 January 2024
Available online 20 January 2024
1569-8432/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Utilizing satellite data to its fullest extent is crucial for flood management because it provides a near real-time and cost-effective method for mapping floods operationally (Giustarini et al., 2016; Pappas et al., 2021). Moreover, satellite-derived flood zone visualizations can significantly improve the ability to predict floods by enabling more precise hydraulic model validation and calibration (Grimaldi et al., 2016; Wood et al., 2016), incorporating flooded area data (Nguyen et al., 2022), or integrating spatially distributed water levels obtained from digital elevation models (DEMs) (García-Pintado et al., 2013; Moya et al., 2022; Surampudi and Kumar, 2023). One critical element in effective flood disaster mitigation is the determination of inundation extent from satellite imagery. With an increasing quantity of satellites and advancements in image analysis technology, this information can be obtained more frequently and with higher quality (Martinis et al., 2022).

The most widely used tool for mapping floods from space is Synthetic Aperture Radar (SAR) equipment carried by satellites, thanks to its ability to collect data 24/7 regardless of weather conditions (Tripathy and Malladi, 2022; Tazmul Islam and Meng, 2022a). SAR is an active system that directs microwave pulses at the target from an oblique angle. Surface roughness is the primary factor affecting the quantity of microwave energy scattered off a feature, while shape and dielectric characteristics have secondary effects (Woodhouse, 2017). Rough terrestrial terrains exhibit high backscatter because they scatter energy in multiple directions, including back toward the sensor. On the other hand, open water has a relatively smooth surface, reflecting radar energy away from the sensor which results in low backscatter, appearing as dark regions (Henderson and Lewis, 2008). Many flood mapping techniques, such as automatic thresholding (Tiwari et al., 2020), region growing (Wan et al., 2019), Bayesian model averaging (Liu and Merwade, 2018), supervised pixel-based classification (Tong et al., 2018), unsupervised pixel-based classification (Zhang et al., 2021c), object-oriented classification (Zhang et al., 2021a), and change detection (Hamidi et al., 2023), leverage this characteristic due to the typically high land-water contrast it provides.

Flood mapping using SAR satellite imagery is a complex and challenging task due to its inherent complexity and the presence of speckle noise. During the imaging process, electromagnetic radiation can be influenced by echo interference when it encounters rough surfaces, leading to phase variations. This interference results in decreased echo intensity and introduces speckle noise into the SAR data. Speckle noise can obscure target details in the SAR image, making it difficult to recognize and identify SAR targets (Zhai et al., 2023). Moreover, various parameters, including the local incidence angle and the geometrical shapes of SAR imaging at the time of acquisition, significantly affect backscatter variation (O'Grady et al., 2013). It should be noted that this study is not focused on addressing the discussed issues inherent in the SAR imaging systems, however, we are interested in significantly improving the segmentation performance of the current state-of-the-art models by incorporating advanced computer vision techniques. Effective emergency and land management depends on a comprehensive understanding of flood dynamics at a large scale.

A practical approach to generating maps of flooded zones over extensive areas is the utilization of machine learning techniques. Recently, traditional machine learning classifiers, such as Random Forest, which are easy to implement, have been widely employed for mapping of flooded areas (Woznicki et al., 2019; Feng et al., 2015; Vamsi et al., 2023). However, traditional algorithms like Random Forest face several key challenges: (1) The accuracy and performance of traditional classifiers heavily depend on manual feature engineering. (2) These techniques often struggle to capture complex feature characteristics, particularly in densely populated and cultivated regions where textual information is intricate. (3) Conventional algorithms overlook spatial information and the spatial correlation among neighboring pixels, relying solely on spectral or backscattering data. Deep Convolutional Neural Networks (DCNNs) address these challenges and have shown astonishing effectiveness in various applications, including image segmentation (Zhang et al., 2023), object detection (Ye et al., 2023), and image classification (Jackson et al., 2023; Qiao et al., 2023). In recent years, there has been a significant rise in the utilization of new computer vision-based techniques in various remote sensing applications (Li et al., 2023; Wu et al., 2023; Yokoya et al., 2022; Chang and Ghamisi, 2023; Yue et al., 2022). For instance, (Li et al., 2023) introduced a low-rank representation (LRR) algorithm (LRR-Net), a hyperspectral anomaly detection model that combines deep learning methods with the LRR approach. The model utilizes prior expertise across the deep network for controlling parameter optimization. Wu et al. (2023) presented an efficient "U-Net in U-Net" deep learning architecture for the detection of small entities in infrared images. The developed model enables multi-scale and multi-level learning of the representation of various features by embedding a tiny U-Net into a larger U-Net backbone. Moreover, to enhance the AI model's capacity for generalization from multi-city locations, (Hong et al., 2023) created a high-resolution domain adaptation network (HighDAN). Through adversarial learning, HighDAN can close the gap caused by the significant variations in RS image representations between various cities while maintaining the geographically topological structure of the examined urban environment. In Persello et al. (2022), the intersection of deep learning and Earth observation and their contribution to sustainable development goals has been explored. Unlike other machine learning classification algorithms, CNNs employ a range of convolutions to establish a hierarchical representation of information and can learn feature representations from large datasets (Chen et al., 2016). This enhanced learning capability results in improved accuracy and performance. For instance, Muñoz et al. (2021) combined dual-polarized SAR and multi-spectral Landsat imagery to assess the efficiency of a data fusion framework based on CNNs for enhanced flood mapping. These studies underscore the crucial role that deep learning algorithms can play in advancing flood classification.

One of the key drawbacks of CNNs is that they can only predict whether a desired feature will appear in an image, but not where it can be recognized. Because of this, Long et al. (2015) suggested a fully convolutional network (FCN) that enables the classification of every pixel as a particular class, i.e., a pixel-based approach for classification known as semantic segmentation, in contrast with image classification, to address this limitation and increase the range of applications of CNNs. One of the many FCN structures follows the U-Net architecture, developed by Ronneberger et al. (2015). This framework concatenates output maps at different levels to boost the effectiveness of segmentation tasks. The primary distinction between the U-Net architecture and conventional FCNs can be understood by the higher number of expansion networks that allow the algorithm to disseminate information through higher resolution levels. Moreover, the CNNs' fundamental constraints in the backbone restrict them from precisely determining the sequential characteristics of an imagery. By effectively utilizing the mechanism of self-attention deployed in vision transformers (ViTs), this particular issue can be dealt with. For example, by creating a UNet-based transformer (UNetFormer) model for urban landscape segmentation, Wang et al. (2022) investigated a lightweight version of ResNet-18 encoder for obtaining a robust global-local attention system for modeling both local and global knowledge in the decoder. Additionally, in order to enhance feature representation capability while achieving better results on segmentation of cardiac images, Yuan et al. (2023) developed a model that combined both Transformers and CNNs.

Recently, due to the complexity and computational expense of the self-attention modules in Vision Transformers (ViTs), simpler architectures based on multi-layer perceptrons (MLPs), such as the MLP-Mixer model (Tolstikhin et al., 2021), have gained popularity in computer vision tasks. These vision MLPs offer greater versatility across tasks and introduce fewer inductive biases compared to Convolutional Neural Networks (CNNs) and ViTs (Tang et al., 2022a). Additionally,

while ViTs require a substantial amount of reference data to maximize their image classification performance, CNNs also demand a significant quantity of ground-truth information to reach their full potential. Advanced MLPs have emerged as viable alternatives to both deep CNNs and ViTs, particularly in situations where limited training data is available (Tolstikhin et al., 2021). In this paper, we present the Residual Wave Vision U-Net, hereafter referred to as WVResU-Net, which is a deep-learning segmentation framework. We utilize advanced vision MLPs and ResU-Net to achieve precise and reliable flood mapping using dual-polarization data from Sentinel-1 SAR. The contributions and novelties of this study can be described as:

- Development of a deep learning-based flood mapping architecture that efficiently incorporates a vision MLP mechanism and residual connections within a U-Net, significantly improving the segmentation accuracy of the base ResU-Net segmentation architecture.
- Incorporating the vision MLP to address a key limitation of CNNs and ViTs, namely their dependency on a large number of labeled data and computational efficiency.
- Utilization of wave vision, which greatly enhances the decoder network's ability to identify where to find relevant discriminative and task-specific information.

The proposed deep learning-based segmentation model of WVResU-Net significantly improves the segmentation capability of current cutting-edge deep learning models, including Swin U-Net (Cao et al., 2021), U-Net+++ (Huang et al., 2020), Attention U-Net (Oktay et al., 2018), R2U-Net (Alom et al., 2018), ResU-Net (Diakogiannis et al., 2020), TransU-Net (Chen et al., 2021), and TransU-Net++ (Jamali et al., 2023), by effectively incorporating the capabilities of advanced vision-based MLPs and residual connections. The obtained segmentation results demonstrate a substantial segmentation enhancement by the proposed model over the other CNN- and ViT-based algorithms in accurately mapping flooded regions by using solely dual-polarized Sentinel-1 SAR imagery.

2. Proposed segmentation framework

The flowchart of the proposed methodology for flood mapping using dual polarized Sentinel-1 SAR imagery is illustrated in Fig. 1. Various polarization features are extracted from the dual-polarized Sentinel-1 SAR imagery. Python TensorFlow GPU 2.6.2 is used for deep learning model development and the performance of each segmentation architecture for flood mapping is assessed. More details on the procedures are discussed in the following sections.

Given a set of data ($\mathbf{X_{SAR}}$, $\mathbf{Y}$), where $\mathbf{X_{SAR}}$ represents the input Synthetic-Aperture Radar (SAR) backscattering coefficients for the segmentation model, and $\mathbf{Y}$ represents binary or flooding maps. Considering the Sentinel-1 SAR imagery, $\mathbf{X_{SAR}} \in \mathbb{R}^{w \times h \times bc}$, where w and h represent spatial width and height, respectively, and $bc = \{2, \dots, n\}$ is the number of input backscattering features (i.e., VV, VH, and the extracted backscattering features), the objective is to generate a segmentation map, $\mathbf{Y} \in \mathbb{R}^{w \times h \times bc}$, where $bc \in \{VV, VH\}$. This can be achieved by estimating the output of the pixel class using $\mathbf{Y} = \mathcal{F}(\mathbf{X_{SAR}})$ for the input $\mathbf{X_{SAR}}$ having the same spatial size as $(w \times h)$. We have developed and introduced the Residual Wave Vision U-Net (WVResU-Net), as depicted in Fig. 2, which is an enhanced ResU-Net framework for the segmentation of flooding regions. The proposed WVResU-Net integrates the characteristics of vision MLPs in the form of a U-shaped ResU-Net architecture. There are several significant advantages to using the WVResU-Net: (1) The skip connections in a residual unit, which convey low-level features to their corresponding high-level feature representations, enhance the propagation of information without degradation. This enables us to build a lower-complexity segmentation model that gains more effective semantic segmentation knowledge with a limited quantity of labeled data. (2)The use of residual learning facilitates efficient network training. (3) The Wave Vision function treats each image patch as a wave operation with two essential key elements: amplitude and phase. Amplitude corresponds to the initial backscattering coefficients, while the phase is estimated as a complex value that varies relative to the semantic information of the input Sentinel-1 SAR backscattering features.

2.1. Residual learning

It utilizes skip connections to facilitate training by allowing the model to focus on learning residual (difference) information. This approach enhances gradient flow and enables the creation of deeper networks. Increasing the number of convolutional layers, thereby creating deeper CNN models, can enhance the projection capabilities of CNNs. However, this approach can sometimes affect the accuracy of the transmission of model information during back-propagation, resulting in the common issue of the vanishing gradient problem (He et al., 2016). To address this shortcoming, traditional convolutional blocks have been replaced with a residual unit. By incorporating skip connections and residual learning, this unit effectively addresses the issue of gradient vanishing in the training phase. The inclusion of a skip connection within a residual unit simplifies the process of converting a low-level feature representation into its relative high-level counterpart.

2.2. Vision multi-layer perceptron

MLP algorithms primarily consist of two separate blocks: the channel-mixing MLP and the token-mixing MLP. Both of these blocks are composed of layers of fully connected and activation functions. The channel-mixing MLP converts the features of every token, while the token-mixing MLP aims to combine features from various tokens. By alternately stacking these two types of MLP blocks, the basic MLP structure can effectively extract features and excel in vision-related tasks (Tang et al., 2022a). Recent research in computer vision has demonstrated that a pure MLP architecture, such as MLP-Mixer (Tolstikhin et al., 2021), primarily comprising fully connected (FC) layers, can rival CNNs and Vision Transformers (ViTs) in terms of its ability to extract effective features. An MLP algorithm essentially constitutes a neural network with FC layers and non-linear activation functions.

In the context of the vision MLP, SAR imagery is initially divided into numerous patches, often referred to as tokens. The features of these tokens are then captured using two key components: the token-FC (TFC) and the channel-FC (CFC), as described in the following sections. Considering the intermediate feature maps with n tokens as $T = [t_1, t_2, \dots, t_n]$, where each token t_j represents a d-dimensional vector, the CFC can be expressed as:

$$CFC(t_j, W^c) = W^c t_j, \quad j = 1, 2, 3, \dots, n \tag{1}$$

where learnable weights are expressed by W^c. In order to obtain the features from every token, the channel-FC performs a separate procedure on its various CFC layers which are typically stacked together with a non-linear activation function, creating a channel-mixing MLP, to improve the transformation capability. To combine information from various tokens, the TFC process is required and defined by:

$$TFC(T, W^m)_j = \sum_k W^m_{jk} \odot t_k, \quad j = 1, 2, 3, \dots, n \tag{2}$$

where weights of token-mixing are expressed as W^m, element-wise multiplication is presented as $\odot$, and the jth output token is defined by index j. By combining characteristics from various tokens, the TFC aims to acquire spatial information. In such a way, we are limiting the power of MLPs by ignoring meaningful semantic information of tokens derived from different input images by such a simple token-mixing operation with fixed weights.

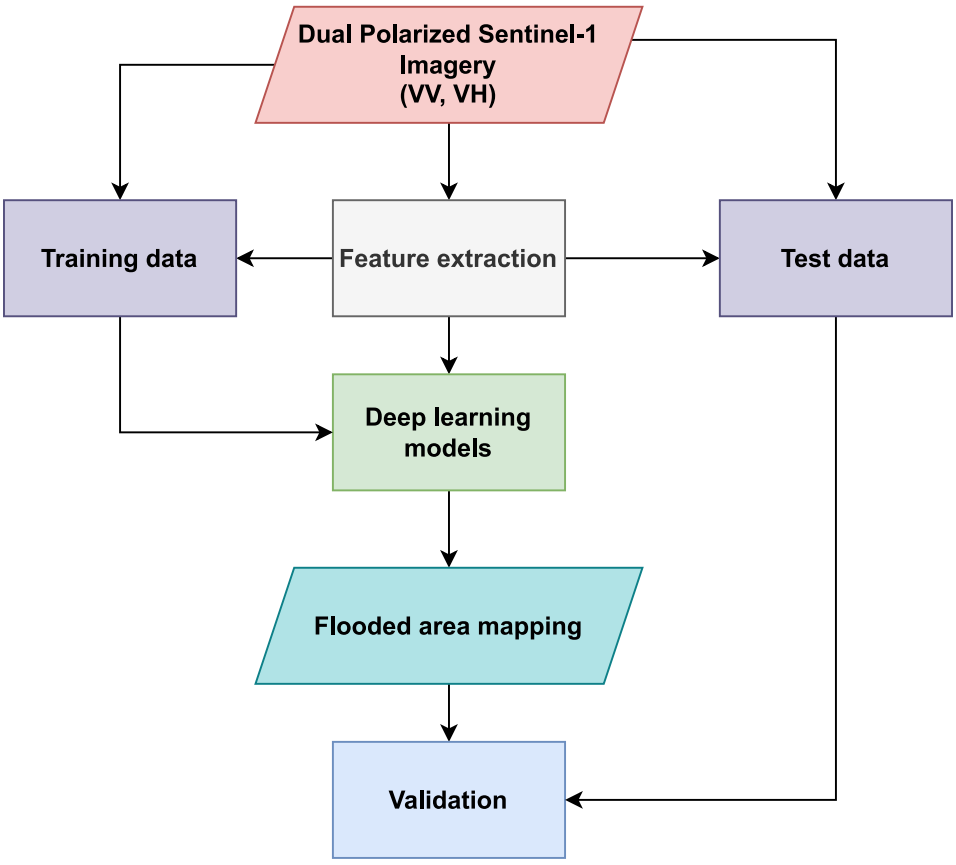


Fig. 1. The flowchart of the proposed methodology for flood mapping using dual-polarized Sentinel-1 SAR imagery.

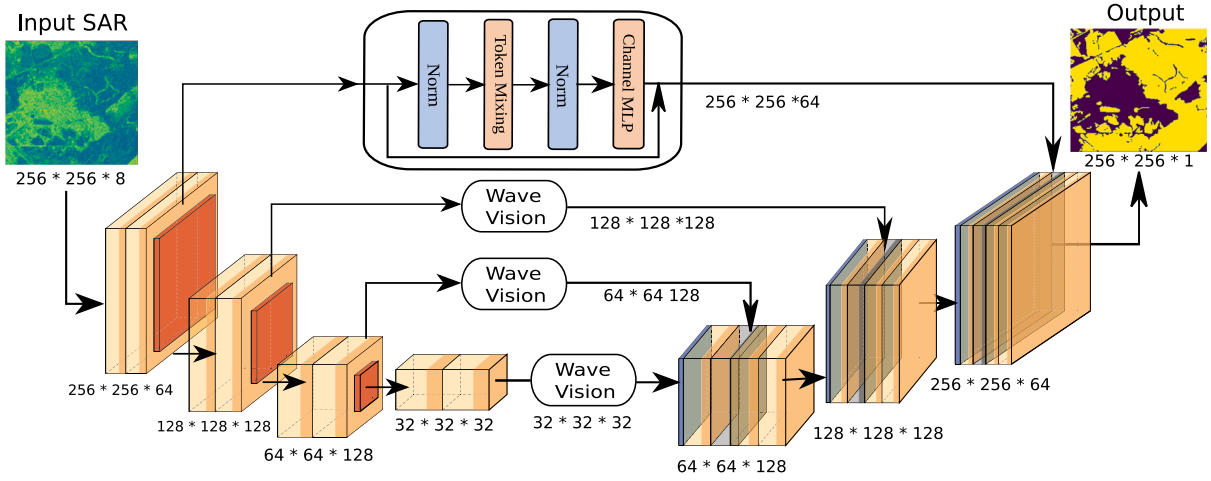


Fig. 2. Overall design of the proposed Residual Weve Vision U-Net architecture.

2.3. Wave vision multi-layer perceptron

As discussed, the current state-of-the-art MLP algorithms directly combine tokens from various input images with fixed weights, disregarding the differing semantic data of the image patches derived from various images. In contrast, we can consider each token as a wave operation with two key elements: amplitude and phase, as suggested by Tang et al. (2022a). The initial feature is considered as the amplitude, and the phase is calculated as a complex value that varies based on the semantic information of the input images (e.g., backscattering coefficients). The phase modulates the MLP interaction between tokens and its fixed weights. The aggregate output of these wave-like tokens is influenced by the variation in phase among them, and tokens with similar phases often enhance each other, as shown in Fig. 3. Thus, the correlation between tokens and predetermined weights in a typical MLP can be dynamically altered effectively by introducing the phase variable. A token can be considered as a wave, denoted as $t_j^{\sim}$, and can be defined as:

$$t_j^{\sim} = |t_j| \odot e^{i\theta_j}, \quad j = 1, 2, 3, \ldots, n \tag{3}$$

where the imaginary unit is expressed as i ($i^2 = -1$). The contextual information of each token is defined by the real value $|t_j|$, and $e^{i\theta_j}$ represents a periodic operation. The phase is expressed as θ_j, indicating the positional information of each token within a wave. Thus, each token

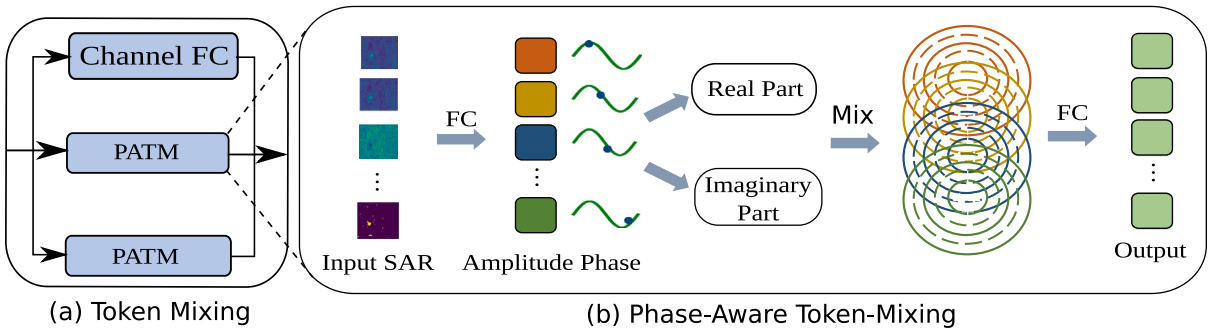


Fig. 3. Unfolding the steps of the Token Mixing module.

$t_j^{\sim}$, comprising amplitude and phase, is considered as a complex-valued entity.

For instance, considering two tokens $t_1^{\sim}$ and $t_2^{\sim}$, the amplitude (t_r) and phase (θ_r) can be defined as:

$$|t_r| = \sqrt{|t_i|^2 + |t_j|^2 + 2|t_i| \odot |t_j| \odot \cos(\theta_j - \theta_i)} \tag{4}$$

$$\theta_r = \theta_i + \text{atan2}(|t_j| \odot \sin(\theta_j - \theta_i), |t_i| + |t_j| \odot \cos(\theta_j - \theta_i)) \tag{5}$$

where $\text{atan2}(x, y)$ defines a two-variable arc tangent operation. Thus, considering the phase difference ($\theta_j - \theta_i$), it has a significant effect on the amplitude of the combined result (t_r).

Considering the output feature maps of the encoders $\widehat{X} = [\widehat{x_1}, \widehat{x_2}, \dots, \widehat{x_n}]$ as input of the wave vision, and the amplitude of t_j is defined by:

$$t_j = CFC(\widehat{x_j}, W^c). \quad j = 1, 2, 3, \dots, n \tag{6}$$

A token as a wave comprising real (amplitude) and complex values (phase) can be rewritten as:

$$t_j^{\sim} = |t_j| \odot \cos(\theta_j) + i|t_j| \sin(\theta_j). \quad j = 1, 2, 3, \dots, n \tag{7}$$

Resultant complex value output tokens can be aggregated through the token-mixing operation, as expressed by:

$$O_j^{\sim} = TFC(T^{\sim}, W^m)_j. \quad j = 1, 2, 3, \dots, n \tag{8}$$

The real value O_j can be estimated through the sum of real and imaginary parts of the $O_j^{\sim}$ as defined by:

$$O_j = \sum_k W_{jk}^m t_k \odot \cos(\theta_k) + W_{jk}^i t_k \odot \sin(\theta_k), \quad j = 1, 2, 3, \dots, n, \tag{9}$$

where learnable weights are presented by W^m and W^i, and phase is expressed as θ_k.

2.3.1. Residual wave vision U-Net model

In order to enhance the robustness of flood mapping, an improved ResU-Net (*WVResU-Net*) algorithm is developed, which leverages the advantages of residual learning and Wave Vision to precisely map flooded regions using Sentinel-1 SAR imagery in a U-shaped model. Consider an 8-channel image dataset, where images of size (256×256×8) pixels serve as input to the proposed deep learning network, *WVResU-Net*. In the encoding part, there are three residual blocks (R) and one bridging (bottleneck) (Br) block, as defined by:

$$R = \text{Conv}(\text{BN}(\text{ReLU}(\text{Conv}(x)))) + x \tag{10}$$

$$Br = \text{BN}(\text{ReLU}(\text{Conv}(\text{BN}(\text{ReLU}(\text{Conv}(x)))))) + x \tag{11}$$

The output maps of the first, second, and third encoding blocks are fed to the second, third, and bridge encoding blocks, respectively. Then, the feature map of the bridge block is fed to the vision block and upsampled in the first decoding block. The resultant feature map is concatenated with the vision block positioned after the third encoding residual block, resulting in feature map X_{de1}. The next operations in the decoding block De_1 can be expressed as:

$$De_1 = \text{BN}(\text{ReLU}(\text{Conv}(\text{BN}(\text{ReLU}(\text{Conv}(X_{de1})))))) + X_{de1}. \tag{12}$$

The resultant feature map is then upsampled, and the output map is concatenated with the vision block positioned after the second encoding block, resulting in feature map X_{de2}. The next operations in the second decoding block De_2 can be defined as:

$$De_2 = \text{BN}(\text{ReLU}(\text{Conv}(\text{BN}(\text{ReLU}(\text{Conv}(X_{de2})))))) + X_{de2}. \tag{13}$$

The output map of the second decoding block is then upsampled, and the feature map is concatenated with the vision block positioned after the first encoding block, resulting in the feature map X_{de3}. The next operations in the third decoding block De_3 can be expressed as:

$$De_3 = \text{BN}(\text{ReLU}(\text{Conv}(\text{BN}(\text{ReLU}(\text{Conv}(X_{de3})))))) + X_{de3}. \tag{14}$$

With a kernel size of (1×1), a 2D convolution layer with a sigmoid activation operation is used at the final stage of decoding to map the resultant feature output of the final decoding residual block into the desired flooded area maps.

2.4. Comparison algorithms

The proposed segmentation model, WVResU-Net, is assessed over various well-known segmentation algorithms of Swin U-Net (Cao et al., 2021), U-Net+++ (Huang et al., 2020), Attention U-Net (Oktay et al., 2018), R2U-Net (Alom et al., 2018), ResU-Net (Diakogiannis et al., 2020), TransU-Net (Chen et al., 2021), and TransU-Net++ (Jamali et al., 2023). The binary cross-entropy was employed as the loss function in this study in all the developed segmentation models as expressed:

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^{|N|} (y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)) \tag{15}$$

where y and p present the flood labels and predicted probability of flood values. Additionally, the learning rate and number of the epoch are set as 0.001, and 40, while the optimizer was set as $Adam$ optimizer. It should be noted that we used the comparison models with their original parameters without any further modification except for the input image size in all the implemented deep learning models (i.e., we used input images of size 256×256).

2.5. Accuracy evaluation metrics

The outcomes of the flood area segmentation for the developed deep learning models are evaluated in overall accuracy (OA), precision, F-1 score, recall, and dice coefficient, which have been used frequently in relevant studies (Cai et al., 2021; Ghorbanzadeh et al., 2022).

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{16a}$$

$$Precision = \frac{TP}{TP + FP} \tag{16b}$$

$$Recall = \frac{TP}{TP + FN} \tag{16c}$$

$$F1\text{-}score = \frac{2 * TP}{2 * TP + FP + FN} \tag{16d}$$

$$Dice\ \ Coefficient = \frac{2 * TP}{(TP + FP) + (TP + FN)} \tag{16e}$$

The FP, TP, and FN show false positive, true positive, and false negative values, respectively.

2.6. Experimental data and settings

In this research, 542 "chips" from flooding occurrences around the world, each containing dual polarization data, including VV and VH, make up the Sentinel-1 SAR data. Each chip includes an associated flood label mask that identifies the water-containing pixels in a scene. 13 flood events' worth of training data are included. Each unique incident has between 15 and 69 chips (30 and 138 images), with half of the events having fewer than 32 chips (64 images). The data spans a broad geographical region. For the flooding that occurred in the United States, Paraguay, India, and Slovakia, there are more than 60 chips (120 images). The data can be accessed through *Microsoft AI for Earth STAC API*.[1] On the one hand, it has been discussed that the classification performance of deep learning models can be significantly enhanced by increasing the number of input features (Tang et al., 2022b; Han et al., 2022; Mayer et al., 2021). On the other hand, using various backscattering coefficients extracted from the dual-polarized Sentinel-1 imagery has been used and discussed to improve the classification/segmentation capability of various classification algorithms in recognizing different features of interest from Sentinel-1 SAR imagery, specifically for flood mapping (Zhang et al., 2021b; Tazmul Islam and Meng, 2022b). For instance, for flood mapping, (Tazmul Islam and Meng, 2022b) generated several polarization combinations from Sentinel-1 imagery, including $VV + VH$, $VH - VV$, $VV * VH$. Thus, to improve the segmentation performance of deep learning algorithms and increase the number of polarization features, from the dual polarization Sentinel-1 SAR imagery, we derived several polarization features, including $VV+VH$, $VH-VV$, $VV * VV$, $VH * VH$, $VV * VH$ and $(VV+VH) * (VH - VV)$, as shown in Fig. 4. It is worth mentioning that for flood mapping, we used all the discussed dual polarization coefficients and extracted features.

3. Segmentation results

As shown in Table 1, the *WVResU-Net* segmentation model achieved the best scores for overall accuracy (96.2%), F1-score (82.03%), recall (69.67%) and dice coefficient (0.7345) over the other segmentation architectures: *Swin U-Net, Attention U-Net, R2U-Net, ResU-Net, TransU-Net* and *TransU-Net++*. However, it is important to note that the *TransU-Net* segmentation model attained the highest precision score at 95.33%. For instance, the developed segmentation algorithm of *WVResU-Net* with a dice coefficient of 0.7345 substantially outperformed the segmentation accuracy of other models, including *SwinU-Net, R2U-Net, U-Net+++, TransU-Net++, AttentionU-Net, TransU-Net* and *ResU-Net*, by 9, 10, 11, 13, 13, 14, and 15 percentage points, respectively. Additionally, the accuracy of segmentation algorithms like *TransU-Nt++, SwinU-Net, ResU-Net, R2U-Net, AttentionU-Net, TransU-Net* and *U-Net+++* was significantly improved by about 5, 12, 13, 13, 15, 16, 19, and 23 percentage points, respectively, in recall value when compared to the results obtained by the *WVResU-Net* segmentation model with a recall value of 69.67%, as shown in Table 1.

[1] https://github.com/radiantearth/stac-api-spec

3.1. Segmentation maps and confusion matrices

Flood segmentation maps and confusion matrices as shown in Figs. 5–8 illustrate that the developed segmentation algorithm of *WVResU-Net* resulted in the best segmentation performance for mapping flooded areas, while the over-segmentation of flooded regions was not significantly high. Interestingly, the *TransU-Net* model resulted in the least over-classification of flooded regions, but it struggled to recognize flooded regions accurately (i.e., it under-classified flooded pixels). The highest over-classification of flooded regions was observed in the results obtained by the *R2-UNet* algorithm, followed by the *TransU-Net++* model. The highest under-classification of the flooded area was seen in the results of the *U-Net+++* model, followed by the segmentation model of *TransU-Net*. In addition to the developed *WVResU-Net* model, the *TransU-Net++* and *SwinU-Net* segmentation algorithms resulted in the highest accuracy in identifying flooded pixels. The most interesting pattern observed in the results of segmentation was the better accuracy of the model based on vision transformers for accurate flood mapping compared to the CNN-based segmentation models. As discussed, CNNs have several significant disadvantages, one of which is that they can only predict whether a desired feature will appear in an image, not where it will be located. The sequential properties of the backscattering coefficients of SAR imagery cannot be precisely determined by CNNs due to their fundamental limitations in the backbone. The self-attention technique utilized in ViTs can effectively address this specific issue. It should be noted that under-classification can occur due to factors like wind, inundated vegetation, and rain, which can make open water regions appear as rough surfaces and alter the backscattering pattern. Conversely, low backscatter from smooth or dark urban surfaces, such as roofs, car parks, concrete, and asphalt, which may resemble water, can lead to over-classification. Furthermore, one of the greatest challenges in flood recognition is interpreting the backscatter responses of various targets within urban and vegetated regions based on the presence or absence of floodwater. For example, it was observed that the deep learning models, specifically CNN-based algorithms had difficulty in differentiating between inundated vegetation and urban areas due to similar backscattering pattern of these two features. The reason is that the double-bounce is the predominant backscattering mechanism in both inundated vegetation and urban areas. Nevertheless, intricate backscatter mechanisms caused by various kinds of buildings and heights, vegetation areas, and distinct road patterns make detecting floods in urban areas difficult for SAR.

3.2. Visualization of vision network feature maps

We further show the feature maps derived from a spatial transformer network, the first layer of the attention gate-assisted decoded feature, and the resulting feature map of four Sentinel-1 SAR imagery produced by the *WVResU-Net* segmentation architecture in order to further clarify how the vision network affects intermediate feature representation. These instances are shown in Fig. 9. As seen in Fig. 9c1-c3, the proposed vision mechanism was able to signify and differentiate the flooded areas from other regions by emphasizing the positional information derived from dual SAR imagery and the extracted SAR backscattering coefficients. Moreover, it is clear from Fig. 9c1-c3 that the vision mechanism puts much less attention on the non-flooded. The maps produced by the vision network indicate how these functions better focus on the areas where different regions are flooded, improving the effectiveness of the developed deep learning architecture for flood mapping based on the SAR backscattering coefficients and polarization features derived from Sentinel-1 satellite imagery across the globe (see Fig. 9). Due to the vision transformer architecture on the skip connection path in decoded features, which facilitated task-specific learning as outlined in the previous subsections, the segmentation accuracy of the *ResU-Net* model were substantially enhanced by the proposed *WVResU-Net* architecture. The results achieved illustrated that the *WVResU-Net*

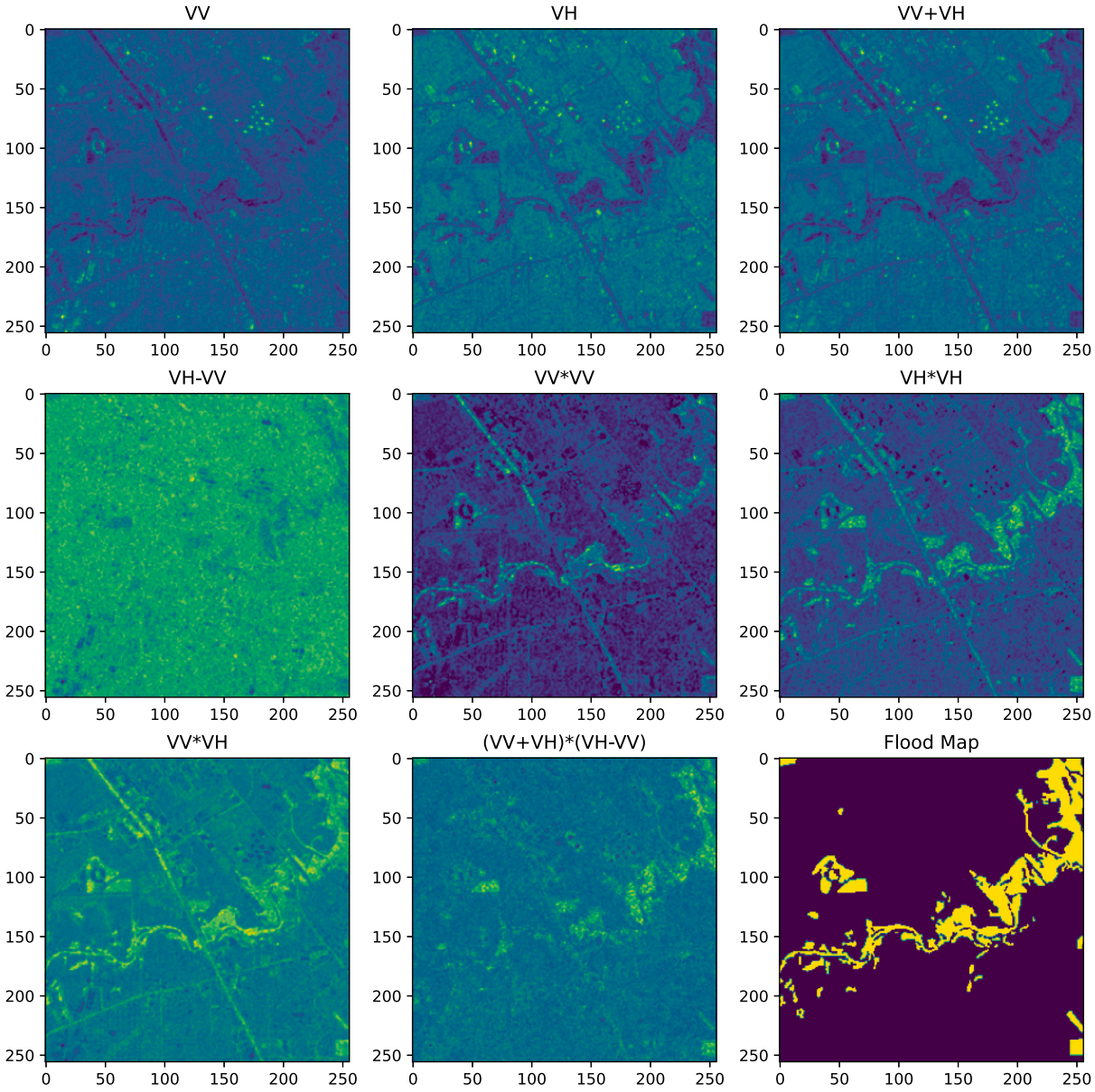


Fig. 4. A randomly selected example of SAR backscattering coefficients and their extracted features.

Table 1
Flood segmentation accuracy of the developed models in Precision (%), OA (%), F1-score (%), Recall (%), and Dice coefficient, respectively.

Algorithm	OA	Precision	Recall	F1-score	Dice coefficient	Time (min)
U-Net+++ (Huang et al., 2020)	93.94	92.80	46.87	62.53	0.6232	18
R2U-Net (Alom et al., 2018)	93.56	77.29	56.28	65.41	0.6320	42
AttentionU-Net (Oktay et al., 2018)	94.49	95.18	51.00	69.20	0.6048	12
SwinU-Net (Cao et al., 2021)	94.91	91.42	57.80	72.63	0.6470	35
TransU-Net (Chen et al., 2021)	94.45	**95.33**	50.47	69.26	0.5994	20
TransU-Net++ (Jamali et al., 2023)	94.56	80.66	64.50	71.27	0.6092	20
ResU-Net (Diakogiannis et al., 2020)	94.40	58.59	57.10	69.56	0.5864	23
WVResU-Net (ours)	**96.20**	92.97	**69.67**	**82.03**	**0.7345**	35

model considerably enhanced the segmentation accuracy of the *ResU-Net* 2, 12, 13, 15 and 34 percentage points in terms of overall accuracy, F1-score, recall, dice coefficient, and precision, respectively.

In addition, the level of measurement of separability is represented by the area under the curve (AUC). It demonstrates how well a segmentation method can discriminate between different classes, and in this case, how well a model can reliably distinguish between flooded and unflooded areas. Greater levels of AUC demonstrate how well a model detects flooded regions. The segmentation model created by *WVResU-Net* considerably outperforms competing algorithms, such as *SwinUNet* (0.786), *TransU-Net++* (0.775), *R2U-Net* (0.772), *AttentionU-Net* (0.753), *TransU-Net* (0.751) and *ResU-Net* (0.780), with an AUC value of 0.845, as seen in Fig. 10. The obtained results illustrated that due to the use of the vision mechanism on the skip connection path

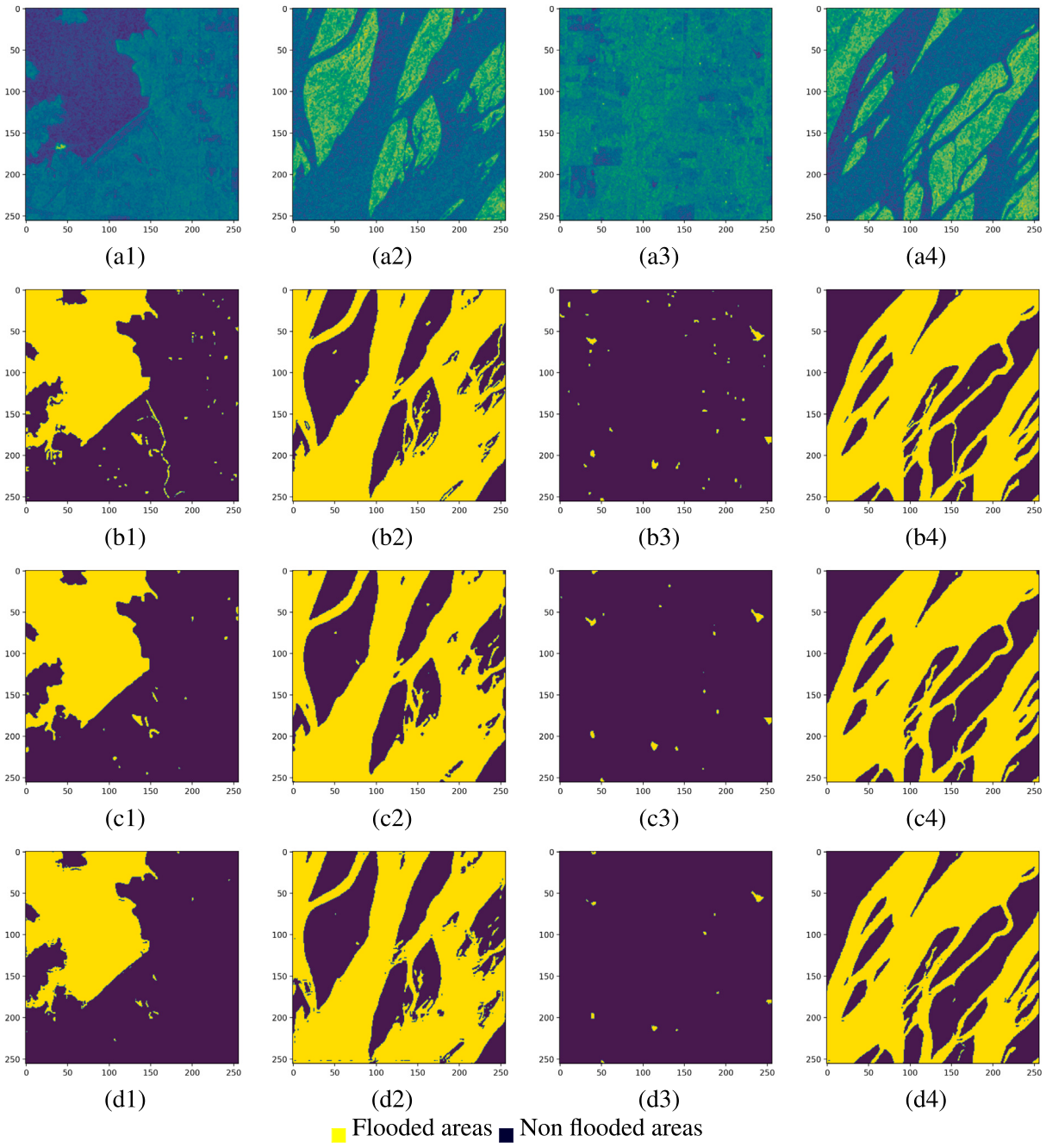


Fig. 5. Segmentation maps of four randomly selected areas obtained using segmentation algorithms of (a1-a4) Sentinel-1 polarization data of VV, (b1-b4) flood masks, (c1-c4) UNet+++ and (d1-d4) AttentionU-Net, respectively.

in decoded features, the AUC value of the base *ResU-Net* segmentation algorithm (0.78) was enhanced by approximately 7 percentage points by the *WVResU-Net* architecture (0.845).

As discussed by Rasti et al. (2022), it is widely acknowledged that during the remote sensing imaging procedure, data typically suffer from a variety of deteriorations, noise effects, or variabilities. For instance, spectral variability in hyperspectral imagery from satellite or aircraft sources is unavoidable, which makes it challenging for spectral unmixing to accurately calculate abundance maps. As discussed in the introduction section, the Sentinel-1 SAR imagery suffers from various similar issues and limitations. For instance, during the imaging process, electromagnetic radiation can be significantly impacted by echo interference when it encounters rough surfaces, leading to phase variations. This interference results in decreased echo intensity and introduces speckle noise into the SAR data. Considering the discussed limitations, specifically the intrinsic speckle noise in SAR imagery, the developed *WVResU-Net* segmentation architecture through the effective use of the wave vision MLP and residual connections could significantly enhance the segmentation capabilities of current state-of-the-art CNN- and ViT-based deep learning models. The obtained segmentation results proved the excellent ability of the developed model in flood mapping even with the presence of speckle noises and similar backscatter responses of various features within urban and inundated vegetated regions.

3.3. Computational cost of the segmentation algorithms

The *R2U-Net* (42 min) and *AttentionU-Net* (12 min) segmentation models had the highest and lowest computation costs in terms of

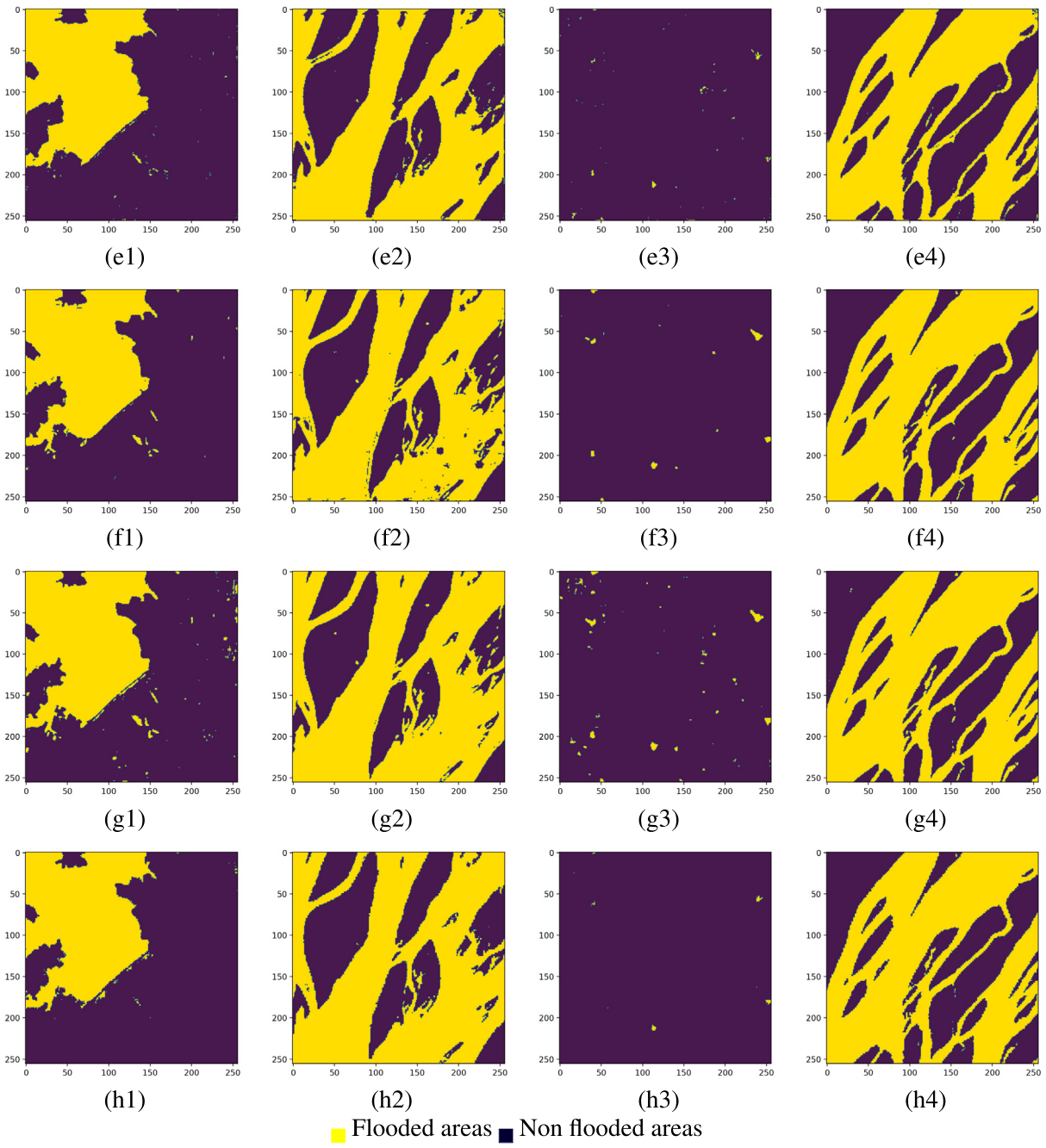


Fig. 6. Segmentation maps of four randomly selected areas obtained using segmentation algorithms of (e1-e4) R2U-Net, (f1-f4) TransU-Net, (g1-g4) TransU-Net++, (h1-h4) UNet+++, (i1-i4) ResU-Net and (j1-j4) the developed WVResU-Net, respectively.

training time, respectively, as shown in Table 1. Due to the embedding of several vision MLPs in the decoding architecture, the developed *WVResU-Net* segmentation model, with a required training time of 35 min, has an increased time complexity compared to the base *ResU-Net* segmentation algorithm (23 min). However, the developed *WVResU-Net* algorithm's ability to recognize flooded areas has been significantly enhanced by the use of a vision mechanism. The experiments were carried out using the programming language Python on an Intel Core-i7 CPU and an NVIDIA RTX 2070 MAX-Q GPU. All the deep learning models are developed using Python TensorFlow GPU 2.6.2.

3.4. Ablation study

To better understand the significance of the wave vision approach in the developed model, we conducted an ablation study. The achieved results illustrated that the use of Wave Vision MLP significantly improved the segmentation capability of the base ResU-Net deep learning model. For example, the segmentation results demonstrate that the *WVResU-Net* algorithm significantly enhances the segmentation accuracy of the *ResU-Net* deep learning model by about 2, 12, 13, 15 and 34 percentage points in terms of overall accuracy, F1-score, recall, dice coefficient, and precision, respectively, as seen in Table 1. The capability of the

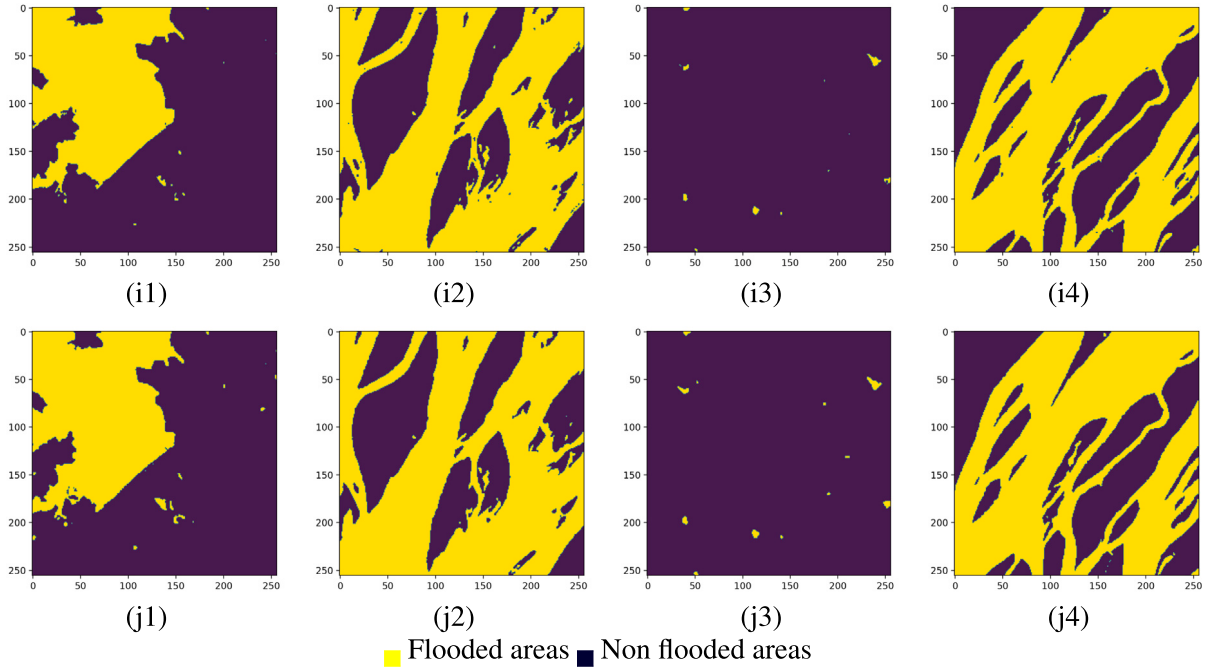


Fig. 7. Segmentation maps of four randomly selected areas obtained using segmentation algorithms of (i1-i4) ResU-Net and (j1-j4) the developed WVResU-Net, respectively.

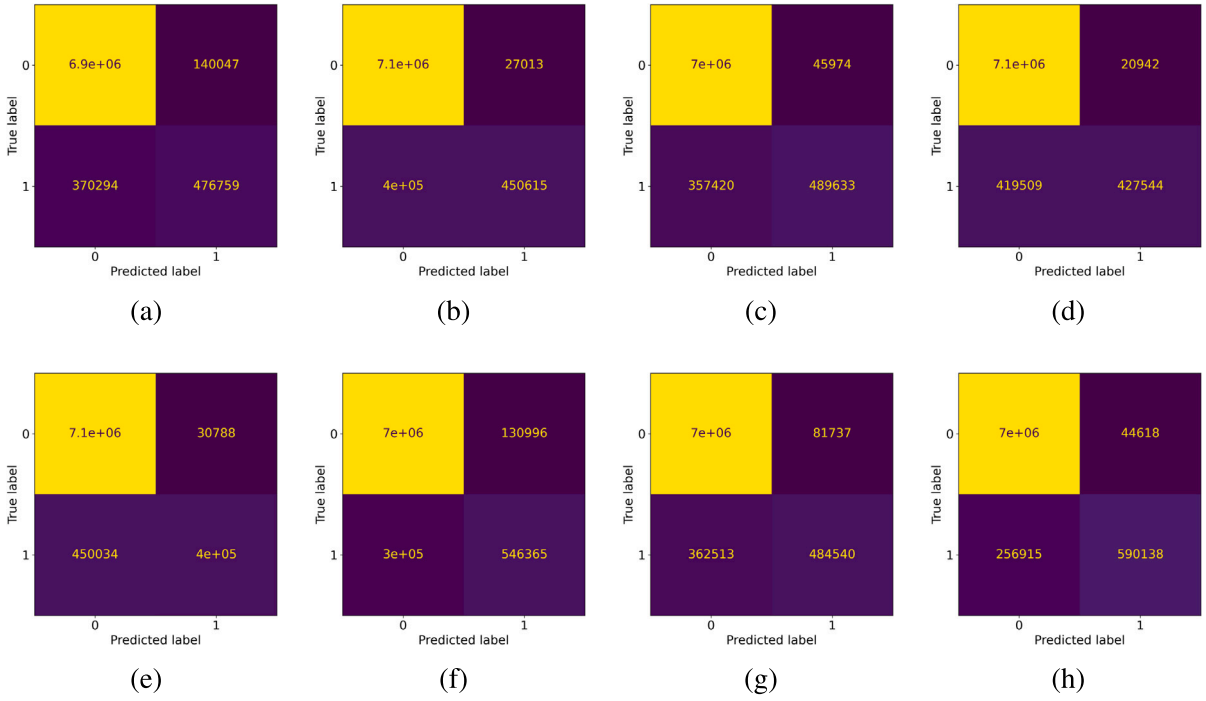


Fig. 8. Confusion matrices obtained using segmentation algorithms of (a) R2U-Net, (b) Attention U-Net, (c) Swin U-Net, (d) TransU-Net, (e) U-Net+++, (f) TransU-Net++, (g) ResU-Net and (h) WVResU-Net, respectively (0 = Non-flooded areas and 1 = flooded areas).

developed model of *WVResU-Net* in recognizing flooded areas was significantly improved over the ResU-Net segmentation algorithm, as illustrated in Fig. 8. Moreover, The segmentation results of the *ResU-Net* algorithm were significantly enhanced by the *WVResU-Net* model in terms of AUC by approximately 7 percentage points, as presented in Fig. 10.

4. Conclusion

In this paper, we introduced WVResU-Net, an advanced deep learning segmentation model that leverages ResU-Net for precise flood mapping using Sentinel-1 dual-polarization data, along with a cost-effective visual MLP based on the concept of wave vision. Our proposed wave

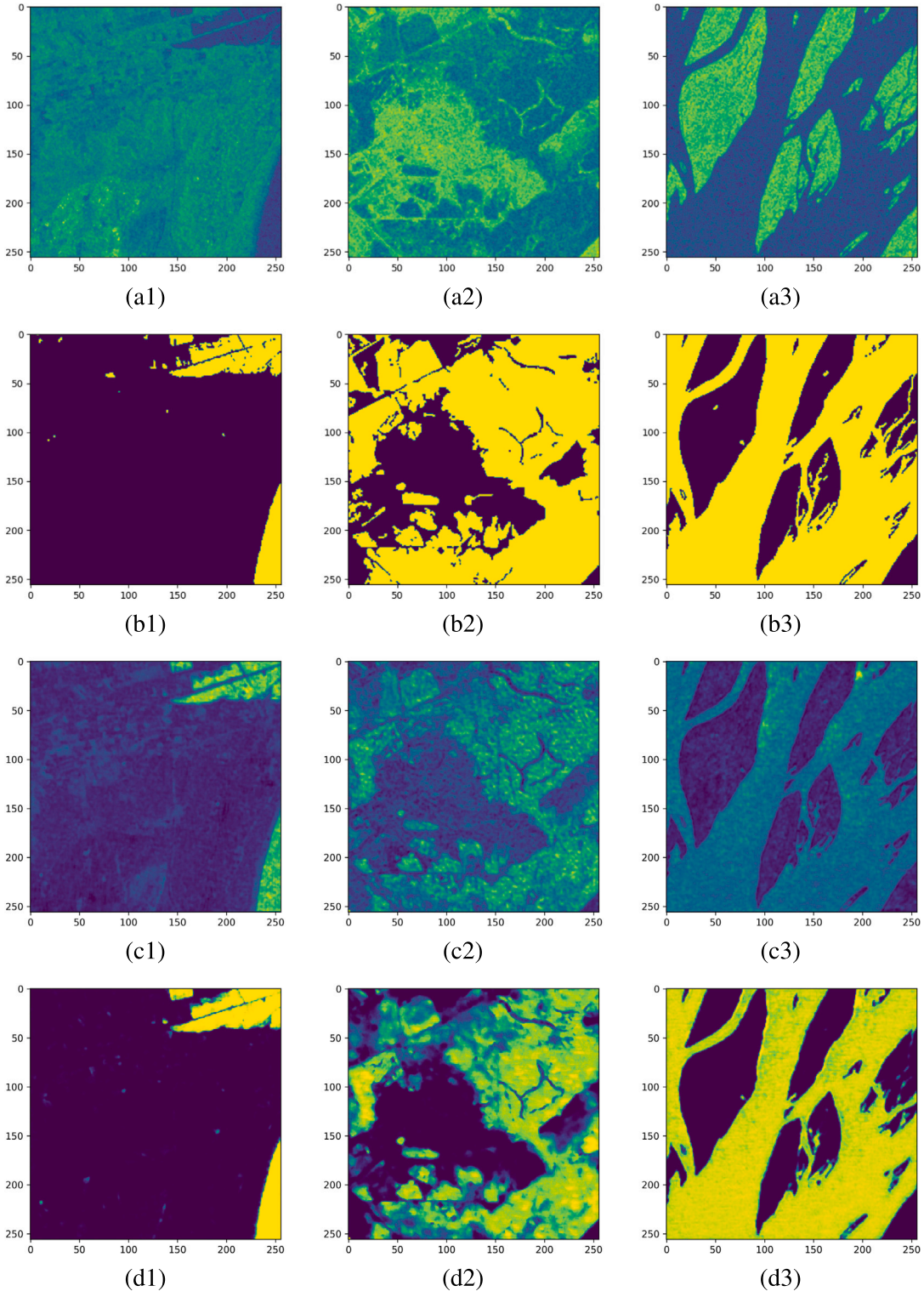


Fig. 9. Feature maps derived from the developed segmentation algorithm of WVResU-Net for three randomly selected areas (a1-a4) SAR polarization data of VV, (b1-b4) flood masks, (c1-c4) derived feature map from last vision network and (d1-d4) last convolutional layer, respectively.

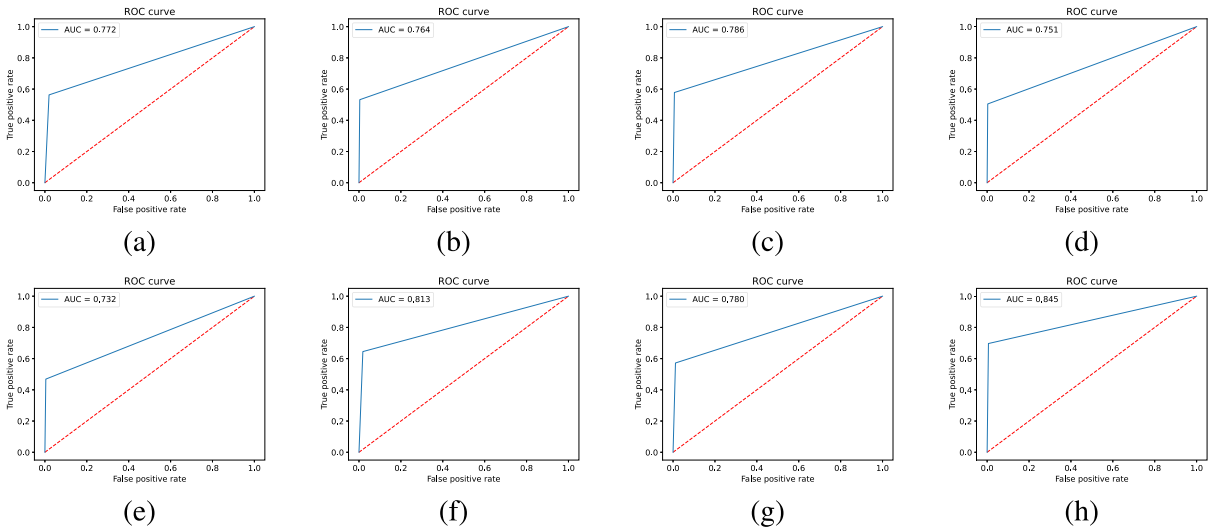


Fig. 10. Area under the ROC Curve (AUC) obtained using segmentation algorithms of (a) R2U-Net, (b) Attention U-Net, (c) Swin U-Net, (d) TransU-Net, (e) U-Net+++, (f) TransU-Net++, (g) ResU-Net, and (h) WVResU-Net, respectively.

vision approach utilizes dynamic weights, incorporating real-value amplitude and complex-value phase modules, a departure from the conventional fixed-weight approaches used in most state-of-the-art MLPs. In contrast to existing algorithms, which often directly combine tokens from various input images with fixed weights, thereby overlooking the varying semantic information of image patches, our developed WVResU-Net takes a more dynamic and adaptive approach. The results of flood segmentation clearly demonstrate that WVResU-Net outperforms several well-known deep learning segmentation models, including Swin U-Net, U-Net+++, Attention U-Net, R2U-Net, ResU-Net, TransU-Net and TransU-Net++. For example, the *WVResU-Net* segmentation model with a dice coefficient of 0.7345 significantly outperformed the segmentation accuracy of other deep learning models, including *SwinU-Net, R2U-Net, U-Net+++, TransU-Net++, AttentionU-Net, TransU-Net* and *ResU-Net*, by about 9, 10, 11, 13, 13, 14 and 15 percentage points, respectively. The feature maps generated by ViTs highlight their ability to effectively focus on regions with flooded pixels, enhancing the efficacy of our deep learning architecture for flood mapping. However, due to the integration of multiple vision MLPs in the decoding section of the architecture, the computational complexity of WVResU-Net, particularly in terms of training time, has increased. The future research will mainly focus on developing new methods to substantially decrease the computational cost of the developed deep learning architecture, while increasing the segmentation capability of the model.

Funding information

This research was partially supported by NOAA, United States (award NA21OAR4590358) and NASA, United States (award 80NSSC23M0051).

CRediT authorship contribution statement

Ali Jamali: Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Swalpa Kumar Roy:** Writing – review & editing, Methodology, Conceptualization. **Leila Hashemi Beni:** Writing – review & editing, Funding acquisition. **Biswajeet Pradhan:** Writing – review & editing. **Jonathan Li:** Writing – review & editing. **Pedram Ghamisi:** Writing – review & editing, Supervision, Project administration, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K., 2018. Recurrent residual convolutional neural network based on U-net (R2U-net) for medical image segmentation. http://dx.doi.org/10.48550/ARXIV.1802.06955.

Cai, H., Chen, T., Niu, R., Plaza, A., 2021. Landslide detection using densely connected convolutional networks and environmental conditions. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 5235–5247. http://dx.doi.org/10.1109/JSTARS.2021.3079196.

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M., 2021. Swin-unet: Unet-like pure transformer for medical image segmentation. arXiv pr doi arXiv:2105.05537.

Chang, S., Ghamisi, P., 2023. Changes to captions: An attentive network for remote sensing change captioning. IEEE Trans. Image Process. 32, 6047–6060. http://dx.doi.org/10.1109/TIP.2023.3328224.

Chen, Y., Jiang, H., Li, C., Jia, X., Ghamisi, P., 2016. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 54 (10), 6232–6251. http://dx.doi.org/10.1109/TGRS.2016.2584107.

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y., 2021. Transunet: Transformers make strong encoders for medical image segmentation. arXiv pr doi arXiv:2102.04306.

Diakogiannis, F.I., Waldner, F., Caccetta, P., Wu, C., 2020. Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. ISPRS J. Photogramm. 162, 94–114. http://dx.doi.org/10.1016/j.isprsjprs.2020.01.013.

Feng, Q., Gong, J., Liu, J., Li, Y., 2015. Flood mapping based on multiple endmember spectral mixture analysis and random forest classifier—The case of Yuyao, China. Remote Sens. 7 (9), 12539–12562. http://dx.doi.org/10.3390/rs70912539.

García-Pintado, J., Neal, J.C., Mason, D.C., Dance, S.L., Bates, P.D., 2013. Scheduling satellite-based SAR acquisition for sequential assimilation of water level observations into flood modelling. J. Hydrol. 495, 252–266. http://dx.doi.org/10.1016/j.jhydrol.2013.03.050.

Ghorbanzadeh, O., Xu, Y., Ghamisi, P., Kopp, M., Kreil, D., 2022. Landslide4sense: Reference benchmark data and deep learning models for landslide detection. IEEE Trans. Geosci. Remote Sens. 60, 1–17. http://dx.doi.org/10.1109/TGRS.2022.3215209.

Giustarini, L., Hostache, R., Kavetski, D., Chini, M., Corato, G., Schlaffer, S., Matgen, P., 2016. Probabilistic flood mapping using synthetic aperture radar data. IEEE Trans. Geosci. Remote Sens. 54 (12), 6958–6969. http://dx.doi.org/10.1109/TGRS.2016.2592951.

Grimaldi, S., Li, Y., Pauwels, V.R.N., Walker, J.P., 2016. Remote sensing-derived water extent and level to constrain hydraulic flood forecasting models: Opportunities and challenges. Surv. Geophys. 37 (5), 977–1034. http://dx.doi.org/10.1007/s10712-016-9378-y.

Hamidi, E., Peter, B.G., Muñoz, D.F., Moftakhari, H., Moradkhani, H., 2023. Fast flood extent monitoring with SAR change detection using Google earth engine. IEEE Trans. Geosci. Remote Sens. 61, 1–19. http://dx.doi.org/10.1109/TGRS.2023.3240097.

Han, K., Wang, Y., Xu, C., Guo, J., Xu, C., Wu, E., Tian, Q., 2022. GhostNets on heterogeneous devices via cheap operations. Int. J. Comput. Vis. 130 (4), 1050–1069. http://dx.doi.org/10.1007/s11263-022-01575-y.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proc. IEEE Comput. Soc.. pp. 770–778.

Henderson, F.M., Lewis, A.J., 2008. Radar detection of wetland ecosystems: A review. Int. J. Remote Sens. 29 (20), 5809–5835. http://dx.doi.org/10.1080/01431160801958405.

Hong, D., Zhang, B., Li, H., Li, Y., Yao, J., Li, C., Werner, M., Chanussot, J., Zipf, A., Zhu, X.X., 2023. Cross-city matters: A multimodal remote sensing benchmark dataset for cross-city semantic segmentation using high-resolution domain adaptation networks. Remote Sens. Environ. 299, 113856. http://dx.doi.org/10.1016/j.rse.2023.113856, URL https://www.sciencedirect.com/science/article/pii/S0034425723004078.

Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., Wu, J., 2020. Unet 3+: A full-scale connected unet for medical image segmentation. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, pp. 1055–1059.

Intergovernmental Panel on Climate Change (IPCC), 2014. Summary for policymakers. In: Climate Change 2013 – the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp. 1–30. http://dx.doi.org/10.1017/CBO9781107415324.004.

Islam, M.T., Meng, Q., 2022. An exploratory study of sentinel-1 SAR for rapid urban flood mapping on Google earth engine. Int. J. Appl. Earth Obs. Geoinf. 113, 103002. http://dx.doi.org/10.1016/j.jag.2022.103002.

Jackson, J., Yussif, S.B., Patamia, R.A., Sarpong, K., Qin, Z., 2023. Flood or non-flooded: A comparative study of state-of-the-art models for flood image classification using the FloodNet dataset with uncertainty offset analysis. Water 15 (5), http://dx.doi.org/10.3390/w15050875.

Jamali, A., Roy, S.K., Li, J., Ghamisi, P., 2023. TransU-Net++: Rethinking attention gated TransU-Net for deforestation mapping. Int. J. Appl. Earth Obs. Geoinf. 120, 103332. http://dx.doi.org/10.1016/j.jag.2023.103332.

Li, C., Zhang, B., Hong, D., Yao, J., Chanussot, J., 2023. LRR-net: An interpretable deep unfolding network for hyperspectral anomaly detection. IEEE Trans. Geosci. Remote Sens. 61, 1–12. http://dx.doi.org/10.1109/TGRS.2023.3279834.

Liu, Z., Merwade, V., 2018. Accounting for model structure, parameter and input forcing uncertainty in flood inundation modeling using Bayesian model averaging. J. Hydrol. 565, 138–149. http://dx.doi.org/10.1016/j.jhydrol.2018.08.009.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: Proc. IEEE Comput. Soc.. pp. 3431–3440.

Martinis, S., Groth, S., Wieland, M., Knopp, L., Rättich, M., 2022. Towards a global seasonal and permanent reference water product from sentinel-1/2 data for improved flood mapping. Remote Sens. Environ. 278, 113077. http://dx.doi.org/10.1016/j.rse.2022.113077.

Mayer, T., Poortinga, A., Bhandari, B., Nicolau, A.P., Markert, K., Thwal, N.S., Markert, A., Haag, A., Kilbride, J., Chishtie, F., Wadhwa, A., Clinton, N., Saah, D., 2021. Deep learning approach for sentinel-1 surface water mapping leveraging Google earth engine. ISPRS J. Photogramm. 2, 100005. http://dx.doi.org/10.1016/j.ophoto.2021.100005, URL https://www.sciencedirect.com/science/article/pii/S2667393221000053.

McCormack, T., Campanyà, J., Naughton, O., 2022. A methodology for mapping annual flood extent using multi-temporal sentinel-1 imagery. Remote Sens. Environ. 282, 113273. http://dx.doi.org/10.1016/j.rse.2022.113273.

Moya, L., Mas, E., Koshimura, S., 2022. Sparse representation-based inundation depth estimation using SAR data and digital elevation model. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 15, 9062–9072. http://dx.doi.org/10.1109/JSTARS.2022.3215719.

Muñoz, D.F., Muñoz, P., Moftakhari, H., Moradkhani, H., 2021. From local to regional compound flood mapping with deep learning and data fusion techniques. Sci. Total Environ. 782, 146927. http://dx.doi.org/10.1016/j.scitotenv.2021.146927.

Nguyen, T.H., Ricci, S., Fatras, C., Piacentini, A., Delmotte, A., Lavergne, E., Kettig, P., 2022. Improvement of flood extent representation with remote sensing data and data assimilation. IEEE Trans. Geosci. Remote Sens. 60, 1–22. http://dx.doi.org/10.1109/TGRS.2022.3147429.

O'Grady, D., Leblanc, M., Gillieson, D., 2013. Relationship of local incidence angle with satellite radar backscatter for different surface conditions. Int. J. Appl. Earth Obs. Geoinf. 24, 42–53. http://dx.doi.org/10.1016/j.jag.2013.02.005.

Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al., 2018. Attention u-net: Learning where to look for the pancreas. arXiv pr doi arXiv:1804.03999.

Pappas, O.A., Anantrasirichai, N., Achim, A.M., Adams, B.A., 2021. River planform extraction from high-resolution SAR images via generalized Gamma distribution superpixel classification. IEEE Trans. Geosci. Remote Sens. 59 (5), 3942–3955. http://dx.doi.org/10.1109/TGRS.2020.3011209.

Persello, C., Wegner, J.D., Hänsch, R., Tuia, D., Ghamisi, P., Koeva, M., Camps-Valls, G., 2022. Deep learning and earth observation to support the sustainable development goals: Current approaches, open challenges, and future opportunities. IEEE Geosci. Remote Sens. Mag. 10 (2), 172–200. http://dx.doi.org/10.1109/MGRS.2021.3136100.

Qiao, X., Roy, S.K., Huang, W., 2023. 3D sharpened cosine similarity operation for hyperspectral image classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 1–12. http://dx.doi.org/10.1109/JSTARS.2023.3337112.

Rasti, B., Chang, Y., Dalsasso, E., Denis, L., Ghamisi, P., 2022. Image restoration for remote sensing: Overview and toolbox. IEEE Geosci. Remote Sens. Mag. 10 (2), 201–230. http://dx.doi.org/10.1109/MGRS.2021.3121761.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp. 234–241.

Schumann, G., Di Baldassarre, G., Bates, P.D., 2009. The utility of spaceborne radar to render flood inundation maps based on multialgorithm ensembles. IEEE Trans. Geosci. Remote Sens. 47 (8), 2801–2807. http://dx.doi.org/10.1109/TGRS.2009.2017937.

Surampudi, S., Kumar, V., 2023. Flood depth estimation in agricultural lands from L and C-band synthetic aperture radar images and digital elevation model. IEEE Access 11, 3241–3256. http://dx.doi.org/10.1109/ACCESS.2023.3234742.

Tang, Y., Han, K., Guo, J., Xu, C., Li, Y., Xu, C., Wang, Y., 2022a. An image patch is a wave: Phase-aware vision MLP. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 10935–10944.

Tang, Y., Han, K., Guo, J., Xu, C., Xu, C., Wang, Y., 2022b. GhostNetV2: Enhance cheap operation with long-range attention. In: Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (Eds.), In: Advances in Neural Information Processing Systems, vol.35, Curran Associates, Inc, pp. 9969–9982, URL https://proceedings.neurips.cc/paper_files/paper/2022/file/40b60852a4abdaa696b5a1a78da34635-Paper-Conference.pdf.

Tazmul Islam, M., Meng, Q., 2022a. An exploratory study of sentinel-1 SAR for rapid urban flood mapping on google earth engine. Int. J. Appl. Earth Obs. Geoinf. 113, 103002. http://dx.doi.org/10.1016/j.jag.2022.103002.

Tazmul Islam, M., Meng, Q., 2022b. An exploratory study of sentinel-1 SAR for rapid urban flood mapping on google earth engine. Int. J. Appl. Earth Obs. Geoinf. 113, 103002. http://dx.doi.org/10.1016/j.jag.2022.103002, URL https://www.sciencedirect.com/science/article/pii/S1569843222001911.

Tellman, B., Sullivan, J.A., Kuhn, C., Kettner, A.J., Doyle, C.S., Brakenridge, G.R., Erickson, T.A., Slayback, D.A., 2021. Satellite imaging reveals increased proportion of population exposed to floods. Nature 596 (7870), 80–86. http://dx.doi.org/10.1038/s41586-021-03695-w.

Tiwari, V., Kumar, V., Matin, M.A., Thapa, A., Ellenburg, W.L., Gupta, N., Thapa, S., 2020. Flood inundation mapping- Kerala 2018; harnessing the power of SAR, automatic threshold detection method and Google earth engine. PLOS ONE 15 (8), 1–17. http://dx.doi.org/10.1371/journal.pone.0237324.

Tolstikhin, I.O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., Lucic, M., Dosovitskiy, A., 2021. MLP-mixer: An all-MLP architecture for vision. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (Eds.), In: Advances in Neural Information Processing Systems, vol.34, Curran Associates, Inc, pp. 24261–24272.

Tong, X., Luo, X., Liu, S., Xie, H., Chao, W., Liu, S., Liu, S., Makhinov, A., Makhinova, A., Jiang, Y., 2018. An approach for flood monitoring by the combined use of landsat 8 optical imagery and COSMO-SkyMed radar imagery. ISPRS J. Photogramm. 136, 144–153. http://dx.doi.org/10.1016/j.isprsjprs.2017.11.006.

Tripathy, P., Malladi, T., 2022. Global flood mapper: A novel Google earth engine application for rapid flood mapping using sentinel-1 SAR. Nat Hazards 114 (2), 1341–1363. http://dx.doi.org/10.1007/s11069-022-05428-2.

Vamsi, C., Amudha, V., Soundharyaa Shri Harini, R., 2023. Accuracy improvement of flooded area detection from satellite images using novel K-nearest neighbors in comparison with support vector machine. In: 2023 International Conference on Advances in Computing, Communication and Applied Informatics. ACCAI, pp. 1–6. http://dx.doi.org/10.1109/ACCAI58221.2023.10199536.

Wan, L., Liu, M., Wang, F., Zhang, T., You, H.J., 2019. Automatic extraction of flood inundation areas from SAR images: A case study of Jilin, China during the 2017 flood disaster. Int. J. Remote Sens. 40 (13), 5050–5077. http://dx.doi.org/10.1080/01431161.2019.1577999.

Wang, L., Li, R., Zhang, C., Fang, S., Duan, C., Meng, X., Atkinson, P.M., 2022. UNetFormer: A unet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. ISPRS J. Photogramm. 190, 196–214.

Wood, M., Hostache, R., Neal, J., Wagener, T., Giustarini, L., Chini, M., Corato, G., Matgen, P., Bates, P., 2016. Calibration of channel depth and friction parameters in the LISFLOOD-FP hydraulic model using medium-resolution SAR data and identifiability techniques. Hydrol. Earth Syst. Sci. 20 (12), 4983–4997. http://dx.doi.org/10.5194/hess-20-4983-2016.

Woodhouse, I.H., 2017. Introduction to Microwave Remote Sensing. CRC Press, p. 400. http://dx.doi.org/10.1201/9781315272573.

Woznicki, S.A., Baynes, J., Panlasigui, S., Mehaffey, M., Neale, A., 2019. Development of a spatially complete floodplain map of the conterminous United States using random forest. Sci. Total Environ. 647, 942–953. http://dx.doi.org/10.1016/j.scitotenv.2018.07.353.

Wu, X., Hong, D., Chanussot, J., 2023. UIU-Net: U-net in U-net for infrared small object detection. IEEE Trans. Image Process. 32, 364–376. http://dx.doi.org/10.1109/TIP.2022.3228497.

Ye, T., Qin, W., Zhao, Z., Gao, X., Deng, X., Ouyang, Y., 2023. Real-time object detection network in UAV-vision based on CNN and transformer. IEEE Trans. Instrum. Meas. 72, 1–13. http://dx.doi.org/10.1109/TIM.2023.3241825.

Yokoya, N., Yamanoi, K., He, W., Baier, G., Adriano, B., Miura, H., Oishi, S., 2022. Breaking limits of remote sensing by deep learning from simulated data for flood and debris-flow mapping. IEEE Trans. Geosci. Remote Sens. 60, 1–15. http://dx.doi.org/10.1109/TGRS.2020.3035469.

Yuan, F., Zhang, Z., Fang, Z., 2023. An effective CNN and transformer complementary network for medical image segmentation. Pattern Recognit. 136, 109228.

Yue, J., Fang, L., Ghamisi, P., Xie, W., Li, J., Chanussot, J., Plaza, A., 2022. Optical remote sensing image understanding with weak supervision: Concepts, methods, and perspectives. IEEE Geosci. Remote Sens. Mag. 10 (2), 250–269. http://dx.doi.org/10.1109/MGRS.2022.3161377.

Zhai, Y., Liao, J., Sun, B., Jiang, Z., Ying, Z., Wang, W., Genovese, A., Piuri, V., Scotti, F., 2023. Dual consistency alignment based self-supervised learning for SAR target recognition with speckle noise resistance. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 16, 3915–3928. http://dx.doi.org/10.1109/JSTARS.2023.3267824.

Zhang, X., Chan, N.W., Pan, B., Ge, X., Yang, H., 2021a. Mapping flood by the object-based method using backscattering coefficient and interference coherence of sentinel-1 time series. Sci. Total Environ. 794, 148388. http://dx.doi.org/10.1016/j.scitotenv.2021.148388.

Zhang, Y., Liu, P., Chen, L., Xu, M., Guo, X., Zhao, L., 2023. A new multi-source remote sensing image sample dataset with high resolution for flood area extraction: GF-FloodNet. Int. J. Digit. Earth 16 (1), 2522–2554. http://dx.doi.org/10.1080/17538947.2023.2230978.

Zhang, H., Qi, Z., Li, X., Chen, Y., Wang, X., He, Y., 2021b. An urban flooding index for unsupervised Inundated Urban Area detection using sentinel-1 polarimetric SAR images. Remote Sens. 13 (22), http://dx.doi.org/10.3390/rs13224511, URL https://www.mdpi.com/2072-4292/13/22/4511.

Zhang, Q., Zhang, P., Hu, X., 2021c. Unsupervised GRNN flood mapping approach combined with uncertainty analysis using bi-temporal sentinel-2 MSI imageries. Int. J. Digit. Earth 14 (11), 1561–1581. http://dx.doi.org/10.1080/17538947.2021.1953160.