IGARSS 2024 - 2024 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-6032-5/24/$31.00 ©2024 IEEE | DOI: 10.1109/IGARSS53475.2024.10640534

DOMAIN KNOWLEDGE-AWARE REMOTE SENSING FOUNDATION MODEL FOR FLOOD DETECTION IN MULTI-SPECTRAL IMAGERY

Yansheng Li, Bo Dang, Fanyi Wei, Jieyi Tan, Yangjie Lin*

School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

ABSTRACT

Obtaining accurate and timely flood information is crucial for effective disaster management and response. To address the limitations of existing methods in terms of accuracy and model robustness, this research significantly improves the accuracy and stability of flood detection by integrating domain knowledge into the Remote Sensing Foundation Model (RSFM). Specifically, we employ advanced RSFM to focus on extracting spatial texture features from images after super-resolution. The Automatic Water Extraction Index (AWEI) is introduced to leverage spectral information from multi-spectral imagery, while model fusion techniques further enhance the accuracy of segmentation results. Moreover, we incorporate prior knowledge such as land use products and Digital Elevation Models (DEM) for knowledge rules-driven post-processing, refining the final flood detection results. Experimental results demonstrate that our approach achieve the second-place ranking in the 2024 IEEE GRSS Data Fusion Contest (DFC) Track 2 test phase (F1: 88.25%), highlighting the effectiveness and competitiveness of our method.

***Index Terms*—** Flood detection, remote sensing foundation model, semantic segmentation network, domain knowledge

1. INTRODUCTION

Floods are a global natural disaster that has significant impacts on human society and the natural environment [1, 2]. They not only result in casualties and property damage but also have the potential to disrupt ecosystems and infrastructure [3]. Therefore, timely and accurate detection and monitoring of floods are crucial for disaster management, emergency response, and recovery efforts.

With the advancements in remote sensing technology and deep learning techniques, utilizing satellite imagery and deep learning models for flood detection has become an effective approach. In particular, the Harmonized Landsat Sentinel-2 images provide multi-spectral imagery covering a wide spectral range from visible to short-wave infrared. These images are ideal data sources for flood detection due to their spatial resolution and observation frequency. Additionally, deep learning-based semantic segmentation models have made significant progress in recent years [4, 5], ranging from CNN-based models [6, 7] to Transformer-based segmentation networks [8], improving the performance of remote sensing image segmentation. These models have also been applied extensively in the extraction of flood information from multi-spectral imagery [1].

Fig. 1. The challenges in extracting flood areas from multi-spectral imagery.

However, there are still challenges in directly extracting flood areas from Harmonized Landsat Sentinel-2 imagery using deep networks, as shown in Figure 1. 1) the presence of cloud cover can deteriorate the quality of the images, thereby affecting the accuracy of flood detection. 2) the limitation in image spatial resolution may result in the loss of details in flood areas, especially for smaller flood events where the accurate delineation of flood boundaries and evolution processes may be challenging. 3) the complex surface features in flooded areas, such as buildings, roads, and vegetation, pose difficulties in model training, potentially leading to the instability of flood detection.

To address the aforementioned challenges, this study proposes a domain knowledge-aware framework based on the Remote Sensing Foundation Model (RSFM) for flood detection in multi-spectral imagery. Our method first applies super-resolution processing to the Harmonized Landsat Sentinel-2 images to improve their spatial resolution and enhance the details of flood areas. Next, we utilize advanced RSFM to extract spatial texture features of flood regions from the super-resolved images. To better utilize spectral information, we introduce the Automatic Water Extraction Index (AWEI) that effectively captures the spectral differences between flood and non-flood areas. To further improve flood detection perfor-

*Corresponding author.

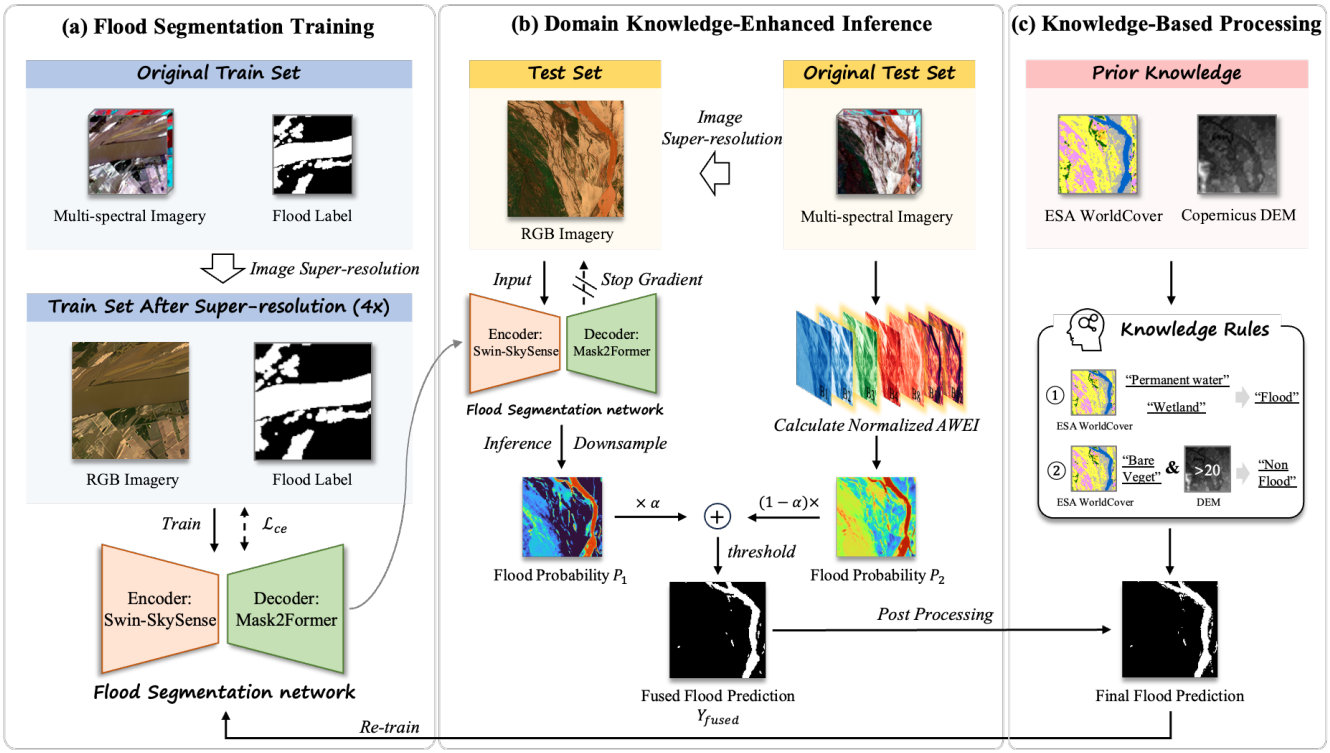

Fig. 2. The overview structure diagram of our proposed domain knowledge-aware framework based on the RSFM.

mance, we employ the model ensemble and post-processing strategy. By fusing multiple flood prediction results and processing them driven by knowledge rules, improving the reliability and accuracy of flood predictions. The experimental results demonstrate the superiority of the proposed method, which achieves an F1-score of 88.25% and ranks the second place in the testing phase of Track 2 of DFC 2024.

2. METHODS

In this section, we introduce the proposed domain knowledge-aware flood detection approach. The overall structure diagram is shown in Figure 2.

2.1. Flood segmentation training

In consideration of the original Harmonized Landsat Sentinel-2 images' limited spatial resolution of 30m, each image presents with a dimension of merely 128×128 pixels. To more effectively mine spatial detail features from these images via deep learning models, we utilize the image super-resolution method SRCNN [9], thereby expanding the size of the images in the RGB band by 4 × and significantly enhancing their visual clarity. Despite a certain degree of spectral information loss in the processed images, the implementation of the super-resolution strategy effectively recovers the critical details that are commonly overlooked at lower resolutions, thereby playing a pivotal role in achieving accurate flood mapping. Concurrently, we upsample the original flood labels to align with new image size, thereby establishing pixel-level image-label pairs.

Our performance in the first validation phase lead us to adopt a semantic segmentation network based on an encoder-decoder architecture for training. Specifically, we utilize the Swin Transformer (SkySense pre-trained [8]) model [10] as the encoder to extract high-dimensional image features. For the decoder, we employ the Mask2former [11] architecture to reconstruct the flood segmentation map. During the training process, we optimize our segmentation network using the cross-entropy loss function.

2.2. Domain knowledge-enhanced inference

During the inference phase, we employ a divide-and-conquer approach, leveraging both a trained segmentation model and an exponential threshold-based model to focus on detailed spatial details and rich spectral information in the images. Specifically, we use the trained model F to process the super-resolved RGB images $x_{super}^{4h\times4w\times3}$, resulting in the flood prediction probability P_1. Simultaneously, we calculate the AWEI using the spectral information from the original multi-spectral images $x^{h\times w\times7}$, including B2, B3, B8, B11 and B12 bands, and normalized it to obtain the flood prediction probability P_2. By introducing an adjustable parameter α, we

Table 1. The details of our ablation study. * denotes the utilization of pre-trained weights from SkySense.

ID	Method	Development phase (F1)	Test phase (F1)
I	Vision Transformer* + UperNet	0.931	-
II	Image Super-resolution + Swin Transformer* + UperNet	0.948	-
III	Image Super-resolution + Swin Transformer* + Mask2Former	0.95363	0.75386
IV	Image Super-resolution + Swin Transformer* + Mask2Former + Domain knowledge-enhanced inference (NDWI ,α=0.8)	-	0.75958
V	Image Super-resolution + Swin Transformer* + Mask2Former + Domain knowledge-enhanced inference (AWEI, α=0.2)	-	0.83426
VI	Image Super-resolution + Swin Transformer* + Mask2Former + Domain knowledge-enhanced inference (AWEI, α=0.2) + Knowledge-based processing (Rules 1) + Re-train#1	-	0.85649
VII	Image Super-resolution + Swin Transformer* + Mask2Former + Domain knowledge-enhanced inference (AWEI, α=0.65) + Knowledge-based processing (Rules 1 & 2) + Re-train#2	-	0.87582
VIII	Image Super-resolution + Swin Transformer* + Mask2Former + Domain knowledge-enhanced inference (AWEI, α=0.8) + Knowledge-based processing (Rules 1 & 2) + Re-train#3	-	0.8825

weight the fusion of P_1 and P_2 to generate the fused flood segmentation map Y_{fused}, as outlined in Eq (1)-(4).

$$P_1 = F(x_{super}^{4h\times 4w\times 3}), \tag{1}$$

$$AWEI = B2 + 2.5 \times B3 - 1.5 \times (B8 + B11) - 0.25 \times B12, \tag{2}$$

$$P_2 = \mathbf{Norm}(AWEI), \tag{3}$$

$$Y_{fused} = \alpha P_1 + (1 - \alpha)P_2. \tag{4}$$

2.3. Konwledge-based processing

Due to the susceptibility of optical imagery to cloud cover and the complex surface features typically found in flood-prone areas, relying solely on optical remote sensing imagery for flood detection is often insufficient. Therefore, we incorporated prior knowledge from ESA WorldCover and Copernicus DEM to develop two rules for further processing of the fused flood segmentation map. The specific rules are as follows: 1. Regions classified as "Permanent water bodies" and "Wetlands" in ESA WorldCover are considered as "Flood". 2. Regions classified as "Bare vegetation" in ESA WorldCover and with a Copernicus DEM value greater than 20 unit are considered as "Non-flood".

Our approach integrates knowledge rules derived from expert insights and hydrological principles. These guiding criteria take into account known water body extents and terrain characteristics associated with flood occurrence, providing extra information for our predictions. Through this diversified approach, we generate flood predictions that are not solely driven by remote sensing imagery, but also incorporate domain-specific knowledge. Our method ultimately produces high-precision outputs that reflect the subtle interplay between model predictions and historical data, with the potential to significantly enhance flood management and response strategies.

To further incorporate domain knowledge into the RSFM, we utilize the post-processed flood segmentation map as pseudo-labels and iteratively feed it into the segmentation network for further training, aiming to enhance the performance of the segmentation model. The details of the iterative process are referenced in Section 3.2.

3. EXPERIMENTS

3.1. Data

The imagery used in our training set consists of Harmonized Landsat Sentinel-2 multi-spectral images with a spatial resolution of 30 meters. The original size of the images is 128×128 pixels. In order to enhance the spatial details of the images, we apply a super-resolution process to the RGB bands, increasing their size by 4×. The labels are processed accordingly. To further improve the recognition performance of our method, we adopt the land cover product at 10 m resolution (i.e., ESA WorldCover) and Copernicus DEM as prior knowledge to finish the knowledge-based processing.

3.2. Implementation details

We select the Swin Transformer branch from the previously proposed RSFM, SkySense, as the encoder, and employ Mask2former as the segmentation decoder. The segmentation

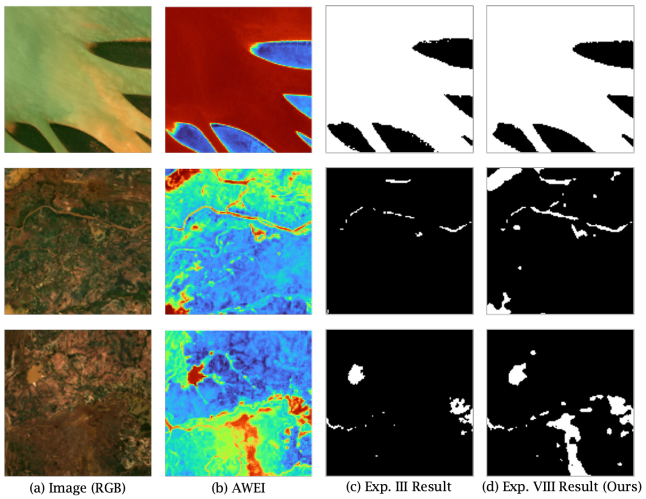

Fig. 3. Visualization of our inference results.

model is trained using the AdamW optimizer, with a base learning rate of 6e-5 and a batch size of 4, on 4 NVIDIA A100 40G GPUs. During the training phase, the main data augmentation strategies employed were random resizing, random cropping, and random flipping. In the inference phase, the value of α is initialized to 0.2. For the re-training process, we perform three iterations, gradually increasing α to 0.8 as the number of iterations increased.

3.3. Ablation study

Table 1 presents the results of our ablation experiments. Our best F1-score achieved is 88.25% (Exp. VIII). It can be observed that relying solely on the original training set yielded unsatisfactory segmentation performance (Exp. III). However, incorporating domain knowledge enhancement significantly improves the performance (Exp. V). Furthermore, employing a post-processing strategy based on knowledge rules further enhanced the overall accuracy of the flood segmentation maps (Exp. VI and VII). As the segmentation network is iteratively trained with domain knowledge, its performance become competitive (Exp. VIII). These experimental results demonstrate the effectiveness of the components in our approach. As shown in Figure 3, the visualized flood detection results also verify the effectiveness of our method.

4. CONCLUSION

This study aims to develop a domain knowledge-aware framework based on the RSFM for extracting flood information from multi-spectral remote sensing images, with the goal of improving the accuracy and robustness of flood detection. By using advanced semantic segmentation models, leveraging spectral information with the AWEI, and incorporating prior knowledge, we achieve competitive results. Experimental results demonstrate the effectiveness of our approach, as evidenced by our second-place ranking in the 2024 IEEE GRSS Data Fusion Contest Track 2 test phase. In the future, we plan to further improve segmentation accuracy by using pre-trained backbones on remote sensing imagery, and attempt to embed domain knowledge into the training loss function of the segmentation model.

5. ACKNOWLEDGEMENT

This work was supported by the National Natural Science Foundation of China under Grants 42371321 and 42030102. The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University. The authors would like to thank the IEEE GRSS Image Analysis and Data Fusion Technical Committee, the Space for Climate Observatory, CNES, NASA, and CERFACS for organizing the Data Fusion Contest.

6. REFERENCES

[1] Goutam Konapala, Sujay V Kumar, and Shahryar Khalique Ahmad, "Exploring sentinel-1 and sentinel-2 diversity for flood inundation mapping using deep learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 180, pp. 163–173, 2021.

[2] Yansheng Li, Bo Dang, Yongjun Zhang, and Zhenhong Du, "Water body classification from high-resolution optical remote sensing imagery: Achievements and perspectives," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 187, pp. 306–327, 2022.

[3] Tawatchai Tingsanchali, "Urban flood disaster management," *Procedia engineering*, vol. 32, pp. 25–37, 2012.

[4] Yansheng Li, Song Ouyang, and Yongjun Zhang, "Combining deep learning and ontology reasoning for remote sensing image semantic segmentation," *Knowledge-based systems*, vol. 243, pp. 108469, 2022.

[5] Yansheng Li, Bo Dang, Wanchun Li, and Yongjun Zhang, "Glh-water: A large-scale dataset for global surface water detection in large-size very-high-resolution satellite imagery," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, vol. 38, pp. 22213–22221.

[6] Yansheng Li, Yuhan Zhou, Yongjun Zhang, Liheng Zhong, Jian Wang, and Jingdong Chen, "Dkdfn: Domain knowledge-guided deep collaborative fusion network for multimodal unitemporal remote sensing land cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 186, pp. 170–189, 2022.

[7] Bo Dang and Yansheng Li, "Msresnet: Multiscale residual network via self-supervised learning for water-body detection in remote sensing imagery," *Remote Sensing*, vol. 13, no. 16, pp. 3122, 2021.

[8] Xin Guo, Jiangwei Lao, Bo Dang, Yingying Zhang, Lei Yu, Lixiang Ru, Liheng Zhong, Ziyuan Huang, Kang Wu, Dingxiang Hu, Huimei He, Jian Wang, Jingdong Chen, Ming Yang, Yongjun Zhang, and Yansheng Li, "Skysense: A multi-modal remote sensing foundation model towards universal interpretation for earth observation imagery," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

[9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[10] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 10012–10022.

[11] Bowen Cheng, Ishan Misra, Alexander G Schwing, Alexander Kirillov, and Rohit Girdhar, "Masked-attention mask transformer for universal image segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 1290–1299.