Rapid Flood Mapping: Outcome of the 2024 IEEE GRSS Data Fusion Contest

Jiepan Li, He Huang, Wei He, Hongyan Zhang, Liangpei Zhang, *Fellow, IEEE,* Ting Liu, Mengke Yuan, Chaoran Lu, Kaixuan Lu, Baochai Peng, Heyang Duan, Mengya Li, Pan Zhang, Tao Wang, Tongkui Liao, Yansheng Li, Bo Dang, Fanyi Wei, Jieyi Tan, Yangjie Lin, Claudio Persello, *Senior Member, IEEE,* Saurabh Prasad, *Senior Member, IEEE,* Gemine Vivone, *Senior Member, IEEE,* Vincent Lonjou, Frédéric Bretar, Raquel Rodriguez-Suquet, Pauline Guntzburger, Vincent Poulain, Jacqueline Le Moigne, *Life Fellow, IEEE,* Benjamin Smith, Sujay Kumar, Thomas Huang, Sophie Ricci, Thanh Huy Nguyen, Andrea Piacentini

***Abstract*—This article presents the scientific outcomes of the 2024 Data Fusion Contest (DFC24) organized by the Image Analysis and Data Fusion Technical Committee (IADF TC) of the IEEE Geoscience and Remote Sensing Society (GRSS), the Space for Climate Observatory (SCO), the Centre national d'etudes spatiales (CNES), the National Aeronautics and Space Administration (NASA), and the Centre Européen de Recherche et de Formation Avancée et Calcul Scientifique (CERFACS). The contest aims to advance image analysis and data fusion algorithms that generate reliable flood maps from multi-modal Earth observation imagery. The DFC24 provides a large-scale, multi-modal flood mapping benchmarking dataset and comprises two challenging competition tracks on the flood mapping task, one based on Synthetic Aperture Radar (SAR) imagery, and another using passive-optical imagery. Additional features, such as a digital terrain model and land-use and water occurrence, are also provided to the participants. This paper presents the methods and results obtained by the first and second-ranked teams of each track. During the development phase, 1935 people registered for the contest, while at the end 46 for Track 1 and 52 for Track 2 teams competed during the test phase in the two tracks, respectively. The data of this contest are openly available to the community for further research, development, and refinement of Geospatial Artificial Intelligence (GeoAI), data fusion, and flood mapping methods.**

***Index Terms*—Transformers, convolutional neural networks, deep learning, data fusion, flood mapping, remote sensing.**

I. Introduction

As a result of climate change, extreme hydro-meteorological events are becoming increasingly frequent, provoking important socio-economic and cultural damages and causing more than 70000 deaths per year. Rapid flood mapping products built on geospatial imaging modalities play an important role in informing flood emergency response and management. These maps are generated from remote sensing data acquired before, during, or after an event to quantify the extent of flooding. The information they provide is crucial for emergency response and damage assessment. There has been a growing interest in generating flood maps from SAR [1]–[5] and passive-optical Earth observations [6], [7], as well as using multi-source Earth observation data [8].

The DFC24, organized by the IADF TC of the GRSS, the SCO, the CNES, the NASA, and the CERFACS, aims to advance image analysis and data fusion algorithms that generate reliable flood maps from multi-modal Earth observation imagery, see Fig. 1. The DFC24 provides a large-scale, multi-modal flood mapping benchmarking dataset and comprises two challenging competition tracks on the flood mapping task, one based on SAR imagery, and another using passive-optical imagery. Additional features, such as a digital terrain model and land-use and water occurrence, are also provided to the participants.

The contest is designed as a benchmark competition following previous editions, e.g., [9]–[17], and consists of two parallel tracks:

- Track 1: Flood mapping from SAR imagery;
- Track 2: Flood mapping from passive-optical imagery.

The reference data is sourced from the labeled flood extent provided by the Copernicus Emergency Management Service (EMS) Rapid Mapping [18] for Track-1 and from the labeled OPERA Dynamic Surface Water Extent CalVal database for Track-2 [19]. Additional reference data for both Track-1 and 2 is obtained from the CERFACS hydrodynamics modeling with data assimilation from in-situ and remote sensing data [20], [21].

Track 1: Flood mapping from SAR data

The focus of Track 1 is to generate water-cover maps from Copernicus Sentinel-1 SAR imagery [22]. All Sentinel-1 images have been processed using S1Tiling to generate time series of calibrated, ortho-rectified and filtered images on any terrestrial region of the Earth [23]. The resulting images were registered to Sentinel-2 optical images, using the same Military Grid Reference System (MGRS) geographic reference. Both polarizations were considered to generate and process VV and VH products that are sensitive to surface scattering reflections such as water bodies. An analysis-ready dataset covering 20 flood sites and events is provided, composed of 2331 patches based on Sentinel-1 images of 512×512 pixels size. This dataset is divided into 1631 patches for the training phase, 349 for the validation phase, and 351 for the test phase. They originate from Copernicus Emergency Management Service dataset (1593 patches) and CERFACS simulations using hydrodynamics modeling with data assimilation over the Garonne river in France (434 patches) and the Ohio river in the US (304 patches).

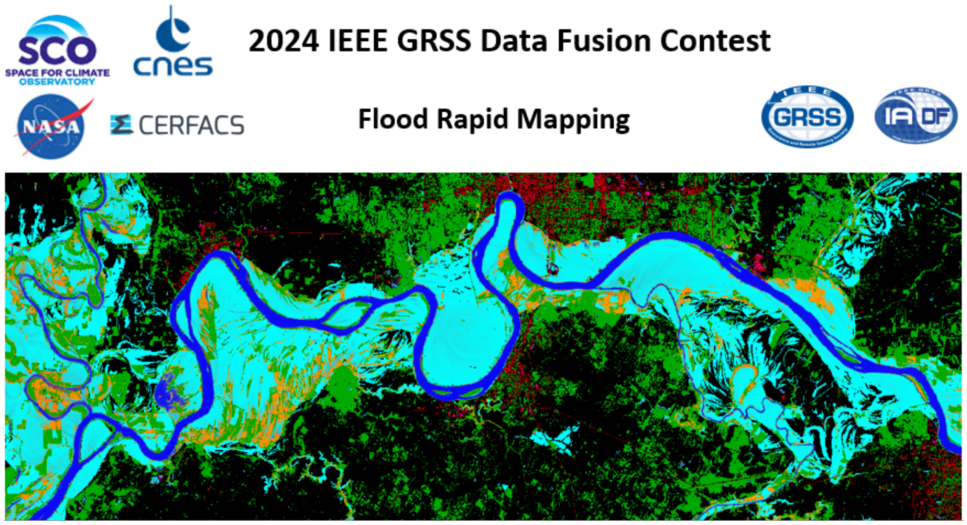


Fig. 1: The banner image of the 2024 IEEE GRSS Data Fusion Contest.

Track 2: Flood mapping from passive-optical imagery

The focus of Track 2 is mapping the water surface from the Harmonized Landsat and Sentinel-2 optical imagery (HLS, [24]). Despite having several multispectral spectral bands and a significantly better signal-to-noise ratio, water reflectance is highly variable in the optical domain. In addition, the availability of "useful" optical data around flooding events is generally poorer due to cloud cover. In this track, images from a set of 72 events are provided over the world, with the goal of accurately determining water vs. non-water pixels in these event areas by fusing data from one or more of the provided data sources. With optical data being less resolved, 30m instead of 10m as in Track-1, and also less availability due to cloud coverage, the patch size has been reduced to 128×128 pixels in Track-2. We obtained a total of 891 patches distributed as follows: 306 training, 125 validation, and 460 test patches. They originate from the OPERA dataset (194), CERFACS simulations over the Garonne river in France (94 patches) and the Ohio river in the US (206 patches) and Copernicus Emergency Management Service dataset (397 patches).

THE DATASET

Images for DFC24 are acquired using Sentinel-1 (SAR), Sentinel-2 (passive-optical) and Landsat-8/9 (passive-optical). Additionally, the following data "sources" are made available to the participants: Digital Elevation Model (DEM) from Merit [25] and Copernicus DEM [26], ESA world cover map [27], and water occurrence probability from the Global Surface Water Dataset [28]. The Water Occurrence dataset reveals the frequency of surface water presence from March 1984 to December 2021, enabling a comprehensive examination of global water dynamics and offering insights into locations featuring permanent water bodies or areas prone to flooding. The data of this contest remains openly available to the community[1].

[1]https://ieee-dataport.org/competitions/2024-ieee-grss-data-fusion-contest-flood-rapid-mapping

II. CONTEST ORGANIZATION AND SUBMISSIONS

The contest consisted of two phases.

- **Phase 1**: Participants are provided training data and additional validation images (without corresponding reference data) to train and validate their algorithms. Participants can submit results for the validation set to the Codalab competition website to get feedback on their performance. The performance of the best submission from each account was displayed on the leaderboard. In parallel, participants are expected to submit a short description of the approach used to be eligible to enter Phase 2.
- **Phase 2**: Participants received the test data set (without the corresponding reference data) and submitted their results within seven days from the release of the test data. After evaluation of the results, three winners for each track were announced.

The training and validation datasets were made available on January 8, 2024 via IEEE DataPort. The evaluation server with a public leaderboard was also open on January 8, 2024 so that participants could submit prediction results for the validation set to the Codalab competition to get feedback on the performance of their approaches. Participants had to submit a short description of the approach used by March 1, 2024, to enter the test phase. The test phase was scheduled from March 11, 2024 through March 17, 2024. The test phase is kept short to ensure an objective and fair comparison among methods. Participants have an opportunity to provide an updated (final) description of their approach. After the final check of the submitted semantic segmentation results, comparing them with the undisclosed ground truth for testing, winners were announced on March 29, 2024.

More information regarding data download and registration to the evaluation server can be found at the IADF TC website (https://www.grss-ieee.org/community/technical-committees/2024-ieee-grss-data-fusion-contest/).

We received 1935 registrations at the CodaLab competition website during the development phase (see Table I). For Track 1, there were 809 unique registrations at the CodaLab

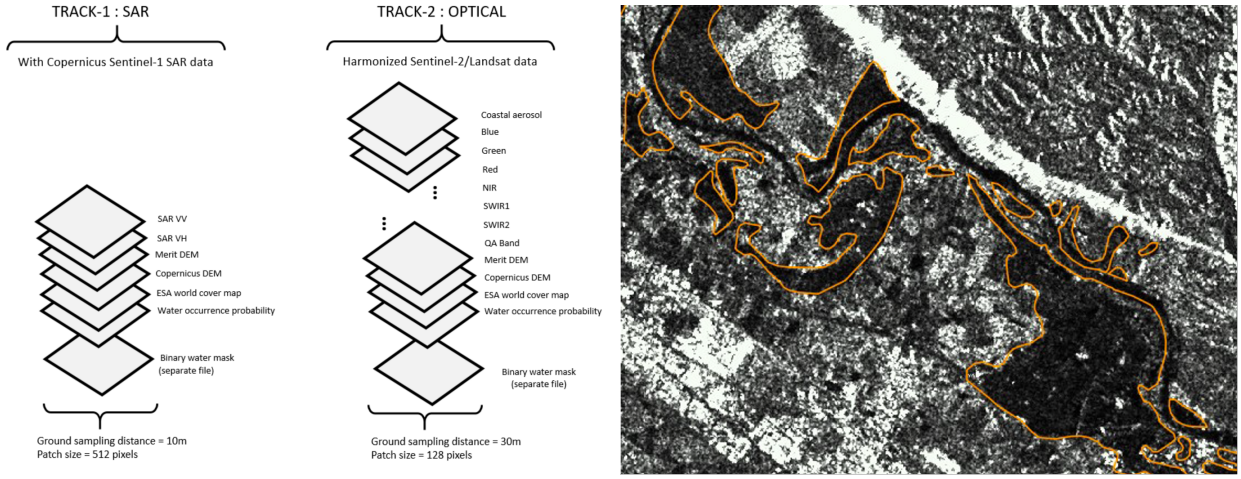


Fig. 2: Illustrating the dataset layers (left), and example SAR image along with annotation of flooded area over Navarre, Spain, 2018-04-13 (right).

TABLE I: Registration and submission statistics of the two tracks

	Track 1	Track 2	Total
Applications	809	1126	1935
Approved applications	772	1093	1865
Teams with successful submissions	46	52	98
submissions	2136	3038	5174

competition website, 772 of which were approved. 46 teams entered the test phase after screening the descriptions of their approaches submitted by the end of the development phase. 2136 submissions were received during the development phase, with active participation from all registered teams. During the test phase, the maximum number of submissions per team was limited to 5 per day.

For Track 2, there were 1126 unique registrations at the CodaLab competition website during, 1126 of which were approved. 52 teams entered the test phase after screening the descriptions of their approaches submitted by the end of the development phase. In total, 3038 submissions were received during the development phase, with active participation from all registered teams. During the test phase, the maximum number of submissions per team was limited to 5 per day.

The first to third-ranked teams in each track were awarded as winners of the DFC2024 for each track and were invited to present their solutions during the 2024 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2024).

In the following, we list the winning teams of the DFC2024 in Track 1

- **1st place**: team *Henryljp*; He Huang, Jiepan Li, Wei He, Hongyan Zhang, Liangpei Zhang from Wuhan University and China University of Geosciences, China.
- **2nd place**: team *tingliu*; Ting Liu, Mengke Yuan, Chaoran Lu, Kaixuan Lu, Baochai Peng, Heyang Duan, Mengya Li, Pan Zhang, Tao Wang , Tongkui Liao from PIESAT Information Technology Co, Ltd., Beijing, China.
- **3rd place**: team *Genshin1/OnePiece*; Shuchang Zou, Qian Yang from PIESAT Information Technology Co, Ltd., Beijing, China.

and in Track 2

- **1st place**: team *Henryljp* team; Jiepan Li, He Huang, Wei He, Hongyan Zhang, Liangpei Zhang from Wuhan University and China University of Geosciences, China.
- **2nd place**: team *bodang1220* team; Yansheng Li, Bo Dang, Fanyi Wei, Jieyi Tan, Yangjie Lin from Wuhan University, China.
- **3rd place**: team *IPIU-XDU*; Xiaoqiang Lu, Tong Gou, Zhongjian Huang, Yuting Yang, Licheng Jiao, Lingling Li, Xu Liu, Fang Liu from Xidian University, China.

The two best-ranked teams in both tracks were invited to provide a detailed presentation on their respective approaches that won the DFC24 in this paper. Details of their methodology are provided in Section III through VI.

III. Track 1 - First Place: Team Henryljp

A. Method

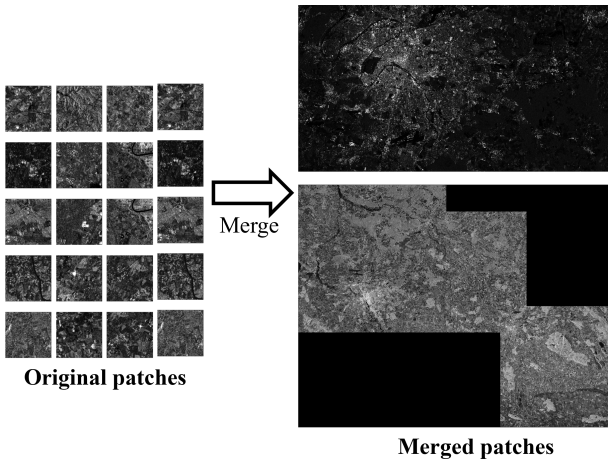


Fig. 3: An illustration of the merging process.

1) Data Preprocessing Strategy: Upon analyzing the dataset provided by the official source, we identified inconsistent degrees of overlap among the original 512×512 images. To address this issue, we adopted a manual visual stitching strategy, dedicating one month to merging the training set (as shown in Fig. 3). This effort resulted in 30 large images encompassing diverse scenes. Subsequently, we employed a five-fold cross-validation approach to partition these large images into training and validation sets, ensuring sufficient generalization capability for the model.

2) Feature Extraction: Given the vast array of land features captured in remote sensing images, the intricate nature of contextual information, and the inherent irregularities in SAR images, we incorporate the concept of uncertainty [32], [33] to tackle these challenges and propose the Uncertainty-Aware Fusion Network (UAFNet). As shown in Fig. 4, we first adopted a general encoder-decoder network to get a relatively uncertain extraction map. Regarding the general encoder-decoder part, we adopted PVT-V2 [34] as the encoder backbone to extract multi-level features from the input image and introduce a multi-branch dilation convolution blocks (MBDC) to enhance the encoded features ($E_i, \{i = 1, 2, 3, 4\}$), and used a typical cross-fusion strategy (Feature Pyramid Network (FPN [35])) to obtain a relatively uncertain extraction map M_4. The whole process can be represented as:

$$\begin{aligned} F_i &= MBDC(E_i), i = 1, 2, 3, 4, \\ M_4 &= FPN(F_1, F_2, F_3, F_4). \end{aligned} \tag{1}$$

Based on the output features ($F_i, \{i = 1, 2, 3, 4\}$) and uncertain extraction map M_4, our UAFNet acts as a decoder strategy to deal with the general flood extraction challenges and output a refined extraction map with low uncertainty.

3) Uncertainty-Aware Fusion Module: Flood regions in RS imagery often present ambiguous boundaries and complex backgrounds, which introduce notable prediction uncertainty. To mitigate this, we adopt an Uncertainty-Aware Fusion Module (UAFM) inspired by UANet [36] that uses pixel-wise uncertainty to guide multi-scale feature fusion.

First, we compute foreground and background uncertainty maps from the extraction map M (after a $Sigmoid$ activation):

$$\begin{aligned} U_f &= Sigmoid(M) - 0.5, \\ U_b &= 0.5 - Sigmoid(M), \end{aligned} \tag{2}$$

where values close to 0 indicate low uncertainty while values near 0.5 indicate high uncertainty for the corresponding perspective.

We then discretize the non-negative part of these uncertainty values into five uncertainty ranks (URA) to emphasize ambiguous pixels in a graded manner. Concretely, for a pixel uncertainty $U \in [0, 0.5]$ we use the following rank mapping (higher rank $\Rightarrow$ higher uncertainty):

$$\mathrm{URA}(U) = \begin{cases} 5, & 0.0 \le U < 0.1, \\ 4, & 0.1 \le U < 0.2, \\ 3, & 0.2 \le U < 0.3, \\ 2, & 0.3 \le U < 0.4, \\ 1, & 0.4 \le U \le 0.5, \\ 0, & U < 0 \end{cases} \tag{3}$$

For implementation we convert the rank $r \in \{0, 1, \ldots, 5\}$ into a fusion weight by a simple linear mapping, e.g., $w = r/5$, so that more uncertain pixels receive larger weights during feature enhancement.

At each fusion stage, these uncertainty-derived weights reweight both high-level and low-level features to highlight ambiguous regions from foreground and background viewpoints. Taking the fusion of F_4 and F_3 as an example, the uncertainty-enhanced high-level feature is

$$F_4^u = Conv_{1\times1}\big(Concat(W_f^4 \odot F_4,\ W_b^4 \odot F_4)\big), \tag{4}$$

where W_f^4 and W_b^4 are the weight maps derived from the URA applied to M_4, $\odot$ denotes element-wise multiplication, and $Conv_{1\times1}$ restores channel dimensions after concatenation. The lower-level feature F_3 is enhanced similarly using upsampled versions of the weight maps, producing F_3^u. The two enhanced features are then fused and decoded:

$$\begin{aligned} G_3 &= Conv_{3\times3}\big(Concat(F_3^u, Up(F_4^u))\big), \\ M_3 &= Conv_{3\times3}(G_3). \end{aligned} \tag{5}$$

This top-down uncertainty-guided fusion is applied iteratively to produce progressively less uncertain extraction maps at deeper decoding stages. All intermediate outputs are supervised by binary cross-entropy:

$$Loss = \sum_{i=1}^{4} BCE(M_i, GT), \tag{6}$$

where M_i denotes the predicted map at stage i and GT is the ground truth.

4) Post-Processing Strategy: To create a three-dimensional input for leveraging the PVT-v2 pre-trained on ImageNet [37] and enhancing the feature representation capability, we concatenated VH, VV, and VH together. Furthermore, by utilizing various versions of PVT as encoders (PVT-V2-B2, PVT-V2-B3, PVT-V2-B4, PVT-V2-B5), we trained four instances of UAFNet and adopted the multi-model fusion strategy to improve the extraction performance. Additionally, upon analyzing the official training data, we discovered the crucial role of provided land cover data in post-processing. To enhance the post-processing process, we merged the predicted results of our strategy on the test set with the designated water regions from the land cover data. Finally, it is worth noting that we applied twelve enhancements to each image in the training set by using six rotation angles and six flip modes, effectively expanding the dataset.

B. Results

1) Visual Comparison: As depicted in Fig. 5, we selected two typical examples to illustrate the effectiveness of our proposed UAFNet. It can be seen that compared to other methods, there are some slight issues in extracting details. Our method can achieve results that are closer to the true labels in terms of details. Due to the limitation of resolution, we can also see from VH and VV images that it is difficult to accurately determine whether certain pixel areas belong to flood. However, to some extent, our uncertainty strategy can overcome this challenge.

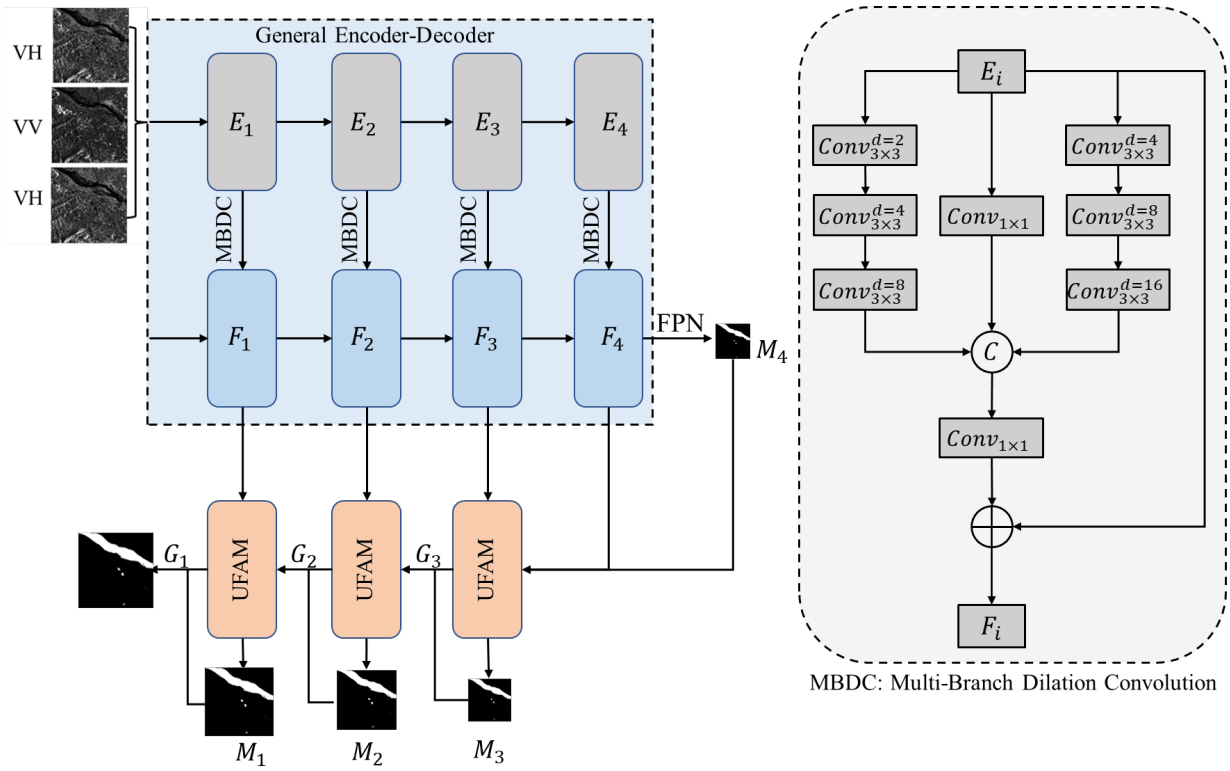
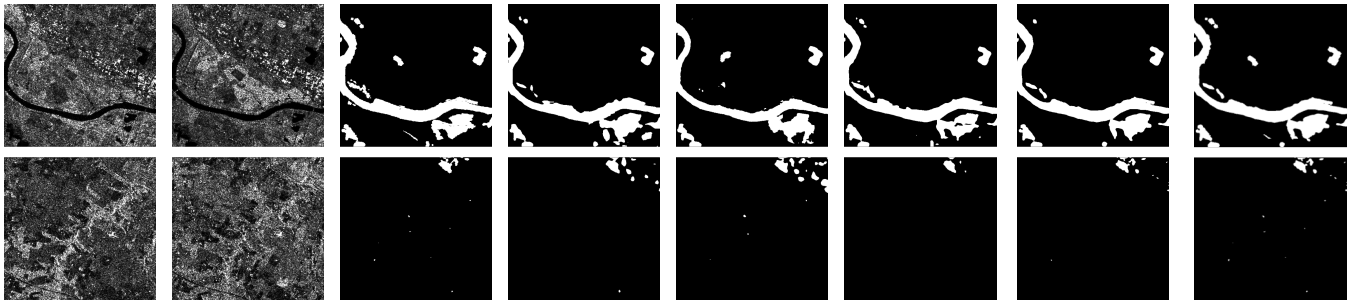


Fig. 4: The structure of the Uncertainty-Aware Fusion Network.



(a) VH (b) VV (c) GT (d) HRNet [29] (e) UNet [30] (f) SF [31] (g) UAFNet (h) Ours

Fig. 5: Visual comparison between UAFNet and the compared methods. SF is short for SegFormer, and Ours indicates the visual results of **Multi-UAFNet+post-processing.**

TABLE II: Quantitative results on the official test dataset. * indicates the model variant without UAFM, and PP denotes the post-processing strategy.

Method	F1 (%)	Params (M)	FLOPS (G)
UNet (R50)	73.935	30.070	33.053
SegFormer (MIT-B5)	74.996	65.846	93.822
UNet (B2)	75.363	40.126	33.053
UAFNet* (B2)	76.808	**25.422**	**23.005**
UAFNet (B2)	78.756	25.589	23.902
UAFNet (B3)	78.773	45.464	38.549
UAFNet (B4)	79.069	62.782	54.837
UAFNet (B5)	80.824	82.182	62.995
Multi-UAFNet	81.890	216.017	180.283
Multi-UAFNet + PP	**82.985**	216.017	360.566

2) Quantitative Comparison: As illustrated in Tab. II, it is evident that the different methods and backbone network versions exhibit varying levels of performance in the flood extraction task. The traditional ResNet-50-based [38] UNet [30] method achieves an F1 score of 73.935%, while SegFormer [31] slightly outperforms it with a score of 74.996%. However, when combining UNet with the PVT-V2-b2 backbone network, there is a noticeable improvement in performance, with an F1 score of 75.363%. This indicates that the PVT series of backbone networks have a positive impact on flood extraction tasks. At the same time, compared to UNet (PVT-V2-B2), UAFNet (PVT-V2-B2) demonstrates a significant improvement in the F1 score, achieving 78.756%. Meanwhile, we can observe that the improvement brought by UAFM can directly enhance the model's performance from an F1 score of 76.808% to 78.756%, fully demonstrating the effectiveness of UAFM. As we upgrade the backbone network version from PVT-V2-B2 to PVT-V2-B5, the performance of UAFNet gradually improves, reaching a peak of 80.824%.

This underscores the superiority of the UAFNet approach in handling flood extraction tasks and its scalability with more capable backbone networks. Furthermore, we explore a multi-model fusion strategy, known as Multi-UAFNet, which combines the results from four models ranging from PVT-V2-B2 to PVT-V2-B5. This strategy further enhances the accuracy of flood extraction, achieving an F1 score of 81.890%, surpassing the best performance of a single model. Finally, by introducing a post-processing strategy, we optimize the results of Multi-UAFNet, achieving the highest F1 score of 82.985%. This not only demonstrates the effectiveness of the UAFNet approach but also highlights the importance of post-processing in enhancing the precision of flood extraction.

In summary, our UAFNet method, coupled with the PVT series of backbone networks, significantly improves the performance of flood extraction tasks. Through multi-model fusion and post-processing, we further enhance the accuracy of flood detection, providing robust technical support for flood monitoring and early warning systems.

3) Computational Efficiency and Practical Implications: To further assess the practicality of the proposed framework for real-time flood mapping, we analyze the computational efficiency of our models and other winning methods, as summarized in Table II. Several top-ranking approaches rely on heavy preprocessing steps, such as manual image stitching and extensive data augmentation. These operations are mainly dataset-specific engineering optimizations adopted during the competition and are difficult to quantify in terms of computational complexity. In contrast, our analysis focuses on model-side factors that can be objectively measured, including the number of parameters (Params) and floating-point operations (FLOPs).

As shown in Table II, the proposed UAFNet series achieves a favorable trade-off between accuracy and efficiency. Compared with standard baselines such as UNet (R50) and SegFormer (MIT-B5), UAFNet not only improves the F1 score but also significantly reduces both Params and FLOPs, demonstrating superior computational efficiency. Additionally, we report the ensemble variant (Multi-UAFNet) and its post-processed version (Multi-UAFNet + PP), which achieve the highest overall accuracy while naturally increasing the computational cost due to multi-model fusion. These results highlight that UAFNet maintains a strong balance between performance and efficiency, making it a promising and scalable solution for large-scale flood mapping applications.

C. Discussion

During the 2024 IEEE GRSS Data Fusion Contest, we proposed the UAFNet to tackle uncertainty challenges in flood mapping using SAR data. Experimental results highlighted the effectiveness of UAFNet, securing first place in Track 1 of the contest [39]. In the future, we will continue to explore the approach to achieve high-precision flood extraction on a global scale.

IV. Track 1 - Second Place: Team tingliu

A. Method

1) Data Normalization Module: We have analyzed the distribution of water and background, and observe that the categories contained in the dataset are unbalanced, water area is accounted for less than 7%. Also, the multi-source data is also analyzed and we conclude that 1) *DTM* (Merit DEM) and *DSM* (Copernicus DEM GLO-30) have the same distribution, and we use the averaged value ($AvgDEM$) for our model to address the invalid data in these data sources. Additionally, the areas where $AvgDEM \leq 250$ contain 99% water area. 2) The value in the water occurrence probability map is below 100 in most areas, with the occurrence probability being less than 1.6%. 3) the multi-modal data sources can be classified into two groups: the polarized VV and VH with low-level semantic information and $AvgDEM$, *LCM* (ESA World Cover Map), and *WOP* (Global Surface Water occurrence probability) with high-level semantic information.

Based on these findings, we will standardize the two input groups as follows. For the low-level semantic data group, we utilize NMT methods [40](N times the mean of non-zero data for truncation and stretching of the original SAR data) to convert the polarized VV, VH into a uint8 data format. Subsequently, by expanding the radar vegetation index, we generate a pseudo-color image. On the other hand, for the high-level semantic data group, we truncate the $AvgDEM$ and standardize the three spectrum bands using the mean and standard deviation.

2) Siamese Network: A Siamese network with dual backbone branches is built for various multi-modal encoders, and utilizes the UperNet [41] decoder for flood semantic segmentation. State-of-the-art backbones like ConvNeXt [42] (CT: ConvNeXt-*Tiny*, CS: ConvNeXt-*Small*, CB: ConvNeXt-*Base*, CL: ConvNeXt-*large*), Swin Transformer [43] (SB: SwinTransformer-*Base*), InternImage [44] (IB: InternImage-*Base*) are employed to explore various backbone combinations for the Siamese network. Three distinct feature interaction strategies at different levels: concatenation (*Cat.*), feature exchange (*Ex.*), and Aggregation-Distribution (*AD.*) [45] are evaluated, and *Cat.* is verified to be the optimal feature extraction approach.

B. Results

1) Implementation Details: We employ a three stages training process: 1) Train with 80% of the training dataset and evaluating on the remaining 20%. In this process, we employ the AdamW optimizer with initial learning rate of 2×10^{-4}, a weight decay of 0.05, a scheduler that uses linear learning rate decay, and a linear warm-up of 1500 iterations. Models are trained on 2 GPUs with 12 images per GPU for 80k iterations. The data augmentations adopted by us include random re-scaling within ratio range [0.5, 2.0], random cropping, random flipping, and random rotate with 90 degree. 2) Fine-tune on the entire dataset using different scales, including 512, 768, 1024. The optimizer, scheduler, and other augmentations are the same. The initial learning rate is 10^{-4} and we just train for 40k iterations. 3) Ensemble the best models and just use

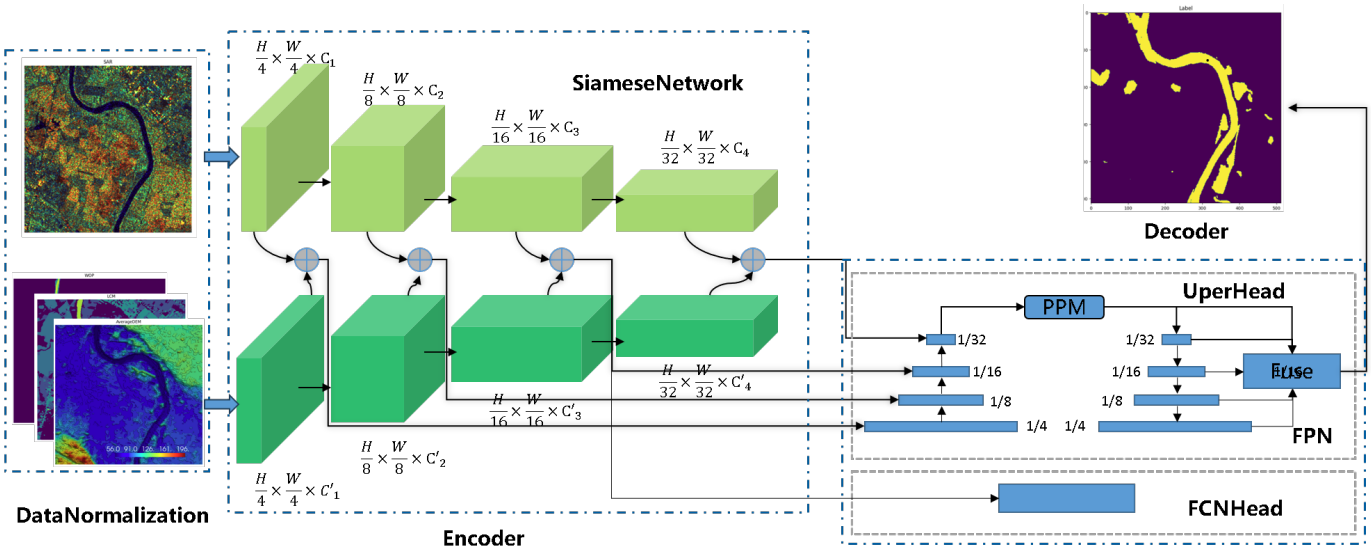


Fig. 6: The Siamese network and the overall network architecture diagram of the proposed water extractor. Features from different semantic levels are visualized using distinct colors in each module.

the predicted result of the test data in the development phase as pseudo label. A pseudo learning is performed on this stage with 20k iterations.

During inference, we employ a fixed data scale of 768 testing approach for the models in the multi-scale training phase. This resolution is maintained throughout the subsequent fine-tuning and pseudo-learning stages. To generate the final output, we combine the outcomes from various models by averaging them, where values exceeding a predetermined threshold are considered indicative of the water region.

2) Main Results and Ablation Study: Table III shows the quantitative results. The highest score of 79.17% is achieved by combining multiple segmentation models with various Siamese networks and iterations. The transformer-based model is not as effective as the convolution-based models in this framework. A brief visual illustration is presented Fig. 7. We found that many results are counter-visual, and other information, such as DEM and LCM, etc., can play a good auxiliary role at this time. We have also explored different input modalities and normalization modules, and discovered that our data normalization module yields the best results. The details are presented in Table IV.

C. Discussion

After conducting thorough experiments, we introduce a Data Normalization Module and a Siamese Network to leverage SAR, DEM, land cover map, and Water Occurrence Probability to tackle the challenges posed by the imaging mechanism of SAR, including shadow, low texture of bare soil, and roads' influence. Future research should focus on developing innovative fusion techniques that can effectively combine information from different modalities in Siamese networks with some strong fusion strategies. This will help in improving their ability to generalize across various tasks and datasets. Overall, continued exploration in this area will be essential to unlock the full potential of these networks in various applications.

V. Track 2 - First Place: Team Henryljp

A. Method

The diverse range of data sources inherently introduces aleatoric uncertainty, primarily due to the inclusion of synthetic and non-authentic data. Furthermore, the extensive coverage of Remote Sensing (RS) imagery, combined with the relatively small proportion of flood-affected areas, leads to a pronounced imbalance between foreground and background classes, thereby amplifying epistemic uncertainty [32]. In addition, the 30-meter resolution of the imagery poses challenges in distinguishing between visually similar features in complex flood scenarios, further intensifying epistemic uncertainty in the models [32]. These factors collectively present significant hurdles to accurate flood detection and analysis. To tackle these uncertainties, we propose an Uncertainty-aware Detail-Preserving Network (UADPNet) designed for rapid flood mapping using multi-source optical data. In the following parts, we will provide a comprehensive overview of our framework, encompassing the data preprocessing strategy, the UADPNet architecture, and the post-processing methodology.

1) Data Preprocessing Strategy: Upon analyzing the dataset provided by the official source, we identified inconsistent degrees of overlap among the original 128×128 images. To address this issue, we adopted a manual visual stitching strategy, (as shown in Fig. 8). This effort resulted in 10 large images encompassing diverse scenes. Subsequently, we employed a five-fold cross-validation approach to partition these large images into training and validation sets, ensuring sufficient generalization capability for the model.

2) Stage A: Aleatoric Uncertainty Estimator: In the realm of flood extraction, our primary objective is to minimize the

TABLE III: The main results of our method. All models utilize the Data Normalization Module and Siamese Network structure, with Bce and BDice loss equally weighted. The notation "X/Y/Z" denotes the number of training iterations for each of the three stages, specifically "X" iterations for stage 1, "Y iterations for stage 2, and "Z" iterations for stage 3.

Model	Scale	Iterations (k)	Development Phase (F1)	Test Phase (F1)
CB+CS	1024	80/40/10	0.94832	0.78347
CB+CS	1024	80/40/12	0.94859	0.7828
CB+CS	1024	80/40/20	0.94858	0.7826
CB+CS	1024	80/40/16	0.9487	0.7817
SB+SS	1024	80/40/20	/	0.78143
IB+IB	1024	80/40/0	/	0.78216
IB+IB	1024	80/40/40	/	0.78502
Model Ensemble			**0.94882**	**0.79170**

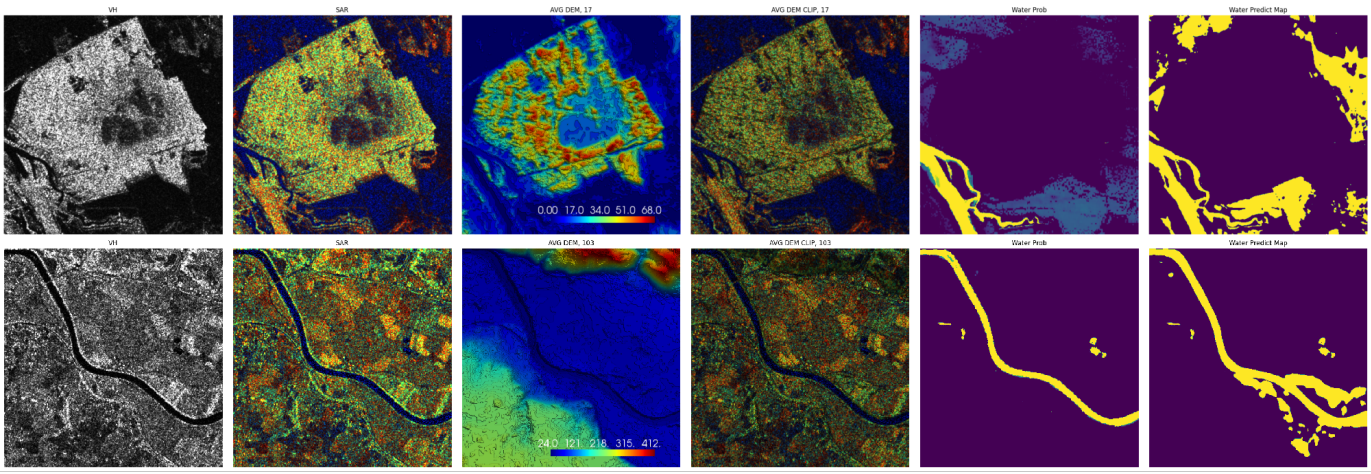


Fig. 7: Two data samples are presented to illustrate the process of our method. The first column displays the original SAR data, represented as VH. The second column showcases the generated pseudo-color image, while the third column features the clipped average DEM. The overlay of the average DEM and SAR is presented in the fifth column. Lastly, the final two columns display the water probability map and the predicted result map.

TABLE IV: The performance of different input modalities. The **bold** value indicates the best performance, which is achieved by our Data Normalization Module.

InputModal	Framework	F1
VV,VH,VV	UperNet_CB	0.904
VV,VH,VV/VH	UperNet_CB	0.901
VV,VH,RVI	UperNet_CB	0.905
VV,VH,VV AvgDEM,LCM,WOP	Upernet_CB + CS	0.920
VV,VH,VV AvgDEM,LCM,WOP	SiameseNet_CB + CS	0.921
VV,VH,RVI AvgDEM,LCM,WOP	**SiameseNet_CB + CS**	**0.926**

Fig. 8: An illustration of the merging process.

overall loss function, which serves as a metric for evaluating the model's performance. This loss function is formally expressed as:

$$\min_{\theta} \mathbb{E}_{X,Y}[\mathcal{L}(f(X;\theta),Y)] \approx \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}(f(x_i;\theta),y_i), \tag{7}$$

where θ denotes the learned parameters by the model $f(\cdot)$, and (x_i, y_i) are N individual samples drawn from the joint data distribution $p(X, Y)$, where X represents the concatenation of the NIR, Green, and Blue bands from the dataset, and Y refers to the corresponding ground truth. The function $\mathcal{L}(\cdot)$ quantifies the loss between the model's predictions and the ground truth.

As mentioned earlier, due to the presence of simulated data in the training dataset, coupled with a certain degree of human error in the real observation data, there is an inevitable uncertainty (aleatoric uncertainty) at the data level. To address this challenge, we utilize the Conditional Variational Autoencoder (CVAE [47]) to estimate this aleatoric uncertainty using maximum likelihood estimation. In CVAE, the prior distribution of latent variables is conditioned on the input data X and follows a Gaussian distribution. A typical conditional generative model comprises three key components: conditional variables X, latent variables z, and output variables Y. The latent variable

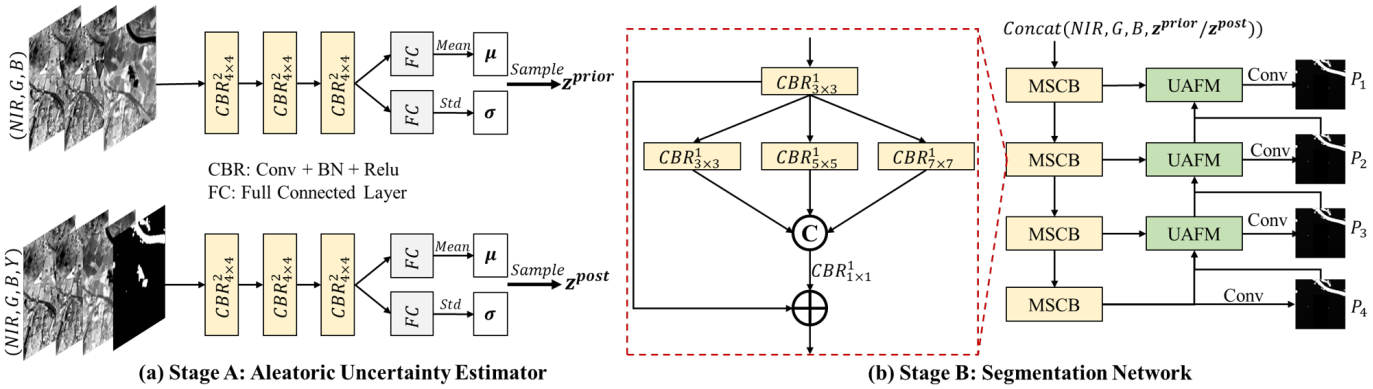
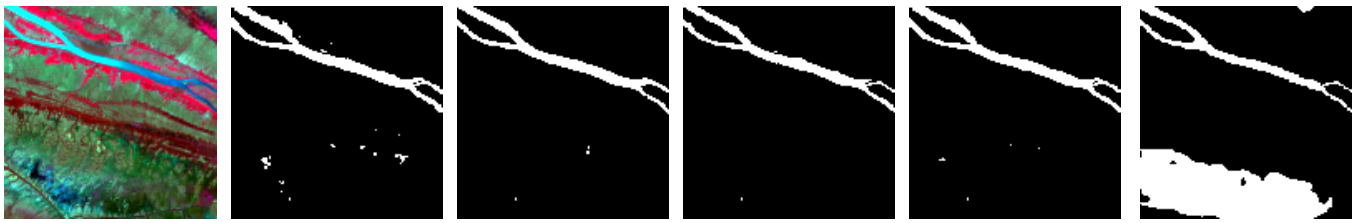


Fig. 9: The pipeline of the proposed Uncertainty-aware Detail-Preserving Network (UADPNet), which is composed of two stages, i.e., an Aleatoric Uncertainty Estimator (AUE) and a Segmentation Network.



(a) RGB Image (b) Deeplab v3+ [46] (c) HRNet [29] (d) UNet [30] (e) SegFormer [31] (f) Ours

Fig. 10: Visual comparison between the proposed framework and the compared methods.

is defined by the conditional distribution $P_\theta(z|X)$, and given the joint input conditional variable X, the distribution of the output variable Y, is obtained as $P_\omega(Y|X,z)$. Additionally, the posterior distribution of z is represented as $Q_\phi(z|X,Y)$. The loss function of CVAE is carefully designed to capture both the reconstruction accuracy and the closeness of the learned latent space to the true prior distribution. It is defined as:

$$L_{CVAE} = \mathbb{E}_{z\sim Q_\phi(z|X,Y)}[-\log P_\omega(Y|X,z)] + \mathbb{D}_{\mathrm{KL}}(Q_\phi(z|X,Y)||P_\theta(z|X)), \tag{8}$$

Here, $P_\omega(Y|X,z)$ represents the likelihood of observing the output variable Y given the latent variable z and the conditional variable X. The second term, $\mathbb{D}_{\mathrm{KL}}(Q_\phi(z|X,Y)||P_\theta(z|X))$, quantifies the closeness between the learned posterior distribution $Q_\phi(z|X,Y)$ and the prior distribution $P_\theta(z|X)$ using the Kullback-Leibler (KL) divergence. By minimizing this loss function, we aim to achieve both accurate reconstructions and a well-regularized latent space, enabling robust flood extraction even in noisy environments.

Adhering to the standard CVAE paradigm, we introduce an Aleatoric Uncertainty Estimator (AUE), whose components are meticulously detailed as follows. Precisely, in this study, we define $P_\theta(z|(X))$ as the prior network, which functions as a mapper, transforming the input image into a compressed latent space. Here, θ represents the parameter set specific to the prior network. Assuming identical network structures and the availability of ground truth Y, we designate $Q_\phi(z|(X),Y)$ as the posterior network, where ϕ comprises the parameters exclusive to the posterior network.

As illustrated in Fig. 9, within the hidden layer networks (both prior and posterior), we concatenate the multiple inputs and employ three 4×4 convolution blocks, each with a stride of 2. This process transforms the concatenated inputs of X (or X and Y in the case of the posterior network) into the latent variable z, which follows a normal distribution $N(\mu, \mathrm{diag}(\sigma^2))$. Here, μ and σ belong to $\mathbb{R}^K$ and represent the mean and standard deviation of the latent variable, respectively. This architecture not only preserves the essence of CVAE but also incorporates uncertainty awareness, rendering AUE a robust and versatile decoder.

3) Stage B: Segmentation Network: As depicted in Fig. 9, our segmentation network follows a U-shaped design akin to UNet. Specifically, we merge the estimated aleatoric uncertainty (EU) with X as the input to the entire network. Subsequently, we employ four Multi-Scale Convolution Blocks (MSCBs) to systematically extract four layers of encoded features.

The MSCB structure is straightforward: it first undergoes a 3×3 operation, followed by three parallel convolution branches (with convolution kernels of sizes 3×3, 5×5, and 7×7) aimed at capturing features with varied receptive fields. These branches are then subjected to dimensional reduction and restoration via a 1×1 convolution, supplemented by a residual connection to preserve the original feature information. This four-layer MSCB processing yields four layers of encoded features ($E_i, i = 1, 2, 3, 4$).

During the decoding phase, we take into account the challenges stemming from class imbalance in the dataset and the minor inter-class disparities in images caused by low resolution, resulting in model-level uncertainties (epistemic uncertainty). To tackle these issues, we employ a progressive decoding strategy utilizing our UAFM introduced in UANet [36]. We leverage the uncertainty evident in the pre-

liminary flood extraction results to guide the model's attention towards challenging samples with higher uncertainty during training. This progressive methodology aims to systematically reduce uncertainties and accomplish precise flood extraction.

Specifically, we first use a 1×1 convolution directly to process E_4 and output P_4. Subsequently, we adopt the uncertainty ranking algorithm from UANet [36] to measure the uncertainty level of each pixel in the feature space and assign different weights based on the measured uncertainty level. In detail, we directly use the $Sigmoid$ function to get the corresponding probabilities of all pixels in the extraction map P_4 from spatial perspective, then we subtract all values of the probability map with 0.5 to measure the uncertainty belonging to foreground flood (U_f) and meanwhile subtract 0.5 with all values of the probability map to measure the uncertainty belonging to background (U_b):

$$\begin{aligned} U_f &= Sigmoid(P_4) - 0.5, \\ U_b &= 0.5 - Sigmoid(P_4). \end{aligned} \tag{9}$$

Subsequently, we rank the uncertainty of the foreground and background into five levels using the URA, which is described as:

$$\mho(i,j) = \begin{cases} \lfloor \frac{0.5 - U_{i,j}}{0.1} \rfloor, U_{i,j} >= 0, \\ 0, U_{i,j} < 0, \end{cases} \tag{10}$$

where $U_{i,j}$ means the pixel in the i_{th} row and the j_{th} column of U_f or U_b.

Afterward, we apply URA to P_4 so that we can get the corresponding foreground uncertainty rank map (R_f^4) and background uncertainty rank map (R_b^4). Then we directly use E_4 and E_3 to multiply with them to highlight the uncertain pixels from both the foreground and background perspectives. Subsequently, we concatenate these enhanced features and recover its original channel to get G_3 by a 1×1 convolution operation, which can be described as:

$$G_3 = Conv_{1\times1}(C(R_f^4 * E_4, R_b^4 * E_4, R_f^4 * E_3, R_b^4 * E_4)), \tag{11}$$

where C is short for Concatenation and G_3 is processed by 1×1 convolution to output P_3 with less uncertainty than P_4. With such a UAFM, we can utilize P_4 to fuse E_4 and E_3 and achieve output G_3 and P_3, utilize P_3 to fuse G_3 and E_2 and achieve output G_2 and P_2, and utilize P_2 to fuse G_2 and E_1 and output P_1, where P_1 can be viewed as the final refined extraction map with the lowest uncertainty.

4) Loss Function: On the whole, we use the simple binary cross-entropy (BCE) loss function to supervise all extraction outputs and use L_{CVAE} to supervise the closeness of the learned latent space to the true prior distribution in AUE. The overall loss is:

$$Loss = \sum_{i=1}^{4} BCE(P_i, GT) + \lambda \cdot L_{CVAE}, \tag{12}$$

where λ is 0.01.

5) Post-Processing Strategy: To enhance the post-processing process, we merged the predicted results of our strategy on the test set with the designated water regions from the land cover data. It is worth noting that we applied twelve enhancements to each image in the training set by using six rotation angles and six flip modes, effectively expanding the dataset.

TABLE V: Quantitative results on the official test dataset.

Method	F1 (%)	Params (M)	FLOPS (G)
UNet	82.093	26.637	**13.196**
HRNet	80.790	41.005	44.337
Deeplab v3+	80.697	10.198	41.742
SegFormer	83.426	33.678	17.577
UNet (MSCB)	84.507	32.745	<u>16.372</u>
UNet++ (MSCB)	84.577	38.640	36.330
Attention UNet (MSCB)	86.033	33.678	47.978
MSCB+UAFM	87.027	**5.423**	23.902
MSCB+UAFM+AUE	87.170	<u>5.942</u>	40.485
Multi-Fusion	<u>89.272</u>	111.005	165.067
Multi-Fusion + Post-Processing	**89.843**	111.005	1,980.804

B. Results

1) Visual Comparison: As illustrated in Fig. 10, we present an example to demonstrate the superiority of our method. It is evident that the lower left portion of the synthetic image is inundated with floodwaters. Nevertheless, none of the contrast methods was able to effectively extract this flood-affected area. Given the constraints of resolution, it becomes challenging to precisely discern flood-prone pixel regions from the synthetic image. However, our uncertainty-based strategy is capable of overcoming this hurdle to a considerable extent.

2) Quantitative Comparison: As shown in Tab. V, we compare various methods and their respective performance metrics in terms of test accuracy and $F1$ score. Methods like UNet, HRNet, Deeplab v3+, and SegFormer achieve accuracies between 80.697% and 83.426%, indicating their effectiveness but with room for improvement. Integrating multi-scale and attention mechanisms leads to better results. Methods like UNet, UNet++, and Attention UNet with Multi-Scale Convolution Block (MSCB) show higher $F1$ scores, ranging from 84.507% to 86.033%. This suggests that leveraging multi-scale features boosts model performance. Furthermore, combining MSCB with techniques like UAFM and AUE further enhances performance, with MSCB+UAFM+AUE achieving an $F1$ score of 87.170%. The most significant improvement comes from the Multi-Fusion strategy, scoring an impressive $F1$ of 89.272%, indicating the benefits of fusing UADPNet trained on different datasets and components. Finally, applying post-processing techniques on Multi-Fusion further improves performance, achieving the highest $F1$ score of 89.843%. This underscores the importance of incorporating advanced techniques and optimizing post-processing for optimal performance in semantic segmentation tasks.

3) Computational Efficiency and Practical Implications: To further evaluate model practicality, we compare F1 scores, parameter counts, and FLOPs in Table V. Conventional CNNs (UNet, HRNet, Deeplab v3+) and transformers (SegFormer) reveal diverse accuracy–efficiency trade-offs: UNet attains 82.09% F1 with 13.20G FLOPs, while SegFormer achieves 83.43% F1 at 17.58G FLOPs. Advanced CNN variants (UNet++, Attention UNet) reach 84–86% F1 but at much higher cost (up to 48G FLOPs). Our proposed

MSCB+UAFM model achieves the best balance, reaching 87.03% F1 with only 5.42M parameters and 23.9G FLOPs. Adding the AUE module further improves performance to 87.17% F1 at modest cost (5.94M, 40.5G FLOPs), showing enhanced representation and robustness. In contrast, ensemble-based Multi-Fusion attains the highest accuracy (89.84%) but requires enormous computation (111M parameters, 1,980G FLOPs), limiting real-world applicability. Overall, the proposed MSCB+UAFM(+AUE) achieves an excellent efficiency–accuracy trade-off, enabling rapid and scalable optical flood mapping.

C. Discussion

From the results of the two tracks, it can be observed that the overall performance in the optical track is higher than in the SAR track. This difference can be attributed to both the scale of the datasets and the intrinsic imaging principles of the sensors. The optical dataset provides richer spectral and textural information, which facilitates learning discriminative features for flood detection. In contrast, SAR data are affected by speckle noise, geometric distortions, and complex backscattering mechanisms, including specular reflection over smooth water surfaces, volume scattering from vegetation, and double-bounce scattering in built-up or flooded areas. These factors increase the difficulty of accurate segmentation in the radar domain and partially explain the performance gap between optical and SAR-based methods.

During the 2024 IEEE GRSS Data Fusion Contest, we addressed uncertainty issues and adopted effective strategies at both the data and model levels, resulting in high-precision and low-uncertainty flood mapping. Consequently, our method achieved first place in Track 2 of the contest [48]. In future work, we aim to further investigate approaches for achieving high-precision flood extraction on a global scale.

VI. Track 2 - Second Place: Team bodang1220

A. Method

Floods are among the most impactful natural disasters, causing significant human and environmental consequences [49]–[52]. Advances in remote sensing and deep learning have enabled effective flood detection using multi-spectral imagery, particularly from Harmonized Landsat Sentinel-2, which offers valuable spatial and temporal resolution. However, challenges such as cloud interference, limited resolution, and complex surface features hinder accurate flood mapping. To address these, we propose a domain knowledge-aware framework based on the Remote Sensing Foundation Model (RSFM), as shown in Figure 11. Our method enhances image resolution, integrates spatial texture features with the Automatic Water Extraction Index (AWEI) to distinguish flood areas, and employs model ensembles and knowledge-driven post-processing to improve prediction accuracy and reliability.

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 [53], 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 the new image size, thereby establishing pixel-level image-label pairs.

Our performance in the first validation phase led us to adopt a semantic segmentation network based on an encoder-decoder architecture for training. Specifically, we utilize the Swin Transformer (SkySense pre-trained [54]) model [43] as the encoder to extract high-dimensional image features. For the decoder, we used the Mask2former [55] architecture to reconstruct the flood segmentation map. During the training process, we optimized our segmentation network using the cross-entropy loss function.

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 (using $\mathbf{Norm}(\cdot)$) to obtain the flood prediction probability P_2. By introducing an adjustable parameter α, we weigh the fusion of P_1 and P_2 to generate the fused flood segmentation map Y_{fused}:

$$P_1 = F(x_{super}^{4h\times4w\times3}), \tag{13}$$

$$AWEI = B2 + 2.5 \times B3 - 1.5 \times (B8 + B11) - 0.25 \times B12, \tag{14}$$

$$P_2 = \mathbf{Norm}(AWEI), \tag{15}$$

$$Y_{fused} = \alpha P_1 + (1-\alpha)P_2. \tag{16}$$

3) Knowledge-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

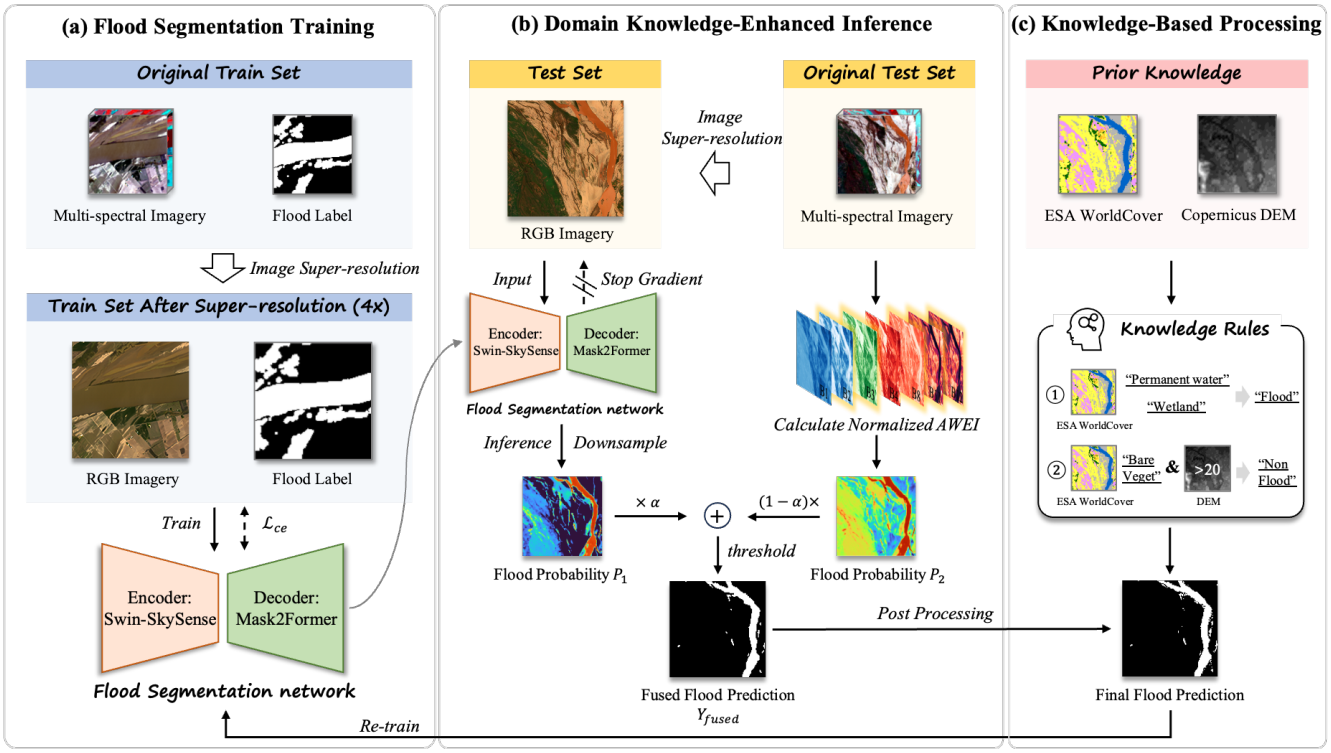


Fig. 11: The overview block diagram of the proposed domain knowledge-aware framework based on the RSFM.

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.

B. Results

1) 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 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.

2) Ablation Study: Table VI 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 becomes competitive (Exp. VIII). These experimental results demonstrate the effectiveness of the components in our approach. As shown in Fig. 12, the visualized flood detection results also verify the effectiveness of our method.

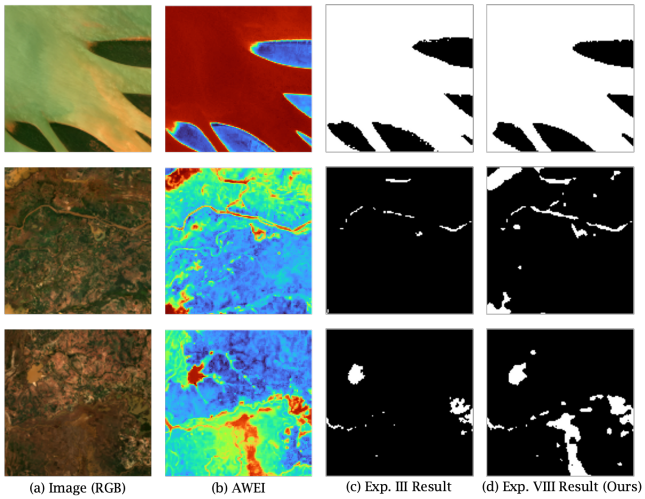


Fig. 12: Visualization of inference maps.

TABLE VI: The details of our ablation study.

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

C. Discussion

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 [56]. 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. 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.

VII. Conclusion

Rapid advancements in Earth observation sensing modalities, coupled with advances in machine learning and computer vision can deliver significant improvements for tasks such as segmentation, object detection, etc. These can play a critical role in tasks of significant societal importance, such as rapid flood mapping in flood emergency response and management. The DFC24 challenge disseminated a unique testbed dataset for rapid flood mapping that is based on SAR and passive optical imagery, and promoted research on advancing flood detection capability. The winning teams leveraged variants of convolutional semantic segmentation networks, Siamese networks, uncertainty estimators and appropriate data preprocessing to advance segmentation with the end goal of rapid flood detection. We envision the data to continue to be useful in years to come to further algorithmic advances that can effectively leverage passive optical and SAR imagery for robust and rapid flood detection.

Acknowledgements

The IADF TC chairs would like to thank the Space for Climate Observatory (SCO), the Centre national d'études spatiales (CNES), the National Aeronautics and Space Administration (NASA), and the CERFACS for providing the data and annotations, as well as the IEEE GRSS for their continued support of the annual Data Fusion Contest through funding and other resources.

References

[1] B. Bauer-Marschallinger, S. Cao, M. E. Tupas, F. Roth, C. Navacchi, T. Melzer, V. Freeman, and W. Wagner, "Satellite-based flood mapping through bayesian inference from a sentinel-1 sar datacube," Remote Sensing, vol. 14, no. 15, p. 3673, 2022.

[2] P. Tripathy and T. Malladi, "Global flood mapper: a novel google earth engine application for rapid flood mapping using sentinel-1 sar," Natural Hazards, vol. 114, no. 2, pp. 1341–1363, 2022.

[3] E. Hamidi, B. G. Peter, D. F. Muñoz, H. Moftakhari, and H. Moradkhani, "Fast flood extent monitoring with sar change detection using google earth engine," IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1–19, 2023.

[4] A. Misra, K. White, S. F. Nsutezo, W. Straka III, and J. Lavista, "Mapping global floods with 10 years of satellite radar data," Nature Communications, vol. 16, no. 1, p. 5762, 2025.

[5] C. Zhou, W. Wu, X. Ke, Y. Song, Y. He, W. Li, Y. Li, R. Jing, P. Song, L. Fu et al., "Assessment of flooding and drought disaster risk in henan, china by a multiscale approach," Geomatics, Natural Hazards and Risk, vol. 16, no. 1, p. 2491474, 2025.

[6] L. Hashemi-Beni and A. A. Gebrehiwot, "Flood extent mapping: an integrated method using deep learning and region growing using uav optical data," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 2127–2135, 2021.

[7] G. I. Drakonakis, G. Tsagkatakis, K. Fotiadou, and P. Tsakalides, "Ombrianet—supervised flood mapping via convolutional neural networks using multitemporal sentinel-1 and sentinel-2 data fusion," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 2341–2356, 2022.

[8] B. Zhao, H. Sui, J. Liu, W. Shi, W. Wang, C. Xu, and J. Wang, "Flood inundation monitoring using multi-source satellite imagery: a knowledge transfer strategy for heterogeneous image change detection," Remote Sensing of Environment, vol. 314, p. 114373, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425724003997

[9] N. Longbotham, F. Pacifici, T. Glenn, A. Zare, M. Volpi, D. Tuia, E. Christophe, J. Michel, J. Inglada, J. Chanussot et al., "Multi-modal change detection, application to the detection of flooded areas: Outcome of the 2009–2010 data fusion contest," IEEE Journal of selected topics in applied earth observations and remote sensing, vol. 5, no. 1, pp. 331–342, 2012.

[10] N. Yokoya, P. Ghamisi, J. Xia, S. Sukhanov, R. Heremans, I. Tankoyeu, B. Bechtel, B. Le Saux, G. Moser, and D. Tuia, "Open data for global multimodal land use classification: Outcome of the 2017 ieee grss data fusion contest," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 11, no. 5, pp. 1363–1377, May 2018.

[11] Y. Xu, B. Du, L. Zhang, D. Cerra, M. Pato, E. Carmona, S. Prasad, N. Yokoya, R. Hänsch, and B. Le Saux, "Advanced multi-sensor optical remote sensing for urban land use and land cover classification: Outcome of the 2018 ieee grss data fusion contest," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 12, no. 6, pp. 1709–1724, June 2019.

[12] S. Kunwar, H. Chen, M. Lin, H. Zhang, P. Dangelo, D. Cerra, S. M. Azimi, M. Brown, G. Hager, N. Yokoya, R. Hansch, and B. Le Saux, "Large-scale semantic 3d reconstruction: Outcome of the 2019 ieee grss

data fusion contest - part a," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, pp. 1–1, 2020.

[13] Y. Lian, T. Feng, J. Zhou, M. Jia, A. Li, Z. Wu, L. Jiao, M. Brown, G. Hager, N. Yokoya, R. Hansch, and B. Le Saux, "Large-scale semantic 3d reconstruction: Outcome of the 2019 ieee grss data fusion contest - part b," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, pp. 1–1, 2020.

[14] C. Robinson, K. Malkin, N. Jojic, H. Chen, R. Qin, C. Xiao, M. Schmitt, P. Ghamisi, R. Hänsch, and N. Yokoya, "Global land-cover mapping with weak supervision: Outcome of the 2020 ieee grss data fusion contest," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 3185–3199, 2021.

[15] Y. Ma, Y. Li, K. Feng, Y. Xia, Q. Huang, H. Zhang, C. Prieur, G. Licciardi, H. Malha, J. Chanussot, P. Ghamisi, R. Hänsch, and N. Yokoya, "The outcome of the 2021 ieee grss data fusion contest - track dse: Detection of settlements without electricity," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 12 375–12 385, 2021.

[16] R. Hänsch, C. Persello, G. Vivone, J. Castillo Navarro, A. Boulch, S. Lefevre, and B. Le Saux, "Report on the 2022 ieee geoscience and remote sensing society data fusion contest: Semisupervised learning," IEEE Geoscience and Remote Sensing Magazine, pp. 2–5, 2022.

[17] C. Persello, R. Hänsch, G. Vivone, K. Chen, Z. Yan, D. Tang, H. Huang, M. Schmitt, and X. Sun, "2023 ieee grss data fusion contest: Large-scale fine-grained building classification for semantic urban reconstruction [technical committees]," IEEE Geoscience and Remote Sensing Magazine, vol. 11, no. 1, pp. 94–97, 2023.

[18] "Copernicus emergency management service," 2023, https://emergency.copernicus.eu/ [Accessed: 21/12/2023].

[19] OPERA, "Opera dynamic surface water extent from harmonized landsat sentinel-2 calval database," 2023. [Online]. Available: https://podaac.jpl.nasa.gov/dataset/OPERA_DSWX-HLS_CALVAL_PROVISIONAL_V1

[20] T. H. Nguyen, S. Ricci, C. Fatras, A. Piacentini, A. Delmotte, E. Lavergne, and P. Kettig, "Improvement of Flood Extent Representation With Remote Sensing Data and Data Assimilation," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–22, 2022.

[21] T. H. Nguyen, S. Ricci, A. Piacentini, E. Simon, R. Rodriguez Suquet, and S. Peña Luque, "Gaussian Anamorphosis for Ensemble Kalman Filter Analysis of SAR-Derived Wet Surface Ratio Observations," IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1–21, 2024.

[22] R. Torres, P. Snoeij, D. Geudtner, D. Bibby, M. Davidson, E. Attema, P. Potin, B. Rommen, N. Floury, M. Brown, I. N. Traver, P. Deghaye, B. Duesmann, B. Rosich, N. Miranda, C. Bruno, M. L'Abbate, R. Croci, A. Pietropaolo, M. Huchler, and F. Rostan, "Gmes sentinel-1 mission," Remote Sensing of Environment, vol. 120, pp. 9–24, 2012, the Sentinel Missions - New Opportunities for Science. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0034425712000600

[23] "S1 tiling code tag version 1.0.0rc2," 2023, https://gitlab.orfeo-toolbox.org/s1-tiling/s1tiling/-/tree/1.0.0rc2?ref_type=tags [Accessed: 21/12/2023].

[24] "Harmonized landsat and sentinel-2 project," 2023, https://hls.gsfc.nasa.gov/ [Accessed: 21/12/2023].

[25] D. Yamazaki, D. Ikeshima, R. Tawatari, T. Yamaguchi, F. O'Loughlin, J. C. Neal, C. C. Sampson, S. Kanae, and P. D. Bates, "A high-accuracy map of global terrain elevations," Geophysical Research Letters, vol. 44, no. 11, pp. 5844–5853, 2017. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL072874

[26] "Copernicus digital elevation model," 2023, https://spacedata.copernicus.eu/collections/copernicus-digital-elevation-model [Accessed: 21/12/2023].

[27] D. Zanaga, R. Van De Kerchove, D. Daems, W. De Keersmaecker, C. Brockmann, G. Kirches, J. Wevers, O. Cartus, M. Santoro, S. Fritz, M. Lesiv, M. Herold, N. Tsendbazar, P. Xu, F. Ramoino, and O. Arino, "Esa worldcover 10 m 2021 v200," 2022.

[28] J.-F. Pekel, A. Cottam, N. Gorelick, and A. S. Belward, "High-resolution mapping of global surface water and its long-term changes," Nature, vol. 540, p. 418–422, 2016.

[29] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 5686–5696.

[30] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, 2015, pp. 234–241.

[31] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Álvarez, and P. Luo, "Segformer: Simple and efficient design for semantic segmentation with transformers," in Neural Information Processing Systems, 2021. [Online]. Available: https://api.semanticscholar.org/CorpusID:235254713

[32] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" Advances in neural information processing systems, vol. 30, 2017.

[33] J. Li, W. He, Z. Li, Y. Guo, and H. Zhang, "Overcoming the uncertainty challenges in detecting building changes from remote sensing images," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 220, pp. 1–17, 2025.

[34] W. Wang, E. Xie, X. Li, D.-P. Fan, K. Song, D. Liang, T. Lu, P. Luo, and L. Shao, "Pvt v2: Improved baselines with pyramid vision transformer," Computational Visual Media, vol. 8, no. 3, pp. 415–424, 2022.

[35] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2117–2125.

[36] J. Li, W. He, W. Cao, L. Zhang, and H. Zhang, "Uanet: An uncertainty-aware network for building extraction from remote sensing images," IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1–13, 2024.

[37] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009, pp. 248–255.

[38] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[39] H. Huang, J. Li, W. He, H. Zhang, and L. Zhang, "Overcoming the uncertainty challenges in flood rapid mapping with sar data," in IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2024, pp. 767–771.

[40] X. Shi, S. Fu, J. Chen, F. Wang, and F. Xu, "Object-level semantic segmentation on the high-resolution gaofen-3 fusar-map dataset," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 3107–3119, 2021.

[41] T. Xiao, Y. Liu, B. Zhou, Y. Jiang, and J. Sun, "Unified perceptual parsing for scene understanding," in Proceedings of the European Conference on Computer Vision (ECCV), September 2018.

[42] Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie, "A convnet for the 2020s," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[43] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 10 012–10 022.

[44] Y. Xiong, Z. Li, Y. Chen, F. Wang, X. Zhu, J. Luo, W. Wang, T. Lu, H. Li, Y. Qiao, L. Lu, J. Zhou, and J. Dai, "Efficient deformable convnets: Rethinking dynamic and sparse operator for vision applications," arXiv preprint arXiv:2401.06197, 2024.

[45] S. Fang, K. Li, and Z. Li, "Changer: Feature interaction is what you need for change detection," IEEE Transactions on Geoscience and Remote Sensing, 2023.

[46] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 801–818.

[47] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," Advances in neural information processing systems, vol. 28, 2015.

[48] J. Li, H. Huang, W. He, H. Zhang, and L. Zhang, "Overcoming the uncertainty challenges in flood rapid mapping with multi-source optical data," in IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2024, pp. 780–784.

[49] T. Tingsanchali, "Urban flood disaster management," Procedia engineering, vol. 32, pp. 25–37, 2012.

[50] Y. Li, B. Dang, Y. Zhang, and Z. 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.

[51] Y. Li, B. Dang, W. Li, and Y. 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, vol. 38, no. 20, 2024, pp. 22 213–22 221.
[52] B. Dang and Y. Li, "Msresnet: Multiscale residual network via self-supervised learning for water-body detection in remote sensing imagery," Remote Sensing, vol. 13, no. 16, p. 3122, 2021.
[53] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 2, pp. 295–307, 2015.
[54] X. Guo, J. Lao, B. Dang, Y. Zhang, L. Yu, L. Ru, L. Zhong, Z. Huang, K. Wu, D. Hu, H. He, J. Wang, J. Chen, M. Yang, Y. Zhang, and Y. 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.
[55] B. Cheng, I. Misra, A. G. Schwing, A. Kirillov, and R. 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.
[56] Y. Li, B. Dang, F. Wei, J. Tan, and Y. Lin, "Domain knowledge-aware remote sensing foundation model for flood detection in multi-spectral imagery," in IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2024, pp. 785–789.