CC-Former: Urban Flood Mapping from InSAR Coherence with Vision Transformer: Libya and Storm Daniel as Test Case

Tamer Saleh[1,2,3], Shimaa Holail[4], Mina Al-Saad[5], Fang Xu[1], Mohamed Zahran[3], Gui-Song Xia[1,2,*]

[1] School of Artificial Intelligence, Wuhan University, Wuhan, 430072, China (tamersaleh; xufang; guisong.xia)@whu.edu.cn

[2] School of Computer Science, Wuhan University, Wuhan 430072, China

[3] Geomatics Engineering Department, Faculty of Engineering at Shoubra, Benha University, Cairo 11629, Egypt (tamer.mohamed; mohamed.zahran01)@feng.bu.edu.eg

[4] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China (shimaaholail@whu.edu.cn)

[5] College of Engineering and IT, University of Dubai, Dubai 14143, United Arab Emirates (malsaad@ud.ac.ae)

Key Words: Urban Flood Mapping, Sentinel-1 SAR, InSAR Coherence, Deep Learning, Vision Transformer, Remote Sensing, Derna-Libya.

Abstract:

Urban flooding is a recurring and distressing issue with severe consequences, including the destruction of densely populated infrastructure and loss of life. Mapping inundated urban areas using synthetic aperture radar (SAR) data is crucial for local authorities to quickly assess risks and coordinate rescue efforts. However, due to the complexity of backscattering mechanisms, SAR-based urban floodwater mapping remains a challenge. In this work, we address this problem by introducing a novel algorithm, coherence-guided change transformer (CC-Former), for urban flood mapping that leverages the coherence of interferometric SAR (InSAR) with vision transformers. Specifically, CC-Former utilizes two Siamese weight-sharing encoders to extract multi-scale features from input InSAR coherence images and employs a decoder to generate final predictions. Additionally, we propose a coherence-based scaling (CoBS) module designed to focus on the acquired coherence features of urban flood classes and mitigate the imbalanced distribution of training classes. For qualitative and quantitative evaluation, the proposed CC-Former model was trained and validated using multi-temporal, dual-polarized Sentinel-1 SAR data to map the flood extent in Derna, Libya, following Tropical Storm Daniel in September 2023. Experimental results demonstrate that the proposed model outperforms state-of-the-art methods, achieving an F1 score of 89.4% and an IoU of 84.4% in both co- and cross-polarization, and an F1 score of 87.9% when integrating intensity and coherence. We conclude that the CC-Former model offers a promising solution for accurate and efficient urban flood mapping from InSAR coherence, with the potential for rapid generalization to other affected areas. As such, it can significantly aid disaster management efforts in vulnerable communities in near real-time.

1. INTRODUCTION

The world is increasingly experiencing natural disasters that affect various regions, including agricultural lands and densely populated coastal areas. Urban flooding, in particular, is a recurring and complex issue, often caused by environmental factors such as heavy rainfall, tropical storms, rising sea levels, and the increased frequency of extreme weather events (WHO, 2024; Bolan et al., 2023). In September 2023, Tropical Storm Daniel struck the eastern coast of Libya, resulting in unprecedented flooding, extensive infrastructure damage, and a tragic death toll, which has since garnered international scientific interest (ReliefWeb, 2023). The disaster was further exacerbated by the collapse of two dams near the Libyan city of Derna, causing deadly floods that devastated roads, bridges, agricultural lands, and large portions of urban areas (HRW, 2023). Urban flood extent maps are crucial for local authorities and crisis management teams to quickly assess damages and coordinate rescue efforts. Satellite Earth observation systems have proven to be time- and cost-effective resources for producing more accurate and efficient urban flood maps (Zhu et al., 2024). Among these systems, Synthetic Aperture Radar (SAR) stands out for its ability to capture images in almost all-weather conditions, day or night, a significant advantage over optical systems. SAR’s side-view geometry also provides unique backscatter mechanisms, such as surface roughness and water permeability, allowing the classification of different flood conditions, including urban, vegetative, and open areas (Amitrano et al., 2024). Figure 1 presents a comparison of pre- and post-flood images of Derna using Sentinel-1 SAR and Sentinel-2 multi-spectral images. Optical images are hindered by cloud cover, while SAR penetrates clouds effectively. Several studies (Igarashi and Wakabayashi, 2024; Lang et al., 2024; Berezowski et al., 2024; Dhanabalan et al., 2021; Saleh et al., 2024a,b) have developed algorithms for flood mapping using SAR intensity data (σ_0), specifically through the analysis of double-backscatter signals (when radar waves reflect between buildings and the ground). Saleh et al. (2024c) employed the multi-temporal S1GFloods dataset, captured by the Sentinel-1 sensor, using a change detection approach to distinguish between standing water and floodwater, while successfully minimizing false detections caused by shadowed areas. Despite these advances, many algorithms face limitations, as building facades often do not align perpendicularly to the satellite’s line of sight, reducing the accuracy of urban floodwater detection. To address this challenge, studies (Lan et al., 2024; Li et al., 2019; Sonobe and Hashiba, 2024; Pulvirenti et al., 2015; Selmi et al., 2014) proposed improving flood mapping by reducing event-concurrent InSAR coherence (ρ) relative to pre-event coherence. For example, Figure 1 shows the intensity and coherence variations for the urban area of Derna under pre-

* Corresponding author

https://doi.org/10.5194/isprs-annals-X-G-2025-745-2025

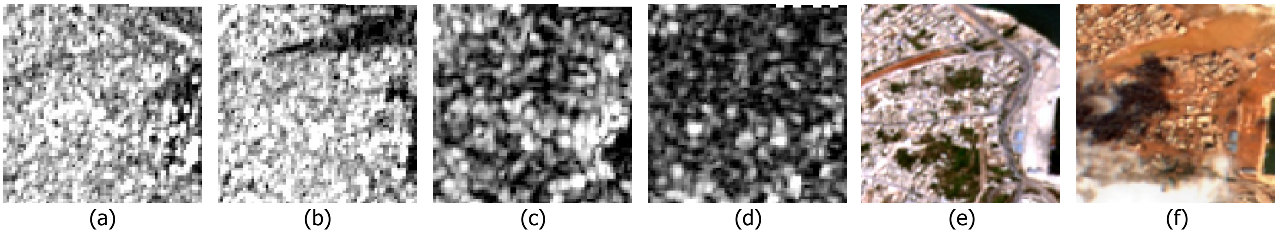

Figure 1. Differences in intensity and coherence from Sentinel-1 data. (a-f) represent, respectively, pre-flood intensity, post-flood intensity, pre-flood coherence, post-flood coherence, pre-flood Sentinel-2 optical data, and post-flood Sentinel-2 cloud cover during the event. This data is copyrighted by Copernicus Sentinel-1© ESA, CC BY-SA 3.0.

and post-flood conditions in Sentinel-1 data. The intensity of the double-bounce backscatter either remains the same or decreases slightly, while the coherence decreases significantly in the case of urban flooding. Considering this, mapping the extent of floods in complex urban areas based on SAR intensity alone is a challenging task. Recent deep-learning studies have shown remarkable progress in general flood detection (Huang et al., 2024; Shahi et al., 2024). For example, Chamatidis et al. (2024) presented a flood detection method that integrates transfer learning with vision transformers pre-trained on Sentinel-1 SAR images. Saleh et al. (2023, 2024c) introduced DAM-Net and PDCA-Former for flood detection from single-polarization SAR intensity images using vision transformers. Jamali et al. (2024) developed WVResU-Net, a model for flood mapping using dual-polarization Sentinel-1 SAR data, based on a vision multi-layer perceptron and ResU-Net. Although these studies have tackled extreme flood issues, to the best of our knowledge, a systematic and quantitative study combining SAR intensity and coherence from dual-polarization data within a foundational model framework for flood mapping in complex urban settings remains absent in the literature. In this study, we evaluate and map the extent of urban flooding in the Libyan city of Derna by integrating SAR intensity and InSAR coherence, applying change detection (CD) within a vision transformer framework. This approach enhances temporal differences in intensity images and identifies spatial pattern changes in coherence images. We used a series of multi-temporal, dual-polarized Sentinel-1 SAR Single Look Complex (SLC) datasets to generate coherence pairs pre- and post-Storm Daniel, which struck the eastern coast of Libya in September 2023. Additionally, high-resolution optical satellite imagery from Maxar's open data program, captured during the event, was used as a reference to assess damage to urban infrastructure, such as buildings and roads. Our proposed framework significantly improves flood extent estimation compared to state-of-the-art methods, providing valuable data essential for reconstruction and regional economic recovery efforts.

The main contributions of this paper are twofold. First, we propose a novel framework, called CC-Former, that integrates SAR intensity and InSAR coherence data to enhance the accuracy of urban flood inundation mapping. A key advantage of this framework is its applicability to vision transformers training. Second, we introduce a new coherence-based scaling (CoBS) module, which focuses on the coherence features of urban flood classes by deriving a coherence mask from InSAR data and mitigating the unbalanced distribution of training classes. To achieve the objectives of this paper, the rest of the paper is structured as follows: Section 2 introduces the proposed CC-Former method and a novel coherence-based scale module. Section 3 describes the study area, dataset, pre-processing, and experimental results. Finally, Section 4 provides conclusions, discusses limitations, and suggests future research directions.

2. PROPOSED METHODOLOGY

2.1 Intensity and coherence

Due to SAR's side-looking geometry, double-bounce scattering, and varying backscatter properties, flooded urban areas can exhibit different appearances across polarization modes (Xiang et al., 2016). For instance, urban floodwaters present different scattering characteristics, with significantly lower (σ_0) in cross-polarization (VH) and higher (σ_0) in co-polarization (VV) when comparing flooded to non-flooded surfaces. Additionally, stagnant floodwaters in built-up areas often manifest as bright linear features due to double-bounce scattering. Interferometric coherence provides valuable information for urban flood mapping, as urban areas typically act as stable targets with high coherence. Conversely, a reduction in InSAR coherence suggests the presence of floodwaters in these regions. We hypothesize that using SAR intensity data alone may underestimate the extent of urban flooding. To address this, we propose a novel method that enhances flood detection by utilizing multi-temporal InSAR coherence derived from both copolarization and cross-polarization channels. First, we generate a building extraction map from multi-temporal SAR images, following the method outlined in (Verma et al., 2023). Next, we identify buildings near floodwaters and compute the InSAR coherence change between two complex images (phase and amplitude) using a square moving window, as described in (1). A 5 $\times$ 5 window size was found to be optimal for urban areas.

$$\gamma = \frac{E(I_1 \cdot I_2^*)}{\sqrt{E(|I_1|^2)E(|I_2^*|^2)}} \tag{1}$$

where I_1 and I_2 are the SAR images forming the InSAR pair, the asterisk denotes complex conjugation, and $E(\cdot)$ represents the expectation value.

In detail, we utilize two pre-flood InSAR coherence images (t_1 and t_2) to calculate (ρ_{pre}), which has values close to zero, as temporal coherence in built-up areas is typically stable, indicating non-flooded conditions. We then compare a pre-flood image (t_1) with a flood image (t_0) to derive (ρ_{co}), where a drop in (ρ_{co}) (relative to ρ_{pre}) signifies flooding. When ($\rho_{pre} > \rho_{co}$), the area is considered affected by urban flooding due to changes in scatterer distribution within the resolution cell. In this study, we map the decrease in InSAR coherence from both VV and VH channels and combine them to enhance urban flood detection. Figure 2 outlines the main pre-processing steps applied to the SLC datasets using the python-snappy[1]. While histogram

[1] https://github.com/puzhao8/snappy_InSAR

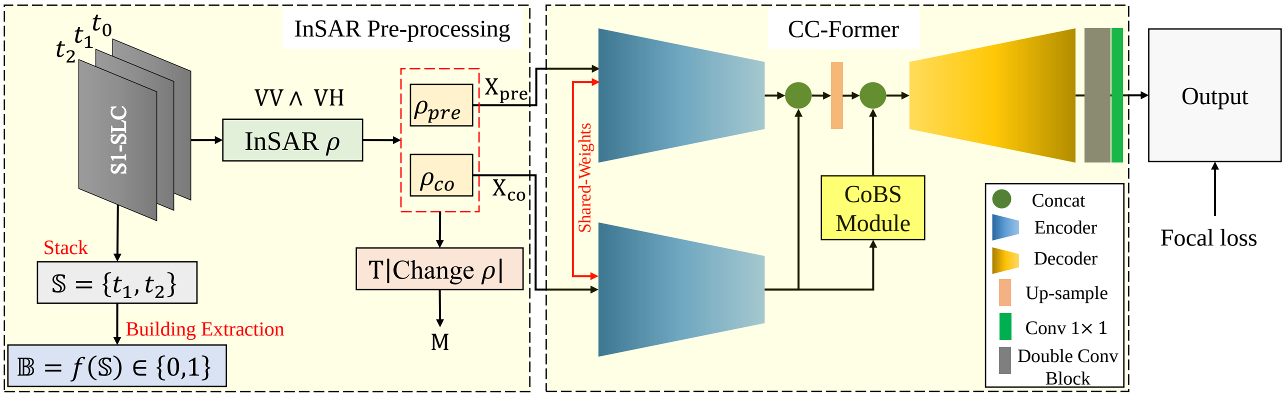

Figure 2. Proposed Coherence Change-Aware Vision Transformer Architecture (CC-Former), featuring the CoBS block, which serves as the Coherence-Based Scaling module.

thresholding is typically employed to distinguish water from non-water areas, its accuracy depends on class overlap within the image. To improve separation accuracy, we adopt a vision transformer-based approach, as described in Section 2.2.

2.2 Overview of the CC-Former

The proposed CC-Former algorithm for urban flood detection is illustrated in Figure 2. This methodology is based on vision transformers, drawing inspiration from (Saleh et al., 2024c). CC-Former is a Siamese network that integrates changes in multi-temporal coherence information from radar images with a coherence-based scale (CoBS) module. It comprises three main components: the encoder, the CoBS module, and the decoder. The encoder utilizes a Siamese weight-sharing configuration, based on the ViTAEv2 (Zhang et al., 2022) architecture as its core structure, to extract robust multi-scale features from bi-temporal SAR coherence images, corresponding to pre-flood and co-flood scenarios. The pre-flood and co-flood features are concatenated, allowing the model to explore the change relationships between them. These features are further refined using the CoBS module, as described in Section 2.3. In the decoder, high-level semantic features extracted by the encoder branches are compared with high-level change features and fused with the enhanced features from the CoBS module through concatenation. Linear interpolation is applied to upsample the high-level features from the encoder and retrieve detailed change features. The decoder consists of two iterations of 3×3 convolutional layers, a batch normalization layer, and a ReLU activation function. Urban flooding is detected by applying an additional convolutional layer. The focal loss function is employed to handle class imbalance, as shown in (2), where the focal factor γ is set to 2, and the weighting factor α is set to 0.3.

$$FL(p) = -\alpha(1-p)^{\gamma}\log(p) \tag{2}$$

where p is the probability of the class.

2.3 CoBS module

In this study, we distinguish between two classes of the urban mask urban flood and non-flood derived from SAR data, as described in Section 2.4. However, the training dataset may exhibit an imbalanced distribution between these two classes, potentially due to limited diversity or size. This imbalance can result in sub-optimal utilization of the urban mask, leading to less accurate predictions. To address this issue, we propose a coherence-based scaling (CoBS) module designed to focus on the learned coherence features of urban flood classes, thereby enhancing the model's generalization capability. The module consists of two components applied sequentially: a channel attention sub-module and a coherence-aware scale sub-module, as illustrated in Figure 3. This module is embedded within a vision transformer network architecture to aid the learning process, ultimately improving the accuracy of urban flood mapping.

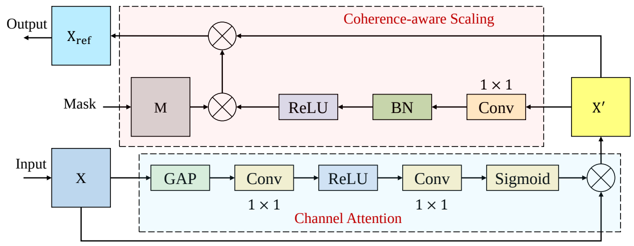

Figure 3. Design of the Coherence-Based Scaling Module.

2.3.1 *Channel attention sub-module:*

Given a multi-layer feature tensor $X \in \mathbb{R}^{W\times H\times C}$ representing the feature map, different attention weights are applied to emphasize various channels. Here, W, H, and C correspond to the width, height, and number of channels, respectively. The channel attention mechanism captures the relationships between feature channels by using a learnable network that assigns weights to each channel based on their significance, generating more informative outputs. Specifically, global average pooling (GAP) is used to compress and pool the spatial dimensions of the input feature map. Following this, two 1×1 convolutional layers, with a rectified linear unit (ReLU) between them, are applied. The Sigmoid function is then employed to calculate the channel attention score (S_{Ac}), with values ranging from 0 to 1. Finally, the refined feature map X' is obtained by multiplying S_{Ac} with the input feature tensor X. This process can be expressed mathematically as:

$$X' = \text{Sigmoid}(\text{Conv}(\text{ReLU}(\text{Conv}(\text{GAP}(X))))) \cdot X \tag{3}$$

2.3.2 *Coherence-aware scale sub-module:*

This submodule incorporates a scale factor (W_s) derived from the optimized intermediate feature X' and the urban flood coherence mask M, as described in Section 2.4. Given the intermediate feature tensor $X' \in \mathbb{R}^{W\times H\times C}$, a 1×1 convolutional layer is applied, followed by batch normalization (BN) and ReLU activation. The resulting output is then multiplied by the M to generate W_s. Finally, the refined feature (X_{refined}) is obtained by multiplying W_s with the optimized intermediate feature X', as shown in (4):

$$X_{\text{refined}} = (\text{ReLU}(\text{BN}(\text{Conv}(X'))) \cdot M) \cdot X' \tag{4}$$

2.4 Urban flood coherence mask

Due to the limited distinct features of urban floodwater classes, we leveraged the double-bounce scattering effect and generated an urban mask from multi-temporal SAR images in both VV and VH polarizations. This was achieved by applying a histogram threshold based on the average intensity backscatter to classify pixels as either flooded or non-flooded. Additionally, the low multi-temporal InSAR coherence in both VV and VH polarizations plays a critical role in the generation of the urban floodwater mask. By combining the urban mask derived from SAR intensity with the InSAR coherence, we aim to minimize false alarms in flood detection, particularly in vegetated areas. The urban mask also provides valuable information about potential construction sites and the geometry of built-up areas. In this study, the urban flood mask, denoted as (M), is used as an input feature for the CoBS module, which is described in Section 2.3.

2.5 Evaluation metrics

To compare the ground truth and predicted change map, five metrics were utilized to validate the accuracy and effectiveness of the proposed method. These metrics include Overall Accuracy (OA), Precision (P), Recall (R), F1-score, and IoU. The formulas for these metrics are provided below:

$$\text{OA} = \frac{TP + TN}{TP + FN + TN + FP} \tag{5}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

$$\text{F1-score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{8}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{9}$$

where TP = total number of true positives for all classes
FN = false negatives
FP = false positives
TN = true negatives

3. EXPERIMENTAL RESULTS

3.1 Study area

The eastern coast of Libya is a key hub for energy exports to Europe. The study area, located in Derna, Libya, lies between latitudes 32.7 to 32.8 degrees N and longitudes 22.6 to 22.7 degrees E, along the northeastern coast of Libya on the Mediterranean Sea was selected. In September 2023, Storm Daniel hit the region, bringing 120 km/h winds and 240 mm of rainfall over 25 hours, causing severe flooding in Derna and other coastal cities. The storm impacted roads, bridges, residential and educational buildings, as well as some industrial and commercial areas (OCHA, 2023a,b). Repairing the coastal damage will cost an estimated US$1.8 billion (ReliefWeb, 2023). Figure 4 shows the Area of Interest (AOI) and the locations of collapsed bridges and dams. Assessing the floodwater's extent is critical for effective management and preventing further humanitarian disasters in this unstable region.

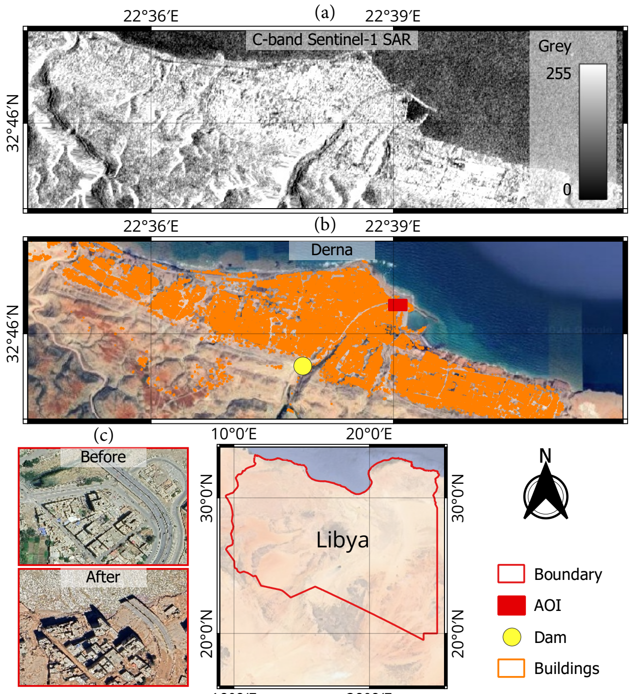

Figure 4. Location map of the study area in Derna, Libya, depicted through multiple views: (a) SAR imagery of the area; (b) buildings overlaid on a Google Earth satellite image as the background; (c) Maxar VHR imagery illustrating the Derna Bridge before and after the flooding. The study area includes a densely populated coastal region and the collapsed Derna Dam, highlighted by the yellow circle.

3.2 Datasets

In this study, we acquired Sentinel-1 SAR Interferometric Wide (IW) data, including ground range detected (GRD) intensity and single-look complex (SLC) image pairs, from the Alaska Satellite Facility (ASF)[2]. The SLC data were employed to detect low coherence in urban areas during the period from July 15 to September 13, 2023. Specifically, Sentinel-1 intensity and coherence (VV-VH polarizations), captured both before and during the flooding event, were utilized. Additionally, high-resolution (0.3 m) optical data from September 13, 2023, sourced through the Maxar open data program[3], was used to manually annotate flood-impacted buildings and roads. These annotations served as ground truth for evaluating our experimental results. The detailed characteristics of the dataset used are summarized in Table 1.

3.3 Pre-processing

The following pre-processing steps were applied to each acquired Sentinel-1 image: TOPS splitting, back-geocoding,

[2] https://search.asf.alaska.edu/
[3] https://xpress.maxar.com/

	Acquisition date	Polarization	Resolution (m) (range × azimuth)	Orbit	Pass	Characteristics
Pre-event	15-Jul-23	VV+VH	10.0 × 10.0	49429	Descending	2 GRD image
	8-Aug-23	VV+VH	10.0 × 10.0	49779		
Co-event	13-Sep-23	VV+VH	10.0 × 10.0	50304	Descending	1 GRD image
Pre-event	15-Jul-23	VV+VH	2.33 × 13.9	49429	Descending	2 SLC images
	8-Aug-23	VV+VH	2.33 × 13.9	49779		
Co-event	13-Sep-23	VV+VH	2.33 × 13.9	50304	Descending	1 SLC image
Co-event	13-Sep-23	RGB	0.30 × 0.30	-	-	1 Maxar image

Table 1. Characteristics of Sentinel-1 GRD and SLC images, as well as the Maxar optical image, used as input datasets in this study.

enhanced spectral diversity (ESD), interferogram formation, Goldstein phase filtering, and range-Doppler terrain correction. For SAR intensity, the Sentinel-1 data were normalized and converted into backscatter coefficients (dB), with a 5 × 5 pixel Lee Sigma filter applied to reduce speckle noise. In addition to SAR intensity, interferometric coherence (ρ) was calculated using SLC image pairs from before and during the flood event, denoted as ρ_{pre} and ρ_{co}, respectively, using a 7 × 7-pixel moving window. All intensity and coherence data were stacked and geocoded to the WGS 1984 UTM zone 34N at a pixel resolution of 10 m, producing images of 8,192 × 10,240 pixels. Using interferometric coherence (ρ) from the pre-event and co-event, coherence change (CC) was generated through a logarithmic ratio to identify regions of change. Flooded pixels were assigned a value of 255, while non-flooded pixels were assigned 0. The pre-processed images were then divided into non-overlapping patches of 512 × 512 pixels. Figure 5 presents the phase images and labels for the changes in SAR data, where red areas indicate flooded regions and cyan areas denote non-flooded regions in the label image. The dataset, consisting of 172 samples obtained after patchifying, was further divided into training, validation, and test sets in a 70:20:10 ratio for training and validating deep learning models.

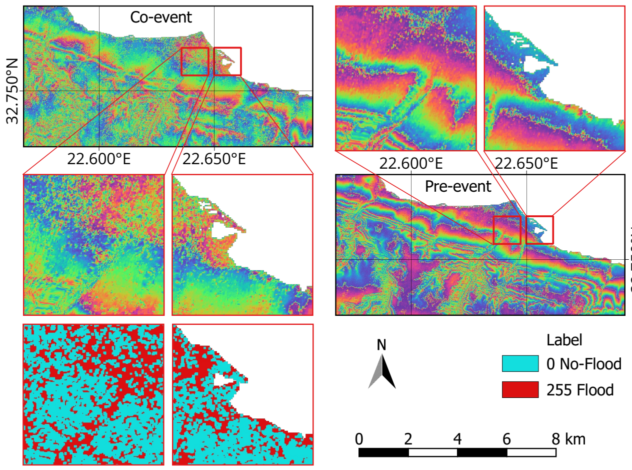

Figure 5. Phase images before the event ($\rho_{pre}^{15/07-08/08/2023}$), during the event ($\rho_{co}^{08/08-13/09/2023}$), and labels showing changes in coherence, with flooded regions in red and non-flooded regions in cyan.

3.4 Implementation details

During training, data augmentation was applied to the input image patches due to the small size of the Derna dataset. This augmentation included random flipping, rotations of 90, 180, and 270 degrees, histogram matching, Gaussian blurring, and random clipping of 128 × 128-pixel patches, with the number of augmented samples generated automatically during model training. The input image size was fixed at 512 × 512 for both the U-Net and CC-Former methods and at 448 × 448 for the PDCA-Former. The CC-Former and other competing methods were trained with a batch size of eight using the Adam optimizer with an initial learning rate of 1×10^{-6} and momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, determined by trial and error. Models were trained for 100 epochs, with weights saved every five epochs, and the best model (highest F1-score) was selected for inference on different test sites. Experiments were run on a virtual machine desktop with a 64-bit Windows 10 Pro OS. The software setup included Python with PyTorch[4] implementation, and the hardware setup featured an NVIDIA GRID RTX8000-8Q GPU (8GB memory), an Intel Xeon CPU E5-2687W v4 @ 3.00GHz, and 28GB of GPU memory for optimal performance.

3.5 Results and analysis

In this study, we trained three deep-learning models—U-Net (Fang et al., 2021), PDCA-Former (Saleh et al., 2023), and our proposed CC-Former—using the same training datasets and strategy for urban flood detection. We then qualitatively and quantitatively evaluated the resulting urban flood maps for the Derna flood event in Libya. The models classified coherence and intensity images for the VV polarization. A sample of the results is presented in Figure 6, where SAR intensity is shown as an RGB composite (pre-flood = R, post-flood = B = G). Flood-induced changes are visible in built-up areas (red and cyan), as floodwaters reduced the backscatter. However, SAR intensity alone has limitations in detecting flooded buildings. To address this, we leveraged InSAR coherence, which typically provides high values due to the temporal stability of buildings. Figure 6(a) displays InSAR coherence over the same area (ρ_{pre} = R, ρ_{co} = B = G), where a noticeable decrease in coherence during the flooding event is represented by red tones across the scene. The results from the U-Net, PDCA-Former, and CC-Former models, which utilize coherence data, are shown in Figure 6(c-e). Additionally, we classified the damage to flooded buildings into three categories: high, moderate, and low. As illustrated in the region of interest (white box), false alarms in urban areas show considerable variation in damage classification across different models. This variation is largely due to the low resolution of InSAR coherence data, which limits the model's ability to differentiate between damage levels. This is-

[4] https://pytorch.org/

sue could be mitigated in the future by using higher-resolution InSAR data, such as TerraSAR-X. Figure 6(f) shows the flood classification from our proposed method using Maxar VHR imagery, where the high resolution allows for the identification of damaged houses flooded buildings and road polygons from UNOSAT[5]. To quantitatively assess the flood maps, we calculated that 713,511 pixels were detected as flooded, while 42,471 pixels were classified as such by our model. We observed a 16.8% increase in the damaged area, which is significant given the vulnerability of urban regions to flooding.

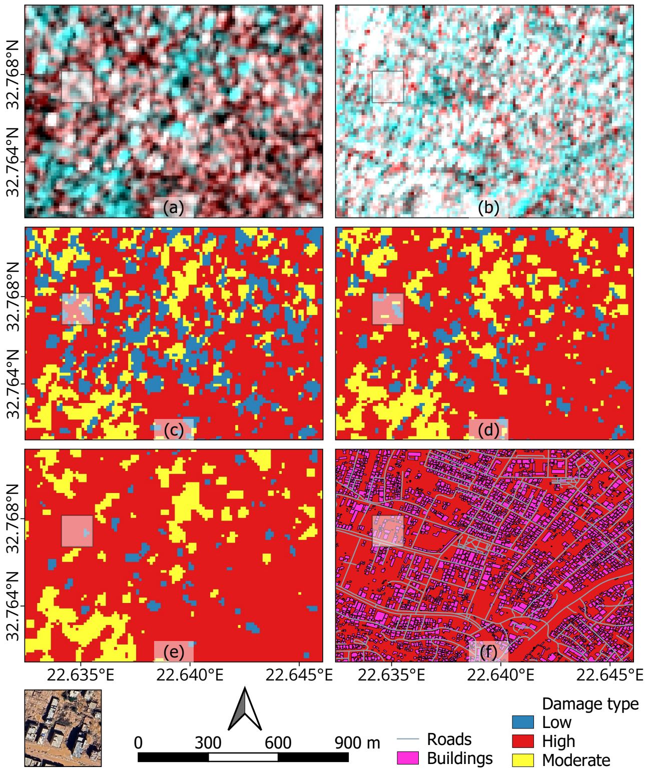

Figure 6. Classification results of various methods for the Derna study area. (a) Coherence RGB composite (R = ρ_{pre}, B = G = ρ_{co}); (b) Intensity RGB composite (R = August 8, 2023, B = G = September 13, 2023); (c-e) Model outputs from the U-Net, PDCA-Former, and CC-Former methods, respectively; (f) Flood reference mask derived from VHR optical imagery provided by Maxar1, along with flooded building polygons from UNOSAT.

Data	OA	R	P	IoU	F1
Intensity	97.1	83.1	89.4	81.4	86.1
Coherence	97.8	85.5	87.9	82.7	86.7
Intensity+Coherence	98.4	84.4	91.7	**84.1**	**87.9**

Table 2. Quantitative evaluation using the proposed method. All results are reported as percentages.

Table 3 summarizes the quantitative evaluation results of the CC-Former, PDCA-Former, and U-Net methods for VV, VH, and VV+VH polarizations. The proposed CC-Former achieves an overall F1-score improvement of 6.6% (from 82.8% to 89.4%), demonstrating greater effectiveness than the other two algorithms, which scored 86.7% and 81.1%, respectively, in the VV+VH scenario. Notably, precision for VV improved from 87.9% to 90.7%, though this was accompanied by a slight decrease in recall from 78.3% to 85.6%. In contrast, both precision and recall increased for VV+VH, as shown in the precision-recall curve (Figure 8). Significant increases in the IoU score were observed for VV (from 79.5% to 83.2%), VH (from 77.6% to 83.5%), and VV+VH (from 78.8% to 84.4%). Moreover, the integration of both intensity and coherence data during the event using the proposed method yields a high F1-score of 87.9%, as presented in Table 2. In this study, we use different combinations of red, green, and blue to enhance the understanding of the theoretical analysis, similar to the role of the normalized difference vegetation index (NDVI) in optical data. Figure 7 shows the contribution of the VV and VH polarization channels to urban flood detection, as well as the behavior of multi-temporal SAR intensities for these channels. It also highlights the significant coherence decreases corresponding to heavily flooded buildings. The widespread reddish tones in the coherence maps, compared to the intensity maps, reflect the effectiveness of the proposed model in detecting urban floods through increased coherence. White indicates high-density buildings resulting from the double-bounce effect, which are not flooded, while cyan suggests that some buildings may be flooded but are challenging to detect through intensity differences in multi-temporal images, as shown in Figure 7(a, d). This improvement highlights the importance of utilizing both polarization channels and integrating intensity and coherence information for SAR-based urban flood mapping across different land cover types.

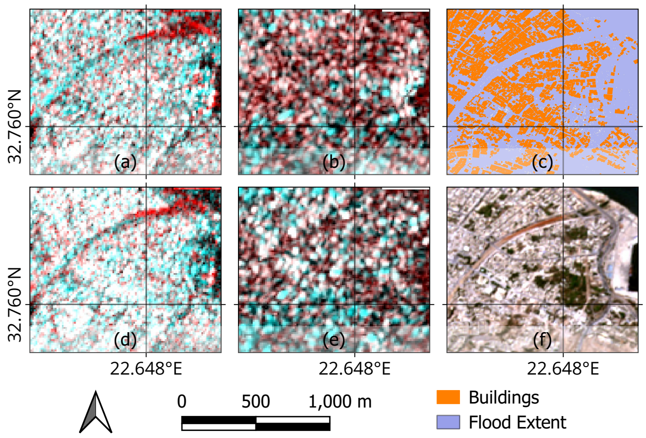

Figure 7. (a, d) Intensity RGB composite (R = August 8, 2023; B = G = September 13, 2023) for VV and VH polarizations, respectively. (b, e) Coherence RGB composite (R = ρ_{pre}; B = G = ρ_{co}) for VV and VH polarizations. (c) Urban flood extent detected by our model. (f) Pre-event Sentinel-2 optical satellite imagery.

To establish the ground truth for each coherence slice, we calculate the proportion of the area occupied by flooded buildings within each slice. This involves computing precision (the ratio of slices classified as flooded that are indeed flooded) and recall (the ratio of actual flooded slices that are classified as flooded) for each method. Given the unbalanced nature of our dataset, where non-flooded areas significantly outnumber flooded areas, we prefer the precision-recall curve over the receiver operating characteristic (ROC) curve for classifier evaluation. In this study, we quantitatively compare various classification methods by calculating the area under the precision-recall curve (PR-AUC), as depicted in Figure 8. The proposed CC-Former method demonstrates a notable improvement, achieving a PR-AUC of 82.0%, indicating that over 80% of the regions classi-

[5] `https://unosat.org/products/3670`

Scenarios	Method	OA	R	P	IoU	F1-score±σ
VV	Unet (Fang et al., 2021)	95.9	78.3	87.9	79.5	82.8±3.6
	PDCA-Former (Saleh et al., 2023)	96.2	79.6	<u>92.3</u>	81.7	85.5±3.1
	Ours	98.4	**85.6**	90.7	83.2	<u>88.1±2.2</u>
VH	Unet (Fang et al., 2021)	95.7	76.7	84.2	77.6	80.3±3.3
	PDCA-Former (Saleh et al., 2023)	96.6	80.1	88.8	81.4	84.2±2.7
	Ours	**98.8**	85.1	90.3	<u>83.5</u>	87.6±2.1
VV+VH	Unet (Fang et al., 2021)	95.4	77.8	84.7	78.8	81.1±3.2
	PDCA-Former (Saleh et al., 2023)	97.2	84.4	89.1	82.2	86.7±2.9
	Ours	<u>98.6</u>	<u>85.5</u>	**93.7**	**84.4**	**89.4±1.7**

Table 3. The average quantitative results of flood extent maps obtained by different methods on the test set in the urban area. The best results are highlighted in bold font, and the second-best results are underlined. σ represents the standard deviation associated with the quantitative results. All values are reported as percentages (%).

fied as flooded are truly flooded. Furthermore, the CC-Former exhibits a recall of 85.5%, compared to 77.8% for the U-Net method, highlighting a significant quantitative advancement.

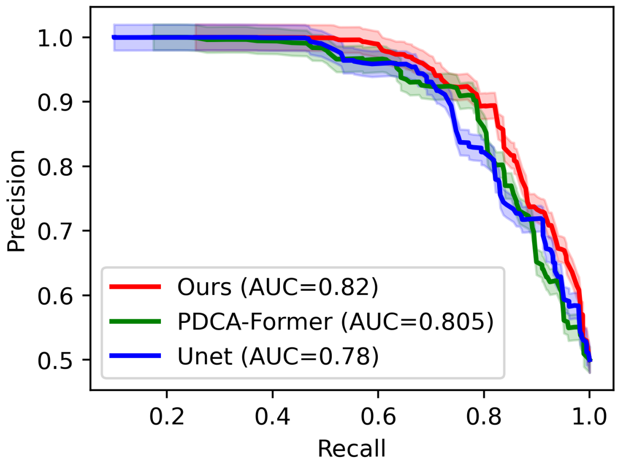

Figure 8. Recall-precision curves for flood mapping using the CC-Former (red line), PDCA-Former (green line), and U-Net (blue line) methods. A larger area under the curve (AUC) for the CC-Former indicates improved performance.

4. CONCLUSIONS

This paper presents an automated vision transformer-based algorithm, CC-Former, for urban flood mapping using dual-temporal Sentinel-1 SAR intensity data, as well as InSAR coherence from both VV and VH polarizations. A novel coherence-based scaling (CoBS) module is introduced, which focuses on the coherence features of urban flood classes by deriving a coherence mask from InSAR data to improve the detection accuracy of urban inundation maps. The performance of CC-Former was evaluated using data captured during Tropical Storm Daniel in September 2023, which caused extensive flooding in the Libyan city of Derna. The results indicate that leveraging both VV and VH polarizations of InSAR coherence increases the ability to detect floodwater around buildings by more than 2.7% compared to using SAR intensity alone. Additionally, the proposed method was compared with other recent techniques, demonstrating superior performance in flood detection tasks. However, some limitations persist, as the CoBS module relies on a coherence mask derived from low-resolution SAR coherence data, leading to over-detection of water around buildings. Further investigation is required with the availability of higher-resolution urban data, such as TerraSAR-X, to mitigate this issue.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (NSFC). Many thanks to Maxar Technologies Inc. for providing satellite imagery.

References

Amitrano, D., Di Martino, G., Di Simone, A., Imperatore, P., 2024. Flood detection with SAR: A review of techniques and datasets. *Remote Sensing*, 16(4), 656.

Berezowski, T., Niemiec, S., Chybicki, A., 2024. Floodsar: Automatic mapping of river flooding extent from multitemporal SAR imagery. *SoftwareX*, 26, 101717.

Bolan, S., Padhye, L. P., Jasemizad, T., Govarthanan, M., Karmegam, N., Wijesekara, H., Amarasiri, D., Hou, D., Zhou, P., Biswal, B. K. et al., 2023. Impacts of climate change on the fate of contaminants through extreme weather events. *Science of The Total Environment*, 168388.

Chamatidis, I., Istrati, D., Lagaros, N. D., 2024. Vision Transformer for Flood Detection Using Satellite Images from Sentinel-1 and Sentinel-2. *Water*, 16(12), 1670.

Dhanabalan, S., Abdul Rahaman, S., Jegankumar, R., 2021. Flood monitoring using Sentinel-1 sar data: A case study based on an event of 2018 and 2019 Southern part of Kerala. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 44, 37–41.

Fang, S., Li, K., Shao, J., Li, Z., 2021. SNUNet-CD: A densely connected siamese network for change detection of VHR images. *IEEE Geoscience and Remote Sensing Letters*, 19, 1–5.

HRW, 2023. Libya: Slow flood recovery failing displaced survivors. `https://www.hrw.org/news/2024/09/10/libya-slow-flood-recovery-failing-displaced-survivors`. Accessed: 2024-09-13.

Huang, B., Li, P., Lu, H., Yin, J., Li, Z., Wang, H., 2024. WaterDetectionNet: A New Deep Learning Method for Flood Mapping with SAR Image Convolutional Neural Network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Igarashi, T., Wakabayashi, H., 2024. Detection of flooded areas caused by Typhoon Hagibis by applying a learning-based method using Sentinel-1 data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Jamali, A., Roy, S. K., Beni, L. H., Pradhan, B., Li, J., Ghamisi, P., 2024. Residual wave vision U-Net for flood mapping using dual polarization Sentinel-1 SAR imagery. *International Journal of Applied Earth Observation and Geoinformation*, 127, 103662.

Lan, Q., Dong, J., Lai, S., Wang, N., Zhang, L., Liao, M., 2024. Flood Inundation Extraction and its Impact on Ground Subsidence using Sentinel-1 Data: A Case Study of the "7.20" Rainstorm Event in Henan Province, China. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Lang, F., Zhu, Y., Zhao, J., Hu, X., Shi, H., Zheng, N., Zha, J., 2024. Flood Mapping of Synthetic Aperture Radar (SAR) Imagery Based on Semi-Automatic Thresholding and Change Detection. *Remote Sensing*, 16(15), 2763.

Li, Y., Martinis, S., Wieland, M., 2019. Urban flood mapping with an active self-learning convolutional neural network based on TerraSAR-X intensity and interferometric coherence. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 178–191.

OCHA, 2023a. Libya: Impact of Storm Daniel in eastern Libya and the collapse of dams in Derna - Briefing note 13 September 2023. `https://reliefweb.int/report/libya/libya-impact-storm-daniel-eastern-libya-and-collapse-dams-derna-briefing-note-13-september-2023`. Accessed: 2024-09-15.

OCHA, 2023b. Tropical Storm Daniel - Sep 2023. `https://reliefweb.int/disaster/fl-2023-000168-lby`. Accessed: 2024-09-15.

Pulvirenti, L., Chini, M., Pierdicca, N., Boni, G., 2015. Use of SAR data for detecting floodwater in urban and agricultural areas: The role of the interferometric coherence. *IEEE Transactions on Geoscience and Remote Sensing*, 54(3), 1532–1544.

ReliefWeb, 2023. Libya: Storm and flooding 2023 - rapid damage and needs assessment. `https://reliefweb.int/report/libya/libya-storm-and-flooding-2023-rapid-damage-and-needs-assessment-enar`. Accessed: 2024-09-13.

Saleh, T., Holail, S., Xiao, X., Xia, G.-S., 2024a. High-precision flood detection and mapping via multi-temporal SAR change analysis with semantic token-based transformer. *International Journal of Applied Earth Observation and Geoinformation*, 131, 103991.

Saleh, T., Holail, S., Zahran, M., Xiao, X., Xia, G.-S., 2024b. LiST-Net: Enhanced Flood Mapping With Lightweight SAR Transformer Network and Dimension-Wise Attention. *IEEE Transactions on Geoscience and Remote Sensing*, 62, 1-17.

Saleh, T., Weng, X., Holail, S., Hao, C., Xia, G.-S., 2024c. DAM-Net: Flood detection from SAR imagery using differential attention metric-based vision transformers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 212, 440–453.

Saleh, T., Zahran, M., Holail, S., Xia, G.-S., 2023. PDCA-former: Prior-diagonal cross attention-guided transformer for flood mapping from SAR imagery: A case in Khartoum. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10, 723–730.

Selmi, S., Ben Abdallah, W., Abdelfatteh, R., 2014. Flood mapping using InSAR coherence map. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40, 161–164.

Shahi, K. R., Camero, A., Eudaric, J., Kreibich, H., 2024. DC4Flood: A deep clustering framework for rapid flood detection using Sentinel-1 SAR imagery. *IEEE Geoscience and Remote Sensing Letters*.

Sonobe, M., Hashiba, H., 2024. Evaluation of flooded buildings by multi-temporal interferometric coherence using l-band sar satellite data. *IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, 3943–3946.

Verma, A., Bhattacharya, A., Dey, S., López-Martínez, C., Gamba, P., 2023. Built-up area mapping using Sentinel-1 SAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 203, 55–70.

WHO, 2024. Floods. `https://www.who.int/health-topics/floods#tab=tab_1`. Accessed: 2024-09-12.

Xiang, D., Tang, T., Ban, Y., Su, Y., Kuang, G., 2016. Unsupervised polarimetric SAR urban area classification based on model-based decomposition with cross scattering. *ISPRS Journal of Photogrammetry and Remote Sensing*, 116, 86–100.

Zhang, Q., Xu, Y., Zhang, J., Tao, D., 2022. Vitaev2: Vision transformer advanced by exploring inductive bias for image recognition and beyond. *arXiv preprint arXiv:2202.10108*.

Zhu, X., Guo, H., Huang, J. J., 2024. Urban flood susceptibility mapping using remote sensing, social sensing and an ensemble machine learning model. *Sustainable Cities and Society*, 108, 105508.