UrbanSARFloods: Sentinel-1 SLC-Based Benchmark Dataset for Urban and Open-Area Flood Mapping

Jie Zhao, Zhitong Xiong, Xiao Xiang Zhu

Department of Aerospace and Geodesy, Data Science in Earth Observation,
Technical University of Munich, Arcisstraße 21, 80333 Munich, Germany

{jie.zhao, zhitong.xiong,xiaoxiang.zhu}@tum.de

Abstract

*Due to its cloud-penetrating capability and independence from solar illumination, satellite Synthetic Aperture Radar (SAR) is the preferred data source for large-scale flood mapping, providing global coverage and including various land cover classes. However, most studies on large-scale SAR-derived flood mapping using deep learning algorithms have primarily focused on flooded open areas, utilizing available open-access datasets (e.g., Sen1Floods11) and with limited attention to urban floods. To address this gap, we introduce **UrbanSARFloods**, a floodwater dataset featuring pre-processed Sentinel-1 intensity data and interferometric coherence imagery acquired before and during flood events. It contains 8,879 512×512 chips covering 807,500 km^2 across 20 land cover classes and 5 continents, spanning 18 flood events. We used UrbanSARFloods to benchmark existing state-of-the-art convolutional neural networks (CNNs) for segmenting open and urban flood areas. Our findings indicate that prevalent approaches, including the Weighted Cross-Entropy (WCE) loss and the application of transfer learning with pretrained models, fall short in overcoming the obstacles posed by imbalanced data and the constraints of a small training dataset. Urban flood detection remains challenging. Future research should explore strategies for addressing imbalanced data challenges and investigate transfer learning's potential for SAR-based large-scale flood mapping. Besides, expanding this dataset to include additional flood events holds promise for enhancing its utility and contributing to advancements in flood mapping techniques. The UrbanSARFloods dataset, including training, validation data, and raw data, can be found at* https://github.com/jie666-6/UrbanSARFloods.

1. Introduction

As one of the most devastating natural disasters worldwide, floods have impacted billions of people [41]. Additionally, it has been reported that the frequency and severity of flooding have increased due to intensified heavy precipitation patterns [9]. Thus, there is a growing demand for flood mapping, monitoring, and forecasting on a global scale, not only for high-profile catastrophic flooding events but also for those unreported floods. Within this context, Earth Observation (EO) data, particularly satellite EO data, plays a critical role in understanding and quantifying the extent and depths of flooding events on a large scale, thus presenting a significant opportunity for global flood mapping and monitoring. Currently, a variety of satellite optical data sources such as MODIS, VIIRS, Landsat constellations, Sentinel-2, and PlanetScope, as well as satellite synthetic aperture radar (SAR) data including Sentinel-1, COSMO-SkyMed, Radarsat Constellation, TerraSAR-X, and GaoFen-3, are widely employed for flood mapping. However, satellite optical data can be easily contaminated by clouds although they offer longer temporal coverage and are easier to interpret. Therefore, satellite SAR data is preferred due to its ability to be acquired regardless of weather conditions (i.e., clouds) and solar illumination. Furthermore, it is worth noting that the biggest difference between large-scale flood mapping and region-scale flood mapping is that the methods should identify floodwater in various land cover classes, including bare soils, sparsely vegetated areas, urban areas, agricultural fields, etc. However, to the best of our knowledge, most large-scale SAR-derived flood mapping studies mainly focus on flooded open areas, i.e., bare soils and sparsely vegetated areas [e.g., 4, 21, 26, 48, 52], while large-scale SAR-derived flood mapping considering both open areas and urban areas has not been well investigated yet. To clarify, in this study, the term "*large-scale urban flood mapping*" is used to denote flood mapping that covers both open areas and urban areas.

Recently, a growing number of studies [e.g. 2, 14, 17, 19]

Sentinel-1 SLC data coverage of 18 selected
Urban flood events

▼ Train/Val sites
● Test sites

Band 1: Pre-event
coherence in VH

Band 2: Pre-event
coherence in VV

Band 3: Co-event
coherence in VH

Band 4: Co-event
coherence in VV

Band 5: Pre-event
Iintensity in VH

Band 6: Pre-event
intensity in VV

Band 7: Post-event
intensity in VH

Band 8: Post-event
intensity in VV

Non-Flooded areas
Flooded open areas
Flooded ubran areas

Semi-Automatic Flood Labels: Beira,
Mozambique (March 20, 2019)

8 Bands of Pre-Processed Sentinel-1 SLC Data
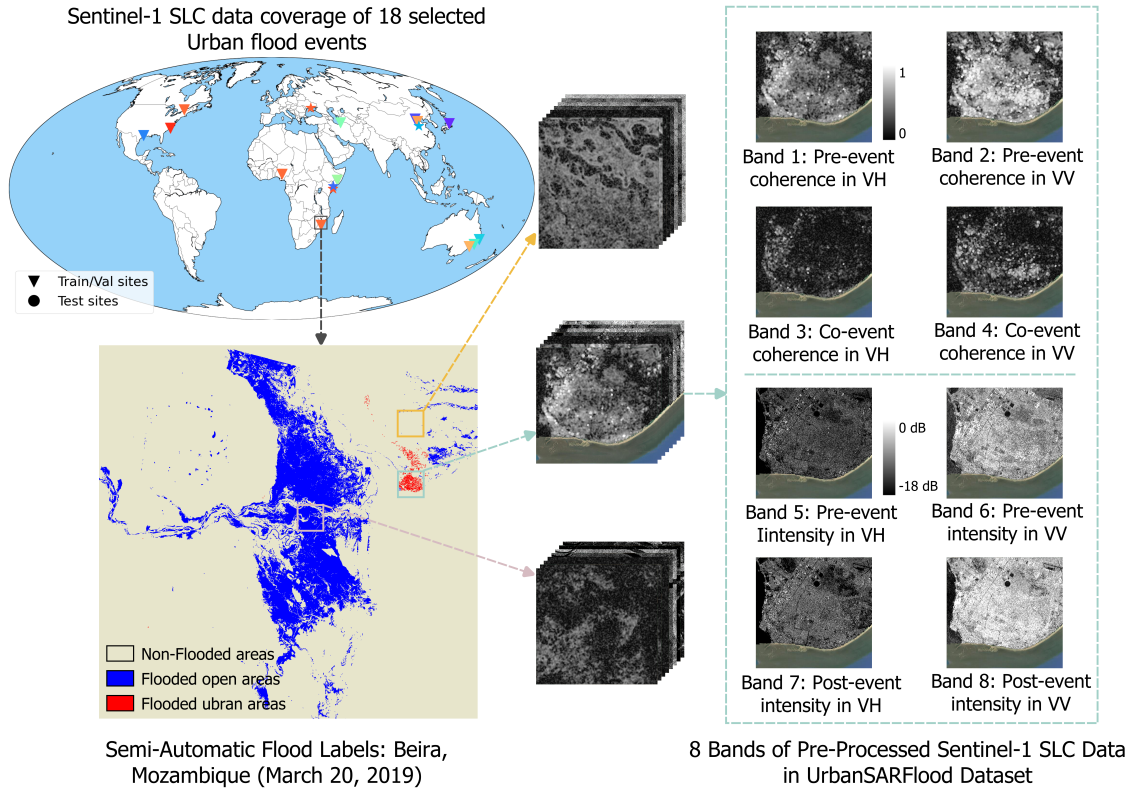in UrbanSARFlood Dataset

Figure 1. Overview of the UrbanSARFloods dataset.

have utilized SAR data, showing its efficacy in identifying floodwater in urban areas. It has been found that both SAR intensity and Interferometric SAR (InSAR) coherence are critical in urban flood mapping [28, 53]. Specifically, in SAR data, the totally submerged areas (such as bare soils and buildings) typically exhibit lower backscatter than the background because the SAR signal has been reflected away from the sensor due to the specular reflection in the calm open water surface. Partially submerged buildings, on the other hand, often show relatively higher backscatter due to the double-bounce effects between the water surface and the buildings' facades. Additionally, the coherence of flooded buildings experiences a sharp drop with the appearance of floodwater, serving as a useful indicator in flood detection, particularly when changes in the intensity of flooded buildings are too small to be detected. Furthermore, it is essential to consider different polarizations simultaneously, as the backscatter is influenced by both the orientation of the buildings and the line-of-sight (LOS) of the sensor [37, 53]. However, since most existing methods have not been tested on large datasets in both temporal and spatial dimensions, their generalizability and robustness in large-scale urban flood mapping remain unknown.

Although deep learning (DL) techniques have a transformative impact on the remote sensing field, their integration into SAR-based large-scale urban flood mapping applications remains limited. Only a handful of studies [28, 49, 53] have ventured into this domain so far. Comparing their generalizability and robustness is challenging due to variations in data, flood events, and data preprocessing methodologies employed across studies. Meanwhile, we have observed that a growing number of DL-based large-scale flood mapping studies [e.g. 16, 23, 24, 27, 33, 47] were carried out thanks to the publication of the georeferenced Sen1Floods11 dataset [5]. This phenomenon of public datasets driving advances across various research fields has also been demonstrated in different domains. For instance, in the field of image classification and object detection in computer vision, ImageNet has provided researchers worldwide with access to vast amounts of data, while its test dataset has served to document the enhancements in computer vision capabilities [5]. Therefore, it is believed that one of the bottlenecks hindering the advancement of DL-based methodologies in the field of SAR-based large-scale urban flood mapping is the absence of a benchmark dataset.

To bridge this gap and engage more AI researchers in advancing large-scale urban flood mapping with SAR data, we developed a georeferenced flood dataset named **UrbanSARFloods**. The dataset comprises flooded urban and open areas, with SAR intensity and InSAR coherence data

obtained pre- and post-event in VV and VH polarizations using Sentinel-1 SLC data(Fig. 1). It offers global coverage, consisting of 8,879 chips of 512×512 20m pixels across 18 flood events, spanning 807,500 km^2. All 8,879 chips underwent semi-automatic labeling via conventional remote sensing methods. Moreover, three subsets from selected events were manually annotated using high-resolution optical data (i.e., 3m PlanetScope and 10cm UAV optical RGB orthophoto). Semi-automatic and manual labels are utilized during testing. Further details are provided in Section 3. Leveraging the UrbanSARFloods dataset, we evaluate flood detection performance of various semantic segmentation models, including both pre-trained and scratch-trained ones. Our findings highlight the persistent challenge in urban flood detection, primarily due to data imbalance and limited dataset availability.

2. Related Work

2.1. Flood-related benchmark datasets

According to several remote sensing data platforms, such as EarthNet (https://earthnets.github.io/) [46], the IEEE platform (https://eod-grss-ieee.com/dataset-search), and SpaceML from Frontier Development Lab (https://spaceml.org/repo), there are several flood-related datasets available. As shown in Tab. 1, it is evident that most flood-related datasets currently focus on flooded open areas. However, only three datasets (SpaceNet, Hurricane Harvey Floods, Flood-Net) provide annotations related to flood-affected buildings. However, the Hurricane Harvey Floods dataset is designed for flood-damaged buildings instead of flooded buildings, while the SpaceNet and FloodNet datasets provide annotations for two flood classes (i.e., flooded buildings and flooded roads) using high-resolution optical data. Besides, another recent S1GFloods dataset, which reportedly includes wetlands, riverine areas, mountainous regions, urban and rural areas, and vegetation [43], annotated flooded areas using Sentinel-1 intensity data only. Therefore, only severely flooded buildings may be included, while most of the flooded built-up areas cannot be effectively included, as it has been established that InSAR coherence is important complementary information with SAR intensity when it comes to detecting flooded buildings using SAR data.

Indeed, there is currently no suitable SAR dataset specifically designed for large-scale urban flood mapping. While it is possible to gather flood labels from various studies, as exemplified by [6] who shared their manually annotated flood labels as a supplementary file, inconsistencies in annotation granularity stemming from variations in spatial resolution and labeling styles may compromise the reliability of conclusions drawn from such data. Additionally, it has been observed that certain non-flooded areas,

such as tarmac and shrubland, exhibit characteristics similar to flooded areas, potentially leading to overestimations in large-scale flood mapping efforts [51]. In comparison to existing datasets, UrbanSARFloods offers a distinctive advantage by encompassing two flooded classes (i.e., flooded open areas and flooded urban areas) characterized by SAR intensity and InSAR coherence data. This unique feature renders UrbanSARFloods particularly suitable for large-scale urban flood mapping applications.

2.2. Semantic Segmentation in flood mapping application

Currently, numerous semantic segmentation methods have recently been applied to flood mapping applications using SAR data. For instance, [35] evaluated the performance of UNet [42], XNet [8] and UNet with ResNet as the backbone on the UNOSAT Flood Dataset (https://unosat-rm.cern.ch/FloodAI/apps/MMR/), stating that the performance of the models did not significantly differ. However, the UNet with ResNet as the backbone was reported as the most favorable due to its greater flexibility in the choice of precision/recall tradeoff. Subsequently, [23] indicated that UNet outperformed Seg-Net [1] using the Sen1Floods11 dataset in flood mapping, attributed to its skip connection architecture. Furthermore, [31] proposed a modified DeepLabV3+ model employing MobileNetv2 as the backbone for detecting flooded open areas using a C-band commercial SAR satellite data Hisea-1, which is surpassed SegNet, UNet, and DeepLabv3+ [12] in both accuracy and inference time [31].

Recently, [20] introduced the Residual Wave Vision U-Net (WVResU-Net), trained and tested on Sentinel-1 data, integrating advanced Vision Multi-Layer Perceptrons (MLPs) and ResU-Net, exhibiting significant superiority over several well-known CNN and ViT DL models. Beside the supervising learning, an unsupervised CNN model has been introduced in SAR-based flood mapping [21], where SAR images are pre-segmented using the graph-based Felzenszwalb and Huttenlocher (Felz) segmentation algorithm, followed by the CNN being employed to generate the final flood map. In recent research, [47] introduced a lightweight unsupervised flood mapping DL model, named Contrastive ConvLSTM Variational AutoEncoder (CLVAE), which employs fully self-supervised training with simplified contrastive learning techniques. However, it should be noted that all the above-mentioned studies were developed for flooded open areas instead of flooded urban areas using SAR data. To the best of our knowledge, only a handful of studies [28, 49, 53] have been conducted specifically focusing on SAR-based urban flood mapping using DL techniques, where many popular semantic segmentation models such as UNet++ and Deeplabv3+ have not been investigated.

Table 1. Summary of current publicly available flood-related datasets.

Dataset	Platform	Modality	Resolution [m]	Flood Labels
MM-Flood [34]	Spaceborne	SAR intensity, DEM, hydrography maps	10	Flooded open areas
Sen1Floods11 [5]	Spaceborne	SAR intensity, Multispectral	10	Flooded open areas
Sen12-Flood [40]	Spaceborne	SAR intensity, Multispectral	10	Flooded open areas
RAPID-NRT [48]	Spaceborne	SAR intensity, topography, water occurrence, land cover classification, river width, hydrography, and water type	10	Flooded open areas
SpaceNet 6 [45]	Spaceborne	Optical data	0.3 - 0.8	Flooded building / Flooded road
EU Flood Dataset [3]	Airborne / Ground collected data	Multispectral data	-	Flooding
Hurricane Harvey Floods [36]	Spaceborne	SAR intensity, InSAR coherence, Optical data	1, 2, 10	Flood Damage building
Flood Area Segmentation [22]	Airborne	Optical data	-	Flood
Roadway flooding image [44]	Ground collected data	Optical data	-	Flood
ML4Floods [38]	Spaceborne	Optical data	10 - 30	Flooded open areas
FloodNet [39]	Airborne	Optical data	-	Flooded building / Flooded road
California flood dataset [30]	Spaceborne	SAR intensity, Multispectral	-	Flooded open areas

3. Datasets

Our dataset contains 18 Sentinel-1-covered urban flood events, in which the changes caused by floodwater in open areas and urban areas could be measured. The geographic distribution of selected flood events is shown in Fig. 1, while more detailed information is listed in Tab. 2. The imagery in our dataset has 8 bands, including 2 bands (VV and VH) for Sentinel-1 intensity acquired pre-event, 2 bands (VV and VH) for Sentinel-1 intensity acquired post-event, 2 bands (VV and VH) for Sentinel-1 coherence acquired pre-event, and 2 bands (VV and VH) for Sentinel-1 intensity acquired co-event.

3.1. Preprocessing

The Sentinel-1 Level-1 Interferometric Wide Swath SLC data, downloaded from Alaska Satellite Facility (ASF), were used to extract interferograms and then calibrated and transformed into intensity (in dB). The multilooking (4 looks in the range and 1 looks in the azimuth) was carried out in order to get the interferogram with the square pixels. A Goldstein filter with a size of 9×9 pixels was applied to the interferogram to reduce noise in the phase, and then the interferometric coherence was estimated by a moving window of 9×9 pixels. All the intensity and coherence data were geocoded to World Geodetic System (WGS) 1984 longitude and latitude with 20 m spatial resolution.

3.2. Label generation

3.2.1 Semi-automatic Labeling

In this study, generating annotations covering the entire Sentinel-1 image frame can only be achieved through conventional remote sensing approaches. This is because no high-resolution optical images acquired on the same acqui-sition dates as Sentinel-1 data are capable of covering the entire image frame ($\sim$43,000 km^2). Three-step annotation is needed: 1) the open flooded areas are extracted by applying a hierarchical Split-based change detection approach (i.e., HSBA) [13] to SAR intensity imagery; 2) the urban floods are extracted using a threshold (fixed at 0.3 based on trial-and-error) on difference interferometric coherence image (i.e., pre-event coherence - co-event coherence). The built-up areas and non-built-up areas are distinguished using World Settlement Footprint 2019 (WSF2019) [32]; 3) isolated objects with a small number of pixels are eliminated, whose threshold for removal was determined on a case-by-case basis by remote sensing analysts. It should be noted that for both flood classes, once a pixel has been annotated as flooded in at least one polarization, it is annotated as flooded.

In the annotation data, flooded open areas are designated with a pixel value of 1, flood urban pixels are assigned a value of 2, and non-flooded regions are represented by a pixel value of 0.

3.2.2 Hand Labeling

For the areas where high-resolution optical data is available, trained remote sensing analysts annotated open floods and urban floods using centimeter-level UAV-based RGB imagery. All annotations were carried out using QGIS software. Due to the spatial resolution difference between RGB imagery and Sentinel-1 data, analysts need to remove some small flooded areas that cannot be recognized in Sentinel-1 data due to its coarse resolution.

3.3. Training, Validation and Testing Data

We aim to achieve diversity in addressing real-world flood issues, thereby advancing the development of effective

Table 2. Detailed description of global flood events in UrbanSARFloods dataset. Events labeled with * signify the inclusion of optical data.

Continent	Location	Event Date	Image Size	Absolute Orbit	Path	Number of tiles		
						NF	FO	FU
North America	Houston, US	30 August 2017	14918 × 12981	7169	143	274	323	129
	Lumberton, US	11 Oct 2016	9931 × 6465	13449	77	147	201	2
	Sainte-Marthe-sur-le-Lac, Canada	02 May 2019	21638 × 11184	27055	33	279	431	2
Africa	Beledweyne, Somalia	08 May 2018	14293 × 11656	21807	35	495	73	4
	Beira, Mozambique	20 March 2019	14904 × 12608	15432	6	133	89	14
	Beledweyne, Somalia	14 Nov 2023	15457 × 12634	51207	35	455	56	20
	Jubba, Somalia*	01 Dec 2023	15548 × 13078	50580	108	400	59	13
			15454 × 12710	51455	108	460	65	10
	Lokoja, Niger	13 Oct 2022	15500 × 12587	44902	30	427	107	1
Asia	Iwaki/Koriyama, Japan	12 Oct 2019	11751 × 10096	18447	46	353	158	19
	Weihui, China*	27 July 2021	18927 × 12245	38962	40	293	221	135
	Aqqala, Iran	29 March 2021	19549 × 12580	26554	57	333	351	25
	Zhuozhou, China	05 August 2023	19906 × 12207	49739	142	332	204	137
	Langfang, China	05 August 2023	19458 × 12220	49739	142	279	269	117
Oceania	Coraki, Australia	2 March 2022	18160 × 13358	42146	74	105	54	11
	Sydney, Australia	24 March 2021	19495 × 13582	37144	147	468	95	6
	Port Macquarie, Australia	19 March 2021	18774 × 13460	37071	74	124	89	26
Europe	NovaKakhovka, Ukraine*	09 June 2023	22596 × 12226	48911	14	103	612	37

flood detection methodologies. Hence, we consider not only the quantity of images but also their representativeness across various flood scenarios during data division. Specifically, we aim for training and validation datasets that encompass floods occurring in various land cover classes and different environmental conditions, while the testing datasets should feature floods from diverse locations, allowing us to evaluate the robustness and transferability of different flood detection methods. Therefore, three flood events located in Africa, Asia, and Europe were selected as testing cases due to the availability of high-resolution optical data, while the remaining 15 flood events were used for model training and validation.

For the 15 flood events utilized for training and validation, all imagery was partitioned into 512×512 pixel non-overlapping chips. In our dataset, we retained all chips where no flood exists, as it is believed that the features of non-flood pixels also contribute to the improved detection of flooded pixels. More specifically, concerning large-scale flood mapping, many non-flooded areas exhibit flood-lookalike characteristics, such as the sparse shrubland near Beledeweyne, Somalia, which may lead to confusion for flood detection models and result in over-detection [51] if no reliable annotation data from those specific areas is involved. Moreover, flood pixels typically constitute only a small fraction of the entire scene, even during catastrophic flood events, especially when dealing with 20m spatial resolution SAR data covering an area of 43,000 km^2. Consequently, flood datasets often exhibit an extreme data imbalance ratio ($\rho > 1000$), especially in urban flood cases. Therefore, the training and validation data division was carried out using stratified sampling strategies based on flood

event cases.

Firstly, all tiles were classified into Non-flooded (NF) tiles, Flooded open areas (FO) tiles, and Flooded urban (FU) tiles by examining the maximal value in the annotation data generated in Sec. 3.2.1: if the maximal value in the annotation data of a tile is 2, then the tile is classified as an FU tile; if the maximal value is 1, then the tile is classified as an FO tile; otherwise, the tile is classified as an NF tile. Then, both the FO and FU tiles were further divided into two subclasses separately: tiles with $\rho > 1000$ were assigned as subclass 1, while tiles with $\rho < 1000$ were assigned as subclass 2. In other words, all tiles were classified into 5 categories: NF, FO_1, FO_2, FU_1, and FU_2. For each category, 70% of the tiles were allocated to the training dataset, while the remaining tiles were utilized for the validation dataset.

Following this scheme, our dataset comprises 8,879 non-overlapping chips covering 807,500 km^2: 2,408 from three selected study sites for testing only; 4,501 for training, and 1,970 for validation across the remaining 15 events.

3.4. Statistics of the UrbanSARFloods Dataset

To have a better understanding of the dataset, we also show the statistics of the land cover classes distribution in training, validation, and testing dataset separately, where the Copernicus Global Land cover map 2019 [7] was involved for such analysis. Fig. 2 demonstrates that both the training set, validation set, and testing set contain different land cover classes and the distribution of land cover classes is similar. Also, we display the label distributions of the training set, validation set, and testing set, where the annotation obeys clear long-tailed distributions and indicates the seri-

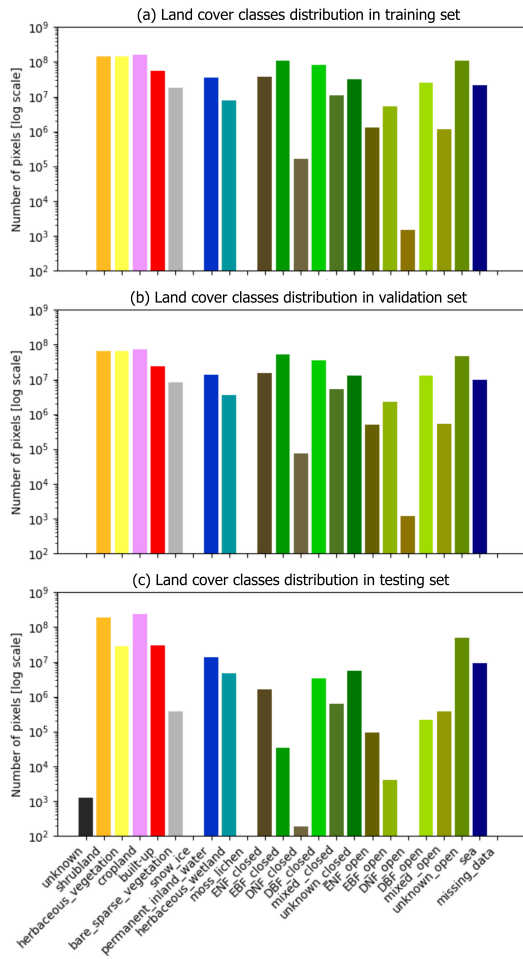ous data imbalance issue (Fig. 3).

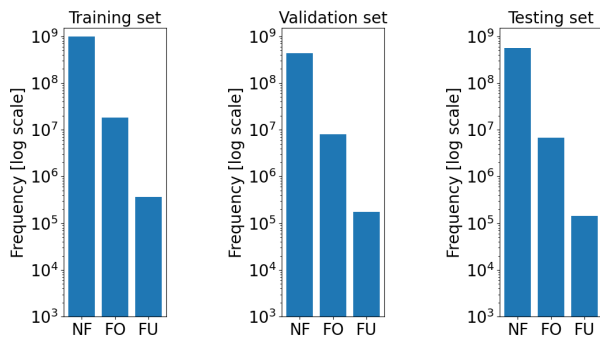Figure 2. Statistics of label distribution of UrbanSARFloods.

Figure 3. Statistics of semi-automatic label distribution of UrbanSARFloods.

4. Experiment and results

In this dataset, we have explicitly provided the official splits for the training, validation, and test subsets. Offering a well-defined dataset and official split is crucial for ensuring reproducibility. In addition, the different semantic segmentation models were trained on UrbanSARFloods dataset, including Unet [42], Unet++ [54], MANet [15], Linknet [10], FPN [29], PSPNet [50], PAN [25], DeepLabV3 [11] and DeeplabV3+ [12].

4.1. Implementation details

All models were implemented using PyTorch and executed on an NVIDIA A40 GPU. The public codebase provided by Segmentation Models PyTorch [18] was employed for this purpose. Given our aim to establish straightforward baselines for the datasets rather than optimize for the best possible models, we did not conduct an exhaustive hyperparameter search. Training for all models was conducted for 100 epochs. The input images, initially sized 512×512, were randomly cropped into 256×256 dimensions with random horizontal and vertical flips, as well as random rotations (i.e., 90°, 180°, 270°), applied for data augmentation. The batch sizes were set to 12, while we utilized the Adam optimizer with an initial learning rate of 1e-5 and a weight decay coefficient of 1e-4. The Weighted Cross-Entropy (WCE) loss was employed to address the class imbalance among the three classes. To evaluate the efficacy of transfer learning in flood mapping, each model underwent two training regimens: one in which the model's weights were initialized via Xavier initialization, and another in which the model was pre-trained using the ImageNet dataset. For all experiments, Precision, Recall, and F1 score of each class were used as the evaluation metrics.

4.2. Evaluation of State-of-the-art semantic segmentation models

We evaluated nine existing semantic segmentation models available in Segmentation Models PyTorch [18]. The results of two flood classes are presented in Tab. 3 and Tab. 4, separately. As is shown in Tab. 3, the F1 score for FO in Weihui and Jubba ranges from 0.51 to 0.77 due to relatively low precision, indicating overestimation of FO. An example of Weihui is shown in Fig. 4, where FO is displayed in blue and FU is displayed in red. Within the RGB combination of pre-event/post-event intensity, FO is cyan and FU is red in Fig. 4 (2). Similarly, within the RGB combination of pre-event and co-event coherence, FU is cyan in Fig. 4 (3). Thus, combining the label data in Fig. 4 (1) and SAR data in Fig. 4 (2-3), it is clear that most overestimation of FO mainly exists in the boundary of FO. However, the F1 source of FO is much lower in the NovaKakhovka with the corresponding precision below 0.2. Compared with the 3m Planetscope data acquired on the same date, there are two

sources of those FO false alarms: non-flooded agriculture fields and wind-affected permanent water surfaces having similar characterises of FO.

When it comes to the quantitative results of FU, the F1 scores and their corresponding precision are lower than 0.1 (Tab. 4), indicating too many false alarms of FU exist. Besides the difficulty in distinguishing flooded urban pixels from other FU-lookalike pixels, some pixels that definitely do not have FU-lookalike features have been wrongly classified as FU, as is shown in the yellow box in Fig. 4. In addition, there is a disparity in performance between FO and FU, which can be attributed to the imbalanced data issue, despite our efforts to address it using the Weighted Cross-Entropy (WCE) loss.

Furthermore, we conducted an evaluation of all flood maps using manually annotated data, and no significant differences were observed. Therefore, it is inferred that the relatively poor performance of all models is attributable to the challenges inherent in large-scale flood mapping applications, rather than being solely attributed to the semi-automatic label data itself.

4.3. Evaluation of transfer learning

Furthermore, all models pretrained on ImageNet were fine-tuned and tested on our dataset. Results are presented in Tables 5 and 6. Comparing with results in Tables 3 and 4, no significant performance differences were observed. This could be attributed to the substantial disparity between the ImageNet dataset and our UrbanSARFloods dataset. ImageNet comprises millions of RGB neutral images, while UrbanSARFloods consists of 8-band SAR data, including SAR intensity and InSAR coherence data. This difference in feature spaces between the input channels of the source domain (ImageNet) and the target domain (UrbanSARFloods) may impede the models' ability to extract relevant features for flood classification in our study. Similar results were obtained when using manually annotated ground truth data for subsets of the entire Sentinel-1 image frame. Hence, further exploration of advanced transfer learning techniques, such as domain adaptation, is essential to address challenges in large-scale urban flood mapping.

5. Conclusion

One of the bottlenecks in integrating DL techniques with large-scale urban flood mapping is the lack of proper open-access datasets. To address this gap, we constructed a pre-processed Sentinel-1 dataset, known as UrbanSARFloods, encompassing both urban and rural floods. This dataset encapsulates two significant challenges encountered in large-scale remote sensing mapping: complex background samples and imbalanced data. We evaluated state-of-the-art methods on the UrbanSARFloods dataset, uncovering the specific challenges posed by UrbanSARFloods. Addition-
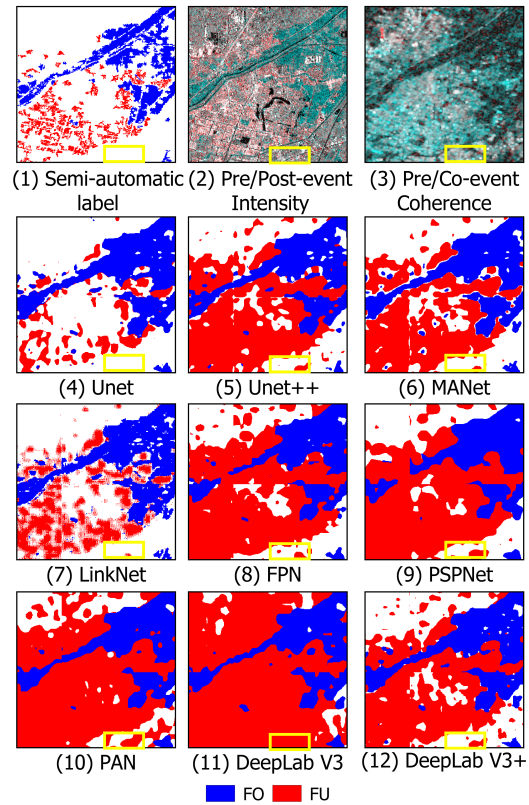

Figure 4. Example of one test site (Weihui): flood label data, SAR data, and generated flood maps using different models trained from scratch.

ally, we conducted transfer learning experiments to explore alternative approaches for overcoming these challenges.

This work offers a free and open dataset to advance large-scale urban flood mapping in the area of microwave remote sensing. We also provide this benchmarked task with two considerable challenges, allowing other researchers to easily build on this work and create new and enhanced capabilities. A potential positive societal impact may arise from the development of generalizable models that can produce large-scale flood maps considering urban floods and rural floods accurately. This could help provide global-scale flood maps using all achieved satellite data.

Acknowledgement

Table 3. Evaluation of results using semi-supervised labels across the entire Sentinel-1 image frame for Flooded open areas (FO)

	Weihui			Jubba			NovaKakhovka		
	Precision ↑	Recall ↑	F1 ↑	Precision ↑	Recall ↑	F1 ↑	Precision ↑	Recall ↑	F1 ↑
Unet	0.54	0.99	0.70	0.38	0.99	0.55	0.20	0.81	0.32
Unet++	0.64	0.98	0.77	0.50	0.99	0.66	0.22	0.89	0.35
MANet	0.50	0.98	0.67	0.45	0.99	0.62	0.12	0.96	0.21
Linknet	0.46	0.99	0.63	0.46	0.99	0.63	0.11	0.93	0.20
FPN	0.56	0.97	0.71	0.46	0.99	0.63	0.19	0.86	0.31
PSPNet	0.40	0.97	0.57	0.34	0.99	0.51	0.12	0.90	0.22
PAN	0.46	0.98	0.62	0.37	0.99	0.54	0.14	0.95	0.24
Deeplab v3	0.49	0.98	0.65	0.34	0.99	0.51	0.15	0.95	0.26
Deeplab v3 +	0.56	0.98	0.72	0.42	0.99	0.59	0.18	0.88	0.30

Table 4. Evaluation of results using semi-supervised labels across the entire Sentinel-1 image frame for Flooded urban areas (FU)

	Weihui			Jubba			NovaKakhovka		
	Precision ↑	Recall ↑	F1 ↑	Precision ↑	Recall ↑	F1 ↑	Precision ↑	Recall ↑	F1 ↑
Unet	0.06	0.32	0.10	0.37	0.79	0.50	0.02	0.41	0.04
Unet++	0.01	0.97	0.02	0.12	0.95	0.21	0.01	0.86	0.02
MANet	0.02	0.93	0.04	0.19	0.83	0.31	0.01	0.61	0.02
Linknet	0.03	0.61	0.07	0.07	0.63	0.12	0.12	0.39	0.03
FPN	0.01	0.90	0.01	0.07	0.90	0.13	0.01	0.75	0.01
PSPNet	0.01	0.91	0.02	0.08	0.95	0.15	0.01	0.81	0.01
PAN	0.01	0.98	0.01	0.07	0.90	0.13	0.01	0.70	0.02
Deeplab v3	0.01	0.88	0.01	0.12	0.74	0.21	0.01	0.77	0.01
Deeplab v3 +	0.01	0.87	0.03	0.13	0.89	0.23	0.01	0.77	0.02

Table 5. Evaluation of results using semi-supervised labels across the entire Sentinel-1 image frame for Flooded open areas (FO) using the pretrained models

	Weihui			Jubba			NovaKakhovka		
	Precision ↑	Recall ↑	F1 ↑	Precision ↑	Recall ↑	F1 ↑	Precision ↑	Recall ↑	F1 ↑
Pretrained Unet	0.65	0.98	0.78	0.60	0.98	0.74	0.19	0.88	0.31
Pretrained Unet++	0.50	0.99	0.66	0.43	0.99	0.60	0.16	0.88	0.27
Pretrained MANet	0.46	0.99	0.63	0.28	0.99	0.44	0.11	0.94	0.20
Pretrained Linknet	0.53	0.99	0.69	0.35	0.99	0.52	0.12	0.94	0.21
Pretrained FPN	0.52	0.99	0.69	0.44	0.99	0.61	0.14	0.96	0.25
Pretrained PSPNet	0.58	0.94	0.71	0.50	0.98	0.66	0.30	0.74	0.43
Pretrained PAN	0.44	0.98	0.61	0.47	0.99	0.64	0.21	0.84	0.34
Pretrained Deeplab v3	0.44	0.95	0.60	0.33	0.99	0.49	0.10	0.92	0.18
Pretrained Deeplab v3 +	0.40	0.99	0.57	0.29	0.99	0.45	0.09	0.94	0.17

Table 6. Evaluation of results using semi-supervised labels across the entire Sentinel-1 image frame for Flooded urban areas (FU) using the pretrained models

	Weihui			Jubba			NovaKakhovka		
	Precision ↑	Recall ↑	F1 ↑	Precision ↑	Recall ↑	F1 ↑	Precision ↑	Recall ↑	F1 ↑
Pretrained Unet	0.02	0.74	0.05	0.15	0.87	0.25	0.01	0.70	0.02
Pretrained Unet++	0.04	0.57	0.07	0.20	0.79	0.32	0.03	0.63	0.05
Pretrained MANet	0.01	0.97	0.02	0.10	0.94	0.19	0.01	0.73	0.01
Pretrained Linknet	0.01	0.96	0.03	0.08	0.87	0.15	0.01	0.59	0.02
Pretrained FPN	0.02	0.85	0.04	0.15	0.72	0.25	0.01	0.50	0.02
Pretrained PSPNet	0.01	0.97	0.01	0.07	0.93	0.12	0.01	0.92	0.01
Pretrained PAN	0.01	0.83	0.02	0.11	0.92	0.20	0.01	0.61	0.03
Pretrained Deeplab v3	0.01	0.98	0.01	0.07	0.92	0.13	0.01	0.67	0.01
Pretrained Deeplab v3 +	0.02	0.69	0.04	0.07	0.70	0.13	0.02	0.51	0.04

References

[1] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017. 3

[2] Shadi Sadat Baghermanesh, Shabnam Jabari, and Heather McGrath. Urban flood detection using TerraSAR-X and SAR Simulated Reflectivity Maps. *Remote Sensing*, 14(23):6154, 2022. 1

[3] Björn Barz, Kai Schröter, Moritz Münch, Bin Yang, Andrea Unger, Doris Dransch, and Joachim Denzler. Enhancing flood impact analysis using interactive retrieval of social media images. *arXiv preprint arXiv:1908.03361*, 2019. 4

[4] Bernhard Bauer-Marschallinger, Senmao Cao, Mark Edwin Tupas, Florian Roth, Claudio Navacchi, Thomas Melzer, Vahid Freeman, and Wolfgang Wagner. Satellite-based flood mapping through bayesian inference from a sentinel-1 sar datacube. *Remote Sensing*, 14(15):3673, 2022. 1

[5] Derrick Bonafilia, Beth Tellman, Tyler Anderson, and Erica Issenberg. Sen1floods11: A georeferenced dataset to train and test deep learning flood algorithms for sentinel-1. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 210–211, 2020. 2, 4

[6] Fabio Brill, Stefan Schlaffer, Sandro Martinis, Kai Schröter, and Heidi Kreibich. Extrapolating satellite-based flood masks by one-class classification—a test case in houston. *Remote Sensing*, 13(11):2042, 2021. 3

[7] Marcel Buchhorn, Bruno Smets, Luc Bertels, Bert De Roo, Myroslava Lesiv, Nandin-Erdene Tsendbazar, Martin Herold, and Steffen Fritz. Copernicus global land service: Land cover 100m: collection 3: epoch 2019: Globe. *Version V3. 0.1*, 2020. 5

[8] Joseph Bullock, Carolina Cuesta-Lázaro, and Arnau Quera-Bofarull. Xnet: a convolutional neural network (cnn) implementation for medical x-ray image segmentation suitable for small datasets. In *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*, pages 453–463. SPIE, 2019. 3

[9] M.A. Caretta, A. Mukherji, M. Arfanuzzaman, R.A. Betts, A. Gelfan, Y. Hirabayashi, T.K. Lissner, J. Liu, E.L. Gunn, R. Morgan, S. Mwanga, S. Supratid, H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, and B. Rama. Water. in: Climate change 2022: Impacts, adaptation and vulnerability. contribution of working group ii to the sixth assessment report of the intergovernmental panel on climate change. Technical report, Cambridge University, 2022. 1

[10] Abhishek Chaurasia and Eugenio Culurciello. Linknet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE visual communications and image processing (VCIP)*, pages 1–4. IEEE, 2017. 6

[11] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017. 6

[12] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018. 3, 6

[13] Marco Chini, Renaud Hostache, Laura Giustarini, and Patrick Matgen. A hierarchical split-based approach for parametric thresholding of sar images: Flood inundation as a test case. *IEEE Transactions on Geoscience and Remote Sensing*, 55(12):6975–6988, 2017. 4

[14] Marco Chini, Ramona Pelich, Luca Pulvirenti, Nazzareno Pierdicca, Renaud Hostache, and Patrick Matgen. Sentinel-1 insar coherence to detect floodwater in urban areas: Houston and hurricane harvey as a test case. *Remote Sensing*, 11(2): 107, 2019. 1

[15] Tongle Fan, Guanglei Wang, Yan Li, and Hongrui Wang. Ma-net: A multi-scale attention network for liver and tumor segmentation. *IEEE Access*, 8:179656–179665, 2020. 6

[16] Shubhika Garg, Ben Feinstein, Shahar Timnat, Vishal Batchu, Gideon Dror, Adi Gerzi Rosenthal, and Varun Gulshan. Cross modal distillation for flood extent mapping. *arXiv preprint arXiv:2302.08180*, 2023. 2

[17] Hideomi Gokon, Fuyuki Endo, and Shunichi Koshimura. Detecting urban floods with small and large scale analysis of alos-2/palsar-2 data. *Remote Sensing*, 15(2):532, 2023. 1

[18] Pavel Iakubovskii. Segmentation models pytorch. `https://github.com/qubvel/segmentation_models.pytorch`, 2019. 6

[19] Pasquale Iervolino, Raffaella Guida, Antonio Iodice, and Daniele Riccio. Flooding water depth estimation with high-resolution SAR. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2295–2307, 2015. 1

[20] Ali Jamali, Swalpa Kumar Roy, Leila Hashemi Beni, Biswajeet Pradhan, Jonathan Li, and Pedram Ghamisi. 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, 2024. 3

[21] Xin Jiang, Shijing Liang, Xinyue He, Alan D Ziegler, Peirong Lin, Ming Pan, Dashan Wang, Junyu Zou, Dalei Hao, Ganquan Mao, et al. Rapid and large-scale mapping of flood inundation via integrating spaceborne synthetic aperture radar imagery with unsupervised deep learning. *ISPRS journal of photogrammetry and remote sensing*, 178:36–50, 2021. 1, 3

[22] Md Faizal Karim, Krish Sharma, and Niyar R Barman. Flood area segmentation, 2023. Accessed on [8 March 2024]. 4

[23] Vaibhav Katiyar, Nopphawan Tamkuan, and Masahiko Nagai. Near-real-time flood mapping using off-the-shelf models with sar imagery and deep learning. *Remote Sensing*, 13 (12):2334, 2021. 2, 3

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

[25] Hanchao Li, Pengfei Xiong, Jie An, and Lingxue Wang. Pyramid attention network for semantic segmentation. *arXiv preprint arXiv:1805.10180*, 2018. 6

[26] Junjie Li, Linyi Li, Yanjiao Song, Jiaming Chen, Zhe Wang, Yi Bao, Wen Zhang, and Lingkui Meng. A robust large-scale surface water mapping framework with high spatiotemporal resolution based on the fusion of multi-source remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 118:103288, 2023. 1

[27] Wenwen Li, Hyunho Lee, Sizhe Wang, Chia-Yu Hsu, and Samantha T Arundel. Assessment of a new geoai foundation model for flood inundation mapping. In *Proceedings of the 6th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, pages 102–109, 2023. 2

[28] Yu Li, Sandro Martinis, and Marc Wieland. 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, 2019. 2, 3

[29] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017. 6

[30] Luigi Tommaso Luppino, Filippo Maria Bianchi, Gabriele Moser, and Stian Normann Anfinsen. Unsupervised image regression for heterogeneous change detection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(12):9960–9975, 2019. 4

[31] Suna Lv, Lingsheng Meng, Deanna Edwing, Sihan Xue, Xupu Geng, and Xiao-Hai Yan. High-performance segmentation for flood mapping of hisea-1 sar remote sensing images. *Remote Sensing*, 14(21):5504, 2022. 3

[32] Mattia Marconcini, Annekatrin Metz-Marconcini, Thomas Esch, and Noel Gorelick. Understanding current trends in global urbanisation-the world settlement footprint suite. *GI_Forum*, 9(1):33–38, 2021. 4

[33] Gonzalo Mateo-Garcia, Joshua Veitch-Michaelis, Lewis Smith, Silviu Vlad Oprea, Guy Schumann, Yarin Gal, Atılım Güneş Baydin, and Dietmar Backes. Towards global flood mapping onboard low cost satellites with machine learning. *Scientific reports*, 11(1):7249, 2021. 2

[34] Fabio Montello, Edoardo Arnaudo, and Claudio Rossi. Mm-flood: A multimodal dataset for flood delineation from satellite imagery, 2022. 4

[35] Edoardo Nemni, Joseph Bullock, Samir Belabbes, and Lars Bromley. Fully convolutional neural network for rapid flood segmentation in synthetic aperture radar imagery. *Remote Sensing*, 12(16):2532, 2020. 3

[36] Orion29. Satellite image segmentation for flood damage analysis. https://github.com/orion29/Satellite-Image-Segmentation-for-Flood-Damage-Analysis. 4

[37] Ramona Pelich, Marco Chini, Renaud Hostache, Patrick Matgen, Luca Pulvirenti, and Nazzareno Pierdicca. Mapping floods in urban areas from dual-polarization insar coherence data. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022. 2

[38] Enrique Portalés-Julià, Gonzalo Mateo-García, Cormac Purcell, and Luis Gómez-Chova. Global flood extent segmentation in optical satellite images. *Scientific Reports*, 13(1):20316, 2023. 4

[39] Maryam Rahnemoonfar, Tashnim Chowdhury, Argho Sarkar, Debvrat Varshney, Masoud Yari, and Robin Roberson Murphy. Floodnet: A high resolution aerial imagery dataset for post flood scene understanding. *IEEE Access*, 9:89644–89654, 2021. 4

[40] Clément Rambour, Nicolas Audebert, Elise Koeniguer, Bertrand Le Saux, Michel Crucianu, and Mihai Datcu. Sen12-flood : a sar and multispectral dataset for flood detection, 2020. 4

[41] Jun Rentschler, Melda Salhab, and Bramka Arga Jafino. Flood exposure and poverty in 188 countries. *Nature communications*, 13(1):3527, 2022. 1

[42] Olaf Ronneberger, Philipp Fischer, and Thomas 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*, pages 234–241. Springer, 2015. 3, 6

[43] Tamer Saleh, Xingxing Weng, Shimaa Holail, Chen Hao, and Gui-Song Xia. Dam-net: Global flood detection from sar imagery using differential attention metric-based vision transformers. *arXiv preprint arXiv:2306.00704*, 2023. 3

[44] Cem Sazara, Mecit Cetin, and Khan Iftekharuddin. Image dataset for roadway flooding. Mendeley Data, V1, doi: 10.17632/t395bwcvbw.1, 2019. 4

[45] Jacob Shermeyer, Daniel Hogan, Jason Brown, Adam Van Etten, Nicholas Weir, Fabio Pacifici, Ronny Hansch, Alexei Bastidas, Scott Soenen, Todd Bacastow, et al. Spacenet 6: Multi-sensor all weather mapping dataset. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 196–197, 2020. 4

[46] Zhitong Xiong, Fahong Zhang, Yi Wang, Yilei Shi, and Xiao Xiang Zhu. Earthnets: Empowering ai in earth observation. *arXiv preprint arXiv:2210.04936*, 2022. 3

[47] Ritu Yadav, Andrea Nascetti, Hossein Azizpour, and Yifang Ban. Unsupervised flood detection on sar time series using variational autoencoder. *International Journal of Applied Earth Observation and Geoinformation*, 126:103635, 2024. 2, 3

[48] Qing Yang, Xinyi Shen, Emmanouil N Anagnostou, Chongxun Mo, Jack R Eggleston, and Albert J Kettner. A high-resolution flood inundation archive (2016–the present) from sentinel-1 sar imagery over conus. *Bulletin of the American Meteorological Society*, pages 1–40, 2021. 1, 4

[49] Qing Yang, Xinyi Shen, Qingyuan Zhang, Sean Helfrich, Josef M Kellndorfer, and Wei Hao. Promoting sar-based urban flood mapping with adversarial generative network and out of distribution detection. In *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*, pages 2336–2338. IEEE, 2023. 2, 3

[50] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In

Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2881–2890, 2017. 6

[51] Jie Zhao, Ramona Pelich, Renaud Hostache, Patrick Matgen, Senmao Cao, Wolfgang Wagner, and Marco Chini. Deriving exclusion maps from c-band sar time-series in support of floodwater mapping. *Remote Sensing of Environment*, 265: 112668, 2021. 3, 5

[52] Jie Zhao, Ramona Pelich, Renaud Hostache, Patrick Matgen, Wolfgang Wagner, and Marco Chini. A large-scale 2005–2012 flood map record derived from envisat-asar data: United kingdom as a test case. *Remote Sensing of Environment*, 256:112338, 2021. 1

[53] Jie Zhao, Yu Li, Patrick Matgen, Ramona Pelich, Renaud Hostache, Wolfgang Wagner, and Marco Chini. Urban-aware u-net for large-scale urban flood mapping using multitemporal sentinel-1 intensity and interferometric coherence. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–21, 2022. 2, 3

[54] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer, 2018. 6