

# STREET TO CLOUD: IMPROVING FLOOD MAPS WITH CROWDSOURCING AND SEMANTIC SEGMENTATION

**Veda Sunkara**

Cloud to Street  
New York, NY 112131, USA  
veda@cloudtostreet.info

**Matthew Purri**

Rutgers University  
New Brunswick, NJ 08854, USA  
matthew.purri@rutgers.edu

**Bertrand Le Saux & Jennifer Adams**

European Space Agency, ESRIN, I-00044  
Frascati (Rome), Italy  
{Bertrand.Le.Saux, Jennifer.Adams}@esa.int

## ABSTRACT

To address the mounting destruction caused by floods in climate-vulnerable regions, we propose Street to Cloud, a machine learning pipeline for incorporating crowdsourced ground truth data into the segmentation of satellite imagery of floods. We propose this approach as a solution to the labor-intensive task of generating high-quality, hand-labeled training data, and demonstrate successes and failures of different plausible crowdsourcing approaches in our model. Street to Cloud leverages community reporting and machine learning to generate novel, near-real time insights into the extent of floods to be used for emergency response.

## 1 INTRODUCTION

The frequency and magnitude of flooding are increasing at an alarming rate (UNDRR, 2015), affecting growing populations of climate-vulnerable people. Flooding affects more people than any other environmental hazard and hinders sustainable development (Hallegatte et al., 2017; CRED, 2019), and research consistently shows that relative property loss for floods are highest in places of social vulnerability (Tellman et al., 2020).

Nowcasting flood extents enables decision makers, relief agencies, and citizens to make informed decisions and provide direct relief where it is needed most. Optical, radar, and microwave satellites make it possible to remotely create scalable, low-cost, and high-quality flood maps and impact assessments. However, there are significant challenges to flood mapping, monitoring, and analyzing based on satellite data. Unique challenges arise from infrequent revisit times, varying resolutions across satellites, adverse and obscuring weather conditions, and difficult to parse images of urban areas where most of the world’s population and assets are concentrated.

Most existing algorithms to process these images, machine learning or otherwise, use finely annotated data that often requires remote sensing expertise to generate. Traditional, threshold-based remote sensing often requires a nontrivial amount of manual quality assurance and parameter tuning from domain experts.

In an effort to develop an algorithm that not only addresses these data issues but also directly engages the communities affected in disaster reporting, we propose a methodology for using crowd-sourced data and simplified flood masks to train a semantic segmentation model to generate high quality flood masks. Using Cloud to Street’s Sen1Floods11 dataset (Bonafilia et al., 2020) of high-quality hand-labeled Sentinel-2 imagery, we created a dataset of simplified flood masks and synthetic crowd-sourced data points. These masks are intended to be simple to generate even without remote sensing expertise, and therefore can be generated easily and at scale. Our synthetic crowdsourced data mirrors two plausible scenarios for aggregating data from the community: passive social media scraping and active data collection by community members or trained data collectors. Leveraging dense and sparse data at the same time is a challenge for segmentation networks that we tackle by adopting a

two-stage process (see Figure 1) in which the second stage is inspired by continual learning. After training our network using these two approaches, we benchmark our results against the models trained on purely hand-labeled and purely simplified training masks.

We expect this research to allow us to provide high quality, rapidly available flood maps for evacuation and aid. In the case of some urban areas, crowdsourcing will enable us to verify flooding on a street-by-street level where remote sensing data alone cannot. Flood waters recede quickly, sometimes before there is a satellite overpass or the clouds clear, rendering optical remote sensing data insufficient for flood detection. Similarly, radar data, which can map water through clouds, is often very noisy in urban areas as signals can bounce off buildings. With street-level crowdsourcing and machine learning, we can train models to do necessary initial inundation detection and compensate for challenges when only using satellite data.

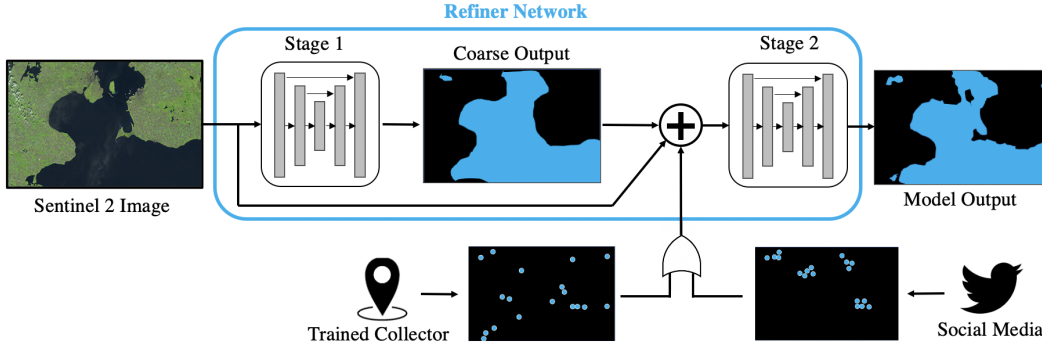


Figure 1: The inference pipeline of our model. The two-stage model first generates a segmentation mask from Sentinel-2 imagery in Stage 1, and then combines Sentinel-2 imagery, initial coarse output, and crowdsourced points in Stage 2 to generate the final segmentation mask. Points collected from either a Trained Collector or Social Media model can be used interchangeably in this model.

In this work we provide a dataset of simplified water masks of flood events, built off of Sen1Floods11, as well as a dataset of synthetic crowdsourced data for each event in a number of plausible collection scenarios. We present Street to Cloud, a multi-modal model framework which combines satellite imagery and in-situ, crowdsourced data in a segmentation and refiner network to produce nowcast flood extent maps for monitoring, aid, and disaster relief.

## 2 RELATED WORK

Prior research using the Sen1Floods11 dataset has demonstrated gains in using Fully Convolutional Neural Networks (FCNNs) to segment Sentinel-2 imagery of floods over threshold-based methods (Bonafilia et al., 2020). Of all the training strategies discussed, the most successful approach required training the network on hand-labeled images of flood events which we use in our work. Other approaches, such as DeepWaterMap (Isikdogan et al., 2020), generate water segmentation maps of Landsat-7 imagery with global surface water labels.

Multi-modal approaches to semantic segmentation of remotely sensed imagery build off of significant prior work geolocating data and incorporating crowdsourcing into disaster tracking. Efforts to geolocalize street-view imagery have shown promising results using feature matching between aerial and ground data (Regmi & Shah, 2019). The methods described can be used to identify a photographer’s angle and location when parsing crowdsourced images. Other work has delved further into flood classification from social media imagery as well as separately in satellite imagery (MediaEval, 2018), providing promising baselines for inferring ground truth data from social media images. There are examples of incorporating crowdsourcing into flood monitoring, including to assess flood depth (Hultquist & Cervone, 2020) and for interactive flood modeling (Gebremedhin et al., 2020).

Exploration into iterative segmentation using human-in-the-loop annotation (Xu et al., 2016; Lenczner et al., 2020) suggests potential gains to be made using ground-truth verified data in addition to initial segmentation masks.

### 3 METHODS

We generated two new datasets to train our model: coarse water masks of flood events and corresponding synthetic crowdsourced data points. To generate the coarse water masks, we used hand labeled Sentinel-2 imagery from the Sen1Floods11 dataset and simplified the water masks using a Gaussian blur with a large kernel. To generate the synthetic crowdsourced data, we sought to emulate two plausible approaches to data collection. The first is to emulate social media scraping, in which we anticipate a significant number of data points coming from the same region in an event (e.g. a populated street, community center). These points have low dispersion. We anticipate these data could come from Tweets, geotagged photos, and other online data sources. The second is to emulate more spread out crowdsourced data that could be obtained by contractors walking the perimeter of an event and providing data points at regular intervals. These points have high dispersion.

The total number of points per image is between 20 and 50, which makes up roughly 0.02% of the total points in the image. We sample these points from the edge of the water annotation in Sen1Floods11, either in clusters or with a higher dispersion factor to emulate these two scenarios. In addition, we incorporate varied amounts of noise into the data to emulate geolocalization and user errors (e.g. distance from a reported flood event boundary). The range of simulated noise from a GPS sensor is 0 to 50 and 0 to 100 meters for low and high noise scenarios, respectively. The points are aligned with the other data modalities by projecting the generated points onto a blank image.

We introduce a two-stage segmentation network to utilize both multispectral Sentinel-2 imagery and crowdsourced points which we call the Refiner Network. The first stage of the network is tasked with generating a coarse water mask as shown in Figure 1. The second stage refines on the coarse prediction by receiving crowdsourced points, the coarse mask, and multispectral imagery to generate a fine-grained output. We compare our Refiner Network to a standard UNet model (Ronneberger et al., 2015).

### 4 RESULTS

We assess model performance with pixel accuracy and mean intersection over union (mIoU). We benchmark against training our segmentation network on coarse labels, on finely annotated labels, and then on coarse labels with varying permutations of synthetic crowdsourced data points and noise.

Our two-stage Refiner segmentation network outperforms the standard UNet architecture for both metrics on coarse and fine annotation levels as shown in Table 1. The difference between these models is greater when trained on the coarse data than when trained on the fine data, suggesting that the refiner model is able to take advantage of more precise annotations. The refiner model, when trained with a small number of synthetic points added to the coarse annotations, nears the performance of the UNet model trained on fine annotations.

Model	Training Labels	Acc	mIoU
UNet	Coarse	95.2	53.8
Refiner	Coarse	95.6	56.5
Refiner	Coarse+Points	97.2	61.8
<i>UNet</i>	<i>Fine</i>	<i>97.0</i>	<i>62.4</i>
<i>Refiner</i>	<i>Fine</i>	<i>98.1</i>	<i>64.9</i>

Table 1: Comparison of model performance across training data annotation granularity and inclusion of crowdsourced points. Coarse training labels are created using Gaussian blur of hand labeled data, fine training labels are reference hand labeled data, and crowdsourced points are synthetically generated. We represent the best performing crowdsourcing scenario, as discussed further in Table 2.

In Figure 2, we show the qualitative improvement of utilizing crowdsourced points. The addition of crowdsourced points during training improves the model’s ability to localize small water areas such as streams or flooded pools. In the bottom of Figure 2, notice the Refiner model with points generated the most complete river compared to the other models. The low cost and minimal additional data from crowdsourced points greatly improves the performance of the network, and nears

the upper bound performance of the UNet model trained on more expensive and time consuming annotations.

We then analyze what form of crowdsourced points improve segmentation performance. In Table 2, we compare crowdsourced points generated from a 'social media' (SM) and 'trained data collector' (TDC) model, or low and high dispersion points respectively, along the fine annotation border. In Table 2, highly dispersed points result in higher model performance compared to the less dispersed points. In any situation the addition of crowdsourced points improves the performance of the refiner model over the baseline trained purely on coarse labels. Highly dispersed points with minimal noise produce the greatest improvement over the coarse, no point baseline. The importance of sensor noise affects each model differently. More severe sensor noise added to the TDC model decreases performance while more noise improves SM models. The additional noise may increase the diversity the low dispersion points, making them appear more like the TDC model.

Dispersion	Noise	Acc	mIoU
No Points	No Points	95.6	56.5
Low	Low	95.9	59.6
Low	High	96.9	61.0
High	Low	<b>97.2</b>	<b>61.8</b>
High	High	97.0	60.9

Table 2: Comparison of accuracy and mIoU across crowdsourcing dispersion and noise levels. Low dispersion corresponds to a social media scraping approach, whereas high dispersion corresponds to a trained data collector approach.

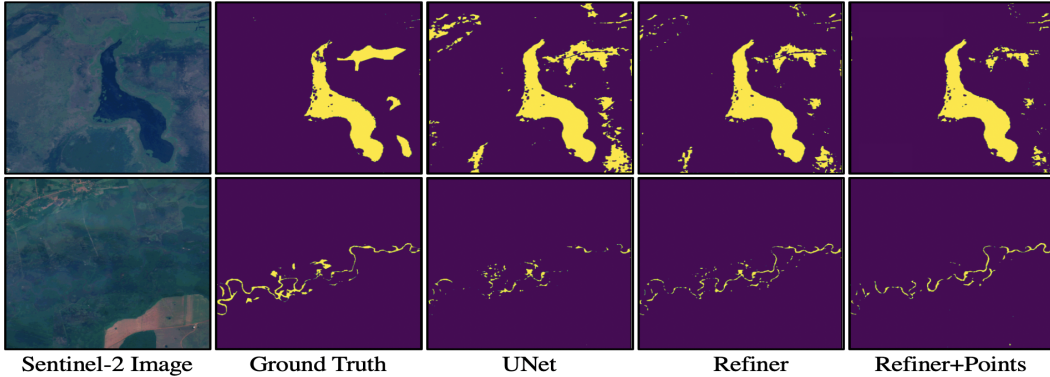


Figure 2: Qualitative results of the UNet, Refiner, and Refiner+Points models trained on coarse annotations. The Refiner+Points model appears to reduce false positives that other models generate.

## 5 DISCUSSION AND CONCLUSION

Given the challenges with generating high quality labels for remote sensing data and the asset of community partners willing to participate in crowdsourcing, we sought to build an algorithm that utilized both modalities of observation to generate better flood segmentation masks for disaster resilience. We developed a multi-modal segmentation network trained on simple, easy to generate training data and synthetic crowdsourced data points. While we found that all types of crowdsourced data points improved a single-input segmentation model, the best results used data points dispersed across the perimeter of the event.

In practice, community members or government employees could provide data points along the perimeter of flood events with which we could train models to nowcast flood extents. Social media scraping, simple WhatsApp bots, and crowdsourcing-specific mobile applications could also be used to collect data and improve segmentation models. Further feasibility studies of these crowdsourcing approaches are required, including analyses of the availability of the necessary data and the temporal frequency of its availability.

Future work should include a sensitivity analysis of the impact of crowdsourced points on the accuracy of Street to Cloud’s predictions to determine how many points are necessary to outperform existing baselines for both crowdsourcing strategies. Additional studies of obtaining and parsing real crowdsourced data to determine the feasibility of both approaches is also required. Our success with a small volume of crowdsourced data suggests that coarse training labels could be generated using unsupervised or weakly supervised learning, which is another avenue to explore when determining how to deploy this algorithm in practice.

Street to Cloud is a prototype for a multi-modal segmentation network that uses crowdsourcing to mitigate the need for finely annotated training data. With further field testing across a variety of urban and rural domains and incorporating real crowdsourced data, we anticipate this model can be used widely to nowcast flood extents, monitor impacts, and inform disaster relief.

## REFERENCES

- 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 (CVPR) Workshops*, June 2020.
- CRED. 2018: Extreme weather events affected 60 million people. Centre for Research on the Epidemiology of Disasters CRED, Jan 2019. URL <https://www.cred.be/2018-review-disaster-events>.
- Eskeidar T. Gebremedhin, Laura Basco-Carrera, Andreja Jonoski, Mark Iliffe, and Hessel Winsemius. Crowdsourcing and interactive modelling for urban flood management, Feb 2020. URL <https://onlinelibrary.wiley.com/doi/full/10.1111/jfr3.12602>.
- Stephane Hallegatte, Adrien Vogt-Schilb, Mook Bangalore, and Julie Rozenberg. *Unbreakable: building the resilience of the poor in the face of natural disasters*. World Bank, 2017.
- Carolynne Hultquist and Guido Cervone. Integration of crowdsourced images, usgs networks, remote sensing, and a model to assess flood depth during hurricane florence, Mar 2020. URL <https://www.mdpi.com/2072-4292/12/5/834>.
- L. F. Isikdogan, A. Bovik, and P. Passalacqua. Seeing through the clouds with deepwatermap. *IEEE Geoscience and Remote Sensing Letters*, 17(10):1662–1666, 2020.
- Gaston Lenczner, Bertrand Le Saux, Nicola Luminari, Adrien Chan-Hon-Tong, and Guy Le Besnerais. Disir: Deep image segmentation with interactive refinement. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, V-2-2020:877–884, Aug 2020. ISSN 2194-9050. doi: 10.5194/isprs-annals-v-2-2020-877-2020. URL <http://dx.doi.org/10.5194/isprs-annals-V-2-2020-877-2020>.
- MediaEval. Mediaeval 2018 multimedia benchmark workshop, 2018. URL <http://ceur-ws.org/Vol-2283/>.
- Krishna Regmi and Mubarak Shah. Bridging the domain gap for ground-to-aerial image matching, 2019.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.
- Beth Tellman, Cody Schank, Bessie Schwarz, Peter D Howe, and Alex de Sherbinin. Using disaster outcomes to validate components of social vulnerability to floods: flood damage and property damage across the usa. SocArXiv, Jun 2020. doi: 10.31235/osf.io/byrgu. URL [osf.io/preprints/socarxiv/byrgu](https://osf.io/preprints/socarxiv/byrgu).
- UNDRR. The human cost of weather-related disasters 1995-2015. Centre for Research on the Epidemiology of Disasters CRED, 2015. URL <https://www.undrr.org/publication/human-cost-weather-related-disasters-1995-2015>.
- Ning Xu, Brian Price, Scott Cohen, Jimei Yang, and Thomas S Huang. Deep interactive object selection. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 373–381. IEEE, 2016.