Flood Detection On Low Cost Orbital Hardware

Gonzalo Mateo-Garcia[*]
University of Valencia
gonzalo.mateo-garcia@uv.es

Silviu Oprea[*]
University of Edinburgh
silviu.oprea@ed.ac.uk

Lewis Smith[*]
University of Oxford
lsgs@robots.ox.ac.uk

Josh Veitch-Michaelis[*]
Liverpool John Moores University
j.l.veitchmichaelis@ljmu.ac.uk

Atılım Güneş Baydin
University of Oxford
gunes@robots.ox.ac.uk

Dietmar Backes
University of Luxembourg
dietmar.backes@uni.lu

Abstract

Satellite imaging is a critical technology for monitoring and responding to natural disasters such as flooding. Despite the capabilities of modern satellites, there is still much to be desired from the perspective of first response organisations like UNICEF. Two main challenges are rapid access to data, and the ability to automatically identify flooded regions in images. We describe a prototypical flood segmentation system, identifying cloud, water and land, that could be deployed on a constellation of small satellites, performing processing on board to reduce downlink bandwidth by 2 orders of magnitude. We target ΦSat-1, part of the FSSCAT mission, which is planned to be launched by the European Space Agency (ESA) near the start of 2020 as a proof of concept for this new technology.

1 Introduction

Floods are among the most destructive extreme weather events, affecting millions of people each year [1]. Satellite imagery is one of the most important sources of information for disaster response. Optical (visible and infrared) and synthetic aperture radar (SAR) imagery are routinely used to determine flood extent [2] and to help direct relief efforts.

Many countries do not have direct access to satellite imagery in the event of a disaster. To address this, organisations such as the International Charter "Space and Major Disasters"[2], initiated by the European Space Agency (ESA), liaise with space agencies and associated commercial organisations to produce free high resolution maps for users in the field. Despite best efforts it can take many days to provide actionable data, due to satellite tasking and image analysis [3]. Commercial organisations are able to provide the highest-frequency (daily) and highest-resolution (sub-metre) images, but these are only freely available for a limited period of time during disaster events. ESA's Copernicus program [4] provides open data globally at $10\,\mathrm{m}$ resolution, but the optical component, Sentinel 2 (S2, [5]), has a worst-case revisit time of around five days at the equator. This leads to wait periods much longer than two days in areas such as Africa where other alternatives for first response are limited.

In this paper we investigate how a constellation of small inexpensive satellites assembled from COTS hardware, also known as CubeSats [6], could be used for disaster response, using flooding as a case study. The main advantage of CubeSats is improved revisit time through larger constellations of satellites. Around 30 CubeSats similar to ESA's upcoming FSSCat mission [7] could be launched for the cost of a single S2 satellite, reducing the nominal revisit time from 5 days to around 8 hours for

---

[*]Equal contribution (listed in alphabetical order)
[2]https://disasterscharter.org

the same cost. However, CubeSats can have very limited downlink bandwidth, in the order of 1Mbps. In order to reduce the amount of data being transferred, we suggest performing flood mapping (an image segmentation task) on-board the satellite and only transmitting the final map.

We optimised our application for ESA's ΦSat-1, part of the FSSCat mission [8], which is scheduled to be launched in 2020[3]. Among other sensors, FSSCat will carry a HyperScout 2 49-band hyperspectral camera ($80\,\mathrm{m}$ ground sample distance) which features an Intel$^{\circledR}$ Movidius$^{\mathrm{TM}}$ Myriad$^{\mathrm{TM}}$ 2 vision processing unit (VPU) as a co-processor for performing on-board computer vision and neural network inference [7, 9]. Using this capability, a 2-bit flood map (up to 4 classes) would reduce the amount of data being downlinked by a factor of 100 (assuming 49 16-bit channels).

The contributions of this paper are as follows:

1. We introduce a new dataset, called *WorldFloods*, containing pairs of Sentinel-2 images and flood extent maps covering 159 global flood events.

2. We train convolutional neural networks for flood segmentation and compare their performance to standard baseline methods like NDWI.

3. We design our models to process large volumes of hyperspectral data, yet fit the constraints of hardware deployed on the satellite, and report test results on such hardware.

2   Flood mapping and related work

Water mapping, of which flood mapping is a special case, is a semantic segmentation task that has been studied for decades. A simple approach to water mapping is to compute indices like the Normalised Difference Water Index (NDWI) [10] which exploits the strong absorption of light by water bodies in the green and infrared part of the electromagnetic spectrum. However, this method can perform poorly because the spectral profile of flood water varies widely due to the presence of debris, pollutants and suspended sediments [11].

More sophisticated segmentation techniques include rule-based classifiers [11] which use fixed or tuned threshold on indices; classical supervised machine learning [2]; and recently deep learning [12, 13]. State-of-the-art results in image segmentation are now routinely achieved using fully convolutional neural networks (FCNNs) [14]. Most segmentation networks can be described as encoder-decoder architectures and include the popular U-Net [15] among many others [16].

3   World Floods dataset

The development and evaluation of flooding response systems has been constrained so far by use of datasets of limited geographical scope, with studies often only considering a single flood event [17]. It is unclear whether such models would accurately generalise to the rest of the world due to variations in topography and landcover. To address this we collated a new global dataset called *WorldFloods*, which we believe is the largest of its kind.

WorldFloods contains 564 flood extent maps created either manually, or semi-automatically, where a human validated machine-generated maps. The dataset covers 159 floods that occurred between November 2015 and March 2019. We sourced all maps from three organisations: the Copernicus Emergency Management Service (Copernicus EMS) [18], the flood portal of UNOSAT [19], and the Global Flood Inundation Map Repository (GLOFIMR) [20]. The geographical distribution of flood maps is shown in Figure 1.

For each flood event we provide the raw 13-band S2 image closest in time after the event, and a rasterised segmentation ground truth (cloud, water and land) at $10\,\mathrm{m}$ resolution. We generated cloud masks using `s2cloudless`.[4] We manually validated the data to account for gross errors such as missing water bodies or invalid intensities. Sources of error include label noise and temporal misalignment, e.g. the closest S2 image may have been acquired 5–6 days after the map was produced. While the labels in the training set may be noisy, we wanted to ensure that the test set provides a fairly clean measurement of the true performance of our system. In this direction, we manually selected

---

[3]FSSCat has been developed by ESA and other partners as a technology demonstrator.

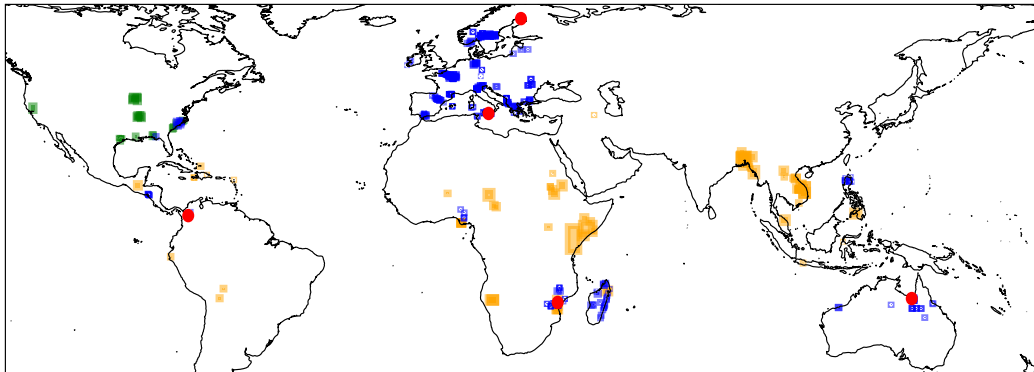[4]https://github.com/sentinel-hub/sentinel2-cloud-detector

Figure 1: Locations of flood events contained in *WorldFloods*. Blue, orange and green areas denote Copernicus EMS, UNOSAT and GLOFIMR data, respectively. Red points denote test regions.

test images from flood extent maps that were derived from S2 images and had no misalignment. To avoid data leakage, we did not include test set countries in the training set. Additionally, our test set does not contain any flood maps from events in the training and validation sets. Table 1 shows the train/validation/test statistics.

Table 1: General statistics of the training, validation and test splits of the *WorldFloods* dataset.

Dataset	Flood events	Flood maps	256x256 patches	Water/flood pixels (%)	Land pixels (%)	Cloud pixels (%)	Invalid pixels (%)
Training	60	378	76,500	2.58	37.19	56.35	3.87
Validation	6	7	1,552	8.48	76.17	13.27	2.07
Test	5	10	2,029	21.14	59.57	15.98	3.31

4   Experiments

In order to demonstrate that a FCNN-based flood detection model can segment floods accurately and could be deployed on ΦSat-1, we first train FCNN models on *WorldFloods* at its original resolution (10 m). We then train models on *degraded* imagery, mimicking the resolution of HyperScout-2 (80 m). We also verify our trained (degraded) models can be run on a Intel® Movidius™ Myriad™ 2 and measure the processing speed.

We focus on the segmentation accuracy of the water/flood class by measuring precision, recall and the Intersection over Union (IoU). Since missing flooded areas (false negatives) is more problematic than over-predicting floods (false positives), high recall is preferred to high precision. Provided the recall was over 95%, we found that the IoU was a good compromise.

As baselines, we use NDWI (S2 band 2 and 8) and a linear model (all S2 bands) trained on *World-Floods*. A range of NDWI thresholds have been suggested [10, 11, 21]. In order to set a stronger baseline, we also compute results where the threshold is tuned for each image, representing the absolute best case performance for the NDWI. We compare our baselines to two FCNNs: a simple CNN (SCNN) comprising four convolutional layers (0.26M parameters) and a U-Net (7.8M parameters, [15]).

Models were trained from scratch for 40 epochs using all 13 S2 bands with input patches of size 256x256 for 10 m data or 64x64 for 80 m data (2.5 km x 2.5 km). In order to achieve models with high recall we used a cross-entropy loss function that weights each class by the inverse of the observed frequency in Table 1, combined with a Dice loss [22]. Augmentation was applied during training including flips and rotations, per-channel jitter, Poisson noise and brightness/contrast adjustments.
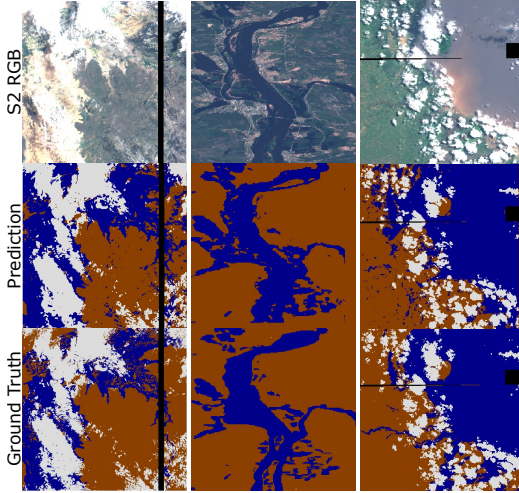
Figure 2: Segmentation results of degraded models run on Intel® Movidius™ Myriad™ 2. Black regions are mosaicing artifacts.
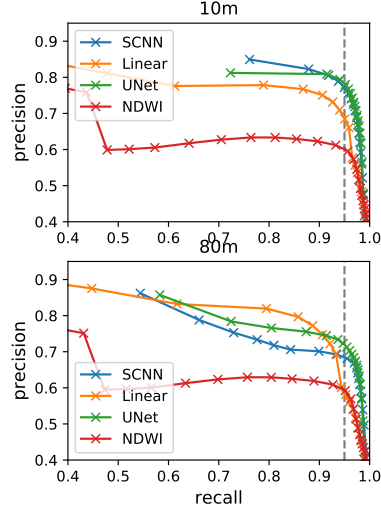

Figure 3: Precision-Recall curves of models trained on the S2 resolution ($10\,\mathrm{m}$) and HyperScout-2 resolution ($80\,\mathrm{m}$). In gray, 95% recall threshold.

Table 2: IoU and recall results for models trained on *WorldFloods*.

	NDWI (thresh 0)	NDWI (tuned)	Linear	SCNN	U-Net
IOU@$10\,\mathrm{m}$	38.37	58.63	63.34	67.54	**69.25**
IOU@$80\,\mathrm{m}$	37.74	58.09	57.25	65.42	**67.51**
Recall@$10\,\mathrm{m}$	43.7	93.26	96.24	**97.43**	97.34
Recall@$80\,\mathrm{m}$	43.13	92.72	95.24	**97.03**	96.56

5   Results

Table 2 shows the IoU and recall for the different models and baselines. Our three models (Linear, SCNN and UNet) all have a recall above 95% whereas the tuned NDWI has a recall of 93.5%; NDWI without tuning generalises poorly, we suspect due to muddy water. FCNN models performed best although there was only a small increase in performance between SCNN and U-Net, despite U-Net having 30x more parameters. The drop in performance from $10\,\mathrm{m}$ to $80\,\mathrm{m}$ is around 2 points for FCNN models which is acceptable taking into account that the spatial resolution is 8 times worse. Figure 3 shows the precision and recall for different thresholds on the water/flood class; again, our trained models beat NDWI and larger models tend to perform better.

The SCNN model was selected to be tested on the Myriad 2 chip due to its lower computational footprint compared to UNet (1FLOPS vs 2.68FLOPS for a 64x64x13 input). Figure 2 shows the images segmented using the Myriad 2. In general, the model over-predicts water content. False positives are mostly clustered in the surroundings of water bodies and in cloud shadows. This model segments a 12Mpx image approximately the size acquired by HyperScout-2 in less than one minute.

6   Conclusion

We have demonstrated that accurate flood segmentation is feasible to perform using low resolution images in orbit using available hardware. Our models outperform standard baselines and are favourably comparable to human annotation, while being efficiently computable on-board with current hardware. We are currently performing a more rigorous hyper-parameter search over a larger number of models and we hope to release both pipeline code and the *WorldFloods* dataset shortly, which we hope will serve as a useful tool to foster research in disaster response.

References

[1] Centre for Research on the Epidemiology of Disasters. The human cost of weather-related disasters 1995-2015. *United Nations Office for Disaster Risk Reduction*, 2015.

[2] S. B. Serpico, S. Dellepiane, G. Boni, G. Moser, E. Angiati, and R. Rudari. Information extraction from remote sensing images for flood monitoring and damage evaluation. *Proceedings of the IEEE*, 100(10):2946–2970, Oct 2012.

[3] Clemens Havas, Bernd Resch, Chiara Francalanci, Barbara Pernici, Gabriele Scalia, Jose Fernandez-Marquez, Tim Van Achte, Gunter Zeug, Maria Mondardini, Domenico Grandoni, et al. E2mc: improving emergency management service practice through social media and crowdsourcing analysis in near real time. *Sensors*, 17(12):2766, 2017.

[4] Michael Berger, Jose Moreno, Johnny A Johannessen, Pieternel F Levelt, and Ramon F Hanssen. ESA's Sentinel missions in support of Earth system science. *Remote Sensing of Environment*, 120:84–90, 2012.

[5] Matthias Drusch, Umberto Del Bello, Sébastien Carlier, Olivier Colin, Veronica Fernandez, Ferran Gascon, Bianca Hoersch, Claudia Isola, Paolo Laberinti, Philippe Martimort, et al. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. *Remote sensing of Environment*, 120:25–36, 2012.

[6] Hank Heidt, Jordi Puig-Suari, Augustus Moore, Shinichi Nakasuka, and Robert Twiggs. CubeSat: A new generation of picosatellite for education and industry low-cost space experimentation. In *14th Annual/USU Conference on Small Satellites*, 2000.

[7] M Esposito, SS Conticello, M Pastena, and B Carnicero Domínguez. In-orbit demonstration of artificial intelligence applied to hyperspectral and thermal sensing from space. In *CubeSats and SmallSats for Remote Sensing III*, volume 11131, page 111310C. International Society for Optics and Photonics, 2019.

[8] Adriano Camps, Alessandro Golkar, Antonio Gutierrez, JA Ruiz de Azua, Joan Francesc Munoz-Martin, Lara Fernandez, Carlos Diez, Andrea Aguilella, Simone Briatore, Rustam Akhtyamov, et al. FSSCAT, the 2017 Copernicus Masters'"ESA Sentinel Small Satellite Challenge" Winner: A Federated Polar and Soil Moisture Tandem Mission Based on 6U Cubesats. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 8285–8287. IEEE, 2018.

[9] Marco Esposito, Nathan Vercruyssen, Chris Dijk, Bernardo Carnicero Domíguez, Massimiliano Pastena, Philippe Martimort, and Pierluigi Silvestrin. Hyperscout 2 highly integration of hyperspectral and thermal infrared technologies for a miniaturized EO imager. In *Living Planet Symposium*, 2019.

[10] S. K. McFeeters. The use of the normalized difference water index (ndwi) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7):1425–1432, 1996.

[11] Akhtar Ali Memon, Sher Muhammad, Said Rahman, and Mateeul Haq. Flood monitoring and damage assessment using water indices: A case study of pakistan flood-2012. *The Egyptian Journal of Remote Sensing and Space Science*, 18(1):99 – 106, 2015.

[12] Tim Rudner, Marc Rußwurm, Jakub Fil, Ramona Pelich, Benjamin Bischke, and Veronika Kopackova. Multi3net: Segmenting flooded buildings via fusion of multiresolution, multisensor, and multitemporal satellite imagery. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33:702–709, 07 2019.

[13] Furkan Isikdogan, Alan C Bovik, and Paola Passalacqua. Surface water mapping by deep learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(11):4909–4918, 2017.

[14] 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.

[15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[16] Alberto Garcia-Garcia, Sergio Orts-Escolano, Sergiu Oprea, Victor Villena-Martinez, and Jose Garcia-Rodriguez. A review on deep learning techniques applied to semantic segmentation. *arXiv preprint arXiv:1704.06857*, 2017.

[17] Guy J-P Schumann. The need for scientific rigour and accountability in flood mapping to better support disaster response. *Hydrological Processes*, 1(5), 2019.

[18] Copernicus Emergency Management System. `https://emergency.copernicus.eu/`. Accessed: 2019-09-15.

[19] UNOSAT. `http://floods.unosat.org/geoportal/catalog/main/home.page`. Accessed: 2019-09-15.

[20] Global Flood Inundation Map Repository. `https://sdml.ua.edu/glofimr/`. Accessed: 2019-09-15.

[21] Stuart K. McFeeters. Using the Normalized Difference Water Index (NDWI) within a Geographic Information System to Detect Swimming Pools for Mosquito Abatement: A Practical Approach. *Remote Sensing*, 5(7):3544–3561, 2013.

[22] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 240–248. Springer, 2017.