POLITECNICO DI TORINO
Repository ISTITUZIONALE

AI-Based Flood Event Understanding and Quantifying Using Online Media and Satellite Data

Original
AI-Based Flood Event Understanding and Quantifying Using Online Media and Satellite Data / Zaffaroni, M., Lopez Fuentes, L., Farasin, A., Garza, P., Skinnemoen, H.. - ELETTRONICO. - 2670:(2019), pp. 1-3. (MediaEval 2019 Multimedia Benchmark Workshop EURECOM - Sophia Antipolis 27/10/2019).

Availability:
This version is available at: 11583/2846167 since: 2022-04-14T09:24:52Z

Publisher:
CEUR-WS

Published
DOI:

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
default_article_editorial [DA NON USARE]

-

(Article begins on next page)

02 July 2026

AI-based flood event understanding and quantification using online media and satellite data

Mirko Zaffaroni[1,4,*], Laura Lopez-Fuentes[2,5,*], Alessandro Farasin[3,4,*], Paolo Garza[3], Harald Skinnemoen[5]
[1] University of Turin, Italy; mirko.zaffaroni@unito.it
[2] University of the Balearic Islands, Spain; l.lopez@uib.es
[3] Politecnico di Torino, Italy; {name.surname}@polito.it
[4] LINKS Foundation, Italy; {name.surname}@linksfoundation.com
[5] AnsuR Technologies, Norway; {name}@ansur.no
* The authors contributed equally to this work.

ABSTRACT

In this paper we study the problem of flood detection and quantification using online media and satellite data. We present a three approaches, two of them based on neural networks and a third one based on the combination of different bands of satellite images. This work aims to detect floods and also give relevant information about the flood situation such as the water level and the extension of the flooded regions, as specified in the three subtasks, for which of them we propose a specific solution.

1 INTRODUCTION

The frequency and the intensity of natural disasters have risen significantly due to climate change. Flood events alone represent about the 39% of the natural disasters occurred worldwide. During this type of natural disasters it is important for emergency responders to have as much information as possible about the magnitude of the disaster, the areas affected and the situation and location of people in danger. In order to extract this information we consider two information sources: online news articles and satellite spectral imagery. Thanks to the rapid access to internet, online news contain information about natural disasters in almost real-time while satellite spectral imagery can give information of the extension of the flood. Using these two information sources, we propose approaches for flood event understanding and quantification:

- An algorithm that determines if an image extracted from an online news article contains relevant information about the flood. For example, images of the flood itself, but also images of emergency responders, people in danger, etc.
- An algorithm that given an image, extracted from online news, determines if there is water in the image and in case of containing water, if the water level is above or below the knee level of the people in the scene, if there are any. It contemplates also the use of news text as additional data for inference.
- An algorithm that given spectral imagery from satellites it segments the water regions of the images and gives a flood/no flood prediction and an estimation of the flood extension.

Copyright 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
MediaEval'19, 27-29 October 2019, Sophia Antipolis, France

This work has been done in the context of MediaEval 2019, as a participation in the Multimedia Satellite task. Detailed information about the task and data can be found in [3].

2 RELATED WORK

Emergency prevention, detection, assistance and understanding through computer vision and image processing techniques is an open problem since the early stages of this field [13]. In particular, in the flood detection domain scientific work mostly focuses on flood detection either in social media or satellite data [1, 2, 4]. Among the latest, several approaches are known in the literature and exploit spectral bands and other sensor measurements [9, 14–16, 19] to retrieve proper indicators.

This work builds on top of our Multi-modal deep learning approach for flood detection [12], which used social media images together with their metadata to determine if a social media post contained visual information about a flood; and our deep learning models for passability detection of flooded roads [5, 11], which went a step further and gave information about the state of the roads during a flood event, information that is of utmost importance during a flood in order to build a map of accessible roads for rescue and supply operations. Moving in this direction, in this paper we aim at giving an estimation of the water level.

3 APPROACH

In this section, each stage of the solution will be briefly introduced.

News Image Topic Disambiguation (NITD). During a flooding the media normally updates the information about the situation to keep the reader updated. Due to the large amount of online newspapers and media, searching for these relevant articles can be time consuming. To optimize the search it is possible to use natural language processing (NLP) algorithms or keyword searches. Since most of these articles contain images, in this first stage, we want to refine the search using a computer vision algorithm to classify those images in flood event related/not flood event related.

In order to train the classifier we use the training set for this task that is composed by 5145 images which have been retrieved from online news as containing information about a flood by an NLP or keyword algorithm and then manually classified. As for the algorithm, we use an ensemble of 4 state-of-the-art networks (InceptionV3 [18], MobileNet [10], VGG16 and VGG19 [17]) and

CEUR-WS.org/Vol-2670/MediaEval_19_paper_47.pdf

cross-validation using two folds. Since the dataset is highly imbalanced we balance the dataset during training by randomly undersampling the negative class for each epoch. This way the dataset stays balanced but we use all the samples from both categories. Finally, we combine the networks by majority voting.

Multimodal Flood Level Estimation (MFLE). Given online articles with visual and textual information we developed a textual, a textual-visual and a only visual model to estimate the flood level by predicting if the water is above or below the knee of the people in the scene. The latter model is composed by two branches: (i) it takes as input image crops of person's knees extracted by a state of the art pose estimator [6], and predicts if the knee is under or above the water; (ii) it takes as input a full image of the scene, and predicts if the image has people with knees underwater.

To create the training data for the first branch we used the pose estimator algorithm to extract a region around all the knees from the training set. The knees from images which were labelled as 0, water below the knee, were labelled as 0 by default, while the ones belonging to images labelled as 1 were manually labelled, since there could be people in the same image with water level above or below the knee. Both networks use a VGG19 [17] pre-trained on ImageNet [7] to extract deep features of the images, followed by a fully-connected (FC). Then the information is concatenated to combine the semantic features of the knee with the context information provided by the full resolution image. This way the first branch gets information about the context while the second branch gets information about the knees. Finally, a FC estimates if the knee is above or below the water and another FC if the water is above or below the knee level. The two branch system is propose because a simple one-branch Convolution Neural Network (CNN) would greedily learn to predict flooded images as "water above the knee" class, since it lacks specific data about the knees in the scene and so it would associate the features of a flooded area as "water above the knee" class, because it solely composed by these examples.

Finally, an image is classified as "water above the knee" if there is at least one knee in the scene that is classified as "water above the knee" by the knee branch and the context branch also classified the image as "water above the knee". We also combined textual data of the articles to verify if it could lead to a better predictor. This was achieved by building an ensemble composed by the previous model and an NLP module. This module is composed by a bidirectional Long Short-Term Memory (LSTM) network. The result of the LSTM is concatenated to the last FC layer of the image classifier. The only textual model is composed by the module described above alone.

City-centered Satellite Sequences (CCSS). Given a sequence of Sentinel-2 satellite images that depict a certain city over a certain length of time, this task aims to classify whether there was a flooding event ongoing in that city at that time.

We built an *expert system* which leverages on both the spectral and the related metadata information. Firstly, it computes a binary mask for each layer, in which white pixels represent areas with presence of water, while black pixels represent the other regions. The binary masks are obtained: (i) by computing, for each pixel, the Modified Normalized Difference Water Index (MNDWI) [8] adapted for Sentinel-2 bands (S2), according to Equation (1); (ii) by setting to white the pixels having $MNDWI_{S2} \geq 0$, black the others.

$$MNDWI = \frac{\rho_{green} - \rho_{swir1}}{\rho_{green} + \rho_{swir1}},\ MNDWI_{S2} = \frac{B03 - B11}{B03 + B11} \quad (1)$$

Assuming that the dataset does not have missing values lasting for the whole time serie, we set the pixels related to uncovered areas to white. Then, we performed the pixel-wise intersection among two sets of layers: (i) the computed binary layers marked as FLOODED and (ii) the ones marked as NON-FLOODED in the metadata file.

The two images depict the water persistence in case of flood and non-flood. Finally, to discriminate flooded regions from normal water-sources (like rivers or lakes) a pixel-wise difference among the two sets is computed. Even if a binary mask representing the residual flood extent is available, to be compliant with the *CCSS* subtask, the approach returns 1 if there is still any white region in the resulting binary mask, 0 otherwise.

4 RESULTS

The results, split by subtask, are reported in Table 1. For the subtasks *NITD* and *MFLE*, the F1-Scores are referred to the 20 % of the development set, used as validation set. Conversely, being the *CCSS* proposed approach an expert system, the whole devset was used. In this latest substask, the confusion matrix on the devset, TP:108, FP:0, FN:33, TN:127, shows that the approach is strong against false positives, having a precision of 1.0.

Table 1: Results per subtask

Subtask	Data	DevSet F-Score	TestSet F-Score
NITD	Visual	0.8062	0.6628
MFLE	Visual	0.7667	0.5428
	Text	0.5213	0.4956
	Visual & Text	0.5454	0.5284
CCSS	Satellite	0.8850	0.9118

5 ANALYSIS AND CONCLUSIONS

Conclusions present our insight on the subtasks. **(NITD)** Balancing the dataset during training and combining different models significantly improves the perfomance. **(MFLE)** (i) Merging global and local classifiers improves the performance; (ii) the text brings some information, but the approach gives better results processing only images; (iii) people water reflection degrades the performance of pose estimation algorithm. (iv) the importance of the two branches is supplied by an ablation study in which the two branch model achieved 0.79 F1-score on validation, while the full image branch alone achieved 0.71 and the branch using the cropped knees achieved 0.76. **(CCSS)** (i) B03 and B11 are highly informative for water segmentation; (ii) the approach is an expert system, therefore there is no need of a training set and it is computationally fast;

ACKNOWLEDGMENTS

This work was supported by the European Commission H2020 SHELTER project, GA no. 821282 and by the Spanish grant TIN2016-75404-P. Laura Lopez-Fuentes benefits from the NAERINGSPHD fellowship of the Norwegian Research Council under the collaboration agreement Ref.3114 with the UIB.

REFERENCES

[1] K. Avgerinakis, A. Moumtzidou, S. Andreadis, E. Michail, I. Gialampoukidis, S. Vrochidis, and I. Kompatsiaris. Visual and Textual Analysis of Social Media and Satellite Images for Flood Detection@ Multimedia Satellite Task MediaEval 2017. In *Proc. of the MediaEval 2017 Workshop* (Sept. 13-15, 2017). Dublin, Ireland.

[2] B. Bischke, P. Bhardwaj, A. Gautam, P. Helber, D. Borth, and A. Dengel. Detection of Flooding Events in Social Multimedia and Satellite Imagery using Deep Neural Networks. In *Proc. of the MediaEval 2017 Workshop* (Sept. 13-15, 2017). Dublin, Ireland.

[3] B. Bischke, P. Helber, S. Brugman, E. Basar, Zx Zhao, M. Larson, and K. Pogorelov. The Multimedia Satellite Task at MediaEval 2019: Estimation of Flood Severity. In *Proc. of the MediaEval 2019 Workshop* (Oct. 27-29, 2019). Sophia Antipolis, France.

[4] B. Bischke, P. Helber, Z. Zhao, J. De Bruijn, and D. Borth. The Multimedia Satellite Task at MediaEval 2018. In *Proc. of the MediaEval 2018 Workshop* (Oct. 29-31, 2018). Sophia Antipolis, France.

[5] B. Bischke, P. Helber, Z. Zhao, J. de Bruijn, and D. Borth. The Multimedia Satellite Task at MediaEval 2018: Emergency Response for Flooding Events. In *Proc. of the MediaEval 2018 Workshop* (Oct. 29-31, 2018). Sophia-Antipolis, France.

[6] Z. Cao, T. Simon, S. Wei, and Y. Sheikh. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 7291–7299.

[7] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*.

[8] G. Donchyts, J. Schellekens, H. Winsemius, E. Eisemann, and N. van de Giesen. 2016. A 30 m resolution surface water mask including estimation of positional and thematic differences using landsat 8, srtm and openstreetmap: a case study in the Murray-Darling Basin, Australia. *Remote Sensing* 8, 5 (2016), 386.

[9] A. Farasin and P. Garza. 2018. PERCEIVE: Precipitation Data Characterization by means on Frequent Spatio-Temporal Sequences. In *ISCRAM*.

[10] A. G Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. (2017).

[11] L. Lopez-Fuentes, A. Farasin, H. Skinnemoen, and P. Garza. Deep Learning Models for Passability Detection of Flooded Roads. In *Proc. of the MediaEval 2018 Workshop* (Oct. 29-31, 2018). Sophia Antipolis, France.

[12] L. Lopez-Fuentes, J. van de Weijer, M. Bolanos, and H. Skinnemoen. Multi-modal Deep Learning Approach for Flood Detection. In *Proc. of the MediaEval 2017 Workshop* (Sept. 13-15, 2017). Dublin, Ireland.

[13] L. Lopez-Fuentes, J. van de Weijer, M. González-Hidalgo, H. Skinnemoen, and A. D. Bagdanov. 2018. Review on computer vision techniques in emergency situations. *Multimedia Tools and Applications* 77, 13 (2018), 17069–17107.

[14] K. Osumi. 2019. Detecting land cover change using Sentinel-2. *Abstracts of the ICA* 1 (2019).

[15] S. Qiu, Z. Zhu, and B. He. 2019. Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. *Remote Sensing of Environment* 231 (2019), 111205.

[16] C. Rossi, F. S. Acerbo, K. Ylinen, I. Juga, P. Nurmi, A. Bosca, F. Tarasconi, M. Cristoforetti, and A. Alikadic. 2018. Early detection and information extraction for weather-induced floods using social media streams. *International journal of disaster risk reduction* 30 (2018), 145–157.

[17] K. Simonyan and A. Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ICLR* (2015).

[18] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2818–2826.

[19] Z. Zhu and C. E. Woodcock. 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote sensing of environment* 118 (2012), 83–94.