Flood Level Estimation from Social Media Images

Julia Strebl[1], Djordje Slijepcevic[1],
Armin Kirchknopf[1], Muntaha Sakeena[1], Markus Seidl[1], Matthias Zeppelzauer[1]
[1]St. Pölten University of Applied Sciences, Austria
{firstname.lastname}@fhstp.ac.at

ABSTRACT

In this paper, we present an approach and first results for the Media-Eval 2019 sub-task on "Multimodal Flood Level Estimation from News" from the "2019 Multimedia Satellite Task". The water level is measured by detecting people standing in water and using the human body as a reference. We focus only on the visual modality and propose a combination of a ResNet-based water detector and pose estimation to solve the task. First results are promising and show that our approach is clearly performing above baseline.

1 INTRODUCTION

The assessment of natural disasters by automated media analysis becomes increasingly important as, on one hand, the amount of user-generated media has been rising with the availability of smart phones and, on the other hand, the likelihood of disasters increases e.g. due to the ongoing climate change. The availability of (social) media data represents an opportunity to automatically detect and assess disasters to better guide first responders and emergency forces. The types of disasters targeted in this work are floods and the particular task to solve is flood level estimation [6]. The task has been formulated in the course of the "2019 Multimedia Satellite Task: Estimation of Flood Severity" conducted in the MediaEval 2019 benchmarking initiative [1]. This paper presents our contribution to the benchmark together with results on the benchmarking test set. The task of flood level estimation is defined as follows: "build a binary classifier that predicts whether or not the image contains at least one person standing in water above the knee"[1]. Input to the classifier can be visual data, textual data or both. The data stems from online news articles and comprises 6172 articles, whereas 1234 articles belong to the test set and 4932 articles to the training set (6 articles, i.e., 598, 3932, 4465, 5019, 5091, and 5419, were excluded due to corrupt image files). There is one image per article. The test labels were not available during development. A major challenge of this task is the strong imbalance between the positive class (people standing in water above the knee) and the negative class. The textual data was only partly available through the links provided by the organizers. For this reason our approach focuses only on the visual domain. Our results show that the approach is clearly above the random baseline and has a good generalization ability.

2 RELATED WORK

Disaster recognition based on social media images has been a rising topic recently. Flood level estimation is technically challenging as

---

[1]http://www.multimediaeval.org/mediaeval2019/multimediasatellite, 01.10.2019

Figure 1: Flood level estimation with our approach using the OpenPose tracker [3].

shown by Hostache et al. [4] and Zwenzner et al. [9]. Both authors proposed a combined method based on SAR images and a Digital Elevation Model (DEM). By using crowd-sourced, non-authoritatively collected data, Schnebele et al. [8], proposed a method to detect flood events on road infrastructure in the US. Pandey et al. [7] used different data modalities, such as MODIS images and TRMM precipitation to detect floodings after a dam breach and could estimate a rise of flood level by 1.0 to 1.4m. Related research as mentioned above fused several information sources, such as aerial images and DEM models to estimate flood levels. The estimation of flood levels from RGB data is a challenging task, as the visual appearance of water varies strongly.

3 APPROACH

Since the provided dataset is multimodal, our initial idea consisted of training two different classifiers for the image data and a separate one for the text data and then fusing the predictions. Due to insufficient text data we decided to provide predictions based only on visual data. We developed two classification approaches. The first approach (see Figure 2a) relies on detecting water within the whole image and detecting at least one person with obscured lower body parts. The second approach (see Figure 2b) performs local water detection. To this end, for each human body detected, a patch that also contains the local neighbourhood of the human body is taken into account for water detection. If at least for one patch in the image our model detects obscured lower extremities and water in the vicinity, the image is assigned to the positive class. Both proposed approaches build upon three main components: (i) a water detector that predicts whether a certain image or image region contains water, (ii) a pose estimator that detects people and fits skeletons into their bodies and (iii) a rule-based fusion module that combines the information from the water detector and the pose estimator to make a final decision.

Water detector: we build upon ResNet50, which is pre-trained on ImageNet and fine-tuned for water/no water detection using images showing either water or not. Images are resized, using nearest-neighbor interpolation, to the network's input size (227x227) while

(a) Approach 1: Global water detection (Run 1 and 4).

(b) Approach 2: Local water detection (Run 5).
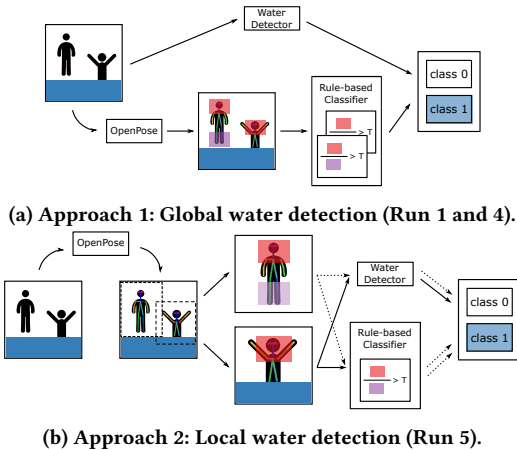
Figure 2: Overview of two water detection approaches.

keeping the original aspect ratio. Horizontal flipping, brightness variations and non-uniform re-scaling of the images are applied for data augmentation. The top five layers are fine-tuned (for 6 epochs, batch size 256) before the whole network (for 10 epochs, batch size 32) is trained using the Adam optimizer (learning rate of 10^{-4}).

Pose estimator (OpenPose): we employ OpenPose [3] to detect body key points from depicted human bodies. To filter out false positive detections and unreliable skeletons, we calculate a confidence score (C_U) from the two most robust upper body parts, i.e. head and chest (OpenPose joint IDs 0 and 1). Only skeletons with an empirically estimated threshold of $C_U > 0.6$ are further considered. To detect whether the lower extremities of a body are visible, we calculate a second confidence score (C_L) as the mean confidence over the lower body parts (OpenPose joint IDs 9, 10, 12, and 13). Note that for missing body parts the confidence is zero.

Rule-based Classifier: to determine whether the lower extremities of a detected skeleton are visible we employ the following heuristic rule: $C_U / \max(C_L, 10^{-4}) > T$, with C_U and C_L being the mean detection confidence for the upper and lower body and T an empirically determined threshold of 1.5. The max operator prevents division by zero.

Final decision rule: A positive detection of a person standing in water is declared when both the rule-based classifier and the (local or global) water detector predict positively.

4 EXPERIMENTAL RESULTS

We train our models on 80% of the training data and use 20% from each class for validation (randomly selected). For Run 1, we use only the data provided by the organizers, but we manually label the images regarding whether or not they contain water. In all other runs, we further use the data from the Multimedia Satellite Task from 2018 [2] (Task: Flood classification for social multimedia; manually labeled to water/no water) to train the water detector [5]. For Run 1 and 4, we employ the classification pipeline depicted in Figure 2a. In Run 5 we evaluate the local approach from Figure 2b. Run 3 was reserved for a multimodal run combining text and image data, which we could not submit due to large amounts of inaccessible text data. Therefore, for Run 3 we perform a majority voting

over the predictions of Runs 1, 4 and 5. Results for the validation and test set are presented in Table 1.

Table 1: Macro-averaged Precision (P), Recall (R), and F1-scores for *Multimodal Flood Level Estimation from News.*

Run	Validation (P	R	F1)		Test (F1)
1	0.58	0.67	0.61	0.61
3	0.59	0.68	0.61	0.61
4	0.55	0.60	0.56	0.59
5	0.58	0.77	0.60	0.59

The performance of our approach is almost the same on our validation and the benchmark test set, which shows that our approach generalizes well. The overall performance is around 60% for all runs and does not show a significant difference between the local and global approaches. Similarly, the fusion of both (Run 3) does not outperform our baseline (Run 1). The random baseline for this task depends on the class cardinalities in the test set and is thus unknown to the authors. It has, however, an upper limit of 50% for the task due to the use of the macro averaged class-wise F1-scores as performance measure. Our approach outperforms this baseline, which shows that it learns useful patterns related to the target task, although there is room for improvement. A closer analysis of the results shows several directions for improvement. While the water detection is quite robust (classification accuracy of 0.88; model is trained on data from last year's task as well as 700 images from this year's task and evaluated on 200 images from this year's task), we observe numerous false and missed detections of OpenPose. Furthermore, reflections of the human body on the water surface represent a problem, i.e. for the detection of lower extremities. In several cases, OpenPose added body parts for the lower extremities, which were actually under water (see right image in Figure 1).

5 DISCUSSION AND OUTLOOK

In this paper, we presented our contribution to the MediaEval 2019 task on flood level estimation from news media images. Our approach combines a pose detector and a water detector to find images showing people standing in water above the knee. First results show a promising generalization ability. Concerning the overall performance, improvements are possible. A promising approach to increase robustness is the use of several human (pose) detectors trained on different data (e.g. urban and rural setting). A limitation of our approach is that not only water but also other objects can obscure the lower extremities of a person or that only torso or head are shown in the picture. As a result, the lower extremities cannot be detected and if water is present, our approach may fail. In order to compensate for these effects, pixel-wise segmentation of water and humans could be advantageous. Additionally, pixel accurate data could help to detect false detections by OpenPose, e.g. detected body parts protruding out of the segmented area, which represents a human body, should not be considered.

REFERENCES

[1] Benjamin Bischke, Patrick Helber, Simon Brugman, Erkan Basar, Zhengyu Zhao, Martha Larson, and Konstantin Pogorelov. Oct. 27-29, 2019. The Multimedia Satellite Task at MediaEval 2019: Estimation of Flood Severity. In *Proc. of the MediaEval 2019 Workshop*. Sophia Antipolis, France.

[2] Benjamin Bischke, Patrick Helber, Zhengyu Zhao, Jens De Bruijn, and Damian Borth. 2018. The multimedia satellite task at MediaEval 2018: Emergency response for flooding events. In *2018 Working Notes Proceedings of the MediaEval Workshop, MediaEval 2018*. CEUR-WS. org, 1–3.

[3] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Re-altime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 7291–7299.

[4] Renaud Hostache, Patrick Matgen, Guy Schumann, Christian Puech, Lucien Hoffmann, and Laurent Pfister. 2009. Water level estimation and reduction of hydraulic model calibration uncertainties using satellite SAR images of floods. *IEEE Transactions on Geoscience and Remote Sensing* 47, 2 (2009), 431–441.

[5] Armin Kirchknopf, Djordje Slijepcevic, Matthias Zeppelzauer, and Markus Seidl. 2018. Detection of Road Passability from Social Media and Satellite Images.. In *MediaEval*.

[6] Victor Klemas. 2014. Remote sensing of floods and flood-prone areas: an overview. *Journal of Coastal Research* 31, 4 (2014), 1005–1013.

[7] Rajesh Kumar Pandey, Jean-François Crétaux, Muriel Bergé-Nguyen, Virendra Mani Tiwari, Vanessa Drolon, Fabrice Papa, and Stephane Calmant. 2014. Water level estimation by remote sensing for the 2008 flooding of the Kosi River. *International journal of remote sensing* 35, 2 (2014), 424–440.

[8] E. Schnebele, G. Cervone, and N. Waters. 2018. Road assessment after flood events using non-authoritative data. *Nat. Hazards Earth Syst. Sci.* 14, 4 (2018), 1007–1015. https://doi.org/10.5194/nhess-14-1007-2014

[9] H. Zwenzner and S. Voigt. 2009. Improved estimation of flood parameters by combining space based SAR data with very high resolution digital elevation data. *Hydrology and Earth System Sciences* 13, 5 (2009), 567.