Multimedia Analysis Techniques for Flood Detection Using Images, Articles and Satellite Imagery

Stelios Andreadis[1], Marios Bakratsas[1], Panagiotis Giannakeris[1], Anastasia Moumtzidou[1], Ilias Gialampoukidis[1], Stefanos Vrochidis[1], Ioannis Kompatsiaris[1]
[1]Centre for Research & Technology Hellas - Information Technologies Institute, Greece
{andreadisst,mbakratsas,giannakeris,moumtzid,heliasgj,stefanos,ikom}@iti.gr

ABSTRACT

This paper presents the various algorithms that the CERTH-ITI team has implemented to tackle three tasks that relate to the problem of flood severity estimation, using satellite images and online media content. Deep Convolutional Neural Networks were deployed to classify articles as flood event-related based on their images, but also to detect flooding events in satellite sequences. Remote sensing indices play a key role in the machine learning approach to identify changes between satellite imagery, while visual and textual features were exploited to estimate whether an image shows people standing in flooded areas.

1 INTRODUCTION

News websites now play a crucial role in the field of public information, turning into a rich and open source of articles and images that cover numerous events. At the same time, the high availability of satellite data induces an alternative source of imagery. This data can be exploited in the domain of natural disasters, e.g. to detect a flooding incident or to estimate the severity of a flood. Several ongoing H2020 projects follow this direction: beAWARE [3] includes the analysis of visual and textual information for disaster forecasting and management, while EOPEN [10] involves Earth Observation and social media data in flood risk monitoring.

The Multimedia Satellite Task is a challenge of MediaEval that consists of the following subtasks. News Image Topic Disambiguation (NITD) entails an image classifier that is able to identify whether or not an image belongs to a flood-related article. Multimodal Flood Level Estimation (MFLE) calls for a classifier that receives visual and/or textual information from articles and predicts whether or not an image contains people standing in water above the knee. Finally, City-Centered Satellite Sequences (CCSS) asks participants to detect a flooding incident by using sequences of satellite images. For further details on the subtasks and the respective data sets, the reader is referred to [1].

The next section presents the algorithms proposed by the CERTH-ITI team for each subtask, followed by the results of their evaluation and a short discussion with conclusions.

2 APPROACH

2.1 News Image Topic Disambiguation (NITD)

We aim to classify news articles' topics judging from the images that appear in them. One challenge of this task is that inside the images flooded areas may be completely out of view. Even more challenging are the instances where a flooded area is clearly shown in the image but the article's topic is not relevant to a flood event. Also, in some instances water is present but not in the context of floods (e.g. a beach). In order to examine the performance of state-of-the-art image classification techniques [11] in this task we deploy a Deep Convolutional Neural Network (DCNN) that was trained on the full development set ("CNN2019"). Another DCNN that was trained on the Mediaeval 2017 development set ("CNN2017") [2] is also tested here in order to evaluate a straight flood/non-flood image classifier and compare both approaches.

We acquire the VGG architecture pre-trained on the Places365 dataset [13] for both cases. The weights of this model are carefully optimized to extract features for scene recognition which is a suitable starting point for our objective [8]. In order to fine-tune the network, 5-fold cross-validation was performed so as to find how many of the final layers to freeze and at which epoch to stop the training. The setting with the highest average accuracy was fine-tuning all fully-connected layers for 35 epochs. The development set is heavily biased towards negative samples (nearly 7 times more negative images), therefore we chose to oversample the set with positive images to balance it.

2.2 Multimodal Flood Level Estimation (MFLE)

The estimation of flood level involves checking whether or not an image contains people standing in water above the knee, and it is realized by considering machine learning techniques on visual and textual information. Regarding the visual information, a 22-layer GoogleNet network was fine-tuned and the dimension of the classification layer was set equal to 345 [9], which equals to the 345 SIN TRECVID concepts. Then a set of 6 concepts were considered as interesting for locating people (being "Adult", "Person", "Two_People") and water ("River", "Waterscape_Waterfront"). The probabilities of each concept appearing in each image were considered as input to a binary Support Vector Machine (SVM) classifier.

Regarding the textual information, we followed a well-established approach in text classification called word2vec [7] that considers word embeddings. In general, word embeddings stand on the concept that similar words tend to occur together and have a similar context (e.g. football and basketball are linked to sports) and they are based on Deep Neural Networks (DNN) [4]. Eventually, a binary SVM classifier is trained using the word2vec text representations.

Finally, a simple late fusion approach was followed in order to consider both visual and textual information, so the outputs of the above two modules are considered for deciding the fused approach prediction. If the output of the two SVM binary classifiers coincide,

MediaEval'19, 27-29 October 2019, Sophia Antipolis, France

then their common label defines the label of the fused module; otherwise, only the output of the visual module is considered.

2.3 City-Centered Satellite Sequences (CCSS)

The first approach to detect flood events using satellite sequences involved the use of a deep learning model which was trained on two different datasets of three-channel images with the differences of two days within an event. The first dataset was created by combining the Red-Green-Blue (B02-B03-B04) bands and the second by combining the Red-Swir-Nir (B02-B03-B04). Then, the three bands were stacked and converted to JPEG. Within each event, the unique differences between its days were calculated. Next, pre-trained networks on ImageNet [6] were fine-tuned in order to learn the new features of our dataset. The last pooling layer was replaced with a densely-connected NN layer with a softmax activation function with 2 outputs. The following parameters were considered: (i) evaluation of the Adam [5] and SGD optimizers, and (ii) evaluation of learning rates 0.1, 0.01, 0.001. Batch size was set to 32.

An additional change detection approach based on the remote sensing water index of MNDWI [12] was implemented. Within each event the MNDWI differences of the consecutive days were calculated. For each difference image, the outliers were estimated as follows: pixel's values that fall within $\left[m - \gamma\sigma, m + \gamma\sigma\right]$ denotes no change. A minimum *water_ratio* needs to be set to characterize the image as changed (i.e. flooded). The method was applied on the dev set to identify the optimum values for *gamma* and *water_ratio*.

As a third approach, outlier detection was also performed on water body masks, produced by zero thresholding of the MNDWI index. Counting the water pixels of each day of an event generated time series of integers. Then, Z-score was calculated per each point as $(x - m) \,/\, \sigma$, where x is the value of each point and m and σ are the mean and standard deviation of all points in the time series. If a point exceeded a threshold γ, it was considered an outlier and thus the complete sequence of images was classified as an event.

3 RESULTS AND ANALYSIS

The complete results in the dev set and the test set for all three subtasks of the Multimedia Satellite Task can be seen in Table 1, where it is evident that the DCNN approaches in NITD and the image differencing technique in CCSS really stood out. In detail:

NITD Examining the errors, we observe that the article classifier is mainly producing False Positives and very little to none False Negatives. Many of the FP cases actually show flooded areas, although the article topic is negative to a flood event. On the test set, the 2019 model performs better than the 2017 model reaching an accuracy of 90.2%. We hypothesise that it is performing better because it has learnt correlations beyond the obvious: a flooded area in an image is a strong sign of flood relevancy in the article but certain groups of people appearing may also be a positive flag, like authorities or politicians. This is expected to hold true, especially if the training and the test set are taken from a single event where the same people appear frequently on the news articles.

MFLE The exploitation of visual information reaches a ~65% F1-score, due to the significant number of FP, since the concept detection focused on the identification of humans and water and it didn't restrict to images of people standing in water above the knee. The textual information features performed slightly lower to the visual ones, while the fusion of visual and textual features performed equally to the visual, which can be easily explained by the aforementioned description of the approach.

Table 1: CERTH-ITI results in all tasks

Run submissions	Dev set F1-Score (%)	Test set F1-Score (%)
News Image Topic Disambiguation		
CNN2019	95.98	90.20
CNN2017	78.46	88.73
Multimodal Flood Level Estimation		
Visual	47.3	64.33
Textual	35.2	57.62
Visual & Textual	47.3	64.33
City-centered satellite sequences		
MNDWI γ=2.1 ratio=0.05	83.54	76.47
VGG16 Red-Green-Blue	60.57	70.58
VGG16 Red-Swir-Nir	60.74	70.58
VGG19 Red-Swir-Nir	60.74	70.58
MNDWI water masks γ=2	57.53	54.41

CCSS Detecting the outliers on the differences of MNDWI consecutive images achieved a 76.47% F1-score. The image differencing technique proved adequate to detect changes relative to flood events, using the σ and minimum *water_ratio* values that were calculated on the annotated dev set. Using DCNN provided decent results (70.58%), showing its ability to learn to detect flood patterns even with a small training set. On the other hand, outlier detection on water masks, using MNDWI index and setting γ to 2, did not accomplish a high F1-score (54.41%), possibly due to the fact that all the remote sensing information was limited to a binary mask.

4 DISCUSSION AND OUTLOOK

Through the participation in the Multimedia Satellite challenge, the CERTH-ITI team gained the opportunity to examine various methodologies for the problem of flood detection. Results for the **NITD** task indicate that it is possible to classify flood event articles with good accuracy using either a generic flood detector or by annotating a specific dataset. However, the second approach looks more promising when dealing with articles concerning a single event. Results of the **MFLE** task show that visual features perform better than the textual ones, but they could be further improved if a segmentation step was applied on top of the proposed approach for recognising whether water covered people below the knee. Finally, results of the **CCSS** demonstrate the ability of the combined method of image differencing and water relative index of MNDWI to detect flood events, showing better robustness with balanced FP and FN rates, compared to the DCNN approach , whereas the three extra layers of VGG19 don't show any impact on the learning process.

ACKNOWLEDGMENTS

This work was supported by EC-funded projects H2020-700475-beAWARE and H2020-776019-EOPEN.

REFERENCES

[1] Benjamin Bischke, Patrick Helber, Simon Brugman, Erkan Basar, Zhengyu Zhao, Martha Larson, and Konstantin 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.

[2] Benjamin Bischke, Patrick Helber, Christian Schulze, Srinivasan Venkat, Andreas Dengel, and Damian Borth. The Multimedia Satellite Task at MediaEval 2017: Emergence Response for Flooding Events. In *Proc. of the MediaEval 2017 Workshop* (Sept. 13-15, 2017). Dublin, Ireland.

[3] H2020 DRS. 2017-2020. *beAWARE project.* https://beaware-project.eu/

[4] Moshe Hazoom. 2018. Word2Vec For Phrases. Learning Embeddings For More Than One Word. (2018). https://bit.ly/32mDMNH

[5] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. 1097–1105.

[7] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*. 3111–3119.

[8] Anastasia Moumtzidou, Panagiotis Giannakeris, Stelios Andreadis, Athanasios Mavropoulos, Georgios Meditskos, Ilias Gialampoukidis, Konstantinos Avgerinakis, Stefanos Vrochidis, and Ioannis Kompatsiaris. 2018. A Multimodal Approach in Estimating Road Passability Through a Flooded Area Using Social Media and Satellite Images.. In *Proc. of the MediaEval*.

[9] Nikiforos Pittaras, Foteini Markatopoulou, Vasileios Mezaris, and Ioannis Patras. 2017. Comparison of fine-tuning and extension strategies for deep convolutional neural networks. In *International Conference on Multimedia Modeling*. Springer, 102–114.

[10] H2020 EO RIA. 2017-2020. *EOPEN project.* https://eopen-project.eu/

[11] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations*.

[12] Hanqiu Xu. 2006. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International journal of remote sensing* 27, 14 (2006), 3025–3033.

[13] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. 2017. Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence* 40, 6 (2017), 1452–1464.