The Multimedia Satellite Task at MediaEval 2018

Emergency Response for Flooding Events

Benjamin Bischke[1, 2], Patrick Helber[1, 2], Zhengyu Zhao[3], Jens de Bruijn[4], Damian Borth[1]
[1] German Research Center for Artificial Intelligence (DFKI), Germany
[2] TU Kaiserslautern, Germany
[3] Radboud University, The Netherlands
[4] VU University Amsterdam, The Netherlands

ABSTRACT

This paper provides a description of the MediaEval 2018 Multimedia Satellite Task. The primary goal of the task is to extract and fuse content associated with events represent in Satellite Imagery and Social Media. Establishing a link from Satellite Imagery to Social Multimedia can yield to a comprehensive event representation which is vital for numerous applications. Focusing on natural disaster events, the main objective of the task is to leverage the combined event representation within the context of emergency response and environmental monitoring. In particular, our task focuses on flooding events and consists of two subtasks. The first *Image Classification from Social Media* subtask requires participants to retrieve images from Social Media that show a direct evidence for road passability during flooding events. The second task *Flood Detection from Satellite Images* aims to extract potentially flooded road sections from satellite images. The task seeks to go beyond state-of-the-art flooding map generation by focusing on information about road passability and the accessibility of urban infrastructure. Such information shows a clear potential to complement information from social images with satellite imagery for emergency management.

1 INTRODUCTION

Recent advances in Earth observation and the access to satellite imagery at a large scale are opening up a new exciting area for the applications of remotely sensed data. A proper analysis of this data source has potential to change how agriculture, urbanization and environmental monitoring will be done in the future. Hand in hand with this development, the Multimedia Satellite Task at MediaEval 2018 addresses natural disaster and environmental monitoring, allowing to improve situational awareness for such events.

One challenge when solely relying on remotely sensed data is the sparsity problem of satellite imagery over time, which often results in a poor event representation. The larger goal of this task is therefore to combine the satellite view with the ground-level perspective represented by images in social media streams in order to obtain a comprehensive picture of disaster events. Such a multi-modal event representation from social media and satellite imagery is of vital importance to achieve situational awareness and to provide support in emergency response, e.g., helping to coordinate rescuer efforts in large scale disasters. It is also important for studying disasters after they have happened, and support planning that will prevent or mitigate the impact of future disasters.

MediaEval'18, 29-31 October 2018, Sophia Antipolis, France

The Multimedia Satellite Task 2018 continues to focus on flooding events as in last year's Task 2017 [2], since, among high-impact natural disasters, flooding events represent according to the United Nations Office for the Coordination of Humanitarian Affairs[1] the most common type of disaster worldwide. This year the task will look at passability, namely whether or not it is possible to travel through a flooded region. Rapid information about road passability and the accessibility of the urban infrastructure is a critical aspect in emergency response. Additionally, passability of roads is also an area in which the information in social images has clear potential to complement the information in satellite images.

2 TASK DETAILS

The main objective of this year's task is to quantify the impact of flooding events on infrastructure. The task involves two subtasks:

Flood Classification from Social Multimedia.

The goal of the first subtask is to retrieve all images from social media that provide direct evidence for passability of roads by conventional means (no boats, off-the-road vehicles, monster trucks, Hummer, Landrover, farm equipment). The objective is to design a system/algorithm/method that (in principle) given any collection of flood related multimedia images and their metadata (e.g., Twitter, Flickr, YFCC100M) is able to identify those images that (1) provide evidence for road passability and (2) discriminate between images showing passable vs. non passable roads. In our context, road passability is related to the water level visible in the image and the surrounding context. Participants are allowed to submit 5 runs:

- Required run 1: using visual data only
- General run 2, 3, 4, 5: everything automated allowed, including using data from external sources (e.g. Twitter, Flickr)

Flood Detection from Satellite Imagery.

Participants receive high resolution satellite imagery for areas in Houston, that have been partially flooded during the hurricane event *Harvey* in 2017 from DigitalGlobe[2]. The goal of this subtask is to move forward the state-of-the-art of flood map generation by concentrating on road passability. In this regard, the challenge of this subtask is to identify sections of roads that are potentially blocked due to high water levels. Participants receive in addition to the very high resolution satellite patches, two pre-defined points on the road network depicted in the image. The task is to decide whether or not it is possible for a vehicle to drive on the road between the two points using the shortest path without passing through potentially flooded sections. Fusion of satellite and social

[1] http://reliefweb.int/disasters
[2] https://www.digitalglobe.com/opendata

Metadata	image_id, image_url, date_taken, date_uploaded, user_nsid, user_nickname, title, text, hashtags, capture_device, latitude, longitude
Visual Features	AutoColorCorrelogram, EdgeHistogram, Color and Edge Directivity Descriptor (CEDD), ColorLayout, Fuzzy Color and Texture Histogram (FCTH), Joint Composite Descriptor (JCD), Gabor, ScalableColor, Tamura

Table 1: Details of provided metadata information and visual features for the Social Images-Dataset

multimedia information is encouraged. Participants are allowed to submit 5 runs:

- Required run 1, 2: using the provided satellite data only
- General run 3, 4, 5: everything automated allowed, including using data from external sources (e.g. Open Street Map, Elevation Maps, Other Satellite Images, Social Media)

3 DATA

Flood Classification from Social Multimedia.

The dataset of the first subtask consists of 7,387 Tweet-Ids (dev-set) and 3.683 Tweet-Ids (test-set). All tweets with the tags *flooding, flood and floods* in the text and an accompanying image have been collected during the three big hurricane events in 2017 (named by *Harvey*, *Irma* and *Maria*) from Twitter. [3] In line with previous research [1], we also observed a large number of (near)-duplicated images in the collected dataset. Therefore, two pre-processing steps have been applied in order to de-duplicated such content. In a first step, perceptual hashing using the pHash function [4] was applied to remove all duplicated images based on the same hash-value. In the second step, near duplicates have been excluded based on the similarity of the deep feature representation of the last fully connected layer of an ImageNet [6] pre-trained ResNet101 [5]. As similarity measure, the cosine distance was used and all image features with a small distance under an empirically determined threshold (t=0.1) were grouped to one cluster of image duplicates.

The ground truth labels of the dataset consists of two classes: (1) one class label for the evidence of road passability for each tweet-Id with respect to the embedded image (0=no evidence/ 1=evidence). Those images that are labeled as showing evidence, have a second class label (2) for the actual road passability (0=not passable/ 1=passable). The images accompanying the text of the tweets were labeled by human annotators in a crowd-sourcing setup on the platform Figure Eight[3].

Participants were asked multiple questions about the image content with respect to the road passability and corresponding evidence for passability. The examples for road passability were available to the annotators in the interface during the entire process. The annotation process was not time restricted. The scores were collected from three annotators and aggregated according to the majority voting.

For each image, classical visual feature descriptors are provided to participants. These features were extracted with the open-source LIRE library[4] using default parameter settings. An overview of the provided features is given in Table 1. The dataset is separated with a ratio of 70/30 into the following two sets:

- **Development-Set** contains 7,387 tweets, along with visual and metadata features as well as two class labels for evidence and road passability
- **Test-Set** contains 3,683 images and features

Flood Detection from Satellite Imagery.

The dataset for the second remote sensing subtask consists of 1,664 satellite image patches that were extracted from DigitalGlobe's WorldView satellite. The imagery has a ground-sample distance (GSD) of about 0.5 meters and was collected from the Houston area during the hurricane event *Harvey* in 2017. The image patches have the spatial resolution of 512 x 512 pixels and show flooded as well as unflooded areas of Houston.

The satellite imagery comes with additional binary annotations for the road passability between two given point locations on the road network. The dataset is separated into the following split:

- **Development-Set** contains 1,438 image patches. For each image patch we provide two points on the road network and an annotation for the passability (1= passable, 0 = non passable).
- **Test-Set** consists of 226 satellite image patches.

4 EVALUATION

Flood Classification from Social Multimedia.

The official metric for evaluating the correctness of classified images from social multimedia is the macro averaged F1-Score. In our problem definition, the metric has to consider the following three classes (C1) images with no evidence on passability, (C2) images with evidence and passable roads as well as (C3) images with evidence and non passable roads. Since this definition extends the binary classification to a multi-label problem, the average of two F1-Scores for class C2 and C3 is computed.

Flood Detection from Satellite Imagery.

In order to assess the performance of the system for the classification of satellite patches that depict potentially blocked road connections between two given points, the metric F1-Score is used. This metric computes the harmonic mean between precision and recall for the non passable road class.

ACKNOWLEDGMENTS

We would like to thank Martha Larson for the very valuable feedback and support during the setup of this task. Additionally, we would like to thank DigitalGlobe for providing us with high-resolution satellite images for this task. This work was partially funded by the BMBF Project DeFuseNN (01IW17002).

[3]https://www.figure-eight.com

[4]LIRE, http://www.lire-project.net/

REFERENCES

[1] Benjamin Bischke, Damian Borth, Christian Schulze, and Andreas Dengel. 2016. Contextual enrichment of remote-sensed events with social media streams. In *Proceedings of the 2016 ACM on Multimedia Conference*. ACM, 1077–1081.

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

[3] Benjamin Bischke, Patrick Helber, Zhengyu Zhao, Jens de Bruijn, and Damian 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.

[4] Mengjuan Fei, Jing Li, and Honghai Liu. 2015. Visual tracking based on improved foreground detection and perceptual hashing. *Neurocomputing* 152 (2015), 413–428.

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[6] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, and others. 2015. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision* 115, 3 (2015), 211–252.