The Multimedia Satellite Task at MediaEval 2019

Estimation of Flood Severity

Benjamin Bischke[1,2], Patrick Helber[1,2], Simon Brugman[3],
Erkan Basar[3,4], Zhengyu Zhao[3], Martha Larson[3], Konstantin Pogorelov[5,6]
[1]German Research Center for Artificial Intelligence (DFKI), Germany [2]TU Kaiserslautern, Germany
[3]Radboud University, Netherlands [4]FloodTags, Netherlands
[5]Simula Research Laboratory, Norway [6]University of Oslo, Norway
{benjamin.bischke,patrick.helber}@dfki.de
{erkan.basar,simon.brugman,z.zhao,m.larson}@cs.ru.nl
konstantin@simula.no

ABSTRACT

This paper provides a description of the Multimedia Satellite Task at MediaEval 2019. The main objective of the task is to extract complementary information associated with events which are present in Satellite Imagery and Social Media. Due to their high socio-economic impact, we focus on flooding events and built upon the last two years of the Multimedia Satellite Task. Our task focuses this year on flood severity estimation and consists of three subtasks: (1) *Image-based News Topic Disambiguation*, (2) *Multimodal Flood Level Estimation from news*, (3) *Classification of city-centered satellite sequences*. The task moves forward the state of the art in flood impact assessment by concentrating on aspects that are important but are not generally studied by multimedia researchers.

Figure 1: Sample images for the Multimodal Flood Level Estimation dataset. The goal of this subtask is to identify persons standing in water above knee level, based on visual and textual information of news articles.

1 INTRODUCTION

Floods can cause loss of life and substantial property damage. Moreover, the economic ramifications of flood damage disproportionately impact the most vulnerable members of society [6]. In order to assess the impact of a flooding event, typically satellite imagery is acquired and remote sensing specialists visually or semi-automatically interpret them to create flood maps to quantify impact of such events. One major drawback of this approach when only relying on satellite imagery are unusable images from optical sensors due to the presence of clouds and adverse constellations of non-geostationary satellites at particular points in time. In order to overcome these drawbacks, we additionally analyse complementary information from social multimedia and news articles. The larger goal of this task is to analyse and combine the information in satellite images and online media content in order to provide a comprehensive view of flooding events. While there has been some work in disaster event detection from social media [1, 3, 5], not much research has been done in the direction of flood severity estimation. In this task, participants receive multimedia data, new articles, and satellite imagery and are required to train classifiers. The task moves forward the state of the art in flood impact assessment by concentrating on aspects that are important but are not generally studied by multimedia researchers. In this year, we are also in particular interested into a closer analysis of both, visual and textual information for severity estimation. In the following, we define three subtasks in the direction of flood severity estimation.

Copyright held by the owner/author(s).
MediaEval'19, 27-29 October 2019, Sophia Antipolis, France

2 TASK DETAILS

2.1 Image-based News Topic Disambiguation

For the first subtask, participants receive links to a set of images that appeared in online news articles (English). They are asked to build a binary image classifier that predicts whether or not the topic of the article in which each image appeared was a water-related natural-disaster event. All of the news articles in the data set contain a flood-related keyword, e.g., "flood", but their topics are ambiguous. For example, a news article might mention a "flood of flowers", without being an article on the topic of a natural-disaster flooding event. Participants are allowed to submit 5 runs:

- Required run 1: using visual information only
- General run 2, 3, 4, 5: everything automated allowed, including using data from external sources

2.2 Multimodal Flood Level Estimation

In the second subtask, participants receive a set of links to online news articles (English) and links to accompanying images. The set has been filtered to include only news articles for which the accompanying image depicts a flooding event. Participants are asked to build a binary classifier that predicts whether or not the image contains at least one person standing in water above the knee. Participants can use image-features only, but the task encourages a combination of image and text features, and even use of satellite imagery. As in the previous task, participants are allowed to submit 5 runs:

- Required run 1: using visual information only
- Required run 2: using text information only
- Required run 3: using visual and text information only
- General run 4, 5: everything automated allowed, including using data from external sources

2.3 City-centered satellite sequences

In this complementary subtask, participants receive a set of sequences of satellite images that depict a certain city over a certain length of time. They are required to create a binary classifier that determines whether or not there was a flooding event ongoing in that city at that time. Because this is the first year we work with sequences of satellite images, the data will be balanced so that the prior probability of the image sequence depicting a flooding event is 50%. This design decision will allow us to better understand the task. Challenges of the task include cloud cover and ground-level changes with non-flood causes. For this subtask, participants are allowed to submit the following five runs:

- Required run 1: using the provided satellite imagery
- General run 2, 3, 4, 5: everything automated allowed, including using data from external sources

3 DATA

3.1 Image-based News Topic Disambiguation

The dataset for this task contains links to images that were accompanying English-language news articles. News articles published in 2017 and 2018, were collected from ten local newspapers for multiple African countries (Kenya, Liberia, Sierra Leone, Tanzania and Uganda) if they contained at least one image and at least one occurrence of the word *flood*, *floods* or *flooding* in the text. This resulted in a set of 17.378 images. We noticed that there is a large number of duplicates in the dataset, therefore we applied a de-duplication algorithm and filtered out images such that we finally obtained a set of unique URLs for a all images in the dataset. This filtering step decreased the size of the dataset to 6.477 images.

The ground truth data of the dataset consists of a class label (0=not flood event related/1=flood event related) for each image. The ground truth was extracted from the corresponding text of the article with an advanced NLP solution [4]. This solution was developed by FloodTags[1] and is used for flood event detection in emergency response applications. We are aware that labels might be slightly noisy due to the automatically generated ground truth. However, by manually checking the labels for random articles we observed a good label quality. The images for this task were divided into a development set (5.181 images) and test set (1.296 images) using stratified sampling with a split ratio of 80/20.

3.2 Multimodal Flood Level Estimation

The dataset the Multimodal Flood Level Estimation task was extracted from the same African newspapers articles that were collected for the above described subtask. However, rather than in the previous task, we provide participants not only with images but rather the complete article. In total we collected 6.166 articles with the word *flooding, floods*.

We annotated the images based on the image content. For the annotation we used the open-source VGG Image Annotator[2] (VIA) from the Visual Geometry Group at Oxford [2]. We drew a bounding box around all people who are depicted with at least one of their feet occluded by water. Children are included in the definition of people, although they are shorter. In order to derive consistent labels, we were in particular interested in persons standing in water, in the sense that the part of the body that is under water, should be in the upright position. For each of the bounding boxes we additionally collected a depth indicator: *feet*, *knee*, *hip* or *chest*. If one knee is occluded by water and not the hip, then we annotated knee, because the highest body part the water has reached is the knee. We follow the same approach as described above to divide the articles into a development set (4.932 articles) and test set (1.234 articles).

3.3 City-centered satellite sequences

The dataset for last subtask was derived from the Sentinel-2 satellite archive of the European Space Agency (ESA) and the Copernicus Emergency Management Service (EMS). We collected satellite images for past flooding events that have been mapped and validated by human annotators from the Copernicus EMS team. Rather than relying on a single satellite image to estimate flood severity, we consider a sequence of images. We also provide all 13-bands of the Sentinel-2 image (L2A), since bands beyond the visible RGB-channels contain vital information about water. For each flooding event, we determine and download the corresponding Sentinel-2 image scences that have been recorded 45 days before and 45 days after the flooding event. We compute the intersection of the satellite images with the ground truth obtained from the EMS service and split the image scences into smaller patches of size 512 x 512 pixels. This resulted in a set of 335 image sequences. Depending on the constallation of the Sentinel-2 satellites, we obtained for each sequence between 4 and 20 image patches. For each image patch, we provide additional metadata such as cloud cover and the amount of black pixels due to errors in the data acquisition. The label is created based on the intersection of the images in each sequence with the manually annotated flood extend of EMS (0=no overlap, 1=overlap with image sequence). We split the sequences with 80/20 into a development set and test set.

4 EVALUATION

In order to evaluate the approaches we will use the metric F1-Score for all three subtasks. The metric computes the harmonic mean between precision and recall for the corresponding class of the task.

ACKNOWLEDGMENTS

This work was supported BMBF project DeFuseNN (01IW17002) and the NVIDIA AI Lab (NVAIL) program. We would like to thank the FloodTags team for giving us access to the collected links of news articles and the NLP based annotations for the first subtask.

[1] https://www.floodtags.com/

[2] https://github.com/multimediaeval/2019-Multimedia-Satellite-Task/raw/wiki-data/multimodal-flood-level-estimation/resources/via.html

REFERENCES

[1] Tom Brouwer, Dirk Eilander, Arnejan Van Loenen, Martijn J Booij, Kathelijne M Wijnberg, Jan S Verkade, and Jurjen Wagemaker. 2017. Probabilistic flood extent estimates from social media flood observations. *Natural Hazards & Earth System Sciences* 17, 5 (2017).

[2] Abhishek Dutta and Andrew Zisserman. 2019. The VIA Annotation Software for Images, Audio and Video. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19)*. ACM, New York, NY, USA, 4. https://doi.org/10.1145/3343031.3350535

[3] Dirk Eilander, Patricia Trambauer, Jurjen Wagemaker, and Arnejan Van Loenen. 2016. Harvesting social media for generation of near real-time flood maps. *Procedia Engineering* 154 (2016), 176–183.

[4] A Hürriyetoğlu, Jurjen Wagemaker, NHJ Oostdijk, and APJ van den Bosch. 2016. Analysing the role of key term inflections in knowledge discovery on twitter. (2016).

[5] Min Jing, Bryan W Scotney, Sonya A Coleman, Martin T McGinnity, Stephen Kelly, Xiubo Zhang, Khurshid Ahmad, Antje Schlaf, Sabine Grunder-Fahrer, and Gerhard Heyer. 2016. Flood event image recognition via social media image and text analysis. In *IARIA conference COGNITIVE*.

[6] Tim GJ Rudner, Marc Rußwurm, Jakub Fil, Ramona Pelich, Benjamin Bischke, Veronika Kopačková, and Piotr Biliński. 2019. Multi3Net: Segmenting Flooded Buildings via Fusion of Multiresolution, Multisensor, and Multitemporal Satellite Imagery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 702–709.