Detection of Road Passability from Social Media and Satellite Images

Armin Kirchknopf[1], Djordje Slijepcevic[1], Matthias Zeppelzauer[1], Markus Seidl[1]
[1] Media Computing Resarch Group, St. Pölten University of Applied Sciences
armin.kirchknopf@fhstp.ac.at,djordje.slijepcevic@fhstp.ac.at,m.zeppelzauer@fhstp.ac.at,markus.seidl@fhstp.ac.at

ABSTRACT

This paper presents the contribution of Team MC-FHSTP to the multimedia satellite task at the MediaEval 2018 benchmark. We present two methods, one for the estimation of the passability of roads from social media images due to flooding and one method that estimates passability from satellite images. We present the results obtained in the benchmark for both methods.

1 INTRODUCTION

Flood threats have prompted many researchers to develop technology-based solutions for the precise and autonomous exploration of flood areas. Such solutions should enable the assessment of the impact of hazards as well as the immediate response to disasters. This can be done by analyzing satellite images. In the work of Pradhan et al. [9], Amitrano et al. [2], and Sumalan et al. [10] satellite images are analyzed with the aim of identifying flood areas. Yamaguchi and Saji [6] propose a method that analyses satellite images and indicates road conditions after an earthquake including a tsunami. The work of Amit et al. [1] utilizes an convolutional neural network (CNN) to extract disaster regions automatically by combining pre-disaster and post-disaster satellite images. Another more recent data source are media from social networks, which provide very prompt and local information about the affected areas. In order to identify bush fires and their effects, Lagerstrom et al. [7] investigated methods that divide Twitter images into two classes, fire-related and non-fire-related content, see furthermore [3]. Yang et al. [11] used images and text data downloaded from Flickr to distinguish between three different disaster classes: hurricane, oil spill, and earthquake. Each of the classes was divided into two subcategories, with floods forming a subcategory of the hurricane class. The authors utilized feature vectors based on the word frequency, and some basic image features prior to an multiple correspondence analysis (MCA). Nguyen et al. [8] have used a CNN to distinguish images from social media of four catastrophic events into three classes of severity. The work of Cervone et al. [5] utilized several data streams including satellite images, aerial images, tweets and images downloaded from Flickr to assess the damage to transportation infrastructure after the Colorado floods in 2013.

In this paper we are aiming at determining road passability from images as proposed in the Multimedia Satellite Task of the MediaEval 2018 benchmark [4]. We present and evaluate different machine learning approaches on two datasets provided for two sub-tasks. The first sub-task comprises the retrieval of images from social media which provide evidence for the passability of roads. The second sub-task addresses the detection of the passability of roads in satellite images.

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

2 APPROACH

The following two sections describe our approaches for both subtasks. The idea behind both approaches is to provide a simple, reproducible and straight-forward baseline for the respective tasks.

2.1 Sub-task 1: Flood classification for social multimedia

The dataset provided for sub-task one consists of twitter text-messages with accompanying images and is therefore a multimodal data-source. Our initial idea consisted of training two different classifiers for the image data and a separate one for the text data and then merging the predictions. This should allow the information from both inputs to be used for the final decision.

For the textual data, we investigated several methods. All tweets that were not in English or Spanish were excluded from further processing. Then the Spanish tweets (a representative minority in the data) were translated into English using the Google Translate API. We generated Bag-of-Words(BOW) descriptors on the $N = 100$ and $N = 50$ most common words and used them separately for training. Since the above mentioned representation did not deliver promising results on the training data, we chose to calculate TFIDF representations.

For the image data, two approaches were developed. First, we extracted a Bag-of-Words descriptor based on the responses (tags) retrieved by the clarifai[1] API for the twitter images and fed it into a Support Vector Machine (SVM). Since preliminary results were, however, little promising on the development data, we excluded the approach from the final evaluation. We assume, that the main problem with the clarifai tags was that they were not specific enough to solve our task. In a second visual classification approach we leveraged a deep neural network for image description and classification (ResNet50 pre-trained on ImageNet[2]). The original network was extended with two densely connected layers (of size 512 and 256, respectively) and with two dropout layers (likelihood 0.5) in between and a softmax layer on top. Initially, in all experiments only the added layers were trained for 20 epochs. Depending on the run (see below), in a further step the whole network was fine-tuned within a variable number of epochs. Prior to network training the provided twitter images were pre-processed in two ways: first, the images were rescaled (non-uniformly) to the required input size of the network (224x224), and second, a central patch was cut out from

[1] https://www.clarifai.com/
[2] http://www.image-net.org/

the images to put more focus on the image center. Furthermore, we applied rotation and mirroring as data augmentation steps.

2.2 Sub-task 2: Detection of road passability in satellite images

For the prediction of road passability between two given points in a satellite image we made a simplifying assumption. We assumed, that at least one of the two given points is under water in case the road segment defined by the two points is not passable. Consequently, we assumed that the patches around points under water share visual properties. We modelled our assumption as follows. *Patches*: From each satellite image, we extracted a patch of 50x50px around each of the two given points. *Visual features*: From each patch, we extracted RGB histograms with 16 bins per channel. *Training and Classification*: We trained SVMs. For the evaluation of our approach on the development data we used 10-fold cross validation. For the test data we trained the SVMs with the complete development data set. We used three aggregation methods for the RGB histograms:

1. *Concatenation (concat):* We concatenated the RGB histograms of the two patches of a satellite image to a feature vector with 96 dimensions and trained an SVM.
2. *Separation (sep):* We trained two SVMs, one with the patches of the first point of each image and the second with the patches of the second points, respectively. We predicted, that a satellite image contains a passable road, if the predictions for both patches of the image are passable.
3. *Joined (join):* As in aggregation (sep), we kept the RGB histograms of the two patches per satellite images separate. Instead of training two SVMs, we trained one SVM with all patches, i.e. we used a training data set with double the number of samples. We predicted as in the (sep) case.

In a first baseline experiment, we evaluated this assumption. The results were surprisingly useful, consequently we decided to use the approach.

3 EXPERIMENTAL RESULTS

3.1 Results of sub-task 1

Since the initial data set was unbalanced, we divided the data into a balanced training set to avoid any bias towards a class during training and an unbalanced validation set (that contains the remaining development data). The training set contained 874 samples per class. The results for the training, validation and test phase in terms of F1-score and classification accuracy (in percent) are summarized in Table 1. In Run 1 we directly used the predictions of the ResNet50 network, which was trained on the balanced training set and evaluated on the unbalanced validation set. We trained the model for five epochs, with a batch size of 32, the Adam optimization algorithm and a learning rate of 0.0001. In Run 2 we employed the SVM model trained with the TFIDF representations which are based on a high min_df value of 120. We skipped Run 3 because we could not gain better results by combining visual and textual information. In Run 4 and Run 5 the predictions where obtained from the ResNet50 network trained on the entire development data for three epochs (Run 4) and six epochs (Run 5). In the training phase Run 2 performed best regarding the F1-score, and Run 5 regarding the accuracy. However, it should be noted that generally a relatively high accuracy was achieved, but a rather low F1-score, which assumably stems from the class imbalance and indicates that the models could not learn all classes equally well. On the test set Run 5 clearly performs best (F1-score = 0.35). The weakest result is obtained by Run 4 which indicates that the training was stopped too early. A prolonged training phase (more than 6 epochs as in Run 5) could further improve the result. Overall, it is notable that the best run (run 5) also has the strongest generalization ability (i.e. F1 on test set is larger than on validation set).

Table 1: F1-scores (macro-averaged over classes 2 (passable with evidence) and 3 (non passable with evidence)) and accuracies. The asterisk (*) is used if no data is available.

Run	Training (F1 | Acc.)		Validation (F1 | Acc.)		Test (F1)
1	34%	83%	28%	86%	20%
2	43%	85%	37%	83%	24%
4	27%	85%	*	*	17%
5	27%	89%	*	*	35%

Table 2: Recall, precision and F1-scores for the non-passable class in the satellite image sub-task.

Run	Training (R | P | F1)			Test (F1)
1 (concat)	73%	85%	79%	57%
2 (sep)	83%	80%	81%	32%
3 (join)	86%	82%	84%	39%
4 (majority)	84%	84%	84%	56%
5 (unanimous)	73%	89%	79%	57%

3.2 Results of sub-task 2

Table 2 contains the results of the second sub-task. We used the runs 1,2 and 3 for each of our aggregation methods introduced in Section 2.2. Runs 4 and 5 do not contain additional data but represent classifier fusions of runs 1,2 and 3, where run 4 is the result of majority voting and run 5 of unanimous voting. The results for runs 1,2 and 5 on the test set are notably better than for runs 2 and 3 which indicates over-fitting for these two runs.

4 CONCLUSION

We have presented two approaches for the detection of road condition (passability) from social media images and satellite images. Both approaches have low complexity and are easy to reproduce and generalize in most cases well on the test data. To improve the visual modality of task 1 a two-step approach, i.e. first classification on evidence and secondly on passability, could be more rewarding. They represent a first baseline for this task that shall be improved in future.

ACKNOWLEDGMENTS

This work was supported by the Austrian Research Promotion Agency (FFG), Project No. 856333.

REFERENCES

[1] Siti Nor Khuzaimah Binti Amit, Soma Shiraishi, Tetsuo Inoshita, and Yoshimitsu Aoki. 2016. Analysis of satellite images for disaster detection. In *Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International*. IEEE, 5189–5192.

[2] D. Amitrano, G. Di Martino, A. Iodice, D. Riccio, and G. Ruello. 2018. Unsupervised Rapid Flood Mapping Using Sentinel-1 GRD SAR Images. *IEEE Transactions on Geoscience and Remote Sensing* 56, 6 (June 2018), 3290–3299. https://doi.org/10.1109/TGRS.2018.2797536

[3] 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.

[4] 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.

[5] Guido Cervone, Elena Sava, Qunying Huang, Emily Schnebele, Jeff Harrison, and Nigel Waters. 2016. Using Twitter for tasking remote-sensing data collection and damage assessment: 2013 Boulder flood case study. *International Journal of Remote Sensing* 37, 1 (2016), 100–124.

[6] Hitoshi Saji Keishi Yamaguchi. 2012. Analysis of road damage after a large-scale earthquake using satellite images. (2012), 8524 - 8524 - 8 pages. https://doi.org/10.1117/12.976288

[7] Ryan Lagerstrom, Yulia Arzhaeva, Piotr Szul, Oliver Obst, Robert Power, Bella Robinson, and Tomasz Bednarz. 2016. Image Classification to Support Emergency Situation Awareness. *Frontiers in Robotics and AI* 3 (2016), 54. https://doi.org/10.3389/frobt.2016.00054

[8] Dat Tien Nguyen, Firoj Alam, Ferda Ofli, and Muhammad Imran. 2017. Automatic image filtering on social networks using deep learning and perceptual hashing during crises. *arXiv preprint arXiv:1704.02602* (2017).

[9] B. Pradhan, M. S. Tehrany, and M. N. Jebur. 2016. A New Semiautomated Detection Mapping of Flood Extent From TerraSAR-X Satellite Image Using Rule-Based Classification and Taguchi Optimization Techniques. *IEEE Transactions on Geoscience and Remote Sensing* 54, 7 (July 2016), 4331–4342. https://doi.org/10.1109/TGRS.2016.2539957

[10] A. L. Sumalan, D. Popescu, and L. Ichim. 2017. Flooded and vegetation areas detection from UAV images using multiple descriptors. In *2017 21st International Conference on System Theory, Control and Computing (ICSTCC)*. 447–452. https://doi.org/10.1109/ICSTCC.2017.8107075

[11] Yimin Yang, Hsin-Yu Ha, Fausto Fleites, Shu-Ching Chen, and Steven Luis. 2011. Hierarchical disaster image classification for situation report enhancement. In *Information Reuse and Integration (IRI), 2011 IEEE International Conference on*. IEEE, 181–186.