Multi-Modal Machine Learning for Flood Detection in News, Social Media and Satellite Sequences

Kashif Ahmad[1], Konstantin Pogorelov [2], Mohib Ullah[3],
Michael Riegler[4], Nicola Conci[5], Johannes Langguth [2], Ala Al-Fuqaha [1]
[1]Hamad Bin Khalifa University, Doha, Qatar, [2] Simula Research Laboratory, Norway
[3]Norwegian University of Science and Technology, Norway, [4] Simula Metropolitan Center for Digitalisation and
Kristiania University College, Norway, [5] University of Trento, Italy
{kahmad,aalfuqaha}@hbku.edu.qa,mohib.ullah@ntnu.no,nicola.conci@unitn.it
{konstantin,michael,langguth}@simula.no

ABSTRACT

In this paper we present our methods for the MediaEval 2019 Multimedia Satellite Task, which is aiming to extract complementary information associated with adverse events from Social Media and satellites. For the first challenge, we propose a framework jointly utilizing colour, object and scene-level information to predict whether the topic of an article containing an image is a flood event or not. Visual features are combined using early and late fusion techniques achieving an average F1-score of 82.63, 82.40, 81.40 and 76.77. For the multi-modal flood level estimation, we rely on both visual and textual information achieving an average F1-score of 58.48 and 46.03, respectively. Finally, for the flooding detection in time-based satellite image sequences we used a combination of classical computer-vision and machine learning approaches achieving an average F1-score of 58.82.

1 INTRODUCTION

When natural disasters occur, an instant access to relevant information may be crucial to mitigate loss in terms of property and human lives, and may result in a speedy recovery [12]. In this regards, social media and remotely sensed information have been proved very effective [1, 3, 12]. Similar to the 2017 [5] and 2018 [6] versions of the task, the MediaEval 2019 Multimedia Satellite task [4] aims to combine the information from the two complementary sources, namely social media and satellites.

This paper provides detailed description of the methods proposed by team UTAOS for the MediaEval 2019 Multimedia Satellite challenge. The challenge consists of three parts, namely (i) News Image Topic Disambiguation (NITD), (ii) Multimodal Flood Level Estimation (MFLE) and (iii) City-centered Satellite Sequences (CCSS).

The first two tasks(NITD and MFLE) are based on social media data aiming to (a) predict whether the topic of the article containing the image was a water-related natural-disaster event or not, and (b) to build a binary classifier that predicts whether or not the image contains at least one person standing in water above the knee.

In the CCSS task, the participants are provided with a set of sequences of satellite images depicting a certain city over a certain length of time, and the they need to propose and develop a framework able to determine whether or not there was a flooding event ongoing in that city at that time.

2 PROPOSED APPROACH

2.1 Methodology for NITD task

Considering the diversity of the content covered by natural disaster-related images, based on our previous experience [2], we utilize a diversified set of visual features including colour, texture, object and scene-level features. The object and scene-level features are extracted through three different Convolutional Neural Network (CNN) models, namely AlexNet[9], VggNet [13] and ResNet [8], pre-trained on the ImageNet dataset [7] and the Places dataset [15]. The models pre-trained on ImageNet correspond to object level information while the ones pre-trained on the Places dataset extract scene level information. For feature extraction from all models, we use the Caffe toolbox[1]. For colour and texture features we rely on the LIRE open source library [10] which we used to extract joint composite descriptor (JCD) features from the images.

In order to combine the features, we use both early and late fusion techniques. For the early fusion, feature vectors are concatenated. For late fusion two different techniques namely (i) simple averaging and (ii) Particle Swarm Optimization (PSO) based technique is used for late fusion. The basic motivation behind PSO based fusion is to assign merit based weights to the deep models. For classification purposes, we rely on Support Vector Machines (SVMs) in all of the submitted fusion runs. Moreover, to deal with class imbalance problem, we use ensemble different re-sampled data sets technique where five different models are trained using all the samples of the rare class and n-differing samples of the abundant class.

2.2 Methodology for MFLE task

For the MFLE task, we proposed two different solutions exploiting both: visual and textual information. For visual features based flood estimation, we proposed a two step framework where as a first step an ensemble of binary image classifiers trained on deep visual features, extracted through AlexNet pre-trained on ImageNet and Places datasets, is used to differentiate between flood and non-flooded images. In the second step, we rely on tracking techniques [14], for which an open source library, namely OpenPose[2], has been used to draw and extract body points on the people in the flood related images. Subsequently, the generated coordinates are analyzed to identify the images having at least one person standing in water and the water level is above the knee height by checking the knee joints at the corresponding index of the generated files

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

[1] https://github.com/BVLC/caffe
[2] https://www.learnopencv.com/tag/openpose/

of the joints extracted for each person. On the other hand, for text analysis we employed two methods, namely (i) Bag-of-words Model (BoW) and (ii) LSTM network. Before applying the methods, the data was pre-processed for tokenization and removing of punctuation.

2.3 Methodology for CCSS task

For CCSS task, first, we tried to employ a recurrent convolutional neural network architecture designed for change detection in multispectral satellite imagery data (ReCNN) [11]. This network was initially designed to solve the task very similar to CCSS task goals, and the results depicted by ReCNN's authors are promising. However, despite high expectations, ReCNN was not able to achieve sufficient and better-than-random performance of detection changes caused by flooding. Our assumption is that was caused by the "real" nature of the dataset provided in CCSS task. Images were taken in different seasons, often partially or fully covered by clouds and sometimes have noticeable pixel offset between each other. After a series of unsuccessful experiments, we decided to use a classical image processing and analysis approach with multi-stage image processing using simple operations. First, we mask-out from the further analysis all cloud-covered image areas by applying simple threshold function. Reference threshold value is computed per-image by averaging the values of the pixels located in monotonically white-colored areas. The same masking is performed for dark underexposured and areas with missing imaging data. Next, we scaled images down to uniform size of $128 * 128$ pixels to reduce noise and soften image-shifting influence. Then, scaled images are converted into hue-saturation-value (HSV) color space, and further analysis is performed on HSV bands. Using the same thresholding methodology, we mask-out pixels with too-low and too-high saturation (S) and Value (V) channel values. The resulting masks are filtered with median filter and processed by dilation filter. Resulting images are compared in sequential pairs within non-masked-out regions using grey level co-occurrence matrix texture feature. Final flooding presence detection is made by using random tree classifier.

3 RESULTS AND ANALYSIS

3.1 Runs Description in NITD Task

For NITD, we submitted total four runs. In run 1, we used the PSO based weight optimization method for assigning weights to each model on merit basis. For run 2, the deep models are treated equally by assigning equal weights to all models. In our run 3, we added colour based features to our pool of features descriptors in a late fusion method where the scores of all models are simply added to obtain the final prediction. Our run 4 is based early fusion where the deep features are simply concatenated for training SVMs. Table 1a provides the experimental results our proposed solutions for NITD task on both development and test sets. Overall better results are obtained with PSO based late fusion which shows the advantage of merit based late fusion of the models. On the other hand, least F1-score is obtained with early fusion. Moreover, the colour based features did not contribute positively in the performance of the framework. This might be due to the fact that the JCD feature is very compressed and does not contain much information that the fusion algorithm could exploit.

Table 1: Evaluation of our proposed approaches for (a) NITD and (b) MLFE tasks in terms of F1-scores.

(a) NITD

Run	Dev. Set	Test Set
Run 1	81.08	82.63
Run 2	79.43	82.40
Run 3	77.77	81.40
Run 4	75.70	76.77

(b) MFLE

Run	Dev. Set	Test Set
Run 1	61.00	58.48
Run 2	56.71	46.03
Run 4	57.56	44.91

3.2 Runs Description in MLFE Task

For MLFE task, we submitted two mandatory and one optional run. The first run is based on visual information where a two phase approach has been proposed for flood level estimation starting with deep features based classification of flooded and non-flooded images, followed by human body points detection via Openpose library in the flood-related images. Our second and third runs are based on textual information where Bag-of-words (BoW) and LSTM based techniques are used for the article classification, respectively. Table 1b shows the experimental results of our solutions for MLFE task on both development and test sets. Overall, better results are obtained with visual information. Moreover, BoW features produce slightly better results over LSTM based approach.

3.3 Runs Description in CCSS Task

For CCSS task we submitted the mandatory run only. Evaluation performed by the task organizers showed F1-score of 58.82% for flooding detection performance on the provided test set. The relatively high performance for our simple detection approach can be explained by the used aggressive image masking technique which allow us to perform comparison of only clearly visible areas. However, our own evaluation shows that our approach is not able to distinguish correctly between image changes caused by flooding and seasonal vegetation grow.

4 CONCLUSIONS AND FUTURE WORK

This year, the social multimedia satellite task introduced a new and important challenges including image based news topic disambiguation (NITD), multi-modal flood level estimation in social media content (MFLE) and predicting a flood event in a set of sequences of satellite images of a certain city over a certain length of time (CCSS). For the NITD task, we mainly relied on ensembles of classifiers trained on deep features extracted through several pre-trained deep models as well as global features (GF). During the experiments, we observed that the object and scene-level features complement each others when jointly utilized in a proper way. Moreover, deep features are proved more effective compared to GF. For MLFE task, we used both textual and visual information where better results were obtained with visual information. However, textual and visual information can complement each other. In the future, we aim to analyze the task with more advanced early and late fusion techniques to better utilize the multi-modal information. Furthermore, we plan to use complex GF. For CCSS task, we used the combination of computer-vision and machine learning approaches. For future results improvement, we will continue investigating recurrent CNN and GAN-based approaches in combination with classical image processing algorithms.

REFERENCES

[1] Kashif Ahmad, Konstantin Pogorelov, Michael Riegler, Olga Ostroukhova, Pål Halvorsen, Nicola Conci, and Rozenn Dahyot. 2019. Automatic detection of passable roads after floods in remote sensed and social media data. *Signal Processing: Image Communication* 74 (2019), 110–118.

[2] Kashif Ahmad, Amir Sohail, Nicola Conci, and Francesco De Natale. 2018. A Comparative study of Global and Deep Features for the analysis of user-generated natural disaster related images. In *2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP)*. IEEE, 1–5.

[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 24th ACM international conference on Multimedia*. ACM, 1077–1081.

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

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

[6] 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.

[7] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on.* Ieee, 248–255.

[8] 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.

[9] 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.

[10] Mathias Lux, Michael Riegler, Pål Halvorsen, Konstantin Pogorelov, and Nektarios Anagnostopoulos. 2016. LIRE: open source visual information retrieval. In *Proceedings of the 7th International Conference on Multimedia Systems*. ACM, 30.

[11] Lichao Mou, Lorenzo Bruzzone, and Xiao Xiang Zhu. 2018. Learning spectral-spatial-temporal features via a recurrent convolutional neural network for change detection in multispectral imagery. *IEEE Transactions on Geoscience and Remote Sensing* 57, 2 (2018), 924–935.

[12] Naina Said, Kashif Ahmad, Michael Riegler, Konstantin Pogorelov, Laiq Hassan, Nasir Ahmad, and Nicola Conci. 2019. Natural disasters detection in social media and satellite imagery: a survey. *Multimedia Tools and Applications* (17 Jul 2019). https://doi.org/10.1007/s11042-019-07942-1

[13] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[14] Mohib Ullah and Faouzi Alaya Cheikh. 2018. A directed sparse graphical model for multi-target tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 1816–1823.

[15] Bolei Zhou, Agata Lapedriza, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. 2014. Learning deep features for scene recognition using places database. In *Advances in neural information processing systems*. 487–495.