Deep learning approaches for flood classification and flood aftermath detection

Naina Said[1], Konstantin Pogorelov[2], Kashif Ahmad[4], Michael Riegler[3],
Nasir Ahmad[1], Olga Ostroukhova [5], Pål Halvorsen[3], Nicola Conci[4]
[1]University of Engineering and Technology, Peshawar, Pakistan
[2]Simula Research Laboratory, University of Oslo, Norway
[3]Simula Metropolitan Center for Digital Engineering, University of Oslo, Norway
[4]University of Trento, Italy
[5]Research Institute of Multiprocessor Computation Systems n.a. A.V. Kalyaev, Russia
naina_dcse@yahoo.com,konstantin@simula.no,kashif.ahmad@unitn.it
michael@simula.no,n.ahmad@uetpeshawar.edu.pk,paalh@simula.no,nicola.conci@unitn.it

ABSTRACT

This paper presents the method proposed by team UTAOS for MediaEval 2018 Multimedia Satellite Task: Emergency Response for Flooding Events. In the first challenge, we mainly rely on object and scene level features extracted through multiple deep models pre-trained on the ImageNet and Places datasets. The object and scene-level features are combined using early, late and double fusion techniques achieving an average F1-score of 65.03%, 60.59% and 63.58%, respectively. For the second challenge, we rely on a convolutional neural networks (CNNs) and a transfer learning-based classification approach achieving an average F1-score of 62.30% and 61.02% for run 1 and run 2, respectively.

1 INTRODUCTION

Natural disasters, such as floods, earthquakes and droughts, may cause significant damage to both human life and infrastructure. In such adverse events, an instant access to information may help to mitigate the damage [1, 2]. In recent years, social media and remotely sensed information have been widely utilized to analyze natural disasters and their potential impact on the environment [4, 9, 11]. Similar to the 2017 version, the MediaEVal 2018 Social Media and Satellite task [5] aims to combine the information from the two complementary sources of information to provide a better view of the underlying natural disaster.

This paper provides detailed description of the methods proposed by team UTAOS for MediaEval 2018 Multimedia Satellite Task: Emergency Response for Flooding Events. The challenge is composed of two parts, namely (i) flood classification for social multimedia and (ii) flood detection in satellite imagery. The first task is further divided in two sub-tasks aiming to predict (a) whether there are evidences of a flood in a given social media image or not and (b) if evidences of flood exists in the image, whether it is possible to pass through the flooded road (passability). The second task aims to analyze the roads from satellite images, and predict whether or not it is possible for a vehicle to pass a road.

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

2 PROPOSED APPROACH

2.1 Methodology for FCSM Task

To tackle the first challenge, based on our previous experience [3], we rely on features extracted through four different Convolutional Neural Network (CNN) models pre-trained on the ImageNet dataset [6] and the Places dataset [18]. These models include two pre-trained models, on the Places dataset (AlexNet[10] and VggNet [15]) and two models (VggNet and ResNet [8]) pre-trained on the ImageNet dataset. 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].

To be able to fuse the complementary information (i.e., object and scene-level features), we use three different fusion techniques, namely early, late and double fusion. In the early fusion, we simply concatenate the features extracted through different models. In the late fusion, we simply average the results obtained through the individual models. In our third fusion technique, we combine the results obtained from the first two techniques in an additional late fusion step by assigning them equal weights. For classification purposes, we rely on Support Vector Machines (SVMs) in all of the submitted fusion runs.

2.2 Methodology for FDSI Task

For the FDSI part of the task, we initially tried to apply the well-performing Generative Adversarial Network (GAN) approach introduced in our previous works for the flood detection satellite imagery [1] and medical imagery [13, 14].

We conducted an exhaustive set of experiments, but unfortunately could not achieve a roads passability detection performance better than random label assignment would achieve. The reason for that is the limited size of the dataset (only 1,437 samples were provided in the development set). This, in combination with the large variety of landscapes, road types, types of obstacles and weather conditions, etc., prevents the GAN-based approach from adequate training and finding key visual features required to reliable distinguish between flooded and non-flooded roads.

Thus, we decided to fall-back to another convolutional neural network (CNN) and a transfer learning-based classification approach,

[1]http://caffe.berkeleyvision.org/

which has been mainly introduced for the medical images classification in our previous work [12]. This approach is based on the Inception v3 architecture [16] pre-trained on the ImageNet dataset [6] and the retraining method described in [7].

For the here presented work, we froze all the basic convolutional layers of the network and only retrained the two top fully connected layers after random initialization of their weights. The fully connected layers were retrained using the RMSprop [17] optimizer which allows an adaptive learning rate during the training process.

As the input for the CNN model, we used the image patches, extracted from the full images using the provided coordinates of the target road end points. Visual inspection of the training-dataset-generated roads' patches showed relatively good coverage for the road-related areas and enough coverage of the neighbourhood areas together give enough visual information for the following CNN-based analysis and classification.

Moreover, in order to increase the number of training samples, we also performed various augmentation operations on the images. Specifically, we performed horizontal and vertical flipping and change of brightness in the interval of ±40%. After the model has been retrained, we used it for a multi-class classifier that provides the probability for each of two classes: passable and non-passable.

3 RESULTS AND ANALYSIS

3.1 Runs Description in FCSI Task

During the experimentation on the development set, we observed that the classifiers trained on scene level features extracted through models pre-trained on the places dataset perform better compared to the ones pre-trained on Imagenet. However, we also observed that combining the object and scene level features leads to better results compared to the individual models. In order to better utilize the scene and object level features, we used three different techniques in each run. Our first run is based on late fusion where we used equal weights for each model. In the second run, we concatenate the features extracted through the individual models. An SVM classifier is then trained on the combined feature vectors. In the final run, we use double fusion by combining the results of the first two runs in a late fusion way. Table 1 provides the evaluation results of our proposed methods in terms of F1 scores for each of the runs. Overall, better results are obtained with late fusion while least results are obtained with an early fusion of the features.

Table 1: Evaluation of our proposed approach for the FCSI task in terms of F1 Scores.

Run	Method	F1 Score
1	Late Fusion	65.03%
2	Early Fusion	60.59%
3	Double Fusion	63.58%

3.2 Runs Description in FDSI Task

For the experimental setup of the FDSI task, we decided to perform only two mandatory runs which are utilizing the task-provided training data only. Due to a limited amount of training samples available, the usage of the additional road network detection methods is not possible for these runs. Thus, we decided to perform two types of training for our transfer-learning detection approach.

Table 2: Evaluation of our proposed approach for the FDSI task in terms of F1 Scores.

Run	Method	F1 Score
1	All-train	62.30%
2	Half-train	61.02%

First, we implemented a pipeline for classification that differs from common procedures. This process was involving all the training samples into the training process as both training and validation sets. Usually, for classification tasks, this would result into overfitting of the model and inability to correctly classify the test samples. However, for this specific task, the limited number of training epochs and significant training data augmentation in conjunction with a high variety of road patch samples resulted in normal training process. This allowed to correctly retrain the last layers of the network and produce reasonable classifiers even on such a limited training set.

The official F1-Score metric (see table 2) on the non-passable road class for the first "All-train" run is 62.30%. To verify our idea of the usability of using all the training data for both training and validation, we also performed a normal network training with a random 50/50 development/validation data split. This second *Half-trained* run resulted in F1-Score of 61.02% which is slightly lower comparing to the *All-trained* run. This is confirming the validity of our idea of using the complete training dataset and heavy data augmentation to improve road patches classification performance.

4 CONCLUSIONS AND FUTURE WORK

This year, the social media and satellite task introduced a new and important challenge of detecting the passibility of roads, which can be vital for the people lining in the flooded regions. In the social media image analysis, we mainly relied on deep features extracted through different pre-trained deep models. During the experiments, we observed that the scene-level information, extracted through models pre-trained on the places dataset perform better compared to the ones pre-trained on Imagenet. However, the object-level information well complement the scene-level features when combined. We also observed that late fusion performs slightly better than the early fusion on the provided dataset. However, it needs to be investigated in more detail. In the future, we aim to analyze the task with more advanced early and late fusion techniques.

On the other hand, in the satellite subtask, we found that just a normal image segmentation approach cannot help, and we implemented an task-oriented CNN and transfer-learning-based approach. This approach was able to classify image patches with roads and achieved an F1-Score of 62.30% for the non-passable road class. In the future, we plan to implement an advanced road network and flooding detection and segmentation using a combined CNN- and GAN-based approach pre-trained on the existing annotated road network and flooded areas datasets.

REFERENCES

[1] Kashif Ahmad, Konstantin Pogorelov, Michael Riegler, Nicola Conci, and Pål Halvorsen. 2018. Social media and satellites. *Multimedia Tools and Applications* (2018), 1–39.

[2] Kashif Ahmad, Michael Riegler, Konstantin Pogorelov, Nicola Conci, Pål Halvorsen, and Francesco De Natale. 2017. Jord: a system for collecting information and monitoring natural disasters by linking social media with satellite imagery. In *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*. ACM, 12.

[3] 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.

[4] Benjamin Bischke, Prakriti Bhardwaj, Aman Gautam, Patrick Helber, Damian Borth, and Andreas Dengel. 2017. Detection of flooding events in social multimedia and satellite imagery using deep neural networks. In *Proceedings of the Working Notes Proceeding MediaEval Workshop, Dublin, Ireland*. 13–15.

[5] 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.

[6] 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.

[7] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. 2014. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition.. In *Proc. of ICML*, Vol. 32. 647–655.

[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] M Jing, BW Scotney, SA Coleman, and others. 2016. Flood Event Image Recognition via Social Media Image and Text Analysis. In *Signals and Systems Conference (ISSC)*. 4–9.

[10] 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.

[11] Keiller Nogueira, Samuel G Fadel, Ícaro C Dourado, Rafael de O Werneck, Javier AV Muñoz, Otávio AB Penatti, Rodrigo T Calumby, Lin Tzy Li, Jefersson A dos Santos, and Ricardo da S Torres. 2018. Exploiting ConvNet Diversity for Flooding Identification. *IEEE Geoscience and Remote Sensing Letters* 15, 9 (2018), 1446–1450.

[12] Konstantin Pogorelov, Sigrun Losada Eskeland, Thomas de Lange, Carsten Griwodz, Kristin Ranheim Randel, Håkon Kvale Stensland, Duc-Tien Dang-Nguyen, Concetto Spampinato, Dag Johansen, Michael Riegler, and others. 2017. A holistic multimedia system for gastrointestinal tract disease detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference*. ACM, 112–123.

[13] Konstantin Pogorelov, Olga Ostroukhova, Mattis Jeppsson, Håvard Espeland, Carsten Griwodz, Thomas de Lange, Dag Johansen, Michael Riegler, and Pål Halvorsen. 2018. Deep Learning and Hand-crafted Feature Based Approaches for Polyp Detection in Medical Videos. In *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 381–386.

[14] Konstantin Pogorelov, Olga Ostroukhova, Andreas Petlund, Pål Halvorsen, Thomas de Lange, Håvard Nygaard Espeland, Tomas Kupka, Carsten Griwodz, and Michael Riegler. 2018. Deep learning and handcrafted feature based approaches for automatic detection of angiectasia. In *Biomedical & Health Informatics (BHI), 2018 IEEE EMBS International Conference on*. IEEE, 365–368.

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

[16] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, and Zbigniew Wojna. 2015. Rethinking the inception architecture for computer vision. *arXiv preprint arXiv:1512.00567* (2015).

[17] Tijmen Tieleman and Geoffrey Hinton. 2012. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural networks for machine learning* 4, 2 (2012).

[18] 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.