

Ensemble and Inference based Methods for Flood Severity Estimation Using Visual Data

Mir Murtaza, Muhammad Hanif, Muhammad Atif Tahir, Muhammad Rafi
National University of Computer and Emerging Sciences, Karachi Campus, Pakistan
{K173029,hanif.soomro,atif.tahir,muhammad.rafi}@nu.edu.pk

ABSTRACT

This paper presents the contribution of NUCES DSGP team for the Multimedia Satellite Task at MediaEval 2019. The essential tasks include News Image Topic Disambiguation (NITD) and Multimodal Flood level Estimation (MFLE) from news images. An ensemble based deep learning method has been applied to the News Image Topic Disambiguation task, where data augmentation and transfer learning were used for binary classification of images. During training, the challenge of class imbalance is managed by using data augmentation technique and selection of equal sample size from each class. For Multimodal Flood Level Estimation task, person's lower body keypoints were detected along with image flood probability scores from two deep convolutional network architectures, namely ResNet50 and VGG19. The confidence scores of detected keypoints and the convolutional networks' output probabilities were combined and were passed to a Random Forest classifier for a final prediction score. The evaluation of the proposed methods for the test set of NITD task revealed a 0.895 F1-Micro score (3rd best score), while the evaluation of MFLE task provided 0.734 F1-Macro score (2nd best score).

1 INTRODUCTION

Effective and efficient flood response system requires timely information about the event. The risk of damage could be reduced by appropriate actions on the basis of inferred information. The data collected through different mediums including news articles is easily available and could be used for disaster response system. The NITD task of "Multimedia Satellite Task at MediaEval, 2019" [1] provided the challenge of developing binary image classifier by using images published in different articles. The classifier method should predict whether or not the topic of article in which particular image was included, discussing any water related disaster.

In the MFLE task, we had to determine whether at least one person in the image is standing in water, and whether or not the level of water is above the knee. To solve this problem, we considered both global and local perspective of an image. Our approach was to first detect and localize all person(s) present in an image. Then a dataset was prepared which contains persons' bounding boxes from all of the images. The ResNet50 Convolutional Neural Network was fine-tuned to extract flood probability of the bounding box. The label of the image was attached to all bounding box(es) of person(s) present in the image. Then we stacked ResNet50 probability with confidence score of four lower body keypoints of interest: namely Right Knee, Left Knee, Right Ankle, Left Ankle of a person

and a score that indicated belief of calculated confidence scores. We also added an image level flood evidence probability obtained from VGG19 pre-trained on the Places365 dataset and fine-tuned on Multimedia Satellite 2017 flood images. This flood probability is same for all the detected person(s) in an image. For inference, we trained a random forest on keypoints scores and flood probability scores obtained for each person bounding box. If the person with the maximum probability exceeds 0.50, the image is classified as positive instance. Our inspiration for this work comes from efforts on text and image based data from different sources to identify type and intensity of disasters [5, 9, 11, 15].

2 APPROACH

Two different deep learning based approaches were used for both tasks. The details of approaches are given below:

News Image Topic Disambiguation (NITD): An ensemble based deep learning approach has been adopted for the task of Image-based News Topic Disambiguation, MediaEval, 2019. The dataset for the task contains 5181 images for training and 1296 images for testing purpose. There were 564 images related to first class and 4617 images for second class. An ensemble based method has been implemented for the specified binary image classification task. Initially, class imbalanced problem has been solved by using data augmentation technique. The Augmentor [3] library has been used to create multiple copies of minority class by using different parameters, including rotate, flip and zoom. After balancing both classes, 3000 images from each class were randomly selected for training. Visual Geometry Group (VGG16) [14] classification model pretrained on Hybrid dataset (Places365 [9] and ImageNet [7]) has been used for the purpose of classification. Visual Geometry Group (VGG16) is a neural network, launched during ImageNet Large Scale Visual Recognition Challenge (ILSVRC), 2014 and secured first and second positions in image localization and image classification tasks, respectively. Dropout ratio was set as 0.3, which was added after the first dense layer. And a batch normalization layer was added after the dropout layer. The last softmax layer was replaced with a sigmoid unit for binary classes. The last 6 layers of the model were retrained on dataset of the task and remaining layers were fixed during the training. The parameters of the model were optimized using the Adam optimizer, with learning rate of $10e-6$. The model was retrained for 40 epochs and finally the trained model was applied for prediction on 1296 test images. Similarly, five models were trained and majority voting was applied for the final prediction, as depicted in Figure 1.

Multimodal Flood Level Estimation from news (MFLE): For the prediction of at least one person standing in water level above

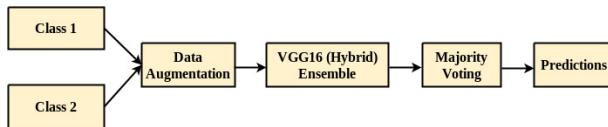


Figure 1: Data Flow diagram showing process for the task of News Image Topic Disambiguation

the knee in an image, we hypothesized that "a detected person in the flooded region with low knees and ankles visibility is a positive case". For this purpose, we prepared a separate dataset which consists of all the detected persons' bounding boxes from their respective images. The label attached to a person bounding box was same as that of the parent image. For person's bounding box detection, the Faster-RCNN model [13] is used, an end-to-end architecture that uses a Region Proposal Network to select regions of interest along with VGG16 Deep CNN for classification. We fine-tuned the last convolutional layer of the ImageNet ResNet50 CNN architecture on these detected person patches of the training images and replaced the last softmax layer with a sigmoid unit. As the dataset was highly imbalanced, balanced weights were used which is an effective method to train Convolutional Neural Network models on imbalanced data [6]. The purpose of using ResNet is to get a local estimate of the flood level in the person's bounding box.

Afterwards, for the persons bounding boxes dataset, we obtained the visibility confidence scores for four lower body keypoints i.e Right Knee, Left Knee, Right Ankle, Left Ankle of a person and an overall score indicating belief in the keypoints calculation. These probabilities help in estimation of the visibility part of our hypothesis. We utilized Fang et. al [8] approach of Regional Multi-person Pose Estimation (RMPE) for keypoint detection. The development dataset had images with no flood evidence. Therefore, we quantified the global flood evidence by acquiring flood evidence probability from Places365 VGG19 [16] fine-tuned on MediaEval Multimedia Satellite 2017 dataset [2]. All persons bounding boxes of an image get same global flood evidence probability.

Then a Random Forest [4] was trained with 321 trees and gini splitting criteria on the feature vectors comprising of ResNet50 local flood probability, Right Knee confidence, Left Knee confidence, Right Ankle confidence, Left Ankle confidence, keypoints Accuracy Score and VGG19 global flood evidence. For inference, we take an image, detect person(s) bounding box(es), then for each person bounding box, pass it through ResNet50 and get a local flood probability. Then we extract confidence probabilities for Right Knee, Left Knee, Right Ankle, Left Ankle of a person using RMPE and the global flood evidence probability using Places365 VGG19. Each person's features set is passed to the trained Random Forest and the final probability of that person being in water above knee-level is obtained. For classification, we select the person in the image having the maximum probability, if the maximum probability is greater than 0.50, the image is classified as positive instance, else a negative one as shown in Figure 2.

3 TEST RESULTS FOR NITD AND MFLE

For all runs of the NITD task, VGG16 ensemble was used. Run 1 and 2 use class weights to handle the class imbalance, Run 3, 4 and 5

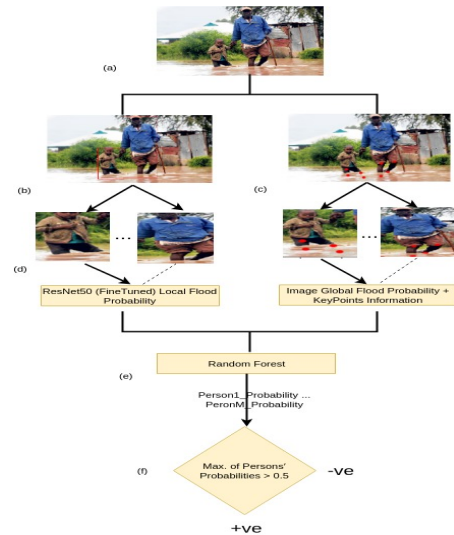


Figure 2: MFLE Inference Pipeline. (a) An image (b) Detected Persons and Bounding Boxes (c) Persons keypoints (d) ResNet50 flood estimation (e) Random Forest Inference of the feature set (f) Max Probability from Persons and Thresholding

use data augmentation. Run 5 uses EasyEnsemble [10, 12] approach for classification. Run 1 of MLFE uses the image feature set and Random Forest classifier.

Table 1 shows the performance of various runs on test set. For NITD, the ensemble of five different VGG16 (Hybrid) models has produced a micro-averaged F1 score of 0.895. For MFLE, the proposed method has produced F1 (macro-averaged) score of 0.734. We got an F1-Score (+ve class) of 0.694 on the MFLE validation set.

Table 1: Test set results from various runs of NITD and MFLE.

Task	Run1	Run2	Run3	Run4	Run5
NITD	0.711	0.718	0.890	0.895	0.747
MFLE	0.734	-	-	-	-

4 CONCLUSION

Various approaches for NITD task have been applied. An approach without data augmentation has been implemented by balancing the minority and majority classes through weights. Also, features with the length of 1365 have been extracted for each image and EasyEnsemble classifier is applied. However, the best performing approach utilized data augmentation technique to increase the quantity of minority class. We did an ablation study of the MFLE inference model. First, when VGG19 global level flood estimation component was removed, the performance deteriorated. Same hold true for the ResNet local level flood estimation and for the Ankle keypoints. For the Random Forest probabilities, we computed different statistics such as mean and median, but maximum of probabilities resulted in better performance on the validation set.

REFERENCES

- [1] Benjamin Bischke, Patrick Helber, Simon Brugman, Erkan Basar, Martha Larson, and Konstantin Pogorelov. 2019. The Multimedia Satellite Task at MediaEval 2019: Estimation of Flood Severity. In Proc. of the MediaEval 2019 Workshop (Oct. 27-29, 2019). Sophia Antipolis, France.
- [2] Benjamin Bischke, Patrick Helber, Christian Schulze, Venkat Srinivasan, Andreas Dengel, and Damian Borth. 2017. 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] Marcus D Bloice, Peter M Roth, and Andreas Holzinger. 2019. Biomedical image augmentation using Augmentor. *Bioinformatics* (04 2019). <https://doi.org/10.1093/bioinformatics/btz259> btz259.
- [4] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.
- [5] Tom Brouwer, Dirk Eilander, Arnejan Van Loenen, Martijn J Booi, 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).
- [6] Mateusz Buda, Atsuto Maki, and Maciej A Mazurowski. 2018. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks* 106 (2018), 249–259.
- [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 *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 248–255.
- [8] Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai, and Cewu Lu. 2017. RMPE: Regional multi-person pose estimation. In *Proceedings of the IEEE International Conference on Computer Vision*. 2334–2343.
- [9] 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.
- [10] Guillaume Lemaitre, Fernando Nogueira, and Christos K. Aridas. 2017. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *Journal of Machine Learning Research* 18, 17 (2017), 1–5. <http://jmlr.org/papers/v18/16-365.html>
- [11] Zhenlong Li, Cuizhen Wang, Christopher T Emrich, and Diansheng Guo. 2018. A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods. *Cartography and Geographic Information Science* 45, 2 (2018), 97–110.
- [12] Xu-Ying Liu, Jianxin Wu, and Zhi-Hua Zhou. 2008. Exploratory under-sampling for class-imbalance learning. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 39, 2 (2008), 539–550.
- [13] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*. 91–99.
- [14] Karen Simonyan and Andrew Zisserman. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations*.
- [15] Nataliya Tkachenko, Stephen Jarvis, and Rob Procter. 2017. Predicting floods with Flickr tags. *PLoS one* 12, 2 (2017), e0172870.
- [16] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. 2017. Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence* 40, 6 (2017), 1452–1464.