Disaster based Visual Sentiment Analysis using Deep Learning

Mohsin Ali, Muhammad Hanif, Muhammad Atif Tahir, Muhammad Nouman Durrani, Muhammad Rafi
National University of Computer and Emerging Sciences, Karachi Campus, Pakistan
{mohsin.ali,hanif.soomro,atif.tahir,muhammad.nouman,muhammad.rafi}@nu.edu.pk

ABSTRACT

In case of a disaster, a large number of relevant and irrelevant images are propagated through social networks. In this case, the sentiment identification of such disasters is important to speed up relief work in the affected region. This paper describes the contribution of FAST-NU-DS team for the Visual Sentiment Analysis: A Natural Disaster MediaEval Use-case [4] held at MediaEval 2021. Various pre-trained deep learning-based models for the single-label classification and multi-label classification tasks have been used for feature extraction and classification. Data augmentation techniques to over-sample minority classes were used to deal with the inherent imbalance nature of the dataset. For a single-label and multi-label classification of tasks, VGG16 proved to be more useful than ResNet50. In this work, we achieved a 0.65 weighted F1 score for the first single label classification subtask and 0.54 and 0.41 weighted-F1 scores for the second and third multi-label subtasks, respectively.

1 INTRODUCTION

Disaster creates a difficult situation to handle, which may harm valuable resources and loss of human lives. Government, NGOs and the public use different social networks to propagate relevant and irrelevant information about the natural calamity in the form of images, videos, and posts to aware others. In literature, researchers have more focused on text-based sentiment analysis using NLP. However, images and videos sentiment analysis using ML models is an open research problem and need attention to identify sentiments communicated through images. These images may reveal emotional responses. Hence, careful identification of disaster the sentiment is important to stop, aware and control any miss lead. For example, this image-based sentiment analysis may be used for rapid identification of situational awareness during disaster and assistance in restoration activities. In addition, the categorization of such images may further be used to understand the adversity of the situation. In this work, we have worked on the "Visual Sentiment Analysis: A Natural Disaster Use-case at MediaEval 2021". The research has performed single-label and multi-label classification to identify visual sentiments that occurred during disasters [4].

2 RELATED WORK

Visual Sentiment Analysis: A Natural Disaster Use-case task of MediaEval, 2021, involves multi-class and multi-label classification tasks. Various similar studies have been performed which has focused on visual sentiment analysis.

MediaEval'21, December 13-15 2021, Online

The research effort has focused on disaster-related images from social media and also implemented various deep learning-based methods [3]. Researchers have used crowd-sourcing to annotate images in multi-label classes, where one image may be a part of one or more classes based on visual sentiments. Moreover, researchers have also implemented various deep learning based models, which were pre-trained on datasets of ImageNet and Places [8] [3].

Another research effort has proposed a framework that considers both text and image-based visual sentiment analysis [1]. The framework has analyzed geo-tagged data objects from disaster-related social media images. The framework is partitioned into sentiment analysis, geo-sentiment modelling, and spatial-temporal partitioning. Moreover, the research has extracted data from Twitter and Flicker, which is related to Napa Earthquake and Hurricane Sandy [1].

Similarly, a class-specific residual attention module (CSRA) has been proposed, which has an extremely simple and efficient model, It requires fewer resources for training and achieved the state of the art results for various datasets of multi-label classification of images[9]. Recently, ensemble-based approaches, such as bagging, boosting and stacking have been discussed by the researchers in image classification [7]

3 PROPOSED APPROACH

The dataset for the task of "Visual Sentiment Analysis A Natural Disaster Use-case" at MediaEval, 2021, contains 2432 images[4]. The dataset has been used for three different tasks. In the first task, single-label classification is performed among three classes: positive, negative, and neutral. While second and third subtasks involve multi-label classification, which contains 7 and 10 classes, respectively.

3.1 Proposed approach for Subtask 1

The proposed method for the first subtask of single label classification has been designed by performing different experiments to select image augmentation technique and appropriate deep learning model.

As the dataset used for the first subtask is imbalanced. The negative class contains 1695 images, and the positive class includes 648 images. However, the neutral class includes 89 images, which is significantly less than the positive and negative classes. Three different methods are utilized to manage the challenge of class imbalance, including weight assignment to classes, oversampling, and image augmentation. In the first attempt, different weights are allocated to all three classes, so that they can be balanced.

- Negative: 0.47
- Positive: 1.26
- Neutral: 9.2

Another class balancing effort has been performed by increasing the number of images by simply using oversampling. The oversampling technique increases copies of instances of minority class and makes them equal to majority class. The method has proved better in comparison to the weight balancing technique. In last, the class imbalance has been reduced by using the data augmentation technique, in which variants of a single image are created by using different augmentation techniques. Moreover, augmentation techniques including random shift, random flip, random brightness, and random zoom are applied to increase the number of images in minority classes and make them equal in quantity. The data augmentation technique has provided the best results in tackling class imbalance problems compared to oversampling and class weight assignment techniques.

There are two deep learning based pre-trained models are selected for the experiments, including Visual Geometry Group (VGG) [6] and ResNet50 [5]. Both of the networks are used by pre-training on the ImageNet [2] dataset. The experiments on the training set revealed that VGG16 had produced a higher F1-score than ResNet50. Hence VGG16 has been selected to be implemented to predict unseen test instances. Moreover, the ImageNet dataset carries weights, which focuses on objects, while the visual sentiment analyzer requires scenario-based information. Due to this, the last six layers of the model are unfrozen so that retraining can be performed using the visual sentiment analysis dataset. Also, remaining of the layers are frozen to avoid their retraining. During training, various hyperparameters have been experimented, and the best combination is applied for the training of the model. The learning rate for the model is set as 10^{-4}, and the *softmax* activation function has been used for the processing. The quantity of epochs is set automatically by applying early-stopping based on the best F1 score.

To improve the efficiency of experiments, all the experiments are initially performed by converting images into grayscale, which has reduced the processing time for the method. After selecting optimal values for hyperparameters, coloured images are used to further improve the method's performance. The trained model is then applied for the prediction of test set instances, which are 1199 images.

3.2 Proposed approach for Subtask 2 and 3

Subtask 2 and *Subtask 3* are aimed to classify images based on multi-label image classification, where one image can be assigned to various classes, according to its depicted emotions. According to subtask 2, the image may belong to one or more classes, including anger, disgust, joy, fear, neutral, surprise, and sadness. However, for subtask 3, the image may acquire one or more classes, and it is also a multi-label classification task. The difference between subtask 2 and subtask 3 is the number of classes containing 7 and 10 classes, respectively.

The dataset is imbalanced for subtask 2 and subtask 3, and few classes contain more images than other classes. The oversampling technique is utilized to increase the number of images in minority classes to reduce the class imbalance.

After balancing the classes, the VGG16 pre-trained on ImageNet is fine-tuned on the dataset for subtask 2 and subtask 3. Moreover, the *sigmoid* activation function is used to predict multi-label classification.

4 RESULTS AND ANALYSIS

At the initial stage, VGG16 and ResNet50 have been experimented for subtask 1. The implementation has been performed on three different data balancing techniques: weight balancing, oversampling, and image augmentation. The weighted F1 scores on training data for the first subtask are shown in Table 1.

Table 1: F1-Score on training data of subtask 1

Balancing Technique	F1 Score(ResNet50)	F1 Score(VGG16)
Weighted Class	60.15	61.93
Over Sampling	63.17	67.33
Image Augmentation	66.85	**67.49**

The training experiments have proved VGG16 as a better pre-trained model and image augmentation as the best class balancing technique. By considering better performance, the VGG16 has been trained on the whole dataset and used to predict test data. The image augmentation technique is used for class balancing in subtask 1. However, for subtask 2 and subtask 3, the oversampling technique increases the number of minor classes. The results for test-set are visualized in Table 2.

Table 2: Results achieved by proposed approach on Test-set

Task	Model	Balancing Technique	F1 Score
Task 1	VGG-16	Image Augmentation	65.37
Task 2	VGG-16	Oversampling	54.24
Task 3	VGG-16	Oversampling	41.74

5 CONCLUSION

The research has proposed a deep learning-based model for single-label and multi-label classification tasks to analyze visual sentiments during disastrous conditions. The approach has tried various class balancing techniques and pre-trained models. Furthermore, the research can be extended by using weights from the Places dataset [8], which involves scene-level information and may produce better performance. Moreover, for multi-label classification, the image augmentation technique may be used to over-sample the minority classes.

ACKNOWLEDGMENTS

This work was supported in part by the Smart Video Surveillance Lab, an affiliated Laboratory of (NCBC), FAST-National University of Computer Emerging Science.

REFERENCES

[1] Abdullah Alfarrarjeh, Sumeet Agrawal, Seon Ho Kim, and Cyrus Shahabi. 2017. Geo-spatial multimedia sentiment analysis in disasters. In *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 193–202.

[2] 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.

[3] Syed Zohaib Hassan, Kashif Ahmad, Steven Hicks, Pål Halvorsen, Ala Al-Fuqaha, Nicola Conci, and Michael Riegler. 2020. Visual sentiment analysis from disaster images in social media. *arXiv preprint arXiv:2009.03051* (2020).

[4] Syed Zohaib Hassan, Kashif Ahmad, Michael Riegler, Steven Hicks, Nicola Conci, Pål Halvorsen, and Ala Al-Fuqaha. 2021. Visual Sentiment Analysis: A Natural Disaster Use-case Task at MediaEval 2021. In *Proceedings of the MediaEval 2021 Workshop, Online*.

[5] 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.

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

[7] Muhammd Waqas, Muhammad Atif Tahir, and Rizwan Qureshi. 2021. Ensemble-Based Instance Relevance Estimation in Multiple-Instance Learning. In *2021 9th European Workshop on Visual Information Processing (EUVIP)*. IEEE, 1–6.

[8] 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.

[9] Ke Zhu and Jianxin Wu. 2021. Residual Attention: A Simple but Effective Method for Multi-Label Recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 184–193.