Sentiment Analysis from Images of Natural Disasters

Syed Zohaib Hassan[1], Kashif Ahmad[2(✉)], Ala Al-Fuqaha[2], and Nicola Conci[1]

[1] University of Trento, Trento, Italy
syedzohaib.hassan@studenti.unitn.it, nicola.conci@unitn.it
[2] Hamad Bin Khalifa University, Doha, Qatar
{kahmad,aalfuqaha}@hbku.edu.qa

Abstract. Social media have been widely exploited to detect and gather relevant information about opinions and events. However, the relevance of the information is very subjective and rather depends on the application and the end-users. In this article, we tackle a specific facet of social media data processing, namely the sentiment analysis of disaster-related images by considering people's opinions, attitudes, feelings and emotions. We analyze how visual sentiment analysis can improve the results for the end-users/beneficiaries in terms of mining information from social media. We also identify the challenges and related applications, which could help defining a benchmark for future research efforts in visual sentiment analysis.

Keywords: Sentiment analysis · Natural disasters · Multi-label classification · CNNs · Social media

1 Introduction

Sudden and unexpected adverse events, such as floods and earthquakes, may not only damage the infrastructure but also have a significant impact on people's physical and mental health. In such events, an instant access to relevant information might help to identify and mitigate the damage. To this aim, information available on social networks can be utilized for the analysis of the potential impact of natural or man-made disasters on the environment and human lives [1].

Social media outlets along with other sources of information, such as satellite imagery and Geographic Information Systems (GIS), have been widely exploited to provide a better coverage of natural and man-made disasters [2,16]. The majority of the approaches rely on computer vision and machine learning techniques to automatically detect disasters, collect, classify, and summarize relevant information. However, the interpretation of *relevance* is very subjective and highly depends on the application framework and the end-users.

In this article, we analyze the problem from a different perspective and focus in particular on sentiment analysis of disaster-related images. Specifically, we

consider people's opinions, attitudes, feelings, and emotions toward the images related to the event by estimating the emotion/perceptual content evoked by a generic image [7,9,14]. We aim to explore and analyze how the visual sentiment analysis of such images can be utilized to provide more accurate description of adverse events, their evolution, and consequences. We believe that such analysis can serve as an effective tool to convey public sentiments around the world while reducing the bias of news organizations. This can lead to new beneficiaries beyond the general public (e.g., online news, humanitarian organizations, non-governmental organizations, etc.).

The concept of sentiment analysis has been utilized in Natural Language Processing (NLP) and in a wide range of application domains, such as education, entertainment, hosteling and other businesses [15]. On the other hand, Visual sentiment analysis is relatively new and less explored. A large portion of the literature on visual sentiment/emotion recognition relies on facial expressions [3], where face-close up images are analyzed to predict a person's emotions. More recently, the concept of emotion recognition has been extended to relatively more complex images having multiple objects and background details. Thanks to the recent advances in deep learning, encouraging results have been recently obtained [6,18].

In this article, we analyze the role of visual sentiment analysis in complex disaster-related images. To the best of our knowledge, no prior work analyzes disaster-related imagery from this prospective. We also identify the challenges and potential applications with the objective of setting a benchmark for future research on visual sentiment analysis.

The main contributions of this work can be summarized as follows:

– We extend the concept of visual sentiment analysis to disaster-related visual contents, and identify the associated challenges and potential applications.
– In order to analyze human's perception and sentiments about disasters, we conducted a crowd-sourcing study to obtain annotations for the experimental evaluation of the proposed visual sentiment analyzer.
– We propose a multi-label classification framework for sentiment analysis, which also helps in analyzing the correlation among sentiments/tags.
– Finally, we conduct experiments on a newly collected dataset to evaluate the performance of the proposed visual sentiment analyzer.

The rest of the paper is organized as follows: Sect. 2 provides detailed description of the related work; Sect. 3 describes the proposed methodology; Sect. 4 provides detailed description of the experimental setup, conducted experiments, and detailed analysis of the experimental results; Sect. 5 provides concluding remarks and identifies directions of future research.

2   Related Work

In contrast to other research domains, such as NLP, the concept of sentiment analysis is relatively new in visual content analysis. The research community

has demonstrated an increasing interest in the topic and a variety of techniques have been proposed with particular focus on the feature extraction and classification strategies. The vast majority of the efforts in this regard aim to analyze and classify face-closeup images for different types of sentiments/emotions and expressions. Busso et al. [3] rely on facial expressions along with speech and other information in a multimodal framework. Several experiments have been conducted to analyze and compare the performance of different sources of information, individually and in different combination, in support of human emotions/sentiment recognition. A multimodal information based approach has also been proposed in [18], where facial expressions are jointly utilized with textual and audio features that are extracted from videos. Facial expressions are extracted through the Luxand FSDK 1.7[1] open source library along with GAVAM features [19]. Textual and audio features are extracted through the Sentic computing paradigm [4] and OpenEAR [8], respectively. Next, different feature and decision-level fusion methods are used to jointly exploit the visual, audio, and textual information for the task.

More recently, the concept of emotion/sentiment analysis has been extended to more complex images involving multiple objects and background details [6,7,12,22]. For instance, Wang et al. [23] rely on mid and low-level visual features along with textual information for sentiment analysis in social media images. Chen et al. [6] proposed DeepSentiBank, a deep convolutional neural network-based framework for sentiment analysis of social media images. To train the proposed deep model, around one million images with strong emotions have been collected from Flickr. In [22], Deep Coupled Adjective and Noun neural networks (DCAN), is proposed for sentiment analysis without the traditional Adjective Noun Pairs (ANP) labels. The framework is composed of three different networks, each aiming to solve a particular challenge associated with sentiment analysis. Some methods also utilized existing pre-trained models for sentiment analysis. For instance, Campose et al. [5] fine-tuned CaffeNet [11], on a newly collected dataset for sentiment analysis conducting experiments to analyze the relevance of the features extracted through different layers of the network. In [17] existing pre-trained CNN models are fine-tuned on a self-collected dataset. The dataset contains images from social media, which are annotated through a crowdsourcing activity involving human annotators. Kim et al. [12] also rely on the transfer learning techniques for their proposed emotional machine. Object and scene-level information, extracted through deep models pre-trained on ImageNet and Places datasets, respectively, have been jointly utilized for this purpose. Color features have also been employed to perceive the underlying emotions.

3   Proposed Methodology

Figure 1 provides the block diagram of the framework implemented for visual sentiment analysis. As a first step, social media platforms are crawled for disaster-related images using different keywords (floods, hurricanes, wildfires, droughts,

---

[1] https://www.luxand.com/facesdk/.

landslides, earthquakes, etc.). The downloaded images are filtered manually and a selected subset of images are considered for the crowd-sourcing study in the second step where a large number of participants tagged the images. A CNN and a transfer learning method is used for multi-label classification to automatically assign sentiments/tags to images. In the next subsections, we provide a detailed description of the crowd-sourcing activity and the proposed visual deep sentiment analyzer.
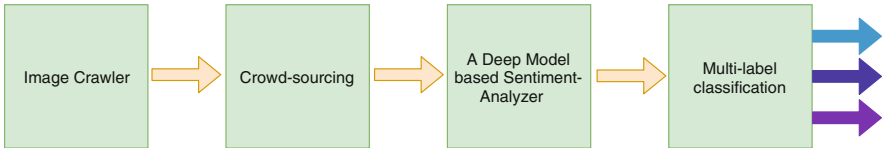

Fig. 1. Block diagram of the proposed framework for visual sentiment analysis.

3.1 The Crowd-Sourcing Study

In order to analyze human's perception and sentiments about disasters and how they perceive disaster-related images, we conducted a crowd-sourcing study. The study is carried out online through a web application specifically developed for the task, which was shared with participants including students from University of Trento (Italy), and UET Peshawar (Pakistan) as well as with other contacts with no scientific background. Figure 2 provides an illustration of the platform we used for the crowd-sourcing study. In the study, participants were provided with a disaster-related image, randomly selected from the pool of images, along with a set of associated tags. The participants were then asked to assign a number of suitable tags, which they felt relevant to the image. The participants were also encouraged to associate additional tags to the images, in case they felt that the provided tags were not relevant to the image.

One of the main challenges in the crowd-sourcing study was the selection of the tags/sentiments to be provided to the users. In the literature, sentiments are generally represented as *Positive*, *Negative* and *Neutral* [15]. However, considering the specific domain we are addressing (natural and man-made disasters) and the potential applications of the proposed system, we are also interested in tags/sentiments that are more specific to adverse events, such as pain, shock, and destruction, in addition to the three common tags. Consequently, we opted for a data-driven approach, by analyzing users' tags associated with disaster images crawled form social media outlets. Apart from the sentimental tags, such as *pain*, *shock* and *hope*, we also included some additional tags, such as *rescue* and *destruction*, which are closely associated with disasters and can be useful in different applications utilized by online news agencies, humanitarian, and non-governmental organizations (NGOs). The option for adding additional tags also helps to take the participants' viewpoints into account.
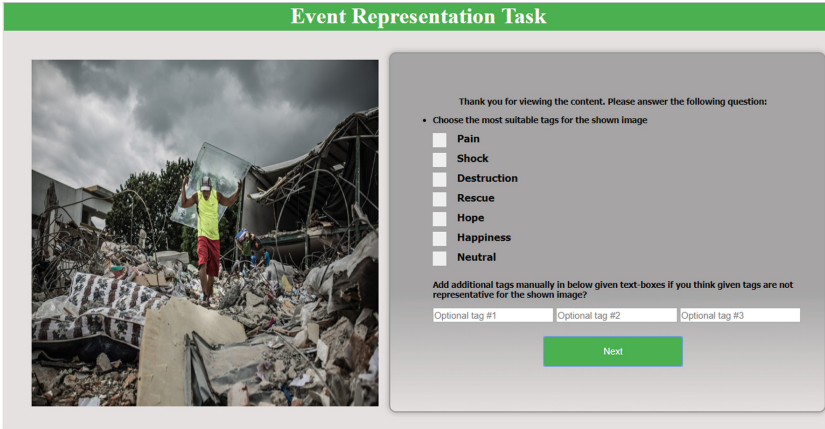
Fig. 2. Illustration of the platform used for the crowd-sourcing study. A disaster-related image and several tags are presented to the users for association. The users' are also encouraged to provide additional tags.

Table 1. Statistics of the crowd-sourcing study in terms of the total number of times each tags has been associated with images by the participants.

Sentiments/tags	Count
Destruction	871
Happiness	145
Hope	353
Neutral	214
Pain	454
Rescue	694
Shock	354

The crowd-sourcing activity was carried out on 400 images related to 6 different types of disasters: earthquakes, floods, droughts, landslides, thunderstorms, and wildfires. In total, we obtained 2,587 responses from the users, with an average of 6 users per image. We made sure to have at least 5 different users for each image. Table 1 provides the statistics of the crowd-sourcing study in terms of the total number of times each tag has been associated with images by the participants. As can be seen in Table 1, some tags, such as *destruction*, *rescue* and *pain*, are used more frequently compared to others.

During the analysis of the responses from the participants, we observed that certain tag pairs have been used to describe images. For instance, pain and destruction, hope and rescue, shock and pain, are used several times jointly. Similarly, shock, destruction and pain have been used jointly 59 times. The three tags: rescue, hope, and happiness, are also used often together. This correlation

among the tag/sentiment pairs provides the foundation for our multi-label classification, as opposed to single-label multi-class classification, of the sentiments associated with disasters-related images. Figure 3 shows the number of times the sentiments/tags are used together by the participants in the crowd-sourcing activity. For final annotation, the decision is made on the basis of majority votes from the participants of the crowd-sourcing study.
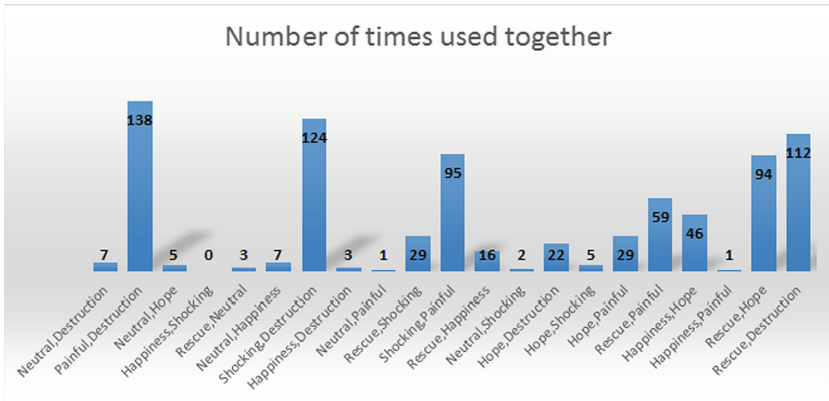
Fig. 3. Correlation of tag pairs: number of times different tag pairs used by the participants of the crowd-sourcing study to describe the same image.

3.2 The Visual Sentiment Analyzer

The proposed framework for visual sentiment analysis is inspired by the multi-label image classification framework[2] and is mainly based on a Convolutional Neural Network (CNN) and a transfer learning method, where the model pre-trained on ImageNet is fine-tuned for visual sentiment analysis. In this work, we analyze the performance of several deep models such as AlexNet [13], VggNet [20], ResNet [10] and Inception v-3 [21] as potential alternatives to be employed in the proposed visual sentiment analysis framework.

The multi-label classification strategy, which assigns multiple labels to an image, better suits our visual sentiment classification problem and is intended to show the correlation of different sentiments. In order for the network to fit the task of visual sentiment analysis, we introduced several changes to the model as will be described in the next paragraph.

---

[2] https://towardsdatascience.com/multi-label-image-classification-with-inception-net-cbb2ee538e30.

3.3   Experimental Setup

In order to fit the pre-trained model to multi-label classification, we create a ground truth vector containing all the labels associated with an image. We also made some modifications in the existing pre-trained Inception-v3 [21] model by extending the classification layer to support multi-label classification. To do so, we replaced the soft-max function, which is suitable for single-label multi-class classification, and squashes the values of a vector into a *[0,1]* range holding the total probability, with a sigmoid function. The motivation for using a sigmoid function comes from the nature of the problem, where we are interested to express the results in probabilistic terms; for instance, an image belongs to the class *shock* with 80% probability and to class *destruction* and *pain* with 40% probability. Moreover, in order to train the multi-label model properly, the formulation of the cross entropy is also modified accordingly (i.e., replacing softmax with sigmoid function). For the multiple labels, we modify the top layer to obtain posterior probabilities for each type of sentiment associated with an underlying image.

The dataset used for our experimental studies has been divided into training (60%), validation (10%), and evaluation (30%) sets.

4   Experiments and Evaluations

The basic motivation behind the experiments to provide a baseline for the future work in the domain. To this aim, we evaluate the proposed multi-label framework for visual sentiment analysis using several existing pre-trained state-of-the-art deep learning models including: AlexNet, VggNet, ResNet, and Inception v3. Table 2 provides the experimental results obtained using these deep models.

Table 2. Evaluation of the proposed visual sentiment analyzer with different deep learning models pre-trained on ImageNet.

Model	Accuracy (%)
AlexNet	79.69
VggNet	79.58
Inception-v3	80.70
ResNet	78.01

Considering the complexity of the task and the limited amount of training data, the obtained results are encouraging. Though there's no significant difference in the performance of the models, slightly better results are obtained with Inception-v3 models. Lowest accuracy has been observed for ResNet, but such reduction in the performance could be due to the size of the dataset used for the study.

In order to show the effectiveness of the proposed visual sentiment analyzer, we also provide some sample output images in Fig. 4, showing the output of

the proposed visual sentiment analyzer in terms of the percentage/probabilities for each label. Table 3 provides the statistics for these samples in terms of the probability for each label and probabilities/percentages computed through human annotators. Due to space limitation, only four samples are provided in the paper to give an idea about the performance of the method. For this particular qualitative analysis, we converted the responses of the participants of the crowd sourcing study into percentages (i.e., the degree to which each image belongs to a particular label) for each label associated with each image. These percentages are different from the ground truth used during training and evaluation where images were assigned labels on a majority voting basis. For instance, the percentages based on the responses of the crowd sourcing study for the first image (leftmost in Fig. 4) are: $destruction = 0.10$, $happiness = 0.0$, $hope = 0.10$, $neutral = 0.0$, $pain = 0.35$, $rescue = 0.30$ and $shock = 0.20$ while the output of the proposed visual sentiment analyzer in terms of probabilities for each label/class are: $destruction = 0.16$, $happiness = 0.04$, $hope = 0.06$, $neutral = 0.02$, $pain = 0.58$, $rescue = 0.28$ and $shock = 0.17$. In most of the cases, the proposed model provides results that are similar to the percentages obtained from the users' responses, demonstrating the effectiveness of the proposed method.

Fig. 4. Some sample output of the proposed visual sentiment analyzer.

Table 3. Sample outputs in terms of ground truth obtained from users in terms of percentage in the crowd-sourcing study vis-a-vis predicted probabilities.

Image	Destruction		Happiness		Hope		Neutral		Pain		Rescue		Shock	
	GT	Pred.	GT	Pred.	GT	Pred.	GT	Pred.	GT	Pred.	GT	Pred.	GT	Pred.
1	0.10	0.16	0.0	0.04	0.1	0.06	0	0.027	0.35	0.58	0.30	0.28	0.20	0.17
2	0.24	0.24	0.0	0.05	0.0	0.08	0.34	0.36	0.429	0.44	0.514	0.59	0.20	0.33
3	0.167	0.23	0.0	0.05	0.10	0.13	0.16	0.17	0.46	0.59	0.33	0.26	0.0	0.13
4	0.10	0.18	0.0	0.03	0.09	0.05	0.20	0.26	0.0	0.33	0.72	0.72	0.0	0.20

5   Conclusions, Challenges and Future Work

In this paper, we addressed the challenging problem of visual sentiment analysis of disaster-related images obtained from social media. We analyzed how people

respond to disasters and obtained their opinions, attitudes, feelings, and emotions toward the disaster-related images through a crowd-sourcing activity. We show that the visual sentiment analysis/emotions recognition, though a challenging task, can be carried out on more complex images using some deep learning techniques. We also identified the challenges and potential applications of this relatively new concept, which is intended to set a benchmark for future research in visual sentiment analysis.

Though the experimental results obtained during the initial experiments on the limited dataset are encouraging, the task is challenging and needs to be investigated in more details. Specifically, the reduced availability of suitable training and testing images is probably the biggest limitation. Since visual sentiment analysis aims to present human's perception of an entity, crowd-sourcing seems to be a valuable option to acquire training data for automatic analysis. In terms of visual features, we believe that object and scene-level features can play complementary roles in representing the images. Moreover, multi-modal analysis will further enhance the performances of the proposed sentiment analyzer. This suggests that within the domain of purely visual information, the conveyed information can differ, suggesting that the interpretation of the image is subject to change depending on the level of detail, the visual perspective, and the intensity of colors. We expect these elements to play a major role in the evolution of frameworks like the one we have presented, and when combined with additional media sources (e.g., audio, text, meta-data), can provide a well rounded perspective about the sentiments associated with a given event.

References

1. Ahmad, K., Pogorelov, K., Riegler, M., Conci, N., Halvorsen, P.: Social media and satellites. Multimed Tools Appl. **78**, 1–39 (2018)
2. Ahmad, K., et al.: Automatic detection of passable roads after floods in remote sensed and social media data. Signal Process.: Image Commun. **74**, 110–118 (2019)
3. Busso, C., et al.: Analysis of emotion recognition using facial expressions, speech and multimodal information. In: Proceedings of the 6th International Conference on Multimodal Interfaces, pp. 205–211. ACM (2004)
4. Cambria, E., Hussain, A., Havasi, C., Eckl, C.: Sentic computing: exploitation of common sense for the development of emotion-sensitive systems. In: Esposito, A., Campbell, N., Vogel, C., Hussain, A., Nijholt, A. (eds.) Development of Multimodal Interfaces: Active Listening and Synchrony. LNCS, vol. 5967, pp. 148–156. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-12397-9_12
5. Campos, V., Salvador, A., Giro-i Nieto, X., Jou, B.: Diving deep into sentiment: understanding fine-tuned CNNs for visual sentiment prediction. In: Proceedings of the 1st International Workshop on Affect & Sentiment in Multimedia, pp. 57–62. ACM (2015)
6. Chen, T., Borth, D., Darrell, T., Chang, S.F.: DeepSentiBank: visual sentiment concept classification with deep convolutional neural networks. arXiv preprint arXiv:1410.8586 (2014)
7. Constantin, M.G., Redi, M., Zen, G., Ionescu, B.: Computational understanding of visual interestingness beyond semantics: literature survey and analysis of covariates. ACM Comput. Surv. (CSUR) **52**(2), 25 (2019)

8. Eyben, F., Wöllmer, M., Schuller, B.: OpenEAR–introducing the Munich open-source emotion and affect recognition toolkit. In: 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, pp. 1–6. IEEE (2009)

9. Gygli, M., Grabner, H., Riemenschneider, H., Nater, F., Van Gool, L.: The interestingness of images. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1633–1640 (2013)

10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

11. Jia, Y., et al.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia, pp. 675–678. ACM (2014)

12. Kim, H.R., Kim, Y.S., Kim, S.J., Lee, I.K.: Building emotional machines: recognizing image emotions through deep neural networks. IEEE Trans. Multimed. **20**, 2980–2992 (2018)

13. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)

14. Machajdik, J., Hanbury, A.: Affective image classification using features inspired by psychology and art theory. In: Proceedings of the 18th ACM International Conference on Multimedia, pp. 83–92. ACM (2010)

15. Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: a survey. Ain Shams Eng. J. **5**(4), 1093–1113 (2014)

16. Nogueira, K., et al.: Exploiting convnet diversity for flooding identification. IEEE Geosci. Remote Sens. Lett. **15**(9), 1446–1450 (2018)

17. Peng, K.C., Chen, T., Sadovnik, A., Gallagher, A.C.: A mixed bag of emotions: model, predict, and transfer emotion distributions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 860–868 (2015)

18. Poria, S., Majumder, N., Hazarika, D., Cambria, E., Gelbukh, A., Hussain, A.: Multimodal sentiment analysis: addressing key issues and setting up the baselines. IEEE Intell. Syst. **33**(6), 17–25 (2018)

19. Saragih, J.M., Lucey, S., Cohn, J.F.: Face alignment through subspace constrained mean-shifts. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 1034–1041. IEEE (2009)

20. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

21. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)

22. Wang, J., Fu, J., Xu, Y., Mei, T.: Beyond object recognition: visual sentiment analysis with deep coupled adjective and noun neural networks. In: IJCAI, pp. 3484–3490 (2016)

23. Wang, Y., Wang, S., Tang, J., Liu, H., Li, B.: Unsupervised sentiment analysis for social media images. In: Twenty-Fourth International Joint Conference on Artificial Intelligence (2015)