2020 IEEE International Conference on Convergence to Digital World – Quo Vadis (ICCDW 2020)

Emotion Detection and Sentiment Analysis of Static Images

Udit Doshi
Department of Information Technology
Atharva College of Engineering
Mumbai, India
udoshi2498@gmail.com

Vaibhav Barot
Department of Information Technology
Atharva College of Engineering
Mumbai, India
vaibhavbarot04@gmail.com

Sachin Gavhane
Department of Information Technology
Atharva College of Engineering
Mumbai, India
sachingavhane@atharvacoe.ac.in

2020 International Conference on Convergence to Digital World - Quo Vadis (ICCDW) | 978-1-7281-4635-5/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICCDW45521.2020.9318713

Abstract—**The usage of social media platform such as Facebook, Instagram, Flicker, etc. is rising day by day wherein images play a major role. It is said "An image is worth a thousand words", people these days upload certain images on these sites to display their sentiments and emotions in the form of picture on almost every occasion. Images play the most important role in today's generation where it has become a major part of everyone's lives. Most of the prevailing research have focused on sentiment analyses of textual data, but only limited researches have focused on analyzing sentiment of visual data. In this project, we have explored the possibilities of Convolutional Neural Networks (CNN) to predict the various emotions (happiness, surprise, sadness, fear, anger and neutral) depicted by an image. These sort of predictions can be useful in applications for automatic tag predictions of the visual data available on social media platforms and understanding sentiments of the people and their emotions.**

Keywords— convolutional neural network, sentiment analysis, emotion detection, data mining.

I. Introduction

People these days share a large number of contents on social media platforms in the form of images, could be a personal image, scenery, or their opinion in the form of memes. Images have become an integral part of the people in today's time. Every minute detail to a huge number of details can be shared in the form of images. Analysing these contents from social media platform or image-sharing websites like Instagram, Twitter, Flickr, Facebook, etc., can give an idea about the general sentiment of people and the emotions they share.

Sentiment analysing is a field of study which helps to analysis opinions, attitudes, sentiments and emotions of people. It plays a major role in business development and marketing strategies in this growing age of internet. Because of this, sentiment analysis is deserving to be paid more attention [1]. As sentiments are the key influencing factor for almost all human behaviours and activities, systems for analysing sentiments are being used in wide areas such as advertisements recommender, blog recommendation, movie recommendation, virtual marketing [2]. Analysing sentiment of textual contents has evolved to be one of the most active areas of research in Natural Language Processing (NLP). But recently, there has been very limited work which is focused on analysing sentiment of visual contents in computer vision. Use of various kinds of images have started to rise. Moreover, among the users of ubiquitous social media platforms, use of visual contents has been more popular than textual contents.

As images are playing a major role in today's society it is advantageous to understand various emotions and sentiment an image can depict and to automatically predict emotional tags on them - like happiness, love, sadness, anger, neutral, etc. As a part of this project, we aim to predict the emotions of an image that falls into the category - Happiness, Surprise, Sadness, Fear, Anger and Neutral. We have achieved this by implementing our custom convolutional neural network and a pretrained Convolutional Neural Network model for the task of predicting the emotions and analysing sentiments of multiple static images using a custom built GUI. The use of CNN has been increased popularly for sentiment analysis in the recent years [3] [4] [5] [6]. The increase in the field of visual data prediction helped us to explore the area of CNN and how it works on visual sentiment detection and emotion prediction in this paper. To use CNN the most important requirement is a large dataset containing various images to provide a better accuracy. The dataset plays the most important role in the training of the model and for the prediction as it is the core element required for further development.

II. Related Work

A number of work have been carried out on images such as rotating and flipping the input images [7], cropping, etc. to make the dataset clean and free of noise. We have created a custom dataset by collecting large number of images available on the internet and social media platforms such as Facebook, Instagram, Flickr, etc. some parts of the dataset is also collected from ImageNet [8]. The CNN model that we have utilized is the VGGNet16 [9] which has been most widely used in the past years. We also created our custom CNN model with seven layers that we will be comparing it with the VGGNet16 model and the ResNet model to compare the accuracies between them.

Analysing sentiment of people on these online platforms have attracted a significant amount of interest amongst researchers in this field, be it in textual format or visual format. This is because it simply provides ample amount of

978-1-7281-4635-5/20/$31.00 ©2020 IEEE

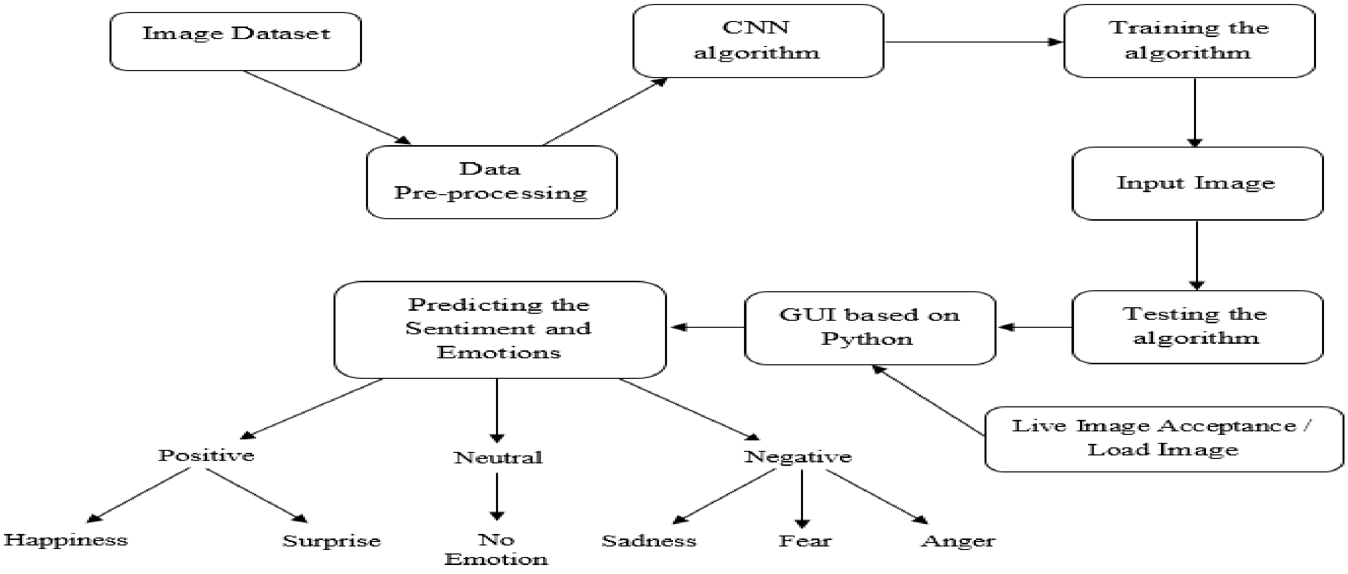


Fig. 1. Overview of the proposed visual sentiment detection and emotion prediction framework

information about a particular person. This is possible because the use of deep learning algorithms has provided a number of accurate results in past several years. Analysing sentiment of textual data have boosted to be one of the most interesting research areas. This is done with the use of NLP [10]. It has provided many convincing results in computing vision and Natural Language Processing (NLP) over past several years. An exhaustive survey on deep learning for sentiment analysis in [10] have also boosted its use. Most researches conducted are basically conducted on textual data rather than visual data. [11] focuses on sentiment analysis of short text that can be beneficial for everyday purposes. They have proposed a framework which combined both CNN and RNN. Paper by Mishra et al. implemented a CNN framework that is able to extract the cognitive features from the human eyes when people are reading certain texts and they have utilized these features with the textual features together to classify sentiments of the reader [12]. [13] constructed a substantial visual sentiment ontology which is consisting of more than 3,000 adjective noun pairs by using web mining technology and psychological theories to detect visual sentiment. [14] uses edge descriptors and MPEG7 colour to execute emotion classification and compares the result with Bag of Emotions method. They introduced a visual sentiment concept classification method based on deep convolutional neutral networks [15]. Taking into consideration it is seen that the research on emotion prediction of visual content have been deeply looked into. Our work focuses on predicting emotions as well as detecting the sentiment depicted by an image and with the use of a GUI which helps in auto tag prediction of any particular image. Details of our work carried out will be explained in the following section.

III. Methodology

In this section, we explain the complete details of our sentiment analysis and emotion predictor framework. As shown in Fig. 1, the dataset is first pre-processed which will help in removing noise and broken data from the large dataset available. Then we adopted the CNN model for the training purpose which will then be able to predict the emotion and sentiment depicted by the images. Then a GUI is used to load a particular image or capture a live image on which the trained model will be used to predict the emotion depicted by the image under the categories of happiness, surprise, sadness, fear, anger and neutral. A detailed information is provided in this section.

A. Data Collection

Data is the most important and crucial part for any sentiment prediction to be carried out. A large amount of data is a must to generate a better accuracy. The data for our model is collected from a number of social media platforms like Facebook, Instagram, Flickr, etc. where images are used on an everyday basis for sharing and passing information rather then the textual data. Some of the images in the dataset are also collected from ImageNet which has a large number of images available to work upon. We have made a custom dataset with these mixture of images available from the online platform.

B. Data Pre-processing

Cleaning the dataset is the first most important task before proceeding with any other steps. So the large number of dataset built had to be completely cleaned first by removing noise from it before moving ahead. We cleaned the damaged images or any repeated images or images which does not specify the specific standards needed. Fig. 2, shows how the pre-processing of the dataset is done and the noise is removed and the dataset can be used for further processing.

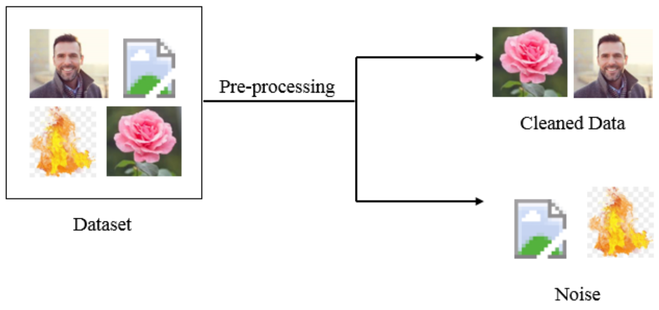


Fig. 2. Pre-processing of data

Further the images with variable sizes are converted into a custom size for better accuracy and faster results. All the images in the dataset are converted in a particular format so that the model can work in an efficient way without any disturbance.

C. CNN Framework

For the training purpose we have used a VGG16 model, ResNet model, both pre-trained and our custom CNN framework trained on our built dataset. Shown in Fig.3 is the VGG16 model which is pre-trained on the ImageNet dataset.

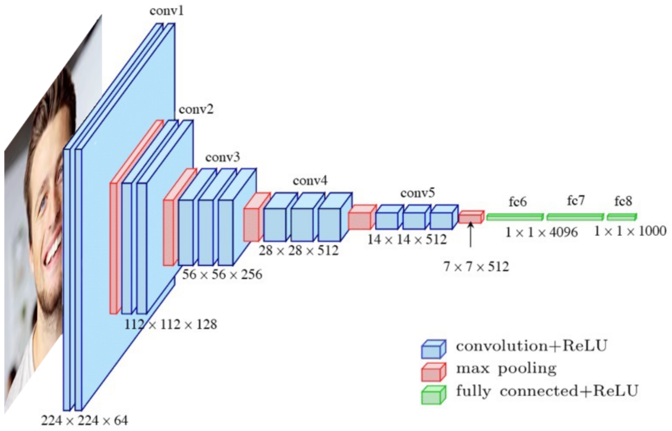


Fig. 3. VGG16 framework

On the other we created our proposed custom CNN model with seven layers to be trained on our own dataset. This model works similar to the VGG16 and the ResNet model but are trained on our custom dataset. The accuracy between these three is taken into consideration and the better performing model will be further used for the prediction of emotion of the visual images through the GUI.

It is also important that more the data a deep learning algorithm uses for training, the more effective are the results. In our custom CNN framework we have used custom dataset to be trained on.

Our custom model uses the CNN framework with additional layers added for a better accuracy for the prediction. This creates better results on the dataset and the prediction is much more accurate.

IV. Experiments and Evaluation

In this section we manifest the experiments conducted on our data. We decided to use the 80-20 principle here, so we split the our training dataset into 80% training dataset and 20% validation dataset. This is done to finally evaluate the accuracy of the model on the data it is never exposed to. This helped us to check whether we are over-fitting on the training dataset and whether we should lower the learning rate and train for more epochs if we get the validation accuracy higher than the training accuracy.

A. Proposed CNN Model with 7 layers

After loading the data and splitting it, we pre-processed them by transforming them into the size needed and what the network expects and scaling them so that all values are in the [0, 1] interval. The data previously consisted of variations in shape and sizes, so we converted them to a standard shape. We observed that with the pre-processing of the dataset increases the accuracy. The CNN with 7 convolutional layer generated an output accuracy of 0.80

B. VGG16 Model

This is a pre-trained model which is trained on the dataset provided by ImageNet. For this model first we poped the last layer and then appended the dense activation (softmax) layer for a better fine tuning. As the model is pre-trained on a ubiquitous range of classes it showed a testing accuracy of 0.35. Which is a bit low then our proposed CNN framework.

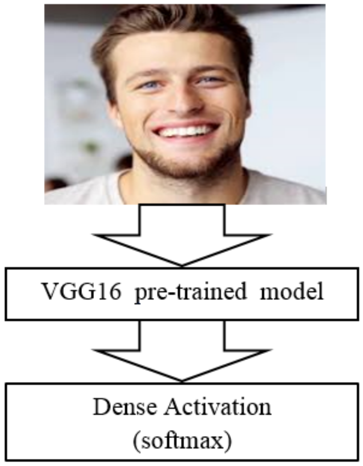


Fig. 4. VGG16 framework with added Dense Activation Layer

C. ResNet-50 Model

The ResNet model is also a pre-tained model which is an effective model to work with when images are taken into consideration. We also experimented with this pre-trained model for comparing the accuracies. This network is very deep as it consists of 50 layers, and it is trained on both object-centric data as well as scene-centric data (MS COCO). This model provided a testing accuracy of 0.48 on our custom dataset.

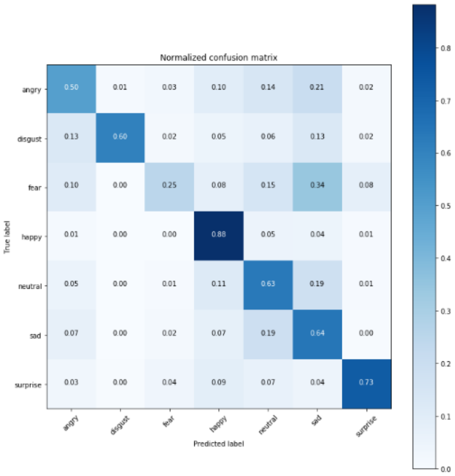


Fig. 5. Confusion Matrix

D. Results

In this subsection, we present the experimental results of our proposed framework of emotion detection and sentiment analysis.

TABLE I
THE ACCURACY OF THE THREE MODELS USED

Model	Accuracies achieved	
	Validation set	Testing set
Proposed Model	0.67	0.80
VGG16	0.40	0.35
ResNet	0.43	0.48

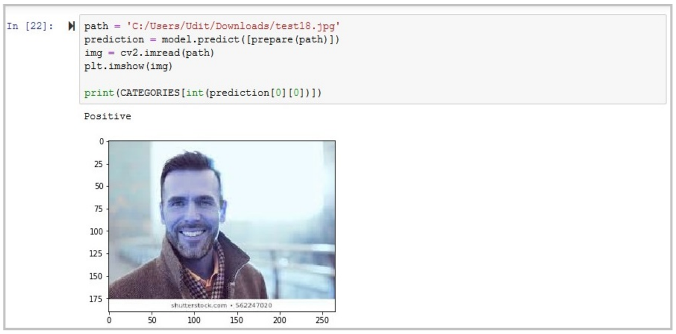
To begin with, Table I shows the accuracies of the proposed model, VGG16 and ResNet. From this experiment we think that our proposed model trained on our custom dataset provides better accuracy than the other two pre-trained model.

Fig.5 shows the confusion matrix obtained when tested on the images from the testing dataset. Fig.6 shows the derived results as a raw output when a particular image from the test dataset is loaded. It shows the sentiment depicted by the particular loaded image.

We created a GUI based on python which is able to upload an image from anywhere as well as capture live image using the webcam. Fig.5 shows the output generated using the GUI which provides the sentiment and the predicted emotion of the uploaded image.

V. Conclusion And Future Works

Prediction of visual sentiments has developed to be one of the most active areas of research in recent years. However, most of the previous works considered CNN features for the classification of images and not predicting emotions to our knowledge. In this paper, a proposed model and two pre-trained models are taken into considerations to compare their accuracies and find out the most promising model for future use. In addition, our results of the proposed model outperforms other model when trained on custom dataset and does provide promising result with better accuracy. This proves very useful form many social media platforms where image tagging is carried out on a regular basis.

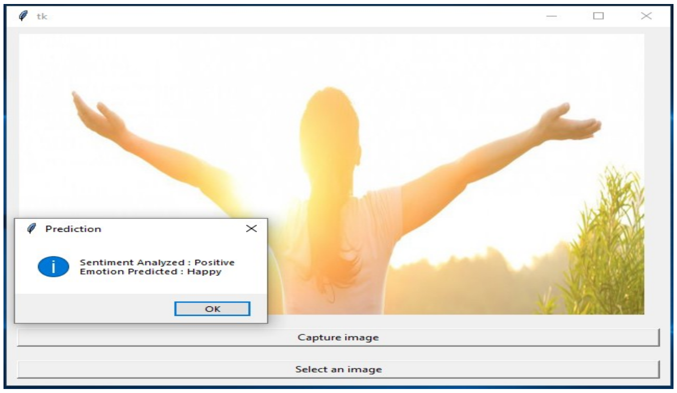


```
In [22]:  path = 'C:/Users/Udit/Downloads/test18.jpg'
          prediction = model.predict([prepare(path)])
          img = cv2.imread(path)
          plt.imshow(img)

          print(CATEGORIES[int(prediction[0][0])])

          Positive
```

Fig. 6. Raw result of the visual data



Sentiment Analyzed : Positive
Emotion Predicted : Happy

Fig. 7. Result using the GUI

There are various future directions that are worth investigating. Firstly, in this paper, we have focused on detecting the sentiment of images and then to predict what emotion the image is depicting that is happiness, surprise, sadness, anger, fear or no emotion. This method can be useful in social media platforms where the use of images is escalating at a rapid pace. This could save time for users on typing or searching for emotion tags. Secondly, it can be taken forward on working with dynamic images (videos) useful for security purpose using CCTV's to detect the actions of a person and predict whether he/she is suspicious or not. Finally, even a single emotion can have several intensity levels such as ”enjoyable”, ”playful”, ”crying”, etc. where we may need a more sophisticated framework and further evaluation.

Acknowledgment

A part of this research was carried out with the support of Atharva College of Engineering and acknowledging the same for providing the infrastructure.

References

[1] B. Liu, “Sentiment analysis: mining opinions, sentiment, and emotions,” The Cambridge University Press, 2015.

[2] D. McDuff, R. El. Kaliouby, J. F. Cohn and P. Picard, “Predicting ad liking and purcha se intent: Large-scale analysis of facial responses to ads,” IEEE Transactions on Affective Computing, 6(3), 2015, pp. 223-235.

[3] Q. You, J. Luo, H. Jin and J. Yang, “Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks,” AAAI. 2015.

[4] V. Campos, A. Salvador, B. Jou, and X. Giro-i Nieto, “Diving Deep into Sentiment: Understanding Fine-tuned CNN for Visual Sentiment Prediction,” arXiv preprint arXiv:1508.05056v2, 2015.

[5] J. Islam and Y. Zhang, “Visual sentiment analysis for social images using transfer learning approach,” Big Data and Cloud Computing (BD-Cloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom), IEEE International Conferences on. IEEE, 2016, pp. 124-130.

[6] V. Campos, B. Jou, and X. Giro-i Nieto, “From pixels to sentiment: Fine-tuning CNN for visual sentiment prediction,” arXiv preprint arXiv:1604.03489, 2016.

[7] J. Wang and L. Perez, “The effectiveness of data augmentation in image classification using deep learning,” Technical report, 2017.

[8] J. Deng, W. Dong, R Socher, L. Li, K. Li and F. Li, “Imagenet: A large-scale hierarchical image database,” Computer Vision and Pattern Recognition, IEEE Conference on. IEEE, 2009, pp. 248-255.

[9] Simonyan, Karen, and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

[10] L. Zhang, S. Wang and B. Liu, “Deep Learning for Sentiment Analysis: A Survey,” arXiv preprint arXiv:1801.07883, 2018.

[11] X. Wang, W Jiang, and Z. Luo, “Combination of convolutional and recurrent neural network for sentiment analysis of short texts,” Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016, pp. 2428-2437, in press.

[12] Mishra, Abhijit, K. Dey, and P. Bhattacharyya, “Learning cognitive features from gaze data for sentiment and sarcasm classification using convolutional neural network,” Proceedings of the 55th Annual Meeting of the Association.

[13] D. Borth, R. Ji, T. Chen, T. Breuel and S. Chang, “Large-scale visual sentiment ontology and detectors using adjective noun pairs,” Proceedings of the 21st ACM international conference on Multimedia. ACM, 2013, pp. 223-232.

[14] Dellagiacoma, Michela, et al. ”Emotion based classification of natural images.” Proceedings of the 2011 international workshop on DETecting and Exploiting Cultural diversiTy on the social web. ACM, 2011.

[15] T. Chen, D. Borth, T. Darrell and S. Chang, “DeepSentiBank: Visual Sentiment Concept Classification with Deep Convolutional Neural Networks,” arXiv:1410.8586v1,2014.