Facial Emotion Recognition Using CNN And Computer Vision

Srinivas Brahmoji

Assistant Professor & Software Trainer

ABSTRACT

It's an intriguing and difficult problem to study how people react emotionally to visual stimuli like pictures and movies, which is called visual sentiment analysis. It makes an effort to comprehend visual data's high-level content. The success of modern models can be linked to the creation of reliable computer vision algorithms. Most of the current models suggest either robust features or more sophisticated models as ways to address the issue. The main suggested inputs are primarily visual elements from the entire image or video. Local areas, which we believe are quite relevant to people's emotional reaction to the entire image, have not received much attention. Image recognition software is used to identify persons in pictures and assess their feelings. The CNN algorithm is used in this project to carry out that task. When given an image, the programme would look for faces, recognise them, place rectangles in the appropriate places, characterise the emotions it discovered, and display the percentage of emotions.

Keywords: Face emotion recognition, computer vision, convolutional neural networks, deep learning, hass arcade, fer-2013.

1. INTRODUCTION

The study of human emotional reactions to visual stimuli like photos and films is known as visual sentiment analysis. It differs from textual sentiment analysis, which focuses on how people emotionally react to textual semantics (Pang and Lee 2008). Visual sentiment analysis recently outperformed textual sentiment analysis in terms of performance.

This is due to deep learning's success in vision tasks, which makes it feasible to grasp high-level visual semantics like image aesthetic analysis and visual sentiment analysis. The design of visual features, from pixel-level through middle attribute level to recent deep visual features, has been the subject of work on visual sentiment analysis. As a result, increasingly rich visual features have led to a progressive improvement in the performance of visual sentiment analysis systems. However, practically all of these methods attempted to elucidate the overall feeling of the entire set of photographs from a global standpoint. The study of how local regions approach the task of visual sentiment analysis and from which local regions we have obtained the sentimental response has received little attention. We are attempting to resolve these two difficult challenges in this work. To discover the link between local picture regions and the sentimental visual qualities, we use the recently published attention model. We can identify the local image regions in this method, which is important for sentiment analysis. A emotion classifier is then constructed on top of the visual features that were taken from these regional areas.

2. LITERATURE SURVEY

Many academics are considering using face emotion recognition to enhance the learning environment (FER). A method that can analyse students' facial expressions and assess the effectiveness of classroom instruction was proposed by Tang et al. [3]. Data gathering, face detection, face recognition, facial expression recognition, and post-processing are the system's five steps. In this method, classification is done using K-nearest neighbours (KNN), while pattern analysis is done using uniform local gabor binary pattern histogram sequence (ULGBPHS). A online programme that analyses the emotions of students taking part in active face-to-face classroom instruction was proposed by Savva et al. [4]. The programme collects live recordings using webcams placed in classrooms, after which machine learning algorithms are applied to the data. Whitehill et al. suggested a strategy in [5] that uses students' facial expressions to determine their level of interest. The method makes use of the SVM algorithm and Gabor characteristics to track student interaction with cognitive skills training software. The authors collected labels from videos that had been annotated by people. When students were interacting with an educational game designed to convey essential concepts of classical mechanics in a school computer lab, the authors in [6] employed computer vision and machine learning approaches to identify the affect of the pupils there. In order to enhance the e-learning environment for better material delivery, the authors of [7] presented a system that recognises and tracks students' emotions while providing feedback in real-time. The technology employs head and eye movements to infer pertinent information about students' moods in an online learning environment. Facial Emotion Recognition System (FERS) was created by Ayyaz et al. [8] to identify students' emotional states and motivations during videoconference-style online learning. The system employs four machine learning algorithms with the highest accuracy rates: SVM, KNN, Random Forest, and Classification & Regression Trees.

Table 1. Literature Survey

S.No	Name of the Author	Year	Technique Used	Advantages	Disadvantages
1	Whitehill	2018	Gabor features and SVM algorithm	The authors collected labels from videos that had been annotated by people.	In a school computer lab, using computer vision and other approaches, students' impact is determined.
2	Ayvaz et al.	2018	4 machine learning algorithms (SVM, KNN, Random Forest and Classification & Regression Trees	developed a Facial Emotion Recognition System (FERS), which detects students' motivation and emotional states during videoconference-style online learning.	KNN and SVM algorithms are not accurate.
3	Kim et al.	2018	RFC algorithm	system which is able of producing real-time recommendation to the teacher.	RFC and SVM algorithms are not accurate.
4	Chiou	2017	wireless sensor network technology	used wireless sensor network technology to create an intelligent classroom management system that aids teachers to modify instruction.	that aids teachers to modify instruction modes rapidly to avert wasting of time.

Table 1. Shows the literature survey done about "Emotion detection using CNN" over the years by various researcher scholars in various techniques.

3. OBJECTIVE

Due to its significance for clinical practice, social robotics, and education, facial expression recognition has received a lot of attention lately. A variety of studies have found that emotion is crucial to teaching. Exams, surveys, and observations are currently used by teachers as sources of feedback; nevertheless, these traditional approaches frequently have low efficiency. The teacher can modify their method and teaching materials based on the students' expressions in order to support learning among the class.

4. EXISTING SYSTEM

There have been a number of recent publications on visual sentiment analysis that have utilised deep visual features and unsupervised frameworks. Initially, pixel-level features were used, followed by mid-level attributes. The majority of the methods now in use focus on either textual or picture analysis.

5. PROPOSED SYSTEM

We concentrate on visual sentiment analysis, as opposed to the extensively researched textual sentiment analysis. Our research focuses first on the local visual regions that emotion-related visual characteristics cause. We base our model on a recently developed attention model that can be taught to interpret students' facial expressions with the use of a Convolutional Neural Network (CNN) architecture. The algorithm first recognises faces in the input image, and then it crops and normalises those faces to a size of 4848. These facial photos are then fed into CNN. The findings of facial expression recognition are the output (anger, happiness, sadness, disgust, surprise or neutral).

6. METHODOLOGY

6.1 SYSTEM ARCHITECTURE

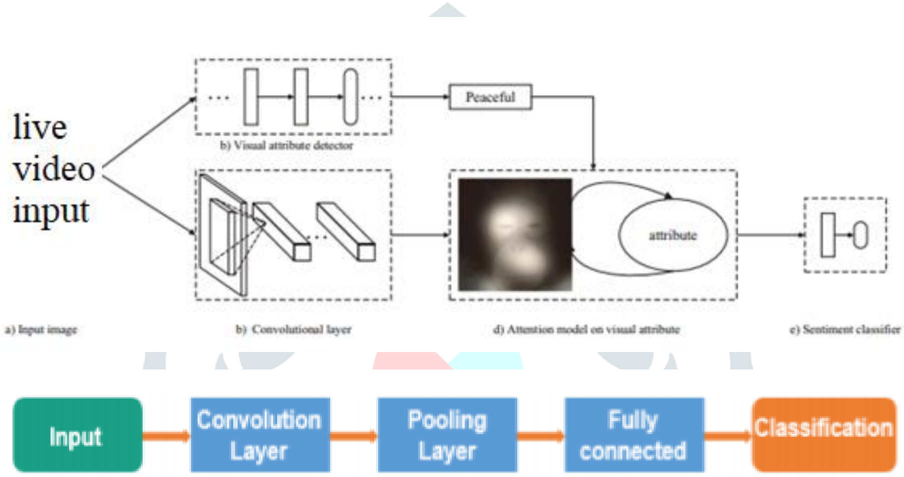

Figure 1: System Architecture

We explain our suggested Convolutional Neural Network (CNN) architecture-based system for analysing students' facial expressions. The algorithm first recognises faces in the input image, and then it crops and normalises those faces to a size of 4848. These facial photos are then fed into CNN. The findings of facial expression recognition are the output (anger, happiness, sadness, disgust, surprise or neutral). Our suggested strategy's structure is shown in Figure 1.

In contrast to conventional image classification techniques, a convolutional neural network (CNN) is a deep artificial neural network that can recognise visual patterns from input images with little pre-processing. This indicates that the filters, which were manually constructed in traditional techniques, are learned by the network. A neuron is a key component of a CNN layer. They are interconnected so that the output of one layer's neurons becomes the input of the following layer's neurons. The backpropagation algorithm is used to calculate the cost function's partial derivatives. Convolution is the process of creating a feature map from an input image by applying a filter or kernel. In actuality, the CNN model has three different kinds of layers, as demonstrated in Figure 1.

6.2 PYTHON MODULES

We can save our code in files with Python (also called modules). For more serious programming, where we don't want to start again on a lengthy function definition only to fix one mistake, this is quite helpful. By doing this, we are essentially creating our own set of modules similar to those found in the Python library. Python includes a feature that allows definitions to be stored in a file and used in a script or an interactive instance of the interpreter to enable this. Definitions from such a file are known as modules, and they can be imported into either the main module or additional modules.

NumPy: Python has a package called NumPy. "Numeric Python" or "Numerical Python" are abbreviations for the name. Additionally, NumPy adds robust data structures to the Python programming language by implementing multi-dimensional arrays and matrices.

Pandas: Python's Pandas is an open-source library. It offers high-performance data structures and tools for data analysis that are ready for use. The widely used Pandas module for data science and analytics operates on top of NumPy.

Open cv: The OpenCV-Python library of Python bindings was created to address issues with computer vision. Numpy, a highly efficient library with a MATLAB-like syntax for numerical calculations, is used by OpenCV-Python. Each and every OpenCV array structure is translated into and out of a NumPy array.

Keras: A simple Python deep learning package called Keras can be used with either Theano or TensorFlow. It was created to make the study and construction of deep learning models as quick and simple as feasible.
Imtulis: With OpenCV and both Python 2.7 and Python 3, the convenience functions known as Imutils make it simpler to perform common image processing tasks including translation, rotation, scaling, skeletonization, and Matplotlib picture presentation.

Pickle: The object data is stored in a file using the dump() function. pickle. The function dump() requires 3 arguments. The object you want to store is the first argument. The file object you receive after opening the requested file in write-binary (wb) mode is the second argument.

6.3 DATA ACQUISITION

We classify facial expressions into seven categories based on basic emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral. CNN is also used to identify distinct emotions or facial expressions. Based on the FER2013 dataset, estimate the amount of facial emotion in Fig. 1.

Figure 2: Variety Facial Emotion Recognition

6.4 DATA STORAGE

The research's dataset was Facial Emotion Recognition 2013 (FER2013). The testing set has 7179 cases, whereas the training set has 28,709 examples. Images of the face in grayscale measuring 48 × 48 pixels make up the data. The faces have been automatically registered such that they are about evenly spaced throughout each image and are somewhat off-center. On the other hand, there are 70 cases in the test data set.

6.5 CNN FOR FACIAL EXPRESSION DETECTION

The structure of CNN is shown in Fig 2. CNN has 12 convolution layers, 6 convolution layers, 2 subsampling layers, and 6 convolution layers.

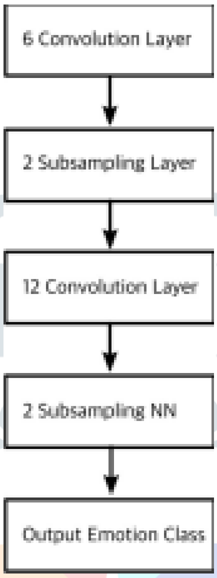

Figure 3: The Diagram of CNN for Facial Expression Detection

6.6 HAAR CASCADE CLASSIFIER

Paula Viola and Michael Jones developed an approach to object detection called the Haar feature-based cascade classifier. They suggested the title "Rapid Object Detection using a Boosting Cascade of Simple Features" for a paper in 2001. A classifier is created by combining a group of Haar Like Features, or "Haar cascade." The feature is the difference between the pixel values in the write and blank areas. The face detector's base measures 24 x 24. There are around 160k potential Haar-Like Features from that base face detector. Not all of these features are utilised, though.

7. EXPERIMENT RESULTS AND DISCUSSION

INPUT SCREENSHOTS:

0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121 119 115 110 98 91 84 84 90 99 110 126 143 153 158 171...	Training
0	151 150 147 155 148 133 111 140 170 174 182 154 153 164 173 178 185 185 189 187 186 193 194 185 183 ...	Training
2	231 212 156 164 174 138 161 173 182 200 106 38 39 74 138 161 164 179 190 201 210 216 220 224 222 218...	Training
4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 19 43 52 13 26 40 59 65 12 20 63 99 98 98 111 75 62 41 ...	Training
6	4 0 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50	Training

0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121 119 115 110 98 91 84 84 90 99 110 126 143 153 158 171...	Training
0	151 150 147 155 148 133 111 140 170 174 182 154 153 164 173 178 185 185 189 187 186 193 194 185 183 ...	Training
2	231 212 156 164 174 138 161 173 182 200 106 38 39 74 138 161 164 179 190 201 210 216 220 224 222 218...	Training
4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 19 43 52 13 26 40 59 65 12 20 63 99 98 98 111 75 62 41 ...	Training
6	4 0 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50	Training

Figure 4: Training Data Set

Fig 4. consists of a screenshot of training data set which is used to train the model in the dataset.

0	254 254 254 254 254 249 255 160 2 58 53 70 77 76 75 78 68 18 32 29 0 54 73 75 72 68 75 77 76 76 75 8...	PublicTest
1	156 184 198 202 204 207 210 212 213 214 215 214 214 213 216 217 218 217 216 214 213 214 213 214 215 ...	PublicTest
4	69 118 61 60 96 121 103 87 103 88 70 90 115 122 123 124 129 132 133 131 131 121 113 110 101 100 99 1...	PublicTest
6	205 203 236 157 83 158 120 116 94 86 155 180 205 231 219 217 190 198 208 174 159 167 211 230 215 209...	PublicTest
3	87 79 74 66 74 96 77 80 80 84 83 89 102 91 84	PublicTest

Figure 5: Testing Data Set

Fig 5. shows the screenshot of the testing data set that we are going to use for the prediction of the emotion.

OUTPUT SCREENSHOTS:

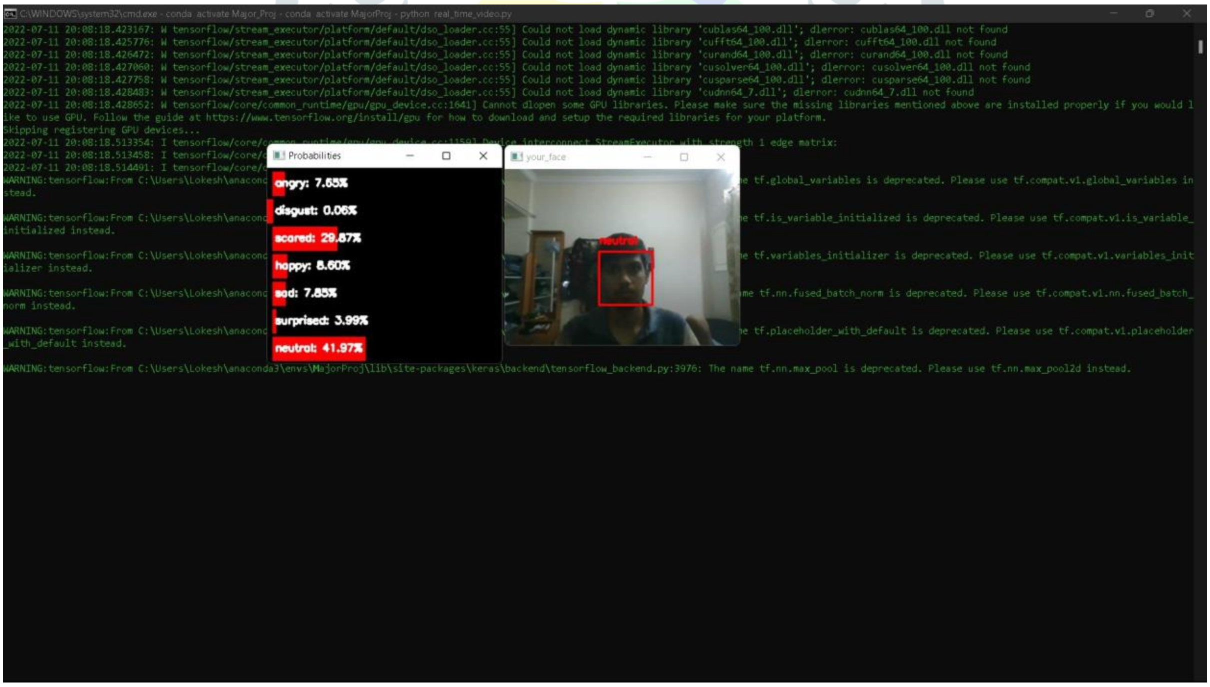

Figure 6: Prediction page with emotion detection

Fig 6. shows the output of the executed program. The result is the detection of emotion with neutral face and emotions percentage window.

8. CONCLUSION

For the purpose of recognising facial expressions in students, we presented a convolutional neural network model. Four convolutional layers, four maximum pooling layers, and two fully linked layers make up the suggested model. Using a Haar-like detector, the system detects faces in the input photographs from the students and categorises them into one of seven facial expressions: surprise, fear, disgust, sad, pleased, furious, or neutral. The FER 2013 database showed a 70% accuracy rate for the proposed model. Our facial expression recognition system can assist the teacher in determining whether or not the class has understood his presentation. In order to extract the emotions of students' faces in 3D using a convolutional neural network model, this will be the main goal of our future study.

9. FUTURE SCOPE

The future enhancements that can be projected for the project are:

- More interactive user interface.
- Can be done as Mobile Application.
- More Details with results applicable to real time applications like classrooms.
- Audio can be used as secondary input for developing more accurate model.
- A dataset with more emotions can be used for building better system with more reach.

REFERENCES

1.R. G. Harper, A. N. Wiens, and J. D. Matarazzo, Nonverbal communication: the state of the art. New York: Wiley, 1978.

2.P. Ekman and W. V. Friesen, “Constants across cultures in the face and emotion,” Journal of Personality and Social Psychology, vol. 17, no 2, p. 124-129, 1971.

3.C. Tang, P. Xu, Z. Luo, G. Zhao, and T. Zou, “Automatic Facial Expression Analysis of Students in Teaching Environments,” in Biometric Recognition, vol. 9428, J. Yang, J. Yang, Z. Sun, S. Shan, W. Zheng, et J. Feng, Éd. Cham: Springer International Publishing, 2015, p. 439-447.

4.A. Savva, V. Stylianou, K. Kyriacou, and F. Domenach, “Recognizing student facial expressions: A web application,” in 2018 IEEE Global Engineering Education Conference (EDUCON), Tenerife, 2018, p. 1459-1462.

5.J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, “The Faces of Engagement: Automatic Recognition of Student Engagementfrom Facial Expressions,” IEEE Transactions on Affective Computing, vol. 5, no 1, p. 86-98, janv. 2014.

6.N. Bosch, S. D'Mello, R. Baker, J. Ocumpaugh, V. Shute, M. Ventura, L. Wang and W. Zhao, “Automatic Detection of Learning-Centered Affective States in the Wild,” in Proceedings of the 20th International Conference on Intelligent User Interfaces - IUI ’15, Atlanta, Georgia, USA, 2015, p. 379-388.

7.Krithika L.B and Lakshmi Priya GG, “Student Emotion Recognition System (SERS) for e-learning Improvement Based on Learner Concentration Metric,” Procedia Computer Science, vol. 85, p. 767-776, 2016.

8.U. Ayvaz, H. Gürüler, and M. O. Devrim, “USE OF FACIAL EMOTION RECOGNITION IN E-LEARNING SYSTEMS,” Information Technologies and Learning Tools, vol. 60, no 4, p. 95, sept. 2017.

9.Y. Kim, T. Soyata, and R. F. Behnagh, “Towards Emotionally Aware AI Smart Classroom: Current Issues and Directions for Engineering and Education,” IEEE Access, vol. 6, p. 5308-5331, 2018.