International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-4, April 2019

Emotion Analysis by Deep Learning Methods using Convolutional Neural Network

Pelash Choudhary, Shravan Vijay, Sathya R

Abstract- In machine learning, CNN uses a variation of multilayer perceptron designed to use relatively little pre-processing compared to other image classification algorithms. This means that the network learns the filtering that was hand-engineered in other algorithms. This independence of human endeavors for feature design is a major advantage due to which it is used in this paper. In the context of machine vision, image recognition is the capability of software to identify objects in images. The algorithm is used to train the model from a data set of around 10000 images and 12 videos. The model will detect and recognize types of feelings through the person's expression, such as anger, fear, happiness, sadness, and surprise. The model gives an accuracy of 67%. This provides a behavioral measure for the study of emotion, cognitive process and social interaction.

Keywords- Emotion analysis, convolutional neural network, facial recognition, Reinforced learning.

I. INTRODUCTION

In machine learning, Convolutional Neural Network (CNN) use a variation of multilayer perceptron design to use relatively little pre-processing compared to other image classification algorithms.This means that the network learns the filtering that has hand-engineered in other algorithms. The self-sustaining human efforts for feature design is a major advantage due to which we are using CNN. In the context of machine vision, image recognition is the capability of software to identify objects in images. Facial recognition draws from many disciplines including science and computational linguistics. In the pursue to fill the gap between security and computational understanding. Facial expression analysis essentially a research topic for psychologists. The human's way of understanding emotions is different from the machine. It took a lot of effort to recognize facial expressions in images. The image usually captures the apex of the expression and the machine learns from every single facial cue in that image. Which in the case of humans is not possible. The facial recognition is the past has been done manually. We automate the process of detecting the facial cues and forming a cluster for it. Image processing has been a concept that is prevailed for years but most of it use the supervised learning methods. In our, model we use reinforcement learning. The reinforced learning has the principle of reward system. The AI rewards itself for every goal it reaches and whenever the machine makes the mistake it takes the feedback to learn and create a knowledge base from it. The downside is that reinforced learning and convolutional neural network together consumes more hardware but also significantly bumps the efficiency and performance of the operation. In this model, we have a data set of 10000 images and 12 videos. This both the static and dynamic data sets are processed. Images and videos can never be classified in a supervised algorithm. When compared to its reinforced algorithm has shown improvement in efficiency and performance in the process. Emotion analysis could help in realizing health issues like depression and anxiety. The model's application can also be extended to crime incidents to reduce the effort in understanding the potential suspects. Automated recognition of emotion from the images provides a behavioral measure for the study of emotion, cognitive process and social interaction.

II. LITERATURE SURVEY

As discussed by Gerard Pons[1], the human faces give the emotional perception. The accurate expression of someone's emotion is in their face. The automated facial recognition algorithm goes for the emotions defined by Ekman and Friesen [2], which had six common emotions, namely anger, sadness, happiness, fear, surprise and neutral. He also defined the Facial Action Coding System(FACS), the methodology thatis used in the present Computer Vision techniques. The supervised learning requires input and output variables. It learns from these variables in order to map the function. So we use reinforced learning to create a reward system for the machine. This way it learns from every random input given to it.Prudhvi et al[3]. (2017) mentioned the creation of 5-layer CNN which was trained to classify the common human emotions. To reduce the redundancy of the learned features the layers are reduced to 4 and the last layer is the Long Short Term Memory(LSTM) which gives feedback on a sequence of data.The research challenges in facial analysis such as Emotional recognition in the wild(EMotiW) and Kaggle's Facial Expression Recognition Challenge which has neutral emotion classified as a part of it.

Manuscript published on 30 April 2019.
* Correspondence Author (s)
Pelash Choudhary*, B.Tech CSE, SRM Institute of Science and Technology, Chennai, India.
Shravan Vijay,B.Tech CSE, SRM Institute of Science and Technology, Chennai, India.
Sathay R, M.tech(Ph.D),Assistant Professor (O.G), CSE, SRM Institute of Science and Technology, Chennai, India.

The CNN classifies the facial expression as an object. Arushi et al[7]. achieved to classify the human faces into discrete emotion categories. It gave a .60 accuracy on the EmotiW data sets.The study of the facial recognition using geometric information of a facial image or using the template vector and neural network was mentioned by Young Hoon et al[4]. Instead of the open computer vision, the optical flow model can be used for recognizing the emotion. The image usually captures the apex of the expression. So Young Hoon et al[4]. proposed an algorithm to detect emotion via the frontal facial image. It had three stages: image processing, feature extraction, and emotion detection. The first stage of the algorithm was processed with help of algorithm developed the past studies. They also proposed the feature extraction method which included the axillary region of the face. It was extracted by matching the geometric and shape information. The patterns of emotion are indistinct. So a fuzzy classifier was used to present the emotion.The internal emotion plays an important role in the facial changes of a human being. Alberto et al[5]. presented a method that has an accuracy larger than 95% even in the controlled environments. A video-based facial recognition was proposed by Byeon et al[6]. The CNN methodology used was using 3D input from 5 continuous frames. Since the input was in three dimensions, it requires the height and width of the images which is multiplied with the
number of frames. It was professed that the deformation of the image was also taken care of by this methodology. The only disadvantage with this method was it depended on sequence containing the video for the expressions. Depending on the histograms of gradients the spatial-temporal framework can process video-based images. The aspects of this method were preprocessing, extraction and classification. The preprocessing of the image reduces the variations in head pose. The extraction phase gives information on the variation of the facial shape. The classifier used is SVM with RBF kernel. The CK+ database gave an accuracy of 83.7% and with MMI database the accuracy was 74.3%. 350ms and 520ms per image was the recognition time for the database. The method operates in real time to recognize the expression. The major factor was the pre-processing operations of the image. The works related to emotion analysis has a common intention of taking advantage of the potential applications of the models in the fields such as customer-focused marketing, health tracking, interactive gaming and it can be applied to robots that have to be sociable by making it emotionally intelligent.

III. PROPOSED SYSTEM

This paper is concerned with upgrading the existing system to better machine learning methods. The supervised learning method used in most of the facial recognition algorithms requires lots of computation time. It might also get a wrong class label during the classification.Hence we substitute supervised learning with reinforcement learning. In simple words, it can be defined as trial and error. It attempts to learn in order to get rewarded. At every training action, it tries to learn and give us the desired outcome. It is rewarded accordingly. Reinforcement learning has to goal to maximize its reward. So the training dataset given to it, it learns from it and performs the action to give us the output as an emotion. The reward functions are SR and NR which are positive and negative rewards respectively. Whenever the machine gets a negative reward, it learns the dataset again with a different approach. The next time when a similar input image is given the accuracy of output is improved. If the machine gets a positive reward and there is no space
for maximizing the reward then it has reached its level of accuracy.

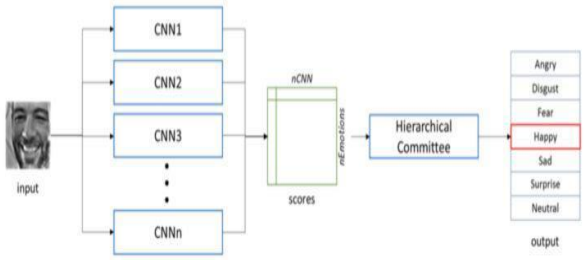

Fig. 1 Example of the CNN architecture

It is not always that the data input we give be static like images. Sometimes it can also be videos. So the algorithm works in a way that it splits the video into single frames which in turn becomes an image itself. The Reinforced learning combined with CNN to perform this operation significantly bumps the efficiency and performance of the process. The convolutional layers use the open computer vision model to identify the object in the image. It is similar to cameras focusing on the face. The CNN sliding window scans the facial features of the image and matches it with the trained dataset using the confusion matrix. It gives a cosine relevance between the input image and the trained dataset. Depending on the cosine relevance in the confusion matrix, it gives an emotion output.

IV. PROPOSED METHOD

The operation of the proposed design begins with the training of the machine. If the trained datasets are not available then it will collect from the fer database. This consists of three phases. These are; Collection, Training, and Evaluation. Here the fer2013 database is used. The database has pictures of human faces. The images in the dataset are pre-processed by understanding the facial cues and features. The foremost goal is to place each image of the dataset. The CNN process the image via the pixel object density. At every layer, CNN tries to understand the images in the dataset. Once the datasets are trained then an input image is fed to the machine. This image is pre-processed using CNN and every facial
expression is handled to help the machine understand the emotion.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

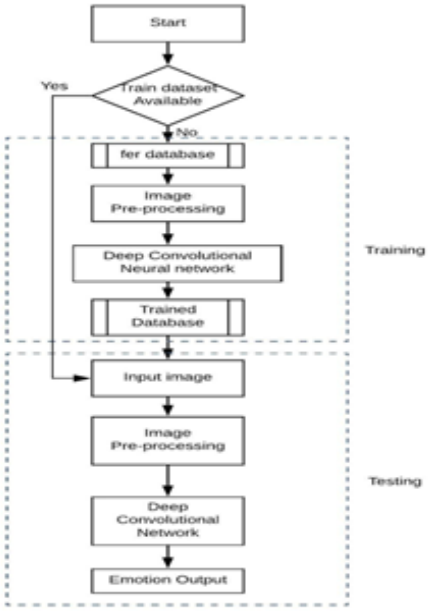

Fig. 2 Block diagram of the facial recognition process

A. Collection

The first and foremost step in facial recognition is having a dataset. The dataset is used to let the machine learn the operation. The machine should acquire enough data in order to process and learn from them. It is expected the dataset should consist only of human faces. Since it would classify each data into one of the categories of emotion. The Collection phase is where the data obtained is labeled and classified. So that machine can understand the data and learn from it. The data is annotated into a collection and append it with the attributes. The data collection can be custom which is done by manually

labeling them. It could be a tedious task. The alternate options for data collection are scraping from the web or use a third party to get the dataset. The methodology to collect dataset can be different but the goal is to train the machine with the dataset. The model's performance depends on the dataset. More the dataset, more learning the machine does and performs better.

Data collection is a part of deep learning. Some of the data in the dataset might not be human faces. It will be images that are similar to human faces. The machine needs to understand the facial patterns and differentiate the actual human faces with images that look like human faces.The Data acquisition of human faces has certain parameters. The image in the dataset will be able to detect the face and also estimate the features like head pose. It then extracts the facial data like appearance, eyes, nose, mouth and, chin. This process is done for every image that is in the dataset. The information extracted from this is then used by the machine to understand and learn human facial expressions.

B. Training

The model starts with checking whether a trained dataset in available or not. If a trained dataset is not available then the machine trains itself from the dataset. The dataset used here is the fer2013 dataset. The dataset is pre-processed to classify the object as a face and also label them according to the facial feature. It is then validated for the training. The requirement here is to use CNN and achieve better performance. For this, it demands more hardware power due to high computational challenges.

Fig. 3 Classification of human facial expressions

In the deep Convolutional Neural Network, we apply the filters and weights with random values. It takes

the input image as an array of pixels and measures its height, width and, dimension. Then the image passes through the layers of the convolutional network which include filter, pooling and also ReLu activation matrix layer. The neural network might have a different number of layers based on the requirements. But the first layer in the network would always be the convolutional layer. This layer extracts the features of the image. In our case, it extracts facial features. The extraction of features is done by taking small squares of the image as inputs. These inputs are combined with the filters to get the required operation from the convolutional layer. There are different filters to get a different kind of convolved image. After the convolutional layer, the image is processed in the pooling layer. This is where the dimensionality size of the image is reduced. Sometimes the images in the dataset are too large. During pooling, it tries to retain the important information of the image.

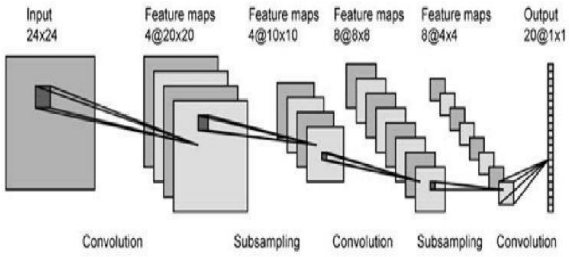

Fig. 4 Layers in the Convolutional Neural Network

Once all the features of the image are extracted and mapped into labels. The matrix is flattened to vectors. This vector is given as an input into the Fully Connected Layer. A model is created when the mapped features of the image are combined. This model gives us an activation function to classify the image as one of the common human emotions.

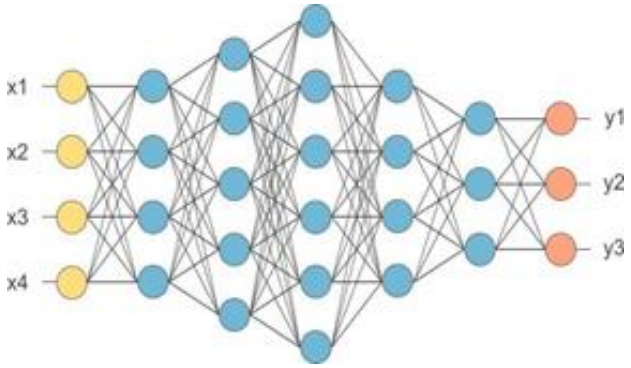

Fig. 5 Fully Connected layer in CNN

These phases of the image processing in the neural network are executed every image in the dataset. More the images in the dataset, more trained dataset under one classification and so more learning the machine does improve the accuracy.

C. Evaluation

Once the model has a trained dataset, then the input image will directly be processed. The facial features of the input image are processed by the convolutional neural network. A confusion matrix for the model is created. This will evaluate the most confused human expression for our trained model.

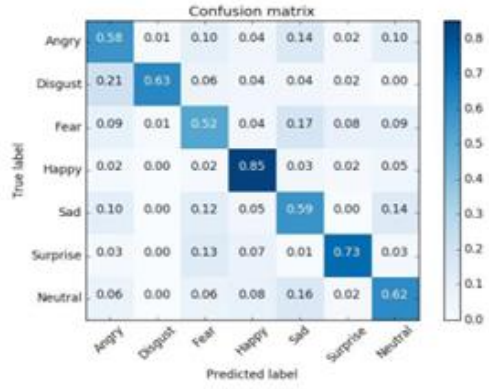

Fig. 6 Confusion matrix

The cosine relevance of the input image with the trained dataset is calculated. The value must be greater than 0.52. Comparing the cosine relevance value with the confusion matrix the input image is classified under the defined human emotions. The machine might make errors during the extraction and mapping of human faces. During the error, it gets a negative reward for the mistake. Then by reinforcement learning process the machine would train itself and learn from the dataset again. So next time when a similar image is given to the model, the accuracy of the analyzing the emotion is improvised. Resulting in a positive reward function.

$$\vec{a} \cdot \vec{b} = \|\vec{a}\| \|\vec{b}\| \cos\theta$$

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|}$$

The image processing time is higher when the image has a higher resolution. To reduce the dimensionality
size of the image, the pooling layer in the neural network requires more time since the pixels in the image is more, it tries to retain the important information in the image. It should not affect the facial features that are necessary for mapping them into one of the defined classes of emotion.

V. EXPERIMENTAL RESULTS

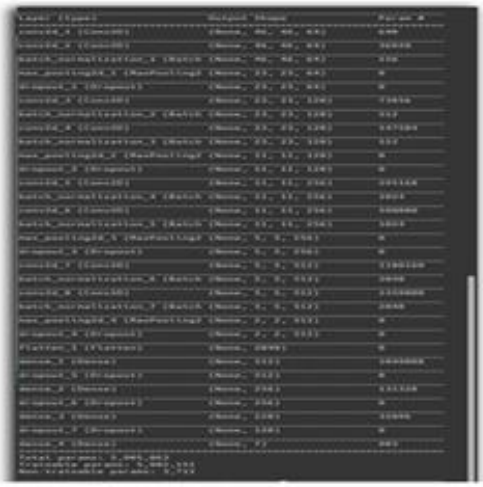

Fig. 7 Classifier of the Convolutional Neural Network

The Layers of the Convolutional Neural Network used in implementing this classifier depends on the trained dataset in the model. It takes around one hour to train the dataset using Intel Core i7-7700K 4.20GHz processor and an Nvidia GeForce GTX 1060 6GB GPU, with tensorflow running on GPU support. If the hardware power is increased the time duration for training can be reduced. The lower the hardware the challenges the model needs to face to process the image and classify it under one of the emotions. This model had an accuracy of 67% on FER2013 dataset of the kagel challenge. Whereas the winner of the challenge had 71% accuracy. So with the reinforced learning approach, we can improve the performance to a reasonable margin.

VI. CONCLUSION

A facial expression can define a person's state of mind. In this paper, we have used a different approach to analyze the emotion of human beings with an image than conventional methods. The model has combined the open-cv and CNN with reinforced learning. We make the process more efficient in terms of performance and accuracy. The major trade for efficiency with this model is that it consumes more hardware. The fer dataset used to train the machine and learn human facial cues. The probability of the machine making errors is high if the dataset is less. But the machine relearns whenever it receives a negative reward for error occurred.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

It tries to understand the image by training itself again. So next time whenever it comes across a similar image, it processes it better than the previous iteration. With the help of this model, we try to understand human psychology. The potential applications built using this is superior to what one can expect.

REFERENCES

1. Gerard Pons and David Masip, Supervised Committee of Convolutional Neural Networks in Automated Facial Expression Analysis, IEEE Trans Affective Computing, Nov. 2017.
2. P. Ekman, W. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement., Consulting Psychologists Press, Palo Alto, 1978.
3. Prudhvi Raj Dachapally, Facial Emotion Detection Using Convolutional Neural Networks and Representational Autoencoder Units, School of Informatics and computing, Indiana University, 2017.
4. Young Hoon Joo, Moon Hwan Kim, Jin Bae Park, Emotion Detection Algorithm Using Frontal Face Image, ICCAS, 2005.
5. Alberto F. De Souza, Thiago Oliveria-Santos, Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and the Training Sample Order, Research Gate, July. 2016.
6. Y.-H. Byeon, K.-C. Kwak*, Facial expression recognition using 3d convolutional neural network, Vol. 5, 2014.
7. Arushi Raghuvanshi, Vivek Choksi, Facial Expression Recognition with Convolutional Neural Networks, Stanford University, 2016.
8. Abir Fathallah, Lotfi Abdi, and Ali Douik, Facial Expression Recognition via Deep Learning, IEEE/ACS, 2017.
9. D. C. Ali Mollahosseini and M. H. Mahoor. Going deeper in facial expression recognition using deep neural networks. IEEE Winter Conference on Applications of Computer Vision, 2016.
10. C. W. Pablo Barros and S. Wermter. Emotional expression recognition with a cross-channel convolutional neural network for human-robot interaction. IEEE 15th International Conference on Humanoid Robots, 2015.
11. P. Viola and M. J. Jones, “Robust real-time face detection,” International journal of computer vision, vol. 57, no. 2, 2004.
12. J.J. Lien, T. Kanade, J.F. Cohn, and C.C. Li, “Automated facial expression recognition based on FACS action units,” Third IEEE International Conference on Automatic Face and Gesture Recognition, pp.390-395, 1998.
13. P. Burkert, F. Trier, M. Z. Afzal, A. Dengel, and M. Liwicki, “Dexpression: Deep convolutional neural
14. network for expression recognition,” arXiv:1509.05371, 2015.
15. A. Mollahosseini, D. Chan, and M. H. Mahoor, “Going deeper in facial expression recognition using deep neural networks,” in Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on. IEEE, 2016, pp. 1– 10.
16. B.-K. Kim, J. Roh, S.-Y. Dong, and S.-Y. Lee, “Hierarchical committee of deep convolutional neural networks for robust facial expression recognition,” Journal on Multimodal User Interfaces, pp. 1–17, 2016.

AUTHORS PROFILE

PelashChoudhary- Undergraduate Student, Computer Science and, Engineering at SRM Institute of Science and Technology(TN)

Shravan Vijay -Undergraduate Student, Computer Science and, Engineering at SRM, Institute of ,Science and Technology(TN)

SathyaR- M.tech(Ph.D),Assistant Professor (O.G), Computer Science and Engineering at SRM, Institute of Science and, Technology(TN)