Detection Of Human Emotion using OpenCV.

Sourav Gope[a] *, Sahil Barua[a], Sankha Subhra Debnath[a]

[a] Department of Computer Science and Engineering,

Techno College of Engineering Agartala, Maheshkhola, Agartala, Tripura (West, Madhuban, Tripura 799004)

1. Abstract:

The utility of facial recognition has been intensively researched in recent years.Applications for facial recognition range from emotion detection to use in security cameras.

Face recognition in general, and emotion detection in particular, offers enormous potential in a variety of industries.There are several procedures that need to be followed in order to customise the programme for emotion detection.

It is necessary to construct two databases, one with positive images (pictures that depict a certain emotion) and the other with negative images (images without a face). The two databases are then used to train a classifier. Then, a function is used to incorporate the classifier into the software. In the end, this procedure ought to make it possible to successfully detect emotions for a variety of applications.

Keywords: Human Emotion Recognition, Emotion detection, Open CV.

2. Introduction:

Numerous studies are now being done on facial recognition technology and its potential.However, utilising facial recognition to detect emotions is not as well discussed.

I'm putting forth a programme that can identify human emotion, which might be useful for many things, such as questioning criminals and creating interactive software that can help autistic kids recognise emotion. The sole goal of this application will be emotion recognition, and it will build on existing research. his application's architecture includes a database for each emotion that contains both real and fake emotions, enabling the software to tell if the emotion on a scanned face is real or fake.An eye must be kept out for specific possible issues such a veiled

face, poor lighting, or a non-frontal face during the creation of these databases. We are able to ignore these issues as individuals, but a machine finds it more harder to do so. For the software to succeed overall, facial representation is crucial.

The total application cannot operate effectively or as precisely without a reliable facial recognition technology.

The choice was taken to do the facial recognition component using OpenCV's open source code directly due to the research's time constraints.

By using this route, more time may be dedicated to research and the project's emotional component.

This essay has been divided into sections depending on many key elements.

3. Literature Review:

In the areas of emotion recognition and feature extraction, several studies have previously been examined.

Here are a few of the crucial techniques that are discussed:

A: Linear analysis to discriminate face data

Finding a linear combination of components that divides or separates additional categories of objects or events is done using the LDA technique.The outcome may be used to create a line arrangement.

On a computer screen, an increased number of pixels are employed to depict emotion.In order to reduce attributes and make it more manageable, analysis by line is used. The template's pixel values are linearly combined to create the new dimension.

B: Analysis of Principle component (Principle Component Analysis)

PCA uses an arithmetic procedure to transform a number of fickle variables that might be related to a few unrelated variables.

The initial main components compute data conflicts, and succeeding components result in more variability.

A frequently utilised tool is data analysis for testing and creating models that predict PCA.With the use of PCA, the eigenvalue computation of the

data matrix covariance or single value matrix data decomposition was carried out.

4. Problem Definition:

Human emotion recognition is crucial for characterising human emotions in systems like conversational interfaces, moving images, video conferences, and live animation from animated images.

For the artificially based analysis of face images, several system models have been developed.

While Turk and Pentland welcomed Sirovich and Kirby's first proposal to put facial coding in it for fragmentation, Eigen's expression obscures important regional or global "facial reasons."

This approach stands out for the following reasons:

As they recorded several "signatures" of the earth's surface and were consequently tolerant and protected in a range of locations, independent facial pictures are shown statistically via eigen vector.

ue to the fact that face recognition is frequently affected by different variations in the appearance of angles and hair, as well as modest closure and unclear images.

Compared to previous modelling techniques that offer face models via geometric measurements of local visual qualities, such as geographic and eye size, nose and mouth distance, Eigen's facial features are more computer-based and physiologically sound. Another alluring method for distributing coded pictures to networks is using Eigen facial traits.

However, the current Eigen face model has a fundamental flaw. Only one-dimensional photographs taken with a small viewing angle—typically front-facing—are available for use in existing models of Eigen's face.This significantly lessens its performance and longevity.

For instance, if the viewing non-identical is large, three face-to-face photos of the same face (i.e., one person) would likely be recognised as by distinct categories.

5. DETECTION PROCESS:

The system must first be able to determine whether or not there is a face in the picture or video.

An existing cascade classifier from OpenCV was used to do this.

If there is a face in the image, this classifier locates it first. It then draws a circle around the face to indicate to the user that it has indeed found one.

After finding a prospective face, the system has to be able to scan the region and check it against a database to see if a face is indeed present.

Once the identity of the face has been established, we may start examining facial characteristics and using that knowledge to ascertain the mood and facial expression of that face.

I decided to employ a haar feature-based cascade classifier for facial identification in this specific experiment.

This classifier, which is currently used to recognise other faces, was trained using a wide range of both positive and negative pictures.

5.1. Facial Recognition:

The accuracy of a facial recognition system depends on how the faces are represented.

It is advisable to save the representations of the faces in a database.

To establish if the image being scanned by the system is a face, this database is then compared to it.

When attempting to identify facial expressions or emotions, creating a data collection is very crucial.

For instance, Curtis Padgett and Garrison Cottrell from the University of California's Department of Computer Science undertook an experiment to use face recognition software to recognise human emotion.

In their initial data collection, there were images of undergrads expressing various emotions. The faces that were acting out an emotion and the faces who were pretending it differed, though.

Due to these variations, emotion recognition was incredibly incorrect.

A verified data set was produced by trained actors that could faithfully represent the emotions in order to address the problem.

Being consistent with your data sets is obviously crucial.

5.2. Facial Detection:

The most crucial phase of facial recognition is probably determining whether or not a face is present.primarily because it effectively marks the start of the first phase in the procedure (besides preparing the data set).OpenCV, the programme used in this research for facial recognition, employs classifiers to find objects.

The programmer trains these classifiers to recognise whatever it is that he or she wants the recognition software to be able to identify. Since the source code and classifier for OpenCV are already available, it was decided to utilise it for facial recognition since this project primarily focuses on facial expressions and emotions.The following part and the design section go into further detail on training classifiers and databases.

Once the data set is prepared, a few typical issues must be taken into account. thresholding algorithm that takes in input from every pixel from the eye region. This algorithm is represented as $T_{local}(x,y) = \mu_{global}(x,y) + k * \sigma_{local}(x,y)$

$$\mu_{global}(x,y) = (1/M*N)[\sum_{j=0}^{N}\sum_{i=0}^{M} f(i,j)]$$

where $T_{local}(x,y)$ is the threshold value of the pixel at (x,y) and $\sigma_{local}(x,y)$ is the standard deviation. μ_{local} is the local mean and μ_{global} is the global mean and $M*N$ is the size of the eye.

Fig 1: This demonstrates how the classifier informs the user that it has located the face.This illustration shows how to utilise a nested-classifier to identify the eyes.

A face that is only partially blocked is a typical issue. When a person views a photograph of another person whose face is partially obscured, they can fill in the blanks.Not all systems are able to complete the gaps, which leads to subpar or nonexistent object detection.Right present, if the face is even slightly obscured, the algorithm is unable to recognise it.The same holds true if the person's face is too much to the side or too far back in the present photograph.Lighting is also another frequent problem.Once more, individuals do not have as much of a problem with this as do machines.

5.3. Feature Extraction

There are several approaches of extracting facial characteristics.
A multimodal learning technique was utilised by Zhang et al. to learn the representations of texture and landmark modality.
The texture modality is a collection of picture patches that emphasise certain features on the face.
The landmark modality shows the major facial points in a series of expressions.
This multimodal learning method functions similarly to an artificial neural network (ANN), stacking inputs such as numbers and modality in a hidden layer before producing a result.

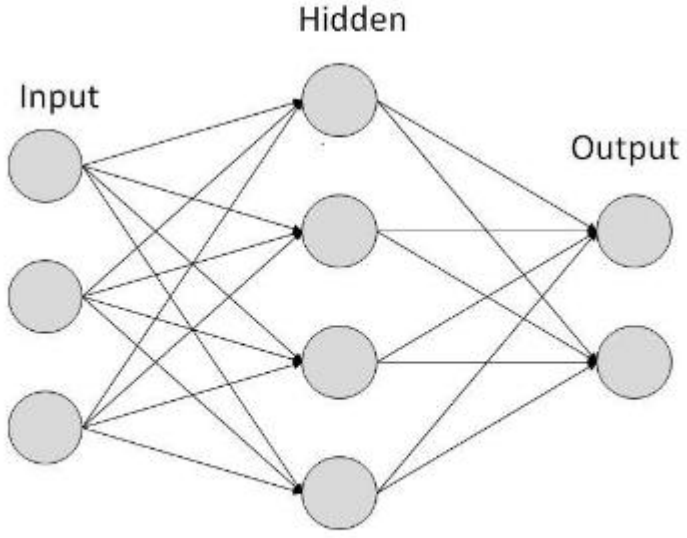

Figure 2: Structure of multimodal algorithm.[5]

5.4. MODEL

Our first model relies on the notion of totally removing the fully linked layers.
The combined depth-wise separable convolutions and residual modules are included in the second design, which includes the elimination of the fully linked layer.
Using the ADAM optimizer, both designs were trained.
Our architecture followed the previous architecture patterns by using Global Average Pooling to totally eliminate any fully linked layers.
In the final convolutional layer, the same amount of feature maps and classes were used to achieve this.
On each reduced feature map, a softmax activation function is applied.
We firstly put forth a typical fully-convolutional neural network with nine convolutional layers, ReLUs , Batch Normalization, and Global Average Pooling as our suggested design.
There are around 600,000 parameters in this model.
It was trained on the IMDB gender dataset, which consists of 460,723 RGB photos classified as either "women" or "men," and on this dataset, it has a 96% accuracy rate.
Additionally, we tested this model with data from FER-2013.
The 35,887 grayscale photos in this dataset are divided into the following classes: "angry," "disgust," "fear," "glad," "sad," "surprise," and "neutral."
In this dataset, our model had a 66% accuracy rate.

Our model is also inspired by the Xception archi-
tecture This design combines the usage of depth-wise separable convolutions with residual modules.
The desired mapping between two further layers is modified by residual modules, making the learnt features the difference between the desired features and the original feature map.
So that the simpler learning issue F (X) may be solved, the desirable characteristics H(x) are altered as follows:
To do this, they first apply a D x D filter to each of the M input channels before combining the M input channels into N output channels using N x 1 x 1 x M convolution filters.
Each value in the feature map is combined using 1 x 1 x M convolutions, but their spatial relationship inside the channel is not taken into account.

We also added to our implementation a real-time guided back-propagation visualization to observe which pixels in the image activate an element of a higher-level feature map. Given a CNN with only ReLUs as activation functions for the intermediate layers, guided-back propagation takes the derivative of every element (x, y) of the input image I with respect to an element (i, j) of the feature map f L in layer L. The reconstructed image R filters all the negative gradients; consequently, the remaining gradients are chosen such that they only increase the value of the chosen element of the feature map. Following a fully ReLU CNN reconstructed image in layer l is given by:

$$R_{i,j}^{l} = (R_{i,j}^{l+1} > 0) * R_{i,j}^{l+1}$$

6. RESULTS

Our complete real-time pipeline including: face detection with emotions An example of our complete pipeline includes the categorization of emotion and face datas.Several frequent misclassifications may be seen, such as anticipating "sad" instead of "fear" and "angry" instead of "disgust."

A comparison of the learned features between several emotions and both of our proposed models can be observed in Figure Ra. The white areas in figure Rb correspond to the pixel values that activate a selected neuron in our last convolution layer. The selected neuron was always selected in accordance to the highest activation.

Fig: Ra: sample of the resultant human emotions

We can observe that the CNN learned
to get activated by considering features such as the frown, the teeth, the eyebrows and the widening of one's eyes, and that each feature remains constant within the same class. These results reassure that the CNN learned to interpret understandable human-like features, that provide generalizable elements.

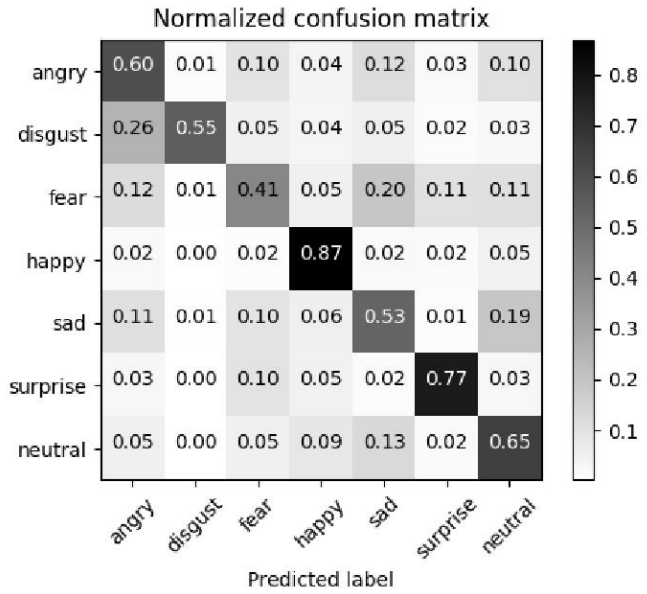

Fig: Rb: Convulation matrix(face data)

7. CONCLUSION

We have put out and evaluated a generic architecture for developing real-time emotion detection using CNN and algorithms like multimodal algo, thresholding algorithm.Our suggested designs have been methodically constructed to minimise the number of parameters.
We began by eliminating completely the fully connected layers and by reducing the amount of parameters in the remaining convolutional layers via depth-wise separable convolutions. We have shown that our proposed models can be stacked for multi-class classifications while maintaining real-time inferences. Specifically, we have developed a vision system that performs face detection, gender classification and emotion classification in a single integrated module. Current results show that the program is accurately able to detect the emotion of the subject distinctively.
Our complete pipeline has been successfully integrated and finally we have presented a visualization of the learned features.This visualization

technique is able to show us the high-level features learned by our models and discuss their interpretability.

REFERENCES

[1] Franc ois Chollet. Xception: Deep learning with depthwise separable convolutions. CoRR, abs/1610.02357, 2016.

[2] Andrew G. Howard et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. CoRR, abs/1704.04861, 2017.

[3] Dario Amodei et al. Deep speech 2: End-to-end speech recognition in english and mandarin. CoRR, abs/1512.02595, 2015.

[4] Ian Goodfellow et al. Challenges in Representation Learning: A report on three machine learning contests, 2013.

[5] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, pages 315–323, 2011.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[7] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning, pages 448–456, 2015.

[8] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[9] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks. International Journal of Computer Vision (IJCV), July 2016.

[10] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[11] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. Striving for simplicity: The all convolutional net. arXiv preprint arXiv:1412.6806, 2014.

[12] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2818–2826, 2016.

[13] Yichuan Tang. Deep learning using linear support vector machines. arXiv preprint arXiv:1306.0239, 2013.