iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.43578

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Deep Learning based Emotion Analysis and Recognition for Movie Review

Kamalesh. T. K[1], Ragul.P[2], Aravindh Siva. S[3], Pradeep Kumar. G[4]
[1, 2, 3]*UG.Students,* [4]*Assistant Professor, Department of ECE, Velammal College of Engineering and Technology, India.*

Abstract: ***In the present world of AI-based systems, image processing is a prominent emerging method. One of the topics covered by artificial intelligence machine learning techniques is emotion recognition. Emotion Recognition is the process of analyzing facial expressions, body posture, gestures, and voice to interpret a person's thinking and current mental state. This allows one to determine whether or not the individual is interested in the continuing action. The Audience Feedback Analysis' fundamental functionality Using Emotion Recognition allows you to collect and analyze every individual's facial expression during or after a seminar, discussion, lecture, or keynote, and provide an analysis of the average emotional state of the audience. The major goal of the proposed system is to offer the event organizer or administrator with accurate and unbiased feedback by analyzing the expressions of the crowd during the duration of the event or at a specific time interval.***
Keywords: **Convolutional Neural Network(CNN), Emotion recognition, Deep-Learning, ReLU.**

I. INTRODUCTION

Because they convey a person's emotional state to observers, facial expressions and emotions are considered the primary means to assess a person's mood and mental state rather than words [1]. Facial expressions are often thought to be universal or global, however many of them have various meanings in different cultures. Facial expressions are widely recognised as a global language for communicating emotional states across cultures [2]. They are regarded as grammatical functions and are included in nonverbal language grammar [3]. Human expressions are important. Human emotions are based on facial expressions, which can be used to assess an individual's feelings. The eyes are an important element of the human face for expressing various emotional states. The pace at which a person's eyes blink is used to determine if they are frightened or dishonest. Similarly, persistent eye contact reveals a person's concentration. Shame, or the admitting of loss, is commonly associated with the eyes and is a fundamental component of neurolinguistic programming (NLP), a method of altering human behavior and life through psychological procedures.

II. EXISTING WORK

Chronicity, sickness, and social skills were all linked to facial perception. People might perceive another feelings and react in a quick manner in specific situations because of emotions. For example, "A man's judgment using psychological study," Facial emotion identification is difficult due to the foggy nature of the task, and extracting effective emotional components is an unresolved question. It is commonly used for security systems, mobile application unlocking systems, and iris scan unlocking systems for high-tech security, such as unique mark or eye iris recognition systems, but only partially because the computers can not understand the feeling states. We can also determine whether or not a man is convinced for the motivational speech. Emotion is a conscious state of mind marked by rapid mental action and a sense of joy or disappointment. Scientific discussions have had varying outcomes, and there is no universal agreement on the definition. They applied sentiment analysis in the existing system based on customer feedback, rating, ranking, review, and emogi data. To begin with, emotion detection is normally inferred at the phrase level, which does not show a user's emotional state over time. Customers may define them for the sake of formality, even if they have other thoughts and ideas about the videos and movies. Only photographs and still images were used by some researchers to study face emotion and sentiment analysis. It was unable to evaluate moving images and produce accurate results.

III. PROPOSED WORK

Propose two models evaluated about their test accuracy and the number of parameters. Both models Developed with the idea of achieving the highest accuracy Over-number parameter ratio. Reduce the number of Parameters help you overcome two important issues. beginning, Use a small CNN to free yourself from poor performance Systems with hardware restrictions such as robot platforms. When Second, reducing the parameters provides a better generalization under the Occam's razor framework. Our first model Relies on the idea of completely eliminating what is completely connected layer.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

The second architecture is Inclusion of fully connected layers and combined ones Folds and residual modules that are separable in the depth direction. both The architecture was trained with the ADAM optimizer.

Fig 1 Samples of the FER-2013 emotion data set

Following the previous architectural scheme, the original architecture used global average pooling to completely eliminate any fully connected layers. As many feature maps as the last convolution layer this was achieved. The number of classes and the application of the soft-max activation function to each map with reduced functionality. The architecture we first proposed is a standard, fully complex neural network consisting of:

1) 9-layer convolution.
2) ReLU.
3) Stack Normalization.
4) Global average pooling.

Fig 2 Samples of the IMDB dataset

This model contains approximately 600,000 parameters. Trained on IMDB Gender dataset containing 460,723 RGB images. Each image belongs to the "Woman" or "Man" class and We achieved 96% accuracy with this dataset. This model of the FER2013 dataset. This record contains 35,887 grayscale images where each image belongs to one following class {"angry", "disgust", "scary", "happiness", “sad, “surprise”, “Neutral”}.

Our initial model achieved 66% accuracy of this data set. Refer to this model .

We designed our second model which is inspired by Xception architecture. This architecture combines the use of two process (residual modules and depth-wise separable convolutions). Residual modules change the desired mapping between two following layers, making the learned features the difference between them. Consequently, the desired features K(x) are modified in order to solve an easier learner problem P(X) such that:

$$K(x)=P(x) + x \longrightarrow 1$$

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

Our initial proposed architecture deleted all last fully connected layers, we reduced further the number of parameters by eliminating them now from the convolution layers. This can be achieved through the use of depth-wise separable convolutions. Depth-wise separable convolutions are composed of two different layers

- Depth-wise convolutions
- Point-wise convolutions

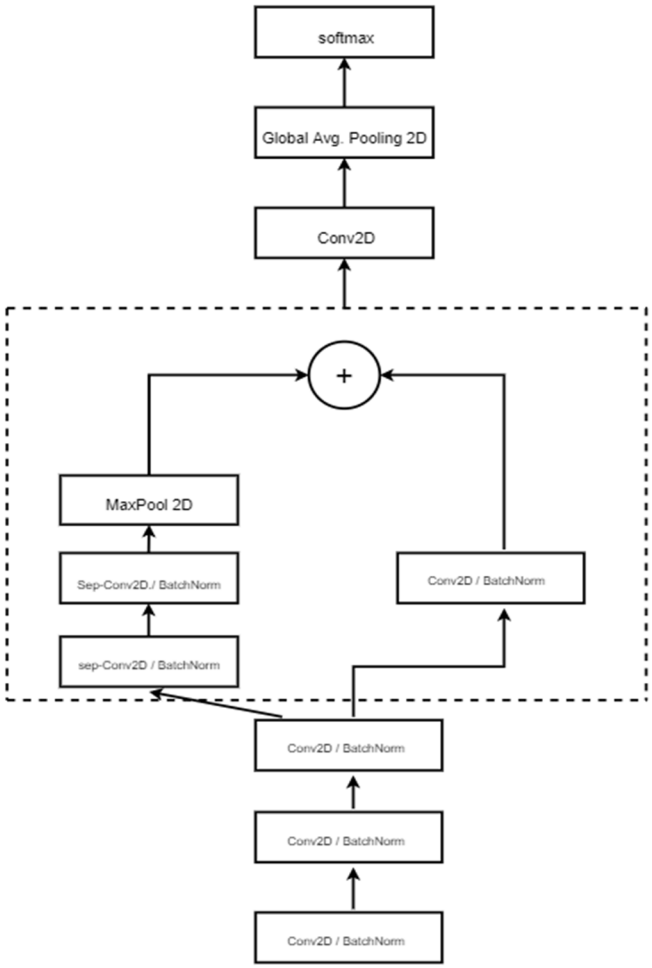
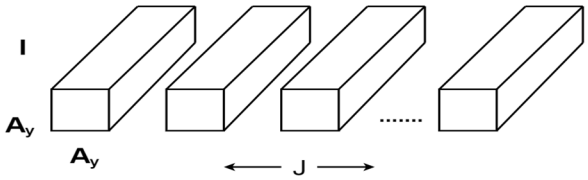


Fig 3 Our proposed model for Emotion Recognition

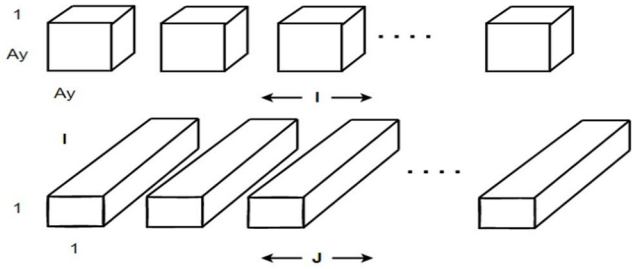
Spatial cross-correlations are separated from channel cross-correlations by these layers. First, D * D filters are applied to every M input channels, followed by N 1 * 1 * M convolution filters to combine the M input channels into N output channels.Convolutions of 1 * 1 * M combine each value in the feature map without considering the spatial relationship between them.

Depth-wise separable convolutions reduces the computation with respect to the standard convolutions by a factor of $x = \frac{1}{J} + \frac{1}{A^2}$. In Figure 4, you can see the difference between a normal Convolution layer and a depth-wise separable Convolution layer.

It is a fully-convolution neural network with four residual depth-wise separable convolutions followed by a batch normalization operation and a ReLU activation function. To produce a prediction, the



(a)



(b)

Fig 4: Difference between (a) Standard convolutions and (b) Depth-wise separable convolutions.

last layer uses a global average pooling function and a soft max activation function. There are approximately 60, 000 parameters in this architecture. This represents a reduction of 10* over our naive implementation, and 80* over the original CNN. Our mini-Xception architecture is shown in Figure 3. This architecture provides 95% accuracy for gender classification tasks. This represents a 1% reduction compared to the first implementation. In addition, we tested this architecture on the FER2013 data set and obtained the same 66% accuracy in the emotion classification task. The final architectural weights can be stored in a 855 kilobyte file. By reducing the computational cost of the architecture, you can now connect both models and use them in sequence in the same image without significantly reducing time. The complete pipeline including open-cv face recognition module, gender classification and emotion classification takes 0.22 ± 0.0003ms on the i54210M CPU. This is 1.5 times faster than Tang's original architecture.

We also added real-time guidance to the implementation Visualization of back-propagation to observe which pixels are in image activates an element in the parent feature map. Given a CNN containing only ReLU as the activation function of The middle tier leads the returned propagation Derivation of each element (x, y) of input image I Regarding the elements (i, j) of the feature map f^l of slice L. The reconstructed image Q excludes all negative gradients. Therefore, the remaining gradients are chosen to increase only the values of the selected elements in the feature map. According to [11], the image of slice 1 completely reconstructed with ReLU CNN is given by the following equation.

$$\longrightarrow \quad Q_{i,j}^{l} = (Q_{i,j}^{l+1} > 0) * Q_{i,j}^{l+1} \qquad 2$$

Emotion recognition is divided in to

A. Data Set

A data set is a collection of related, distinct pieces of data that can be accessed individually or in combination or managed as a single entity.

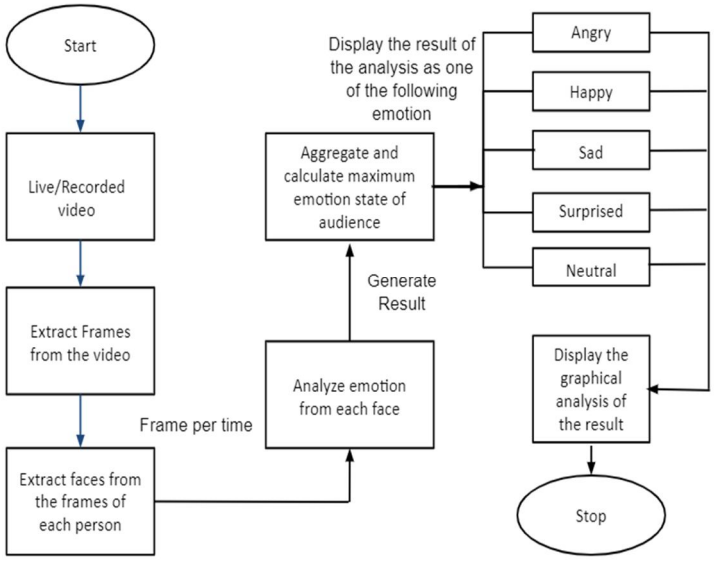


Fig 4: Process of Emotion detection

A data set is structured into a data structure of some sort. A data set in a database, for example, could contain a collection of business data (names, salaries, contact information, sales figures, and so forth). The database itself, as well as bodies of data inside it relating to a certain sort of information, such as sales data for a specific corporate department, can be deemed data sets.
We have taken the date set from FER-2013 data set and IMDB data set.

- https://www.kaggle.com/datasets/msambare/fer2013
- https://data.vision.ee.ethz.ch/cvl/rrothe/imdb-wiki/

B. Input

An input is what we supply to any programme or action as the first data to be processed in order to get the desired outcome. In other words, input is the information that is fed into the programme that is being run. The input for this project can be provided in a variety of ways. A video stream or a picture comprising the faces of persons to be examined can be used as the input. Because the project is built on analyzing audience reaction during any seminar or talk, the input should ideally be a recorded video of the entire event or even a live video stream of the current event. The video input can be of any size or length. The application accepts all video formats, including mp4, mkv, and others, in any duration.

C. Frame Extraction

In general, video is a large volume item with high redundancy and insensitive data, with a complex structure that includes scene, shot, and frame. The key-frame extraction is a vital item in the structure analysis; it allows us to summarizes videos and browse enormous collections of videos. A key-frame is a frame or group of frames that accurately depicts the complete content of a short video clip. It must include the majority of the video clip's most important characteristics. Extraction Faces The cropping of the faces of each of the individuals present in the frame from all of the frames extracted during the previous step is the next stage of the programme. These chopped faces will be saved in a buffer file as an image and used to analyse emotions. After the faces have been cropped, they will be turned into grey scaled photos, as previously discussed, for improved accuracy and analysis of the emotional state. It makes no difference whether a person's face has been cropped from a previous frame or not; if a face appears in any of the frames, it is treated as a new image that must be analyzed to determine emotional condition.

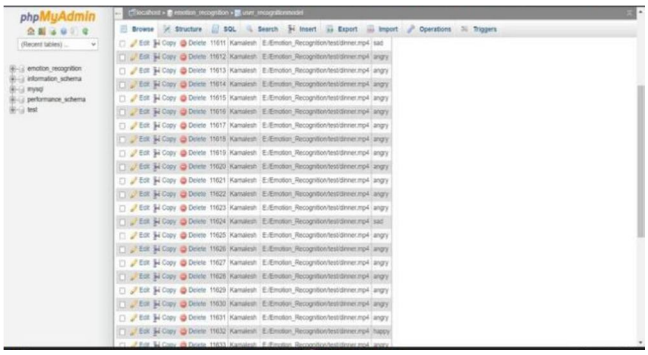
D. Emotion Extraction

The program's final step is to generate emotional value from the clipped photos of faces. The software runs in a loop, analyzing all extracted or cropped faces and generating emotion for each image using the convolution neural network technique. These photos are saved in a different folder called Emotion output, which is located in the same directory as the programme. The photographs will be saved with Emotion as the image's name. That emotion will be in relation to the emotion depicted in the image.

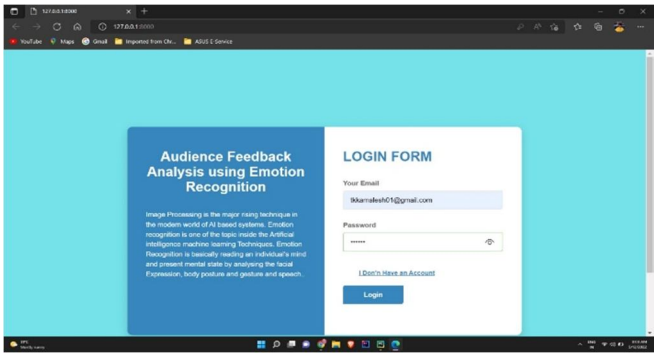
IV. RESULTS AND DISCUSSIONS

From this project, we observed various emotions of a person or a group of people is analyzed. These detected emotions are classified in to any one of these classes{"angry", "disgust", "scary", "happiness", “sad, “surprise”, “Neutral”}. Finally, it generates graphical representation of those emotions classes. The desired response for the movie based on deep learning will be the ensuing average output.

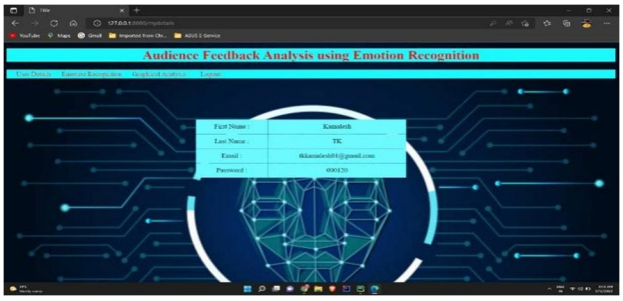
A. Database Storage



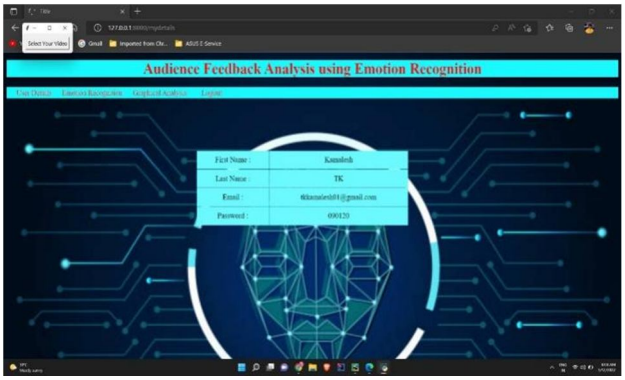
B. Login Screen



C. Home Screen

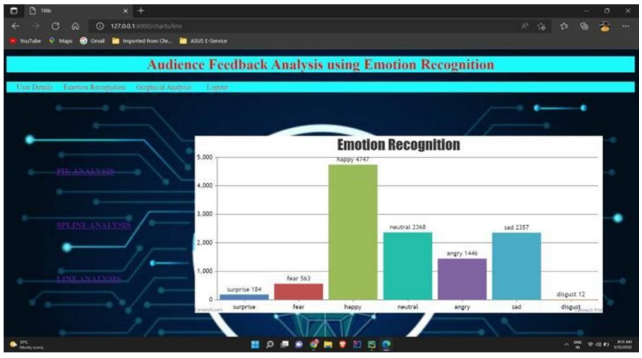


D. Video Selection Window



E. Frame Emotional Recognition

F. Bar Graph



G. Pie Chart

H. Spline Graph

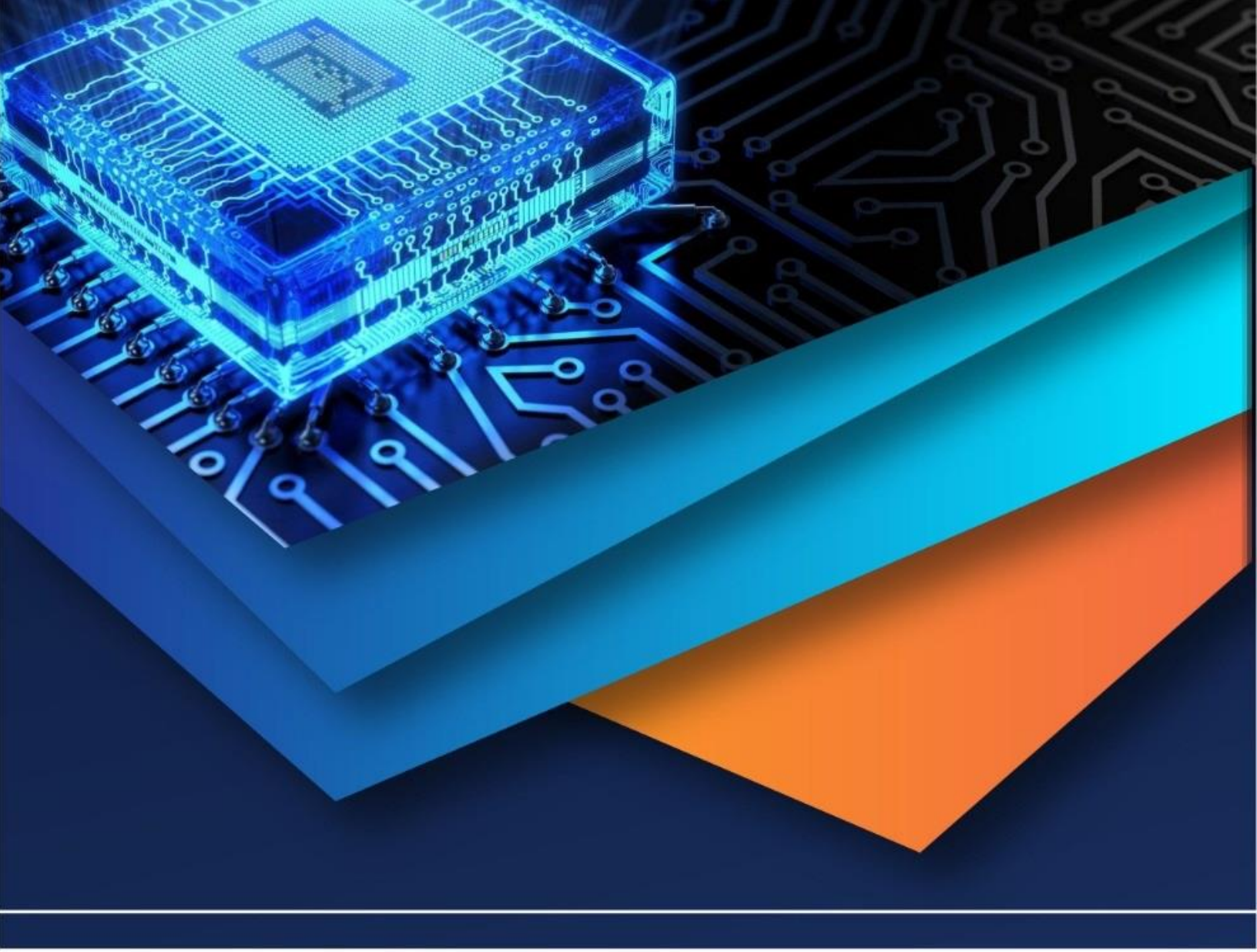
I. Emotion Recognition

V. CONCLUSION

Automatic face detection is difficult, if not impossible, to describe since human forms can appear in a variety of ways in photographs. Not only because of their physiognomy qualities and attributes, but also because of the differences in their 2-D perception related to position (orientation/leaning, scaling). Automatic face detection in digital photos is a capability in digital cameras, and facial recognition applications are included in security systems, to name a few examples. In our project, we attempted to incorporate the entire ongoing process into various modules, which assisted us in obtaining high accuracy for the project's stated goal. The training of the algorithm to detect emotional states was the section where we ran into some issues. The software is used to detect emotional states based on facial traits, specifically the shape of the face, the position of the mouth, and the shape of the nose. We plan to expand the application in the future to include more cues for emotion recognition, such as voice and body position, and to compare our final solution to existing solutions in terms of performance and emotion detection accuracy.

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.
[14] T. Shiva, T. Kavya, N. Abhinash Reddy, Shahana Bano, “Calculating The Impact Of Event Using Emotion Detection”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-7 May, 2019.
[15] Debishree Dagar, Abir Hudait, H. K. Tripathy, M. N. Das, “Automatic Emotion Detection Model from Facial Expression”, 2016 International Conference on Advanced Communication Control and Computing Tech- nologies (ICACCCT)
[16] Loredana Stanciu, Florentina Blidariu “Emotional States Recognition by Interpreting Facial Features,” in The 6th IEEE International Conference on E-Health and Bioengineering - EHB 2017
[17] Binh T. Nguyen,Minh H. Trinh, Tan V. Phan and Hien D. Nguyen, “An Efficient Real-Time Emotion Detection Using Camera and Facial Landmarks,”in The 7th IEEE International Conference on Information Science and Technology Da Nang, Vietnam; April 16-19,2017.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY