IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁒ Impact Factor 7.105 ⁒ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.9685

A Small Convolutional Neural Network for Detecting Real-Time Facial Expressions

Bhavana M P[1], Prof. K. Sharath[2]

Student, Department of MCA, Bangalore Institute of Technology, Bangalore, India[1]

Professor, Department of MCA, Bangalore Institute of Technology, Bangalore, India[2]

Abstract: In this paper, our team suggests and creates a convolutional neural network (CNN) that is lightweight and can distinguish between face expressions of emotion gradually and in mass to achieve a superior classification impact. We can determine the success of our model by developing a consistent vision framework. This framework uses multiple task flowed to finish face localization and pass the gathered face directions to the facial feelings categorization model that we previously created. neural networks (MTCNN). The feeling classification assignment is then completed. Flowed convolutional networks have an outpouring recognition include, one of which can be used alone, reducing the control of memory assets. Global Average Pooling is used to replace the fully related layer in the conventional deep convolution neural network model. The entirely related layer's discovery abilities are somewhat reduced because the feature map's channels are connected to the appropriate class. Simultaneously, the remaining modules and depth-wise separate convolutions are combined in our model removing massive amounts the model's boundaries and compressing it. Then, using the FER-2013 dataset, we run our model. The task of characterizing looks can be completed with just 0.496GB, or 3.1% of the 16GB memory. Our model has a 67 percent precision on the FER-2013 dataset and fits in an 872.9 kilobyte file. On non-dataset figures, it also has significant discovery and recognition effects.

Index Terms: Emotion recognition, lightweight CNN, consistency, and articulation recognition

PRESENTATION

Individuals can convey complex work to computers to satisfy particular market and life needs, thanks to rapid advancements in human-PC communication and example recognition, as well as rapid updates in PC equipment. It is very beneficial to humanity.A recent intelligent human-computer interaction method is facial expression recognition. It has numerous applications, including VR games, clinical consideration, online schooling, driving, and security. Many cameras now have grin mode, which means that a picture without the subject's consent is automatically taken when the camera detects a smile the need for the client to physically press the shade, improving the client experience. Look recognition is used in a few European countries to capture and assess the mood swings of primary school students in class. Several models of Nizam Uddin Ahamed were in charge of organising the audit of this original copy and approving it for distribution as the partner supervisor.

To detect fatigued driving and avoid traffic accidents, Lexus, Toyota's premium brand, monitors the driver's expressions and eyes. People's appearance is one of the most crucial means by which they express their emotions. It is simple to infer someone's inner thoughts from his or her outward appearance.
Facial expression's fundamental ability is to capture the subject's emotional change through facial feelings.
Appearances vary more than other modes of communication. On the spur of the moment, it is easier to express one's true feelings. Ekman [1] was the first to identify six essential articulation structures: bitterness, satisfaction, dread, revulsion, shock, and outrage. An ordinary articulation is now included in the FER-2013 dataset[2].
Fig.1showsthesamplesoftheexpressionsfromtheFER-2013dataset[2]. As we can see, physically determining them is difficult. Furthermore, among the seven emotions, people can accurately classify facial photos with a 63.5% accuracy. Image classification and item location are two of the most advanced image-management strategies.
convolutionneuralnetworks. LaerenceandGiles [3]Shinetal proposes a cross-bred brain network for human face recognition that combines nearby self-organizing map (SOM) neural network, convolutional brain network, and image testing. [4] Consider PC-supported discovery issues using deep convolutional Their model has 5,000 to 160 million boundaries and requires a lot of computer hardware. [5] created a convolutional neural network to remove highlights from information images. The difficulty A mindful classification calculation reduces the complexity of look recognition brought on by ecological factors by dividing the dataset into a basic classification test subspace and a complicated classification test subspace. [9] so that eigenvalues can be separated more vigorously. The CNN design of these endeavours is expected to have a large number of constraints [10], making Sending on inserted devices is challenging. Large convolution parts on high-layered highlight charts used to merely reduce dimensionality do not lead to unnecessary

DOI: 10.17148/IARJSET.2022.9685

calculations in GoogleNet [11] and AlexNet [12]. Further reducing model complexity and the number of parameters is possible by employing continuous big convolution kernels as opposed to tiny convolution kernels. We suggest and create a real-time convolution neural network architecture for face emotion recognition. OurmodelemploysaGlobalAverage Instead of a fully associated layer, the pooling layer combines the lingering To gradually eliminate several borders and streamline our organisational structure, module- and profundity-wise divisible convolution was used. Furthermore, the FER-2013 acknowledgement rate dataset is 67% accurate.

FIGURE 1: FER-2013 emotion dataset samples

I. CONNECTED WORK

The two main categories of face expression recognition techniques used nowadays are deep learning network models and manual approaches.

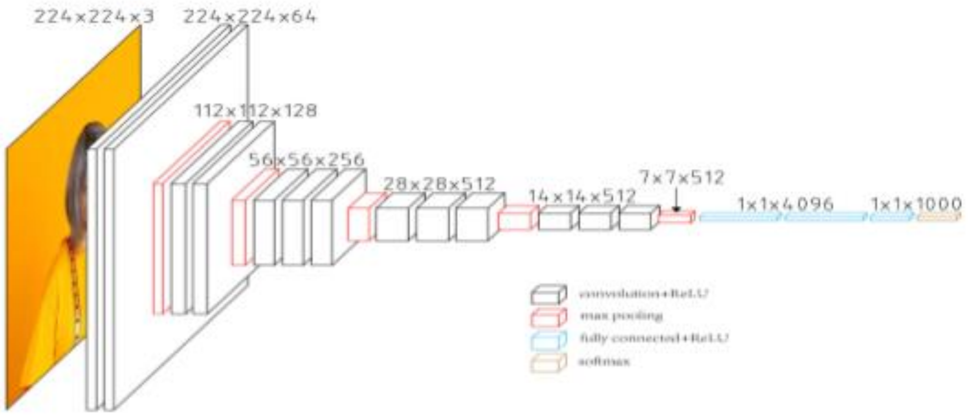
The classical approach is often applied, although its functional applications are quite constrained [13], [14].

In order to solve the issue of glance recognition, Xiao et al.
[26] integrated the Region of Interest (ROI) and K-Nearest Neighbour computation profound brain networks' limited speculation capacity due to limited information.

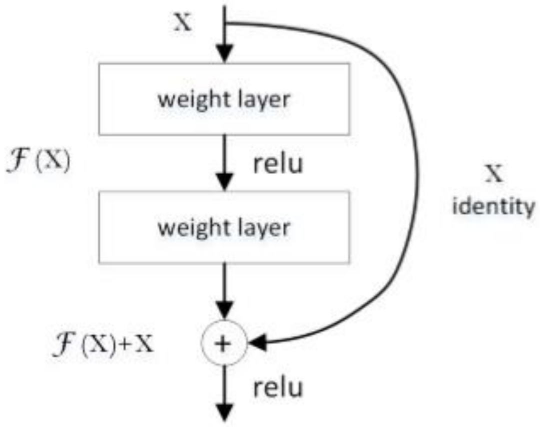
[27] Liu et al. suggested a deep learning method. An approach using a geometric formula for identifying facial expressions model of the facial region is used. Under normal conditions, A simple articulation identification approach that can address the postpone issue was presented by Zhao et al.

[28]. Using a neural network model for face credit identification that was based on movement, with the goal of determining how to group faces based on normal facial highlights. The part used for highlight extraction in some standard exemplary CNN models typically contains a slew of completely near the end, related layers Furthermore, the number of Boundaries is extremely low in fully connected layers. For example, the final fully linked layers of VGG Net

[30] comprise around 90% of all of their borders. VGG16's primary goal is to demonstrate that increasing the. For a given ROI, stacking because several nonlinear layers can improve the depth of the image, using tiny convolution bits instead of stacking big convolution bits organisation, resulting in more complicated designs and fewer boundaries in the learning strategy. Figure 2 depicts the VGG16 organisation design, which consists of There are 13 convolution three totally interconnected layers and levels. The system as a whole uses the same convolution kernel size (33) and maximum pooling size (22). Furthermore, in comparison to a single big filter convolution layer, a combination of several tiny filter convolution layers (33) performs better (55 or 77). Third, it is demonstrated that by continuously developing the organisation, the presentation can be improved.



VGG16 architecture is depicted in Figure 2. VGG-Net, on the other hand, has flaws. It uses more calculating resources and employs more boundaries, which results in higher memory utilisation. The first fully associated layer is responsible for the majority of the boundaries. Inception V3 [31], By incorporating a Global Average Pooling activity, an open source model recently reduced the quantity of final layer parameters. Layers that are inextricably linked coordinate and result in the component depiction. This activity reduces the effect of element area on classification significantly. It does, however, have a few drawbacks, such as a vast number of parameters, which slows down training speed and makes overfitting easy. By taking the average of all components in the element, Each component image is reduced to a scalar value by Global Average Pooling. The organisation can Using the standard activity, take the supplied image and extract global features. Substantial residual learning [33] in addition to depth-wise distinct convolutions [34]are two deep residual learning examples and learning are used in Xception [32], a modern CNN design, to further reduce the number of parameters. It can be improved by isolating the convolution layer's element extraction and arrangement cycles. Increasing the organization's scope is something that both VGG16 and Inception V3 are striving to do in order to increase their precision. The first problem with deeper organisational structures is that these additional layers are indicators of boundary refreshes. Because the inclination is created in reverse, the slope of the front layers will be minimal as the network depth increases. This implies that these layers' learning problem more difficult. As a result, simply increasing the organization's depth leads to more training mistakes The birth of ResNet has solved this difficulty ResNet's central concept, Shortcut Connection,is depicted in Figure 3.ResNet has finally adopted a Global Average Pooling layer, as has GoogLeNet. The lingering module allows you to build a 152-layer remaining organisation. Our model combines GoogLeNet capability in conjunction with the Layer of Global Average Pooling in the



3. ILLUSTRATION

Alternate way association.

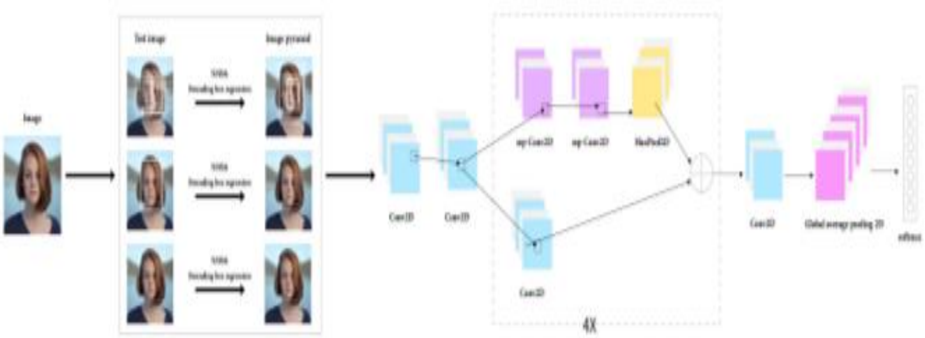
End, as well as reducing the number of organisational layers.' A 2-standard is included to regulate the weight coefficient. These enhancements will provide the model areas of strength in terms of capacity as well as a reasonable recognition rate

DOI: 10.17148/IARJSET.2022.9685

III. METHOD

A. DATASET

ThispaperadoptstheopensourcedatasetFER-2013[2]. We really want to use due to the unique dataset, pandas was used to parse and extract the photos is in CSV format. After parsing, the dataset has 35,887 looks. 28709 is the train set, 3589 is the set for private validation and 3589 is the set for public approval. An picture in grayscale with a set size of 48 48 pixels makes up each figure. Ones that seven articulations that correspond to advanced names 0-6: 0 = outrage; 1 = disdain; 2 = dread; 3 = cheerful; 4 = unhappy; 5 = astonished; 6 = ordinary. The The train set includes 3995, 436, 4097, 7215, 4830, 3171, and 4965 figures with seven different types of articulations. WIDERFACE is a face recognition benchmark dataset with 32203 images and 393,703 faces appearances.



4th FIGURE Tests the dataset WIDER FACE

The scale, presence, and occlusion of these expressions vary greatly. WIDERFACE chose images primarily derived from the open data collection WIDER. The Chinese University of Hong Kong's founders also selected 61 WIDER occasion categories. As the preparation, validation, and testing sets, one of 40%, 10%, and 50% is randomly selected for each categorization. A sample from the WIDER FACE dataset is shown in Figure 4. The Karolinska Directed Emotional Faces (KDEF) are 4900 images of human faces that are part of a dataset of small information tests. This image collection includes 70 people, each with a distinct profound articulation, such as impartial, blissful, furious, apprehensive, nauseated, miserable, and shocked. There are 35 men and 35 women among them.

Figure 5 KDEF dataset testing is depicted.

B. MODELING

This paper proposes and plans a sensation recognition model based on the MTCNN [35] discovery strategy. Instead of using the conventional OpenCV face detection, we switch to MTCNN, which employs and has detection cascade methods recently demonstrated good identification results.

DOI: 10.17148/IARJSET.2022.9685

In the final exploratory test, we had excellent results. We remove the impedance variables of the image's various countenances to greatly improve the impact of emotion recognition. We benefit from the possibility of Xception [32] in the expression acknowledgment model, which the usage of profound lingering learning and depthwise separable convolutions is combined. The primary objective of this design method is to obtain the highest possible identification precision across a wide range of border proportions. To totally eliminate the linked layer, our first model implements GlobalAveragePooling. This is done by inserting a feature map proportional to the number of classes in the final convolution layer and dealing with the classification problem using the SoftMax initiation function. The ADAM enhancer [36] was applied to our model. Figure 6 depicts the model development for our demeanour classification.

FIGURE 6. The design of our model.

TheNetworkinNetworkmodelproposedbyLinetal.[37] replaces the completely linked layer in the classic deep convolution brain network model with the Global Average Pooling technique and obtains good performance on the CIFAR100 dataset The model of applying global normal pooling brings indisputable value to the organisation yield layer channels. It assigns the corresponding classification category to each channel of the element map. This technique, to some extent, kills the completely associated layer's black box qualities [38].

As a result, this paper extends on the concept by using the Global Average Pooling Operation to average each component guide of the element combination and using it as another component map. Worldwide Average Pooling can be linked to global data, strengthening the link between spatial data and learning more point by point and broad look highlights. The pooling layer, on the other hand, has no boundaries, reduces network boundaries, and prevents overfitting. The pooling layer, however, has no bounds, which decreases inhibits overfitting and defines network boundaries. (1)

defines the final picture's size in the convolution activity, where W is the framework's width, H is the matrix's height, Fi is the convolution kernel's width and height, Pispadding is its value, and Si is its step size.

Because the element map's size may remain unchanged after convolution due to the cushioning action of a similar mode, the padding mode the language used in this study is the same. (2) determines the size of the result picture when using the Pooling activity, and the final outcome is lessened. (W + F + 2P) S= (H F +2P)+1S (1) (W F) S +1 (W F) S +1(W F) S +1 (2) The orange appears in the same mode The illustration section is blue, the filter portion is blue, and the backdrop portion is white.

When the filter is active, the focus point (K) corresponds to one and the filter will execute a convolution operation on the image's sides. As may be observed, the range of motion has been reduced compared to before. During forward proliferation, the same mode can keep the element map's size constant.

When the filter's focal point (K) corresponds to a side the filter then applies a convolution procedure on the picture .As may be observed, the range of motion has been reduced compared to before. During forward proliferation, the same mode can keep the element map's size constant.

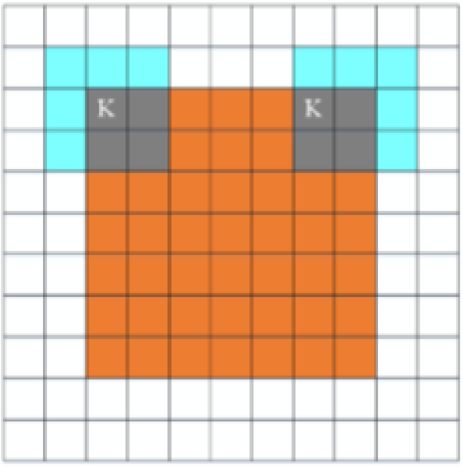


FIGURE 7. Same mode.

Despite the fact that our collection has increased the dataset, the number of images remains small, and there are numerous duplicatefaces. At the same time, the pixel density of the images is extremely low, and the FER-2013 dataset contains some noisy images. We add '2-standard to the weight coefficient to avoid overfitting. We selected '2-standard over '1-standard because '2-standard can obtain boundaries with minor qualities. Furthermore, '2-standard can keep our ideal arrangement steady and quick while preventing overfitting. In general, the fitting system will keep the loads as low as possible and eventually build a model with as few parameters as possible .The model's anti-disturbance capacity is solid when the parameters are small enough. '2-standard calamity work' is (3), and it can be written as (4). (5) indicates that the slope plummet technique is used to refresh the boundaries, no matter how small. It is equivalent to multiplying each framework by a factor of one (1/m), where m is a positive number. As a result, The term "2-standard" has also been used. "Weight Loss."

P I = 1 L(y (I), y I + 2m kwk 2 (2) J0 = J0 + 2mkwk2 (4) J w = J0 w plus mw w0 = w J w = w J0 w m w = (1 m) wJ0 w (5)

By including '2-standard,' you can keep the weight framework from becoming excessively large the obvious understanding is to set the weight of numerous secret units to zero if the regularisation is sufficiently enough and the weight network W is near to zero, effectively killing many of these secret units' effects. The extraordinarily simplified brain organisation in this case will be Despite being several organisational layers deep, it is as tiny as a logistic regression unit. As a result of the organization's proximity to the provinces of "High fluctuation" and "High predilection," respectively, when a median value is provided, the organisation may be regarded as "just right," as illustrated in Fig. 8.

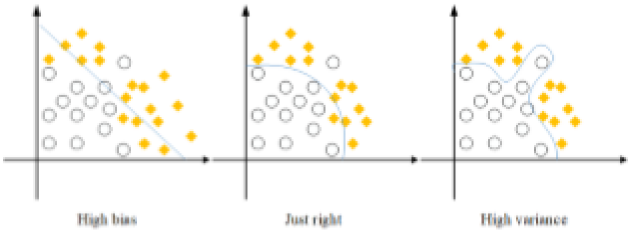


FIGURE 8. Network effects of the "2" standard.

As a result, if is sufficiently expanded, W will approach zero, but this will not occur. To simplify the organisation and eventually get closer to calculated relapse, we try to remove or reduce the effects of stowed away units. We have a strong intuition that there are a large number of hidden units have been completely removed. In reality, all secret brain units will always exist, but their effects will diminish as the brain network becomes less difficult, making overfitting less likely. This model also incorporates lingering depth-wise separable convolutions and modules. Traditional convolutions are combined with a depthwise convolution and an 11 convolution to form depth-wise separable convolutions. A typical convolution is shown in Figure 9(a). Accept that the convolution bit is DK DK M in size and the feature map input is DF DF M in size. The outcome includes The normal convolution layer is DK DK M N in size, whereas the map is DF DF N. Convolution is shown profundity-wise in Figure 9 (b) and point-by-point in Figure 9 (c).

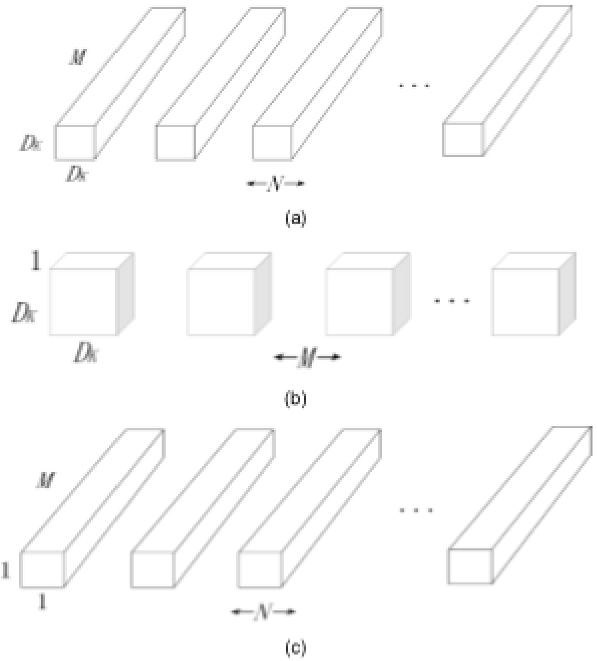


FIGURE 9.

Convolution channels (a)Traditional convolution channels (b)Depth-aware convolution channels
(c)Convolution channels are referred to as point wise convolution in profundity wise detachable convolution.

The union of these two convolutions is referred to as profundity wise separableconvolution. Filtering is handled by depth-wise convolution, with the size Following up on each information channel are the numbers (DK,DK,1) and M.

The channel change is handled by point-by-point convolution, N is the number, size is (1,1,M), and results include profundity wise convolution planning. Because the number of profundity wise distinct convolution boundaries is the standard convolution the number of profundity-wise unique convolutions is (1/N + 1/D2 K) boundaries is (1/N + 1/D2 K).

We used a brain network with four lingering profundity wise distinguishable convolutions and the appearance was created using an MTCNN finding. Immediately after each of these four convolutions are clusters of standardised activity and ReLU [39, 40] initiation work.

The final layer includes both the Global Average Pooling layer and a delicate maximum enactment work for classification. There are 58423 boundaries in this design, 56951 of which are teachable. Our tests are performed based on the FER-2013 dataset Opinion precision The categorization rate is 67 per cent. Additionally, we may lighten the burden of our final product acknowledgment design in an 872.9 kilobyte file.

RESULTS

Our investigation is powered by an Intel (R) Core (TM) i5-8400 16GB of RAM, an NVIDIA GeForce GTX 1060 GPU, a CPU running at 2.80 GHz, and the Ubuntu 16.04 LTS activity framework We use the WIDER FACE dataset to retrain the MTCNN model while keeping the weight bounds that were established for face detection in the chart file.

The figure 10 shows the MTCNN's genuine upward trend which can reach around 95 percent. We achieved excellent The MTCNN model was use of OpenCV, the industry-standard Application Programming Interface (API), to get results in face recognition. The effects of utilising OpenCV to detect faces and MTCNN to detect faces, respectively, are depicted in Fig. 11 in (a) and (b), respectively. Clearly, (b) has had a location impact is investigated.

DOI: 10.17148/IARJSET.2022.9685

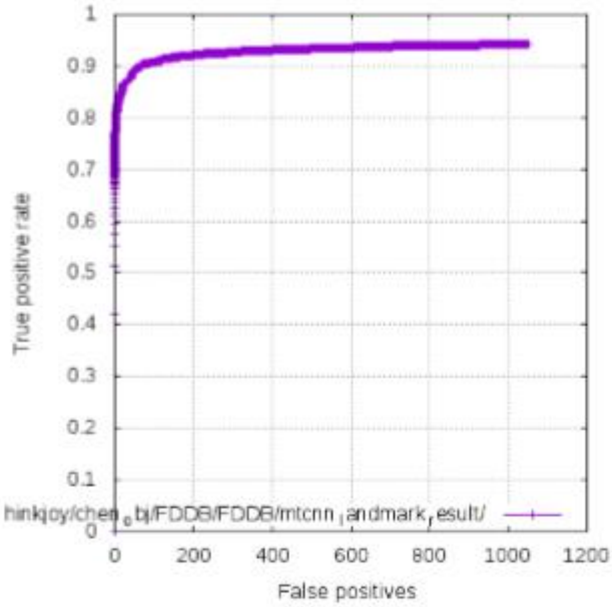
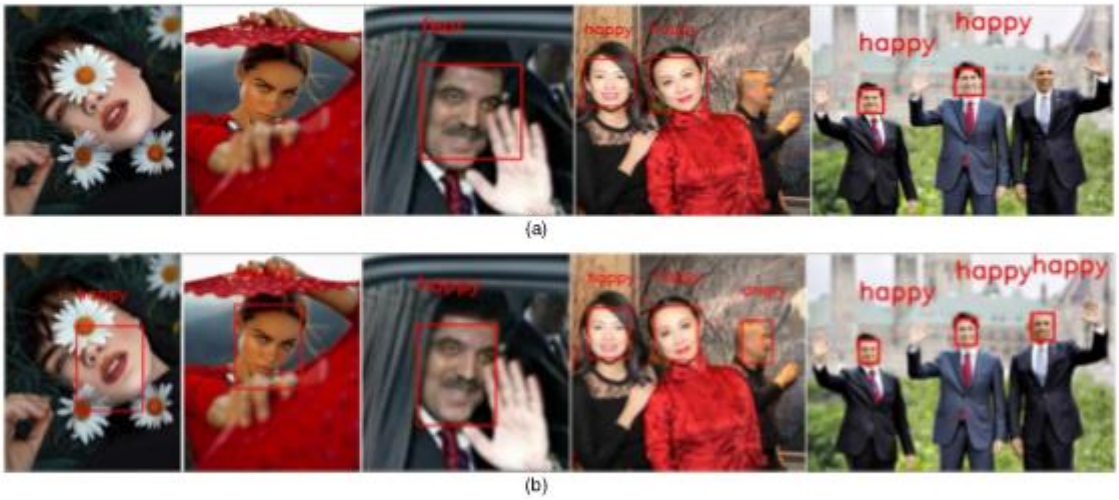


Figure 10 depicts the MTCNN's genuine positive pace.

The recognition rate of the seven articulations is depicted in Table 1 based on our organisation engineering's normalised disarray grid. The accuracy in perceiving the fear category remains flawed, as should be obvious. The primary cause of this outcome is that the number of ears in the dataset has a low diversity. It implies that the majority of them have European faces and that there are no information tests of various kinds.

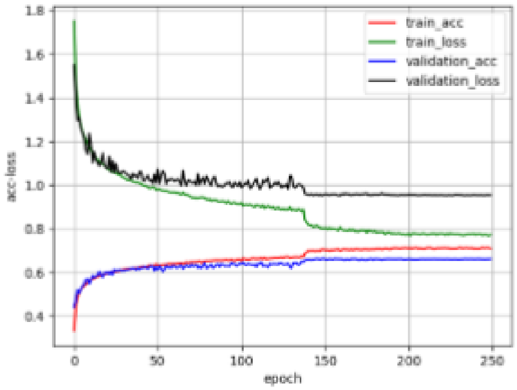


11th Figure. Correlation illustrating the consequences of two unique discovery modules

Our model	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	0.62	0.01	0.10	0.03	0.12	0.13	0.10
Disgust	0.26	0.55	0.05	0.04	0.06	0.02	0.03
Fear	0.12	0.01	0.40	0.05	0.20	0.11	0.11
Happy	0.02	0.00	0.02	0.88	0.02	0.02	0.05
Sad	0.10	0.01	0.10	0.09	0.53	0.01	0.15
Surprise	0.03	0.00	0.10	0.04	0.02	0.75	0.03
Neutral	0.04	0.00	0.05	0.08	0.11	0.02	0.65

DOI: 10.17148/IARJSET.2022.9685

TABLE 1: Our model's aftereffect testing on the FER-2013 dataset.



PICTURE 12: The FER-2013 dataset in our model shows an acc-misfortune bend.

The Loss is depicted in Figure 12, worth and Accuracy worth change curves after 250 epochs of the four models on the FER-2013 data set. As illustrated in According to In Figure 12, the accuracy rate on the train set may reach around 71 percent, whereas the accuracy rate on the approval set may be as low as 67 percent. Our final model's impact on the test set is shown in Figure 13 at the same time with a precision ranging from 66.8 to 67.0 percent. It should be clear that our model has a deep residual learning after consolidating separate profundity-wise convolutions somewhat serious level of precision. Since we utilise Global

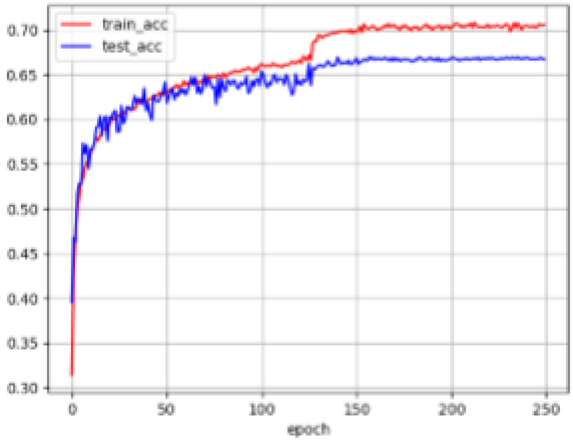


FIGURE 13. Expectation to learn and adapt on a railroad set and in testing set.

We use Global Normal because The model's boundary count is lowered. by pooling instead of the completely associated layer and adding '2-standard, making our model more versatile. Table 2 looks at the trial outcomes of our model and a few other models. We are unable to retrain their models because the code is not distributed in some form. Their literature can be used to determine the correctness on the test set of their models. Figure 14 is plotted through Table 2, to more naturally demonstrate the disparity in precision between these models 11th Figure: Correlation of two distinct discovery modules' effects.

CONCLUSION

This study presents and develops a lightweight convolutional neural network for face expression recognition developed by our group. In our network model, the fully connected layer is eliminated, the remaining depth-wise separable convolution is combined, and the "2-norm regularisation term" is added to minimise the number of parameters in the convolutional layer. Moreover, our model's detection and classification capabilities are not noticeably harmed. By identifying images that are not in the dataset, our model achieves good detection results, demonstrating that the multiclassification of facial expressions is appropriate for the model established in this research. We developed a visual system that may generally be included into low-power computer devices in order to classify facial expressions and narrow down a large number of parameters achieves the model created in this research is suitable for facial expression multiclassification, as shown by successful detection results by detecting photographs outside of the dataset. We

DOI: 10.17148/IARJSET.2022.9685

developed a visual system that may generally be included into low-power computer devices in order to classify facial expressions and narrow down a huge number of variables. Our model's accuracy is higher when compared to recent models, Moreover, based on experimental findings, it has produced effective detection outcomes in photos outside the dataset. Despite the fact that our model generated some results, there might be a lot of noise in the pictures taken in reality. For example, pictures with too strong or too dark lights, blurry pictures, pictures where the majority of the face is blocked, and other things that are not good for pictures could all cause noise accurate results.

REFERENCES

1. P. Ekman and W. V. Friesen, "Face and Emotion Constants Across Cultures," Journal of Personality and Social Psychology, vol. 17, no. 2, 1971, pp. 124–129.
2. "Three Representation Learning Machine Learning Competitions," I. J. Goodfellow et al., Neural Netw., vol. 64, April 2015, pp. 59–63.Face recognition using a convolutional neural network, or
3. IEEE Transactions on Neural Networks, S. Lawrence, C. L. Giles, A. Chung Tsoi, and A. D. Back , vol. 8, no. 1, Jan. 1997, pp. 98-113.

I. Nogues, J. Yao, D. Mollura, M. Gao, L. Lu, Z. Xu, H. R. Roth, H.-R.M. Summers, C. Shin, and C. Shin, "Deepconvolutional CNN designs, dataset characteristics, and transfer learning in neuralnetworks for computer-aided detection," IEEE Transactions on Neural Networks, vol. IEEETrans.

4. Med.Imag., volume 35, issue 5, pages 1285–1298.A complexity perception classification method for identifying facial expressions 2018 [Online] T. Chang, G. Wen, Y. Hu, and J. Ma arXiv:1803.00185 You may access it at: http://arxiv.org/abs/1803.00185C. IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1285–1298, May 2016. "Deepconvolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics, and transfer learning."
5. T. Chang, G. Wen, Y. Hu, and J. Ma developed a complexity perception classification technique for identifying facial emotions in [Online] arXiv:1803.00185. You may access it at: http://arxiv.org/abs/1803.00185.
6. M.-I. Wang, "Local learning using deep and tailored features for facial emotion recognition." IEEE Access, vol. 7, no. 7, 2019, pp. 64827–64836, by R.T. Georgescu, Ionescu, and M. Popescu.
7. Du, vol. 5, pp. 15750–15761, "Multi-scale convolutional neural network-based IEEE Access," S. Gao and C., 2017. picture segmentation-based multi-focus image fusion.
8. M. Z. Uddin, M. M. Hassan, A. Almogren, A. Alamri, M. Alrubaian, and G. Fortino, "Face emotion recognition using strong local direction-based signals and a deep belief network," IEEE Access, vol. 5, pp. 4525-4536, 2017.
9. M. Z. Uddin, W. Khaksar, and J. Torresen, "Identification of face emotions using salient features and a convolutional neural network." IEEE Access, vol. 5, pp. 26146-26161, 2017. M. Z. Uddin, W. Khaksar, and J. Torresen. Deep Speech 2: End-to-end Speech Recognition in English and Mandarin
10. D. Amodei et al., Proceedings of the National Academy of Sciences, vol. Int. Conf. Machine Learning, June 2016, pp. 173–182.
11. "Inception-v4, inception-ResNet, and the influence of residual connections on learning," in Proceedings of the 31st AAAI Conf. Artif. Intell., 2016, p. 1. C. Szegedy, S. Ioffe, and V. Vanhoucke.A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification using deep convolutional neural networks," in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2012, pp. 1097–1105
12. Multi-class classification of lung endomicroscopic pictures by M. R. Koujan, A. Akram, P. McCool, J. Westerfeld, D. Wilson, K. Dhaliwal, S. McLaughlin, and A. McLaughlin, Perperidis, in Proceedings of the IEEE 15th International Symposium on Biomedical Imaging (ISBI), pp. IEEE International Conference on Acoustics, Speech Signal Processing (ICASSP), May 2019. International Conference on Acoustics, 5866-5870.
13. Frontiers in Human Neuroscience, vol. 7, no. 8, December 2013, p. 810, "Face, body, and scene emotional signals alter viewers' facial expressions, fixations, and pupil size." K. Roelofs, J. J. Stekelenburg, B. de Gelder, M. E. Kret, and Deep affect prediction in the wild: An aff-wild study by S. Zafeiriou, D. Kollias, P. Tzirakis, M. A. Nicolaou, A. Papaioannou, G. Zhao, B. Schuller, I. Kotsia, and deep architectures, the database challenge, and beyond "Affective Behavior Analysis in the First ABAW 2020 Competition," Int.
14. D. Kollias, S. Zafeiriou, E. Hajiyev, and A. Schulc. 2020, arXiv:2001.11409. [Online] http://arxiv.org/abs/2001.11409 is the URL to access it.
15. W.Y. Choi, K.Y. Song, and C.W. Lee, "Convolutional attention networks for multimodal emotion identification from voice and text data," in Proc. Grand Challenge Workshop Hum. Multimodal Lang. (Challenge-HML), 2018, pp. 28-34.