WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

FACIAL EXPRESSION RECOGNITION BASED ON MOBILENETV2

Ishan Tharwal, Omkar Patil, Pramod Yadav, Sankirti Shiravale

Marathwada Mitra Mandal's College of Engineering

ABSTRACT

Facial Expression Recognition is an important part of human-machine interaction and has piqued the curiosity of many computer scientists. Convolution Neural Network is used to construct a facial expression recognition system (CNN). In this research, a Kaggle face expression dataset with seven facial emotion labels (happy, sad, surprise, fear, rage, disgust, and neutral) is utilized. MobileNetV2 – a pre-trained model from keras library is used and transfer learning is implemented. MobileNetV2, which is quicker and more accurate, is implemented in a real-time framework that produces accurate and timely results. Additional thick layers are added to the model, as well as the ReLU and SoftMax activation functions for the seven classes of happiness, sadness, neutrality, anger, fear, surprise, and disgust. Images for classification can be provided via code or through the webcam of the computer. For face detection, Haar Cascade is used.

Keywords*: Facial Expression Recognition (FER), Convolutional Neural Network (CNN).*

1. INTRODUCTION

Facial expression is one of the most important cues for identifying human emotion and intention in people. An autonomous system for precise and reliable facial expression analysis, driven by recent breakthroughs in human-centered computing, has growing applications such as interactive gaming, online/remote education, and entertainment.

In this contemporary world, deep-learning is the fundamental approach to analysis and classification problems. The development of Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), and many more neural networks paved the path for design variants and enhancements such as MobileNetV2, AlexNet, VGG, and ResNet. They are often trained in a variety of classes. MobileNetV2 and AlexNet, for example, were trained on 1000 different classes of images. This makes the model, in theory, develop universal features, and better at differentiating between objects. Emotion classification introduces a new set of challenges, since the model needs to be able to differentiate interfacial features. We use DL for, both, feature selection and expression classification. MobileNetV2 is an excellent starting point for Transfer Learning. Because of the many classes the model was trained on, there is an expectation that universal features had been learned by the network. The model will be trained or tuned using 224x224 photos as input. Then, we will classify each image as one of seven different expressions classes.

Emotion recognition has several applications. By monitoring client behavior, emotion detection will aid in the improvement of products and services. Emotions are important in psychiatric issues, and this strategy can aid with that.

2. LITERATURE SURVEY

- Face Recognition Based on Convolution Neural Network. In this work, A face recognition method based on convolution neural network (CNN) is presented in this paper. The recognition rate is 98.95%. Furthermore, the network has great convergence and resilience.

- Convolutional Neural Network (CNN) for Image Detection and Recognition. The study investigates many deep learning methods, including CNN, and achieves image recognition and detection on MNIST and CIFAR –10 datasets using just the CPU unit. MNIST accuracy is good, however CIFAR-10 accuracy can be enhanced by training with bigger epochs and on a GPU unit. MNIST estimated accuracy is 99.6 percent, whereas CIFAR-10 calculated accuracy is 80.17 percent.

- Design and implementation of a convolutional neural network on an edge computing smartphone for human activity recognition. They demonstrated a deep convolutional neural network model for classifying five daily activities utilizing raw accelerometer and gyroscope data from a wearable sensor as input. The accuracy of the proposed model was 96.4 percent.

- Facial Emotion Recognition: A Survey. After analyzing many techniques of Facial emotion detection, they came to a conclusion that the AlexNet CNN solution is the best way to do facial emotion detection compared to others on the basis of accuracy and pricing. After researching numerous ways that combine Histogram of Oriented Gradients with different classifiers and HOG with other methods, HOG-

ESRs algorithm provides greater accuracy with increased robustness than previously available methods. Among all datasets CK+ gives the best result in many cases.

3. DATASET AND FEATURES

We utilised a dataset given by the Kaggle website for this research, which comprises of around 28,709 well-structured 48 x 48-pixel gray-scale photos of faces. The images are processed such that the faces are almost centered and each face fills around the same amount of space in each image. Each image has to be categorized into one of the six classes that express different facial emotions. These facial emotions have been categorized as: 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad and 5=Surprise. Augmentation is done by resizing the 48 x 48 pixel images to 224 x 224 pixel. Each image, then, underwent a standard array of preprocessing and data augmentation that is recommended for use in CNN.

A. Preprocessing

For face identification and cropping, we chose the OpenCV Haar cascade function. Normalization is done by dividing the image array with 255.

B. Augmentation

Resizing is done as the images in FER2013 dataset are 48 x 48 pixel and the MobileNetV2 requires image pixel size as 224 x 224.

C. Normalization

Normalization is carried out by dividing the images array by 255.0. Dividing by 255 expresses a 0-1 representation since 255 is the maximum value. Each channel (red, green, and blue) is 8 bits, so they are each restricted to 256, since 0 is included, 255 is used in this case.

4. METHODS

Using the transfer learning approach, a pre-trained network (MobileNetV2) was pre-trained using the FER2013 dataset. First, we describe the network architecture and how it was tailored to the purpose of emotion recognition. We next go over the learning algorithms that were utilized to train these networks on our new datasets in depth.

A. Network Architectures: Before discussing the structure and scale of the architecture, we will provide a broad overview of the subcomponents employed in this network.

1) CNN components:

- **Convolutional Layers**: CNNs are distinguished by its use of convolutional layers, which distinguishes them from other typical neural network architectures such as multi-layer perceptions. A number of 2-dimensional "feature maps" or images layered on a third dimension or "depth" are typically given into a particular input layer, and a smaller rectangular "kernel" is convolved with the image. The explicit formulation of this procedure is given below, where h is the input layer, an is one kernel, and the output is one feature map layer. There are various reasons to use this image convolution method. Convolution provides "temporal" independence for pixel characteristics. i.e., extracted features are only functions of nearby pixels because pixels in images are typically only related to pixels in nearby regions and independent of pixels far away. Another reason to use convolution is that the resulting feature map is usually smaller than the original input, as shown by the output size relationships in the following equations.

  $$Wo = (Win - Fw + 2P)/S + 1$$

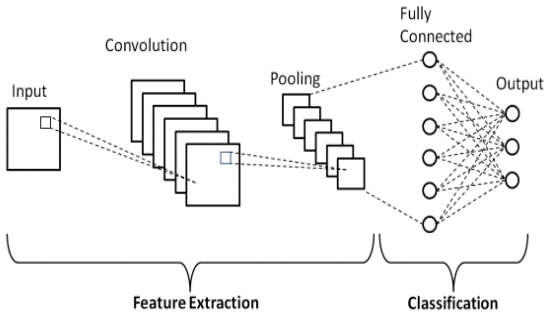
  $$ho = (hin - Fh + 2P)/S + 1$$

  Where wo represents the output width, win represents the input image width, fw represents the kernel width, p represents the padding size, and s represents the stride used. Height is also stated in the latter equation. This reduction in the size permits a method of dimensionality reduction, which is essential when reducing the number of features in huge images (curse of dimensionality).

- Linear Rectified Units (ReLU). ReLU is a non-linear activation function that is employed at the output of each convolution layer. The network can learn nonlinear functions thanks to ReLU. The reason for using ReLU over other activation functions is that it decreases the "vanishing and exploding" gradient problem that deep neural networks experience when back propagating. This unit allows for 6 times the training speed of tanh while retaining consistency.

- Maximum Pooling. Max pooling reduces dimensionality by lowering the output size of each layer and providing network translational invariance. The latter is significant because if a feature is slightly displaced by a few pixels, it is collected in a "maxpool" and routed to succeeding layers of a bigger perimeter. This has been found to perform better than other pooling approaches such as "average pooling.".

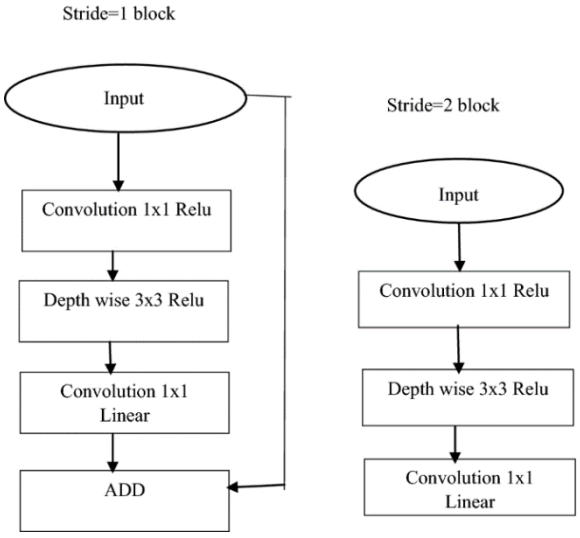
- Drop Out. In CNNs, this strategy reduces overfitting by limiting the number of trainable parameters and momentarily blocking brain connections. As neurons specialize for certain features, randomly deactivating weights causes the network to build new routes and feature representations across the network.



Architecture of CNN

2) Architectures & Adaptations: The CNN was trained with pretrained weights, then tweaked for our unique job and fine-tuned on our new dataset.

MobileNetV2: In MobileNetV2, there are two sorts of blocks. The first one is a residual block with a stride of 1. Another alternative for downsizing is a block with a stride of 2. Both blocks have three levels each. The first layer is 11 convolutions with ReLU6, the second layer is depth wise convolution, and the third layer is another 11 convolutions with no non-linearity. If ReLU is applied again, deep networks are said to have just the power of a linear classifier on the non-zero volume part of the output domain.



MobileNetV2

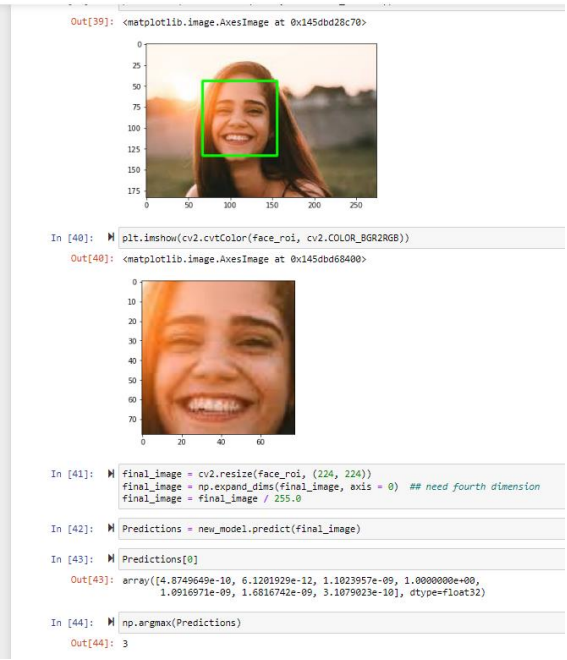
5. TRAINING AND TESTING

The pre-trained MobileNetV2 is used to train on FER2013 dataset and as the pre-trained model contains 1000 predictions. The last layer is removed and additional layers are added. After the output of global layer, the layers added are in the following sequence: dense layer, relu layer, dense layer, relu layer and finally SoftMax function layer with our 7 classes.

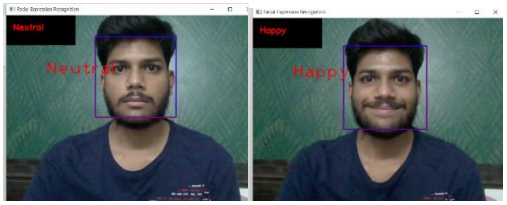
Following is the table of Performance of MobileNetV2: The model has training accuracy of 95.05% at 25th Epoch.

Epoch	Training Loss	Training Accuracy
1	1.3649	0.4821
2	1.1500	0.5713
3	1.0591	0.6016
4	0.9819	0.6330
5	0.9235	0.6510
6	0.8292	0.6921
7	0.7588	0.7215
8	0.6827	0.7491
9	0.5982	0.7829
10	0.5190	0.8141
11	0.4701	0.8272
12	0.4037	0.8567
13	0.3547	0.8722
14	0.3232	0.8857
15	0.2739	0.9021
16	0.2484	0.9148
17	0.2331	0.9171
18	0.2135	0.9261
19	0.1951	0.9333
20	0.1620	0.9442
21	0.1735	0.9382
22	0.1640	0.9418
23	0.1547	0.9475
24	0.1485	0.9466
25	0.1415	0.9505

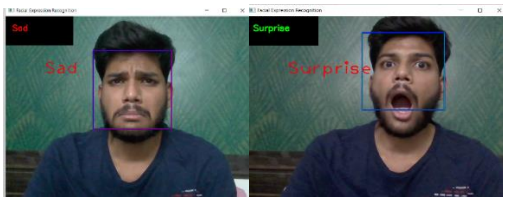
For detecting faces, Haar cascade is used. Testing is carried out by giving a random image downloaded from google. A happy girl image which is not present in the dataset is supplied and the model predicted the class label - ‘3’ which belongs to happy.



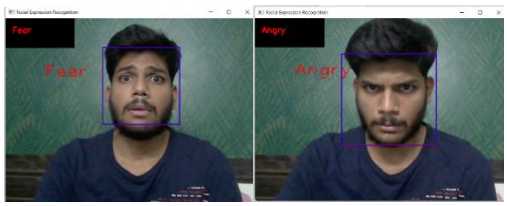
After testing the model on a single image, by using laptop's camera or webcam, live image feed is captured and supplied to the model for predicting. Again, Haar Cascade plays a vital role here.



(a) Neutral **(b) Happy**



(c) Sad **(d) Surprise**



(e) Fear (f) Angry

(g) Disgust

6. COMPARISON WITH ANOTHER RESEARCH

Following tables shows comparison of training accuracy achieved between our model and other models, which have used the same dataset i.e., FER2013. In the first research mentioned in the table, normal CNN is used and Decision tree is used at the end to classify the output. On the other hand, transfer learning used in our project gives more accuracy than the previous method.

Research	Accuracy
Human Facial Expression Recognition Using TensorFlow and OpenCV - Saransh Srivastava	94.93%
MobileNetV2 (Our model)	95.05%

7. CONCLUSION

In this research, we implemented facial expression recognition using transfer learning. Transfer Learning was performed using the MobileNetV2 pre-trained model. MobileNetV2 is one of the models that provides high accuracy and accurate prediction. Furthermore, by applying batch normalization, the accuracy may be enhanced, and more accurate results can be achieved, which we are contemplating for future study.

REFERENCES

[1] Yan, K., Huang, S., Song, Y., Liu, W., & Fan, N. (2017). Face recognition based on convolution neural network. 2017 36th Chinese Control Conference (CCC). doi:10.23919/chicc.2017.8027997

[2] Prakash, R. M., Thenmoezhi, N., & Gayathri, M. (2019). Face Recognition with Convolutional Neural Network and Transfer Learning. 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT). doi:10.1109/icssit46314.2019.8987

[3] Guo, S., Chen, S., & Li, Y. (2016). Face recognition based on convolutional neural network and support vector machine. 2016 IEEE International Conference on Information and Automation (ICIA). doi:10.1109/icinfa.2016.7832107

[4] Wang, D., Yu, H., Wang, D., & Li, G. (2020). Face Recognition System Based on CNN. 2020 International Conference on Computer Information and Big Data Applications (CIBDA). doi:10.1109/cibda50819.2020.00111

[5] Chauhan, R., Ghanshala, K. K., & Joshi, R. . (2018). Convolutional Neural Network (CNN) for Image Detection and Recognition. 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC). doi:10.1109/icsccc.2018.8703316

[6] Yang, J., & Li, J. (2017). Application of deep convolution neural network. 2017 14th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). doi:10.1109/iccwamtip.2017.830148

[7] Zebin, T., Scully, P. J., Peek, N., Casson, A. J., & Ozanyan, K. B. (2019). Design and implementation of a convolutional neural network on an edge computing smartphone for human activity recognition. IEEE Access, 1–1. doi:10.1109/access.2019.2941836

[8] Ajit, A., Acharya, K., & Samanta, A. (2020). A Review of Convolutional Neural Networks. 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE). doi:10.1109/ic-etite47903.2020.049

[9] Blot, M., Cord, M., & Thome, N. (2016). Max-min convolutional neural networks for image classification. 2016 IEEE International Conference on Image Processing (ICIP). doi:10.1109/icip.2016.7533046

[10] Nisa, S. U., & Imran, M. (2019). A Critical Review of Object Detection using Convolution Neural Network. 2019 2nd International Conference on Communication, Computing and Digital Systems (C-CODE). doi:10.1109/c-code.2019.8681010

[11] Sultana, F., Sufian, A., & Dutta, P. (2018). Advancements in Image Classification using Convolutional Neural Network. 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN). doi:10.1109/icrcicn.2018.8718718

[12] D, K., Hebbar, R., P V, V., M P, H., L, J., & S H, M. (2018). CNN Based Technique for Systematic Classification of Field Photographs. 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C). doi:10.1109/icdi3c.2018.00021

[13] Mujawar, S., Kiran, D., & Ramasangu, H. (2018). An Efficient CNN Architecture for Image Classification on FPGA Accelerator. 2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC). doi:10.1109/icaecc.2018.8479517

[14] Thai, Le Hoang, Nguyen Do Thai Nguyen, and Tran Son Hai. "A facial expression classification system integrating canny, principal component analysis and artificial neural network." arXiv preprint arXiv: 1111.4052 (2011).

[15] Wei, X., Chen, Y., & Zhang, Z. (2020). Comparative Experiment of Convolutional Neural Network (CNN) Models Based on Pneumonia X-ray Images Detection. 2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI). doi:10.1109/mlbdbi51377.2020.0009.

[16] S. Zhang, X. Zhao, and B. Lei, "Facial expression recognition based on local binary patterns and local fisher discriminant analysis," WSEAS transactions on signal processing, vol. 8, no. 1, pp. 21-31, 2012.

[17] F. Y. Shih, C.-F. Chuang, and P. S. Wang, "Performance comparisons of facial expression recognition in JAFFE database," International Journal of Pattern Recognition and Artificial Intelligence, vol. 22, no. 03, pp. 445-459, 2008.

[18] P. Kumar, S. Happy, and A. Routray, "A real-time robust facial expression recognition system using HOG features," in 2016 International Conference on Computing, Analytics and Security Trends (CAST), 2016: IEEE, pp. 289-293.

[19] P. Liu, S. Han, Z. Meng, and Y. Tong, "Facial expression recognition via a boosted deep belief network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1805-1812.

[20] D. Arumugam and D. S., Emotion Classification Using Facial Expression, International Journal of Advanced Computer Science and Applications, vol. 2, no. 7, 2011.