Facial Expression Detection using Convolutional Neural Network

*

Mudit
Department of Computer Science
Chandigarh University
Gharuan, Punjab, India
muditms1150@gmail.com

Divya K
Department of Computer Science
Chandigarh University
Gharuan, Punjab, India
divya.e12116@cumail.in

*Sanjeev Kumar Joshi
Uttaranchal University
Dehradun, India
248007
mehod@uttaranchaluniversity.ac.in

Sahil Verma
Faculty of Computer Science and Engineering
SGT University
Gurugram, 122505, India
sahilverma@ieee.org

2022 International Conference on Cyber Resilience (ICCR) | 978-1-6654-6122-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICCR56254.2022.9995993

***Abstract*—A face expression is a clear representation of an individual's affective state, cognitive activity, intention, personality, psychopathology, and it serves as a means of communication in interpersonal relationships. Automatic facial expression recognition is a critical segment of usual interface between human and a machine, and it must be used for behavioristic psychology therapeutic practice as well. Face identification and placement in a chaotic scene, facial feature extraction, and facial feature categorization are all tasks that an autonomous face features Recognition system must complete. Convolution Neural Networks are used to create a facial expression detection system. LeNet Architecture is the foundation of the CNN model. During this experiment, we have used the Kaggle facial features dataset in which we acquire seven types of face expressions which indicates our emotions swiftly. Kaggle facial feature data set with 7 face expressions labeled as (happy, sad, surprise, fear, rage, disgust, neutral) was taken for the research. Through this paper we are able to obtain 56.7% accuracy and 0.576 precision while testing the dataset.**

***Index Terms*—Convolution Neural Network, facial feature extraction, facial features, dataset**

I. Introduction

Facial expression is the most powerful form of nonverbal communication, conveying information about one's spirit, thinking, and intention. Facial expressions where influence the discussion's flow, they also allow its listeners to communicate a plethora of informative data to the speaker without saying anything. When the facial expression does not sync with the uttered words, the data that has been conveyed by the face has a greater influence on comprehending the information [24], [26]‘. Face expression analysis that is automated aids in human-machine interaction. However, it is not easy to achieve [10], [17]. Many facial expression traits are extracted and assessed for fine sentiments analysis utilising deep learning CNN [3], [23]. CNN is a feedforward artificial neural network pattern of connectivity between neurons that is taken by the organization of animal visual cortex [**?**], [25] in the field of machine learning. Each cortical neuron responds to stimuli in a small area known to be receptive field. These fields having different neurons partially overlaps, resulting in a visual field tile. Convolutional networks are multilayer perceptron versions that employ minimum pre-processing and are inspired by biological processes [1], [18]. Before machines take over more of our lives, human-machine contact should be improved to more closely resemble human-to-human interaction. Face detection, facial feature points extraction, and facial feature categorization are the three primary components of an automatic countenance identification system. The endeavor requires the system to collect an input image and apply image processing algorithms to it in order to locate the facial region. Face localization is the term used in static photos, whereas face tracking is used in videos [15], [24]. Face detection and tracking, feature extraction, and expression categorization are the three most prevalent phases in facial emotion identification. The stage where face detection is done searches for the face' region in input photos or do the sequencing without the aid for human's intervention. The next stage that has been worked upon is to get discriminative information created by facial expression after face will be positioned. The final level of the system is face feature recognition. The alterations in the face are frequently classified as prototypic emotions or facial action units. [7], [9] Section II is about the related work of different author about face expression deep learning and face expression recognition. Section III explains the methodology for the paper Section IV explain the results and the analysis and followed by conclusion.

978-1-6654-6122-1/22/$31.00 ©2022 IEEE

Authorized licensed use limited to: Florida Institute of Technology. Downloaded on January 17,2024 at 07:57:33 UTC from IEEE Xplore. Restrictions apply.

II. RELATED WORK

As Face expression detection is the most vast topic available here is the list of various works done on this field. Das, Sumit, et al [6] presents the real time emotion recognition system using machine learning techniques and implemented it using python. The system can be used in the health care field and further aims to implement the method to display emoji for the emotion. The result shows 70% accuracy and it has been noticed that positive emotions are more accurate then negative emotions. Matsumura, Naoki, et al [21] proposed one novel structuring the scarce fully-connected layer (FCL) in CNN. It uses cuBLAS to implement the suggested sparse FCL on GPU. FCLs obtain a speed-up factor of 14.97 and 16.67 for forward and backward propagation, respectively. Kumar, B. K., Swaroopa, K., Balaga, T. R et al [17] implemented the facial expression detection using sliding window approach and support vector machines. The result investigates the topic of face emotion analysis. Alqumboz, M. N. A., Abu-Naser et al [4] proposed his study on avacado using deep learning. Natarajan, V. Anantha, et al [23] focuses on the segmentation of nuclei using Convolutional Deep Neutral Architecture. The segment accuracy, as measured by the Dice Co-efficient, is used to evaluate the deep learning model's performance. It outperforms a fully Convolutional Neural Network in terms of performance. Jie, H. J., Wanda et al [13] introduced dynamic pooling layer renowned as run pool for training of CNN architecture. The outcome reveals that the planned CNN can produce the greatest score of 94 percent. Das, S., Sanyal, M. K., Kumar Upadhyay et al [4] focuses on Machine learning was used in this work to predict cardiac disease. SVM, Decision Tree Classifier, Random Forest Classifier, and k Nearest Neighbour Classifier were the algorithm employed in this study. The results suggest that K Nearest Neighbour Classifier fared best in terms of accuracy score. It received a score of 86.84 percent. Das, S., Sanyal, M. K et al [5] suggested a prototype known as a machine intelligence diagnostic system (MIDs) that can learn, think, reason, and manage ambiguity in the same way as a real-world clinician can. The outcome demonstrates the intelligent result of MIDs, which are able to function to be doctor to some measure for compensating the world's doctor's shortage. Das, S., Sanyal, M. K., Datta, D et al [5] demonstrates progress of a diagnostic model for the therapy of pain in lower back, which includes one mathematical model that specifies root of disease as well as AI-assisted hardware implementation. The suggested idea learns through examples from the data from non-linear regression, as illustrated in results. Deshmuk,Renuka, et al [7]] proposed a survey to report an illustrative study of most popular emotion recognition methods which are used in emotion recognition problems.It gives the brief overview towards the process , different techniques and application of facial emotion reognition system. Fu, Gang, et al [8] presents the upgraded FCN model has been used to provide precised classification strategy in high resolutions remote sensing data. The avg. precision, recall Kappa coeff. for aor method shows 0.81, 0.784 0.83, respectively. Perveen, Nazia, et al [24] proposed that a face expression is believed to be compromised of disfigurement face parts changes in face pigmentation. Six types of training data set has been taken and RST - Invariant features are used for better results.The result shows maximum accuracy of 90% and surprise has better result of accuracy then others. LeCun, Y. et al [18] explained about Le-Net-5 architecture of CNN. LeNEt-5 is designed for handwrittten and machine-printed character recognition. Soo, Sander et al [27] offered a case study on a vehicle detection and counting system and the opportunities it will bring in a semi-enclosed region - both statistically and for the average person. The results reveal that the classifier correctly detects some vehicles, but that it also incorrectly classifies some sections of the pavement and some grass as cars.

Yin, Lijun, et al [30] attempts to create a 3d face expression database for research into affective computing and detailed 3d structure in human emotions the results reveal that the number of expression types is still limited to the prototypic expression space, implying that additional spontaneous expressions are needed to analyse spontaneously occurring emotions.

Matsugu, Masakazu, et al [20] shows a rule-based algorithm for robust facial expression recognition in addition with robust face detection with CNN. The result presents genuine detection of smile having recognition rate upto 97.6% for 5600 static pictures of about more ten subjects

Sundaram, N. M., Sivanandam, S. N et al [29] investigates the usefulness of the SoftMax function in Neural network as an activation function for multi-class classification issues

III. METHODOLOGY

Convolutional neural networks are applied to create the facial expression detection system. The following is a block diagram:

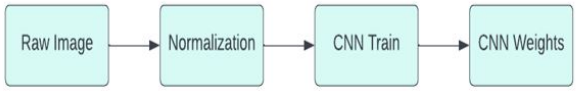

Fig. 1. Training Phase

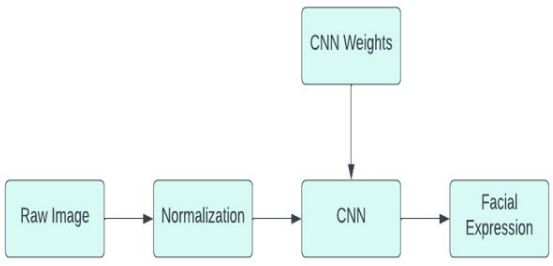

Fig. 2. Testing Phase

Authorized licensed use limited to: Florida Institute of Technology. Downloaded on January 17,2024 at 07:57:33 UTC from IEEE Xplore. Restrictions apply.

Figure 1 and 2 shows training and testing phase. While training, the system obtains training data in the form of greyscale photographs of the faces and expression labels linked with them, as well as set of weights needed for network. A photograph featuring a face was used as input for the training step. Subsequently, the image is subjected to intensity normalization. After that, the normalized images are utilised to help train the Convolution Networks [19]. To avoid order of presentation of the examples affecting training performance, a validation data set is used to determine the ultimate best set of weight from a set of training with samples displayed on multiple orders. The weights' set which have produced the best results for training data is output for training step. In this way system gets greyscale picture of face throughout the test then outputs expected expression by using final network weight that are learned. The result will be a single number derived from the seven fundamental expressions [29].

A. Dataset

The dataset for training testing is used through a challenge called Kaggle Expression, Detection Challenge [3]. It contains pre-cropped grayscale images labeled with seven classes of emotions. Dataset consists training set with 35910 images along their labels. There is issue of class imbalance as examples of some classes is larger than others which is balanced increasing the numbers of classes which are in minorities. The balanced data set have 40284 images and out of them 29340 are for training,6000 for testing and 4944 for validation. Figure 3 shows distribution data for training, testing and validation.

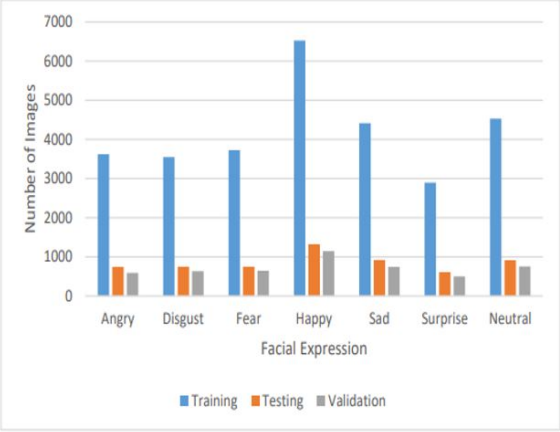

Fig. 3. Training, Testing and Validation Data distribution B. Architecture of CNN

IV. Architecture of CNN

An input layer, many convolut. layers, fully connected layers, and one output layer make up the CNN architecture. The CNN based on the LeNet Architecture [30] with several modifications. The CNN employed in this model's architecture. [2], [6], [28]

A. Input-Layer

The image must be pre-processed before being sent into this layer, which has a already determined, already fixed dimension. The training, validation, and testing are all done with Kaggle [20] normalised grayscale photos of size 48/48 pixels. Propose laptop webcam photos to test them, where the face is recognised, cropped, and normalised with the aid of OpenCV Haar Cascade Classifier [27]. [6], [31]

B. Convolution Pooling Layer

Convolution and pooling [13] have been completed and are now ready for execution. Here each one of the groups have N no of photos, and the Convo. network filter weight on those batches modified. Each convolution layer accepts a four-dimensional image group input of N * Color Channel * Width * Height. The input feature map number, output feature map number, filter's width, and filter's height are all four-dimensional in the feature map or filter for convolution. 4Dimensional convolutions are calculated b/w picture group and mapping is performed in each convolution layer. The only argument that changes after convolution is the picture's width and height. For dimensionality reduction, down-sampling/sub-sampling is conducted after each convolution layer. Pooling is the name given to this procedure. Two well-known techniques of pooling are maximum-pooling mean pooling. After convolution, maximum pooling is completed throughout this operation. The pool's size is 2/2, it divides the image in a grid of 2*2 blocks, every one of them using a maximum of 4 pixels. Only the height and breadth are modified prior to pooling. The architecture employs 2 convolution layers and one pooling layer. The image batch size is N*1*48*48 at principal convo. layer size of input. The picture batch-size dimensions are N in this case. The number of color channels are one, image's height and width are both 48 pixels. The convolution consisting a 1*20*5*5 features map resultant batch of image has a dimension N*20*44*44. Convolutional pooling consisting pool having size 2/2 is completed, resulting in one picture's group having size N*20*22*22. That is accompanied with a 2nd convolutional layer having 20*20*5*5 feature map, resulting in a photo batch of size Nx20x18x18. That frequently accompanied with one pooling layer having a pool size of 2/2, that concludes with N*20*9*9 picture group [6].

C. Fully Connected Layer

The way neurons transmit impulses across the brain stimulates the Fully Connected Layer [2][34]. This layer takes the amount of input qualities and transforms them into features using layers with trainable weight vectors. The fully-connected layer has 2 layers, first of which has five hundred units in size and 2nd of which has three hundred units of size. Feed forward propagation is used to train the weights of CNN, and subsequently feed backward propagation has been utilised for transmit mistakes on the behind. Back propagation begins by computing diff. b/w the forecast and true values, as well as load fine tuning required by every layer, that is capable of managing the training's speed , as a result, design's complexity

Authorized licensed use limited to: Florida Institute of Technology. Downloaded on January 17,2024 at 07:57:33 UTC from IEEE Xplore. Restrictions apply.

	Precision	Recall	F1-Score
Anger	0.39	0.42	0.42
Disgust	0.95	0.99	0.97
Fear	0.45	0.38	0.39
Happy	0.68	0.69	0.69
Sad	0.44	0.38	0.41
Surprise	0.69	0.65	0.67
Neutral	0.45	0.49	0.47
Average	0.57	0.57	0.57

TABLE I
PRECISION, RECALL AND F1-SCORE

after fine tuning excited parameters. Learning rate, momentum., regularisation parameters, and decay are examples of these layers [6], [12], [14], [22].

D. Output Layer

The output derived by second hidden layer is linked with output layer, that contains seven separate classifications. The result is derived considering odds of all 7 classes by the help of the Soft-max activation algorithm [20, 29, 30,32]. Category with highest likelihood is expected class prediction. [6], [11], [16]

V. RESULT AND ANALYSIS

Human emotions are extremely important in this advanced AI era, and machine learning, a subset of AI, is an important part of medical diagnosis [4]. The facial expression detection CNN architecture was created in Python using the Python programming language, NumPy, imultis, OpenCV, tensor flow, scikit-learn, and CUDA libraries. The batch size for training images is 30, and the filter map dimensions for both layers are 20 x 5 x 5. For the process of validating the training's process validation set has been taken in the paper. Validation cost, validation error, and training errors is determined for last batch of each epoch. Image set with accompanying output label is the training's input parameters. On the basis of hyper parameters including learning rate, regularization, momentum and decay, training process updates weight of hidden layers and also the feature maps. The learning rate per batch is 10e-5, the momentum is 0.9, the regularisation is 10e-7, and the decay is 0.9999. 6000 images are used to carry out the testing of model and the accuracy provided by the classifier is 56.7% The precision, recall F1 score of every expression is described using the relation below in Table 1: So, the overall precision and recall that we get is 0.57 and 0.57 respectively. There is high precision scores for happy and surprised as the model does a great job on classifying positive emotions. Due to oversampling disgust has the highest precision and recall of 0.95 and 0.99. The negative emotion seems to have weaker performance. Especially sad emotion has low precision of 0.44 and recall is 0.38. The least expressive faces like sad and neutral makes the prediction most difficult. The overall F1-score is noted as 0.57 which is highest for disgust and fear has the least among all which is 0.39. Thus, the emotion detection system may prove to be helpful in monitoring regulating patients' conditions, as facial expressions can be utilised to intelligently analyse their state by a computer [4] to minimise danger in the case of acute lower back pain [5].

VI. CONCLUSION

The emotions on our face plays the most important part for the purpose of communication and it is very necessary to find the appropriate expression along with what is being said. This paper shows a way to differentiate different categories of emotions. Through this paper is became possible to achieve fine face detection and emotions extraction by just face images. This technology can be applied in places like video surveillances, digitalised camera, security human machine interface. In this paper we have used LeNet architecture based six-layer Convolutional Neural Network and has classified six different human expressions that are surprise, happy, fear, sad, anger, neutral and disgust. The parameter at which this paper is evaluated are Precision, Accuracy, Recall and F1 score. As a result we are able to achieve accuracy of 56.7% ,precision 0f 0.57,recall 0.57 and F1-score is 0.57. Future work: For future progress we are looking forward to implement some methods so that system can show color images and an emoji can be shown for each type of emotion.

REFERENCES

[1]
[2] Munir Ahmad, Taher M Ghazal, and Nauman Aziz. A survey on animal identification techniques past and present. *International Journal of Computational and Innovative Sciences*, 1(2):1–7, 2022.
[3] A. Singh al. A new clinical spectrum for the assessment of nonalcoholic fatty liver disease using intelligent methods. *in IEEE Access*, 8:38470–13848, 2020.
[4] S. Das and M. K. Sanyal. Machine intelligent diagnostic system (mids): An instance of medical diagnosis of tuberculosis. *Neural Computing and Applications*, 32(19):15585–15595, 2020.
[5] S. Das, M. K. Sanyal, and D. Datta. Artificial intelligent embedded doctor (aiedr.): A prospect of low back pain diagnosis. *International Journal of Big Data and Analytics in Healthcare (IJBDAH)*, 4(2):34–56, 2019.
[6] S. Das, M. K. Synyal, S. K. Upadhyay, and S. (2021 Chatterjee. February). an intelligent approach for predicting emotion using convolution neural network. *In Journal of Physics: Conference Series (Vol.*, 1797:1.
[7] R. Deshmukh and M. E. Scholar. A comprehensive survey on techniques for facial emotion recognition. *International Journal of Computer Science and Information Security*, 15:3, 2017.
[8] G. Fu, C. Liu, R. Zhou, T. Sun, and Q. Zhang. Classification for high resolution remote sensing imagery using a fully convolutional network. *Remote Sensing*, 9(5):498, 2017.
[9] G. Ghosh, D. Anand, et al. A review on chaotic scheme-based image encryption techniques. In Hsieh SL., Gopalakrishnan SY., and Duraisamy and S., editors, *Peng*. Intelligent Computing and Innovation on Data Science. Lecture Notes in Networks and Systems, vol 248. Springer, Singapore.
[10] G. Ghosh et al. Secure surveillance system using chaotic image encryption technique. *in IOP Conference Series: Materials Science and Engineering*, 993(1):012062, 2020.
[11] Mamoona Humayun, Farzeen Ashfaq, Noor Zaman Jhanjhi, and Marwah Khalid Alsadun. Traffic management: Multi-scale vehicle detection in varying weather conditions using yolov4 and spatial pyramid pooling network. *Electronics*, 11(17):2748, 2022.

Authorized licensed use limited to: Florida Institute of Technology. Downloaded on January 17,2024 at 07:57:33 UTC from IEEE Xplore. Restrictions apply.

[12] N. Z. Jhanjhi, Sarfraz Nawaz Brohi, Nazir A. Malik, and Mamoona Humayun. Proposing a hybrid rpl protocol for rank and wormhole attack mitigation using machine learning. In *2020 2nd International Conference on Computer and Information Sciences (ICCIS)*, pages 1–6. IEEE, 2020.

[13] H. J. Jie and P. Wanda. Runpool: A dynamic pooling layer for convolution neural network. *Int. J. Comput. Intell. Syst.*, 13(1):66–76, 2020.

[14] M. Kaur, R. Bajaj, and N. A Kaur. Review of mac layer for wireless body area network. In *J*, pages 767–804. Med. Biol. Eng. 41, 2021.

[15] Manjit Kaur et al. Flying ad-hoc network: Challenges an routing protocols. *in Journal of Computational and Theoretical Nanoscience*, 17(6):7, June 2020.

[16] Muhammad Ibrahim Khalil, Mamoona Humayun, N. Z. Jhanjhi, M. N. Talib, and Thamer A. Tabbakh. Multi-class segmentation of organ at risk from abdominal ct images: A deep learning approach. In Intelligent Computing and, editor, *and Innovation on Data Science*, pages 425–434. Springer, Singapore, 2021.

[17] B. K. Kumar, K. Swaroopa, and T. R. Balaga. Facial emotion recognition and detection using cnn. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(14):5960–5968, 2021.

[18] Y. LeCun. Lenet-5, convolutional neural networks. *URL: lecun. com/exdb/lenet*, 20(5):14, 2015.

[19] D. Lekhak. *A Facial Expression Recognition System Using Convolutional Neural Network*. Tribhuwan University, Institute of Engineering, 2017.

[20] M. Matsugu, K. Mori, Y. Mitari, and Y. Kaneda. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, 16(5-6):555–559, 2003.

[21] N. Matsumura, Y. Ito, K. Nakano, A. Kasagi, and T. Tabaru. A novel structured sparse fully connected layer in convolutional neural networks. *Concurrency and Computation: Practice and Experience, e*, 6213, 2021.

[22] S. K. Mishra, S. Mishra, and A. Alsayat. ...sahoo, k. *S., Luhach, A.K., Energy-aware task allocation for multi-cloud networks, in IEEE Access*, 8:78825–17883, 2020.

[23] V. A. Natarajan, M. S. Kumar, R. Patan, S. Kallam, and M. Y. N. (2020 Mohamed. September). segmentation of nuclei in histopathology images using fully convolutional deep neural architecture. In *InInternational Conference on Computing and Information Technology (ICCIT-1441)*, pages 1–7. IEEE, 2020.

[24] N. Perveen, N. Ahmad, M. A. Q. B. Khan, R. Khalid, and S. Qadri. Facial expression recognition through machine learning. *International Journal of Scientific Technology Research*, 5:03, 2016.

[25] Sowjanya Ramisetty et al. The amalgamative sharp wireless sensor networks routing and with enhanced machine learning. *in Journal of Computational and Theoretical Nanoscience*, 16(9):4, September 2019.

[26] S. Rani, D. Koundal, et al. An optimized framework for wsn routing in the context of industry 4.0. *Sensors*, 21:19, 2021.

[27] S. Soo. Object detection using haar-cascade classifier. *Institute of Computer Science, University of Tartu*, 2(3):1–12, 2014.

[28] Tariq Rahim Soomro and Mumtaz Hussain. Social media-related cybercrimes and techniques for their prevention. *Appl. Comput. Syst.*, 24(1):9–17, 2019.

[29] N. M. Sundaram and S. N. Sivanandam. Soft max activation function for Neural Network Multi class classifiers.

[30] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. (2006 Rosato. April). In *A 3D Facial Expression Database for Facial Behavior Research*, pages 211–216. In 7th International Conference on Automatic Face and Gesture Recognition (FGR06) . IEEE.

[31] N. Zaman, T. J. Low, and T. Alghamdi. Enhancing routing energy efficiency of wireless sensor networks. In *International Conference on Advanced Communication Technology*, pages 587–595, 7224928, 2015. ICACT 2015-August.

Authorized licensed use limited to: Florida Institute of Technology. Downloaded on January 17,2024 at 07:57:33 UTC from IEEE Xplore. Restrictions apply.