

Facial Expression Recognition with Histogram of Oriented Gradients using CNN

Sahar Zafar Jumani¹, Fayyaz Ali², Subhash Gururo¹, Irfan Ali Kandhro¹, Asif Khan¹ and Adnan Zaidi³

¹Department of Computer Science, Sindh Madressatul Islam University, Karachi, Pakistan; sahar@smiu.edu.pk, subhash@smiu.edu.pk, irfan@smiu.edu.pk, asifkhan@smiu.edu.pk

²Department of Computer Science, Sir Syed University of Engineering and Technology, Karachi, Pakistan; Fayyaz54@gmail.com

³Department of Computer Science, Muhammad Ali Jinnah University, Karachi, Pakistan; fa17phcs0003@maju.edu.pk

Abstract

Objectives: A new method is introduced in this study for Facial expression recognition using FER2013 database consisting seven classes consisting (Surprise, Fear, Angry, Neutral, Sad, Disgust, Happy) in past few decades, Exploration of methods to recognize facial expressions have been active research area and many applications have been developed for feature extraction and inference. However, it is still challenging due to the high-intra class variation. **Methods/Statistical Analysis:** we deeply analyzed the accuracy of both handcrafted and leaned aspects such as HOG. This study proposed two models; (1) FER using Deep Convolutional Neural Network (FER-CNN) and (2) Histogram of oriented Gradients based Deep Convolutional Neural Network (FER-HOGCNN). the training and testing accuracy of FER-CNN model set 98%, 72%, similarly Losses were 0.02, 2.02 respectively. On the other side, the training and testing accuracy of FER- HOGCNN model set 97%, 70%, similarly Losses were 0.04, 2.04. **Findings:** It has been found that the accuracy of FER- HOGCNN model is good overall but comparatively not better than Simple FER-CNN. In dataset the quality of images are low and small dimensions, for that reason, the HOG loses some important features during training and testing. **Application/Improvements:** The study helps for improving the FER System in image processing and furthermore, this work shall be extended in future, and order to extract the important features from images by combining LBP and HOG operator using Deep Learning models.

Keywords: Deep Learning, Emotion Recognition, Facial Expression, CNN, FER, HOG

1. Introduction

There are important applications of automatic facial expression detection in enormous areas like HCI (human computer interaction) but it is challenging problem though interesting one. Human emotions understandable intelligent robots can be built by automatic facial expression application. This intelligence is also handy in various real-world applications like interactive game development and call center. There are six universal motional expressions according to Ekman. These are named as

fear, disgust surprise, anger, sadness and happiness¹. Face variances can be observed to recognize these expressions. For instance, a signal of happiness can be identified as a gesture of smile by tightened eyelids and raised mouth corners. A person internal affective states, social communication and intentions are indicated by change of facial expressions. Effective states of human beings are expressed predominantly by human face. Many applications in many areas like human emotions analysis, natural HCI, image retrieval and talking heads have diverse effect of automatic facial expression detection. This type of

*Author for correspondence

detection has gained significant attraction and has been an impacting issue in the science community for over the last decade as human beings find facial expressions one of the most natural, immediate and powerful means to express their intentions and emotions². Last stage of this system is facial expression detection. In³ states that there are basically three stage training procedure in expression recognition systems named as classifier construction, feature learning and feature selection. Feature learning stage is first, feature selection is second and classifier construction is last one. Only related facial expressions variations among all features are extracted by feature learning stage. Facial expression is then represented by the best features which are chosen by feature selection. Not only maximizing inters class variation but they also should minimize the intra class variations of expressions. Because same expressions of different individuals in image are far from each other in pixel's space so minimizing the intra class variation of expressions is a problem. Because images of different expressions of same person may be in very close vicinity to each other in the pixel's space, maximizing the inter class variation is not an easy task as well. One or more classifiers with one for each expression are utilized to deduct the facial expressions in the end of process, provided the selected features⁴. The core goal of this research is to extract data set of facial expression based on Histogram Gradient using Convolutional neural network. To improve the understanding and learning of CNN, HOG feature map is fed as CNN's input. Following sections comprises this research paper. Section 2 most recent relevant work in the area of Facial Emotion Detection in Deep Learning Models. In Section 3 explained proposed methodology where HOG feature descriptor is combined with Facial Emotion detection using CNN. The outcome of experimental results is discussed in Section 4. Section 5 compares the result of presented methodology (FER-HOGCNN) with conventional (FER-CNN). Last Section 6 comprises conclusion of this paper.

In visual classification and detection, the performance CNN learned features outclass compare HOG Handcrafted features. This paper compared two chips designed HOG and CNN models and find out the difference between two techniques. These approaches provide potential intuition about optimization and even the accuracy of learned features is twice. It shows (large 311 x to 13486 x above in energy unitization although basic computation overhead exists. Combined approaches highlighted in this paper and potentially reduce memory size and energy and help

to reduce the gap between handcrafted and learned features⁵. Eigen faces algorithm was proposed in 1991 and it was the first face recognition approach introduced successfully. Its foundation laid on PCA⁶. In our proposal, we used 10 different facial images for each 20 individuals (AT&T face database was utilized⁷). It comprises different illuminations and facial expressions on 200 images in total of 92-b-112 pixels⁸. Classifiers as well as face recognition both are extensively utilized by feature extraction. Precision of classifiers are conditionally dependent on the robust selectivity of features from face image and speed of classifiers. Two methods based on edge and local feature extraction have been utilized to counter the noise produced by variations in face pose and illuminations. Four popular methods are local binary pattern (L.B.P)^{9,10}, Gabor filters^{11,12}, Scale Invariant Feature Transform (S.I.F.T)^{13,14} and HOG^{15,16}. Utilizing Eigen faces show less successful results as compared to these four methods. But these methods are unable to handle face pose or illustrations variance unless fed by the additional training samples and input data processing in HCI and other areas, novel application is enabled automatically by the features to identify facial expressions. In^{17,18} the result, many recent works use Convolutional Neural Networks by active researchers in this field for feature identification, processing and generation of output. Because of different characteristics and architecture of CNN, these works differ prominently with each other. As we compare experimental results, the performance significance of these parameters is unclear conducted widespread test on seven free available FER databases facial expression databases, viz., Mulri PIE, MMI, CK+, DISFA, FERA, SFEW, FER2013¹⁹. another study designed method, the modern technique in image-based facial expression identification by utilizing CNN and performance impact of algorithmic differences are highlighted. This will be the base to set direction to move ahead in this research field by identifying current shortcomings. Moreover, we will show how this can play an important role to increase performance substantially by utilizing CNN basic architecture (comparatively) in this field to handle one of these shortcomings. 75.2% accuracy of FET2013 test is achieved with an entity of modern deep CNN by us. It means even without face registration or auxiliary training data need; this eclipse former works Prudhvi and his team designed 8-layer CNN system for pose Emotion (JAFPE) dataset²⁰.

The Histogram of Oriented Gradient is a feature descriptor in image processing and object selection,

and it achieved amazing results on feature detection of pedestrian. Nevertheless, this approach takes significant amount of time. To overcome this issue, the updated and modified version of HOG is designed to reduce the features' dimension. Based on gradient orientation interval, nine individual HOG channels (HOG-C) are obtained for the purpose of analyzing HOG process. A combination of HOG channels (CHOG-C) aspect is devised based on statistical regularities to evaluate the performance of (HOG-C) for individual detection of pedestrian. CHOG-C feature shows significant performance on the thorough experimentation on database of INRIA. Without decreasing accuracy, dimension is reduced as the experimental outcome. (Ke Shan and junqi Guo) developed deep DL model for extraction of deeper feature in FER, and then compared results simple KNN approach shan li and his team build deep pipeline standard system (FER) developed for detecting pose various angles. (Thanh VO and trang Nguyen) developed two separate models; the first model is Race Recognition (RR) with CNN and second RR-VGG²¹. Local Binary Pattern (HOG-LBP) was proposed as a faster method for Histogram of Oriented Gradients. It is based on pedestrian recognizer. So, it had foundation on two stage back to back structure. Rather than to extract the features from all the area inside the recognizer window (same as old method), only best characterized features from the area were extracted for pedestrian as the first stage evaluation. Each candidate evaluation process is boosted up because of the reduction in features processes. Ada Boost classifier was trained to learn the selectivity of the blocks who's Support Vector Machines responded the pedestrian samples differently than non-pedestrians. This was done to identify which area was best to characterize the pedestrian. Conventional HOG-LBP classifier was utilized as a second stage to re-process the candidates which were gone through the first stage processes. HOG-LBP SVM algorithm was founded almost three times slower than the detection algorithm as the experimental results suggest. For efficient video saliency framework, oriented gradients (HOG) implementation by a relatively low-complex histogram was proposed. For video saliency framework's, an algorithm and optimized HOG flow was presented to show important computational bottleneck elimination for visual understanding pipes which were closely related to original HOG before. Optimized HOG flow and algorithm was able to decrease computational demands without compromise on algorithm

performance. Moreover, to improve system efficiency, optimal memory usage was explained with simplified light-weight computational requirements and scanning of reusable data. It was proved on analysis and testing, that the proposed HOG implementation, optimized not only processing performance and complexity but also maintained capability on the video saliency algorithm focused RDM, DBN and SAE+SM learns the representation of pixels and predicts the emotion²². In the area of human-computer interaction, gesture recognition technology has remarkable importance, visual sensitivity of the gesture recognition technology has impact on the experimental environment lightning and in the result, a great change will be witnessed in recognition. Because of this, these are the most challenging topics related to this technology. To maintain a better invariance on optical and geometric deformation, a successfully applied HOG feature was operating on the local grid unit of image to detect pedestrian detection. Gradient direction histogram (HOG) features of gestures were extracted, after this; these features were trained by utilizing Support Vector Machines. Gesture recognition rate in various illumination conditions was compared after decision was taken utilizing the previously learned SVMS at testing time. It was concluded experimentally that higher recognition rate was achievable by using multivariate SVM classification and HOG feature extraction methods. For illumination, these methods had better robustness. Byoung chul KO used hybrid DL methodology merge CNN with spatial features of separate frames and Long Short-Term Memory (LSTM) for temporal features.

2. Proposed Methodology

2.1 Dataset

The FER2013 dataset was designed during the ICML challenges in representation learning in 2013. The dataset large scale free database collected with the help of Google image API. And it consists of 35000 thousand samples of grayscale images, the 28,709 thousand samples for training images, 3,589 validation and 3,589 testing. The dimension of dataset images consists of 48 x 48 pixel and faces are dynamically registered so that face shapes are adjusted centered and occupies same amount of space. The Emotion task is to classify each face based on the stored emotion shown in facial expression seven categories. The

labels of categories are (0-6) where 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). Figure 1 shows 48 x 48 images frequency of 2304 pixels.



Figure 1. Random images illustrate facial expressions from the FER2013.

2.2 Convolutional Neural Networks

The Convolutional neural network was introduced. Having various types' information for processing data, the main functionality of CNN creates complex structures for representation of complex data. We designed FED system and the Architecture of Face Emotion Detection-CNN model is shown Figure 2. The input of model is 48 x 48 grayscale images with HOG features and output shows classification result in single class from (0-6) categories. There are three Convolutional layers from C1 to C3, Three MaxPooling layers from P1 to P3 and four leakyReLU activation layers from R1 to R4. Besides, all these layers are fully connected between input and output layers as shows in Figures 3, 4. The layer Convolutional C1 filters the input image 48 x 48 with learnable kernels of size 3 x 3 to give 32 metrics the size of 62 x 62. The results of previous layer passed through the leakyReLU activation layer which basically converts the metrics in small range when gradient is non-zero and unit is not active or disabled. The results of leakyReLU layer passed to P1 MaxPooling layer and 32 learnable kernels setup with 2 x 2. The result of MaxPooling P1 matrices of size 48 x 48. These results are sent to the next Convolutional layer C2 where again 64 learnable kernels are configured with the size of the 3x3 and get 64 matrices of the size of the 24x24. Then R2 setup with alpha 0.1 and sent R2 results to MaxPooling

layer P2 with 64 automatic tunable kernels with the size of 2x2, continuous process and get the compiled results 64 metrics of size 24x24. Next, C3 and P3 are continuous with 128 automatic learnable kernels where size of kernels 3x3 and 2x2. Finally, they sent to the flatten layer which gives 4608 values, three fully connected layers from, first layer with 2304 hidden units and other with 1152, last, the output layers applies seven classes.

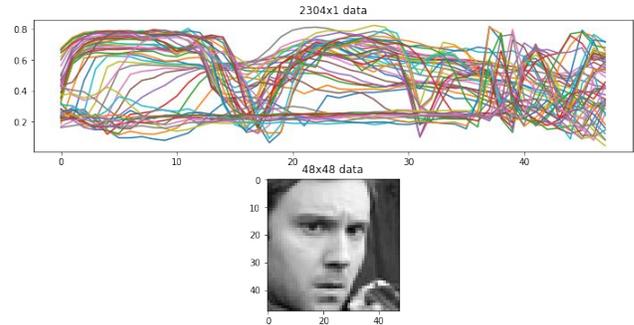


Figure 2. Image Frequency of 48 x 48 pixel of the FER2013.

Layer (type)	Output Shape
conv2d_1 (Conv2D)	(None, 48, 48, 32)
leaky_re_lu_1 (LeakyReLU)	(None, 48, 48, 32)
max_pooling2d_1 (MaxPooling2)	(None, 24, 24, 32)
conv2d_2 (Conv2D)	(None, 24, 24, 64)
leaky_re_lu_2 (LeakyReLU)	(None, 24, 24, 64)
max_pooling2d_2 (MaxPooling2)	(None, 12, 12, 64)
conv2d_3 (Conv2D)	(None, 12, 12, 128)
leaky_re_lu_3 (LeakyReLU)	(None, 12, 12, 128)
max_pooling2d_3 (MaxPooling2)	(None, 6, 6, 128)
flatten_1 (Flatten)	(None, 4608)
dense_1 (Dense)	(None, 128)
leaky_re_lu_4 (LeakyReLU)	(None, 128)
dense_2 (Dense)	(None, 7)

Figure 3. HOG + CNN model configuration.

2.3 Feature Extraction with HOG

Computer vision has feature selection, a most probable key point of recognition process. To diminish the dimensionality curse and at the same time to attain high recognition rate, there is dire need to select the most

informative and relevant feature. There are many available features to address general recognition's problem. If lighting variations are considered, color-based cues are found to be less robust than edges and gradients. Because of the object detection, tremendous attention has been given to HOG descriptors recently. Description of the shape of facial expression detection and local appearance are considerably described by HOG descriptors. A histogram of gradient directions (edge orientations) of targeted image is computed by dividing it into small connected regions (cells) to implement these descriptors. The descriptor is represented by combination of these histograms. The algorithm is divided into multiple stages. Gradient values are computed as the first stage. Cell histograms are created as second stage. Based on the values found in the gradient computation for an orientation-based histogram channel, each pixel within the cell casts a weighted vote. The components of the normalized histograms are concatenated by the HOG descriptor as vector from all the block regions. There are several parameters must consider: the number of orientations bins, the size of the cell in pixels and the number of cells per block. A very large feature vector is resulted by utilizing implementation of HOG blocks over whole image. So important question is raised in selection of most informative HOG blocks²³. Extraction method of HOG feature produces a lot of data. Edge directions or distribution of intensity gradients are used to determine complex shapes of structures as shown in Figure 5. Descriptor is produced by concatenation of generated pixel-wise histograms of gradient directions of HOG. HOG features are comprised of gradients angles and distribution of magnitude. Because of this fact, they are naturally adaptable to variation in color and lighting fluctuations. This fact witnesses their robustness in visual data. Firstly, all pixels' gradients are processed. Gradient estimation filters for an image I as

And Where $*$ represents convolution. Each pixel's gradient value can be computed as:

Estimation of dominant gradient direction can be computed at each pixel as

Cell histograms are created after this. Histogram channel is created as a weighted vote is casted by each point in a cell. Earlier computed gradient values have basis of these votes. Distribution of this orientation-based channel has span from 0° to 180° . These cells are categorized into normalized locally and spatially connected larger blocks to provide invariance in fluctuations in contrast and illumination. HOG feature vector is produced by concatenation of these normalized cell histograms

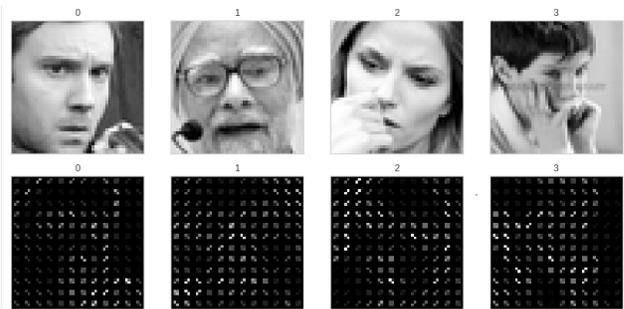


Figure 5. Hog features on random images.

3. Results and Discussion

Emotion Repression Recognition (FER) models tested through the FER2013 database and results are evaluated with different evaluation metrics, namely, Confusion matrix, Precision, Recall, Support, Micro avg and weighted avg, F1 Score and in last, training and testing accuracies. Precision shows the positive predictive value, and recall captures the sensitivity and true positive rate of the model. To compute the overall precision and recall, we used micro-averages to combine the results across the seven categories. Figures 6 and 7, classification report of precision (P) and recall (R) rate of overall classes, (Figure 6), The P and R of label happy P (81), R (89) which is best, and class fear P(65) and sad R(54) are low, (Figure 7), The P and R of classes disgust P (93), R (71) and happy P (84),R (83) which comparatively good. We split dataset 70%, 15%, 15% split for the training, validation and testing. To further understand and asses the models, we examined the metrics for each emotion as well as confusion matrix. See Figures 8 and 9. FER-CNN Confusion matrix of seven class's facial expression recognition results obtained by FER2013 database. The results of con-

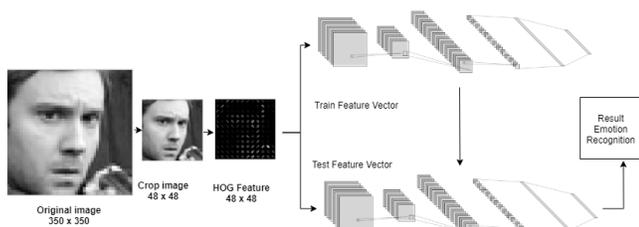


Figure 4. System architecture of Mo.

And ,Suppose gradient matrices and are computed as

fusion showing the prediction of seven classes. Where, (527) true labels are predicted of happy class, which is comparatively, better than other classes. Figures 10 and 11 shows the behavior of model (1) and (2) with accuracy and loss during training and testing on 50 epochs and 32 batch size parameters. Furthermore, the curves shows, the accuracies of the models are increasing with respect of epochs. Similarly, losses are displayed during the learning are decreasing in each epoch.

	precision	recall	f1-score	support
0	0.71	0.69	0.70	386
1	0.93	0.71	0.81	38
2	0.62	0.70	0.66	406
3	0.84	0.83	0.83	640
4	0.64	0.59	0.61	448
5	0.84	0.81	0.82	325
6	0.65	0.68	0.67	451
micro avg	0.72	0.72	0.72	2694
macro avg	0.75	0.72	0.73	2694
weighted avg	0.72	0.72	0.72	2694

Figure 6. Shows precision, recall, F1-score and support of FER-CNN.

0	0.67	0.67	0.67	386
1	0.71	0.63	0.67	38
2	0.64	0.69	0.67	406
3	0.81	0.89	0.85	640
4	0.68	0.54	0.60	448
5	0.82	0.84	0.83	325
6	0.73	0.71	0.72	451
accuracy			0.73	2694
macro avg	0.72	0.71	0.71	2694
weighted avg	0.73	0.73	0.73	2694

Figure 7. Shows precision, recall, F1-score and support of FER-HOGCNN.

Table 1, summarized the results of both approaches on FER2013 dataset. Figures 10 and 11 the training Accuracy of FER-CNN model is 98% and validation accuracy is 72%, similarly, training Loss 0.05 and validation loss are 2.01. Figures 12 and 13, shows training and testing accuracy 97% and 70% respectively on FER-HOGCNN. In the same way, training and validation losses are 0.07 and 2.02 respectively. Due to the small size of images and less sharp picture quality, the results of HOG feature descriptor are not better than basic model. Comparative results are less efficient as shown in Table 1. In case of noisy/unclear images of dataset then it is recommended to use simple convolutional neural network.

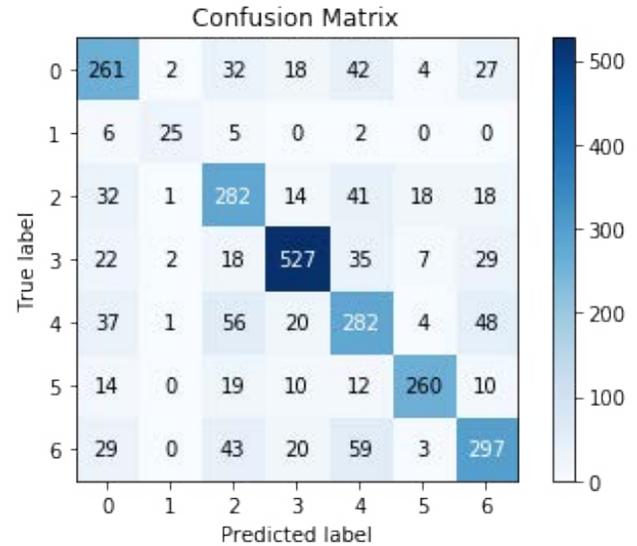


Figure 8. FER-CNN model confusion matrix of 7-class facial expression recognition results obtained by FER2013 database.

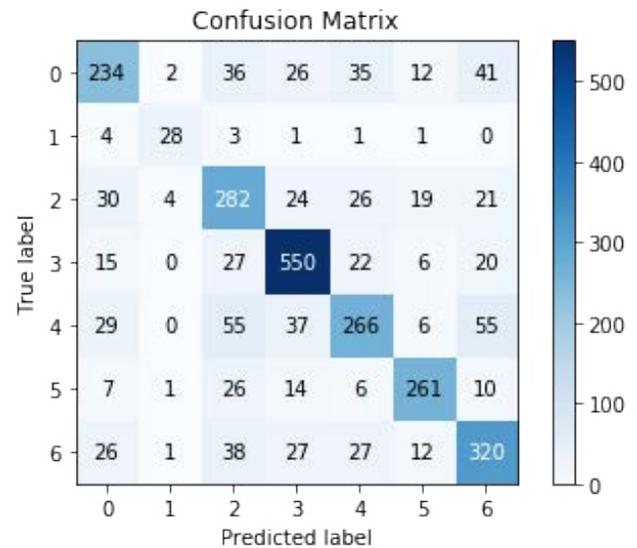


Figure 9. CNN + HOG model confusion matrix of 7-class facial expression recognition results obtained by FER2013 database.

Table 2 summarizes the results, FER-CNN and FER-HOGCNN with different configurations. The epochs are (15, 20, 25, 30, 35, 40, 45, and 50) and batch sizes (8, 16, 32, 64, 128, 256, 512, and 1024). Various combinations of epochs and batch sizes depict the different results. It shows the performance comparison with different Hyper-parameters. The significant performance is witnessed in

both models at batch 32 and 64. In batch 32, FER-CNN accuracy is 98%, while validation accuracy is 97%.

Table 1. Accuracy and loss FER-CNN and FER-HOGCNN

Model	Acc	Loss	Val_Acc	Val_Loss
FER-CNN	98%	0.05	72%	2.01
FER-HOGCNN	97%	0.07	70%	2.02

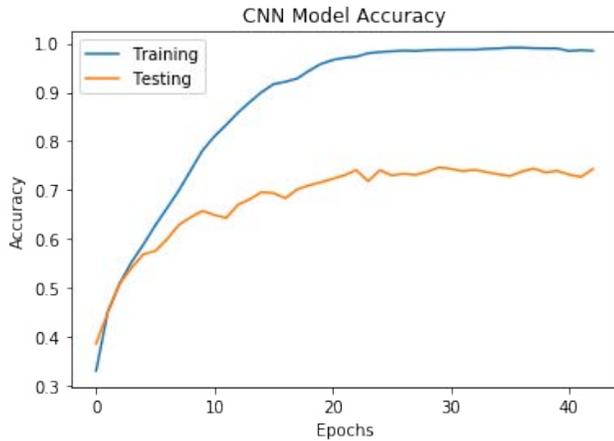


Figure 10. Training and testing accuracy FER-CNN.

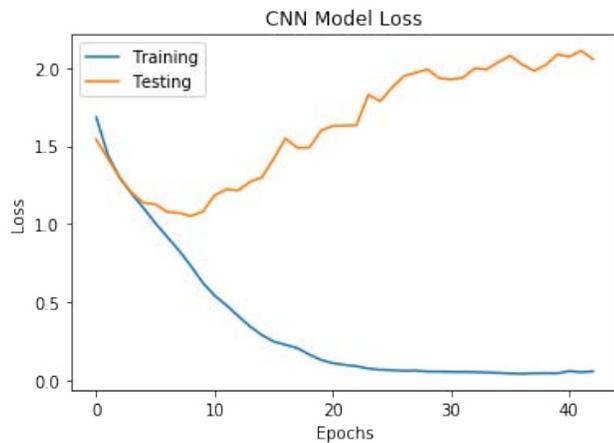


Figure 11. Training and testing loss of FER-CNN.

Table 2. Results with various hyper-parameters

Epochs	Batch Size	FER-CNN Acc	FER-HOGCNN Acc	FER-CNN Loss	FER-HOG CNN Loss
15	8	95%	91%	0.08	2.04
20	16	97%	93%	0.08	2.04
25	32	98%	97%	0.05	2.02

30	64	98%	97%	0.05	2.02
35	128	96%	93%	0.06	2.03
40	256	96%	91%	0.07	2.04
45	512	96%	92%	0.07	2.05
50	1024	96%	93%	0.07	2.05

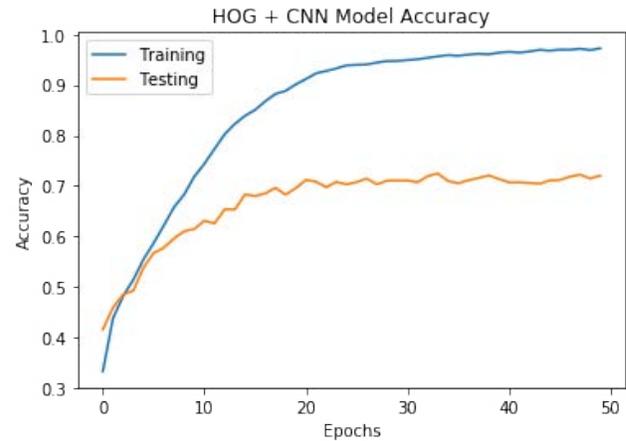


Figure 12. Training and testing accuracy of FER-HOGCNN.

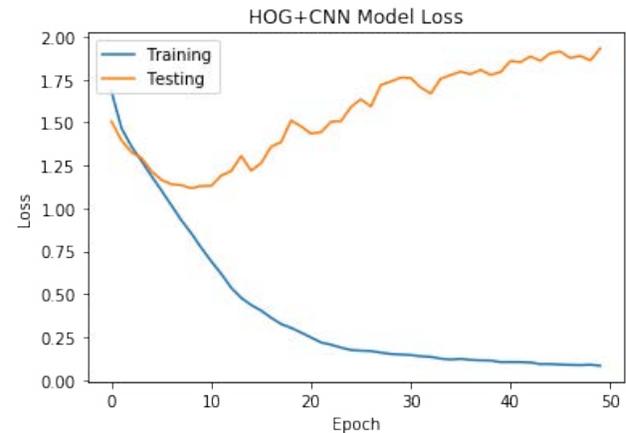


Figure 13. Training and testing accuracy of FER-HOGCNN.

4. Conclusion

In this study, a new deep neural network architecture is introduced to recognize human based facial expression using HOG with CNN. In deep learning field the model CNN has one of the most presentable network architectures for image processing and FER system, to find more precise and efficient result, various techniques have been used in identifying face expressions. In our model, the feature extraction is the first step. HOG describes the important features of an image in a gray scale range, and in

this paper, FER2013 Dataset is used and it splits training, testing and validation samples. There are 28,709 training samples and 3,589 samples for testing and validation. We trained two models, FER-CNN and FER-HOGCNN. The second model utilizing the Feature Extraction operator HOG, so the initially, the 48x48 image pass from feature extraction operator, then output of operator maps the input of CNN network, classifier to classify input face expressions. We conducted comprehensive experiment using Facial Expression database and 97% accuracy for training, 70% for testing and losses 0.05 and 2.01 respectively. It was found that when image size is small and picture quality is unclear then results of HOG operator are not effective than FER-CNN model.

5. Future Work

In future, this work shall be extended, and order to extract the important features from images by combining Local Binary Pattern (LBP) and HOG operator using Deep Learning models.

6. References

- Lopes AT, Aguiar ED, Oliveira-Santos T. A facial expression recognition system using convolutional networks. 28th SIBGRAPI Conference on Graphics, Patterns and Images; 2015. <https://doi.org/10.1109/SIBGRAPI.2015.14>
- Zhao X, Zhang S. Facial expression recognition based on local binary patterns and kernel discriminant isomap. *Sensor*. 2011; 11(10):9573-88. <https://doi.org/10.3390/s111009573> PMID:22163713 PMID:PMC3231257
- Liu P, Han S, Meng Z, Tong Y. Facial expression recognition via a boosted deep belief network. In 2014 IEEE Conference on Computer Vision and Pattern Recognition; 2014. p. 1805-2. <https://doi.org/10.1109/CVPR.2014.233>
- Shan C, Gong S, McOwan PW. Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing*. 2009; 27(6):803-16. <https://doi.org/10.1016/j.imavis.2008.08.005>
- Suleiman A. Towards closing the energy gap between HOG and CNN features for embedded vision. IEEE International Symposium on Circuits and Systems; 2017. <https://doi.org/10.1109/ISCAS.2017.8050341> PMID:28217180 PMID:PMC5301302
- Islam KTo, Raj RG, Al-Murad A. Performance of SVM, CNN, and ANN with BOW, HOG, and image pixels in face recognition. 2nd International Conference on Electrical and Electronic Engineering; 2017. <https://doi.org/10.1109/CEEE.2017.8412925>
- Pramerdorfer C, Kampel M. Facial expression recognition using convolutional neural networks: state of the art. arXiv preprint arXiv:1612.02903; 2016.
- Weixing L. A fast pedestrian detection via modified HOG feature. 2015 34th Chinese Control Conference; 2015. <https://doi.org/10.1109/ChiCC.2015.7260236>
- Shan K. Automatic facial expression recognition based on a deep convolutional-neural-network structure. IEEE 15th International Conference on Software Engineering Research, Management and Applications; 2017. <https://doi.org/10.1109/SERA.2017.7965717> PMID:28163057
- Li S, Deng W. Deep facial expression recognition: A survey. arXiv preprint arXiv:1804.08348; 2018.
- Park W. Fast human detection using selective block-based HOG-LBP. 19th IEEE International Conference on Image Processing; 2012. <https://doi.org/10.1109/ICIP.2012.6466931> PMID:PMC3572427
- Lee T. Low-complexity HOG for efficient video saliency. IEEE International Conference on Image Processing; 2015. <https://doi.org/10.1109/ICIP.2015.7351505>
- Dhall A, Goecke R, Lucey S, Gedeon T. Collecting large, richly annotated facial-expression databases from movies. *IEEE Multimedia*. 2012; 19(3):34-41. <https://doi.org/10.1109/MMUL.2012.26>
- Ghayoumi M. A quick review of deep learning in facial expression. *Journal of Communication and Computer*. 2017; 14(1):34-8. <https://doi.org/10.17265/1548-7709/2017.01.004>
- Feng K, Yuan F. Static hand gesture recognition based on HOG characters and support vector machines. 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation; 2013. <https://doi.org/10.1109/IMSNA.2013.6743432>
- Ko B. A brief review of facial emotion recognition based on visual information. *Sensors*. 2018; 18(2):401. <https://doi.org/10.3390/s18020401> PMID:29385749 PMID:PMC5856145
- Byeon YH, Kwak KC. Facial expression recognition using 3d convolutional neural network. *International Journal of Advanced Computer Science and Applications*. 2014; 5(12). <https://doi.org/10.14569/IJACSA.2014.051215>
- Mahoor M, Behzad H. Facial expression recognition using enhanced deep 3D convolutional neural networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*; 2017.
- Mollahosseini A, Chan D, Mahoor MH. Going deeper in facial expression recognition using deep neural networks. *IEEE Winter Conference on Applications of Computer Vision*; 2016. <https://doi.org/10.1109/WACV.2016.7477450>
- Dachapally PR. Facial emotion detection using convolutional neural networks and representational autoencoder units. arXiv preprint arXiv:1706.01509; 2017.

21. Vo T, Nguyen T, Le C. Race recognition using deep convolutional neural networks. *Symmetry*. 2018; 10(11):564. <https://doi.org/10.3390/sym10110564>
22. Das D, Chakrabarty A. Emotion recognition from face dataset using deep neural nets. *International Symposium on Innovations in Intelligent Systems and Applications*; 2016. <https://doi.org/10.1109/INISTA.2016.7571861>
23. Nandi D. Traffic sign detection based on color segmentation of obscure image candidates: a comprehensive study. *International Journal of Modern Education and Computer Science*. 2018; 10(6):35. <https://doi.org/10.5815/ijmecs.2018.06.05>
24. Orrite C, Ganan A, Rogez G. Hog-based decision tree for facial expression classification. *Iberian Conference on Pattern Recognition and Image Analysis*. Berlin, Heidelberg: Springer; 2009. p. 176-83. https://doi.org/10.1007/978-3-642-02172-5_24