FACIAL EMOTION DETECTION THROUGH DEEP COVOLUTIONAL NEURAL NETWORKS

AYMUN SAIF DAR¹, SHERAZ NASEER¹, AIHTSHAM ALI¹, ISHMAL RAUF¹, MUHAMMAD AHSAN¹

School of Systems and Technology, University of Management and Technology, Lahore, Pakistan
 sheraz.naseer@umt.edu.pk

ABSTRACT. *Our society has evolved to a threshold where use of machines to automate mundane tasks is constantly increasing in daily life. Providing machines with capability to develop perception from their environment can lead them to perform a great variety of tasks. Facial emotion detection is crucial sub-part of machine perception development. In this article we present a deep learning based approach for Facial emotion Detection. Our model uses Convolutional Neural Network (CNN) to learn deep features for classification of facial images into one of 22 emotion (Basic 7 + Compound 15) categories considered in this study. We trained our CNN model with the images dataset from Martinez et al. Our Facial Emotion Detection model was developed using keras with theano backend and implemented on a GPU-powered testbed. Our model achieved 67.6% accuracy for basic emotions and 33% accuracy for compound emotions.*

1. Introduction: In information age, Computing is expected to help in every walk of life whether it is technical or social. For greater social acceptance of machines, the interaction between human and machine has to become smoother and more natural as machine's exposure with humans increase. To achieve this capacity, machines must be able to learn from their surroundings and especially from human interactions. When term machine is used it comprises of computers and robots. Humans use their senses to learn from environment therefore machine perception tries to mimic the human senses. One of the important senses which need to be mimicked is vision whereby machines can process visual stimulus to learn from their surroundings. A granular discipline of machine vision is facial emotion detection. Emotion detection is process of making machines capable of recognizing the mood and future perception of the individual from visual stimulus. The human face has several emotions in which 7 are basic emotions and 15 are compound emotions. The basic emotions are sadness, fear, surprise, anger, and disgust, happiness and the last one is the neutral and the compound emotions are those which are made up of the combination of the basic emotions. It is cleared by the example that a person see something unexpected then he gets surprised and when he faces some scared environment or some unpredictable than he get fearful so when these both things happen together than a new emotion comes into existence that is fearfully surprised this emotion is called the compound emotion because it is made up of the two basic emotions. Emotion detection and recognition is an emerging area of machine vision for last few years. Nowadays machines can capture visual stimulus in the form of pictures and videos using cameras. Hence this information can be used with machine learning algorithms to generate effective machine perception and computer vision. Effective computing requires that emotion detection is necessary for machines to better serve their purpose. For example use of the robots for house-hold, hospitals or elderly care requires emotional understanding of humans. Facial information delivers information about innerstate of a human. Capturing this information in form of images and applying deep learning algorithms

Deep Learning is an area of Machine Learning which applies neuron like mathematical structures [1] for learning tasks. Neural Networks have been around for many decades [2] and have been gaining and losing the favor of research community. The latest rise of this technology is attributed to Alexnet [3] a Deep Neural Network, which won the ImageNet classification challenge. Alexnet achieved top-1 and top-5 error rates of 37.5% and 17.0% on ImageNet [4] Dataset which were considerably better than the previous state-of-the-art mechanisms. Since then, Deep Neural Networks (DNNs) have attracted the attention of research community once again and multiple DNN structures including Convolutional neural networks (CNNs) [5] Recurrent Neural networks (LSTM) [6], Deep belief

nets (DBNs) and different types of Autoencoders have been proposed. These DNN structures have been successfully applied to devise state of art solutions in multiple disciplines.

In this paper we developed and implemented a deep CNN model to address the problem of emotion detection. Fine tuning and hyper-parameter optimization of proposed DCNN was performed using RandomizedSearch [7] over configuration space. Classification metrics including confusion matrix, accuracy, precision, recall and F1 scores were used to assess the performance of model.

I. Table 1 Categories of Basic and Compound Emotions available in selected Dataset

Basic Emotions	Compound Emotions	
Neutral	Happily surprised	Fearfully disgusted
Happy	Happily disgusted	Angrily surprised
Sad	Sadly fearful	Angrily disgusted
Fearful	Sadly angry	Disgustedly surprised
Angry	Sadly surprised	Appalled
Surprised	Sadly disgusted	Awed
Disgusted	Fearfully angry	Hatred
	Fearfully surprised	

For this study, we used dataset devised by Martinez et al [8] Dataset consists of 5060 images of 1000x750 pixels. These are colored images, which contains 07 basic emotions and 15 compound emotions as shown in table.

METHODOLOGY:

Preprocessing: Complete Dataset was processed in grayscale and resized in images of 64x64 pixels. Raw pixel data was read in numpy arrays to allow further processing by model. After reading the raw pixel data, we normalized them by subtracting mean of the images from each image. For both Basic emotions and compound emotions, Data was divided into training and testing partitions. In order to classify the expressions, mainly we used the features generated by convolution layers using the raw pixel data give them as input features into Fully Connected (FC) layers.

Model Architecture: Proposed IDS approach uses a DCNN with an input layer, 3 pairs of conv-subsample-dropout layers, 2 fully connected layers and an output layer with 7 unit and softmax non-linearity. The input plane receives preprocessed training dataset images in the form of 64x64 greyscale images. With local receptive fields, earlier layer neurons can extract elementary features which are combined by subsequent CNN layers to form higher-order features. These higher order features are passed to fully connected multi-level perceptron layers for classification. Each layer consists of trainable parameters and nodes as described in Table x. We use LeNet-5 nomenclature to name layers of DCNN for description purpose where convolution layers are labeled as Cx, subsampling layers as Sx, dropout layers as Dx and Fully connected layers FCx.

Experimental Setup: Hardware setup used for implementing proposed model included:

- CPU: Intel Core I5 E-1650 Quad Core
- RAM: 16 GB
- GPU: nVidia Geforce 940M with 684 CUDA cores and cuda 8.0

Software toolchain used to implement the model consist of jupyter development environment using Keras 2.0 on Theano [9] backend and nVidia cuda 8.0 [10] Training and testing data is manipulated in form of numpy arrays.

Model for Basic Emotions: Model for Basic emotions classification was built using CNN defined in table 2. We had 10 3×3 filters, with the stride of size 1, along with dropout and Maxpooling and Rectified Linear Unit (ReLU) as the activation function in the first convolutional layer while in second convolution layer we had 10 5×5 filters, with the stride of size 1, along max-pooling with a filter size of 2×2 and Rectified Linear Unit (ReLU) as the activation function. Third and fourth convolution layers used respectively 10 and 64 feature maps along with kernel size of 3, with the stride of size 1, max-pooling with a filter size 2×2 and Rectified Linear Unit (ReLU) as the activation function. In the FC layer, we placed a hidden layer with 512 neurons and Rectified Linear Unit (ReLU) as the activation function and Softmax as the loss function. Model was trained using training Dataset of Basic Emotions with 60 epochs and a batch size of 32 and cross-validated using hyper-parameters of the model with different values for regularization and the number of hidden neurons. To validate our model in each iteration, we used the validation set and to evaluate the performance of the model, we used the test set. We achieved 67% accuracy from this DCNN, its and great progress of 22%.

Table 2: Architecture of Convolutional Neural Network used to develop Basic and Compound Emotion Detection Models

Layer Name : Function	Output Shape	Trainable (Weights) Parameters
C1: Convolution layer with 3x3 kernels and 10 feature maps	10, (64,64)	$((3*3) + 1)*10 = 100$
S1: Subsampling layer with 2x2 non-overlapping kernel	10, (32,32)	0 (Sub-sampling)
D1: Dropout layer with 0.25 drop probability		Dropout Layer
C2: Convolution layer with 5x5 kernels and 64 feature maps	10, (28,28)	$((5*5)*10)+1)*10 = 2510$
S2: Subsampling layer with 2x2 non-overlapping kernel	10, (14,14)	0 (Sub-sampling)
D2: Dropout layer with 0.25 drop probability		Dropout Layer
C3: Convolution layer with 3x3 kernels and 64 feature maps	10, (12,12)	$((3*3)*10)+1)*10 = 910$
S3: Subsampling layer with 2x2 non-overlapping kernel	10, (6,6)	0 (Sub-sampling)
D2: Dropout layer with 0.25 drop probability		Dropout Layer
C4: Convolution layer with 3x3 kernels and 64 feature maps	64, (4,4)	$((3*3)*64)+1)*64 = 5824$
S3: Subsampling layer with 2x2 non-overlapping kernel	64, (2,2)	0 (Sub-sampling)
Model Flattening	256,1	Not Applicable
D1: Dropout layer with 0.5 drop probability		Dropout Layer
FC7: Fully connected layer	512 ,1	$(256 + 1) * 512 = 131584$
FC8: Fully connected layer	7 ,1	$(512 + 1) * 7 = 3591$
Output: Fully connected layer	1	$(7 + 1) * 1 = 8$

Analysis of Compound Emotions: Model for compound Emotions was built using a CNN of 4 convolutional layers of depths(10x7x7,32x5x5,64x3x3 & 128x3x3) along with Max pooling of 2x2 and dropout and Relu as activation function, we used 128 neurons for fully connected layer, the results we achieved are quite good as these are compound emotions and they take quite long time to train and greater the training samples,greater will be the results, the results we achieved from this analysis is 33 % .

Evaluations and Results: This section presents the results of implemented models. We used well-known classification quality metrics including Accuracy, Precision, Recall and F1 score to present the performance of models. These evaluation metrics are calculated using confusion matrix which presents four measures as follows:

- True Positive: if an anomaly is classified by model as an anomaly, result is accepted as TP
- False Positive: if a normal instance is classified by model as an anomaly, result is accepted as FP
- True Negative : if an anomaly is classified by model as normal instance, result is accepted as TN
- False Negative: if a normal instance is classified by model as normal instance, result is accepted as FN

Accuracy is defined as ratio of the number of correctly classified anomalous and normal instances to total number of all instances.

Figure 1: Accuracy Graph of Basic emotion Model

Figure 2 Loss Graph of Basic Emotions Model

We achieved 67.62% Accuracies for Basic Facial Emotions test Dataset and 33% accuracy for compound Facial Emotions test Dataset. This section provides various visualization regarding the performance of both models.

Table 3: Precision, Recall, F1-Score of Basic Emotions Model

	Precision	Recall	F1-Score	Support
Neutral(0)	0.60	0.80	0.69	15
Happy(1)	0.75	1.00	0.86	12
Sad (2)	0.75	0.20	0.20	15
Fearful(3)	0.75	0.33	0.33	18
Angry(4)	0.62	0.73	0.73	11
Surprised(5)	0.75	0.90	0.90	20
Disgusted(6)	0.60	0.86	0.86	14
Avg / Total	0.69	0.68	0.64	105

Figure 3: Confusion Matrix Heatmap Visualization of Basic Emotions Model

Fig 1 shows the progression of training and validation accuracy of Basic Emotions Model with respect to training epochs. Fig 2 shows the mitigation of training and validation loss with respect to training epochs.

Figure 4 Accuracy Graph of Compound emotions Model

Figure 5: Training and Validation Loss Graph of Compound Emotions Model

Table 3 shows the Precision, Recall and F1 score for all seven emotion classes in basic emotions. It can be seen from the table that Basic Emotions Model predicted 4 emotions i.e happy, sad, fearful and surprised with maximal precision. For Recall the best score was achieved for basic emotion “happy”. Fig 3 shows the heatmap visualization of confusion matrix for basic emotions model.

Table 4: Precision, Recall, F1-Score of Compound Emotions Model

	Precision	Recall	F1-Score	Support
HS(0)	0.58	0.81	0.68	47
HD(1)	0.42	0.73	0.53	37
SF(2)	0.13	0.18	0.15	38
SA(3)	0.15	0.17	0.16	40
SS(4)	0.38	0.31	0.34	45
SD(5)	0.18	0.19	0.18	43

FA(6)	0.29	0.29	0.24	45
FS(7)	0.20	0.25	0.22	52
FD(8)	0.41	0.24	0.30	62
AS(9)	0.63	0.47	0.54	55
AD(10)	0.30	0.21	0.25	43
DS(11)	0.49	0.70	0.57	50
APPALLED(12)	0.15	0.12	0.13	43
AWED(13)	0.19	0.13	0.16	38
HATRED(14)	0.35	0.31	0.33	52
Avg / Total	0.34	0.34	0.33	690

Fig 4 shows the progression of training and validation accuracy of Compound Emotions Model with respect to training epochs. Fig 5 shows the training and validation loss with respect to training epochs.

Figure 6: Confusion Matrix Heatmap Visualization of Basic Emotions Model

Table 4 shows the Precision, Recall and F1 score for all fifteen emotion classes in compound emotions. It can be seen from the table that Compound Emotions Model predicted disgustedly surprised with maximal precision followed by Fearfully Disgusted. The best recall score was for Happily surprised and happily disgusted. Best F1-measure also turned out to be for Happily surprised Fig 3 shows the heatmap visualization of confusion matrix for compound emotions model.

Conclusions: Emotion detection is helpful for many applications like driver monitoring, surveillance systems, pain detection as well as in other medical fields and suspicious person detection. In this article we used Deep convolution neural networks for the problem of Human Facial Emotion Detection. Stack of Convolution layers of the models extracted higher-order features which were used by fully connected portion of CNN to perform classification using Softmax classifier. The Deep features extracted using convolution layers can be used in different conventional machine learning algorithms for solving problems like surveillance system, Safety driving, suspicious person Detection and pain Detector. By taking large date set of images we can enhance the percentage accuracy of basic and compound emotions.

REFERENCES

- [1] Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). Deep learning (Vol. 1). Cambridge: MIT press.
- [2] Minsky, M., & Papert, S. (1969). Perceptron Expanded Edition.

- [3] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- [4] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Berg, A. C. (2015). Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3), 211-252.
- [5] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324..
- [6] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.
- [7] Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb), 281-305.
- [8] Du, S., & Martinez, A. M. (2015). Compound facial expressions of emotion: from basic research to clinical applications. *Dialogues in clinical neuroscience*, 17(4), 443.
- [9] Team, T. T. D., Al-Rfou, R., Alain, G., Almahairi, A., Angermueller, C., Bahdanau, D., ... & Belopolsky, A. (2016). Theano: A Python framework for fast computation of mathematical expressions. *arXiv preprint. arXiv preprint arXiv:1605.02688*.
- [10] Nickolls, J., Buck, I., Garland, M., & Skadron, K. (2008, August). Scalable parallel programming with CUDA. In *ACM SIGGRAPH 2008 classes* (p. 16). ACM.
- [11] Hoo-Chang, S., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., ... & Summers, R. M. (2016). Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5), 1285.
- [12] Toshev, A., & Szegedy, C. (2014). Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*(pp. 1653-1660).