Real Time Human Emotions Recognition using Auto-Multiplexers based Deep Convolutional Neural Networks

T. Muni Reddy[1], Dr. R.P. Singh[2]
[1]Research Scholar (SSSEC1525), ECE Dept., SSSUTMS, Bhopal, (India)
[2]Vice-Chancellor & Professor, ECE Dept., SSSUTMS, Bhopal, (India).

Abstract

Now a day's one of the unsolved problem in computer vision is recognizing or understanding other people's emotions and feelings. Deep Convolutional Neural Networks (CNN) has tried to be economical in feeling recognition issues. The good degree of performance achieved by these classifiers can be attributed to their ability to self-learn a down-sampled feature vector that retains abstraction info through filter kernels in convolutional layers. In this paper we have a tendency to explore the impact of coaching the initial weights in associate unsupervised manner. we have a tendency to study the result of pre-training a Deep CNN as a Convolutional Auto-Multiplexer (CAM) in a very greedy layer-wise unsupervised fashion for emotion recognition mistreatment facial features pictures. once trained with at random initialized weights, our CNN feeling recognition model achieves a performance rate of 92.16% on the Karolinska Directed Emotional Faces (KDEF) dataset. In distinction, By using this pre-trained, the performance will increase to 93.52%. Pre-training our CNN as a CAM conjointly reduces coaching time marginally.

Keywords - Emotion; Face Expression; Convolutional Auto-Multiplex(CAM); MAM Pooling, DCNN

I. INTRODUCTION

Emotion recognition sometimes involves analyzing a person's facial expressions, visual communication, or speech signals and classifying them as a selected feeling. it's been declared that feeling recognition is essential for everyday living and is essential for interaction with others [1], [2]. In the work mentioned during this paper, we have a tendency to compare the performance of a Deep Convolutional Network once trained with a haphazardly initialized set of weights and once pretrained as a Stacked Convolutional Auto-Encoder to classify facial expression pictures from the KDEF [3] dataset. Bengio [4] suggests that random initialization of a network can lead to convergence on local minima, and thus result in poor classification. To avoid this difficulty, In this paper employ Auto-Multiplexers and MAM Pooling to pre-train in unsupervised manner[5],[6]. which leads to improved feature extraction and classification performance.

The structure of this paper is as follows: Section II introduces existing state-of-the art emotion recognition approaches based on DL and some previous work on Auto-Encoders used as a pre-training method.

II. LITERATURE SURVEY/RELATED WORK

Due to the natural non-linearity of profound systems, observational preparing strategies, for example, Stochastic Gradient Decent (SGD) may fall flat if the parameters are not instated fittingly or on the other hand if the system topology isn't perfect. Loose arrange setups can prompt substantial or little angles and issues in acquiring an arrangement of weights that give ideal speculation of the preparation information. Where the topology or parameters of the system are not perfect, it regularly requires a long preparing process, especially for profound models. Irregular weight instatement is regularly the favored decision among analysts and is expected to give the system with a weight conveyance that does not support a specific class.

A. Weight Initialization Methods

Initializing a network with the right weights is one of the difficult one. Krahenbuhl et al. [5] introduced a data-dependent initialization method for CNNs. Mishkin and Matas [8] proposed an initialization method,which they refer to as layer-sequential unit-variance (LSUV).

B. Deep CNN Normalization

Rectified Linear Unit (ReLU) layers, along with MAM Pooling, have become essential additives of Convolutional Networks. Most, if no longer all, latest Deep CNN architectures use rectifier neurons to normalize the output of convolutional Layers. He et al. [9], in the form of the Parametric Rectified Linear Unit (PReLU); Maas et al. [10] who introduced leaky ReLU; and Xu et al. [11] who proposed the Randomized leaky ReLU. Munireddy et al.[12] who introduced MAM pooling.

One of the most recent improvements to deep networks is Batch Normalization (BN) which normalizes the distribution of each input feature at every layer [13].

C. Unsupervised Pre-Training

Restricted Boltzmann Machines (RBM) have frequently been used to pre-educate Deep perception [15] and CNN models [16].

Auto-Multiplexers are used for data dimensionality reduction, are trained in an unsupervised greedy layer-wise manner and learn to encode the input vector into a down-sampled representation of the input. In this paper, we follow Masci et al.[17] approach and use SCAM to pre-train a CNN for emotion recognition.

D. Emotion Recognition Using CNNs

Convolutional networks have an capacity to self-research important Functions vital for category even as preserving Spatial statistics. Burkert et al. [18] have devised an emotion recognition version,Which they discuss with as DeXpression.

III. METHODOLOGY AND EXPERIMENTAL DESIGN

In this paper we try to find the right balance between classification performance and prediction time. We develop a SCAM with reduced number of deep learning layers for emotion recognition and compare this with a conventional CNN for emotion recognition.

A. Facial Expression Corpus

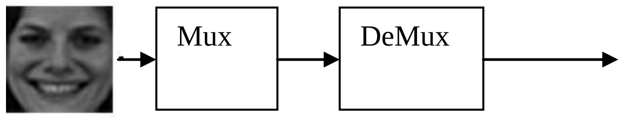
This work utilizes the Karolinska Directed Emotional Faces database (KDEF) [3] because of the high number of members it contains; 35 guys and 35 females, and mulling over that it was made to be especially appropriate for recognition, consideration, feeling, memory and in reverse covering tests [3]. Each participant illustrates the following emotional states: sad, surprised, neutral, happy, fear, disgust,and angry.

B. Convolutional Neural Networks with Batch Normalization

In this our CNN is composed of Convolutional, BN, ReLU, and MAM Pooling layers, except for the last block which does not have a MAMPooling layer. The first two convolutional layers use kernels of 5 × 5 and the last two convolutional layers use kernels of size
3 × 3 The last block is connected to a fully connected layer which is a Multilayer Perceptron (MLP) , also with BN and ReLU layers.

First Stage

(Reduce image reconstruction error.)



Second Stage

(1. Initialize CNN and MLP with Multiplexer weights.

2. Fine-tune CNN and MLP.)

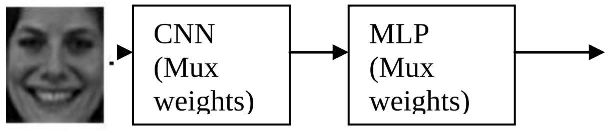


Fig. 2. Illustration of the SCAM architecture.

C. Stacked Auto-Multiplexers

While trying to enhance preparing time and order execution of our CNN feeling acknowledgment show, we chose to pre-prepare it as a SCAM. Basically, each convolutional layer and its ensuing layers: BN, ReLU, and Max Pooling, are dealt with as a solitary square and an Auto-Multiplexer is made for every single one of these squares shown in Fig.3.

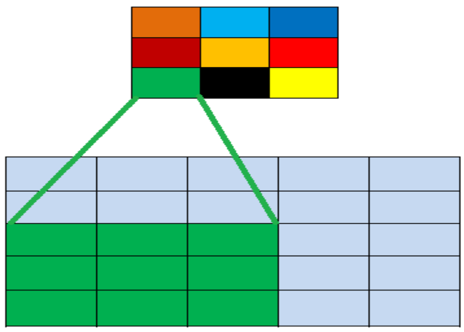


Fig.3. convolution between a 5x5x1 input and a 3x3x1 convolutional filter. The result is a 3x3x1 activation map. (Source)

a.Fear

b.Sad

Fig. 3. Sample output of first convolutional layer of the emotion recognition model pre-trained as a SCAEM and fine tuned as a CNN.

IV. RESULTS AND DISCUSSION

The CNN with BN and the SCAE emotion recognizers are trained and tested using the KDEF [3] dataset.**Table :1** illustrates the confusion matrix of this model when pre-trained as a SCAM.

Label	An	Di	Fe	Ha	Ne	Sa	Su	Total
An	39	1	0	0	1	1	0	91.49
Di	1	39	0	0	0	3	0	91.58
Fe	1	0	35	0	0	5	1	83.33
Ha	0	0	0	41	0	1	0	97.62
Ne	0	0	1	0	40	1	0	96.24
Sa	0	0	2	0	0	40	0	96.20
Su	0	1	0	0	0	1	40	96.46
								93.52

TABLE I: SCAM CONFUSION MATRIX: LEFT TO RIGHT; ANGRY, DISGUST, FEAR,HAPPY, NEUTRAL, SAD, SURPRISE. RIGHT MOST COLUMN DENOTES AVERAGE ACCURACY RATE PER CLASS AND TOTAL AVERAGE.

V. CONCLUSION

In this work we have proposed two CNN models: A CNN model that combines BN and fewer layers than an empirical CNN, and a SCAE that pre-trains the weights to the CNN element using Auto-Multiplexers we also plan to explore the effect of pre-training Auto-Encoders as a single unit rather than layer by layer.

REFERENCES

[1] M.Lewis, J. M. Haviland-Jones, and L. F. Barrett, “Handbook of Emotions.” Contemporary Sociology, vol. 24, no. 3, p. 298, may 1995.

[2] A.Chavhan, S. Chavan, S. Dahe, and S. Chibhade, “A Neural Network Approach for Real Time Emotion Recognition,”Ijarcce, vol. 4, no. 3, pp. 259–263, 2015.

[3] D.Lundqvist, A. Flykt, and A. O¨ hman, “The Karolinska Directed Emotional Faces - KDEF CD ROM from Department of Clinical Neuroscience, Psycology section,” Karolinska Institutet, pp. 3–5, 1998.

[4] Y.Bengio, “Learning Deep Architectures for AI,” Foundations and Trends R _ in Machine Learning, vol. 2, no. 1, pp. 1–127, 2009.

[5] T.Muni Reddy and R.P. Singh, “Deep Convolutional Auto-Multiplexers for Emotion Recognition from Facial Expressions”, Journal of VLSI Design and its Advancement, pp. 8-12, Vol. - 6, Issue-4, July, 2018.

[6] T.Muni Reddy and R.P. Singh, “Real Time Human Emotion Recognition Using MAM

[7] Pooling Based Convolutional Neural Networks”, Journal of VLSI Design and its Advancement, pp. 17-21Volume 1 Issue 2, 2018.

[8] P.Kr¨ahenb¨uhl, C. Doersch, J. Donahue, and T. Darrell, “Data-dependent Initializations of Convolutional Neural Networks,” International Conference on Computer Vision, pp. 1–12, nov 2015

[9] D.Mishkin and J. Matas, “All you need is a good init,” Computers & Mathematics with Applications, vol. 31, no. 11, p. 135, nov 2015.

[10] K.He, X. Zhang, S. Ren, and J. Sun, “Delving Deep into Rectifiers:Surpassing Human-Level Performance on ImageNet Classification,” feb 2015. [Online].

[11] A.L.Maas, A. Y. Hannun, and A. Y. Ng, “Rectifier Nonlinearities Improve Neural Network Acoustic Models,” Proceedings of the 30 th International Conference on Machine Learning, vol. 28, p. 6, 2013.

[12] B.Xu, N. Wang, T. Chen, and M. Li, “Empirical Evaluation of Rectified Activations in Convolutional Network,” ICML Deep Learning Workshop,pp. 1–5, may 2015.

[13] T.Muni Reddy and R.P Sing,” Real Time Human Emotion Recognition Using Artificial Neural Networks” ijarcsse ,Vol.8, Issue.4,April 2018.

[14] S.Ioffe and C. Szegedy, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,” feb 2015.

[15] G.E.Hinton, S. Osindero, and Y.-W. Teh, “A fast learning algorithm for deep belief nets.” Neural computation, vol. 18, no. 7, pp. 1527–54, jul 2006.

[16] O.Abdel-Hamid, L. Deng, and D. Yu, “Exploring convolutional neural network structures and optimization techniques for speech recognition,”in INTERSPEECH, 2013.

[17] J.Masci, U. Meier, D. Cirean, and J. Schmidhuber, “Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction,”

[18] in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics),2011, vol. 6791 LNCS, no. PART 1, pp. 52–59.

[19] P.Burkert, F. Trier, M. Z. Afzal, A. Dengel, and M. Liwicki,“DeXpression: Deep convolutional neural network for expression recognition,” arXiv preprint, pp. 1–8, 2015.