Deep Convolutional Neural Networks for Facial Expression Recognition

Ayşegül Uçar
Department of Mechatronic Engineering
Fırat University
Elazığ, Turkey
agulucar@firat.edu.tr

***Abstract*—Facial expression recognition is a very active research topic due to its potential applications in the many fields such as human-robot interaction, human-machine interfaces, driving safety, and health-care. Despite of the significant improvements, facial expression recognition is still a challenging problem that wait for more and more accurate algorithms. This article presents a new model that is capable of recognizing facial expression by using deep Convolutional Neural Network (CNN). The CNN model is generated by using Caffe in Digits environment. Moreover, it is trained and tested on NVIDIA Tegra TX1 embedded development platform including a 250 Graphics Processing Unit (GPU) CUDA cores and Quadcore ARM Cortex A57 processor. The proposed model is applied to address the facial expression problem on the publicly available two expression databases, the JAFFE database and the Cohn-Kanade database.**

Keywords— Convolutional Neural Networks; Deep Learning, Embedded Developlement Platform with GPU.

I. Introduction

Facial expression recognition has attracted considerable attention in the past ten years due to their potential applications, such as human-robot interaction [1], human-machine interface [2] surveillance [3], driving safety [4] and health-care [5]. There are many works for recognizing the emotions, such as sadness, surprise, anger, happiness, fear, disgust, and neutrality [6-12]. However, each person gives same emotion in a different way and even the imaging conditions changes facial expression appearance. Hence facial expression recognition is still challenging problem. To get rid of the challenge and to achieve more accurate results, it is looked at robust features.

Facial expression recognition methods are separated into two categories: geometric-based methods [13-17] and appearance-based methods [18-23].

Geometric-based methods concern about the feature vectors encoding some facial geometric properties such as distance, angle, and position to determine the shapes and locations of the invariance points of face. For instance, in [16], 34 invariance points belong to a face image were extracted for facial expression recognition. 20 invariance points were derived from 74 separate landmarks in [17]. Success of the methods depends on powerful face component detection methods to set facial invariance points, which provides a few difficult at real life applications.

Appearance-based methods use the features extracted directly from the images but does not include an information relating to the facial points. There are a lot of Appearance-based methods. The most important ones are Local Binary Pattern (LBP) [18], Gabor Wavelets [19], Local Gabor Binary Patterns (LGBP) [20-21], Scale Invariant Feature Transform (SIFT) [22], Histogram of Oriented Gradient (HOG) [23], and Curvelet Transform [23]. Facial expressions make the certain regions of face change, which causes interest in just the special regions. In [24-26], the salient features were extracted from local patches.

Facial expression recognition problem constructs a classification algorithm to separate emotions into the classes such as sadness, surprise, anger, happiness, fear, disgust, and neutrality by using the extracted features. In the literature, it was obtained promising results by using Artificial Neural Networks (ANNs) [27], Support Vector Machines, (SVMs) [28], Spherical Classifiers [29], Hidden Markov Models (HMMs) [30], K-Nearest Neighbors (KNNs) [31], and compressive sensing based sparse classifiers [32] for recognizing facial expressions.

Classifiers receives descriptive features obtained from above methods as its inputs. Hence the classifier performance depend on the quality of feature vectors. Moreover the methods extract low-level features belong to facial expression but don't extract high-level feature. However recently the methods as called as deep learning have appeared as a promising one to perform facial expression recognition. Deep learning methods extract both low-level and high level features without a method. Thanks to the capability, deep learning methods has applied successfully in both signal processing and image processing [33-36].

This article proposes a new architecture of Deep Convolutional Neural Networks (CNNs) for recognizing facial expression. In addition, the proposed system is trained and tested on NVIDIA Tegra TX1 embedded development platform including a 250 Graphics Processing Unit (GPU) CUDA cores and Quadcore ARM Cortex A57 processor in order to provide that the proposed model is usable ate real life applications. Thanks to the salient features and structure of

978-1-5090-5795-5/17/$31.00 ©2017 IEEE

proposed CNN model, it is obtained both high accuracy in a fast way for facial expression recognition

In the rest of this article is organized as follows. In Section II, the CNNs are shortly reviewed. In Section III, the proposed CNN model is introduced. The experimental results are shown in Section IV. Finally, the article is concluded in Section V.

II. Deep Convolutional Neural Networks

CNN is an updated version of multi-layer neural networks [33-35]. CNNs receive as the input the images unlike conventional neural networks. The basic layers of CNNs are called as convolutional, pooling, rectified linear units, fully connected and loss layer. An example CNN structure is shown in Fig.1. The input size is reduced to the end layers. In the sequential layers, the properties from low-level and high level features are extracted.

In convolutional layer of CNNs, each input image is convoluted with kernels. The obtained output image are imported to the second layer.

In the pooling layer of CNNs, it is selected the salient features in the previously determined regions. In other way, each region is down–sampled by a non-linear down-sampling operations such as maximum, minimum or average.

In the Rectified Linear Units Layer of CNNs, it is employed the rectifier, f(x)=max(0,x) in the neural networks literature.

The fully connected layer of CNNs is located after the above defined layers. All neurons at the layer are fully connected to all activations in the previous layer.

In the loss layer of CNN, different loss function is applied. For instance, softmax loss function is used for multi class classification problems.

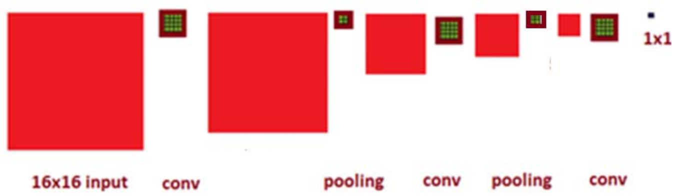

Fig. 1. An example process steps of Convolutional Neural Networks

III. The Proposed Algortihm

The graphical pipeline of the proposed CNN is given in Fig. 2. The model architecture is applied at nine layers:

(1) Input images are given.
(2) Convolutional layer is applied. Kernel size is 5x5, stride is 1, and pad is 2.
(3) Maximum pooling layer is applied. Kernel size is 3x3, stride is 2, and pad is 1.
(4) Convolutional layer is applied. Kernel size is 5x5, stride is 1 and pad is 2.
(5) Maximum pooling layer is applied. Kernel size is 3x3, stride is 1, and pad is 1.
(6) Convolutional layer is applied. Kernel size is 5x5, stride is 1 and pad is 2.
(7) Maximum pooling layer is applied. Kernel size is 3x3, stride is 2, and pad is 1.
(8) Convolutional layer is applied. Kernel size is 2x2, stride is 1 and pad is 0.
(9) Convolutional layer is applied. Kernel size is 1x1, stride is 1 and pad is 0.
(10) Finally fully connected layer is added.

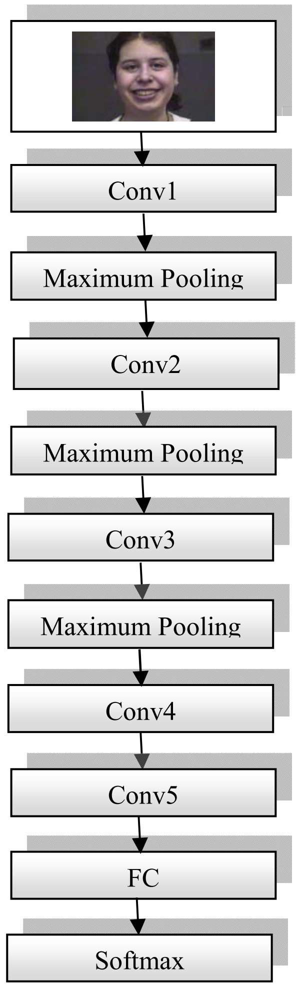

Fig. 2. The graphical pipeline of the proposed CNN model.

IV. Experiments

In this section, the effectiveness of proposed CNN model is evaluated on two datasets: JAFFE database [36] and the Cohn-Kanade database [37]. The experiments were carried out by using Caffe software developed by University of California, Berkeley in Digits environment [38-39].

In the first evaluation, the JAFFE database is used [36]. The JAFFE database consists of 10 Japanese women. Each person has two to four different images with the resolution pixel of 256 x 256 for an expression, giving a total number of 213 facial images. The database includes seven facial expression categories consisting of sadness, happy, fear, disgust surprise, angry, and neutrality. Fig. 3 gives seven facial expression images of a women from the JAFFE database.

In the second evaluation, Cohn-Kanade database was employed [36]. The database includes of 2,105 images relating to 182 students from 18 to 30 years, 15 % of which is African-American, 15 % are female, and 3 % is Asian or Latino. Images have the resolution pixel of 640 x 480 or 640 x 490. All image sequences includes a neutral and apex expressions. Especially the last frames cover the most discriminative image. This article evaluate six class classifications problem the images of the databases. Fig. 4 shows some samples of six basic facial expressions from the Cohn-Kanade database.

In the experiments, the image size was resized into16x16. The proposed model was trained for 30 epochs. The learning rate was selected as 0.001 for first eleven epochs and 0.002 for epochs 13-29 and 0.00001 for the end epoch. Batch size is selected as 40. An example visualization of the proposed CNN model is illustrated in Fig. 5

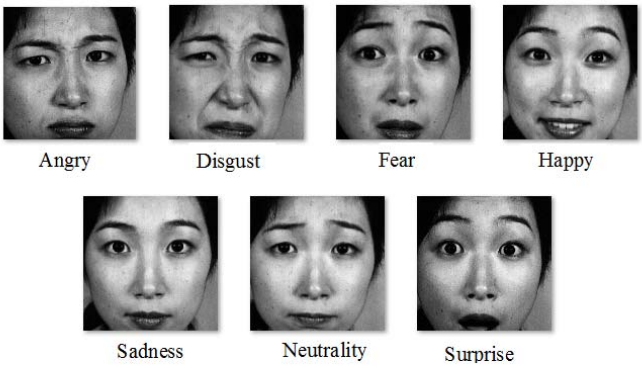

Fig. 3. Some examples from JAFFE dataset [36].

Fig. 4. Some examples from Cohn-Kanade dataset [37].

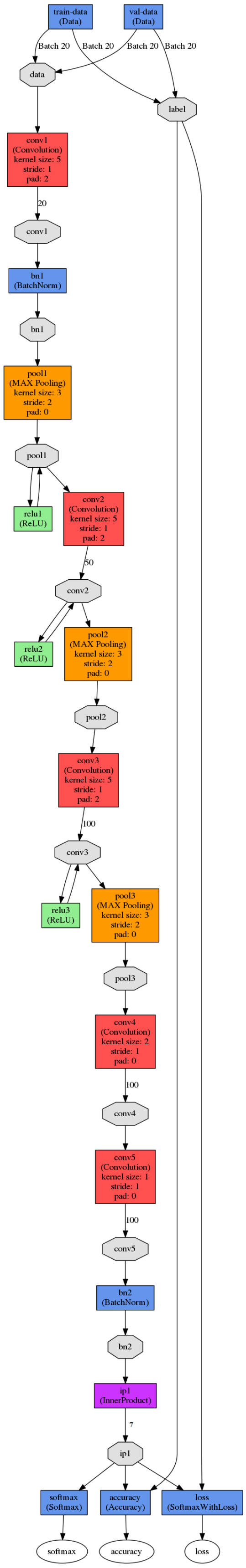

Fig. 5. An example visualization of the proposed model in digit environment.

Table I tabulates comparatively classification performances with art of state algorithms in the literature for JAFFE database. For JAFFE databases, top 1 and top5 errors were obtained as 96.10 and 100, respectively.

Table II tabulates comparatively classification performances with art of state algorithms in the literature for Cohn-Kanade database. Top 1 and top5 errors were obtained as 98.70 and 99.10, respectively. As can be seen from Table I and II, the obtained results are the best one and the proposed model outperform the others.

TABLE I. PERFORMANCE COMPARISON ON JAFFE DATASET

Study	Method	Recognition (%)
The proposed	CNN model	Top1:96.10 Top 5:100
[23], 2016	Local curvelet transform	94.65
[46], 2015	Deep belief network	90.95
[20], 2012	Radial encoded Gabor jets	89.67
[40], 2011	Patch-based-Gabor	91.00
[41], 2010	Salient feature vectors	85.92
[42], 2008	WMMC	65.77
[43], 2007	KCCA	67.00
[44], 2008	DCT	79.30
[45], 2010	FEETS + PRNN	83.84

TABLE II. PERFORMANCE COMPARISON ON COHN-KANADE DATASET

Study	Method	Recognition (%)
The proposed	CNN model	Top 1: 98.70 Top 5: 99.10
[23], 2016	Local curvelet transform	95.17
[46], 2015	Deep belief network	98.57
[19], 2009	Boosted LBP	95.15
[19], 2009	LBP	92.60
[37], 2011	Patch-based-Gabor	91.00

V. CONCLUSIONS

In this article, a new facial expression recognition algorithm is proposed based on CNN. An appropriate CNN architecture was selected. Convolutional layer numbers, kernel and stride sizes, and pooling structures were determined. Experimental were constructed on JAFFE database and the Cohn-Kanade database. The obtained results showed that the proposed system outperforms the geometric and appearance abed methods approaches in the literature.

The proposed system can be applied at real life applications. In order to prove the capability, the proposed system was trained and tested on NVIDIA Tegra TX1 embedded development platform.

REFERENCES

[1] K. Berns, J. Hirth, “Control of Facial expressions of the humanoid robot head roman,” in: IEEE International Conference on Intelligent Robots and Systems, pp.3119–3124, 2006.

[2] M. S. Bartlett, G. Littlewort, I. Fasel, J.R. Movellan, “Real time face detection and facial expression recognition: development and applications to human computer interaction,” in: IEEE Conference on Computer Vision and Pattern Recognition, pp. 53, 2003.

[3] Z. Song, B. Ni, D. Guo, T. Sim, S. Yan, “Learning universal multi-view age estimator using video context,” in: Proceedings of IEEE International Conference Comput. Vis., pp. 241–248, 2011

[4] E. Vural, M. Cetin, A. Ercil, G. Littlewort, M. Bartlett, J. Movellan, “Automated drowsiness detection for improved driver safety-comprehensive databases for facial expression analysis” in: International Conference on Automotive Technologies, 2008.

[5] P. Lucey, J. Cohn, S. Lucey, I. Matthews, S. Sridharan, K. Prkachin, “Automatically detecting pain using facial actions,” in: IEEE International Conference on Affective Computing and Intelligent Interaction, pp.1–8, 2009

[6] Y. Tian, T. Kanade, J.F. Cohn, Facial expressions analysis, in: Stan Z. Li, Anil K. Jain (Eds.), Handbook of Face Recognition, Springer-Verlag, 2004.

[7] H. Ali , M. Hariharan, S. Yaacob, A. Hamid-Adom, “Facial recognition using empirical mode decomposition,” Expert Syst. Appl., Vol. 42, no. 3, pp. 1261–1277, 2015.

[8] Y. Luo, C.Wu, Y. Zhang, “Facial expression recognition based on fusion features of PCA and LPB with SVM,” Optik, Vol. 124, no.17. pp. 2767–2770, 2013.

[9] K. Buciu, I. Pitas, “An analysis of facial expression recognition under partial face image occlusion,” Image Vision Comput., Vol. 26, no.7, pp. 1052–1067, 2008.

[10] Y. Miyakoshi, S. Kato, “Facial emotion detection considering partial occlusion of face using Bayesian network,” in: Proceedings of IEEE Symposium on Computers & Informatics (ISCI), pp. 96–101, 2011.

[11] L. Zhang, D. Tjondronegoro, V. Chadran, “Gabor based templates for facial expression recognition in images with facial occlusion,” Neurocomputing, Vol. 145, no.5, pp. 451–464, 2014.

[12] Z. Zhang , L. Wang , Q. Zhi , S. Chen , Y. Chen, “Pose invariant face recognition using facial landmarks and Weber local descriptor,” Knowl. Based Syst., Vol. 84, pp. 78–88, 2015.

[13] N. Perveen, S. Gupta, K. Verma, “Facial expression recognition using facial characteristic points and Gini index, in: IEEE Students Conference on Engineering and Systems, pp.1–6, 2012.

[14] S. He, S. Wang, Y., Lv. “Spontaneous facial expression recognition based on feature point tracking,” in: IEEE2011 Sixth International Conference on Image and Graphics (ICIG), pp.760–765, 2011.

[15] A.Lorincz, L.A.Jeni, Emotional expression classification using time-series kernels, in: IEEE Conference on Computer Vision and Pattern Recognition, pp.889–895, 2013.

[16] S. Bashyal, and G. Venayagamoorthy, “Recognition of facial expressions using Gabor wavelets and learning vector quantization,” Eng. Appl. Artif. Intell., Vol. 21, no. 7, pp.1056_64, Oct. 2008.

[17] T. H. Zavaschi, A. S. Britto Jr, L. E. Oliveira, A. L. Koerich, “Fusion of feature sets and classifiers for facial expression recognition,” Expert Syst. Appl., Vol. 40, no. 2, pp. 646–655, Feb. 2013.

[18] G. Littlewort, M.S. Bartlett, I. Fasel, J. Susskind, J. Movellan, “Dynamics of facial expression extracted automatically from video”, Image Vis. Comput., Vol.24, no.6, 615–625, 2006.

[19] C. Shan, S.Gong, P.W. McOwan, “Facial expression recognition based on local binary patterns: a comprehensive study,” Image Vis. Comput., Vol. 27, no.6, pp. 803–816, 2009.

[20] G. W, Xiang C, Venkatesh YV, D. Huang, H. Lin, “Facial expression recognition using radial encoding of local Gabor features and classifier synthesis,” Pattern Recognit., Vol. 45, no. 1, pp. 80–91, 2012.

[21] X. Sun, H. Xu, C. Zhao, J.Yang, “Facial expression recognition based on histogram sequence of local Gabor binary patterns,” in: IEEE Conference on Cybernetics and Intelligent Systems, pp.158–163, 2008.

[22] U. Tariq, K.H. Lin, Z. Li, X. Zhou, Z. Wang, V. Le, T.S. Huang, X. Lv, T.X. Han, "Emotion recognition from an ensemble of features," in: IEEE International Conference on Automatic Face and Gesture Recognition and Workshops (FG), pp.872–877, 2011.
[23] Y. Hu, Z. Zeng, L.Yin, X. Wei, X. Zhou, T.S. Huang, "Multi-view facial expression recognition," in: IEEE International Conference on Automatic Face and Gesture Recognition and Workshops (FG) , pp.1–6, 2008.
[24] A. Uçar, Y. Demir, C. Güzeliş, "A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering," Neural Comput. and App., Vol. 27. pp. 131-142, 2016.
[25] P. Yang, Q. Liu, D.N. Metaxas, "Exploring facial expressions with compositional features," in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.2638–2644, 2010.
[26] L. Zhong, Q. Liu, P. Yang, B. Liu, J. Huang, D. N. Metaxas, "Learning active facial patches for expression analysis", in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.2562–2569, 2012.
[27] S. Haykin, Neural Networks and Learning Machines, Prentice Hall: 3rd ed. Upper Saddle River, USA, 2008.
[28] C. Cortes, V.N. Vapnik, Support vector networks, Mach. Learn. 20 (1995) 273–297.
[29] A. Uçar, Y. Demir, C. Güzeliş, "A penalty function method for designing efficient robust classifiers with input–space optimal separating surfaces," Turk J Elec Eng & Comp Sci., Vol. 22, pp. 1664–1685, 2014.
[30] L. Rabiner, B. Juang, "An introduction to hidden Markov models," IEEE ASSP Mag., Vol. 3, no. 1, pp. 4_16, Jan. 1986.
[31] T. Cover, P. Hart, "Nearest neighbor pattern classification," IEEE Trans. Inform. Theor., Vol. 13, no. 1, pp. 21_7, Jan. 1967.
[32] S. Zhang, X. Zhao, B. Lei, "Robust facial expression recognition via compressive sensing," Sensors, Vol. 12, no. 3, pp. 3747–3761, Mar. 2012.
[33] L. Deng, D. Yu, "Deep learning: methods and applications," Foundations and Trends in Signal Processing, Vol.7. pp. 3–4, 2014.
[34] S. Christian, A. Toshev, D. Erhan. "Deep neural networks for object detection," Advances in Neural Information Processing Systems, 2013.
[35] A. Krizhevsky, I. Sutskever, G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Neural Information Processing Systems, pp. 1097–1105, 2012.
[36] M. Lyons, J. Budynek, S. Akamatsu, "Automatic classification of single facial images," IEEE Trans Pattern Anal Mach Intell., Vol. 21, no.12, pp. 1357–1362, 1999.
[37] T. Kanade, J.F. Cohn, T. Yingli, "Comprehensive database for facial expression analysis," in: IEEE 4th international conference on automatic face and gesture recognition. Pittsburgh, PA, USA, pp. 46–53, 2000.
[38] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, "Caffe: convolutional architecture for fast feature embedding," arXiv preprint arXiv:1408.5093, 2014.
[39] Deep Learning GPU Training System https://developer.nvidia.com/digits, 2017.
[40] L. Zhang, D. Tjondronegoro, "Facial expression recognition using facial movement features. IEEE Trans Affect Compt., Vol. 2, no. 4, pp. 219–229, 2011.
[41] M. Kyperountas, A. Tefas, I. Pitas, Salient feature and reliable classifier selection for facial expression classification. Pattern Recognit., Vol. 43, no.3, pp. 972–986, 2010.
[42] C. Zhengdong, S. Bin, F. Xiang, Z. Yu-Jin, "Automatic coefficient selection in weighted maximum margin criterion," in: 19th International conference on pattern recognition. Tampa, FL, pp. 1–4, 2008.
[43] Y. Horikawa, "Facial expression recognition using KCCA with combining correlation kernels and Kansei information," in: International conference on computational science and its applications. Kuala Lampur, Malaysian, pp. 489–498, 2007.
[44] J. Bin, Y. Guo-Sheng, Z. Huan-Long, "Comparative study of dimension reduction and recognition algorithms of DCT and 2DPCA," in: International conference on machine learning and cybernetics. Kunming, China, pp 407–410, 2008.
[45] J. Wong, S.A. Cho, "Face emotion tree structure representation with probabilistic recursive neural network modeling," Neural Comput Appl., Vol. 19, no. 1, pp. 33–54, 2010.
[46] X. Zhao, X. Shi, S. Zhang, Facial expression recognition via deep learning, IETE Technical Review, Vol. 32, no.5, 2015.