Deep Learning for Facial Expression Recognition: *A step closer to a SmartPhone that Knows your Moods.*

Shabab Bazrafkan[1], Tudor Nedelcu[1], Pawel Filipczuk[2], Peter Corcoran[1] *Member, IEEE*
1: Center for Cognitive, Connected & Computational Imaging, College of Engineering & Informatics,
NUI Galway, Galway, Ireland
2: Fotonation LTD, Galway, Ireland
E-mails: {s.bazrafkan1 , t.nedelcu1}@nuigalway.ie, pfilipczuk@fotonation.com,
peter.corcoran@nuigalway.ie

***Abstract*—By growing the capacity and processing power of the handheld devices nowadays, a wide range of capabilities can be implemented in these devices to make them more intelligent and user friendly. Determining the mood of the user can be used in order to provide suitable reactions from the device in different conditions. One of the most studied ways of mood detection is by using facial expressions, which is still one of the challenging fields in pattern recognition and machine learning science.**

Deep Neural Networks (DNN) have been widely used in order to overcome the difficulties in facial expression classification. In this paper it is shown that the classification accuracy is significantly lower when the network is trained with one database and tested with a different database. A solution for obtaining a general and robust network is given as well.

I. Introduction

Today's handheld devices are growing in their capacity to interact with end-users. They have access to an ever-growing range of network based services and their sensing capabilities of the location and local environment continue to grow in scope. One remaining challenge for today's devices is to sense and determine the emotional state of the user. This introduces new challenges [1], [2] and requires a range of sophisticated edge technologies that can capture and analyze information from the user on the device. One example is the real-time analysis of speech patterns for detecting emotion [3], [4]. More recently researchers in this field have turned to deep learning techniques [5]. Facial expression analysis is also well known in the literature [6]–[11].

But it is computationally complex and it is challenging to achieve high recognition rates using conventional feature extraction and classification schemes. In this paper we follow the trend from the speech recognition field and explore facial emotion recognition using deep learning techniques. The goal is to demonstrate the potential for high performance solution that can run on relative lightweight convolutional neural networks that can be efficiently implemented in hardware or on a GPU. Such a solution could realistically enable a new generation of smartphones that can understand the moods of their owners.

A. Facial Expression Classification

This In recent years the facial expressions classification has attracted a lot of attention because of it's various potential applications including psychology, medicine, security [12], man-machine interaction and surveillance [13]. There are two main approaches to investigate the facial expression in a systematic way: Action Unit (AU) based and appearance based methods.

AU model introduces the Facial Action Coding System (FACS) which has been developed by Carl-Herman Hjortsjö in 1969 [14]. This technique described the facial expression as a composition of Action Units which are describing the facial muscle motions. This method takes advantage of the strong support of the psychology and physiology sciences since it uses the facial muscle movements for modeling different expressions [13]. The AU based methods suffer from the difficulties such as dependencies on invisible muscle motions [13] which makes it extremely difficult to model the FACS system using machines.

In contrast the appearance based methods are using feature extraction, feature selection and classification methods [15] in order to determine the expression in the face. In this approach different kinds of features have been used so far including, Local Binary Pattern [16], Scale Invariant Feature Transform (SIFT) [17] and Histogram of Oriented Gradient [18]. There are several main difficulties in facial expression detection, like, changes in the appearance and the shape of the face in unexpected ways (because of the non-rigidity of the face) [19], imaging conditions, and inter-person differences in facial expressions. Because of these basic problems there are no golden methods which can be called as a standard for automatic facial expression classification.

B. Deep Learning

In recent years by emerging powerful parallel processing hardware, Deep Neural Networks (DNN) become a hot topic in pattern recognition and machine learning science. Deep Learning (DL) scheme is based on the consecutive layers of signal processing units in order to mix and re-orient the input data to their most representative order correspond to a specific application.

Facial expression classification has taken advantage of DNN classifiers in recent years. In [13] an AU inspired Deep Network has been proposed which uses the DNN to extract the most representative features.in [20] a DNN with five layers

This research is funded under the *SFI Strategic Partnership Program* by Science Foundation Ireland (SFI) and FotoNation Ltd. Project ID: 13/SPP/I2868 on *Next Generation Imaging for Smartphone and Embedded Platforms.*

This work is supported by the Enterprise Partnership Scheme program of the Irish Research Council

978-1-5090-5544-9/17/$31.00 ©2017 IEEE

and 65K neurons has been designed to classify the expression into five categories (Neutral, happy, sad, angry and surprised). In [15], a Boosted Deep Belief Network (BDBN) has been proposed and implemented using joint fine tune process in BDBN framework to classify the facial expression.

In all DNN based investigations on facial expression, the proposed network is trained and tuned for a specific database and the test data is drawn from the same database as well. Since the network is biased for a specific data type, the result on the test data of that dataset will give surprisingly low error rate. But if the designed network for database A would be tested on database B the results will be shockingly bad. Therefore, these classifiers would fail to work in wild environments.

The main goal of this paper is to investigate the amount of error caused by network trained with one database and tested with other and also design and implement a network which can overcome the problem with mixing databases for training stage.

In the next section an introduction to Deep Neural Networks (DNN) is presented, and also Databases and database expansion is presented. In the Third section the networks designed for single and multi database purposes are given and results and discussion is given in section 4.

II. Methodology

The goal of this research is to investigate the amount of error that occurs from training a DNN network for a database and test it on other database. This inter-database investigation can give a general perspective on design and train networks for wild applications and costumer device implementations.

A. Deep Neural Networks (DNN)/Deep Learning (DL)

Deep Neural Networks training also known as Deep Learning is one of the most advanced machine learning techniques trending in recent years due to appearance of extremely powerful parallel processing hardware and Graphical Processing Units (GPU). Several consecutive signal processing units are set in serial/parallel architecture mixing and re-orienting the input data in order to result in most representative output considering a specific problem. The most popular image/sound processing structure of DNN is constructed by three main processing layers: Convolutional Layer, Pooling Layer and Fully Connected Layer. DNN units are described below:

Convolutional Layer: This layer convolves the (in general 3D) image “I” with (in general 4D) kernel “W” and adds a (in general 3D) bias term “b” to it. The output is given by:

$$P = I*W + b, \tag{1}$$

where * operator is nD convolution in general. In the training process the kernel and bias parameters are selected in a way to optimize the error function of the network output.

Pooling Layer: The pooling layers applies a non-linear transform on the input image which reduce the neuron numbers after the operation. It’s common to put a pooling layer between two consecutive convolutional layers. This operation also reduces the unit size which will lead to less computational load and also prevents the over-fitting problem.

Fully Connected Layer: Fully connected layers are exactly same as the classical Neural Network (NN) layers where all the neurons in a layer are connected to all the neurons in their subsequent layer. The neurons are triggered by the summation of their input multiplied by their weights passed from their activation functions.

B. Databases

Three databases has been used in the research. Radboud Faces Database (RaFD) [21], Cohn-Kanade AU-Coded Facial Expression Database Version 2 (CK+) [22] and The Japanese Female Facial Expression (JAFFE) Database [23].

RaFD: The Radboud Faces Database is a set of 67 persons with different gender and different races (Caucasian and Moroccan Dutch), both children and adults. This database displays 8 different emotions which in the presented work seven of them are used.

CK+: Cohn-Kanande version 2 (known as CK+) database is a facial expression database including both posed and non-posed expressions wherein the subject changes emotion in several sequences from neutral to one of seven different expressions. In the presented work just the non-posed data has been used.

JAFFE: The Japanese Female Facial Expression Database is made of 213 images of 7 expressions contains faces of 10 Japanese female models. Since the number of images in this database is not sufficient to train a DNN, this database is eliminated from our inter-database investigations and is used just for multi database network training.

C. Database Expansion

Since the number of image sin each database is not enough in order to train a DNN a database expansion scheme has been used to overcome the problem which includes flipping images and rotating them by [-3,-2,-1,1,3,3] degrees and put them back in the dataset. Using this approach, we ended up with large number of images for each dataset shown in table 1.

Table1: Number of images in each database and each dataset

Database\Dataset	Train	Validation	Test
RaFD	13160	312	146
CK+	14392	304	187
JAFFE	1960	42	22

As it has been shown in table 1, we can see that there are not enough images in JAFFE database to train a DNN. This database is used to train the multi-database network. Just note that the database expansion applied to training samples and Validation and Test samples remain unchanged.

978-1-5090-5544-9/17/$31.00 ©2017 IEEE

III. Deep Neural Networks for Expression Classification

A. Single database networks

For each of databases RaFD and CK+ a DNN has been designed and trained and the results for testing on each database is calculated as well. The networks designed for each database are given in the following list.

DNN on RaFD: The architecture for network trained on RaFD database is given in table 2.

Table 2: Network configuration for RaFD database

layer	Kernel/Units	Size/Dropout Probability
Convolutional	16	3x3
Maxpool	N/A	2x2
Convolutional	8	3x3
Maxpool	N/A	2x2
Convolutional	8	3x3
Maxpool	N/A	2x2
Fully Connected	15	Dropout p=0.8
Fully Connected	7	Dropout p =0.5

DNN on CK+: The architecture for network trained on CK+ database is given in table 3.

Table 3: Network Configuration for CK+ database.

Layer	Kernel/Units	Size/Dropout Probability
Convolutional	8	3x3
Maxpool	N/A	2x2
Convolutional	8	3x3
Maxpool	N/A	2x2
Convolutional	8	3x3
Maxpool	N/A	2x2
Fully Connected	7	Dropout p =0.5

B. Multi-Database Network

A network has been designed and trained for a mixture of all three databases. The architecture of the network is given in table 4.

Table4: Network Configuration for mixed Dataset

Layer	Kernel/Units	Size/Dropout Probability
Convolutional	16	3x3
Maxpool	N/A	2x2
Convolutional	13	3x3
Maxpool	N/A	2x2
Convolutional	10	3x3
Maxpool	N/A	2x2
Fully Connected	7	Dropout p =0.5

IV. Results and Discussion

Three networks explained in the previous section have been implemented using Lasagne library (a Theano based DNN library in python). The mean categorical cross-entropy loss function has been used with Nestrov momentum optimization. First and second networks shown in tables 2 and 3 are trained by RaFD and CK+ datasets respectively and the third network shown in table 4 is trained using a mixture of three databases RaFD, CK+ and JAFFE. The main goal of this work is to present the cross-database error which is accomplished by calculating the amount of error for each network using all databases. The classification error is presented in table 5.

Table5: Error given from each network for each dataset

Network\error	Error for RaFD	Error for CK+	Error for JAFFE
Network 1	6.84%	59.59%	72.73%
Network 2	21.23%	19.59%	50%
Network 3	4.1%	16.04%	13.36%

The first two rows of the table 5 are associated with the classification test error values for networks trained just with one database; RaFD for network 1 and CK+ for network 2. Since these networks are trained with just one database, they are tuned for that specific database and the test error for that specific network is less than the errors for other databases. Network 3 is trained with a mixture of three databases given in section II.B.

The most important result is that in network 3 the test error for each database is even lower than the value of the error from the network which is trained by that specific database.

While implementing a solution in consumer devices, it is crucial to provide algorithms which are robust to environment and condition changes. In this work it has been shown that in order to obtain a more general and robust DNN, one of the solutions is to mix as much data as possible. In fact adding a wide range of samples drawn from different conditions and properties to DNN training sets, will lead to a more reliable network for wild use cases which is one of the most important considerations in consumer electronic devices.

Reference

[1] V. Pejovic and M. Musolesi, "Anticipatory mobile computing: A survey of the state of the art and research challenges," *ACM Comput. Surv.*, vol. 47, no. 3, pp. 1–29, 2015.

[2] R. Rana, M. Hume, J. Reilly, R. Jurdak, and J. Soar, "Opportunistic and Context-Aware Affect Sensing on Smartphones," *IEEE Pervasive Comput.*, vol. 15, no. 2, pp. 60–69, 2016.

[3] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognit.*, vol. 44, no. 3, pp. 572–587, 2011.

[4] C. N. Anagnostopoulos, T. Iliou, and I. Giannoukos, "Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011," *Artif. Intell. Rev.*, vol. 43, no. 2, pp. 155–177, 2012.

[5] R. Rana, R. Jurdak, X. Li, and J. Soar, "Emotion Classification from Noisy Speech-A Deep Learning Approach," *arXiv Prepr. arXiv*, 2016.

[6] J. Sung and D. Kim, "Pose-Robust Facial Expression Recognition Using View-Based 2D + 3D AAM," *Syst. Man Cybern. Part A Syst. Humans, IEEE Trans.*, vol. 38, no. 4, pp. 852–866, 2008.

[7] I. Bacivarov and P. Corcoran, "Facial expression modeling using component AAM models—Gaming applications," in *Games Innovations Conference, International IEEE Consumer Electronics Society's. (ICE-GIC 2009)*, 2009, pp. 1–16.

[8] G. Mancini, S. Agnoli, B. Baldaro, P. E. R. Bitti, and P. Surcinelli, “Facial expressions of emotions: recognition accuracy and affective reactions during late childhood.,” *J. Psychol.*, vol. 147, no. 6, pp. 599–617, 2013.

[9] S. L. Happy and A. Routray, “Automatic facial expression recognition using features of salient facial patches,” *IEEE Trans. Affect. Comput.*, vol. 6, no. 1, pp. 1–12, 2015.

[10] L. Ding and A. M. Martinez, “Features versus context: An approach for precise and detailed detection and delineation of faces and facial features.,” *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 32, no. 11, pp. 2022–38, Nov. 2010.

[11] I. Bacivarov, “Advances in the Modelling of Facial Sub-Regions and Facial Expressions using Active Appearance Techniques,” National Uinversity of Ireland Galway, 2009.

[12] O. Rudovic, M. Pantic, and I. Patras, “Coupled Gaussian processes for pose-invariant facial expression recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1357–1369, 2013.

[13] M. Liu, S. Li, S. Shan, and X. Chen, “AU-inspired Deep Networks for Facial Expression Feature Learning,” *Neurocomputing*, vol. 159, no. 1, pp. 126–136, 2015.

[14] C.-H. Hjortsjö, *Man’s face and mimic language*. Studen litteratur, 1969.

[15] P. Liu, S. Han, Z. Meng, and Y. Tong, “Facial Expression Recognition via a Boosted Deep Belief Network,” *2014 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1805–1812, 2014.

[16] C. Shan, S. Gong, and 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.

[17] U. Tariq, K.-H. Lin, Z. Li, X. Zhou, Z. Wang, V. Le, T. S. Huang, X. Lv, and T. X. Han, “Emotion recognition from an ensemble of features,” in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, 2011, pp. 872–877.

[18] Y. Hu, Z. Zeng, L. Yin, X. Wei, X. Zhou, and T. S. Huang, “Multi-view facial expression recognition,” in *Automatic Face & Gesture Recognition, 2008. FG’08. 8th IEEE International Conference on*, 2008, pp. 1–6.

[19] Y. Wu, Z. Wang, and Q. Ji, “Facial feature tracking under varying facial expressions and face poses based on restricted boltzmann machines,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, pp. 3452–3459, 2013.

[20] I. Song, H. J. Kim, and P. B. Jeon, “Deep learning for real-time robust facial expression recognition on a smartphone,” *Dig. Tech. Pap. - IEEE Int. Conf. Consum. Electron.*, pp. 564–567, 2014.

[21] O. Langner, R. Dotsch, G. Bijlstra, D. H. J. Wigboldus, S. T. Hawk, and A. van Knippenberg, “Presentation and validation of the Radboud Faces Database,” *Cogn. Emot.*, vol. 24, no. 8, pp. 1377–1388, 2010.

[22] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, 2010, pp. 94–101.

[23] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, “Coding facial expressions with gabor wavelets,” in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, 1998, pp. 200–205.

978-1-5090-5544-9/17/$31.00 ©2017 IEEE