Journal of Computers Vol. 29 No. 6, 2018, pp. 132-139
doi:10.3966/199115992018122906012

Good Portrait Selection Based on Deep Learning Using Facial Expressions

Umer Waqas, Tae-Young Choe*

Department of Computer Engineering, Kumoh National Institute of Technology,
Gumi, South Korea
{umerwaqas, choety}@kumoh.ac.kr

Received 9 January 2018; Revised 22 January 2018; Accepted 22 January 2018

Abstract. This paper suggests an algorithm to choose good pictures among a bundle of pictures. Since it is very hard to collect private selections of good images from people, the paper divides facial expressions into ten categories, and add 'good' tags to photos included in some categories like wink or grin. The proposed algorithm uses convolution neural networks (CNN) to classify the pictures as good or not. The experimental results show that the accuracy of the algorithm is 97.15%, and average execution time is 0.3 seconds. For application purposes, the proposed algorithm is further applied to a group photo in order to count the number of faces that look good.

Keywords: convolution neural networks, deep learning, facial expressions, image processing, photo psychology, picture selection

1 Introduction

Increased number of digital devices having cameras provided access to everyone for making photographs. In recent years cameras became cheaper and people are capable of having huge collection of pictures. Another advancement in recent years are social media platforms where people love to share their pictures. In such situations, they want to choose a good picture out of many pictures in their collection. This task is not easy for them because of various factors like phycology, time and how huge collection is. This motivates the idea of developing an algorithm which handles such situations, such that a user does not need to choose good pictures by himself. Instead, the algorithm suggests good pictures from a collection of pictures. Our determination is to help people finding good pictures from their collection in a very short time. Furthermore, in this paper facial expressions are considered. Facial expression recognition has gained much research focus in recent days. Since facial expressions provide an important behavioral measure for the study of emotion, cognitive processes, and social interaction [1], it has several practical applications in many fields. Most of the works in this field focus on improving the recognition accuracy. However, there is still plenty of room available for better practical usability of facial expressions. The objective of our research is not to recognize facial expressions, but to present an algorithm that classifies a good picture based on facial expressions.

Choosing a picture as good is not an easy job because preference and taste vary from people to people. It raises the problem of first defining what a good picture really is. The development of a solution to classify good pictures is extremely important because of its broad spectrum of applicability in a variety of fields, such as phototherapy, smart phones, virtual reality, wearables, etc.

In this work, the process of choosing a good picture is dependent on the information that is retrieved from facial expressions. There are always some pictures that are liked by majority of the people. Especially, pictures that show happiness are quite popular. Since, facial expressions contain information, as Alice said, "The most remarkable aspect of the human face is the diversity of information it provides to the human observer, both perceptually and socially" [2]. Cohen et al. said, "The most expressive way

* Corresponding Author

humans display emotions is through facial expressions" [3]. Emotion is one of the important information obtained from a facial expression. While one can investigate the recognition of good pictures using multiple ways like facial expressions, posture of bodies, and backgrounds, this work focuses on facial expressions.

Rest of the paper is structured as follows. Section 2 describes the previous works. Section 3 describes the idea and scheme proposed in this paper. The section also contains explanation of psychological background, dataset used to train and test the algorithm, convolutional neural network model, and the algorithm flowchart. Section 4 provides results of the proposed algorithm, which include training loss, training accuracy, validation loss, and validation accuracy in the form of graphs. Section 5 provides one possible application of this algorithm that counts the number of good faces in a group photo. Finally, section 6 concludes the paper.

2 Previous Works

Rodden study described some insight into how computer-based systems to help people organize images might be designed [4]. They interviewed people and asked about their collection of personal photographs, how it is organized, what they use their photographs for, and how they go about the task of searching for particular photographs. They proposed a shoebox system that resulted in the creation of a number of content-based multimedia indexing and retrieval tools. This technique keeps track of the pictures that users like and they have already chosen by themselves. But our objective is to design an algorithm to choose good pictures among number of pictures. In this case even users don't have to choose the pictures by themselves. So, an efficient automated way is required to choose good pictures.

Andreas et.al. proposed three visual interfaces that use keyframes—still images automatically pulled from video—to provide access points for efficient navigation of recorded contents. The output images can help in identifying potentially useful or relevant video segments. They used hierarchical agglomerative clustering algorithm to cluster video frames and segments by color similarity to select and to present keyframes [5].

Another prominent line of research is facial expression which is related to our work. In studies of facial expression, researchers have addressed how emotions develop, to what extent the information they convey is best captured verbally by discrete categories or scalar dimensions, whether emotions have distinct biological substrates, and the extent to which facial expressions of emotion are universal [6]. Furthermore, American psychologist Paul Ekman did extensive research in facial expressions and emotions. He discussed important questions about facial expressions and emotions. The most relevant question to us is "What information does an expression typically convey?" [7]. Explanation of this question shows that it is possible to gather different type of information from various facial expressions. Moreover, in the past, facial expression analysis was primarily a research subject for psychologists, but already in 1978, a preliminary investigation on automatic facial expression analysis from an image sequence was presented [8].

Over the last decade, automatic facial expression analysis has become an active research area that finds potential applications in areas such as human–computer interfaces, image retrieval talking heads, and human emotion analysis. Facial expressions reflect not only emotions, but also social interaction, mental activities and physiological signals [9]. A lot of research has been carried out to detect facial expressions. An algorithm to recognize six universal facial expressions with 78% accuracy was proposed in [10]. A cluster based approach to recognize three facial expressions was proposed in [11]. The facial expressions were neutral, smile and anger [11]. Similarly, various experiments in facial expression recognition have been carried out over the past decade but they had a common goal of "better facial recognitions" by using different methods. They classified facial expressions in multiple categories by using different datasets. A key difference in the present study is that we are presenting an algorithm using facial expression where users are not required to choose good picture by themselves.

3 Proposed Scheme

Facial expression recognition is important as facial expressions convey emotions, where happiness is a key emotion for selecting a good picture. This motivates the idea of selecting good pictures using facial expressions in this paper. In this work, deep learning techniques are used to train the considered model. There are two output classes to select a picture as good or not by looking at facial expressions. To the best of author's knowledge, using facial for good picture selection is not addressed in the existing literature. Facial expressions used in this work are: smile, grin, tongue, kiss, wink, sad, surprise, angry, disappointed, and cry. These expressions cover many points of concern while choosing a good picture. The dataset used in this work is from flicker. The dataset was used in a project "the face we make". The purpose of this project was to understand how well people identify emotions represented by the emoticons that we used every day in our digital life. The data was collected by asking strangers on the streets of New York to perform the emotions suggested by ten common emoticons. This dataset is from people's real life and it contains frequently used facial expressions. In this dataset, every person has ten pictures with ten facial expressions. The dataset contains around 1960 images where we used 80% for training the model and 20% for testing. We used CNN to train our dataset. The CNN categorizes images into two types i.e., "good" or "not".

To classify an image as good, it is required to train the CNN with some guidelines such as a component that affects classification of images. Two approaches were considered in order to set the guidelines: analytical or sensuous. Analytical approach was chosen in this work, which suggests some components in an image. One component is facial expressions. There is diverse research on relations between facial expression and good photo in psychology area. So, we consulted psychology in order to make guidelines which are compulsory for a good picture. It was noted that wink, grin, tongue, and kiss are robust facial expressions that contribute in a selection of good picture.

3.1 Psychological Background

In order to choose a good picture, guidelines needed to be defined. Fortunately, many researches are processed on 'good picture' in psychology area. Daniel said "The field of photo psychology tells us that we're also reading pictures very differently" [12]. It means that every human has a different way of perceiving things and making choice. People like some photos and at the same time some people may not like the same photos. Anastasia said "The best definition so far that captures the essence of the psychology of choice is that choice is the purest expression of free will" [13]. Since, everyone has a free will to choose a photo depending on their taste, we had to standardize some basic and important features that most of the human likes. Ming Thein said "According to the psychology a strong image has to tell a story: that's secondary evidence of the viewer being able to make some inferences from the visual cues" [14]. It is observed that inferences upon looking at photo can be made and some cues to choose features that helps to define a good photo. Jane et al. said "Human face processing skills can make simultaneous use of a variety of information from the face, including information about the age, sex, race, identity, and even current mood of the person" [15]. In other words, mood identification of a person by looking at face is possible. Lauren Lim said "According to Photography techniques 'happiness is the key point of a good picture'" [16].

According to Rubin et al. the muscles around the mouth creates smile by opening and closing the lips thus showing the teeth, which creates nasolabial fold [17]. This forms grin and which is the strongest cue of happiness.

Similarly, tongue and wink also show nasolabial folds. As grin, tongue, wink and kiss are the indications of person's good mood or happiness that leads to a good picture selection. So, we standardize these four facial expressions for our algorithm. In this way when our algorithm look at these four facial expressions, it knows that there is a high probability of choosing it as a good picture. Hence our algorithm classify them as good pictures.

3.2 Dataset and PreProcessing

For training the CNN we used 1568 images where every person has ten facial expressions. For testing the model we have 392 images. In our dataset, we have total 190 different people and they all have ten facial expressions. The original image size is 500*500. The original images are in RGB (Colored). Fig. 1 shows the ten facial expressions of a single person. Similar to Fig. 1 all other 190 people have images with the ten facial expressions.

Fig. 1. Sample images with ten emotions

Images were preprocessed as follows:

· Image Dimensions were reduced to 200* 200.

· Color images were converted to gray.

· Assigned labels for training.

· The dataset was shuffled.

· Normalized the dataset by dividing pixel value by 255 to keep the scale between 0 and 1 for fast processing.

3.3 CNN Model

Keras [18] is used as front end and Theano [19] as back end. Keras Sequential Convolutional Neural Network Model was applied. The model has 4 convolution layers, 2 max pooling layers, 1 dense layer and 1 output layer. Batch size of 32 was used during training. 30 number of epochs were considered, after 30 epochs validation loss did not decrease any further. Fig. 2 shows the diagram of the CNN model in our scheme.

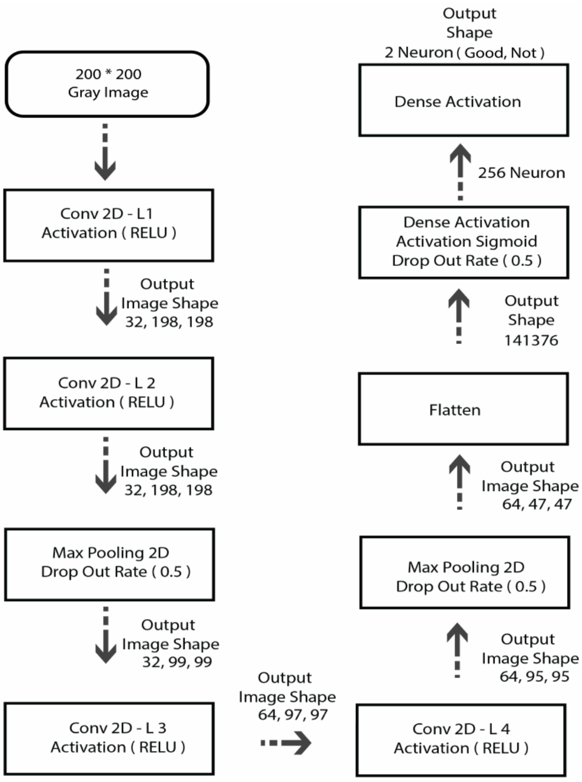

Fig. 2. CNN Model Diagram

In the first layer, input image was provided with 200*200 dimensions. In the first and second convolutional layer, 32 filters were used. Filter size was 3*3 and stride rate was (1, 1). ReLU (Rectified Linear Units) source for the activation function [20] was applied. After passing through first convolutional layer the output image size was same and after passing through second layer the image size was decreased to 198*198.

In next step Max Pooling Layer was used with stride of (2, 2). Pooling size was 2*2 and the output image size was reduced to 99*99. Dropout with rate 0.5 was applied.

We again passed the 99*99 image to 3rd and 4th convolution layer. In 3rd and 4th convolutional layer we increased the number of filters from 32 to 64. Filter size was same 3*3 and stride rate was also same (1, 1). ReLU activation function was used in these layers. After passing from 3rd convolutional layer the image size was reduced to 99*99. After passing through 4th convolutional layer the image size was reduced to 97*97.

Max Pooling layer was applied again with stride of (2, 2) and pooling size of 2*2. The output image was reduced to 47*47. The same dropout rate of 0.5 was used. Then this output was passed to flattened layer where we got 141376 parameters.

Output of flattened layer was connected to a dense layer. 254 neurons were used in dense layer which were connected to final output layer. In final output layer we had 2 neurons which trigger one if it classify image as good and trigger zero if it classify image as not good. We named our CNN model as DGF (Detecting Good Faces). Section 5 shows one of its application.

3.4 Algorithm Flowchart

The overall flow of the algorithm is shown in Fig. 3. Input image was converted into gray scale and then its dimensions were reduced. Algorithm accepts the dimensions of 200*200 so after reducing the dimensions the image was passed to CNN. The CNN classified the image as good or not and predicted the classes in form of zero or one. An extra module is added to show an application of this algorithm. In this module a group photo is given as input to algorithm, first it detects every face in the picture then crops it and pass it to algorithm where algorithm looks at facial expression of every face and classifies who is looking good in picture.

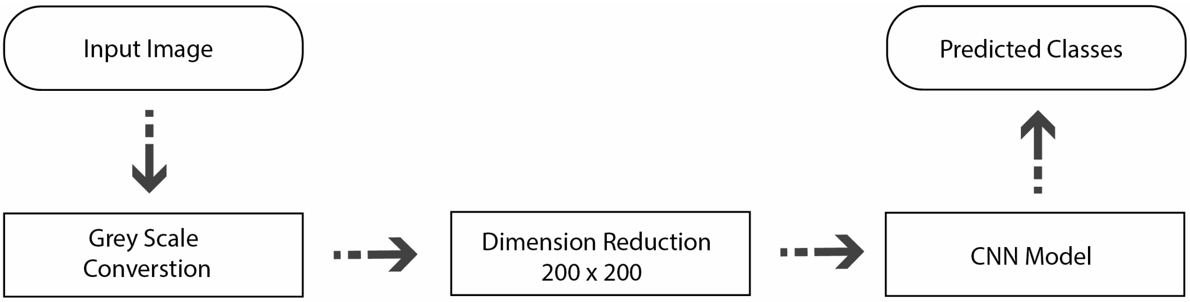

Fig. 3. Algorithm flowchart

4 Experimental Results

We monitored the change in loss and accuracy curves by increasing the number of epochs in order to verify algorithm is improving or not. Algorithm parameters were adjusted like dropout rate after looking at different variations in graphs. Fig. 4 shows the training loss vs validation loss curves. Since training and validation loss was decreasing normally and there was not too much variation between them, this meant that underfitting or overfitting problem can be negligible.

Fig. 5 compares the training accuracy vs validation accuracy. The accuracy started to improve from epoch 23 till epoch 30. After that there was not much improvement because algorithm achieved the maximum training accuracy of 98% and validation accuracy was 97.30%.

In our experiment different drop rates were used and obtained different variations in curves. By reducing dropout rate slowly to 20%, validation curves became worse. Similarly by increasing upto 80% the validation curves were disturbed a lot. Since, dropout rate of 0.5 provided smooth curves and prevented underfitting and overfitting so we fixed it in our final experiment.

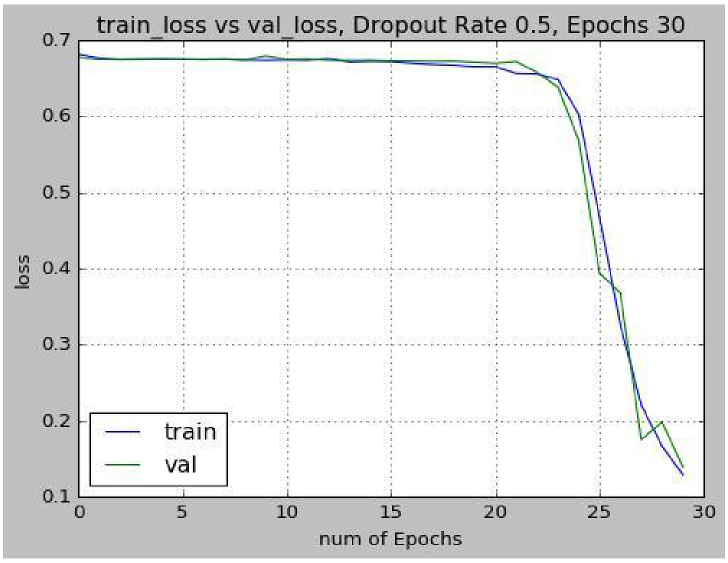

Fig. 4. Training Loss vs Validation Loss

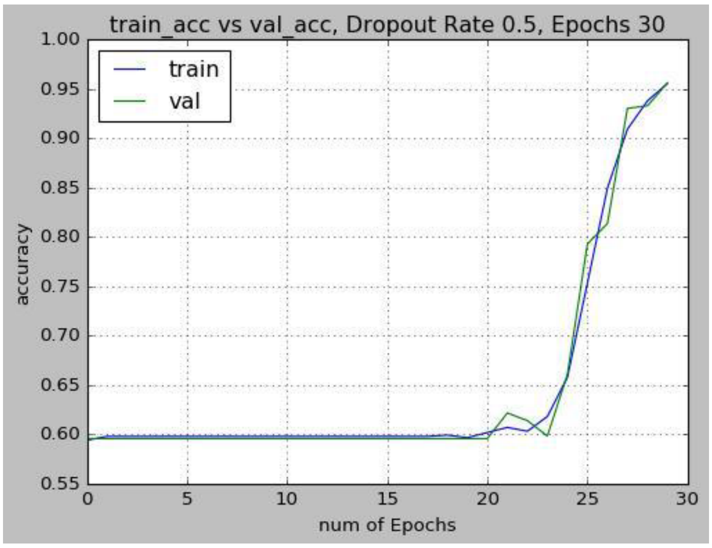

Fig. 5. Training accuracy vs. validation accuracy

The performance of algorithm is also tested against test set and we achieved the testing accuracy of 97.15%. In our final test we again provided ten images of different persons with ten facial expressions. These images were not present in training or testing set. When we provided these totally unseen images to algorithm, it correctly classified these ten images as good or not.

5 Application of Algorithm on Group Photos

The proposed algorithm can be used for various applications. However, in this work, for application purposes, the algorithm is applied to recognize and count the number of good faces in a particular group photo. OpenCV library is used to detect faces in a group photo. Once the faces are detected, they are cropped and passed to the developed algorithm. The algorithm analyze all faces in the group photo one by one, and gives an array of 0s and 1s as output. Each element of this array represents the classification of each individual in the group, where 1 represents “good” and 0 represents not good”. Fig. 6 shows the application flow chart.

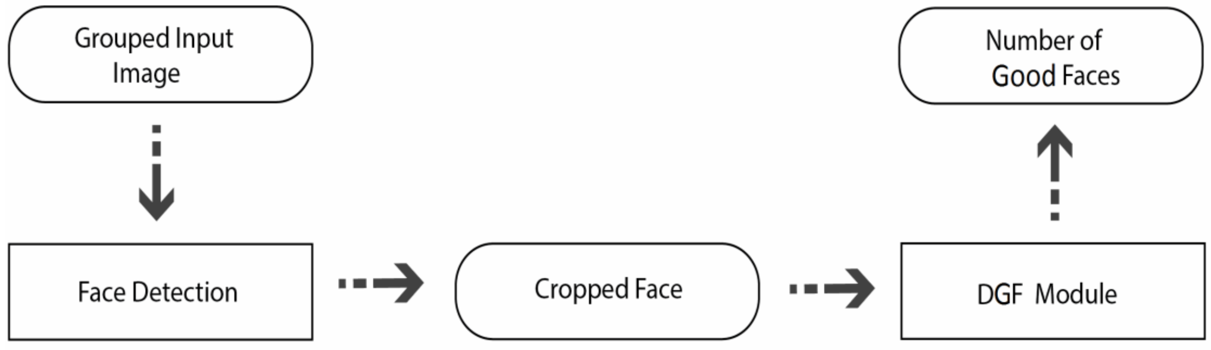

Fig. 6. Application flowchart

Fig. 7 shows the detected faces in a picture by the algorithm. Faces highlighted using red boxes are classified as good, for which, the algorithm generated an output of 1. So, by counting the number of 1s, the total number of good faces in the picture were computed. In Fig. 7, there are 18 people, where the total number of good faces are 11.

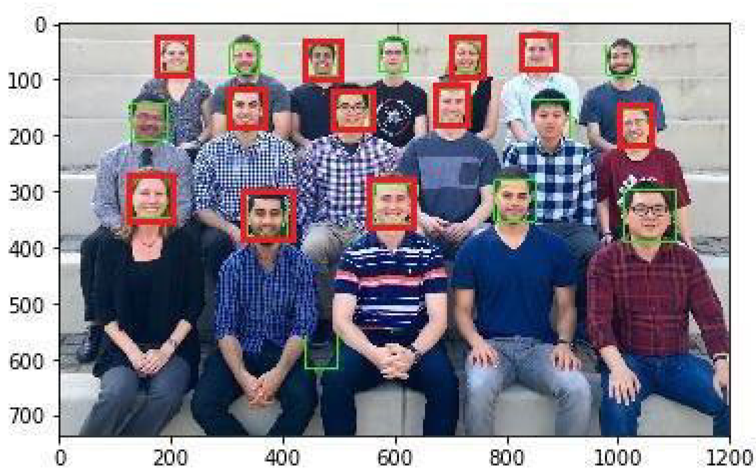

Fig. 7. Red boxes shows the good faces

6 Conclusion

The approach presented here is differed from other computer based picture retrieval techniques and facial expression recognition systems. In this paper we focused on classifying the images into two categories good or not rather than classifying facial expression categories. Users are no longer required to choose good pictures instead algorithm will perform this task.

In our current work the algorithm is limited to classify pictures based on facial expression only. In future work we want to consider body gestures and background. We also want to investigate the problem of choosing one good picture among multiple pictures having same facial expression. It is possible that a person have many pictures with same facial expression and he wants to choose one good picture among them. In this case we want to explore the issue and optimal solution for it. We are looking forward to utilize this algorithm in smart phones. So, when people make photos in their daily life they will be able to get suggestions from our algorithm whether the picture they taken is good or not.

Acknowledgements

This paper was supported by Kumoh National Institute of Technology.

References

[1] M.S. Bartlett, J.C. Hager, P. Ekman, T.J. Sejnowski, Measuring facial expressions by computer image analysis, Psychophysiology 36(2)(1999) 253-263.

[2] A.J. O'Toole, S.Z. Li, A.K. Jain, Psychological and neural perspectives on human face recognition, in: S.Z. Li, A.K. Jain (Eds.), Handbook of Face Recognition, Springer, New York, 2005, pp. 349-369.

[3] I. Cohen, N. Sebe, A. Garg, L.S. Chen, Facial expression recognition from video sequences: temporal and static modeling, Computer Vision and image understanding 91(1-2)(2003) 160-187.

[4] K. Rodden, How do people organise their photographs?, in: Proc. BCS-IRSG Annual Colloquium on IR Research 1999, 1999.

[5] A. Girgensohn, J. Boreczky, L. Wilcox, Keyframe-based user interfaces for digital video, Computer 34(9)(2001) 61-67.

[6] D. Keltner, P. Ekman, G.C. Gonzaga, J. Beer, Facial expression of emotion, in: R.J. Davidson, K.R. Scherer, H.H. Goldsmith (Eds.), Series in Affective Science. Handbook of Affective Sciences, Oxford University Press, New York, 2003.

[7] P. Ekman, Facial expression and emotion, American Psychologist 48(4)(1993) 384.

[8] M. Suwa, N. Sugie, K. Fujimora, A preliminary note on pattern recognition of human emotional expression, in: Proc. International Joint Conference on Pattern Recognition, 1978.

[9] B. Fasel, J. Luettin, Automatic facial expression analysis: a survey, Pattern Recognition 36(1)(2003) 259-275.

[10] A. Savoiu, J. Wong, Recognizing facial expressions using deep learning. <http://www.semanticscholar.org/paper/Recognizing-Facial-Expressions-Using-Deep-Learning-Savoiu-Wong/59b504e396e657627cc503a047475b>, 2017.

[11] X.-W. Chen, T. Huang, Facial expression recognition: a clustering-based approach, Pattern Recognition Letters 24(9)(2003) 1295-1302.

[12] D. Kestenholz, The psychology of photography. <http://the.me/the-psychology-of-photography/>, 2012.

[13] Anastasia, Everything you need to know about the psychology of choice. <http://www.cleverism.com/psychology-of-choice/>, 2015.

[14] M. Thein, Photography and psychology: it's all a mind game. <http://www.huffingtonpost.com/ming-thein/photography-and-psychology-part-1_b_4380400.html>, 2013.

[15] A.K. Jain, S.Z. Li, Handbook of Face Recognition, Springer, New York, 2011.

[16] L. Lim, The science of facial expressions and how it can improve your photography. <http://www.photographyconcentrate.com/science-facial-expressions-and-how-it-improves-your-photography/>, 2017.

[17] L.R. Rubin, Y. Mishriki, G. Lee, Anatomy of the nasolabial fold: the keystone of the smiling mechanism, Plastic and Reconstructive surgery 83(1)(1989) 1-8.

[18] F. Chollet, Keras. <http://www.github.com/fchollet/keras>, 2015.

[19] R. Al-Rfou, G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas, F. Bastien et al., A Python framework for fast computation of mathematical expressions. <http://arxiv.org/abs/160502688>, 2016.

[20] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Proc. the 27th International Conference on Machine Learning (ICML-10), 2010.