J Electr Eng Technol.2018; 13(1): 485-492
http://doi.org/10.5370/JEET.2018.13.1.485

ISSN(Print) 1975-0102
ISSN(Online) 2093-7423

Facial Expression Classification Using Deep Convolutional Neural Network

In-kyu Choi*, Ha-eun Ahn* and Jisang Yoo†

Abstract – In this paper, we propose facial expression recognition using CNN (Convolutional Neural Network), one of the deep learning technologies. The proposed structure has general classification performance for any environment or subject. For this purpose, we collect a variety of databases and organize the database into six expression classes such as 'expressionless', 'happy', 'sad', 'angry', 'surprised' and 'disgusted'. Pre-processing and data augmentation techniques are applied to improve training efficiency and classification performance. In the existing CNN structure, the optimal structure that best expresses the features of six facial expressions is found by adjusting the number of feature maps of the convolutional layer and the number of nodes of fully-connected layer. The experimental results show good classification performance compared to the state-of-the-arts in experiments of the cross validation and the cross database. Also, compared to other conventional models, it is confirmed that the proposed structure is superior in classification performance with less execution time.

Keywords: Convolutional neural network, Facial expression, Data augmentation, Database

1. Introduction

Computers are not only an important part of human daily life. In addition, they also offer convenience in various forms. In the future, the intimacy and interaction between computers and humans will continue to increase. As a result, research on Human-Computer Interaction(HCI) has been conducted in various disciplines such as ergonomics, industrial engineering, psychology and computer science. For a natural interaction between users and computers, the computer must comprehensively judge the user's intention and react accordingly. Emotion is the most important factor that can be expressed in form of the psychological state of the human being. In order to maximize the satisfaction of the user, the emotion recognition of the user is most important. One of the important means of expressing emotions is facial expressions, and therefore techniques for classifying facial expressions are also important.

With the development of hardware and the construction of big data, deep learning technology which learns the suitable pattern for the purpose itself has attracted the attention. Deep learning can overcome the technical limitations of existing machine learning where the performance is drastically deteriorated for complex problems by using deep neural networks to extract high-level features appropriate to the given data. Convolutional Neural Networks(CNN) are developed to imitate human visual cognition processes in deep learning technology and has been widely applied to the field of image recognition and shows high performance.

In the annual ILSVRC (ImageNet Large Scale Visual Recognition Competition), the top teams with good grades since 2012 have used CNN-based techniques. Microsoft Research's ResNet which won the 2015 competition, reduced the top5 error rate to 3.54% for a 1000-class classification problem[1]. In addition, Facebook's Deep Face which is a face recognition algorithm based on CNN, has an accuracy of 97.25% which is close to human average accuracy that is (97.53%) [2].

The basic CNN structure consists of convolutional layers and fully-connected layers. By passing through multiple convolutional layers sequentially, it is found to extract high-level features. With the extracted high-level features, the final classification result is determined in the Fully-connected layer.

Facial expression recognition using CNN has been studied extensively. In order to improve facial recognition performance, a study has been conducted combining a network extracting temporal appearance features from image sequences and a network extracting temporal geometry features from temporal facial landmark points [3]. A simple solution using a combination of the convolution neural network and a specific image preprocessing step was also proposed for facial expression recognition [4]. In addition, research has been carried out to apply the inception module, which has good performance in object recognition [5]. However, the above-mentioned studies show a high accuracy for a specific database, but since they use a small database, it is necessary to conform that whether they have general classification performance or not. Studies using various databases have also been conducted, but the classification performance is significantly lowered in the cross database experiment [6].

† Corresponding Author: Dept. of Electronic Engineering, Kwangwoon Univerity, Korea. (jsyoo@kw.ac.kr)
* Dept. of Electronic Engineering, Kwangwoon Univerity, Korea. ({cig2982, mysco226}@kw.ac.kr)
Received: May 1, 2017; Accepted: October 24, 2017

In order to recognize facial expressions based on CNN, a much amount of well-separated training database is needed. In this paper, 10 different databases were collected to form a well-classified high quality database for each facial expression. and each database is as follows. Amsterdam Dynamic Facial Expression Set(ADFES) [7], Chicago Face Database [8], Cohn-Kanade AU-Coded Facial Expression (CK+)[9], EU-Emotion Stimulus Set [10], ESRC 3D Face Database [11], FACE DATABASE [12], Karolinska Directed Emotional Faces (KDEF) [13], Radboud Faces Database [14], Web Search Database and Warsaw Set of Emotional Facial Expression Pictures (WSEFEP) [15]. And we reorganize the database by selecting facial expressions that are commonly included in 10 different databases. The 6 expressions to be classified are 'neutral', 'happy', 'sad', 'angry', 'surprised' and 'disgusted'. CNN architecture with smaller execution time and training parameters should be designed. Two experiments of subject-independent cross-validation and cross-database are used to evaluate whether the proposed CNN structure has high classification performance with generalization.

The paper is organized as follows. Section 2 introduces the process of collecting database, and data preprocessing and augmentation techniques for improving classification performance. Next, section 3 describes the process of designing an optimal training structure to reduce the execution time and at the same time, improving the classification performance. Section 4 evaluates the classification performance of the proposed structure through two experiments of subject-independent cross-validation and cross-database. And compares the execution time and classification accuracy with other CNN models. Finally, Section 5 summarizes the article and concludes the article.

2. Database Configuration

2.1 Database collection

First, a database containing a large amount of facial images is required for recognition of facial expressions with high accuracy. The images of the database should consist of facial images representing emotions. The database used in the 'Facial Expression Recognition Challenge' held in Kaggle in 2013(FER2013) consists of 37,000 facial images of seven facial expressions[16]. However, the image resolution is very low at 48x48 pixels and contains incorrectly labeled images. If a CNN structure is designed for a low-resolution input image, the high-resolution input image must be resized to fit the structure. In this process, the classification performance is decreased because the image ratio is altered and blur occurs. Also incorrect labeling deteriorate the classification performance.

Fig. 1 shows some of the FER2013 database that were mislabeled with different facial expressions. To overcome

Fig. 1. Examples of images classified as incorrect faces in FER 2013 database

these problems, we use the following ten high quality databases published by universities and research institutes around the world.

① Amsterdam Dynamic Facial Expression Set (ADFES): It is a rich stimulus set of 648 filmed emotional expressions. It features the displays of nine emotions: the six basic emotions (anger, disgust, fear, joy, sadness, and surprise), as well as contempt, pride and embarrassment. Expressions are displayed by 22 models (10 female, 12 male). It included North-European and Mediterranean models.

② Chicago Face Database(CFD): It includes neutral face images for 597 people between 17 and 65 years of age. It consists of diverse races, and includes face images of 'happy', 'angry', and 'fear' for 158 people.

③ Cohn-Kanade AU-Coded Facial Expression(CK+): It includes 593 video sequences for 123 people between 18 and 30 years of age. Among them, 309 sequences show expressions of 'happiness', 'sadness', 'angry', 'surprise', 'fear', 'disgust' and 'contempt'.

④ EU-Emotion Stimulus Set: It consists of expressions such as 'neutral', 'happiness', 'sadness', 'angry', 'surprise', and 'disgust' for 19 actors between the ages of 10 and 70. The actors were employed in drama schools and professional acting agencies in the UK.

⑤ ESRC 3D Face Database: It includes images taken at various angles and lighting using four cameras for 45 males and 54 females. It consists of the expressions of 'happiness', 'sadness', 'angry', 'surprise', and 'disgust'.

⑥ Lifespan database: A database of 575 individual faces ranging from ages 18 to 93. It was developed to be more representative of age groups across the lifespan, with a special emphasis on recruiting older adults

⑦ Karolinska Directed Emotional Faces(KDEF): It includes 4,900 images taken at five angles of -90°, -40°, 0°, +45° and +90° for 35 women and 35 men between 20 and 30 years of age. It consists of seven expressions of 'neutral',

'happiness', 'sadness', 'angry', 'surprise', 'fear', and 'disgust'.

⑧ Radboud Faces Database(RaFD): It is a set of pictures of 67 models (including Caucasian males and females, Caucasian children, both boys and girls, and Moroccan Dutch males) displaying 8 emotional expressions. The RaFD is a high quality faces database, which contain pictures of eight emotional expressions : Anger, disgust, fear, happiness, sadness, surprise, contempt, and neutral. Each emotion was shown with three different gaze directions and all pictures were taken from five camera angles simultaneously.

⑨ Web Search Database: A database acquired through web search.

⑩ Warsaw Set of Emotional Facial Expression Pictures (WSEFEP): It contains 210 of high-quality pictures of 30 individuals. They display six basic emotions (enjoyment, fear, disgust, anger, sadness, surprise) and neutral display

The type of facial expression to be recognized is selected by the number of facial expressions that are similar to each database and is as follows: neutral, happy, sad, angry, surprised and disgusted. The remaining facial expressions not mentioned above are excluded because of their small amount of data. Each facial expression data is selected for various races regardless of gender and age.

2.2 Data preprocessing and augmentation

Since human facial expression recognition requires only facial region information, it is necessary to preprocess to detect and cut only the face region of the training image. In this paper, we use a method based on Haar feature to detect and cut face region [17]. Based on our experience of recognizing facial expressions regardless of skin color, facial expression recognition through CNN also needs to be performed irrespective of the color information of the input image. In this paper, we simply convert the input image into a 1-channel gray image and do not use techniques such as intensity normalization.

Fig. 2 shows the result of detecting and cutting out the face region from the original image and converting it to gray image.

If the number of training images is insufficient compared

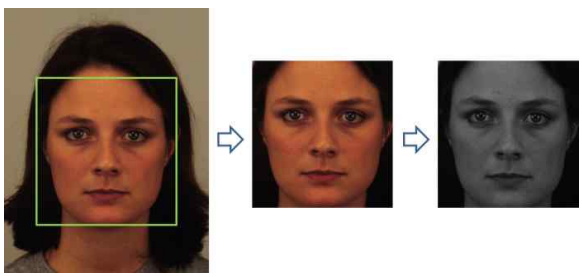

Fig. 2. The result of converting cut-out face region image into gray image

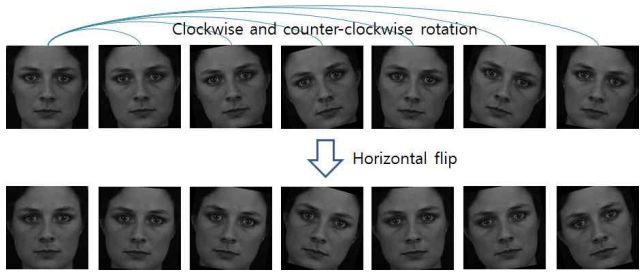

Fig. 3. The result of applying data augmentation technique

to the training parameters of CNN, over-fitting problem may occur and classification performance is reduced. To solve this problem, data augmentation technique which increases the number of training images is used. As shown in Fig. 2, the facial images of the database are mostly taken with the neck straightened to upward direction. In real life, facial images may change depending on the position of the camera or the posture of the person, so these points are considered in determining the data augmentation technique.

First, images obtained by rotating the reference image by 5°, 10° and 15° in the clockwise direction and after while, counterclockwise direction respectively are acquired. Then, the rotated images and the reference image are horizontally flipped to make them a total of fourteen images. By generating synthetic images with simple rotating and flipping operations, the amount of the database is increased and training with this database improves the generalization of the CNN model. Fig. 3 shows the result of applying the data augmentation technique.

3. Proposed CNN Architecture

The architecture of proposed CNN (Convolutional Neural Network) is represented in Fig. 4. As depicted in Fig. 4, the net contains eight layers. The first five are convolutional layer (C1-5) and the remaining three are fully-connected layer (FC6-8). The output of the last fully-connected layer is supplied to a 6-way softmax which produces a distribution over the 6 class labels. Max-pooling layers follow first, second and fifth convolutional layer. The ReLU(Rectified Linear Unit) non-linearity is applied to the output of every convolutional and fully-connected layer.

The first convolutional layer filters 227×227 input image

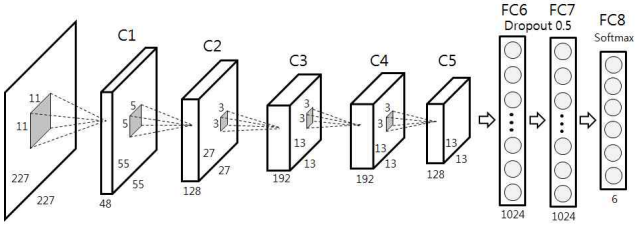

Fig. 4. The proposed CNN architecture

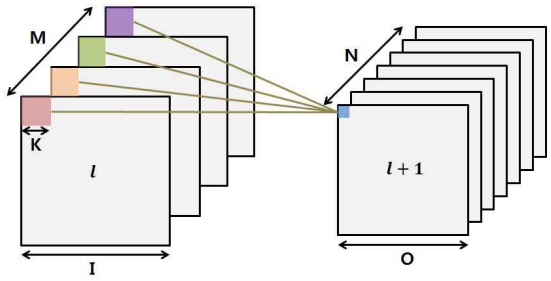

Fig. 5. Computational relationship between consecutive convolutional layers

with 96 kernels of size 11×11 with a stride of 4 pixels. The second convolutional layer takes as input the output of the first convolutional layer and filters it with 128 kernels of size 5× 5×48. The third, fourth, and the fifth convolutional layers are connected to one another without any intervening pooling or normalizing layers. The third convolutional layer has 192 kernels of size 3×3×128 connected to the outputs of the second convolutional layer. The fourth convolutional layer has 192 kernels of size 3×3×192, and the fifth convolutional layer has 128 kernels of size 3×3×192. The fully-connected layers have 1024 neurons each. To prevent over-fitting, the dropout technique is applied to the first two fully-connected layers [18].

The proposed structure is similar to AlexNet [19]. In the facial expression recognition, the number of channels in the convolutional layer and the number of nodes in the fully-connected layer are reduced in order to select an optimal structure with superior classification performance, less execution time and less training parameters. As we can see in Fig. 5, when the number of channels of consecutive convolutional layers are reduced, the amount of computation is quadratically reduced. Thus, the number of channels of each convolutional layer is reduced by 1/2, which greatly reduces the amount of computation. Unlike AlexNet, which is classified into 1000 classes, the proposed structure should classify only six expressions. Therefore, the number of nodes in fully-connected layer can be greatly reduced by 1/4. In Section 4, we compare the performance of the proposed structure with that of AlexNet.

4. Experiments and results

The implementation of pre-processing steps was done using OpenCV and Python. All the experiments were carried out using a GPU based CNN library(Theano)[20] with an Intel Core i5 3.4 GHz and a NVIDA GeForce GTX 980 Ti CUDA Capable that has 6 GB of memory in the GPU. The environment of the experiment is Window7, with the NVIDIA CUDA Framework 7.0 and the cuDNN library installed.

The batch size is of 128 and the training parameters are updated using back propagation method of stochastic gradient descent with momentum 0.9. The total epoch is 60 and the learning-rate is reduced from the initial value 0.01 by 1/10 when the epoch is 20 and 40, respectively.

Table 1. Accuracy comparison of data augmentation techniques

Preprocessing	Accuracy (%)
Without Data augmentation	89.07
With Data augmentation	93.95

Based on the empirical opinion that color information is not important for face expression recognition, 3-channel color image cut out after face area detection is converted into 1-channel gray image as described in section 2. The preprocessed gray training images are rotated at angles of –15 °, -10 °, -5 °, + 5 °, + 10 °, and + 15 °, and they are flipped horizontally and finally we obtain fourteen images. Then, we compare the facial expression recognition accuracy with and without data augmentation. Table 1 shows the effect of augmentation techniques on the recognition accuracy in the optimal structure (48, 128, 192, 192, 128, 1024, 1024, 6). As shown in Table 1, when the data augmentation technique is applied, the recognition accuracy increases significantly.

We have evaluated the accuracy of the proposed deep neural network architecture in two different experiments such as cross validation and cross database. In the cross validation experiment, we used the K-fold cross validation technique with K=10. The database was separated in 10 groups without subject overlap between the groups. This methodology ensures the generalizability of classifiers.

Table 2 gives the average accuracy when classifying the images into the six expressions. The average confusion matrix on cross validation experiments can be seen in Table 3. In Table 2, the results of the cross validation experiment are better compared to the current state-of-the-arts. In already published papers, there are no results using CFD, EU-Emotion Stimulus Set, ESRC and WSEFEP database. Web Search is also not a comparable database because each facial expression image is retrieved and acquired individually. As a result of the strict subject-independent cross validation experiment, it can be seen that the proposed structure has excellent general classification

Table 2. Average Top 1 Accuracy(%) on cross validation

Evaluation database	Proposed	State-of-the-arts
ADFES	100.00	96.30[21]
CFD	97.64	-
CK+	96.83	96.76[4]
EU-Emotion Stimulus Set	79.17	-
ESRC	84.76	-
FACE DATABASE	91.29	89.81[22]
KDEF	91.79	89.00[23]
RafD	99.25	93.96[24]
Web Search	81.97	-
WSEFEP	96.11	-

Table 3. Average confusion matrix on cross validation (%)

	NE	HA	SA	AN	SU	DI
NE	97.15	0.65	1.36	0.58	0.19	0.06
HA	2.30	96.19	0.16	0.40	0.48	0.48
SA	5.18	0.43	87.19	4.60	1.15	1.44
AN	1.86	0.74	6.20	86.37	0.12	4.71
SU	1.93	0.64	0.26	0.13	96.79	0.26
DI	0.74	1.78	3.56	3.12	0.59	90.21

Table 4. Average Accuracy(%) on cross database

Evaluation database	Top-1 Accuracy	State-of-the-arts
ADFES	99.24	-
CFD	92.16	-
CK+	92.61	64.20[6]
EU-Emotion Stimulus Set	73.96	-
ESRC	72.11	-
FACE DATABASE	78.32	79.27[22]
KDEF	78.81	-
RafD	94.94	-
Web Search	82.79	-
WSEFEP	95.56	-

performance in terms of subject in the same environment (lighting or pose). Table 3 shows a high accuracy of at least 86.37% for the six expressions to be classified as the confusion matrix for the cross validation experiment.

In the cross-database experiment, one database is used for evaluation and the rest of databases are used for training the network. Cross-database is a difficult task because each database has different lighting, human posture, camera angle and emotional expression. Table 4 gives the average cross-database accuracy when classifying six expressions. Cross-database experiments have not been studied much yet. So compare only CK+ and FACE DATABASE. The result of [6] is the classification accuracy for seven facial expressions including 'Fear'. Although there are six expressions to classify, we can see that our results are superior.

The result of [22] is the classification accuracy using Support Vector Machine with Radial Basis Function kernel (SVM + RBF). However, the best results were obtained by changing the training database and classifier's parameters. Although we can expect our results to be improved if we find the optimal training database for the test database, this paper shows only comprehensive results of learning all the databases except the test database. Nevertheless, classification accuracy is a little low.

Through the cross validation and cross-database experiments, we can confirm that the proposed structure is suitable for the generalization of the facial expression recognition in the field.

For comparison with other CNN models, cross validation experiment was performed for AlexNet, VGGNet(11-layer) [25], OverFeat(fast model) [26] and CNN using inception module [5,6]. We measure the training and testing time when a batch passes through each model. Table 5 shows that the proposed structure takes much less in both training and test time.

Table 5. Training and testing time for each model (batch : 128)

Model	Training time (sec / batch)	Test time (sec / batch)
AlexNet	0.325	0.068
OverFeat	0.593	0.125
VGGNet	2.128	0.569
Inception Module[5]	0.413	0.125
Inception Module[6]	0.756	0.178
Proposed	0.131	0.027

Table 6. Accuracy(%) for each model

Model	Top-1 Accuracy
AlexNet	93.55
OverFeat	93.55
VGGNet	91.60
Inception Module[5]	93.55
Inception Module[6]	93.15
Proposed	93.95

We trained each CNN model from scratch using the same protocol which are used to train our own network (as opposed to fine-tuning an already trained network). Table 6 shows the classification accuracy of each model. It can be confirmed that the classification performance of the proposed structure is the best.

5. Conclusion

In this paper, we propose a deep CNN for automated facial expression recognition algorithm for six expressions of ‘neutral’, ‘happy’, ‘sad’, ‘angry’, ‘surprised’ and ‘disgusted’. The structure of the proposed algorithm has good generality and classification performance.

First, we collect a variety of well-classified, high-quality databases. Then, in order to remove unnecessary information, the face region is detected, cut and converted into a gray image of one channel. In the proposed algorithm, the data augmentation which increases the number of training images is applied to solve the over-fitting problem which degrades classification performance.

In the existing CNN structure, the optimal structure for reducing the execution time and improving the classification performance was determined by adjusting the number of feature maps in the convolutional layer and the number of nodes in the fully-connected layer. Experimental results confirmed the effectiveness of data preprocessing and augmentation techniques. Cross validation and cross database experiments showed that the proposed structure has better classification performance and better generality than other state-of-the-arts. In comparison with other CNN models, our proposed algorithm has a short execution time and excellent classification performance.

Appendix

See Tables 7-16.

Table 7. Confusion matrix on ADFES (%)

	NE	HA	SA	AN	SU	DI
NE	100.00	0.00	0.00	0.00	0.00	0.00
HA	0.00	100.00	0.00	0.00	0.00	0.00
SA	0.00	0.00	100.00	0.00	0.00	0.00
AN	0.00	0.00	0.00	100.00	0.00	0.00
SU	0.00	0.00	0.00	0.00	100.00	0.00
DI	0.00	0.00	0.00	0.00	0.00	100.00

Table 8. Confusion matrix on CFD (%)

	NE	HA	SA	AN	SU	DI
NE	99.33	0.17	0.17	0.17	0.00	0.17
HA	2.93	97.07	0.00	0.00	0.00	0.00
SA	-	-	-	-	-	-
AN	5.84	0.65	0.00	92.21	0.00	1.30
SU	-	-	-	-	-	-
DI	-	-	-	-	-	-

Table 9. Confusion matrix on CK+ (%)

	NE	HA	SA	AN	SU	DI
NE	-	-	-	-	-	-
HA	0.00	100.00	0.00	0.00	0.00	0.00
SA	0.00	0.00	92.86	7.14	0.00	0.00
AN	4.44	0.00	2.22	93.33	0.00	0.00
SU	0.00	0.00	1.20	0.00	98.80	0.00
DI	1.69	1.69	0.00	1.69	0.00	94.92

Table 10. Confusion matrix on EU-Emotion Stimulus Set (%)

	NE	HA	SA	AN	SU	DI
NE	88.24	0.00	11.76	0.00	0.00	0.00
HA	6.67	93.33	0.00	0.00	0.00	0.00
SA	13.33	0.00	73.33	13.33	0.00	0.00
AN	12.50	0.00	6.25	62.50	0.00	18.75
SU	6.25	6.25	0.00	6.25	81.25	0.00
DI	5.88	11.76	0.00	5.88	0.00	76.47

Table 11. Confusion matrix on ESRC (%)

	NE	HA	SA	AN	SU	DI
NE	-	-	-	-	-	-
HA	2.04	90.31	1.02	2.55	1.53	2.55
SA	1.64	0.00	81.97	11.48	3.28	1.64
AN	0.57	2.29	18.29	70.29	0.00	8.57
SU	0.00	1.06	0.00	0.00	97.87	1.06
DI	0.00	4.37	7.65	4.37	1.64	81.97

Table 12. Confusion matrix on FACE DATABASE (%)

	NE	HA	SA	AN	SU	DI
NE	95.87	1.77	1.38	0.39	0.59	0.00
HA	5.12	94.49	0.00	0.00	0.39	0.00
SA	26.56	3.13	62.50	1.56	1.56	4.69
AN	10.00	0.00	40.00	40.00	10.00	0.00
SU	10.53	2.63	1.32	0.00	85.53	0.00
DI	33.33	0.00	16.67	16.67	0.00	33.33

Table 13. Confusion matrix on KDEF (%)

	NE	HA	SA	AN	SU	DI
NE	94.29	0.00	4.29	1.43	0.00	0.00
HA	0.71	97.86	0.00	0.00	1.43	0.00
SA	4.29	0.71	90.00	1.43	0.71	2.86
AN	0.00	0.71	5.71	82.86	0.00	10.71
SU	4.29	0.00	0.00	0.00	95.71	0.00
DI	0.00	0.71	4.29	4.29	0.71	90.00

Table 14. Confusion matrix on RafD (%)

	NE	HA	SA	AN	SU	DI
NE	99.50	0.00	0.50	0.00	0.00	0.00
HA	0.00	100.00	0.00	0.00	0.00	0.00
SA	1.99	0.00	97.51	0.50	0.00	0.00
AN	0.00	0.00	1.49	98.51	0.00	0.00
SU	0.00	0.00	0.00	0.00	100.00	0.00
DI	0.00	0.00	0.00	0.00	0.00	100.00

Table 15. Confusion matrix on Web Search (%)

	NE	HA	SA	AN	SU	DI
NE	76.67	0.00	13.33	10.00	0.00	0.00
HA	3.70	92.59	0.00	0.00	0.00	3.70
SA	16.67	0.00	66.67	16.67	0.00	0.00
AN	0.00	0.00	7.14	78.57	0.00	14.29
SU	0.00	0.00	0.00	0.00	100.00	0.00
DI	6.25	0.00	18.75	12.50	0.00	62.50

Table 16. Confusion matrix on WSEFEP (%)

	NE	HA	SA	AN	SU	DI
NE	96.67	0.00	0.00	3.33	0.00	0.00
HA	0.00	100.00	0.00	0.00	0.00	0.00
SA	6.67	0.00	90.00	3.33	0.00	0.00
AN	0.00	0.00	0.00	96.67	0.00	3.33
SU	0.00	0.00	0.00	0.00	100.00	0.00
DI	0.00	0.00	0.00	6.67	0.00	93.33

Acknowledgements

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No.R0132-15-1005, Content visual browsing technology in the online and offline environments)

References

[1] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[2] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[3] H. Jung, S. Lee, J. Yim, S. Park and J. Kim, "Joint fine-tuning in deep neural networks for facial expression recognition," *Proceedings of the IEEE International Conference on Computer Vision*, 2015.

[4] AT Lopes, E de Aguiar and AF De Souza, "Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order," *Pattern Recognition*, vol. 61, pp.610-628, 2017.

[5] P. Burkert, F. Trier, M.Z. Afzal, A. Dengel and M. Liwichki, "Dexpression: Deep convolutional neural network for expression recognition," *arXiv preprint arXiv:1509.05371*, 2015.

[6] A. Mollahosseini, D. Chan and M.H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*, IEEE, 2016.

[7] J. Van der Schalk, S. T. Hawk, A. H. Fischer, and B. J. Doosje, "Moving faces, looking places: validation of the Amsterdam Dynamic Facial Expression Set (ADFES)," *Emotion*, vol. 11, pp. 907-910, 2011.

[8] D.S. Ma, J. Correll and B. Wittenbrink, "The Chicago face database: A free stimulus set of faces and norming data," *Behavior research methods*, vol. 47, no. 4, pp.1122-1135, 2015.

[9] P. Lucey, J. 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," *Computer Vision and Pattern Recognition Workshops* (*CVPRW*), *2010 IEEE Computer Society Conference on*, IEEE, 2010.

[10] H. O'Reilly, D. Pigat, S. Fridenson, S. Berggren, S. Tal, O. Golan, S. B"olte, S. Baron-Cohen and D. Lundqvist, "The EU-emotion stimulus set: a validation study," *Behavior research methods*, vol. 48, no. 2, pp. 567-576, 2016.

[11] ESRC 3D Face Database. http://pics.stir.ac.uk/ESRC/

[12] M. Minear and D.C. Park, "A lifespan database of adult facial stimuli," *Behavior Research Methods, Instruments, &Computers*, vol. 36, no. 4, pp. 630-633, 2004.

[13] D. Lundqvist, A. Flykt, and A.Öhman, "The Karolinska Directed Emotional Faces(KDEF)," *CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet*, 1998.

[14] O. Langner, R. Dotsch, G. Bijlstra, D.H. Wigboldus, S.T. Hawk and A. van Knippenberg, "Presentation and validation of the Radboud Faces Database," *Cognition and emotion*, vol. 24, no. 8, pp.1377-1388, 2010.

[15] M. Olszanowski, G. Pochwatko, K. Kuklinski, M. Scibor-Rylski, P. Lewinski and RK. Ohme, "Warsaw set of emotional facial expression pictures: a validation study of facial display photographs," *Frontiers in psychology*, vol. 5, no. 1516, pp.1-8, 2015.

[16] Learn facial expressions from an image. https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data

[17] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Computer Vision and Pattern Recognition, 2001, CVPR 2001. Proceedings of the 2001, IEEE Computer Society Conference on*, vol. 1, IEEE, 2001.

[18] N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929-1958, 2014.

[19] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, 2012.

[20] R. Al-Rfou, G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas and Y. Bengio, "Theano: A Python framework for fast computation of mathematical expressions," *arXiv preprint arXiv:1605.02688*, 2016.

[21] K. Sikka, T. Wu, J. Susskind and M. Bartlett, "Exploring Bag of Words Architectures in the Facial Expression Domain," *Computer Vision–ECCV 2012, Workshops and Demonstrations*, Springer Berlin/ Heidelberg, 2012.

[22] J. Bekios-Calfa, JM. Buenaposada and L. Baumela, "Revisiting linear discriminant techniques in gender recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp.858-864, 2011.

[23] M.J. Den Uyl and H Van Kuilenburg, "The FaceReader: Online facial expression recognition," *Proceedings of measuring behavior*, vol. 30, 2005.

[24] M. Ilbeygi and H. Shah-Hosseini, "A novel fuzzy facial expression recognition system based on facial feature extraction from color face images," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 1, pp. 130-146, 2012.

[25] K. Simonyan, and A. Zisserman. "Very deep convolutional networks for large-scale image recognition," *in Proc. International Conference on Learning Representations*, http://arxiv.org/abs/1409.1556, 2014.

[26] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus and Y. LeCun, "OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks," *In Proc, ICLR*, 2014.

[27] B. Sun, L. Li, G. Zhou and J. He, "Facial expression recognition in the wild based on multimodal texture features," *Journal of Electronic Imaging*, vol. 25, no. 6, pp,061407-061407, 2016.

[28] N. Mousavi, H. Siqueira, P. Barros, B. Fernandes and S. Wermter, "Understanding how deep neural networks learn face expressions," *Neural Networks*

(IJCNN), 2016 International Joint Conference on, IEEE, 2016.
[29] M. Z. Uddin, M. M. Hassan, A. Almogren, M. Zuair, G. Fortino and J. Torresen, "A facial expression recognition system using robust face features from depth videos and deep learning," *Computers & Electrical Engineering*, 2017.
[30] V. Mayya, R. M. Pai and M. M. Pai, "Automatic Facial Expression Recognition Using DCNN," *Procedia Computer Science*,k vol. 93, pp.453-461, 2016.
[31] Z. Meng, P. Liu, J. Cai, S. Han and Y. Tong, "Identity-Aware Convolutional Neural Network for Facial Expression Recognition," *Automatic Face & Gesture Recognition* (*FG 2017*), *2017 12th IEEE International Conference on*, IEEE, 2017.

In-Kyu Choi He received the B.S degree at department of electrical engineering, KwangWoon University, Wolgye-dong, Nowon-gu, Seoul 01897, Republic of Korea in 2014. He received the M.S. degree at department of electrical engineering, KwangWoon University, Wolgye-dong, Nowon-gu, Seoul 01897, Republic of Korea in 2016. He is currently pursuing the Ph.D degree at Kwangwoon University. His research interests include computer vision, image processing, signal processing and deep-learning.

Ha-eun Ahn He received the B.S degree at department of electrical engineering, KwangWoon University, Wolgye-dong, Nowon-gu, Seoul 01897, Republic of Korea in 2014. He received the M.S. degree at department of electrical engineering, KwangWoon University, Wolgye-dong, Nowon-gu, Seoul 01897, Republic of Korea in 2016. He is currently pursuing the Ph.D degree at Kwangwoon University. His research interests include computer vision, image processing and deep-learning.

Jisang Yoo He received the B.S degree at department of electrical engineering, Seoul national university, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea in 1985. He received the M.S. degree at department of electrical engineering, Seoul National University, 1, Gwanak-ro, Gwanak-gu, Seoul 08826, Republic of Korea in 1987. He received the Ph.D degree at department of electrical engineering, Purdue University, 610 Purdue Mall, West Lafayette, IN 47907, the United States of America in 1993. He is currently a professor with the department of electronics engineering, KwangWoon University, Wolgye-dong, Nowon-gu, Seoul 01897, Republic of Korea. His research interests include computer vision, image processing, signal processing and deep learning.