Proc. 1st International Conference on
Machine Learning and Data Engineering (iCMLDE2017)
20-22 Nov 2017, Sydney, Australia
ISBN: 978-0-6480147-3-7

High Efficiency Facial Expression Recognition based Active Semi-Supervised Learning

Kim Jin Woo[1], Minhaz Uddin Ahmed[1], Md Rezaul Bashar[2] and Phill Kyu Rhee[1]

[1] Department of Computer Engineering, Inha University, Incheon, South Korea

[2]Science,Technology and Management Crest, Sydney, Australia

Corresponding author's E-mail:pkrhee@inha.ac.kr

Abstract

Face expression recognition is an important area of computer vision research which has been extensively used for HCI (human computer interaction), identity confirmation, persons emotion recognition and different access control. It is difficult to recognize Human's facial expression in the real-world due to various circumstances, such as face dissimilarity in pose, ages, lighting condition, occlusion and face motion. Large amount of labeled and trained facial expression data can be one of the key idea. If the training dataset and test dataset collect from a particular person and environment the chances of getting higher accuracy is possible due to similar variation in facial expression. In this paper, we propose a method for gathering many facial expression data within an indoor environment and train them through ASSL(Active Semi-Supervised Learning) framework to gain high accuracy. We gather a large number of unlabeled facial expression data from intelligent technology Lab members of Inha University and BU-3DFE (Binghamton University 3D Facial Expression benchmark datasets. We train our initial model with the ASSL framework using deep learning network VGG(visual geometry group, University of Oxford) model. Our framework adopted MTCNN(Multi-task Cascaded Convolutional Networks) detector for face detection and we also modify the last two layer of VGG network for better performance. Repeat this entire process support us to get better performance improvement. Therefore using the ASSL method, we gain better performance and higher accuracy with less labor force. Our experimental result shows the high efficiency with various training data.

***Keywords*:** Expression recognition, emotion classification, face detection, convolutional neural network, face recognition.

1. INTRODUCTION

As technology advances, machines become increasingly important in modern society. Especially, in the area of simple labor, many machines are replacing human labor, and as machine learning technology develops, more and more complicated tasks are also becoming the domain of machines. But there are still some tasks where machines cannot perform well. The reason is because there is a crucial and large barrier

of emotion between man and machine. It is necessary to go beyond the high obstacle for smooth communication between human and computer.

The emotion of a person varies where facial expressions, voices, tone, gestures, words and phrases are also important factors. Therefore, machines need to observe and learn these factors in order to grasp human emotions. A person tries to understand other person's emotion by judge the surrounding situation and face expression. Whether it is for a short period of time but it is natural for every human being. But this is not a natural ability. It is because the human brain has learned how to grasp the emotions of the human beings little by little. People live in various situation and learn such process naturally whether the person is angry or the person who glares the eyes is unpleasant or the person who laughs wide open.

Yet machines do not follow the human brain's abilities. Early psychologist A. Mehrabian (1968) found that 7% of the person's emotions expressed in the language, 38% of the sentences delivered through the speech, and 55% of the emotions conveyed by the human facial expressions. Therefore, facial expressions can be very valuable information that can examine human behavior, consciousness, and mental activity. Thus, research has long been carried out to recognize facial expressions more automatically, efficiently, and accurately by Xiaoming Zhao, Shiqing Zhang (2016).

The research on facial expression recognition can be divided into two parts a) frontal facial expression recognition and b) non-frontal facial expression recognition. Among them, facial expression recognition is actively studied and has a high recognition rate. Especially since the convolution neural network is applied, the recognition rate goes up by Changxing Ding (2015).We propose a new method to obtain better recognition rate with less effort even change of environment does not affect the performance.

We organize this paper as follows. In Section 2, we explain the background information of incremental active semi-supervised learning with facial expression recognition system. In Section 3, we describe the facial expression recognition methods. In Section 4 and Section 5, we have explained details about the dataset and experimental environment with results. Finally, in section 6 we discuss the conclusions drawn from the experiments and future work.

2. BACKGROUND

A special kind of neural network known as CNN (convolution neural networks) has been widely used to solve face recognition problem with many different computer vision problems by Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton (2012). When training new data through CNN, overfitting problems are likely to occur. Overfitting problem such as where the learned model classifies well in the similar dataset but cannot classify well in new dataset. In addition over fitting has various drawbacks such as excessive training and tuning etc. However, the simplest solution is to create a training model using diverse and large amounts of training data so that new data can be classified well.

There are different types of face expression data coming out every day, it is almost impossible to perform supervise learning by labeling all these things. It requires too much labor and cost. Labeling a large volume of data is time consuming and tiring work for human but it is convenient for the computer due to repetitive work. On the other hand when overfitting problem occurs computer cannot tackle correctly. In order to solve this problem, a better learning approach is labeling new data using active semi-supervised learning method.

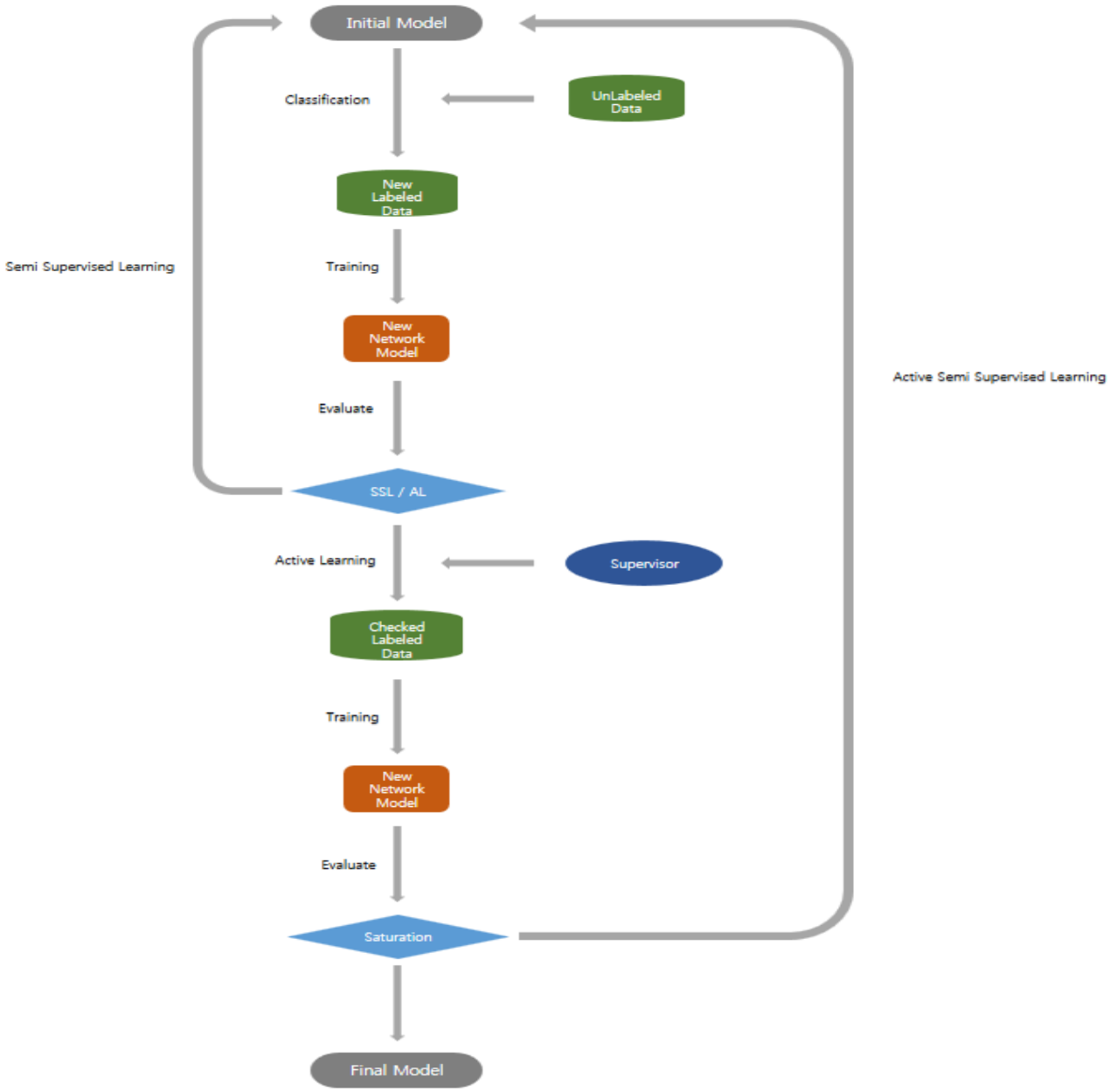

Figure 1: active semi-supervised learning model

Figure 1. shows the overall design of the propose system. First of all, we train the BU3DFE database to create a pre-learning model with excellent vector extraction capabilities because 2D-based analysis is difficult to handle large pose variation and slight facial expression recognition by Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm and Michael Reale (2008). Next, an initial learning model is created through active learning based on a pre-learning model. Subsequently, we use the semi-supervised learning with active learning where active learning reduces the wrong face identification and gradually improve the performance of the learning model. If the performance is lower than that of the previous training result, the training data is discarded and the next training data is used to perform the training again.

Active learning is performed when the result of the learning is not improved smoothly. We compare threshold value with estimated labeled score from face detector generate through semi-supervised learning. If the labeled score is below threshold value we do not consider for the active learning process. A confidence value is between 0 and 1 whereas threshold value is 0.9 because lower threshold value biases the outcome in our experiment.

3. METHODS

In facial expression recognition research, seven facial expressions are considered such as angry, disgust, fear, happy, sadness, surprise and neutral. However, different level of intensity increases the variation of facial expressions let alone these seven expressions. American psychologist Paul Ekman (1993) proposed the concept of basic six emotions, which depicts the six facets of anger, dislike, fear, happiness, sadness, and amazement by H. Rowley, S. Baluja, and T. Kanade(1995). In fact, the expression of a person is not only affect by emotion but also various factors such as appearance, age, sex, and culture. For example, in case of the West and the East, the expression is same but the emotion is different. Western people usually express joy as it is, but in case of oriental considers it humbly. This is a result of learning by a cultural difference between western and oriental. Also, young people express their emotion differently compared to elder people and handsome people are more confident than those who are unattractive.

The overall procedure of facial expression recognition is shown in Fig.1. First of all, a face image is capture by webcam as input then normalization and segmentation is performed. Normalization removes the changes of illumination or face location so that the facial image has uniform characteristics. Partitioning is a technique for extracting only a part of a given image including a necessary facial image. This preprocessing process has a great effect on accuracy because it plays a role in remove noise which obstructs recognition by B. Fasel, Juergen Luettin (2012), Andre Teixeira Lopes, Edilson de Aguiar, Alberto F. De Souza, Thiago Oliverira-Santos(2017).

We can easily distinguish the facial expression at a glance by looking at the face of an angry expression, a smiling expression or a surprised expression. The reason for this is that we can quickly identify the features of the raised eyebrows, tight lips, and wide mouths. The characteristic vector is the symbol of the characteristic of each facial expression so that the machine can recognize it like a person. Therefore, in order to recognize facial expressions, it is necessary to extract feature vectors that can be clearly distinguished by applying feature extraction techniques such as Deformation Extraction or Motion Extraction for the result of preprocessing adopted by B. Fasela and Juergen Luettin (2003). The transformation extraction is to express an image or a model including a geometric transformation of a facial expression as a feature vector, and an operation extraction is a feature vector expressing a change amount of a facial expression according to the passage of time.

The extracted feature vectors are classified by modifying the layer of VGG network so that they could be identify through facial feature representation. After that, classifiers are used to classify facial feature expressions according to the input labels. After that, the classifier has a value that can distinguish the facial feature representation for each label. The classifier classifies the extracted feature vectors when receiving a new face image with this value, where 1 denotes an angry expression, 2 denotes a disgusting expression, 3 denotes a fearful expression, 4 denotes a smiling expression, 5 denotes a sad expression, 6 denotes surprised facial expression, and 7 is expressionless.

Convolutional neural networks are one of the new machine learning schemes that have received attention in recent years due to better performance in resolving computer vision problems. The convolution neural network extracts and classifies the feature vectors according to the network structure determined by the user. CNN extracts appropriate feature vectors from each image and classifies them into the label.VGG very deep 16 network was used for learning facial expression recognition. The network adds a very small (3x3) convolution filter to the existing VGG network, resulting in better performance for large image recognition by Ken Chatfield, Karen Simonyan, Andrea Vedaldi, Andrew Zisserman (2014), Karen Simonyan, Andrew Zisserman (2015).

4. DATA SET

The BU3DFE database is used to create a pre-learning model for this experiment. The database includes diverse races, including West, East, East-Asian, Middle-east Asian, Indian, Hispanic Latino, and others. The BU3DFE dataset provides facial expressions for 100 people, ranging from 56% women and 44% men between the ages of 18 and 70. The types of facial expressions provided by Paul Ekman (1993) basic 6 emotions are added to the expression, there are seven labels of anger, disgust, fear, happiness, sadness, surprise and expressionless. There are a number of emotion levels by Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, Michael Reale (2008). We selected 700 images (7 x 100) of images with the strongest emotion level of the frontal pose and conducted preliminary learning.

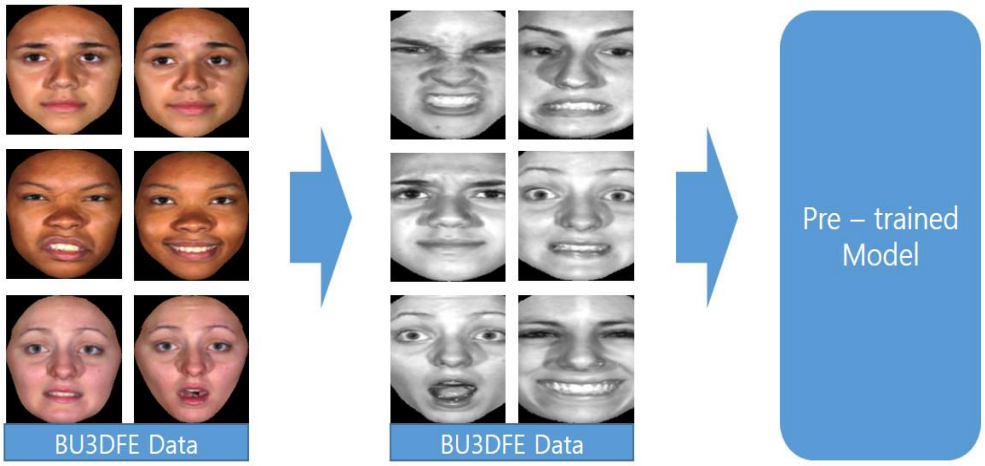

Figure 2: Pre-trained model

Figure 2. shows the BU3DFE dataset that use to transform pre-trained model before image preprocessing technique such as noise elimination and intensity standardization applied to increase the facial expression recognition rate. Then, the modified data is learned by the VGG-based deep learning network model, and an initial learning model for semi-supervised active learning method is created. In order to measure the performance, five experimental data consisting of 35 images is prepared separately for each emotion. Through this, the performance is continuously measured while learning through the active learning method, and the average is obtained. We have continuously measured the performance change of the previous learning model for the new environment by changing the experimental environment whenever performance converges in the process.

5. EXPERIMENTAL RESULT

Our experiments use publicly available VGG net that has 5 convolutional layers and 3 fully connected layers. This networks are pre-trained ILSVRC-2013 dataset by A. Berg, J. Deng, and L. Fei-Fei. (2010), which includes 1.2M training images , labeled into 1000 classes. For the detection system we use MTCNN by K. Zhang and Z. Zhang and Z. Li and Y. Qiao Joint (2016) that is convolutional neural network based face detector and ASSL frameworks is implemented on the popular deep learning library matconvnet and Caffe by Yangqing Jia,Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick Sergio Guadarrama and Trevor Darrel (2014). All implementations are on a single server with cuDNN S. Chetlur and C. Woolley (2014) and a single NVIDIA GeForce GTX 970 graphics card with operating system ubuntu 14.04.

We decide the order for six different environments. We use the initial learning model to measure the initial performance in each environment. Then we measure initial performance in ascending order to start learning from the environment with the lowest initial performance. Table 1 shows the order of data.

Table 1 Initial performance and order of data for 6 environments

Distinction	**Front**	**Left**	**Right**	**Glasses**	**Hat**	**Hat & Glasses**	**Mask**
Initial performance	71.4	51.4	85.7	34.2	94.2	65.7	31.4
Order	5	3	6	2	7	4	1

The reason for setting the learning sequence in this way, because of active learning starts from an environment with a high initial performance. If we start active semi-supervised learning from high initial performance environment than semi-supervised learning predict incorrect labels with high confidence. Therefore, to minimize this process, we gradually attempted to learn from the environment most similar to the data used to construct the initial learning model.

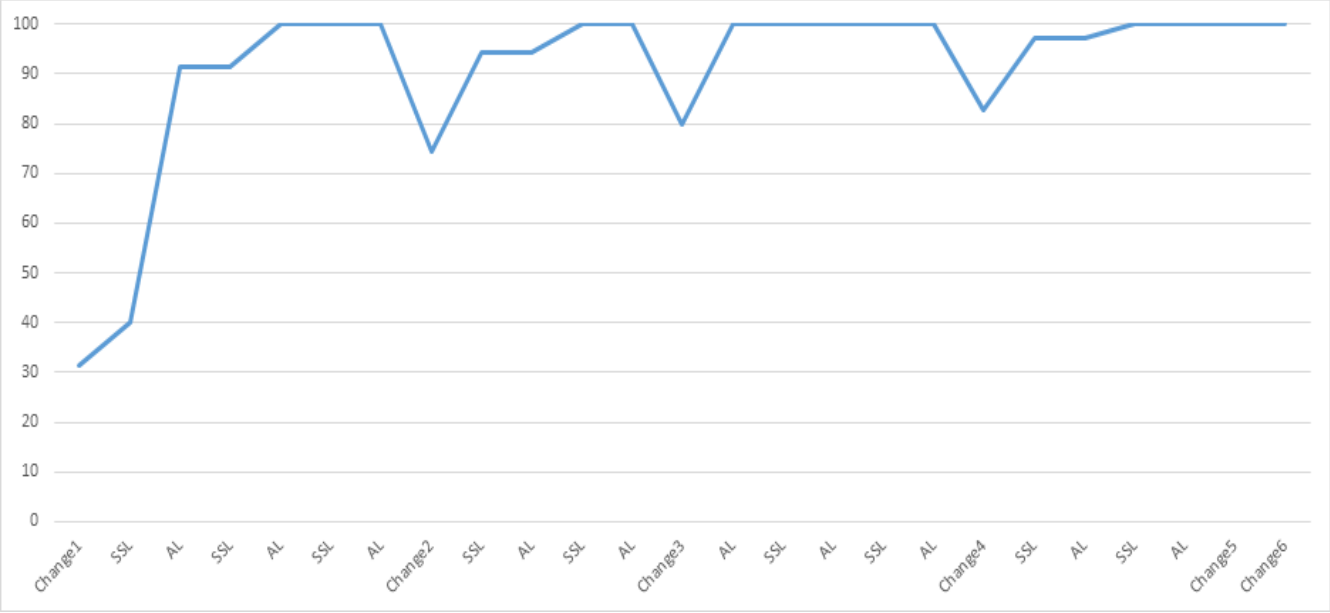

Figure 3: Performance of High-Efficiency Adaptive Facial Expression Recognition System by Active Learning

Experimental results show that the convergence of the initial performance decreases as the environment changes. This is because the learning of various environment data is automatically learned through active learning. Figure 3. shows the initial performance changes due to environmental changes.

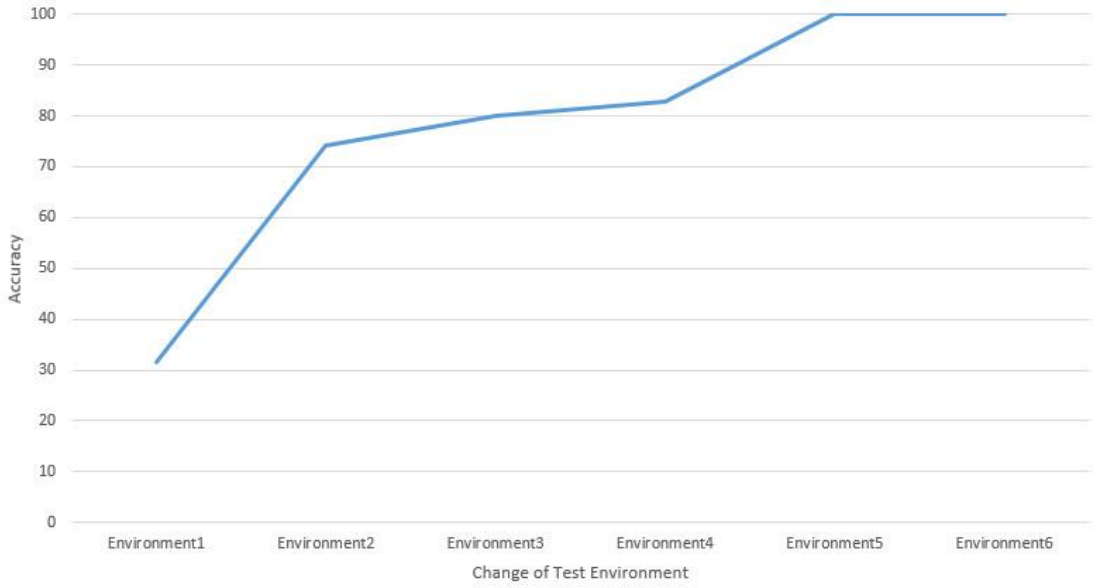

Figure 4: Initial recognition rate increase in graph due to environmental change

This experiment also requires much less manpower than labeling all the collected data. As shown in Figure 4, it can be seen that the labor required for manual labeling is 3.5 times more than the active semi-supervised learning method.

6. CONCLUSION

In this paper, we propose a semi-supervised learning framework for face expression recognition that can adapt to various environmental changes with the low labor force. In the proposed experiment, the undesirable side of this framework is label with wrong expression with high reliability which caused the problem for performance improvement due to the wrong learning result. However, by adjusting the good dataset order by comparing the initial performance, this problem can be solved by excluding the data from the experiment when the performance is not good. The proposed system is expected to help a lot when requires a large amount of training data in various environments. Our future research direction is to find a way to train sequence of face expression images with fewer images. In addition, by reducing the number of erroneous labels with high reliability can minimizing the discarded learning data and maximize efficiency.

7. ACKNOWLEDGEMENTS

This work was supported by the ICT R&D program of MSIP/IITP. [2017-0-00543], Development of Precise Positioning Technology for the Enhancement of Pedestrian's Position/Spatial Cognition and Sports Competition Analysis].The GPUs used in this research were generously donated by NVIDIA.

REFERENCES

A. Mehrabian (1968). Communication without words, Psychol. Today, Vol. 2, pp. 53_5

Xiaoming Zhao, Shiqing Zhang, (2016). A Review on Facial Expression Recognition - Feature Extraction and Classification, IETE Technical Review.

Changxing Ding (2015). Robust Face Recognition Via Multimodal Deep Face Representation,IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 17, NO. 11.

B. Fasel, Juergen Luettin (2012). Automatic facial expression analysis: a survey, International Journal of Computer Science & Engineering Survey (IJCSES), Vol.3, No.6.

H. Rowley, S. Baluja, T. Kanade (1995). Human face detection in visual scenes. Technical Report CMU-CS-95-158R, School of Computer Science, Carnegie Mellon University.

Andre Teixeira Lopes, Edilson de Aguiar, Thiago Oliveira-Santos (2015). A Facial Expression Recognition System Using Convolutional Networks, 28th SIBGRAPI Conference

Ken Chatfield, Karen Simonyan, Andrea Vedaldi, Andrew Zisserman (2014). Return of the Devil in the Details: Delving Deep into Convolutional Nets, arXiv:1405.3531v4 [cs.CV].

Karen Simonyan, Andrew Zisserman (2015).Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556v6 [cs.CV].

Cohn, D., Ghahramani, Z., I. Jordan M (1999). Active learning with statistical models. Journal of Artificial Intelligence Research, Vol.4, pp. 129-145,

Riccardi, G. and Hankkani-Tur, D (2005). Active learning: theory and applications to automatic speech recognition, IEEE Transactions on Speech and Audio Processing, Vol.13, No.4, pp. 504-511

Andre Teixeira Lopes, Edilson de Aguiar, Alberto F. De Souza, Thiago Oliverira-Santos (2017). Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order, Pattern Recognition Volume 61, Pages 610–628.

Gil-Jin Jang, Ahra Jo, Jeong-Sik Park, Yong-Ho Seo (2014). Video-based Facial Emotion Recognition using Active Shape Models and StatisticalPattern Recognizers, JIIBC, VOL. 14 NO. 3, pp.139-146.

Sinno Jialin Pan, Qiang Yang (2010). A Survey on Transfer Learning, IEEE Trans. Knowl. Data Eng. 22, 1345-1359

Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, Michael Reale (2008). A High-Resolution 3D Dynamic Facial Expression Database, The 8th International Conference on Automatic Face and Gesture Recognition.

K. Zhang and Z. Zhang and Z. Li and Y. Qiao Joint (2016). Face Detection and Alignment Using Multitask Cascaded Convolutional Networks, IEEE Signal Processing Letters

Steve Lawrence, C. Lee Giles, Ah Chung Tsoi, and Andrew D. Back (1997). Face Recognition: A Convolutional Neural-Network Approach, IEEE Transactions on Neural Network

Lijun Yin; Xiaochen Chen; Yi Sun; Tony Worm; Michael Reale (2008). A High-Resolution 3D Dynamic Facial Expression Database" The 8th International Conference on Automatic Face and Gesture

Ekman, Paul (1993). Facial expression and emotion. American Psychologist, Vol 48(4), 384-392.

B. Fasela and Juergen Luettin (2003). Automatic facial expression analysis: a survey, Elsevier, pattern recognition, 36 (2003) 259 – 275

A. Berg, J. Deng, and L. Fei-Fei. (2010). Large scale visual recognition challenge (ILSVRC), URL http://www.image-net.org/challenges/LSVRC/2010/.

Yangqing Jia,Evan Shelhamer, Jeff Donahue, Sergey Karayev, Jonathan Long, Ross Girshick

Sergio Guadarrama and Trevor Darrel (2014). Caffe: Convolutional Architecture for Fast Feature Embedding, ACM international conference on Multimedia

S. Chetlur and C. Woolley (2014). cuDNN: Efficient Primitives for Deep Learning, arXiv Prepr. arXiv. 1–9

Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton (2012) ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems 25