

PAPER • OPEN ACCESS

## Real Time Emotion Detection of Humans Using Mini-Xception Algorithm

To cite this article: Syed Aley Fatima *et al* 2021 *IOP Conf. Ser.: Mater. Sci. Eng.* **1042** 012027

View the [article online](#) for updates and enhancements.



The Electrochemical Society  
Advancing solid state & electrochemical science & technology



18th

### 239th ECS Meeting with IMCS18

DIGITAL MEETING • May 30-June 3, 2021

Live events daily • Free to register



Register now!

# Real Time Emotion Detection of Humans Using Mini-Xception Algorithm

Syed Aley Fatima, Ashwani Kumar\*, Syed Saba Raoof  
Research Scholar, Department of Computer Science and Engineering  
Vardhaman College of Engineering, Hyderabad, India.

\*Department of Computer Science and Engineering  
Vardhaman College of Engineering, Hyderabad, India.

[\\*ashwani.kumarcse@gmail.com](mailto:ashwani.kumarcse@gmail.com)

**Abstract.** In the past few decades there has been operating analysis on emotion expression recognition due to the great intra-class deviation also it is still challenging. Maximum number of research work performs the best on controlled datasets (i.e., small datasets with less features), whereas it fails to operate well and it's still challenging on datasets varies variations in images and even in partial faces. In modern years, many works have introduced an end-to-end plan for emotion expression recognition, utilizing deep learning models. Although emotion recognition is a great task, there still seems to be a huge area for development. In this paper, we developed a mini-Xception based on Xception and Convolution Neural Network (CNN), which is easy to concentrate on great parts like the face, and conclude important improvements to earlier works. We validated our model by creating a real-time vision system which accomplishes the tasks of face detection, and emotion classification simultaneously in one blended step using our proposed mini-Xception architecture. We still utilize a visualization technique that is ready to detect important face sectors because recognizing various emotions, based on the classifier's output. For experimental analysis we had used FER-2013 dataset and results manifest that the proposed method can efficiently perform all the tasks like detection and classification with seven different emotions using with Mini-Xception algorithm and achieved accuracy around 95.60%.

**Keywords:** *Mini-Xception, Emotion expression, Convolution neural network (CNN), Face detection.*

## 1. Introduction

"Emotion" is described by many research studies but there is no particular definition in all the literatures about the emotion. Reflection or actualization of feeling can be defined as "Emotion". It can be either sham or real not only feeling [1]. For instance, feeling of pain straight forwardly describe feeling, where as emotions cannot be felt completely and exactly. An emotion describes the inner esoteric internal state of affairs. In various fields of research like, psychology, health, biomedical engineering, and even in neuroscience emotion play a vital role and it has become an immense research subject. Emotion detection is an important research area in biomedical engineering [2]. Recent research studies in this field focus one motion detection and diagnosing psychological disorders by automated computer-aided systems. Various methods were implemented by researchers for emotion detection like multimodal, Galvanic Skin Response (GSR), Electro Encephalography (EEG), Facial Expression; visual scanning behavior etc. [3], over past few years, familiarization of deep learning has been a great progress in the field of image classification. Convolutional neural networks (CNNs) are one of the mostly used and popular deep learning algorithms used for implementation of image segmentation, recognition and classification.

Deep learning based algorithms are one of the mostly used methods to detect the state of emotion of human beings (e.g., anger, disgust, fear, happy, sad, surprise, and neutral) [4]. We had developed deep learning based algorithm Mini-Xception. The main aim of the proposed model is to



automatically detect emotions and to predict emotional condition with high accuracy. In this method, for analyzing experimental results we had used labeled facial expression images from FER dataset. These images are given as input to the model created and it is trained by these images. Then, the proposed model makes a determination which facial expression is performed.

This research article is organized as follows: Section 1 indicates the introduction of emotion detection with deep learning algorithm. Section 2 discusses related work. Section 3 describes methodology and proposed architecture. Section 4 explains about dataset, and experimental results. Section 5 describes conclusion and future work.

## 2. Related Work

Global survey has been done on DL approaches where various learning techniques, models, and recently proposed training approaches are discussed. Shervin Minaee [5] implemented a model on Facial expression recognition using Attention Convolution Network. FER-2013, CK+, FERF JAFFE dataset are used for training models and 93.3% accuracy is achieved. D Y Lilianasuntha et al. [6] developed a model on Emotion recognition from facial expression using deep Convolutional neural network. CK+ dataset are used for training models and 92.81% accuracy is achieved. Jyostna Devi Bodapati, N. Veeranjanyulu. [7], proposed a Facial Emotion Recognition Using Deep CNN Based Features. For feature extraction pre-trained convolution neural network model (VGG16) is used. CK+ dataset are used for training model and 86.04% accuracy is achieved. Nithya Roopa.s [8], in this paper they proposed an Emotion Recognition from Facial Expression using deep learning. The proposed model is Inception Net v3 is applied and Kaggle's Facial Expression Recognition Challenge and Karolinska Directed Emotional Faces (KDEF) datasets are used for training model and 39% accuracy is achieved. Panagiotis Giannopoulos [9], in this paper they proposed "Deep Learning approaches for Facial Emotion Recognition". In this paper they proposed, the performance of two known deep learning approaches (Google Net and Alex Net) on facial expression recognition. FER-2013 dataset are used for training models and 83% accuracy is achieved.

Junnan Li and Edmund Y. Lam. [10] in this paper they proposed a "Facial Expression Recognition using Deep Neural Networks". CK+ dataset are used for training and testing model and 91.9% accuracy is achieved. Arushi Raghuvanshi [11] "Facial Expression Recognition with Convolutional Neural Networks". VGG16, the model from Caffe's Model Zoo which was trained on Image Net, and VGG Face, a network trained on a facial recognition data set. They trained and tested these models on the data set from the Kaggle Facial Expression Recognition Challenge, and FER dataset are used for training and testing model and 0.60% accuracy is achieved. K. Sravanthi, G. Jaya suma [12] developed a method "A Prediction of Emotions for Recognition of Facial Expressions Using Deep Learning, CK+, JAFE and BU-3DFE dataset are used for training and testing model and 100% accuracy is achieved. VICTOR-EMIL NEAGOE [13] developed "A Deep Learning Approach for Subject Independent Emotion Recognition from Facial Expressions". They have focused on two "deep" neural models: Convolutional Neural Networks (CNN) and Deep Belief Networks (DBN). One has selected the Support Vector Machine (SVM) as a benchmark algorithm. JAFE dataset are used for training and testing model and 95.71% accuracy is achieved. Fatima Zahra Salman [14] proposed an "Emotion Recognition from Facial Expression Based on Fiducial Points Detection and using Neural Network". They detect and track 49 Fiducial points using a powerful and recent Supervised Decent Method (SDM). The proposed approach has achieved an emotion recognition accuracy of 99% on the CK+ database, 84.7% on the Oulu-CASIA VIS database, and 93.8% on the JAFFE database. The following table shows the comparative study of related work.

**Table I:** Summary of related work using different algorithm & methods.

S.NO.	Authors	Methods Used	Dataset Used	Accuracy Measures
1.	Shervin Minaee	Attention Convolution Network	FER-2013, CK+, FERG & JAFFE	93.3%
2.	D Y Lilianasuntha et al.	Deep Convolutional Neural Network.	CK+	92.81%
3.	Jyostna Devi Bodapati, N. Veeranjanyulu	Deep CNN Based Features, (VGG16).	CK+	86.04%
4.	Nithya Roopa.s	Deep Learning	Karolinska Directed Emotional Faces (KDEF)	39%
5.	Panagiotis Giannopoulos	Google Net and Alex Net.	FER-2013.	83%
6.	Junnan Li and Edmund Y. Lam	Deep Neural Networks.	CK+.	91.9%
7.	Arushi Raghuvanshi	Convolutional Neural Networks.	Kaggle Facial Expression Recognition and FER.	0.60%
8.	K.Sravanthi, G.Jaya suma	Deep Learning.	CK+, JAFE and BU-3DFE.	100%
9.	Victor-Emil Neagoe	Support Vector Machine (SVM).	JAFE.	95.71%
10.	Fatima Zahra Salman	Supervised Decent Method (SDM).	CK+, Oulu-CASIA VIS, JAFFE.	99%

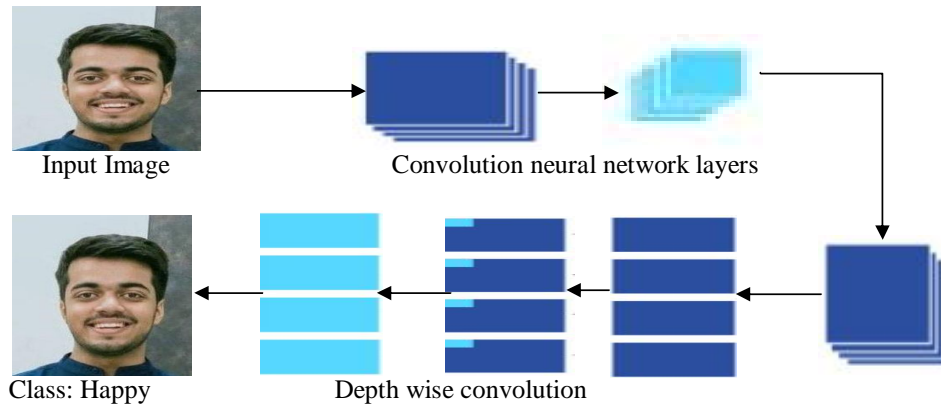
### 3. Proposed Method

We had implemented the proposed model based on improved deep learning approaches for emotion recognition, classification, and detection. An enhanced recognition model has been developed utilizing residual convolution networks for recognizing the emotions. Recognition is done through Mini-Xception based models. Comparison of the proposed model along with other models demonstrates greater performance against other algorithms. By using Xception model we had proposed a mini-Xception for emotion detection. Our proposed model is inspired by the Xception algorithm. This architecture combines the use of residual modules [6] and depth-wise separable convolutions. Residual modules modify the desired mapping between two subsequent layers, so that the learned features become the difference of the original feature map and the desired features. Consequently, the desired features  $H(x)$  are modified in order to solve an easier learning problem  $F(X)$  such that:

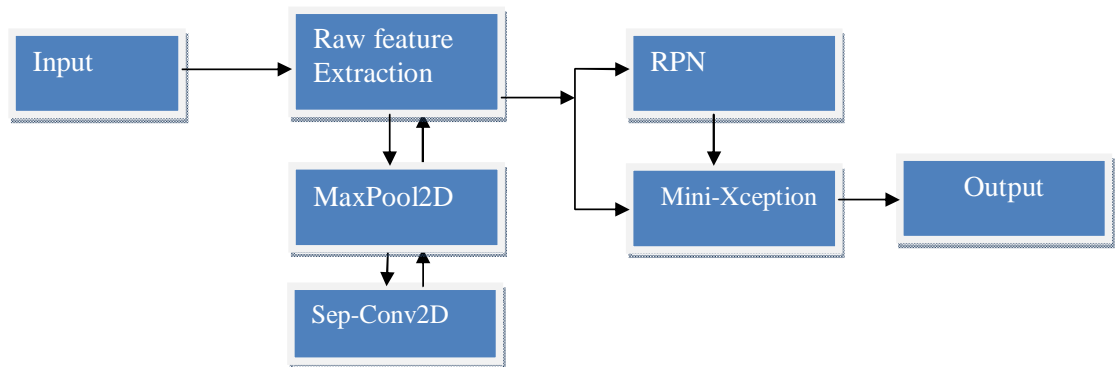
$$H(x) = F(x) + x \quad (1)$$

In proposed model parameters are reduced from the Convolutional layers, which was done by using depth-wise separable convolutions. Depth-wise separable convolutions are composed of two different layers: depth-wise convolutions and point wise convolutions.

The main purpose of these layers is to separate the spatial cross-correlations from the channel cross correlations [1]. They do this by first applying a  $D \times D$  filter on every  $M$  input channels and then applying  $N \times 1 \times M$  convolution filters to combine the  $M$  input channels into  $N$  output channels. Applying  $1 \times 1 \times M$  convolutions combines each value in the feature map without considering their spatial relation within the channel. Depth-wise separable convolutions reduces the computation with respect to the standard convolutions by a factor of  $1/N + 1/D^2$  [2]. A visualization of the difference between a normal Convolution layer and a depth-wise separable convolution. The proposed architecture contains 4 residual depth-wise separable convolutions where each convolution is followed by a batch normalization operation and a ReLUs activation function. The last layer applies a global average pooling and a soft-max activation function to produce a prediction. This architecture has approximately 60,000 parameters. Figure 1(a) (b) displays our complete architecture which is referred as mini-Xception.



**Figure 1(a).** Architecture of Proposed Model



**Figure 1(b).** Architecture of Proposed Model

We tested this architecture in the FER-2013 dataset and obtained the accuracy of 89.96% for the emotion detection task. Real-time image and video analysis is also done based on back-propagation. Given a mini-Xception with ReLU function for the intermediate layers, back propagation takes the derivative of every element  $(x, y)$  of the input image  $I$  with respect to an element  $(i, j)$  of the feature map  $f^L$  in layer  $L$ . The reconstructed image  $R$  filters all the negative gradients; consequently, the remaining gradients are chosen such that they only increase the value of the chosen element of the feature map. Following a fully ReLU mini-Xception reconstructed image in layer  $L$  is given by:

$$Rl_{i,j} = (Rl + 1 \text{ if } i,j > 0) * Rl + 1 \text{ if } i,j \quad (2)$$

### Algorithm

Description of our proposed model is mentioned below:

Method: Emotion Expression and Recognition using Mini-Xception.

**Input:** FER-2013 Dataset.

**Output:** Classification of Emotions.

- Step 1:** Load the FER-2013 dataset.
- Step 2:** Partition the dataset into training and testing.
- Step 3:** Apply the preprocessing techniques.
- Step4:** Build the model using Mini-Xception algorithm.
- Step5:** Test data is given to Mini-Xception for classification.
- Step6:** Calculate the Accuracy.

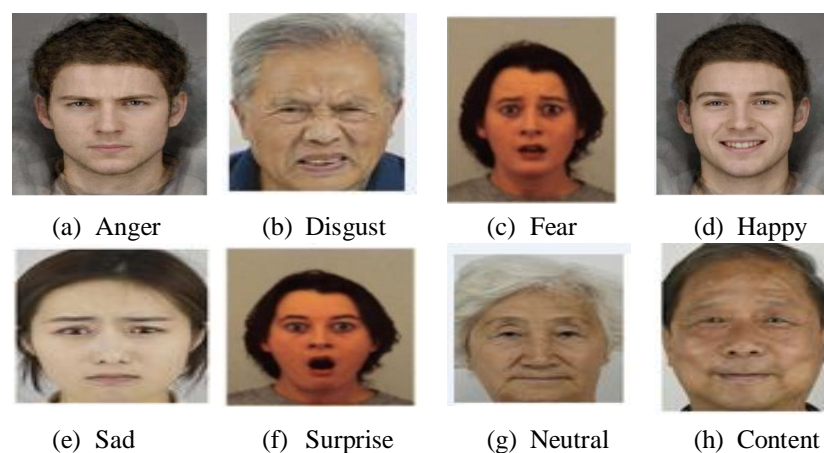
#### 4. Methodology

Initially the face region is extracted from the given face images using proposed face detection algorithm. Then from the cropped face regions, deep features are extracted from the last fully connected layer of mini-Xception. Based on the experimental results we observed that the proposed method is simple and gives better performance compared to the other methods. Mini-Xception, a pre-trained convolution model, [21] is known for its state of the art performance in various applications [21] like image classification, object detection etc. It is trained on Image Net dataset. After training of proposed architecture, the trained model was tested in real time. First of all, human faces were detected. After that, the detected images were sent to the model and the classes they belong to were queried. This process was performed on every 30 frames that occurred every second of the camera image obtained in real time.

#### 5. Experimental Analysis

The experimental analysis is evaluated using Python IDLE tool. Mini-Xception algorithm is applied on emotion detection dataset. Dataset contains 35,887 images. Accuracy is calculated from the confusion matrix.

*Dataset:* To train our models we used dataset: FER-2013. The FER-2013 dataset had been created by acquiring and combining the result of Google image search for every particular emotion [22]. Each and every image in FER dataset is labeled image and it consists of seven emotion images i.e., anger, disgust, fear, happy, sad, surprise, and neutral. This dataset consists of 35,887 images in the training set, 3,500 labeled images in the test set, and 3,500 images in the development set. It consists of pair of posed and un-posed identification images, these images are in grayscale and pixel value is 48x48. Even Though this dataset consists of labeled emotions, it was discovered that the proposed model achieved better accuracy. Here are some examples of the FER 2013 dataset recognized using proposed algorithm i.e., Mini-Xception. Figure 2 represents emotions like anger, disgust, fear, happy, sad, surprise, and neutral respectively.



**Figure 2.** Emotions recognized using Mini-Xception proposed algorithm.

Accuracy is calculated from the confusion matrix. For quantitative analysis of the experimental results, the following performance metrics are considered, including accuracy (AC). To do this we also use the variables True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Confusion Matrix: It is a particular table format that allows the performance evaluation of an algorithm [23-25].

**Table 2.** The following table shows confusion matrix of proposed algorithm.

Expression	Anger	Disgust	Fear	Happy	Sad	Surprise	Neutral
<b>Anger</b>	60.0	0.1	0.0	0.0	25.0	0.0	10
<b>Disgust</b>	0.0	55.0	0.5	0.7	11.9	25.0	0.0
<b>Fear</b>	12.9	0.01	65.78	0.5	0.12	17.8	0.0
<b>Happy</b>	0.02	0.5	0.02	95.55	0.08	14.5	0.78
<b>Sad</b>	25.0	11.9	0.12	0.08	85.0	0.05	7.50
<b>Surprise</b>	0.0	25.0	17.8	14.5	0.05	73.68	11.9
<b>Neutral</b>	10.0	0.0	0.0	0.78	7.50	11.9	89.66
<b>Average Accuracy =95.60</b>							

The overall accuracy is calculated using.

$$AC = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

**Table 3.** The following table demonstrates the performance measures calculated by using accuracy, precision, and recall.

Test size	Accuracy	Precision	Recall
90%-20%	95.60	93	90

## 6. Conclusion

In this paper, we implemented Mini-Xception model which is an enhanced model of Xception architecture using residual networks for Emotion expression and Recognition. Using the FER-2013 database gives better performance than the existing method. There are seven types of emotion expressions such as (e.g., anger, disgust, fear, happy, sad, surprise, and neutral.) we tried to recognize. In modern years, many works have introduced an end-to-end plan for emotion expression recognition, utilizing deep learning models. Although emotion recognition is a great task, it still seems emotion huge area for development. The accuracy obtained for Emotion expression and recognition using Mini-Xception is 95.60% and precision and recall rate is 93% and 90% respectively. Further the accuracy can be increased by training the Mini-Xception algorithm using the original image dimensions of 48x48. The number of Convolutional layers and the size of filter can also be increased to improve accuracy.

## References

- [1] S. Minaee and A. Abdolrashidi. Deep-emotion: Facial expression recognition using Attentional Convolutional network. arXiv preprint arXiv:1902.01019, 2019. 1, 2.1, 5.1, 5.2.
- [2] D Y Liliana [2019] "Emotion recognition from facial expression using deep Convolutional neural network". ". 2018 International Conference of Computer and Informatics Engineering (IC2IE)
- [3] Jyostna Devi Bodapati, N. Veeranjanyulu "Facial Emotion Recognition Using Deep Cnn Based Features", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8Issue-7, May 2019.
- [4] Nithya Roopa. S " Emotion Recognition from Facial Expression using Deep Learning", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-8 Issue-6S, August 2019.
- [5] Panagiotis Giannopoulos, Isidoros Perikos and Ioannis Hatzilygeroudis,"Deep Learning Approaches for Facial Emotion Recognition", Springer International Publishing AG 2018.



- [6] Junnan Li and Edmund Y. Lam, "Facial Expression Recognition using Deep Neural Networks. 978-1-4799-8633-0/15/\$31.00 ©2015 IEEE.
- [7] Arushi Raghuvanshi, Vivek Choksi, (2017) "Facial Expression Recognition with Convolutional Neural Networks". 978-1-7281-3783-4/20/\$31.00 ©2020 IEEE.
- [8] K.Sravanthi, G.Jaya Suma,"A Prediction of Emotions for Recognition of Facial Expressions Using Deep Learning", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-2S11, September 2019.
- [9] A VICTOR-EMIL NEAGOE, ANDREI-PETRU BĂRAR, NICU SEBE, PAUL ROBITU," A Deep Learning Approach for Subject Independent Emotion Recognition from Facial Expressions".978-1-7281-3783-4/20/\$31.00 ©2020 IEEE.
- [10] Fatima Zahra Salmam, Abdellah Madani, and Mohamed Kissi. "Emotion Recognition from Facial Expression Based on Fiducial Points Detection and using Neural Network", International Journal of Electrical and Computer Engineering (IJECE) Vol. 8, No. 1, February 2018, pp. 52 – 59.
- [11] Syed Aley Fatima, Ashwani Kumar, Ajay Pratap, Syed Saba Raoof, " Object recognition and detection in remote sensing images". 2020 International Conference on Artificial Intelligence and Signal Processing (AISP) 978-1-7281-4458-0/20/\$31.00 ©2020 IEEE.
- [12] Jiaxing Lia, Dexiang Zhanga, Jingjing Zhanga, Jun Zhanga, Teng Lia, Yi Xiaa, Qing Yana, and Lina Xuna. "Facial Expression Recognition with Faster R-CNN", International Congress of Information and Communication Technology (ICICT 2017).
- [13] Najmeh Samadiani, , Guangyan Huang , Borui Cai , Wei Luo , Chi-Hung Chi , Yong Xiang and Jing He "A Review on Automatic Facial Expression Recognition Systems Assisted by Multimodal Sensor Data". International Congress of Information and Communication Technology (ICICT 2017).
- [14] Mengyi Liu, Ruiping Wang, Shaoxin Li, Shiguang Shan, Zhiwu Huang, Xilin Chen "Combining Multiple Kernel Methods on Riemannian Manifold for Emotion Recognition in the Wild". Sensors 2019, 19, 1863; 10.3390/s19081863 www.mdpi.com/journal/sensors.
- [15] Asim Munira, Ayyaz Hussaina,d, Sajid Ali Khan, Muhammad Nadeema, Sadia Arshid "Illumination invariant facial expression recognition using selected merged binary patterns for real world images". <http://dx.doi.org/10.1145/2663204.2666274>.
- [16] Junkai Chen, Zenghai Chen, Zheru Chi, and Hong Fu "Facial Expression Recognition in Video with Multiple Feature Fusion", IEEE Transactions on Affective Computing • Optik 158 (2018) 1016–1025.
- [17] Gil Levi, Tal Hassner "Emotion Recognition in the Wild via Convolutional Neural Networks and Mapped Binary Patterns", November 9, 2015.
- [18] P. Shen, ShangfeiWang, and Zhilei Liu "Facial Expression Recognition from Infrared Thermal Videos", Intelligent Autonomous Systems 12, AISC 194, pp. 323–333.
- [19] André TeixeiraLopes, EdilsondeAguiar, AlbertoF.DeSouza, ThiagoOliveira-Santos "Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order", 19July2016.
- [20] Kaihao Zhang, Yongzhen Huang, Yong Du, and Liang Wang, "Facial Expression Recognition Based on Deep Evolutional Spatial-Temporal Networks", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 26, NO. 9, SEPTEMBER 2017.
- [21] Shangfei Wang • Chongliang Wu • Menghua He • Jun Wang • Qiang Ji "Posed and spontaneous expression recognition through modeling their spatial patterns", Machine Vision and Applications (2015) 26:219–231. ", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 26, NO. 9, SEPTEMBER 2017.
- [22] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. Image net classification with deep Convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012.
- [23] Andrej Karpathy. Cs231n Convolutional neural networks for visual recognition. <http://cs231n.github.io/convolutional-networks/>, 2015. [Online; Accessed 24 May 2016].
- [24] Andrej Karpathy, George Toderici, Sanketh Shetty, Thomas Leung, Rahul Sukthankar and Li Fei-Fei. Large-scale video classification with Convolutional neural networks. In Proceedings of International Computer Vision and Pattern Recognition (CVPR 2014), 2014.
- [25] Challenges in representation learning: Facial expression recognition challenge: <http://www.kaggle.com/c/challengesin-representation-learning-facial-expression-recognition>



challenge. ", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 26, NO. 9, SEPTEMBER 2017.

- [26] Liu, Ping, Shizhong Han, Zibo Meng, and Yan Tong. "Facial expression recognition via a boosted deep belief network." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1805-1812. 2014.