

# ## Title: Extended deep neural network for facial emotion recognition

JavaScript is disabled on your browser. Please enable JavaScript to use all the features on this page. [Skip to main content](#)[Skip to article](#)

ScienceDirect

\* Journals & Books

\* Help

\* Search

Gergo Gyori

IT University of Copenhagen

\* View \*\*PDF\*\*

\* Download full issue

Search ScienceDirect

## Outline

1. Highlights
2. Abstract
3. 4. Keywords
5. 1\ Introduction
6. 3\ Datasets and pre-processing
7. 4\ Proposed model
8. 5\ Experiment and evaluation
9. 6\ Conclusion
10. References

Show full outline

## Cited by (291)

## Figures (8)

1. 2. 3. 4. 5. 6.

Show 2 more figures

## Tables (3)

1. Table 1
2. Table 2
3. Table 3

## Pattern Recognition Letters

Volume 120, 1 April 2019, Pages 69-74

# Extended deep neural network for facial emotion recognition

Author links open overlay panelDeepak Kumar Jain a, Pourya Shamsolmoali b, Paramjit Sehdev c

Show more

Outline

Add to Mendeley

Share

Cite

<https://doi.org/10.1016/j.patrec.2019.01.008>Get rights and content

## Highlights

\* ?

A new Deep Fully Connected model for facial emotion recognition.

\* ?

The model evaluated based on multiple hyperparameters and real world datasets.

\* ?

The application of the model on the real-time events.

## Abstract

Humans use facial expressions to show their emotional states. However, facial expression recognition has remained a challenging and interesting problem in

computer vision. In this paper we present our approach which is the extension of our previous work for facial emotion recognition [1]. The aim of this work is to classify each image into one of six facial emotion classes. The proposed model is based on single Deep Convolutional Neural Networks (DNNs), which contain convolution layers and deep residual blocks. In the proposed model, firstly the image label to all faces has been set for the training. Secondly, the images go through proposed DNN model. This model trained on two datasets Extended Cohn+Kanade (CK+) and Japanese Female Facial Expression (JAFPE) Dataset. The overall results show that, the proposed DNN model can outperform the recent state-of-the-art approaches for emotion recognition. Even the proposed model has accuracy improvement in comparison with our previous model.

\* Previous article in issue

\* Next article in issue

## ## Keywords

Facial emotion recognition

Deep neural network

Fully convolution network

## ## 1\ Introduction

Facial expression is one of the most important features of human emotion recognition. As persons we can assume the impression of someone's emotions by observing their face. Facial emotion recognition has several applications in computer vision, non-verbal human behavior and human-computer interaction. It is a challenging task due to several issues for example similarity of actions, large head poses and etc. Recently, Facial emotion recognition in images has attracted growing attention [5], which is for instances more complicated due to the backgrounds and low-resolution faces. Emotion refers to the feeling, energy, and dispositional effects. The goal of this work is to classify the facial emotion as sad, angry, happy, surprise, fear, disgust, and neutral. Facial expression is one of the most powerful, natural and common signals for human beings to convey their emotional states and intentions.

It is computationally complex and challenging to achieve high recognition rates using conventional feature extraction and classification schemes. This paper proposed a new deep learning model for the emotional recognition, to classify facial emotion from the images. Several techniques have been established in this regards, but, most current works are appeared focusing on hand-engineered features [2], [3], [10]. Currently, due to the size and variety of datasets, deep neural network is the most appropriate techniques in all the image processing and computer visions tasks [4], [5], [8]. Deep conventional networks have the ability to simply handle spatial image [6], [16], [17]. The main commitment of this work is to propose a deep neural network model which contains several convolution layers and deep residual blocks for facial emotion recognition [14], [15]. The proposed model is able to learn the subtle features that discriminate the six different facial expressions [9], [11], [13]. The rest of the paper is organized as follow: next section delivers the related works. Section 3 presents the proposed network. The results and experiments are presented in Section 4. At the end, we have concluded our observation in Section 5.

## ### 2\ Related work

The researches in the area of facial emotion detection focused on recognizing human emotion based on image or video records. Some recent work pursue to recognize faces in images or video records [20], however, these approaches did not use neural network strategy to extract features from the images.

Based on the features extracted for the recognition, we can differentiate two fundamental methodologies: geometric based approaches and appearance

approaches. For the first scenario, the model focus on limiting and tracking specific facial standards, in order to train the model to classify based on relevant positions of these standards. In [26] the authors proposed a model to track a set of points to classify emotions from sequences. In the similar way [24] the authors applied a transformation on a reduced subset of 117 landmarks for emotion recognition. In form based emotion recognition, a set of features is extracted from pixel images to train the classification model. Ko [27] presented a review article which focuses on recent hybrid deep-learning models. Krestinskaya and James [28], proposed an emotion recognition model based on min-max similarity which reduces the problem of interclass pixel mismatching while classification. This paper analyzes the strengths and the limitations of systems based only on facial expressions or acoustic information. It also discusses two approaches used to fuse these two modalities: decision level and feature level integration. Busso et al. [29] evaluate the strengths and the boundaries of models based on facial expressions or acoustic material. In addition, the authors discussed two models that used to fuse decision level and feature level integration. Lopes et al. [31] proposed a simple way for facial expression recognition by combining Convolutional Neural Network and some image pre-processing methods. Regardless of the current progresses, still emotion recognition remains an open problem for the computer vision community. EmotionNet is the largest Challenge in terms of data available, with more than 1 million images (2000 labeled emotions and 950 K unlabeled samples). The first version of the challenge determined that non-frontal faces still pose main problem for the classification algorithms. So, the recognition rates decrease as a function of pitch and yaw rotations [5].

For the last couple of years, Convolutional Neural Networks (CNNs) is one of the most popular approaches in the computer vision. For example, C. Szegedy et al. [20] proposed a model called as GoogLeNet which has numerous convolution layers and this model had novelty due to usage of inception blocks. In every inception block there are several convolution layers which are individually connected to each other and at the last stage the results of convolution layers concatenated and pass to the next convolution layer out of inception block.

### ## 3\ Datasets and pre-processing

#### ### 3.1. Dataset details

To train a neural network, a huge amount of labeled data is required to handle the curse of dimensionality [5]. There are several facial expressions datasets are publicly available such as Cohn?Kanade (CK+) [7] and Japanese Female Facial Expression (JAFPE) datasets and were used in this paper. A big part of datasets with emotion labels have facial pose expressions that were recorded in a precise environment, with head pose and constant lighting. The models that trained on this kind of dataset mostly have low performance on the data which have different conditions, particularly in the live and outdoor environment. We performed training on non-similar datasets to have better generalization on the performance.

The CK dataset involves of 486 video sequences from 97 posers, in CK+ the number of sequences has been extended by 22%. The data has been scaled to 640 × 480 or 640 × 490 pixels and all recorded in grayscale. Entire images were captured with the equal lighting and poses. The dataset is quite appropriate to use for all the models of feature extraction. The JAFPE dataset holds 213 images in total of seven facial expressions (six basic facial expressions and one neutral) posed by ten Japanese female models. The images are in size of 256 × 256 pixels and the expresser have 2?4 samples for every expression. The

samples presented in Fig. 1.

- 1. Download: Download high-res image (137KB)
- 2. Download: Download full-size image

Fig. 1. Sample of dataset for four type of emotion (Sur, Ang, Hap, Neu).

The experimental dataset contains in total 8363 images. 8200 images are used for training, and the rest of images for testing and validation. The dataset details are in Table 1.

Table 1. Dataset description.

Base| Number of images

---|---

Training| CK+| 8000

| JAFFE| 200

Testing| CK+| 150

| JAFFE| 13

### 3.2. Data pre-processing

Image normalization is important pre-processing technique to decrease the inner-class feature mismatch which could be observed as intensity offsets. As the intensity offsets are constant in the local region, Gaussian normalization and standard deviation has been used. The input image is denoted as  $x(p, q)$ , and  $y(p, q)$  is the normalized output image, while  $p$  and  $q$  are the row and column number of the processed image. The normalized resulted image is computed by Eq. (1) [19], where  $\mu$  is a local mean and  $\sigma$  is a local standard deviation calculated over a window of  $M \times M$  size

$$[12].(1) y(p, q) = \frac{x(p, q) - \mu(p, q)}{\sigma(p, q)}$$

The proposed model has the ability to handle different image sizes without human intermediation. The cropping section is delimited by a vertical factor of 4.7 (considering 1.4 for the above the eyes area and 3.3 for the area below) useful to the gap between the eyes middle point and the center of right eye. The horizontal cropping section is delimited by a factor of 2.5 functional to the similar distance. These factor values were determined empirically. A sample of this process is presented in Fig. 2.

- 1. Download: Download high-res image (129KB)
- 2. Download: Download full-size image

Fig. 2. Image cropping example. This action aims to remove all non-expression features, such as background and hair.

The image intensity and contrast can differ even in the same person images and in the same expression, hence, increasing the variation in the feature vector which could increase the complexity of the problem that the classifier has to solve for each expression. To reduce these problems intensity normalization been used. This technique adapted from [30], which named contrastive equalization. In fact, the normalization process is a two-step scheme: firstly subtract local contrast; next, uses divisive local contrast normalization. Initially, the value of all pixels is subtracted from a Gaussian-weighted average of its neighbors. Later, each pixel is divided by the standard deviation of its own neighbor pixels. An example of this procedure is presented in Fig. 3.

- 1. Download: Download high-res image (106KB)
- 2. Download: Download full-size image

Fig. 3. The intensity normalization. The original intensity image (left) and its intensity normalized form (right).

Eqs. (2) and (3) are used to calculate the factors of  $\mu$  and  $\sigma$  while  $a = (M - 1)/2$ .

$$(2) \mu(p, q) = \frac{1}{M^2} \sum_{i=-a}^a \sum_{j=-a}^a x(p+i, q+j) \quad (3) \sigma(p, q) = \sqrt{\frac{1}{M^2} \sum_{i=-a}^a \sum_{j=-a}^a [x(p+i, q+j) - \mu(p, q)]^2}$$

### 3.3. Feature detection

The feature parts that valuable for the facial emotion recognition are

forehead, eyebrows, eyes, cheeks and mouth areas. Here, we show the feature detection by computing the local deviation of already normalized image by a window of  $N \times N$  size. Eq. (4) used for the feature detection with  $b = (N - 1)/2$ . (4)  $\mu(p, q) = \frac{1}{N^2} \sum_{k=-b}^b \sum_{n=-b}^b y(p+k, q+n)$  Parameter  $\mu$  in Eq. (4) denotes the mean of the normalized image  $y(p, q)$  and can be computed by Eq. (5). (5)  $\sigma(p, q) = \frac{1}{N^2} \sum_{k=-b}^b \sum_{n=-b}^b [y(p+k, q+n) - \mu(p, q)]^2$  In the proposed model the convolution layers extracts features hierarchically, and the fully-connected layer and the softmax layer used for indicating 6 expression classes. The input of the network is  $128 \times 96 \times k$  for all patches, where  $k = 3$  for color patches and  $k = 1$  for gray patches. The final output is one of the 6 expression classes.

#### 4. Proposed model

In the following subsections the details of proposed model presented.

##### 4.1. Convolution neural network architecture

Firstly, we introduce the use of face detector.

For the face detection we proposed a deep convolution neural network as shown in Fig. 4 which contains six convolution layers (the green blocks) and two blocks of deep residual learning (purple blocks) the details presented in the next section. Right after each convolution layer there is a max pooling layer (orange blocks). The deep residual blocks implemented after 2nd and 4th convolution layer. There are also 2 Fully Connected layers (FC), each with a ReLU activation function, and dropout for training [18], [22], [23].

1. Download: Download high-res image (221KB)

2. Download: Download full-size image

Fig. 4. Architecture of the proposed DNN model.

Finally we used Softmax for the classification [21]. The details of proposed network presented in Table 2.

Table 2. Details of proposed network.

Type	Filter size / stride	Output size
Conv1	$5 \times 5 / 2$	$64 \times 48 \times 32$
Maxpool1	$2 \times 2 / 2$	$32 \times 24 \times 32$
Conv2	$3 \times 3 / 1$	$32 \times 24 \times 64$
Res1	4 Conv	?
Conv3	$3 \times 3 / 1$	$32 \times 24 \times 128$
Maxpool2	$2 \times 2 / 2$	$16 \times 12 \times 128$
Conv4	$3 \times 3 / 1$	$16 \times 12 \times 128$
Res2	4 Conv	?
Conv5	$3 \times 3 / 1$	$16 \times 12 \times 256$
Maxpool3	$2 \times 2 / 2$	$8 \times 6 \times 256$
Conv6	$3 \times 3 / 1$	$8 \times 6 \times 512$
FC1	1024	1024
FC2	512	512

Furthermore, we performed regularization for each weight matrix  $W$  that limits the size of the weights at individual layer by adding a term to the loss equal to a fixed hyper parameter. We explain these in Eq. (1), where  $x$  be the output of a particular neuron in the network and  $p$  the dropout possibility.

The DNN is used for feature extraction and we just used the extra dataset for the training.

##### 4.2. Residual network

In the proposed model we implemented 2 residual blocks. Each residual block has four convolution layers two short connections and one skip connection. The first convolution layer in residual block has size of  $1 \times 1 \times 64$ . The second

one has size of  $3 \times 3 \times 64$ . The third one has size of  $3 \times 3 \times 128$  and the last one has size of  $1 \times 1 \times 256$ .

### ### 4.3. Regression DNN

Firstly we used a single DNN model to train the datasets. At each time trained a single image, the corresponding image passed through the DNN model, the details of the model shown in Fig. 4.

Two fully-connected layers with 250 hidden units for the approximation of the valence label have been used. For the cost function the mean squared error has been used. For the network training stochastic gradient descent while the batch size sets to 64 and the weight decay sets to  $1E-5$ . Moreover, the learning rate at the beginning sets to  $4e-2$  which decrees by 0.005 every 10 epochs.

### ## 5\.. Experiment and evaluation

For all the pre-training settings described in this paper, global average pooling at the last layer has been applied to decrease spatial dimensionality of data before passing to the fully connected layers. In addition, batch normalization is used to improve generalization and optimization.

For the data preprocessing, we initially identify the face in every outline utilizing face and point of interest finder. Then map the distinguished landmark points to characterized pixel areas in a request to guarantee correspondence concerning outlines. After the normalization the forehead, eyes, nose, mouth while processing each face image through the subtraction and contrast normalization. For tested the proposed DNN model we used a workstation with Intel(R) Core(TM) i7-8700 K and 32GB memory.

Fig. 5 shows the prediction accuracy of the proposed DNN model for training vas validation on the CK+ dataset. These charts clearly show the smooth performance of the proposed model.

1. Download: Download high-res image (246KB)

2. Download: Download full-size image

Fig. 5. Loss and the prediction accuracy for the proposed model.

Fig. 6 presents the Roc curve and the Precision-Recall curve of the proposed model. As it is visible the proposed model has the ability to with the least number of errors and high performance for the face emotion recognition.

1. Download: Download high-res image (175KB)

2. Download: Download full-size image

Fig. 6. Roc and the precision-recall curve.

The confusion matrices of Proposed DNN model on the JAFFE dataset presented in Fig. 7 and its performance on CK+ dataset presented in Fig. 8. The proposed model emotion recognition can reach to 95%. As it is visible in the Figs. 7 and 8 the best recognition are for the Happy, Angry, Neutral, and Surprise emotions.

1. Download: Download high-res image (266KB)

2. Download: Download full-size image

Fig. 7. Confusion matrices on JAFFE Datasets.

1. Download: Download high-res image (261KB)

2. Download: Download full-size image

Fig. 8. Confusion matrices on CK+ Datasets.

Table 3 illustrated the overall performance of proposed DNN model as compared to other deep learning recent models on the CK+ and JAFFE datasets. The proposed DNN model gain higher results in comparison with five other models [[1], [24], [25], and [26]].

Table 3. Proposed model versus other models Performance Comparison on JAFFE and MMI dataset.

Method| Accuracy JAFFE| Accuracy CK+

---|---|---

Zhang et al.[24]| 94.89%| 92.35%

Khorrami et al. [25]| 82.43%| 81.48%

Chernykh et al. [26]| 73%| 70.12%

Jain et al. [1]| 94.91%| 92.71%

Krestinskaya and James [28]| 94.89%| 92.74%

Lopes et al. [31]| 94.86%| 92.73%

Proposed model| \*\*95.23%\*\*| \*\*93.24%\*\*

Our recent approach has better performance in emotion recognition in comparison with our previous model; in Table 3 we present the performance of recent approaches on the whole JAFFE and CK+ dataset.

As the results shows the proposed model on the JAFFE dataset has 0.32%, 0.34% better performance as compared to Jain et al. [1] and Zhang et al. [24] respectively. On the CK+ dataset the proposed model has 0.53% better performance in comparison with Jain et al. [1] and 0.89% better performance while comparing with Zhang et al. [24] model.

### ## 6\ Conclusion

In this paper, we present a fully deep neural network model for facial emotion recognition and the model has been tested on two public datasets to assess the performance of the proposed model. Particularly, it has been found that the combination of FCN and residual block cloud considerably improve the overall result, which verified the efficiency of the proposed model.

Special issue articlesRecommended articles

### ## References

#### 1. [1]

N. Jain, S. Kumar, A. Kumar, P. Shamsolmoali, M. Zareapoor

Hybrid deep neural networks for face emotion recognition

Pattern Recognit. Lett. (2018)

<https://doi.org/10.1016/j.patrec.2018.04.010>

Google Scholar

#### 2. [2]

S.E. Kahou, P. Froumenty, C. Pal

Facial expression analysis based on high dimensional binary features

ECCV Workshop on Computer Vision with Local Binary Patterns Variants (2014)

Google Scholar

#### 3. [3]

C. Shan, S. Gong, P.W. McOwan

Facial expression recognition based on local binary patterns: a comprehensive study

Image Vision Comput., 27 (May (6)) (2009), pp. 803-816

View PDFView articleView in ScopusGoogle Scholar

#### 4. [4]

N. Kalchbrenner, E. Grefenstette, and P. Blunsom. A convolutional neural network for modelling sentences. arXiv:1404.2188, 2014.

Google Scholar

#### 5. [5]

A. Krizhevsky, I. Sutskever, G.E. Hinton

Imagenet classification with deep convolutional neural networks

Advances in neural information processing systems (2012), pp. 1097-1105

Google Scholar

#### 6. [6]

S.E. Kahou, C. Pal, X. Bouthillier, P. Froumenty, C. Gulcehre, \_et al.\_

Combining modality specific deep neural networks for emotion recognition in video

International Conference on Multimodal Interaction, 13, ICMI (2013)

[Google Scholar](#)

7. [7]

M. Liu, R. Wang, S. Li, S. Shan, Z. Huang, X. Chen

Combining multiple kernel methods on riemannian manifold for emotion recognition in the wild

International Conference on Multimodal Interaction, 14, ICMI (2014), pp. 494-501

[Crossref](#)[View in Scopus](#)[Google Scholar](#)

8. [8]

S.E. Kahou, X. Bouthillier, P. Lamblin, C. Gulcehre, V. Michalski, \_et al.\_

Emonets: multimodal deep learning approaches for emotion recognition in video

J. Multimodal User Interf. (2015), pp. 1-13

[View in Scopus](#)[Google Scholar](#)

9. [9]

E. Sariyanidi, H. Gunes, A. Cavallaro

Automatic analysis of facial affect: a survey of registration, representation, and recognition

IEEE Trans. Pattern Anal. Mach. Intell., 37 (6) (2015), pp. 1113-1133

[View in Scopus](#)[Google Scholar](#)

10. [10]

N. Dalal, B. Triggs

Histograms of oriented gradients for human detection

Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 1, IEEE (2005), pp. 886-893

[Crossref](#)[Google Scholar](#)

11. [11]

Y. Zhu, L.C. De Silva, C.C. Ko

Using moment invariants and hmm in facial expression recognition

Pattern Recognit. Lett., 23 (1) (2002), pp. 83-91

[View PDF](#)[View article](#)[View in Scopus](#)[Google Scholar](#)

12. [12]

Y. Sun, X. Chen, M. Rosato, L. Yin

Tracking vertex flow and model adaptation for three-dimensional spatiotemporal face analysis

IEEE Trans. Syst. Man Cybern. Part A: Syst. Humans, 40 (3) (2010), pp. 461-474

[View in Scopus](#)[Google Scholar](#)

13. [13]

N. Sebe, M.S. Lew, Y. Sun, I. Cohen, T. Gevers, T.S. Huang

Authentic facial expression analysis

Image Vision Comput., 25 (12) (2007), pp. 1856-1863

[View PDF](#)[View article](#)[View in Scopus](#)[Google Scholar](#)

14. [14]

M. Zareapoor, P. Shamsolmoali, J. Yang

Learning depth super-resolution by using multi-scale convolutional neural network

Journal of Intelligent & Fuzzy Systems (2018), pp. 1-11, 10.3233/JIFS-18136

[Google Scholar](#)

15. [15]

P. Shamsolmoali, M. Zareapoor, J. Yang

Convolutional neural network in network (CNNiN): hyperspectral image classification and dimensionality reduction

IET Image Processing (2018), 10.1049/iet-ipr.2017.1375

[Google Scholar](#)



16. [16]

A. Krizhevsky, I. Sutskever, G.E. Hinton

Imagenet classification with deep convolutional neural networks

Advances in Neural Information Processing Systems (2012), pp. 1097-1105

Google Scholar

17. [17]

C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V.

Vanhoecke, A. Rabinovich

Going deeper with convolutions

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition

(2015), pp. 1-9

CrossrefGoogle Scholar

18. [18]

M. Zareapoor, D.K. Jain, J. Yang

Local spatial information for image super-resolution

Cogn. Syst. Res., 52 (2018), pp. 49-57

View PDFView articleView in ScopusGoogle Scholar

19. [19]

A. Graves, A. r. Mohamed, G. Hinton

Speech recognition with deep recurrent neural networks

Proc. IEEE International Conference on Acoustics, Speech and Signal Processing

(2013), pp. 6645-6649

View in ScopusGoogle Scholar

20. [20]

A. Graves, N. Jaitly

Towards end-to-end speech recognition with recurrent neural networks

Proc. International Conference on Machine Learning (2014), pp. 1764-1772

Google Scholar

21. [21]

A. Sanin, C. Sanderson, M.T. Harandi, B.C. Lovell

Spatiotemporal covariance descriptors for action and gesture recognition

IEEE Workshop on Applications of Computer Vision (2013)

Google Scholar

22. [22]

P. Shamsolmoali, M. Zareapoor, D.K. Jain, V.K. Jain, J. Yang

Deep convolution network for surveillance records super-resolution

Multimed Tools Appl. (2018), pp. 1-16, 10.1007/s11042-018-5915-7

View in ScopusGoogle Scholar

23. [23]

F. Visin, K. Kastner, K. Cho, M. Matteucci, et al., Renet: a recurrent neural

network based alternative to convolutional networks. arXiv preprint

arXiv:1505.00393, 2015.

Google Scholar

24. [24]

T. Zhang, W. Zheng, Z. Cui, Y. Zong, Y. Li

Spatial-temporal recurrent neural network for emotion recognition

IEEE Trans. Cybern. (99) (2018), pp. 1-9

arXiv:1705.04515

Google Scholar

25. [25]

P. Khorrami, T.L. Paine, K. Brady, C. Dagli, T.S. Huang

How deep neural networks can improve emotion recognition on video data

IEEE Conference on Image Processing (ICIP) (2016)

Google Scholar

26. [26]

V. Chernykh, G. Sterling, P. Prihodko, Emotion recognition from speech with recurrent neural networks, arXiv:1701.08071v1 [cs.CL], 2017.

Google Scholar

27. [27]

B.C. Ko

A brief review of facial emotion recognition based on visual information

Sensors, 18 (2018), p. 401

CrossrefView in ScopusGoogle Scholar

28. [28]

O. Krestinskaya, A.P. James

Facial emotion recognition using min-max similarity classifier

Conference on Advances in Computing, Communications and Informatics (ICACCI)

(2017), pp. 752-758

CrossrefView in ScopusGoogle Scholar

29. [29]

C. Busso, Z. Deng, S. Yildirim, M. Bulut, C.M. Lee, A. Kazemzadeh, S. Lee, U.

Neumann, S. Narayanan

Analysis of emotion recognition using facial expressions

Speech and Multimodal Information, International Conference on Multimodal

Interfaces (2004), pp. 205-211

CrossrefView in ScopusGoogle Scholar

30. [30]

B.A. Wandell

Foundations of Vision

(First ed.), Sinauer Associates Inc, Sunderland, Mass (1995)

Google Scholar

31. [31]

A.T. Lopes, E. d. Aguiar, A.F. De Souza, T. O.Santos

Facial expression recognition with convolutional neural networks: coping with few data and the training sample order

Pattern Recognit., 61 (2017), pp. 610-628

View PDFView articleView in ScopusGoogle Scholar

## Cited by (291)

\* ### A systematic review on affective computing: emotion models, databases, and recent advances 2022, Information Fusion

Show abstract

Affective computing conjoins the research topics of emotion recognition and

sentiment analysis, and can be realized with unimodal or multimodal data,

consisting primarily of physical information (e.g., text, audio, and visual)

and physiological signals (e.g., EEG and ECG). Physical-based affect

recognition caters to more researchers due to the availability of multiple

public databases, but it is challenging to reveal one's inner emotion hidden

purposefully from facial expressions, audio tones, body gestures, etc.

Physiological signals can generate more precise and reliable emotional

results; yet, the difficulty in acquiring these signals hinders their

practical application. Besides, by fusing physical information and

physiological signals, useful features of emotional states can be obtained to

enhance the performance of affective computing models. While existing reviews

focus on one specific aspect of affective computing, we provide a systematical

survey of important components: emotion models, databases, and recent

advances. Firstly, we introduce two typical emotion models followed by five

kinds of commonly used databases for affective computing. Next, we survey and

taxonomize state-of-the-art unimodal affect recognition and multimodal

affective analysis in terms of their detailed architectures and performances. Finally, we discuss some critical aspects of affective computing and its applications and conclude this review by pointing out some of the most promising future directions, such as the establishment of benchmark database and fusion strategies. The overarching goal of this systematic review is to help academic and industrial researchers understand the recent advances as well as new developments in this fast-paced, high-impact domain.

\* ### Automated emotion recognition: Current trends and future perspectives  
2022, Computer Methods and Programs in Biomedicine

Show abstract

Human emotions greatly affect the actions of a person. The automated emotion recognition has applications in multiple domains such as health care, e-learning, surveillance, etc. The development of computer-aided diagnosis (CAD) tools has led to the automated recognition of human emotions.

This review paper provides an insight into various methods employed using electroencephalogram (EEG), facial, and speech signals coupled with multi-modal emotion recognition techniques. In this work, we have reviewed most of the state-of-the-art papers published on this topic.

This study was carried out by considering the various emotion recognition (ER) models proposed between 2016 and 2021. The papers were analysed based on methods employed, classifier used and performance obtained.

There is a significant rise in the application of deep learning techniques for ER. They have been widely applied for EEG, speech, facial expression, and multimodal features to develop an accurate ER model.

Our study reveals that most of the proposed machine and deep learning-based systems have yielded good performances for automated ER in a controlled environment. However, there is a need to obtain high performance for ER even in an uncontrolled environment.

\* ### A survey on facial emotion recognition techniques: A state-of-the-art literature review  
2022, Information Sciences

Show abstract

In this survey, a systematic literature review of the state-of-the-art on emotion expression recognition from facial images is presented. The paper has as main objective arise the most commonly used strategies employed to interpret and recognize facial emotion expressions, published over the past few years. For this purpose, a total of 51 papers were analyzed over the literature totaling 94 distinct methods, collected from well-established scientific databases (ACM Digital Library, IEEE Xplore, Science Direct and Scopus), whose works were categorized according to its main construction concept. From the analyzed works, it was possible to categorize them into two main trends: classical and those approaches specifically designed by the use of neural networks. The obtained statistical analysis demonstrated a marginally better recognition precision for the classical approaches when faced to neural networks counterpart, but with a reduced capacity of generalization. Additionally, the present study verified the most popular datasets for facial expression and emotion recognition showing the pros and cons each and, thereby, demonstrating a real demand for reliable data-sources regarding artificial and natural experimental environments.

\* ### Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review  
2020, Information Fusion

Citation Excerpt :

They combined semantic and sentiment feature representations to classify three classes of emotions (i.e., happy, sad and angry) in the texts. In [170] Jain et al. utilized an extended deep convolutional neural network for facial

emotion classification. The deep network architecture contains several convolution layers and deep residual blocks.

Show abstract

In recent years, the rapid advances in machine learning (ML) and information fusion has made it possible to endow machines/computers with the ability of emotion understanding, recognition, and analysis. Emotion recognition has attracted increasingly intense interest from researchers from diverse fields. Human emotions can be recognized from facial expressions, speech, behavior (gesture/posture) or physiological signals. However, the first three methods can be ineffective since humans may involuntarily or deliberately conceal their real emotions (so-called social masking). The use of physiological signals can lead to more objective and reliable emotion recognition. Compared with peripheral neurophysiological signals, electroencephalogram (EEG) signals respond to fluctuations of affective states more sensitively and in real time and thus can provide useful features of emotional states. Therefore, various EEG-based emotion recognition techniques have been developed recently. In this paper, the emotion recognition methods based on multi-channel EEG signals as well as multi-modal physiological signals are reviewed. According to the standard pipeline for emotion recognition, we review different feature extraction (e.g., wavelet transform and nonlinear dynamics), feature reduction, and ML classifier design methods (e.g., k-nearest neighbor (KNN), naive Bayesian (NB), support vector machine (SVM) and random forest (RF)). Furthermore, the EEG rhythms that are highly correlated with emotions are analyzed and the correlation between different brain areas and emotions is discussed. Finally, we compare different ML and deep learning algorithms for emotion recognition and suggest several open problems and future research directions in this exciting and fast-growing area of AI.

\* ### Deep convolution network based emotion analysis towards mental health care  
2020, Neurocomputing

Citation Excerpt :

We would argue that static approaches show promising, but require rigorous and various testing conditions for robust outcomes. Deepak et al., for example, proposed a convolution neural network based algorithm to recognize facial expression with high accuracy [7]. However, their algorithm was tested on only two small facial expression datasets.

Show abstract

Facial expressions play an important role during communications, allowing information regarding the emotional state of an individual to be conveyed and inferred. Research suggests that automatic facial expression recognition is a promising avenue of enquiry in mental healthcare, as facial expressions can also reflect an individual's mental state. In order to develop user-friendly, low-cost and effective facial expression analysis systems for mental health care, this paper presents a novel deep convolution network based emotion analysis framework to support mental state detection and diagnosis. The proposed system is able to process facial images and interpret the temporal evolution of emotions through a new solution in which deep features are extracted from the Fully Connected Layer 6 of the AlexNet, with a standard Linear Discriminant Analysis Classifier exploited to obtain the final classification outcome. It is tested against 5 benchmarking databases, including JAFFE, KDEF, CK+, and databases with the images obtained in the wild such as FER2013 and AffectNet. Compared with the other state-of-the-art methods, we observe that our method has overall higher accuracy of facial expression recognition. Additionally, when compared to the state-of-the-art deep learning algorithms such as Vgg16, GoogleNet, ResNet and AlexNet, the

proposed method demonstrated better efficiency and has less device requirements. The experiments presented in this paper demonstrate that the proposed method outperforms the other methods in terms of accuracy and efficiency which suggests it could act as a smart, low-cost, user-friendly cognitive aid to detect, monitor, and diagnose the mental health of a patient through automatic facial expression analysis.

\* ### Facial emotion recognition using deep learning: Review and insights  
2020, Procedia Computer Science

Show abstract

Automatic emotion recognition based on facial expression is an interesting research field, which has presented and applied in several areas such as safety, health and in human machine interfaces. Researchers in this field are interested in developing techniques to interpret, code facial expressions and extract these features in order to have a better prediction by computer. With the remarkable success of deep learning, the different types of architectures of this technique are exploited to achieve a better performance. The purpose of this paper is to make a study on recent works on automatic facial emotion recognition FER via deep learning. We underline on these contributions treated, the architecture and the databases used and we present the progress made by comparing the proposed methods and the results obtained. The interest of this paper is to serve and guide researchers by review recent works and providing insights to make improvements to this field.

View all citing articles on Scopus

View Abstract

© 2019 Elsevier B.V. All rights reserved.

## Part of special issue

Pattern Recognition Techniques for Non Verbal Human Behavior

Edited by Wankou Yang, Su-Jing Wang, Pritee Khanna, Xiong Li

View special issue

## Recommended articles

\* ### Random walk-based feature learning for micro-expression recognition

Pattern Recognition Letters, Volume 115, 2018, pp. 92-100

Deepak Kumar Jain, ?, Kaiqi Huang

View PDF

\* ### Facial expression recognition based on deep convolution long short-term memory networks of double-channel weighted mixture

Pattern Recognition Letters, Volume 131, 2020, pp. 128-134

Hepeng Zhang, ?, Guohui Tian

View PDF

\* ### Facial expression recognition via learning deep sparse autoencoders

Neurocomputing, Volume 273, 2018, pp. 643-649

Nianyin Zeng, ?, Abdullah M. Dobaie

View PDF

\* ### Multi angle optimal pattern-based deep learning for automatic facial expression recognition

Pattern Recognition Letters, Volume 139, 2020, pp. 157-165

Deepak Kumar Jain, ?, Kaiqi Huang

View PDF

\* ### Human emotion recognition using deep belief network architecture

Information Fusion, Volume 51, 2019, pp. 10-18

Mohammad Mehedi Hassan, ?, Giancarlo Fortino

View PDF

\* ### Spatio-temporal convolutional features with nested LSTM for facial expression recognition

Neurocomputing, Volume 317, 2018, pp. 50-57

Zhenbo Yu, ?, Jiankang Deng

[View PDF](#)

[Show 3 more articles](#)

**## Article Metrics**

Citations

\* Citation Indexes: 288

Captures

\* Readers: 303

[View details](#)

\* [About ScienceDirect](#)

\* [Remote access](#)

\* [Shopping cart](#)

\* [Advertise](#)

\* [Contact and support](#)

\* [Terms and conditions](#)

\* [Privacy policy](#)

Cookies are used by this site. [Cookie Settings](#)

All content on this site: Copyright © 2024 Elsevier B.V., its licensors, and contributors. All rights are reserved, including those for text and data mining, AI training, and similar technologies. For all open access content, the Creative Commons licensing terms apply.

**## Cookie Preference Center**

We use cookies which are necessary to make our site work. We may also use additional cookies to analyse, improve and personalise our content and your digital experience. For more information, see our [Cookie Policy](#) and the list of [Google Ad-Tech Vendors](#).

You may choose not to allow some types of cookies. However, blocking some types may impact your experience of our site and the services we are able to offer. See the different category headings below to find out more or change your settings.

Allow all

**### Manage Consent Preferences**

**#### Strictly Necessary Cookies**

Always active

These cookies are necessary for the website to function and cannot be switched off in our systems. They are usually only set in response to actions made by you which amount to a request for services, such as setting your privacy preferences, logging in or filling in forms. You can set your browser to block or alert you about these cookies, but some parts of the site will not then work. These cookies do not store any personally identifiable information.

[Cookie Details List?](#)

**#### Functional Cookies**

Functional Cookies

These cookies enable the website to provide enhanced functionality and personalisation. They may be set by us or by third party providers whose services we have added to our pages. If you do not allow these cookies then some or all of these services may not function properly.

[Cookie Details List?](#)

**#### Performance Cookies**

Performance Cookies

These cookies allow us to count visits and traffic sources so we can measure and improve the performance of our site. They help us to know which pages are the most and least popular and see how visitors move around the site.

[Cookie Details List?](#)

**#### Targeting Cookies**

Targeting Cookies

These cookies may be set through our site by our advertising partners. They may be used by those companies to build a profile of your interests and show you relevant adverts on other sites. If you do not allow these cookies, you will experience less targeted advertising.

Cookie Details List?

Back Button

### Cookie List

Search Icon

Filter Icon

Clear

checkbox label label

Apply Cancel

Consent Leg.Interest

checkbox label label

checkbox label label

checkbox label label

Confirm my choices