Outline

Neurocomputing
Volume 388, 7 May 2020, Pages 212-227
Deep convolution network based emotion analysis towards mental health care
Author links open overlay panelZixiang Fei a, Erfu Yang a, David Day-Uei Li b, Stephen Butler c, Winifred Ijomah a, Xia Li d, Huiyu Zhou e

https://doi.org/10.1016/j.neucom.2020.01.034Get rights and content
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.
* Previous article in issue
* Next article in issue
Keywords
Facial expression recognition
Deep convolution network
Mental health care
Emotion analysis
1\. Introduction
Understanding people's emotions plays an important role in daily human communications. For advanced human-computer interaction in many emerging applications, recognizing users? emotions is also vital. Currently, there are many approaches for automated emotional recognition, including the recognition of facial expression and analysis of voice tone [1], which exist alongside more conventional physiological measures such as measurements of blood pressure, pulse rate or skin conductivity.
Automated facial expression recognition has found many practical applications such as in e-learning and health care systems [2,3]. For example, facial expressions are considered as an important feedback mechanism for teachers in terms of monitoring levels of students? understanding [3]. Within this domain, Mau-Tsuen et al. proposed an automatic system to identify how well students were learning, by means of the analysis of videos taken from learners [3], and demonstrated that such a system could help improving teaching effectiveness and efficiency. Additionally, the user interface developed in such a system could work as a cognitive tool [4] to support and understand the users? mental state better when and where it was appropriate without external actuation. Further applications of facial expression analysis include its use in the fields of human-computer interaction and interface optimization. Within these domains, Bahr et al. proposed a novel method to analyze postural and facial expressions to guide interface actuation and actions [4], whilst Pedro et al. employed electromyogram (EMG) sensors to investigate the relationship between users? facial expressions and adverse-event occurrences [5]. Moreover, health and medical applications of automated facial expression analysis are also being investigated, for example in the area of diagnostics related to developmental disorders such as autism. Here, promising work has been

conducted by means of a system of facial expression analysis during social interactions, where many individuals with such disorders find challenging [6]. Modern techniques in facial expression recognition systems play important roles in these practical applications. These techniques typically involve multiple components, including face localization and facial component alignment, facial feature extraction and facial feature classification. As a simple dichotomy, facial expression analysis systems can be divided into two groups: those using static images and those using continuous video frames. Static algorithms enjoy the advantage of facial expression recognition from a single image or a video frame. In this paper, we describe an algorithm which has been tested for facial expression recognition accuracy from single images mainly acquired through webcams and mobile phone cameras. 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. Conversely, Jie et al. also proposed three novel convolution neural network (CNN) models with different architectures [8], and tested their algorithms on several datasets such as CK+ and FER2013 datasets. Within the three architectures they presented, the data suggested that the pertained CNN with 5 convolution layers had the best performance. It should be noted however that not all researchers restrict their work to 2D images. For instance, Chenlei et al. very recently proposed a new 3D facial expression modeling method based on facial landmarks [9]. On the other hand, other researchers have moved beyond the analysis of static images and proposed facial expression recognition systems using continuous video frames. For instance, Zia, Lee and other researchers worked on facial expression recognition using temporal dynamics [10,11]. Their proposed system used Fisher Independent Component Analysis as a feature extractor and Hidden Markov Models to learn the features of six different expressions. The experiment results showed that the recognition rate was about 92.85%.

Despite advances in the recognition accuracy, there remain significant issues to be addressed in the field of facial expression recognition systems. The first issue is related to the datasets employed in testing such systems. Some facial expression recognition systems may have good performance in some image datasets, but perform poorly in the others. For instance, deep CNN approaches often need to determine large amount of weights in the training phase. Consequently it is observed that such approaches suffer from performance decrements when being trained on a small image dataset [8]. A second important issue is ecological validity, in that most of the existing systems use lab-posed facial expressions images. This is potentially problematic, as it ignores real-world problems such as lighting conditions, image quality and background complexity. It is fatal to address such issues for real-world applications. Facial expression recognition in the wild is a challenging topic due to such issue as well as others such as variance in poses [8]. Thirdly, both traditional approaches and deep learning based approaches have inherent weaknesses. Traditional approaches such as Local Binary Pattern (LBP), Scale Invariant Feature Transform (SIFT) and Support Vector Machine (SVM) employ manually designed features, which can result in poor performance with unseen images. Deniz et al., for instance, observed that traditional approaches performed weakly when being presented with variations in pose, and instead employed a deep CNN based approach to recognize facial expressions, which they argued resulted in improved performance [12]. Approaches such as AlexNet, Vgg, and GoogleNet have long training and testing time as well as an enormous

amount of memory resource [13] and require hardware incorporating Graphics Processing Units (GPU's), whilst it is also essential in these approaches. Extensive literature provides a comprehensive review of the relative advantages and weaknesses of the existing facial expression systems, we would recommend those seeking further information to consult one of the several good reviews of the area (see [14], [15], [16], [17], [18]). Indeed, even more recently Byoung, reviewed approaches to facial emotion recognition including conventional facial expression recognition, deep learning based facial expression recognition, well-known facial expression datasets and has also examined some common performance evaluation methods for automated facial expression recognition [19].

Whilst there are many notable developments within the field of automatic face recognition systems, there remain urgent priorities to be addressed to allow these technologies to flourish within the field of mental health care. We would argue that there is vast untapped potential for the field in this area of healthcare. Globally, there are rising numbers of people suffering from cognitive impairment, and consequently there is an urgent need to develop and deploy user-friendly, low-cost and effective facial expression analysis systems for mental health care. Within mental health care, there is enormous potential for automatic facial expression recognition systems to assist clinicians working within mental health settings. Whilst it is uncontroversial to suggest that facial expressions can reflect people's mental states [20], it has also been reported that the patients with cognitive impairments may express abnormal facial expressions [21] that are quantitatively different from those of healthy elderly people. Burton and Kaszniak, for example, reported abnormal corrugator activity in individuals with cognitive impairment when compared to a control group, when these participants were exposed to image or video stimuli [21]. The forehead muscle is integral to the emitted emotions related to frowning [22]. Work in line with this was conducted by Henry et al., who invited 20 cognitive impaired people and 20 healthy people to watch videos in order to compare their facial reactions when viewing such stimuli, reporting that the group of participants with cognitive impairment demonstrated difficulty both in facial muscle control and in the amplification of expressed emotion [23]. Similarly, Smithfound cognitive impaired people demonstrated more negative emotions when they were exposed to negative image stimuli, which they argued was indicative of reduced emotional control [24]. Building upon such initial findings, this paper presents a proposed system that has the potential to be applied to the detection and monitoring of cognitive impairments such as dementia through the application of machine learning techniques to the automatic analysis of people's facial expressions. Our proposed smart system is able to label and quantify the users? emotions automatically, through continuous focus upon the evolution of facial expressions over a period of time.

In order to develop an efficient and effective framework for recognizing facial expression towards mental health care, this paper presents a facial expression recognition framework which effectively exploits the features extracted from the Fully Connected Layer 6 of AlexNet and the Linear Discriminant Analysis Classifier (LDA). We present the findings from a series of experiments where the performance of our proposed method are compared to that of traditional approaches such as SVM, LDA, the K-nearest Neighbor Classifier (KNN) and made further comparisons to deep learning based approaches such as AlexNet, Vgg, GoogleNet and ResNet. Within our comparisons, we have employed a broad range of well-known facial expression databases including the Karolinska Directed Emotional Faces (KDEF) Database [25], the

Japanese Female Facial Expression (JAFFE) Database [26], the extended Cohn-Kanade dataset (CK+) [27,28], the Facial Expression Recognition Challenge (FER2013) [29] and the AffectNet Database [30]. From our experiments it would suggest that exploiting the features extracted from the Fully Connected Layer 6 of AlexNet with the LDA classifier can achieve the best performance in all the five testing databases.

The major contributions of this paper are as follows:

* 1.
First, our novel framework for facial expression analysis to support mental health care is presented. Our proposed system extracts deep features from the Fully Connected Layer 6 of the AlexNet, and uses a standard Linear Discriminant Analysis Classifier to train these deep features. As it is known that patients with cognitive impairments may express abnormal facial expressions when exposed to emotional visual stimuli, we would argue that our proposed facial analysis system has excellent potential to detect cognitive impairment at the early stage.
* 2.
We present the findings from the tests of our proposed framework against both across databases with small number of images such as the JAFFE, KDEF and CK+ databases, and databases with images obtained ?in the wild? such as the FER2013 and the AffectNet databases.
* 3.
We present the findings from the comparisons of our proposed method with both traditional methods and state-of-the-art methods proposed by other researchers. It is observed that our method has better accuracy facial expression recognition. More importantly, we have also observed that our proposed method has much less computing time and lower device requirements than the other state-of-the-art deep learning algorithms such as Vgg16, GoogleNet, and ResNet. We would argue that these characteristics support our view that the proposed method is both competent and suitable for a facial analysis system that can be employed within mental health care settings.
* 4.
A system which can analyze facial expressions from video stimuli, and subsequently produce an accurate evaluation of the facial expressions detected in an automated system is presented.

This paper is organized as follows. Following our general overview in Section 1, we present our proposed deep learning-based framework in Section 2. Section 3 then introduces the experimental set-up for the evaluation of the proposed system to obtain the results. Our findings are then discussed within Section 4. Finally, Section 5 provides our conclusions.

2\. Materials and methods

2.1. Overview of the whole system structure

In the current research, the inputs to the system are videos of people's frontal faces. Image pre-processing is then applied to the video streams. The analysis of the acquired videos is carried out using the deep convolution network AlexNet combined with traditional classifiers such as SVM, LDA and KNN. Finally, the system we present analyzes the emotions acquired in the videos and reports the evolutions of the emotions detected over a period of time. The general structure of the system is shown in Fig. 1.

1. Download: Download high-res image (960KB)
2. Download: Download full-size image

Fig. 1. Structure of the whole system [34].

In the system, the first stage is the system input. Here we take video frames of facial expressions from 120 s of video with a rate of 30 frames per second.

As a result, a 120-second video is thus converted to 3600 images. The image pre-processing technique can then be applied to such input image frames. This stage has two main operations: removal of unnecessary image parts such as background environmental aspects of the image, and removal of hair; and resizing of the images. Within our approach, firstly a face detector is used to locate the position of the face in the images using the standard Viola-Jones algorithm [31,32]. Next, the appropriate face part is cropped from the images. Finally, the cropped face part is resized to the required size, which is decided by the input of the deep learning network. In our proposed system, we utilize well-established and reliable AlexNet to extract the deep features from the images which have been applied with the aforementioned image pre-processing techniques.

Finally, at this stage the image will be also converted into RGB format by concatenating the arrays [33] to meet the requirement of the deep learning network if the video frames are grayscale.

In the third stage, facial expression analysis techniques are applied to the pre-processed images. These techniques use a combination of deep learning networks i.e. AlexNet, and traditional classifiers such as SVM, LDA and KNN, which will be introduced in detail in the next section. Facial expressions in the video stimuli are classified into five facial expression groups, namely neutral, happy, sad, angry and surprised. In addition, the probability of each facial expression category for every frame is established. By combining all the facial expression recognition results in the image frames, we obtain the evolution of the probability of each type of facial expression over a period of time.

2.2. Convolution neural networks and proposed framework

2.2.1. Overview of convolution neural networks

Convolution Neural Networks (CNNs) are a type of deep learning network that needs less image pre-processing compared to other traditional image classification algorithms [35]. CNNs have an advantage in that they do not need prior knowledge and manually design the features. They have many applications in various domains including natural language processing and computer vision [36]. A typical CNN has an input layer, an output layer, and hidden layers such as convolution layers, pooling layers and fully connected layers.

2.2.2. AlexNet

AlexNet is a type of CNN that was designed by Alex Krizhevsky in the SuperVision group [37]. The AlexNet competed in the ImageNet Large Scale Visual Recognition Challenge in 2012, and achieved a top-five error of 15.3%. It has eight major layers in total including five convolution layers and three fully connected layers. The detailed network structure is shown in Fig. 2. The original AlexNet was trained on a subset of the ImageNet database which contains more than one million images, and was able to classify images into 1000 object categories [34]. In the AlexNet, the input receives images and the output includes the label of the images and the probabilities for each of the object categories. In our experiments, the AlexNet is run in Matlab. Moreover, the transfer learning strategy is used to reduce the training time of the network [34]. By using the transfer learning, a much smaller number of training images are needed. There are several steps needed for transfer learning in a pre-trained AlexNet.

1. Download: Download high-res image (317KB)
2. Download: Download full-size image

Fig. 2. Structure of the AlexNet [34,61].

There are 25 layers in the AlexNet, which begins with an image input layer.

Next, there are five convolution layers to extract facial features. The output of the activation function forms the neurons of the current layer. As a result, it will form the feature map of the current convolution layer. The calculation can be described in the following function [37], [38], [39]:(1)Xjl=F(?i?MjXil?1*wijl+wb)Where, Xjl?1 is the feature map of the output of the L-1 layer, * depicts the convolution operation, wijl and _w b_represents weight and the bias, respectively. Each convolution layer is followed by the ReLU (Rectified Linear Units) layer in order to increase the nonlinear properties of the network. The ReLU is a half-wave rectifier function with the advantage of reducing the training time whilst preventing overfitting [40]. In addition, ReLU layers can prevent the gradient vanishing problem and are much faster than other logistic function [41]. The ReLU layers for input _x_ can be described as [37], [38], [39]:(2)F(x)=max(0,x)
In addition, convolution layers are also followed by the max pooling layers. The max pooling layers are used for down-sampling. The number of feature maps will not' be changed by the down-sampling process. On the other hand, the down-sampling process removes unnecessary information and reduces the number of the parameters of the feature map. The down-sampling layer can be described by Krizhevsky et al. [37], [38], [39]:(3)Xjl=F(down(Xjl?1)+wb)where Xjl?1 represents the _j_ feature map of the pooling layer _l_ , and _w b_ depicts the offset term of the down-sampling layer.
Finally, in the AlexNet, after the five convolution layers, there are three fully connected layers: FC6, FC7 and FC8. The fully connected layers can be considered as a convolution layer. Also, its convolution kernel size and the input data size should be consistent with those used in the convolution layer. The Fully Connected Layer can be described by Krizhevsky et al. [37], [38], [39]:(4)Xjl=F(?iwijlXjl?1+wbl)
2.2.3. Deep feature extraction
In this context, the AlexNet works as a deep feature extractor to extract image features. We then use these features to train traditional classifiers such as SVM, LDA and KNN. In spite of the possible improvement in recognition accuracy, feature extraction remains fast and relatively simple, using the representational power of the pre-trained deep networks [42]. In addition, as feature extraction only requires a single pass through the data, a CPU is also able to do this work. In this work, a pre-trained CNN works as the deep feature extractor. The advantages of using a pre-trained CNN to extract deep features, compared to the training of a new CNN, are: (1) Less computational power is needed and (2) less data is needed to achieve high recognition accuracy [43]. The work we present used a pre-trained AlexNet, which was trained with more than a million images so that the network model has learned rich feature representations.
To demonstrate what features can be learned by the AlexNet, we need to visualize the deep features extracted by the network. In our work, we employed the T-SNE (Stochastic Neighbor Embedding) approach to visualize the features extracted by the AlexNet. The T-SNE - (TSNE) is an algorithm for dimensionality reduction [44,45]. This algorithm allows us to visualize the high-dimensional data of the facial images. The T-SNE function will convert high-dimensional data into low dimensional data. Generally, distant points in high-dimensional space will be converted into distant embedded low-dimensional points and nearby points in the high-dimensional space will be converted into nearby embedded low-dimensional points. As a result, we can visualize the low-dimensional points to find the clusters in the original high-dimensional data.
The features extracted by the AlexNet which have transfer-learned facial images from the KDEF dataset are shown in Fig. 3.

1. Download: Download high-res image (238KB)
2. Download: Download full-size image

Fig. 3. Using the T-SNE to visualize the features extracted by the AlexNet.

2.2.4. Layer selection

In this study, we aimed to achieve the best performance by selecting the most appropriate layer to extract the features including the Fully Connected Layer 6 (FC6), Fully Connected Layer 7 (FC7) and Fully Connected Layer 8 (FC8) from the AlexNet. While extracting the features from deep CNNs, deeper layers contain higher level features and earlier layers produce lower-level features [42]. In order to obtain the feature representations of the training and testing images, we can use the activations on FC7 or we can obtain lower-level representations of the images from FC6. Meanwhile, the earlier layers may learn features like colors and edges [46]. However, in the deeper layers, the network may learn complicated features such as eyes. We will do experiments to explore, evaluate and compare which layer in the AlexNet is most appropriate to extract the deep features in our study.

2.2.5. Traditional classifiers

The deep features or ?activations? can provide the input of traditional classifiers such as SVM, LDA and KNN in order to further improve the recognition accuracy. Therefore, we also investigated which traditional classifier combined with the AlexNet can achieve the best recognition accuracy. Each of the traditional classifiers has its own characteristics. For example, SVM can use kernels to transform many feature representations into a higher dimensional space in order to classify multiple classes [43]. In addition, SVM has good performance in object classification and face detection applications.

On the other hand, the LDA approach is able to find the optimal transformation which can better separate different classes [47]. The LDA has wide applications such as face recognition and image retrieval [48]. In the LDA, when _W_ represents an optimal set of discriminant projection vectors [48], the LDA can be represented as a function of _W_ :(5)J(W)=|WTSBW||WTSWW| In (5), _S B_ and _S W_ are between class scatter matrix and within class scatter matrix, respectively. They are given by(6)Sb=?i=1NMi(mi?m)(mi?m)T(7)SW=?i=1NMiwhere(8)Mi=?X?xi(x??i)(x??i)T The comparison of these traditional classifiers are presented in the experimental results section.

3\. Results

In order to evaluate and identify the approach with the best recognition accuracy, different methods were tested in our experiments. To this end, we tested the pre-trained deep CNN AlexNet with the initial learning rate 0.0003, minimum batch size 5 and maximum epochs 10 [34]. The test of the traditional classifiers included multiclass model for SVM, the LDA with linear Discriminant-Type and the KNN with Euclidean Distance and 1 neighbor number. We also conducted some experiments to test the influence of the hyperparameters in the training stage. For the proposed method (AlexNet + FC6 + LDA), we use the LDA classifier to train and classify the deep features extracted from the AlexNet. We have tried to use the ?OptimizeHyperparameters? function in Matlab to find the optimized hyperparameters. We found that the time cost outweighs the small performance improvement. In the LDA, the function will try to optimize the performance by changing hyperparameters Delta and Gamma automatically. By using the OptimizeHyperparameters? function, we found the operating time is increased by 100 times and there is limited improvement in recognition accuracy in the JAFFE dataset for the LDA classifier. We found the ?OptimizeHyperparameters? function has the drawback

of greatly increasing the operating time, which is not appropriate in our research. As a result, we use the default hyperparameters for the LDA classifier.

We also tested the combination of the AlexNet with traditional classifiers [49]. As the layer selected to extract the features affects the recognition result, we conducted the experiments using different layers to extract the deep features. The explored layers are the FC6, FC7, and FC8 of the AlexNet. We first tested the combination of the AlexNet and the LDA using FC6, FC7 and FC8 to investigate which layer may have the best recognition accuracy. We subsequently discovered that FC6 demonstrated the highest recognition accuracy. We then used FC6 to extract the deep features and experimentally tested which classifier demonstrated the highest recognition accuracy among SVM, LDA and KNN. We found that the LDA showed the best recognition performance. In summary, we tested the recognition accuracy of facial expressions using the following methods: deep convolution neural networks AlexNet, traditional classifiers SVM, LDA and KNN and the combination of the AlexNet and traditional classifiers using FC8 and LDA, FC7 and LDA, FC6 and LDA, FC6 and SVM, and FC6 and KNN, respectively.

Additionally, in order to compare the recognition accuracy among the nine methods, we conducted our experiments on different facial expression databases in this research. Specifically, we used JAFFE, KDEF, CK+, FER2013 and AffectNet Datasets. Moreover, we used the proposed system to quantify the evolution of facial expressions from a video of facial expressions emitted by the first author as the ground truth is known in this case. The detailed experimental outcomes of these comparisons are reported in the following sections.

3.1. Experiment using the online datasets

3.1.1. JAFFE dataset experiments

To begin with, the JAFFE is a facial expression database with only 213 static images. By using the JAFFE Dataset, we aim to test the influence of small number of images in training the system using different methods. From the JAFFE Dataset, we selected 202 images that were all processed using the image preprocessing techniques mentioned above (it was observed that the JAFFE dataset included some incorrectly labeled facial expressions, which were removed [26]). There are 7 different facial expressions in this dataset: angry, happy, neutral, surprised, sad, afraid, and disgusted. In each test, 70% of the images were randomly selected as the training images, with the rest of the images serving as testing images. Table 1 compares the 5-fold Cross Validation Error Rate for the facial expressions using the JAFFE Dataset with each different method. The first column in the table shows the method used for facial expressions recognition. The second column shows the 5-fold Cross Validation Error Rate for each method.

Table 1. Comparison of cross-validation error rate using different methods for the JAFFE dataset.

Method	Error Rate (%)
AlexNet	24.8
LDA	26.7
SVM	24.8
KNN	39.6
AlexNet + FC8 + LDA	18.8
AlexNet + FC7 + LDA	9.4
AlexNet + FC6 + LDA	5.9
AlexNet + FC6 + SVM	10.9

AlexNet + FC6 + KNN| 15.4
As the table shows, our proposed method (AlexNet + FC6 + LDA) reaches the lowest error rate of 5.9%. As the number of images in the JAFFE Database is quite small, the deep CNN AlexNet does not demonstrate satisfactory performance. Also, the performance of the AlexNet appears to be quite near to that of the traditional classifiers such as LDA and SVM.
Furthermore, Table 2 also shows the recognition accuracy for each emotion and the overall recognition accuracy rates obtained when using each different method for the JAFFE Dataset. The first column in the table shows the method used for facial expressions recognition whilst the second to the eighth column shows the recognition accuracy of each method for angry, disgusted, fearful, happy, neutral, sad and surprised faces, respectively. The ninth column shows the overall recognition accuracy of each method. Also, the last row shows the average recognition accuracy using all 9 methods for each emotion. The method using AlexNet, FC6 and LDA has the highest overall recognition accuracy, but does not demonstrate good performance in recognizing ?surprise?. In general, ?happy? is the most easily recognized emotion by this method, while ?surprised? is the most difficult emotion category to be recognized.
Table 2. Comparison of recognition accuracy of each emotion and overall recognition accuracy using different methods for the JAFFE dataset.
Method	Angry (%)	Disgust (%)	Fear (%)	Happy (%)	Neutral (%)	Sad (%)	Surprise (%)	Overall Accuracy for Each Method (%)
AlexNet	88.9	37.5	33.3	100	77.8	87.5	87.5	73.3
LDA	66.7	62.5	55.6	88.9	55.6	37.5	25.0	56.7
SVM	88.9	62.5	66.7	88.9	100	62.5	25.0	71.7
KNN	77.8	62.5	33.3	77.8	100	75.0	37.5	66.7
AlexNet + FC8 + LDA	100	87.5	88.9	88.9	88.9	75.0	50.0	83.3
AlexNet + FC7 + LDA	100	87.5	88.9	100	100	87.5	25.0	85.0
AlexNet + FC6 + LDA	100	87.5	100	100	100	100	75.0	95.0
AlexNet + FC6 + SVM	88.9	87.5	88.9	100	100	87.5	37.5	85.0
AlexNet + FC6 + KNN	77.8	87.5	44.4	88.9	100	87.5	62.5	78.3
Average Accuracy for Each Emotion	87.7	73.6	66.7	92.6	91.4	77.8	47.2	77.2
3.1.2. KDEF dataset experiments
The KDEF Dataset contains 4900 facial expression images. As our system aims to quantify the evolution of emotion from front-view videos of facial expressions, we have only used front-view images. From the KDEF Dataset, we selected all 980 front-view images of facial expressions that were all processed using the images preprocessing techniques outlined above [25]. There are seven different facial expressions in this dataset: angry, happy, neutral, surprised, sad, fearful, and disgusted. In the experiment, 70% of the images were again randomly selected as the training images, with the rest of the images used as the testing images. Table 3 compares the 5-fold Cross Validation Error Rate for the facial expressions using the KDEF Dataset with each different method. The first column in the table shows the method used for facial expressions recognition, whilst the second column shows the 5-fold Cross Validation Error Rate for each method.
Table 3. Comparison cross-validation error rate using different method for the KDEF dataset.
Method	Error Rate (%)
AlexNet	15.1
LDA	36.3

SVM| 32.1
KNN| 56.0
AlexNet + FC8 + LDA| 17.2
AlexNet + FC7 + LDA| 12.0
AlexNet + FC6 + LDA| 11.6
AlexNet + FC6 + SVM| 12.6
AlexNet + FC6 + KNN| 32.6
As the table shows, our proposed method has the lowest error rate at 11.6%. However, as the number of images in the KDEF was greatly increased compared to the JAFFE Database, the performance of the AlexNet can be seen to be significantly improved, with an error rate now 4% higher than the proposed method. On the other hand, the traditional classifiers such as LDA and SVM did not show such a good performance when there were tested with such a large image database.
In addition, Table 4 shows the recognition accuracy for each emotion category, and the overall recognition accuracy when using each different method for the KDEF Dataset. The first column in the table shows the method used for facial expressions recognition whilst the second to eighth columns show the recognition accuracy of each method for angry, disgusted, fearful, happy, neutral, sad and surprised faces respectively. The ninth column shows the overall recognition accuracy of each method. Also, the last row shows the average recognition accuracy using all 9 methods for each emotion. The proposed method of using AlexNet, FC6 and LDA clearly demonstrates the highest recognition accuracy. Moreover, we can observe that ?happy? is still the easiest emotion to be recognized, while ?sad? and ?angry? are the two most difficult emotions to be recognized.
Table 4. Comparison of recognition accuracy of each emotion and overall recognition accuracy using different methods for the KDEF dataset.
Method	Angry (%)	Disgust (%)	Fear (%)	Happy (%)	Neutral (%)	Sad (%)	Surprise (%)	Overall Accuracy for Each Method (%)
AlexNet	76.2	81.0	92.9	97.6	88.1	81.0	76.2	84.7
LDA	47.6	69.0	50.0	83.3	71.4	50.0	71.4	63.3
SVM	57.1	66.7	66.7	81.0	54.8	52.4	76.2	65.0
KNN	38.1	35.7	31.0	47.6	50.0	35.7	57.1	42.2
AlexNet + FC8 + LDA	83.3	69.0	83.3	97.6	88.1	73.8	90.5	83.7
AlexNet + FC7 + LDA	83.3	81.0	85.7	100	85.7	78.6	85.7	85.7
AlexNet + FC6 + LDA	78.6	85.7	83.3	100	92.9	83.3	90.5	87.8
AlexNet + FC6 + SVM	78.6	83.3	90.5	97.6	88.1	78.6	88.1	86.4
AlexNet + FC6 + KNN	40.5	64.3	45.2	88.1	78.6	54.8	76.2	64.0
Average Accuracy for Each Emotion	64.8	70.6	69.8	88.1	77.5	65.4	79.1	73.6
3.1.3. CK+ dataset experiments
The CK+ Dataset consists of 593 sequences of facial expressions [27,28]. Each video sequences can be regarded as a few continuous video frames. As a result, this is a large database of around 10,000 images of facial expressions taken from 123 models. As these image sequences are continuous, there are many similar images. In our experiment, after removing similar images, 693 images were selected and processed using the image preprocessing techniques described. We selected images with seven different facial expressions in the dataset: angry, happy, neutral, surprised, sad, fearful, and disgusted. We again randomly selected 70% of the images as the training images, with the remainder employed as the testing images. Table 5 shows the 5-fold Cross Validation Error Rate for the facial expressions from the CK+ Dataset with

each different method. The first column in the table shows the method used for facial expressions recognition, whilst the second column shows the 5-fold Cross Validation Error Rate for each method.

Table 5. Comparison cross-validation error rate using different methods for the CK dataset.

Method	Error Rate (%)
AlexNet	10.9
LDA	10.7
SVM	8.4
KNN	24.2
AlexNet + FC8 + LDA	8.0
AlexNet + FC7 + LDA	5.2
AlexNet + FC6 + LDA	3.6
AlexNet + FC6 + SVM	6.7
AlexNet + FC6 + KNN	21.6

As the table shows, our proposed method still obtains the lowest error rate at 3.6%. Generally, in the CK+ Dataset, two images are selected for each emotion for each subject, with one image being the frame when the emotion begins to be expressed whilst the other image is the frame in the image sequence when the emotion reaches the peak of its expression. All the methods appear to have good performance in this dataset.

Additionally, Table 6 shows the recognition accuracy for each emotion, and the overall recognition accuracy rate when employing each different method for the CK+ Dataset. The first column in the table shows the method used for facial expressions recognition, with the next seven columns showing the recognition accuracy for each method for angry, disgusted, fearful, happy, neutral, sad and surprised facial images respectively. The ninth column shows the overall recognition accuracy of each method tested. Also, the last row shows the average recognition accuracy using all the 9 methods for each emotion. Within this dataset, ?sad? seems to be the most difficult emotion to be recognized, with several methods wrongly classifying the sad emotion as ?neutral?. The difficulty of recognition of sad emotion will be discussed in the discussion part.

Table 6. Comparison of recognition accuracy of each emotion and overall recognition accuracy using different method for the CK dataset.

Method	Angry (%)	Disgust (%)	Fear (%)	Happy (%)	Neutral (%)	Sad (%)	Surprise (%)	Overall Accuracy for Each Method (%)
AlexNet	95.7	85.7	81.8	97.4	76.2	33.3	97.8	86.4
LDA	78.3	88.6	72.7	97.4	88.1	33.3	91.1	85.4
SVM	73.9	91.4	81.8	92.1	95.2	66.7	88.9	87.9
KNN	65.2	77.1	63.6	65.8	59.5	33.3	64.4	64.1
AlexNet + FC8 + LDA	87.0	77.1	63.6	97.4	95.2	33.3	91.1	85.4
AlexNet + FC7 + LDA	91.3	82.9	72.7	97.4	97.6	33.3	93.3	88.3
AlexNet + FC6 + LDA	100	91.4	100	100	97.6	58.3	95.6	94.7
AlexNet + FC6 + SVM	91.3	80.0	81.8	97.4	97.6	41.7	97.8	89.8
AlexNet + FC6 + KNN	73.9	85.7	63.6	68.4	66.7	33.3	71.1	69.9
Average Accuracy for Each Emotion	84.1	84.4	75.7	90.4	86.0	40.7	87.9	83.5

3.1.4. FER2013 dataset experiments

The FER2013 is an online open dataset for facial expressions [29]. This dataset consists of more than 30,000 48×48 pixel grayscale images of faces. There are also seven types of emotions in this dataset: angry, disgusted,

fearful, happy, sad, surprised and neutral. However, unlike the JAFFE and KDEF datasets that are lab-posed facial expressions, the FER2013 dataset consists of facial images taken from the internet. In the JAFFE and KDEF databases, facial expression of the same emotional category are similar to each other. However, in the FER2013 dataset, very large variations in facial expression can be observed, within the same category of emotion. In the current experiment, about 30,000 images were selected. Image preprocessing techniques were not applied to this dataset, as the original 48×48 pixel grayscale images are already quite small. We tested with all seven categories of facial emotion available in the database, again randomly selecting 70% of the images as the training set and using the remainder of the dataset as the testing images.

Table 7 compares the 5-fold Cross Validation Error Rate for the facial expressions using the FER2013 Dataset with each different method. The first column in the table shows the method used for facial expressions recognition, whilst the second column shows the 5-fold Cross Validation Error Rate for each method tested. As the table shows, the proposed method still obtains the lowest error rate at 43.5%. However, as the FER2013 dataset is primarily drawn from the internet, with a broad range of individual variation in terms of expression within the same category of emotion, we can see from the data that all approaches demonstrated difficulty in accurate classification, with no method demonstrating good levels of performance.

Table 7. Comparison cross-validation error rate using different methods for the FER2013 dataset.

Method	Error Rate (%)
AlexNet	44.6
LDA	68.6
SVM	73.4
KNN	61.7
AlexNet + FC8 + LDA	50.2
AlexNet + FC7 + LDA	46.5
AlexNet + FC6 + LDA	43.5
AlexNet + FC6 + SVM	48.2
AlexNet + FC6 + KNN	49.6

Table 8 shows the recognition accuracy for each emotion category, and the overall recognition accuracy using each particular method with the FER2013 Dataset. The first column in the table shows the method used for facial expressions recognition, whilst the next seven columns show the recognition accuracy of each method for angry, disgusted, fearful, happy, neutral, sad and surprised faces respectively. The ninth column shows the overall recognition accuracy of each method. Also, the last row shows the average recognition accuracy using all 9 methods for each emotion. In this dataset, happy and surprised facial categories seem to be the simplest emotions to be recognized. However, even the method using the AlexNet, FC6 and LDA fails to show good performance in recognizing fear emotion.

Table 8. Comparison of recognition accuracy of each emotion and overall recognition accuracy using different methods for the FER2013 dataset.

Method	Angry (%)	Disgust (%)	Fear (%)	Happy (%)	Neutral (%)	Sad (%)	Surprise (%)	Overall Accuracy for Each Method (%)
AlexNet	42.0	43.8	34.2	81.3	58.2	43.0	65.6	56.3
LDA	18.7	14.6	18.4	47.6	28.9	21.1	43.9	30.8
SVM	14.7	27.0	13.0	38.0	20.5	25.6	40.1	26.2
KNN	28.0	54.0	31.7	38.0	38.8	30.9	52.4	36.5

AlexNet + FC8 + LDA| 37.5| 25.5| 24.6| 70.9| 50.1| 41.5| 65.3| 49.8
AlexNet + FC7 + LDA| 42.3| 32.1| 28.5| 72.3| 55.8| 46.4| 67.2| 53.5
AlexNet + FC6 + LDA| 43.8| 36.5| 30.7| 74.9| 59.9| 52.1| 67.5| 56.4
AlexNet + FC6 + SVM| 42.2| 41.6| 36.9| 70.0| 45.6| 40.3| 69.5| 51.7
AlexNet + FC6 + KNN| 40.1| 59.9| 42.1| 61.6| 43.0| 39.5| 66.9| 49.5
Average Accuracy for Each Emotion| 34.5| 35.8| 28.9| 62.1| 44.3| 37.5| 60.9| 45.8

3.1.5. AffectNet dataset experiments

The AffectNet is also an online dataset which consists of about one million images of facial expressions which were again collected from the Internet, through three major search engines and 1250 emotion related keywords [30]. This dataset provides eleven emotion and non-emotion labels including: Neutral, Surprised, Happy, Sad, Fearful, Disgusted, and Angry, whilst additionally providing the categories of Contemptuous, None, Uncertain and No-Face. In line with the FER2013 Dataset, the AffectNet contains the images of facial expressions which are naturally occurring rather than containing the images posed within a lab. In the current study, about 16,000 images were selected. Image preprocessing techniques were not applied to this dataset. We selected the seven emotional expressions that were in line with the previous four datasets tested and again we randomly selected 70% of the images to serve as the training images, while the remaining images were again employed as testing images. Table 9 compares the 5-fold Cross Validation Error Rate for the facial expressions using the AffectNet Dataset with each different method. The first column in the table shows the method used for facial expressions recognition, whilst the second column shows the 5-fold Cross Validation Error Rate for each method.

Table 9. Comparison cross-validation error rate using different methods for the AffectNet dataset.

Method	Error Rate (%)
AlexNet	40.81
LDA	59.97
SVM	60.92
KNN	68.54
AlexNet + FC8 + LDA	48.02
AlexNet + FC7 + LDA	43.55
AlexNet + FC6 + LDA	39.43
AlexNet + FC6 + SVM	45.28
AlexNet + FC6 + KNN	59.81

As the table shows, the proposed method has the lowest error rate at 39.43%. As the AffectNet dataset is similar to the FER2013 dataset in employing images of facial expressions taken from the Internet, the huge variance within the same class of emotion can again be seen to result in low recognition accuracy rates.

Table 10 shows the recognition accuracy of each emotion category, and the overall recognition accuracy using each different method with the AffectNet Dataset. The first column in the table shows the method used for facial expressions recognition, whilst the second to the eighth columns show the recognition accuracy for each method tested for angry, disgusted, fearful, happy, neutral, sad and surprised emotional expressions respectively. The ninth column shows the overall recognition accuracy for each method. Also, the last row shows the average recognition accuracy using all 9 methods for each emotion. In general, it can be seen that these methods show their best performance in recognizing happy emotions, which is similar to the performance

observed with the FER2013 dataset. On the other hand, each method demonstrates its worst performance in recognizing the emotion of surprise.

Table 10. Comparison of recognition accuracy of each emotion and overall recognition accuracy using different methods for the AffectNet dataset.

Method| Angry (%)| Disgust (%)| Fear (%)| Happy (%)| Neutral (%)| Sad (%)| Surprise (%)| Overall Accuracy for Each Method (%)
---|---|---|---|---|---|---|---
AlexNet| 58.2| 22.0| 14.3| 85.6| 50.5| 17.0| 36.3| 58.6
LDA| 17.8| 4.3| 8.9| 60.8| 38.5| 14.3| 14.4| 38.9
SVM| 20.8| 9.1| 12.5| 61.8| 37.3| 19.4| 16.1| 40.3
KNN| 14.5| 8.1| 15.6| 43.1| 34.0| 19.4| 7.6| 30.7
AlexNet + FC8 + LDA| 25.1| 5.4| 19.6| 78.8| 53.1| 17.0| 26.9| 52.1
AlexNet + FC7 + LDA| 33.1| 11.8| 22.8| 78.6| 59.2| 25.1| 34.0| 56.0
AlexNet + FC6 + LDA| 36.2| 12.4| 25.4| 83.2| 64.1| 32.7| 32.0| 60.1
AlexNet + FC6 + SVM| 38.0| 15.6| 31.3| 77.2| 49.1| 32.3| 32.0| 54.6
AlexNet + FC6 + KNN| 18.8| 8.6| 19.2| 59.4| 35.6| 21.4| 17.6| 39.2
Average Accuracy for Each Emotion| 29.2| 10.8| 18.8| 69.8| 46.8| 22.1| 24.1| 47.8

3.2. Experiment using the author's facial expressions

In this experiment, our proposed framework was used to recognize facial expression taken from a video of facial expressions emitted by the first author as the ground truth of emotions are perfectly known in this case. Within the video stimulus there were 898 continuous frames, which were preprocessed by the techniques previously outlined. The framework mainly aimed to recognize 5 kinds of emotion in the video stimulus: angry, happy, neutral, surprised, and sad. The training images used 20% of the frames from the video, combined with another dataset emitted by the author containing about 2600 images, whilst the remaining 80% images from the video acted as the testing images in this case. In the subsequent testing, the proposed system successfully classified the images into five facial expression categories: angry, happy, neutral, sad and surprised. In this experiment, as Fig. 4 illustrates, the accuracy was about 96.0%.

1. Download: Download high-res image (195KB)
2. Download: Download full-size image

Fig. 4. Recognition results using the proposed framework.

Fig. 5 shows the probability of each type of emotion, which is made up of predicted results for each frame in the video. The graph displays the evolution of facial expressions from the video and changes in facial expressions over time. In the graph, the probability of the five facial expression categories for every frame was predicted by the proposed framework. Each line represents the evolution of one emotion. In Fig. 5, the five lines show the evolution of five emotions over a period of time. In addition, the plot was smoothed by calculating the average of the recent frames.

1. Download: Download high-res image (611KB)
2. Download: Download full-size image

Fig. 5. Facial expression analysis for the video stimulus of the author own emotional expressions.

As an electrocardiogram can reflect the electrical activity of the heart, Fig. 5 reflects the mental state of the patient/ user over a period of time. As shown in Fig. 5, around the 300th frame, there are three emotions: happy, neutral and sad. The evolutions of the three emotions at that time are also shown. In addition, this figure illustrates when the user was said to be happy, the duration of the emotion, and how quickly it reached the peak of the emotion. By data analysis, the plot also shows the relative percentage of time

for each emotion over the period. As a result, in the area of mental health care, the proposed system clearly has the potential to identify the emotional state of the user. As patients with severe cognitive impairments may express abnormal facial expressions [21], the proposed system may have the potential to be used diagnostically in the future.

4\. Discussion

Facial expressions are an important component of human social interaction, which reflect people's emotions, attitudes, social relations and physiological state. The automated facial expression analysis can play an important role in human-computer interaction in many important applications. On the other hand, deep learning is a dynamic and vibrant topic, encompassing diverse and useful applications, such as the use of A DSAE-based deep learning framework for facial expression recognition and the use of a deep-belief-network-based particle Filter for Analysis of Gold Immunochromatographic Strips [50], [51], [52]. In this work, we have proposed a deep learning based facial expression recognition framework towards mental health care.

In the experiments presented here we have tested the facial expression recognition performance of nine methods, including deep convolution neural network AlexNet, the traditional classifiers SVM, LDA and KNN and the combination of the AlexNet and traditional classifiers against five datasets. The overall performance can be seen in Table 11, which shows the 5-fold cross validation error rate of each method for the five datasets employed in our testing. The first column in the table shows the method used for facial expressions recognition, while the second column to the sixth columns show the error rate of each method for JAFFE, KDEF, CK+, FER2013 and AffectNet respectively.

Table 11. Comparison cross-validation error rate using different methods over 5 databases.

Method	JAFFE	KDEF	CK+	FER2013	AffectNet
AlexNet	24.8	15.1	10.9	44.6	40.8
LDA	26.7	36.3	10.7	68.6	60.0
SVM	24.8	32.1	8.4	73.4	60.9
KNN	39.6	56.0	24.2	61.7	68.5
AlexNet + FC8 + LDA	18.8	17.2	8.0	50.2	48.0
AlexNet + FC7 + LDA	9.4	12.0	5.2	46.5	43.6
AlexNet + FC6 + LDA	5.9	11.6	3.6	43.5	39.4
AlexNet + FC6 + SVM	10.9	12.6	6.7	48.2	45.3
AlexNet + FC6 + KNN	15.4	32.6	21.6	49.6	59.8

Our experiments demonstrate that in a small dataset like the JAFFE and CK+ datasets, the deep convolution neural network AlexNet and some traditional classifiers like SVM and LDA have similar facial expression recognition accuracy. However, by combining the AlexNet with traditional classifiers like SVM and LDA, the recognition accuracy increases, especially when we extract the deep features from FC6. The experiments show that the method that extracts features from FC6 has better performance than FC7 and FC8 in the five online datasets. We use the AlexNet which is pre-trained with one million natural images from ImageNet to extract features. The features extracted from FC7 are more in line with the classification attribute of the training set of natural images but less in accordance with the dataset of facial expressions [53]. As a result, the methods that extract the features from FC6 have better performance.

Additionally, we have observed that using the LDA to classify the deep features results in better performance. When the number of the images in the

training dataset increased, such as in the case of the KDEF dataset (which contains about 1000 images), the AlexNet seemed to have better recognition accuracy relatively to its performance when tested with a small image database. On the other hand, we have observed that classifying facial expressions with traditional classifiers did not show good overall performance. As a result, in the KDEF dataset, the recognition accuracy of the combination of the AlexNet and traditional classifiers only demonstrated slightly better performance than that of the AlexNet. In addition, it is noticed that the overall recognition accuracy for the FER2013 and AffectNet databases are lower than that of the JAFFE, KDEF and CK+ databases. We would argue that this is mainly as a result of the FER2013 and AffectNet databases containing the more challenging facial expressions sourced from the Internet, which have huge variance within a given class of emotion for example the difference in image sizes and lighting situations. These facial expressions are more natural and diverse, resulting in stimuli that are more difficult to recognize than the lab-based, controlled and actor-generated facial expressions. Although the recognition performance for the AffectNet is worse than the performance in the JAFFE, KDEF and CK+ databases, the proposed framework has a relatively good performance for the AffectNet database compared to the other state-of-art facial expression recognition algorithms [54,55]. Fig. 6 shows the accumulative recognition error rate for the nine methods over the five databases.

1. Download: Download high-res image (160KB)
2. Download: Download full-size image

Fig. 6. Comparison of accumulative cross-validation error rates using different methods over 5 databases.

In general, the method that extracts the deep features using FC6 from the AlexNet and classifies with LDA showed the best overall performance in the five databases tested using these nine methods. Our experiment result showed a facial expression recognition accuracy of 94.1% on the JAFFE database and 88.4% on the KDEF database, which is a relatively good performance compared to other facial expression recognition algorithms tested on the same databases, as shown in Table 12. The recognition accuracy is calculated from the error rate.

Table 12. Recognition accuracy from published papers on KDEF and JAFFE datasets.

System	Accuracy for KDEF	Accuracy for JAFFE
DeepPCA [56]	83.0%	/
AAM+SVM [57]	74.6%	/
Feature+SVM [58]	82.4%	/
C+CNN [59]	/	91.6%
HF [60]	/	87.1%
Proposed Method (5-fold cross validation recognition accuracy)	88.4%	94.1%

In order to further estimate the performance of the proposed method, we compared the proposed method (AlexNet + FC6 + LDA) with some state-of-the-art deep CNN including pure AlexNet, VGG16, GoogleNet and ResNet. We estimated the performance mainly with regard to the operating time of training the network, recognizing the facial expressions and in terms of the recognition accuracy of the facial expression categories. Three facial expression datasets including the JAFFE, KDEF and CK+ Datasets were used in this estimation. Fig. 7 shows both the recognition accuracy and the operating time for each method with the JAFFE, KDEF and CK+ Datasets. In Fig. 7, the recognition accuracy is shown as clustered columns, with the operating time superimposed as a line chart. We

can observe that the proposed method has high recognition accuracy compared to the other deep learning algorithms, but slightly lower than that of the ResNet. However, this should be considered in light of the clear reduction in operating time, as the operating time of the proposed method is around 100 times shorter than that of the ResNet. In addition, it is important to note that deep learning algorithms have high device requirements relating to GPU resources and local dynamic random-access memory requirements. Indeed in the current assessment, the Vgg16 failed to produce a recognition result in the KDEF and CK+ datasets due to insufficiency in memory resource to complete the task. In general, the proposed method can be seen to have relatively good recognition accuracy, much shorter operating time and low device requirements compared to the state-of-art deep learning algorithms.

1. Download: Download high-res image (521KB)
2. Download: Download full-size image

Fig. 7. Comparison of recognition accuracy and operating time for different methods in the JAFFE, KDEF and CK+ datasets.

Another important aspect is the ratio of training images to testing the images employed in assessments. This can enable us to determine the influence of training image ratios for different algorithms on the selected dataset. In the experiments presented, we selected 70% of the images randomly from the whole dataset as the training dataset and used the remaining images as the testing dataset. Here we choose to test the JAFFE Database. Fig. 8 shows the different ratios of the training images with the different methods for the JAFFE Database and the resulting recognition accuracy. We observe that when the training images ratios increases, the recognition accuracy for all the methods is increased. The result also suggests that the proposed method using the AlexNet, FC6 and LDA demonstrates the best performance, regardless of the training images ratio examined. Indeed, when we select 90% of the images randomly from the JAFFE Database to act as the training images, and use the remaining images as the testing dataset, the recognition accuracy of the proposed method reaches 97.0%. We can assume that when the training images ratios increase, the recognition performance for other datasets will also be improved.

1. Download: Download high-res image (301KB)
2. Download: Download full-size image

Fig. 8. The influence of the training images ratios on recognition accuracy with the different algorithms for the JAFFE database.

In the experiment, the first phase of the proposed method involves image pre-processing, including aspects such as removing unnecessary environment aspects in the image, identification of the facial region in the image, and some additional essential pre-processing steps. To test the efficiency of this phase we also conducted experiments to compare the 5-fold cross validation error rate using each particular method, for both processed and unprocessed images from the CK+ dataset. The results of these experiments are shown in Table 13 below. The first column in the table shows the method used for facial expressions recognition, with the second and third columns showing the 5-fold Cross-Validation Error Rate of each method for both processed and unprocessed images, respectively.

Table 13. Comparison cross-validation error rate using different methods for processed and unprocessed images from the CK+ dataset.

Method	Error Rate (%) for Processed Data	Error Rate (%) for Unprocessed Data
AlexNet	10.90	12.72
LDA	10.74	26.42

SVM| 8.42| 23.80
KNN| 24.24| 20.46
AlexNet + FC8 + LDA| 7.98| 11.76
AlexNet + FC7 + LDA| 5.22| 8.56
AlexNet + FC6 + LDA| 3.63| 3.92
AlexNet + FC6 + SVM| 6.68| 8.85
AlexNet + FC6 + KNN| 21.63| 19.74

As Table 13 shows, the error rate is decreased in all the algorithms for the processed dataset, which proves the efficiency of the image-preprocessing phase. Additionally, we can observe that the image pre-processing phase improves the error rate performance of the traditional classifier to a greater extent, but the combination of the AlexNet and the traditional classifiers seems less dependent on this phase.

Finally, we report on the performance relating to the recognition of each kind of facial expression category. Confusion matrices (a), (b), (c), (d) and (e) in Fig. 9 show the data relating to the AlexNet + FC6 LDA for the Datasets of JAFFE, KDEF, CK+, FER2013 and AffectNet respectively. Here, it can be observed that the emotional category ?happy? appears to be the easiest emotion to be recognized. On the other hand, some emotional categories like ?sad? and ?disgust? seems to be much more difficult to be recognized. We would argue that the sad emotion is currently relatively hard to be recognized for the following two reasons. The first reason is due to the quantity of images depicting the sad emotion within the training dataset. We have observed that in most of the datasets, there are notably fewer images of sad emotions, which consequently increases the recognition difficulty. For example, within the CK+ dataset there are 693 images in total, whilst only 40 images depict sad expressions. Additionally, there are relatively small visual differences between sad emotions and neutral emotions, and consequently sad expressions may be recognized as neutral by the system without using sufficient training data.

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

Fig. 9. The confusion matrix using the AlexNet + FC6 LDA for the datasets of JAFFE, KDEF, CK+, FER2013 and AffectNet are shown in (a), (b), (c), (d) and (e).

Furthermore, we notice that in the datasets using the images from the Internet, such as the FER2013 and AffectNet datasets, the number of images for each emotion category is uneven. The predominant emotion type contained within these two datasets are happy images, which consequently reduces the difficulty of recognition of the happy emotion. However, in the case of the other databases, which contain a more balanced range of emotions, the emotional category with the lowest error rate varies.

As the developed facial expression recognition method aims to benefit the mental health care, we further test the proposed method with the datasets that contain the images of facial expressions from the patients with cognitive impairment. We tested the proposed method in the following three datasets: a dataset with facial expressions from the patient with cognitive impairment, a dataset that combines the images from the JAFFE dataset and the images from patients and a dataset that combines the images from the KDEF dataset and the images form patients. We used 70% of images as the training dataset and 30% of images as the testing dataset. Table 14 shows the recognition accuracy using the proposed method in the three datasets. The experimental result shows that the proposed method also has a good performance for the facial expressions from the patients with cognitive impairment.

Table 14. The recognition accuracy using the proposed algorithm (AlexNet + FC6 + LDA) in the datasets containing facial expressions from patient with cognitive impairment.

Dataset	Recognition Accuracy (%)
Patient with Cognitive Impairment Dataset	85.13
JAFFE + Patient Dataset	89.90
KDEF + Patient Dataset	89.33

In our examination of the data relating to recognition when employing the first author's facial expressions, drawn from a series of continuous video frames, the accuracy rate was approximately 96.0%.We would argue that the system achieved strong performance in recognizing facial expressions taken from videos in this context. It should be noted that one factor that contributed to the high accuracy rate may be that some of the training images and the testing images were selected from a video that had the similar lighting condition and viewpoint. In addition, there was only one participant and one viewpoint. However, the analysis of facial expressions from the videos in our testing appears to be able to evaluate the evolution of the facial expression over a period of time, with changes of the expression emitted, by recognizing facial expressions for each frame in the video. Additionally, the system was also able to quantify the extent of the facial expression. However, there remain some practical issues to be considered. First, it is noticed that the differences in facial expression for happy and neutral, neutral and sad, sad and angry are small. Moreover, for the video of the facial expressions emitted by the first author, the facial expression is currently labelled manually, and there may be potential problems in labelling some frames during a change of expressions. In addition, as stated previously, it was noted that in the JAFFE dataset, some facial expressions appeared to be labelled incorrectly and were removed prior to testing. In the experiment, 202 images of facial expressions were used from the JAFFE, while the original JAFFE dataset has about 213 images. Finally, there are some practical problems for the proposed facial expression recognition system that remain to be addressed in future, including a need for enhanced stability with regard to how the system crops the head area in the images.

From the perspective of clinical practice, in order to use the system within elderly people to detect mental health issues like cognitive impairment, the framework needs to be trained with large samples of natural facial expressions taken from elderly people, in order to achieve optimal performance. In recent work, we have collected circa 100, 000 images of facial expressions from elderly people. The recognition of naturally occurring facial expression is more challenging. We should also note that work to complete the processing of these images is ongoing, and we will report in a future work regarding satisfactory verification of these images as a suitable training set for our network.

5\. Conclusion

This paper has presented a novel emotion analysis framework that is able to understand and automatically recognize users? emotions by analyzing the users? facial expression from their facial images. The system consists of three parts: input of the videos of facial expressions, the image pre-processing technique, and automatic facial expression analysis. The system is able to successfully conduct image pre-processing and facial expression analysis. Additionally, after facial expression analysis has been undertaken, it is able to understand the evolution of facial expression over a period of time and can quantify the extent of the emotions detected from a facial video. As facial

expressions reflect people's mental health state, the proposed framework has great potential to be employed within mental health care.

For the facial expression analysis, this paper has also proposed a new solution that extracts the deep features from the FC6 of the AlexNet whilst the standard LDA is exploited to train these deep features. The proposed solution shows promising and has stable performance on all the five tested datasets for the nine methods studied. Additionally, we would argue that the proposed method has relatively good recognition accuracy, much less operating time and lower device requirements compared to the other current state-of-the-art deep learning algorithms. Furthermore, the analysis of facial expressions when taken from the videos demonstrated that our proposed approach is able to report the evolution of facial expression over a period of time, and reliably detect the changes of expression by means of the recognition of facial expression within each frame in the video.

CRediT authorship contribution statement

Zixiang Fei: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Erfu Yang:** Conceptualization, Methodology, Investigation, Resources, Writing - review & editing, Supervision, Funding acquisition. **David Day-Uei Li:** Writing - review & editing, Supervision. **Stephen Butler:** Writing - review & editing, Supervision. **Winifred Ijomah:** Writing - review & editing, Supervision. **Xia Li:** Investigation, Resources, Supervision. **Huiyu Zhou:** Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

The research work is funded by Strathclyde's Strategic Technology Partnership (STP) Programme with CAPITA (2016?2019). We thank Dr. Neil Mackin and Miss Angela Anderson for their support. The contents in the paper are those of the authors alone and don't stand for the views of CAPITA plc. Huiyu Zhou was partly funded by UK EPSRC under Grant EP/N011074/1, and Royal Society in Newton Advanced Fellowship under Grant NA160342. Winifred Ijomah is supported under EPSRC (Grant no. EP/N018427/1). The authors thank Shanghai Mental Health Center for their help and support. We also thank Dr. Fei Gao from Beihang University, China for his kind support and comment.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors without the proper ethical approval.

Recommended articles

References

1. [1]
N. Yildirm, A. Varol
A research on estimation of emotion using EEG signals and brain computer interfaces
Proceedings of the 2nd International Conference on Computer Science and Engineering UBMK'20 (2017), pp. 1132-1136, 10.1109/UBMK.2017.8093523
View in ScopusGoogle Scholar

2. [2]
A. Oliveira, C. Pinho, S. Monteiro, A. Marcos, A. Marques
Usability testing of a respiratory interface using computer screen and facial expressions videos
Comput. Biol. Med., 43 (2013), pp. 2205-2213, 10.1016/j.compbiomed.2013.10.010
View PDFView articleView in ScopusGoogle Scholar

3. [3]
M.-T. Yang, Y.-J. Cheng, Y.-C. Shih, Facial expression recognition for learning status analysis, 2011. doi:10.1007/978-3-642-21619-0_18.
Google Scholar
4. [4]
G.S. Bahr, C. Balaban, M. Milanova, H. Choe, Nonverbally smart user interfaces: postural and facial expression data in human computer interaction, 2007.
Google Scholar
5. [5]
P. Branco, P. Firth, L.M. Encarnação, P. Bonato
Faces of emotion in human-computer interaction
Proceedings of the Conference on Human Factors in Computing Systems (2005), pp. 1236-1239, 10.1145/1056808.1056885
View in ScopusGoogle Scholar
6. [6]
O. Grynszpan, J.-C. Martin, J. Nadel, Using facial expressions depicting emotions in a human-computer interface intended for people with autism, 2005. doi:10.1007/11550617_41.
Google Scholar
7. [7]
D.K. Jain, P. Shamsolmoali, P. Sehdev
Extended deep neural network for facial emotion recognition
Pattern Recognit. Lett., 120 (2019), pp. 69-74, 10.1016/J.PATREC.2019.01.008
View PDFView articleView in ScopusGoogle Scholar
8. [8]
J. Shao, Y. Qian
Three convolutional neural network models for facial expression recognition in the wild
Neurocomputing, 355 (2019), pp. 82-92, 10.1016/j.neucom.2019.05.005
View PDFView articleView in ScopusGoogle Scholar
9. [9]
C. Lv, Z. Wu, X. Wang, M. Zhou
3D facial expression modeling based on facial landmarks in single image
Neurocomputing, 355 (2019), pp. 155-167, 10.1016/J.NEUCOM.2019.04.050
View PDFView articleView in ScopusGoogle Scholar
10. [10]
M.Z. Uddin, J.J. Lee, T.-S. Kim
An enhanced independent component-based human facial expression recognition from video
IEEE Trans. Consum. Electron., 55 (2009), pp. 2216-2224, 10.1109/TCE.2009.5373791
View in ScopusGoogle Scholar
11. [11]
J.J. Lee, Z. Uddin, T.-S. Kim
Spatiotemporal human facial expression recognition using fisher independent component analysis and Hidden Markov Model
Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (2008), pp. 2546-2549
EMBS?08 - "Personalized Healthc. through Technol
CrossrefView in ScopusGoogle Scholar
12. [12]
D. Engin, C. Ecabert, H.K. Ekenel, J.P. Thiran
Face frontalization for cross-pose facial expression recognition

Eur. Signal Process. Conf., European Signal Processing Conference, EUSIPCO (2018), pp. 1795-1799, 10.23919/EUSIPCO.2018.8553087
Google Scholar
13. [13]
W. Sun, H. Zhao, Z. Jin
An efficient unconstrained facial expression recognition algorithm based on Stack Binarized Auto-Encoders and Binarized Neural Networks
Neurocomputing, 267 (2017), pp. 385-395, 10.1016/J.NEUCOM.2017.06.050
View PDFView articleView in ScopusGoogle Scholar
14. [14]
A. Samal, P.A. Iyengar
Automatic recognition and analysis of human faces and facial expressions: a survey
Pattern Recognit. (1992), p. 25, 10.1016/0031-3203(92)90007-6
Google Scholar
15. [15]
B. Fasel, J. Luettin
Automatic facial expression analysis: a survey
Pattern Recognit., 36 (2003), 10.1016/S0031-3203(02)00052-3
Google Scholar
16. [16]
G. Sandbach, S. Zafeiriou, M. Pantic, L. Yin
Static and dynamic 3D facial expression recognition: a comprehensive survey
Image Vis. Comput., 30 (2012), 10.1016/j.imavis.2012.06.005
Google Scholar
17. [17]
Z. Zeng, M. Pantic, G.I. Roisman, T.S. Huang
A survey of affect recognition methods: audio, visual, and spontaneous expressions
IEEE Trans. Pattern Anal. Mach. Intell., 31 (2009), 10.1109/TPAMI.2008.52
Google Scholar
18. [18]
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 (2015), 10.1109/TPAMI.2014.2366127
Google Scholar
19. [19]
B. Ko, B.Chul Ko
A brief review of facial emotion recognition based on visual information
Sensors, 18 (2018), p. 401, 10.3390/s18020401
View in ScopusGoogle Scholar
20. [20]
S. Baron-Cohen, A. Riviere, M. Fukushima, D. French, J. Hadwin, P. Cross, C. Bryant, M. Sotillo
Reading the mind in the face: a cross-cultural and developmental study
Vis. Cogn., 3 (1996), pp. 39-59
CrossrefView in ScopusGoogle Scholar
21. [21]
K. Burton, A. Kaszniak
Emotional experience and facial expression in Alzheimer's disease
Aging, Neuropsychol. Cogn., 13 (2006), 10.1080/13825580600735085
Google Scholar
22. [22]

S. Passardi, P. Peyk, M. Rufer, T.S.H. Wingenbach, M.C. Pfaltz
Facial mimicry, facial emotion recognition and alexithymia in post-traumatic stress disorder
Behav. Res. Ther., 122 (2019), 10.1016/j.brat.2019.103436
Google Scholar
23. [23]
J.D. Henry, P.G. Rendell, A. Scicluna, M. Jackson, L.H. Phillips
Emotion experience, expression, and regulation in Alzheimer's disease
Psychol. Aging., 24 (2009), 10.1037/a0014001
Google Scholar
24. [24]
M.C. Smith
Facial expression in mild dementia of the Alzheimer type
Behav. Neurol., 8 (1995), pp. 149-156
View in ScopusGoogle Scholar
25. [25]
A. Lundqvist, D. Flykt, A. Öhman, The Karolinska Directed Emotional Faces KDEF, CD ROM from Department of Clinical Neuroscience, Psychology section, Karolinska Institutet, 1998.
Google Scholar
26. [26]
M. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, Coding facial expressions with Gabor wavelets, in: Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition, IEEE Computer Society, n.d.: pp. 200?205. doi:10.1109/AFGR.1998.670949.
Google Scholar
27. [27]
T. Kanade, J.F. Cohn, Yingli Tian, Comprehensive database for facial expression analysis, in: Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580), IEEE Computer Society, n.d.: pp. 46?53. doi:10.1109/AFGR.2000.840611.
Google Scholar
28. [28]
P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews
The Extended Cohn?Kanade Dataset (CK+): a complete dataset for action unit and emotion-specified expression
Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE (2010), pp. 94-101, 10.1109/CVPRW.2010.5543262
View in ScopusGoogle Scholar
29. [29]
Challenges in representation learning: facial expression recognition challenge | Kaggle, (n.d.). https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/overview(accessed April 15, 2019).
Google Scholar
30. [30]
A. Mollahosseini, B. Hasani, M.H. Mahoor
AffectNet: a database for facial expression, valence, and arousal computing in the wild
IEEE Trans. Affect. Comput., 10 (2019), pp. 18-31, 10.1109/TAFFC.2017.2740923
View in ScopusGoogle Scholar
31. [31]
P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, n.d.: p. I-511-I-518.

doi:10.1109/CVPR.2001.990517.
Google Scholar
32. [32]
Detect objects using the Viola-Jones algorithm - MATLAB - MathWorks United Kingdom, (n.d.).
https://uk.mathworks.com/help/vision/ref/vision.cascadeobjectdetector-system-object.html(accessed January 29, 2019).
Google Scholar
33. [33]
Concatenate arrays - MATLAB cat - MathWorks United Kingdom, (n.d.).
https://uk.mathworks.com/help/matlab/ref/cat.html(accessed May 7, 2019).
Google Scholar
34. [34]
Pretrained AlexNet convolutional neural network - MATLAB alexnet - MathWorks United Kingdom, (n.d.).
https://uk.mathworks.com/help/deeplearning/ref/alexnet.html;jsessionid=c4669357f290858d36ed1bcbf8cf(accessed
September 21, 2018).
Google Scholar
35. [35]
M. Matsugu, K. Mori, Y. Mitari, Y. Kaneda
Subject independent facial expression recognition with robust face detection using a convolutional neural network
Neural Netw., 16 (2003), pp. 555-559, 10.1016/S0893-6080(03)00115-1
View PDFView articleView in ScopusGoogle Scholar
36. [36]
R. Collobert, J. Weston
A unified architecture for natural language processing
Proceedings of the 25th International Conference on Machine Learning - ICML ?08, New York, New York, USA, ACM Press (2008), pp. 160-167, 10.1145/1390156.1390177
Google Scholar
37. [37]
A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, n.d.http://code.google.com/p/cuda-convnet/(accessed September 21, 2018).
Google Scholar
38. [38]
X. Chen, X. Yang, M. Wang, J. Zou
Convolution neural network for automatic facial expression recognition
Proceedings of the International Conference Applied System Innovation, IEEE (2017), pp. 814-817, 10.1109/ICASI.2017.7988558
View in ScopusGoogle Scholar
39. [39]
K. Shan, J. Guo, W. You, D. Lu, R. Bie
Automatic facial expression recognition based on a deep convolutional-neural-network structure
Proceedings of the IEEE 15th International Conference on Software Engineering Research, Management and Application, IEEE (2017), pp. 123-128, 10.1109/SERA.2017.7965717
View in ScopusGoogle Scholar
40. [40]
X. Han, Y. Zhong, L. Cao, L. Zhang
Pre-trained alexnet architecture with pyramid pooling and supervision for high

spatial resolution remote sensing image scene classification
Remote Sens. (2017), p. 9, 10.3390/rs9080848
View in ScopusGoogle Scholar
41. [41]
U. Chavan, D. Kulkarni
Optimizing deep convolutional neural network for facial expression recognitions
Advances in Intelligent Systems and Computing, Springer Verlag (2019), pp. 185-196, 10.1007/978-981-13-1402-5_14
View in ScopusGoogle Scholar
42. [42]
Feature Extraction Using AlexNet - MATLAB & Simulink - MathWorks United Kingdom, (n.d.). https://uk.mathworks.com/help/deeplearning/examples/feature-extraction-using-alexnet.html;jsessionid=a9dd0dd508fd2b96854cd6f11d5c(accessed January 23, 2019).
Google Scholar
43. [43]
P. McAllister, H. Zheng, R. Bond, A. Moorhead
Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets
Comput. Biol. Med., 95 (2018), pp. 217-233, 10.1016/J.COMPBIOMED.2018.02.008
View PDFView articleView in ScopusGoogle Scholar
44. [44]
Visualize High-Dimensional Data Using t-SNE - MATLAB & Simulink - MathWorks United Kingdom, (n.d.). https://uk.mathworks.com/help/stats/visualize-high-dimensional-data-using-t-sne.html(accessed April 15, 2019).
Google Scholar
45. [45]
L. van der Maaten, Barnes-Hut-SNE, (2013).
http://arxiv.org/abs/1301.3342(accessed April 15, 2019).
Google Scholar
46. [46]
Visualize Activations of a Convolutional Neural Network - MATLAB & Simulink - MathWorks United Kingdom, (n.d.).
https://uk.mathworks.com/help/deeplearning/examples/visualize-activations-of-a-convolutional-neural-network.html(accessed January 23, 2019).
Google Scholar
47. [47]
Q. Ye, N. Ye, T. Yin
Fast orthogonal linear discriminant analysis with application to image classification
Neurocomputing, 158 (2015), pp. 216-224, 10.1016/J.NEUCOM.2015.01.045
View PDFView articleView in ScopusGoogle Scholar
48. [48]
N.A.A. Shashoa, N.A. Salem, I.N. Jleta, O. Abusaeeda
Classification depend on linear discriminant analysis using desired outputs
Proceedings of the 17th International Conference Science Technology and Automation Control Computing Engineering, IEEE (2016), pp. 328-332, 10.1109/STA.2016.7952041
View in ScopusGoogle Scholar
49. [49]
Compute convolutional neural network layer activations - MATLAB activations - MathWorks United Kingdom, (n.d.).
https://uk.mathworks.com/help/deeplearning/ref/activations.html(accessed

January 30, 2019).
Google Scholar
50. [50]
N. Zeng, H. Zhang, B. Song, W. Liu, Y. Li, A.M. Dobaie
Facial expression recognition via learning deep sparse autoencoders
Neurocomputing, 273 (2018), pp. 643-649, 10.1016/j.neucom.2017.08.043
View PDFView articleView in ScopusGoogle Scholar
51. [51]
N. Zeng, Z. Wang, H. Zhang, K.-E. Kim, Y. Li, X. Liu
An improved particle filter with a novel hybrid proposal distribution for quantitative analysis of gold immunochromatographic strips
IEEE Trans. Nanotechnol., 18 (2019), pp. 819-829, 10.1109/tnano.2019.2932271
View in ScopusGoogle Scholar
52. [52]
N. Zeng, Z. Wang, H. Zhang, W. Liu, F.E. Alsaadi
Deep belief networks for quantitative analysis of a gold immunochromatographic strip
Cognit. Comput., 8 (2016), pp. 684-692, 10.1007/s12559-016-9404-x
View in ScopusGoogle Scholar
53. [53]
L. Ding, H. Li, C. Hu, W. Zhang, S. Wang
Alexnet feature extraction and multi-kernel learning for object-oriented classification
Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences- ISPRS Archives., International Society for Photogrammetry and Remote Sensing (2018), pp. 277-281, 10.5194/isprs-archives-XLII-3-277-2018
View in ScopusGoogle Scholar
54. [54]
Y. Li, J. Zeng, S. Shan, X. Chen
Occlusion aware facial expression recognition using CNN with attention mechanism
IEEE Trans. Image Process., 28 (2019), pp. 2439-2450, 10.1109/TIP.2018.2886767
View in ScopusGoogle Scholar
55. [55]
S. Jyoti, G. Sharma, A. Dhall
A single hierarchical network for face, action unit and emotion detection
Proceedings of the Digital Image Computing: Techniques and Applications, IEEE (2018), pp. 1-8, 10.1109/DICTA.2018.8615852
Google Scholar
56. [56]
K. Rujirakul, C. So-In
Histogram equalized deep pca with ELM classification for expressive face recognition
Proceedings of the International Workshop on Advanced Image Technology, IEEE (2018), pp. 1-4, 10.1109/IWAIT.2018.8369725
View in ScopusGoogle Scholar
57. [57]
P. Lucey, J.F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, I. Matthews
The extended Cohn?Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression
Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops CVPRW (2010), 10.1109/CVPRW.2010.5543262
Google Scholar

58. [58]
T. Jabid, M.H. Kabir, O. Chae
Robust facial expression recognition based on local directional pattern
ETRI J., 32 (2010), pp. 784-794, 10.4218/etrij.10.1510.0132
View in ScopusGoogle Scholar
59. [59]
A.T. Lopes, E. de Aguiar, A.F. De Souza, T. Oliveira-Santos
Facial expression recognition with convolutional neural networks: coping with few data and the training sample order
Pattern Recognit. (2017), p. 61, 10.1016/j.patcog.2016.07.026
Google Scholar
60. [60]
G. Fanelli, A. Yao, P.-L. Noel, J. Gall, L. Van Gool
Hough Forest-Based Facial Expression Recognition from Video Sequences
Springer, Berlin, Heidelberg (2012), pp. 195-206, 10.1007/978-3-642-35749-7_15
View in ScopusGoogle Scholar
61. [61]
NN SVG. n.d.. [accessed 14 December 2019].
Google Scholar
Cited by (110)

* ### A discriminatively deep fusion approach with improved conditional GAN (im-cGAN) for facial expression recognition
2023, Pattern Recognition
Show abstract
Considering most deep learning-based methods heavily depend on huge labels, it is still a challenging issue for facial expression recognition to extract discriminative features of training samples with limited labels. Given above, we propose a discriminatively deep fusion (DDF) approach based on an improved conditional generative adversarial network (im-cGAN) to learn abstract representation of facial expressions. First, we employ facial images with action units (AUs) to train the im-cGAN to generate more labeled expression samples. Subsequently, we utilize global features learned by the global-based module and the local features learned by the region-based module to obtain the fused feature representation. Finally, we design the discriminative loss function (D-loss) that expands the inter-class variations while minimizing the intra-class distances to enhance the discrimination of fused features. Experimental results on JAFFE, CK+, Oulu-CASIA, and KDEF datasets demonstrate the proposed approach is superior to some state-of-the-art methods.

* ### Negative emotions detection on online mental-health related patients texts using the deep learning with MHA-BCNN model
2021, Expert Systems with Applications
Show abstract
Mining the emotions in the text related to mental health-care oriented is a challenging aspect, especially dealing with a long-text sequence of data. The extraction of emotions depends upon the various psychological depression factors like negative and ambiguity. Identifying these factors is the most perplexing task for every psychiatrist to treat their patients. Our study includes the deep learning (DL) models with global vector representations (GloVe) embeddings to capture the text sequence of data. We proposed a model multi-head attention with bidirectional long short-term memory and convolutional neural network (MHA-BCNN) is a pre-eminent mechanism that outperforms better than past research works for capturing the negative text-based emotions. In this paper, by using DL extracted the various negative mental-health emotions like addiction, anxiety, depression, insomnia, stress,

and obsessive cleaning disorder (OCD). By using the GloVe embeddings and handled the ambiguity factors like multiple emotion words in a certain sequence. As we proposed a vigorous appliance in our research to capture and hoard the long-term dependencies. We extracted the questions related to mental health issues were posted by the patients in an online mental healthcare-oriented platform. We efficaciously handled both negative and ambiguity factors at the document level. Our suggested exemplary MHA-BCNN surmounts various aspects from preceding research works and ensued preeminent performance. Experimental results show that our proposed framework MHA-BCNN outperformed than the erstwhile research works.

* ### Dynamic multi-channel metric network for joint pose-aware and identity-invariant facial expression recognition
2021, Information Sciences
Citation Excerpt :
Compared to MPCNN [11], which has the best performance among the other methods, our method improves FER accuracy by more than 1.3% and achieves the highest pose estimation accuracy. Since some state-of-the-art methods only use frontal images from the KDFE dataset for FER, Table 4 shows the frontal accuracy on the KDEF dataset using our method, CNN [37], GAN [38], AlexNet [39], and RCFN [40]. Compared to the state-of-the-art methods, our method also achieved an increase of up to 3.5%.
Show abstract
Facial expression recognition (FER) is challenging because the appearance of an expression varies significantly depending on head pose and inter-subject characteristics. With existing techniques, it is often difficult to learn both pose-aware and identity-invariant representations of facial expressions effectively due to the complex distribution of intra-class variation and similarity caused by these two factors. In this study, we propose a dynamic multi-channel metric learning network for pose-aware and identity-invariant FER, called DML-Net, which can reduce the effects of pose and identity for robust FER performance. Specifically, DML-Net uses three parallel multi-channel convolutional networks to learn fused global and local features from different facial regions. Then it uses joint embedded feature learning to explore identity-invariant and pose-aware expression representations from fused region-based features in an embedding space. DML-Net is end-to-end trainable by minimizing deep multiple metric losses, FER loss, and pose estimation loss with dynamically learned loss weights, thereby suppressing overfitting and significantly improving recognition. We evaluate DML-Net on three widely-used multi-view facial expression datasets, namely, KDEF, BU-3DFE, and Multi-PIE, as well as a wild dataset SFEW2.0. Extensive experiments demonstrate that our approach outperforms several other popular methods with accuracies of 88.2% on KDEF, 83.5% on BU-3DFE, 93.5% on Multi-PIE, and 54.36% on SFEW.

* ### GA-SVM-Based Facial Emotion Recognition Using Facial Geometric Features
2021, IEEE Sensors Journal

* ### Facial Expression Recognition Based on Deep Learning Convolution Neural Network: A Review
2021, Journal of Soft Computing and Data Mining

* ### COVIDetectioNet: COVID-19 diagnosis system based on X-ray images using features selected from pre-learned deep features ensemble
2021, Applied Intelligence
View all citing articles on Scopus
Zixiang Fei received his Bachelor degree in Liverpool John Moores University and Master Degree in University of York. He is currently a Ph.D. student in University of Strathclyde. His major research interests include

computer vision, machine learning, object recognition and deep learning.
Erfu Yang is a Lecturer in University of Strathclyde under Strathclyde Chancellor's Fellowship Scheme. In 2008, he received his Ph.D. degree in robotics in the interdisciplinary area of robotics and autonomous Systems from the University of Essex. His main research interests include robotics, autonomous systems, computer vision, image/signal processing, mechatronics, data analytics, etc.
David Day-Uei Li received the Ph.D. degree in electrical engineering from the National Taiwan University, Taipei, Taiwan, in 2001. He then joined the Industrial Technology Research Institute, Taiwan, working on CMOS communication chipsets. From 2007 to 2011 he worked at the University of Edinburgh before he took the lectureship in biomedical engineering at the University of Sussex. In 2014 he joined the University of Strathclyde, Glasgow, as a Senior Lecturer. His research interests include mixed signal circuits, CMOS sensors and systems, embedded systems, FLIM systems & analysis, and optical communications.
Stephen Butler received an MA and Ph.D. at the University of Glasgow. He is the director of the Strathclyde Oculomotor Lab. His research interests, employing eye tracking technology lie in the cognitive neuroscience of attention and the neuropsychology of eye movements. He has applied his research skills to theoretical work in areas including face processing, addiction, biases in attention and brain injury, and has a broad range of experience in applying his research skills in applied fields from shipping to interface design.
Winifred Ijomah is a Reader in University of Strathclyde. She gained a Ph.D. in Remanufacturing from the University of Plymouth in 2002. Her research focuses on sustainable design and manufacturing, and to date has focused on product end-of-life, particularly on remanufacturing.
Xia Li is a psychiatry specialist for older adults in Shanghai Mental Health Center, Shanghai Jiaotong University, school of medicine. She has been engaged in psychogeriatric clinical practice and research for 16 years. Her interest mainly includes Dementia, behavioral problems and Depression in the elderly.
Huiyu Zhou received a Ph.D. in Computer Vision from Heriot-Watt University. He is a Reader in University of Leicester. He currently heads the Applied Algorithm and AI (AAAI) Theme and is leading the Biomedical Image Processing Lab at University of Leicester. His research interests includes machine learning, computer vision and artificial intelligence.
View Abstract

Recommended articles
* ### Automatic recognition of schizophrenia from facial videos using 3D convolutional neural network
Asian Journal of Psychiatry, Volume 77, 2022, Article 103263
Jie Huang, ?, Shuping Tan
View PDF
* ### Computerized analysis of facial expressions in serious mental illness
Schizophrenia Research, Volume 241, 2022, pp. 44-51
Tovah Cowan, ?, Alex S. Cohen
View PDF
* ### Landmark guidance independent spatio-channel attention and complementary context information based facial expression recognition
Pattern Recognition Letters, Volume 145, 2021, pp. 58-66
Darshan Gera, S Balasubramanian
View PDF

* ### Emotion Recognition from Physiological Signal Analysis: A Review
Electronic Notes in Theoretical Computer Science, Volume 343, 2019, pp. 35-55
Maria Egger, ?, Sten Hanke
View PDF
* ### Human emotion recognition by optimally fusing facial expression and speech feature
Signal Processing: Image Communication, Volume 84, 2020, Article 115831
Xusheng Wang, ?, Congjun Cao
View PDF
* ### Raspberry Pi assisted facial expression recognition framework for smart security in law-enforcement services
Information Sciences, Volume 479, 2019, pp. 416-431
Muhammad Sajjad, ?, Sung Wook Baik
View PDF
Show 3 more articles
Article Metrics
Citations
* Citation Indexes: 107
Captures
* Readers: 132
Social Media
* Shares, Likes & Comments: 1
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

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