International Journal of All Research Education and Scientific Methods (IJARESM),
ISSN: 2455-6211, Volume 11, Issue 7, July-2023, Available online at: www.ijaresm.com

Facial Feature Extraction and Emotional Analysis Using ML

Kalash[1], Palak Chhajed[2], Shashin Bhaskar[3], Dayanithi Elumalai[4], Rahul P G[5], Tushar R Bharadwaj[6], Manasi S[7], Himavarsha Yerrapothu[8]

ABSTRACT

Facial expression recognition has been an active research area over the past few decades, and it is still challenging due to its high intra-class variation. Most of these works perform reasonably well on datasets of images captured in a controlled condition but fail to perform as good on more challenging datasets with more image variation and partial faces. In recent years, several works proposed a framework for facial expression recognition, using deep learning models. Despite the better performance of these works, there is still room great for improvement. In this work, we propose a deep learning approach that is based on an attentional convolutional network, which can focus on important parts of the face and achieves significant improvement over various datasets, including FER-2013, CK+, FERG, and JAFFE. We also use a visualization technique that is able to find important face regions for detecting different emotions, based on the user's output. Through the experimental results, we show that different emotions seem to be reactive to different parts of the face.

INTRODUCTION

Emotional expressions are a very important part of human interactions. As technology has found everywhere in our society and is taking on coaching roles in education, emotion recognition from facial expressions has become an important part of human- computer interaction. However, human facial expressions change so insignificantly that automatic facial expression recognition has always been a challenging task. In this work, we propose a deep learning approach based on an attentional convolutional network, which is able to focus on salient parts of the face and achieves significant improvement over previous models on multiple datasets, including FER-2013, CK+, FERG, and JAFFE. We also use a visualization technique that is able to find salient face regions for detecting various emotions, based on the user's output. Through experimental results, we show that different emotions seem to be reactive to different parts of the face.

Convolutional Neural Networks

Artificial Intelligence has been facing a monumental growth in bridging the gap between the capabilities of humans and machines. Researchers work on numerous aspects of the field to make amazing things happen. One of many such areas is the domain of Computer Vision.

The main reason for this field is to enable machines to view the world as humans do, perceive it in a similar manner and even use the knowledge for various tasks such as Image & Video recognition, Image Analysis & Classification, Media Recreation, Recommendation Systems, Natural Language Processing, etc. The advancements in Computer Vision with Deep Learning has been constructed and perfected with time, basically over one particular algorithm

Convolutional Neural Network

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm that can take in an input image, and give importance (learnable weights and biases) to various aspects in the image and be able to differentiate one from the other.

The preprocessing required in a ConvNet is much lower as compared to other classification algorithms While in basic methods filters are hand-engineered, with enough training, Conv Nets have the ability to learn these characteristics.

The structure of a ConvNet is parallel to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a particular region of the visual field known as the Receptive Field. A collection of such fields clashes to cover the entire visual area.

CNN's, like other neural networks, are made up of neurons with known weights and biases. Each neuron receives several inputs, takes a weighted sum over them, pass it through an activation function and gives an output. The whole network has a loss function and all the tips that are developed for ANN will still apply to CNNs.

International Journal of All Research Education and Scientific Methods (IJARESM),
ISSN: 2455-6211, Volume 11, Issue 7, July-2023, Available online at: www.ijaresm.com

FACE DETECTION USING ARTIFICIALNEURAL NETWORK APPROACH

Abstract
A face detection system using an artificial neural network is presented. In this, The system used the integral image for image representation which allows fast computation of the features used. The system also implements the AdaBoost learning algorithm to select a small number of critical visual features from a very large set of potential features. Besides that, it also used to combine the classifiers algorithm which allows background parts of the image to be quickly discarded while spending more computation on face regions. Further, a set of experiments in the domain of face detection is done. The system yields a good face detection performance.

System Overview
The architecture of the proposed face detection system given. There are four (4) major modules in the proposed face detection system, which are as follows .data acquisition, image pre-processing, feature extraction, and classification.

EXPERIMENTAL RESULTS

The purpose of the experiment is to compare the performance of the proposed face detection using the two classifiers structures, the traditional and a modified structure of the neural network, and the method of comparing the threshold. The traditional structure involves too many hidden nodes, which increase the computation time for every hidden node. The modified structure will decrease the computation time and increase perception accuracy.

CONCLUSION

In this research, we have proposed a face detection system using the traditional and modified ANN approaches. We use an integral image to compute the features to decrease the computation time. The resulting classifier is computationally efficient since only a small number of features need to be evaluated during run time. We also used the combining of classifiers, which will reduce the computation time, and increase the detection accuracy. The cascade presented is in a similar structure, and has the advantage of making simple trade-offs between processing time and detection performance. The experimental results show that the improved ANN, using the voting approach increases the detection accuracy to 96% whereas only 78% using the traditional ANN approach.

SYSTEM ANALYSIS

Existing System
Effective expression analysis hugely depends upon the accurate representation of facial features. **Facial Action Coding System (FACS)** [3] represents the face by measuring all visually observable facial movements in terms of **Action Units (AUs)** and associates them with the facial expressions. Accurate detection of AUs depends upon proper identification and tracking of different facial muscles irrespective of the pose, face shape, illumination, and image resolution. According to Whitehill et al. [4], the detection of all facial fiducial points is even more challenging than expression recognition itself. Therefore, most of the existing algorithms are based on geometric and appearance-based features.

The models based on geometric features track the shape and size of the face and facial components such as eyes, lip corners, eyebrows etc., and categorize the expressions based on the relative position of these facial components. Moreover, the distance between facial landmarks varies from person to person, thereby making the person independent expression recognition system less reliable.

Facial expressions involve a change in local texture. The appearance-based methods generate a high dimensional vector which is further represented in lower-dimensional subspace by applying dimensionality reduction techniques, such as **principal component analysis (PCA),** linear discriminant analysis (LDA) etc. Finally, the classification is performed in the learned subspace. Although the time and space costs are higher in appearance- based methods, the preservation of discriminative information makes them very popular.

Drawbacks
Extraction of facial features by dividing the face region into several blocks achieves better accuracy as reported by many researchers. However, this approach fails with improper face alignment and occlusions. Some earlier works on the extraction of features from specific face regions mainly determine the facial regions which contribute more toward discrimination of expressions based on the training data.

In these approaches, the positions and sizes of the facial patches vary according to the training data. Therefore, it is difficult to conceive a generic system using these approaches.

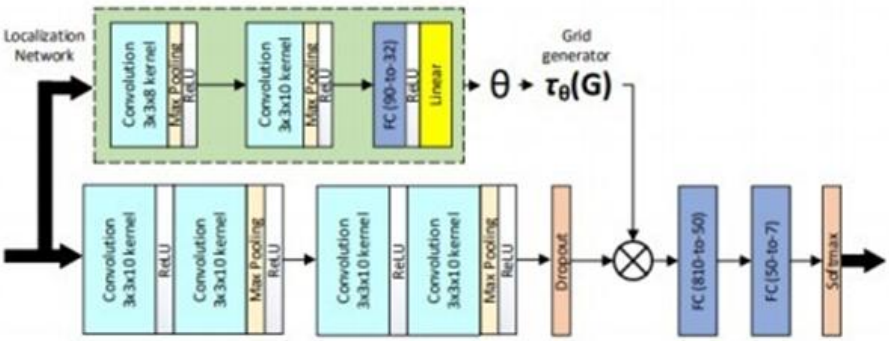
Proposed System

We propose a deep learning-based framework for facial expression recognition, which takes the above observation into account, and uses an attention mechanism to focus on the salient part of the face. We show that by using the attentional convolutional network, even a network with few layers (less than 10 layers) is able to achieve a very high accuracy rate. More specifically**,** this paper presents the following contributions

We propose an approach based on an **attentional convolutional network**, which can focus on feature-rich parts of the face, and yet, outperform remarkable recent works inaccuracy.

Implementation

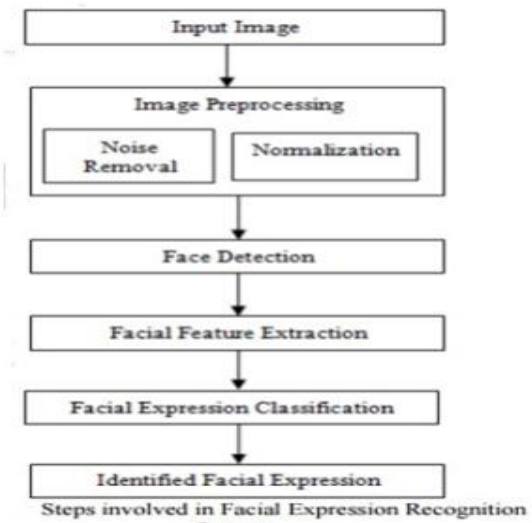
We propose an end-to-end deep learning framework, based on an attentional convolutional network, to classify the underlying emotion in the face images. Oftentimes, improving a deep neural network relies on adding more layers/neurons, facilitating gradient flow in the network. Given a face image, it is clear that not all parts of the face are important in detecting a specific emotion, and in many cases, we only need to attend to the specific regions to get a sense of the underlying emotion. Based on this observation, we add an attention mechanism, through spatial transformer the network into our framework to focus on important face regions.



Description of the corpus the evaluation of the proposed method for facial expression recognition was performed on two databases:

The JAFFE database (The Japanese Female Facial Expression) [LBA99a]: is widely used in the facial expressions research community. It is composed of 213 images of 10 Japanese women displaying seven facial expressions: the six basic expressions and the neutral one. Each subject has two to four examples for each facial expression.

The KANADE database [KCT00]: is composed of 486 video sequences of people displaying 23 facial expressions within the six basic facial expressions. Each sequence begins with a neutral expression and finishes with the maximum intensity of the expression. For a fair comparison between KANADE and JAFFE databases, we selected from the KANADE database the first image (neutral expression) and the last three images (with the maximum intensity of the expression) of 10 people chosen randomly. Moreover, we selected the six basic facial expressions and the neutral one.



Steps involved in Facial Expression Recognition System

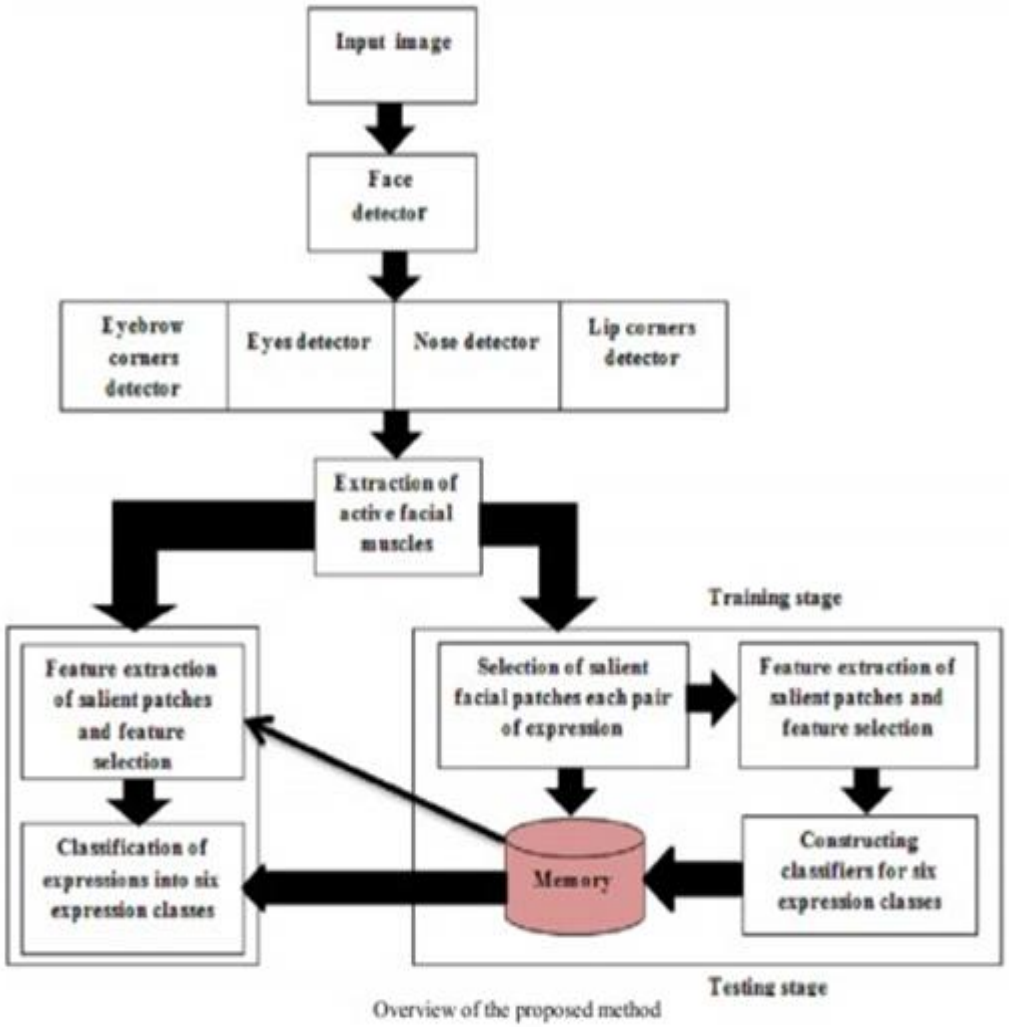
Facial Expression Recognition basically performed in three major steps:

- Face detection
- Feature Extraction
- Facial Expression Classification

The primary need for a Face Expression Recognition system **is Face Detection** which is used to detect the face.

The next phase is **feature extraction** which is used to select and extract relevant features such as eyes, eyebrow, nose and mouth from the face. It is very essential that only those features should be extracted from an image that has a high contribution to expression identification.

The final step is **facial expression classification** that classifies the facial expressions based on the extraction of relevant features.



Overview of the proposed method

MODULES

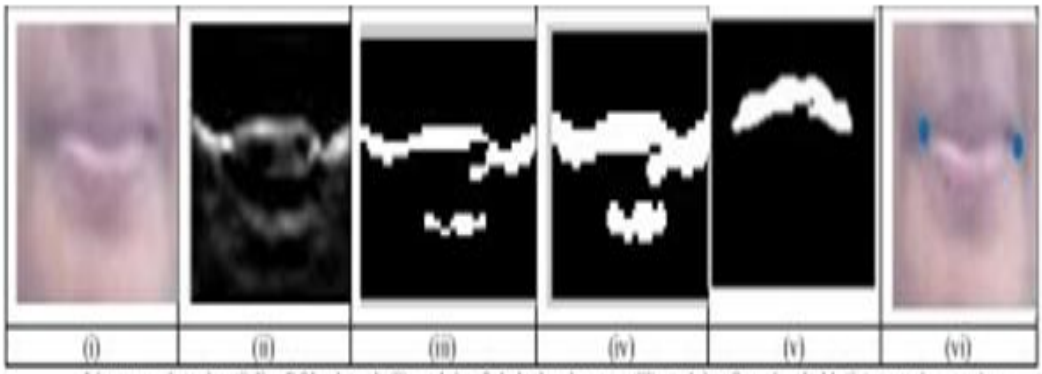
Pre-Processing

Image pre-processing often takes in the form of signal conditioning with the segmentation, location, or tracking of the face or facial parts. A low pass filtering condition is performing using a 3x3 Gaussian mask and it removes noise in the facial images followed by face detection for face localization. Viola-Jones technique of Haar- like features is used for face detection. It has lower computational complexity and sufficiently accurate detection of near -upright and near-frontal face images. Using integral image, it can detect face scale and location in real-time. The localized face images are extracted. Then localized face image scaled to bring it to a common resolution and this made the algorithm shift-invariant.

Eye and Nose Localization

The coarse regions of interests are eyes and nose selected using the geometric position of the face. Coarse region of interests can also use to reduce computational complexity and false detection. Haar classifiers used to both eyes are detected separately and then haar classifier trained for each eye. The classifiers are returns to the vertices of the rectangular area of detected eyes. The centres of both eyes are computed as the mean of these coordinates. The position of the eyes does not change with facial expressions. Similarly, Haar cascades are used to detected nose position. In this case, eyes or nose was not detected using Haar classifier s. In our experiment, for more than 99.6 per cent of cases, these parts were detected correctly.

Lip Corner Detection



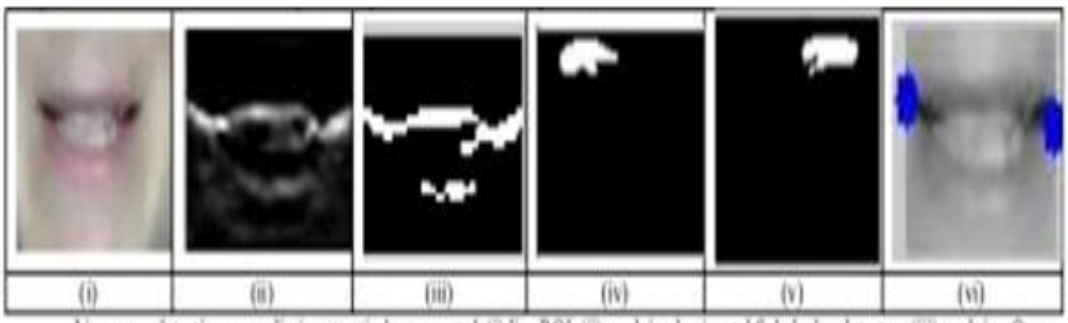
Lip corner detection (i) lips ROI selected, (ii) applying Sobel edge detector, (iii) applying Otsu threshold, (iv) removing spurious components, (v) applying morphological operations, (vi) finally lip corners detection

Algorithm for Lip Corner Detection

1. Step 1. Select coarse lips ROI using face width and nose position.
2. Step 2. Apply 3x3 Gaussian mask to the lips ROI.
3. Step 3. Apply horizontal Sobel operator for edge detection.
4. Step 4. Apply Otsu-thresholding.
5. Step 5. Apply morphological dilation operation.

Step 6 Find the connected components in images.

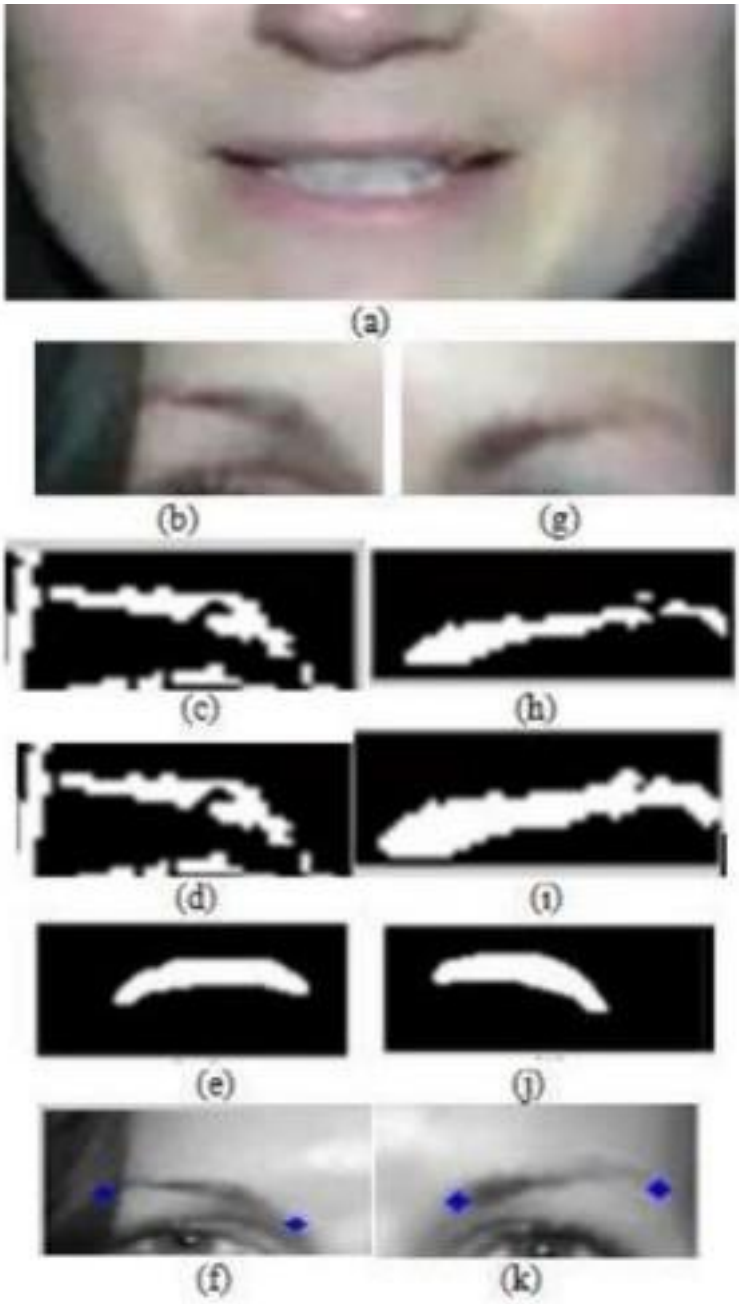
1) Step 7. Remove the spurious connected components using the threshold techniqueto the number of pixels.
2) Step 8. Scan the image from the top and select the first connected component as upper lip position.
3) Step 9. Locate the left and rightmost positions of the connected component as lip corners.



Lip corner detection upper lip is not entirely connected, (i) lips ROI, (ii) applying horizontal Sobel edge detector, (iii) applying Otsu threshold, (iv) and (v) selection of two connected components, (vi) finally lip corner detection.

Eyebrow Corner Detection

The coarse region interest is used to both eyebrows are selected. The eyebrows corner detection following the same procedure as that of upper lip detection. An adaptive threshold operation is applying before the horizontal Sobel operator it improved the accuracy of eyebrow corner detection. The different stage of the process is shown figure below.



Extraction of Active Facial Patches

Facial expressions are usually formed by local facial appearance variations. However, it is more difficult to automatically localize these local active areas on a facial image. Firstly the facial image is divided into N local patches, and then **local binary pattern (LBP)** features are used to represent the local appearance of the patch. During an expression, the local patches are extracted from the face image depending upon the position of active facial muscles. In our experiment are used active facial patches as shown in Fig. 6. Patches are assigned by number. The patches do not have a very fixed position on the face image. The position of patches varying with a different expression. The location depends on the positions of facial landmarks. The sizes of all facial patches are equal and it was approximately one-eighth of the width of the face. In Fig 1P,4P,18P and 17P are directly extracted from the positions of lip corners and inner eyebrows respectively. 15P was at the centre of both the eyes; and 16P was the patch occur just above 15P. 19P and 6P are located in the midway of the eye and nose. 3P was located just below 1P.9Pwas located at the centre of the position of 3P and 8P. 14P and 11P are located just below the eyes. 2P,10P, and 7Pare clubbed together and located at one side of the nose position. It is similar to 5P 12P, and 13P are located.

Feature Extraction

The local binary pattern is used robust illumination of feature extraction. The original LBP operator labels the pixels of an image with decimal numbers called Local Binary Patterns which encode the local texture around each pixel. It proceeds thus, as illustrated in

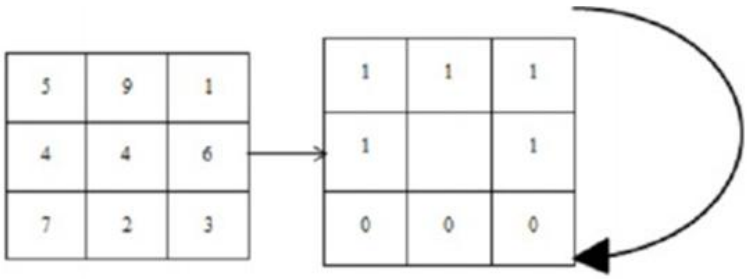


Fig. 7: An example of the basic LBP operator.

PRINCIPAL COMPONENT ANALYSIS FOR FACE RECOGNITION

Algorithm Description

The face is the primary focus of attention in our society. The ability of human beings to remember and recognize faces is quite a difficult task. Automation of this process finds practical application in various tasks such as criminal identification, security systems and human-computer interactions. Various attempts to implement this process have been made over the years. Eigenface recognition is one such technique that relies on the method of principal components. This method is based on an information-theoretical approach, which treats faces as intrinsically 2-dimensional entities spanning the feature space. The method works well under carefully controlled experimental conditions but is error-prone when used in a practical situation. This method functions by projecting a face onto a multi-dimensional feature space that spans the gamut of human faces. Any face in the feature space is then characterized by a weight vector obtained by projecting it onto the set of basis images. When a new face is presented to the system, its weight vector is calculated and **a CNN based classifier is used for classification.**

RESULT

Method	Success rate (%)	Error rate(%)
Traditional CNN	78	22
Modified CNN	98	4

CONCLUSION

The face is the primary focus of attention in society. The ability of human beings to remember and recognize faces is quite robust. Automation of this process finds practical application in various tasks such as criminal identification, security systems and human- computer interactions. Various attempts to implement this process have been made over the years. Eigenface recognition is one such technique that relies on the method of principal components. This method is based on an information-theoretical approach, which treats faces as intrinsically 2-dimensional entities spanning the feature space. The method works well under carefully controlled experimental conditions but is error-prone when used in a practical situation. This method functions by projecting a face onto a multi-dimensional feature space that spans the gamut of human faces. A set of basis images is extracted from the database presented to the system by Eigenvalue-Eigenvector decomposition. Any face in the feature space is then characterized by a weight vector obtained by projecting it onto the set of basis images. When a new face is presented to the system, its weight vector is calculated and a CNN based classifier is used for classification.

REFERENCES

[1]. Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: a survey," 2019, https://arxiv.org/abs/1905.05055.
[2]. View at: Google Scholar
[3]. S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, pp. 91–99, 2015.
[4]. View at: Google Scholar
[5]. S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 8759–8768, Salt Lake City, UT, United States, 2018.

[6]. View at: Google Scholar
[7]. K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.
[8]. View at: Publisher Site | Google Scholar
[9]. K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn.," in *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, Venice, Italy, 2017.
[10]. View at: Google Scholar
[11]. W. Liu, D. Anguelov, D. Erhan et al., "Ssd: single shot multibox detector," in *Computer Vision – ECCV 2016*, Lecture Notes in Computer Science, Springer, 2016.
[12]. View at: Publisher Site | Google Scholar
[13]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, Las Vegas, NV, United States, 2016.
[14]. View at: Google Scholar
[15]. J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263–7271, Honolulu, HI, United States, 2017.
[16]. View at: Google Scholar
[17]. J. Redmon and A. Farhadi, "Yolov3: an incremental improvement," 2018, https://arxiv.org/abs/1804.02767.
[18]. View at: Google Scholar
[19]. A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, "Yolov4: optimal speed and accuracy of object detection," 2020, https://arxiv.org/abs/2004.10934.
[20]. View at: Google Scholar