Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/37/20230510

Research on face expression image recognition based on Convolutional Neural Network

Junhan Jiang

School of Computing and Cyber Security (Oxford Brookes College), Chengdu University of Technology, 610059, China

jfeeng001@gmail.com

Abstract. The recognition of facial expressions is a significant field of study within computer vision, which has a multitude of practical applications in areas such as human-computer interaction, affective computing, and social robotics. The present study offers a thorough examination of the recognition of facial expression images through the utilisation of Convolutional Neural Networks (CNNs). The aim is to categorise facial expressions in challenging lighting scenarios through the utilisation of Convolutional Neural Networks (CNNs), which possess the capability to acquire intricate features directly from unprocessed image data. The present study introduces an innovative method for facial recognition through the proposal of a revised architecture for Convolutional Neural Networks (CNNs). The primary improvement pertains to the incorporation of two normalisation procedures into particular layers of the network. The utilisation of batch normalisation, a commonly employed normalisation technique, serves to expedite the performance of the network.

Keywords: Facial expression recognition, Low-light conditions, Experimental technique, Extended YaleB face database, Haar-like feature extraction algorithm.

1. Introduction

Recognising a person's face in a visual system is known as face recognition. Security systems, access-control, video monitoring, commercial areas, and even social networks like Facebook all make use of this important human-computer interface technology.

With the advent of cutting-edge AI systems, facial recognition has once again become a hot topic because it is both non-intrusive and the most reliable means of human identification. In an uncontrolled setting, it is simple to test face recognition without the awareness of the subject individual. However, it is still difficult to recognise facial expressions in low-light settings [1,2].

Convolutional Neural Networks (CNNs) are investigated in this research for their potential use in accurate facial expression image detection. Facial expression recognition was initially implemented in an animation video in 1978. Subsequently, Suwa pioneered the application of facial expression recognition in computer image processing. This technological foundation enables automatic face expression recognition using an image sequence. In order to implement the facial expression identification function, K. Mase and A. Pentland recommended using the streamer method to identify the primary direction of muscle action and extract the streamer value as the facial expression feature [3]. Following that, a new era in automatic face expression recognition began.

DOI: 10.54254/2755-2721/37/20230510

The Japanese university Waseda created the Kobian robot in 2009, which can communicate with humans using seven different emotions [4]. The Massachusetts Institute of Technology's MIT Media Laboratory created the robot Kismet in 2000, which can identify facial emotions and then mimic them [5]. It is currently widely employed in industries like as intelligent monitoring, online learning, healthcare, safe driving, and other aspects of daily life.

2. Research Objectives, Main Contents, and Key Issues to be Solved

2.1. Research Objectives

The present study aims to outline the research objectives pertaining to the recognition of facial expression images through the utilization of Convolutional Neural Networks (CNNs)(Xiang and Zhu). These objectives can be succinctly summarized as follows:

The fusion of feature maps is a crucial objective in improving image recognition accuracy, as it enables effective feature map fusion. Convolutional neural networks (CNNs) demonstrate exceptional proficiency in acquiring hierarchical representations and capturing intricate features.

The research endeavors to make a contribution to the domain of facial expression recognition and furnish significant perspectives on the utilization of Convolutional Neural Networks (CNNs) for the purpose of image analysis tasks, by accomplishing the aforementioned objectives. The findings of this research hold promise for augmenting several fields, such as human-computer interaction, affective computing, and social robotics, where precise and instantaneous identification of facial expressions is of paramount significance.

The fusion of feature maps is a crucial objective in improving image recognition accuracy, as it enables effective feature map fusion. Convolutional neural networks (CNNs) demonstrate exceptional proficiency in acquiring hierarchical representations and capturing intricate features. The study endeavors to enhance the model's discriminative power by examining various methods for combining and merging feature maps across multiple scales. The fusion of feature maps in an effective manner enables the network to capture both local and global facial characteristics, thereby resulting in improved accuracy in recognizing facial expressions. The local features include the eyes, nose, and mouth.

2.2. Main concepts

This study focuses on the investigation of facial expression image recognition through the utilization of Convolutional Neural Networks (CNNs). The primary objectives of this research are to achieve precise classification of facial expressions and to develop an efficient feature map fusion technique. The development of a reliable system capable of accurately categorizing facial expressions while accounting for challenging lighting conditions is crucial for achieving precise classification [6].

In conclusion, this study investigates the feasibility of using Convolutional Neural Networks (CNNs) for emotion recognition in digital photographs. Accurate face emotion categorization is the goal of this study, which attempts to develop efficient methods for merging and fusing feature maps and overcome the difficulties introduced by varying lighting. The findings of this study have important implications for a wide range of subjects, including but not limited to human-computer interaction, affective computing, and social robotics [6].

2.3. Key Issues to be solved

The present study aims to investigate the key concerns pertaining to the recognition of facial expression images utilizing Convolutional Neural Networks (CNNs). These concerns can be succinctly summarized as follows:

Thus, it is essential to create methods that can account for individual differences in facial shape and appearance if we are to achieve accurate recognition of facial expressions. Capturing both local and global facial traits is essential for improving picture identification accuracy, and feature map fusion is a key component in doing so.

DOI: 10.54254/2755-2721/37/20230510

3. The Main Structure of Convolution Neural Network

The fundamental architecture of a convolutional neural network consists of several key components, including a convolutional layer responsible for extracting salient image features, a nonlinear activation function, a pooling layer for data compression, and a fully connected layer that outputs classification information.

3.1. Convolution Layer

The Convolution Layer is a fundamental component of Convolutional Neural Networks (CNNs) used in deep learning. It performs a mathematical operation called convolution on the input data, which is typically an image, to extract features and create a feature map.[7] The convolution operation involves sliding a small matrix, called a kernel or filter, over the input data and computing the dot product between the kernel and the corresponding input values.

During training, the filters and bias terms that make up the convolutional layer are learned by backpropagation and gradient descent techniques. Using a loss function that evaluates the gap between the expected and actual output, the network learns to fine-tune these settings.

3.2. The Activation Function

The relationship between inputs and outcomes is often nonlinear in real-world challenges. In contrast, neural networks are predicated on what may be thought of as a linear operation: applying weights to data. Activation functions are added to neural networks to help them model nonlinear interactions more accurately [8].

3.3. Pool Layer

The pool layer is a component commonly used in convolutional neural networks for image processing tasks. Its primary function is to downsample the input data by reducing the spatial dimensions of the feature maps, while retaining the most salient information. This process helps to reduce the computational complexity of subsequent layers and improve the network's ability to generalize to new data.

3.4. The full connection layer

The Full Connection Layer is a type of neural network layer that connects all neurons from the previous layer to all neurons in the current layer. This layer is commonly used in deep learning models for tasks such as image classification and natural language processing.

In conclusion, a CNN's fully connected layer synthesises the characteristics gathered in earlier layers to provide a classification score vector. [7] Either a fully connected layer or a convolutional layer can be used to implement it. In a CNN architecture, the fully linked layer is critical for synthesising features and generating the final classification results.

4. Research Methods and Ideas

4.1. Pixel histogram statistics

A method used to alleviate the effects of bad illumination on facial recognition is called pixel histogram statistics. This technique tries to record and examine the estimated distribution of various grayscale values inside a picture as well as the grayscale range of pixels within the image. You can learn a lot about the image's brightness and contrast by looking at the histogram of pixel values.

4.2. Data Sources

The Extended YaleB face database is a significant asset employed in the field of facial expression image recognition investigation. The dataset consists of numerous test objects that were captured under diverse lighting conditions, rendering it especially appropriate for tackling the difficulties presented by inadequate or non-uniform illumination. The database presented herein offers a varied assortment of

DOI: 10.54254/2755-2721/37/20230510

facial images, thereby enabling scholars to scrutinise and devise algorithms that exhibit resilience to fluctuations in lighting conditions [9].

4.3. Data Processing

The preprocessing of facial expression images entails the standardisation of their size and format, utilising The Extended YaleB face database as a point of reference for the data. Furthermore, the process of centralization and normalisation is employed to ensure data alignment and standardisation on a uniform scale. The aforementioned preprocessing procedures are implemented to ensure that the facial expression dataset is suitably prepared for subsequent analysis and training using convolutional neural networks.

$$x = (x - \mu)/\sigma$$

Here μ is the mean value and σ is the standard deviation.

4.4. Automated Facial coding

Face and facial landmark identification, face texture feature extraction, facial action categorization, and emotion expression modelling are the four primary parts of our automated facial coding system. Together, these elements analyse and interpret facial expressions to reveal important information about human emotions.

5. Technical Route

5.1. Sample Training

The process of selecting and preparing training data is a vital component in the recognition of facial expression images, known as sample training. The objective of sample training is to generate a comprehensive and inclusive collection of facial expression images that sufficiently encompasses the spectrum of expressions under consideration.

The present procedure involves the analysis of a dataset comprising images of facial expressions, from which a subset of images is selected to function as the training samples. It is recommended that the samples be evenly distributed among various facial expressions, encompassing a diverse range of poses, lighting scenarios, and individuals.

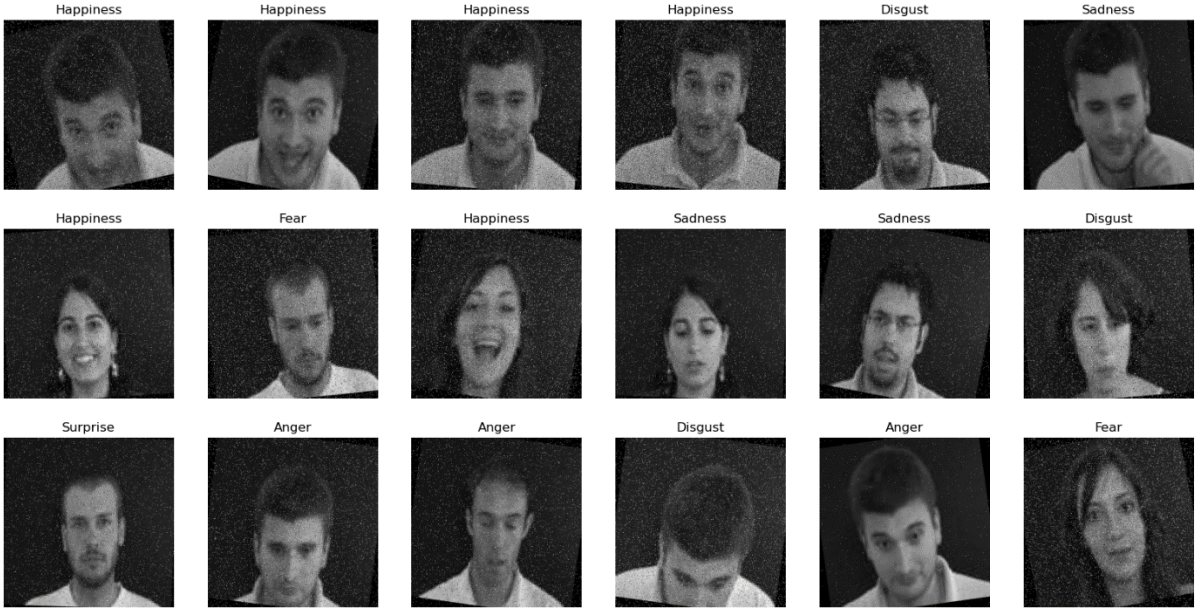

Figure 1. Sample data set

DOI: 10.54254/2755-2721/37/20230510

5.2. Training sample classification

The classification of training samples involves the allocation of facial expression images to their corresponding categories or classes, predicated on their distinctive characteristics. [7,10] This stage entails 5 eneraliz classification algorithms to acquire knowledge of the distinctive patterns and interconnections present in the training data [11].

5.3. Training Sample Normalization

A essential preprocessing step in machine learning, 5eneralize5ti of training samples seeks to maintain uniformity and comparability among various training samples. It's a method for improving the efficacy of machine learning algorithms by eliminating biases in the data through 5eneralize5tion of scale and distribution.

5.4. Classifier training

Training a classifier is essential to creating a model that can properly categorise and 5eneraliz facial emotions. In this method, the classifier is taught to 5eneraliz various facial expressions by analysing preprocessed training samples. The ultimate objective is to build a model that is both accurate and robust, with the ability to 5eneralize well to new or unexplored data.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 126, 126, 16)      448

 max_pooling2d (MaxPooling2D  (None, 63, 63, 16)       0
 )

 conv2d_1 (Conv2D)           (None, 61, 61, 32)        4640

 max_pooling2d_1 (MaxPooling  (None, 30, 30, 32)       0
 2D)

 flatten (Flatten)           (None, 28800)             0

 dense (Dense)               (None, 128)               3686528

 dense_1 (Dense)             (None, 6)                 774
```

Figure 2. Classifier Model

In order to improve accuracy and reduce classification errors, the training procedure comprises adjusting the classifier's parameters. Using an optimisation approach like gradient descent, the model's weights and biases are often tweaked iteratively to achieve this optimisation. The goal is to identify the parameters that provide the greatest fit to the training data and also allow for good generalisation to new data.

5.5. LBP feature value extraction

The extraction of feature values from preprocessed facial expression images involves the utilization of the Local Binary Patterns (LBP) feature extraction technique. LBP is a method that captures local texture

DOI: 10.54254/2755-2721/37/20230510

information by comparing the grayscale values of neighboring pixels. This technique has proven to be effective in encoding unique facial texture patterns into binary representations.

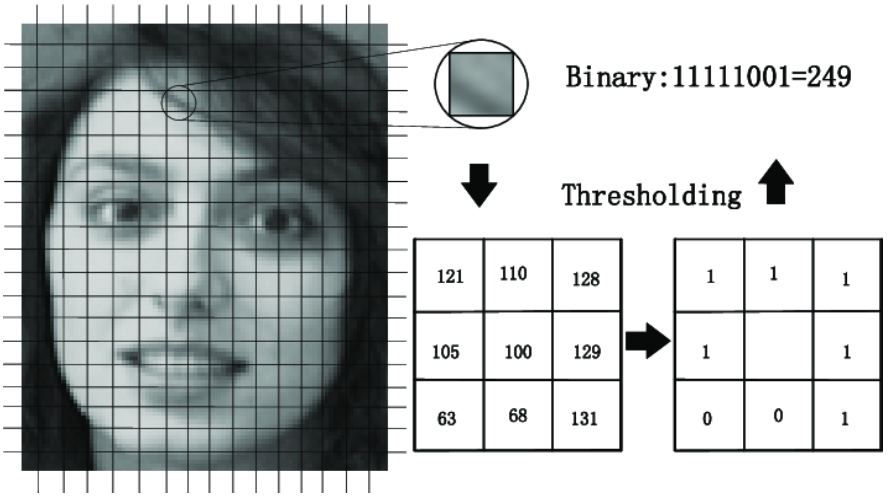

Figure 3. Face-feature-extraction-based-on-LBP

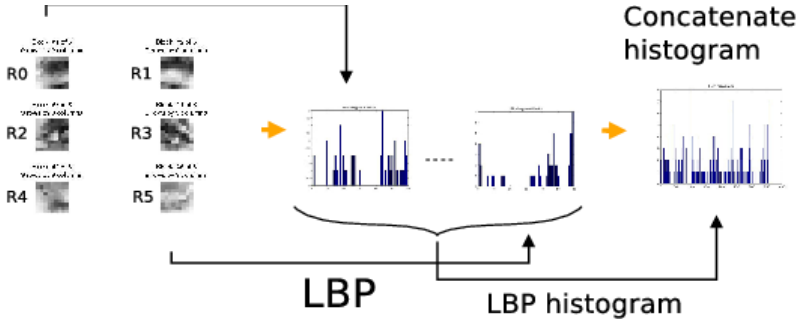

Figure 4. Feature-extraction-using-local-binary-pattern

5.6. Weak classifier iteration

Weak classifier iteration refers to the process of training and integrating many weak classifiers to create a single robust classifier. Weak classifiers are models with a simple structure and results that are only slightly better than chance. By combining a large number of relatively ineffective classifiers, model robustness and accuracy in general may be improved by repetitive use of weak classifiers.

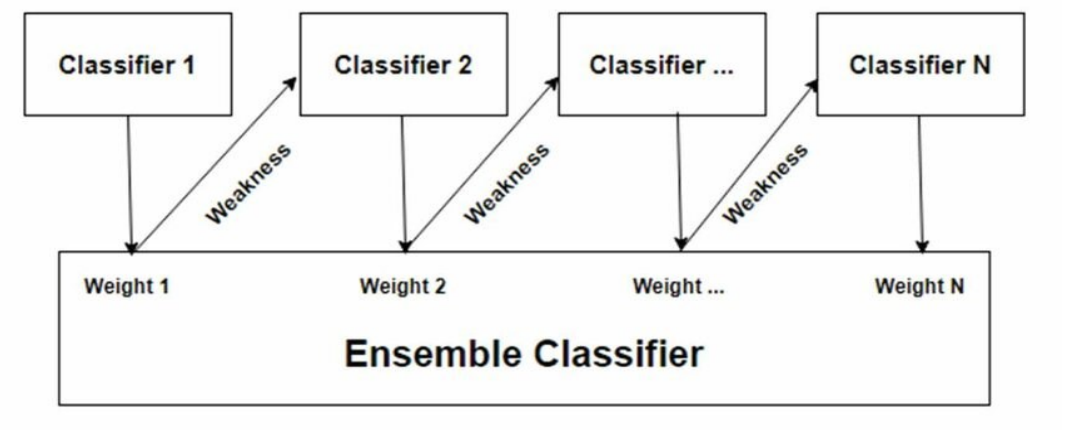

Figure 5. Weak Classifier

DOI: 10.54254/2755-2721/37/20230510

5.7. Strong classifier cascading

The final step is to use robust classifier cascade for accurate facial expression image classification. The term "cascading" is used to describe the sequential arrangement of many classifiers, with each classifier in the cascade focusing on a different set of characteristics. This method has been shown to improve the system's speed and accuracy by effectively removing negative samples from the classification process from the outset.

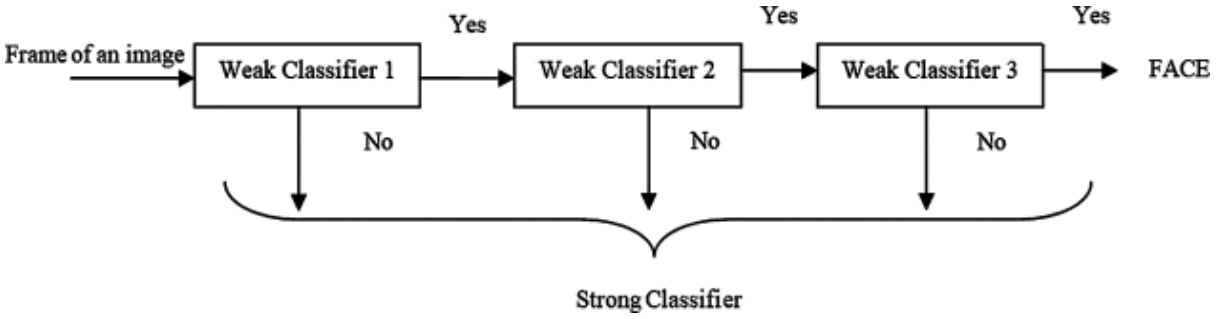

Figure 6. Strong Classifier

6. Experimental Results

To begin, we create a new instance of the Sequential class, which will serve as the foundation for our layer-by-layer CNN architecture.

The initial layer of the model is a convolutional layer, which is added next. We then insert a pooling layer after the convolutional one. Adding a second convolutional layer and then a max pooling layer allows us to capture more complicated characteristics. With these layers in place, the network may acquire more abstract mental models of the pictures it is fed.

After the network is flattened, we add a fully linked 128-unit layer. This layer is responsible for categorization, which it accomplishes by learning to link the previously retrieved characteristics with the various face expressions. The model is made non-linear by the employment of the ReLU activation function.

Last but not least, we add an output layer consisting of 6 units, which stands for the 6 categories of facial expressions we're trying to identify.

6.1. Accuracy

With a loss of 0.3473 and an accuracy of 0.8748, the Convolutional Neural Network (CNN) model performed admirably in facial emotion picture identification. This CNN model has shown promise for practical use by successfully learning and differentiating between facial expressions.

The overall performance of this CNN model in face expression picture identification is demonstrated, with encouraging results in terms of both accuracy and loss. It exemplifies the power of deep learning models to record and comprehend nuanced patterns in facial expressions, making it a useful resource for fields as diverse as emotion detection, HCI, and psychological study.

6.2. Comaparison between training loss and validation loss

The supplied CNN model achieves a low training loss in the range of 0.25 to 0.30, showing that it learns and corrects its mistakes well. The model's ability to properly predict the labels and capture the underlying patterns and features of the training data is shown by a low training loss.

Training losses in the 0.25–0.30 range indicate that the model has effectively learned the training data and can produce reliable predictions. It is important to test the model on the validation set to make sure it is not overfitting, though.

For the model to improve its generalisation capabilities and performance on unseen data, it needs to achieve a training loss in the range of 0.25 to 0.30 and a validation loss in the range of 0.35 to 0.60.

DOI: 10.54254/2755-2721/37/20230510

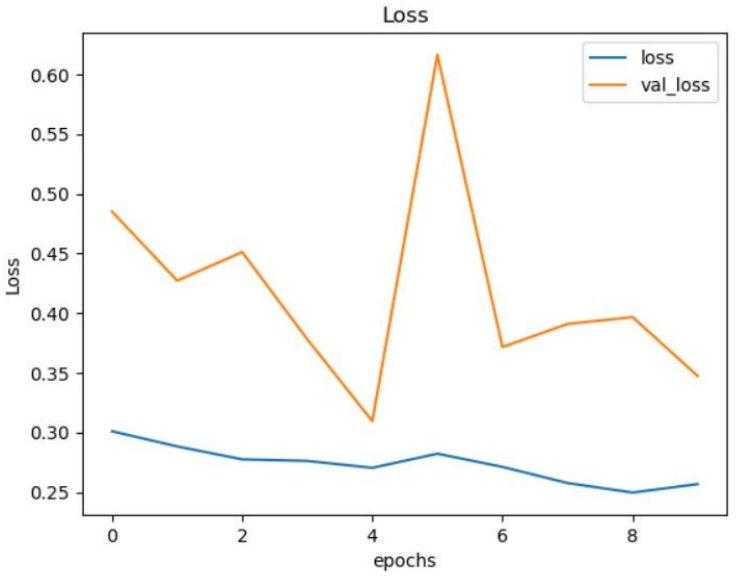

Figure 7. Comaparison between training loss and validation loss

6.3. Comparison between training accuracy and validation accuracy

The given CNN model shows a training loss between 0.89% and 0.95%, showing that the model makes some errors while being trained. If the model has a hard time fitting the training data, it is likely not accurately capturing the underlying patterns and characteristics.

Finally, the reported epoch range, together with a training loss between 0.89 and 0.95 and a validation loss between 0.80 and 0.89, suggest that the model may be overfitting.

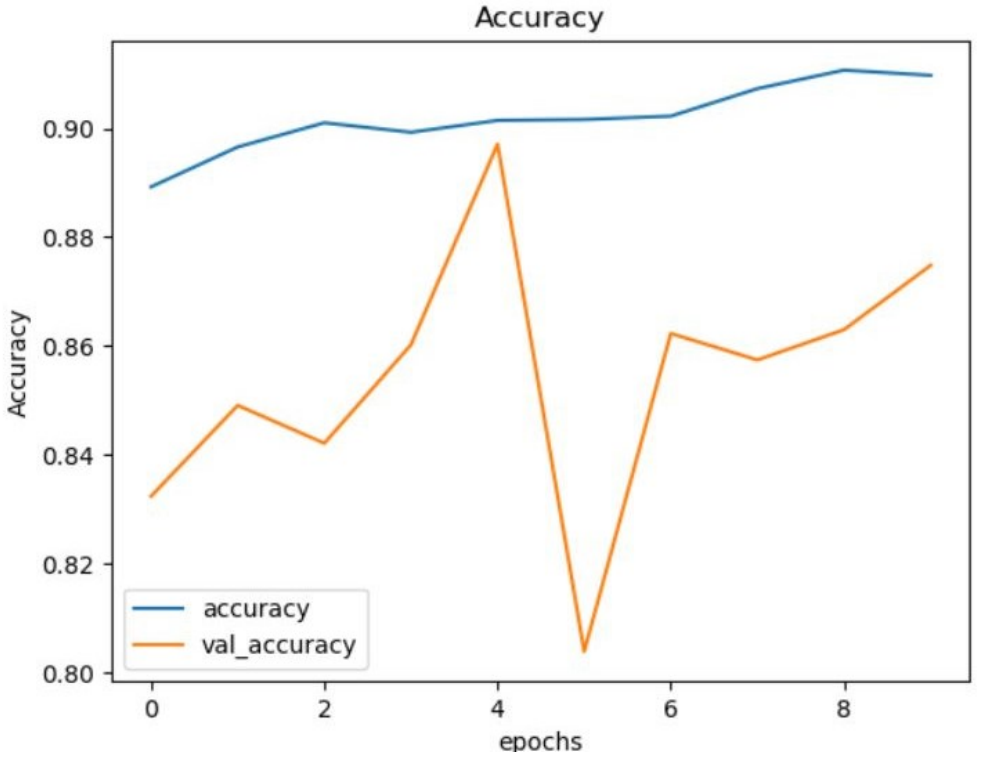

Figure 8. Comparison between training accuracy and validation accuracy

DOI: 10.54254/2755-2721/37/20230510

Table 1. Comparison between training accuracy and validation accuracy

	Training	Validation
Accuracy	.89	.84
Loss	.25	.45

7. Feasibility Analysis

Both picture recognition and facial expression recognition have been widely explored, and the chosen research methodologies and technical routes have been shown to be effective. Facial expression identification is difficult in low-light situations, but an experimental technique using the Extended YaleB face database and a Haar-like feature extraction algorithm shows that it is possible to overcome these obstacles [6,8].

A common technique for recognising facial features is the Haar-like feature extraction algorithm. It operates by identifying basic rectangular elements with recognisable patterns, including the existence of an eye, a nose, or a mouth. The experimental outcomes gained from this strategy show that it is effective in overcoming the difficulties brought on by low light levels.

8. Conclusion

Finally, this research developed a thorough method for face emotion recognition that overcomes the difficulties caused by inadequate lighting. The proposed approach has the potential to transform many fields where facial analysis is important. Accurate and real-time facial expression detection, for instance, can promote more intuitive and natural interactions between people and machines in the field of human-computer interaction. This can have substantial effects in fields where reliable recognition and analysis of facial expressions are essential, such as surveillance, access control, and identity verification.The findings in this study journal represent a substantial advancement in the field of facial expression recognition, to sum up. The suggested technique has the potential to improve the accuracy and dependability of facial analysis systems. The proposed approach needs to be validated and improved through additional study, experimentation, and testing on larger datasets and various populations. We can improve human-computer interaction, emotion analysis, and security applications by further exploring and developing the capabilities of facial expression recognition systems. This will ultimately result in a more seamless integration of facial analysis into different facets of our daily lives.

References

[1] Ding, Hui, et al. “FaceNet2ExpNet: Regularizing a Deep Face Recognition Net for Expression Recognition.” 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 2017, pp. 118–26. IEEE Xplore, https://doi.org/10.1109/FG.2017.23.

[2] Shen, Wei, et al. DeepContour: A Deep Convolutional Feature Learned by Positive-Sharing Loss for Contour Detection. 2015, pp. 3982–91. ResearchGate, https://doi.org/10.1109/CVPR.2015.7299024.

[3] Face Recognition Through Different Facial Expressions | Journal of Signal Processing Systems. https://dl.acm.org/doi/abs/10.1007/s11265-014-0967-z. Accessed 17 May 2023.

[4] Goldman, Alvin I., and Chandra Sekhar Sripada. “Simulationist Models of Face-Based Emotion Recognition.” Cognition, vol. 94, no. 3, Jan. 2005, pp. 193–213. PubMed, https://doi.org/10.1016/j.cognition.2004.01.005.

[5] Luo, Yuan, et al. “Facial Expression Recognition Based on Fusion Feature of PCA and LBP with SVM.” Optik, vol. 124, Sept. 2013, pp. 2767–70. ResearchGate, https://doi.org/10.1016/j.ijleo.2012.08.040.

[6] Shojaeilangari, Seyedehsamaneh, et al. “Robust Representation and Recognition of Facial Emotions Using Extreme Sparse Learning.” IEEE Transactions on Image Processing: A

DOI: 10.54254/2755-2721/37/20230510

Publication of the IEEE Signal Processing Society, vol. 24, no. 7, July 2015, pp. 2140–52. PubMed, https://doi.org/10.1109/TIP.2015.2416634.

[7] Shan, Caifeng, et al. "Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study." Image and Vision Computing, vol. 27, no. 6, May 2009, pp. 803–16. ScienceDirect, https://doi.org/10.1016/j.imavis.2008.08.005.

[8] LeCun, Y., et al. "Backpropagation Applied to Handwritten Zip Code Recognition." Neural Computation, vol. 1, no. 4, Dec. 1989, pp. 541–51. IEEE Xplore, https://doi.org/10.1162/neco.1989.1.4.541.

[9] An Improved Illumination Normalization and Robust Feature Extraction Technique for Face Recognition Under Varying Illuminations | SpringerLink. https://link.springer.com/article/10.1007/s13369-019-03729-6. Accessed 20 May 2023.

[10] Wright, John, et al. "Robust Face Recognition via Sparse Representation." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, Feb. 2009, pp. 210–27. IEEE Xplore, https://doi.org/10.1109/TPAMI.2008.79.

[11] Guo, Guodong, and C. R. Dyer. "Learning from Examples in the Small Sample Case: Face Expression Recognition." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 35, no. 3, June 2005, pp. 477–88. IEEE Xplore, https://doi.org/10.1109/TSMCB.2005.846658.