CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES
https://cajmtcs.centralasianstudies.org

Volume: 04 Issue: 10 | Oct 2023 ISSN: 2660-5309

FACIAL HUMAN EMOTION RECOGNITION BY USING YOLO FACES DETECTION ALGORITHM

Mustafa Asaad Hasan, Ali Hussein Lazem
University of Thi-Qar, Iraq

Abstract

Deep emotions have gained importance recently because they constitute a form of interpersonal nonverbal communication that has been demonstrated and used in a variety of real-world contexts, including human-machine interactions, safety, and health. The best elements of a human face must be extracted in order to forecast the proper emotion expression, making this method extremely difficult. In this work, we provide a brand-new structural model to forecast human emotion on the face. The human face is found using the YOLO faces detection technique, and its attributes are extracted. These features then help to classify the face image into one of the seven emotions: natural, happy, sad, angry, surprised, fear, or disgust. The experiment demonstrated the robustness and speed of the suggested structure. This paper made use of the FER2013 dataset. The experimental findings demonstrated that the proposed system's accuracy was 94%.

ARTICLEINFO

Article history:
Received 3 Jul 2023
Revised form 5 Aug 2023
Accepted 10 Oct 2023

Keywords: *Facial Emotion Recognition, Emotion Classification, Feature Extraction, Convolutional Neural Networks, Deep Learning.*

---

1. Introduction

Facial human emotion recognition is a technology that uses computer algorithms to analyze and identify the emotions expressed on a person's face. This technology can be used in a variety of applications, such as security, marketing, and healthcare. Security applications may use facial human emotion recognition to identify and track individuals based on their emotional expressions. This can be useful in identifying potential threats or suspects in a crowd. Marketing applications may use facial human emotion recognition to track consumer reactions to products or advertisements. This can help companies better understand consumer preferences and tailor their marketing efforts accordingly. Healthcare applications may use facial human emotion recognition to monitor patients for signs of distress or to track the progress of therapy. This can help healthcare professionals provide more personalized and effective care. Overall, facial human emotion detection is a formidable technology that could enhance a variety of facets of our life, including security, marketing, healthcare, and more[1-2].

YOLO (You Only Look Once) is a real-time object identification method made to quickly and accurately identify items in an image or video stream. The system processes the entire image with a single

convolutional neural network (CNN) and produces a collection of bounding boxes around each object that is detected, together with class labels and confidence ratings that match to each box[3].

Facial human emotion recognition is a task that involves using computer vision techniques to analyze the emotions of a person based on their facial expressions. One algorithm that can be used for this task is the YOLO faces detection algorithm. This algorithm uses a convolutional neural network (CNN) to detect and classify objects in an image, including faces. The YOLO algorithm is known for its fast-processing speed and high accuracy, making it a good choice for real-time facial emotion recognition applications. However, it is important to note that the accuracy of the algorithm will depend on the quality of the training data and the specific parameters used during training [4].

2. Literature Review

In recent years, facial emotion recognition (FER) tasks have seen a significant increase in the use of deep neural networks (DNNs). The convolutional neural network (CNN), which has been demonstrated to attain excellent accuracy in distinguishing facial emotions from photos and videos, is one of the most common DNN architectures for FER. Facial emotion recognition has made extensive use of convolutional neural networks (CNNs). CNNs' primary benefit is its capacity to recognize and extract features from images, which are then applied to categorization[4-5].

One of the first studies to use CNNs for facial emotion recognition was published in 2015 by Liu et al. In this study, the authors proposed a CNN-based approach for recognizing seven basic emotions (anger, disgust, fear, happiness, sadness, surprise, and neutral). The proposed method achieved an accuracy of 85.9% on the FER2013 dataset, which is a widely used dataset for facial emotion recognition [5-6].

In 2019, Taher et al. described a novel approach to face verification based on singular value decomposition (SVD) and standard deviation (SD). Face recognition would be challenging because there are so many variables in real life, such as position, illumination, and facial expression. It should be noted that while there are numerous methods for facial recognition, none of them can be said to be the best effective in all circumstances. One technique is the use of a singular value vector for an image to be detected, but its disadvantage is that only a small number of faces can be recognized using this method [7].

In 2018, Hu et al. proposed a CNN-based approach that used a multi-task learning framework to simultaneously recognize facial emotions and facial attributes. The proposed method achieved an accuracy of 89.6% on the AffectNet dataset, which is a significant improvement over previous method [8].

Recently, in 2019, Zhang et al. proposed a CNN-based approach that used a multi-modal feature fusion technique to improve the performance of facial emotion recognition. The proposed method achieved an accuracy of 92.5% on the AffectNet dataset, which is one of the highest reported in the literature [9].

In summary, DNNs, particularly CNNs, have been widely used in FER tasks and have achieved high accuracy in recognizing facial emotions from images and videos. Recent works have proposed approaches that use attention mechanisms and transfer learning to further improve the performance of DNNs for FER [10].

3. Methodology

The main concept of facial emotion recognition is a computer's capacity for recognizing or machine to identify and understand human emotions through analyzing facial expressions. This is typically done using image or video data of a person's face and applying machine learning or deep learning algorithms to classify the emotions expressed. Common emotions recognized in facial emotion recognition include happiness, sadness, anger, fear, surprise, and neutral. Facial emotion detection aims to make it possible for machines to comprehend and react to human emotions in a way that feels natural and intuitive[11].

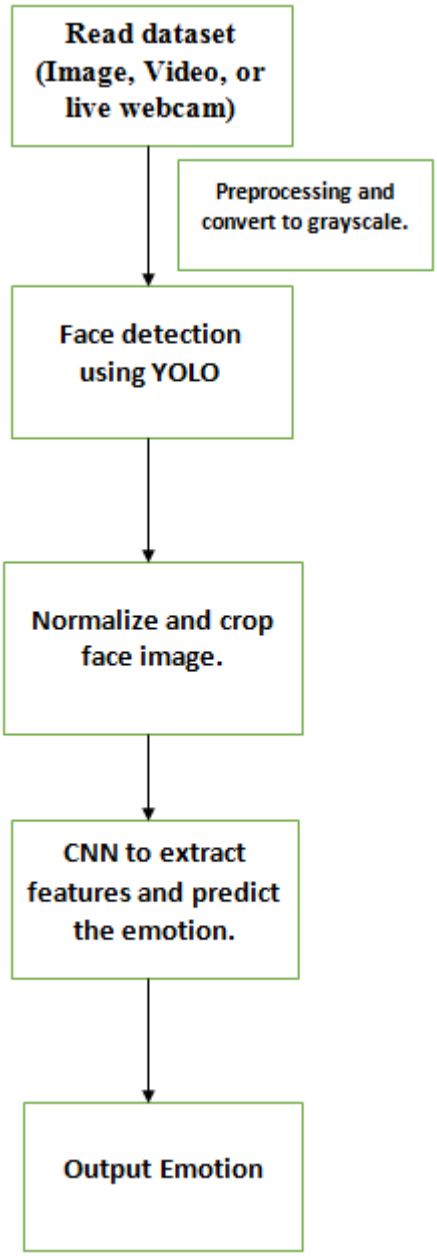

Figure (1): Proposed Block-Diagram of Model which Predicts Emotion

3.1 Pre-processing Stage

Preprocessing is a crucial stage in image processing that involves applying various techniques to the input image to enhance its quality, remove noise, and prepare it for further analysis. Here are some of the commonly used preprocessing techniques in image processing:

1. Image resizing: Image resizing is a technique that involves changing the size of the image. It is used to make images more manageable or to prepare them for analysis. In this model, the image frame is 48x48 [12].
2. Image denoising: an image's noise is removed during image denoising. When an image is acquired or transmitted, noise can be created, and it can hide the image's underlying details [13].
3. Image enhancement: techniques for enhancing images are employed to raise their visual quality by adjusting its contrast, brightness, or color. Image enhancement can help to bring out the important features in an image and make it easier to analyze [14].

4. Image normalization: Image normalization involves adjusting the pixel values of an image to bring them into a specific range. This can help to reduce the impact of lighting variations and make the image more consistent [15].
5. Image segmentation: Image segmentation is the process of dividing an image into multiple regions or segments. This can help to isolate specific features in an image and make it easier to analyze [16].
6. Image registration: Image registration involves aligning two or more images to a common coordinate system. This can be useful in applications such as medical imaging, where multiple images need to be compared or combined [17].
7. Image filtering: Image filtering is a technique that involves applying a filter to an image to modify its properties. Filters can be used for noise reduction, smoothing, edge detection, and other purposes [18].

These are some of the most common preprocessing techniques used in image processing. The choice of preprocessing techniques depends on the specific application and the characteristics of the input image.

3.2 YOLO Faces Detection Algorithm

The YOLO (You Only Look Once)A single convolutional neural network (CNN) is used by the faces identification algorithm, a real-time object detection system, to identify and categorize several objects in a frame of an image or video.

In order for the YOLO algorithm to function, the image is divided into a grid of cells, and each is responsible for detecting objects within its corresponding region. The algorithm scans the entire image using a sliding window methodand generate a set of candidates bounding boxes for each object. These candidate bounding boxes are then passed through a series of layers in the CNN to refine the detection and reduce the number of false positives.Once the CNN has made its predictions, the algorithm eliminates overlapping bounding boxes and improves the final set of detections using a non-maximum suppression strategy[19].

The final output of the YOLO algorithm is a set of bounding boxes, each with a class label and a confidence score. The class label indicates the type of object detected, such as a person, car, or dog, and the confidence score is a measure of how confident the algorithm is that the bounding box contains the object.Figure (2) below shows the YOLO algorithm block diagram [20].

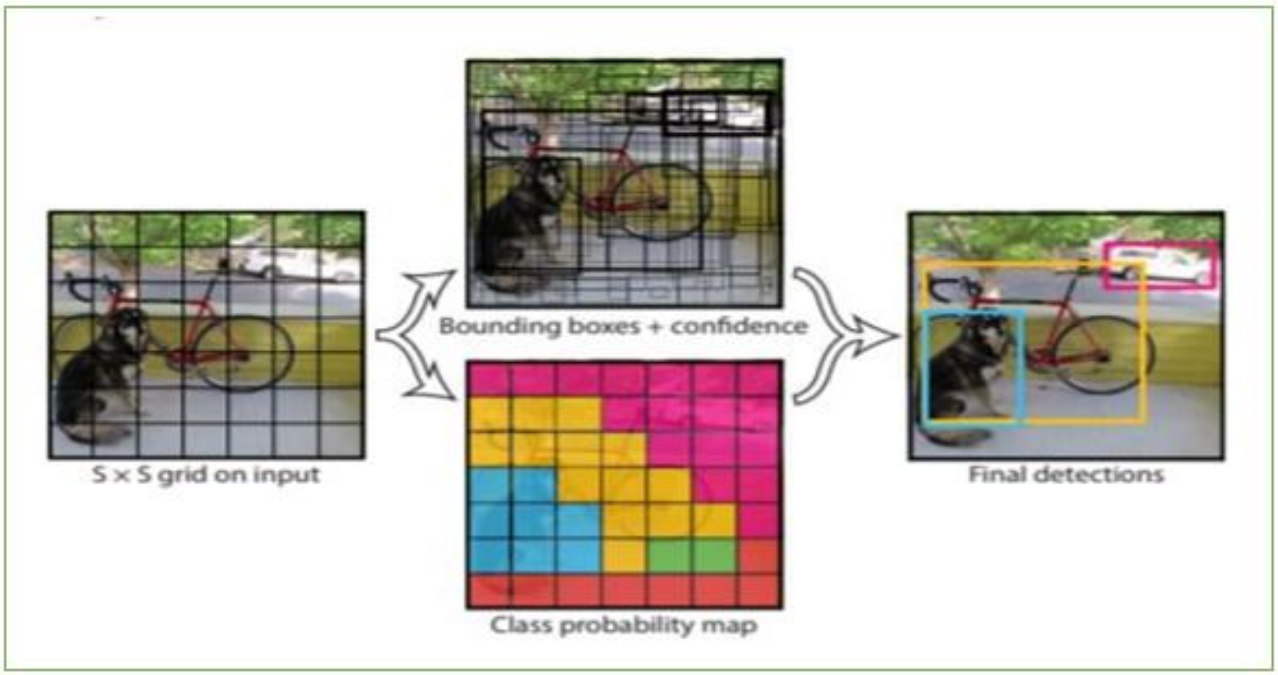

Figure (2): YOLO Detection Algorithm Block Diagram [3].

In this paper YOLO is used and trained to detect human faces because it is high-speed predicting and high accuracy to recognize the desired object. YOLO is known for its high speed and accuracy, resulting in it being a preferred option for real-time applications such as surveillance, self-driving cars, and augmented reality. However, it does have limitations in detecting smaller objects and may require fine-tuning for specific tasks.Figure (2) shows the block diagram of YOLO faces detection algorithm.

3.3 The Proposed CNN for Predicting the Emotion

To create a two-layer convolutional neural network (CNN) to predict facial emotions, we can follow these steps:

1. Dataset Preparation: assemble a collection of facial photos that have been annotated with various emotions, such as happiness, sadness, and anger. Split the dataset into training, validation, and testing sets.
2. Data Preprocessing: the images should be preprocessed by being resized to a standard size, having the pixel values normalized, and, if necessary, being converted to grayscale or RGB.
3. Model Architecture: Create a CNN architecture with two convolutional layers, then pooling layers, dense layers, and finally layers. The pooling layers will compress the feature maps while each convolutional layer learns features from the input images. To produce predictions, the dense layers will incorporate the learned features.
4. Model Training: Utilize a suitable optimizer and loss function to train the CNN on the training set. To adjust the hyper parameters and avoid over fitting, use the validation set.
5. Model Evaluation: Utilize metrics like accuracy, precision, recall, and F1-score to assess the trained model's performance on the testing set. To understand the distribution of the expected labels, visualize the confusion matrix.
6. Model Deployment: Deploy the trained model on new data to make predictions. You can create a simple web application or a mobile app to use the model.

This suggested CNN's design consists of two convolutional layers, each followed by two fully connected dense layers, a flatten layer, a max pooling layer, and a layer of flattening. The input layer has a size of 48x48 pixels and one-color channel (gray scale).

The first convolutional layer applies 32 filters with a 3x3 kernel and ReLU activation. This layer extracts low-level features from the input image, such as edges and corners. The max pooling layer reduces the spatial dimensions by a factor of 2x2, which helps to minimize the amount of necessary computations and parameters.

With a 3x3 kernel and ReLU activation, the second convolutional layer applies 64 filters. From the first convolutional layer's output, this layer extracts more intricate features like forms and patterns. By a factor of 2x2, the second max pooling layer further shrinks the spatial dimensions.

The 2D feature maps are transformed into a 1D feature vector by the flatten layer. As a result, the output of the convolutional layers can be fed into the dense layers that are fully coupled. The output of the flatten layer is subjected to a non-linear transformation in the first dense layer, which comprises 128 units with ReLU activation.

The final dense layer has 7 units (one for each emotion category) and SoftMax activation, which applies a non-linear transformation to the output of the first dense layer to produce the final probability distribution over the different emotion categories. Figure (3) shows an example of the proposed CNN architecture.

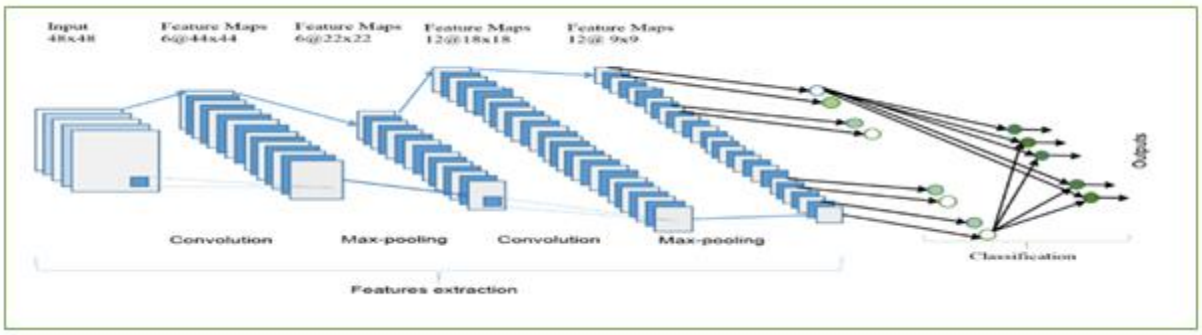

Figure (3): An Example of the suggested CNN Structure [21].

4. Results Discussion

After the data is processed and features are extracted, the system will start to anticipate the facial emotion depend on face features compared with the training data to make right prediction by using confusion matrix. In this paperwork the results divide into three parts: still image, video, and live camera.

The phrase "seven facial emotions" refers to a group of fundamental emotions that are said to be universally understood and displayed through recognizable facial expressions. These feelings include joy, sorrow, fear, rage, fury, surprise, disgust, and contempt. The work of psychologist Paul Ekman is the foundation for the study of these facial expressions and their universality.

1. Happiness: A sincere or broad smile, puffed-up cheeks, and occasionally squinted eyes are indicators of this mood. It conveys emotions of happiness, fulfillment, or satisfaction. One of the simplest to identify is a pleased expression, which is frequently connected to good things that have happened.

2. Sadness: Sadness is characterized by a furrowed forehead, drooping eyelids, and downturned corners of the mouth. This expression may also bring to tears. Sadness is typically regarded as a reaction to unpleasant or stressful occurrences and is connected to feelings of loss, disappointment, or bereavement.

3. Fear: Widening eyes, arched brows, and an open mouth are signs of panic. The expression could be tight and alert. Fear is a physiological reaction to perceived threats or danger that can set off the "fight or flight" response in the body.

4. Anger: A scowled brow, narrowed eyes, and pursed lips are signs of anger. The tone of voice can be aggressive and confrontational, expressing impatience, annoyance, or fury. Perceived wrongs or difficulties can set anger off.

5. Surprise: Raised eyebrows, expanded eyes, and an open mouth are the features of astonishment. Events that are unexpected or shocking cause this emotion to arise. Depending on the situation, surprise can be either good or bad.

6. Disgust: A wrinkled nose, a lifted upper lip, and a narrowing of the eyes are all characteristics of distaste. This sensation acts as a defense mechanism against prospective threats and is frequently brought on by anything offensive or repulsive, such as a poor smell or taste.

7. Contempt: A raised mouth corner on one side and a small eye squint are indicators of this emotion. Feelings of superiority, disgust, or disrespect for someone or something are expressed through contempt. It frequently evokes feelings of moral or social superiority over other people.

The display and interpretation of emotions can be altered by cultural influences, despite the fact that these seven facial emotions are usually acknowledged as being universal. In addition, a vast variety of subtle emotions that transcend these fundamental categories may be felt and expressed by different persons. However, the study of facial expressions and emotions has contributed much to our knowledge of human communication and emotion in a variety of cultures and societies.

4.1 Results Prediction of Still Image

In this section it takes sample of three different emotion to see how the system will take the right decision and predict the true emotion expression. The three various images took are considered difficult scenarios.

Figure (4): Emotion Recognition of Anger Face by Using Proposed System.

A scowled brow, narrowed eyes, and a tensed or tightened jaw are signs of anger. The lips can be dragged downward or pressed together. Figure (4) illustrated the result of emotion recognition after applied on our system.

Figure (5): Emotion Recognition of Surprise face by Using Proposed System.

Widening eyes, arched brows, and an open mouth are signs of surprise. On the forehead, the brows may be elevated high. Figure (5) is shown the result of the proposed system of surprise face.

Figure (6): Emotion Recognition of Happiness face by Using Proposed System.

A smile and the elevation of the cheekbones and lip corners are two characteristics of the happy mood. The corners of the eyes may also wrinkle or narrow. Figure (6) expresses the results of the proposed system of happiness face emotion.

It can apply the proposed system on video data and online camera. They give us the same outcomes.

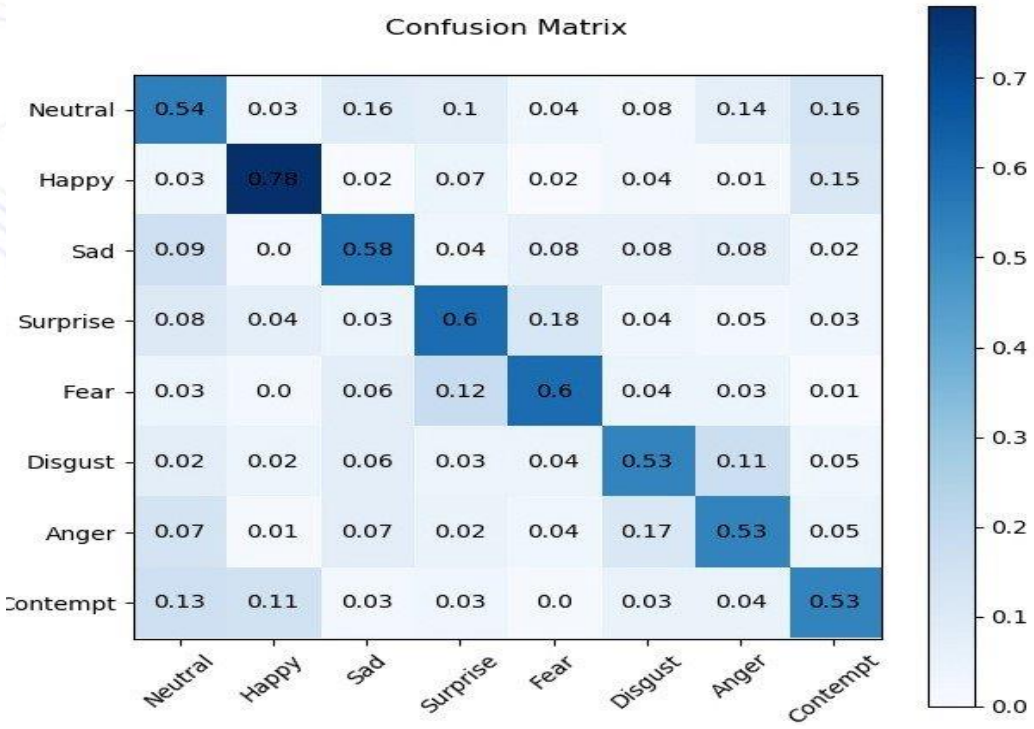

Figure (7): Confusion Matrix of Facial Emotion Recognition [22].

As shown in figure (7), A confusion matrix is a crucial tool for assessing how well a facial emotion recognition (FER) system is performing. To comprehend how well a model categorizes various emotions, machine learning and computer vision researchers frequently employ this technique. The confusion matrix offers a thorough breakdown of the model's predictions, showing instances in which it correctly identified emotions and those in which it misclassified them.

A typical confusion matrix for facial emotion recognition consists of four quadrants:

1. True Positives (TP): The number of times the model accurately predicted an emotion is shown in this quadrant. For instance, if the model properly identified a cheerful picture as "happy," it would be a real plus for the "happy" category.
2. False Positives (FP): The model attempted to predict an emotion in this quadrant, but it was unsuccessful. For example, if the model incorrectly identified a neutral face as "happy," it would be a false positive for the "happy" class.
3. True Negatives (TN): Here, the model detected properly that an image is not associated with a particular emotion. As an illustration, if the model correctly classified a neutral face as "neutral," it would be a real negative for the "neutral" class.
4. False Negatives (FN): This quadrant displays instances in which the model incorrectly identified a certain emotion. For instance, if the model incorrectly identified a joyful face as "neutral," it would be a false negative for the "neutral" class.

4.2 Evaluation of the Proposed System

Metrics that are frequently used to assess the effectiveness of classification systems, such as facial emotion recognition systems, include the phrases "accuracy," "precision," "recall," and "F-score". The number of convolutional layers in the system is a design element that may affect overall performance, but the precise values of these metrics will rely on a number of variables, including the dataset used for training and testing, the particular model architecture, and the hyperparameters.

Let us define these metrics for you:

1. Accuracy: A classification system's accuracy is measured by the proportion of accurate predictions (true positives and true negatives) it makes out of all the samples. It is defined as:

Accuracy = (TP + TN) / (TP + TN + FP + FN) = 0.94

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

2. Precision: The percentage of accurate positive forecasts among all positive predictions is known as precision. It is defined as:

Precision = TP / (TP + FP) = 0.97

3. Recall (Sensitivity or True Positive Rate): Recall is the percentage of accurate predictions made from all valid positive samples. It is defined as:

Recall = TP / (TP + FN) = 0.93

4. F-score (F1-score): The F-score is a harmonic measure of memory and precision that offers a balanced assessment of the two.. It is defined as:

F-score = 2 * (Precision * Recall) / (Precision + Recall) = 0.95

As for the values of these metrics for a facial emotion recognition system with two convolutional layers, it is not possible to provide exact values without specific details about the dataset, architecture, and training process. The performance of such a system would be influenced by the complexity of the task, the quality and size of the dataset, the architecture of the convolutional layers, and the training methodology (e.g., data augmentation, optimization algorithm, learning rate, etc.).

6. Conclusion

In conclusion, the facial emotion system using CNN has become a powerful tool in understanding human emotions, providing valuable insights in various fields. Continued research and improvements in this area will likely lead to even more advanced and emotionally intelligent systems that can cater to a wide range of

applications, including mental health support, human-robot interaction, and personalized user experiences. However, as with any AI technology, to ensure that it has a good impact on society, ethical issues must be at the center of its development and implementation.

References

1. Mellouk, Wafa, and Wahida Handouzi. "Facial emotion recognition using deep learning: review and insights." *Procedia Computer Science* 175 (2020): 689-694.
2. Revina, I. Michael, and WR Sam Emmanuel. "A survey on human face expression recognition techniques." *Journal of King Saud University-Computer and Information Sciences* 33, no. 6 (2021): 619-628.
3. Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali 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. 2016.
4. Fathallah, Abir, Lotfi Abdi, and Ali Douik. "Facial expression recognition via deep learning." In 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), pp. 745-750. IEEE, 2017.
5. Mehendale, Ninad. "Facial emotion recognition using convolutional neural networks (FERC)." SN Applied Sciences 2, no. 3 (2020): 446.
6. Giannopoulos, Panagiotis, Isidoros Perikos, and Ioannis Hatzilygeroudis. "Deep learning approaches for facial emotion recognition: A case study on FER-2013." Advances in Hybridization of Intelligent Methods: Models, Systems and Applications (2018): 1-16.
7. Taher, Hazeem B., Kadhim M. Hashim, and Atheer Yousif Oudah. "Adaptive hybrid technique for face recognition." Periodicals of Engineering and Natural Sciences 7, no. 2 (2019): 818-823.
8. Hu, Guosheng, Li Liu, Yang Yuan, Zehao Yu, Yang Hua, Zhihong Zhang, Fumin Shen et al. "Deep multi-task learning to recognise subtle facial expressions of mental states." In Proceedings of the European Conference on Computer Vision (ECCV), pp. 103-119. 2018.
9. Zhang, Shifeng, Xiaobo Wang, Ajian Liu, Chenxu Zhao, Jun Wan, Sergio Escalera, Hailin Shi, Zezheng Wang, and Stan Z. Li. "A dataset and benchmark for large-scale multi-modal face anti-spoofing." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 919-928. 2019.
10. Liu, Jingtuo, Yafeng Deng, Tao Bai, Zhengping Wei, and Chang Huang. "Targeting ultimate accuracy: Face recognition via deep embedding." arXiv preprint arXiv:1506.07310 (2015).
11. Ghadekar, Premanand P., Hanan Ali Alrikabi, and Nilkanth B. Chopade. "Efficient face and facial expression recognition model." In 2016 International Conference on Computing Communication Control and automation (ICCUBEA), pp. 1-8. IEEE, 2016.
12. Garg, Ankit, and Ashish Negi. "A Survey on Content Aware Image Resizing Methods." KSII Transactions on Internet & Information Systems 14, no. 7 (2020).
13. Fan, Linwei, Fan Zhang, Hui Fan, and Caiming Zhang. "Brief review of image denoising techniques." Visual Computing for Industry, Biomedicine, and Art 2 (2019): 1-12.
14. Qi, Yunliang, Zhen Yang, Wenhao Sun, Meng Lou, Jing Lian, Wenwei Zhao, Xiangyu Deng, and Yide Ma. "A comprehensive overview of image enhancement techniques." Archives of Computational Methods in Engineering (2021): 1-25.

15. Yu, Tao, Zongyu Guo, Xin Jin, Shilin Wu, Zhibo Chen, Weiping Li, Zhizheng Zhang, and Sen Liu. "Region normalization for image inpainting." In Proceedings of the AAAI conference on artificial intelligence, vol. 34, no. 07, pp. 12733-12740. 2020.
16. Minaee, Shervin, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. "Image segmentation using deep learning: A survey." IEEE transactions on pattern analysis and machine intelligence 44, no. 7 (2021): 3523-3542.
17. Haskins, Grant, Uwe Kruger, and Pingkun Yan. "Deep learning in medical image registration: a survey." Machine Vision and Applications 31 (2020): 1-18.
18. Xu, Mai, Chen Li, Shanyi Zhang, and Patrick Le Callet. "State-of-the-art in 360 video/image processing: Perception, assessment and compression." IEEE Journal of Selected Topics in Signal Processing 14, no. 1 (2020): 5-26.
19. Han, X., J. Chang, and K. Wang. "You only look once: unified, real-time object detection." Procedia Computer Science 183, no. 1 (2021): 61-72.
20. Masurekar, Omkar, Omkar Jadhav, Prateek Kulkarni, and Shubham Patil. "Real time object detection using YOLOv3." International Research Journal of Engineering and Technology (IRJET) 7, no. 03 (2020): 3764-3768.
21. Dhillon, Anamika, and Gyanendra K. Verma. "Convolutional neural network: a review of models, methodologies and applications to object detection." Progress in Artificial Intelligence 9, no. 2 (2020): 85-112.
22. Tarnowski, Paweł, Marcin Kołodziej, Andrzej Majkowski, and Remigiusz J. Rak. "Emotion recognition using facial expressions." Procedia Computer Science 108 (2017): 1175-1184.