Project name: Face Expression Recognition

Team Members	ID
Mohamed Ashraf Hassan	20190421
Mahmoud Kamal Ahmed	20190501

1. Introduction

This report presents a scientific analysis of a Convolutional Neural Network (CNN) model for emotion recognition. The model aims to classify facial expressions into seven different emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral. Emotion recognition plays a crucial role in various fields, including human-computer interaction, psychology, and artificial intelligence.

2. Dataset

The dataset used for training and testing the model consists of facial images collected from diverse sources. The images are carefully curated to cover a wide range of individuals, demographics, and expressions. The dataset is divided into a training set and a validation set, ensuring a balanced distribution of samples across different emotion categories. Ethical considerations are taken into account to protect the privacy and consent of the individuals in the dataset.

3. Preprocessing

Preprocessing of the images is an essential step to ensure optimal input for the model. The images undergo several preprocessing techniques, including:

Image decoding: The images are read and decoded using appropriate libraries, maintaining the integrity of the image data.

Color space conversion: The images are converted from the default color space (BGR) to the RGB color space, which is more suitable for human perception and subsequent analysis.

Image enhancement: To improve the quality and clarity of the images, techniques such as Gaussian blur are applied. This helps reduce noise and enhance important facial features.

4. Face Detection

Accurate face detection is crucial for isolating the facial region of interest in the images. The MTCNN (Multi-task Cascaded Convolutional Networks) algorithm is employed for robust and efficient face detection. This algorithm utilizes a cascaded architecture to detect faces at different scales and accurately localize facial landmarks.

5. Image Cropping

After face detection, the images are cropped to extract the facial region. This step focuses on isolating the essential facial features relevant for emotion recognition. The cropping process ensures that the model receives input containing only the facial area, eliminating irrelevant background information.

6. Model Architecture

The Convolutional Neural Network (CNN) model employed for emotion recognition utilizes a deep learning architecture that leverages multiple layers to extract meaningful features from facial images. The architecture consists of the following key components:

- Convolutional Layers: The model starts with multiple convolutional layers that perform spatial feature extraction. These layers apply a set of learnable filters to the input image, capturing various low-level and high-level features such as edges, textures, and patterns.
- Batch Normalization: Batch normalization layers are inserted after each convolutional layer. They normalize the activations of the previous layer, reducing internal covariate shift and helping the model converge faster and generalize better.

- Max Pooling Layers: Max pooling layers follow the convolutional layers and downsample the feature maps, reducing their spatial dimensions while retaining the most salient features. This helps in reducing computational complexity and focusing on the most relevant information.
- Dropout: Dropout layers are applied after some convolutional and fully connected layers. Dropout randomly deactivates a fraction of the neurons during training, which prevents overfitting and encourages the model to learn more robust and generalizable features.
- Fully Connected Layers: The convolutional layers are followed by a series of fully connected layers. These layers integrate the extracted features and perform higher-level feature representation. They capture global relationships and patterns in the input data.
- Softmax Activation: The final layer of the model utilizes the softmax activation function, which converts the raw output of the model into a probability distribution over the seven emotion classes. It assigns a probability to each class, indicating the model's confidence in the predicted emotion label.
- Model Compilation: The model is compiled using the Adam optimizer, which adapts the learning rate dynamically during training. The categorical cross-entropy loss function is used as the optimization criterion, measuring the discrepancy between the predicted probabilities and the true emotion labels.

7. Model Training and Performance

The model undergoes training using the curated training set, and its performance is evaluated on the validation set. The training process involves iteratively updating the model's parameters to minimize the loss function. The training is performed over 50 epochs, which refers to the number of times the model has seen the entire training dataset.

During the initial stages of training, the model's accuracy starts at a relatively low level of 23%. However, as the training progresses, accuracy gradually improves and reaches a commendable level of 95% after 50 epochs. This improvement indicates that the model learns to effectively recognize and classify emotions from facial expressions, capturing the underlying patterns and variations across different individuals and emotions.

The significant increase in accuracy demonstrates the model's ability to extract discriminative features from facial images and generalize well to unseen data. It highlights the model's capacity to capture the complex relationships between facial expressions and emotions, enabling accurate predictions.

It is important to note that these results are specific to the dataset provided and training process. Generalization and performance on other datasets may vary, and further evaluation on diverse datasets is recommended to assess the model's robustness and effectiveness in real-world scenarios.

Overall, the achieved accuracy of 95% showcases the potential of the model for emotion recognition tasks and emphasizes the significance of deep learning techniques in analyzing facial expressions.

8. Model Evaluation

The trained model is evaluated using a separate validation set that was not seen during the training phase. Performance metrics such as accuracy, precision, recall, and F1-score are computed to assess the model's ability to correctly classify emotions. Additionally, confusion matrices and classification reports are generated to provide detailed insights into the model's performance across different emotion categories.

9. Image Enhancement

To showcase the model's capabilities, an example image (p066st0k.jpg) is used for image enhancement. The grayscale image is loaded and processed using histogram equalization techniques. Histogram equalization enhances the image's contrast and improves its visual appearance, potentially aiding the model's performance in emotion recognition tasks.

10. Prediction and Analysis

The enhanced image is resized to the required input shape and passed through the trained model for emotion prediction. The model predicts the most probable emotion label for the given facial expression. The prediction is then compared to the ground truth label to assess the model's accuracy and its ability to recognize the underlying emotions. The grayscale image, along with the predicted emotion label, is displayed for visual analysis.

11. Conclusion

The presented CNN model demonstrates a scientific approach to emotion recognition from facial expressions. Through rigorous preprocessing, model architecture design, and training procedures, the model achieves promising results in classifying facial emotions. The evaluation metrics provide insights into the model's strengths and areas for improvement. Further research and evaluation on larger and more diverse datasets are recommended to validate the model's performance in real-world scenarios and its generalizability across different populations.