**Facial Emotion Detection**

Facial Emotion Detection using Convolutional Neural Networks (CNNs) for detecting emotions in images.

Team Members

* Mohamed Hisham
* Abdelrahman Mostafa
* Abdelrahman Amin

**Introduction**

Emotions are fundamental to human communication and interaction. They play a pivotal role in conveying one's feelings, intentions, and responses in social contexts. Recognizing and understanding these emotional cues from facial expressions has been a long-standing pursuit in psychology, neuroscience, and computer science. In recent years, advancements in computer vision and machine learning have made it possible to automate this process with high accuracy, opening a multitude of practical applications.

This project aims to contribute to the advancement of facial emotion detection by creating a robust and efficient model that can accurately recognize a range of emotional states from facial images. We believe that a reliable and automated emotion recognition system has the potential to revolutionize several industries, from entertainment and marketing to healthcare and beyond.

Emotional recognition from facial expressions is a complex task. It involves capturing and analyzing subtle changes in facial features, such as eyebrow movement, eye-widening, or mouth curvature, that are associated with various emotions, including happiness, sadness, anger, surprise, fear, and more. By creating a model that can reliably distinguish these emotional cues, we can pave the way for transformative applications. For example, educational software might adapt its content and pace to a student's emotional state, or customer service chatbots could offer more empathetic responses based on the user's mood.

In the realm of computer vision and artificial intelligence, the ability to discern human emotions from facial expressions is an essential and captivating domain of research. The human face is a rich source of emotional cues, expressing a myriad of sentiments, from unbridled joy to profound sorrow. Understanding these emotions from facial features not only finds applications in human-computer interaction but also plays a crucial role in fields such as psychology, marketing, and user experience design. Facial emotion detection holds the potential to transform industries, from gauging audience responses in marketing campaigns to improving mental health through emotion-aware applications. It is an interdisciplinary field at the intersection of computer vision, machine learning, and psychology, offering solutions to practical challenges and a deeper understanding of human behavior.

The primary objective of this project is to design and implement a facial emotion detection system using convolutional neural networks (CNNs). CNNs have demonstrated remarkable success in a wide range of computer vision tasks, making them the prime choice for the intricate task of recognizing emotions from facial images. This project leverages the power of deep learning to create a model capable of identifying and classifying a range of human emotions based on intricate patterns and variations in facial expressions.

Facial emotion detection using CNN is an emerging research field that has gained immense success in various areas of implementation, such as classification, recommendation models, object recognition, etc. Facial emotion detection and recognition are important aspects of human-computer interaction. Identifying various emotions is sometimes a challenging job as no specified prototype or framework is there to differentiate the various kinds of sentiments, and there are also various complications while recognizing the facial emotion expression. Machine learning techniques, deep learning, and neural network algorithms are used for emotion recognition. The proposed techniques use convolutional neural networks (CNNs) to detect anger, disgust, happiness, fear, sadness, calmness, and surprise. The accuracy of the system can be improved by deploying various pre-processing and feature extraction techniques on the input images. The dataset used for testing and training is FER2013, and the proposed model gives an accuracy of 73%. This dataset consists of thousands of facial images, each labeled with one of several emotions, including happiness, sadness, anger, surprise, and more. The diverse nature of this data set ensures that the model is exposed to a broad spectrum of emotions and expressions, enhancing its ability to generalize. The FER model can be beneficial for business promotions, lie detection, and areas requiring additional security.

By the end of this endeavor, we aim to achieve a robust and accurate emotion recognition model. This report will walk through the key components of the project, from data collection and preprocessing to the intricacies of the CNN architecture, training, evaluation, and, if applicable, real-time emotion detection. We will also discuss the challenges faced during the project, its broader implications, and potential avenues for future research and development.

Facial emotion detection using CNN faces several challenges, including:

* Variations in facial expressions across individuals: Facial expressions can vary significantly across individuals, making it difficult to develop a universal prototype or framework to differentiate between different emotions.
* Environmental factors: The accuracy of facial emotion detection can be affected by environmental factors such as background noise, lighting, and other people.
* Dataset quality: Having access to a diverse and comprehensive dataset is essential for training algorithms that can accurately detect emotions in real time. The quality of the dataset used for training and testing can significantly impact the accuracy of the system.
* Facial feature recognition: Emotion detection through facial feature recognition is an active domain of research in human-computer interaction (HCI). However, facial feature recognition is a challenging task, and the accuracy of the system can be affected by factors such as occlusion, pose, and illumination.
* Integration with other modalities: To improve the accuracy of facial emotion detection, it is often necessary to integrate it with other modalities, such as voice recognition. The integration of voice and facial recognition algorithms enhances the reliability and accuracy of emotion detection, making the system more useful and effective in real-world applications.

**Data Description**

The Face Expression Recognition Dataset is a valuable resource in computer vision and emotion analysis. It consists of grayscale images of human faces, carefully processed to ensure that each face is centered and occupies a huge portion of the image. FER2013 Dataset will be used in this project. This dataset serves to train and test machine learning models for facial expression recognition.

Here are the key details of this dataset:

* Image Characteristics: The dataset comprises 48x48 pixel grayscale images of human faces. The grayscale format simplifies the data by using only shades of gray, eliminating the complexities of color analysis. This makes it an efficient choice for emotion recognition tasks.
* Emotion Categories: Each image in the dataset is associated with one of seven distinct emotions. These emotions are anger, disgust, fear, happiness, sadness, surprise, and neutral. This labeling allows researchers and machine learning practitioners to train models to identify and categorize human emotions accurately.
* Data Split: The dataset is divided into two main subsets—a training set and a test set. The training set contains 28,709 images, while the test set contains 7,178 images. This separation of data is essential for training and evaluating the performance of machine learning algorithms. The model learns from the training set and is tested on the separate test set to assess its ability to generalize and recognize emotions in new, unseen data.
* Availability: As of the last update on January 3, 2019, the dataset was accessible on Kaggle. This means that it was publicly available to the data science and machine learning community, making it a valuable resource for researchers and developers working on emotion recognition tasks.
* Use Cases: The Face Expression Recognition Dataset has been used in various machine learning projects and research studies. For instance, it has been employed in the "Facial Emotion Recognition" project on Kaggle, where participants have developed machine learning models, particularly convolutional neural networks (CNNs), to recognize facial expressions from these images. CNNs are well-suited for image-based tasks like this due to their ability to automatically learn features from the data, and we could apply transfer learning to increase the accuracy of the model.
* Research Studies: In addition to machine learning projects, the dataset has also been utilized in academic research on facial expression recognition. One study mentioned the dataset as the basis for developing a deep learning model specifically designed for facial expression recognition. This highlights the dataset's importance in advancing the state of the art in emotion analysis and computer vision.

**Conclusion**

The Face Expression Recognition Dataset is a comprehensive collection of grayscale facial images categorized according to the emotional expressions they display. With almost 36,000 images labeled with seven distinct emotions, it serves as a valuable resource for training and evaluating machine learning models and advancing research in the field of facial expression recognition.

**Related Work**

The related work provided discusses numerous studies on Facial Emotion Recognition using convolutional neural networks (CNNs). Here is a summary of the key points from each study:

* Yu et al. [39] proposed a biometric quality assessment (BQA) method using a light CNN, which showed high effectiveness in FR applications.
* Sun et al. [40] used a hybrid deep learning approach combining a ConvNet based on the restricted Boltzmann machine (RBM) model for face verification, achieving excellent performance compared to other methods.
* Singh and Om [41] used a deep CNN to identify individuals from newborn infant datasets, finding that increasing the number of hidden layers did not improve identification accuracy.
* Guo et al. [42] developed a CNN-based model that used both visible light and near-infrared images for facial recognition, achieving enhanced performance and robustness to illumination variation.
* Hu et al. [43] investigated the performance of CNNs on 2D and 3D FR systems, finding better accuracy with their CNN-2 model on both types of recognition.
* Nam et al. [44] proposed a CNN model named PSI-CNN for face recognition, which outperformed the VGG-Face model and maintained stable performance with low-resolution images.
* Prasad et al. [7] studied deep learning-based face representation for various FR challenges, showing that VGG-Face and lightened CNN approaches provided satisfactory results.
* Khan et al. [45] proposed a face recognition system using portable smart glasses based on CNN, achieving high-detection and accuracy rates.
* Qin et al. [46] developed a recognition algorithm based on deep CNNs, which performed well on recognizing faces with various poses in an indoor environment.
* Menotti et al. [47] investigated deep learning approaches for iris spoof detection and fingerprint variations, emphasizing the need for comprehensive spoofing detection frameworks.
* Simón et al. [48] proposed a multimodal facial recognition approach using CNNs fused with different feature descriptors, significantly reducing the recognition error rate.
* Parkhi et al. [49] proposed the VGG-Face system, a 16-layer CNN trained on a large image database, achieving improved results in face recognition.
* Zhenyao et al. [50] used a deep network to warp faces into a standard frontal view and performed face verification using PCA and an ensemble of SVMs (Support Vector Machines).
* Guo et al. [51] developed an FR system based on CNNs and optimized training techniques, achieving high-recognition rates in less training time.
* Deep-Face [13] trained an eight-layer CNN architecture on a large face database, achieving exemplary performance in face recognition benchmarks.
* DeepID [53] used an ensemble of small CNNs and network fusion for face verification, achieving high accuracy on the LFW dataset.
* DeepID2+ [55] improved upon DeepID and DeepID2 by using a larger training set and enhancing the number of filters in all layers, resulting in sparse, selective, and robust face representations.
* Lu et al. [56] proposed the Deep Coupled ResNet (DCR) model, consisting of a trunk network and two branch networks, which achieved better performance than state-of-the-art models on the LFW and SCface datasets.

These studies demonstrate the effectiveness of CNNs for various FR tasks, including face quality assessment, verification, identification, and recognition. The researchers used different CNN architectures, optimization techniques, and datasets to improve performance and address specific challenges in FR.

**Methodology**

Two approaches were followed in the study:

* Using pre-trained CNNs (AlexNet and ResNet-50) for feature extraction and SVM for classification.
* Applying transfer learning from AlexNet for feature extraction and classification.

The stages of the study included pre-processing, face representation using CNNs, and classification. SVM was chosen as the classifier due to its effectiveness in handling nonlinear data and pattern recognition problems. The SVM classification process involved finding a hyperplane that maximizes the margin between two input classes. The study did not focus on optimizing SVM strategies.

**Experiments**

The experiments were conducted on a Windows platform with an Intel Core i7 CPU, 16 GB RAM, and NVIDIA GEFORCE GTX 1050TI. MATLAB 2018a was used for evaluating the method, feature selection, and classification. Pre-processing involved resizing images to suitable sizes (227x227 for AlexNet and 224x224 for ResNet-50). The performance of the pre-trained CNN system was evaluated based on recognition accuracy. Three datasets were used in the study:

* ORL database: It contains 10 images of 40 individuals with variations in face angles, expressions, and details.
* GTAV face database: It includes images of 44 individuals with different pose views and illuminations.
* Georgia Tech face database: It consists of images of 50 individuals taken in different sessions, orientations, and scales.

Overall, the study aimed to investigate face recognition performance using CNNs, SVM, and transfer learning. The effectiveness of different approaches and the comparison between SVM and transfer learning from pre-trained AlexNet were analyzed using multiple datasets.

Experiment 1: Pre-Trained CNN AlexNet with SVM

The experiment used a pre-trained CNN model called AlexNet. The images were resized to 227x227 pixels and converted to RGB. The data was split into 80% training and 20% test sets. Features were extracted from different layers of the network, specifically fc6, fc7, and fc8. SVM (Support Vector Machine) was trained using the extracted features. The highest recognition accuracy was achieved with features extracted from the 'fc7' layer. The model achieved high accuracy on various datasets: 100% on YTF, 99.55% on GTAV face, 99.17% on ORL, and 98% on F\_FLW. Lower accuracy was obtained on FEI (97.50%), Georgia Tech face (96%), and LFW (94%) datasets.

Experiment 2: Pre-Trained ResNet-50 Model with SVM

The experiment used a pre-trained CNN model called ResNet-50. The images were resized to 224x224 pixels and converted to RGB. The data was split into 70% training and 30% validation sets. Features were extracted from the 'fc1000' layer. SVM was trained using the extracted features. The ResNet-50 model achieved high accuracy on several datasets: 100% on GTAV face, ORL, and YTF, 98.50% on FEI, and 96% on Georgia Tech face. The specific accuracy for the LFW dataset is not mentioned in the information provided.

Summary of Experiment 2:

Both experiments used pre-trained CNN models (AlexNet and ResNet-50) for feature extraction, followed by training SVM classifiers. The experiments achieved high-recognition accuracy on various datasets, with the best results obtained using features from specific layers of the networks.

Experiment 3: Transfer learning from AlexNet for feature extraction and classification:

* Model Architecture: AlexNet, a pre-trained network with 25 layers, was used. The first 23 layers were used for feature extraction, while the last three layers were for classifying the features into 1000 groups.
* Transfer Learning: The last three layers of AlexNet were removed, and a new fully connected layer was added to match the number of classes in the new dataset. This transferred network was then trained in the new classification task.
* Training Setup: The model was trained with different numbers of epochs, and the highest accuracy was achieved with 20 epochs. A mini-batch size of 20 was used, and the network was validated every 3 training steps. The data was divided into 70% for training and 30% for validation.
* **Performance Analysis**:
  + **Accuracy**: The transfer learning model achieved the highest accuracy of 100% on the Georgia Tech face dataset, GTAV face dataset, and YouTube face dataset. The AlexNet + SVM model achieved 100% accuracy on the YTF dataset, while the ResNet-50 + SVM model achieved 100% accuracy on the GTAV face, ORL, and YTF datasets.
  + **Precision**: All approaches (AlexNet + SVM, ResNet-50 + SVM, transfer learning from AlexNet) achieved precision values between 92% and 99.50%
  + **Recall**: All approaches achieved high-recall results between 93% and 99.98%
  + **F-Measure**: The transfer learning model achieved the highest F-Measure results overall, but the ResNet-50 + SVM model obtained the highest value with the ORL dataset.
* **Testing Time**: The ResNet-50 + SVM model exhibited faster testing times compared to the other networks with all datasets. The transfer learning model using AlexNet took less time than the AlexNet + SVM model.
* **Dataset Performance**:
  + FEI Faces: All three models achieved good accuracy, with the transfer learning model achieving the highest accuracy of 98.7%
  + ORL Dataset: All three models outperformed previous state-of-the-art models, with accuracy ranging from 99.17% to 100%
  + YouTube Face Dataset: All models achieved 100% accuracy, surpassing previous results.
  + LFW Dataset: The transfer learning model achieved an accuracy of 95.63%, while the other models achieved 94%. These results were higher than some previous methods that used pre-processing techniques.

Summary of Experiment 3:

The transfer learning model from AlexNet showed improved performance compared to the AlexNet + SVM model in terms of accuracy, precision, recall, and F-Measure. It achieved high accuracy on various datasets and demonstrated better time efficiency compared to other models.