Project Title: Expression Classification from Facial Images

Name: Ali Ahmad Jan

Email: aliahmadjan141@gmail.com

Github: https://github.com/alijaan110

Summary:

The facial emotion detection project utilizes the MobileNetV2 model to recognize emotions in human faces from images. It involves data collection, annotation, and preprocessing to train the model on various emotions. MobileNetV2's efficiency makes it suitable for resource-constrained devices. Through transfer learning, the model is fine-tuned for emotion recognition. Evaluation on a validation set ensures accuracy and performance. Deployment in real-time applications enables emotion detection from live camera feeds or images. The project's significance spans human-computer interaction, psychology, and affective computing, enhancing emotion analysis capabilities for practical applications.

Project Details:

Facial emotion detection involves using machine learning to recognize emotions from human facial expressions. Potential applications include enhancing human-computer interaction, improving marketing strategies, aiding in healthcare and mental health monitoring, optimizing education and training, enabling emotion-aware robots, enhancing security and surveillance, and creating interactive experiences in entertainment and gaming. Emotion detection technology can improve user experiences, inform decision-making, and enable more empathetic interactions across various domains.

Research Paper related Facial Emotion Recognition

- **FACIAL EMOTIONAL RECOGNITION USING MOBILENET BASED TRANSFER LEARNING**

Paper link: file:///C:/Users/DELL/Downloads/102pm_39.EPRA%20JOURNALS%2013862.pdf

In this paper, facial emotion recognition is a challenging task that requires identifying subtle differences in expressions. Transfer learning with MobileNet, an efficient and accurate pre-trained CNN, enhances performance. EmoNet utilizes MobileNet and transfer learning for real-time, user-friendly facial expression classification on mobile devices. Using stochastic gradient descent and evaluation metrics, EmoNet achieves high estimation values and superior classification compared to other models. It offers an effective solution for precise and efficient facial expression recognition, applicable to emotion detection, human-computer interaction, and social robotics.

Table: 2 CK+ dataset and FER 2013 dataset for proposed and existing methods

Methods	Accuracy	
	CK+ dataset (%)	FER 2013 (%)
CNNEELM [7]	96.23	98.11
CNN [13]	90.98	89.2
Proposed EmoNet	97.62	99.17

- Emotion Recognition from Spatio-Temporal Representation of EEG Signals via 3D-CNN with Ensemble Learning Techniques

Paper link: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10136603/

This study adopted the success of 3D-CNNs in video analysis, owing to their capacity to extract and learn temporal features in addition to spatial features. Using 3D-CNN, the EEG signals are represented in 3D spatio-temporal space by first converting the 1D raw EEG streams into 2D spatial streams and then stacking the 2D spatial streams into 3D EEG block streams. A 3D MobileNet network with transfer learning was used to extract and learn features from 3D EEG blocks. Additional pools and dense layers were added to the CNN network to enhance classification capabilities. In the SEED-IV dataset, four classes of samples were classified: happiness, sadness, fear, and neutral, with an accuracy of 78.32%. The SEED dataset showed an accuracy of 88.58% for classifying the samples into three groups: positive, neutral, and negative.

I used MobileNetV2 for facial emotion detection. Here's an overview of the MobileNetV2 model:

Architecture: MobileNetV2 is a lightweight deep learning model designed for efficient image classification tasks. It is based on a streamlined version of the Inverted Residuals with Linear Bottlenecks architecture. The model employs depth-wise separable convolutions to reduce the number of parameters and computations while maintaining high accuracy.

Diagram: The detailed diagram of MobileNetV2 may vary based on the number of layers and configurations, but in essence, it consists of multiple layers of depth-wise separable convolutions, followed by batch normalization and ReLU activation functions. It also incorporates shortcut connections and expansion convolution layers.

Main Components: The main components of MobileNetV2 include:

Depth-wise separable convolutions: A combination of depth-wise convolutions and point-wise convolutions, which reduces computational complexity.

Inverted Residuals: Bottleneck layers with inverted connections that prevent overfitting.

Linear Bottlenecks: Bottleneck layers with linear activation functions for better efficiency.

Global Average Pooling: Replacing fully connected layers with global average pooling to reduce model size.

Parameters: The number of parameters in MobileNetV2 can vary depending on the model size and complexity. The original MobileNetV2 architecture has approximately 3.4 million parameters, making it suitable for deployment on resource-constrained devices.

Overall, MobileNetV2's architecture and design choices make it an excellent choice for tasks like facial emotion detection, where computational efficiency and accuracy are crucial factors.

Dataset:

The dataset selected for this task is the **Expression in-the-Wild (ExpW)**[1] dataset. The dataset can be downloaded from the following link:

https://drive.google.com/drive/folders/1SDcI273EPKzzZCPSfYQs4alqjL01Kybq

The **Expression in-the-Wild (ExpW)** dataset is for facial expression recognition and contains 91,793 faces manually labeled with expressions (Figure 1). Each of the face images is annotated as one of the seven basic expression categories: “angry (0)”, “disgust (1)”, “fear (2)”, “happy (3)”, “sad (4)”, “surprise (5)”, or “neutral (6)”.

I split dataset into Training and Validation

80% data split in Training and 20% in Validation

Results and Evaluations:

In training your facial emotion detection model using MobileNetV2! An accuracy of 86% on the validation set is quite promising, indicating that your model has learned to recognize facial emotions with a relatively high level of accuracy.

How I Improve further Result?

Improving the results of my facial emotion detection model involves a combination of data, model, and training strategies. Here are some key approaches to enhance your model's performance:

Data Augmentation: Increase the diversity of your training data by applying various data augmentation techniques, such as rotation, zooming, flipping, and brightness adjustments. This helps the model generaliz better to different variations of facial expressions.

Data Balance: Ensure that your dataset has a balanced representation of all emotion classes. If some emotions are underrepresented, consider collecting more data or using techniques like oversampling or class weighting to address the class imbalance.

Hyperparameter Tuning: Experiment with different learning rates, optimizers, batch sizes, and other hyperparameters to find the best combination for your specific dataset.

Transfer Learning: Instead of training from scratch, consider using a pre-trained model like MobileNetV2 and fine-tuning it on your dataset. This leverages the knowledge learned from a large-scale dataset and can lead to better results.

Model Architecture: Experiment with different model architectures or modify MobileNetV2 to suit your specific needs. You can try deeper or wider models, or explore other efficient architectures designed for image classification tasks.

Regularization: Apply regularization techniques like dropout or L2 regularization to prevent overfitting and improve generalization.

Learning Rate Schedule: Implement a learning rate schedule to adjust the learning rate during training. Lowering the learning rate as training progresses can help the model converge better.

Ensemble Methods: Consider using ensemble techniques like model averaging or stacking multiple models to combine their predictions and achieve better accuracy.

Error Analysis: Analyze the model's mistakes by studying the misclassified samples or confusion matrix. This can help identify patterns of misclassification and guide improvements.