

CS766 Project Proposal

Xiaohu Zhu

Introduction

Facial expression recognition is a challenging problem in computer vision that aims to identify the emotion expressed by a person's face. The problem involves detecting and analyzing facial features such as eye movement, eyebrow position, and mouth shape to classify the emotion being expressed. The goal of our project is to develop a deep learning-based facial expression recognition system that can accurately classify facial expressions in images and videos.

Facial expression recognition is a crucial aspect of understanding human emotions, as it involves interpreting the movements of facial muscles that convey various feelings. By accurately identifying and classifying facial expressions, computers can better understand and respond to human emotions, leading to improved interactions and experiences. Additionally, incorporating facial expression recognition with the face shape and identity recognition may enhance the accuracy of face recognition systems. The combination of these features can provide a more comprehensive understanding of the individual, allowing for more accurate and reliable recognition of their face. Therefore, developing a facial expression recognition system that works in tandem with facial shape and identity recognition is a promising direction in computer vision research.

Motivation

Facial expression recognition has many potential applications in areas such as human-computer interaction, emotion detection, and mental health. An accurate facial expression recognition system can be used to enhance the user experience in video games, virtual reality, and other interactive applications. It can also be used in healthcare to detect and diagnose mental health disorders such as depression, anxiety, and bipolar disorder. The project is of particular interest to us because it combines computer vision and deep learning techniques to solve a challenging real-world problem.

State-of-the-Art

The current state-of-the-art for facial expression recognition is based on deep learning models such as CNNs and RNNs. Deep Emotion Recognition on Small Datasets using Transfer Learning (FER2013) by A. Mollahosseini, D. Chan, and M. H. Mahoor (2017): This approach uses a pre-trained deep neural network (such as VGGNet or ResNet) for feature extraction and fine-tuning it on the FER2013 dataset to perform facial expression recognition.[1] The method achieves state-of-the-art performance on the FER2013 dataset. Existing approaches use pre-trained models such as VGGNet or ResNet to extract features from facial images and then train a classifier on those features. Facial Expression Recognition with Deep Convolutional Neural Networks (FER2013 and CK+) by J. H. Yang and J. Lu (2018): This approach proposes a CNN-based model for facial expression recognition that uses multiple convolutional layers and fully connected layers to learn discriminative features from facial images. The model is trained end-to-end on the FER2013 and CK+ datasets and achieves state-of-the-art performance on both datasets.[2] However, these approaches suffer from limitations such as overfitting, the need for large amounts of training data, and difficulty in capturing the temporal evolution of emotions in video sequences.

Proposed Approach

Our proposed approach is to develop a CNN-based facial expression recognition system in PyTorch that can learn to extract features directly from raw input images. The network architecture will be designed to capture the spatial and temporal features of facial expressions using multiple convolutional layers and fully connected layers. We will use data augmentation techniques such as random cropping, flipping, and rotation to increase the diversity of the training data and prevent overfitting. We will also investigate the use of attention mechanisms to focus on the most discriminative facial regions.

Novelty

The proposed approach differs from existing methods in that it does not rely on pre-trained models for feature extraction, allowing the network to learn more discriminative features from the raw input data. Additionally, we will investigate the use of attention mechanisms to focus on the most informative regions of the face. Finally, we will explore the use of transfer learning techniques to leverage pre-trained models trained on related tasks such as face recognition.

Evaluation

We will evaluate the performance of our proposed approach on two standard datasets, FER2013 and CK+, which contain labeled facial expression images and videos. We will compare the performance of our model with the state-of-the-art approaches using standard evaluation metrics such as accuracy, F1-score, and confusion matrix. We will also investigate the robustness of our model to variations in lighting, pose, and occlusions. The timeline for this project will be as follows:

Feb 24: Project proposal

- 1) Data collection and preprocessing - 1 month;
- 2) Model design and implementation - 2 months;

Apr 4: Mid-term report

- 3) Model training and hyperparameter tuning - 2 months;
- 4) Evaluation and comparison with state-of-the-art approaches - 1 month.

May 5: Final presentation

Reference

- [1] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Deep Emotion Recognition on Small Datasets using Transfer Learning," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017, pp. 10-19.
- [2] J. H. Yang and J. Lu, "Facial Expression Recognition with Deep Convolutional Neural Networks," in Proceedings of the 2018 IEEE International Conference on Multimedia and Expo (ICME), 2018, pp. 1-6.