Facial Emotion Recognition and Classification Using the Convolutional Neural Network-10 (CNN-10)

1. Summary

1.1. Motivation

The paper under review is about facial emotion recognition. It has a wide range of applications in various fields. A good example is the health sector since an individual's health condition is reflected in his facial expression. There are quite a few challenges in facial expression detection, such as 3D face posture, noise, opacity, and different lighting conditions.

1.2. Contribution

This paper presents a facial recognition method, built on top of previous works, aiming to reduce computational complexity and thus increase performance. It employs techniques like convolutional neural network-10 (CNN-10) and vision transformation (ViT), a revolutionary attention-based facial expression recognition technique. The detection accuracy with the CK+ (Cohn-Kanade+), FER-2013, and JAFFE datasets were 99.9%, 84.3%, and 95.4% respectively.

1.3. Methodology

The high-level block diagram of the CNN-10, Vision Transformer (ViT), InceptionV3, and VGG19 models utilized for this paper's analysis of facial expressions are described in this section. The data collection, cross-validation, and data augmentation methods applied in this study are among the significant elements incorporated in the diagram.

1. Dataset: Pictures of human facial expressions have been gathered and are part of a professionally managed dataset. This dataset contains labeled pictures of countless expressions on the face, including happiness, sadness, rage, and more. The input data used to train and evaluate the CNN models are these photos.
2. Data Augmentation: Data augmentation techniques are used to boost the dataset's diversity and robustness. These methods entail applying adjustments to the existing photos, such as rotation, scaling, flipping, and cropping. Data augmentation makes it easier for CNN to generalize to new images and increases the variability of the training data.
3. Cross-Validation: Cross-validation is used to evaluate the generalization and performance capabilities of the CNN models. There are numerous folds, or subsets, inside the dataset. Each fold serves as the validation set once during the multiple iterations of the training and evaluation procedure. This method ensures a thorough assessment of the CNN models' performance across various data subsets.

4. CNN Architecture: The CNN-10, vision transformer (ViT), InceptionV3, and VGG19 models are precisely chosen for their efficacy in image analysis tasks, such as facial expression recognition, and are included in the high-level block diagram.
5. Facial Expression Analysis: The facial expressions in the input images are analyzed using the trained CNN models (CNN-10, ViT, InceptionV3, and VGG19), which have been trained to classify facial expressions into various categories based on the features extracted from the images.

1.4. Conclusion

This study provided the CNN-10 techniques for facial emotion classification and compared it with other approaches like INCEPTIONV3, VGG19, and ViT. In particular, the categorization strategy used by the CNN-10 models is more accurate. In order to identify facial image data, CNN-10 is a reliable and effective computer-assisted diagnostic technique that can successfully improve the classification accuracy of facial expressions. The Kaggle dataset's enhanced images were used for classifier performance analysis, validation, and training. CNN-10 can identify with ease the photographs depicting anger, disdain, disgust, fear, happiness, and sadness that are included in the collection.

2. Limitations

2.1. First limitation

The limited collection of data in terms of both time and space is one of the work's shortcomings. Given its architecture for capturing spatial data, CNN-10 may have trouble accurately capturing the temporal instabilities in facial emotions. CNN-10 may not be able to detect these temporal fluctuations effectively, leading to less precise emotion detection since emotions are often expressed through little changes over time. Moreover, the suggested methodology imposes a limited focus on the worldwide viewpoint. The primary focus of CNN-10 receptive fields is on local spatial features. They may not correctly capture the interplay between different face regions and the overall context, which are essential for deciphering facial expressions.

2.2. Second limitation

The absence of data for some expressions is an additional issue. There is typically a restricted sample size and a wide variety of facial expression data sets. Some emotions, like disgust or disdain, might not be present in everyday life, which means there isn't enough data to create models that accurately represent these feelings. This could affect CNN-10's capacity to identify nuances or less prevalent emotions. CNN-10 is also hard to understand and interpret. Given that CNNs are often viewed as "black-box" models, understanding and clarifying their decision-making process can be challenging. Facial emotion recognition systems may require discussions or understandings of which facial areas or features connect to different emotion predictions. Due to its inherent lack of comprehensibility, CNN-10 is less useful in particular contexts, such as those requiring transparency or accountability.

3. Synthesis

Future research will focus on the selection of transfer learning-based facial expression elements. In other future research directions for this work, researchers will examine how facial emotion recognition can be used to depict people's mental health, health status, and internal wellbeing as well as how appropriate care and therapy can be facilitated to improve people's mental and general health.