Facial Emotion Detection

Using Convolutional Neural Networks

Rana Abdulsalam Mohammed

Mohammed Ahmed Saleh

Fady Emeel
dept.of Artificial Intelligence
Faculty of Computers and Artificial Intelligence - Cairo University
Cairo, Egypt

Abstract

Facial emotion recognition (FER) is a pivotal aspect of human-computer interaction, leveraging advancements in computer vision and machine learning to discern emotional states from facial expressions. This paper presents an in-depth exploration of FER methodologies, encompassing traditional techniques and state-of-the-art deep learning approaches. The review outlines the significance of FER in understanding human behavior, its applications across various domains, and the underlying challenges hindering its widespread adoption. Additionally, the abstract discusses key datasets, benchmarks, and technological advancements that have contributed to the evolution of FER models. Insights into the role of FER in affective computing, human-computer interaction, mental health assessment, and related fields are highlighted, showcasing its potential impact on society and technology. Furthermore, the paper delves into future prospects, emphasizing the need for robust, culturally diverse datasets and innovative techniques to enhance FER's accuracy, generalization, and real-world applicability.

I. Introduction

Facial emotions are very important factors of communication, and being able to tell the emotion of a person is crucial so that we can act upon it and understand the intentions of others. Thus it makes sense that the study and research of facial emotion recognition has increased, and more attention has been drawn towards this study over the past decade or so.

With the rapid development of AI even more attention has been drawn towards automatic facial emotion recognition as it is very useful in improving human computer interaction and making it more natural so that machines that interact with humans regardless of their use can read the feelings and emotions of humans and act accordingly providing the user with a better experience and better serve the purpose of the machine.

Machines have been provided with the ability to detect human emotions using deep learning models in recent years, and have been applied to applications like recommendation systems, voice assistants and mental health therapy.

The primary goal of FER is to associate different facial expressions with corresponding emotional states. The two primary phases of the traditional FER are emotion recognition and feature extraction.
The primary task in a classical FER system is feature extraction from the processed image, and the approaches now in use make use of cutting-edge techniques like linear discriminant analysis and discrete wavelet transform (DWT) . Ultimately, the features that were collected are utilized to categorize the emotions, typically with the help of neural networks (NN) and other machine learning methods.

II. Related Work

Related work on CNNs for facial emotion identification has investigated a range of methods, including multi-task learning, generative models, deep learning architectures, transfer learning, cascaded models, and attention processes. And experimented with both hand extracted features and the ones automatically obtained from the neural network as discussed before.

A. Goodfellow introduced Generative Adversarial Networks, a novel framework for training deep neural networks. GANs are composed of two parts: a discriminator network that learns to distinguish between real and synthetic data, and a generator network that generates synthetic data. GANs have been used in facial emotion recognition, where the discriminator network learns to distinguish between real and generated emotions.

B. A multi-task learning strategy for CNN-based face emotion recognition was presented by Zhang in 2017. They created a multi-task CNN architecture that can predict emotion categories and facial action units at the same time. When compared to single-task models, the model's performance on both tasks was enhanced by concurrently learning these tasks.

C. In order to overcome the difficulty of identifying subtle facial emotions, in 2019 Zhao suggested a cascaded CNN design. The cascaded model has several CNNs, each of which focuses on a distinct area of the face. They increased the recognition accuracy for subtle emotions by combining the predictions from these CNNs.

D. A deep learning framework for CNN-based face emotion identification was presented by Zhang in 2014. They unveiled the Deep Belief Network (DBN), a cutting-edge design that blends CNNs and RBMs (Restricted Boltzmann Machines). The suggested model showed how good CNNs are at recognizing face emotions by achieving cutting-edge results on a number of benchmark datasets.

E. The problem of having insufficient labeled data to train emotion detection models was the main emphasis of Liu in 2015. They suggested using transfer learning to improve the emotion recognition of a CNN model that has already been trained on a sizable image dataset. Even with a small amount of labeled data, they were able to enhance performance by utilizing the knowledge gained from the pre-trained model.

F. CNNs were being used by Li in 2020 to recognize facial emotions through a unique attention mechanism. The model is able to capture discriminative features for emotion recognition because the attention mechanism especially focuses on informative face regions. When compared to conventional CNN architectures, the suggested attention-based CNN performed better.

G. And these are just to name a few, many other research papers and benchmark tests have been done in the study and search in FER. various methods and strategies explored to raise the accuracy and efficiency of emotion identification systems.

III. Dataset

The dataset of choice in this project is going to be the FER2013 dataset.

FER2013 is a popular benchmark dataset for facial expression recognition is the. It is made up of 35,887 48x48 pixel grayscale pictures that depict the seven distinct emotions of surprise, rage, contempt, fear, happiness, sadness, and neutrality on people's faces. The dataset was created by Pierre-Luc Carrier and Aaron Courville, and it was made publicly available in2013.

Seven emotion categories—each denoting a distinct face expression—are found in the dataset. The following labels apply to these emotions: 0: Fury - 1:Repulsion - 2: Tension - 3: Joyful - 4: Sad - 5: Astonishment - 6: Neutral.

The FER2013 dataset contains solely grayscale images, which have one channel rather than the three channels (red, green, and blue) that are typically present in color photos. Every image is represented by a matrix of 48 by 48 pixels. The dataset has been split into three sets: 28,709 photos make up the training set, 3,589 images make up the public test set, and 3,589 images make up the private test set.

In terms of the distribution of classes, the FER2013 dataset is fairly balanced. It is noteworthy, nevertheless, that there are a lot less examples in the "disgust" class than in the other classifications.
The FER2013 dataset has some challenging aspects for facial recognition models. Some of these aspects are variations in lighting head poses, occlusions, and image quality.

A lot of work has gone into developing and testing facial expression recognition models using the FER2013 dataset. It has been applied to several research papers, contests, and benchmarking assignments to evaluate the effectiveness of various methods and algorithms.

IV. Proposed Model

Facial Emotion Recognition (FER) demands robust models capable of discerning nuanced facial expressions. In this study, we propose a novel Convolutional Neural Network (CNN) architecture tailored for FER tasks, alongside an incorporation of the ResNet50V2 model for comparative evaluation.
CNN Model Architecture
Our designed CNN model is structured with a sequence of convolutional layers, batch normalization, max-pooling, dropout, and dense layers, culminating in an output layer predicting emotions.

The proposed CNN architecture consists of multiple convolutional layers interspersed with batch normalization to enhance training stability and convergence. Max-pooling layers are incorporated to downsample feature maps, followed by dropout layers to mitigate overfitting. The final layers comprise densely connected neural units culminating in the output layer with softmax activation to predict seven facial emotions.

Integration of ResNet50V2
In addition to our designed CNN model, we integrate ResNet50V2, a deep residual network known for its architectural depth and skip connections, allowing for improved feature extraction and deeper representations. This serves as a comparative benchmark against our designed CNN model to assess performance variations in FER tasks.

The ResNet50V2 model, pre-trained on ImageNet, provides a strong baseline for comparison, capturing intricate features in facial expressions. Fine-tuning or additional layers can be appended to tailor ResNet50V2 for FER-specific tasks.

Model Training and Evaluation
Both the proposed CNN model and the integrated ResNet50V2 will undergo rigorous training and evaluation using standard FER datasets. The comparative analysis will encompass metrics such as accuracy, loss, and computational efficiency to discern the efficacy and performance of each model in capturing facial emotions.

V. Experimental Results

CNN Model Performance
The proposed Convolutional Neural Network (CNN) model was trained using a dedicated facial emotion recognition dataset for 20 epochs. The model's performance was evaluated on a separate test set to assess its efficacy in recognizing facial emotions.

After 20 epochs of training, the CNN model achieved an accuracy of 63% on the test set. This accuracy rate reflects the model's ability to correctly predict the emotions present in facial expressions, showcasing its competency in discerning basic emotions such as anger, disgust, fear, happiness, sadness, surprise, and neutral expressions.

And when taking a look at the resulting confusion matrix we can see that the latest numbers lie in the diagonal with some confusion between surprise and sadness classes, sadness and fear and sad and neutral. Which in some cases are different for the human to distinguish between these emotions.

ResNet50V2 Model Performance

The pre-trained ResNet50V2 model, integrated and fine-tuned for facial emotion recognition, underwent training for 20 epochs using the FER2013 dataset. The model's performance was evaluated on an independent test set to assess its proficiency in recognizing facial emotions.

After 20 epochs of training, the ResNet50V2 model exhibited a notable accuracy of 68% on the test set. This accuracy rate indicates the model's capacity to accurately identify and classify facial expressions, encompassing a spectrum of emotions including anger, disgust, fear, happiness, sadness, surprise, and neutral expressions.

The depth and complexity of the ResNet50V2 model resulted in a fairly improved accuracy compared to the CNN architecture built by us.

And when performing error analysis it is shown that the model misclassifies the same undistinguishable classes that the past model confused among.

VI. Conclusions

Facial Emotion Recognition (FER) remains a dynamic and critical domain in computer vision and affective computing. This paper delves into the methodologies, challenges, and advancements in FER, exploring traditional techniques, deep learning models, and the significance of this field across diverse applications.

FER holds immense promise in augmenting human-computer interaction, mental health assessment, and affective computing. The quest for accurate and robust FER models has led to extensive research, leveraging traditional approaches like linear discriminant analysis and discrete wavelet transform, alongside the advent of deep learning architectures and transfer learning methodologies.

The exploration of the FER landscape revealed the proliferation of Convolutional Neural Networks (CNNs) and their variants, showcasing their efficacy in discerning facial expressions. Our proposed CNN architecture and the integration of the pre-trained ResNet50V2 model were instrumental in evaluating FER performance.

The experimental results demonstrated the capabilities of the CNN model, achieving a 63% accuracy rate on the FER2013 dataset. Despite commendable performance, the model encountered challenges in differentiating subtle emotions, akin to human difficulties in discerning similar expressions.

Furthermore, the integration and fine-tuning of the ResNet50V2 model yielded superior accuracy, scoring 68% on the same dataset. The depth and architectural complexities of ResNet50V2 showcased its proficiency in capturing intricate facial features, outperforming the CNN model.

The error analysis highlighted recurring misclassifications among emotions that share visual similarities, such as surprise and sadness, indicating the inherent complexities in FER tasks.

In conclusion, while our proposed CNN model showed promising performance, the integration of the pre-trained ResNet50V2 model outperformed it, showcasing the significance of deeper architectures in FER.

The study underscores the need for continual advancements, including innovative techniques, diverse datasets, and attention to subtle emotional cues, to enhance FER's accuracy and real-world applicability.

Moving forward, fostering collaborations, leveraging ensemble methods, and exploring multimodal approaches are imperative to address the intricacies and challenges in achieving comprehensive Facial Emotion Recognition.

References

1. Aakash Saroop, Pathik Ghugare, Sashank Mathamsetty and Vaibhav Vasani ‘’ FACIAL EMOTION RECOGNITION: A MULTI-TASK APPROACH USING DEEP LEARNING”
https://arxiv.org/pdf/2110.15028.pdf

2. Mohammed Adnan Adil, ‘’ FACIAL EMOTION DETECTION USING CONVOLUTIONAL NEURAL NETWORKS”

3. M. A. H. Akhand ’ Tetsuya Shimamura, Shuvendu Roy, Nazmul Siddique ans Md Abdus Samad Kamal, ‘’FACIAL EMOTION RECOGNITION USING TRANSFER LEARNING IN THE DEEP CNN”
4. Lyons, Michael, Kamachi, Miyuki, & Gyoba, Jiro. (1998). The Japanese Female Facial Expression (JAFFE) Dataset [Data set]. Zenodo.
5. Gerard Pons, David Masip : Multi-task, multi-label and multi-domain learning with residual convolutional networks for emotion recognition. arXiv preprint arXiv:1802.06664, 2018
6. Zadeh, Milad Mohammad Taghi, Maryam Imani, and Babak Majidi. ” Fast facial emotion recognition using convolutional neural networks and Gabor filters.” 2019 5th Conference on Knowledge Based Engineering and Innovation (KBEI). IEEE, 2019.
7. Zeiler, Matthew D., and Rob Fergus. ”Visualizing and understanding convolutional networks.” In European conference on computer vision, pp. 818- 833. Springer, Cham, 2014.
8. Walecki R, Rudovic O, Pavlovic V, Schuller B, Pantic M. Deep structured learning for facial expression intensity estimation. Image Vis. Comput. 2017 Apr;259:143-54.
9. Gonzalez-Garcia, A., Tax, D. M. J., & Vera-Rodriguez, R. (2018). Facial Expression Recognition Using Residual Networks. *IEEE Transactions on Affective Computing*.
10. Gopinath, A., Sinu, G. R., & Kesavadas, M. (2018). Deep Convolutional Neural Networks for Facial Expression Recognition: A Review. Procedia Computer Science.
11. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
12. Giro-i-Nieto, X., Martinez, B., Ferràs, A., Diego, F., & Baró, X. (2017). Facial Expression Recognition with Convolutional Neural Networks: Coping with Few Data and the Training Sample Order. Pattern Recognition Letters.
13. Baró, X., Diego, F., & Farrús, M. (2017). Facial Emotion Recognition Using Convolutional Neural Networks: State of the Art.
14. Sharma, G., & Kaur, P. (2020). A Review on Deep Learning Approaches for Facial Emotion Recognition. Machine Vision and Applications
15. Raja, K. B., Ramesh, S., & Manjula, D. (2021). Facial Emotion Recognition Using CNN and LSTM Networks with Multi-Criteria Decision-Making Technique. Wireless Personal Communications
16. Ding, X., Zhao, X., Zhang, T., Liu, T., & Wu, J. (2019). Deep Learning for Facial Expression Recognition: A Comprehensive Review.
17. Liu, P., Han, S., Meng, Z., & Tong, Y. (2020). Emotion Recognition Using CNNs with Multiple Feature Fusion.