Peter the Great St. Petersburg Polytechnic University
Institute of Computer Science and Cybersecurity
Higher School of Artificial Intelligence Technologies

Facial Expression Recognition

Report on Group Project

Group: *5130203/20102*
Development team members:
Zhivotovsky Dmitry *(Project Manager)*
Butyrevskaya Galina *(ML Engineer)*
Kravchenko Sofia *(Frontend)*
Gorelkin Alexey *(Backend)*

Saint-Petersburg
2024

Contents

Problem Statement

Facial Expression Recognition: Importance and Applications

Facial expression recognition (FER) employs artificial intelligence to identify and interpret human emotions through facial cues. By analyzing features such as muscle movements, eye gaze, and overall facial structure, FER can classify emotions like happiness, sadness, anger, and fear. This technology is gaining traction due to its diverse applications across multiple fields.

In customer service, FER enhances human-computer interaction by enabling systems to assess a user's emotional state during video calls. This allows for tailored responses, ultimately leading to improved customer satisfaction and more effective conflict resolution. In the realm of mental health, therapists can leverage FER to gain deeper insight into patients' emotions, even when verbal communication may fall short. This capability can facilitate personalized treatment plans and enhance the overall therapeutic experience.

Security applications of FER are also noteworthy. By monitoring emotional responses in public spaces, security systems can identify potential threats based on sudden changes in behavior or expressions indicative of distress or aggression. This proactive approach can significantly enhance public safety in crowded venues, airports, and other sensitive areas.

The entertainment industry is increasingly adopting FER technology as well. Video games and virtual reality experiences can adapt in real-time to players' emotional reactions, creating more engaging and immersive experiences that resonate on a deeper emotional level.

Despite its numerous benefits, the implementation of facial expression recognition raises ethical considerations, particularly regarding privacy and consent. The collection and analysis of facial data necessitate stringent regulations to protect individuals from potential misuse. As technology advances, it is crucial to establish guidelines that ensure ethical practices while maximizing its positive impact.

In summary, facial expression recognition technology holds significant promise across various sectors, from enhancing customer service and mental health care to boosting security and enriching entertainment. However, navigating the ethical landscape carefully is essential to ensure that these advancements serve humanity positively and responsibly.

Methodology of Solution

This chapter presents a comprehensive examination of the methodologies employed for facial expression recognition (FER), with a particular focus on two prominent artificial intelligence models: Convolutional Neural Networks (CNN) and Support Vector Machines (SVM). These models have gained significant traction in the domains of computer vision and machine learning, owing to their efficacy in classifying and recognizing complex patterns within data.

Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNNs) represent a specialized class of deep learning models designed specifically for the processing and analysis of visual data. Their architecture is particularly well-suited for image recognition tasks, as they possess the ability to automatically learn spatial hierarchies of features from input images. The fundamental components of a CNN architecture include:

Convolutional Layers: These layers perform convolution operations on the input images using a set of learnable filters or kernels. Each filter is designed to detect specific features such as edges, textures, and other salient patterns within the image. The output of this operation is a series of feature maps that encapsulate the presence of these features.

Activation Function: Following the convolutional operation, an activation function—commonly the Rectified Linear Unit (ReLU) is applied. This function introduces non-linearity into the model, enabling it to learn complex patterns that are not linearly separable.

Pooling Layers: Pooling layers serve to reduce the spatial dimensions of the feature maps, thereby alleviating the computational burden and mitigating the risk of overfitting. Max pooling, which retains the most significant features while discarding less critical information, is a widely utilized technique in this context.

Fully Connected Layers: After a series of convolutional and pooling layers, the high-level features are flattened and forwarded through fully connected layers. These layers ultimately perform the classification task based on the learned feature representations.

CNNs demonstrate exceptional performance in FER tasks, as they can autonomously extract relevant features from raw pixel data. This capability obviates the necessity for manual feature extraction, allowing for the capture of intricate details inherent in facial expressions, thus enhancing their effectiveness in this application.

Support Vector Machines (SVM)

Support Vector Machines (SVM) constitute a class of supervised learning models utilized for both classification and regression tasks. The operational principle of SVM revolves around identifying the optimal hyperplane that effectively separates different classes within the feature space. The key components of the SVM framework include:

Hyperplane: In an n-dimensional feature space, a hyperplane is defined as a flat affine subspace that partitions the data points into distinct classes. The objective of SVM is to ascertain the hyperplane that maximizes the margin between the classes, thereby ensuring optimal separation.

Support Vectors: Support vectors are the data points that reside closest to the hyperplane and are pivotal in determining its position. The SVM model is predominantly influenced by these support vectors, rendering it robust to outliers and enhancing its generalization capabilities.

Kernel Trick: SVMs possess the capacity to manage non-linear data through the application of the kernel trick. This technique involves transforming the original feature space into a higher-dimensional space, wherein a linear hyperplane can be employed for effective separation. Commonly utilized kernel functions include linear, polynomial, and radial basis function (RBF) kernels.

SVMs exhibit considerable efficacy in FER tasks, particularly when integrated with feature extraction methodologies. For instance, following the application of CNNs for the extraction of pertinent features from facial images, SVMs can be employed as classifiers to differentiate between various emotional states based on these extracted features. This synergistic approach harnesses the strengths of both models, thereby facilitating accurate and efficient facial expression recognition.

Results

The observed accuracy rates of 33% for the Support Vector Machine (SVM) model and 64% for the Convolutional Neural Network (CNN) model can be attributed to several factors, primarily related to the quality of the dataset utilized in the study.

Firstly, the dataset comprised low-resolution black-and-white images, which inherently limits the amount of visual information available for the models to learn from. Low-resolution images often lack the necessary detail for accurate feature extraction, particularly in tasks such as facial expression recognition, where subtle nuances in facial features are critical for effective classification.

Additionally, manual verification of the dataset revealed discrepancies, with some images not accurately corresponding to their assigned classes. Such inconsistencies can severely impact the training process, as the models may learn from erroneous labels, leading to reduced performance and accuracy.

Furthermore, the constraints of time prevented the selection of a more suitable dataset, which may have provided higher-quality images and more reliable labels. The inability to curate an optimal dataset further exacerbates the challenges faced by both models, particularly the SVM, which may be more sensitive to data quality compared to the CNN.

Given these considerations, it is essential to interpret the accuracy results with caution. The accuracy rates should not be regarded as definitive measures of performance but rather as relative indicators of the effectiveness of the two models in the context of the specified dataset. The CNN's higher accuracy suggests a better capability to extract meaningful features from the available data compared to the SVM. Nonetheless, both results should be viewed in light of the dataset's limitations, emphasizing the need for further investigation with improved data quality to draw more conclusive insights.

Conclusion

In conclusion, the results obtained from the application of the Support Vector Machine (SVM) and Convolutional Neural Network (CNN) models for facial expression recognition present a rather ambiguous picture. While the CNN demonstrated a superior accuracy of 64% compared to the SVM's 33%, the overall efficacy of both models remains challenging to assess. This uncertainty is primarily due to the limitations of the dataset, which included low-resolution black-and-white images and instances of misclassified data. These factors significantly hinder the ability to make a sober and informed judgment regarding the applicability of the models to the problem of facial expression recognition. Consequently, further research utilizing higher-quality datasets and more rigorous validation methods is essential to derive more definitive conclusions about the potential of these models in accurately interpreting human emotions through facial expressions.