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)*

Contents

[Problem Statement 3](#_Toc187740629)

[Methodology of Solution 4](#_Toc187740630)

[Results 6](#_Toc187740631)

[Conclusion 8](#_Toc187740632)

# 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.

Residual Networks (ResNet)

Residual Networks (ResNet) represent a groundbreaking advancement in the architecture of deep learning models, particularly for image classification tasks. Introduced to address the challenges of training very deep networks, ResNet employs a novel approach known as "skip connections" or "residual connections." This architecture facilitates the training of networks with hundreds or even thousands of layers by allowing gradients to flow more effectively during backpropagation. The essential components of a ResNet architecture include:

*Residual Blocks*: At the core of ResNet are residual blocks, which consist of two or more convolutional layers. Each block includes a shortcut connection that bypasses one or more layers, enabling the network to learn the residual mapping rather than the original unreferenced mapping. This design mitigates the vanishing gradient problem, allowing for improved performance in deeper networks.

*Batch Normalization*: ResNet incorporates batch normalization layers, which standardize the inputs to each layer. This normalization process stabilizes the learning process and accelerates convergence, resulting in faster training times and improved overall model performance.

*Activation Function*: Typically, the Rectified Linear Unit (ReLU) is utilized as the activation function within the residual blocks. This non-linear function enhances the network's ability to capture complex patterns in the data.

ResNet has achieved remarkable success in various image recognition benchmarks, demonstrating its robustness and efficiency. By enabling the construction of deeper networks without the degradation of performance, ResNet has established itself as a foundational architecture in the field of deep learning, paving the way for further innovations in neural network design.

# Results

To summarize, on the same dataset, the CNN model achieved 83% accuracy, while the ResNet model achieved 84% accuracy. This indicates that both models are well suited for image classification tasks. Detailed accuracy data for the models is provided below.

Results for CNN:

Изображение выглядит как текст, снимок экрана, Шрифт, дизайн

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, Шрифт, График

Автоматически созданное описание

Изображение выглядит как текст, Шрифт, снимок экрана

Автоматически созданное описание

Results for ResNet:

Изображение выглядит как текст, снимок экрана, диаграмма, линия

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, диаграмма, График

Автоматически созданное описание

Изображение выглядит как текст, Шрифт, снимок экрана

Автоматически созданное описание

# Conclusion

In conclusion, the results obtained from the application of Residual Networks (ResNet) and Convolutional Neural Network (CNN) models for facial expression recognition highlight their effectiveness in this domain. The CNN model achieved an accuracy of 83%, demonstrating its capability to learn essential features from facial images. In contrast, ResNet slightly outperforms CNN with an accuracy of 84%. Both models show significant promise for applications in areas such as human-computer interaction and mental health assessment. While the accuracy difference is minimal, it illustrates the importance of model selection based on task requirements.