Emotion Detection Using Visual Data

Detecting Emotion from Facial Expressions Using CNNs

Abdullah Bilici
Undergraduate at ITU
Istanbul, Turkiye
bilicia20@itu.edu.tr

Ridha Alrubaye
Undergraduate at ITU
Istanbul, Turkiye
alrubaye21@itu.edu.tr

***Abstract*—Emotion detection is increasingly pivotal in enhancing human-computer interaction. This project aims to develop a convolutional neural network (CNN) model capable of recognizing four key emotions - anger, sadness, happiness, and shock - from visual data, including images and videos. The model was developed using Python and employed a dataset compiled from various sources: contributed images from friends, AI-generated images, and datasets available on the internet. This report details the methodology of data collection, preprocessing techniques, model training and evaluation, and the development of a real-time emotion detection application. The outcomes show the model's accuracy in identifying emotions, demonstrating its potential application in various domains such as psychological analysis, user experience enhancement, and interactive gaming.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

This document is a model and instructions for LaTeX. Please observe the conference page limits.

A. Problem Statement

Emotion detection seeks to bridge the gap between human emotional expression and computer understanding. Despite advancements in facial recognition and machine learning, accurately detecting and interpreting human emotions from visual data remains a complex challenge. This project focuses on building a model capable of identifying four distinct emotions - anger, sadness, happiness, and shock - from facial expressions in images and videos. The complexity lies in the subtle features of facial expressions and the need for a model that can generalize across diverse datasets.

B. Project Significance

The significance of this project extends to several domains, including psychological analysis in patients and children, user experience design, and interactive media. In psychological analysis, accurate emotion detection can lead to breakthroughs in understanding human behavior and mental health. It can also help monitor child emotional and mental growth. In the realm of user experience, this technology can tailor interactions based on the user's emotional state, enhancing engagement and satisfaction. Furthermore, in interactive media and gaming, emotion detection can revolutionize user interaction, creating more immersive and responsive experiences. Therefore, the successful development of an emotion detection model has far-reaching implications, setting the stage for more intuitive human-computer interactions.

II. Related Work

The field of emotion detection using visual data, particularly facial expressions, has seen substantial research, using deep learning and specifically Convolutional Neural Networks (CNNs). This report gets inspiration from existing studies, including the use of facial recognition and emotion detection for enhancing online learning experiences. The work by (Sarmah et al.) is particularly relevant, where they proposed a framework utilizing CNN models for emotion detection, demonstrating the efficiency of deep learning in interpreting complex facial expressions.

Despite the advancements in emotion detection, there remains a gap in applying these technologies in real-world, diverse settings, particularly with limited and varying datasets. This project aims to address this gap by developing a CNN model that is robust across different data sources, including self-collected data, AI-generated images, and internet-sourced datasets. The goal is to enhance the model's applicability and effectiveness in real-time emotion detection, a relatively less explored area in emotion recognition research.

III. Methodology

A. Data Collection and Preprocessing

1) Data Collection and Sources: he project's dataset was compiled to include a broad spectrum of emotional expressions, totaling 238 images. This dataset was collected through three main methods:

- **Friends and Personal Contributions:** A big part of the dataset comprised images obtained from personal networks, including friends and family. This part of the collection involved 20 individuals, each contributing 4 images corresponding to the emotions of anger, sadness, happiness, and shock, amounting to 80 images.
- **AI-Generated Imagery:** 80 AI-generated images were included. These images, split evenly across the four target emotions, were created using advanced AI tools like DALL-E, ensuring a realistic portrayal of emotions with an equal representation of genders.

Fig. 1. Sample AI-generated image (Happy)

- **Online Datasets:** 78 images were collected from various online sources. This addition aimed to introduce more variability and real-world scenarios into the dataset.

2) Data Preprocessing: The preprocessing of the collected data was a critical step to ensure consistency and optimize the model's learning process:

- **Image Cropping and Resizing:** Each image was cropped to focus solely on the face, eliminating background distractions. The images were then resized to a uniform dimension of 128x128 pixels. This size was chosen to balance between minimizing data loss and maintaining computational efficiency.
- **Monochromatic Conversion:** To reduce complexity and focus on facial features rather than color information, the images were converted to grayscale. This step helps the model concentrate on structural and expression-related aspects of the faces.
- **Normalization:** Given the disparate sources of the images, normalization was essential to standardize the lighting conditions and camera characteristics across the dataset. This process aids in reducing variances that are not relevant to emotion detection.
- **Data Augmentation:** To address the challenge of a limited dataset and enhance the model's ability to generalize, various data augmentation techniques were applied. These included rotations, translations, Gaussian blur, and minor adjustments in contrast and brightness. Notably, normalization was performed post-augmentation to ensure consistency in the final dataset.

B. Model Architecture

The core of the project is a convolutional neural network designed for emotion detection. This model architecture is defined as follows:

- **Convolutional Layers:** The model begins with two convolutional layers, each followed by batch normalization and a ReLU activation function. These layers are designed to extract and learn features from the facial images. The first convolutional layer has 8 filters, and the second has 16 filters, both with a kernel size of 5x5 and padding of 2.

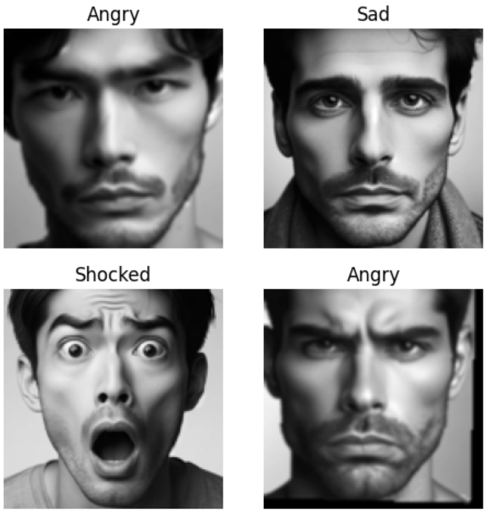

Fig. 2. Images after preprocessing and data augmentation

- **Pooling Layers:** Following each convolutional layer is a max pooling layer with a 2x2 window and stride of 2. These layers reduce the spatial dimensions of the feature maps, helping to decrease computational load and mitigate overfitting.
- **Fully Connected Layers:** The network has two fully connected layers. The first, with an input size calculated based on the output from the preceding layers, has 64 neurons. It is followed by a dropout layer with a dropout rate of 0.35 to further prevent overfitting. The final layer corresponds to the number of emotions (4 in this case) to be classified.

C. Training Process

The training of the model was executed with the following key characteristics:

- **Dataset and Batching:** The training utilized a custom DataLoader, handling the preprocessed dataset with batch size settings and transformation applications. The DataLoader ensured efficient feeding of data into the model during training.
- **Epochs and Optimization:** The model was trained over 200 epochs, using the Adam optimizer with a learning rate of 3e-5. This choice of optimizer and learning rate was aimed at achieving a balance between efficient convergence and avoiding local minima.
- **Loss Function:** The CrossEntropyLoss function was employed, a standard choice for multi-class classification problems. It combines LogSoftmax and NLLLoss in one single class.
- **Performance Tracking:** Throughout the training process, the model's performance was continuously monitored, tracking the training and testing losses. This monitoring was vital for understanding the model's learning progress and making necessary adjustments.

D. Model Evaluation

The evaluation of the model involved both quantitative and qualitative assessments:

- **Accuracy and Loss Metrics:** The primary quantitative metrics were accuracy and loss. Accuracy provided a straightforward measure of the model's performance across the four emotions. The loss, specifically the test loss, offered insights into the model's effectiveness in generalizing beyond the training data.
- **Confusion Matrix:** A confusion matrix was generated to give a detailed view of the model's performance across different emotions. This matrix helped identify which emotions were most accurately recognized and which ones posed challenges, indicating potential areas for model improvement.
- **Real-Time Testing:** A critical aspect of the evaluation involved real-time emotion detection tests using a device camera. This practical test allowed for the assessment of the model's applicability in real-world scenarios, observing its responsiveness and accuracy in a dynamic environment.

IV. Results

- **Accuracy and Loss:** The model achieved a significant level of accuracy in identifying the four emotions. The best accuracy was around 92%.The training and testing loss metrics showed a consistent decrease over the 200 epochs, indicating effective learning and adaptation by the model. The final test loss reached an impressively low value, suggesting good generalization capabilities.

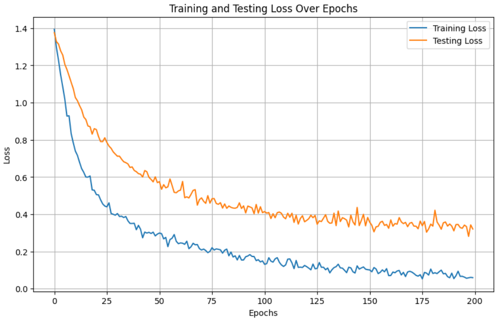

Fig. 3. Loss against Epoch

- **Confusion Matrix Analysis:** The confusion matrix revealed insightful patterns. The model showed high accuracy in detecting emotions like happiness and shock but encountered some challenges with more nuanced emotions like sadness and anger. This variation in performance across different emotions highlighted areas where the model could be further fine-tuned.
- **Real-Time Detection:** In real-time tests, the model demonstrated a promising capability to detect emotions from live camera feeds. While the response time and accuracy were very good, occasional misclassifications indicated room for improvement in real-world application scenarios.

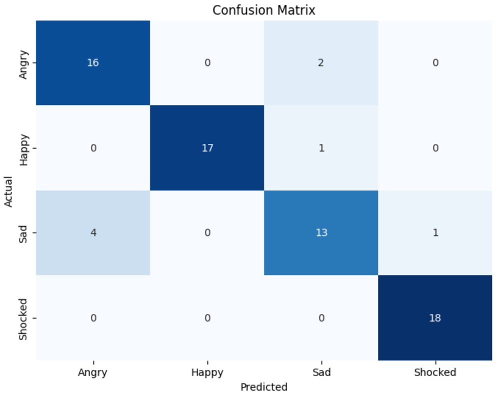

Fig. 4. Confusion Matrix

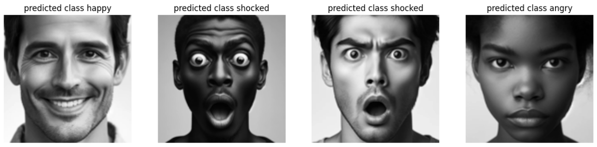

Fig. 5. Sample Predicted Results

V. Discussion

A. Training and Testing Loss

The graph showing training and testing loss over 200 epochs shows effective learning. The sharp decline in both training and test loss during the initial epochs suggests that the model quickly found the fundamental features necessary for emotion classification. As epochs progress, the training loss continues to decrease steadily, indicating ongoing learning and model refinement. The testing loss, while showing slight fluctuations, exhibits a general downward trend and stabilizes, which indicates good model generalization without significant overfitting.

B. Confusion Matrix

The confusion matrix provides a view of the model's predictive capabilities across the four emotional states:

- **Angry:** The model has high predictive accuracy for the ”Angry” emotion, with only two instances being confused with ”Sad”.
- **Happy:** ”Happy” is the emotion the model identifies most accurately, with only one instance being misclassified as ”Sad”.
- **Sad:** While the model performs well in identifying ”Sad”, there are instances where it has been confused with ”Angry” and ”Shocked”, suggesting a resemblance
- **Shocked:** This emotion is perfectly predicted with no misclassifications, which might suggest distinctive features that are well captured by the model, or it could

be an indication of a less varied representation of this emotion in the dataset.

The few instances of misclassification between ”Angry” and ”Sad” could suggest that the subtler distinctions between these emotions are not as easily captured by the model, potentially due to similarities in facial expression features

VI. Conclusion

In conclusion, the project succeeded in developing a convolutional neural network capable of detecting emotions from facial expressions. The model, through rigorous training and evaluation, has demonstrated the viability of using CNNs for emotion recognition tasks. The study revealed the model’s strengths in accuracy and its potential for real-time emotion detection, highlighting the transformative possibilities in fields such as interactive media, mental health, and user interface design. However, the journey does not end here. Future research will aim to expand the dataset, enhance the model’s accuracy, and refine its real-time detection capabilities, thus further bridging the gap between human emotions and machine understanding. This project serves as a stepping stone for deeper exploration, paving the way for more intuitive and responsive technological solutions.

References

[1] Priyanshu Sarmah, Rupam Das, Sachit Dhamija, Saurabh Bilgaiyan, Bhabani Shankar Prasad Mishra, ”Facial identification expression-based attendance monitoring and emotion detection—A deep CNN approach”, https://www.sciencedirect.com/science/article/pii/B9780323852098000018