

Final Report - Facial Expressions classification

Moshe Ofer
Talor Langnas
Avichai Mizrachi

March 11, 2024

Abstract

Emotion recognition technology has garnered significant attention owing to its diverse applications in understanding human behavior, enhancing user experience, and facilitating personalized interactions in fields such as security, and human-computer interaction. In this report, we explore the implementation of facial expression recognition using convolutional neural networks (CNNs). We present our approach, detailing the dataset used, training process, model architecture, and performance evaluation. Through experiments, we demonstrate the effectiveness of our CNN-based approach for emotion classification in facial images. Our findings indicate promising results, with the model achieving a test accuracy of 0.61 on a diverse dataset. Visual comparisons between predicted and actual labels provide insights into the model's performance and areas for improvement. Overall, this report contributes to the ongoing research and development of facial recognition systems using deep learning techniques.

1 Introduction

With the increasing demand for efficient and accurate facial recognition systems, researchers are constantly exploring new methods and techniques to improve the performance of such systems.

In this report, we present our approach to facial recognition using deep learning, along with the experiments conducted to evaluate its performance. We compare the results obtained with our approach to those achieved using logistic regression, and MLP, providing insights into the effectiveness of deep learning for this task.

2 Required Background in Machine Learning Concepts

- **Supervised Learning:** Supervised learning is a type of machine learning where the model is trained on a labeled dataset, meaning that each example in the dataset is associated with a corresponding label or output. During training, the model learns to map input data to the correct output labels by minimizing a loss function that measures the difference between predicted and true labels.
- **Convolutional Neural Networks (CNNs):** Convolutional neural networks are a class of deep learning models specifically designed for processing structured grid data such as images. CNNs are characterized by their use of convolutional layers, which apply filters to input data to extract features hierarchically, and pooling layers, which downsample feature maps to reduce dimensionality. CNNs have demonstrated state-of-the-art performance in various computer vision tasks, including image classification, object detection, and facial recognition.
- **Loss Functions and Optimization:** Loss functions measure the difference between predicted and true labels and serve as the objective function that the model aims to minimize during training. Common loss functions for classification tasks include categorical cross-entropy and binary cross-entropy. Optimization algorithms such as stochastic gradient descent (SGD) and Adam are used to iteratively update the parameters of the model to minimize the loss function.

- **Data Augmentation:** Data augmentation is a technique used to artificially increase the size and diversity of a training dataset by applying transformations to input data. Common data augmentation techniques for image data include random rotations, scaling, and flipping.

3 Project Description

3.1 Dataset Description

In this project, we utilized a dataset comprising images of cropped faces representing six distinct face expression: Happy, Angry, Sad, Neutral, Surprise, and Ahegao. Each image is in RGB format and is tagged by a corresponding label indicating the true expression. We obtained the dataset from kaggle.com as recommended.

The dataset provides a diverse range of facial expressions, allowing our model to learn and distinguish between various emotional states.

3.2 Training Process

We trained the CNN model over 10 epochs using a dataset comprising images of facial expressions. For the training process, we sampled 15,454 images and ran the training for 10 epochs. Each image was normalized to a size of 222×222 pixels.

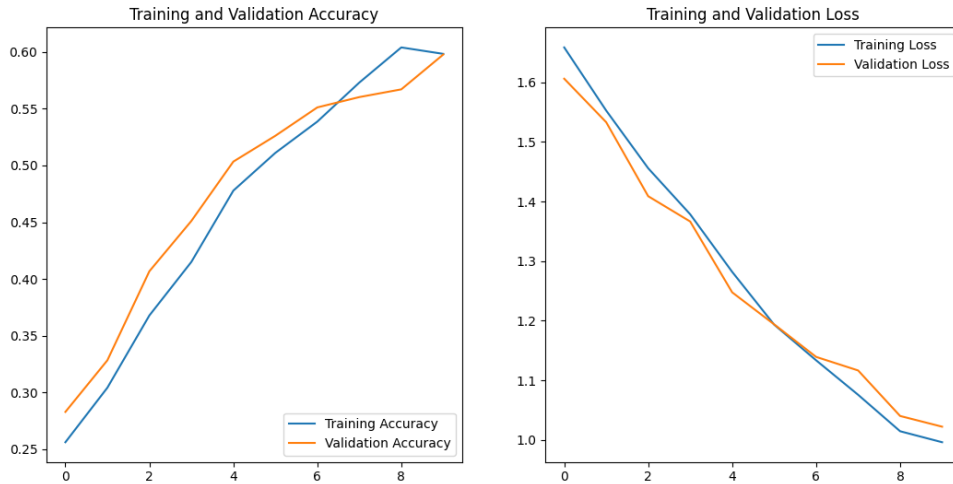


Figure 1: Training and validation accuracy and loss trends across 10 epochs

To augment the dataset and improve the model's robustness, we employed data augmentation techniques using the Keras library. These augmentation techniques help introduce variations in the training data, such as horizontal flipping, random rotation (up to 10 degrees), and random zooming (up to 10 presents), thereby enhancing the model's ability to generalize to unseen data.

Additionally, to ensure unbiased evaluation of the model's performance, we divided the dataset into training, validation, and test sets. We allocated 70% of the data for training, 15% for testing, and 15% for validation. This division strategy helps prevent overfitting and provides a reliable assessment of the model's generalization capabilities.

The training process involved optimizing the categorical cross-entropy loss function using the Adam optimizer. The training and validation accuracies and losses across epochs are summarized in Figure 1.

3.3 Model Architecture

Our CNN model architecture is outlined as follows:

Model: "faceRecoModel"

Layer	Output Shape	# of Param
sequential	(None, 222, 222, 3)	0
conv2d	(None, 222, 222, 32)	896
max_pooling2d	(None, 111, 111, 32)	0
conv2d_1	(None, 109, 109, 64)	18496
max_pooling2d_1	(None, 54, 54, 64)	0
conv2d_2	(None, 52, 52, 64)	36928
max_pooling2d_2	(None, 26, 26, 64)	0
conv2d_3	(None, 24, 24, 64)	36928
max_pooling2d_3	(None, 12, 12, 64)	0
flatten	(None, 9216)	0
dense	(None, 128)	1179776
dense_1	(None, 64)	8256
dense_2	(None, 6)	390
Total params: 1281670		

3.4 Performance Evaluation

Following training, the model was evaluated on a separate test dataset, yielding a test accuracy of 0.61. When we compared our model to a simple logistic regression and a simple multi-layer perception (MLP) model, we got test accuracy of 0.36, and 0.52 respectively.

3.5 Predicted vs. Actual Labels

Figure 2 presents a visual comparison of the predicted and actual labels for a selection of images from the test dataset.

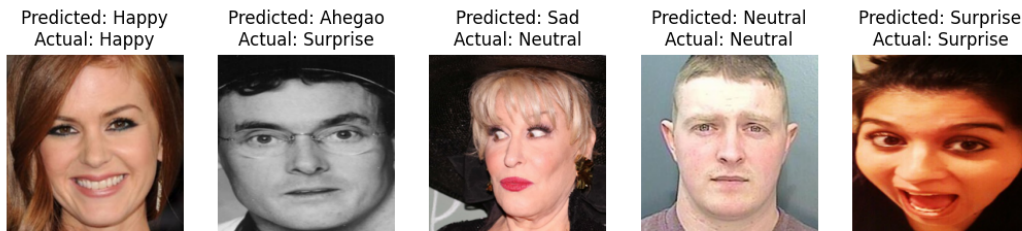


Figure 2: Comparison of Predicted and Actual Labels for Selected Images

The visual comparison allows for a qualitative assessment of the model's performance in accurately classifying facial expressions. Discrepancies between the predicted and actual labels may indicate areas for improvement in the model or highlight challenging instances where emotion recognition is inherently ambiguous. We assume that the model may have difficulty with gray images (2) and those that aren't taken from a straight angle (3).

As we can expect, just after the softmax function, higher values in the arrays in figure 3 correspond to confident predictions, while lower values indicate uncertainty. Close values are observed for uncertain predictions.

	Ahegao	Angry	Happy	Neutral	Sad	Surprise
0	0.00047	0.05059	0.01495	0.42261	0.50487	0.00652
1	0.04216	0.31112	0.05846	0.22826	0.25067	0.10933
2	0.01323	0.01084	0.18796	0.00410	0.31035	0.47353
3	0.00357	0.30980	0.11825	0.26611	0.26304	0.03922
4	0.00351	0.05106	0.20858	0.50650	0.20793	0.02243
5	0.72401	0.07257	0.02762	0.02456	0.06496	0.08628






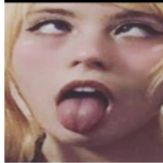
Predicted: Sad Actual: Neutral	Predicted: Angry Actual: Angry	Predicted: Surprise Actual: Happy	Predicted: Angry Actual: Neutral	Predicted: Neutral Actual: Neutral	Predicted: Ahegao Actual: Ahegao
					
0	1	2	3	4	5

Figure 3: Comparison of Predicted and Actual Labels for Selected Images Alongside Model Prediction Arrays Before Argmax just after Softmax

4 Previous Attempts

In our initial attempts, we adopted a simpler architecture for our facial recognition model, aiming to minimize computational complexity and training time. This initial architecture consisted of fewer convolutional layers and parameters, reflecting a more straightforward approach to the task.

However, the performance of the simplified model was less than our expectations. It struggled to capture the nuances present in facial expressions, leading to suboptimal accuracy and limited generalization capabilities. Recognizing the need for a more sophisticated and complex model, we increased the complexity of the architecture. We introduced additional convolutional layers, leveraging deeper networks.

Simultaneously, we expanded our dataset and introduced more diverse samples to improve the model’s ability to generalize to unseen data. By augmenting the dataset, we achieved that. Despite these improvements, we encountered overfitting issues, where the model performed well on the training data but poorly on unseen data. To overcome this challenge, we adjusted our training strategy by limiting the number of epochs to 10 and increasing the size of the training dataset. This approach helped overcome overfitting and improved the model’s performance on unseen data. As a result of these adjustments, we observed significant enhancements in the model’s performance, with higher accuracy and improved generalization capabilities.

5 Conclusions

In conclusion, our work on facial recognition using neural networks highlights the importance of model complexity and dataset quality in achieving accurate results. By iteratively refining our approach and addressing challenges such as overfitting, we have demonstrated significant improvements in model performance. Our findings underscore the potential of deep learning techniques in advancing facial recognition technology.

6 Future Work

Facial recognition using deep learning has shown promising results, but there are several avenues for future exploration and improvement:

- **Enhanced Data Augmentation:** Further investigation into advanced data augmentation techniques could improve the model’s ability to generalize across diverse facial expressions and envi-

ronmental conditions. Techniques such as elastic deformations, color jittering, and style transfer could be explored. All of this to make our dataset be bigger and more diverse.

- **Architecture Optimization:** Experimenting with different CNN architectures, may lead to improvements in accuracy and efficiency. Additionally, exploring more techniques could enhance the model's ability to capture subtle facial features.