Solution Summary Report

Image Classification Model Development

Problem Statement:

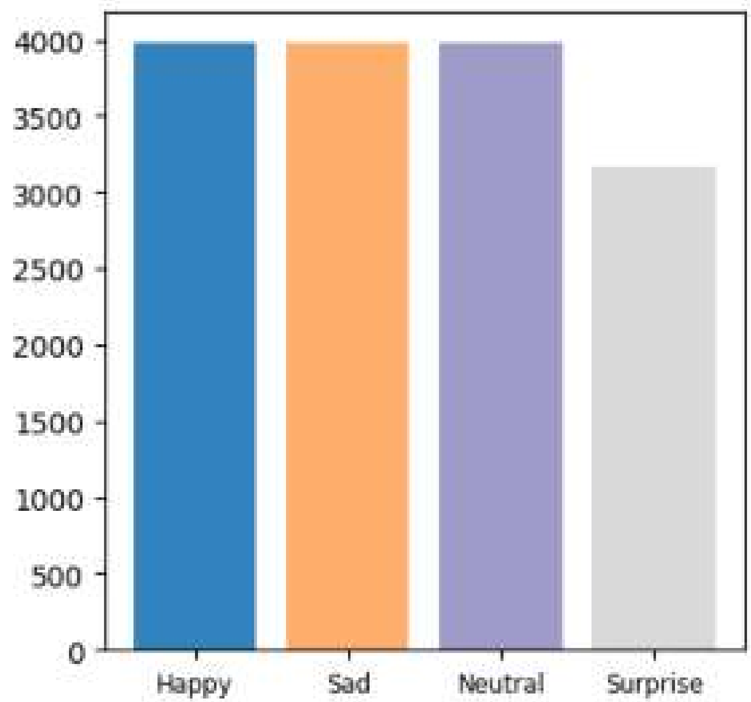
The project we face, Facial emotion recognition with four categories (Happy, Sad, Neutral, Surprise) presents specific challenges. Recognizing not only the face in a picture, but also its specific emotion can be a challenging task. Depp Learning ca definitely help in this regard.

There is definitely a data imbalance in the 'Surprise' category having only 3200 images compared to 4000 in others and can lead to model bias, reduced accuracy for less represented emotions.

The model presented here will accurately distinguishing subtle features that differentiate emotions, especially between similar ones like Sad and Neutral. The challenge here is misclassification, especially from those picture which are not taken from the front of the face, but from the side or 3/4 view. Similarly, the model should perform consistently across different ages, ethnicities, and lighting conditions ensuring consistency even where features can be different.

Another challenge is that emotional complexity: Accounting for the complexity and subjective nature of human emotions. Not all people smile the same way, or the category sad and neutral can be confused with each other. I specifically talk about the facial features which can be confusing and probably even interchangeable between categories at times. This can lead to misclassification or the models can overfit easily.

Data description



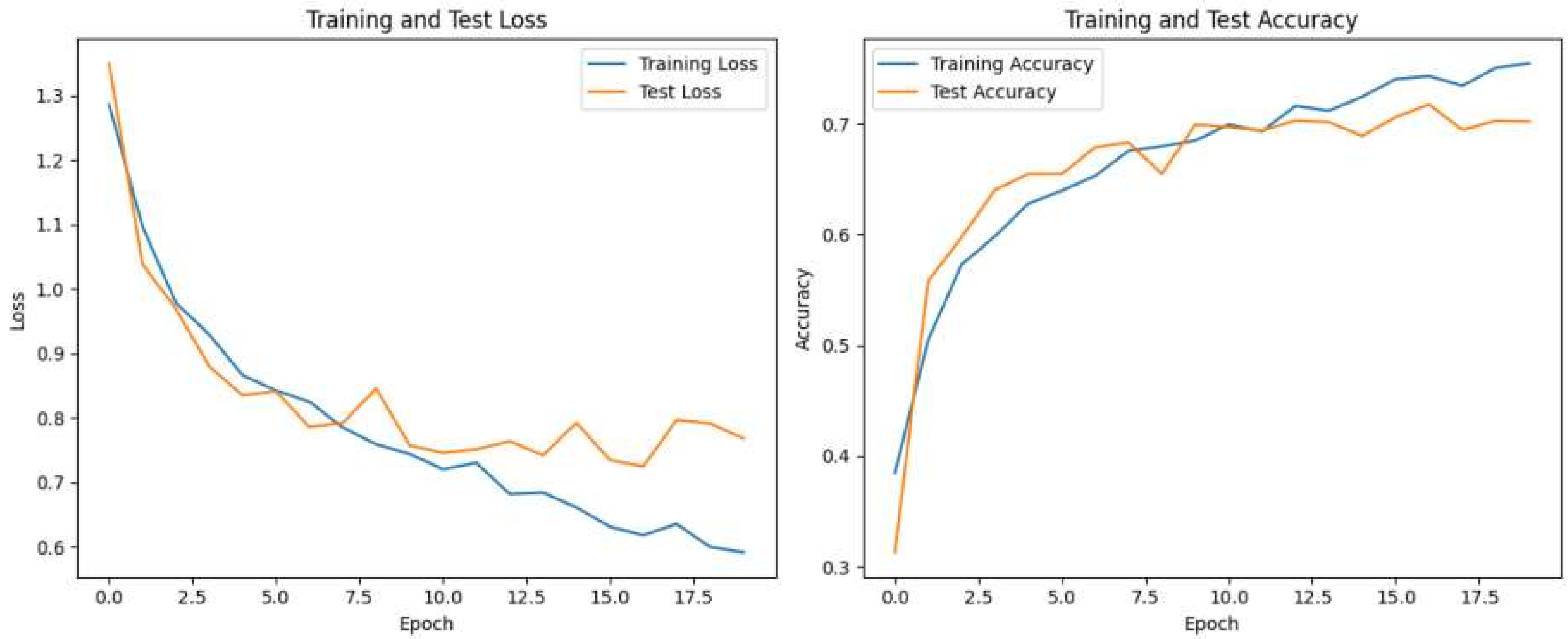
We were presented with 4 categories of face emotions: Happy, Sad, Neutral, Surprise. The amount of images in each category was about 4000 except the surprise category with only had about 3200 images in it. The images were in black and white (grayscale), and the sets of images for training and test purposes were randomly distributed. The imbalance in the surprise category will need to be taken into consideration.

Methodology:

In the "Model Exploration" phase, the approach was methodical and thorough, utilizing a blend of Convolutional Neural Networks (CNNs) and Transfer Learning models. This diverse methodology enabled a comprehensive understanding of model efficiencies in facial emotion recognition. Initial tests with Simple CNNs in both RGB and grayscale modes yielded accuracies of 64% and 75%, respectively, indicating the influence of color information suggesting that color models might not perform well when grayscale images are supplied to the model.

Base First CNN (RGB) model with 65% accuracy

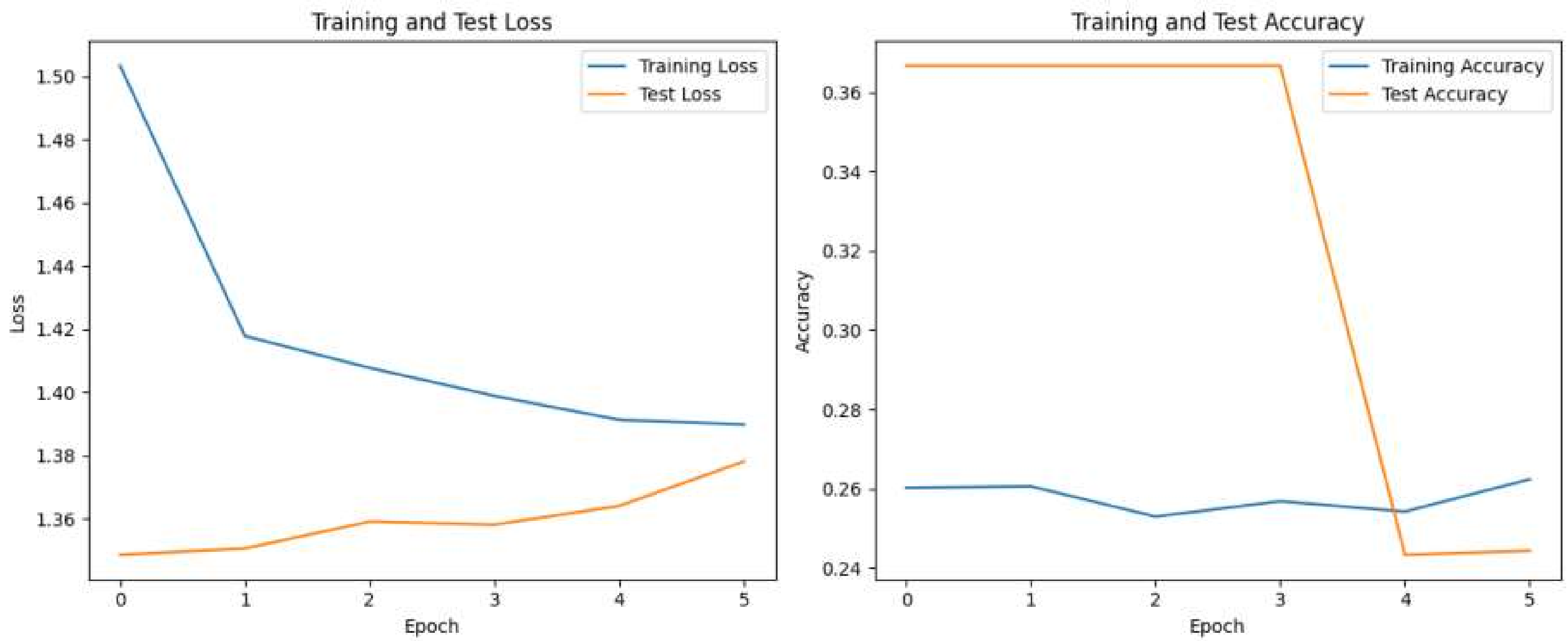
Second CNN (GRAYSCALE) model with 75% accuracy



It appears that the Grayscale model has higher accuracy.

Advanced Transfer Learning models like VGG16, ResNetV2, and EfficientNet were also evaluated, with their accuracies (49%, 61%, 67%) highlighting the varying effectiveness of different architectures in handling the nuanced task of emotion classification.

Accuracy for ResNetV2 (61%) and EfficientNet (67%). VGG16 performed at 49%


Training and Test Loss
Training Loss
Test Loss
Loss
Epoch
Training and Test Accuracy
Training Accuracy
Test Accuracy
Accuracy
Epoch

Models like VGG16, ResNetV2, and EfficientNet performed poorly with grayscale images due to their design and pre-training on color datasets. These models are typically pre-trained on large datasets like ImageNet, which are predominantly color (RGB) images. This pre-training embeds an understanding of features that are inherently dependent on color data. When supplied with grayscale images, the models lose the color information, which is a crucial feature set for these models to detect and differentiate various elements. The absence of color data can lead to a significant reduction in their ability to accurately classify or recognize patterns as effectively as they would with color images. The clues of the final model came from this realization: A well constructed Convolutional Neural Network will be the best solution for grayscale images. A custom CNN can be optimized from the ground up to focus on the patterns and features present in grayscale imagery, potentially leading to better performance for this specific task.

Final Model Architecture

This model consists of three sets of convolutional layers followed by max-pooling and batch normalization. The flattened output is then passed through a dense layer before the final output layer for classification.

First line adds a **2D convolutional layer** with 32 filters, each of size (3,3), using the ReLU activation function. The input shape is set to (48,48,1), indicating a grayscale image with dimensions 48x48 pixels.

In the next line a **2D max-pooling layer** is added with a pool size of (2,2), reducing the spatial dimensions of the previous layer by half.

A **batch normalization layer** is added, which normalizes the activations of the previous layer to improve training stability.

Follows the **same set 32 filters**

Follows another **layer with 128 filters**.

So far the activation function was ReLU.

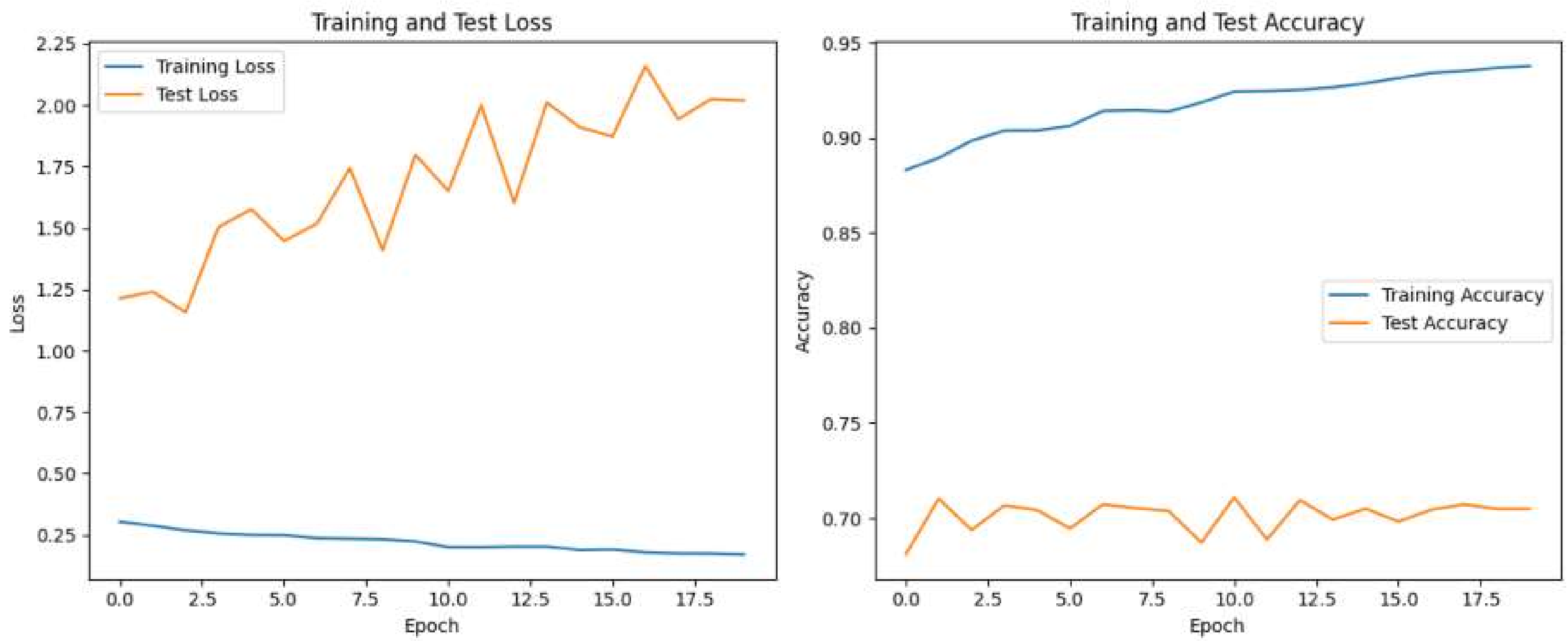
Then a new layer is added, **flatten layer**, converting the 3D output to a 1D vector, preparing it for the fully connected layers.

Next a **fully connected (dense) layer** is added with 128 neurons and a ReLU activation function.

And another **fully connected layer is added with 4 fully connected neurons** The **final dense layer** a softmax activation function is suitable for multi-class classification.

It is a relatively simple model but it has proven to be the most effective with high Accuracy.

Performance of the final model:



The result of this model is much better. It achieved over 90% accuracy, a significant improvement over initial models. The iterative process led to an optimized CNN model with high accuracy.

Future work could explore data augmentation for the underrepresented Surprise category to further enhance model robustness.

Conclusions:

The iterative exploration of various models, including transfer learning, that did not achieve optimal performance, culminated in the development of a highly accurate custom CNN model. This journey underscores the criticality of tailoring model architectures to specific dataset characteristics, particularly when faced with unique challenges like data imbalances in the Surprise category. The eventual success of the custom CNN over pre-trained models illustrates the potential superiority of bespoke solutions in certain contexts, emphasizing the importance of custom design and adaptation in the field of deep learning. This process has highlighted the nuanced understanding required in model selection and the pivotal role of dataset-specific characteristics in determining model efficacy, particularly in the realm of facial emotion recognition where subtle differences in expressions can be significant. The experience gained from the less successful models provided invaluable insights, reinforcing the concept that in machine learning, the path to the optimal solution often involves navigating through a series of iterative improvements and adaptations by trial and error.

Future Recommendations:

Future recommendations involve a comprehensive approach to enhancing the model's robustness and applicability. There's a potential to further refine the model through advanced techniques such as data augmentation, especially for underrepresented categories like 'Surprise' which was underrepresented. This could involve generating synthetic data or modifying existing data to create a more balanced dataset. Additionally, exploring new convolutional architectures and implementing more rigorous validation methods could ensure the model's effectiveness across diverse scenarios. Ethical and privacy considerations are paramount, especially in applications involving human subjects, necessitating a careful and responsible approach to model deployment in real-world situations. These steps not only aim to improve the model's accuracy but also its ethical and practical applicability.