Image Classification

1. Introduction

This report focuses on image recognition, specifically classifying images based on the experts who handled them. Our objective is to categorize the images into five classes: **expA, expB, expC, expD, and expE**.The outcomes showcased in this report are produced using Python scripts within the PyTorch framework.

2. Dataset

The dataset comprises 25,000 samples, with 20,000 samples allocated to the training set and 5,000 to the test set. All images have a resolution of 256 x 256 pixels and are in RGB color space.

3. Preprocessing

PyTorch was utilized for conducting data augmentation tasks including image resizing, flips, and normalization.

4. Training

For the training phase, we employed various strategies including fine-tuning, feature extraction, and training from scratch. Specifically, we utilized ResNet for training the models, implementing the aforementioned strategies. The models (ResNet 18, ResNet 34, and ResNet 50) were trained for up to 40 epochs, starting with a learning rate of **0.01** that decayed over time (every 10 epochs). ResNet 50 exhibited the highest training accuracy, reaching up to 65%, although it displayed signs of overfitting, as we will discuss in the model evaluation section. Generally, beyond 20 epochs, most training models showed no significant improvement in accuracy, as illustrated in **Fig. 1**.

Resnet50 Fine Tuning

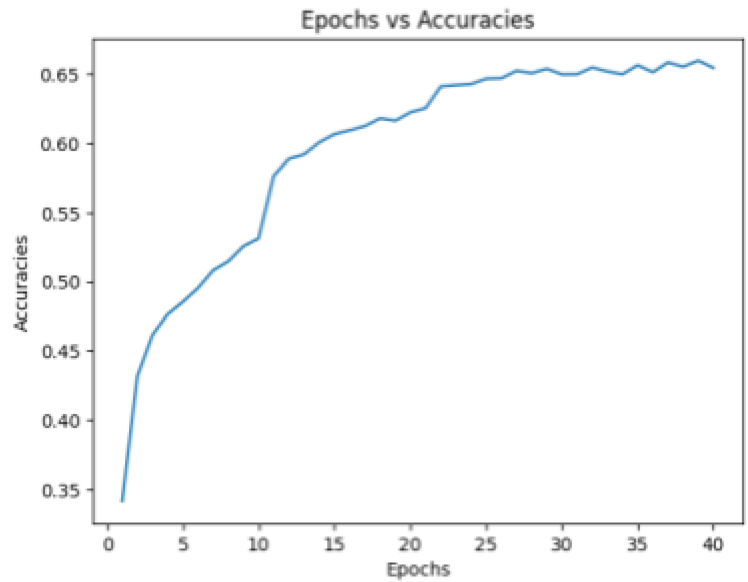

Fig. 1. Accuracy Graph depicting the Model Trained on ResNet50 with Fine Tuning across Epochs

5. Evaluation

The evaluation of the models is illustrated in the table provided below. It is evident from the results that fine-tuning ResNet 50 exhibits clear signs of overfitting, whereas training ResNet 50 from scratch emerges as the most promising model.

	Fine Tuning	Feature Extractor	Scratch
ResNet 18	Best Training Acc: 58.4% Test Acc: 42%	Best Training Acc: 34.91% Test Acc: 26%	Best Training Acc: 45.3% Test Acc: 40%
ResNet 34	Best Training Acc: 60.9% Test Acc: 40%	Best Training Acc: 35.7% Test Acc: 34%	Best Training Acc: 45.49% Test Acc: 44%
ResNet 50	Best Training Acc: 65.99% Test Acc: 44%	Best Training Acc: 39.8% Test Acc: 40%	Best Training Acc: 43.8% Test Acc: 48%

Table 1. Displays various ResNet models trained through fine-tuning, feature extraction, and training from scratch, along with their corresponding top training accuracies and test accuracies.

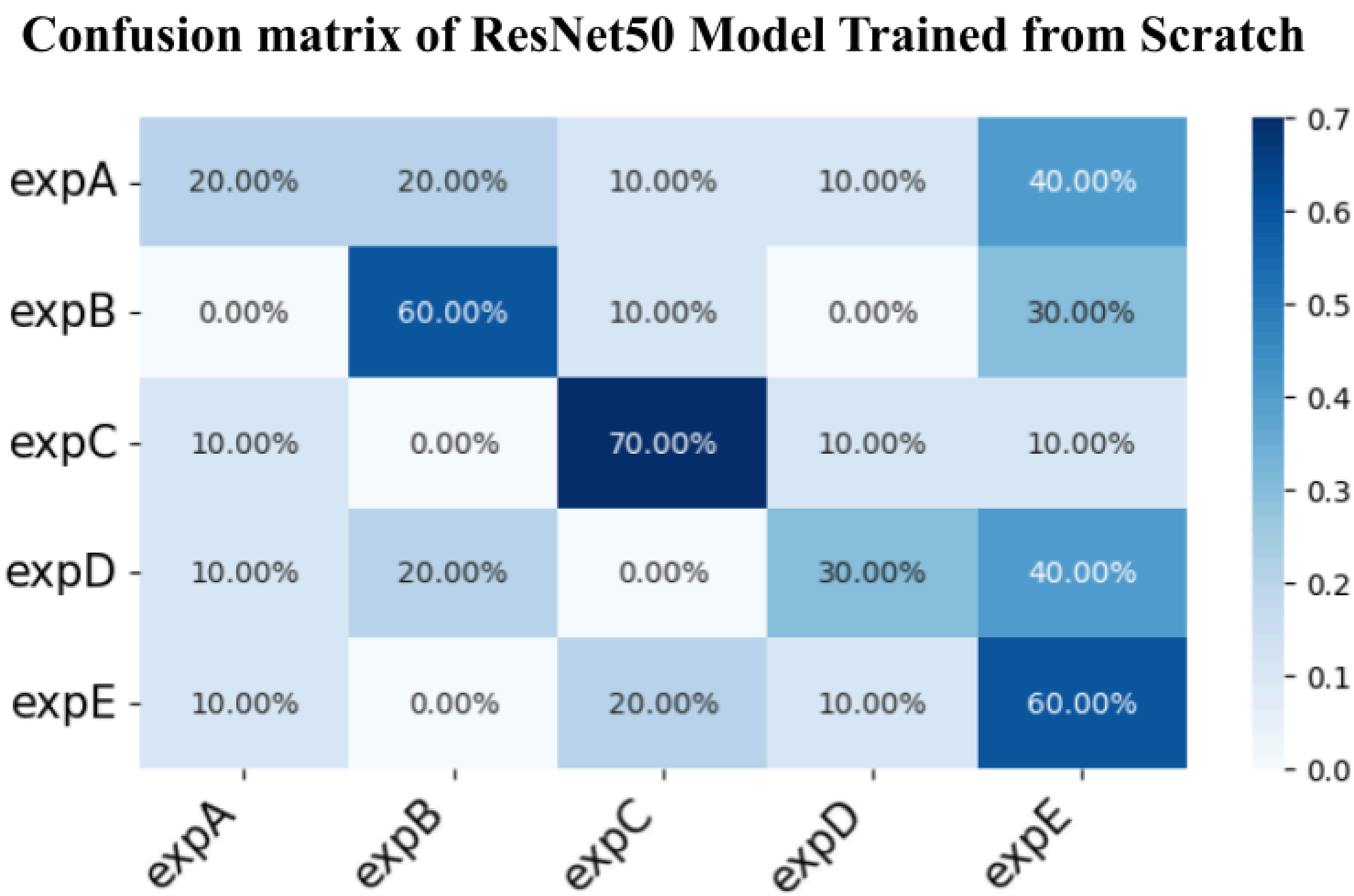

Fig. 2. The confusion matrix illustrates a high classification rate for the "expC" class and a low classification rate for "expA".

6. **Best Model With Variations**

We used our Best Model ResNet 50 trained from scratch, for the further exploratory analysis. We performed two experiments.

1. **Initial Learning Rate (0.5)**

We trained the model with an initial learning rate **0.5** that decays after a certain number of epochs. This time we trained this model till **80** numbers of epochs. But in this case the best training accuracy was **37%** with a test accuracy around 32% as shown in **Fig. 3.**

Fig. 3. Accuracy Graph depicting the Model Trained from scratch using ResNet50 with starting learning rate of 0.5

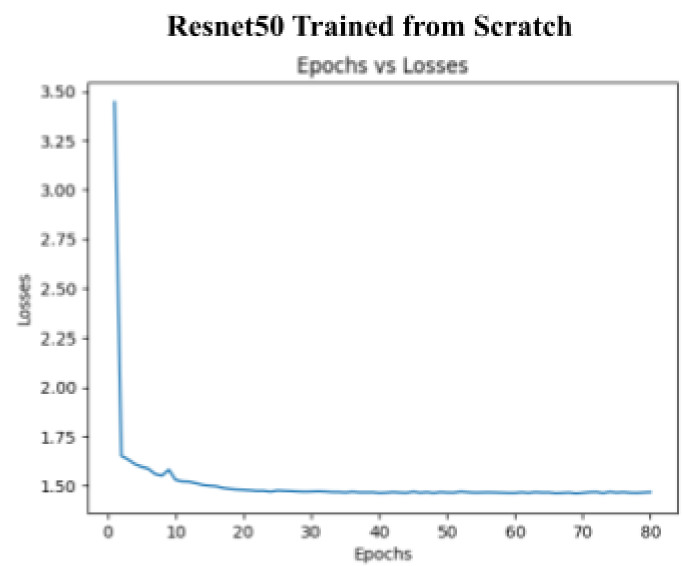

Fig. 4. Loss Graph depicting the Model Trained from scratch using ResNet50 with starting learning rate of 0.5

2. **Initial Learning Rate (0.01)**
This time we used the learning rate as we initially used for our experiment specifically for this model. But we increased the number of epochs from **40 to 80** in order to see if there is any further improvement in the model. The best training accuracy stayed the same **(44%)** as shown in **Fig. 5.** while the test accuracy dropped a little from **48% to 46%.**

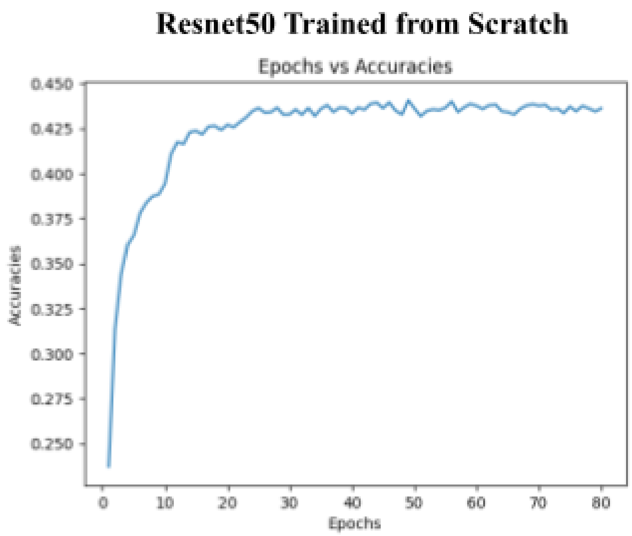

Fig. 5. Accuracy Graph depicting the Model Trained from scratch using ResNet50 with starting learning rate of 0.01

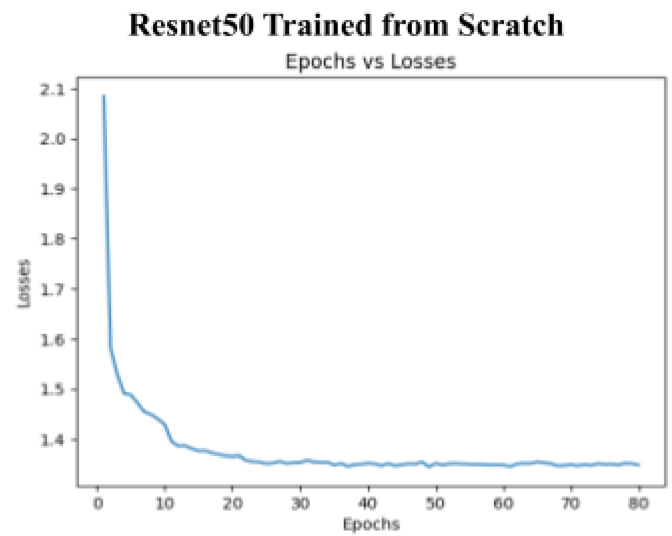

Fig. 6. Loss Graph depicting the Model Trained from scratch using ResNet50 with starting learning rate of 0.01

7. **Model Trained on the Concatenated 6-Channel Image**
We combined the touched (3-channel) images with raw (3-channel) images by merging their channels/depth, resulting in a 6-channel image. This combined image was then used as input for training the **ResNet50** Model from scratch, starting with an initial learning rate of **0.01**, which decayed over time.Notably, there was a substantial improvement in both training and

test accuracy. The final test accuracy reached **57.16%**, while the training accuracy achieved **57.32%**. Below are the training and loss curves associated with this model.

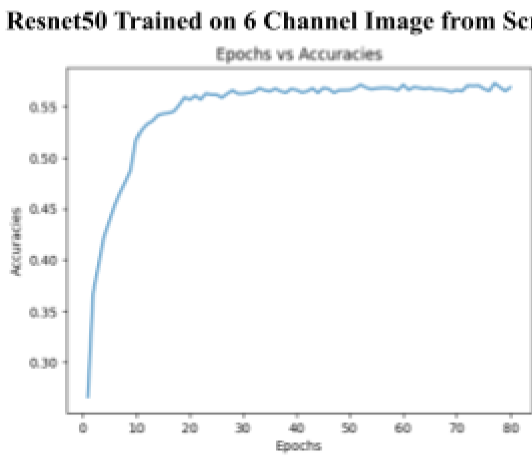

Fig. 7. Accuracy Graph depicting the Model Trained on 6 channel image from scratch using ResNet50 with starting learning rate of 0.01

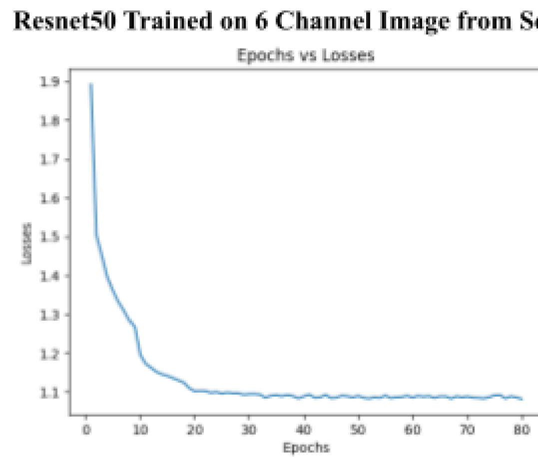

Fig. 8. Loss Graph depicting the Model Trained on 6 channel image from scratch using ResNet50 with starting learning rate of 0.01

1. **Normalization**

   We computed the mean and standard deviation across all images in our dataset for each of the six channels, resulting in these values: **[0.4397, 0.4234, 0.3911, 0.2279, 0.2017, 0.1825] and [0.2306, 0.2201, 0.2327, 0.1191, 0.1092, 0.1088]** respectively. We utilized **ResNet18** and **ResNet34** for training our models due to their efficiency and comparable performance. However, there were no significant differences in their accuracies; Resnet18 had test accuracy **59.44%,** while Resnet34 had test accuracy **59.2%.** Hence, we concluded that, in our case, the initial layers play a crucial role in distinguishing the differences between the various expert images. So for this purpose all of the next experiments were performed using the **Renset18** Model.

2. **Data Augmentation**

All the previous experiments were performed using two augmentation operations: **RandomResizedCrop** with a size of **224** and **RandomHorizontalFlip**. Additional augmentation was performed to assess any potential improvement in model accuracy. These augmentations are detailed below:

   1. Random Resized Crop with varying sizes such as **128x128** and **160x160**, resulting in a decrease in accuracy to **53.12%** and **56.6%** respectively.
   2. Elastic Transformation was applied to the images, resulting in a decline in accuracy down to **57.24%**.
   3. Rotation transformation ranging from **+10** to **-10** degrees was performed, but it did not significantly impact the accuracy of the model. As a result, it ended up with **59.5%** as a final test accuracy.

All of these augmentations are performed on the normalized values of the images.

3. **Learning Rate Optimization**

We changed the milestones for the step learning rate scheduler so that the higher number of epochs are dedicated to higher learning rates. Previous milestones were set at **[9, 18, 34, 50, 70]** having a decay factor **0.1.** Now**,** we tried the new milestones **[19, 36, 45, 54, 63, 72]** so that the model trains on higher learning rates for most of the epochs. Consequently, we recorded improvements in both test accuracy (**63.14%**) and training accuracy (**65.94%**). Below are the training and loss curves for this model.

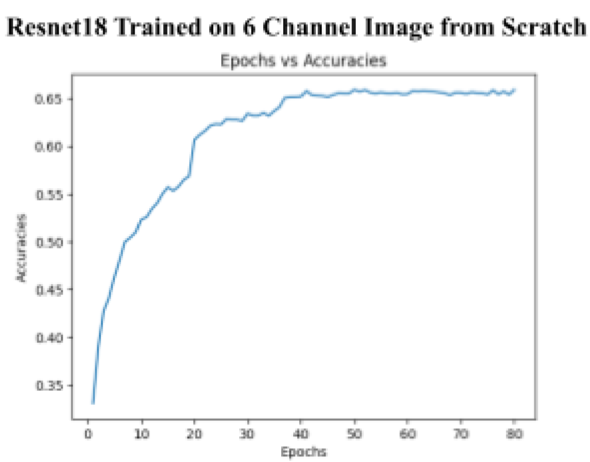

Fig. 9. Accuracy Graph depicting the Model Trained on 6 channel image from scratch using ResNet50 with optimized learning rate

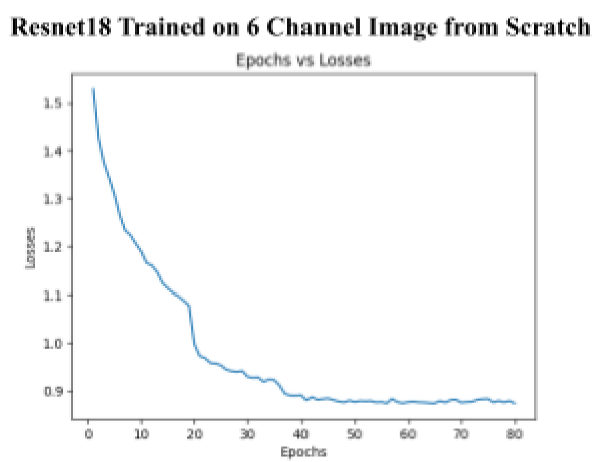

Fig. 10. Accuracy Graph depicting the Model Trained on 6 channel image from scratch using ResNet50 with optimized learning rate

Fig. 11. shows the improved confusion matrix, which is generated after the optimization of the learning rate.

Confusion matrix of ResNet18 Model Trained on 6 Channel Image from Scratch

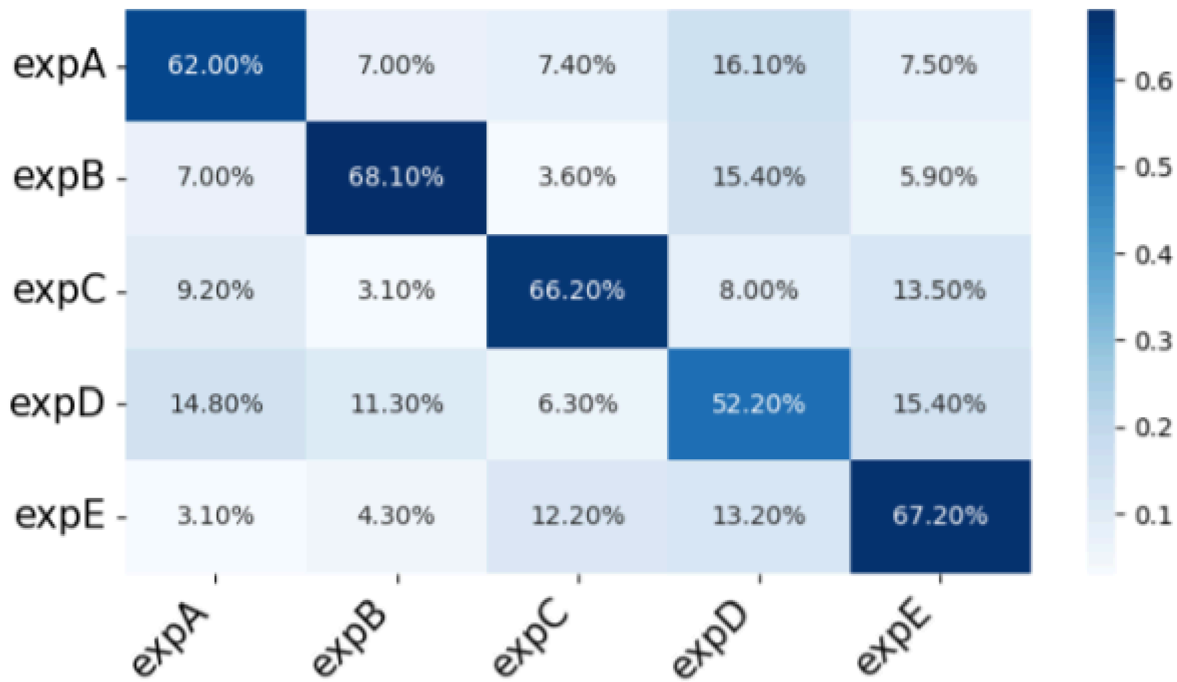

Fig. 11. The confusion matrix illustrates a high classification rate for the "expB" class and a low classification rate for "expD".