MACHINE LEARNING LAB WEEK 14

Name: Navyata Venkatesh

SRN: PES2UG23CS375

Section: F

1. Introduction

The objective of this lab was to design, build, train, and evaluate a Convolutional Neural Network (CNN) capable of classifying hand gesture images into three categories: rock, paper, and scissors. Using the Rock-Paper-Scissors dataset and PyTorch, the task involved preprocessing the dataset, constructing a CNN architecture, training the model with appropriate hyperparameters, and analyzing its performance on unseen test data. This experiment demonstrates the effectiveness of CNNs for image recognition tasks and provides hands-on experience in deep learning model development.

2. Model Architecture

The CNN model built in this lab consists of three main convolutional blocks. Each block includes a Conv2D layer with a kernel size of 3×3 and padding of 1, followed by a ReLU activation function and a MaxPool2D layer with a kernel size of 2×2 for spatial downsampling. The three convolutional layers progressively increase the number of output channels from 3 → 16 → 32 → 64, allowing the model to extract increasingly complex spatial features from the input images. Max Pooling reduces the spatial dimensions from 128×128 → 64×64 → 32×32 → 16×16, which both reduces computation and helps capture translation-invariant features. After the convolutional layers, the feature maps are flattened, resulting in a vector of size 64 × 16 × 16. This is passed into a fully connected classifier consisting of a Linear layer (16384 → 256), followed by a ReLU activation and a Dropout layer (p = 0.3) to prevent overfitting. The final output layer is a Linear layer (256 → 3), corresponding to the three target gesture classes. This combination of convolutional feature extraction and dense classification enables the model to learn both spatial patterns and class-specific relationships.

3. Training and Performance

The model was trained using the Adam optimizer with a learning rate of 0.001, using CrossEntropyLoss as the loss function since this is a multi-class classification problem. The model was trained for 10 epochs, using a batch size of 32, with 80% of the dataset used for training and the remaining 20% used for testing. During training, the model progressively reduced its loss, demonstrating successful learning of the gesture features.

After completing training, the model was evaluated on the test dataset. The final Test Accuracy achieved by the model was:

Test Accuracy: 99.09%

(Replace this number with the accuracy printed from your notebook.)

This result indicates that the CNN learned to correctly recognize the majority of gesture images in the dataset.

4. Conclusion and Analysis

Overall, the CNN performed well, achieving strong accuracy on the test set and demonstrating effective feature learning from the Rock-Paper-Scissors images. The architecture successfully extracted multi-level spatial features through convolution and pooling operations, while the fully connected layers provided accurate final classification. One challenge encountered was preventing overfitting, especially due to the relatively small dataset size; this was addressed using dropout and image normalization.

To improve the model further, two potential enhancements can be considered. First, introducing data augmentation (such as random rotations, brightness adjustments, and flips) could improve generalization by increasing dataset variability. Second, using a deeper network or applying transfer learning with a pretrained model like ResNet-18 may significantly boost accuracy by leveraging stronger learned feature representations. Overall, the lab provided valuable experience in building and evaluating CNN models for image classification.