Machine Learning ~ Lab12

Name : Nagula Anish

SRN : PES2UG23CS358

Section : F

1. Introduction

The objective of this lab was to design, build and train a Convolutional Neural Network (CNN) using the PyTorch library to accurately classify images of hand gestures. The target classes for the classification were 'rock', 'paper', and 'scissors'. The model was trained using the "Rock Paper Scissors" dataset downloaded from Kaggle, which contains over 2,000 labeled images organized into sub-folders for each class.

2. Model Architecture

The implemented CNN architecture consists of two main components: a convolutional feature extractor and a fully connected classifier.

- **Convolutional Blocks:** The model utilizes three sequential convolutional blocks. Each block consists of 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 and stride of 2.
  - **Block 1:** Expands input channels from 3 to 16.
  - **Block 2:** Expands channels from 16 to 32.
  - **Block 3:** Expands channels from 32 to 64.
- **Fully Connected Classifier:** After the third pooling layer, the feature map (sized 64×16×16) is flattened into a 1D vector of size 16,384. This vector is passed through a classifier block containing:
  1. A Linear layer reducing dimensions to 256.
  2. A ReLU activation function.
  3. A Dropout layer with p=0.3 to prevent overfitting.

4. A final Linear layer mapping the 256 features to the 3 output classes.

3. Training and Performance

The model was trained using the following hyperparameters and configurations:

- **Optimizer:** Adam
- **Loss Function:** CrossEntropyLoss
- **Learning Rate:** 0.001
- **Epochs:** 10
- **Batch Size:** 32

Performance Results:
The training loss decreased consistently, reaching a value of approximately 0.0032 by the 10th epoch. Upon evaluation with the unseen test dataset, the model achieved a final **Test Accuracy of 98.86%**.

4. Conclusion and Analysis

The CNN model performed really well, achieving high accuracy and successfully classifying the test images with almost perfect precision. The use of three convolutional layers allowed the network to effectively learn spatial hierarchies in the hand gesture images.

Challenges:
One challenge faced was ensuring the dimensions of the flattened tensor correctly matched the input requirement of the first linear layer (16,384). This required calculating the reduction in image size through the three MaxPool layers.