

Lab- 14

Name: Nidhi N | SRN: PES2UG23CS384 | SEC: F

1. Introduction

The objective of this lab was to build and train a Convolutional Neural Network (CNN) to classify hand gesture images from the Rock-Paper-Scissors dataset.

The task aimed to introduce basic deep learning workflows, including preprocessing images, designing a CNN architecture, training it using PyTorch, evaluating its performance on unseen data, and using the trained model to play a simple Rock-Paper-Scissors game.

Through this lab, I learned how CNNs extract spatial features from images and how different layers contribute to classification accuracy.

2. Model Architecture

CNN Feature Extractor (Convolutional Block)

The CNN architecture consists of three convolutional blocks, each containing:

- A Conv2D layer
 - Kernel size: 3×3
 - Padding: 1
 - Channels:
 - Block 1: 3 -> 16
 - Block 2: 16 -> 32
 - Block 3: 32 -> 64
- ReLU activation after each convolution
- MaxPool2D with kernel size 2, which reduces spatial dimensions by half
 - Image size progression: 128 -> 64 -> 32 -> 16

These layers help the model learn edges, textures, and higher-level patterns from images.

Fully Connected Classifier (FC Block)

After feature extraction, the output is flattened (size = $64 \times 16 \times 16$).

The classifier has:

- Linear layer: $64 \times 16 \times 16 \rightarrow 256$ units
- ReLU activation
- Dropout ($p=0.3$) to prevent overfitting
- Final Linear layer: 256 -> 3 units (for rock, paper, scissors)

3. Training and Performance

Hyperparameters Used

- Optimizer: Adam
- Loss Function: CrossEntropyLoss
- Learning Rate: 0.001
- Epochs: 10
- Batch Size: 32
- Train/Test Split: 1750 (80%) / 438 (20%)

Training Loss Across Epochs

- Epoch 1/10, Loss = 0.6543
- Epoch 2/10, Loss = 0.1729
- Epoch 3/10, Loss = 0.0835
- Epoch 4/10, Loss = 0.0410
- Epoch 5/10, Loss = 0.0186

- Epoch 6/10, Loss = 0.0130
- Epoch 7/10, Loss = 0.0203
- Epoch 8/10, Loss = 0.0334
- Epoch 9/10, Loss = 0.0282
- Epoch 10/10, Loss = 0.0050

Test Accuracy: 97.49%

4. Conclusion and Analysis

The CNN model performed strongly, achieving nearly 97.5% accuracy, indicating that the chosen architecture was effective at learning the visual patterns for rock, paper, and scissors. The training loss decreased smoothly, showing stable learning without overfitting.

Challenges Faced

- Ensuring that training and prediction used the same transforms.
- Fixing device movement issues (CPU/GPU mismatches).
- Understanding correct model output shapes for prediction and evaluation.

Possible Improvements

- Data augmentation (random flips, rotation, color jitter) to further increase generalization.
- Deeper network or pretrained models like MobileNet or ResNet for even higher accuracy.
- Learning rate scheduling or training for more epochs for smoother convergence.