ML Lab

Week 14

CNN Image Classification

Medha Raj

PES2UG23CS334

‘F’

1. Introduction

The objective of this lab was to build, train, and evaluate a Convolutional Neural Network (CNN) capable of classifying images of hand gestures into three categories: rock, paper, and scissors. The workflow included dataset preparation, CNN architecture construction, model training, evaluation on a test set, and performing prediction on unseen images.

2. Model Architecture

Convolutional Backbone

The CNN consists of three convolutional blocks, each including a convolution layer, ReLU activation, and MaxPooling:

Blocks- All 3 blocks have the same block architecture

- Conv2d: 3 → 16 channels, kernel size 3×3, padding 1
- ReLU
- MaxPool2d(2)

After three pooling layers, the input size reduces from 128×128 to 16×16, producing feature maps of size 64 × 16 × 16.

Fully-Connected Classifier

The classifier consists of:

- Flatten layer
- Linear: 64*16*16 → 256
- ReLU
- Dropout (p = 0.3)
- Linear: 256 → 3 (output logits for rock, paper, scissors)

---

3. Training and Performance

Hyperparameters

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

Final Test Accuracy - After training for 10 epochs, the model achieved:

Test Accuracy: 97.49%

4. Conclusion and Analysis

The CNN performed well for a simple three-class classification task, showing that even a compact architecture can learn visual patterns in gesture images effectively. Challenges included ensuring correct data preprocessing (resizing, normalization) and tuning the architecture to avoid overfitting.

Potential Improvements

- Data Augmentation: Adding random flips, rotations, and lighting changes would help the model generalize better.
- Deeper Architecture: Adding more convolutional layers or using a pretrained model (e.g., ResNet18) could further improve accuracy.