Lab Report: CNN Image Classification (Rock, Paper, Scissors)

Name: Preksha Kamalesh
SRN: PES2UG23CS902
Section: F

---

1. Introduction

The objective of this lab was to design, build, and train a Convolutional Neural Network (CNN) using PyTorch to classify hand-gesture images into three categories: **rock**, **paper**, and **scissors**.

For this lab we used the *Rock Paper Scissors* dataset containing over **2,000 labelled images**. A complete pipeline was implemented, including data loading, preprocessing, CNN architecture design, training, and performance evaluation. The aim was to assess how effectively the model can learn and recognise gesture patterns.

---

2. Model Architecture

The designed CNN model (**RPS_CNN**) consists of two main components:

A. Convolutional Block (Feature Extraction)

Three convolutional layers were used sequentially:

- **Layer 1**
    - Conv2d: *3* → *16* channels
    - Kernel size: 3, Padding: 1

    - Activation: ReLU
    - MaxPool2d: kernel size = 2
- **Layer 2**
    - Conv2d: *16* → *32* channels
    - Kernel size: 3, Padding: 1
    - Activation: ReLU
    - MaxPool2d: 2
- **Layer 3**
    - Conv2d: *32* → *64* channels
    - Kernel size: 3, Padding: 1
    - Activation: ReLU
    - MaxPool2d: 2

Spatial size reduction:
128×128 → 64×64 → 32×32 → 16×16

B. Fully Connected Block (Classifier)

- Flattened feature size: **64 × 16 × 16 = 16,384**
- **Hidden Layer:**
    - Linear: 16,384 → 256
    - Activation: ReLU
- **Dropout (p = 0.3)** for regularization
- **Output Layer:**

- Linear: 256 → 3 classes (rock, paper, scissors)

---

3. Training and Performance

Training Hyperparameters:

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

Training Behavior:

- Initial training loss (Epoch 1): **0.6433**
- Final training loss (Epoch 10): **0.0172**
- Loss consistently decreased, showing healthy convergence.

Final Evaluation:

- **Test Accuracy: 97.72%** on unseen test data

This indicates that the model successfully learned gesture-specific features such as hand shape, finger position, and contour.

---

4. Conclusion and Analysis

The CNN model demonstrated strong performance, achieving nearly **98% accuracy**.
Its architecture effectively captured the spatial features needed to distinguish between rock, paper, and scissors gestures.
The training curve suggests balanced learning without significant overfitting, likely due to the moderate complexity and appropriate use of dropout.

Future Improvements

To improve robustness and further boost accuracy:

1. Data Augmentation

- Random horizontal flips
- Small rotations
- Minor scaling

These would improve the model's ability to generalise to gestures at different angles or lighting conditions.

2. Learning Rate Scheduling

Using schedulers such as **StepLR** or **ReduceLROnPlateau** could help refine training in later epochs, when progress slows, and potentially improve accuracy further.

1.