ML Lab Week 14

CNN Image Classification

NAME: NEHA HARISH

SRN: PES2UG23CS378

SECTION: F

1.INTRODUCTION

The objective of this lab is to design, 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 from Kaggle and the PyTorch deep learning framework, the model is trained to recognize gesture patterns through convolutional feature extraction and supervised learning. The lab focuses on building the data pipeline, implementing the CNN architecture, training the network, evaluating its performance, and testing it on individual images as well as a simulated game scenario.

2.MODEL ARCHITECTURE

CNN architecture:

The CNN is built using three sequential convolutional blocks designed to extract visual features from the input images. Each block includes a convolution layer followed by a ReLU activation and a MaxPooling layer. This structure allows the network to learn low-level features such as edges and shapes in early layers, and progressively more complex gesture-related patterns in later layers.

Key Parameters:

Each convolution layer uses a 3×3 kernel with padding of 1 to preserve spatial dimensions before pooling. The number of feature channels increases across the layers: 16 channels in the first block, 32 in the second, and 64 in the third, allowing richer feature representation. After each convolution, a 2×2 MaxPooling layer reduces the height and width by half, controlling model complexity and helping the network learn hierarchical features.

Fully-connected classifier:

After feature extraction, the final feature map of size 64 × 16 × 16 is flattened and passed into a fully connected classifier. This classifier consists of a Linear layer (16384 → 256), followed by a ReLU activation and a Dropout layer (p=0.3) to prevent overfitting. The final Linear layer maps the output to 3 neurons, corresponding to the three classes: rock, paper, and scissors.

3.TRAINING AND PERFORMANCE

Training Hyperparameters:

The model was trained using the Adam optimizer with a learning rate of 0.001, which provides adaptive gradient updates for faster convergence. The loss function used was CrossEntropyLoss, appropriate for multi-class classification tasks. Training was conducted for a total of 10 epochs, with a batch size of 32 and data shuffled at each epoch to improve learning stability.

Model Performance:

After completing the training process, the model was evaluated on the unseen test dataset. The final Test Accuracy achieved was 97.72%, demonstrating that the CNN was able to effectively learn and generalize hand gesture patterns across the rock, paper, and scissors classes.

4.CONCLUSION AND ANALYSIS

Model Performance:

The model performed exceptionally well, achieving a high test accuracy of 97.72%, which indicates that the CNN successfully learned the distinguishing features of rock, paper, and scissors gestures. The predictions on single images and random game simulations further confirmed that the model generalized effectively to new inputs.

Challenges Faced:

One of the main challenges was ensuring that the dataset was properly loaded and pre-processed, especially when working in Google Colab. Maintaining consistent image transformations between training and testing was also crucial, as mismatched preprocessing can lead to incorrect predictions. Setting up the CNN architecture and confirming the correct output dimensions after pooling required careful attention.

Potential Improvements:

The model could be improved by adding simple data augmentation like flips or rotations to make it more robust. Another easy improvement would be trying a slightly deeper CNN or increasing training epochs to see if the accuracy can go even higher.