ML Lab Week – 14 Report
(CNN Image Classification)

Name: Gautam Menon SRN:PES2UG23CS196

Section : C Date:18/11/2025

1.Introduction

The object of this lab is to design,implement and train a CNN using pytorch's framework to solve a multi class classification problem. In this case we specifically were classifying images of hand gestures into 3 categories (Rock , Paper, Scissor).

We had to download the dataset , preprocess the image (resize and normalizing) , defining a 3 layer CNN architecture and training the model on a GPU to achieve high accuracy on unseen test data.

2. Model Architecture

The model consists of a feature extraction section and a classification section:

- Convolutional Layers : The architecture uses three convolutional blocks.
- Input -> RGB image of 128x128px
- Channels -> the depth increases sequentially 16->32->64
- Parameters -> each convolution uses a kernel size of 3x3 with a padding of 1 to maintain spatial dimensions before poling
- Pooling : A max pooling layer follows each convolution effectively halving the height and width of the feature maps (128->64->32->16)

Fully Connected Classifier:

- Flattening : The 3d feature maps are flattened into a 1d vector of size 16,384 (64 channels x 16)
- Hidden Layer : A linear layer maps inputs to 256 neurons, followed by a ReLU activation

• Regularization : A dropout layer with a probability of p=0.3 is used to prevent overfitting

• Output : The final linear layer maps the 256 features to the 3 output classes.

3. Training and Performance

The model was trained using the these hyperparameters and configuration:

• Optimizer: Adam

• Loss Function : CrossEntropyLoss

• Learning Rate : 0.001

• Number of Epochs : 10

• Batch size : 32

The training loss decreases consistently from 0.5818 in Epoch 1 to 0.0012 in Epoch 10 , indicating the model converged well.

Final Test accuracy : 98.17%

4.Conclusion and Analysis

Results: The model performed exceptionally well , achieving a test accuracy of 98.17% . The rapid decrease in training loss suggests that the architecture is well suited for a dataset with this complexity.

The model successfully classified the test image correctly and demonstrated the correct logic in the simulated game against itself

Challenges: One potential challenge in designing CNN's is calculating the correct input size for the first fully connected layer after flattening. Ensuring the spatial dimensions were tracked correctly through the 3 Maxpool layers was critical to defining the linear layer input of 64x16x16

Future Improvements:

1.Data Augementation

2.Early Stopping