

ML Lab Week 14: CNN Image Classification

Namritha Diya Lobo	PES2UG23CS362	Section F
--------------------	---------------	-----------

Introduction:

The objective of this lab was to design, implement, and train a Convolutional Neural Network (CNN) using PyTorch to classify hand gesture images into three classes - rock, paper, and scissors. The dataset was loaded using PyTorch's ImageFolder utility, preprocessed using image transforms, and split into training and testing sets. The goal was to complete the missing sections in the provided boilerplate notebook, train the CNN model, evaluate its performance, and summarize the results in this report.

Model Architecture:

Convolutional Layers

The model consists of three convolutional blocks, each containing a convolution layer, a ReLU activation function, and a MaxPooling layer

1. Block 1:

- conv2d(3 -> 16)
- Kernel size: 3x3
- Padding: 1
- ReLU
- Maxpool2d(2) : reduces spatial size from 128x128 to 64x64

2. Block 2:

- conv2d(16 -> 32)
- Kernel size: 3x3
- Padding: 1
- ReLU
- Maxpool2d(2) : reduces spatial size from 64x64 to 32x32

3. Block 3:

- conv2d(32 -> 64)
- Kernel size: 3x3
- Padding: 1
- ReLU
- Maxpool2d(2) : reduces spatial size from 32x32 to 16x16

After all three pooling layers, the input image (128×128) is reduced to 64 channels of size 16×16, giving a flattened feature vector of size 16384.

Fully connected classifier:

The classifier has two fully connected layers:

- A linear (16384 to 256) layer
- ReLU activation
- Dropout = 0.3 - to reduce overfitting
- A final linear (256 to 3) output layer for predicting rock, paper or scissors

Training and Performance

Training Hyperparameters

- Optimizer: Adam
- Learning Rate: 0.001
- Loss Function: CrossEntropyLoss
- Epochs: 10
- Batch Size: 32
- Transforms: Resize to 128×128, Tensor conversion, Normalization (mean=0.5, std=0.5)

Performance Metrics

After training the model for 10 epochs, it achieved a Test Accuracy of **98.40%**. The model was able to correctly classify most images in the test set, indicating that the convolutional architecture successfully learned useful features from the dataset.

Conclusion and Analysis

The CNN model performed well on the Rock–Paper–Scissors classification task and achieved strong test accuracy, showing that the three-layer convolutional architecture was effective in learning visual features from the dataset.

Challenges Faced

- Matching transforms during training and prediction.
- Handling tensor reshaping for the fully connected layers.
- Slower training when running on CPU.

Possible Improvements

- Add data augmentation to improve generalization.
- Increase model depth or number of filters.
- Use a learning rate scheduler for smoother training.

Screenshots

```
... RPS_CNN(  
    (conv_block): Sequential(  
        (0): Conv2d(3, 16, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))  
        (1): ReLU()  
        (2): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)  
        (3): Conv2d(16, 32, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))  
        (4): ReLU()  
        (5): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)  
        (6): Conv2d(32, 64, kernel_size=(3, 3), stride=(1, 1), padding=(1, 1))  
        (7): ReLU()  
        (8): MaxPool2d(kernel_size=2, stride=2, padding=0, dilation=1, ceil_mode=False)  
    )  
    (fc): Sequential(  
        (0): Flatten(start_dim=1, end_dim=-1)  
        (1): Linear(in_features=16384, out_features=256, bias=True)  
        (2): ReLU()  
        (3): Dropout(p=0.3, inplace=False)  
        (4): Linear(in_features=256, out_features=3, bias=True)  
    )  
)
```

```
... Epoch 1/10, Loss = 0.5819  
Epoch 2/10, Loss = 0.1789  
Epoch 3/10, Loss = 0.0805  
Epoch 4/10, Loss = 0.0437  
Epoch 5/10, Loss = 0.0265  
Epoch 6/10, Loss = 0.0174  
Epoch 7/10, Loss = 0.0075  
Epoch 8/10, Loss = 0.0079  
Epoch 9/10, Loss = 0.0035  
Epoch 10/10, Loss = 0.0039  
Training complete!
```

... Test Accuracy: 98.40%

```
... Randomly selected images:  
Image 1: /content/dataset/rock/zrrYLvM2uywyFwRy.png  
Image 2: /content/dataset/paper/FVEAFUE9uhfJPi2f.png  
  
Player 1 shows: rock  
Player 2 shows: paper  
  
RESULT: Player 2 wins! paper beats rock
```