Report: Comparison of MLP and CNN on CIFAR-10 Dataset

B23DCCE039 Nguyen Duc Ha Hung, B23DCCE099 Trinh Hoang Viet

May 17, 2025

1 Introduction

1. Introduction

The CIFAR-10 dataset is a widely used benchmark for image classification, consisting of 60,000 32x32 color images across 10 classes: plane, car, bird, cat, deer, dog, frog, horse, ship, and truck. It includes 50,000 training images and 10,000 test images. This report evaluates and compares the performance of two neural network architectures—a Multilayer Perceptron (MLP) and a Convolutional Neural Network (CNN)—trained on this dataset for image classification. The analysis includes training performance, test accuracies, and visualizations such as learning curves and confusion matrices.

2. Model Descriptions

2.1 Multilayer Perceptron (MLP)

The MLP is a fully connected feedforward neural network with the following architecture:

Input Layer: Flattens the 32x32x3 (3072-dimensional) image into a single vector.

Hidden Layers:

First layer: 512 neurons with ReLU activation.

Second layer: 256 neurons with ReLU activation.

Output Layer: 10 neurons, one for each class.

The MLP processes images without preserving spatial structure, treating them as flat arrays of pixel values.

2.2 Convolutional Neural Network (CNN)

The CNN is designed for image data, leveraging convolutional layers to capture spatial features. Its architecture includes:

Convolutional Layers:

First layer: 3 input channels, 32 output channels, 3x3 kernel, padding=1, ReLU activation, followed by 2x2 MaxPooling.

Second layer: 32 input channels, 64 output channels, 3x3 kernel, padding=1, ReLU activation, followed by 2x2 MaxPooling.

Third layer: 64 input channels, 64 output channels, 3x3 kernel, padding=1, ReLU activation.

Fully Connected Layers:

Flattens the feature maps (64x8x8 = 4096 dimensions).

First layer: 512 neurons with ReLU activation.

Output layer: 10 neurons for classification.

The CNN uses convolutions and pooling to exploit local spatial correlations, making it well-suited for image tasks.

3. Training Process

Both models were trained with the following hyperparameters:

Optimizer: Adam, learning rate = 0.001.

Loss Function: Cross-Entropy Loss.

Batch Size: 64.

Epochs: 10.

Device: CUDA (if available) or CPU.

Data preprocessing normalized pixel values to the range [-1, 1] using mean (0.5, 0.5, 0.5) and standard deviation (0.5, 0.5, 0.5). Training performance was tracked for loss and accuracy per epoch.

3.1 Training Results

MLP Training

Epoch 1: Loss: 1.638, Accuracy: 41.79

Epoch 2: Loss: 1.428, Accuracy: 49.62

Epoch 3: Loss: 1.319, Accuracy: 53.48

Epoch 4: Loss: 1.226, Accuracy: 56.60

Epoch 5: Loss: 1.144, Accuracy: 59.49

Epoch 6: Loss: 1.068, Accuracy: 62.02

Epoch 7: Loss: 0.992, Accuracy: 64.59

Epoch 8: Loss: 0.920, Accuracy: 67.35

Epoch 9: Loss: 0.856, Accuracy: 69.55

Epoch 10: Loss: 0.784, Accuracy: 72.00

CNN Training

Epoch 1: Loss: 1.324, Accuracy: 52.49

Epoch 2: Loss: 0.908, Accuracy: 67.90

Epoch 3: Loss: 0.705, Accuracy: 75.18

Epoch 4: Loss: 0.549, Accuracy: 80.70

Epoch 5: Loss: 0.397, Accuracy: 85.96

Epoch 6: Loss: 0.264, Accuracy: 90.90

Epoch 7: Loss: 0.164, Accuracy: 94.30

Epoch 8: Loss: 0.112, Accuracy: 96.21

Epoch 9: Loss: 0.087, Accuracy: 97.04

Epoch 10: Loss: 0.076, Accuracy: 97.42

The CNN shows faster convergence, with lower loss and higher accuracy compared to the MLP, especially in later epochs.

4. Results

4.1 Test Accuracies

After training, the models were evaluated on the test set:

MLP Test Accuracy: 52.56

CNN Test Accuracy: 73.66

The CNN outperforms the MLP by approximately 21.1 percentage points, highlighting its effectiveness for image classification.

4.2 Learning Curves

Learning curves for training loss and accuracy are shown below:

Training Loss: The CNN's loss decreases more rapidly, reaching 0.076 by epoch 10, compared to the MLP's 0.784.

Training Accuracy: The CNN achieves 97.42

The CNN demonstrates superior learning efficiency and performance.

4.3 Confusion Matrices

Confusion matrices for the test set predictions are presented below:

MLP Confusion Matrix:

CNN Confusion Matrix:

The CNN's confusion matrix exhibits higher diagonal values (correct predictions) and fewer off-diagonal errors, indicating better classification across all classes compared to the MLP.

5. Conclusion

The CNN significantly outperforms the MLP on the CIFAR-10 dataset, achieving a test accuracy of 73.66

The CNN converges faster, with lower training loss and higher accuracy throughout training.

The CNN generalizes better to the test set, as evidenced by its higher accuracy and more accurate confusion matrix.

The MLP struggles with complex image patterns, resulting in lower performance.

For image classification tasks like CIFAR-10, CNNs are the preferred architecture due to their ability to model spatial relationships, making them far more effective than MLPs.