Executive Summary

Introduction

This project provides a practical comparison of Multi-Layer Perceptron (MLP) and Convolutional Neural Network (CNN) architectures for image classification using the Fashion-MNIST dataset. Through hands-on implementation of both models, I gained experience with essential deep learning workflows including data preprocessing, augmentation, and model evaluation using PyTorch.

The comparative analysis revealed clear performance differences between the two approaches, with CNNs demonstrating superior capability for image recognition tasks. These findings offer valuable insights into optimal network structures and training strategies, providing guidance for selecting appropriate architectures in real-world computer vision applications.

Technical Approach

I started by preparing the Fashion-MNIST dataset, normalizing the pixel values using mean=0.2859 and std=0.3530, and scaling them from 0-255 to 0-1. To make the model more robust, I added data augmentation including random horizontal flips and rotations. I used a batch size of 64 for training.

For the models, I built two different architectures:

- MLP: A simple 4-layer network with sizes 784 → 512 → 256 → 128, using ReLU activation
- CNN: A more sophisticated design with 3 convolutional layers, maxpooling for downsampling, batch normalization, and adaptive average pooling to handle different image sizes

I trained both models using CrossEntropyLoss with the Adam optimizer (learning rate=0.003) for 15 epochs each.

Key Findings

After model training, the results clearly showed that CNNs are better suited for image classification:

- The MLP reached 87.4% accuracy on the validation set
- The CNN achieved 92.2% accuracy - a significant 4.8% improvement

The CNN's convolutional layers were much better at capturing spatial patterns and features in the images compared to the MLP's fully-connected layers.

Recommendations and Conclusions

Based on these results, I would recommend using CNNs over MLPs for any image classification task. The CNN's architecture is specifically designed to work with image data, and the performance improvement is substantial enough to justify the slightly more complex implementation.

For anyone starting with computer vision projects, begin with a CNN architecture similar to what I used here. CNNs provides a solid foundation that can be adapted to various image recognition problems. While MLPs are good for learning the basics of neural networks, CNNs are definitely the way to go for practical image analysis applications.

This project gave me the practical experience I was looking for and clearly demonstrated why CNNs have become the standard for computer vision tasks.