P2899592
Sneha Yenugula

Neural Networks
04 November 2025

A Neural Network Model for Fashion-MNIST Image Classification

Executive Summary

This report compares an Artificial Neural Network (ANN) and a Convolutional Neural Network (CNN) on the **Fashion-MNIST** dataset. Both were implemented in `PyTorch` and trained under identical conditions to analyze model capacity, convergence, and accuracy. The CNN achieved superior performance across all 5/8/10 epochs, which highlighting the benefit of spatial feature learning in image classification tasks.To compare the performance of an ANN and a CNN on the Fashion-MNIST dataset, evaluating accuracy, convergence, and robustness. To investigate the impact of model architecture, regularization, and hyperparameter tuning on classification effectiveness.

Data Processing

The dataset of 70,000 grayscale 28×28 fashion images was normalized to $[-1, 1]$ and split into 48k training, 12k validation, and 10k testing samples. Batches of 128 were used for efficient GPU utilization. `DataLoader` ensured randomized shuffling for training and deterministic ordering for validation/testing. Each batch was converted to tensors using a composed transform pipeline: `ToTensor()` and `Normalize((0.5,), (0.5,))`. This uniform preprocessing supported reproducible and balanced experiments for both ANN and CNN models.

Model Architectures and Overview

ANN Overview: A 4-layer fully connected MLP: flattened 784-pixel input $\rightarrow$ 512 $\rightarrow$ 256 $\rightarrow$ 128 $\rightarrow$ 10. ReLU activations add non-linearity; Dropout(0.2) prevents overfitting. Tuned hidden sizes (512/256/128) and dropout via validation; trained with lr=0.001, reaching 88% validation accuracy at 10 epochs. Treats images as flat vectors, ignoring spatial structure.

CNN Overview: Two convolutional blocks (1→32→64, 3×3 kernels, padding=1) with BatchNorm, ReLU, MaxPool, Dropout2D(0.25), followed by Dense 512→10. Tuned conv channels, dropout, and dense size; trained with lr=0.001, achieving >92% validation accuracy at 10 epochs. Exploits weight sharing and receptive fields for hierarchical spatial features.

Major Findings

Overall Accuracy and Performance Comparison

Epochs	ANN Acc. (%)	CNN Acc. (%)	Gap (pp)
5	85.90	92.25	+6.4
8	87.14	92.43	+5.3
10	87.37	92.54	+5.2

Validation Trends: ANN validation accuracy increased from 82.6% to 86.5% (5 epochs), stabilizing near 88.3% at 10 epochs. CNN validation accuracy grew from 89.0% to 92.3% and peaked near 92.9% with lower loss (0.29→0.21). CNN learning curves were smoother, suggesting improved generalization and spatial feature retention.

Test Evaluation: The CNN consistently outperformed the ANN by 5–6 percentage points. At 10 epochs, Shirt→Coat misclassifications fell from 92 (ANN) to 54 (CNN), reducing confusion by 41%. Both models converged effectively without signs of overfitting.

Confusion Matrix Comparison (ANN vs CNN, 10 Epochs)

Both models achieved good separation among visually distinct classes such as *Sandal*, *Bag*, and *Ankle Boot*. However, the ANN showed considerable overlap between similar apparel types, particularly *Shirt* and *Coat*, with 92 misclassifications. The CNN reduced this confusion to 54, demonstrating improved spatial discrimination through convolutional filters.

The CNN confusion matrix displayed stronger diagonal dominance, indicating more confident and correct predictions. Minor residual errors occurred between *T-shirt/Top* and *Shirt* due to texture and shape similarity. Overall, the CNN achieved higher per-class precision and recall across all categories, validating its superior representation of spatial hierarchies in image data.

Notable Innovations

- Structured preprocessing with reproducible dataset splits and normalization.
- Layer-wise Dropout (0.2–0.5) and weight decay (1×10^{-4}) for strong regularization.
- Integrated visualization suite: training/validation curves and confusion matrix generation.

Conclusions and Insights

The CNN exhibited faster convergence, higher accuracy, and stronger resilience to inter-class ambiguity, confirming convolutional representations as superior for image classification. Future enhancements include data augmentation (rotations, flips), cosine learning-rate annealing, and experimentation with deeper architectures (e.g., ResNet-18) to push accuracy beyond 95%.