Neural Network Model for Fashion – MNIST Image Classification.

Purpose

This project implements and evaluates artificial neural networks (ANNs) for classifying grayscale clothing images from the Fashion-MNIST dataset (60,000 training, 10,000 testing images; 28×28 pixels; 10 classes including T-shirt, trouser, and coat). Using PyTorch, the goal was to design, train, and compare a multilayer perceptron (ANN) and convolutional neural network (CNN), incorporating preprocessing, hyperparameter tuning, and innovations to achieve high accuracy while analysing model limitations, such as confusions between visually similar classes (e.g., shirt vs. coat).

Key Implementation and Findings

Data Handling

- Loaded dataset via torchvision; normalized pixel values (mean=0.286, std=0.353).
- Split into training (80%, 48,000 samples), validation (20%, 12,000 samples), and test sets (10,000 samples).
- Applied data augmentation for training: horizontal flip (p=0.5), random rotation (±10°), random crop (padding=2), Gaussian blur (kernel=3, σ=0.1-2.0) for improved generalization.

ANN Model Overview

- Architecture: Three fully connected layers (input 784 → 512 → 256 → output 10 classes); ReLU activations.
- Regularization: Dropout rate=0.3 (tuned via grid search).
- Tuned hyperparameters: Learning rate=0.001; validation accuracy post-tuning: 81.96%.

CNN Model Overview

- Architecture: Three convolutional layers (1→32→64→128 channels, 3×3 kernels, padding=1, max-pooling 2×2); followed by fully connected layers (2304→256→10); ReLU activations.
- Regularization: Dropout rate=0.3 (tuned via grid search).
- Tuned hyperparameters: Learning rate=0.001; validation accuracy post-tuning: 87.26%.

Training Setup

- Optimizer: Adam (lr=0.001, selected via grid search on lr/dropout values); loss function: CrossEntropyLoss.
- Scheduler: StepLR (step size=7, γ=0.1) for learning rate decay; batch size=64; 10 epochs.
- Ensemble: Soft-voting by averaging softmax outputs from ANN and CNN for combined predictions.
- Convergence: Stable training (e.g., CNN final validation loss=0.263).

Test Results

- ANN: 83.99% accuracy (F1-macro=0.84).
- CNN: 90.64% accuracy (F1-macro=0.91).
- Ensemble: 89.92% accuracy (F1-macro=0.90).
- Class-wise strengths: High F1-scores (>0.97) for trousers and sandals.

Performance Comparison

- CNN outperformed ANN by 6.65% due to effective spatial feature extraction.
- Reduced confusions in visually similar classes (e.g., shirt vs. coat: 153 misclassifications in ANN vs. 121 in CNN).
- Visualizations: Loss/accuracy curves (Matplotlib plots); confusion matrices (Seaborn heatmaps); detailed classification reports.
- Weaknesses: Lower performance on shirts (CNN F1=0.75).

Innovations

Grid search tuned hyperparameters for optimal generalization. Data augmentation and deeper CNN architecture enhanced performance beyond baselines. Dropout (0.3) prevented overfitting, while ensemble soft-voting improved robustness (e.g., +0.93% over ANN). Visualizations (curves, matrices) aided analysis.

Conclusions and Insights

CNNs excel for image tasks by capturing spatial hierarchies ANNs miss, while ensembles mitigate weaknesses cost-effectively. Key insights: Augmentation reduces confusions in similar classes; tuning yields ~2-3% gains. Future: ResNet blocks or advanced scheduling for >92% accuracy in scalable fashion AI.