Executive Summary

This project compares two deep learning architectures—an Artificial Neural Network (ANN) and a Convolutional Neural Network (CNN)—for classifying images in the Fashion-MNIST dataset. The aim was to determine which model achieves greater accuracy, efficiency, and generalization in visual pattern recognition. The study enhances understanding of how neural design choices affect predictive reliability in computer vision.

Project Context and Rationale

Image recognition is a leading application of artificial intelligence, with neural networks providing the computational foundation. While ANNs capture complex nonlinear relationships, they struggle to represent spatial hierarchies essential in image data. CNNs overcome this through convolutional layers that automatically extract localized spatial features, producing more efficient and hierarchical representations (Goodfellow et al., 2016).

The research addressed a central question:

To what extent does a convolutional network outperform a fully connected network in visual classification tasks such as Fashion-MNIST?

The dataset contains 70,000 grayscale images across ten clothing and footwear categories. The classification task mirrors real-world conditions where systems must distinguish subtle differences—such as between shirts, T-shirts, and coats—under limited resolution.

Methodology

Both models were implemented in **PyTorch** under identical experimental settings. Data were normalized and split into 60,000 training and 10,000 testing samples. The ANN used fully connected layers with ReLU activation and dropout regularization, while the CNN integrated convolutional and pooling layers followed by dense classification layers. Evaluation metrics included **accuracy**, **precision**, **recall**, and **F1-score**, supported by **confusion matrices** and **per-class accuracy charts** for detailed analysis.

Key Findings

Results showed the CNN's superior capacity to learn spatial features and generalize to unseen data:

- **CNN:** 92.5 % accuracy | Macro F1 = 0.9246
- **ANN:** 88.8 % accuracy | Macro F1 = 0.8869

The CNN achieved balanced precision and recall above 0.90 for most categories, including trousers, sandals, and bags. The ANN performed well overall but confused visually similar items such as shirts and T-shirts. Training curves indicated that although the ANN converged faster, it overfitted more readily, whereas the CNN demonstrated smoother, more stable learning.

Discussion

Findings align with prior research showing CNNs outperform ANNs in image-based tasks due to spatial feature extraction and weight sharing (LeCun et al., 2015). Although CNNs demand more computation, their accuracy advantage justifies the cost where precision is essential. ANNs remain suitable for simpler or low-resource applications.

Recommendations

1. **Data Augmentation:** Apply rotations, flips, and brightness adjustments to enhance robustness.
2. **Deeper Architectures:** Add convolutional blocks or batch normalization to approach benchmark accuracy (~94–95 %).
3. **Learning-Rate Scheduling:** Use adaptive optimizers (e.g., ReduceLROnPlateau) for smoother convergence.

Conclusion

CNNs significantly outperform ANNs on the Fashion-MNIST dataset, offering higher accuracy, balanced class performance, and stronger generalization. The study validates convolutional networks as the preferred architecture for image classification and demonstrates the importance of structured experimentation in deep learning research. Future extensions applying the recommended enhancements can further align results with state-of-the-art performance in computer vision.