Executive Summary – Deep Learning for Fashion-MNIST Image Classification

Introduction: Problem and Context

The growing integration of artificial intelligence in visual perception systems has transformed industries such as fashion, e-commerce, and manufacturing. Businesses today need exact and effectual ways to classify and manage huge volumes of product images. The key problem this coursework addresses is: how can deep learning models be calculated and enhanced to automatically classify fashion images with high accuracy while continuing computationally efficient?

The project uses the **Fashion-MNIST dataset**, a modern target for computer vision research containing grayscale images of clothing items like shirts, dresses, and shoes. This dataset offers a balanced and consistent testbed for evaluating different neural network architectures. The motivation overdue this work was to understand how variable network depth and learning strategies influence presentation on image recognition tasks. The research definitely compared three models: **Multilayer Perceptron (MLP)**, **Convolutional Neural Network (CNN)**, and **ConvNeXt-Tiny** model based on **transfer learning**.

This work was conducted as part of a deep learning course project, pretending an applied set-up in which a retail or tech organization might seek to automate product identification using AI. The broader aim was to find a balance between accuracy, generalization, and computational cost—factors crucial for organizing models in real-world systems.

Key Findings and Research Insights

The project was carried out using **Python and PyTorch** on a GPU environment. The workflow followed a standard deep learning pipeline, including data preprocessing, augmentation, model training, and evaluation.

Dataset Preparation:
Fashion-MNIST consists of **70,000 images (60,000 for training, 10,000 for testing)**, each 28×28 pixels and belonging to one of ten clothing categories. The dataset was normalized and augmented using random flips and rotations to increase diversity and prevent overfitting. For transfer learning with ConvNeXt, the images were resized to **224×224 pixels** and normalized using ImageNet statistics to match pretrained model requirements.

Models and Methods:

1. **Multilayer Perceptron (MLP):** A simple feedforward network with two hidden layers using SiLU activation and dropout. It served as a baseline model.
2. **Convolutional Neural Network (CNN):** A deeper architecture with convolutional, pooling, and dropout layers to capture spatial features such as edges, textures, and shapes.
3. **ConvNeXt-Tiny (Transfer Learning):** A pretrained model fine-tuned on the Fashion-MNIST dataset. The model's input layer was modified for single-channel images and its output layer replaced with ten neurons for classification.

All models were trained with **cross-entropy loss**, using the **AdamW optimizer** and learning-rate scheduling for optimal convergence. Performance was measured with accuracy, confusion matrices, and classification reports across ten categories.

Results:

- The **MLP** achieved approximately **85% accuracy**, performing adequately but failing to recognize fine visual details.
- The **Custom CNN** improved accuracy to around **88%**, showing stronger feature extraction and generalization capabilities.
- The **ConvNeXt-Tiny** model outperformed both, reaching **95% accuracy** on the test set. Its performance confirmed the effectiveness of transfer learning for image classification tasks.

Analysis of confusion matrices revealed that the most common errors occurred between visually similar items like *shirts* and *T-shirts/tops*, while easily distinguishable items such as *sneakers* and *bags* were recognized with near-perfect accuracy.
In addition, the learning curves demonstrated smooth convergence for the CNN and ConvNeXt models, while the MLP plateaued early, reflecting its limited capacity to capture spatial relationships.

Recommendations

Based on the findings, several actionable recommendations can guide future work or real-world adoption:

1. **Adopt Convolutional Architectures:**
   CNN-based models should be the foundation for any image classification pipeline, as they inherently capture local spatial hierarchies essential for accurate recognition.
2. **Utilize Transfer Learning for Small Datasets:**
   Pretrained models like ConvNeXt provide superior accuracy with less training data and time. Fine-tuning them can quickly achieve high performance without extensive computational resources.
3. **Maintain Regularization and Augmentation:**
   Applying dropout, normalization, and data augmentation consistently reduces overfitting and improves generalization, especially in visually repetitive datasets.
4. **Balance Efficiency with Accuracy:**
   While ConvNeXt delivered the best accuracy, its computational cost may be high for real-time applications. Smaller CNNs can provide an efficient trade-off between performance and resource use.
5. **Explore Emerging Architectures:**
   Future research should investigate Vision Transformers (ViT), lightweight CNN variants, or hybrid models combining convolution and attention mechanisms for even better results.
6. **Real-World Integration:**
   For commercial use, these models can be integrated into retail cataloging systems, automated inventory management, or visual search applications. Implementation should also include ethical data handling and ongoing model monitoring to maintain reliability over time.

Conclusion and Call to Action

This coursework successfully demonstrated that **deep learning models, particularly CNNs and pretrained architectures, can achieve high accuracy in image classification tasks**. The results confirmed that transfer learning is a practical and powerful approach, enabling strong performance even with limited training data and computational capacity.

For organizations or researchers interested in deploying visual recognition systems, adopting transfer learning with careful preprocessing and augmentation provides an efficient path toward reliable automation. The **ConvNeXt-Tiny** model serves as an ideal blueprint for modern image recognition, while the **Custom CNN** highlights how simpler architectures can still achieve strong results when optimized well.

Going forward, continued experimentation with newer architectures and efficient training techniques could enhance both speed and accuracy. This project highlights the real-world applicability of deep learning in visual recognition and underscores the importance of strategic model selection when balancing performance with practical constraints.

Ultimately, this work contributes to the growing understanding of how neural networks can automate visual classification tasks across industries, offering a scalable and intelligent solution to one of modern computing's most relevant challenges.