Model Architecture and Performance Analysis

In this work, a deep convolutional neural network based on **ResNet50V2** was employed for the Fashion-MNIST classification task. Since the original dataset consists of grayscale images with a resolution of **28×28**, the images were first resized to **85×85** pixels and converted to three channels to meet the input requirements of the pretrained ResNet architecture. **Transfer learning** was applied by initializing the model with ImageNet pretrained weights, followed by training a custom classification head.

Training Methodology

The training process was performed in two phases:

1. **Phase 1 (Feature Extraction):** The ResNet backbone was **frozen**, and only the newly added classification layers were trained.
2. **Phase 2 (Fine-Tuning):** The entire network was **unfrozen** and fine-tuned using a very low learning rate, allowing the model to adapt to the dataset while preserving previously learned features.

Data augmentation techniques such as rotation, shifting, and horizontal flipping were used to improve generalization.

Performance Analysis

The final model achieved a test accuracy of **94%** with a loss value of **0.1826**, demonstrating strong performance for the Fashion-MNIST dataset. These results confirm the effectiveness of transfer learning with deep residual networks, even when applied to relatively low-resolution grayscale data.

Hardware Constraint Note

It is important to note that the performance of the model is constrained by hardware limitations. All experiments were conducted using the free Google Colab GPU environment, which has limited GPU memory. Increasing the input image resolution beyond **85 × 85pixels** resulted in out-of-memory crashes, preventing further experimentation with higher resolutions. With access to stronger hardware resources (such as GPUs with larger VRAM), higher input resolutions and larger batch sizes could be used, which is expected to further improve model accuracy and stability.