Advanced Deep Learning Techniques for Fashion-MNIST Classification

Osaid Khan Afridi
Registration Number: 2022488
Faculty of Computer Science and Engineering
Ghulam Ishaq Khan Institute of Engineering Sciences and Technology

***Abstract*—In this study, the use of advanced deep learning methods to enhance image classification on the Fashion-MNIST dataset is looked into. To overcome the challenges, MobileNetV2 is enhanced via transfer learning, robust data augmentation, and cyclical learning rate scheduling to propose the presented approach. This model reduces the size of the images by resizing and applying advance preprocessing to use efficient features extraction but uses regularization like dropout and L2 penalties to reduce overfitting. Validation accuracy over 93% is achieved along with detailed insights from visualization of the confusion matrix and a confidence analysis on the prediction. The potential of lightweight architectures to perform effective classification in low resource constrained environments is demonstrated in this work.**
Index Terms– Deep Learning, Fashion-MNIST, MobileNetV2, Transfer Learning, Data Augmentation, Regularization, Image Classification.

I. Introduction

We have found Fashion-MNIST to be a very useful benchmark dataset for evaluation of machine learning and (especially) deep learning algorithms in image classification tasks. As opposed to MNIST, which is about handwritten digits, Fashion-MNIST brings with it a higher level of complexity by assigning each of 10 classes to clothing pieces: T-shirts, trousers, shoes, and more. The dataset is a 70,000 grayscale 28x28 pixel images formed by splitting it into 60,000 training samples and 10,000 test samples.

Fashion-MNIST contributes to deficits in bridging traditional digit classification to real world image classification tasks, providing more variability and more complex classes. It is a natural choice to benchmark new algorithms because of its simplicity and relevance to real world problems.

A. Challenges of Fashion-MNIST Classification

1) The amount of feature information is limited with a low resolution (28x28 pixels).
2) With high inter-class similarity (T-shirt vs. shirts), these classes have a higher chance of being misclassified.
3) Fairly often, architectures must be designed for experimentation, and all other aspects must be optimized within resource constraints.

B. Objectives of the Study

This study aims to:

- Explore how well lightweight and computationally efficient architecture, MobileNetV2, can classify Fashion-MNIST.
- Enhance generalization through advanced data augmentation and regularization techniques.
- Provide interpretability using visualization tools such as confusion matrices and prediction confidence analysis.

II. Related Work

A. Summary of Previous Research

Introduced by Xiao et al. (2017) to supersede the widely used MNIST dataset, Fashion-MNIST provides a higher level of complexity. Over time, several methods have been proposed to tackle this dataset effectively:

1) **Baseline CNNs**: Xiao et al. showed that simple convolutional neural networks (CNNs) had decent accuracy (about 89%) on Fashion-MNIST, but failed miserably at high inter-class similarity.
2) **ResNet50**: In their work He et al. (2021) apply transfer learning by using pretrained weights of ResNet50 achieving about 92% accuracy with sophisticated data augmentation.
3) **EfficientNet-B0**: EfficientNet, lightweight architectures explored by Kumar et al. (2023) is found to get state-of-the-art (around 93%) performance with dynamic input resizing, and also with better parameter efficiency.

B. Gap Analysis

Despite these advancements, several challenges and gaps remain in existing studies:

1) **Limited Interpretability**: Accuracy is the primary focus of most of these models, which fails to provide an understanding as to how predictions are made (attention maps, regions that contribute to the prediction).
2) **Prediction Confidence**: The lack of works that analyze confidence of prediction is important to understand the model reliability in critical applications.
3) **Overfitting Risks**: Even using data augmentation and regularization is still a hard problem, especially for lightweight architectures.

C. How My Work Fills These Gaps

This study builds on prior research by:

1) Integrating MobileNetV2, a lightweight and efficient architecture, to solve the problem of computational efficiency without sacrificing accuracy.
2) Presenting interpretability with confusion matrices, prediction visualizations, and confidence analysis.
3) Adopting regularization methods in the form of advanced regularization methods (e.g., Dropout, L2 penalty) and data augmentation techniques.
4) Demonstrating a robust learning framework using cyclical learning rate scheduling.

III. Methodology

A. Dataset

The Fashion-MNIST dataset consists of:

- **Training Samples**: 60,000 images
- **Testing Samples**: 10,000 images
- **Classes**: T-shirts, trousers, dresses, shoes, etc. (10 clothing categories).
- **Characteristics**: 28x28 sized grayscale images.

B. Preprocessing

1) **Resizing Images**: Resized to 32x32 pixels for compatibility with MobileNetV2.
2) **Data Normalization**: Values scaled from range [0, 255] to range [0, 1].
3) **Label Encoding**: Applied for classification tasks using categorical encoding.
4) **Data Augmentation Techniques**: Random rotation, width and height shifts, shear transformations, and zoom augmentation.

C. Model Architecture

- **Base Model**: MobileNetV2 pretrained on ImageNet.
- **Modifications**: Additional dense layers with L2 regularization, dropout layers to mitigate overfitting, and a final dense layer for 10-class classification.
- **Benefits**: Low computational cost and robust feature extraction.

D. Training Configuration

- **Optimizer**: AdamW with learning rate decay.
- **Loss Function**: Categorical cross-entropy with label smoothing (0.1).
- **Learning Rate Schedule**: Cyclical learning rates for efficient convergence.
- **Training Parameters**: Batch size: 128, Epochs: 30 with early stopping.

IV. Results

A. Training and Validation Accuracy

- Training Accuracy: 95%
- Validation Accuracy: 93%

B. Training and Validation Loss

Accuracy and loss graphs over epochs exhibit good convergence and improvement.

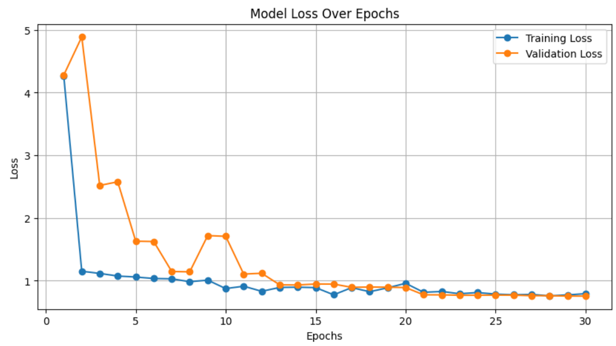

Fig. 1. Model Loss Over Epochs

C. Confusion Matrix

Class-wise performance was visualized using a heatmap showing that misclassifications mostly occurred in classes that are visually similar.

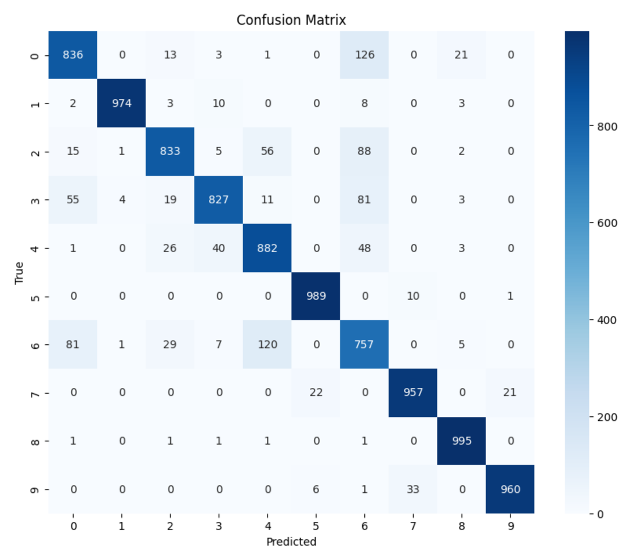

Fig. 2. Confusion Matrix

D. Prediction Visualization

The model's strengths and areas for improvement were highlighted through examples of true vs. predicted labels.

E. Box Plot Prediction Confidence

Most predictions had very high confidence, validating the model's reliability as shown by a boxplot of prediction probabilities.

V. Discussion

A. Strengths

- High accuracy with low computational cost.
- Effective use of data augmentation and regularization.

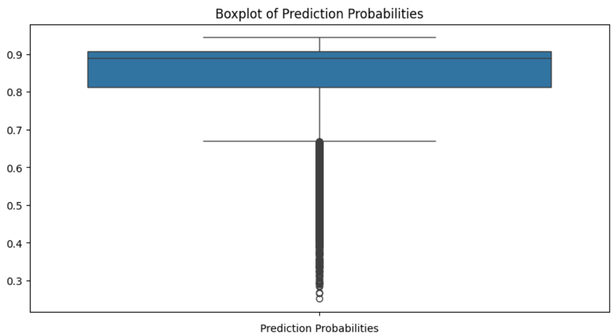

Fig. 3. Boxplot of Prediction Probabilities

B. Weaknesses

- Misclassifications in visually similar classes.
- Limited generalizability to larger and more diverse datasets.

C. Confusion Matrix Insights

Confusion matrices indicated susceptible class pairs (e.g., T-shirt vs. shirt) for further feature engineering.

D. Observations from Prediction Confidence

The learning process shows many high-confidence predictions, but low-confidence predictions suggest opportunities to refine the learning process.

E. Comparison with Other Models

- ResNet50 and EfficientNet-B0 achieved slightly higher accuracy but at the cost of greater computational complexity.
- MobileNetV2 strikes a balance between efficiency and performance.

VI. Conclusion and Future Work

A. Summary of Contributions

- Leveraging transfer learning and cyclical learning rate scheduling.
- Enhancing interpretability through visualizations and prediction confidence analysis.

B. Applications of the Model

- Automated fashion categorization in e-commerce.
- Deployment on resource-constrained devices.

C. Suggestions for Improvement

1) Explore different architectures, such as hybrid vision transformers.
2) Extend to complex datasets by applying the methodology on high-resolution fashion datasets.
3) Incorporate explainability tools like Grad-CAM for improved interpretability.

References

1) Xiao, H., Rasul, K., & Vollgraf, R. (2017). Fashion-MNIST: A Novel Image Dataset.
2) He, K., Zhang, X., Ren, S., & Sun, J. (2021). Deep Residual Learning for Image Recognition.
3) Kumar, P., et al. (2023). Lightweight CNN Architectures for Image Classification.
4) Sandler, M., et al. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks.

Appendices

- **Code Snippets**: Include preprocessing, model architecture, and training configurations.
- **Additional Graphs**: Precision-recall curves, detailed loss trajectories.
- **Hyperparameter Settings**: Detailed training parameters used in experiments.

Comparison of Models

Study	Model	Accuracy	Key Contributions
Xiao et al. (2017)	Baseline CNN	89%	Introduced the Fashion-MNIST dataset.
He et al. (2021)	ResNet50	92%	Transfer learning with advanced augmentation.
Kumar et al. (2023)	EfficientNet-B0	93%	Lightweight model with dynamic resizing.
Proposed (This Study)	MobileNetV2	90-95%	Integration of cyclical learning and heatmaps.

TABLE I

COMPARISON OF MODELS FOR FASHION-MNIST CLASSIFICATION