Image Classification Using Multiple Convolutional Neural Networks on the Fashion-MNIST Dataset

Pankaj Kumar Chaudhary (Roll No. 34)
Priyanshu Gupta (Roll No. 39)

Department of Computer Science
University of Delhi

Academic Year 2025

Guide: Prof. Dilip Senapati

Reference Paper: Davide Nitti, Francesco Leotta, Mauro Mecella

Sensors (MDPI), 2022, Volume 22, Issue 23, Article 9544

Introduction

- The rising elderly population has increased reliance on caregivers, posing long-term sustainability challenges.
- Automated assistance systems and service robots can help with domestic tasks like clothing handling.
- Clothing classification is crucial for intelligent robotic manipulation.
- Traditional feature extraction (HOG, SURF, SIFT, FAST + SVM/KNN) was computationally expensive and less robust.
- Deep learning, especially Convolutional Neural Networks (CNNs), enables automatic and accurate feature extraction.

Related Work

- **Datasets:** Fashion-MNIST, DeepFashion-C, AG, and IndoFashion are widely used for fashion image classification.
- **DeepFashion-C:** Introduced attention-based and semi-supervised networks (AHBN, dual-attention models).
- **IndoFashion:** Over 100,000 ethnic wear images for fine-grained classification.
- **Fashion-MNIST:** CNNs such as GoogLeNet, VGG, ResNet, and WRN achieve accuracy above 90%.
- Recent studies incorporated dropout, batch normalization, and augmentation for improved accuracy.

Datasets Used

1. Fashion-MNIST Dataset

- Contains 60,000 training and 10,000 testing grayscale images (28×28 pixels).
- 10 clothing categories such as shirts, trousers, coats, sandals, and bags.
- Serves as the benchmark dataset for CNN evaluation.

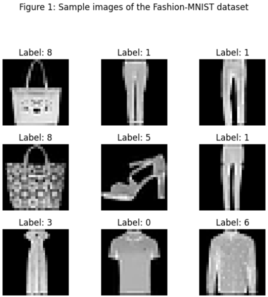


Figure: Sample images from the Fashion-MNIST dataset.

Fashion-Product Dataset

- Dataset sourced from Kaggle (Fashion-Product Dataset).
- Over 44,000 RGB images; 5,000 used for testing.
- Preprocessing pipeline:
  - Resize to 28×28 pixels
  - Convert to grayscale
  - Normalize pixel values to [1, 1]
- Used to evaluate cross-domain generalization of the trained models.

Proposed MCNN Model

- Proposed architecture: **Multiple Convolutional Neural Network (MCNN)**.
- Consists of 3 convolutional blocks + batch normalization, ReLU, and max-pooling.
- Two fully connected layers + dropout at output layer.
- Cross-Entropy Loss function:

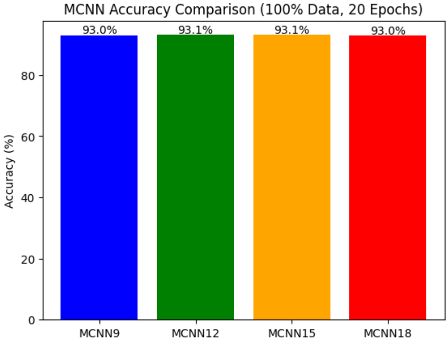
$$\text{Loss} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{K} [y_{ij} \log(p_{ij})]$$

- Tested variants: **MCNN9, MCNN12, MCNN15, MCNN18**.

Hyperparameter Optimization

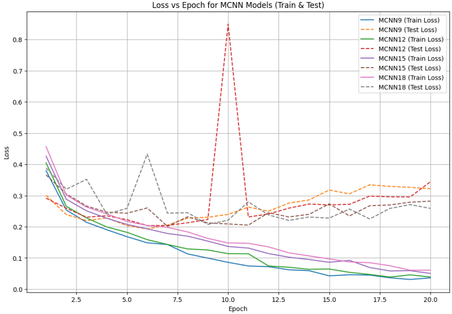
- Optimization frameworks used: Ray Tune, HyperOpt, Optuna, and PBT.
- Parameters tuned:
  - Number of filters and neurons
  - Batch size and kernel size
- Optimizer: Adam (learning rate = 0.001)
- Dropout rate: 0.3; Training epochs: 20 (CPU)

Results: Model Performance



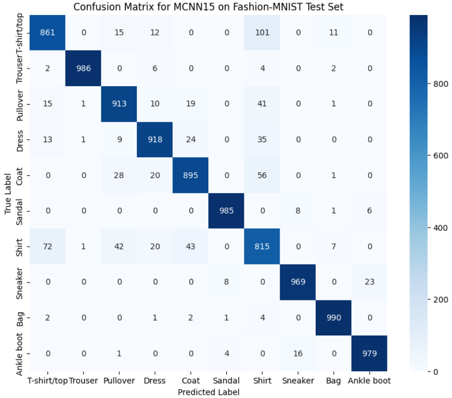
MCNN15 achieved the best accuracy of **93.1%**, showing an optimal balance between network depth and feature extraction.

Results: Loss Curve



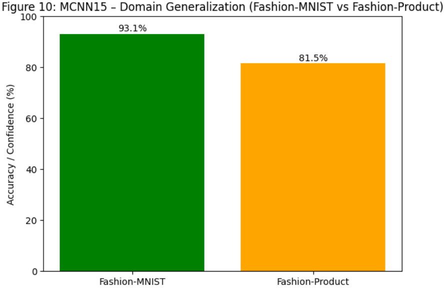
MCNN15 and MCNN12 converged faster with lower final loss. MCNN18 showed instability, suggesting deeper networks aren't always better.

Results: Confusion Matrix



MCNN15 achieved strong class discrimination, especially for sneakers, bags, and trousers. Misclassifications occurred mainly between visually similar classes like shirts and T-shirts/Tops.

Results: Cross-Dataset Comparison



Accuracy: **93.1% (Fashion-MNIST)** vs. **81.5% (Fashion-Product)**. The accuracy drop shows domain shift effects but demonstrates MCNN15's generalization strength.

Discussion

- MCNN15 was the most balanced model in terms of depth and performance.
- MCNN18 showed overfitting; deeper layers caused feature loss on small images.
- Regularization helped reduce overfitting.
- Framework differences (TensorFlow vs. PyTorch) affected results slightly.
- Future improvements: residual blocks (ResNet) and hybrid Vision Transformer (ViT) models.

Conclusions

- MCNN15 achieved:
  - **93.1% accuracy** on Fashion-MNIST.
  - **81.5% confidence** on Fashion-Product dataset.
- The model is efficient, accurate, and generalizable.
- Future work:
  - Hyperparameter optimization
  - GANs and self-supervised learning for better feature extraction
  - Testing on more diverse datasets

References

- Nitti, D., Leotta, F., Mecella, M. (2022). *Image Classification Using Multiple CNNs for Fashion-MNIST. Sensors (MDPI), 22(23):9544.*
- **Access the Original Paper: https://www.mdpi.com/1424-8220/22/23/9544**
- He, K. et al. (2016). Deep Residual Learning for Image Recognition. *CVPR.*
- Donati, L. et al. (2019). Fashion Product Classification. *Applied Sciences.*
- Paszke, A. et al. (2019). PyTorch: An Imperative Deep Learning Library. *NeurIPS.*

Authors of the Reference Paper

Davide Nitti

Francesco Leotta

Mauro Mecella

University of Rome "La Sapienza"

Published in: Sensors (MDPI), Volume 22, Issue 23, Article 9544 (2022)

Click here to view paper online