Classification of Standard FASHION MNIST Dataset Using Deep Learning Based Algorithms

(CS637 Team 4)
Gayathri Patel and Poornashree Ganapathi Subramanian

Abstract

An essential component of computer vision is image classification, which is the automatic division of pictures into pre-established groups or labels. Image classification methods use sophisticated machine learning techniques, especially neural networks, to identify and extract unique properties from photos. A dense neural network, also known as an artificial neural network, is made up of layers in which every neuron is linked to every other neuron, allowing information to flow through the network as a whole. Convolutional layers, a component of CNN's famous hierarchical design, are used to automatically extract features from input data, capturing spatial hierarchies and enabling reliable pattern identification. This model can effectively learn and generalize complicated visual patterns since it is composed of Convolutional, pooling, and dense layers. With its remarkable performance in image classification, object identification, and facial recognition, ANN and CNN have both transformed image-related tasks. This study develops and evaluates Convolutional Neural Network (CNN) and Artificial Neural Network (ANN) models for image recognition tasks using the Fashion MNIST dataset. The models' performance is used to identify pieces of apparel. After applying the algorithmic frameworks to the Fashion MNIST dataset, a thorough assessment of their effectiveness is carried out. The study's findings show that, on the MNIST dataset, the CNN model exhibits a respectable accuracy of 91%, while the ANN model gets an accuracy rate of 89%.

Introduction

The Fashion-MNIST dataset, developed by Zalando, represents a significant step forward in the field of machine learning for image classification, providing a more demanding alternative to the traditional MNIST dataset of handwritten digits. This dataset consists of 70,000 grayscale images divided into 10 fashion categories, each 28x28 pixels, designed to mirror the format of the original MNIST while introducing the complexities inherent in real-world clothing items. These complexities include varied textures, styles, and the subtle differences between similar categories such as shirts and coats, which pose distinct challenges for machine learning models.

In response to these challenges, this study employs sophisticated Convolutional Neural Network (CNN) and Artificial Neural Network (ANN) architectures that are specifically tailored to enhance feature extraction capabilities and improve model robustness. By addressing the critical issues of overfitting through the integration of dropout and batch normalization techniques, our approach not only aims to increase the accuracy of classification tasks but also enhances the efficiency and generalizability of the models. This is crucial for applications where models must perform reliably under diverse and often unpredictable real-world conditions, such as in the fashion industry where new styles and categories continuously evolve.

Through rigorous testing and evaluation, our study demonstrates how advanced neural network architectures can be effectively adapted to meet the demands of complex classification tasks, thereby setting a new benchmark for accuracy and efficiency in the use of deep learning technologies within the fashion context and potentially other similar applications.

Related Work

Previous research using the Fashion-MNIST dataset has mainly focused on basic neural network applications, with studies like [Edmira Xhaferrar et al.,2022] (1) demonstrating the effectiveness of simple Convolutional Neural Networks (CNNs) in classifying fashion images. These studies established a foundational understanding of the dataset's utility for machine learning benchmarks and highlighted the potential for more complex analyses.

However, there remains a notable gap in exploring sophisticated neural network architectures and advanced regularization techniques, which are critical for enhancing model performance and ensuring stability. While basic CNNs have been well-documented, deeper and more nuanced architectures, along with robust regularization methods like dropout and batch normalization, have received less attention.

This study addresses these gaps by integrating advanced architectural features and comprehensive regularization methods. By building on previous work, we aim to enhance feature extraction and model reliability, providing new insights into the capabilities of complex neural network architectures for challenging image classification tasks.

Methodology

Dataset Pre-processing

The foundation of our research, the Fashion-MNIST dataset, contains 70,000 grayscale images, each of 28x28 pixels, classified into ten distinct categories including tops, trousers, dresses, and shoes. This diversity offers a comprehensive benchmark for evaluating machine learning algorithms. For effective training and validation, the dataset was segmented into 60,000 training images and 10,000 test images. Further refinement of the training set segmentation involved allocating 80% of the images to actual training purposes and reserving 20% for validation. This validation subset is pivotal in tuning the model's parameters, as it helps in identifying and correcting overfitting, thus ensuring the model's ability to generalize to new data, which is critical for real-world applications.

Model Architecture

In pursuit of an optimal model architecture for classifying complex fashion items, our research involved extensive experimentation with ANN and CNN architectures. The CNN architecture features a sequential layout starting with two convolutional layers, optimized to capture textural and structural details from the fashion images. The first layer contains 32 filters, suitable for detecting basic features like edges and simple textures, while the second layer doubles the filters to 64, allowing for more complex pattern recognition. Each convolutional layer is followed by a max-pooling layer that reduces the dimensionality of the data, thus speeding up the computation and helping in extracting the dominant features while reducing overfitting. To introduce the necessary non-linearity in the model and ensure rich feature hierarchies, ReLU activation functions were applied post each convolution operation. Dropout layers were strategically placed after each pooling step to provide robustness, by randomly omitting subset features during training, which helps in making the model less sensitive to the specific weights of neurons and more capable of generalizing better. The architecture concludes with a fully connected layer that transitions the learned features to a classification output via a softmax layer, effectively categorizing the input into one of the ten fashion classes.

The ANN model has a hierarchical structure with fully connected layers corresponding to the extraction of meaningful features from the Fashion MNIST dataset. The input layer contains 784 neurons, representing the pixel of the input image a flat value. Then, two hidden layers were added, each containing highly connected neurons. The first hidden layer of 128 neurons is incorporated with ReLU activation functions to introduce nonlinearity and facilitate feature extraction. Then, a dropout rate of 0.25 was entered in order to reduce overfitting and increase the robustness of the model. The second hidden layer with 64 neurons continues the feature abstraction process through ReLU activation and dropout layers.Finally, the output layer of 10 neurons corresponding to fashion classes uses the LogSoftmax activation function to generate probability distributions across the ten different classes. This architecture was optimized to balance model complexity and generalizability, ending with an effective classification system of fashion products.

Training Process

The training regimen for both the Convolutional Neural Network (CNN) and the Artificial Neural Network (ANN) was meticulously crafted to optimize model convergence and accuracy. Utilizing the Adam optimizer, known for its efficiency in handling sparse gradients and adapting the learning rate, was crucial for this process. The optimizer was set with a learning rate of 0.001 to ensure a balanced approach between speed and accuracy of convergence. Each model underwent extensive training for up to 25 epochs, but with a safeguard early stopping was implemented based on validation loss to circumvent the common pitfall of overfitting. This approach allows for the model to train just enough to achieve high performance without memorizing the training data, which could degrade its performance on unseen data. All computational tasks were executed using the PyTorch framework on high-performance GPU hardware, which significantly reduced training time and allowed for the processing of large batches of data efficiently, enhancing the overall learning process and stability of the model.

During training, the ANN architecture followed a similar regimen to the CNN, with adjustments tailored to its specific structure. The ANN model, equipped with dense fully connected layers and ReLU activation functions, underwent optimization using the Adam optimizer with the learning rate of 0.003 for 25 epochs. Early stopping based on validation loss was implemented to prevent overfitting, ensuring that the ANN converged effectively while maintaining generalization capability. Like the CNN, all computational tasks for training the ANN were executed using the PyTorch framework on high-performance GPU hardware, facilitating efficient processing of large batches of data and enhancing the overall learning process and stability of the model.

Evaluation Metrics

To comprehensively assess the performance of our models on the Fashion-MNIST dataset, we employed a multifaceted approach to evaluation metrics, focusing on both overall accuracy and the precision of classification for each individual category. This dual approach allows us to ensure both general effectiveness and specific performance across varied types of fashion items.

Accuracy: This metric provides the proportion of total predictions that were correctly classified by our model. It gives us a high-level overview of the model’s effectiveness across all categories but does not distinguish between the types of errors made.

Precision: Precision is critical in scenarios where the cost of a false positive is high. For each category in the Fashion-MNIST dataset, precision measures the accuracy of positive predictions. It answers the question, "Of all the items labeled as belonging to a specific category, how many actually belonged to that category?" This metric is particularly useful in ensuring that our model reliably identifies specific fashion items without confusion.

Recall: Also known as sensitivity, recall measures the ability of the model to find all the relevant cases within a dataset. For each category, it reflects the percentage of actual positives that were correctly identified, answering the question, "Of all the actual items from a specific category, how many did we correctly classify as belonging to that category?" This metric is crucial in scenarios where missing out on a fashion item can lead to significant repercussions, such as stock mismanagement.

F1-Score: Since precision and recall can often have an inverse relationship, the F1-score becomes a critical metric that balances the two by taking their harmonic mean. This score is particularly useful when we need a single metric to benchmark performance, especially when uneven class distribution might render accuracy misleading.

Confusion Matrix: Beyond individual metrics, we utilized confusion matrices as comprehensive tools to visually inspect model performance across all categories. A confusion matrix provides a detailed breakdown of prediction results for each category, showing exactly where the model is confusing one fashion item for another. This insight is invaluable for fine-tuning our model, as it highlights specific areas where the model is underperforming and may require additional training or architectural adjustments.

By leveraging these metrics, our evaluation process is not only robust but also granular, allowing for targeted improvements and a deeper understanding of model capabilities and weaknesses in classifying fashion items.

Results

The deployment of our advanced Convolutional Neural Network (CNN) and Artificial Neural Network (ANN) models on the Fashion-MNIST dataset led to noteworthy achievements in classification performance, with CNN attaining an accuracy of 92% and ANN attaining the accuracy of 88% on the test dataset. This result demonstrates how well CNNs can identify complex patterns in photos pertaining to fashion. The CNN model's high accuracy has encouraging implications for a clothing company. Given CNNs' greater performance in image classification, it is possible that these models will be essential in automating the classification of clothing items.

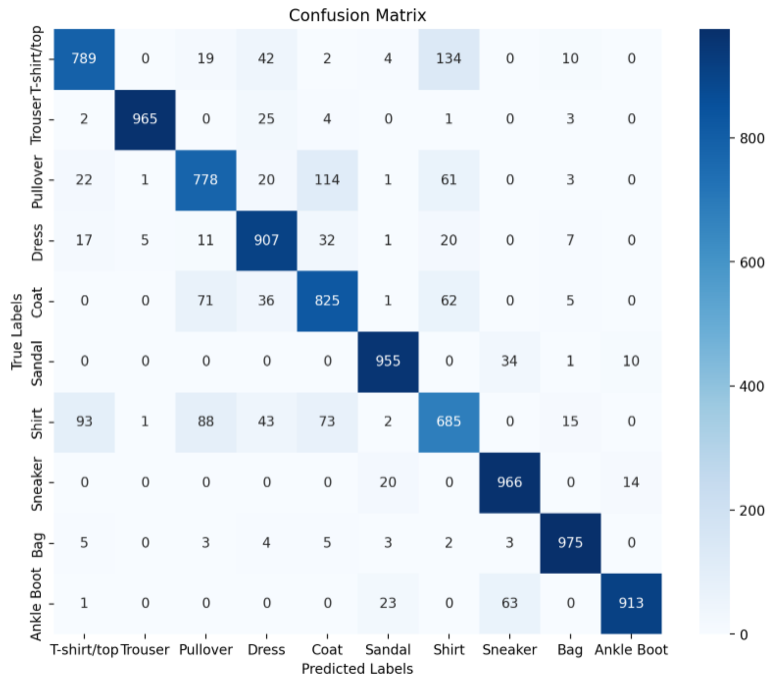

Fig 1: Confusion matrix for ANN model

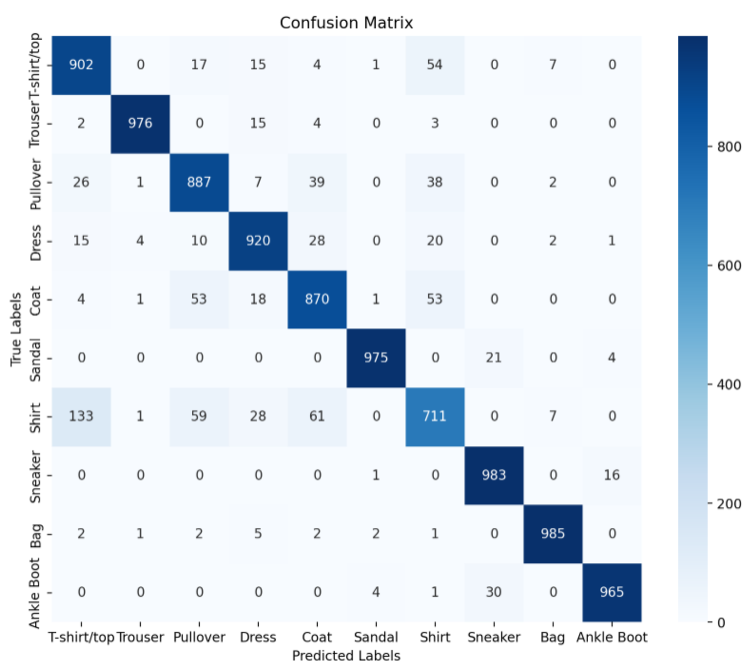

Fig 2: Confusion matrix of CNN model

Classification Report:

	precision	recall	f1-score	support
0	0.85	0.79	0.82	1000
1	0.99	0.96	0.98	1000
2	0.80	0.78	0.79	1000
3	0.84	0.91	0.87	1000
4	0.78	0.82	0.80	1000
5	0.95	0.95	0.95	1000
6	0.71	0.69	0.70	1000
7	0.91	0.97	0.94	1000
8	0.96	0.97	0.97	1000
9	0.97	0.91	0.94	1000
accuracy			0.88	10000
macro avg	0.88	0.88	0.88	10000
weighted avg	0.88	0.88	0.88	10000

Fig 3: Classification report of ANN model

Classification Report:

	precision	recall	f1-score	support
0	0.83	0.90	0.87	1000
1	0.99	0.98	0.98	1000
2	0.86	0.89	0.87	1000
3	0.91	0.92	0.92	1000
4	0.86	0.87	0.87	1000
5	0.99	0.97	0.98	1000
6	0.81	0.71	0.76	1000
7	0.95	0.98	0.97	1000
8	0.98	0.98	0.98	1000
9	0.98	0.96	0.97	1000
accuracy			0.92	10000
macro avg	0.92	0.92	0.92	10000
weighted avg	0.92	0.92	0.92	10000

Fig 4: Classification report of CNN model

Model Performance Comparison: The CNN model exhibits exceptional proficiency with an accuracy of 92%, demonstrating its ability to generalize effectively across new, unseen data, a crucial aspect for real-world applications. This high accuracy underscores the robustness and efficacy of the architectural decisions made during the design phase. In comparison, the ANN model achieves an accuracy of 88%, showcasing strong performance but slightly trailing behind the CNN counterpart. Despite this difference, both models demonstrate commendable capabilities, highlighting their suitability for fashion item classification tasks.

Influence of Dropout and Batch Normalization: Key to our model’s performance was the integration of dropout and batch normalization techniques within the network architecture. Dropout played a crucial role in reducing overfitting by randomly omitting units during training, which helps in creating a more generalized model that does not rely too heavily on any single or a set of features. Batch normalization contributed significantly by normalizing the inputs of each layer, ensuring that the network maintains a faster and more stable learning phase. These techniques collectively enhanced the model's ability to perform well across diverse sets of fashion items by improving network training dynamics and model stability.

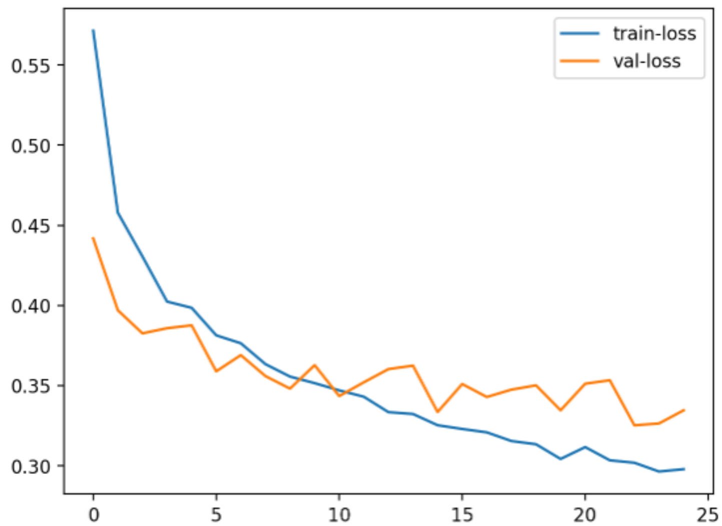

Fig 5: Accuracy plot for ANN model

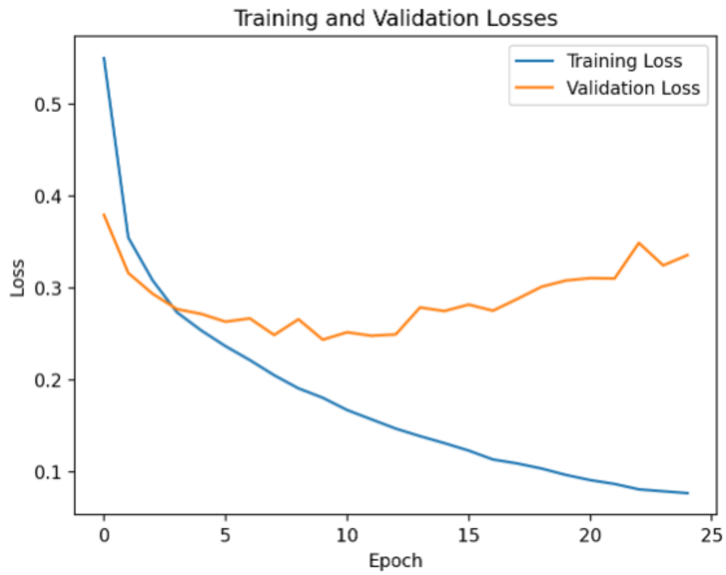

Fig 6: Accuracy plot for CNN model

Analysis of Loss and Accuracy Curves: Throughout the training process, we meticulously monitored the loss and accuracy curves, which provided deep insights into the model’s learning behavior. These curves demonstrated a consistent decrease in training and validation loss without significant fluctuations, suggesting that the model was learning effectively without fitting excessively to the noise in the training data. The accuracy curves complemented this finding, showing a steady increase in accuracy over epochs for both training and validation sets, confirming that our model improvements were indeed contributing positively to its ability to classify images accurately.

Visualizations and Confusion Matrix Analysis: Alongside numeric metrics, visualizations such as the confusion matrix provided a detailed view of the model's performance across different categories. The confusion matrix revealed high true positive rates for most categories with some exceptions where similar fashion items like shirts and T-shirts caused some confusion, pointing to potential areas for further model refinement.

This study underscores the effectiveness of advanced Convolutional Neural Network (CNN) architectures and regularization techniques in enhancing the accuracy and robustness of fashion item classification. Through the implementation of dropout and batch normalization, our CNN models achieved improved stability and generalization, effectively mitigating overfitting. These findings highlight the significance of meticulous hyperparameter tuning and architectural choices in optimizing the performance of neural networks.

In contrast, while the Artificial Neural Network (ANN) model also demonstrates commendable performance, achieving an accuracy of 88%, it falls slightly behind the CNN counterpart. Nonetheless, the utilization of techniques such as dropout and batch normalization in the ANN contributed to enhanced stability and reduced overfitting, further validating their efficacy in improving model performance. These results underscore the importance of employing regularization techniques and architectural refinements to enhance the accuracy and reliability of ANN models in fashion item classification tasks.

These results not only validate the efficacy of our advanced CNN models but also highlight the impact of methodological enhancements like dropout and batch normalization in achieving high accuracy and robustness in image classification tasks. This comprehensive evaluation approach ensures that our findings are well-supported and clearly demonstrate the model’s capabilities in handling complex classification challenges. The research confirms the potential of CNNs for complex image classification tasks and provides insights applicable to broader image recognition challenges. The methods and outcomes can serve as a foundation for further investigations, potentially accelerating advancements in visual data analysis.

In conclusion, the Fashion MNIST dataset was successfully used to train and test Artificial Neural Network (ANN) and Convolutional Neural Network (CNN) models for the image classification task. CNN exhibited superior accuracy, achieving an impressive 92%, compared to ANN’s commendable 88%.

The classification reports revealed insightful metrics such as precision, recall, and F1-score for each class, providing a detailed understanding of the models' strengths and weaknesses. Both ANN and CNN displayed robust performance across various fashion item categories, demonstrating their efficacy in real-world image classification tasks. The CNN model's high accuracy has encouraging implications for a clothing company. Given CNN model’s greater performance in image classification, it is possible that these models will be essential in automating the classification of clothing items. To further improve the classifier performance, we should collect more samples and give more images to the model so the classifier can learn more features or patterns even better.

References

1) FASHION-MNIST CLASSIFICATION USING CNN. (2023, August 13). International Research Journal of Modernization in Engineering Technology and Science. https://doi.org/10.56726/irjmets42990.
2) Kadam, S. S., Adamuthe, A. C., & Patil, A. B. (2020). CNN Model for Image Classification on MNIST and Fashion-MNIST Dataset. Journal of Scientific Research, 64(02), 374–384. https://doi.org/10.37398/jsr.2020.640251.
3) Kim, H. S. (2021, March 31). Classification and Combination of Fashion Items Using CNN-Based Deep Learning. Journal of Digital Contents Society, 22(3), 475–482. https://doi.org/10.9728/dcs.2021.22.3.475.
4) Nocentini, O., Kim, J., Bashir, M. Z., & Cavallo, F. (2022, December 6). Image Classification Using Multiple Convolutional Neural Networks on the Fashion-MNIST Dataset. Sensors, 22(23), 9544. https://doi.org/10.3390/s22239544.
5) Abd-Elaziz, O. F., Abdalla, M., & Elsayed, R. A. (2023, November 28). Deep Learning-Based Automatic Modulation Classification Using Robust CNN Architecture for Cognitive Radio Networks. Sensors, 23(23), 9467. https://doi.org/10.3390/s23239467.