Image Classification Accuracy on Fashion-MNIST Dataset Using ANN, CNN, and Conditional GAN Augmentation

Abstract

In this project, we have investigated the performance of Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) to classify images using the Fashion-MNIST dataset, and we extended this study by utilizing conditional Generative Adversarial Networks (cGAN) used as synthetic data to augment our data set. This study examines the effect of model architecture and synthetic data on classification accuracy, generalization, and class balance. This study was written using the TensorFlow/Keras framework and executed on a T4 GPU. The results show that CNNs outperform ANNs in image visual tasks, and while GAN methods that create augmentations to the existing image quality limits, may improve generalization.

Introduction

The objective of this coursework is to compare the classification performance of ANN and CNN models on a standardized visual dataset and to explore whether data augmentation using a Conditional GAN (cGAN) can enhance CNN learning on underrepresented classes. We selected the Fashion-MNIST dataset because it contains 70,000 grayscale images across 10 fashion categories, offers moderate difficulty and is well-accepted as a benchmark for comparing deep learning models.

Learning about ANN versus CNN capabilities contributes to the community's scientific discourse about enhancing computer vision models or approaches for overcoming drawbacks of imbalanced or insufficient datasets for supervised learning. Specifically, GAN-based augmentation offers a unique approach for producing synthetic but still relevant data, especially when real-world data collection is expensive or limited.

Experimental Methodology

The experiment was conducted in three stages:

1. **Model Development:**
    - The ANN utilized three dense layers (with layer sizes of 512, 256, 128 respective neurons), implemented the ReLU activation function, and used dropout for bounds on overfitting
    - The CNN was grounded in three convolutional layers (with respective filters sizes of 32, 64, 64 each) with max pooling on the convolutional side and a dense classification head with 128 neurons.
    - Both models were optimized with Adam optimizer with (learning rate = 1e-3) and early stopping to prevent the overfitting.
2. **Dataset Preparation:**
    - The Fashion-MNIST dataset was normalized to interval [0,1] and stratified split between training (54,000), validation (6,000), and testing (10,000) datasets.

- Training data was re-scaled to interval [-1,1] in order to make it more stable and the Conditional GAN used a strategy of training for between 6-20 epochs to generate synthetic data for the selected class labels (T-shirt/top, Pullover, Coat, Shirt) in the GAN stage.

3. **Evaluation Metrics:**

- Model performance during evaluation is provided in terms of accuracy, precision, recall, and F1-score (macro-averaged, without support).

- Visual assessments of evaluations occurred with training curves, bar charts, and confusion matrices through analysis of per-class performance and patterns of misclassification.

Results

- The **ANN** yielded an impressive test accuracy of **88.9%**, alongside macro precision, recall, and F1-score values around 0.889.

- The **CNN** exhibited an impressive accuracy of slightly higher (**92.2%**), and macro precision, recall and F1-score of (**0.9217**, **0.9221**, and **0.9215**, respectively), again showing it fared superiorly in learning hierarchical spatial features when compared to image recognition models using **ANN**.

- The Conditional GAN was able to generate synthetic samples but the generated images rang slightly blurry because it was trained with few epochs and only a lightweight architecture.

- The GAN-augmented CNN achieved **89.7%** accuracy, showing improved robustness for underrepresented classes but a small overall accuracy decrease compared to the baseline CNN.

- Confusion matrices revealed that misclassifications were primarily between **visually similar classes**, such as Shirt vs. T-shirt and Coat vs. Pullover.

- The cGAN achieved lower accuracy and produced slightly blurry images mainly because it was trained for fewer epochs (6–20) and used smaller network layers with limited pixel detail learning. The [-1,1] pixel normalization and short training prevented the generator from capturing fine textures, leading to softer and less realistic outputs.

Conclusion and Future work

The study concludes that CNN models outperform ANNs for image-based classification due to their convolutional feature extraction and spatial awareness. While GAN-based augmentation did not improve the overall accuracy beyond the baseline CNN, it enhanced model generalization for specific classes, indicating potential for further refinement.

Future work should focus on improving GAN training stability and image quality, extending the training epochs, and exploring deeper architectures (e.g., DCGAN or StyleGAN). Incorporating additional augmentation techniques and transfer learning could further optimize accuracy and reduce dependency on synthetic data.

This experiment demonstrates a complete, reproducible deep learning workflow integrating classification and generative modeling, reinforcing the value of GAN-augmented CNNs for addressing data imbalance in computer vision.