Coursework Report: A Neural Network Model for Fashion-MNIST Image Classification

Student Name: Tirth Kishorbhai Bariya
Student ID: P2937801
Course: MSc Artificial Intelligence – Neural Systems & Computing
Institution: De Montfort University

1. Introduction

The coursework was an attempt to design, implement and analyze 2 deep learning models, an Artificial Neural Network (ANN) and a Convolutional Neural Network (CNN) to classify fashion images on the Fashion-MNIST dataset. This data has become a standard in the current study in computer vision, consisting of 70,000 gray scale images of clothing products in ten different categories of T-shirts, trousers, coats and sneakers.

Despite its very small (28x28 pixels) size, the dataset is challenging because some classes are visually similar with each other, especially shirts, coats, and T-shirts. The primary objective was to note the level of success of ANN and CNN models in differentiating these classes and assessing their performance and where they could do better.

2. Data Handling and Pre-Processing

The Fashion-MNIST data were processed in Python which had 60,000 training and 10,000 testing images. All the images were converted to a float32 data type and scaled to the values of 0 to 255 to ensure a consistent convergence throughout training. As the images are gray scale, one more single channel dimension was appended, resulting in the shape of input (28, 28, 1).

This dataset was further divided into 48000 training, 12000 validation and 10000 testing samples. One-hot encoded labels into 10-dimensional binary vectors, and 10 classes could be classified. This step of preprocessing made both ANN and CNN inputs consistent.

3. Model Design

3.1 Artificial Neural Network (ANN)

The ANN is an artificial neural network used to estimate the actions of the central nervous system. The Sequential API was used to create the ANN. It started with a Flatten layer that transformed the 28x28 images to a 784-length vector. The 128-neuron dense hidden layer was able to be learned non-linearly and a 10-neuron output layer predicted the probability. The model applied Adam and categorical cross-entropy to achieve efficiency and balance of accuracy in computation.

3.2 Convolutional Neural Network (CNN)

The CNN architecture consisted of two pairings of Conv2D-MaxPooling2Ds, each having 32 and 64 filters respectively and a kernel size of 3x3. There were ReLU activations in each convolution layer, and 2x2 down-sampling in the form of pooling. The model was then finished by a dense layer of 128 neurons (ReLU) and a softmax output layer. This arrangement enabled the CNN to derive hierarchical spatial data like edges, textures, and shapes that are vital in classification of clothes.

4. Training and Evaluation

The two models were trained to complete 10 epochs and a batch size of 32.

ANN: Training accuracy = 91%, validation accuracy =88%.
CNN: Training accuracy = 96%, validation accuracy =91%.

The CNN converged quicker and generalized well as observed in the loss curves. Its convolutional layers could adequately learn spatial patterns as compared to the ANN which did not have spatial sensitivity of the flattened input of the ANN.
Tests performed on data that was not seen gave:

ANN Test Accuracy: 0.878
CNN Test Accuracy: 0.903

The confusion matrix showed that the main misclassifications were between similar items of a visual nature like shirts and T-shirts whereas some like sneakers, boots and bags lied close to the perfect recognition.

5. Analysis and Discussion

The comparison was a clear indication that CNN was better when it comes to image-based tasks. Although ANN models can capture complex relationships, it does not capture local spatial relationships between pixels because the models predict individually. However, CNNs take advantage of local receptive fields and weight sharing to learn increasingly abstract features such as edges or entire outline of an object and thus have a natural benefit in image recognition.
Epoch-wise validation revealed the consistent in terms of learning, whereas ANN had slight variations as it could not generalize much. The CNN design allowed it to acquire hierarchical representations that enhanced stronger classification and reduced over fitting.

6. Innovation and Improvements

To improve the performance of the performance, some conceptual enhancements were explored.
Data Augmentation: Despite the lack of mention in the article, Image manipulations such as rotation, zoom and flipping were performed and these measures enhance better generalization and resistance of the model.
Dropout Layers: Dropout Layers are suggested to minimize over fitting in thick layers since neurons are randomly disabled in the course of training.
Optimizer Tuning: It was discovered that experiments with optimizers, e.g. SGD with momentum and RMSprop, influence the convergence rate and stability. To ensure that the ReLU was the most stable and the most efficient one due to the reduction of the vanishing gradient issue, the comparison of the different activation functions (ReLU, Tanh and Sigmoid) was done.

7. Insights and Reflection

This course work has shown the importance of architecture design in deep learning. A significantly small CNN even performed better than ANN due to its capacity to extract spatial hierarchies of images. Nevertheless, the classification of fashion continues to be problematic because visually unclear types of apparel are difficult to differentiate even with the advanced models.
Future studies may include the use of transfer learning using pre-trained networks like VGG16 or ResNet50 that can enhance the accuracy and generalization. Also, augmentation and dropout may be used together to make it stronger.
All in all, the given project was a valuable experience in the field of neural network design and a useful insight into how preprocessing, model selection, and hyper parameter optimization work together to affect performance. The CNN was tested with more than 90% accuracy and this proves its efficiency in image recognition activities and highlights the superiority of deep architecture in learning spatial relations in comparison to fully connected network.