Individual Report on FASHION-MNIST

STUDENT NAME: SHEHRYAR SHEHRYAR

STUDENT ID: P2952028

MODULE TITLE: NEURAL SYSTEM

WORD COUNT: 821

Shehryar Shehryar – P2952028
DE MONTFORT UNIVERSITY, LEICESTER

In this project, I used the four deep learning models that I learn during the term time – Artificial Neural Network (ANN), Convolutional Neural Network (CNN), U-Net, and Vision Transformer – are trained through the Fashion-MNIST dataset, and also after that I compared their results, which model show the best accuracy. The purpose of this project is to identify how each model learns visual patterns well, how it consumes training data, and how good it is at predicting unseen images. The dataset Fashion-MNIST is a widely used in computer vision experiments. It contains black-and-white photographs of clothing items such as shirts, pants, shoes, dresses, and bags. It's harder than your classic handwritten digit dataset, so it'll be the perfect bridge between theory and real-world computer vision applications.

The First step in this project I do is to download and prepare the dataset. I normalize images to stabilize learning and then split dataset into train, validation and test. This part is very important, as splits provide you with assurance, that the model is not overfitting and learning rate is good. The training set educates the network, the validation set makes it smarter, and the test set shows us how well it can do when it comes down to it. Otherwise, the model might create good results in theory and fail miserably in practice.

The first trained model I used in my project was the ANN (Artificial Neural Network). The architecture was to flatten the whole 28×28 image into a vector and pass it through the fully-connected layers. It was used as a baseline for the rest. Another drawback is the absence of image spatial structure, which means that especially on mammographic images, the networks are less capable of capturing the visual patterns than CNNs or U-Nets. Nevertheless, the ANN was able to learn important features and achieve decent accuracy, which just shows that even the most networks can show remarkable performance given enough data. Although, when it is in comparison to deeper networks when images look alike and are easy to confuse, such as shirts and coats.

The second model i trained was a CNN (Convolutional Neural Network). It is specially designed for image tasks because CNN uses convolutional layers to find edges, textures, shapes, and more complex patterns. As shown in the output of my project, CNN it seems to produce considerably better results than an ANN, and performs more stable loss curves. In general, it solved noisy and overlapped patterns. For example, CNN able to draw a clear boundary between shoes and sandals. Therefore, it was once again proven that CNNs could serve as an effective body for image classification, it is also the reason why CNNs found their way into being implemented in such places as medical or traffic systems, and face recognition.

The third model was I used is U-Net. Although U-Net was initially designed to tackle different issues like segmentation problems, the model in question was modified for classification. U-Net consists of the encoder-decoder architecture and skip connections that give it the ability to keep low-level output details while still acquiring high-level

characteristics. Because of this model U-net it easily handles the image and texture very well. That's why due to this fact, U-Net exhibited the best ability in processing textures and edges. So, this experiment demonstrated that U-Net architecture is not only a strong candidate for doing segmentation but also, if adjusted correctly, can do classification accurately.

Other than that, I used the Vision Transformer which is by far the best and state of the art architecture. The concept behind Vision Transformer is that while CNNs scan images individually, Vision Transformer divides images into tiny patches and then views these patches like phrases in a sentence. It learns context globally rather than concentrating on local patterns in the image. The architecture is generally taken from transformers in NLP tasks, such as text translation and chatbots. The vision transformer demonstrated good generalization and distinguished classes effectively, especially when clothing items had the same shape and differed in the texture or style.

In conclusion, this project perfectly illustrated how various deep learning methodologies respond when we give the same data. ANN served as a great primitive starting point. CNN demonstrated why it remains a tremendously impactful architecture for images. U-Net exhibited its versatility and utilization outside the boundaries of segmentation. Finally, Vision Transformer displayed in which direction the field is moving. All combined, these models painted a perfect roadmap of how neural networks begin from basic perceptron's, to convolutional models.

In the Future, this work can be enhanced by incorporating data augmentation, trial of deeper transformer variants, employing dropout scheduling, or giving transfer learning with pre-trained models. It could also be improved by using real-world datasets such as CIFAR-10, CIFAR-100, or even medical imaging data.