NS ASSIGNMENT ON THE FASHION-MNIST DATASET

Name: Venkata Sai Tharun Susarla

ID: P2930380

Overview: The report's primary goal is to display the differences in an Artificial Neural Network (ANN) and a Convolutional Neural Network (CNN) performance by the scores attained on the Fashion-MNIST dataset. The dataset consists of 70,000 grayscale images that are categorized into 10 different classes of clothing, and it is a popular choice for the evaluation of image recognition models because of its acceptance in the community. The purpose of the study is to investigate the differences between the two models in terms of their architecture, training, and accuracy in classification.

Data Preparation: The dataset was loaded through the torchvision.datasets.FashionMNIST function and it was subdivided in such a way that 48,000 images were set aside for training, 12,000 for validation and 10,000 for testing. Images were transformed to tensors and normalized using the transforms.ToTensor() function. Data loaders of batch size 128 were created, with shuffling turned on for the training set to ensure better robustness of the model. Correct image-label alignment was verified through visual inspection of sample batches.

Model Architectures:

Artificial Neural Network (ANN):

An ANN was able to use a 784-dimensional input that was a flattening of the images consisting of 28×28 pixels. There were three fully connected layers of 512, 256, and 128 neurons respectively and all of them had ReLU activation after that. For regularization, dropout rates of 0.3 and 0.2 were used, and a linear layer at the end assigned the outputs to 10 classes.

Convolutional Neural Network (CNN):

The CNN was made up of three convolutional blocks with filters of size 3×3 and several filters equal to 32, 64, and 128. Each block had ReLU activation and batch normalization applied. Max-pooling (2×2) operation was performed in the first two blocks, and it decreased the spatial dimensions. Dropout (0.25) was used for the convolutional layers and then the flattened output went through the dense layers of 512 and 256 neurons with ReLU and Dropout (0.5) before the classification.

Training Methodology: Cross-Entropy loss was the main criterion for both models, and the Adam optimizer was employed with 0.001 as the learning rate and 1e-5 for weight decay. The ReduceLROnPlateau learning rate scheduler worked when the validation loss remained unchanged. Training lasted for a total of 15 epochs during which both the training and validation accuracies were monitored. The models switched between training and testing modes and the validation was performed without allowing gradient computation for faster operation.

Evaluation and Analysis: The assessment of performance was made using the test accuracy and confusion matrices which were generated through sklearn.metrics.confusion_matrix. The heat map visualizations via seaborn.heatmap showed the misclassification trends among the classes that looked similar, e.g., Shirt vs T-shirt/top and Coat vs Pullover. Detailed analysis on the Shirt class spotted the common confusions and pointed out the feature extraction differences of the models.

Results and Discussions: The Convolutional Neural Network (CNN) consistently achieved superior results over the Artificial Neural Network (ANN) on the Fashion-MNIST dataset. By using a better learning rate scheduler, the CNN model reached a test accuracy of 93.41% as against 88.38% for the ANN, while in the first runs without the scheduler it reached 94.27% compared to 88.94%. The ReduceLROnPlateau scheduler was found to be a good way of not only making training stable but also of slowing down learning rate when there was no improvement in validation loss, thereby preventing overfitting and making convergence faster. The accuracy of the models during different runs bore slight differences, however, the CNN continued to be the first runner showing the capability of convolutional layers in capturing spatial hierarchies.

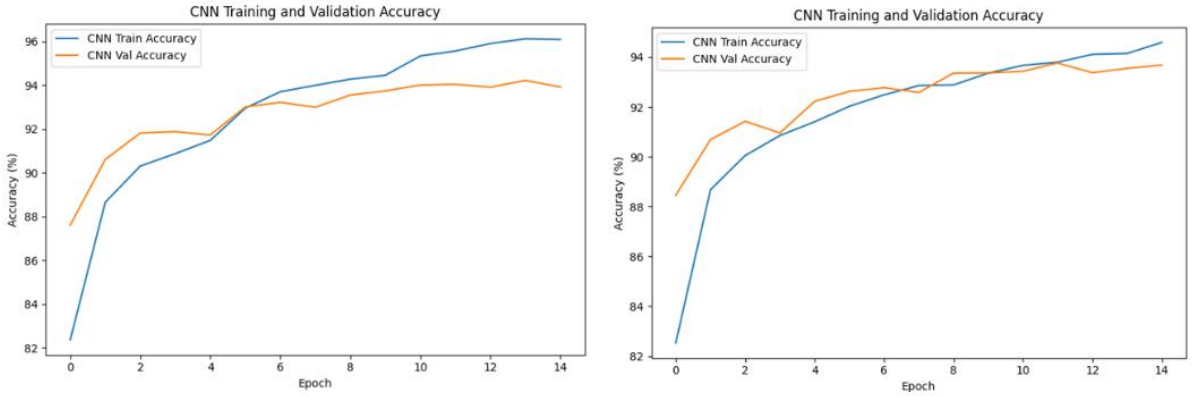

Figure: In the left diagram it is seen that a CNN was trained without any learning rate scheduling. The accuracies for both, the training and the validation reach their maximum value after 10 epochs (the maximum for validation ~93%). The right plot, using the ReduceLROnPlateau scheduler, presents a progression that is smoother but with similar peak validation accuracy, showing the positive effect of the training stability and generalization.

The CNN, in addition to providing higher accuracy, achieved a significant reduction in class confusion, which was particularly evident in visually similar categories. The number of misclassifications between Coat and Pullover dropped by 78.1% and between Shirt and Pullover by 60.4%. The gains made also affected the problem of distinguishing between the different kinds of footwear, such as Sandal, Sneaker, and Ankle boot. A close look at the Shirt class showed that the CNN was able to find more relevant spatial features, and thus there were fewer confusion errors while the ANN's use of flattened inputs resulted in information loss and misclassifications. All in all, the CNN provided a more powerful learning process, better feature extraction, and wider generalization across all configurations.

Conclusion: The CNN considerably surpassed the ANN in image classification tasks done on the Fashion-MNIST dataset. The CNN, by taking up spatial hierarchies and making the performance more stable through convolution, pooling, and dropout, not only gained higher accuracy but also less confusion among similar classes. These findings support the view that CNNs are more appropriate than other methods for visual recognition tasks that deal with spatially structured data.