Fashion Image classification using Convolutional Autoencoders and Transfer learning

Yeshwanth Thota
Faculty of science and Engineering
Yt250@student.aru.ac.uk

***Abstract* – Learning the features of lower-level dimensions for a high-level dimensional data is a challenge. Primarily the aspects of reducing dimensions and model training for collecting important features on the image instances. One such example for this type of instances are fashion Images which has varied number of features to interpret them while making predictions.**

In this paper, we Introduce a convolutional Autoencoder model and transfer learning to solve this challenge. Autoencoders are used for reconstruction of the input image instances and it learns the parameters to train the fully connected classifier neural network. Also, autoencoders are used in image denoising and dimensionality reduction. The reason choosing of Autoencoders but not PCA is that PCA is primarily used for linear transformations but autoencoders are capable for building complex nonlinear functions.

In this paper, we build a two deep neural network which includes two things. Firstly, a four-block convolutional autoencoder with following Maxpooling layer for each block. A classifier network with a fully connected three layered Dense networks for classifying Images. Secondly, VGG16 model is used with custom classifier network. The primary objective is to analyse both the models and their performance on real world fashion images.

The model is trained with 60,000 images and evaluated and tested with 10,000 images for measuring performance of the model. Both models are trained with different optimizers such as Adam, RMSprop, Adadelta and different learning rates. The best performance is achieved in autoencoder model with training accuracy 99.64% and test accuracy of 92.34%. The experimental analysis of results obtained from using different optimizers and learning rates are given in detail in this paper.

***Keywords* – Convolutional neural network, deep learning, Convolutional Encoder, Autoencoders, transfer learning.**

I INTRODUCTION.

Artificial Intelligence (AI) is transforming the fashion industry in composing, Manufacturing, and providing diversified and personalized recommendations for the customers and increasing the sales in industry. (Workman, 2020) examines the customer attitudes and purchasing intention towards AI devices. A conceptual model is designed and trained according to the customers attitudes and purchase intentions.

Application of AI in fashion industry would be a best solution for business. It provides customers varied personalised choices from large collections and filters based on their likelihood of shapes and colours. The results obtained from this can be used for manufacturing the creative designs to improve the sales. This also gives the companies to design new styles and improve their brand value. Machine learning algorithms has the capability of remembering the shopper's history, their events on the web pages and analyse this information for giving next best recommendation to the customer. Applications of AI can be applied to small retailers, and to retailers with high customer count. Autoencoders in classifying the fashion data plays a key role in image reconstruction and feature extraction. Autoencoder mainly comprised of three parts. Firstly a Encoder module which compress the given input image data into an encoded form which is many times smaller in magnitude when compared to original input image data. Secondly, the bottleneck module which has the knowledge obtained from the encoder in compressed format. This part in the network is one of the important modules in the Autoencoder

architecture. This module prevents the unnecessary information to flow through decoder from encoder. Finally, Decoder module that decompresses the knowledge representations and reconstructs the data back to its original form from its encoded format.

II LITERATURE SURVEY

(Baldi, 2012) provides the fundamentals and the role that autoencoders plays in deep learning architectures. Also describes the mathematical framework for the study of nonlinear and linear autoencoders. Baldi gives an overview of different types of autoencoders and their learning complexities. (Welling, 2019) provides extensions to variational autoencoders and its frameworks for deep learning. The traditional autoencoder is an ANN which focus on reproducing the same input as output. (Creswell, 2019) suggests the adversarial autoencoders for denoising and regularisation. It also gives an analysis for denoising to be incorporated in the training of autoencoders. (Manning-Dahan, 2018) provides PCA exploration and autoencoders for image data clustering and finds the predicting capability of each technique when seen with multivariate logistic regression. The reconstruction and classification errors are compared using the fashion image dataset. (Spigler, 2020) gives a unique way to calculate the confidence metric using the reconstructing error of the denoising autoencoders and provides how it is correctly finding the areas in the input that is close to training distribution. (Rezaabad, 2020) proposed a simple variational autoencoders that results meaningful feature representations of input images. This approach is combined with information maximization to increase the inference in autoencoders using the information techniques.

III DATASET

The fashion Mnist (Modified National Institute of standards and Technology database) is an open-source dataset used in this project. Fashion Mnist is a dataset of 70,000 fashion images in total from ten categories. Each image is of 28 * 28 size, and all are grayscale images. Fashion Mnist is the replacement dataset for original Mnist dataset. This dataset can be either downloaded or can be loaded directly from TensorFlow or keras framework to your editor or Jupiter notebook. These 70,000 fashion images are divided into two datasets. One is for training the model called Train dataset of total 60,000 images and for testing the model called test dataset of total 10,000 images. The ten fashion image categories we classify from this dataset are Sandals, Shirt, Trousers, T-shirt/Top, Pullover, Dress, Coat, Sneaker, Bag, Ankle Boot. Some of the example images of this dataset are shown in below figure.

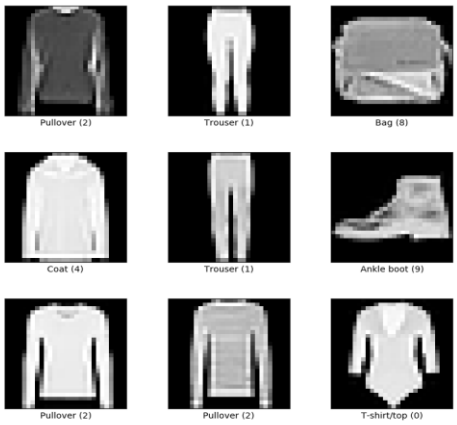

Figure 1: (Vollgraf, 2017) Sample Fashion Mnist Images.

IV DATA PREPROCESSING

The images are grayscale with pixel values between 0 to 255 and the dimensions of the images are 28 * 28. Before fitting the model with this data, pre-processing of data is done that includes converting every image from training and testing set into 28 * 28 * 1 dimensional images. We must convert the training and testing Integer NumPy arrays into float32 form. The next step is to normalise the pixel values into the range between 0 to 1 inclusively. Also, we rescale the train and test data with maximum pixel value of test and train data so that the maximum value for both training and testing data should be rescaled to 1. Good Partitioning of dataset into train and test dataset is very important. We follow a benchmark approach in partitioning the data

into 80% for training and 20% for testing the model.

V MODEL ARCHITECTURE

The model architecture is comprised of two steps. 1) Building Autoencoder for reconstructing the images, and 2) a Dense neural network for classifying the images. Firstly, building a Autoencoder includes two things Encoder and Decoder. Encoder part of Autoencoder includes four convolutional blocks with each block having convolutional layer followed by pooling layer and batch normalization layers for first two convolution blocks. The activation function used is ReLU for every convolution block and padding is 'same'. The kernel size used for the four blocks are 3 * 3 size. And filters are used for four convolutional blocks with pool-size of 3 * 3. Decoder part of the autoencoder has three convolutional blocks followed by a batch normalization and up sampling layers for second and third convolutional blocks. The same ReLU activation is used in decoder. The pool size is 3 * 3. The last layer of the decoder has one filter with size 3 * 3 that will reconstruct the input back. Down sampling and up sampling the input is done two times.

The model is compiled with parameters including 'mean-squared-error' as loss function and 'adam' as optimizer. The is trained by using fit() function in keras. The model is trained for 20 epochs and 1 as verbose to display the console information. As the task is to classify the fashion mnist images by using the above trained encoder part of the autoencoder, we save the encoder weights to use in classification task. Then we add few fully connected Dense layers to the encoder part of the autoencoder to achieve the task of classifying the fashion mnist images. The fully connected layers include Flatten layer and two densely connected layers with ReLU activation function. The last layer of the network is output layers with ten nodes and SoftMax as activation function. SoftMax activation function outputs a vector of values which are probabilities of the target classes which sum up to one when added. Then we get the maximum value from this vector of values which gives our target class. The full network is available to perform the classification task. We make the layers of encoder part False because it is already trained, and we train the fully connected layers. As a final step we train the full model and evaluate, test the performance of the model. Some real time images are used in making individual predictions on the model in which the results are positives as true value. Second model is vgg16 model with custom classifier network to classify the images. The results of the models on real time images are shown below.

Figure: Real time fashion image (amazon, 2020)

VI EXPERIMENTS & RESULTS

	T-shirt/top	Trouser	Pullover	Dress	Coat	Sandal	Shirt	Sneaker	Bag	Ankle boot
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Figure: Autoencoder model prediction result

	T-shirt/top	Trouser	Pullover	Dress	Coat	Sandal	Shirt	Sneaker	Bag	Ankle boot
0	0.000055	9.364586e-07	0.000003	9.273665e-08	9.598078e-07	0.000369	0.000002	0.000032	0.999537	3.145106e-11

Figure: vgg16 model prediction result.

The above two results shows that both the model's predictions are true with respect to their actual values. Any real-world fashion image belonging to these ten classes can be predicted using the proposed two models. The below graph shows both models accuracy and loss curves and confusion matrix. The best performance for autoencoder model is training

accuracy as 99.48% and testing accuracy of 92.02%.

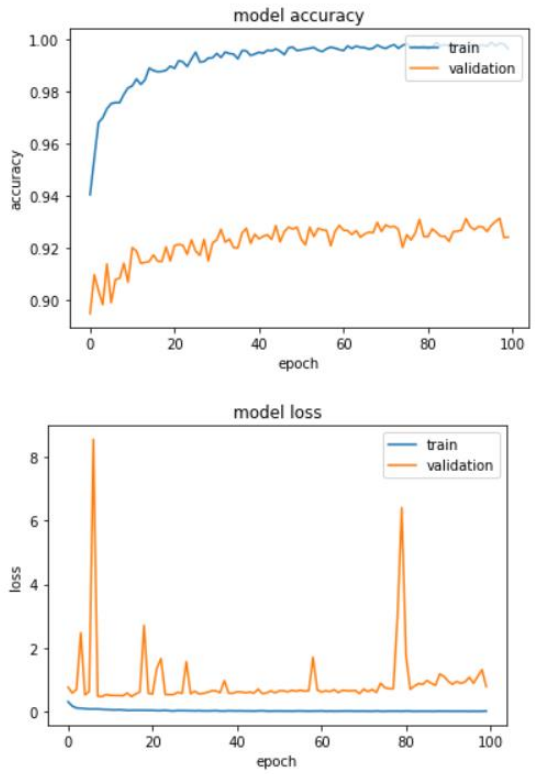

Figure: Autoencoder model accuracy and loss graph.

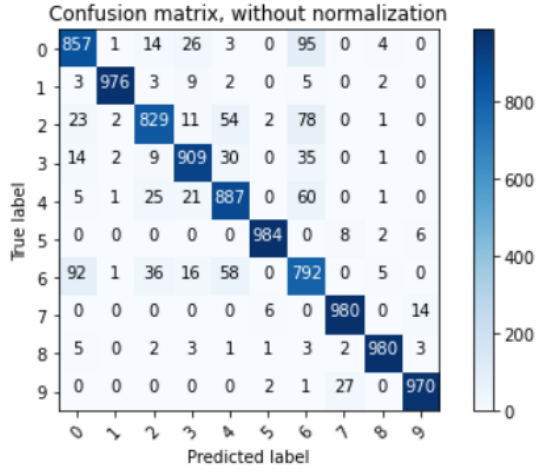

Figure: Confusion matrix

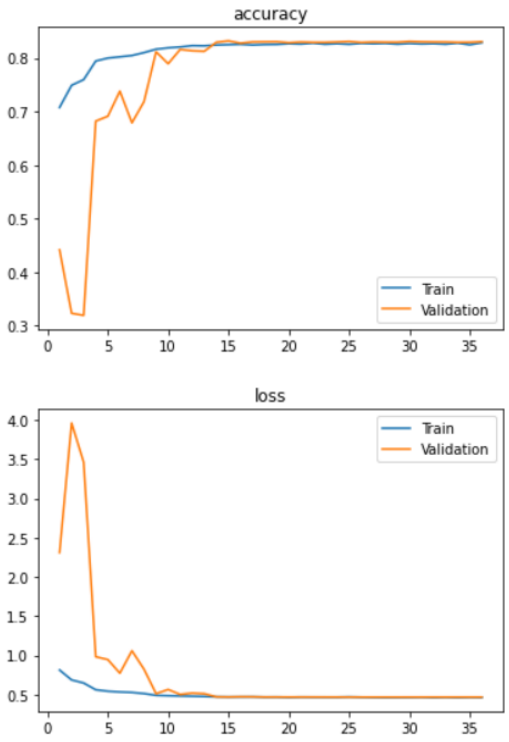

Figure: vgg16 model accuracy and loss graph.

The best performance in vgg16 model is training accuracy with 83.87% and testing accuracy is 82.02%.
Both models' performance from the above graph indicates that the models are generalised with both training and testing datasets and actual and predicted outcomes are same in predicting real time fashion images.

VII CONCLUSION

In this paper, we have discussed the challenge involved in classifying the fashion images and explained how autoencoders can solve this problem. About Dataset and their pre-processing involved in this also discussed. Then both autoencoder model and transfer learning model and their architectures are given. Finally, the experimental results with different optimizers and hyper parameters and their results are given. The real time image predictions are done on both the models and their results are shown.

VIII FUTURE WORK

Collecting large amount of real time fashion images from different ecommerce websites and other sources with different dimensions and properties and testing these images with the model will be future work because dealing with different higher dimensional images will be the challenge due varied hidden features in the images which is difficult for models to assess and predict the results.

References

amazon, 2020. *amazon.* [Online]
Available at:
https://www.amazon.co.uk/Handbags-Leather-Capacity-Handbag-Shoulder/dp/B07GSZCZ8Q?th=1
[Accessed friday march 2022].

Baldi, P., 2012. *Autoencoders, Unsupervised Learning, and Deep Architectures.* Bellevue, Washington, USA, PMLR.

Creswell, A. a. B. A. A., 2019. Denoising Adversarial Autoencoders. *IEEE Transactions on Neural Networks and Learning Systems,* Volume 30, pp. 968-984.

Manning-Dahan, T., 2018. PCA and Autoencoders. *Montreal: Concordia University, INSE,* Volume 6220.

Rezaabad, A. L. a. V. S., 2020. Learning Representations by Maximizing Mutual Information in Variational Autoencoders. *2020 IEEE International Symposium on Information Theory (ISIT),* Volume {}, pp. 2729-2734.

Spigler, G., 2020. Denoising Autoencoders for Overgeneralization in Neural Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Volume 42, pp. 998-1004.

Vollgraf, H. X. a. K. R. a. R., 2017. *TensorFlow.* [Online]
Available at:
https://dblp.org/rec/bib/journals/corr/abs-1708-07747
[Accessed Wednesday March 2022].

Welling, D. P. K. a. M., 2019. An Introduction to Variational Autoencoders. *CoRR,* Volume abs/1906.02691.

Workman, Y. L. a. S.-H. L. a. J. E., 2020. Implementation of Artificial Intelligence in Fashion: Are Consumers Ready?. *Clothing and Textiles Research Journal,* Volume 38, pp. 3-18.