Fashion and Apparel Classification

[1]Riddhi Pawar
Department of Information Technology
Vishwakarma Institute of Information Technology
Pune, India
Riddhi.22020020@viit.ac.in

[2]Anand Shirole
Department of Information Technology
Vishwakarma Institute of Information Technology
Pune, India
Anand.21910161@viit.ac.in

[3]Namrata Thakur
Department of Information Technology
Vishwakarma Institute of Information Technology
Pune, India
namrata.22020010@viit.ac.in

[4]Ishank Sharma
Department of Information Technology
Vishwakarma Institute of Information Technology
Pune, India
ishank.21910221@viit.ac.in

[5] Ekta Mulkalwar
Department of Information Technology
Vishwakarma Institute of Information Technology
Pune, India
ekta.21910986@viit.ac.in

[6]Pravin Futane
Department of Information Technology
Vishwakarma Institute of Information Technology
Pune, India
pravin.futane@viit.ac.in

***Abstract*- In this research, we have attempted to solve the challenge that the e-commerce fashion business is facing. The issue is that the customer may not always know the correct keywords to use when searching for or describing the item he is looking for. To address this problem, a deep learning-based Convolutional Neural Network (CNN) model was created to classify fashion apparel images. The model was trained on the Fashion Product Images (Small) dataset which consists 44k-image. The model was able to reach an accuracy of 85 percent. After that, we deployed this model into a fastapi web application that can categorize various apparel images uploaded by users.**

Keywords- neural networks, multiclass classification, convolutional neural network, fashion image classification

I. INTRODUCTION

Deep learning models are frequently used to classify problems like image classification, object detection, and many more. Because it is good at dealing with image classification and recognizing difficulties and has improved the accuracy of many machine learning tasks. The convolution neural network (CNN) has indeed evolved into a powerful and widely used deep learning model. It is frequently used in deep learning to analyze visual pictures. The CNN model is made up of several convolutional layers.

Apparel in many cultures reflects characteristics such as age, social status, lifestyle and gender. So, apparel is an important descriptor in identifying humans. In an unlabeled image, predicting the clothing details can facilitate the discovery of the most similar fashion items in an e-commerce database. Similarly, classification of a user's favorite fashion images can drive an automated fashion stylist, which would provide outfit recommendations based on the predicted style of a user. Depending on the particular application of fashion classification, the most relevant problems to solve will differ. There are generally two categories of photographs. The first displays products against a plain background. The second depicts a person or parts of a person wearing the products such as pants, t-shirt, shoes, and belt. As a person wearing various products, it introduces noise whereas products in front of plain background lowers the semantic noise in the photos as makes it easy to give single label.

This paper aims to solve the potentially challenging problem faced by the e-commerce fashion industry. Whenever a customer wants to buy a product typically, he searches for it using text-based search. But many of the times, the customer may not know the right keywords to search or describe the item he is looking for. This limits customer's ability to search for a specific product and results in decrease in demand. This problem can be solved using visual search which allows customers to search for a specific product using its image instead of keywords. Customers can upload the product's image to the visual search engine and get its details and similar products. This can help customers while placing orders online or offline i.e., purchasing from nearby shops when he may not know the right keywords but has a visual impression of the product he wants to buy. The customers can easily mimic their favorite influencer/celebrity styles from social media, movies, etc. This project focuses on the classification of fashion and apparel based on its types and

attributes. Among different techniques for image classification ranging from image processing to machine learning approaches, this project uses Convolutional Neural Network (CNN) model. It also uses transfer learning method to train the neural network. Then the trained model is deployed on fastapi based web application. The distinct feature of this project is that, after uploading the image it displays the exact name and related keywords of the product. This helps users to search for the product using its name or related keywords on multiple online platforms including those not having visual search feature as well as nearby shops.

II. Literature Survey

Among different machine learning techniques that are previously used for the purpose of image classification includes custom neural network models and transfer learning. Many architectures and approaches were presented such as GoogLeNet [1], Deep Residual Networks (ResNets) [2] or the Inception Architecture [1]. A survey includes Deep learning and CNN which is fully surveyed in [3]. Many CNN architectures have been used in image classification: LeNet [4], Alex Net [5], Google Net [6], VGGNet [7] and ResNet [8].

K. Chatfield *et. al.* published a study in which they evaluated CNN based methods and traditional shallow feature encoding methods for image classification. The PASCAL VOC-2007 dataset was used to compare the performance. It demonstrated that deep architectures outperform the traditional shallow feature encoding methods [9].

Fengzi Li *et. al.* published a paper proposing neural network models for image classification and image search. They used the Fashion Product Images (Small) dataset having 44k product images. For image classification transfer learning technique was used with pre-trained models. The VGG-19 model performed best for classification with accuracy of 87.1% for gender & master category, 95.7% for sub-category and 84.1% for article type. CNN-based and ResNet-50 based autoencoders and cosine similarity was used to search similar images [10].

Alexander Schindler *et. al.* presented a study of methods to classify fashion and apparel images using different CNN architectures. The proposed model for detecting persons provided 91.07% accuracy. The pre-trained and fine-tuned model produce 88% accuracy for gender prediction and performed better than clean models by 5.9% for VGG-16, 7.9% for InceptionV3 model [11].

Alex Krizhevsky *et. al.* published a paper in which they proposed a deep convolutional neural network to classify 1.2 million images into 1000 classes of ImageNet LSVRC-2010 contest. They achieved top-1 and top-5 error rates of 37.5% and 17.0% respectively [12].

A. S. Henrique *et. al.* published a paper that proposes four convolutional neural network models for classifying garments images from Fashion-MNIST dataset. Comparing the classification results, highest accuracy is 99.1% using dropout technique [13].

Greeshma K V *et. al*. published a paper that explores impact of hyper-parameter optimization methods and regularization techniques on deep neural networks trained on Fashion-MNIST dataset. They got maximum 93.99% accuracy by tuning hyperparameters such as optimizers, batch size, epochs and regularization methods like image augmentation, dropout [14].

S. Bhatnagar *et. al.* published a paper that proposes three CNN models for classification of fashion article images trained on Fashion-MNIST dataset. They have used batch normalization and residual skip connections for ease and acceleration of learning process. The model reports 2% improved accuracy over the other literary systems [15].

Mohammed Kayed *et. al*. published a paper that proposed convolutional neural network based LeNet-5 architecture for image classification. They used Fashion-MNIST dataset having 70k images of 10 categories. The LeNet-5 model achieved accuracy over 98% [16].

Shu Shen published a paper which proposed Long Short-Term Memory Networks to build a model for classifying Fashion-MNIST dataset images. The results showed that the LSTM model can fit the dataset with 88.26 % best precision [17].

Yue Zhang published a paper which evaluates the performance of four convolutional neural networks that are LeNet-5, AlexNet, VGG-16 and ResNet. Among these ResNet was best suited for classifying Fashion-MNIST dataset [18].

Chao Duana *et. al.* published a paper in which they used VGG-11 network with batch normalization layer after the pooling layer for classifying fashion images in fashion-MNIST dataset. The classification accuracy is 91.5% [19].

Table 1. Literature Survey Consolidation

Sr. No.	Title	Algorithm	Accuracy
1	Neural Networks for Fashion Image Classification and Visual Search	VGG-19	95.7%
2	Fashion and Apparel Classification using Convolutional Neural Networks	CNN	88%
3	Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture	LeNet-5	98%
4	Image Classification of Fashion-MNIST Dataset Using Long Short-Term Memory Networks	LSTM	88.26%
5	Image Classification of Fashion-mnist Data Set Based on VGG Network	VGG-11	91.5%

III. METHODOLOGY

A. *Dataset*

For collecting data to train the model for multiclass classification, three main options are available. First is to use readymade data from Kaggle or buy it from third party vendors. Second option involves collecting and annotating data manually. Third one is to write web scraping scripts to collect data i.e. images from internet.

This paper uses Fashion Product Images (Small) dataset [20] which is readily available on Kaggle. This dataset consists of 44k high resolution color images as opposed to Fashion MNIST dataset which has greyscale images. This makes the Fashion Product Images dataset suitable for real-life business problems. This dataset comprises of 44,441 images in jpg format of size 80 x 60 pixels in 3 color channels identified by numeric-id. This numeric-id is mapped to the multiple label attributes describing the product in styles.csv file. These attributes include Gender, Master Category, Sub Category, Article Type, Base Color, Season, Year, Usage, Product Display Name etc.

B. *Data Preprocessing*

The Fashion Product Images (Small) dataset is highly unbalanced. This gives rise to the need of data pre-processing to address the data imbalance in this dataset. For training purpose, among all the available attributes of images, Sub Category attribute is used to classify the fashion product images. To avoid biased results by machine learning algorithms while classifying minority classes and false impression of high accuracy of the model, the classes with lesser images are removed.

For further balancing the dataset, Data Augmentation techniques are used. Data Augmentation techniques are used to increase the amount of data by adding slightly modified copies of already existing data or newly created synthetic data from the existing data. It acts as a regularizer and helps reduce overfitting when training a machine learning model [21]. It includes cropping, rotation, zoom, horizontal or vertical flipping and shifting images. ImageDataGenerator class [22] in Keras is extremely helpful for generating batches of tensor image data with real- time data augmentation.

Table 2. Data after pre-processing

Sub-categories	**No. of labels**
Kurti	153
Saree	426
Bottomwear	409
Topwear	416
Loungewear and Nightwear	443
One Piece Dress	460
Total	**2307**

C. *Convolutional Neural Network*

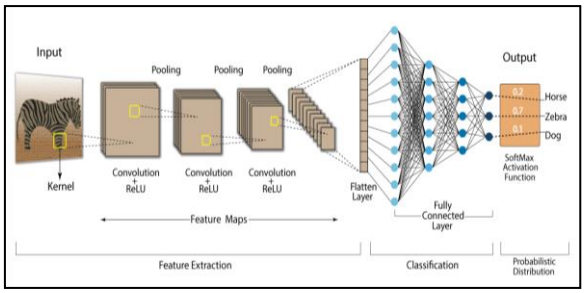

Figure 1. Convolutional Neural Network Working [23]

CNN stands for Convolutional Neural Network. It is a deep learning algorithm that identifies feature set in an input image using kernels. CNN model is used to classify fashion apparel. The key advantage of using CNN is that it automatically learns and extracts features from the input data. The Softmax regression allows to distribute the probability of apparel categories among 6 different categories.

Convolutional layer. It is also called as feature extractor layer. It extracts the features of an image received from the input layer. Convolution operation involves calculating the dot product between a local region of the input image and the filter. This is repeated for the whole image. Mostly, ReLU activation is used in this layer.

Pooling layer. It is used to reduce spatial volume of images while preserving the important characteristics. This layer is mainly used between two convolutional layers and helps to lower the computational power required for data processing. Spatial pooling is also known as down sampling as it reduces size of features extracted by convolution layers. Pooling can be of three different types. First is Max pooling which selects the maximum value. Second is Average pooling, it selects the average of all values in the feature map. Third one is Sum pooling which sums all elements in the feature map.

Fully Connected Layer. This is always the last layer of a neural network. It is not the characteristic of CNN. Fully connected layers are used to classify image. It returns a vector as output. The vector consists of probability of the input image to belong to a specific class. The class with high probability value is the predicted class of the image.

IV. Experimentation and Results

A. *Model Analysis*

The proposed CNN model's first layer is the convolutional layer, Conv2D which receives the input of shape 60x60x3. This layer has 32 filters, kernel size as 3 with ReLU activation. Its output is received by MaxPooling2D layer, which down samples it. Again, a convolutional layer with configuration similar to previous one and a pooling layer are applied. Then the output is flattened and supplied to the dense layers. Finally, the model contains six dense layers. The output dimensions of Dense layers are 32, 64, 128, 256, 256 respectively with a ReLU activation function. The last dense layer has 6 as output dimension and applies softmax activation to predict multinomial probability distribution

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 58, 58, 32)	896
max_pooling2d (MaxPooling2D)	(None, 19, 19, 32)	0
conv2d_1 (Conv2D)	(None, 17, 17, 32)	9248
max_pooling2d_1 (MaxPooling2	(None, 5, 5, 32)	0
flatten (Flatten)	(None, 800)	0
dense (Dense)	(None, 32)	25632
dense_1 (Dense)	(None, 64)	2112
dense_2 (Dense)	(None, 128)	8320
dense_3 (Dense)	(None, 256)	33024
dense_4 (Dense)	(None, 256)	65792
dense_5 (Dense)	(None, 6)	1542

Total params: 146,566
Trainable params: 146,566
Non-trainable params: 0

Figure 2. Convolutional Neural Network Model Architecture

B. *Hyperparameter Tuning*

Optimizers. Optimizers are algorithms responsible for minimizing losses and generating the most accurate results possible by adjusting the weights and learning rate of the neural network. The CNN model in this paper uses Adam optimizer for training.

Batch size And Number of Epochs. The batch size determines how many samples must be processed before the internal model parameters are updated. It is the number of samples from the training dataset that are used to estimate the error gradient. This study uses Minibatch Gradient Descent, with batch sizes more than one but lesser than the total number of samples in the training dataset. The number of

epochs is the number of times the learning algorithm will loop through the full training dataset. An epoch is a single iteration across the full training dataset. In an epoch, there are one or more batches. The CNN model is trained with a batch size of 32 for 100 epochs with callbacks. ReduceLROnPlateau, EarlyStopping and ModelCheckpoint are among the callbacks applied to these models to avoid model overtraining.

Activation Function. The activation function determines whether or not a neuron should be activated by calculating a weighted sum and then adding bias to it. The weighted sum of the input is converted into an output from a node or nodes in a layer of the network using an activation function. In this paper ReLU activation is applied to the hidden layers, while Softmax to the output layer.

C. *Performance Metrics*

The accuracy metric is used to evaluate the model's performance across all classes. It is calculated using the ratio of the number of correct predictions to the total number of predictions made by the model. It is useful when all of the data classes are roughly balanced.

D. *Results*

The data is split into 80-20 for training and validation purpose. With hyperparameter optimization and regularization techniques used, the CNN model was capable of attaining accuracy of 85% during testing the model. The accuracy and loss during the training and validation of the CNN model is as shown in Figure 3 And Figure 4 respectively.

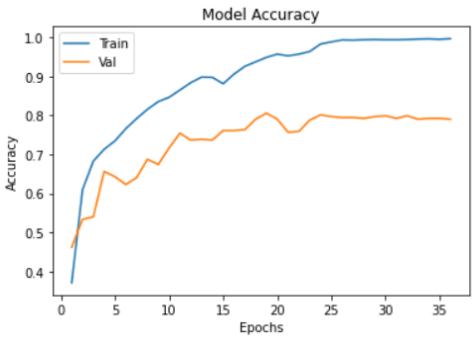

Figure 3. Training and Validation Accuracy

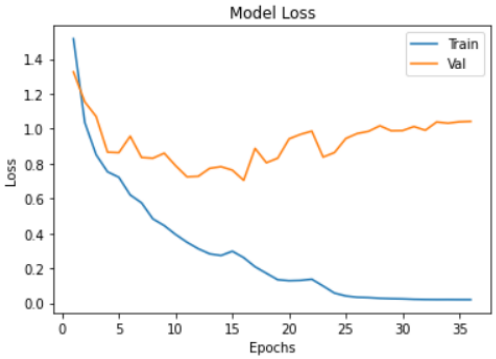

Figure 4. Training and Validation Loss

V. Conclusion and Future Scope

Machine learning models have a particularly exciting use case that is visual search, which overcomes the problem of looking for fashion products when the shopper does not know the correct terms. The image classification feature allows you to classify a given image and then display the apparel's specific name and relevant keywords. It improves the shopping experience by allowing users to find out the label and category of clothing items by simply clicking a photo of the item.

Machine learning models can also help sellers improve their experience while listing products on the site. Sellers can upload images of their fashion items, and image-to-text machine learning algorithms will automatically generate appropriate tags to categorise them. This can help to eliminate product labelling mistakes, which can have a negative impact on demand because the products aren't shown appropriately in search results. To accomplish this, CNN models must be integrated with natural language processing (NLP) techniques like Word2vec to predict text content from visual data and characteristics.

References

[1] Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, 2015

[2] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, 2016

[3] Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Asari, V. K., A state-of-the-art survey on deep learning theory and architectures. Electronics, 8(3), 292. , 2019

[4] LeCun, Y., Bottou, L., Bengio, Y., Haffner P., Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324. , 1998

[5] Krizhevsky, A., Sutskever, I., Hinton, G. E., Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105), 2012

[6] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Rabinovich, A., Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-9), 2015
[7] Simonyan, K., Zisserman, A., Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. ,2014
[8] Li, X., & Cui, Z., Deep residual networks for plankton classification. In OCEANS 2016 MTS/IEEE Monterey (pp. 1-4). IEEE, 2016
[9] K. Chatfield, K. Simonyan, A. Vedaldi, A. Zisserman, Return of the Devil in the Details:Delving Deep into Convolutional Nets, 2014
[10] Fengzi Li, Shashi Kant, Shunichi Araki, Sumer Bangera, Swapna Samir Shukla, Neural Networks for Fashion Image Classification and Visual Search, 2020
[11] Alexander Schindler, Thomas Lidy, Stephan Karner, Matthias Hecker, Fashion and Apparel Classification using Convolutional Neural Networks, 2018
[12] Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, 2012
[13] S. Henrique, A. Fernandes, R. Lyra, V. Leithardt, S. D. Correia, P. Crocker, R. Dazzi, Classifying Garments from Fashion-MNIST Dataset Through CNNs, 2021
[14] Greeshma K V, Sreekumar K, Hyperparameter Optimization and Regularization on Fashion-MNIST Classification, 2019
[15] S. Bhatnagar, D. Ghosal, M. H. Kolekar, Classification of fashion article images using convolutional neural networks, 2017
[16] Mohammed Kayed, Ahmed Anter, Hadeer Mohamed, Classification of Garments from Fashion MNIST Dataset Using CNN LeNet-5 Architecture, 2020
[17] Shu Shen, Image Classification of Fashion-MNIST Dataset Using Long Short-Term Memory Networks, 2018
[18] Yue Zhang, Evaluation of CNN Models with Fashion MNIST Data, 2019
[19] Chao Duana, Panpan Yinb, Yan Zhic, Xingxing Li, Image Classification of Fashion-mnist Data Set Based on VGG Network, 2019
[20] Fashion Product Images (Small), https://www.kaggle.com/paramaggarwal/fashion-product-images-small
[21] Data Augmentation, https://en.wikipedia.org/wiki/Data_augmentation
[22] ImageDataGenerator in Keras, https://www.tensorflow.org/api_docs/python/tf/keras/preprocessing/image/ImageDataGenerator
[23] Convolutional Neural Network Layers Working, https://developersbreach.com/convolution-neural-network-deep-learning/