VISHWAKARMA INSTITUTE OF INFORMATION TECHNOLOGY, PUNE

A Report on

AI for fashion and Apparel Classification

Guided By:

Dr. Pravin Futane sir

Prepared by:

Riddhi Pawar (332070)

Namrata Thakur (332068)

Anand Shirole (332059)

Ishank Sharma (332057)

Ekta Mulkalwar (332039)

Abstract:

In this project, 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. After that, we deployed this model into a fastapi web application that can categorize various apparel images uploaded by users.

Introduction:

This document is report for the Group Mini project "AI for Fashion and Apparel Classification". 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*. 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.

Background:

Image classification is the process of categorizing and labeling groups of pixels or vectors within an image based on specific rules. The categorization law can be devised using one or more spectral or textural characteristics.

Two general methods of classification are:

1) Unsupervised classification method- It is a fully automated process without the use of training data. Using a suitable algorithm, the specified characteristics of an image is detected systematically during the image processing stage. The classification methods used in here are 'image clustering' or 'pattern recognition'. Two frequent algorithms used are called 'ISODATA' and 'K-mean'.

2) Supervised classification method- It is the process of visually selecting samples (training data) within the image and assigning them to pre-selected categories in order to create statistical measures to be applied to the entire image. 'maximum likelihood' and 'minimum distance' are two common methods to categorize the entire image using the training data. For example, 'maximum likelihood' classification uses the statistical characteristics of the data where the mean and standard deviation values of each spectral and textural indices of the image are computed first. Then, considering a normal distribution for the pixels in each class and using some classical statistics and probabilistic relationships, the likelihood of each pixel to belong to individual classes is computed. Finally, the pixels are labeled to a class of features that show the highest likelihood.

Implementation:

In this project we have used Fashion Product Images (Small) dataset 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.

Overall project splits into following components:

1. Data Preprocessing -
   a) 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.
   b) For training purpose, among all the available attributes of images, Sub Category attribute is used to classify the fashion product images.
   c) 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.
   d) 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.
   e) It acts as a regularizer and helps reduce overfitting when training a machine learning model. It includes cropping, rotation, zoom, horizontal or vertical flipping and shifting images.
   f) ImageDataGenerator class in Keras is extremely helpful for generating batches of tensor image data with real- time data augmentation.

2. CNN model use for classifying fashion apparel -
   a) The first layer of this model i.e. convolutional layer, receives the input of shape 60x60x3. This layer has 32 filters, kernel size as 3 with ReLU activation.
   b) The output is received by MaxPooling2D layer.
   c) Further, 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.
   d) Finally, the model contains six dense layers. The output dimensions of dense layers are 32, 64, 128, 256, and 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.

3. Hyperparameter turning -
    a) 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.
    b) The batch size determines how many samples must be processed before the internal model parameters are updated.
    c) 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.
    d) 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.

Challenges faced and their solutions:

The datasets for fashion available on internet are all either not well designed or are incomplete, it took us a long time to find perfect dataset for the fashion and apparel classification.

The Fashion Product Images (Small) dataset is highly unbalanced. Which needed of data pre-processing to address the data imbalance in this dataset. The classes with lesser images are removed to avoid biased results by machine learning algorithms while classifying minority classes and false impression of high accuracy of the model. For further balancing the dataset, Data Augmentation techniques were 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.

Future work:

As we are going to continue explore this topic, In future we want to integrate CNN models with natural language processing (NLP) techniques like Word2vec to predict text content from visual data and characteristics. This 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.

Conclusion:

Hence, we have implemented a deep learning-based Convolutional Neural Network (CNN) model for classifying fashion apparel images.