Fashion Recommendation System

*VGG 16 Architecture

Ajay Kumar Kotha
ECE
Hyderabad, India
2110040024@klh.edu.in

Srujan Kurma
ECE
Hyderabad, India
2110040042@klh.edu.in

***Abstract*—Developed for the e-commerce domain, FashionVision introduces a pioneering computer vision system tailored specifically for the detection and classification of clothing items in images. Leveraging the robust VGG-16 architecture, the system adeptly tackles both detection and classification tasks. Through a combination of a subset of the DeepFashion dataset and manually collected data for training and testing, FashionVision accurately pinpoints clothing items using bounding boxes and proceeds to classify their colors. Experimental results substantiate the effectiveness of the VGG-16 architecture, validating its efficiency and proficiency in clothes detection and classification. This study underscores the transformative potential of deep learning methodologies in reshaping workflows within the fashion industry, ultimately enhancing the e-commerce experience for users globally.**

***Index Terms*—FashionVision,VGG-16,Validation**

I. Introduction

In the dynamic realm of fashion, the fusion of computer vision and machine learning has ushered in transformative breakthroughs, particularly in the realm of personalized style exploration. This conference paper presents a pioneering effort in constructing a Fashion Recommendation System leveraging image features, harnessing the robust capabilities of the Python programming language. Our endeavor begins with the meticulous curation of a diverse fashion dataset, encompassing a myriad of colors, patterns, and styles, laying the groundwork for an enriched recommendation model. Emphasizing the standardization of image formats and resolutions ensures a cohesive dataset conducive to streamlined preprocessing, a pivotal step in preparing images for feature extraction. Central to our methodology is the adoption of a pre-trained Convolutional Neural Network (CNN) model—VGG16, capitalizing on its proficiency in extracting intricate feature representations. Subsequent stages involve image processing, metric definition for similarity computation, and the strategic ranking of images based on their resemblance to an input, ultimately culminating in a comprehensive function encapsulating the entire recommendation pipeline. This conference paper invites readers to delve into the intricate synergy between computer vision and machine learning, offering insights into the innovative domain of personalized fashion recommendation systems. By sharing our methodologies and insights, we aim to contribute to the collective advancement of fashion technology, fostering a more immersive and tailored fashion discovery experience for users worldwide

II. Methodology

Building a fashion recommendation system using image features involves several key steps, leveraging both computer vision and machine learning techniques. Below is a detailed process you can follow to build a fashion recommendation system using image features:

- Assemble a diverse dataset of fashion items. This dataset should include a wide variety of items with different colours, patterns, styles, and categories.
- Ensure all images are in a consistent format (e.g., JPEG, PNG) and resolution.
- Implement a preprocessing function to prepare images for feature extraction.
- Choose a pre-trained CNN model such as VGG16, ResNet, or InceptionV3. These models, pre-trained on large datasets like ImageNet, are capable of extracting powerful feature representations from images.
- Pass each image through the CNN model to extract features.
- Define a metric for measuring the similarity between feature vectors.
- Rank the dataset images based on their similarity to the input image and recommend the top N items that are most similar.
- Implement a final function that encapsulates the entire process from pre-processing an input image, extracting features, computing similarities, and outputting recommendations.

So, the process starts with collecting a dataset of images based on fashionable outfits.

A. Block Diagram

- Data Collection Module: In this module data is collected from the Kaggle dataset which consists of required details for the fashion recommendation process.
- Pre-processing Module: In this module the data is cleaned by removing missing values, redundancy duplicate values.
- Feature Extraction and Selection Module: In this module the feature are selected for the system through which we are able to recommend the fashion.
- Recommendation Module: In this module after training the collected data we will apply machine learning algorithms (VGG-16) the system will recommend the types of items that a user or consumer may prefer.

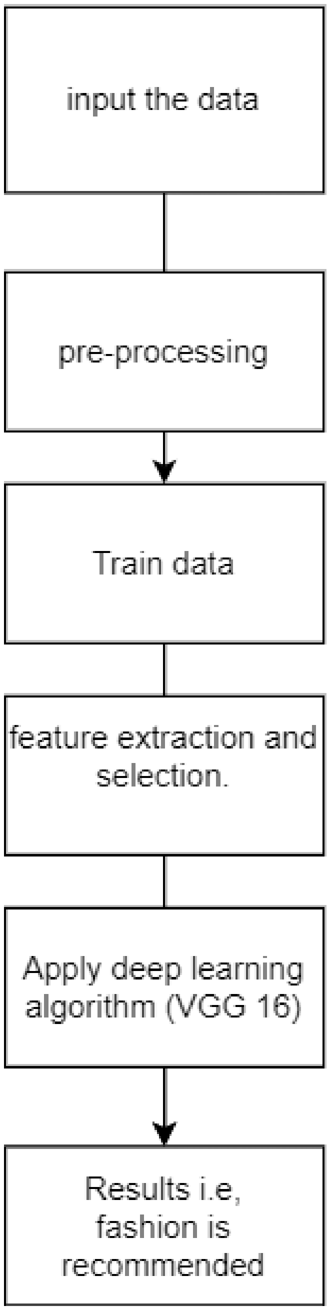

Fig. 1. Block diagram.

B. Dataset and classification

In this project, we worked with the Deep Fashion dataset, which is gathered from researchers from the Chinese Hong Kong University. It has over one million diverse trend pics and wealthy annotations with additional data about landmarks, categories, pairs etc. The dataset consists of 5 distinct types of predicting subsets that are tailor-made towards their tasks. One subset, known as Attribute Prediction, can be used for apparel category and attribute prediction. From almost 290,000 photos of 50 apparel categories and 1,000 apparel attributes, we randomly picked 18k images from different categories

Fig. 2. Fashion dataset.

and then we classified them for training and testing. The distribution of labels is presented in Figure 1.

C. Design of deep learning module

VGG 16 architecture :

A convolutional neural network is also known as a ConvNet, which is a kind of artificial neural network. A convolutional neural network has an input layer, an output layer, and various hidden layers. VGG16 is a type of CNN (Convolutional Neural Network) that is considered to be one of the best computer vision models to date. The creators of this model evaluated the networks and increased the depth using an architecture with very small (3×3) convolution filters, which showed a significant improvement on the prior-art configurations. They pushed the depth to 16–19 weight layers making it approx — 138 trainable parameters.

- The 16 in VGG16 refers to 16 layers that have weights. In VGG16 there are thirteen convolutional layers, five Max Pooling layers, and three Dense layers which sum up to 21 layers but it has only sixteen weight layers i.e., learnable parameters layer.
- VGG16 takes input tensor size as 224, 244 with 3 RGB channel
- Most unique thing about VGG16 is that instead of having a large number of hyper-parameters they focused on having convolution layers of 3x3 filter with stride 1 and always used the same padding and maxpool layer of 2x2 filter of stride 2.
- The convolution and max pool layers are consistently arranged throughout the whole architecture
- Conv-1 Layer has 64 number of filters, Conv-2 has 128 filters, Conv-3 has 256 filters, Conv 4 and Conv 5 has 512 filters.
- Three Fully-Connected (FC) layers follow a stack of convolutional layers: the first two have 4096 channels each, the third performs 1000-way ILSVRC classification and thus contains 1000 channels (one for each class). The final layer is the soft-max layer.

VGG16 is object detection and classification algorithm which is able to classify 1000 images of 1000 different categories with 92.7% accuracy. It is one of the popular algorithms for image classification and is easy to use with transfer learning.

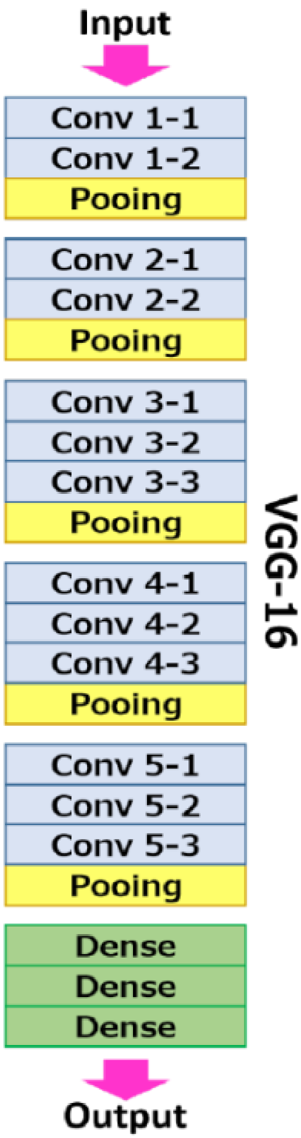

Fig. 3. VGG 16 architecture.

There are many classification algorithms or classifiers in use today. The most notably and the most implemented classifiers are Vgg-16, Vgg-19, AlexNet, BN-Inception, ResNet etc. In our system, the Vgg-16 classifier and feature extractor are implemented to solve a problem of cloth / fashion recommendation process.
The core network of our model is VGG16 as shown in Figure 4. VGG16 was projected by Simonyan, K. and Zisserman, A. who presented a convolutional neural network in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition", at the University of Oxford.". Then model is checked for top-5 accuracy on ImageNet. which contains 14 million images datasets belonging to one thousand classes and achieves the value of 92.7.

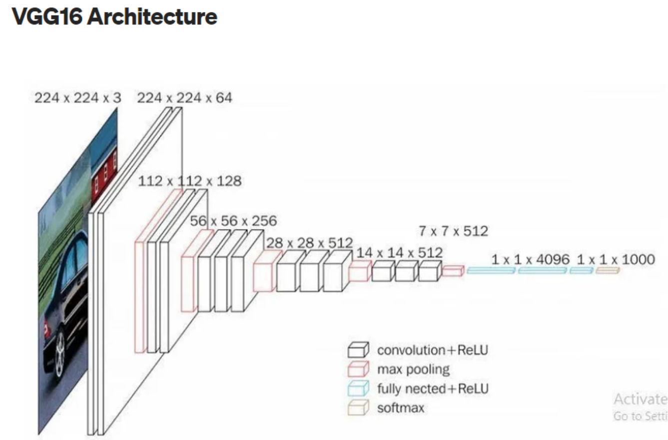

Fig. 4. VGG 16

D. Equations

$$N_i(Wi,.) \triangleq (Vi, Fi, (Ui,.)) \text{for } i = 1,2$$

Here wi = (ui,vi) are weight vectors, si(vi,.) are fully connected softmax output layers that actually perform classification and fi(ui,.) are the CNN without the last layer. They are used as a feature extractors.

E. Design of visual recommendation module

The fashion domain is a very popular playground of machine learning and computer vision. The main problem of this domain is produced by the high level of subjectivity and the semantic complexity of the features involved. Recent work has focused on a variety of approaches including attribute recognition, clothing retrieval, image generation and visual recommendation. fig 5 shows different distance measurement formulas for image feature vector similarity, and definitions for the mentioned similarity measures as presented by (Iglesias and Kastner, 2013).
First closeness measure between style dataset and client input First, we need to build a style profile for the client, which is then refined through taking at least one of the client's photographs from his/her ideal clothing objects. Then, the design vectors are entered and fostered. These vectors are then joined to shape the framework of the style profile for each individual. The component vectors prepare the information from the dataset.

- set train data = get
- for each model in the list (Vgg16) do
- for each distance metric in (Similarity) do
- train new nn model (train data, distance metric, neighbors=5)

Name	Formula
Manhattan (City Block)	$\mathbf{d_{CB}} = \sum_{i-1}^{d} \mid \mathbf{P_i - Q_i} \mid$
Euclidean	$\mathbf{d_{Euc}} = \sqrt{\sum_{i-1}^{d} \mid \mathbf{P_i - Q_i} \mid^2}$
Chebyshev	$\mathbf{d_{Cheb}} = \max_{i} \mid \mathbf{P_i - Q_i} \mid$
Hammington	Compare the first two bits in each string. If same. Record a "0", else "1"
Cosine	$\mathbf{S_{Cos}} = \frac{\sum_{i-1}^{d} \mid \mathbf{P_i - Q_i} \mid}{\sqrt{\sum_{i-1}^{d} \mathbf{P_i^2}} \sqrt{\sum_{i-1}^{d} \mathbf{PQ_i^2}}}$

Fig. 5. Distance measures available for image feature vector similarity

Presently we utilize a similitude calculation to assess the design vector of each image in the archive with the style profile lattice. This offers us a score dependent on the wide assortment of component matches - the higher the score the nearer an image is to the individual's style profile. At that point, we rank the photos arranged in their classification and show, as proposals, the pictures with the highest rank.
set test data = get feature vectors of the test data from the repository database

- for each model in range(n) do
- for each image in test data do
- extract neighbors top five from model
- store results in the database

F. output:

Fig. 6. output

our model can capture the style with high accuracy, meaning that our system achieves its purpose. It can be noticed that our system can perform for all the involved categories like pattern, style, fabric etc.

III. Applications :

Whenever the client doesn't have the foggiest idea what to search for, you can involve it as a pursuit channel or search technique. In internet business, the framework might request the sort of thing that the client or purchaser can browse. Necessities can be founded on purchasers or pre-deals. Individuals can enroll online to become individuals from the House, and everybody can send an advanced duplicate of their closet. They can purchase their garments at closeout or at a decent cost. Each client can settle in to get their number one garments without any problem.

IV. Conclusion :

The present paper presents the development of a system that recognizes fashion similar images. We accomplish this by implementing an already existing VGG 16 model with transfer learning for cloth image recognition using different libraries. For this purpose, we created a plan for collecting data and for developing the steps needed for preprocessing and cleaning up the data. We took into account features like patterns, machine, fabric, style etc. After extensive preprocessing and cleaning of data in a dataset, we constructed the model of VGG 16 to predict the features specific to these attributes and to train the models with the dataset to generate accurate predictions regarding almost all forms of images. A VGG 16 was used and implemented, with the help of this algorithm through which the system can recommend similar images This is the last test to assess if deep learning for style recovery is at a high development and can be utilized in making fashion choices.
So, this is how you can build a Fashion Recommendation System using Image Features using the Python programming language. A Fashion Recommendation System using Image Features leverages computer vision and machine learning techniques to analyze fashion items' visual aspects (like colour, texture, and style) and recommend similar or complementary products to users.

References

[1] Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J. Devin, M. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems, pp. 265–283. arXiv preprint arXiv:1603.04467.
[2] Alkhawlani, M., Elmogy, M. EL Bakry, H. (2015). Text-based, content-based, and semantic-based image retrievals: a survey. Int. J. Comput. Inf. Technol, 4, pp. 58-66.
[3] Bobadilla, J., Hernando, A., Ortega, F. Bernal, J. (2011). A framework for collaborative filtering recommender systems. Expert Systems with Applications, 38, pp. 14609-14623.
[4] Burke, R. (2002). Interactive critiquing forcatalog navigation in e-commerce. Artificial Intelligence Review, 18, pp. 245-267.