Fashion Advisory System-A Complementary Apparel Recommendation System using VGG Model

Tarun Dubey, Shruti Mundargi, Jhanavi Chadha, Mrs. Asmita Deshmukh
Department of Computer Engineering,
K.C. College of Engineering & Management Studies Research,
Kopri, Thane (E) - 400 603, India.

Abstract- *Fashion is inherently social and public. Fashion is like eating, you shouldn't stick to the same menu. We dress not only for ourselves but also for the appropriateness of the environment we are in. In order to look extra ordinary, having a unique fashion style is a common thing observed these days. People try to copy celebrities and the ongoing fashion trends. But it's not always possible to buy new clothes. Rather the idea of sustainable fashion is a better option. Reusing the same apparel, a number of times in a different style can be adopted as it also shows one's creativity. At times, it is hard to decide the proper outfit and what accessories will match with the same. Instead of having to try each and every garment it's better to have a personal advisor who can make your job pretty simple on the click of a button. We proposed a full stack web application that can readily output the fashion items due to its ability to label based on the pre-trained generalized attribute. Specific identified attributes can further be categorized as styles of clothing where further user recommendations can be made. We used a VGG model pre-trained on ImageNet with transfer learning changing the output layer to include various patterns and clothing types. Transfer learning is a method of reusing pre-trained model knowledge for another task.*

Index Terms *-Deep Convolutional Neural Network, Image Classification, Machine Learning, Transfer learning, VGG-16.*

I. INTRODUCTION

Convolutional networks have recently got a great success in large-scale image which has become possible due to the large public image repositories, such as ImageNet and high-performance computing systems, such as GPUs or large-scale distributed clusters. ImageNet Large-scale Visual recognition challenge played an important role in the advance of deep visual recognition architectures, which has served as a testbed for a few generations of images of large-scale classification systems, from high- dimensional shallow feature encodings. Manually labeling clothing styles is often subjective. By using image recognition or image classification techniques, the style recognition of apparel images can be achieved. In the field of image processing, CNNs (Convolutional Neural Networks) have been widely used and have features of the detection area, which can be combined with the recognition of clothing images. Fashion-Net and Alex-Net were used in the classification of clothing categories. This turns out to be certain that convolutional neural network is an efficient tool for the image recognition [1].

The proposed system provides recommendations and labeling of the clothing image that has been passed or uploaded on to the engine by identifying the pattern and clothing type of the image giving the highest confidence level achieved by the CNN recognition. However, deep learning models during training need enough labeled data. Hence it is nearly impossible to create a machine learning based model for a target domain which consists of very few labeled data for supervised learning. In such cases, Transfer learning would significantly improve the performance of learning. The fascinating idea behind transfer learning is to borrow labeled data or knowledge extracted from some related domains to help a machine learning algorithm to achieve greater performance in the domain of interest.

The overall idea is to help you save your time and energy by introducing you to our model which in just few steps gives you an promising cross complementary apparel products, that you can decide by looking from the recommendations displayed by our model and help yourself with the idea of what looks best with your outfit eliminating the fact of going around the shops and actually looking for the apparel that best matches your outfit.

II. BACKGROUND AND RELATED WORK

The practice of transfer learning is motivated by the fact that people can intelligently apply knowledge learned previously to solve new problems faster or with better solution. A learning method to transfer learning is the multi-task learning framework, which tries to learn multiple tasks simultaneously even when they are different. A typical approach for multi-task learning is to uncover the common (latent) features that can benefit each task [2].
Early research related to this topic was finding frequent item sets that generate match items according to purchase history (Han, Pei, and Yin 2000[13]. But as the advancements in the field increased, more efficient methods have been introduced. In 2015, the idea of Recommendation of Complementary Garments Using Ontology was proposed (Deepti Goel, Santanu Chaudhury, Hiranmay Ghosh) [14], they proposed a new way of recommendation by using a model that encoded the subjective knowledge of clothing using Multimedia Web Ontology Language.

Although, our model is more related to Image-based Recommendations on Styles and Substitutes (Julian McAuley, Christopher Targett, Qinfeng ('Javen') Shi and Anton van den Hengel, 2015), [12] where they propose a model that can recommend accessories according to human visual preferences to an image query, A system that can decide which apparel or accessories is visually suited with a particular apparel. Generally, when it comes to complementary recommendation the dataset for apparels are limited in terms of size. Using small dataset can lead to the problem of over-fitting. Thus, in order to overcome that problem, a proposal of using a larger dataset even if it is tangentially related to the goal. Their dataset included 180 million relationships with 6 million different objects. Prediction of suitability was done using Convolutional Neural Network and the calculation of probabilities and distances between items of datasets are done using Weighted Nearest Neighbors and distance between each item was calculated using Mahalanobis Distance. To train the model, dataset was divided into several categories (books, men's clothing etc.). Our dataset has also been structured in the a similar way, difference being that our dataset is divided on the basis of pattern, clothing type and occasion (Floral, beach wear, skirts etc.).

SenseFashion [16] was another model we referred to while working on our project. They used several trained neural network models in order to identify images. VGG model has been trained on ImageNet and the image recognition has been done based on pattern, fabric and clothing type. If the image matches a specific style set of clothing, complementary recommendations are displayed. Our model has similar concept regarding the use of pre-trained VGG models for image recognition.

III. SYSTEM MODEL

We proposed a CNN model by using python language. Open-source software library called TensorFlow was applied which is widely used for machine learning applications such as neural network and used Keras, which works as wrapper and it is a high-level neural network library that's built on top of TensorFlow.[2]. We used AWS (Amazon Web Service) resource i.e. S3 storage for uploading the objects into the drop zone JavaScript upload engine through which we categorized and predicted the recommendations of the suiting apparel. We used Boto3 which is a python SDK for AWS. It lets you to directly create, update, and delete AWS resources from your Python scripts. We wrote a python script for that in such a way that after each upload of the image a session folder is created and with every uploading being done the object folder gets wiped off. So technically we are using the S3 storage to store the object for a temporary basis which gets wiped off of the S3 bucket as soon as the task is executed taking away the thought of being charged for using the resource.

While finalizing the model we employed data augmentation technique such as rotation, brightness, zoom, and size to address over-fitting and increased the accuracy of the model to a desirable more than 60%. The bonus help which we got using Vgg-16 model is, it has a feature to identify any generic image that is being uploaded although the accuracy of the generic image detection is not that great but gives a promising result. Our model predicts the image uploaded by identifying and displaying the type of clothing that has the highest confidence level and recommends the apparel that best matches the outfit uploaded by for which you wanted to get recommendations, saving your lot of time and energy thus by changing the traditional way and using modern solution for that problem.

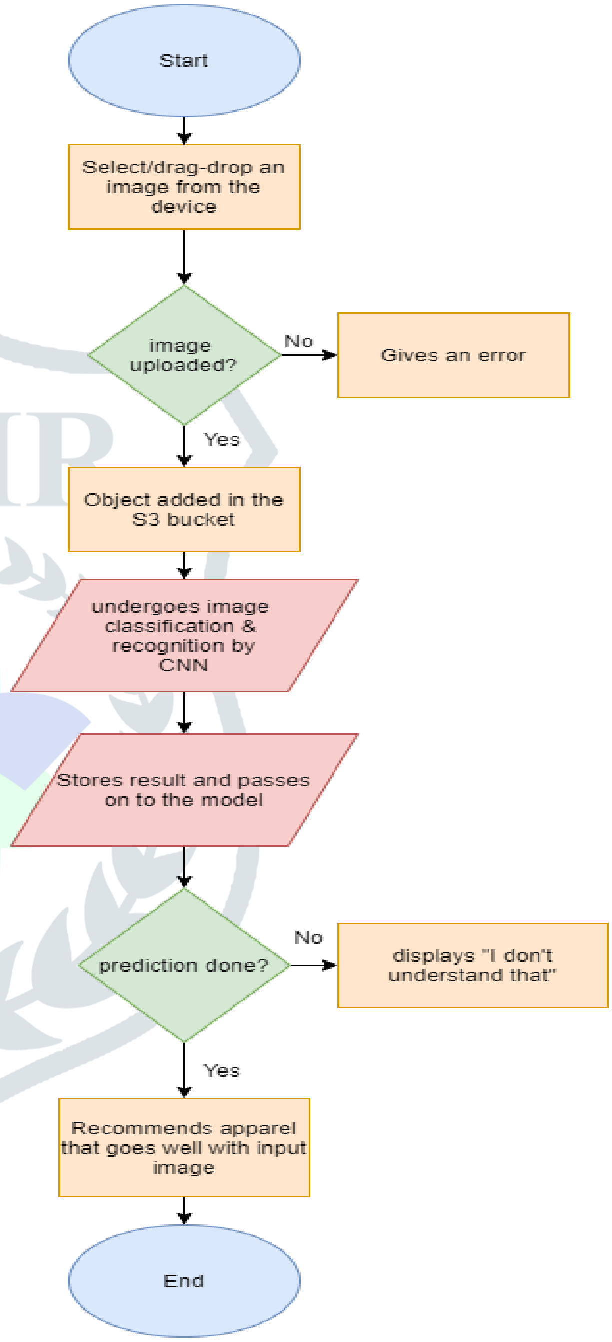

Figure 1 Basic flow our model

VGG model**:**

VGG-16 network is trained on ImageNet dataset which has over 14 million images and 1000 classes, and achieves 92.7% top-5 accuracy. It surpasses Alex-Net network by replacing large filters of size 11 and 5 in the first and second convolution layers with small size 3x3 filters.

VGG-16 is convolutional neural network architecture; its name VGG-16 comes from the fact that it has 16 layers. Its layers consist of Convolutional layers, Max Pooling layers, Activation layers, fully connected layers.

There are in total 13 convolutional layers, 5 Max Pooling layers and 3 Dense layers which sums up to 21 layers but only 16 weight layers. Conv 1 has number of filters as 64 while Conv 2 has 128 filters, Conv 3 has 256 filters while Conv 4 and Conv 5 has 512 filters.

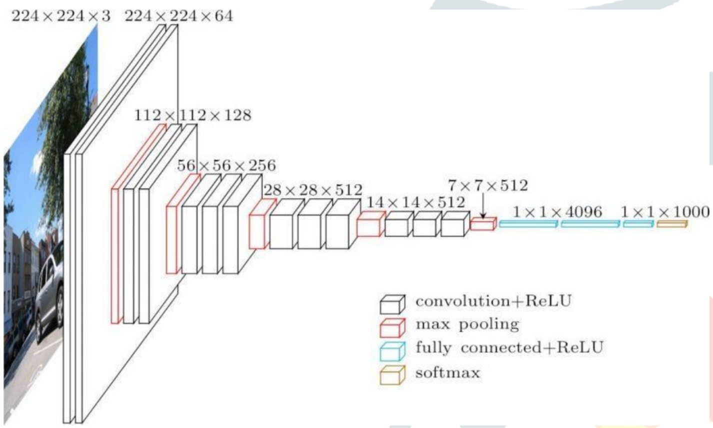

Figure 2 Vgg16 Architecture (image referred from: https://towardsdatascience.com/step-by-step-vgg16-implementation-in-keras-for-beginners-a833c686ae6c)

The input to cov1 layer is of fixed size 224 x 224 RGB image. The image is passed through a stack of convolutional (conv.) layers, where the filters were used with a very small receptive field: 3×3.[3]

Multiple pre-trained models used in transfer learning are based on large **convolutional neural networks (CNN)** (Voulodimos et al. 2018). CNN was shown to excel in a wide range of computer vision tasks (Bengio 2009). Its promising performance and its easiness in training are two of the main factors driving the popularity of CNN over the last years.

CNN has two parts:

1. **Convolutional base** is composed by a stack of convolutional and pooling layers. The important goal of the convolutional base is to generate features from the image. For an intuitive explanation of convolutional and pooling layers, please refer to Chollet (2017).
2. **Classifier** is usually composed by fully connected layers. The goal of the classifier is to classify the image based on the detected features. A layer which is connected fully is a layer whose neurons have full connections to all activation in the previous layer.

Figure 1 shows the **architecture of a model based on CNN**. Note that this is a simplified version, which fits the purposes of this text. The architecture of this type of model is more Complex than what we suggest here [5].

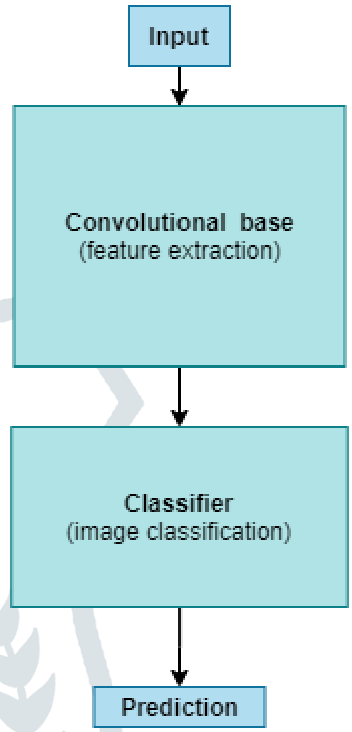

Figure 3 Architecture of a model based on CNN.

Dataset

We point to user created outfit as positive outfit. We also created a set of neutral outfits for each user by randomly mixing the tops, bottoms and shoes.

Although for styling we manually created the datasets that may best match the predicted outfit and add on the styling for the identified apparel and displayed the result. However, the datasets which we used for styling are not big enough which may cause inappropriate results.

IV. METHODOLOGY/SYSTEM FLOW

When an image or several images are added as an input query to the model, the following steps take place.

1. The three pre-trained models recognize the pattern (floral, solid, stripped etc.), clothing type (Chiffon, Denim, Cotton etc.) and most importantly clothing category (Jeans, Blazer, skirt etc.).
2. Now that recognition part is done, recommendation comes next. For complementary recommendation,

images are fetched from our output dataset, considering the cloth pattern and cloth type.

3. Most suited complementary apparels including clothing, footwear and accessories are then displayed to the user.

We used a VGG model pre-trained on ImageNet with transfer learning changing the output layer to include 11 patterns and 17 clothing types. In total we used 200 images per class. For improving the accuracy of the engine, we increased the data to 1500 image per class, as a result classes decreased to 9 patterns and 12 clothing items.

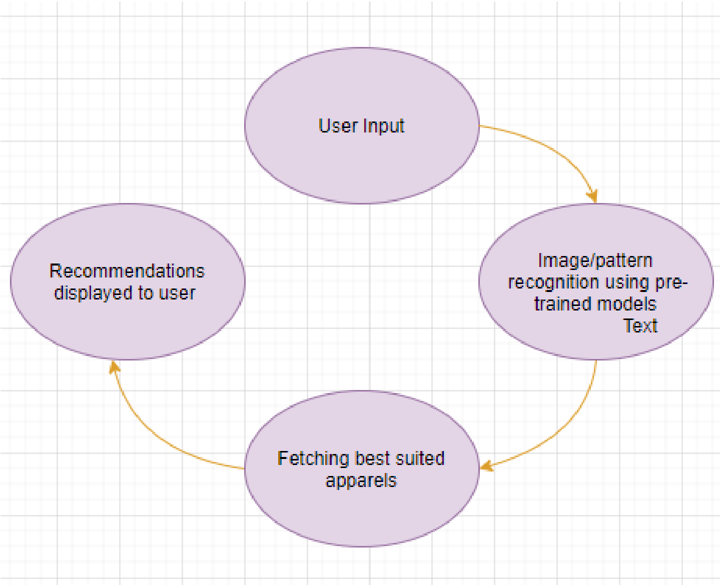

Figure 4 Work flow of our model

V RESULTS

Following is the result of the functionality of the system. This image shows the result of image detection. When an image of a printed floral skirt is the input query, the model detected the apparel as a hoopskirt with a confidence level of 80.4 %.

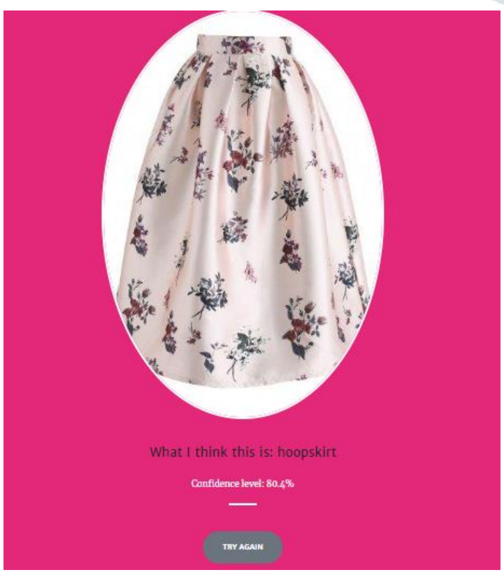

Figure 5 Example of image detection (1)

This image of a plaid skirt was rightly detected as miniskirt with 75% confidence.

Figure 6 Example of image detection (2)

The complementary recommendations for the above apparels are:

Input Image

Complementary Recommendations

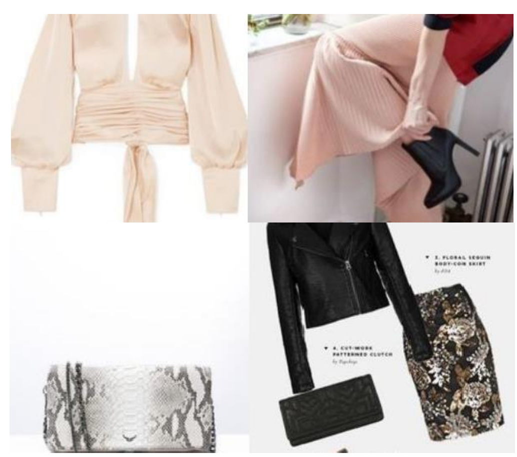

Figure 7 Example of complementary recommendation (1)

Here is another example of our model:

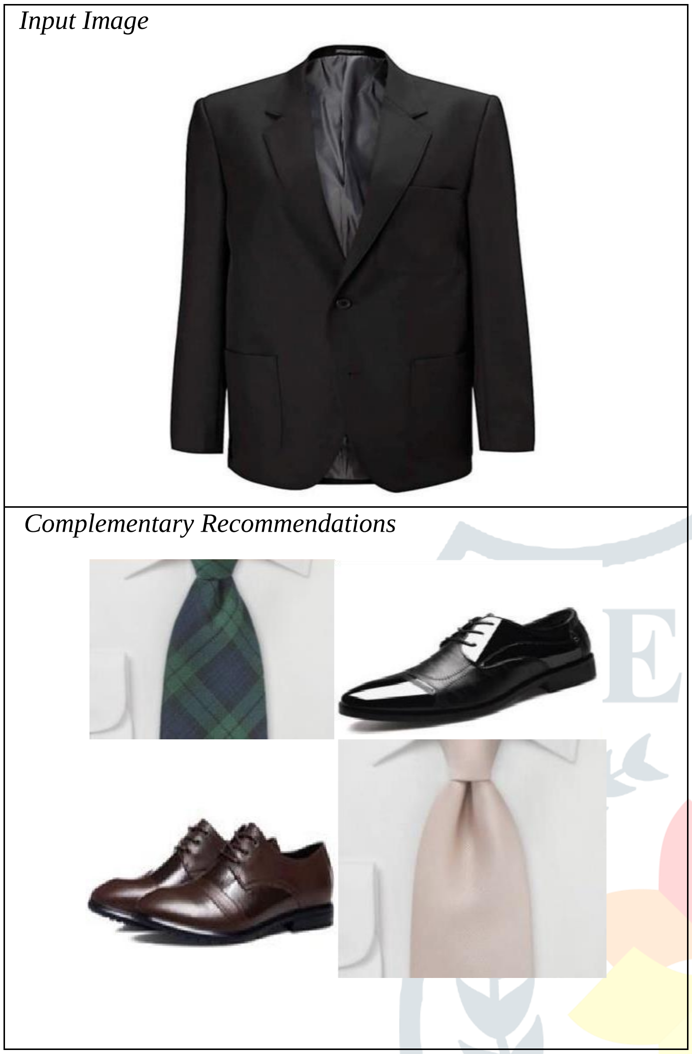

Figure 8 Example of complementary recommendation (2)

Further, trying the testing process with different images we could conclude that the model is capable of detecting the cloth type and recommending the complementary apparel as well as accessories with approximately more than 60% accuracy.

For every top or bottom wear that is given as input, there will be four apparels recommended; these recommendations could either be of the complementary apparel or any kind accessories that will suit the outfit, giving different recommendations each time on same image.

VI CONCLUSION

Our model is a representative application of object set recommendation. We explored the use of deep neural networks, which jointly perform representation learning and compatibility modeling, for this particular task. Our model successfully detects input image and recommends complementary outfits and accessories, based on pattern, occasion and clothing type.

As of now the dataset used in this project is relatively small. Thus, we envision increasing size of our dataset in future, for more efficient and accurate result. Unfortunately, talking about VGGNet, it is painfully slow to train because of its depth and number of fully-connected nodes. VGG16 is over 533MB. This makes deploying VGG a tiresome task. Vgg16 is applied in many deep learning image classification problems. Although it is a great building block for learning purpose as it is easy to implement.

ACKNOWLEDGEMENT

We sincerely thank Mr. Edmund Chan for his guidance and help with our project.

REFERENCES

[1] Karen Simonyan & Andrew Zisserman.: "Very Deep Convolutional Networks for Large-Scale Image Recognition".

[2] Srikanth Taminna.: "Transfer Learning using VGG-16 with Deep Convolutional Neural Network for classifying images".

[3] https://neurohive.io/en/popular-networks/vgg16/

[4] https://machinelearningmastery.com/use-pre-trained-vgg-model-classify-objects-photographs/

[5] https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751

[6] Liu, S., Song, Z., Liu, G., Xu, C., Lu, H., Yan, S.: Street-to-Shop: Cross-Scenario Clothing Retrieval via Parts Alignment and Auxiliary Set. CVPR (2012).

[7] Karen Simonyan, Andrew Zisserman "Very Deep Convolutional Networks for Large-Scale Image Recognition" 2014

[8] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 2012.

[9] Razavian, Ali Sharif, et al. "CNN Features offthe-shelf: an Astounding Baseline for Recognition." Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on. IEEE, 2014.

[10] Pan, Sinno Jialin, and Qiang Yang. "A Survey on Transfer Learning." IEEE Transactions on Knowledge and Data Engineering 22.10(2010): 1345-359.

[11] https://riptutorial.com/keras/example/32608/transfer-learning-using-keras-and-vgg

[12] J. J. McAuley, C. Targett, Q. Shi, and A. van den Hengel. Image-basedrecommendationsonstylesandsubstitutes.In SIGIR,2015

[13] Han, J.; Pei, J.; and Yin, Y. 2000. Mining frequent patterns without candidate generation. In ACM SIGMOD Record, volume 29, 1–12. ACM

[14] D. Goel, S. Chaudhury and H. Ghosh, "Recommendation of complementary garments using ontology," 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Patna, 2015.

[15] Wang-Cheng Kang and Eric Kim and Jure Leskovec and Charles Rosenberg and Julian McAuley "Complete the Look: Scene-based Complementary Product Recommendation" 2018

[16] Edmund Chan, Eloise Rousseau, Danny Vu, Leena Sherif "SenseFashion"