Fashion Recommender System

Rakesh Rawat, P Srinivas Rao, and Suman Kumar

Indraprastha Institute of Information Technology, Delhi
rakesh17046@iiitd.ac.in, ponnaganti17035@iiitd.ac.in,
suman17060@iiitd.ac.in

Abstract. The paper presents a fashion recommendation system trained on the DeepFashion dataset. In the existing datasets the amount of annotated images is very less, whereas the DeepFashion dataset consists of over 800k images which are richly annotated. We constructed a deep learning model which is trained on the Category and Attribute Prediction part of the DeepFashion dataset.

Keywords: DeepFashion · Category and Attribute Prediction · CNN · ResNet · kNN

1 Introduction

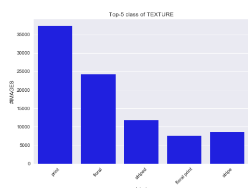
In the fashion domain, the recent research has been devoted to the clothes classification, attribute prediction and clothing item retrieval. The paper presented a novel dataset and using that dataset they proposed their model FashionNet. The main contribution of the author to the community was the DeepFashion dataset.

- In the clothes classification task, the clothes are classified into multiple categories such as rampers, hoodies etc.
- In the attribute prediction, certain attributes of the clothes are predicted. The attributes of a cloth can be of many types. For a cloth the different types of attributes are as follows:
  - Texture Attributes: Palm, color blocked, stripes, etc.
  - Fabric Attributes: Leather, tweed, etc.
  - Shape Attributes: Crop, Midi, etc.
  - Part Attributes: Bow-F, Fringed-H, etc.
  - Style Attributes: Mickey, baseball, etc.
- The clothing retrieval earlier relied on the handcrafted features such as color histograms, Hog and SIFT. But with the introduction of deep learning models this task can be addressed easily by learning more discriminative representations.

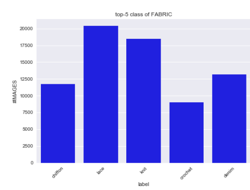
The clothing retrieval task inherently requires addressing the cloth classification and attribute prediction task. We approached the problem of cloth retrieval as the recommendation system task. Not much work has been done on the DeepFashion Dataset. This dataset is introduced in the paper and they proposed a model FashionNet which used the deepfashion dataset for learning. We propose a cloth recommender system which uses the category and the attributes of a cloth for learning and generate the cloth images for a given input cloth image.

2 Related Work

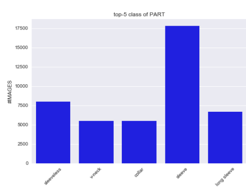
The similar work has been done by the Where To Buy It (WTBI) [2] and Dual Attribute-aware Ranking Network [3]. The WBTI method implemented a multilayer perceptron layer on top of the pretrained ImageNet models. The Darn method used a two stream CNN which are attribute regularized. The first stream handles street images and the other stream handles shop images.



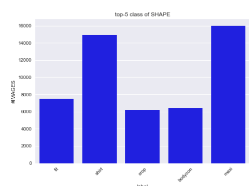
(a) Texture



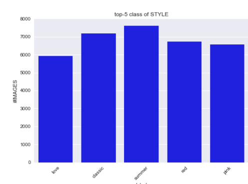
(b) Fabric



(c) Part



(d) Shape



(e) Style

Fig. 1. Attributes Distribution of Style, Texture, Fabric, Part and Shape

3 Dataset

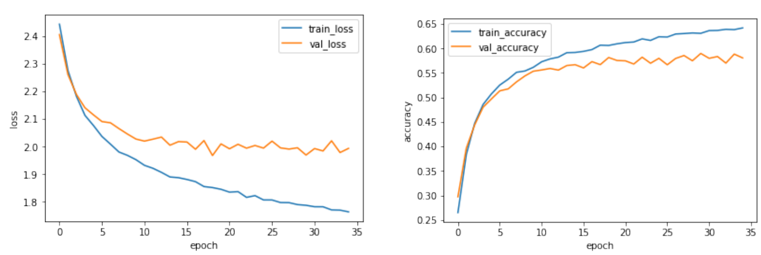
We used DeepFashion dataset as it considered as the standard in the fashion domain. The DeepFashion dataset contains over 800K images which are further annotated with 1000 attributes and 50 categories.The features labels are broadly divided in two parts one is clothing category and attributes of the clothes. The clothing category, mainly contains the noun names e.g. "dress". For attributes, mainly contains the combination of adjectives e.g. "animal print". The attributes were further divided into five groups like texture, fabric, shape, part and style.

For clothing category, we picked the top 10 classes and ... For attributes, as its divided into five groups, We picked the top 5 classes from texture, fabric, shape, part and style respectively which are as follows:

- In texture, we used the floral, floral print, print, stripe and striped.
- In fabric, we used the chiffon, crochet, denim, knit and lace.
- In shape, we used the bodycon, crop, fit, maxi and shirt.
- In part, we used the collar, long sleeve, sleeve, sleeveless and v-neck.
- In style, we used the classic, love, pink, red and summer.

As each classes contain varying number of images, we picked the count of fifth class images and try to keep other four class count similar to the fifth class count. So this way each group contain images which are having approximate same count.

- From texture group, each class contain around 7000 images so in total images 35,000 out of which 26233 used for training and 8742 for validation. The texture attribute distribution is shown in fig. 1(a)
- From fabric group, each class contain around 8000 images so in total images 40,000 out of which 29970 used for training and 9990 for validation. The fabric attribute distribution is shown in fig. 1(b)
- From part, each class contain around 5500 images so in total images 28,000 out of which 21687 used for training and 7230 for validation. The part attribute distribution is shown in fig. 1(c)
- From shape, each class contain around 6000 images so in total images 30,000 out of which 22477 used for training and 7493 for validation. The shape attribute distribution is shown in fig. 2(b)
- From style, each class contain around 6000 images so in total images 30,000 out of which 22428 used for training and 7477 for validation. The style attribute distribution is shown in fig. 1(e)



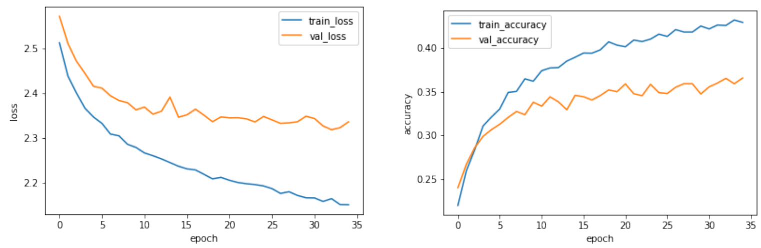
(a) Loss in Shape (b) Accuracy on Shape

Fig. 2. Loss and Accuracy on the ResNet Models of Shape Attribute

4 Methodologies

4.1 Author's Method

The authors proposed their method FashionNet(a deep learning architecture) which is trained completely on the DeepFashion dataset from scratch. The archi-



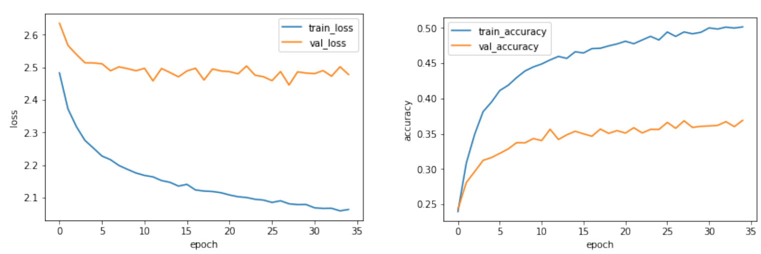
(a) Loss in Part (b) Accuracy on Part

Fig. 3. Loss and Accuracy on the ResNet Models of Part Attribute

tecture of FashionNet model is similar to VGG-16. The VGG-16 is very efficient in many vision tasks for example in object recognition and segmentation. Specifically, the structures of FashionNet below the penultimate (i.e. from top to bottom) convolutional layer are the same as VGG-16, except the last convolutional layer.

4.2 Our Method

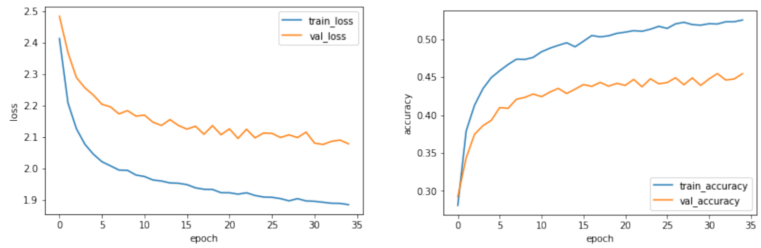
Brief Explanation: Due to the limited computing resources we didn't trained the model from scratch. Unlike authors we proceeded with the different approach. We used the combination of the Where To Buy It (WTBI) method [2] and Dual Attribute-aware Ranking Network [3] method.This combined approach was less computational intensive and yielded similar results. The properties we took from each method is as follows:



(a) Loss in Style (b) Accuracy on Style

Fig. 4. Loss and Accuracy on the ResNet Models of Style Attribute

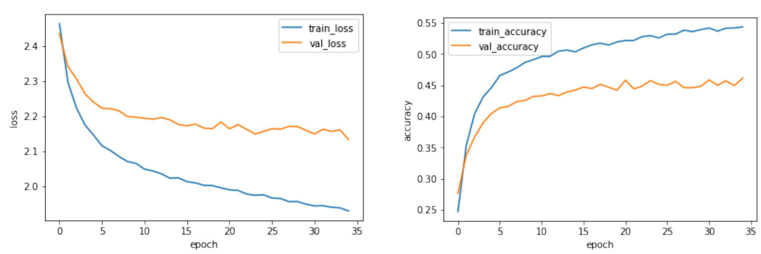
- Similar to Where To Buy It (WTBI) method [2] we used the pretrained model. Unlike WTBI, we used Pretrained ResNet model which is trained on the ImageNet dataset.



(a) Loss in Texture (b) Accuracy on Shape

Fig. 5. Loss and Accuracy on the ResNet Models of Texture Attribute

– Similar to the Stream in Dual Attribute-aware Ranking Network(DARN) method [3]. We also used six streams of ResNet. All these ResNets are learn independently on the categories, and on the five attributes of the clothes images. Instead of training on the street images and shop images we trained our models on the DeepFashion images.



(a) Loss in Fabric (b) Accuracy on Shape

Fig. 6. Loss and Accuracy on the ResNet Models of Fabric Attribute

Detailed Explanation: We used pretrained ResNet Model which was trained on the ImageNet dataset. The ResNet Model was modified so that it can be used to train on the DeepFashion dataset. The details of the modified ResNet is as follows:

– The ResNet model which we are using is pretrained. We clipped off the last two layers and replaced with two new fully connected layers.
– The weights on the two new fully connected layers are randomly initialized.
– During the backpropagation the weights of only last 12 layers are updated and rest all weights are kept unchanged.

There are total six ResNets with the above explained architecture. All the six ResNets are trained independently. The details of each stream is as follows:

- The First ResNet is trained on the features named categories. The output of this ResNet is the 10 dimensional vector in which each dimension represents one feature out of the 10 features.
- The Remaining five ResNets are trained on fabric, shape, style, part and texture features of the deepFashion dataset. The output of the each ResNet is a 5 dimensional vector in which each dimension represents one feature out of 5 features.
- From the remaining 5 ResNets we get 5 five dimensional vectors.
- All the vectors are concatenated. In total, we have 35 dimensional vector. 10 dimensional from the category and the other 25 dimensional vector from the other features.

The codes of our method is on GitHub [4]. From the 35 dimensional vector of the image. We then find the top 4 images which can be recommended for the given input image.

5 Results

For all the images in the training set the 35 dimensional vector is generated. And during the testing a input image of the dress/cloth is given. The test input image is then converted into the 35 dimensional vector. For this test vector we find the nearest neighbourhood image vectors. And in order to recommend the images we recommend the top four images for a given input test image.

The accuracy of the ResNet models is not the appropriate measure for our recommender system as the output of each ResNet model is not used for classification it is further used as the input to the KNN. Hence loss is the better metric to measure the performance of all the six models.

We determined the accuracy and the loss for each models. The results are shown in the following table:

Table 1. The loss and Accuracy of the Models

Model	Loss	Accuracy
Category	2.0486	0.5642
Fabric	0.1232	0.463
Style	2.34	0.375
Shape	2.053	0.574
Texture	2.14	0.453
Part	2.3578	0.365

The results of our model is shown in fig. 10 . The first column of the two rows represents the input images and the other two columns represents the recommended images.

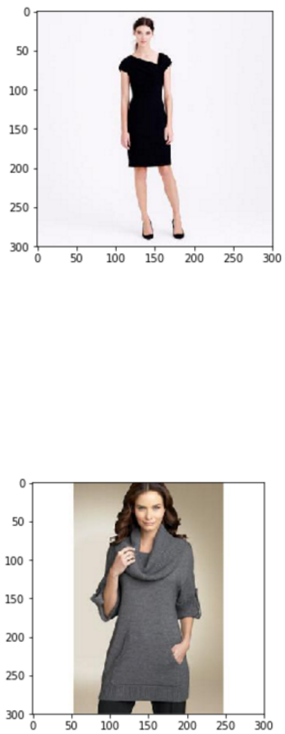


Fig. 7. Input Images

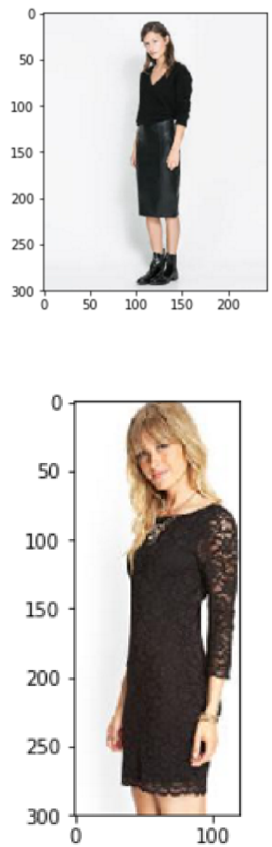


Fig. 8. First Recommended Image

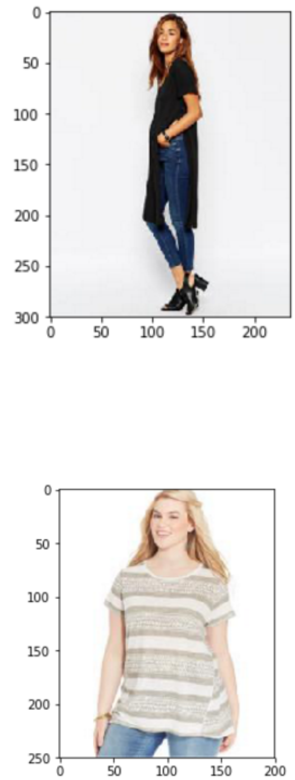


Fig. 9. Second Recommended Image

Fig. 10. Results of our Model

References

1. Ziwei, Luo, Qiu,Wang. Xiaoou Tang. DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. Conference on Computer Vision and Pattern Recognition 2016.
2. M. H. Kiapour, X. Han, S. Lazebnik, A. C. Berg, and T. L. Berg. Where to buy it: Matching street clothing photos in online shops. In ICCV, 2015.
3. J. Huang, R. S. Feris, Q. Chen, and S. Yan. Cross-domain image retrieval with a dual attribute-aware ranking network. In ICCV, 2015.
4. Github Link, `https://github.com/suman9868/CF_project`.