Learning Attribute Representations with Localization for Flexible Fashion Search -Supplementary Material-

Table 1. Some samples of attributes and their associated labels are shown for Shopping100k dataset.

Attribute Categories	Attribute Labels (total number)
Category	Shirt, Dress, Trouser, ... (16)
Collar	Polo, Round, Mandarin, ... (17)
Color	Wine, Navy, Blue,... (19)
Fabric	Denim, Jersey, Sweat, ... (14)
Fastening	Zip, Belt, Button, ... (9)
Fit	Skinny, Loose, Straight, ... (15)
Gender	Male, Female (2)
Length	Normal, 3/4, Short, ...(7)
Neckline	Henley, Boat, Envelope, ... (11)
Pattern	Animal, Plain, Photo, ... (15)
Pocket	Side, Flap, Zip, ... (7)
Sleeve length	Long, Short, Sleeveless, ... (9)
Sport	Running, Yoga, Football, ... (15)

1. Datasets

1.1. Shopping100k Dataset [1]

Shopping100k [1] dataset consists of nearly 100,000 images along with, 12 fashion attributes. In total, the dataset contains 151 number of labels as shown in Table 1. Some examples are given in Figure 1 to demonstrate the detail level of the dataset.

1.2. Deepfashion Dataset [2]

For Deepfashion dataset [2], we choose to use category, texture and shape attributes. Some labels of attributes are for; **(1) Category:** Anorak, Blazer, Blouse, Bomber, Capris, Chinos, Culottes, Cutoffs, Caftan, Cape, Coat, Coverup, ... **(2) Texture:** Baroque, Butterfly, Brocade, Chevron, Clean, Colorblock, Contrast, Daisy, Diamond, ... **(3) Shape:** Crop, Midi, Cutoff, Drapey, Fit, Foldover, High-rise, Knee-length, Round, ...

Figure 1. Some example images and their attributes from the Shopping100k dataset.

2. Visual Examples of the Attribute Manipulation

In figure 2 and 3, some examples are given for conducting attribute manipulation on the query images with the proposed FashionSearchNet for Shopping100k [1] and Deepfashion [2] datasets respectively. We also provide heatmaps related to the attribute manipulation operation to show where the network focuses on when changing an attribute. First 3 columns represent query image, attribute manipulation, attribute activation map of manipulated attributed respectively, and the last 4 columns are top-4 retrieved images, ranked from left to right. Correct predictions are shown with the green bounding box.

Here we interpret some of our results for Shopping100k dataset in Figure 2. FashionSearchNet generally satisfies the requested attribute manipulation but note that another objective to make a successful retrieval is to keep attributes of the query image. As each image has around 6-8 different attributes, that makes the image retrieval problem very challenging. For example, for the 1st-row, retrieved images that do not satisfy the requirement miss the fabric attribute. In the 2nd row, attributes of the 3rd and the 4th retrieved images do not confirm the ”set of desired attributes” because of the different color and pattern attributes respectively. In addition, heatmaps are usually correlated with the attribute manipulation. Note that, another reason why only a few images have the green bounding-box might be because there are no more images which satisfy the wanted attributes. More examples are given in the other rows using different types of attributes.

For DeepFashion dataset, we provide some results using the texture attribute manipulation in Figure 3. Note that for this set of examples, attributes to be satisfied includes texture, shape and category. Therefore, FashionSearchNet aims to find images which satisfy the desired texture attribute while keeping clothing category and shape. For example, in the first row of Figure 3, the 1st and the 2nd retrieved images satisfy these conditions but the 2nd image actually has a different collar attribute. As DeepFashion dataset does not include collar attributes, FashionSearchNet is not aware of it. Therefore, Shopping100k dataset is more suitable in order to conduct a more detailed fashion search. However, by using DeepFashion dataset, our method is proven to be working on the real-world images as well. More examples are given for further investigation in the next rows.

References

[1] K. E. Ak, J. H. Lim, J. Y. Tham, and A. A. Kassim. Efficient multi-attribute similarity learning towards attribute-based fashion search. In *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on*, pages 1671–1679. IEEE, 2018.

[2] Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *IEEE Conference Computer Vision Pattern Recognition (CVPR)*, pages 1096–1104, 2016.

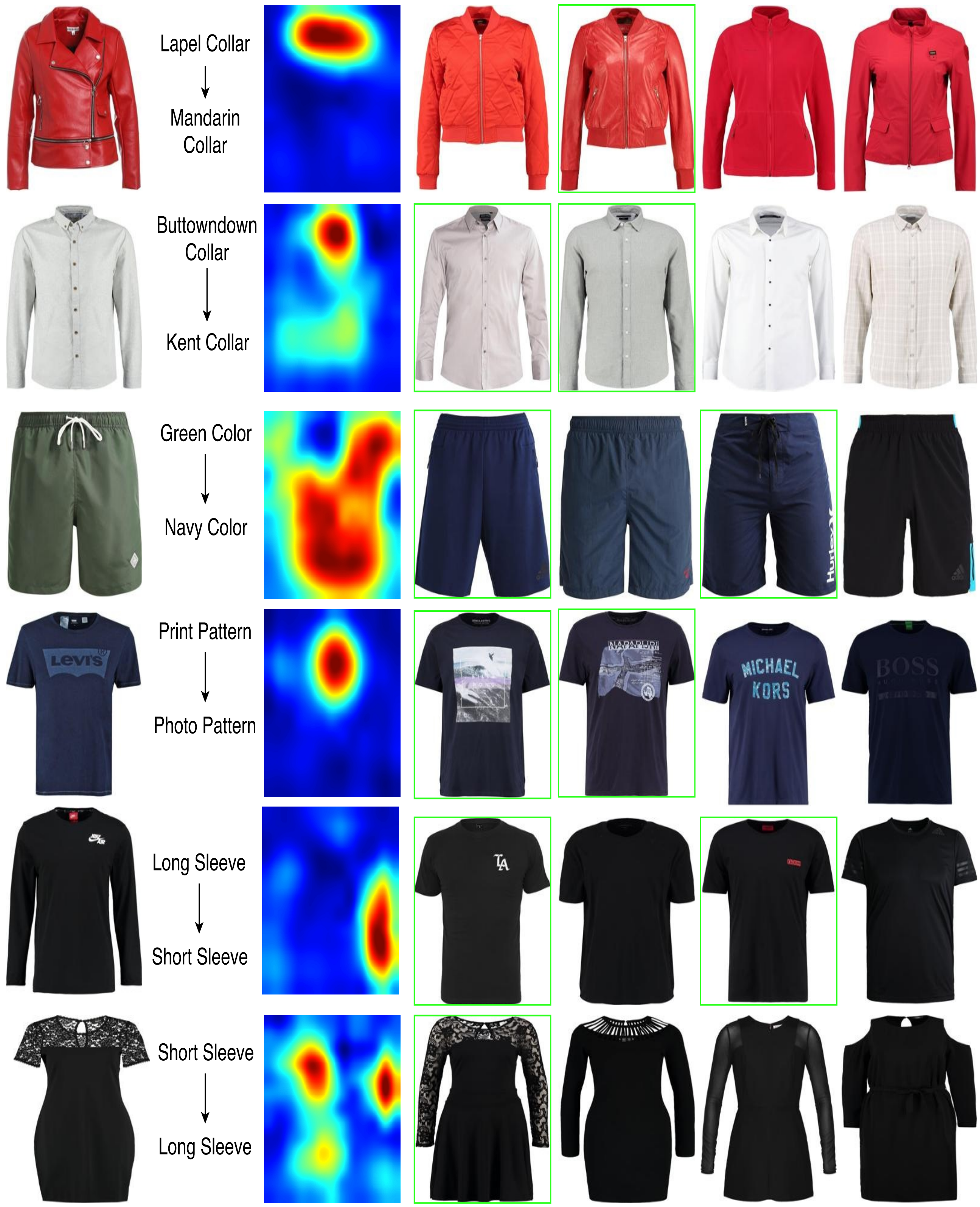

Figure 2. Visual results of attribute manipulations on the Fashion100k dataset. The first 3 columns represent query image, attribute manipulation, attribute activation map of manipulated attributed respectively, and the last 4 columns are top-4 retrieved images, ranked from left to right. Correct predictions are shown with the green bounding-box.

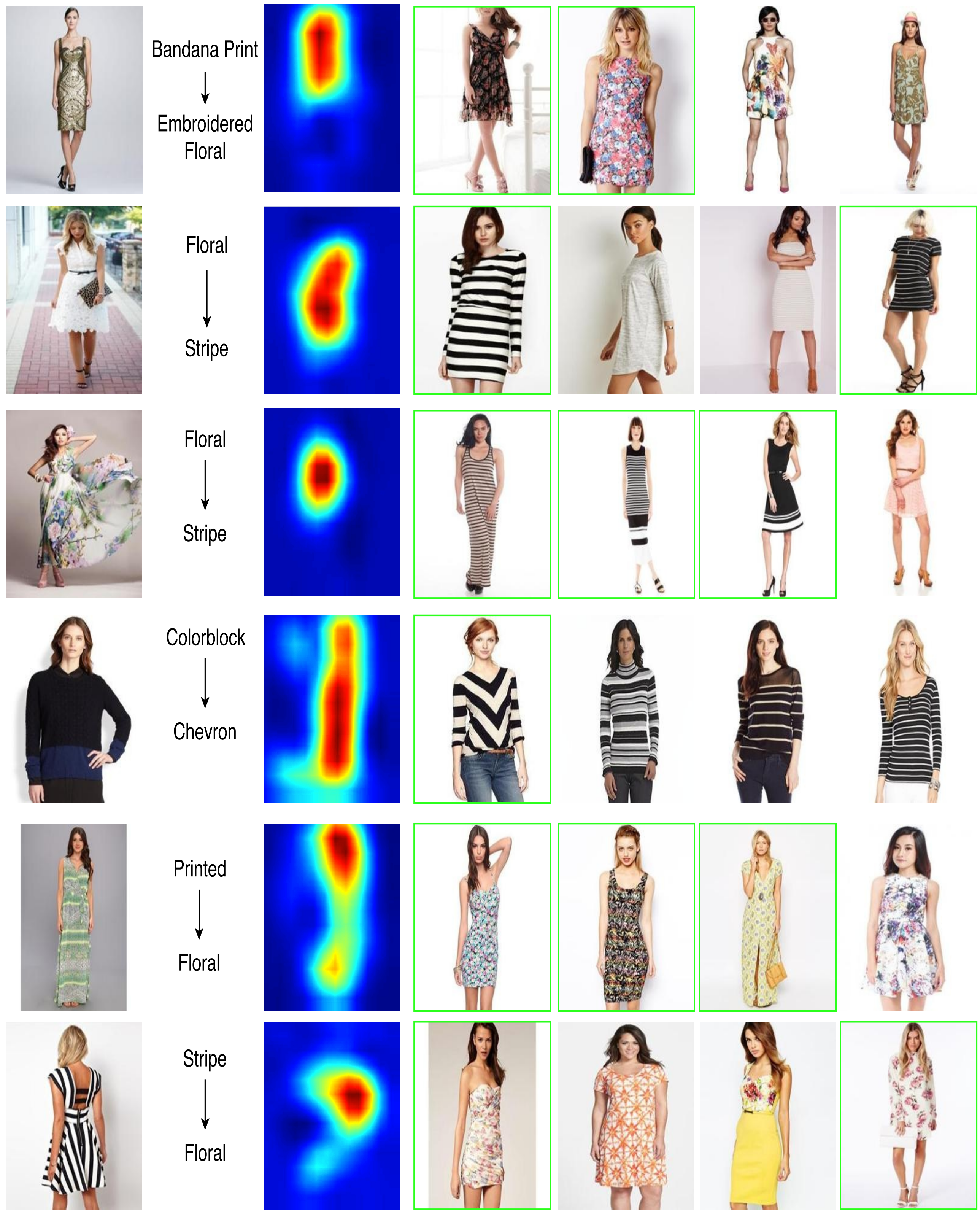

Figure 3. Visual results of attribute manipulations on the DeepFashion dataset's category and attribute prediction benchmark. The first 3 columns represent query image, attribute manipulation, attribute activation map of manipulated attributed respectively, and the last 4 columns are top-4 retrieved images, ranked from left to right. Correct predictions are shown with the green bounding-box.