Learning Multifunctional Binary Codes for Both Category and Attribute Oriented Retrieval Tasks

Haomiao Liu[1,2], Ruiping Wang[1,2,3], Shiguang Shan[1,2,3], Xilin Chen[1,2,3]
[1]Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, 100190, China
[2]University of Chinese Academy of Sciences, Beijing, 100049, China
[3]Cooperative Medianet Innovation Center, China
haomiao.liu@vipl.ict.ac.cn, {wangruiping, sgshan, xlchen}@ict.ac.cn

This document gives details about the attributes defined on ImageNet-150K, and additional real retrieval results. This material is best viewed in color.

1. Example Images of Attributes

In this section, we provide example images of each attribute defined on ImageNet-150K (25 attributes, including color, texture, shape, material, and structure). The attributes were defined and annotated mainly based on the ImageNet-attribute [S1] and Animals with Attribute (AwA) [S2] datasets. Compared to [S1], our dataset covers much more categories (1000 vs 384) and images (50,000 vs 9,600). For each attribute, three positive samples along with three negative samples are shown in Figure 1, 2, and 3 (the leftmost three in each row are positive samples and the rest are negative samples). In our experiments, the attributes are binary, namely, an image either has or does not have the attribute.

2. Real Retrieval Cases

This section gives more real retrieval cases on the attribute-oriented retrieval tasks described in Sections 4.4 and 4.5 of the main paper (the results were obtained with 256-bit binary codes).

2.1. Results on CFW-60K

The results of task II and task III on CFW-60K are shown in Figure 4 and 5 respectively. In task II, the system is required to retrieve images of subjects with the same gender, race, and age group as the subject in the query image. As we can see from the failed cases (Figure 4(b)), for each query image, though the top feedbacks fail to match the exact attributes of the query, all of them have the same gender, race, and age group. By further investigating the failed cases, we found that the main cause is the incorrect attribute predictions of the query image. In task III, either inaccurate attribute prediction or the incapability of binary codes in preserving category similarity would result in failed cases. Here we only show the successful cases to demonstrate the potential of our method in this challenging realistic retrieval scenario.

2.2. Results on ImageNet-150K

The results on ImageNet-150K are shown in Figure 6 and 7. For this dataset, since there are only two images from each category in the "Test" set, to better evaluate our method for qualitative demonstration, in this supplemental experiment we used the "Test" set as query images, and retrieved images from both "Train-Both" and "Train-Attribute" sets. Some successful retrieval results on task II and task III are provided, suggesting that our method has the potential to be applied in these two realistic yet very challenging object retrieval scenarios.

References

[S1] O. Russakovsky, L. Fei-Fei. Attribute Learning in Large-scale Datasets. In European Conference on Computer Vision workshop (ECCVW), 2010, pages 1-14.

Figure 1. Example images of attributes on ImageNet-150K. For each attribute, three positive samples (the leftmost three) and three negative samples (the rightmost three) are shown in this figure.

[S2] C. H. Lampert,H. Nickisch,S. Harmeling. Learning to Detect Unseen Object Classes by Between-class Attribute Transfer. In Computer Vision and Pattern Recognition (CVPR), 2009, pages 951-958.

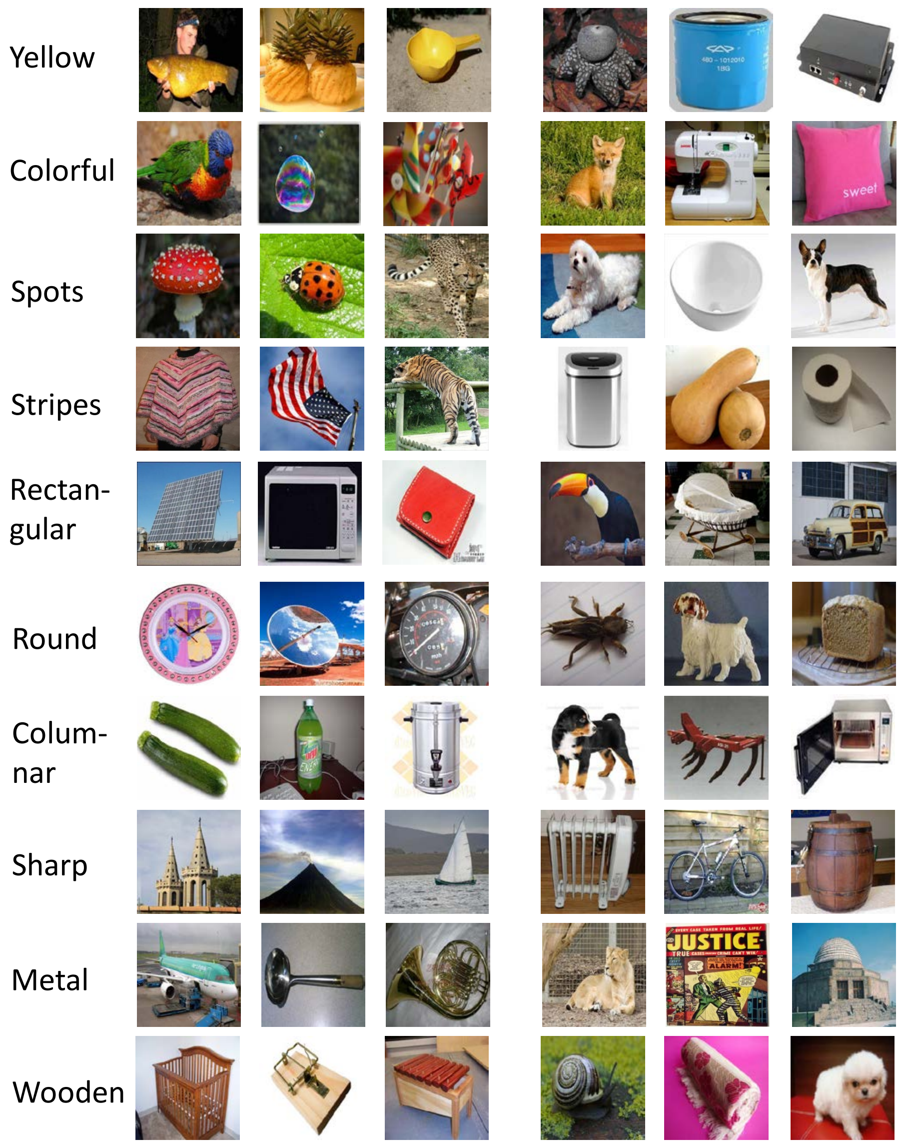


Figure 2. Example images of attributes on ImageNet-150K. For each attribute, three positive samples (the leftmost three) and three negative samples (the rightmost three) are shown in this figure.

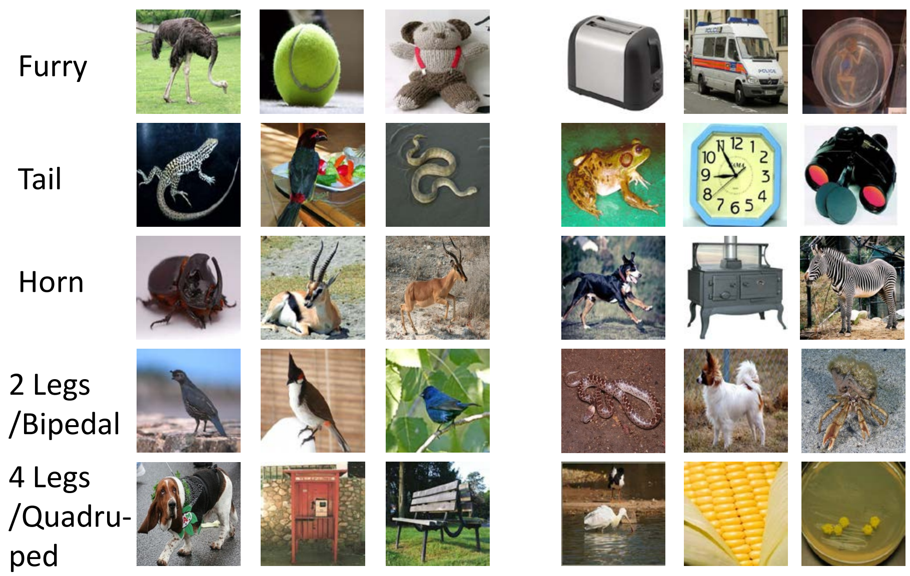
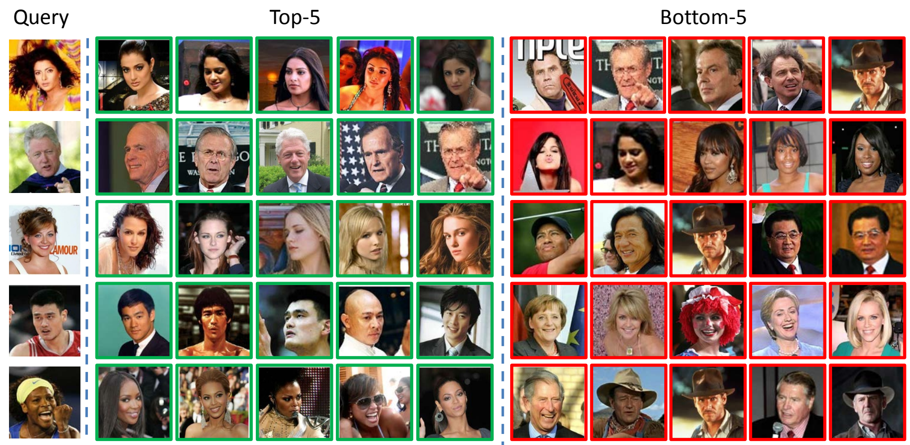


Figure 3. Example images of attributes on ImageNet-150K. For each attribute, three positive samples (the leftmost three) and three negative samples (the rightmost three) are shown in this figure.



(a)

(b)

Figure 4. Some real retrieval results of task II on CFW-60K (retrieving images of subjects with the same gender, race, and age group as the subject in the query image). The results were obtained with 256-bit binary codes. The notations are consistent with the main paper (please refer to Figure 1(b) in the main paper for details). (a) successful cases, (b) failed cases. In the failed cases, the predicted gender, race, and age group of the 3 query images are: 1) male + white + young (groundtruth: male + white + mid-aged), 2) female + Asian + young (groundtruth: female + white + young), 3) male + Asian + young (groundtruth: male + white + young). Note that as mentioned in our main paper, in this task II the predicted attributes of all images (both query and database images) were used for retrieval, while the evaluation is performed on the ground-truth attribute labels. As a result, both wrong predictions of the query image and the database images would cause a mismatch.

Figure 5. Some real retrieval results of task III on CFW-60K. The notations are consistent with the main paper (please refer to Figure 1(b) in the main paper for details).

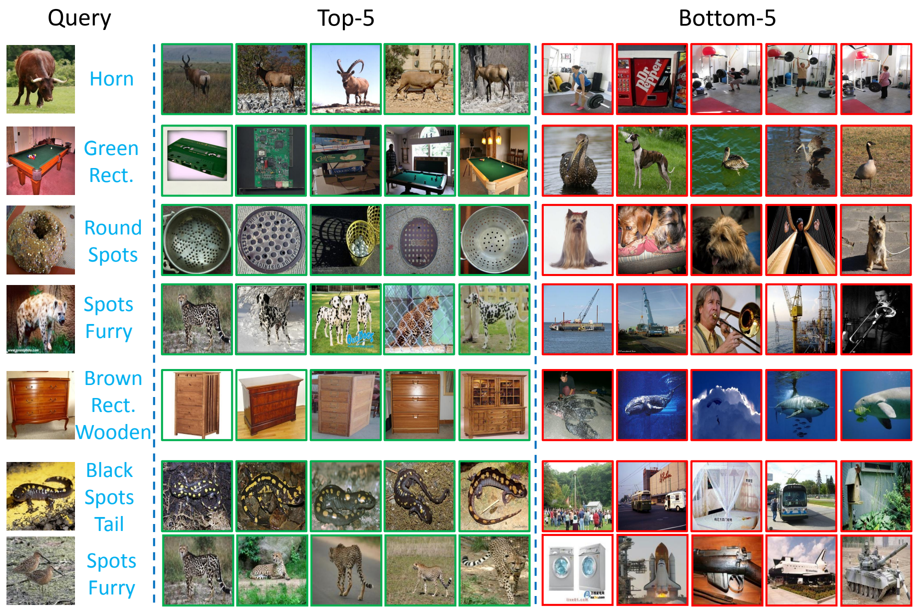


Figure 6. Some real retrieval results of task II on ImageNet-150K, the interested attributes are listed on the right side of the query images, including color, texture, shape, material, and structure. Images in the "Test" set were used as queries, and the "Train-Both" set and "Train-Attribute" set were used as database. The results were obtained with 256-bit binary codes. The notations are consistent with the main paper (please refer to Figure 1(b) in the main paper for details).

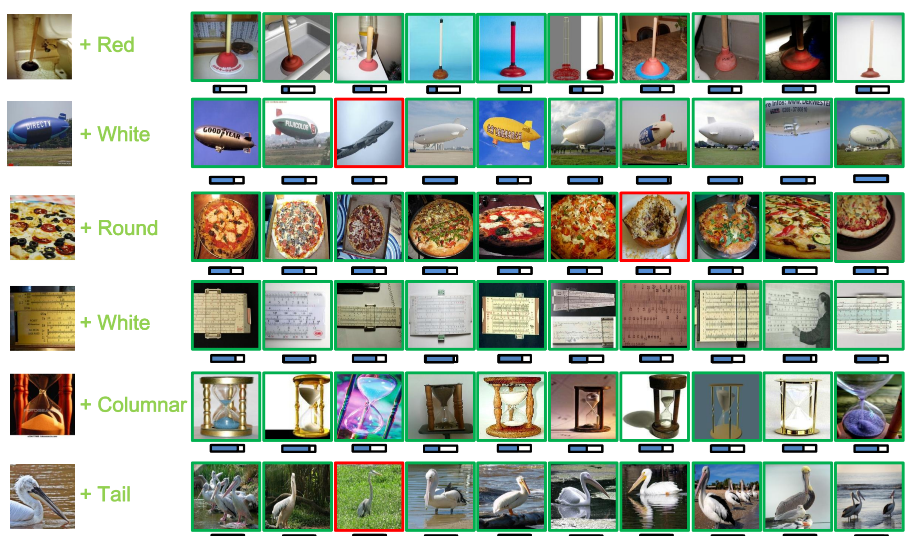


Figure 7. Some real retrieval results of task III on ImageNet-150K. Images in the "Test" set was used as queries, and the "Train-Both" set and "Train-Attribute" set were used as database. The notations are consistent with the main paper (please refer to Figure 1(b) in the main paper for details). Note that for the last query, one of the top-10 feedbacks (that is bounded by red box) does not match the query in terms of category (the ground-truth category of the query is "pelican" respectively, while the ground-truth category of the wrong feedback is "crane"). We can see that the wrong feedback looks very similar to the query image, even humans would have some difficulty to perform such a fine-grained categorization task, thus it is reasonable that the retrieval system made such a mistake.