Table 2: Mean average precision (MAP) of rankings based on HOG features (HOG) and attribute-based features (ATT)

Category	HOG	ATT
Elephant	0.591	**0.872**
Horse	0.377	**0.686**
Leopard	0.577	**0.870**
Rhino	0.408	**0.713**
Cat face	0.588	**0.735**
Dog face	0.592	**0.763**
Leopard face	0.584	**0.735**
Cow head	0.795	**0.885**
Motorbike	0.975	**1.000**
Bike	0.768	**0.962**
Car	0.940	**0.999**
Cannon	0.256	**0.481**
All	0.621	**0.808**

We selected two pairs of categories from the Caltech101 dataset: Emu and Flamingo (different kinds of birds), as well as Windsor chair and Chair (one is a subset of the other). We learned an attribute definition matrix from positive (source images) and negative (background images) examples seperately. Given the attributes of the two, we inferred the attribute weight vectors of target images using the Gibbs samplers. We then trained SVM classifiers with the RBF kernel on $2 - 5$ examples of target images and tested on the held-out 30 images. There are three choices of image representations: HOG features (our baseline), attribute weight features, and a concatenation of the two. We report the average area under ROC curves over 15 trials in Figure 10. As shown, our attribute-based features give a higher detection accuracy in both cases.

The result suggests that attributes learned from images in one class are generic enough to be transferred to the learning of similar images in the other class. However, it is worth mentioning that the improvements in this transfer learning task are not always seen for many pairs of categories in the Caltech101 dataset. We suspect that this could be due to several reasons. For example, since all images in one category are roughly aligned, a few labeled images in the target category are enough to train an accurate detector, and thus the attributes do not really provide any additional information. In addition, we find out that, in some cases, attributes learned from the source images are too specific (such as fixed at some specific locations). These attributes cannot be transferred to the target images very well, even though they appear somewhere in the target images. Again, this could potentially be caused by the source images being roughly aligned. Finally, we discover that the detection accuracies are quite sensitive to the choices of attribute weight vectors inferred by the samplers. In the scenarios where we train our detectors using only a few examples, a wrong choice of attribute weight vectors can largely degrade the detection performance.

7 Conclusion and Future Work

We have proposed a nonparametric approach to the unsupervised learning of image attributes using the infinite sparse factor analysis. The Indian buffet process provides a Beyasian prior that introduces sparsity to our model. Also, the number of attributes does not have to be specified in advance due to the nonparametric property of the IBP. Experimental results show that attributes learned by our model are meaningful.

For future work, we plan to investigate into other inference algorithms. Moreover, the fact that our attributes are not spatially localized motivates us to model attribute positions, orientations, and correlations. We believe that doing so will make it far easier for us to perform transfer learning tasks. The transformed Indian buffet process introduced by Austerweil and Griffiths [1] seems to provide a good framework for such extension.