Cross-Domain Object Recognition Using Object Alignment

Pengcheng Liu
pengcheng.liu@ia.ac.cn

Chong Wang
cwang@nlpr.ia.ac.cn

Peipei Yang
ppyang@nlpr.ia.ac.cn

Kaiqi Huang
kqhuang@nlpr.ia.ac.cn

Tieniu Tan
tnt@nlpr.ia.ac.cn

Center for Research on Intelligent Perception and Computing,
National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences,
Beijing, China

Figure 1: Comparison of methods for a four-class problem on domain adaptation. (a) Existing methods minimize the domain difference based on two ambiguous image feature spaces: X_S and X_T. (b) We minimize the domain difference based on two more similar feature spaces consist of aligned objects: $\widetilde{X}_S$ and $\widetilde{X}_T$. Best viewed in color.
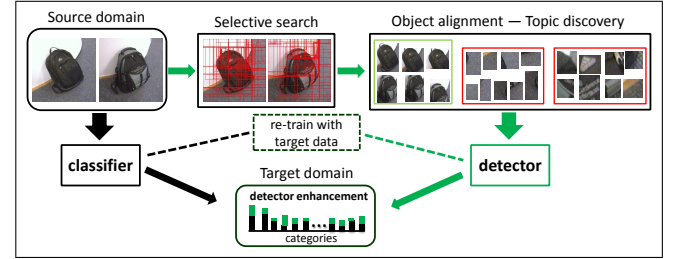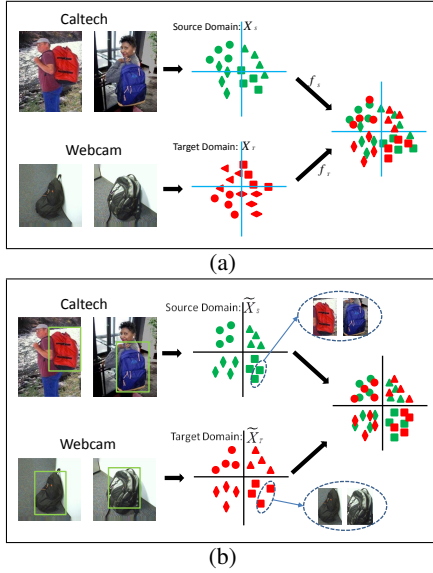

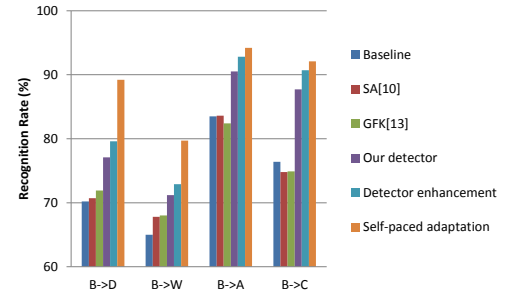Figure 2: Pipeline of our method. Best viewed in color.

Figure 3: Performance of different methods on Bing-Caltech based cross-domain object recognition. This figure is best viewed in color.

In this paper, we focus on the problem of cross-domain object recognition [4], which has long been one of the challenging problems in computer vision. This problem typically arises when training (source domain) and test (target domain) samples are drawn from different distributions. In the problem of object recognition, this case is usually caused by the situation that training and test samples are acquired under different sets of background, lighting, view point, resolution conditions, etc.

One popular solution to the problem of cross-domain object recognition is minimizing the difference between the source and target distributions. Existing methods are devoted to minimizing that domain difference in a complex image space, which makes the problem hard to solve because of background influence, as shown in Figure 1 (a). Since the object and background are twisted in that image feature space, the discrepancy caused by background is difficult to eliminate, which makes it hard to learn optimal f_S and f_T for minimizing $D(f_S(X_S), f_T(X_T))$. To discount the influence of the background, we propose to minimize that difference using object alignment. As shown in Figure 1 (b), we minimize the domain difference by transferring to the feature space of aligned objects $\widetilde{X}_S$ and $\widetilde{X}_T$, but not the image feature space having background influence.

The key insight of our approach is that the difference between the source and target distributions can be reduced by discounting the influence from the ambiguous background. We define the semantic object as the object that occurs in all the images of one class. To discount the background influence, our primary goal is to automatically localize the semantic object so that the irrelevant background can be eliminated. Then based on the semantic object regions, we can learn an object detector that is robust to the influence of the irrelevant background and makes the cross-domain object recognition much easier than before. In addition, since our detectors are learned in a weakly supervised way, we utilize the classifica-

tion results from the image feature space to enhance our object detectors to avoid performance degradation. Finally, based on the detectors, a self-paced adaptation method is introduced to link the source and target data to further reduce the domain difference. The pipeline of our method is summarized in Figure 2.

Extensive experiments are implemented to evaluate the effectiveness of our method on cross-domain object recognition. Here, we simply present our experimental results on the data from the 10 common classes among Bing-Caltech [1] (denoted by **B**) and the four domains (denoted by **A**, **D**, **W** and **C**) in Office-Caltech [2, 4]. For each image, we extract deep convolutional features based on an ImageNet pre-trained CNN [3]. We denote a cross-domain image recognition problem by the notation S→ T, where S and T denote the source and target domains respectively. Figure 3 shows a comparison of the results of different methods on the 4 domain pairs. As can be seen, our method (Self-paced adaptation) and its two variants (Our detector and Detector enhancement) show significant improvement compared to other related domain adaptation method, which validates the effectiveness of our method.

[1] Lorenzo Torresani Alessandro Bergamo. Exploiting weakly-labeled web images to improve object classification: a domain adaptation approach. In *NIPS*, 2010.

[2] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. *Technical Report*, 2007.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

[4] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *ECCV*, 2010.