

Image Domain Adaptation Using Metric Learning

Kate Saenko, Brian Kulis, Mario Fritz and Trevor Darrell
UC Berkeley EECS and ICSI

We propose a novel metric-based approach to domain adaptation, with an application to visual object recognition. Domain adaptation aims to transfer classifiers learned on one domain to another domain (e.g. [1]). In object recognition, this could mean learning object category classifiers on existing image datasets, and adapting them to images from a particular environment where an object recognition system is deployed. Since the user cannot be expected to provide enough training examples for all object categories in that particular environment, it would be desirable to utilize auxiliary datasets from the web and other domains to learn category models. However, large object category datasets are typically mined from internet search engines (e.g. *Caltech 256*, *ImageNet*), and the images they contain are not necessarily representative of the types of images captured in real-world surroundings¹ (see Figure 1.)

We propose compensating for image domain mismatch by applying machine-learning methods to automatically adapt existing object models to new domains. We introduce a novel adaptation technique, which leverages labeled source domain data together with a small amount of labeled target domain data to learn a distance metric that maps between the two domains. The target domain labels could come from the user as part of the initial system “enrollment” in a new domain, or could be transparently collected as part of user-robot interaction. In addition to being one of the first studies of domain adaptation for object recognition, this work develops a general adaptation technique that could be applied to non-image data.

Our method performs *adaptive transfer* of category knowledge from labeled datasets acquired in one domain to other environments. We assume n tasks d_i , which consist of categorical labels y assigned to input observations \mathbf{x} . In our setting, we have many such labeled tasks in the source domain, $D^s = \{d_1^s, \dots, d_n^s\}$, and a subset of the tasks in the target domain, $D^t = \{d_1^t, \dots, d_m^t\}$, where $m \ll n$. Our goal is to adapt the predictive function $f^s(\mathbf{x}) = y$ trained on the tasks $m+1, \dots, n$, which only have source domain labels, to obtain the function $f^t(\mathbf{x})$, which minimizes the predictive error on the target domain by accounting for the domain shift. Without restricting the form of the predictive function, we only assume that it operates over distances between examples, and propose to

¹Web images are typically shot in commercial settings, have little blurring or occlusions, and the objects are usually centered in the image in some canonical pose; they also usually have studio lighting, blank backgrounds, and bright colors. On the other hand, a robot in the real world would probably encounter images with poor lighting, blurring, clutter, etc.

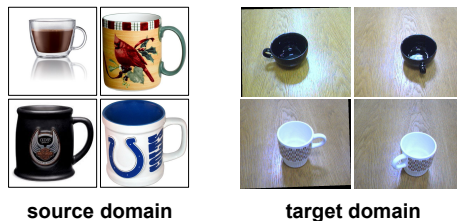


Figure 1: Example web (source) and office (target) images of the *cup* category.

learn a distance metric that places examples from different domains that belong to the same category closer together. Our approach can be thought of as a form of knowledge transfer from the source to the target domain. However, in contrast to many existing transfer learning paradigms (e.g. [3]), we do not presume any degree of relatedness between the tasks that are used to learn the transferred structure, $\{d_i^s, d_i^t\}$, $i \leq m$, and the tasks to which the structure is transferred, $\{d_j^t\}$, $j > m$. (Individual tasks are related across domains, of course; the key point is that we are transferring the structure of the domain shift, not transferring structures common to related categories.)

We learn a representation which minimizes the effect of shifting between source and target domains using a novel metric learning approach. Given the m tasks with labels in both domains, we generate constraints between pairs of object examples (one from the source and one from the target) that should be considered either similar or dissimilar, and learn a metric that appropriately satisfies such constraints. We adapt the information theoretic method of [2] to this problem. This method learns a Mahalanobis distance function $d_A(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T A (\mathbf{x}_i - \mathbf{x}_j)$, parametrized by a positive semi-definite matrix A , such that the constraints are satisfied. This method uses a particular regularizer for positive semi-definite matrices called the LogDet divergence, which has been shown to have desirable properties for metric learning, and the algorithm can easily be applied over large-scale data. We compare the performance of raw vs. adapted source domain classifiers on the tasks for which there is no target domain labeled data.

References

- [1] J. Blitzer, M. Dredze, and F. Pereira. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL*, 2007.
- [2] J. Davis, B. Kulis, P. Jain, S. Sra, and I. Dhillon. Information-Theoretic Metric Learning. In *Proc. ICML*, 2007.
- [3] M. Stark and M. Goesele and B. Schiele. A Shape-Based Object Class Model for Knowledge Transfer. In *Proc. ICCV*, 2009.

Topic: visual processing and pattern recognition; Preference: oral