

Towards Learning Free Naive Bayes Nearest Neighbor-Based Domain Adaptation

Faraz Saeedan¹ and Barbara Caputo^{1,2}(✉)

¹ Sapienza University, Rome, Italy
caputo@dis.uniroma1.it

² Idiap Research Institute, Martigny, Switzerland

Abstract. As of today, object categorization algorithms are not able to achieve the level of robustness and generality necessary to work reliably in the real world. Even the most powerful convolutional neural network we can train fails to perform satisfactorily when trained and tested on data from different databases. This issue, known as domain adaptation and/or dataset bias in the literature, is due to a distribution mismatch between data collections. Methods addressing it go from max-margin classifiers to learning how to modify the features and obtain a more robust representation. Recent work showed that by casting the problem into the image-to-class recognition framework, the domain adaptation problem is significantly alleviated [23]. Here we follow this approach, and show how a very simple, learning free Naive Bayes Nearest Neighbor (NBNN)-based domain adaptation algorithm can significantly alleviate the distribution mismatch among source and target data, especially when the number of classes and the number of sources grow. Experiments on standard benchmarks used in the literature show that our approach (a) is competitive with the current state of the art on small scale problems, and (b) achieves the current state of the art as the number of classes and sources grows, with minimal computational requirements.

Keywords: Naive Bayes Nearest Neighbor · Domain adaptation · Transfer learning

1 Introduction

In the last years the attention of the visual recognition community has been driven towards the existence of differences across predefined image datasets, and the necessity to recompose these idiosyncrasies. The main reason behind this need is the increasing amount of available image data sources and the absence of a unique general learning method that can perform well across all of them. In practice, training a classifier on a dataset (e.g. Flickr photos) and testing on another (e.g. images captured with a mobile phone) produces very poor results although the task (i.e. the set of depicted object categories) is the same.

In this context the notion of *domain*, already used in machine learning for speech and language processing, has been extended to visual problems. A source

domain (S) usually contains a large amount of labeled images, while a target domain (T) refers broadly to a dataset that is assumed to have different characteristics from the source, and few or no labeled samples. Formally we can say that two domains differ when for their probability distributions it holds $P_S(x, y) \neq P_T(x, y)$, where $x \in \mathcal{X}$ indicates the generic image sample and $y \in \mathcal{Y}$ the corresponding class label. Specific annotator tendencies may influence the conditional distributions, implying $P_S(y|x) \neq P_T(y|x)$. Other typical causes of visual domain shift include changes in the acquisition device, image resolution, lighting, background, viewpoint and post-processing [27]. Most of these information are directly encoded in the descriptor space \mathcal{X} chosen to represent the images and may induce a difference among the marginal distributions $P_S(x) \neq P_T(x)$.

In 2013, Tommasi and Caputo showed that by casting the domain adaptation problem into the Naive Bayes Nearest Neighbor framework (NBNN, [5]) one could achieve a very high level of generalization, thanks to the intrinsic properties of NBNN classifiers [23]. The proposed approach used distance metric learning to leverage over the source knowledge at the local patch level. This brought strong results in the semi-supervised and unsupervised domain adaptation scenarios, but the method is computationally expensive and thus not suitable to work on real-time systems, like smartphones or robots.

Here we propose a simple, learning free domain adaptation method that makes it possible to exploit the generalization power of NBNN in the domain adaptation setting. We leverage over the source patches by randomly selecting a subset of them, and adding them to the target patches. To further increase the descriptive power of the descriptors, we perform data augmentation both on the source and the target data, as it is standard practice in the convolutional neural network literature [8]. The combined effect of these two simple actions is remarkable: on commonly used benchmark databases, our approach is on par with the current state of the art when there is a single source from which to adapt, and when the number of classes is limited. In the more challenging and more realistic settings of multiple sources, combined with a large number of classes, our algorithm achieves the state of the art.

The rest of the paper is organized as follows: after reviewing previous work (section 2) we revise the basic definitions for domain adaptation (section 3.1) and the NBNN framework (section 3.2). Section 4 introduces our approach, while section 5 presents a thorough experimental evaluation. We conclude with a summary discussion and outlining possible future avenues for research.

2 Related Work

The problem of domain adaptation stems from the fact that supervised learning methods fail to generalize across datasets [27]. Although this problem exists in various applications [2, 4, 9, 20], the visual recognition community has just recently shown interest in dealing with it [3, 14, 21, 25]. Failure to generalize across datasets has been attributed to the mismatch among various characteristics of the considered databases, and is usually referred to as the ‘dataset bias’ problem [27]. The fact that different image datasets vary considerably in quality,

point of view and image contents, reveals that addressing the domain adaptation problem can significantly improve the performance of visual recognition applications.

Several approaches have been adopted for reducing the distance between datasets. These approaches vary from transferring source data to the target domain [21] or transferring both source and target to a third space [14]. Unfortunately, despite all efforts, [12] showed that sophisticated learning methods, aiming at learning the optimal subset of source data to transfer to the target set, in practice do not offer significant improvement over random transfers. Alternative approaches to the instance-based strategy have been explored in [6, 10, 16, 28].

While the image-to-image paradigm is the dominant approach in the methods mentioned above, recent work suggested that an alternative classification paradigm, called image-to-class, achieves higher generalization across databases compared to the very popular Bag-of-Words (BoW) approach [17, 19]. This classification paradigm, whose most popular algorithm is the Naive Bayes Nearest Neighbor classifier, has been tested on several visual learning applications. Still, its use in domain adaptation has been limited. Only in 2013, Tommasi and Caputo [23] exploited its potential in a metric learning approach, and showed that using NBNN, one can easily surpass the state of the art among BOW-based algorithms presented so far. A drawback of the algorithm is its computational complexity: once the amount of classes, the number of sources and the number of data for each class and source grow, using DA-NBNN becomes computationally prohibitive. Our approach overcomes these computational limitations while preserving, and often significantly surpassing, the performance of DA-NBNN, proposing the first learning free NBNN-based domain adaptation method in the literature.

3 Problem Setting and Definitions

In this section we set the scene by introducing formal definitions for the domain adaptation problem (section 3.1) and NBNN (section 3.2).

3.1 Domain Adaptation

Domain Adaptation is the problem where knowledge from the source domain \mathcal{D}^s is used to enrich and hence improve the performance in the target domain \mathcal{D}^t . The source knowledge might be in the form of instances or data, or model parameters, or metric induced by the source. It is usually implicitly assumed that labeled data on the target domain does not exist (unsupervised setting) or it is scarce (semi-supervised setting). Although the source and the target domains are different, they use equal label sets $\mathcal{Y}^s = \mathcal{Y}^t$ [26].

The core cause of mismatch between the two domains is attributed to the difference in the distribution of these labels. The conditional probability of the labels, for a given feature representation, are not completely coincident

$P^s(Y|X) \sim P^t(Y|X)$; the marginal data distributions are not equal either $P^s(X) \neq P^t(X)$. In this paper, we will focus exclusively on the semi-supervised setting.

3.2 Naive Bayes Nearest Neighbor

In the NBNN classification framework, it is assumed that for each class there exists a distribution from which local descriptors are drawn independently of one another. This leads to the use of a Naive Bayes maximum a posteriori classifier [5] where each feature m votes for one of the classes in $c = \{1, \dots, C\}$. This voting is realized using the local distance between each feature and its nearest neighbor in class c . $D_f 2C(m, c) = \|f_m - f_m^c\|$. The generalization of this distance concept to image to class distance is straightforward:

$$D_I 2C(F_i, c) = \sum_{m=1}^{M_i} D_f 2C(m, c).$$

The output of the classifier would then be

$$p = \underset{c}{\operatorname{argmin}} D_I 2C(F_i, c) \quad (1)$$

The distance to this optimum class p is called the positive distance while the distances to the rest of the classes $n : \{c \neq p\}$, $D_I 2c(F_i, n)$ are called the negative distances.

4 Learning Free NBNN-Based Domain Adaptation

As outlined above, the problem of domain adaptation emerges when the training data for the target task is scarce. Should it not be the case, any supervised learning algorithm would be capable of learning a classifier, according to its learning abilities. It is also assumed that there exists at least another dataset with enough samples to learn a good classifier (the source), but since the two datasets have been acquired in two different domains, the performance obtained training on the source and testing on the target is weak.

The NBNN algorithm builds support sets for each class from the collection of all the features computed from patches extracted from each of the training examples. Due to the scarcity of the data on the target, the support sets that can be built solely using features from the target samples will not contain enough features to guarantee a solid performance. In order to enrich these support sets, *our proposal is to use also features extracted from the patches of the source images.*

How to select such patches-based features? In [12], the authors investigated a domain adaptation approach based on the idea of landmark samples from the source domain, which are relevant for the modeling of the target classifier.

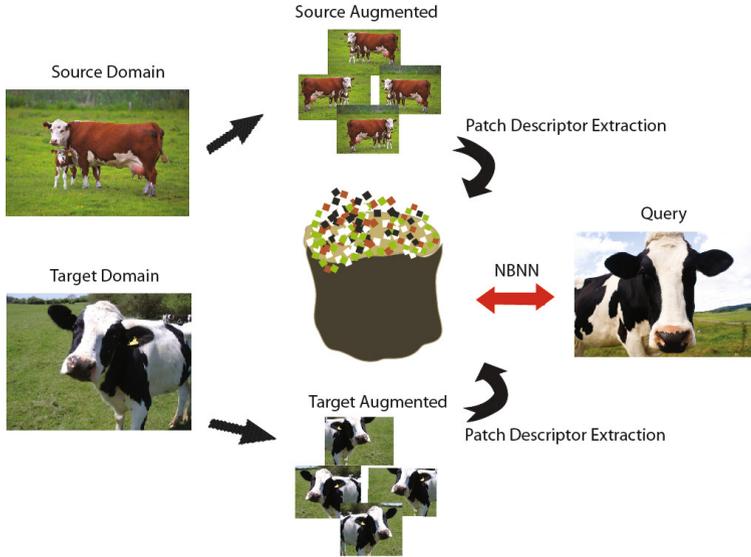


Fig. 1. An overview of our proposed learning free, NBNN-based domain adaptation approach for the class ‘cow’: after performing data augmentation on both the source and target data, patches-based features are extracted from both, and a new target data set is created by merging the whole patches-based features extracted from the target with a fraction of those of the source, randomly selected from the whole sample data. This new pool of patches-based features is then used to build an NBNN classifier in the target domain.

Although their approach is theoretically sound, experiments show that the learning method proposed to select such landmarks is often statistically on par, and otherwise within a two percent range of performance, with a random selection of the learning samples. Motivated by this result, we apply the same philosophy here to the patches-based features, and we propose to achieve domain adaptation in an NBNN-based framework by randomly sampling a percentage of the patches-based features from the source, adding them to the patches-based features of the target. We will show with experiments in the next section that this extremely simple and learning free strategy achieves amazingly good results on standard domain adaptation benchmark databases, while being reasonably stable with respect to the amount of features to be sampled.

To further improve performance, we have tested the effect of performing data augmentation on the source and target data. Data augmentation is a technique that, since the spectacular success of convolutional neural network in the visual classification arena, has been shown to be very effective in general for any classification algorithm [8]. Again, our experiments confirm the effectiveness of this strategy, even more so combined with the instance-based domain adaptation approach based on random sampling of patches-based features from the source. A schematic representation of the overall approach for the class ‘cow’ is given in

figure 4. Note that adding the data augmentation step to our overall approach does not significantly increase the almost non-existent computational load in training. This characteristic, combined with the remarkably good performances achieved especially as the number of classes and sources grow, makes our approach potentially attractive for applications where computational complexity should be low, like mobile robot or online, wearable systems. To the best of our knowledge, there are no previous instance-based, NBNN-based domain adaptation methods in the literature, nor the random sampling strategy has been ever tested in the NBNN learning framework for any learning to learn approach.

5 Experiments

In this section we describe the experiments we performed to assess our approach. We first describe the data, features and experimental setup used (section 5.1), then we report the results obtained (section 5.2). We discuss our findings in section 5.3.

5.1 Datasets, Features and Setup

Datasets. We used the Office dataset, the standard test bed in domain adaptation which addresses the problem of object categorization between any two datasets of objects usually found in offices [21]. This test bed consists of three domains namely Amazon, Webcam and Dslr. The Amazon dataset contains images obtained from online merchants. The images are centered and usually on a white background. Webcam and Dslr are respectively low resolution and high resolution images obtained from web cam and SLR cameras. Unlike Amazon, they could be subject to various environmental disturbances such as lighting or background changes. The Office dataset contain 31 classes of images for each domain.

Having chosen 10 of the original 31 classes from office, [13] suggested that we can add images of the same 10 classes from Caltech-256 [15] and form the Office+Caltech test bed in order to add a fourth domain in the office dataset.

Features. Following the protocol of [23], images were all resized to a common width (256px) and then converted to grayscale. SURF features were extracted according to [1]. The final result was a set of features of length 64 that were consequently fed to a 1-nearest neighbor classifier.

The effect of data augmentation on both domains has also been studied. To this end, we have duplicated the exact procedure suggested in [8] and each image is converted into 10 images through the procedure of cropping and flipping.

Setup. Different pairs of datasets are chosen to act as the source and the target from the Office+Caltech group. From the source dataset, 20 images were selected to represent the source data but only 3 were chosen from the target in every class. When the target was Webcam, 15 images were selected instead of 20 as described in [23]. At this stage, since the Dslr dataset behaves very similarly to Webcam

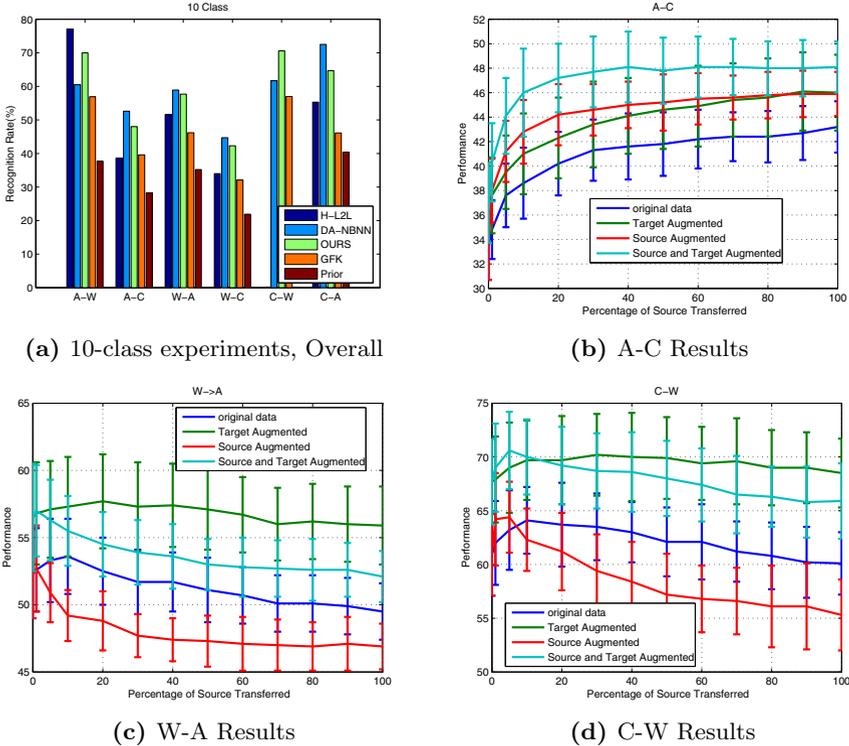


Fig. 2. Results for the 10 class experiments. Figure 2a shows the overall results obtained by our method compared against state of the art algorithms. Figure 2b shows the change in recognition rate on the Amazon-Caltech experiment of our method as the percentage of source data transferred to the target set increases, for the cases no augmentation, only source data augmentation, only target data augmentation and both source and target data augmentation. Analogous results are shown in figure 2c and figure 2d for the Webcam-Amazon and Caltech-Webcam settings respectively.

and it contains a lower number of images, we decided not to include it in our benchmarking. The same sample selection protocol has been adopted for the 31 class adaptation experiments. The third setup that we considered is domain adaptation from more than one source with one target. To this end, all possible combinations of two sources to one target have been examined and benchmarked against the existing reported results in the literature.

5.2 Results

The first set of experiments was done on a subset of Office+Caltech consisting of 10 classes as explained in [23]. Figure 2 shows the results in comparison to the state of the art and some baseline algorithms.

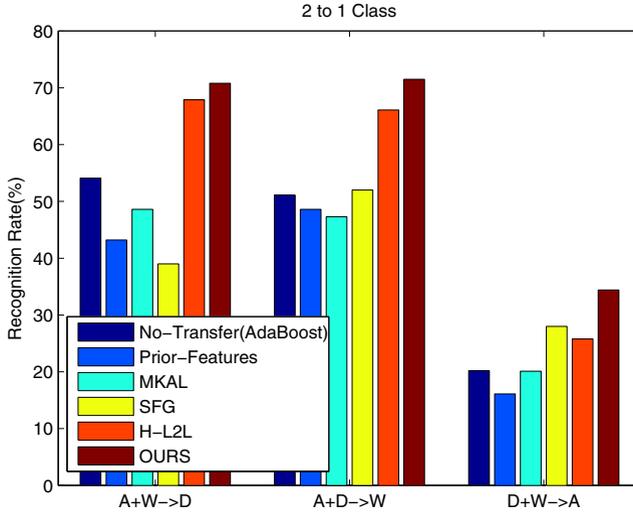


Fig. 3. Accuracy on target domains with multiple sources (A:Amazon, W:Webcam, D:DSLR), 31 class, semi-supervised

Figures 2b, 2c and 2d show the changes in the recognition rate with the increase of the percentage of descriptors, randomly transferred to the target from the source. For a better understanding of the effects of different factors, four cases have been demonstrated together. Original data is where there is no augmentation done neither on the target nor on the source domains. The cases where only the source and only the target domains have been augmented are referred to as Source augmented and Target Augmented respectively. Source and Target Augmented is where both domains have been over-sampled.

The second set of experiments is done on the 31 class Office dataset. The experiments are done exactly inline with what explained and done in [23]. Table 1 shows the results with comparison to the state of the art both using NBNN and the state of the art based on a method other than NBNN. Some further baselines are also included for better comparison.

The Third and last set of experiments are those run using more than one source domain. The Results can be seen in Figure 3. Not all Algorithms can be extended to cover the case of several sources and so only those who had this advantage were included in the comparison. For the experiments the exact test set of [18] has been used.

5.3 Discussion

The biggest advantage of our proposed method is its simplicity combined with its strong performance over growing number of classes and source domains. It also performs surprisingly well in comparison to other algorithms. The results

Table 1. 31 class Office dataset experiments, semi-supervised setting

Algorithm	$A \rightarrow W$	$W \rightarrow D$	$D \rightarrow W$
BOW	34.9 ± 0.6	48.9 ± 0.5	38.4 ± 0.4
GFK	46.4 ± 0.5	66.3 ± 0.4	61.3 ± 0.4
NBNN	40.0 ± 2.0	67.2 ± 2.5	70.7 ± 1.2
I2CDML	47.9 ± 1.3	72.8 ± 2.1	73.8 ± 1.6
$H - L2L(hp - \beta)$	76.2 ± 0.02	67.8 ± 0.05	66.0 ± 0.01
DA-NBNN	52.8 ± 3.7	76.2 ± 2.5	76.6 ± 1.7
OURS	55.0 ± 3.3	77.5 ± 2.0	78.2 ± 1.4

in Figure 2 show that while different algorithms have varying performances on various test settings, our method is never worse than the second best. In particular, compared to DA-NBNN [23] (which is the state of the art among all the methods that exploit an NBNN approach), our method outperforms it in 2 cases (A-W and C-W), while DA-NBNN performs better in two cases (C-A and A-C). In the remaining two cases (W-A, W-C) their performance is close. In fact, the p test shows that in these two experiments there is no statistical evidence of superiority for either of the algorithms.

Our method performs significantly better than L2L [18] where L2L is the state of the art among methods that do not use NBNN. In four of the experiments, L2L achieves inferior results than ours, while only in one setting shows superiority. Note that the accuracy values reported for L2L have been taken from [18], where no result was reported for the C-W experiment.

Using the 31 class Office setting, one can study and compare the scalability of the algorithms with respect to the number of classes. Addressing this type of scalability for our method appears very straightforward. The fact that there is no training, makes things very easy and faster. Table 1 shows that, performance-wise, our method scores higher than DA-NBNN in all three experiments and better than L2L in two out of three cases.

Figure 3 compares the recognition rate for all possible combinations of two sources and one target in the Office dataset. For DA-NBNN it is not clear how it could be extended to this case and no experiments of the kind have been reported by its authors. L2L supports this case and it has been included in the benchmark. It can be seen that our method outperforms all the others for all three cases of experiments. All methods used on the setting $D + W \rightarrow A$ achieve a performance lower than what obtained in the other two settings. This is due to the similarities among the D , W domains, both picturing objects in domestic settings, as opposed to the A domain, which contains images of objects from a commercial website.

An open issue in our method is of course which percentage of the source data should be randomly selected and then added to the target data, in relation to the data augmentation procedure. Results shown in figures 2b-2d show that in general the combination of source plus target data augmentation and random sampling of around 20% of patches-based features from the source seems to

achieve strong performance, always better than the original data. Still, as it can be seen from the figures, the actual optimal performance might vary in terms of percentage of sampling and/or data augmentation strategy for different settings. Although accuracy results are on average quite stable, and therefore the algorithm could be used in online systems even in its current form with good expectations about performance, it would be desirable to explore further the issue of the data selection and find principled ways of selecting the patches to transfer from the source to the target so to have guarantees about the optimality of the procedure. This could be done by learning about the most informative patches in the source, borrowing ideas from [22, 24]. An alternative route might be to borrow acritically all patches, and then use enhanced versions of NBNN to learn what patches are informative for the task at hand, as in [11]. Of course, that would come at the expenses of the current negligible computational cost of the approach.

6 Conclusions

The contribution of this paper is a learning free Naive Bayes Nearest Neighbor based domain adaptation method that is competitive with the current state of the art on the standard Office-Caltech benchmark database, and that achieves the state of the art when the number of classes and sources grows. The method consists in performing a random selection of patches-based local features from the source to the target, combined with a data augmentation strategy mutated from the CNN literature. The resulting algorithm is extremely simple but also remarkably effective, especially when the number of classes and sources grows. An open challenge is how to select the best percentage of source data to add to the target: even though our experimental evaluation indicates that as a rule of thumb sampling around twenty percent of the overall sample data (i.e. after data augmentation) in general leads to very good results, future work will focus on how to determine how much to sample in a principled manner, while at the same time not increasing excessively the computational cost of the approach. This will be investigated on domain adaptation scenarios of increasing complexity, as those presented in [7], which are closer to application needs.

References

1. Bay, H., Ess, A., Tuytelaars, T., Gool, V.: SURF: Speeded up robust features. *CVIU* **110**, 346–359 (2008)
2. Ben-David, S., Blitzer, J., Crammer, K., Pereira, F.: Analysis of representations for domain adaptation. In: *NIPS* (2007)
3. Bergamo, A., Torresani, L.: Exploiting weakly-labeled web images to improve object classification: a domain adaptation approach. In: *NIPS* (2010)
4. Blitzer, J., McDonald, R., Pereira, F.: Domain adaptation with structural correspondence learning. In: *EMNLP* (2006)
5. Boiman, O., Shechtman, E., Irani, M.: In defense of nearest-neighbor based image classification. In: *CVPR* (2008)

6. Bruzzone, L., Marconcini, M.: Domain adaptation problems: A DASVM classification technique and a circular validation strategy. *IEEE PAMI* **32**(5), 770–787 (2010)
7. Caputo, B., Müller, H., Martinez-Gomez, J., Villegas, M., Acar, B., Patricia, N., Marvasti, N., Üsküdarlı, S., Paredes, R., Cazorla, M., Garcia-Varea, I., Morell, V.: ImageCLEF 2014: overview and analysis of the results. In: Kanoulas, E., Lupu, M., Clough, P., Sanderson, M., Hall, M., Hanbury, A., Toms, E. (eds.) *CLEF 2014*. LNCS, vol. 8685, pp. 192–211. Springer, Heidelberg (2014)
8. Chatfield, K., Simonyan, K., Vedaldi, A., Zisserman A.: Return of the Devil in the Details: Delving Deep into Convolutional Nets
9. Daume III, H.: Frustratingly easy domain adaptation. In: *ACL* (2007)
10. Duan, L., Tsang, I.W.-H., Xu, D., Maybank, S.J.: Domain transfer svm for video concept detection. In: *CVPR* (2009)
11. Fornoni, M., Caputo, B.: Scene recognition with naive bayes non-linear learning. In: *Proc. ICPR* (2014)
12. Gong, B., Grauman, K., Sha, F.: Connecting the dots with landmarks: discriminatively learning domain-invariant features for unsupervised domain adaptation. In: *JMLR* (2013)
13. Gong, B., Shi, Y., Sha, F., Grauman, K.: Geodesic flow kernel for unsupervised domain adaptation. In: *CVPR* (2012)
14. Gopalan, R., Li, R., Chellappa, R.: Domain adaptation for object recognition: an unsupervised approach. In: *ICCV* (2011)
15. Griffin, G., Holub, A., Perona, P.: Caltech 256 object category dataset. Technical Report UCB/USD-04-1366, California Institute of Technology (2007)
16. Khosla, A., Zhou, T., Malisiewicz, T., Efros, A.A., Torralba, A.: Undoing the damage of dataset bias. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) *ECCV 2012, Part I*. LNCS, vol. 7572, pp. 158–171. Springer, Heidelberg (2012)
17. Ni, J., Qiu, Q., Chellappa, R.: Subspace interpolation via dictionary learning for unsupervised domain adaptation. In: *CVPR* (2013)
18. Patricia, N., Caputo, B.: Learning to learn, from transfer learning to domain adaptation: a unifying perspective. In: *CVPR* (2014)
19. Qiu, Q., Patel, V.M., Turaga, P., Chellappa, R.: Domain adaptive dictionary learning. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) *ECCV 2012, Part IV*. LNCS, vol. 7575, pp. 631–645. Springer, Heidelberg (2012)
20. Quionero-Candela, J., Sugiyama, M., Schwaighofer, A., Lawrence, N.: *Dataset Shift in Machine Learning*. The MIT Press (2009)
21. Saenko, K., Kulis, B., Fritz, M., Darrell, T.: Adapting visual category models to new domains. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) *ECCV 2010, Part IV*. LNCS, vol. 6314, pp. 213–226. Springer, Heidelberg (2010)
22. Tommasi, T., Caputo, B.: The more you know, the less you learn: from knowledge transfer to one-shot learning of object categories. In: *Proc. BMVC* (2009)
23. Tommasi, T., Caputo, B.: Frustratingly easy NBNN domain adaptation. In: *ICCV* (2013)
24. Tommasi, T., Orabona, F., Caputo, B.: Learning categories from few examples with multi model knowledge transfer. *IEEE Transaction on PAMI* **36**(5), 928–941 (2014)
25. Tommasi, T., Orabona, F., Castellini, C., Caputo, B.: Improving control of dexterous hand prostheses using adaptive learning. *IEEE Transaction on Robotics*, pp. 1–13 (2013)

26. Tommasi, T., Quadrianto, N., Caputo, B., Lampert, C.H.: Beyond dataset bias: multi-task unaligned shared knowledge transfer. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012, Part I. LNCS, vol. 7724, pp. 1–15. Springer, Heidelberg (2013)
27. Torralba, A., Efros, A.A.: Unbiased look at dataset bias. In: CVPR (2011)
28. Yang, J., Yan, R., Hauptmann, A.G.: Cross-domain video concept detection using adaptive svms. In: ACM Multimedia (2007)