Learning Partially Shared Dictionaries for Domain Adaptation

Viresh Ranjan[1] Gaurav Harit[2] C. V. Jawahar[1]

[1]CVIT, IIIT Hyderabad, India [2]IIT Jodhpur, India

Abstract. Real world applicability of many computer vision solutions is constrained by the mismatch between the training and test domains. This mismatch might arise because of factors such as change in pose, lighting conditions, quality of imaging devices, intra-class variations inherent in object categories etc. In this work, we present a dictionary learning based approach to tackle the problem of domain mismatch. In our approach, we jointly learn dictionaries for the source and the target domains. The dictionaries are partially shared, i.e. some elements are common across both the dictionaries. These shared elements can represent the information which is common across both the domains. The dictionaries also have some elements to represent the domain specific information. Using these dictionaries, we separate the domain specific information and the information which is common across the domains. We use the latter for training cross-domain classifiers i.e., we build classifiers that work well on a new target domain while using labeled examples only in the source domain. We conduct cross-domain object recognition experiments on popular benchmark datasets and show improvement in results over the existing state of art domain adaptation approaches.

1 Introduction

Visual object recognition schemes popularly use feature descriptor such as SIFT [1], HOG [2] followed by a classification strategy such as SVMs [3]. They train on a set of annotated training set images and evaluate on a set of similar images for quantifying the performance. However, such object recognition schemes may perform badly in the case of large variations between the source domain and the target domain [4]. Variations between the source and target domain might arise from changes in pose, illumination or intra-class variations inherent in object categories. In Figure 1, we show sample images of the categories chair and bottle from three different domains, namely Amazon, DSLR and Webcam [5]. The domain Amazon is visually very different from the other two domain, the reason being large intra-class variations. The difference between the domains DSLR and Webcam arises because of change in pose, camera quality and lighting conditions.

To tackle the issue of variations across the source and target domains, various domain adaptation (DA) techniques have been proposed in the natural language processing as well as computer vision communities. In Figure 2, we present the overall idea behind a general DA approach. The figure depicts the idea that a

Fig. 1. Sample images of categories "bottle" and "chair" from the domains Amazon, DSLR and Webcam [5]. Images from Amazon are visually very different in comparison to the other two domains. Visual mismatch between DSLR and Webcam is relatively less and arises from factors such as changes in pose, image resolution and lighting conditions.

classifier trained on the source domain may need further adaptation in order to perform well on the target domain. In the natural language processing community, DA techniques have been applied for tasks such as sentiment classification, parts of speech tagging etc. Blitzer *et al.*[6] present a DA technique to modify discriminative classifiers trained on the source domain to classify samples from the target domain. The primary aspect of their work is identifying the *pivot* features, i.e. those features which occur frequently and behave similarly across the two domains. Hal Daume [7] presents a feature augmentation approach where source, target and a common domain representation are obtained by replicating the original feature. Jiang and Zhai [8] present an instance weighting approach where they prune misleading examples from the source domain and give more weight to the labeled examples from the target domain.

In recent years, there has been a surge of interest in the visual domain adaptation task. Several DA strategies have been proposed which adapt either the feature representation or the classifier. These strategies are semi-supervised or unsupervised depending on whether some labeled data from the target domain is available or not. Utilizing labeled examples from source as well as target domains, Saenko *et al.*[5] learn a transformation to map vectors from one domain to another. This transformation tries to bring closer the intra-class vectors from the two domains and push the inter-class vectors farther apart. In a similar feature transformation based approach, Kulis *et al.* [9] do an extension of this work in which the vectors in the two domains can have different dimensions. Unlike the previous two works, the feature transformation based approach presented by Fernando *et al.* [10] is completely unsupervised. They model the source subspace by the eigenvectors obtained by doing PCA over the source domain and similarly for the target subspace. They align the source subspace with the target subspace by learning a transformation matrix. The source and the target domain samples are then projected to their corresponding aligned subspace. Gopalan *et al.* [11] present a feature augmentation based DA approach where the source and the

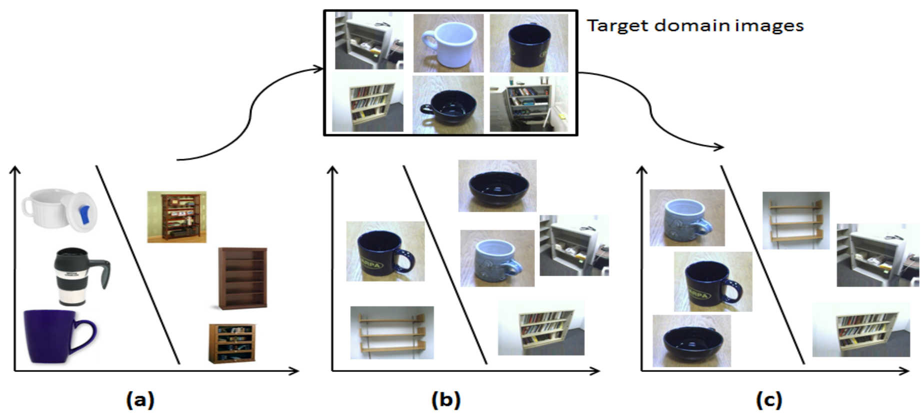

Fig. 2. Overall idea behind a Domain Adaptation approach is shown. Source and target domains are Amazon and Webcam respectively. The two object categories are mug and bookcase. Fig(a) shows a classifier which perfectly separates the two object categories in the source domain. Fig(b) shows the same classifier misclassifies images from the target domain. Fig(c) shows the scenario after domain adaptation, the classifier now correctly classifies the target domain images. The target domain images aid the DA strategy. These examples can be labeled or unlabeled depending on whether the DA approach is semi-supervised or unsupervised.

target subspaces are modeled as points on a Grassmann manifold. They sample points along the geodesic between the source subspace and the target subspace to obtain intermediate subspaces. The data points are projected along all the intermediate subspaces to obtain a domain independent representation. Gong *et al.* [12] propose a geodesic flow kernel based approach and instead of sampling finite number of subspaces along the geodesic from source subspace to target subspace, they integrate over infinite number of intermediate subspaces. Jhuo *et al.* [13] present a semi-supervised DA approach based on low-rank approximation. The samples from the source domain are mapped to an intermediate representation where the transformed source samples can be expressed as a linear combination of target samples. The authors consider single source domain as well as multiple source domain scenarios in this work. Apart from the feature adaptation based DA techniques, classifier adaptation based DA techniques have also been proposed. Yang *et al.* [14] present a classifier adaptation based DA approach where they adapt a source domain SVM classifier by using few labels from the target domain.

Recently, sparse representation has been used for various visual DA tasks such as object recognition, face recognition [15, 16] and action recognition [17]. Zheng *et al.* [17] propose a dictionary learning approach for doing cross-domain action recognition. Given correspondence between videos from two domains, i.e. videos of same action shot from two different views, they learn two separate dictionaries while forcing the sparse representation for corresponding video frames from the two domains to be same. Using this view independent representation, action model learned from the source view video can be directly applied on the target view video. Ni *et al.* [16] present a dictionary learning based DA approach when correspondence information across domains is not available. Given a dictionary in one domain, say source, they iteratively modify the dictionary to be suitable for the target domain. They store all the intermediate dictionaries and use all of them to obtain a view independent representation of images from both the domains. Shekhar *et al.* [15] present a dictionary learning based approach where they map samples from both the domains to a low dimensional subspace and learn a common dictionary by minimizing reconstruction error for the projected samples in the low dimensional subspace.

In our current work, we present a dictionary learning approach for learning partially shared dictionaries across different domains. We learn separate dictionaries for the source and target domains. These dictionaries have some shared atoms which represent the common information which is present in both the domains. The dictionaries also have some domain specific atoms to represent the domain specific information. We show the effectiveness of our dictionary learning strategy by using it for the cross-domain classification task. The domain specific information can cause confusion while doing cross-domain classification. Hence, we ignore the domain specific information and use the representation obtained from the common dictionary elements for training cross-domain classifier. The highlights of our approach are

1. We present a strategy for jointly learning partially shared dictionaries across domains.
2. We design the dictionaries to have two types of elements, i.e. domain specific elements and domain independent elements. As the name suggests, the domain specific atoms represent the domain specific information whereas the domain independent elements capture the information common to both the domains.
3. A form of selective block sparsity arises naturally from the partially shared dictionary learning formulation. More specifically, depending on the underlying domain of the signal, a specific block of sparse coefficients is forced to consist only of zeros. A simple strategy for obtaining sparse representation in presence of selective block sparsity is given.
4. Our dictionary learning approach can be seen as making few modifications over an existing dictionary learning approach [18]. However, using this simple approach, we obtain comparable results to the state of the art visual DA approaches.

2 Domain Adaptation using Partially Shared Dictionaries

2.1 Method Overview

A dictionary learned from the source domain might not be suitable for representing signals from the target domain. Using such a dictionary for representing

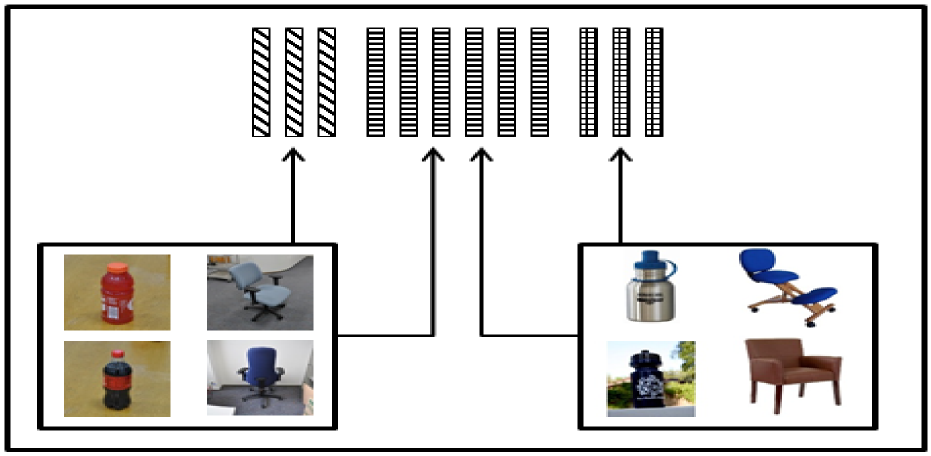

Fig. 3. Overview of partially shared dictionary learning. Dictionary for each domain consists of two types of atoms, domain specific atoms and atoms shared across the domains. Shared atoms are learned using samples from both the domains whereas domain specific atoms are learned using samples from the corresponding domain.

both the domains might result in a scenario where the sparse representation obtained for the same class signals from the two domains are very different. Clearly, such a representation will lead to poor cross-domain classification performance. Hence, while designing dictionaries in the presence of domain mismatch, further steps are required to accommodate signals from the new domains. We have presented a partially shared dictionary learning strategy to tackle the issue of domain mismatch. Our strategy is based on the idea that there could be some commonalities between the source and target domains. The same set of dictionary atoms can be used to represent this common information. Also, the signals from a domain will have certain domain specific aspects. This domain specific information can be represented well by dictionary atoms which are exclusive to the particular domain. Using the common atoms for sparse decomposition will lead to similar representation for the same class signals from both the domains. Hence, such a representation is more suited for the cross-domain classification task. The overview of our approach is shown in Figure 3. As shown in the figure, some atoms are shared across the dictionaries from the source and the target domains. Apart from these common atoms, the dictionaries also have some domain specific atoms.

2.2 Sparse Representation of Signals

A signal $y \in R^n$ can be sparsely represented using a dictionary $D \in R^{n \times K}$, consisting of K atoms or prototype signals. The atoms of D can be pre-defined using discrete cosine transform basis [19], wavelets [20] or they can be learned from the available signals. The learned dictionaries have been shown to perform better than pre-defined dictionaries for tasks such as reconstruction [21]. For learning dictionaries from the data, several efficient dictionary learning strategies such as K-SVD [22] and MOD [18] have been proposed in the past. These dictionary learning techniques solve the following optimization problem

$$\min_{D,A} \|Y - DA\|_F^2 \quad \text{subject to} \quad \forall i, \quad \|a^i\|_0 \leq T_0. \tag{1}$$

Here the signals are arranged along the columns of Y and the columns of A, i.e. a^i, contain the corresponding sparse representation. The dictionary learning techniques solve this problem by alternating between solving for A, i.e. sparse coding step and updating D, i.e. dictionary update step. In [18], the dictionary update consists of updating all the dictionary elements while keeping the sparse representation unchanged. The dictionary learning approach given in [22], however, updates a single dictionary atom at a time. The sparse coefficients also change during the update so that the number of nonzero coefficients further reduces or remains the same. The sparse decomposition problem with the l_0 penalty is NP hard and greedy algorithms are used to solve this. When D is fixed, sparse representation a^i can be obtained using greedy pursuit algorithms such as OMP [23]. Sparse decomposition can also be done by relaxing the l_0 penalty and using a l_1 penalty in its place [24].

2.3 Partially Shared Dictionary Learning

Dictionary learned from one visual domain might not be suitable for representing signals from another visual domain. Hence, we propose a dictionary learning strategy which jointly learns a dictionary which is suitable for both the source as well as target visual domains. We believe that examples from any domain can be represented effectively using a dictionary which has some domain specific atoms as well as some domain independent atoms, i.e. which are common across domains. This assumption is supported by the observation that instances of same category across different domains generally have some similarity between them. Hence, we represent the source domain dictionary D_s and the target domain dictionary D_t as

$$D_s = [D_{src} \ \ D_c] \ ; \quad D_t = [D_{tgt} \ \ D_c], \tag{2}$$

where D_{src}, D_{tgt} are source and target domain specific atoms and D_c are the common atoms across the two domains. Also, we represent the combined dictionary D as

$$D = [D_{src} \ \ D_c \ \ D_{tgt}]. \tag{3}$$

The objective for jointly learning D is given as given as

$$\begin{aligned} \min_{D,A,B} \quad & \|[Y_s \ \ Y_t] - D[A \ \ B]\|_F^2, \\ \text{subject to} \quad & a^i_{tgt} = [0\ 0 \ .... \ \ 0]^T, \ b^j_{src} = [0\ 0 \ .... \ \ 0]^T, \\ & \|a^i\|_0 \leq T_0, \ \|b^j\|_0 \leq T_0, \end{aligned} \tag{4}$$

where $a^i = \begin{bmatrix} a^i_{src} \\ a^i_{com} \\ a^i_{tgt} \end{bmatrix}$, $b^j = \begin{bmatrix} b^j_{src} \\ b^j_{com} \\ b^j_{tgt} \end{bmatrix}$, index i corresponds to the $i-th$ source domain sample, i.e. $i-th$ column of Y_s, similarly index j corresponds to the $j-th$ target domain sample. Both the sparse coefficient vectors a^i and b^j can be seen as a concatenation of three blocks of coefficient vectors. Depending upon the underlying domain of the corresponding signal, one of these three blocks, i.e. a^i_{tgt} or b^j_{src}, is forced to have all elements as zero. The equality constraints thus give rise to a specific form of block sparsity [25], which we call selective block sparsity. The above optimization problem allows for jointly learning the source as well as the target domain dictionaries. The equality constraint $a^i_{tgt} = [0\ 0\ ....\ 0]^T$ makes sure that the dictionary atoms D_{tgt} are used only for representing the target domain signals Y_t. Hence, D_{tgt} captures only the target domain information. Similarly, the equality constraint $b^j_{src} = [0\ 0\ ....\ 0]^T$ makes sure that D_{src} captures only the source domain information. The block of sparse coefficients a^i_{com} and b^j_{com} correspond to the common dictionary atoms D_c. As both a^i_{com} and b^j_{com} can have non-zero terms, the dictionary atoms D_c are used while representing signals from both the domains, hence, these atoms capture the common information across the source and target domains.

To effectively solve the optimization problem given in Equation 4, we rewrite it as

$$\begin{aligned} \min_{D_s, D_t, A_s, B_t} \quad & \|Y_s - D_s A_s\|_F^2 \quad + \quad \|Y_t - D_t B_t\|_F^2, \\ \text{subject to} \quad & \|a^i_s\|_0 \le T_0,\ \|b^j_t\|_0 \le T_0, \end{aligned} \tag{5}$$

where $a^i_s = \begin{bmatrix} a^i_{src} \\ a^i_{com} \end{bmatrix}$, $b^j_t = \begin{bmatrix} b^j_{tgt} \\ b^j_{com} \end{bmatrix}$, a^i_s is the $i-th$ column of A_s, b^j_t is the $j-th$ column of B_t. We would like to point out here that to learn D_s and D_t using Equation 5, one might be tempted to use MOD and alternate between sparse coding and dictionary update by taking derivative of the energy term with respect to D_s and D_t. However, such an approach would not ensure the structure we desire to be present in the dictionaries D_s and D_t, as presented in Equation 2. We take a short digression to describe how to solve Equation 4 in case the dictionary structure given in Equation 2 is not present. In such a scenario, we can rewrite the optimization problem given in Equation 4 as Equation 5. For dictionary learning, we can use MOD. Obtaining a^i_s and b^j_t is straightforward and these can be obtained via OMP. If a^i_s and b^j_t are available, a^i and b^j can be obtained trivially by concatenating a vector of zeros at appropriate position.

Now we get back to our original dictionary learning formulation. To maintain the desired structure in the two dictionaries D_s and D_t, we further couple the two dictionaries D_s and D_t using the following relation between the two

$$D_t = D_s P, \tag{6}$$

where P is a square matrix. Since we want some elements to be common among D_s and D_t, we fix a set of the columns of P, i.e. Q, to have a single element as 1 and remaining elements as 0. The location of 1 in each column of Q corresponds

Algorithm 1 Partially Shared Dictionary Learning(PSDL)

Input: source domain vectors Y_s, target domain vectors Y_t, n
Output: D, A_s, B_t
  initialize: D_s, $P = [R \ \ Q]$
  for i= 1 **to** n **do**
    $A_s \leftarrow OMP(Y_s, D_s)$
    $\begin{bmatrix} B_{tgt} \\ B_{com} \end{bmatrix} = B_t \leftarrow OMP(Y_t, D_s P)$
    $D_s \leftarrow (Y_s A_s^T + Y_t B_t^T P^T)(A_s A_s^T + P B_t B_t^T P^T)^{-1}$
    $R \leftarrow (D_s^T D_s)^{-1} D_s^T E_t B_{tgt}^T (B_{tgt} B_{tgt}^T)^{-1}$
  end for

to location of common atoms D_c. Hence P can be represented as

$$P = [R \quad Q]. \tag{7}$$

Using Equation 7, Equation 5 can be rewritten as

$$\begin{aligned} \min_{D_s,R,A_s,B_t} \quad & \|Y_s - D_s A_s\|_F^2 \quad + \quad \|Y_t - D_s Q B_{com} - D_s R B_{tgt}\|_F^2, \\ \text{subject to} \quad & \|a_s^i\|_0 \le T_0, \ \|b_t^j\|_0 \le T_0, \end{aligned} \tag{8}$$

where $B_t = \begin{bmatrix} B_{tgt} \\ B_{com} \end{bmatrix}$.

To solve this optimization problem, we alternate between updating D_s and R followed by sparse coding step. We set the first order derivative with respect to D_s equal to zero and obtain the following closed form expression for D_s

$$D_s = (Y_s A_s^T + Y_t B_t^T P^T)(A_s A_s^T + P B_t B_t^T P^T)^{-1}. \tag{9}$$

Similarly, the update for R is done using the following closed form expression.

$$R = (D_s^T D_s)^{-1} D_s^T E_t B_{tgt}^T (B_{tgt} B_{tgt}^T)^{-1}, \tag{10}$$

where $E_t = Y_t - D_s Q B_{com}$. In the sparse coding step, D_s and R is kept fixed and OMP is used to obtain the sparse representation. We summarize our partially shared dictionary learning (PSDL) strategy in Algorithm 1.

2.4 Cross-Domain Classification using PSDL

Using unlabeled data from the source and the target domains, the dictionary D is learned as described in the previous section. The dictionary atoms which are common across the two domains, i.e. D_c, are then used to obtain sparse representations for signals from both the domains. The sparse decomposition using D_c maps signals from both the domains to a common subspace. The sparse representation of samples from source and target domain, thus obtained, are used directly for doing cross-domain classification. The coefficients corresponding to D_{src} and D_{tgt} are ignored while doing cross-domain classification.

The dictionary atom subsets D_{src} and D_{tgt} represent the domain specific information, hence, using their coefficients also for the cross-domain classification task will create confusion for the classifier. By using just the coefficients corresponding to D_c, we effectively extract only the common information which is shared across the source and target domains. This results in similar sparse representation for same class signals across the two domains. Clearly, such a representation is better suited for the crosss-domain classification task.

As stated before, we use just the coefficients corresponding to D_c for representing signals from the source and the target domains. The classifiers are trained using the sparse representation for plenty of labeled data from the source domain as well as a small amount of labeled data from the target domain. We use SVMs for the cross-domain classification task, as in [16].

3 Results and Discussions

We validate our approach by conducting object recognition experiments in a cross-domain setting on benchmark datasets. We conduct the experiments using the same experimental setup as in [12, 16]. Our dictionary learning approach PSDL is unsupervised, and does not use any label information from the source or the target domains. We compare our dictionary learning based DA approach with various baseline approaches as well as a recently proposed dictionary learning based DA approach [16]. We also compare our approach with other DA techniques [11, 12].

3.1 Dataset and Representation

We conduct object recognition experiments on 4 datasets, i.e. Amazon(images downloaded from online merchants), Webcam(images taken by a low resolution webcam), DSLR(images taken by a digital SLR camera) and Caltech(images taken from the Caltech-256 [26] dataset). The first three datasets were introduced in [5] whereas the fourth one was first studied by [12]. Each of the dataset are considered as a separate domain. Datasets consist of images pertaining to the following 10 classes BACKPACK, TOURING-BIKE, CALCULATOR, HEADPHONES, COMPUTER-KEYBOARD, LAPTOP, COMPUTER-MONITOR, COMPUTER-MOUSE, COFFEE-MUG, VIDEO-PROJECTOR. There are atleast 8 images and a maximum of 151 images per category in each domain. The datasets consist of a total of 2533 images.

Scale invariant interest points were detected in the images using the SURF detector [27]. A 64 dimensional SURF descriptor was used to describe the patch around the interest points. A codebook consisting of 800 visual words was constructed by clustering random descriptors from the Amazon dataset, using k-means clustering. A histogram representation was obtained for each of the images by obtaining the count of each of the visual words in the image. All the histograms were z-score normalized to have zero mean and unit deviation along each dimension.

3.2 Experiments

For experiments, two domains are picked from the datasets. We use one of them as the source domain and the other is used as the target domain. Goal of the experiments is to classify target domain data points. We conduct experiments in unsupervised setting as well as semi-supervised setting. In the unsupervised setting, labeled examples are present only in the source domain. In the semi-supervised setting, along with the labeled source examples, we also sample few labeled examples from the target domain. When Webcam or DSLR are the source domains, we sample 8 labeled points from them. In case Amazon or Caltech are the source domains, 20 labeled examples are sampled. In semi-supervised setting, 3 labeled examples are sampled from the target domain. For dictionary learning using PSDL, we utilize unlabeled samples from both the domains. The optimal parameters for PSDL(number of dictionary atoms) and for the SVM classifier are obtained by empirically searching over the parameter space. Sparse representation is obtained using Orthogonal Matching Pursuit(OMP) [23]. Following the previous works [12, 16, 11] , all the experiments are repeated 20 times and the mean classification accuracy over the 20 trials is reported in each case.

Table 1. Classification accuracies for PSDL is compared with baseline approaches as well as the DA approaches given in [11, 12, 16]. For baseline approaches we learn source and target domain dictionaries using [18]. The acronyms A, C, D, W represent the domains Amazon, Caltech, DSLR and Webcam respectively. In the notation C $\rightarrow$ A, C is the source domain and A is the target domain. Similar notation is followed for the other dataset pairs. Experiments are done in semi-supervised setting.

Method	C $\rightarrow$ A	C $\rightarrow$ D	A $\rightarrow$ C	A $\rightarrow$ W	W $\rightarrow$ C	W $\rightarrow$ A	D $\rightarrow$ A	D $\rightarrow$ W
SVM$_S$	40.6	45.7	36.5	26.5	23.2	30.6	35.3	69.7
SVM$_{ST}$	46.4	52.1	38.9	38.7	31.0	40.2	42.3	74.1
MOD$_{source}$	44.9	50.5	39.2	46.6	27.3	38.5	37.6	67.2
MOD$_{target}$	49.2	53.6	39.4	50.7	34.2	44.4	44.3	72.0
SGF[11]	40.2	36.6	37.7	37.9	29.2	38.2	39.2	69.5
GFK[12]	46.1	55.5	39.6	56.9	32.8	46.2	46.2	80.2
Ni *et al.*[16]	50.0	57.1	41.5	57.8	40.6	51.5	50.3	87.8
PSDL	53.9	59.4	41.8	57.7	37.0	46.8	48.9	83.3

In Table 1, we compare our approach with baseline approaches as well as other popular DA approaches. This experiment is done in a semi-supervised setting, i.e. we use few labeled examples from the target domain along with the labeled examples from the source domain. SVM$_S$, SVM$_{ST}$ and MOD based dictionary learning approaches are taken as the baseline in this experiment. SVM$_S$

Table 2. Classification accuracies for PSDL is compared with baseline approaches as well as the DA approaches given in [11, 12, 16]. The acronyms A, C, D, W represent the domains Amazon, Caltech, DSLR and Webcam respectively. In the notation C $\rightarrow$ A, C is the source domain and A is the target domain. Similar notation is followed for the other dataset pairs. Experiments are done in unsupervised setting.

Method	C $\rightarrow$ A	C $\rightarrow$ D	A $\rightarrow$ C	A $\rightarrow$ W	W $\rightarrow$ C	W $\rightarrow$ A	D $\rightarrow$ A	D $\rightarrow$ W
MOD$_{source}$	39.8	42.1	37.0	36.2	19.8	26.8	30.1	55.3
MOD$_{target}$	44.4	44.0	36.8	38.2	30.5	35.4	34.5	69.5
SGF[11]	36.8	32.6	35.3	31.0	21.7	27.5	32.0	66.0
GFK[12]	40.4	41.1	37.9	35.7	29.3	35.5	36.1	79.1
Ni *et al.*[16]	45.4	42.3	40.4	37.9	36.3	38.3	39.1	86.2
PSDL	47.6	48.5	39.8	38.9	31.8	36.0	37.9	79.1

refers to SVM classifiers learned using only the source examples, SVM$_{ST}$ refers to SVM classifiers learned using source as well as target domain examples. We use MOD for learning dictionaries from the source as well as the target domains. For MOD$_{source}$, sparse decomposition of signals from both the domains is done using the dictionary learned from just the source domain. Similarly, for MOD$_{target}$, the dictionary learned from the target domain is used for sparse representation. We also present the results given in the dictionary learning based DA approach given in [16]. We also compare our results with two other DA approaches [11, 12]. Our dictionary learning approach as well as [16], always outperform the baseline approaches as well as the DA approaches given in [11, 12]. For the dataset pair Webcam and DSLR, all the approaches perform better compared to the other dataset pairs. The reason for this high classification accuracy is the high similarity across these two domains, i.e. these datasets consist of images of the same object instances obtained using different imaging devices. On the other hand, all the approaches tend to show low accuracy for some dataset pairs, for example Amazon and Caltech. This can be explained by the large variations across these two domains. We observe that for the first four dataset pairs, our method performs almost as good or better than [16]. For the remaining dataset pairs, [16] outperforms our method. In Table 2, results are reported for experiments in unsupervised setting. For the unsupervised setting, PSDL as well as [16] outperform the baseline approaches as well as [11, 12]. Like semi-supervised setting, here also PSDL lags behind [16]for the last four cases. For these cases, training domain has 8 labeled examples whereas for the first four cases, it has 20 examples. In the objective given in Eqn 8, the target reconstruction term may dominate over the source term when the source has fewer samples than the target domain. For tasks 5-8 (in Table 1 and Table 2), there are only 8 samples in the source domain. We believe this leads to a common dictionary which represent the target domain well but not the source domain.

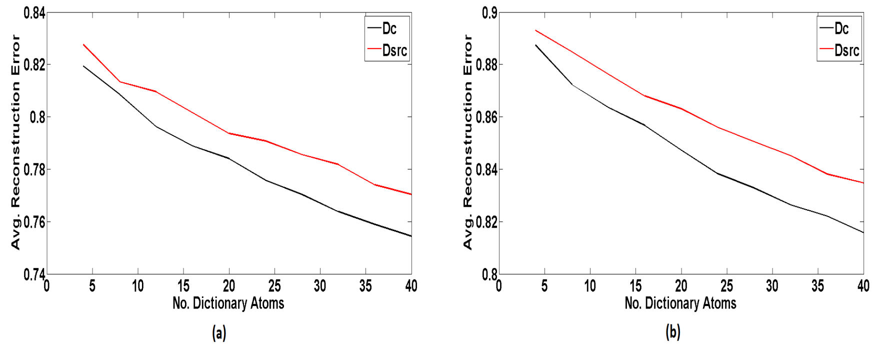

Fig. 4. Average reconstruction error for target dataset is plotted as a function of number of dictionary atoms. Dc represents the shared dictionary and $Dsrc$ the source domain dictionary. The shared dictionary Dc does a better reconstruction of the target domain samples in comparison to $Dsrc$. For (a), the source and target domains are Amazon and Caltech respectively and for (b), Caltech and Amazon.

In Figure 4, we show the comparison between the average reconstruction error obtained by dictionaries D_c and D_{src} while representing the target domain. Irrespective of the dictionary size, D_c results in less reconstruction error, thus showing that it is more suited for representing the target domain. In Figure 5, we show three test images from the target domain and corresponding top five nearest neighbors from the source domain. We provide nearest neighbors for two cases, i.e. results obtained via our dictionary learning based approach and no adaptation case using original features. In all the three examples, improvement because of our approach is clearly observable.

4 Conclusion

We present a partially shared dictionary learning approach. Our approach allows dictionaries from the source and the target domains to share some atoms. We use the shared atoms to represent signals from both the domains. This results in similar sparse representation for same class signals across the two domains. Our results show that such a representation is better suited for cross-domain visual object recognition task.

We show the effectiveness of our approach by performing cross-domain object recognition on the benchmark datasets. We also compare our approach with various existing approaches and show improvement in results over the existing state of art approaches. For our future work, we plan to learn discriminative shared dictionaries by utilizing available label information.

Fig. 5. Test images from the DSLR domain on the left and two adjacent rows of corresponding nearest neighbors from the Amazon domain. The top row adjacent to each query image is obtained using our dictionary learning approach, bottom row adjacent to each query corresponds to nearest neighhbors obtained with original features.

References

1. Lowe, D.G.: Object recognition from local scale-invariant features. In: Computer vision, IEEE international conference on, IEEE (1999)
2. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition. CVPR. IEEE Computer Society Conference on, IEEE (2005)

3. Perronnin, F., Sánchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification. In: ECCV. (2010)
4. Torralba, A., Efros, A.A.: Unbiased look at dataset bias. In: Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, IEEE (2011)
5. Saenko, K., Kulis, B., Fritz, M., Darrell, T.: Adapting visual category models to new domains. In: ECCV. Springer (2010)
6. Blitzer, J., McDonald, R., Pereira, F.: Domain adaptation with structural correspondence learning. In: Proceedings of the conference on empirical methods in natural language processing, Association for Computational Linguistics (2006)
7. Daumé III, H.: Frustratingly easy domain adaptation. In: ACL. (2007)
8. Jiang, J., Zhai, C.: Instance weighting for domain adaptation in nlp. In: ACL. (2007)
9. Kulis, B., Saenko, K., Darrell, T.: What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. In: Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, IEEE (2011)
10. Fernando, B., Habrard, A., Sebban, M., Tuytelaars, T., et al.: Unsupervised visual domain adaptation using subspace alignment. Proceedings of ICCV (2013)
11. Gopalan, R., Li, R., Chellappa, R.: Domain adaptation for object recognition: An unsupervised approach. In: Computer Vision (ICCV), IEEE International Conference on, IEEE (2011)
12. Gong, B., Shi, Y., Sha, F., Grauman, K.: Geodesic flow kernel for unsupervised domain adaptation. In: Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, IEEE (2012)
13. Jhuo, I.H., Liu, D., Lee, D., Chang, S.F.: Robust visual domain adaptation with low-rank reconstruction. In: Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, IEEE (2012)
14. Yang, J., Yan, R., Hauptmann, A.G.: Cross-domain video concept detection using adaptive svms. In: Proceedings of the 15th international conference on Multimedia, ACM (2007)
15. Shekhar, S., Patel, V.M., Nguyen, H.V., Chellappa, R.: Generalized domain-adaptive dictionaries. In: Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, IEEE (2013)
16. Ni, J., Qiu, Q., Chellappa, R.: Subspace interpolation via dictionary learning for unsupervised domain adaptation. In: Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, IEEE (2013)
17. Zheng, J., Jiang, Z., Phillips, P.J., Chellappa, R.: Cross-view action recognition via a transferable dictionary pair. In: BMVC. (2012)
18. Engan, K., Aase, S.O., Hakon Husoy, J.: Method of optimal directions for frame design. In: Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on, IEEE (1999)
19. Ahmed, N., Natarajan, T., Rao, K.R.: Discrete cosine transform. Computers, IEEE Transactions on (1974)
20. Mallat, S.: A wavelet tour of signal processing. Academic press (1999)
21. Elad, M., Aharon, M.: Image denoising via learned dictionaries and sparse representation. In: Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, IEEE (2006)
22. Aharon, M., Elad, M., Bruckstein, A.: -svd: An algorithm for designing overcomplete dictionaries for sparse representation. Signal Processing, IEEE Transactions on (2006)
23. Tropp, J.A., Gilbert, A.C.: Signal recovery from random measurements via orthogonal matching pursuit. Information Theory, IEEE Transactions on (2007)

24. Donoho, D.L.: Compressed sensing. Information Theory, IEEE Transactions on (2006)
25. Eldar, Y.C., Kuppinger, P., Bolcskei, H.: Block-sparse signals: Uncertainty relations and efficient recovery. Signal Processing, IEEE Transactions on (2010)
26. Griffin, G., Holub, A., Perona, P.: Caltech-256 object category dataset. (2007)
27. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). (Computer vision and image understanding)