Uniform Low-Rank Representation for Unsupervised Visual Domain Adaptation

Pengcheng Liu, Peipei Yang, Kaiqi Huang, Tieniu Tan
Center for Research on Intelligent Perception and Computing,
National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences
pengcheng.liu@ia.ac.cn, {ppyang,kqhuang,tnt}@nlpr.ia.ac.cn

Abstract

Visual domain adaptation aims to adapt a model learned in source domain to target domain, which has received much attention in recent years. In this paper, we propose a uniform low-rank representation based unsupervised domain adaptation method which captures the intrinsic relationship among the source and target samples and meanwhile eliminates the disturbance from the noises and outliers. In particular, we first align the source and target samples into a common subspace using a subspace alignment technique. Then we learn a domain-invariant dictionary with respect to the transformed source and target samples. Finally, all the transformed samples are low-rank represented based on the learned dictionary. Extensive experimental results show that our method is beneficial to reducing the domain difference, and we achieve the state-of-the-art performance on the widely used visual domain adaptation benchmark.

1. Introduction

Visual domain adaptation is devoted to addressing the problem originated from a distribution mismatch between the source (training) and target (testing) data, which has long been one of the challenging problems in computer vision. Domain adaptation is crucial for the success of adapting a model learned in the source domain to the target domain, which presents significant value both in theory and practice. Taking the object recognition task for example, the distribution mismatch problem is usually caused by the situation that training and test samples are acquired under different sets of background, lighting, view point, resolution conditions, etc. As a consequence, the performance of the model learned in the source domain degrades considerably on the target domain.

In the last few years, many solutions for visual domain adaptation have been proposed to overcome that problem. We roughly divide them into three categories. One kind of approach is to learn a new domain-invariant feature representation by exploring a common projection space [14, 1, 19, 10, 16, 18]. Methods like manifold-alignment [20, 21] and Maximum Mean Discrepancy (MMD) [7] are also introduced to learn that new representation. Another kind of approach, sample re-weighting [11] or selection [4, 2], devotes to assigning the optimal weights to the source samples so as to minimize the difference between the target and weighted source distribution. In another approach, to capture the intrinsic domain shift between the source and target domains, a mapping function [3] between the source and target subspaces or a set of intermediate subspaces [6, 5, 9] is learned to link the two domains.

The previous proposals typically reduce the domain distribution difference by exploiting the source and target samples separately without accounting for the mutual dependency among them. This may cause the adapted distribution to be arbitrarily scattered and the structure information among the samples in both source and target domains may become undermined. In addition, most of them blindly learn the adapted distribution based on all the samples including the noises and possible outliers. Thus, it is hard for them to learn an optimal adaptation so that the source model can be applied in the target domain without significant performance degradation.

In this paper, we propose a uniform low-rank representation based visual domain adaptation method which not only tries to explore the intrinsic structure information of the samples in the source and target domains but also is robust to the influence from the noises and outliers. Work [12] is the most related to ours. However, it is specially designed for semi-supervised classification task while our method is unsupervised. First, we project both source and target samples into a common subspace using a subspace alignment technique, which can preliminarily reduce the domain difference. Then we learn a domain-invariant dictionary based on the transformed samples. After that, both the transformed source and target samples can be linearly reconstructed based on the common dictionary. Upon this linear

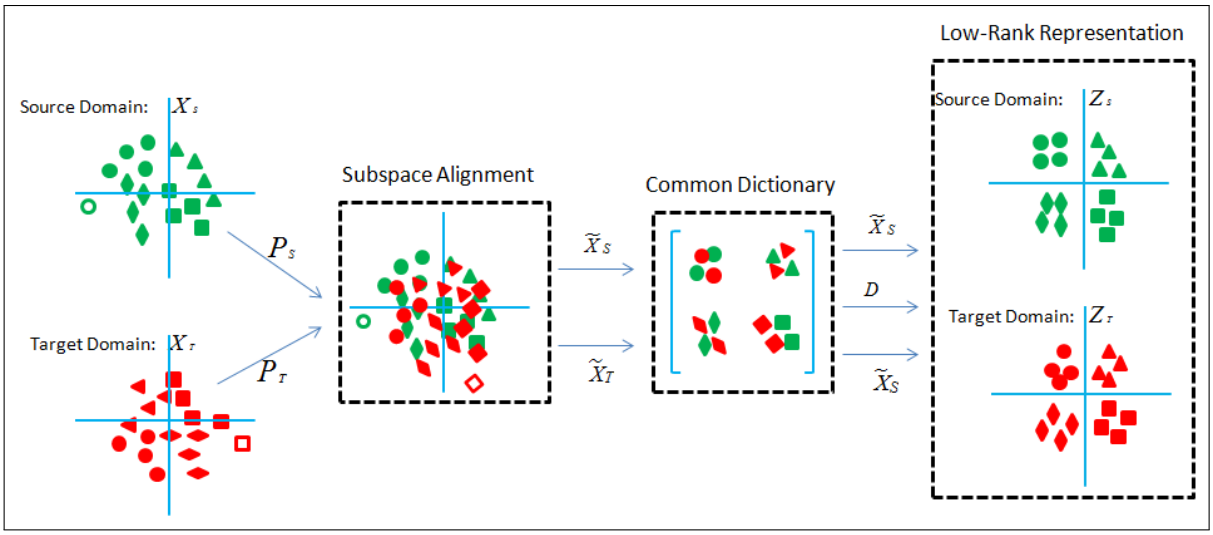

Figure 1. Pipeline of our method. Best viewed in color.

reconstruction, we adopt a low-rank structure to capture the intrinsic relationship among the source and target samples and meanwhile eliminate the disturbance from the noises and outliers. The whole procedure of learning the low-rank representation is unsupervised because we do not utilize any label information. Extensive experimental results demonstrate that our method is beneficial to reducing the domain difference, and shows the state-of-the-art performance on the widely used domain adaptation benchmark dataset.

2. Proposed method

In this section, we will present our uniform low-rank representation based domain adaptation method in detail. The pipeline of our method is summarized in Figure 1. In order to generate the low-rank representation which not only captures the intrinsic structure information of the source and target samples but also eliminates the disturbance from the noises and outliers, we firstly project both source and target samples into an aligned subspace so that we can obtain a domain-invariant dictionary. Then the transformed source and target samples are reconstructed in the low-rank subspace based on the dictionary. The rest of this section will describe our method step by step.

2.1. Subspace alignment and dictionary generation

In order to obtain a domain-invariant dictionary, we first project both source and target samples into an aligned subspace, which can also preliminarily reduce the distribution difference between the source and target domains. Denote a set of n_s samples in source domain as $X_S \in \mathbb{R}^{D\times n_s}$, and a set of n_t samples in target domain as $X_T \in \mathbb{R}^{D\times n_t}$, where D is the dimension of the feature vector, and each column of X_S or X_T represents a sample. Inspired by the work [3], using PCA, we project X_S and X_T to their respective subspaces by the operations $P_S^\top X_S$ and $P_T^\top X_T$ ($P_S, P_T \in \mathbb{R}^{D\times d}$), where P_S and P_T consist of d eigenvectors corresponding to the d largest eigenvalues. Then, we aim to find a linear transformation matrix so that the bases of the two subspaces are aligned. Denoting M as the transformation matrix from P_S to P_T, our subspace alignment is equivalent to solving the following minimization problem

$$M^* = \arg\min_M \frac{\lambda}{2}\|M\|_F + \frac{1}{2}\|P_S M - P_T\|_{\mathrm{F}}^2,$$

where $\|.\|_{\mathrm{F}}$ is the Frobenius norm and λ is the tradeoff parameter to avoid overfitting.

Since the above formulation is a least-square problem, it is simple to get the optimal solution $M^* = (\lambda I + P_S^\top P_S)^{-1} P_S^\top P_T$. Projecting the source data X_S and target data X_T into the aligned subspace using $P_S M^*$ and P_T respectively, we can get the new representations of the transformed source data $\widetilde{X}_S = (P_S M^*)^\top X_S$ and target data $\widetilde{X}_T = P_T^\top X_T$.

The subspace alignment presented above can reduce the domain difference in some degree. However, the mutual dependency among source and target samples has not been taken into account. Thus, as shown in Figure 1, a domain-invariant dictionary D is generated, which would be beneficial to exploring the intrinsic structure information among the samples. Here, we get the dictionary D using K-Means clustering on the new representation $\widetilde{X}_S$ and $\widetilde{X}_T$ in the domain invariant space. Finally, $D \in \mathbb{R}^{d\times K}$ consists of the K clusters.

2.2. Low-rank representation for unsupervised domain adaptation

In this section, we aim to capture the intrinsic structure information among the source and target samples and mean-

while eliminate the disturbance from the noises and outliers. Assuming each observed sample to be composed of the clean sample reconstructed with D and the noise, the aim becomes to the estimation of both the parts. To this end, we firstly try to find a low-rank reconstruction coefficient matrix Z so that the transformed source and target samples reconstructed with the common dictionary D are highly correlated. This naturally establishes a link between the representations of the related source and target samples. As a consequence, it can capture the intrinsic structure information among all the source and target samples. Secondly, with respect to the clean dictionary D, we are devoted to decomposing the noise and outlier information in both source and target domains into an error matrix E. Since only a small number of the source and target samples are noises or outliers in practice, it is reasonable to ensure that the column vectors of E are as sparse as possible.

To summarize in mathematical, we formulate the estimation of the representations and noises as the following rank minimization problem:

$$\begin{aligned} &\min_{Z,E} \ rank(Z) + \alpha \|E\|_{2,0}, \\ &s.t. \ \widetilde{X} = DZ + E, \end{aligned} \tag{1}$$

where $\widetilde{X} = [\widetilde{X}_S \ \widetilde{X}_T]$, $Z = [Z_S \ Z_T]$ and $E = [E_S \ E_T]$, $rank(\cdot)$ denotes the rank of a matrix, $\|E\|_{2,0}$ denotes the $\ell_{2,0}$ norm, which is adopted to encourage the column sparsity of the error term E, and α is the tradeoff parameter. The optimal solution of variable Z is the lowest-rank representation of the transformed data $\widetilde{X}$ with respect to the domain-invariant dictionary D. It captures the intrinsic structure information among all the source and target samples, and meanwhile it is robust to the noises and outliers.

However, the optimization problem (1) is hard to solve due to the discrete nature and the non-convexity of the rank function and the $\ell_{2,0}$ norm. As suggested in [13], we can replace the rank function in (1) with the nuclear norm and relax the $\ell_{2,0}$ norm to the $\ell_{2,1}$ norm, resulting in the following approximate convex optimization problem:

$$\begin{aligned} &\min_{Z,E} \ \|Z\|_* + \alpha \|E\|_{2,1}, \\ &s.t. \ \widetilde{X} = DZ + E. \end{aligned} \tag{2}$$

The problem (2) can be solved by the Inexact ALM algorithm, which is detailed in [13], and omitted here due to the limit of space. The solution to (2) is a good surrogate solution to (1).

Once we obtain the optimal solution $\widetilde{Z} = [\widetilde{Z}_S \ \widetilde{Z}_T]$ for variable $Z = [Z_S \ Z_T]$, we can learn a source classification model based on the new source representation $\widetilde{Z}_S$. Since the model captures the intrinsic structure information among the source and target samples and meanwhile eliminates the disturbance from the noises and outliers, it would perform well on the new target representation $\widetilde{Z}_T$.

3. Experiments

In this section, we evaluate our method on the widely-used Office-Caltech benchmark [1, 2, 3, 5, 6, 4] for unsupervised and semi-supervised cross-domain object recognition. We compare the proposed method with several competitive ones. Experimental results show that our method is effective for cross-domain object recognition, and we achieve the state-of-the-art performance.

3.1. Dataset and data preparation

The widely-used Office-Caltech dataset for cross-domain image recognition is composed of four domains: Amazon (denoted by $\mathbf{A}$), DSLR (denoted by $\mathbf{D}$), Webcam (denoted by $\mathbf{W}$) and Caltech (denoted by $\mathbf{C}$). The first three domains are from the Office dataset [17], and share 31 common object categories, while the Caltech domain is introduced in [8] and there are 10 common classes among the four domains. The number of images per class ranges from 8 to 151. As suggested by the standard protocol presented in previous studies [17, 5, 4], each image is represented by SURF features encoded with a visual dictionary of 800 words, which were computed via K-means on a subset of Amazon images.

3.2. Parameter settings

In our method, there are 2 key parameters need to be evaluated. First, the dimensionality d for subspace alignment. Second, the number of clusters K for generating the common dictionary. We take two representative domains $\mathbf{A}$ (with clean background) and $\mathbf{C}$ (with complex background) for example, and set the two parameters in turn by 5-fold cross-validation. We first evaluate the parameter d based on our subspace alignment method and an one-vs-all SVM with linear kernel. The corresponding experimental results on two pairs ($\mathbf{A} \rightarrow \mathbf{C}$ and $\mathbf{C} \rightarrow \mathbf{A}$) of cross-domain object recognition are shown in Figure 2(a). According to these results, we set $d = 128$ which performs best on both domain pairs. Then we fix the parameter d, and evaluate the parameter K based on our proposed method with a similar procedure. The recognition accuracies with respect to different K are shown in Figure 2(b). While $K = 32$, our method shows the best performance on both domain pairs. For simplicity, we set $d = 128$ and $K = 32$ for all the domain pairs under an unsupervised domain adaptation setting.

3.3. Experiment on unsupervised domain adaptation

In this section, we focus on the unsupervised domain adaptation setting (no labeled target sample is available while learning the classification model) among the four domains. Since there are only a few samples in DSLR, it is not used as a source domain in most of the previous work. In

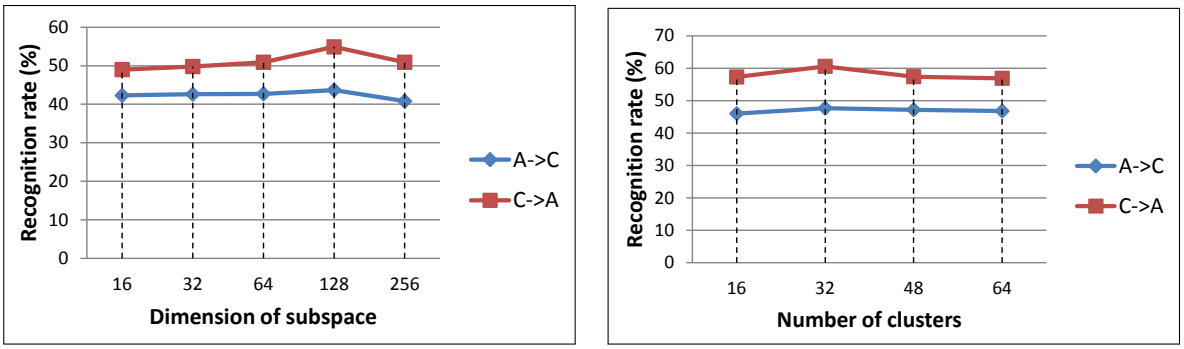

Figure 2. Recognition performance under different parameter settings: (a) dimension of subspace for subspace alignment, (b) number of clusters for the common dictionary.

Method	A→C	A→D	A→W	C→A	C→D	C→W	W→A	W→C	W→D
Baseline	41.7	41.4	34.2	51.8	54.1	46.8	31.1	31.5	70.7
GFS [6]	39.2	36.3	33.6	43.6	40.8	36.3	33.5	30.9	75.7
GFK [5]	42.2	42.7	40.7	44.5	43.3	44.7	31.8	30.8	75.6
LMS [4]	45.5	47.1	46.1	56.7	57.3	49.5	40.2	35.4	75.2
DIP [1]	47.4	50.3	47.5	55.7	60.5	**58.3**	42.6	34.2	88.5
DIP-CC [1]	47.2	49.0	47.8	58.7	61.2	58.0	40.9	37.2	91.7
SIE [2]	**48.2**	49.1	48.1	56.7	61.2	58.0	42.7	38.6	**93.0**
SIE [2]	47.6	49.0	47.8	57.6	61.2	57.3	42.4	36.2	**93.0**
Only SA	43.5	44.0	42.4	54.9	57.9	47.8	38.5	36.0	86.6
Only LRR	45.2	47.1	49.5	58.5	61.1	56.6	41.4	40.1	89.2
Our method	47.7	**51.6**	**52.9**	**60.6**	**64.9**	**58.3**	**43.4**	**40.7**	**93.0**

Table 1. Recognition accuracies on 9 pairs of unsupervised cross-domain object recognition. We evaluate on the 10 common classes using the standard experimental protocol of [5].

this paper, we also focus on the remaining 9 pairs of source (S) and target (T) domains. We denote a cross-domain image recognition problem by the notation S → T. The recognition accuracies on the 10 common classes are shown in Table 1.

The first set of results in Table 1 are quote directly from their papers. It is worth noting that the baseline method is a classification model trained with source data only, i.e., without the domain adaptation.

The second set of results in Table 1 report the recognition performance of our method and its two variants. We use the one-vs-all SVM with linear kernel as the classifier for cross-domain image recognition. The recognition accuracies of only our subspace alignment (SA) based domain adaptation method are presented in the first row, which are comparable to most of the previous reported performance. In the second row, we report the experimental results of only our low-rank representation (LRR) based domain adaptation method which learns the dictionary with respect to the original source and target samples, i.e., our method without the SA part. The recognition accuracies are much better than the results of the only SA based domain adaptation method, and comparable to the state-of-the-art. It means that the LRR based domain adaptation method has the benefits of capturing the intrinsic structure information among all samples and eliminating the disturbance from the noises and outliers. In the last row, our method shows the best performance on most of the domain pairs, which validates the effectiveness of the low-rank representation on solving the domain adaptation problem.

3.4. Experiment on semi-supervised domain adaptation

In this section, we evaluate our method in the semi-supervised way (only a few labeled target samples are available while learning the classification model). We follow the standard experimental protocol from [17, 12, 15], and use the Office dataset consisting of three domains: **A**, **D** and **W**. There are 31 common object categories among the three domains.

In order to learn the classification models, we randomly select 20 samples per category from the source domain **A** and 8 samples per category from **D** or**W** as the source domain while 3 labeled target samples per category are ran-

Method	W→D	A→W	D→W
Baseline	49.6±0.03	50.7±0.03	49.6±0.03
Metric [17]	48.1±0.60	34.5±0.70	36.9±0.80
RDALR [12]	32.9±1.20	50.7±0.80	36.9±1.90
GFS [6]	61.0±0.50	37.4±0.50	55.2±0.60
GFK [5]	66.3±0.40	46.4±0.50	61.3±0.04
H-L2L(LP-β) [15]	67.8±0.05	58.8±0.03	66.0±0.03
Our method	**71.0**±0.02	**59.1**±0.02	**69.1**±0.02

Table 2. Recognition accuracies on 3 pairs of semi-supervised cross-domain object recognition over 31 categories.

domly selected. We evaluate our method based on 10 random splits across all domain adaptation pairs. The average cross-domain image classification accuracies and the standard deviation are reported in Table 2. For all the domain adaptation pairs, we set parameters $d = 256$ and $K = 128$ in a similar cross-validation way introduced before.

In Table 2, the classification model (linear-SVM) of the baseline method is directly trained with the labeled source and target samples, i.e, without domain adaptation method. The recognition results of the compared methods are directly quoted from their papers. Compared with other methods, the proposed method shows the best performance on all the three domain pairs, which validates the effectiveness of our method once again.

4. Conclusion

In this paper, we have presented a novel domain adaptation method. The proposed method captures the intrinsic structure information among all samples and meanwhile eliminates the disturbance from the noises and outliers using the uniform low-rank representation, which is beneficial to reducing the domain difference. Extensive experiments have proved the effectiveness of our method on the cross-domain object recognition. Experimental results show that our method achieve the state-of-the-art performance on the widely used visual domain adaptation benchmark.

Acknowledgement

This work is funded by the National Basic Research Program of China (Grant No. 2012CB316302), National Natural Science Foundation of China (Grant No. 61135002 and Grant No. 61403388), and the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant XDA06040102).

References

[1] M. Baktashmotlagh, M. T. Harandi, B. C. Lovell, and M. Salzmann. Unsupervised domain adaptation by domain invariant projection. In *ICCV*, 2013. 1, 3, 4

[2] M. Baktashmotlagh, M. T. Harandi, B. C. Lovell, and M. Salzmann. Domain adaptation on the statistical manifold. In *CVPR*, 2014. 1, 3, 4

[3] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars. Unsupervised visual domain adaptation using subspace alignment. In *ICCV*, 2013. 1, 2, 3

[4] B. Gong, K. Grauman, and F. Sha. Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation. In *ICML*, 2013. 1, 3, 4

[5] B. Gong, Y. Shi, F. Sha, and K. Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *CVPR*, 2012. 1, 3, 4, 5

[6] R. Gopalan, R. Li, and R. Chellappa. Domain adaptation for object recognition: An unsupervised approach. In *ICCV*, 2011. 1, 3, 4, 5

[7] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola. A kernel two-sample test. *J. Mach. Learn. Res.*, 13:723–773, 2012. 1

[8] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. *Technical Report*, 2007. 3

[9] J. Hoffman, T. Darrell, and K. Saenko. Continues manifold based adaptation for evolving visual domains. In *CVPR*, 2014. 1

[10] J. Hoffman, E. Rodner, J. Donahue, K. Saenko, and T. Darrell. Efficient learning of domain-invariant image representations. In *International Conference on Learning Representations*, 2013. 1

[11] J. Huang, A. J. Smola, A. Gretton, K. M. Borgwardt, and B. Schölkopf. Correcting sample selection bias by unlabeled data. In *NIPS*, 2007. 1

[12] I.-H. Jhuo, D. Liu, D. T. Lee, and S.-F. Chang. Robust visual domain adaptation with low-rank reconstruction. In *CVPR*, 2012. 1, 4, 5

[13] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(1):171–184, 2013. 3

[14] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang. Domain adaptation via transfer component analysis. In *IJCAI*, 2009. 1

[15] N. Patricia and B. Caputo. Learning to learn, from transfer learning to domain adaptation: A unifying perspective. In *CVPR*, 2014. 4, 5

[16] Q. Qiu, V. Patel, P. Turage, and R. Chellappa. Domain adaptive dictionary learning. In *ECCV*, 2012. 1

[17] K. Saenko, B. Kulis, M. Fritz, and T. Darrell. Adapting visual category models to new domains. In *ECCV*, 2010. 3, 4, 5

[18] S. Shekhar, V. M. Patel, H. V. Nguyen, and R. Chellappa. Generalized domain-adaptive dictionaries. In *CVPR*, 2013. 1

[19] Y. Shi and F. Sha. Information-theoretical learning of discriminative clusters for unsupervised domain adaptation. In *ICML*, 2012. 1

[20] C. Wang and S. Mahadevan. Manifold alignment without correspondence. In *IJCAI*, 2009. 1

[21] C. Wang and S. Mahadevan. Heterogeneous domain adaptation using manifold alignment. In *IJCAI*, 2011. 1