Learning Cross-Domain Landmarks for Heterogeneous Domain Adaptation

Yao-Hung Hubert Tsai[1], Yi-Ren Yeh[2], and Yu-Chiang Frank Wang[1]

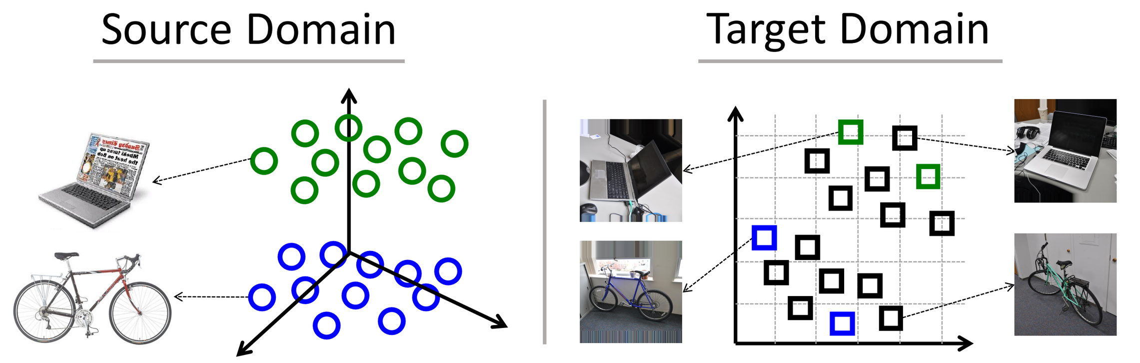
[1]Research Center for IT Innovation, Academia Sinica, Taipei, Taiwan ; [2]Dept. Mathematics, National Kaohsiung Normal University, Kaohsiung, Taiwan

IEEE 2016 Conference on Computer Vision and Pattern Recognition

CVPR 2016

Introduction

- **Domain Adaptation**
  - Aim to **transfer knowledge** across domains (from **source** to **target**)
  - Typically, **labeled** data available in the source domain
  - **Only few or no** labeled data in the target domain
- **Heterogeneous Domain Adaptation (HDA)**
  - Cross-domain data with **distinct** feature representations (e.g., **BoW** v.s. **CNN**)



- **Our Method: Cross-Domain Landmark Selection (CDLS)**
  - We propose to exploit **heterogeneous** source and target-domain data for **learning cross-domain landmarks**.
  - By learning the **adaptability of cross-domain data** (including the unlabeled target-domain data), we derive a **domain-invariant feature space** for HDA.

Related Works

- **Instance Re-weighting or Landmark Selection**
  - TJM [1] and LM [2] (only **homogeneous DA** considered)
- **Supervised HDA**
  - Only **labeled source and target-domain data** available for adaptation
  - HeMap [3], DAMA [4], ARC-t [5], HFA [6], MMDT [7], and SHFR [8]
- **Semi-supervised HDA**
  - **Unlabeled target-domain data** can be jointly exploited during adaptation.
  - HTDCC [9], SHFA [10], SSKMDA [11], SCP [12], and **Ours**

Notations and Motivations

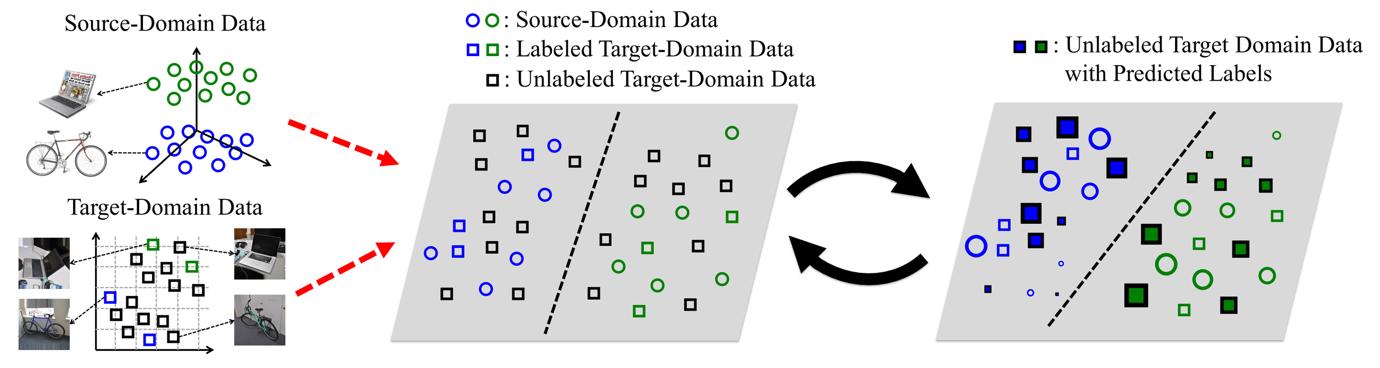
- **Notations**

$$\text{Labeled Source-Domain Data } \mathcal{D}_S = \{(\mathbf{x}_s^1, y_s^1), \dots, (\mathbf{x}_s^{n_s}, y_s^{n_s})\} = \{\mathbf{X}_S, \mathbf{y}_S\}$$
$$\text{Labeled Target-Domain Data } \mathcal{D}_L = \{(\mathbf{x}_l^1, y_l^1), \dots, (\mathbf{x}_l^{n_l}, y_l^{n_l})\} = \{\mathbf{X}_L, \mathbf{y}_L\}$$
$$\text{Unlabeled Target-Domain Data } \mathcal{D}_U = \{(\mathbf{x}_u^1, y_u^1), \dots, (\mathbf{x}_u^{n_u}, y_u^{n_u})\} = \{\mathbf{X}_U, \mathbf{y}_U\}$$
$$\mathcal{D}_T = \mathcal{D}_L \cup \mathcal{D}_U, n_t = n_l + n_u, \text{and } d_s \neq d_t$$

- **Motivations**
  - Learn a **domain-invariant mapping** $\mathbf{A} \in \mathbb{R}^{m \times d_s}$.
  - Determine **landmark weights** $\boldsymbol{\alpha} \in \mathbb{R}^{n_S}$ for $\mathbf{X}_S$ and $\boldsymbol{\beta} \in \mathbb{R}^{n_U}$ for $\mathbf{X}_U$.
  - View $\mathbf{X}_L$ as the **most representative landmarks** (i.e., weights $= \mathbf{1}$).
- **Matching Cross-Domain Data Distributions as in JDA [13]**
  - **Marginal distributions**: $P(\mathbf{X}_T)$ and $P(\mathbf{A}^\top \mathbf{X}_S)$
  - **Conditional distributions**: $P(\mathbf{X}_T|\mathbf{y}_T)$ and $P(\mathbf{A}^\top \mathbf{X}_S|\mathbf{y}_S)$

Cross-Domain Landmarks Selection (CDLS)

- **Illustration**



- **Supervised CDLS**
  - Only $\mathbf{X}_S$, $\mathbf{X}_L$ available and **no landmarks/weights** to be learned
  - Objective function:

$$\min_{\mathbf{A}} E_M(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L) + E_C(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L) + \lambda \|\mathbf{A}\|^2$$

  * Cross-domain marginal data distributions:

$$E_M(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L) = \left\| \frac{1}{n_S} \sum_{i=1}^{n_S} \mathbf{A}^\top \mathbf{x}_s^i - \frac{1}{n_L} \sum_{i=1}^{n_L} \mathbf{x}_l^i \right\|^2$$

  * Cross-domain conditional data distributions:

$$E_C(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L) = \sum_{c=1}^{C} \left\| \frac{1}{n_S^c} \sum_{i=1}^{n_S^c} \mathbf{A}^\top \mathbf{x}_s^{i,c} - \frac{1}{n_L^c} \sum_{i=1}^{n_L^c} \mathbf{x}_l^{i,c} \right\|^2 + \frac{1}{n_S^c n_L^c} \sum_{i=1}^{n_S^c} \sum_{j=1}^{n_L^c} \left\| \mathbf{A}^\top \mathbf{x}_s^{i,c} - \mathbf{x}_l^{j,c} \right\|^2$$

- **Semi-supervised CDLS**
  - Jointly exploit $\mathbf{X}_S$, $\mathbf{X}_L$, and $\mathbf{X}_U$ for **learning cross-domain landmarks** $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$
  - Need to assign **pseudo-labels** $\{\tilde{\mathbf{y}}_u^i\}_{i=1}^{n_U}$ for $\mathbf{X}_U$.
  - Objective function:

$$\min_{\mathbf{A}, \boldsymbol{\alpha}, \boldsymbol{\beta}} E_M(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L, \mathbf{X}_U, \boldsymbol{\alpha}, \boldsymbol{\beta}) + E_C(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L, \mathbf{X}_U, \boldsymbol{\alpha}, \boldsymbol{\beta}) + \lambda \|\mathbf{A}\|^2$$
$$\text{s.t. } \{\alpha_i^c, \beta_i^c\} \in [0,1] \text{ and } \frac{\boldsymbol{\alpha}^{c\top}\mathbf{1}}{n_S^c} = \frac{\boldsymbol{\beta}^{c\top}\mathbf{1}}{n_U^c} = \delta$$

  * $\delta \in [0,1]$ controls the portion of cross-domain data in each class for adaptation.
  * Cross-domain marginal data distributions:

$$E_M(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L, \mathbf{X}_U, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \left\| \frac{1}{\delta n_S} \sum_{i=1}^{n_S} \alpha_i \mathbf{A}^\top \mathbf{x}_s^i - \frac{1}{n_L + \delta n_U} \left(\sum_{i=1}^{n_L} \mathbf{x}_l^i + \sum_{i=1}^{n_U} \beta_i \mathbf{x}_u^i \right) \right\|^2$$

  * Cross-domain conditional data distributions:

$$E_C(\mathbf{A}, \mathcal{D}_S, \mathcal{D}_L, \mathbf{X}_U, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \sum_{c=1}^{C} E_{cond}^c + \frac{1}{e^c} E_{embed}^c$$

with $e^c = \delta n_S^c n_L^c + \delta n_L^c n_U^c + \delta^2 n_U^c n_S^c$

$$E_{cond}^c = \left\| \frac{1}{\delta n_S^c} \sum_{i=1}^{n_S^c} \alpha_i \mathbf{A}^\top \mathbf{x}_s^{i,c} - \frac{1}{n_L^c + \delta n_U^c} \left(\sum_{i=1}^{n_L^c} \mathbf{x}_l^{i,c} + \sum_{i=1}^{n_U^c} \beta_i \mathbf{x}_u^{i,c} \right) \right\|^2$$

$$E_{embed}^c = \sum_{i=1}^{n_S^c} \sum_{j=1}^{n_L^c} \left\| \alpha_i \mathbf{A}^\top \mathbf{x}_s^{i,c} - \mathbf{x}_l^{j,c} \right\|^2 + \sum_{i=1}^{n_L^c} \sum_{j=1}^{n_U^c} \left\| \mathbf{x}_l^{i,c} - \beta_j \mathbf{x}_u^{j,c} \right\|^2 + \sum_{i=1}^{n_U^c} \sum_{j=1}^{n_S^c} \left\| \beta_i \mathbf{x}_u^{i,c} - \alpha_j \mathbf{A}^\top \mathbf{x}_s^{j,c} \right\|^2$$

where E_{embed}^c enforces the **projected** (cross-domain) data similarity

Pseudo-Code

Input: Labeled source and target-domain data $\mathcal{D}_S = \{\mathbf{x}_s^i, \mathbf{y}_s^i\}_{i=1}^{n_S}$, $\mathcal{D}_L = \{\mathbf{x}_l^i, \mathbf{y}_l^i\}_{i=1}^{n_L}$; unlabeled target-domain data $\{\mathbf{x}_u^i\}_{i=1}^{n_U}$; feature dimension m; ratio δ; parameter λ

Derive an m-dimensional subspace via PCA from $\{\mathbf{x}_l^i\}_{i=1}^{n_L}$ and $\{\mathbf{x}_u^i\}_{i=1}^{n_U}$

Initialize $\mathbf{A}$ by **Supervised CDLS** and pseudo-labels $\{\tilde{\mathbf{y}}_u^i\}_{i=1}^{n_U}$

while not converge **do**

  Update **transformation** $\mathbf{A}$

  Update **landmark weights** $\{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$

  Update **pseudo-labels** $\{\tilde{\mathbf{y}}_u^i\}_{i=1}^{n_U}$

end while

Output: Predicted labels $\{\mathbf{y}_u^i\}_{i=1}^{n_U}$ of $\{\mathbf{x}_u^i\}_{i=1}^{n_U}$

Experiment Setup

- **Object Recognition: Office and Caltech-256 Datasets** [14, 15]
  - Images from Amazon (A), Webcam (W), DSLR (D), Caltech-256 (C)
  - Feature: **DeCAF$_6$** (4096 dimensions) versus **SURF** (800 dimensions)
  - **10** overlapping object categories
  - Randomly select **3** images per category for labeled target-domain instances $\mathbf{X}_L$.
- **Text Categorization: Multilingual Reuters Dataset** [16]
  - Source domain: English, French, German, and Italian
  - Target domain: Spanish
  - Feature: **BoW** + TF-IDF with 60% energy preserved via **PCA**
  - **6** categories in **5** languages
  - Randomly select **5/10/15/20** articles per category for $\mathbf{X}_L$.

Evaluation I - Object Recognition

- **Across Feature Spaces**

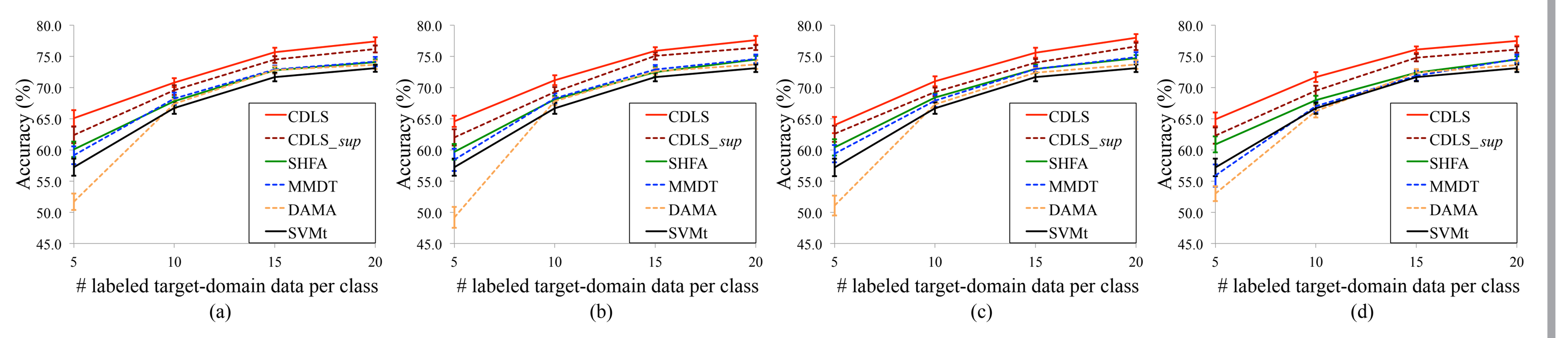
S,T	SVM$_t$	DAMA	MMDT	SHFA	CDLS_*sup*	CDLS
SURF to *DeCAF$_6$*						
A, A	87.3±0.5	87.4±0.5	89.3±0.4	88.6±0.3	86.7±0.6	**91.7±0.2**
W, W	87.1±1.1	87.2±0.7	87.3±0.8	90.0±1.0	88.5±1.4	**95.2±0.9**
C, C	76.8±1.1	73.8±1.2	80.3±1.2	78.2±1.0	74.8±1.1	**81.8±1.1**
DeCAF$_6$ to *SURF*						
A, A	43.4±0.9	38.1±1.1	40.5±1.3	42.9±1.0	45.6±0.7	**46.4±1.0**
W, W	57.9±1.0	47.4±2.1	59.1±1.2	62.2±0.7	60.9±1.1	**63.1±1.1**
C, C	29.1±1.5	18.9±1.3	30.6±1.7	29.4±1.5	31.6±1.5	**31.8±1.2**

- **Across Features & Datasets**

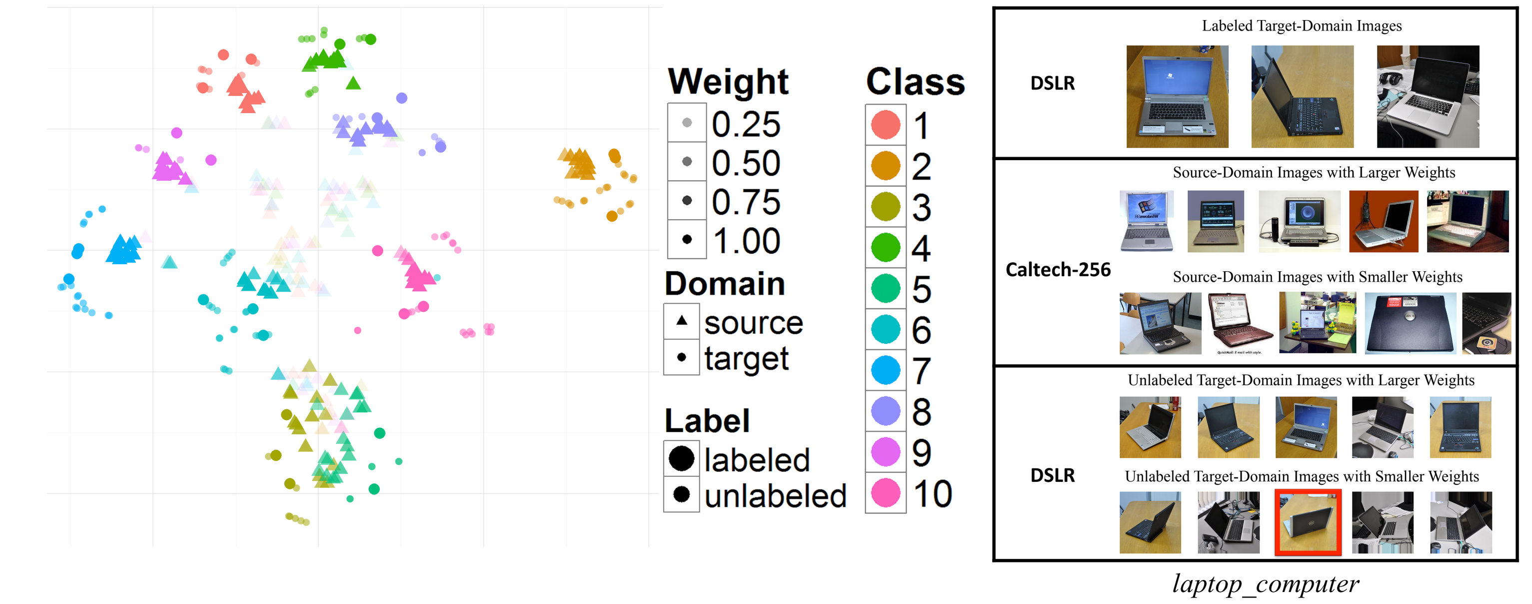
S,T	SVM$_t$	DAMA	MMDT	SHFA	CDLS_*sup*	CDLS
SURF to *DeCAF$_6$*						
A, D	90.9±1.1	91.5±1.2	92.1±1.0	93.4±1.1	92.0±1.2	**96.1±0.7**
W, D		91.2±0.9	91.5±0.8	92.4±0.9	91.0±1.1	**95.1±0.8**
C, D		91.0±1.3	93.1±1.2	93.8±1.0	91.9±1.3	**94.9±1.5**
DeCAF$_6$ to *SURF*						
A, D	54.4±1.0	53.2±1.5	53.5±1.3	56.1±1.0	54.3±1.3	**58.4±0.8**
W, D		51.6±2.2	54.0±1.3	57.6±1.1	57.8±1.0	**60.5±1.0**
C, D		51.9±2.1	56.7±1.0	57.3±1.1	54.8±1.0	**59.4±1.2**

Evaluation II - Text Categorization

- **Text Categorization**
  - Source domain: (a) English, (b) French, (c) German, and (d) Italian



- **Visualizing through t-SNE & Example Images/Landmarks**
  - for Caltech-256 → DSLR (with SURF to DeCAF$_6$)
  - Target-domain image bounded by the red rectangle was mis-classified as *monitor*.



Conclusion

- We presented **Cross-Domain Landmarks Selection (CDLS)** for HDA.
- Our CDLS is able to learn **representative cross-domain landmarks** for **deriving a proper feature subspace** for adaptation and classification purposes.
- Our CDLS **performs favorably** against state-of-the-art HDA approaches.

References

[1] M. Long, *et al.* Transfer joint matching for unsupervised domain adaptation. In *IEEE CVPR*, 2014.
[2] B. Gong, *et al.* Connecting the dots with landmarks: Discriminatively learning domain-invariant features for unsupervised domain adaptation. In *ICML*, 2013.
[3] X. Shi *et al.* Transfer learning on heterogeneous feature spaces via spectral transformation. In *IEEE ICDM*, 2010.
[4] C. Wang and S. Mahadevan. Heterogeneous domain adaptation using manifold alignment. In *IJCAI*, 2011.
[5] B. Kulis *et al.* What you saw is not what you get: Domain adaptation using asymmetric kernel transforms. In *IEEE CVPR*, 2011.
[6] L. Duan *et al.* Learning with augmented features for heterogeneous domain adaptation. In *ICML*, 2012.
[7] J. Hoffman *et al.* Efficient learning of domain-invariant image representations. In *ICLR*, 2013.
[8] J. T. Zhou *et al.* Heterogeneous domain adaptation for multiple classes. In *AISTATS*, 2014.
[9] X. Wu *et al.* Cross-view action recognition over heterogeneous feature spaces. In *IEEE ICCV*, 2013.
[10] W. Li *et al.* Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation. In *IEEE T-PAMI*, 2014.
[11] M. Xiao *et al.* Feature space independent semi-supervised domain adaptation via kernel matching. In *IEEE T-PAMI*, 2015.
[12] M. Xiao *et al.* Semi-supervised subspace co-projection for multi-class heterogeneous domain adaptation. In *ECML*, 2015.
[13] M. Long, *et al.* Transfer feature learning with joint distribution adaptation. In *IEEE ICCV*, 2013.
[14] K. Saenko, *et al.* Adapting visual category models to new domains. In *ECCV*, 2010.
[15] G. Griffin, *el al.* Caltech-256 object category dataset. 2007.
[16] N. Ueffing, *el al.* NRC's PORTAGE system for WMT 2007. In *ACL workshop*, 2007.