Learning an Invariant Hilbert Space for Domain Adaptation

Samitha Herath
samitha.herath@data61.csiro.au
Australian National University & Data61-CSIRO

Mehrtash Harandi
mehrtash.harandi@data61.csiro.au
Australian National University & Data61-CSIRO

Fatih Porikli
fatih.porikli@anu.edu.au
Australian National University

Introduction

- We present an algorithm to address both *unsupervised* and *semi-supervised* domain adaptation problems.
- Our goal is to learn a latent space , $\mathcal{H} \subset \mathbb{R}^p$ in which domain disparities are minimized.
- We show such a space can be learned by **1.** matching the statistical properties of the projected domains (*e.g.*, covariance matrices), **2.** adapting the Mahalanobis metric of the latent space to the labeled data.

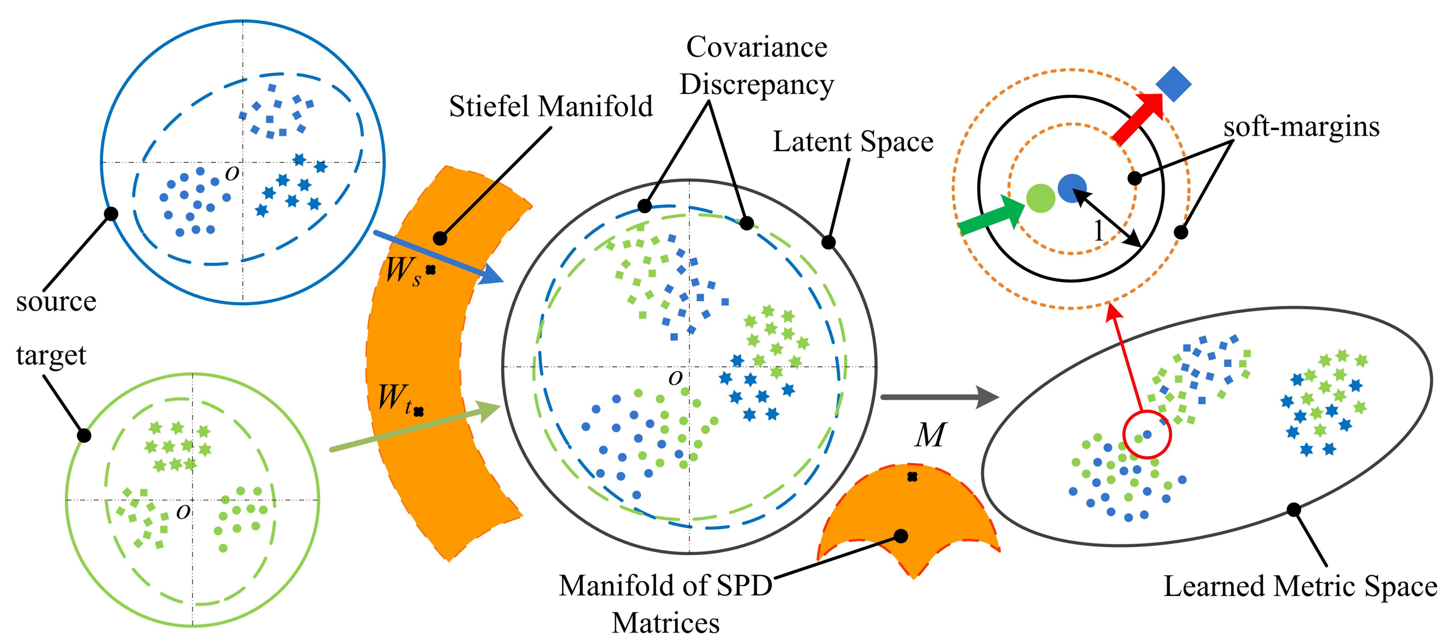


Figure: A diagram of our proposal. The marker shapes represents instance labels and color represents their original domain.

Our Model

- To learn $W_s \in \mathbb{R}^{s \times p}$, $W_t \in \mathbb{R}^{t \times p}$ and $M \in \mathcal{S}_{++}^p$ we propose to minimize a loss function in the form,

$$\mathcal{L} = \mathcal{L}_d + \lambda \mathcal{L}_u; . \quad (1)$$

  - $\mathcal{L}_d$: a discriminative loss defined on labeled similar and dissimilar pairs.
  - $\mathcal{L}_u$: a loss indicating statistical disparity between source and target domains.

- **Discriminative Loss ($\mathcal{L}_d$)**
  - Defined on labeled pairs, $(z_{1,k}, z_{2,k})$; $k = 1, 2, \cdots, N_p$ with, $z_{1,k}, z_{2,k} \in \mathcal{H}$ and a label $y_k \in \{+1, -1\}$.
  - Pulls similarly labeled pairs (*i.e.* $y_k = +1$) closer and pushes dissimilarly labeled pairs (*i.e.* $y_k = -1$) apart.
- We see improvements when,
  - Using a differentiable approximation to the hinge-loss.
  - Appending with softer margins.
- Our discriminative loss term takes the form,

$$\mathcal{L}_d = \frac{1}{N_p} \sum_{k=1}^{N_p} \ell_\beta(M, y_k, z_{1,k} - z_{2,k}, 1 + y_k \epsilon_k) + r(M) + \frac{1}{N_p} \sqrt{\sum \epsilon_k^2}, \quad (2)$$

with,

$$\ell_\beta(M, y, x, u) = \frac{1}{\beta} \log\left(1 + \exp\left(\beta y (x^T M x - u)\right)\right). \quad (3)$$

Here, $r(M)$ is the regularizer for the metric, M and $\epsilon_k \in \mathbb{R}$, $k = 1, 2, ...N_p$ are slack-variables.

- **Loss over Domain Statistics ($\mathcal{L}_u$)**
  - First Order Statistics : Use centered data.
  - Second Order Statistics : Match source and target covariances.

When $\Sigma_s \in \mathcal{S}_{++}^s$ and $\Sigma_t \in \mathcal{S}_{++}^t$ being the covariance matrices of the source and target domains,

$$\mathcal{L}_u = \frac{1}{p} \delta_s(W_s^T \Sigma_s W_s, W_t^T \Sigma_t W_t), \quad (4)$$

with δ_s being the Stein divergence,

$$\delta_s(P, Q) = \log\det\left(\frac{P + Q}{2}\right) - \frac{1}{2} \log\det(PQ), \quad (5)$$

for $P, Q \in \mathcal{S}_{++}$.

Optimization

We optimize the loss function in an **alternating** fashion w.r.t. each model parameter (*i.e.* W_s, W_t, M and the slack variables).

- **Constraints on the Loss Function**
  - **Orthogonality on projection matrices** W_s and W_t is motivated by the common practice in dimensionality reduction.
  - Must follow $M \in \mathcal{S}_{++}$ for the Metric.
- **Using Riemannian Optimization Techniques**
  - **The Stiefel Manifold,** $\mathrm{St}(p, n)$:The set of $(n \times p)$-dimensional matrices, $p \leq n$, with orthonormal columns.
  - **The SPD Manifold,** $\mathcal{S}_{++}^p$:The set of $(p \times p)$ dimensional real, Symmetric Positive Definite matrices.
- **A Faster Solution with Product Topology**

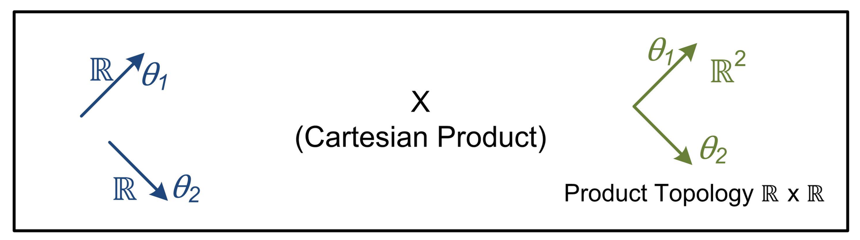
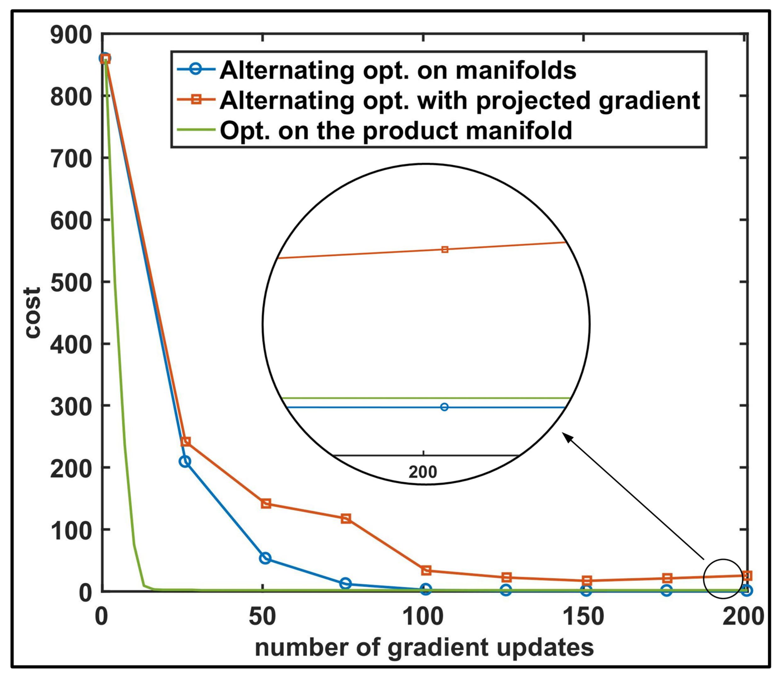


Figure: A demonstration on forming of a product topology.

We form the product topology,

$$\mathcal{M}_{prod.} = \mathrm{St}(p, s) \times \mathrm{St}(p, t) \times \mathcal{S}_{++}^p \times \mathbb{R}^{N_p}, \quad (6)$$

to avoid alternating optimization on model parameters.



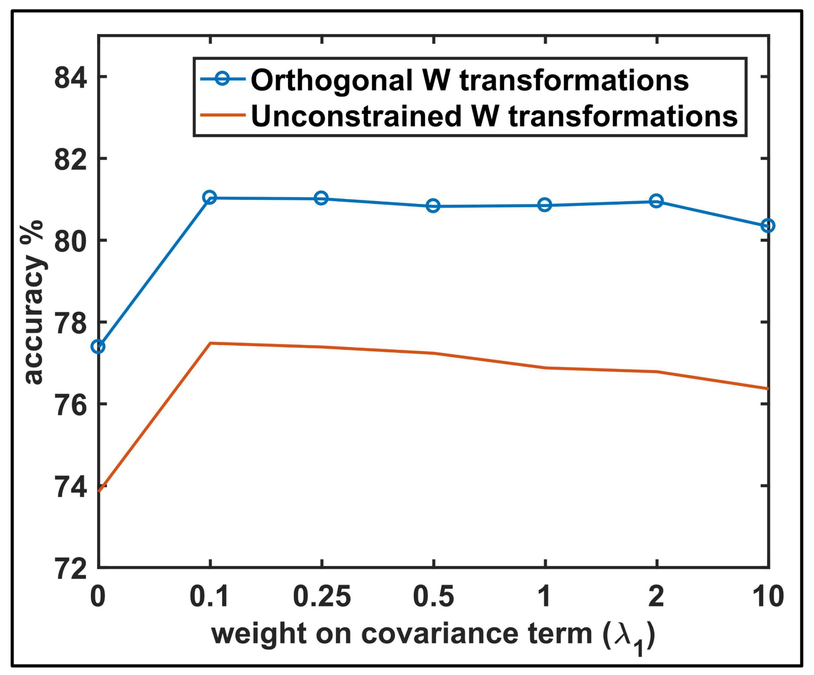


Figure: **(Left)** Comparison of optimization methods. **(Right)** Importance of the orthogonality constraint.

Experimental Results

- **Experiments on Office-Caltech10 Dataset.**
  - Average performances on all 12 transformation sets in Office-Caltech10 dataset.

	SURF	VGG-FC6	VGG-FC7
NN-t	29.0	75.6	76.2
SVM-t	48.9	84.8	83.8
HFA[1]	48.1	83.7	83.0
MMDT[2]	52.5	80.8	78.1
CDLS[3]	53.5	85.9	85.4
ILS(1-NN)	**55.6**	**88.5**	**86.4**

Table: Semi-super. DA on Office-Calt.10

	SURF	VGG-FC6	VGG-FC7
1-NN-s	23.5	65.2	66.7
SVM-s	38.9	70.2	70.6
GFK[4]	42.5	76.1	73.2
SA[5]	44.2	75.3	73.1
CORAL[6]	**46.7**	78.3	76.1
ILS(1-NN)	46.0	**80.8**	**78.8**

Table: Unsuper. DA on Office-Calt.10

- Performance gain when PCA subspaces are replaced

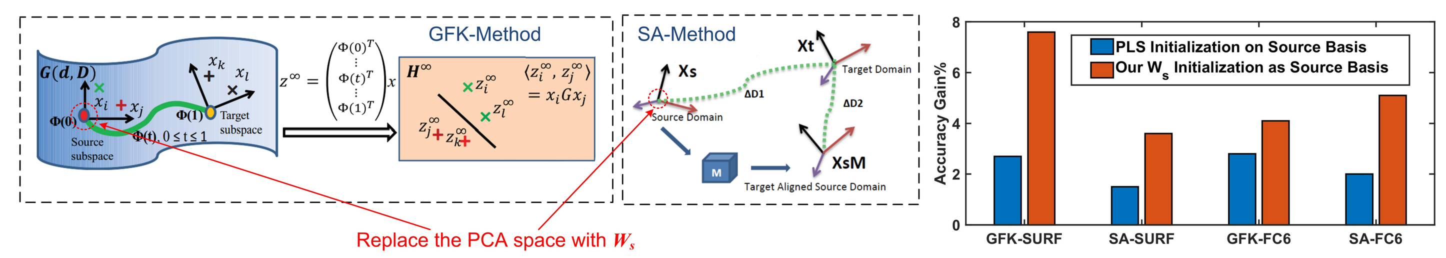
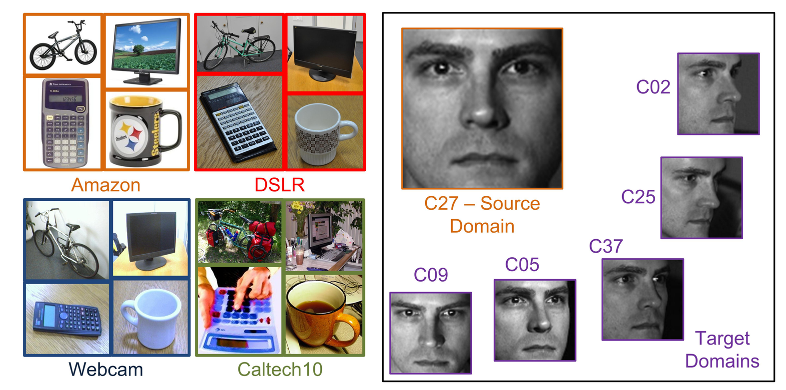


Figure: Performance improvement in GFK[4] and SA[5] methods when their PCA subspaces are replaced with our W_s matrix.

- **Recognizing Faces with Pose Variations.**

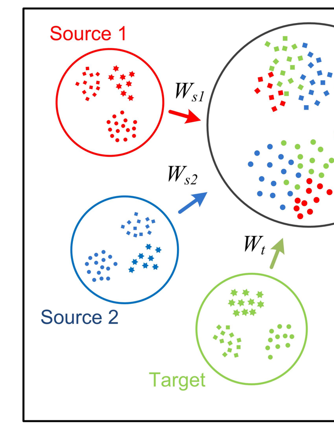


Camera Pose	C09	C05	C37	C25	C02
1-NN-s	92.5	55.7	28.5	14.8	11.0
SVM-s	87.8	65.0	35.8	15.7	16.7
GFK-PLS[4]	92.5	74.0	32.1	14.1	12.3
SA[5]	**97.9**	85.9	47.9	16.6	13.9
CORAL[6]	91.4	74.8	35.3	13.4	13.2
ILS(1-NN)	96.6	**88.3**	**72.9**	**28.4**	**34.8**

Table: **(Left)**Samples of domain images for Office-Caltech10 and PIE faces. **(Right)** Performance on face recognition with pose differences on Multiview-PIE dataset.

Future Extensions

- Our solution could be extended to **multiple source domain adaptation**.
- The latent space is compatible with **Heterogeneous domain adaptation**.
- **Stochastic optimization** on Riemannian manifolds for **large scale experiments**.



Link to our code

Figure: **(Left)** Multiple sources extension. **(Right)** Link to our code.

References.

[1] Duan, Lixin, Dong Xu, and Ivor W. Tsang. ”Learning with Augmented Features for Heterogeneous Domain Adaptation.” Proceedings of the 29th International Conference on Machine Learning (ICML-12). 2012.

[2] Hoffman, J., Rodner, E., Darrell, T., Donahue, J., & Saenko, K. Efficient learning of domain-invariant image representations. In In Proc. ICLR.2013.

[3] Hubert Tsai, Yao-Hung, Yi-Ren Yeh, and Yu-Chiang Frank Wang. ”Learning cross-domain landmarks for heterogeneous domain adaptation.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[4] Gong, Boqing, et al. ”Geodesic flow kernel for unsupervised domain adaptation.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2012.

[5] Fernando, Basura, et al. ”Unsupervised visual domain adaptation using subspace alignment.” Proceedings of the IEEE international conference on computer vision. 2013.

[6] Sun, Baochen, Jiashi Feng, and Kate Saenko. ”Return of Frustratingly Easy Domain Adaptation.” AAAI. Vol. 6. No. 7. 2016.