



迁移学习和领域自适应方法及应用

Transfer learning/Domain adaptation: Methods and Application

Lei Zhang (张 磊)

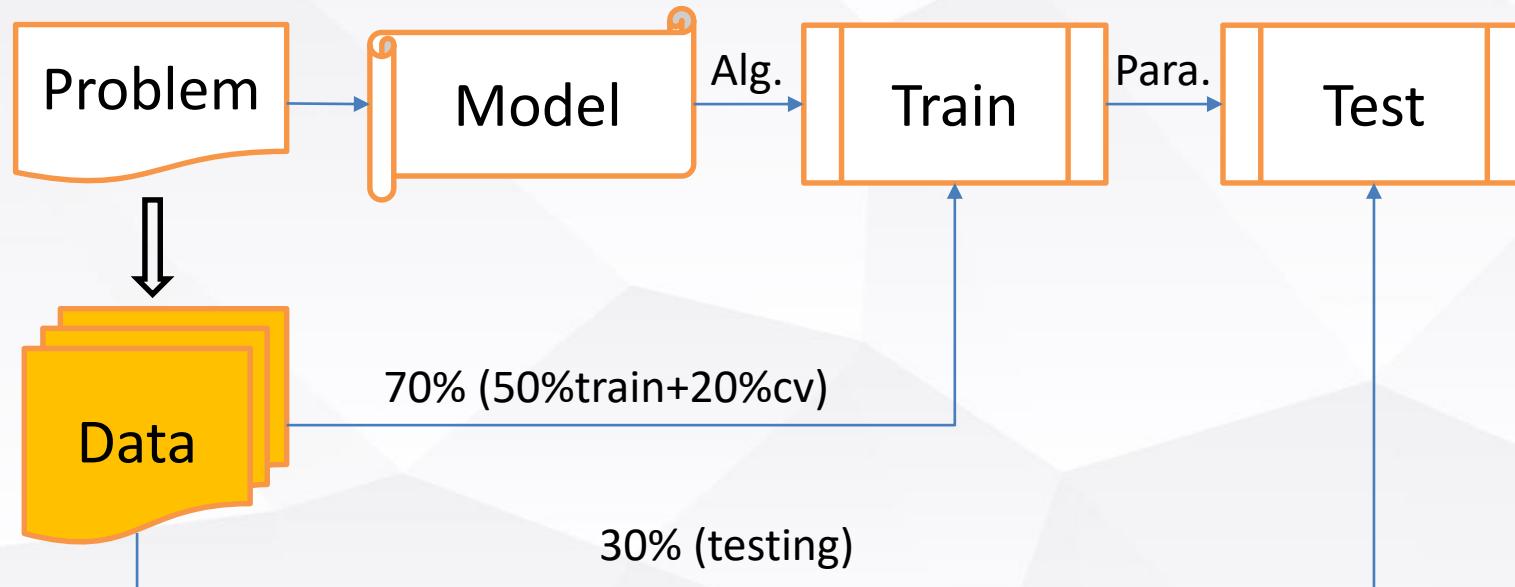
重庆市“生物感知与智能信息处理”重点实验室

重庆大学 微电子与通信工程学院

Website: <http://www.leizhang.tk/>



General route of machine learning app.

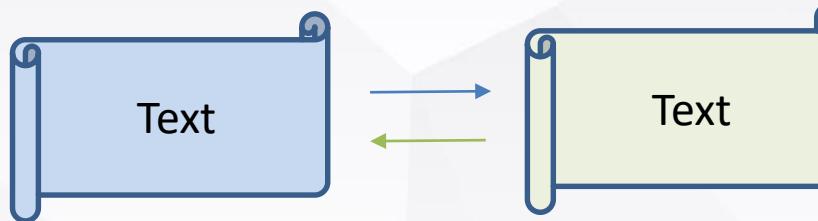


Condition: Independent Identical distribution (i.i.d.)!!

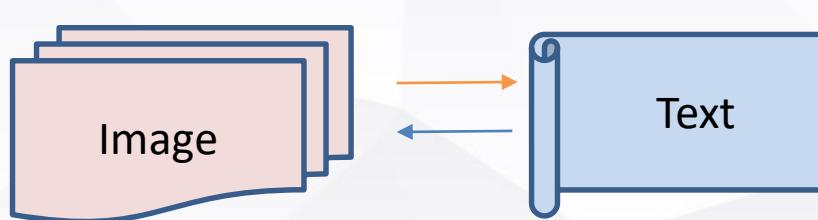


Scenarios of Non. i.i.d.

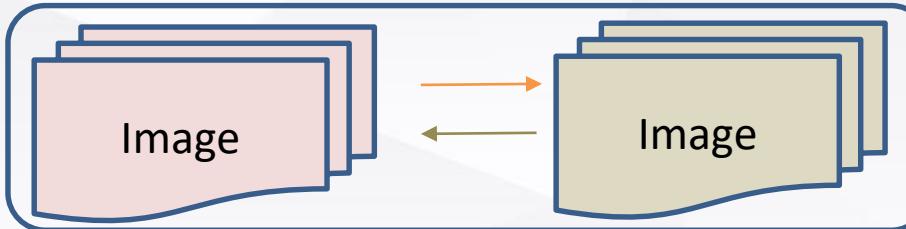
Data of Heterogeneity
(language, blur, etc.)



Data of Heterogeneity
(Media)



Data of Heterogeneity
(background, viewpoint, pose
, modality, etc.)





Weak Learning (弱学习)

The concept of “weak learning” originates from the era of Boosting and AdaBoost (30 years ago).

Amazingly, **the past** “weak learning” is equivalent to “strong learning”. There is a sentence:

“A problem can be weak-learned if and only if it can be strong-learned.”

Currently, the weak learning is really a weak problem instead of strong problem.

1. Weakly-supervised learning (周志华)-不完全、不确切、不准确 (label)
2. Transfer learning (杨强)-边缘分布、类条件分布 (data vs. label)
3. Domain adaptation (Kate Saenko)-知识共享 (data)

前百度首席科学家、斯坦福大学教授，吴恩达 (Andrew Ng) “迁移学习将会是继监督学习之后的下一个机器学习商业成功的驱动力”，NIPS’16.



Problem Proposal: Transfer learning is everywhere!



+



Transfer data

Source data

2018/9/29

Modeling

Classifier?
Feature?



Target data

?

Dog
Cat
...

Conventional Transfer learning Problem!



Problem Proposal: Transfer learning is everywhere!



Big Source data

+



Transfer data

Modeling

Classifier?
Feature?



Target data

?

Dog
Cat
...

Big Data Conditioned Transfer learning Problem!



Transfer learning

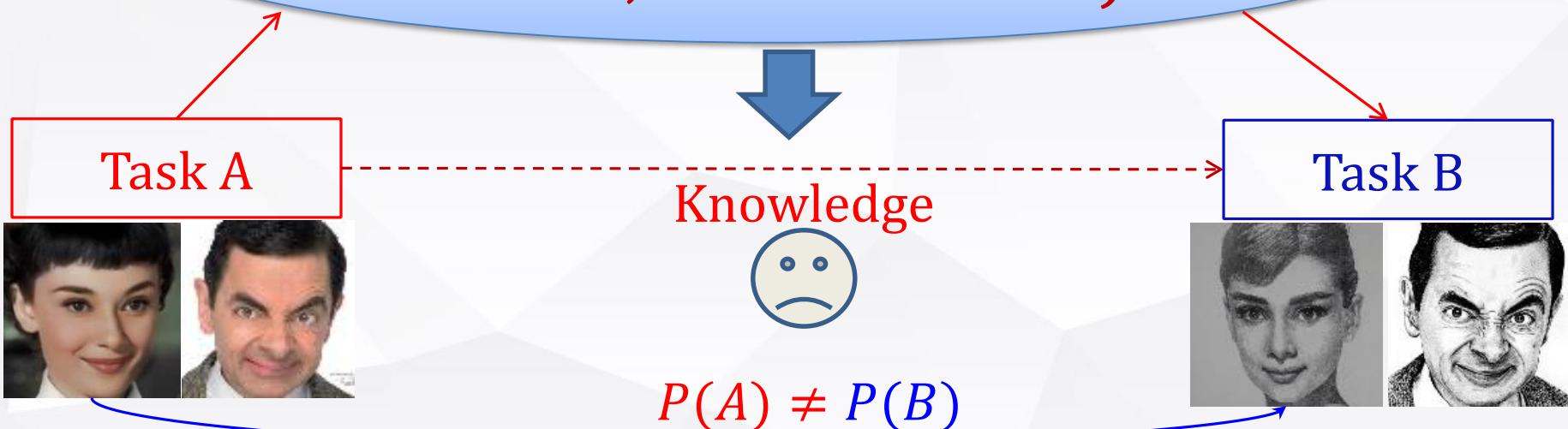
- ❑ What is transfer learning?
- ❑ Why transfer?
- ❑ How to transfer

human like learning



What is transfer learning?

Model parameters (classifier, neural network, transformation etc.)

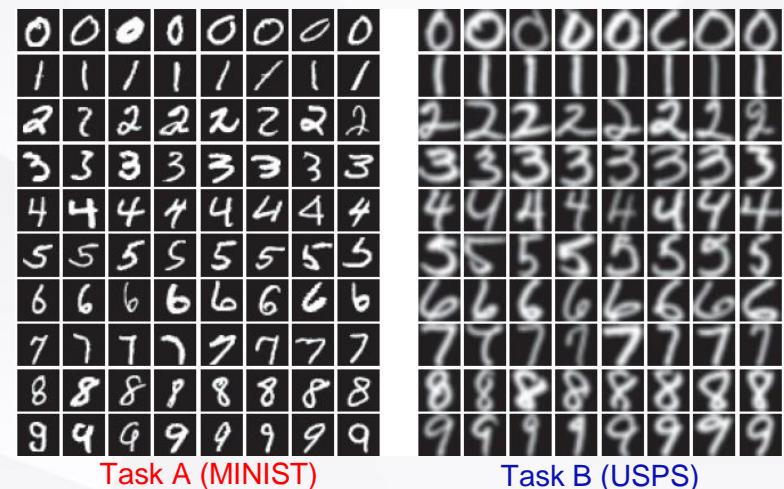
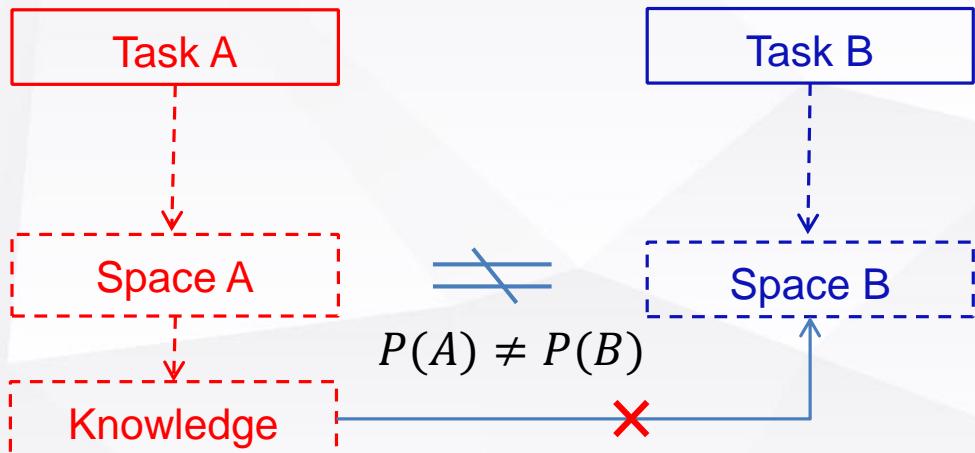


Related but different domain



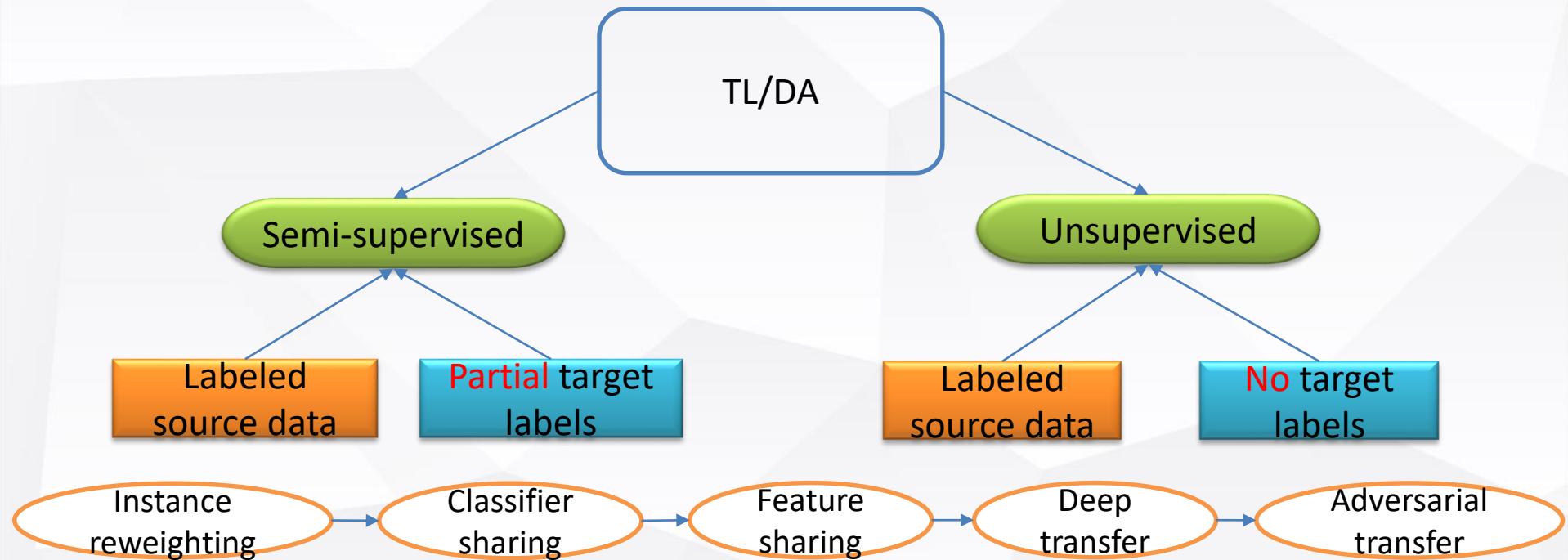
Why transfer learning (domain adaptation)?

- The data (feature) probability distribution generated from **Task A** and **Task B** is different, such that the learning parameters in raw data space are not “generalized” (e.g. computer vision).
- The implied basic assumption of machine learning is that the training and testing data should hold similar distribution, *i.e.*, independent identical distribution (*i.i.d*), which is **violated**.



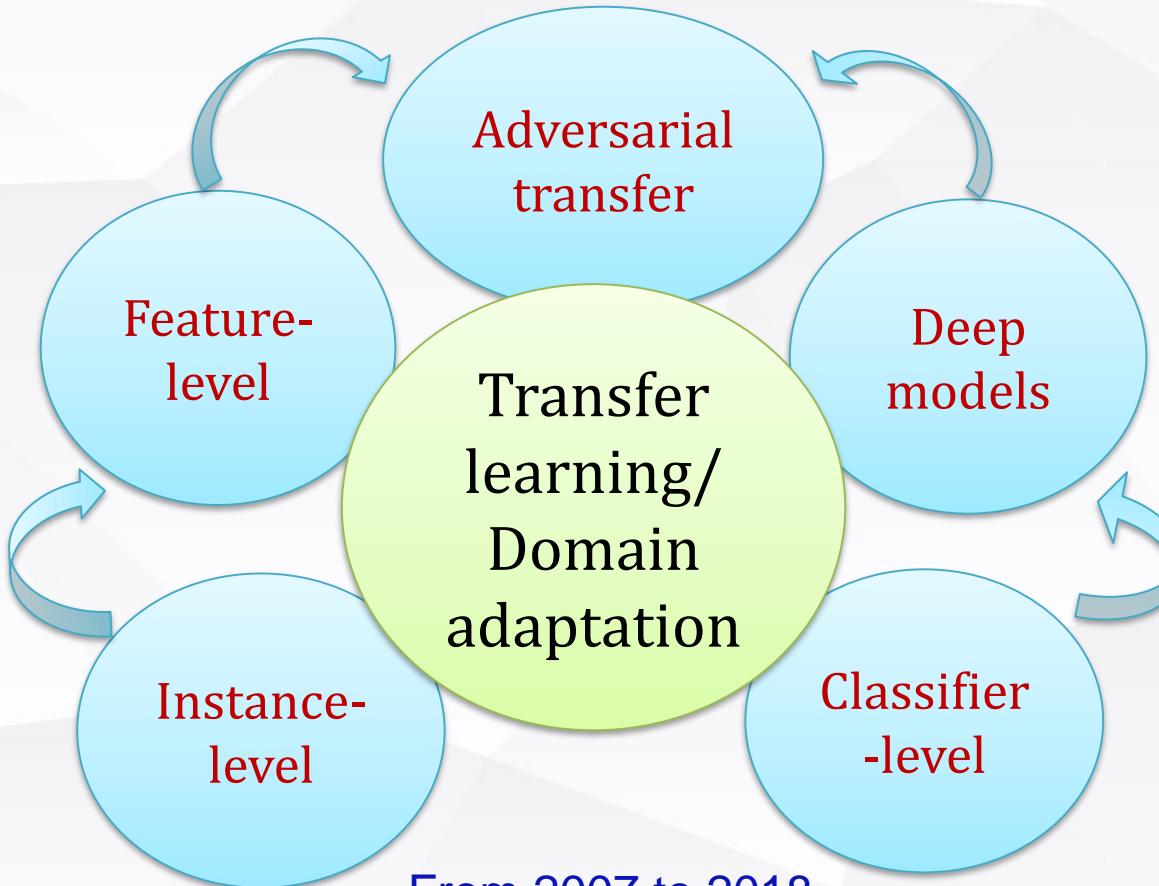


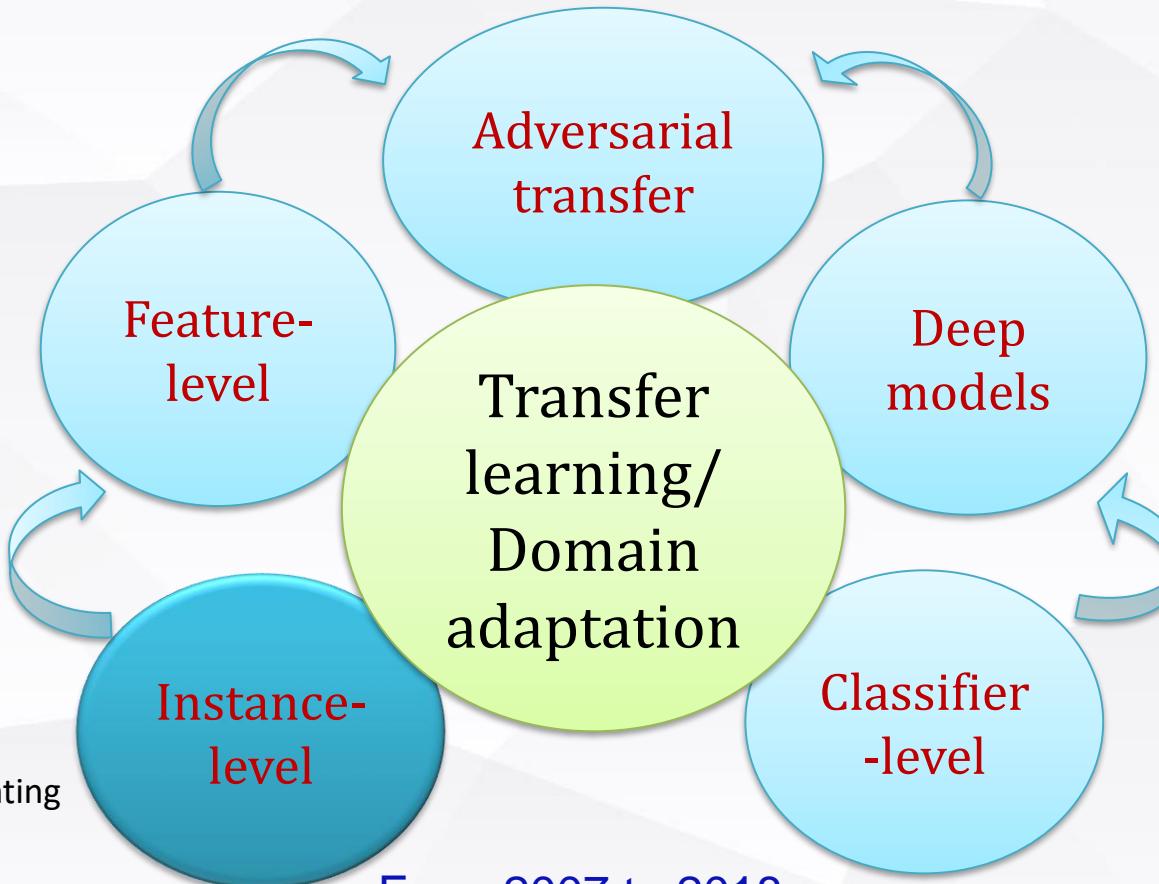
How transfer learning? Categorization: TL/DA



2007

2018







How to transfer learning (domain adaptation)?

- **[Instance Level]** Learn instance weights, such that **Task A** and **Task B** have less data disparity (Jiang and Zhai, ACL 2007; Huang et al. NIPS 2007).

Generally, a learning model minimizes expected risk:

$$R[\Pr, \theta, l(x, y, \theta)] = \mathbf{E}_{(x,y) \sim \Pr} [l(x, y, \theta)]$$

But the training data only comes from a subset, so the average empirical risk is minimized:

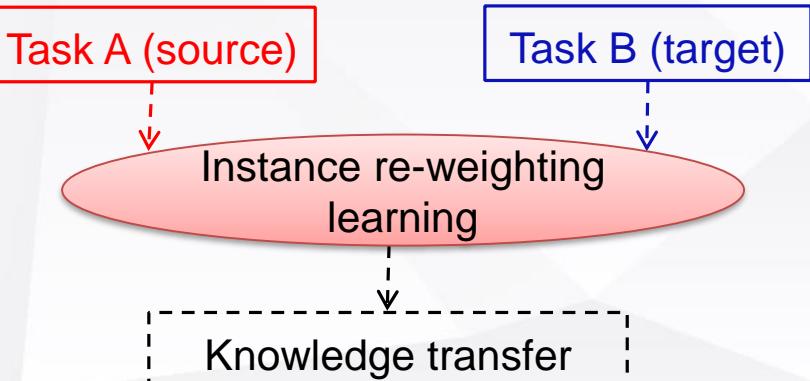
$$R_{\text{emp}}[Z, \theta, l(x, y, \theta)] = \frac{1}{m} \sum_{i=1}^m l(x_i, y_i, \theta)$$

$$R_{\text{reg}}[Z, \theta, l(x, y, \theta)] := R_{\text{emp}}[Z, \theta, l(x, y, \theta)] + \lambda \Omega[\theta]$$

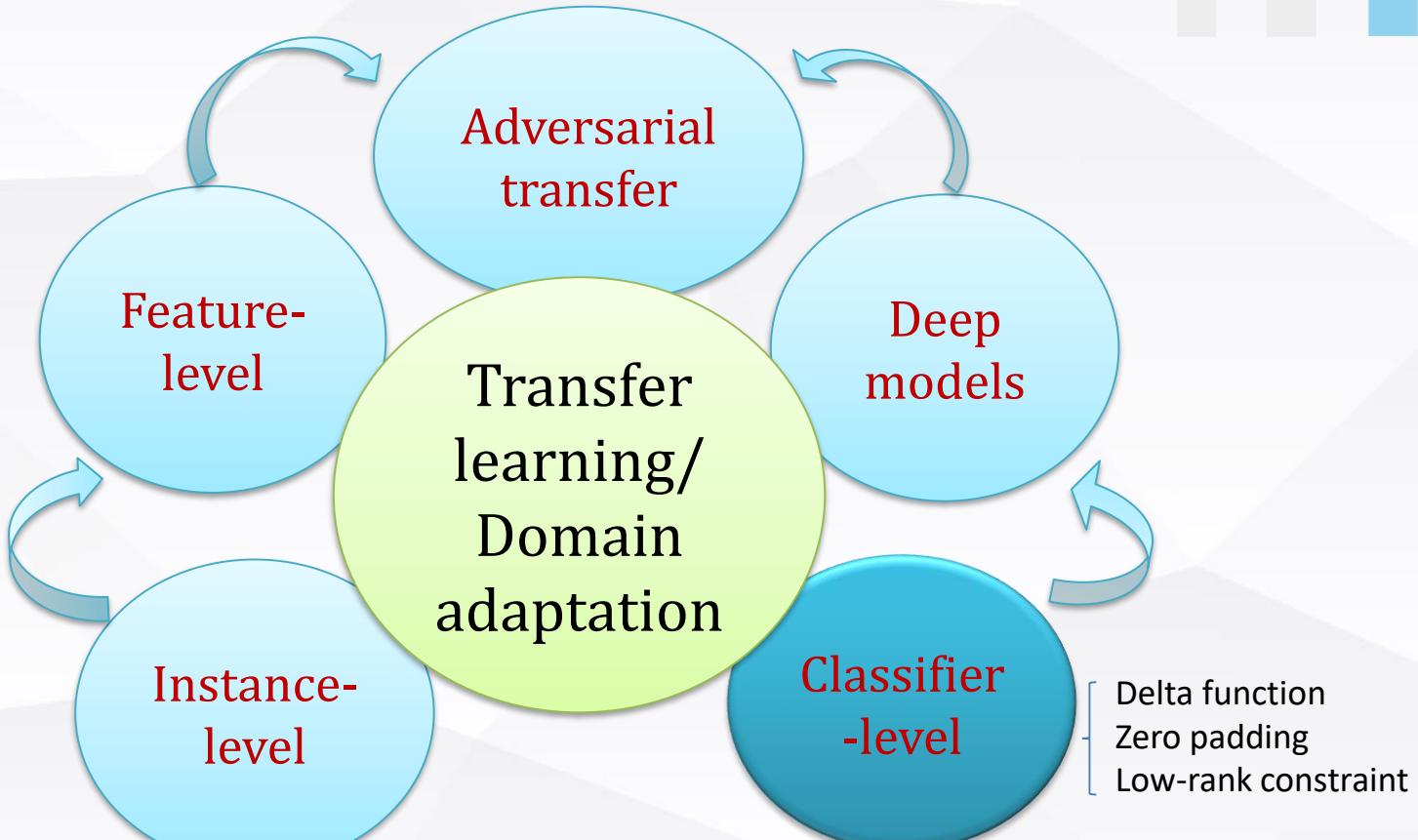
Actually, we focus on the performance of testing data

$$R[\Pr', \theta, l(x, y, \theta)] = \mathbf{E}_{(x,y) \sim \Pr'} [l(x, y, \theta)] = \mathbf{E}_{(x,y) \sim \Pr} \left[\underbrace{\frac{\Pr'(x,y)}{\Pr(x,y)}}_{:=\beta(x,y)} l(x, y, \theta) \right]$$

$$R_{\text{reg}}[Z, \beta, l(x, y, \theta)] := \frac{1}{m} \sum_{i=1}^m \beta_i l(x_i, y_i, \theta) + \lambda \Omega[\theta]$$



Instance Level





How to transfer learning (domain adaptation)?

- ▣ [Classifier Level] Learn a common classifier on **Task A**, by leveraging a few labeled/unlabeled target samples from **Task B**. (Yang et al. ACM MM'07; Duan et al. CVPR'12, TPAMI'12; Wang et al. ACM MM'18)

Assumption: There exists a *delta function* between the auxiliary classifier (source) f_a and the new classifier (target) f .

$$f(\mathbf{x}) = f^a(\mathbf{x}) + \Delta f(\mathbf{x}) = f^a(\mathbf{x}) + \mathbf{w}^T \phi(\mathbf{x})$$

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad \text{Standard SVM}$$

$$\text{s.t. } \xi_i \geq 0, \quad y_i \mathbf{w}^T \phi(\mathbf{x}_i) \geq 1 - \xi_i, \quad \forall (\mathbf{x}_i, y_i) \in \mathcal{D}_l^p$$

$$\min_{\mathbf{w}} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^N \xi_i \quad \text{ASVM}$$

$$\text{s.t. } \xi_i \geq 0$$

$$y_i f^a(\mathbf{x}_i) + y_i \mathbf{w}^T \phi(\mathbf{x}_i) \geq 1 - \xi_i.$$



$$\text{DIST}_k(\mathcal{D}^A, \mathcal{D}^T) = \left\| \frac{1}{n_A} \sum_{i=1}^{n_A} \varphi(\mathbf{x}_i^A) - \frac{1}{n_T} \sum_{i=1}^{n_T} \varphi(\mathbf{x}_i^T) \right\|_{\mathcal{H}}$$

$$f^T(\mathbf{x}) = \sum_{p=1}^P \beta_p f_p(\mathbf{x}) + \underbrace{\sum_{m=1}^M d_m \mathbf{w}'_m \varphi_m(\mathbf{x}) + b}_{\Delta f(\mathbf{x})}$$

$$\min_{\mathbf{d} \in \mathcal{M}} G(\mathbf{d}) = \frac{1}{2} \Omega^2(\mathbf{d}) + \theta J(\mathbf{d}),$$

where

$$J(\mathbf{d}) = \min_{\mathbf{w}_m, \beta, b, \xi_i} \frac{1}{2} \left(\sum_{m=1}^M d_m \|\mathbf{w}_m\|^2 + \lambda \|\beta\|^2 \right) + C \sum_{i=1}^n \xi_i,$$

s.t. $y_i f^T(\mathbf{x}_i) \geq 1 - \xi_i, \xi_i \geq 0,$



How to transfer learning (domain adaptation)?

- [Classifier Level] Learn a common classifier on **Task A**, by leveraging a few *labeled/unlabeled* target samples from **Task B**.

Representative work (zero padding feature augmentation, low-rank solution and delta function):

- ① Daumé III, et al. ACL'07(Frustrating Easy Adaptation, EA)
- ② Li, et al. TPAMI'14 (HFA)
- ③ Li, et al. TPAMI'18 (LRE-SVMs)
- ④ Zhang, et al. IEEE Sens.'17 (MFKS)

$$\Phi^s(x) = \langle \boxed{x}, x, 0 \rangle, \quad \Phi^t(x) = \langle \boxed{x}, 0, x \rangle$$

$$\begin{aligned}\Phi^s(x) &= \langle \Phi(x), \Phi(x), \mathbf{0} \rangle \\ \Phi^t(x) &= \langle \Phi(x), \mathbf{0}, \Phi(x) \rangle\end{aligned}$$

kernelize

$$R_{reg}[W, l(X_S, X_T, W)] = \sum R_{emp}[w_i, l(X_S, X_T, w_i)] + \|[w_1, w_2, \dots, w_D]\|_*$$

- ⑤ Joachmis, ICML'1999 (T-SVM)
- ⑥ Yang, et al. ACM MM'07 (ASVM)
- ⑦ Duan, et al. TPAMI'12 (AMKL)
- ⑧ Duan, et al. TPAMI'13 (DTSVM, DTMKL)

Examples in Re-ID (WeiShi Zheng and Jianhuang Lai):

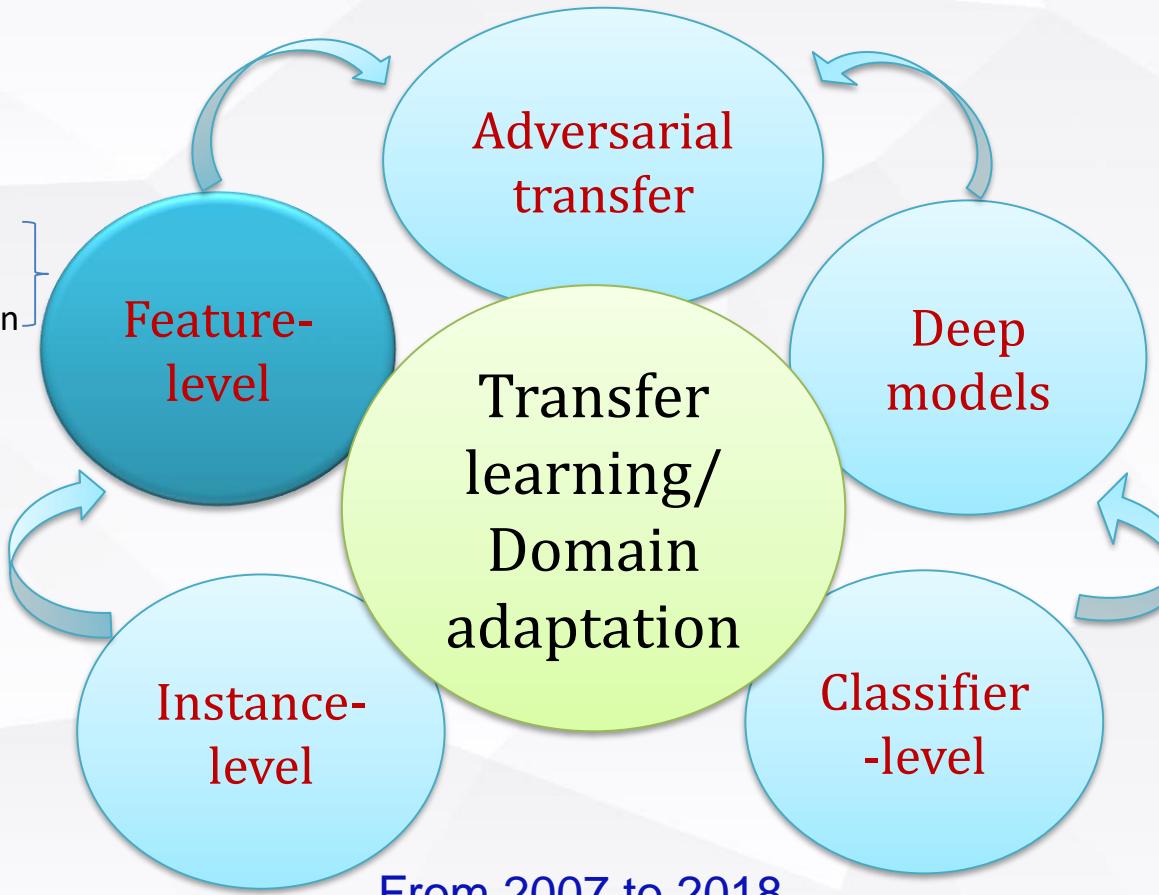
View-specific transform for Re-ID (IJCAI'15, TPAMI'18)

Deep zero padding

$$f(\mathbf{x}) = f^a(\mathbf{x}) + \Delta f(\mathbf{x}) = f^a(\mathbf{x}) + \mathbf{w}^T \phi(\mathbf{x})$$



Subspace unification
Manifold alignment
Subspace reconstruction





How to transfer learning (domain adaptation)?

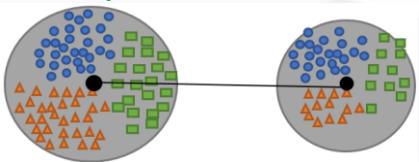
- [Feature Level] Learn a common subspace on both **Task A** and **Task B** with *domain discrepancy minimization*. (Pan et al. TKDE'10; TNNLS'11; Hoffman et al. IJCV'14; Kan et al. IJCV'14)

General paradigm:

$$R_{reg}[W, l(X_S, X_T, W)] = R_{emp}[W, l(X_S, X_T, W)] + \Omega[W]$$

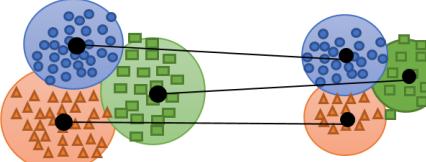
Marginal distribution consistency

$$P(\phi(X_S)) \approx P(\phi(X_T))$$



Conditional distribution consistency

$$P(\phi(X_S^i)|y_S^i) \approx P(\phi(X_T^i)|y_T^i), i = 1, \dots, C$$



| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |

Task A (source)

Common Subspace

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |

Task B (target)

“Borrow” data

Classifier Learning

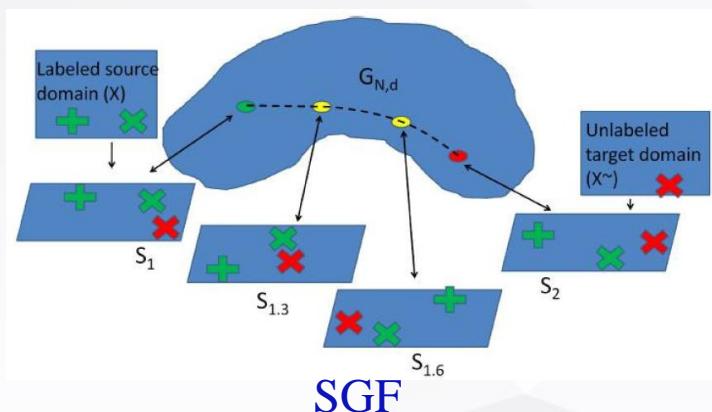
Label Prediction



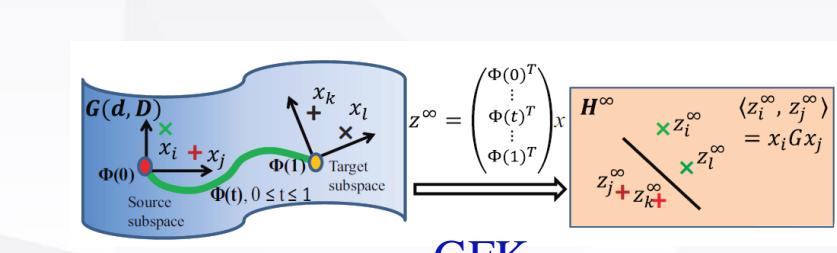
How to transfer learning (domain adaptation)?

□ [Feature Level] Learn an aligned subspace on both **Task A** and **Task B** with *alignment*.

(Gopalan et al. ICCV'11, SGF; Gong, et al. CVPR'12, GFK; Fernando, et al. ICCV'13, SA)

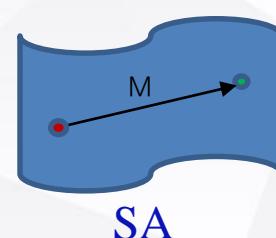


find some intermediate representation along the geodesic path



GFK

construct kernels along the geodesic path



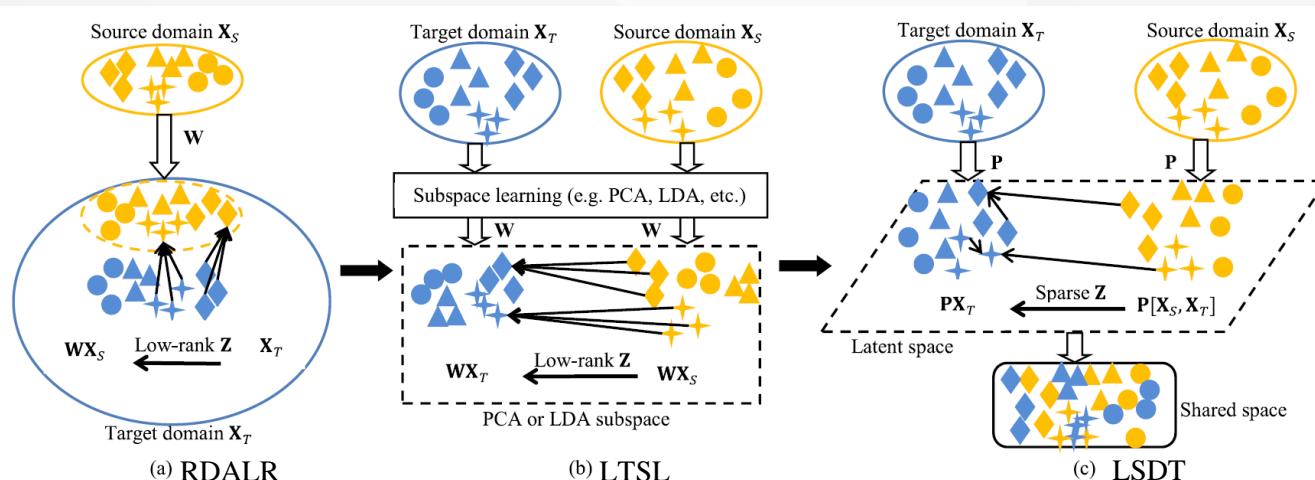
$$F(M) = \|X_S M - X_T\|_F^2$$

$$M^* = \operatorname{argmin}_M F(M)$$



How to transfer learning (domain adaptation)?

- [Feature Level] Learn a common subspace on both **Task A** and **Task B** with *domain reconstruction and representation*. (Jhuo, et al. CVPR'12, RDALR; Shao, et al. IJCV'14, LTSI; Zhang et al. TIP'16, LSDT; Xu et al. TIP'16, DTSL)



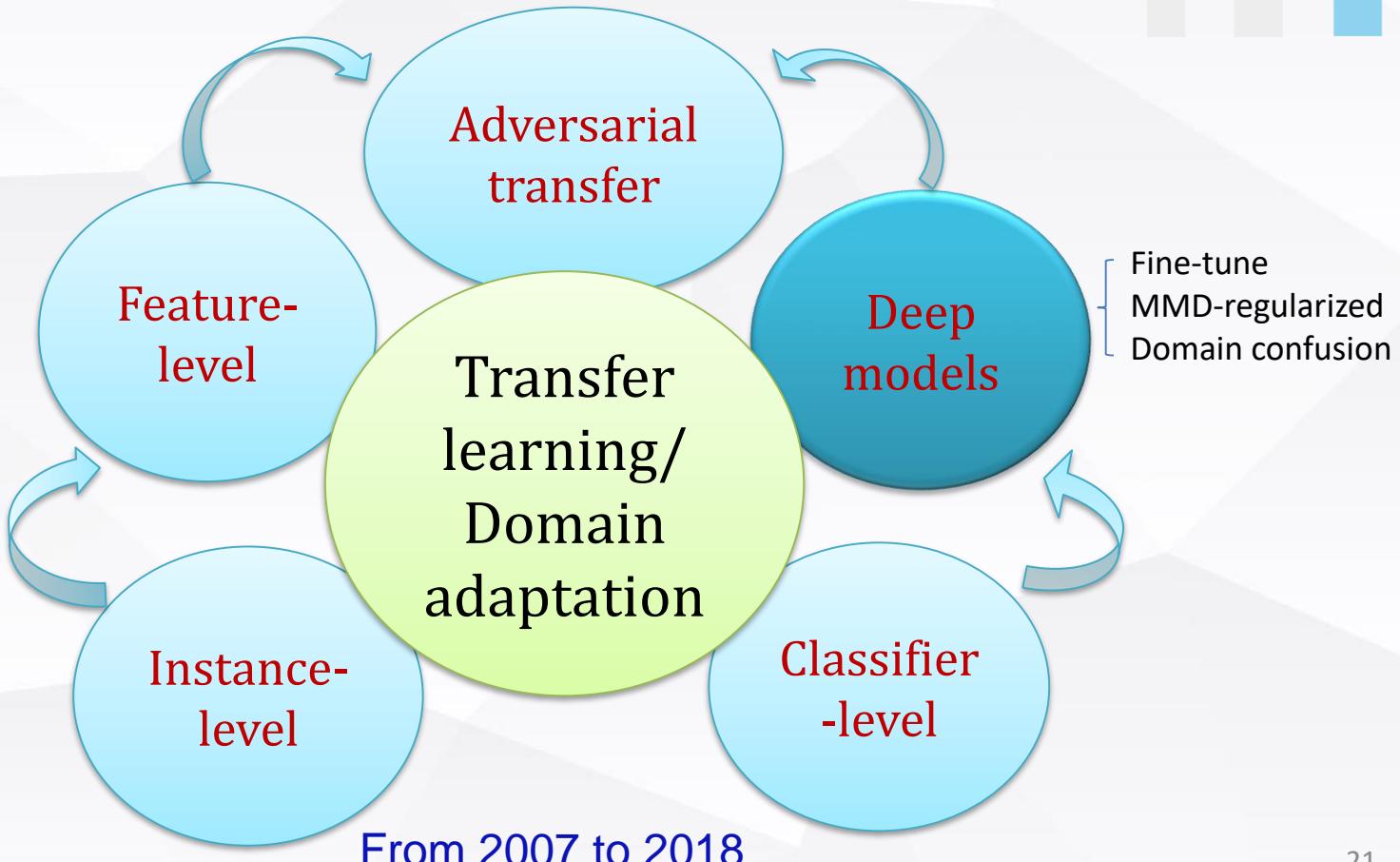
$$\begin{aligned} \min_{W, Z, E} & F(W) + \Re(Z) + \Omega(E) \\ \text{s.t. } & f(X_T) = f(X_S)Z + E \end{aligned}$$

$F(\cdot)$ is subspace learning fun.
 $f(\cdot)$ is transformation fun.

For better basis:
 Domain adaptive dictionary
 (Rama Chellappa. CVPR'13, SDDL)

LRR: strength(better locality of data, block-wise structure, neighbor to neighbor reconstruction)

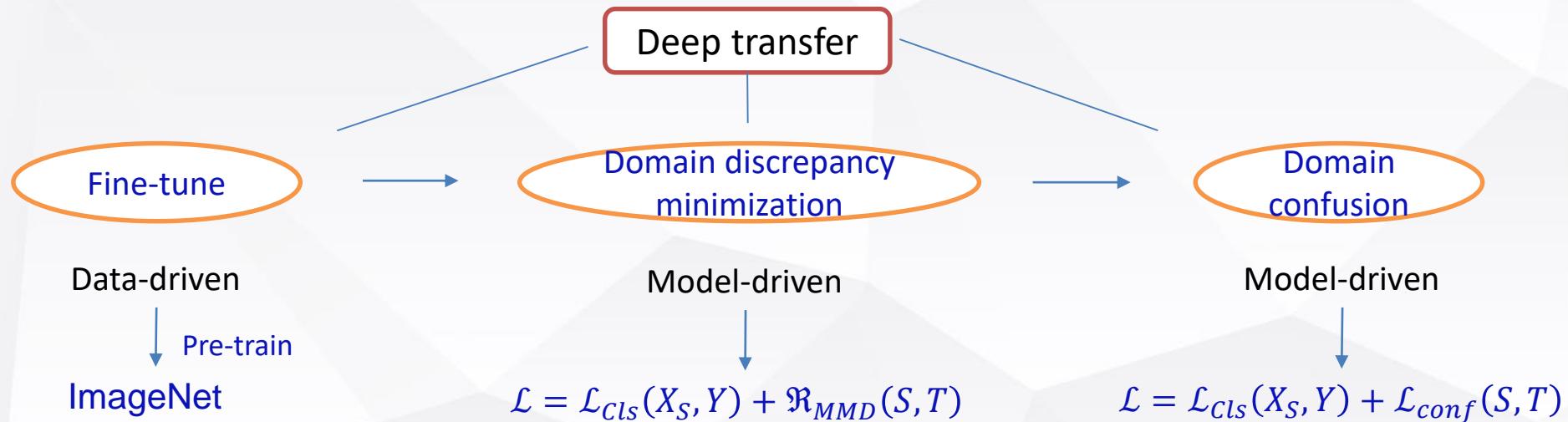
weakness(strong assumption of independent subspaces and sufficient data, easy to get trivial solution)





How to transfer learning (domain adaptation)?

- [Deep models] Learn general feature representation with CNN models



Objective: Small-sample learning problem in big data (大数据中的小样本学习问题)



Deep learning belongs to Transfer learning

- **[Deep models]** Learn general feature representation with *fine-tuning* (AlexNet, NIPS'12)

Example: General deep learning (self-contained multi-source data)

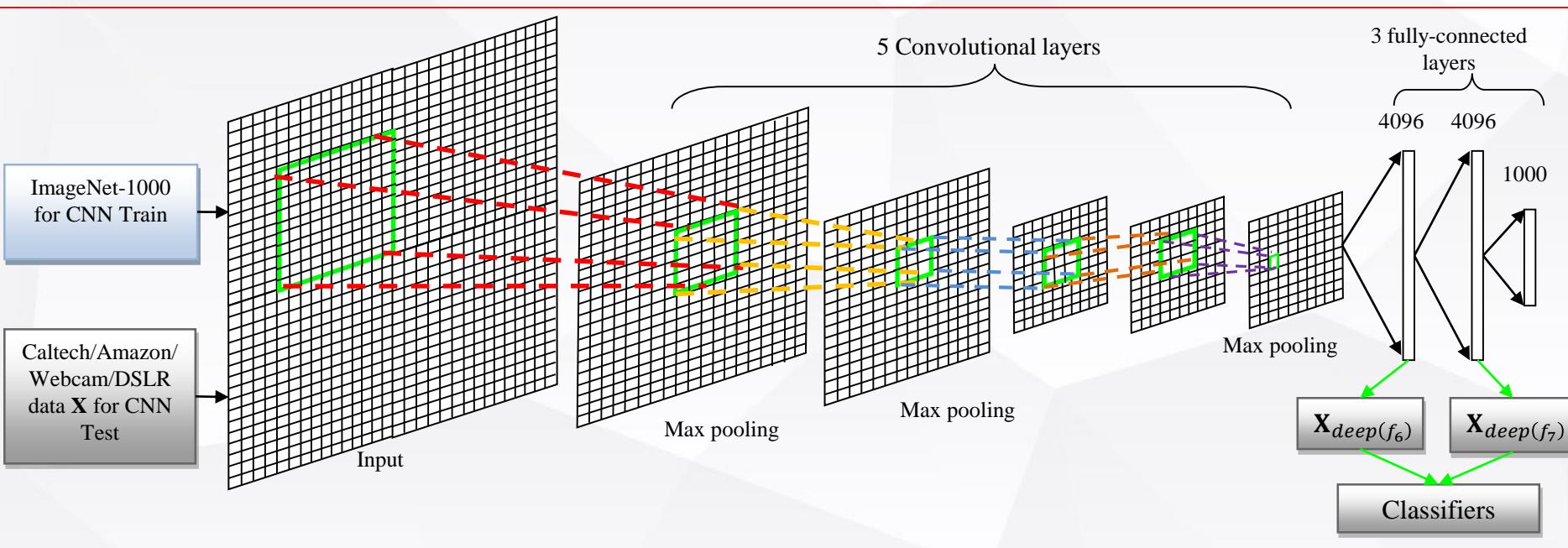


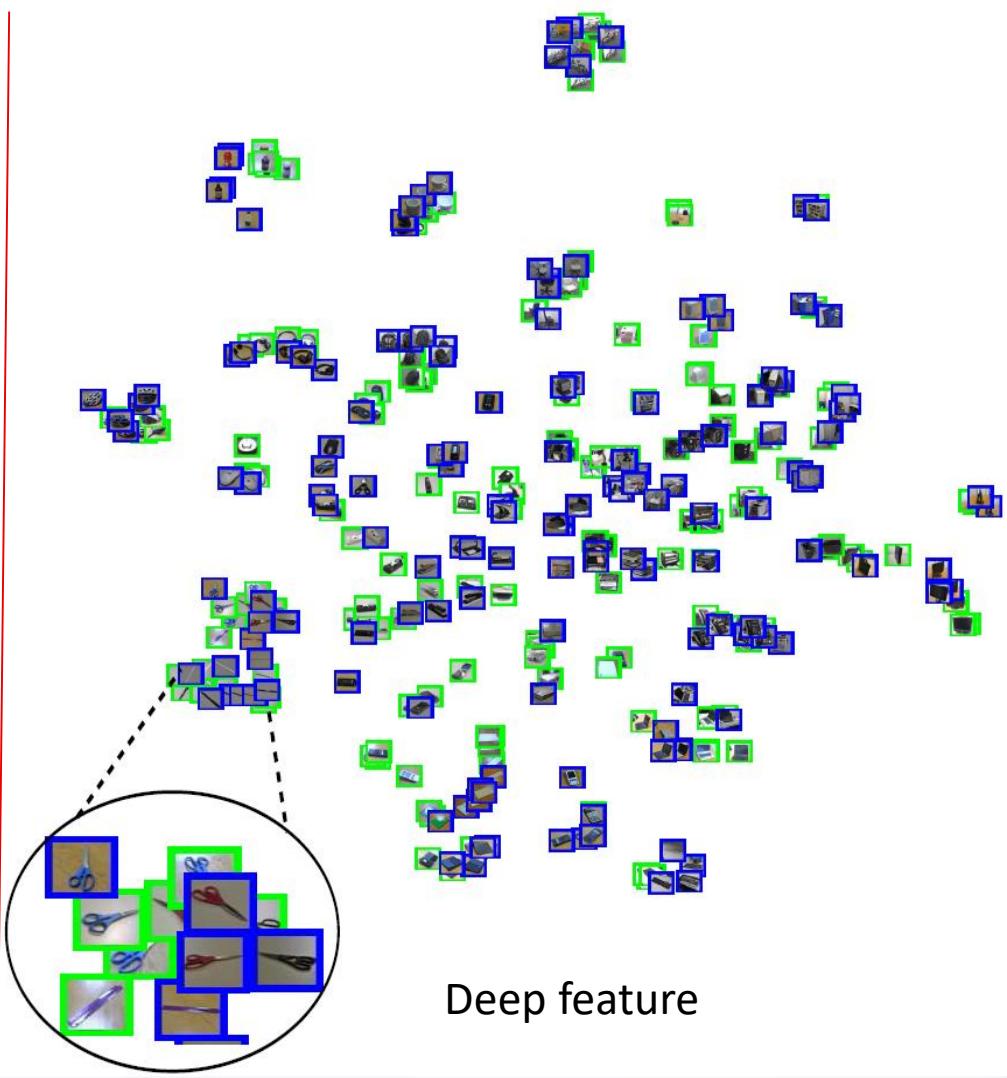
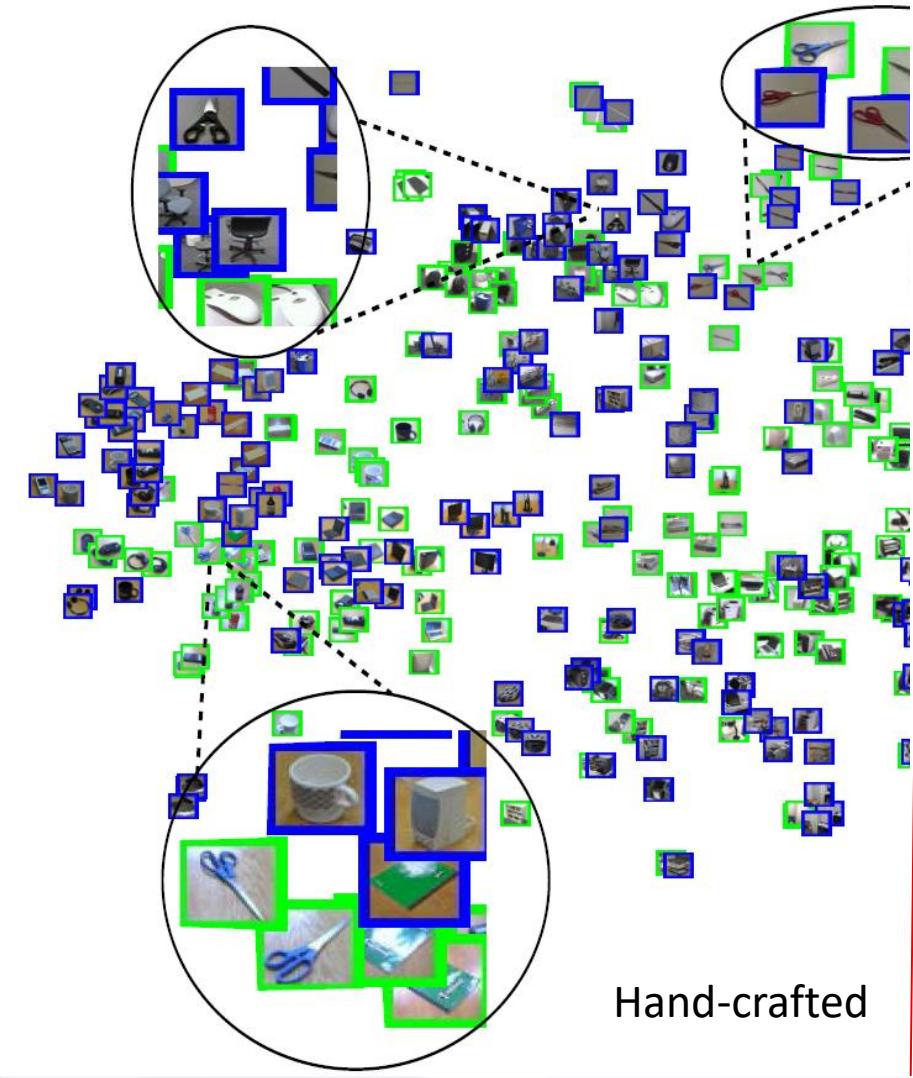
ImageNet: Large-scale Visual Recognition Challenge (ILSVRC)



Deep learning belongs to Transfer learning

- [Deep models] Learn general feature representation with *fine-tuning* (AlexNet, NIPS'12)







Deep learning vs. Transfer learning

- Deep transfer learning (learning general feature representation)



transfer

New fields with limited
training data (i.e. medical,
satellite, agriculture,
smart grid)

ImageNet: Large-scale Visual Recognition Challenge



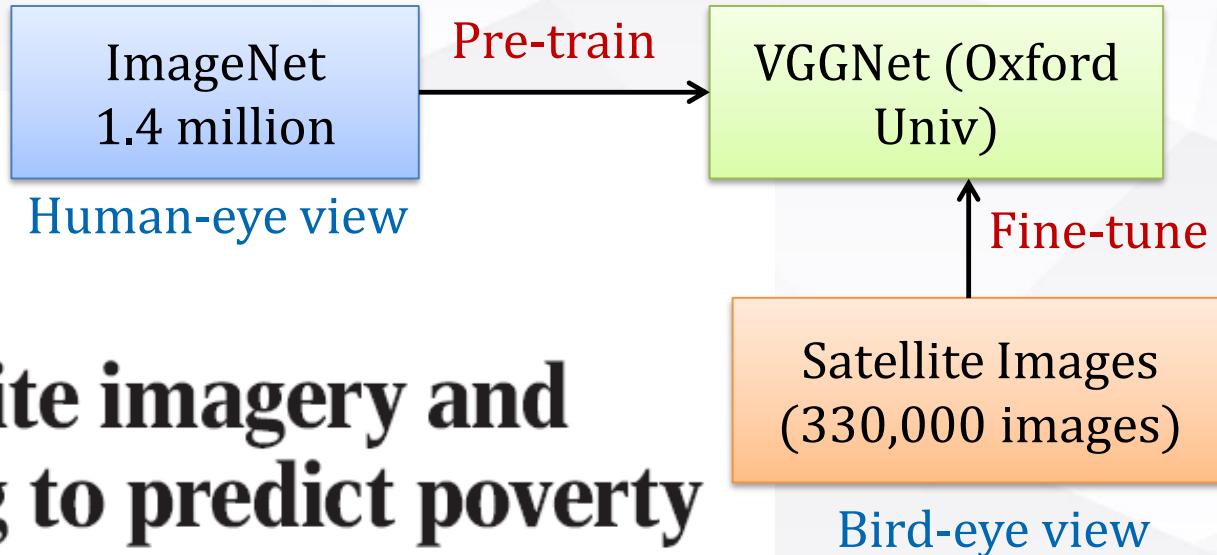
Deep learning vs. Transfer learning

RESEARCH ARTICLES

ECONOMICS

Combining satellite imagery and machine learning to predict poverty

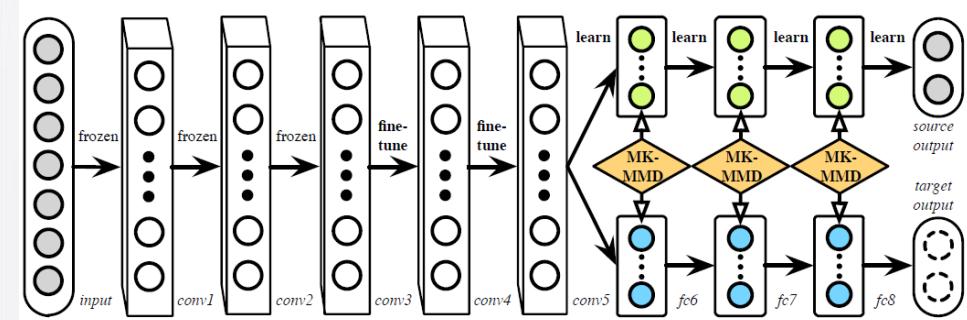
Neal Jean,^{1,2*} Marshall Burke,^{3,4,5*}† Michael Xie,¹ W. Matthew Davis,⁴
David B. Lobell,^{3,4} Stefano Ermon¹





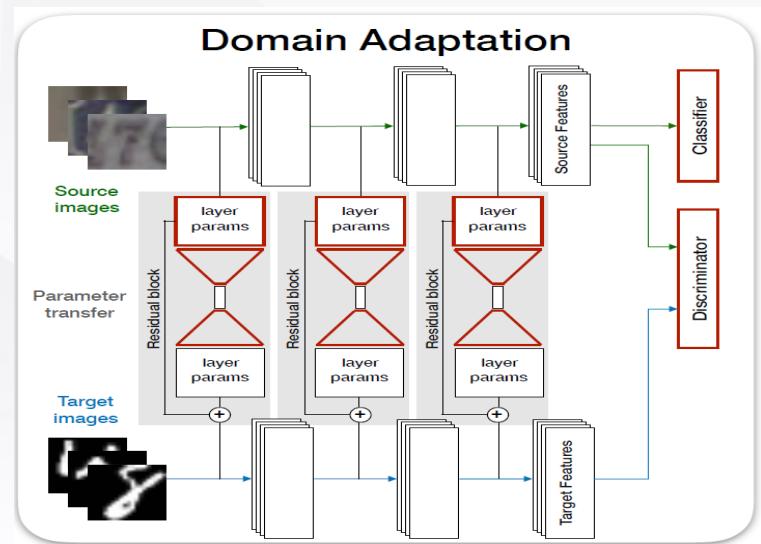
How to transfer learning (domain adaptation)?

- [Deep models] Learn general feature representation with *domain discrepancy minimization* in supervised manner (Tzeng, arXiv'14; Long et al. ICML'15, NIPS'16; Yan, et al. CVPR'17; Rozantsev et al. CVPR'18)



One-stream (shared, Long ICML'15)

$$\mathcal{L}_{\text{fixed}} = \mathcal{L}_{\text{class}} + \mathcal{L}_{\text{disc}}$$

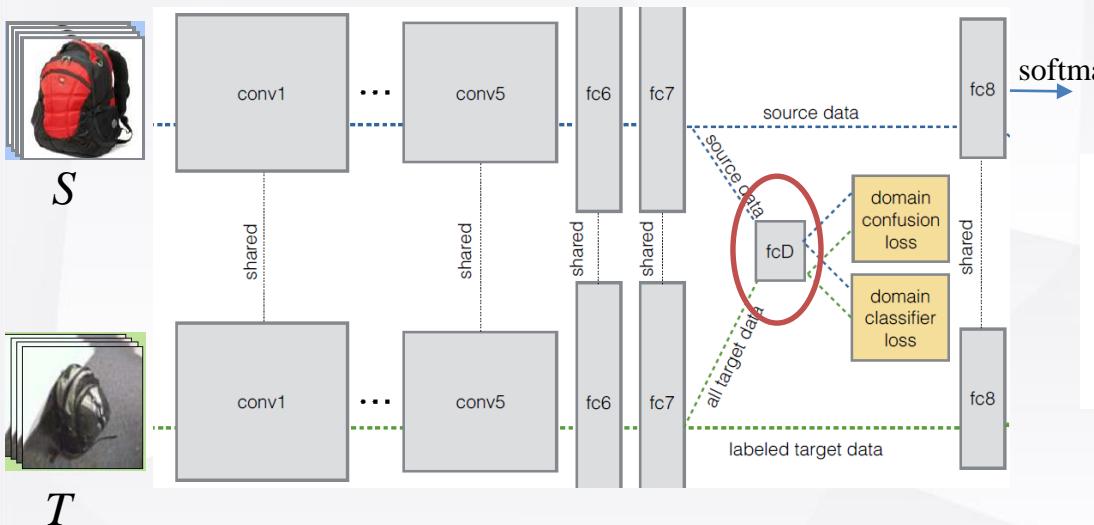


Two-stream (not shared, CVPR'18)



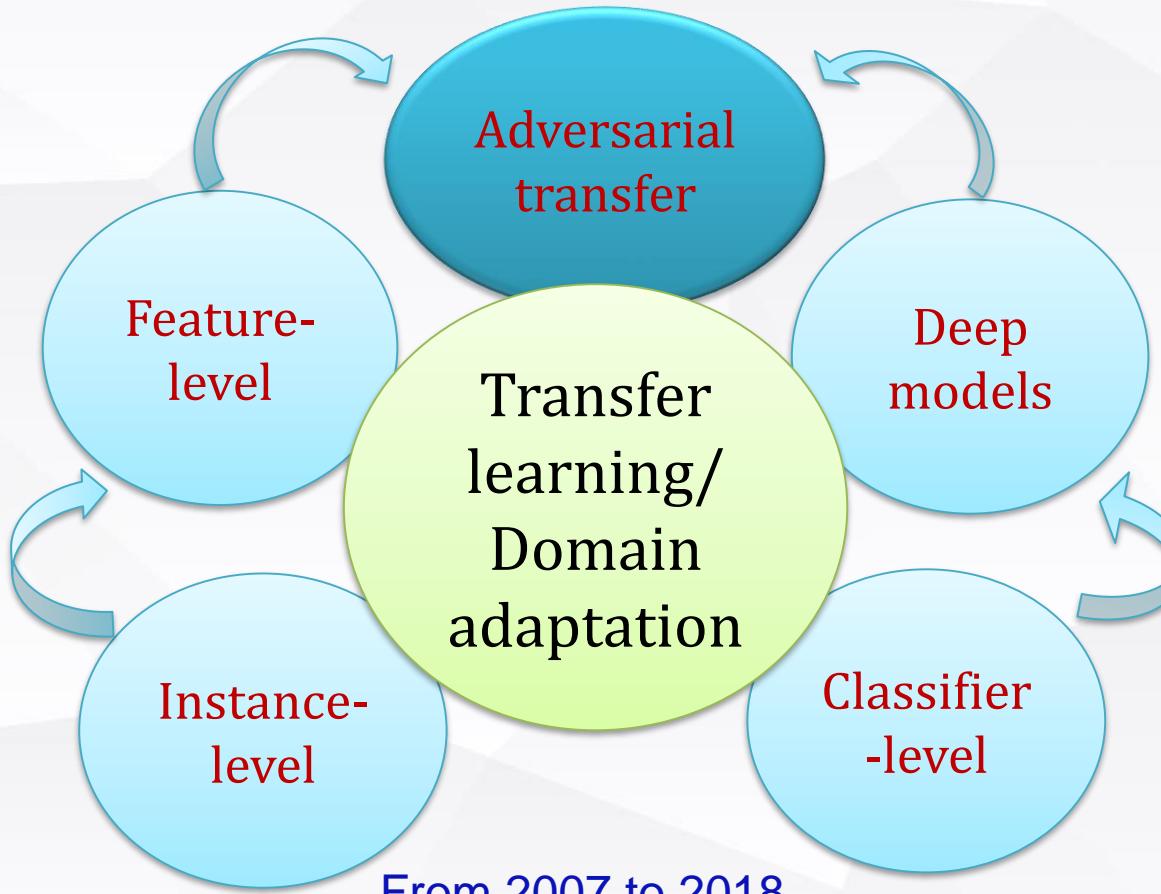
How to transfer learning (domain adaptation)?

- [Deep models] Learn general feature representation with *domain confusion maximization* in supervised manner (Ajakan et al. NIPS'14, DANN; Tzeng et al. ICCV'15, DDC; Murez et al. CVPR'18)



$$\begin{aligned}\mathcal{L}(x_S, y_S, x_T, y_T, \theta_D; \theta_{\text{repr}}, \theta_C) = \\ \mathcal{L}_C(x_S, y_S, x_T, y_T; \theta_{\text{repr}}, \theta_C) \\ + \lambda \mathcal{L}_{\text{conf}}(x_S, x_T, \theta_D; \theta_{\text{repr}}) \\ + \nu \mathcal{L}_{\text{soft}}(x_T, y_T; \theta_{\text{repr}}, \theta_C).\end{aligned}$$

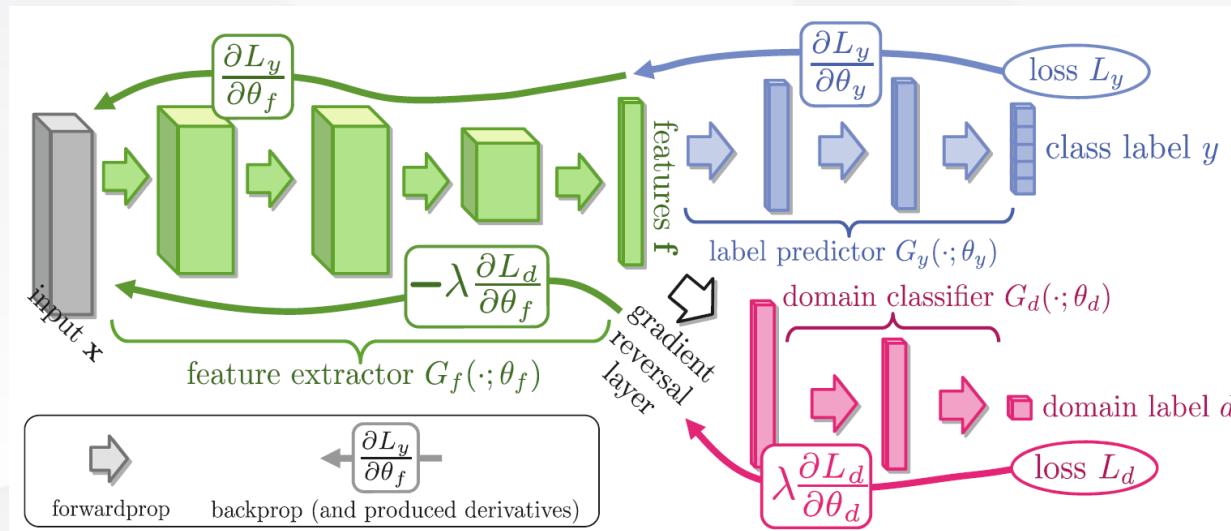
Goal: learning domain-invariant representation





How to transfer learning (domain adaptation)?

- **[Adversarial transfer]** Learn feature generation model with *domain confusion* (Ganin et al. JMLR'16; Tzeng et al. CVPR'17, ADDA; Chen et al. CVPR'18, RAAN; Saito et al. CVPR'18, MCD; Pinheiro, CVPR'18)



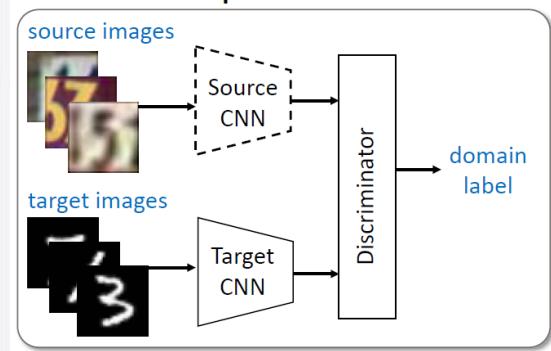
Ganin et al. JMLR'16, Gradient Reversal (GradRev)



How to transfer learning (domain adaptation)?

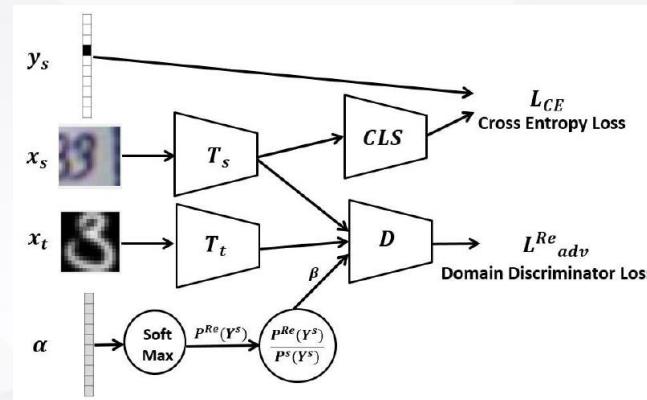
- **[Adversarial transfer]** Learn feature generation model with *domain confusion* (Ganin et al. JMLR'16; Tzeng et al. CVPR'17, ADDA; Chen et al. CVPR'18, RAAN; Saito et al. CVPR'18, MCD; Pinheiro, CVPR'18 , Cao et al., ECCV'18)

Adversarial Adaptation



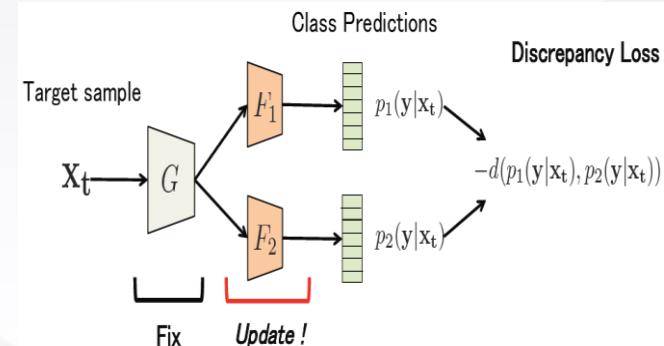
ADDA

Adversarial Discriminative Domain Adaptation



RAAN

Re-weighted Adversarial Adaptation Network



MCD

Maximize Classifier Discrepancy

Note: TL/DA in pose, identity face/person synthesis in Face Recognition/Re-ID are not included here



Maximum Mean Discrepancy (MMD)

Gretton et al. NIPS'06, NIPS'09, JMLR'12 from MPI, Germany proposed MMD. A non-parametric statistic for testing whether two distributions are different.

By using smooth functions “Rich” and “Restrictive”.

1. MMD vanishes if and only if $p=q$.
2. MMD empirical estimation can easily converge to its expectation.

In MMD, the unit balls in universal reproducing kernel Hilbert space are used as smooth functions.

Gaussian and Laplacian kernels are proved to be universal.

<http://www.gatsby.ucl.ac.uk/~gretton/mmd/mmd.htm>



Maximum Mean Discrepancy (MMD)

Definition 2 Let \mathcal{F} be a class of functions $f : \mathcal{X} \rightarrow \mathbb{R}$ and let p, q, X, Y be defined as above. Then we define the maximum mean discrepancy (MMD) and its empirical estimate as

$$\text{MMD} [\mathcal{F}, p, q] := \sup_{f \in \mathcal{F}} (\mathbf{E}_{x \sim p}[f(x)] - \mathbf{E}_{y \sim q}[f(y)]), \quad (1)$$

Arbitrary Function Space:

$$\text{MMD} [\mathcal{F}, X, Y] := \sup_{f \in \mathcal{F}} \left(\frac{1}{m} \sum_{i=1}^m f(x_i) - \frac{1}{n} \sum_{i=1}^n f(y_i) \right). \quad (2)$$

Theorem 3 Let \mathcal{F} be a unit ball in a universal RKHS \mathcal{H} , defined on the compact metric space \mathcal{X} , with associated kernel $k(\cdot, \cdot)$. Then $\text{MMD} [\mathcal{F}, p, q] = 0$ if and only if $p = q$.

Using $\mu[X] := \frac{1}{m} \sum_{i=1}^m \phi(x_i)$ and $k(x, x') = \langle \phi(x), \phi(x') \rangle$, an empirical estimate of MMD is

RKHS:

$$\text{MMD} [\mathcal{F}, X, Y] = \left[\frac{1}{m^2} \sum_{i,j=1}^m k(x_i, x_j) - \frac{2}{mn} \sum_{i,j=1}^{m,n} k(x_i, y_j) + \frac{1}{n^2} \sum_{i,j=1}^n k(y_i, y_j) \right]^{\frac{1}{2}}.$$

<http://www.gatsby.ucl.ac.uk/~gretton/mmd/mmd.htm>



Maximum Mean Discrepancy (MMD)

Publications with MMD:

Classifier
-level

Duan, et al. TPAMI'12 (AMKL, DTSVM)
Wang et al. ACM MM'18 (MEDA)

Feature-
level

Zhang, et al. CVPR'17 (JGSA)
Long, et al. ICCV'17 (JDA)
Ghifary et al. TPAMI'17 (SCA)
Deng et al. TNNLS'18 (EMFS)

Deep
transfer

Tzeng, et al. Arxiv'14 (DDC)
Yan, et al. CVPR'17 (WDAN)
Wu, et al. CVPR'17 (CDNN)
Long, et al. ICML'15, '17 (DAN, JAN)
Long, et al. NIPS'16 (RTN)

Other distribution measures other than MMD:

1. HSIC criterion (Gretton et al. ALT'05; Yan et al. TCYB'17, Wang et al. ICCV'17, CRTL)
2. Bregman divergence (Si et al. TKDE'10, TSL)
3. Manifold criterion (Zhang et al. TNNLS'18, MCTL):
4. Second-order statistic (Herath et al. CVPR'17, ILS; Sun et al. arXiv'17, CORAL)

$$W = \arg \min_{W \in R^{D \times d}} F(W) + \lambda D_W(P_L \| P_U)$$

$$\mathcal{L}_u = \frac{1}{p} \delta_s(\mathbf{W}_s^T \boldsymbol{\Sigma}_s \mathbf{W}_s, \mathbf{W}_t^T \boldsymbol{\Sigma}_t \mathbf{W}_t)$$



Our Recent Works



Table of Contents

Part I: Classifier-level Domain Adaptation

- [1] L. Zhang and D. Zhang, IEEE Trans. Image Processing, 2016.
- [2] L. Zhang and D. Zhang, IEEE Trans. Multimedia, 2016.

Part II: Feature-level Transfer Learning

- [3] L. Zhang, W. Zuo, and D. Zhang, IEEE Trans. Image Processing, 2016.
- [4] L. Zhang, J. Yang, and D. Zhang, Information Sciences, 2017.
- [5] S. Wang, L. Zhang, W. Zuo, ICCV W 2017.
- [6] L. Zhang, Y. Liu and P. Deng, IEEE Trans. Intru. Meas. 2017.
- [7] L. Zhang, S. Wang, G.B. Huang, W. Zuo, J. Yang, and D. Zhang, IEEE Trans. Neural Networks and Learning Systems, 2018.

Part III: Self-Adversarial Transfer Learning

- [8] Q. Duan, L. Zhang, W. Zuo, ACM MM, 2017.
- [9] L. Zhang, Q. Duan, W. Jia, D. Zhang, X. Wang, IEEE Trans. Cybernetics, 2018. in review

Part IV: Guide Learning (A try for TL/DA)

- [10] J. Fu, L. Zhang, B. Zhang, W. Jia, CCBR oral, 2018.
- [11] L. Zhang, J. Fu, S. Wang, D. Zhang, D.Y. Dong, C.L. Philip Chen, IEEE Trans. Neural Net. Learn. Syst. 2018. in review.



Cross-domain Classifier Model (EDA, TIP'16)

Common Classifier Learning: Semi-supervised Joint empirical risk for domain sharing.

$$R_{reg}[\theta, l(x, y, \theta)] = R_{emp}[\theta, l(x, y, \theta)] + \Omega[\theta]$$



$$R_{reg}[\theta, l(x, y, \theta)] = R_{emp}[\theta, l_S(x, y, \theta)] + \mu R_{emp}[\theta, l_T(x, y, \theta)] + \Omega[\theta]$$

Graph manifold
preservation

Label correction

Task A (source)

Cross-domain
classifier

Task B (target)

“Borrow” auxiliary data

Knowledge transfer

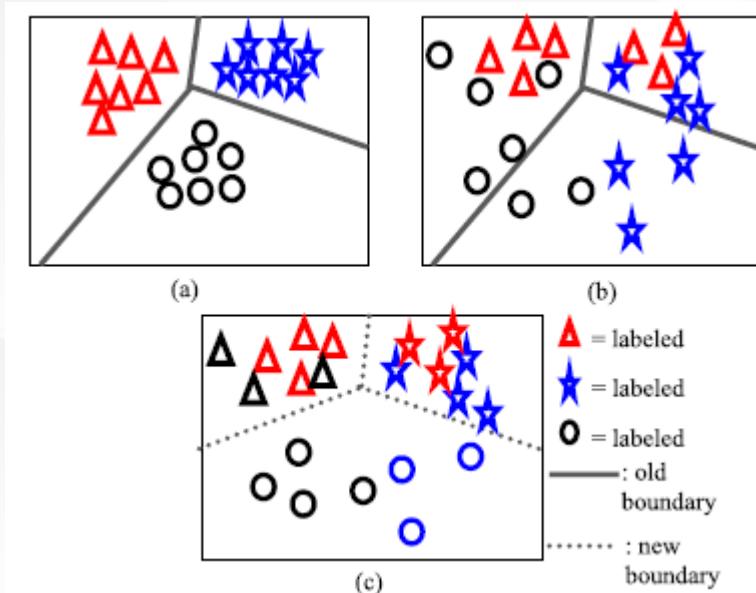
Classifier Level



Cross-domain Classifier Model (EDA)

By formulating a least-square loss function and a category transformation,

$$\begin{aligned} \min_{\beta, \Theta, \xi_S^i, \xi_T^j, \xi_{T,u}^k} & \|\beta\|_{2,1} + C_S \sum_{i=1}^{N_S} \|\xi_S^i\|_2^2 \\ & + C_T \sum_{j=1}^{N_T} \|\xi_T^j | \beta, \Theta\|_2^2 + \gamma \|\Theta - I\|_F^2 \\ & + \tau \sum_{k=1}^{N_{T,u}} \|\xi_{T,u}^k | \beta, \phi_\varphi\|_2^2 + \lambda \cdot \text{tr}(F^T \mathcal{L} F) \\ \text{s.t. } & \begin{cases} H_S^i \beta = t_S^i - \xi_S^i, & i = 1, \dots, N_S \\ H_T^j \beta = (t_T^j)^T \circ \Theta - \xi_T^j | \beta, \Theta, & j = 1, \dots, N_T \\ H_{T,u}^k \beta = \phi_{\varphi, T,u}^k - \xi_{T,u}^k | \beta, \phi_\varphi, & k = 1, \dots, N_{T,u} \\ F = H \beta \end{cases} \end{aligned}$$





Mv-EDA (Multi-view extension)

$$\begin{aligned}
 & \min_{\beta_v, \Theta_v, \alpha_v, \xi_{S,v}^i, \xi_{T,v}^i, \xi_{T_u,v}^k} \sum_{v=1}^V \|\beta_v\|_{2,1} \\
 & + C_S \sum_{i=1}^{N_{S\ell}} \sum_{v=1}^V \alpha_v \left\| \xi_{S,v}^i \right\|^2 \\
 & + C_T \sum_{j=1}^{N_{T\ell}} \sum_{v=1}^V \alpha_v \left\| \xi_{T,u}^j | \beta_v, \Theta_v \right\|^2 \\
 & + \gamma \sum_{v=1}^V \alpha_v \left\| \Theta_v - \mathbf{I} \right\|_F^2 \\
 & + \tau \sum_{k=1}^{N_{T_u}} \sum_{v=1}^V \alpha_v \left\| \xi_{T_u,v}^k | \beta_v, \phi_p^v \right\|^2 \\
 & + \lambda \cdot \text{tr} \left(\sum_{v=1}^V \alpha_v^r \mathbf{F}_v^T \mathcal{L}_v \mathbf{F}_v \right) \quad (22)
 \end{aligned}$$

s.t.

$$\begin{cases}
 \mathbf{H}_{S,v}^i \beta_v = \mathbf{t}_S^i - \xi_{S,v}^i, & i = 1, \dots, N_{S\ell} \\
 \mathbf{H}_{T,v}^j \beta_v = (\mathbf{t}_T^j)^T \circ \Theta_v - \xi_{T,v}^j | \beta_v, \Theta_v, & j = 1, \dots, N_{T\ell} \\
 \mathbf{H}_{T_u,v}^k \beta_v = \phi_{\emptyset, T_u}^{k,v} - \xi_{T_u,v}^k | \beta_v, \phi_p^v, & k = 1, \dots, N_{T_u} \\
 \mathbf{F}_v = \mathbf{H}_v \beta_v \\
 \sum_{v=1}^V \alpha_v = 1, & 0 < \alpha_v < 1 \\
 r > 1, \gamma, \lambda, \tau, C_S, C_T > 0
 \end{cases}$$

Algorithm 3 Complete EDA

Input:

1: Training samples $\{\mathbf{X}_{S,v}, \mathbf{T}_{S,v}\} = \{\mathbf{x}_{S,v}^i, t_{S,v}^i\}_{i=1}^{N_S}$ of the source domain S w.r.t. the v -th modality, $v = 1, \dots, V$;

2: Labeled guide samples $\{\mathbf{X}_{T,v}, \mathbf{T}_{T,v}\} = \{\mathbf{x}_{T,v}^j, t_{T,v}^j\}_{j=1}^{N_{T\ell}}$ of the target domain T w.r.t. the v -th modality, $v = 1, \dots, V$;

3: Unlabeled samples $\{\mathbf{X}_{T_u,v}, \mathbf{T}_{T_u,v}\} = \{\mathbf{x}_{T_u,v}^j, t_{T_u,v}^j\}_{j=1}^{N_{T_u}}$ of the target domain T w.r.t. the v -th modality, $v = 1, \dots, V$;

4: The trade-off parameters;

Output: β_v and α_v ($v = 1, \dots, V$)

Procedure:

Stage 1. EDA Network Initialization.

5: Initialize the EDA network of L hidden neurons with randomly selected input weights \mathbf{W} and hidden bias \mathbf{B} with 0-1 uniform distribution;

Stage 2. EDA Feature Mapping and Graph Construction.

6: Calculate the hidden layer output matrix $\mathbf{H}_{S,v}$, $\mathbf{H}_{T,v}$, $\mathbf{H}_{T_u,v}$ as $\mathbf{H}_{S,v} = \mathcal{H}(\mathbf{W} \cdot \mathbf{X}_{S,v} + \mathbf{B})$, $\mathbf{H}_{T,v} = \mathcal{H}(\mathbf{W} \cdot \mathbf{X}_{T,v} + \mathbf{B})$ and $\mathbf{H}_{T_u,v} = \mathcal{H}(\mathbf{W} \cdot \mathbf{X}_{T_u,v} + \mathbf{B})$, respectively;

7: Compute \mathbf{H}_v w.r.t. all the instances in target domain;

8: Compute the graph Laplacian matrix \mathcal{L}_v ;

Stage 3. Learning algorithm.

9: if $V < 2$ then

10: Call Algorithm 1.

11: else Call Algorithm 2.

12: end if



Results For Video Event Recognition



TABLE II

MEAN AVERAGE PRECISION (%) OF THREE CASES WITH DIFFERENT NUMBER OF LABELED TARGET TRAINING DATA ($m = 1, 3, 5, 7, 10$)

| # m | Methods | SVM_T | SVM_ST | FR | MKL | DTMKL | A-MKL | EDA _{AMKL} | MvEDA | Improvement |
|-------|---------|----------------|----------------|----------------|-----------------------|----------------|-----------------------|-----------------------|-----------------------|-------------|
| 1 | SIFT | 38.8 \pm 4.8 | 49.4 \pm 3.2 | 47.3 \pm 0.4 | 43.9 \pm 2.4 | 48.7 \pm 1.4 | 51.6 \pm 1.4 | 51.9 \pm 1.0 | - | 0.3% |
| | ST | 27.3 \pm 3.8 | 23.9 \pm 1.2 | 28.8 \pm 2.1 | 35.1 \pm 1.9 | 33.4 \pm 1.0 | 37.6 \pm 1.7 | 38.7 \pm 1.6 | - | 1.1% |
| | SIFT+ST | 36.9 \pm 7.3 | 33.9 \pm 1.8 | 43.9 \pm 3.4 | 45.3 \pm 2.1 | 48.8 \pm 1.6 | 52.2 \pm 1.0 | - | 56.0 \pm 1.4 | 3.8% |
| 3 | SIFT | 42.3 \pm 5.5 | 53.9 \pm 5.6 | 50.0 \pm 5.6 | 47.2 \pm 2.6 | 52.4 \pm 1.9 | 57.1 \pm 2.3 | 57.2 \pm 2.1 | - | 0.1% |
| | ST | 32.6 \pm 2.1 | 24.7 \pm 2.2 | 28.4 \pm 2.6 | 35.3 \pm 1.6 | 31.1 \pm 2.6 | 37.2 \pm 1.6 | 39.0 \pm 1.8 | - | 1.8% |
| | SIFT+ST | 42.0 \pm 4.9 | 36.2 \pm 3.4 | 44.1 \pm 3.6 | 46.9 \pm 2.5 | 53.8 \pm 2.9 | 58.2 \pm 1.9 | - | 60.3 \pm 1.8 | 2.1% |
| 5 | SIFT | 46.8 \pm 4.1 | 54.9 \pm 5.2 | 53.3 \pm 5.9 | 49.0 \pm 8.1 | 54.8 \pm 7.6 | 57.4 \pm 9.0 | 57.4 \pm 8.1 | - | 0% |
| | ST | 35.4 \pm 3.6 | 25.1 \pm 2.1 | 29.6 \pm 2.2 | 37.7 \pm 2.3 | 33.3 \pm 3.1 | 41.6 \pm 7.0 | 43.1 \pm 6.4 | - | 1.5% |
| | SIFT+ST | 48.4 \pm 3.4 | 39.2 \pm 2.4 | 48.8 \pm 4.3 | 44.2 \pm 6.0 | 58.1 \pm 8.4 | 57.7 \pm 9.0 | - | 62.4 \pm 7.9 | 4.7% |
| 7 | SIFT | 66.5 \pm 2.7 | 71.8 \pm 3.9 | 71.9 \pm 3.8 | 62.1 \pm 2.2 | 71.6 \pm 4.5 | 72.6 \pm 4.4 | 72.8 \pm 3.5 | - | 0.2% |
| | ST | 42.2 \pm 3.2 | 24.9 \pm 1.3 | 30.4 \pm 0.7 | 46.3 \pm 2.0 | 37.4 \pm 1.6 | 41.0 \pm 2.6 | 49.8 \pm 4.6 | - | 3.5% |
| | SIFT+ST | 63.8 \pm 2.5 | 54.0 \pm 4.0 | 67.8 \pm 2.3 | 58.4 \pm 3.7 | 72.9 \pm 4.5 | 73.2 \pm 4.6 | - | 76.5 \pm 4.4 | 3.3% |
| 10 | SIFT | 68.7 \pm 2.7 | 73.3 \pm 3.5 | 74.0 \pm 3.8 | 66.0 \pm 2.9 | 73.6 \pm 2.6 | 74.4 \pm 2.3 | 74.7 \pm 2.5 | - | 0.3% |
| | ST | 46.0 \pm 4.7 | 25.2 \pm 3.4 | 30.5 \pm 3.0 | 45.6 \pm 4.0 | 39.2 \pm 3.9 | 42.0 \pm 8.2 | 48.0 \pm 3.5 | - | 2.4% |
| | SIFT+ST | 66.0 \pm 5.6 | 59.7 \pm 2.3 | 69.1 \pm 2.5 | 58.5 \pm 4.0 | 76.5 \pm 2.2 | 74.9 \pm 2.1 | - | 77.3 \pm 2.9 | 2.4% |



Results For Object Recognition on 4DA Office Dataset

TABLE V

RECOGNITION ACCURACIES (%) FOR ALL METHODS ON THE 4DA EXTENDED *Office* DATASET WITH LOW-LEVEL SURF FEATURE

| Method | SVM_S | SVM_T | LandMark | SGF | GFK | HFA | ARC-t | Symm | MMDT | EDA _{SVM} | Improvement |
|---------|----------|----------|----------|----------|----------|----------|----------|----------|----------|--------------------|-------------|
| C→D | 35.6±0.7 | 55.8±0.9 | 57.3 | 50.2±0.8 | 57.7±1.1 | 51.9±1.1 | 50.6±0.8 | 48.6±1.1 | 56.5±0.9 | 59.0±1.2 | 1.3% |
| C→W | 30.8±1.1 | 60.3±1.0 | 49.5 | 54.2±0.9 | 63.7±0.8 | 60.5±0.9 | 55.9±1.0 | 50.5±1.6 | 63.8±1.1 | 67.3±0.8 | 3.5% |
| C→A | 35.9±0.4 | 45.3±0.9 | 56.7 | 42.0±0.5 | 44.7±0.8 | 45.5±0.9 | 44.1±0.6 | 43.8±0.6 | 49.4±0.8 | 53.5±0.5 | - |
| A→C | 35.1±0.3 | 32.0±0.8 | 45.5 | 37.5±0.4 | 36.0±0.5 | 31.1±0.6 | 37.0±0.4 | 39.1±0.5 | 36.4±0.8 | 43.8±0.4 | - |
| A→W | 33.9±0.7 | 62.4±0.9 | 46.1 | 54.2±0.8 | 58.6±1.0 | 61.8±1.1 | 55.7±0.9 | 51.0±1.4 | 64.6±1.2 | 68.9±1.0 | 4.3% |
| A→D | 35.0±0.8 | 55.9±0.8 | 47.1 | 46.9±1.1 | 50.7±0.8 | 52.7±0.9 | 50.2±0.7 | 47.9±1.4 | 56.7±1.3 | 57.6±1.0 | 0.9% |
| W→C | 31.3±0.4 | 30.4±0.7 | 35.4 | 32.9±0.7 | 31.1±0.6 | 29.4±0.6 | 31.9±0.5 | 34.0±0.5 | 32.2±0.8 | 38.6±0.5 | 3.2% |
| W→A | 35.7±0.4 | 45.6±0.7 | 40.2 | 43.4±0.7 | 44.1±0.4 | 45.9±0.7 | 43.4±0.5 | 43.7±0.7 | 47.7±0.9 | 52.4±0.9 | 4.7% |
| W→D | 66.6±0.7 | 55.1±0.8 | 75.2 | 78.6±0.4 | 70.5±0.7 | 51.7±1.0 | 71.3±0.8 | 69.8±1.0 | 67.0±1.1 | 73.8±0.8 | - |
| D→C | 31.4±0.3 | 31.7±0.6 | - | 32.9±0.4 | 32.9±0.5 | 31.0±0.5 | 33.5±0.4 | 34.9±0.4 | 34.1±0.8 | 38.0±0.4 | 3.1% |
| D→A | 34.0±0.3 | 45.7±0.9 | - | 44.9±0.7 | 45.7±0.6 | 45.8±0.9 | 42.5±0.5 | 42.7±0.5 | 46.9±1.0 | 50.4±0.8 | 3.5% |
| D→W | 74.3±0.5 | 62.1±0.8 | - | 78.6±0.4 | 76.5±0.5 | 62.1±0.7 | 78.3±0.5 | 78.4±0.9 | 74.1±0.8 | 84.1±0.6 | 5.5% |
| Average | 40.0±0.6 | 48.5±0.8 | 50.3 | 49.7±0.7 | 51.0±0.7 | 47.4±0.8 | 49.5±0.6 | 48.7±0.9 | 52.5±1.0 | 57.3±0.8 | 4.8% |



Table of Contents

Part I: Classifier-level Domain Adaptation

- [1] L. Zhang and D. Zhang, IEEE Trans. Image Processing, 2016.
- [2] L. Zhang and D. Zhang, IEEE Trans. Multimedia, 2016.

Part II: Feature-level Transfer Learning

- [3] L. Zhang, W. Zuo, and D. Zhang, IEEE Trans. Image Processing, 2016.
- [4] L. Zhang, J. Yang, and D. Zhang, Information Sciences, 2017.
- [5] S. Wang, L. Zhang, W. Zuo, ICCV W 2017.
- [6] L. Zhang, Y. Liu and P. Deng, IEEE Trans. Intru. Meas. 2017.
- [7] L. Zhang, S. Wang, G.B. Huang, W. Zuo, J. Yang, and D. Zhang, IEEE Trans. Neural Networks and Learning Systems, 2018.

Part III: Self-Adversarial Transfer Learning

- [8] Q. Duan, L. Zhang, W. Zuo, ACM MM, 2017.
- [9] L. Zhang, Q. Duan, W. Jia, D. Zhang, X. Wang, IEEE Trans. Cybernetics, 2018. in review

Part IV: Guide Learning (A try for TL/DA)

- [10] J. Fu, L. Zhang, B. Zhang, W. Jia, CCBR oral, 2018.
- [11] L. Zhang, J. Fu, S. Wang, D. Zhang, D.Y. Dong, C.L. Philip Chen, IEEE Trans. Neural Net. Learn. Syst. 2018. in review.



Counterparts

Subspace Transfer

- SA, Fernando et al., ICCV'13;
- TCA, Pan et al., TNNLS '11;
- MMDT, Hoffman et al., IJCV'14;
- Kulis et al., CVPR'12;
- SGF, Gopalan et al., ICCV'11;
- GFK, Gong et al., CVPR'12;

Reconstruction Transfer

- LTSL, Shao et al., IJCV'14;
- RDALR, Jhuo et al., CVPR'12;
- DTSL, Fang et al., TIP'16;



Our Work

- **CDSL**: Cross-domain discriminative subspace learning ([T-IM'17](#))
- **LSDT**: Latent sparse domain transfer learning ([T-IP'16](#))
- **DKTL**: Discriminative kernel transfer learning ([Info. Sci.'17](#); [IJCNN'16](#))
- **CRTL**: Class-specific Reconstruction transfer learning ([ICCV'17](#))
- **MCTL**: Manifold Criterion Guided transfer learning ([T-NNLS'18](#))



CDSL Model

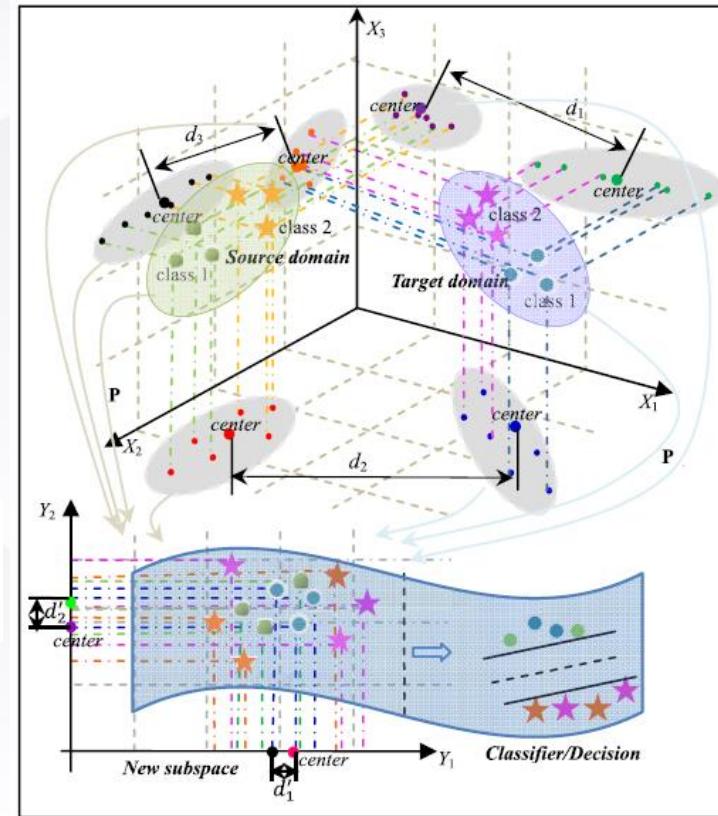
- CDSL: Cross-domain discriminative subspace learning (T-IM'17)

Class discrimination (源域数据类间判别性)

Energy preservation (目标域数据能量保持)

Domain mean discrepancy (域间中心差异最小)

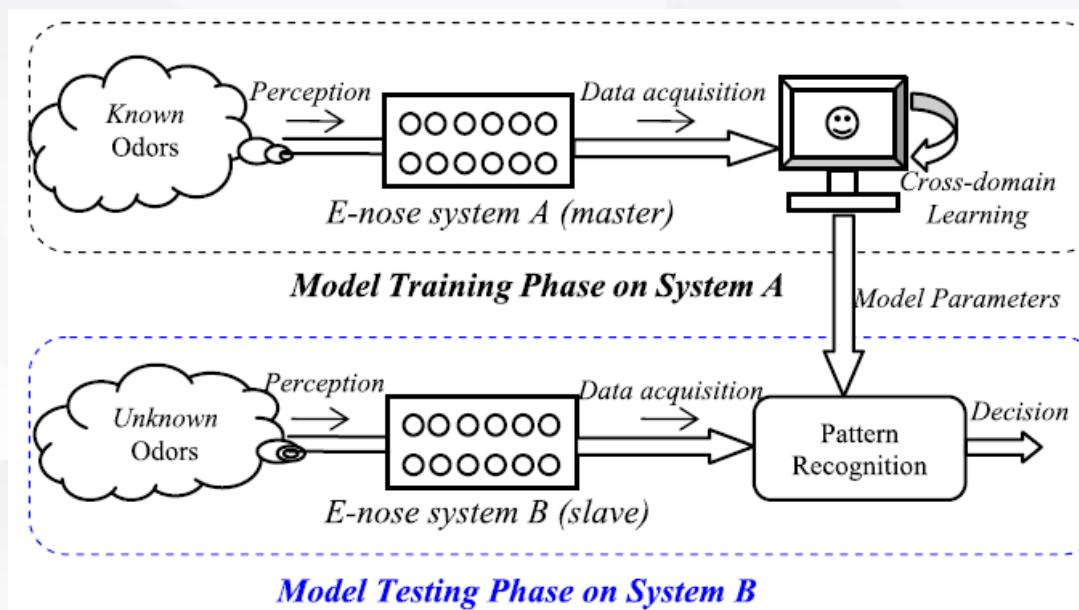
$$\max_{\mathbf{P}} \frac{\text{Tr}(\mathbf{P}^T \mathbf{S}_B^S \mathbf{P}) + \lambda_0 \cdot \text{Tr}(\mathbf{P}^T \mathbf{X}_T \mathbf{X}_T^T \mathbf{P})}{\text{Tr}(\mathbf{P}^T \mathbf{S}_W^S \mathbf{P}) + \lambda_1 \cdot \sum_{i=1}^M w_i \left\| \frac{1}{N_S} \sum_{j=1}^{N_S} \mathbf{P}^T \mathbf{x}_S^{ij} - \frac{1}{N_T} \sum_{k=1}^{N_T} \mathbf{P}^T \mathbf{x}_T^{ik} \right\|_2^2}$$





CDSL Model

$$\max_{\mathbf{P}} \frac{\text{Tr}(\mathbf{P}^T \mathbf{S}_B^S \mathbf{P}) + \lambda_0 \cdot \text{Tr}(\mathbf{P}^T \mathbf{X}_T \mathbf{X}_T^T \mathbf{P})}{\text{Tr}(\mathbf{P}^T \mathbf{S}_W^S \mathbf{P}) + \lambda_1 \cdot \sum_{i=1}^M w_i \left\| \frac{1}{N_S} \sum_{j=1}^{N_S} \mathbf{P}^T \mathbf{x}_S^{ij} - \frac{1}{N_T} \sum_{k=1}^{N_T} \mathbf{P}^T \mathbf{x}_T^{ik} \right\|_2^2}$$





Results for Cross-system Application

TABLE III
COMPARISONS OF RECOGNITION ACCURACY ON TWO TASKS (SETTING 1)

| Cross-domain task | SVM | PCA | LDA | LPP | NPE | NCA | MDS | LFDA | GFK | SGF | SA | OSC | DS | GLSW | CDSL |
|---------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------------|
| Source domain → target domain 1 | 51.97 | 51.97 | 51.97 | 53.95 | 53.62 | 41.28 | 51.15 | 61.84 | 33.88 | 55.10 | 41.10 | 34.38 | 45.00 | 40.46 | 71.88 |
| Source domain → target domain 2 | 60.59 | 60.59 | 56.77 | 57.81 | 54.69 | 33.85 | 58.51 | 61.63 | 32.81 | 57.49 | 31.12 | 36.46 | 42.62 | 53.65 | 71.88 |

TABLE V
COMPARISONS OF RECOGNITION ACCURACY ON SETTING 2 (TASK 1: SOURCE DOMAIN → TARGET DOMAIN 1) WITH A DIFFERENT NUMBER OF LABELED TARGET DATA PER CLASS

| No. of labeled target data per class (k) | 1 | 3 | 5 | 7 | 9 | Average |
|--|--------------|--------------|--------------|--------------|--------------|--------------|
| SVM | 59.14 | 63.22 | 62.80 | 70.49 | 70.76 | 65.28 |
| PCA | 59.14 | 63.39 | 65.40 | 70.49 | 70.76 | 65.84 |
| LDA | 67.77 | 71.36 | 74.22 | 75.09 | 76.17 | 72.92 |
| LPP | 65.46 | 69.83 | 71.45 | 72.08 | 71.48 | 70.06 |
| NPE | 64.78 | 64.07 | 63.49 | 71.55 | 71.84 | 67.15 |
| NCA | 52.49 | 50.85 | 53.81 | 50.00 | 63.00 | 54.03 |
| MDS | 61.13 | 64.75 | 65.57 | 70.32 | 72.92 | 66.94 |
| LFDA | 62.13 | 67.12 | 71.63 | 76.86 | 74.91 | 70.53 |
| GFK | 33.39 | 34.07 | 35.81 | 37.63 | 38.45 | 35.87 |
| SGF | 66.61 | 66.95 | 67.13 | 70.14 | 72.74 | 68.71 |
| SA | 49.67 | 58.64 | 59.69 | 58.83 | 61.91 | 57.75 |
| OSC | 41.03 | 45.25 | 45.85 | 43.29 | 45.13 | 44.11 |
| DS | 54.69 | 62.56 | 59.78 | 59.09 | 59.49 | 59.12 |
| GLSW | 52.99 | 51.69 | 69.90 | 60.42 | 59.75 | 58.95 |
| CDSL ($M=1$) | 72.76 | 77.97 | 79.07 | 81.80 | 83.21 | 78.96 |



LSDT (TIP'16)

Latent Sparse Domain Transfer (LSDT)

Difference from:

RDALR, Jhuo et al., CVPR'12;
LTSL, Shao et al., IJCV'14;

RDALR:

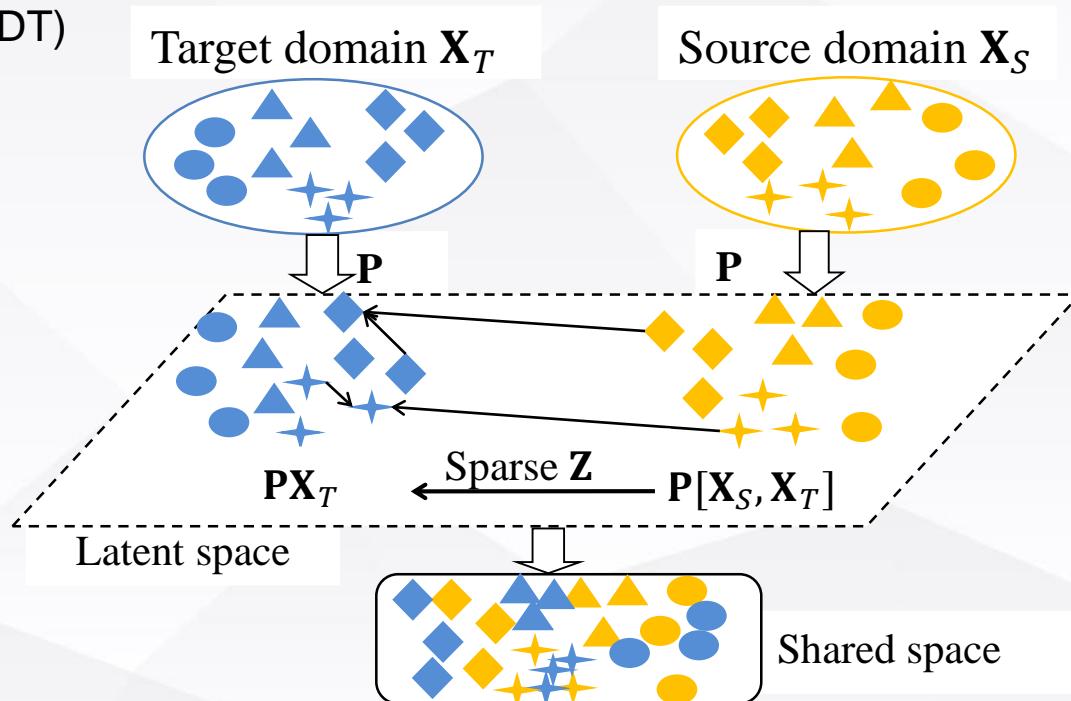
$$\min_{W, Z, E} \text{rank}(Z) + \alpha \|E\|_{2,1}$$

$$\text{s.t. } W X_S = X_T Z + E, \quad W W^T = I$$

LTSI:

$$\min_{W, Z, E} F(W, X_S) + \lambda_1 \text{rank}(Z) + \lambda_2 \|E\|_{2,1}$$

$$\text{s.t. } W^T X_T = W^T X_S Z + E, \quad W^T U_2 W = I$$



在重建迁移过程中，学习共同子空间

Idea of LSDT



LSDT (TIP'16)

LSDT



$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{P}} \quad & \|\mathbf{Z}\|_1 + \lambda_1 \|\mathbf{P} \mathbf{X}_T - \mathbf{P} [\mathbf{X}_S, \mathbf{X}_T] \mathbf{Z}\|_F^2 \\ & + \lambda_2 \left\| [\mathbf{X}_S, \mathbf{X}_T] - \mathbf{P}^T \mathbf{P} [\mathbf{X}_S, \mathbf{X}_T] \right\|_F^2 \\ \text{s.t. } & \mathbf{P} \mathbf{P}^T = \mathbf{I}, \quad \mathbf{1}_{N_S+N_T}^T \mathbf{Z} = \mathbf{1}_{N_T}^T, \quad Z_{N_S+i,i} = 0, \\ & \forall i = 1, \dots, N_T \end{aligned}$$

NLSDT

$$\begin{aligned} \min_{\mathbf{Z}, \Phi} \quad & \|\mathbf{Z}\|_1 + \lambda_1 \left\| \Phi^T \mathcal{K}_T - \Phi^T \mathbf{K} \mathbf{Z} \right\|_F^2 \\ & + \lambda_2 \operatorname{Tr} \left(\left(\mathbf{I} - \Phi \Phi^T \mathcal{K} \right)^T \mathcal{K} \left(\mathbf{I} - \Phi \Phi^T \mathcal{K} \right) \right) \\ \text{s.t. } & \Phi^T \mathcal{K} \Phi = \mathbf{I}, \quad \mathbf{1}_{N_S+N_T}^T \mathbf{Z} = \mathbf{1}_{N_T}^T, \\ & Z_{N_S+i,i} = 0 \quad \forall i = 1, \dots, N_T. \end{aligned} \tag{14}$$



LSDT (TIP'16)

Model Solution:

Ease Implementation

- ◆ Solve Z : ADMM algorithm

Variable alternating optimization



- ◆ Solve Φ : Eigenvalue decomposition algorithm



- ◆ Converge (over)

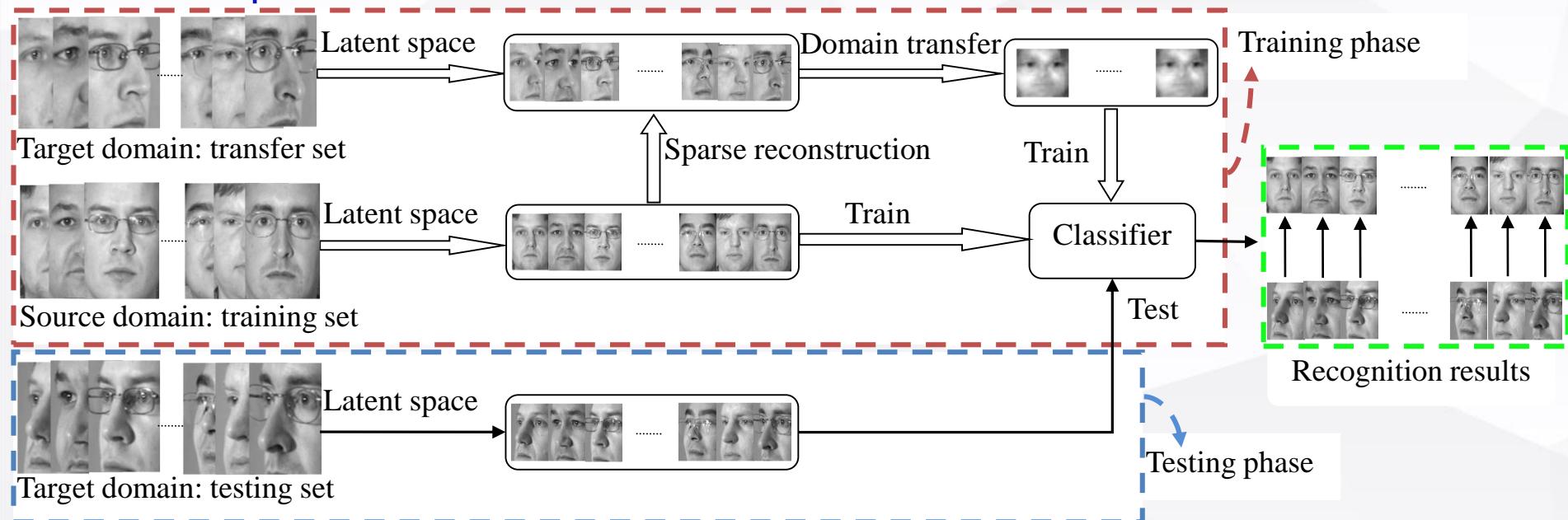


Iteration



LSDT (TIP'16)

Pipeline of our LSDT





Cross-domain Experiments



4DA office objects



CMU PIE Faces



Consumer videos & YouTube videos:

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 |
| 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 6 | 6 | 6 | 6 | 6 | 6 | 6 | 6 |
| 7 | 7 | 7 | 7 | 7 | 7 | 7 | 7 |
| 8 | 8 | 8 | 8 | 8 | 8 | 8 | 8 |
| 9 | 9 | 9 | 9 | 9 | 9 | 9 | 9 |

Handwritten digits



Experiment on Multi-task Object Recognition

| Domains | | Compared methods | | | | | | | Our method | |
|---------------|--------|------------------|----------|----------|----------|-----------|-------------|--------------|-----------------|-----------------|
| Source | Target | ASVM [12] | GFK [10] | SGF [8] | SA [41] | RDALR [2] | LTSI-PCA[1] | LTSI-LDA [1] | LSDT | NLSDT |
| Amazon | Webcam | 42.2±0.9 | 46.4±0.5 | 45.1±0.6 | 48.4±0.6 | 50.7±0.8 | 49.8±0.4 | 53.5±0.4 | 50.0±1.3 | 56.3±0.7 |
| DSLR | Webcam | 33.0±0.8 | 61.3±0.4 | 61.4±0.4 | 61.8±0.9 | 36.9±1.9 | 62.4±0.3 | 54.4±0.4 | 69.4±0.7 | 69.9±0.3 |
| Webcam | DSLR | 26.0±0.7 | 66.3±0.4 | 63.4±0.5 | 65.7±0.5 | 32.9±1.2 | 63.9±0.3 | 59.1±0.5 | 72.6±0.9 | 74.6±0.5 |
| Amazon+DSLR | Webcam | 30.4±0.6 | 34.3±0.6 | 31.0±1.6 | 54.4±0.9 | 36.9±1.1 | 55.3±0.3 | 30.2±0.5 | 69.0±0.8 | 66.1±0.7 |
| Amazon+Webcam | DSLR | 25.3±1.1 | 52.0±0.8 | 25.0±0.4 | 37.5±1.0 | 31.2±1.3 | 57.7±0.4 | 43.0±0.3 | 67.5±1.8 | 65.7±0.9 |
| DSLR+Webcam | Amazon | 17.3±0.9 | 21.7±0.5 | 15.0±0.4 | 16.5±0.4 | 20.9±0.9 | 20.0±0.2 | 17.1±0.3 | 22.0±0.1 | 23.2±0.6 |



Experiment on Multi-task Object Recognition (AlexNet)

| Method | Layer | A→D | C→D | W→D | A→C | W→C | D→C | D→A | W→A | C→A | C→W | D→W | A→W |
|--------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| SourceOnly | f ₆ | 80.8±0.8 | 76.6±2.2 | 96.1±0.4 | 79.3±0.3 | 59.5±0.9 | 67.3±1.2 | 77.0±1.0 | 66.8±1.0 | 85.8±0.4 | 67.5±1.6 | 95.4±0.6 | 70.5±0.9 |
| | f ₇ | 81.3±0.7 | 77.6±1.1 | 96.2±0.6 | 79.3±0.3 | 68.1±0.6 | 74.3±0.6 | 81.8±0.5 | 73.4±0.7 | 86.5±0.5 | 67.8±1.8 | 95.1±0.8 | 71.6±0.6 |
| NaïveComb | f ₆ | 94.5±0.4 | 92.9±0.8 | 99.1±0.2 | 84.0±0.3 | 81.7±0.5 | 83.0±0.3 | 90.5±0.2 | 90.1±0.2 | 89.9±0.2 | 91.6±0.8 | 97.9±0.3 | 90.4±0.8 |
| | f ₇ | 94.1±0.8 | 92.8±0.7 | 98.9±0.2 | 83.4±0.4 | 81.2±0.4 | 82.7±0.4 | 90.9±0.3 | 90.6±0.2 | 90.3±0.2 | 90.6±0.8 | 98.0±0.2 | 91.1±0.8 |
| SGF [8] | f ₆ | 90.5±0.8 | 93.1±1.2 | 97.7±0.4 | 77.1±0.8 | 74.1±0.8 | 75.9±1.0 | 88.0±0.8 | 87.2±0.5 | 88.5±0.4 | 89.4±0.9 | 96.8±0.4 | 87.2±0.9 |
| | f ₇ | 92.0±1.3 | 92.4±1.1 | 97.6±0.5 | 77.4±0.7 | 76.8±0.7 | 78.2±0.7 | 88.0±0.5 | 86.8±0.7 | 89.3±0.4 | 87.8±0.8 | 95.7±0.8 | 88.1±0.8 |
| GFK [10] | f ₆ | 92.6±0.7 | 92.0±1.2 | 97.8±0.5 | 78.9±1.1 | 77.5±0.8 | 78.8±0.8 | 88.9±0.3 | 86.2±0.8 | 87.5±0.3 | 87.7±0.8 | 97.0±0.8 | 89.5±0.8 |
| | f ₇ | 94.3±0.7 | 91.9±0.8 | 98.5±0.3 | 79.1±0.7 | 76.1±0.7 | 77.5±0.8 | 90.1±0.4 | 85.6±0.5 | 88.4±0.4 | 86.4±0.7 | 96.5±0.3 | 88.6±0.8 |
| SA [41] | f ₆ | 94.2±0.5 | 93.0±1.0 | 98.6±0.5 | 83.1±0.7 | 81.1±0.5 | 82.4±0.7 | 90.4±0.4 | 89.8±0.4 | 89.5±0.4 | 91.2±0.9 | 97.5±0.7 | 90.3±1.2 |
| | f ₇ | 92.8±1.0 | 92.1±0.9 | 98.5±0.3 | 83.3±0.2 | 81.0±0.6 | 82.9±0.7 | 90.7±0.5 | 90.9±0.4 | 89.9±0.5 | 89.0±1.1 | 97.5±0.4 | 87.8±1.4 |
| LTSL-PCA [1] | f ₆ | 94.6±0.6 | 93.4±0.6 | 99.2±0.2 | 85.5±0.3 | 82.0±0.5 | 84.7±0.5 | 91.2±0.2 | 89.5±0.2 | 91.3±0.2 | 90.2±0.8 | 97.0±0.5 | 89.4±1.2 |
| | f ₇ | 95.7±0.5 | 94.6±0.8 | 98.4±0.2 | 86.0±0.2 | 83.5±0.4 | 85.4±0.4 | 92.3±0.2 | 91.5±0.2 | 92.4±0.2 | 90.9±0.9 | 96.5±0.2 | 91.2±1.1 |
| LTSL-LDA [1] | f ₆ | 95.5±0.3 | 93.6±0.5 | 99.1±0.2 | 85.3±0.2 | 82.3±0.4 | 84.4±0.2 | 91.1±0.2 | 90.6±0.2 | 90.4±0.1 | 91.8±0.7 | 98.2±0.3 | 92.2±0.4 |
| | f ₇ | 94.5±0.5 | 93.5±0.8 | 98.8±0.2 | 85.4±0.1 | 82.6±0.3 | 84.8±0.2 | 91.9±0.2 | 91.0±0.2 | 90.9±0.1 | 90.8±0.7 | 97.8±0.3 | 91.5±0.5 |
| LSDT | f ₆ | 96.4±0.4 | 95.4±0.5 | 99.4±0.1 | 85.9±0.2 | 83.1±0.3 | 85.2±0.2 | 92.2±0.2 | 91.0±0.2 | 92.1±0.1 | 93.3±0.8 | 98.7±0.2 | 92.1±0.8 |
| | f ₇ | 96.0±0.4 | 94.6±0.5 | 99.3±0.1 | 87.0±0.2 | 84.2±0.3 | 86.2±0.2 | 92.5±0.2 | 91.7±0.2 | 92.5±0.1 | 93.5±0.8 | 98.3±0.2 | 92.9±0.8 |
| NLSDT | f ₆ | 96.4±0.4 | 95.7±0.5 | 99.5±0.1 | 85.8±0.2 | 83.3±0.3 | 85.3±0.2 | 92.3±0.2 | 91.1±0.2 | 91.9±0.1 | 92.9±0.7 | 98.6±0.2 | 94.2±0.4 |
| | f ₇ | 96.0±0.4 | 94.4±0.8 | 99.4±0.2 | 86.9±0.2 | 84.3±0.3 | 86.2±0.2 | 92.5±0.2 | 91.9±0.2 | 92.3±0.1 | 93.2±0.8 | 98.1±0.3 | 94.1±0.4 |



Experiment on Cross Video Event Recognition (跨视频事件识别)

CLASSIFICATION ACCURACY (%) OVER 6 VISUAL EVENTS WITH DIFFERENT NUMBER OF LABELED TARGET VIDEOS PER EVENT

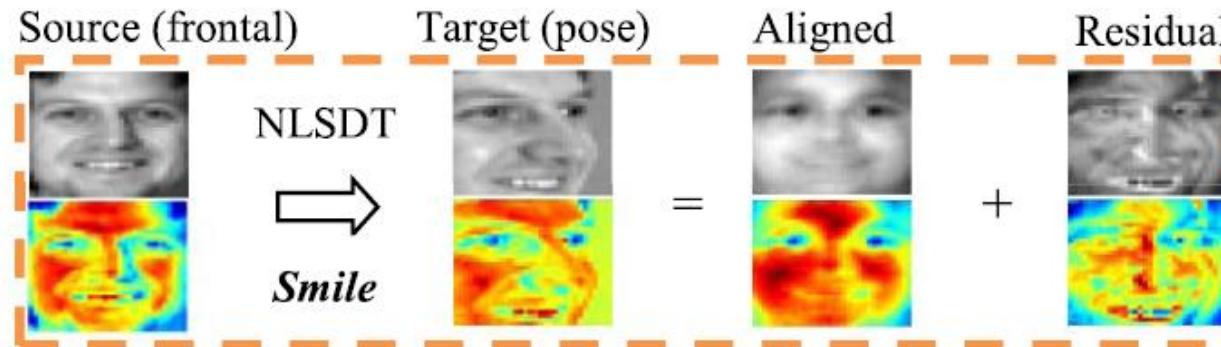
| m | Feature | Compared methods | | | | | | | Our methods | | |
|----|---------|------------------|------------|-----------------|----------|----------|----------|-----------------|--------------|-----------------|-----------------|
| | | Naïve Comb | DTSVM [16] | A-MKL [17] | GFK [10] | SGF [8] | SA [41] | LTSL-PCA [1] | LTSL-LDA [1] | LSDT | NLSDT |
| 1 | SIFT | 55.4±0.5 | 50.9±0.9 | 57.0±0.8 | 24.0±0.4 | 31.4±3.6 | 43.5±1.4 | 57.8±0.8 | 55.7±0.5 | 56.3±0.2 | 58.3±0.1 |
| | ST | 43.1±0.3 | 45.3±0.2 | 49.2±0.4 | 29.7±0.9 | 40.5±0.4 | 45.5±0.6 | 48.2±0.3 | 47.5±0.1 | 48.4±0.4 | 49.2±0.3 |
| | SIFT+ST | 46.8±0.6 | 52.8±0.4 | 57.7±0.8 | 21.4±1.0 | 42.2±2.1 | 46.0±1.1 | 59.6±1.3 | 60.7±0.7 | 60.8±0.8 | 62.0±0.6 |
| 3 | SIFT | 59.1±1.2 | 56.0±0.5 | 61.4±1.3 | 32.3±1.6 | 43.6±1.6 | 50.3±2.5 | 62.3±0.9 | 61.3±0.4 | 58.2±0.6 | 63.8±1.1 |
| | ST | 45.2±0.9 | 45.4±0.7 | 48.3±0.4 | 26.4±0.7 | 44.3±0.5 | 44.3±0.5 | 49.3±0.1 | 48.4±0.3 | 49.3±0.3 | 49.7±0.2 |
| | SIFT+ST | 51.2±1.6 | 55.4±0.0 | 61.5±1.0 | 27.0±1.6 | 40.1±4.2 | 46.4±0.6 | 62.8±1.0 | 60.5±0.6 | 61.8±1.2 | 67.1±0.6 |
| 5 | SIFT | 63.8±2.0 | 60.0±1.6 | 65.0±2.2 | 34.2±0.9 | 36.5±2.5 | 60.4±2.9 | 64.7±1.0 | 63.0±1.8 | 63.3±1.3 | 67.9±1.8 |
| | ST | 48.7±1.4 | 47.0±0.6 | 51.5±0.8 | 26.4±0.9 | 42.4±0.0 | 45.9±0.3 | 50.8±0.5 | 51.3±0.3 | 51.4±0.2 | 51.5±0.3 |
| | SIFT+ST | 57.5±1.8 | 58.8±0.6 | 65.3±2.2 | 27.6±0.9 | 47.9±2.4 | 53.2±2.8 | 62.3±0.7 | 64.1±2.1 | 66.1±1.3 | 71.4±2.0 |
| 7 | SIFT | 66.9±2.3 | 62.9±1.9 | 67.3±2.3 | 44.2±1.0 | 36.5±3.7 | 54.2±1.8 | 65.7±1.4 | 67.3±2.3 | 67.3±1.7 | 70.0±2.1 |
| | ST | 47.8±1.3 | 48.2±1.2 | 49.7±1.0 | 31.4±0.8 | 43.1±0.1 | 45.8±0.6 | 51.6±0.4 | 51.6±0.3 | 51.9±0.4 | 52.3±0.4 |
| | SIFT+ST | 57.9±1.5 | 60.5±1.2 | 67.7±2.4 | 36.0±0.9 | 50.8±2.8 | 53.2±2.9 | 68.6±1.8 | 66.9±2.4 | 70.3±1.5 | 73.0±2.4 |
| 10 | SIFT | 72.4±2.1 | 67.3±2.3 | 72.4±2.2 | 46.9±1.6 | 46.0±1.3 | 64.2±2.8 | 72.9±2.2 | 73.5±2.1 | 78.3±1.4 | 76.7±2.3 |
| | ST | 51.5±0.8 | 48.3±1.5 | 51.7±1.3 | 32.7±1.1 | 46.6±0.5 | 48.9±0.6 | 53.4±0.7 | 52.7±0.6 | 54.4±0.6 | 54.2±0.8 |
| | SIFT+ST | 65.0±0.9 | 67.2±1.7 | 72.4±2.3 | 41.0±1.1 | 54.7±2.0 | 56.5±2.3 | 69.6±1.0 | 75.1±1.9 | 80.9±1.7 | 79.0±1.5 |



Experiment on Cross-pose Face Recognition (跨姿态人脸识别)

COMPARISON WITH OTHER METHODS FOR FACE RECOGNITION ACROSS POSES

| Domains | | | Compared methods | | | | | | | Our method | |
|---------------|-----------|-----------|------------------|-----------|---------|----------|---------|--------------|--------------|------------|-------------|
| Tasks | Source | Target | NaïveComb | ASVM [12] | SGF [8] | GFK [10] | SA [41] | LTSI-PCA [1] | LTSI-LDA [1] | LSDT | NLSDT |
| Session 1 | frontal | 60° pose | 61.0 | 57.0 | 53.7 | 56.0 | 51.3 | 55.7 | 56.0 | 59.7 | 63.7 |
| Session 2 | frontal | 60° pose | 62.7 | 62.7 | 55.0 | 58.7 | 62.7 | 58.7 | 60.7 | 63.3 | 70.7 |
| Session 1+2 | frontal | 60° pose | 60.2 | 60.1 | 53.8 | 56.3 | 61.7 | 57.8 | 60.7 | 61.7 | 67.5 |
| Cross session | Session 1 | Session 2 | 93.6 | 94.3 | 92.5 | 96.7 | 98.3 | 96.7 | 96.7 | 95.8 | 99.4 |





Discriminative Kernel Transfer Learning (DKTL, InfoSci'17)

□ Idea:

The key idea behind is to realize robust transfer by simultaneously integrating discriminative subspace learning based on the proposed *domain-class-consistency* (DCC) metric, *kernel learning* in reproduced kernel Hilbert space, and *representation learning* between source domain and target domain via $l_{2,1}$ -norm minimization.

➤ Domain-class-consistency (DCC)----maximization:

Domain consistency: measure the *between-domain* distribution discrepancy;

Class consistency: measure the *within-domain* class separability;

➤ Domain-class-inconsistency (DCIC)----minimization:



➤ Subspace Transfer Reconstruction

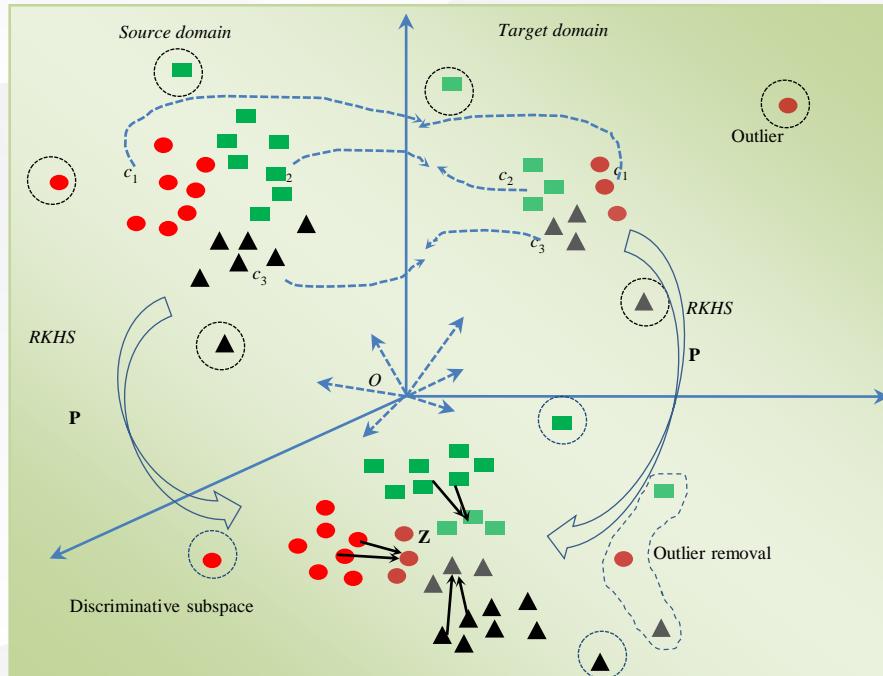
For domain adaptation, *source* data is used to reconstruct the *target* data

➤ Kernel mapping for handling nonlinear transfer

Reproduced Kernel Hilbert Space



DKTL (判别核迁移学习)



Schematic diagram of the proposed DKTL method

DKTL model:

$$\begin{aligned} & \min_{\mathbf{P}, \mathbf{Z}} E(\mathbf{X}_S, \mathbf{X}_T, \mathbf{P}, \mathbf{Z}) + \lambda \cdot \Omega(\mathbf{P}) + \tau \cdot R(\mathbf{Z}) \\ & \text{s.t. } \mathbf{P}^T \mathbf{P} = \mathbf{I}, \lambda, \tau > 0 \end{aligned}$$

where $E(\cdot)$ represents the *domain-inconsistency* term (i.e. cross domain representation or reconstruction error), $\Omega(\cdot)$ denotes the *class-inconsistency* term (i.e. discriminative regularizer) among multiple domains, $R(\cdot)$ represents the model regularization term of the representation coefficients with robust outlier removal



DKTL (判别核迁移学习)

DKTL model:

$$\begin{aligned} \min_{\mathbf{P}, \mathbf{Z}} \quad & E(\mathbf{X}_S, \mathbf{X}_T, \mathbf{P}, \mathbf{Z}) + \lambda \cdot \Omega(\mathbf{P}) + \tau \cdot R(\mathbf{Z}) \\ \text{s.t. } \quad & \mathbf{P}^T \mathbf{P} = \mathbf{I}, \lambda, \tau > 0 \end{aligned}$$

Suppose \mathbf{P} be represented by a linear combination of the transformed training samples $\varphi(\mathbf{X}) = [\varphi(\mathbf{X}_S), \varphi(\mathbf{X}_T)]$ via $\varphi(\cdot)$, as

$$\mathbf{P} = \varphi(\mathbf{X})\Phi$$

$$\begin{aligned} E(\mathbf{X}_S, \mathbf{X}_T, \mathbf{P}, \mathbf{Z}) &= \left\| \mathbf{P}^T \varphi(\mathbf{X}_T) - \mathbf{P}^T \varphi(\mathbf{X}_S) \mathbf{Z} \right\|_F^2 \\ &= \left\| \Phi^T \varphi(\mathbf{X})^T \varphi(\mathbf{X}_T) - \Phi^T \varphi(\mathbf{X})^T \varphi(\mathbf{X}_S) \mathbf{Z} \right\|_F^2 \end{aligned}$$

The second term $\Omega(\mathbf{P})$ pursues a discriminative subspace where the domain-class-inconsistency (DCIC) is minimized

2018/9/29

$$\begin{aligned} \Omega(\mathbf{P}) &= \sum_{c=1}^C \left\| \mathbf{P}^T \varphi(\mathbf{\mu}_S^c) - \mathbf{P}^T \varphi(\mathbf{\mu}_T^c) \right\|_2^2 - \sum_{t \in \{S, T\}} \sum_{c, k=1, c \neq k}^C \left\| \mathbf{P}^T \varphi(\mathbf{\mu}_t^c) - \mathbf{P}^T \varphi(\mathbf{\mu}_t^k) \right\|_2^2 \\ &= \sum_{c=1}^C \left\| \Phi^T \varphi(\mathbf{X})^T \varphi(\mathbf{\mu}_S^c) - \Phi^T \varphi(\mathbf{X})^T \varphi(\mathbf{\mu}_T^c) \right\|_2^2 - \\ &\quad \sum_{t \in \{S, T\}} \sum_{c, k=1, c \neq k}^C \left\| \Phi^T \varphi(\mathbf{X})^T \varphi(\mathbf{\mu}_t^c) - \Phi^T \varphi(\mathbf{X})^T \varphi(\mathbf{\mu}_t^k) \right\|_2^2 \end{aligned}$$

where

$$\varphi(\mathbf{\mu}_S^c) = \frac{1}{N_S^c} \sum_{i=1}^{N_S^c} \varphi(\mathbf{X}_{S,i}^c) \quad \text{and} \quad \varphi(\mathbf{\mu}_T^c) = \frac{1}{N_T^c} \sum_{i=1}^{N_T^c} \varphi(\mathbf{X}_{T,i}^c)$$

The third term $R(\mathbf{Z})$ in Eq.(1) is a robust sparse constraint on the transfer coefficients \mathbf{Z} for regularization. Generally, it can be formulated as follows

$$R(\mathbf{Z}) = \|\mathbf{Z}\|_{q,p}$$

where $\|\cdot\|_{q,p}$ represents $l_{q,p}$ -norm



DKTL (判别核迁移学习)

DKTL model:

$$\begin{aligned} \min_{\Phi, \mathbf{Z}} & \left\| \Phi^T \varphi(\mathbf{X})^T \varphi(\mathbf{X}_T) - \Phi^T \varphi(\mathbf{X})^T \varphi(\mathbf{X}_S) \mathbf{Z} \right\|_F^2 + \lambda \cdot \\ & \left(\sum_{c=1}^C \left\| \Phi^T \varphi(\mathbf{X})^T \varphi(\boldsymbol{\mu}_S^c) - \Phi^T \varphi(\mathbf{X})^T \varphi(\boldsymbol{\mu}_T^c) \right\|_2^2 - \right. \\ & \left. \sum_{t \in \{S, T\}} \sum_{c=1, c \neq k}^C \left\| \Phi^T \varphi(\mathbf{X})^T \varphi(\boldsymbol{\mu}_t^c) - \Phi^T \varphi(\mathbf{X})^T \varphi(\boldsymbol{\mu}_t^k) \right\|_2^2 \right) + \tau \cdot \|\mathbf{Z}\|_{2,1} \\ s.t. \quad & \Phi^T \varphi(\mathbf{X})^T \varphi(\mathbf{X}) \Phi = \mathbf{I}, \lambda, \tau > 0 \end{aligned}$$

$$\begin{aligned} \min_{\Phi, \mathbf{Z}} & \left\| \Phi^T \mathbf{K}_T - \Phi^T \mathbf{K}_S \mathbf{Z} \right\|_F^2 + \lambda \cdot \left(\frac{1}{C} \sum_{c=1}^C \left\| \Phi^T \mathbf{K}_{\mu, S}^c - \Phi^T \mathbf{K}_{\mu, T}^c \right\|_2^2 - \right. \\ & \left. \frac{2}{C(C-1)} \sum_{t \in \{S, T\}} \alpha_t \sum_{c, k=1, c \neq k}^C \left\| \Phi^T \mathbf{K}_{\mu, t}^c - \Phi^T \mathbf{K}_{\mu, t}^k \right\|_2^2 \right) + \tau \cdot \|\mathbf{Z}\|_{2,1} \\ s.t. \quad & \Phi^T \mathbf{K} \Phi = \mathbf{I}, \alpha_S + \alpha_T = 1, \lambda, \tau, \alpha_S, \alpha_T > 0 \end{aligned}$$

where there are two variables in the proposed DKTL model, and it is convex to each variable. Therefore, a **variable alternating optimization method is used**

Gaussian kernel function is used in this paper.

$$\kappa(\mathbf{x}, \mathbf{y}) = \exp\left(-\|\mathbf{x} - \mathbf{y}\|_2^2 / 2\sigma^2\right)$$

| | |
|--|--|
| $\mathbf{K} = \varphi(\mathbf{X})^T \varphi(\mathbf{X}) = \kappa(\mathbf{X}, \mathbf{X})$ | $\mathbf{K}_T = \varphi(\mathbf{X})^T \varphi(\mathbf{X}_T) = \kappa(\mathbf{X}, \mathbf{X}_T)$ |
| $\mathbf{K}_S = \varphi(\mathbf{X})^T \varphi(\mathbf{X}_S) = \kappa(\mathbf{X}, \mathbf{X}_S)$ | kernel Gram matrix |
| $\mathbf{K}_{\mu, T}^c = \varphi(\mathbf{X})^T \varphi(\boldsymbol{\mu}_T^c) = \kappa(\mathbf{X}, \boldsymbol{\mu}_T^c)$ | $\mathbf{K}_{\mu, S}^c = \varphi(\mathbf{X})^T \varphi(\boldsymbol{\mu}_S^c) = \kappa(\mathbf{X}, \boldsymbol{\mu}_S^c)$ |
| kernel mean vectors | |



Optimization

Update Φ :

By fixing the variable Z , the problem with respect to Φ then becomes

$$\begin{aligned} \min_{\Phi} & \left\| \Phi^T \mathbf{K}_T - \Phi^T \mathbf{K}_S Z \right\|_F^2 + \lambda \cdot \left(\frac{1}{C} \sum_{c=1}^C \left\| \Phi^T \mathbf{K}_{\mu,S}^c - \Phi^T \mathbf{K}_{\mu,T}^c \right\|_2^2 - \right. \\ & \left. \frac{2}{C(C-1)} \sum_{t \in \{S,T\}} \alpha_t \sum_{c,k=1, c \neq k}^C \left\| \Phi^T \mathbf{K}_{\mu,t}^c - \Phi^T \mathbf{K}_{\mu,t}^k \right\|_2^2 \right) \\ \text{s.t. } & \Phi^T \mathbf{K} \Phi = \mathbf{I}, \alpha_S + \alpha_T = 1, \lambda, \alpha_S, \alpha_T > 0 \end{aligned}$$



$$\begin{aligned} \min_{\Phi} & Tr(\Phi^T A \Phi) \\ \text{s.t. } & \Phi^T \mathbf{K} \Phi = \mathbf{I} \end{aligned}$$

where

$$\mathbf{A} = \mathbf{A}_1 + \lambda \cdot \mathbf{A}_2 - \lambda \cdot \mathbf{A}_3$$

$$\begin{aligned} \mathbf{A}_1 &= (\mathbf{K}_T - \mathbf{K}_S Z)(\mathbf{K}_T - \mathbf{K}_S Z)^T \\ \mathbf{A}_2 &= \frac{1}{C} \sum_{c=1}^C (\mathbf{K}_{\mu,S}^c - \mathbf{K}_{\mu,T}^c)(\mathbf{K}_{\mu,S}^c - \mathbf{K}_{\mu,T}^c)^T \\ \mathbf{A}_3 &= \frac{2}{C(C-1)} \sum_{t \in \{S,T\}} \alpha_t \sum_{c,k=1, c \neq k}^C (\mathbf{K}_{\mu,t}^c - \mathbf{K}_{\mu,t}^k)(\mathbf{K}_{\mu,t}^c - \mathbf{K}_{\mu,t}^k)^T \end{aligned}$$

Algorithm 1. Solving Φ

Input: kernel gram matrix and vectors , λ , d;

Procedure:

1. Initialize ;
2. Compute \mathbf{A}_1 , \mathbf{A}_2 and \mathbf{A}_3 , respectively;
3. Compute \mathbf{A} ;
4. Perform Eigen-value decomposition of (9);
5. Get Φ consisting of Eigen-vectors w.r.t. the d smallest Eigen-values;

Output: Φ



Optimization

Update Z:

By fixing Φ , the problem is transformed into the following problem

$$\min_{\mathbf{Z}} \left\| \Phi^T \mathbf{K}_T - \Phi^T \mathbf{K}_S \mathbf{Z} \right\|_F^2 + \tau \cdot \|\mathbf{Z}\|_{2,1}$$

where $\|\mathbf{Z}\|_{2,1} = \text{Tr}(\mathbf{Z}^T \Theta \mathbf{Z})$

where Θ is a diagonal matrix, whose the i -th diagonal element is calculated as

$$\Theta_{ii} = \frac{1}{2\|\mathbf{Z}_i\|_2}$$



$$\min_{\mathbf{Z}} \left\| \Phi^T \mathbf{K}_T - \Phi^T \mathbf{K}_S \mathbf{Z} \right\|_F^2 + \tau \cdot \text{Tr}(\mathbf{Z}^T \Theta \mathbf{Z})$$

It can be easily solved as

$$\mathbf{Z} = (\mathbf{K}_S^T \Phi \Phi^T \mathbf{K}_S + \tau \cdot \Theta)^{-1} \mathbf{K}_S^T \Phi \Phi^T \mathbf{K}_T$$

Algorithm 2. Solving Z

Input: kernel gram matrix and vectors , Φ ;

Procedure:

1. Initialize $\mathbf{Z} = \mathbf{K}_S^T \mathbf{K}_T$;
2. Compute Θ ;
3. Compute \mathbf{Z} ;

Output: Z

Algorithm 3. DDKL

Input: kernel gram matrix and vectors , λ, τ, d, T ;

Procedure:

1. Initialize $\mathbf{Z} = \mathbf{K}_S^T \mathbf{K}_T$ and $t=1$;
2. **While** not converge ($t < T$) **do**
3. Update Φ by calling **Algorithm 1**;
4. Update \mathbf{Z} by calling **Algorithm 2**;
5. **Until** Convergence;

Output: Z and Φ



Experiments

- ❑ Object Recognition Across Domains
- ❑ Face Recognition Across Poses and Expression
- ❑ Handwritten Digits Recognition Across Tasks



Object Recognition Across Domains

□ Results on 3DA data

| Tasks | ASVM [8] | GFK [19] | SGF [4] | RDALR [22] | SA [20] | LTSL [21] | DKTL |
|-----------------|----------|----------|----------|------------|----------|-----------|-----------------|
| Amazon → Webcam | 42.2±0.9 | 46.4±0.5 | 45.1±0.6 | 50.7±0.8 | 48.4±0.6 | 53.5±0.4 | 53.0±0.8 |
| DSLR → Webcam | 33.0±0.8 | 61.3±0.4 | 61.4±0.4 | 36.9±1.9 | 61.8±0.9 | 62.4±0.3 | 65.7±0.4 |
| Webcam → DSLR | 26.0±0.7 | 66.3±0.4 | 63.4±0.5 | 32.9±1.2 | 63.4±0.5 | 63.9±0.3 | 73.3±0.5 |

| Tasks | ASVM [8] | GFK [19] | SGF [4] | RDALR [22] | SA [20] | LTSL [21] | DKTL |
|--------------------|----------|----------|----------|------------|----------|-----------|-----------------|
| Amazon+DSLR→Webcam | 30.4±0.6 | 34.3±0.6 | 31.0±1.6 | 36.9±1.1 | 54.4±0.9 | 55.3±0.3 | 60.0±0.5 |
| Amazon+Webcam→DSLR | 25.3±1.1 | 52.0±0.8 | 25.0±0.4 | 31.2±1.3 | 37.5±1.0 | 57.7±0.4 | 63.7±0.7 |
| DSLR+Webcam→Amazon | 17.3±0.9 | 21.7±0.5 | 15.0±0.4 | 20.9±0.9 | 16.5±0.4 | 20.0±0.2 | 22.0±0.4 |



Object Recognition Across Domains

□ Results on 4DA data

| Method | A→D | C→D | A→C | W→C | D→C | D→A | W→A | C→A | C→W | A→W |
|------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| NaïveComb | 94.1±0.8 | 92.8±0.7 | 83.4±0.4 | 81.2±0.4 | 82.7±0.4 | 90.9±0.3 | 90.6±0.2 | 90.3±0.2 | 90.6±0.8 | 91.1±0.8 |
| SGF [4] | 92.0±1.3 | 92.4±1.1 | 77.4±0.7 | 76.8±0.7 | 78.2±0.7 | 88.0±0.5 | 86.8±0.7 | 89.3±0.4 | 87.8±0.8 | 88.1±0.8 |
| GFK [19] | 94.3±0.7 | 91.9±0.8 | 79.1±0.7 | 76.1±0.7 | 77.5±0.8 | 90.1±0.4 | 85.6±0.5 | 88.4±0.4 | 86.4±0.7 | 88.6±0.8 |
| SA [20] | 92.8±1.0 | 92.1±0.9 | 83.3±0.2 | 81.0±0.6 | 82.9±0.7 | 90.7±0.5 | 90.9±0.4 | 89.9±0.5 | 89.0±1.1 | 87.8±1.4 |
| LTS defense [21] | 94.5±0.5 | 93.5±0.8 | 85.4±0.1 | 82.6±0.3 | 84.8±0.2 | 91.9±0.2 | 91.0±0.2 | 90.9±0.1 | 90.8±0.7 | 91.5±0.5 |
| DKTL | 96.6±0.5 | 94.3±0.6 | 86.7±0.3 | 84.0±0.3 | 86.1±0.4 | 92.5±0.3 | 91.9±0.3 | 92.4±0.1 | 92.0±0.9 | 93.0±0.8 |



Object Recognition Across Domains

□ Results on 4DA data

Deep transfer models

- AlexNet, Krizhevsky et al., NIPS'12
- DAN, Long et al., ICML'15;
- RTN, Long et al. , NIPS '16;

Table 4

Comparisons with deep transfer learning methods on 4DA dataset.

| Method | A→D | C→D | A→C | W→C | D→C | D→A | W→A | C→A | C→W | A→W | Average |
|--------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| AlexNet [19] | 88.3 | 87.3 | 77.9 | 77.9 | 81.0 | 89.0 | 83.1 | 91.3 | 83.2 | 83.1 | 84.2 |
| DDC [38] | 89.0 | 88.8 | 85.0 | 78.0 | 81.1 | 89.5 | 84.9 | 91.9 | 85.4 | 86.1 | 86.0 |
| DAN [28] | 92.4 | 90.5 | 85.1 | 84.3 | 82.4 | 92.0 | 92.1 | 92.0 | 90.6 | 93.8 | 89.5 |
| RTN [27] | 94.6 | 92.9 | 88.5 | 88.4 | 84.3 | 95.5 | 93.1 | 94.4 | 96.6 | 97.0 | 92.5 |
| DKTL | 96.6 | 94.3 | 86.7 | 84.0 | 86.1 | 92.5 | 91.9 | 92.4 | 92.0 | 93.0 | 91.0 |

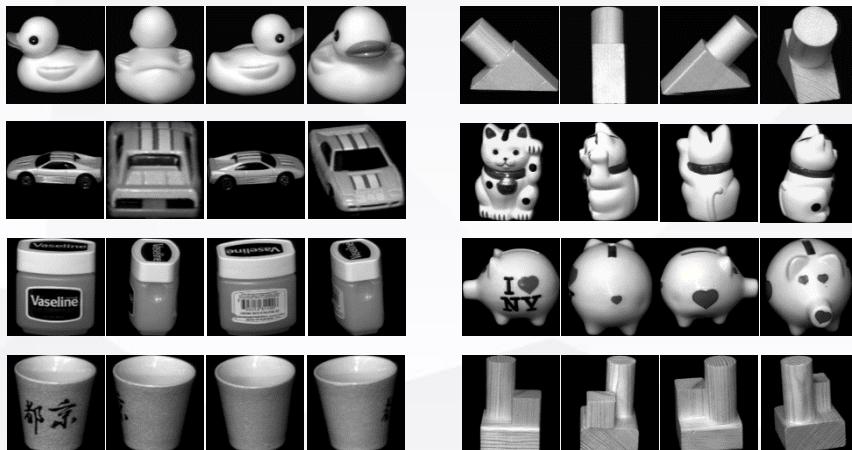


Object Recognition Across Domains

□ COIL-20 data: *Columbia Object Image Library (Nene et al.)*

The COIL-20 dataset contains 1440 gray scale images of 20 objects (72 images with different poses per object). Each image has 128×128 pixels with 256 gray levels per pixel. For experiments, the size of each image is adjusted as 32×32 .

The dataset is partitioned into four subsets, i.e. COIL 1, COIL 2, COIL 3 and COIL 4 according to the directions. $[0^\circ, 85^\circ]$, $[180^\circ, 265^\circ]$, $[90^\circ, 175^\circ]$, $[270^\circ, 355^\circ]$. 360 samples are included for each domain.



Several objects from COIL-20 data



Object Recognition Across Domains

□ Results on COIL-20 data (12 settings)

| Tasks | ASVM [8] | GFK [19] | SGF [4] | SA [20] | LTSI (IJCV'16) | DKTL |
|-----------------|----------|----------|---------|---------|----------------|-------------|
| COIL 1 → COIL 2 | 79.7 | 81.1 | 78.9 | 81.1 | 79.7 | 83.8 |
| COIL 1 → COIL 3 | 76.8 | 80.1 | 76.7 | 75.3 | 79.2 | 79.7 |
| COIL 1 → COIL 4 | 81.4 | 80.0 | 74.7 | 76.7 | 81.4 | 80.0 |
| COIL 2 → COIL 1 | 78.3 | 80.0 | 79.2 | 81.1 | 76.4 | 81.1 |
| COIL 2 → COIL 3 | 84.3 | 85.0 | 79.7 | 81.9 | 86.4 | 85.6 |
| COIL 2 → COIL 4 | 77.2 | 78.9 | 74.4 | 78.3 | 77.2 | 79.7 |
| COIL 3 → COIL 1 | 76.4 | 79.7 | 71.1 | 78.9 | 76.4 | 80.8 |
| COIL 3 → COIL 2 | 79.6 | 83.0 | 81.1 | 80.3 | 79.7 | 82.8 |
| COIL 3 → COIL 4 | 74.2 | 73.3 | 73.3 | 76.1 | 74.2 | 75.8 |
| COIL 4 → COIL 1 | 81.9 | 81.1 | 72.5 | 79.4 | 81.9 | 81.7 |
| COIL 4 → COIL 2 | 77.5 | 79.2 | 71.1 | 72.8 | 77.8 | 78.6 |
| COIL 4 → COIL 3 | 74.8 | 75.6 | 76.7 | 78.3 | 74.7 | 79.2 |



Face Recognition Across Poses and Expression

□ Results on CMU Multi-PIE face data

| Cross domain tasks | NaïveComb | ASVM [8] | SGF [4] | GFK [19] | SA [20] | LTSI [21] | DKTL |
|--------------------------------------|-----------|----------|---------|----------|---------|-----------|-------------|
| Session 1: Frontal → 60° pose | 52.0 | 52.0 | 53.7 | 56.0 | 51.3 | 61.0 | 66.0 |
| Session 2: Frontal → 60° pose | 55.0 | 56.7 | 55.0 | 58.7 | 62.7 | 62.7 | 71.0 |
| Session 1+2: Frontal → 60° pose | 54.5 | 55.1 | 53.8 | 56.3 | 61.7 | 60.2 | 69.5 |
| Cross session: Session 1 → Session 2 | 93.6 | 97.2 | 92.5 | 96.7 | 98.3 | 97.2 | 99.4 |



Handwritten Digits Recognition Across Tasks

□ Results across datasets

| Cross domain tasks | NaïveComb | A-SVM [8] | SGF [4] | GFK [19] | SA [20] | LTSI [21] | DKTL |
|--------------------|-----------|-----------|----------|----------|----------|-----------|-----------------|
| MINIST → USPS | 78.8±0.5 | 78.3±0.6 | 79.2±0.9 | 82.6±0.8 | 78.8±0.8 | 78.4±0.7 | 88.0±0.4 |
| SEMEION → USPS | 83.6±0.3 | 76.8±0.4 | 77.5±0.9 | 82.7±0.6 | 82.5±0.5 | 83.4±0.3 | 85.8±0.4 |
| MINIST → SEMEION | 51.9±0.8 | 70.5±0.7 | 51.6±0.7 | 70.5±0.8 | 74.4±0.6 | 50.6±0.4 | 74.9±0.4 |
| USPS → SEMEION | 65.3±1.0 | 74.5±0.6 | 70.9±0.8 | 76.7±0.3 | 74.6±0.6 | 64.5±0.7 | 81.6±0.4 |
| USPS → MINIST | 71.7±1.0 | 73.2±0.8 | 71.1±0.7 | 74.9±0.9 | 72.9±0.7 | 71.2±1.0 | 79.0±0.6 |
| SEMEION → MINIST | 67.6±1.2 | 69.3±0.7 | 66.9±0.6 | 74.5±0.6 | 72.9±0.7 | 66.8±1.2 | 77.3±0.7 |



Class-specific Reconstruction Transfer (CRTL, ICCV W'17)

- Class imbalance induced class-specific Reconstruction (类不平衡，类特定重构)
- Projected Hilbert-Schmidt Independence Criterion (pHSIC独立性)
- Low-rank and sparse constraint for global and local preservation

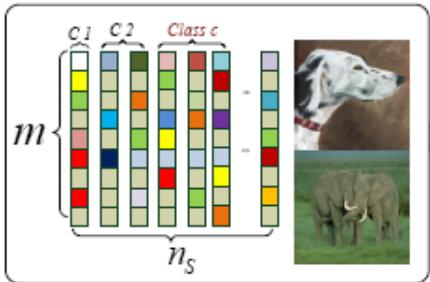
[**HSIC**]: A. Gretton, et al. Measuring statistical dependence with Hilbert-Schmidt norms. ALT, 2005

[**HSIClasso**]: High-dimensional feature selection by Feature-Wise Kernelized Lasso. Neural Computation, 2014.

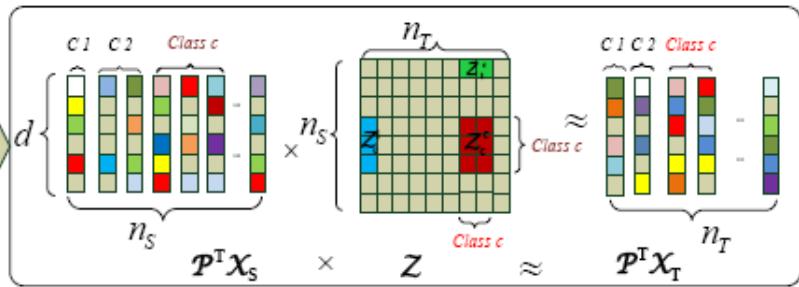


CRTL (类特定重建迁移学习)

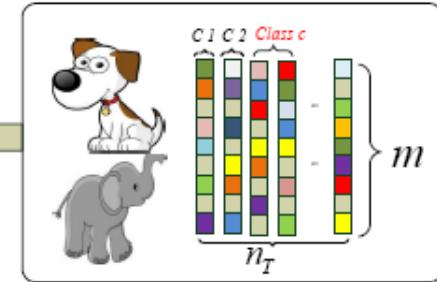
Source Domain \mathcal{X}_S



Common Subspace



Target Domain \mathcal{X}_T



$$\begin{aligned}
 & \min_{\mathcal{P}, \mathcal{Z}} \sum_{c=1}^C (\beta_1 \| \mathcal{P}^T \varphi(\mathbf{X}_T^c) - \mathcal{P}^T \varphi(\mathcal{X}_S^c) \mathcal{Z}_c^c \|_F^2 \\
 & + \sum_{c=1}^C \sum_{k=1, k \neq c}^C \beta_2 \| \mathcal{P}^T \varphi(\mathcal{X}_S^c) \mathcal{Z}_k^c \|_F^2 \\
 & - \frac{1}{(N-1)^2} \text{Tr}(k(\mathcal{P}^T \varphi(\mathcal{X}), \mathcal{P}^T \varphi(\mathcal{X})) \mathbf{H} \mathcal{L} \mathbf{H}) \\
 & + \lambda_1 \| \mathcal{Z} \|_* + \lambda_2 \| \mathcal{Z} \|_1
 \end{aligned}$$

$$s.t. \mathcal{P}^T \mathcal{P} = \mathbf{I}, \mathcal{X} = [\mathcal{X}_S, \mathcal{X}_T], \mathbf{1}^{1 \times n_S} \mathcal{Z} = \mathbf{1}^{1 \times n_T}$$

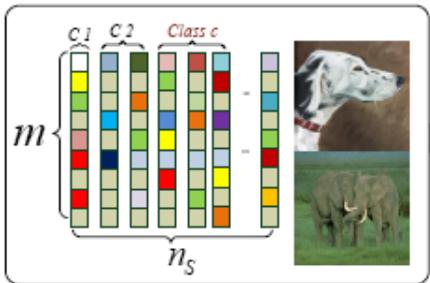
(5)

$$(\mathcal{P}^*)^T = \Phi^T \varphi(\mathcal{X})^T$$

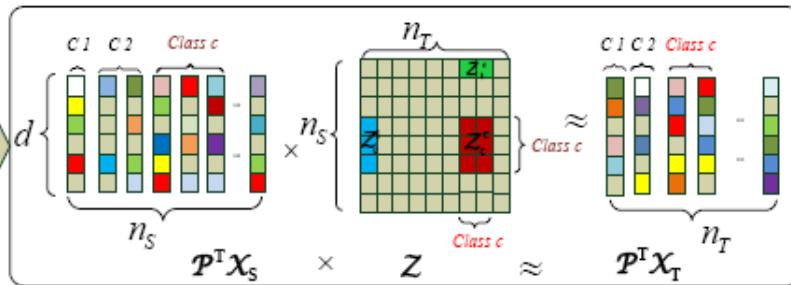


CRTL (类特定重建迁移学习)

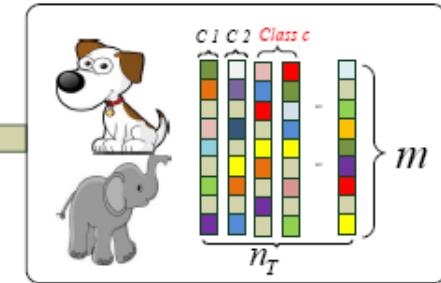
Source Domain \mathcal{X}_S



Common Subspace



Target Domain \mathcal{X}_T



$$\min_{\Phi, \mathcal{Z}} \sum_{c=1}^C (\beta_1 \| \Phi^T \mathcal{K}_T^c - \Phi^T \mathcal{K}_S^c \mathcal{Z}_c \|_F^2)$$

$$+ \sum_{c=1}^C \sum_{k=1, k \neq c}^C \beta_2 \| \Phi^T \mathcal{K}_S^c \mathcal{Z}_k \|_F^2$$

$$- \frac{1}{(N-1)^2} \text{Tr}(\Phi^T \mathcal{K} \mathbf{H} \mathcal{L} \mathcal{H} \mathbf{H} \mathcal{K} \Phi)$$

$$+ \lambda_1 \| \mathcal{Z} \|_* + \lambda_2 \| \mathcal{Z} \|_1$$

$$s.t. \Phi^T \mathcal{K} \Phi = \mathbf{I}, \mathbf{1}^{1 \times n_S} \mathcal{Z} = \mathbf{1}^{1 \times n_T}$$

(6)

ALM and Gradient descent can be used for OPTIMIZATION



□ Experiments





☐ Experiments

| Tasks | SourceOnly | | Naive Comb | | SGF [10] | | GFK [9] | | SA [7] | | LTSL [27] | | LSDT [37] | | CRTL | |
|-------------------|------------|-------|------------|-------|----------|-------|---------|-------|--------|-------|-----------|-------|-------------|-------------|-------------|-------------|
| | f_6 | f_7 | f_6 | f_7 | f_6 | f_7 | f_6 | f_7 | f_6 | f_7 | f_6 | f_7 | f_6 | f_7 | f_6 | f_7 |
| $A \rightarrow D$ | 80.8 | 81.3 | 94.5 | 94.1 | 90.5 | 92.0 | 92.6 | 94.3 | 94.2 | 92.8 | 95.5 | 94.5 | 96.4 | 96.0 | 96.4 | 95.8 |
| $C \rightarrow D$ | 76.6 | 77.6 | 92.9 | 92.8 | 93.1 | 92.4 | 92.0 | 91.9 | 93.0 | 92.1 | 93.6 | 93.5 | 95.4 | 94.6 | 95.2 | 94.8 |
| $W \rightarrow D$ | 96.1 | 96.2 | 99.1 | 98.9 | 97.7 | 97.6 | 97.8 | 98.5 | 98.6 | 98.5 | 99.1 | 98.8 | 99.4 | 99.3 | 99.4 | 99.3 |
| $A \rightarrow C$ | 79.3 | 79.3 | 84.0 | 83.4 | 77.1 | 77.4 | 78.9 | 79.1 | 83.1 | 83.3 | 85.3 | 85.4 | 85.9 | 87.0 | 86.2 | 87.0 |
| $W \rightarrow C$ | 59.5 | 68.1 | 81.7 | 81.2 | 74.1 | 76.8 | 77.5 | 76.1 | 81.1 | 81.0 | 82.3 | 82.6 | 83.1 | 84.2 | 83.6 | 84.9 |
| $D \rightarrow C$ | 67.3 | 74.3 | 83.0 | 82.7 | 75.9 | 78.2 | 78.8 | 77.5 | 82.4 | 82.9 | 84.4 | 84.8 | 85.2 | 86.2 | 85.5 | 86.4 |
| $D \rightarrow A$ | 77.0 | 81.8 | 90.5 | 90.9 | 88.0 | 88.0 | 88.9 | 90.1 | 90.4 | 90.7 | 91.1 | 91.9 | 92.2 | 92.5 | 92.5 | 92.7 |
| $W \rightarrow A$ | 66.8 | 73.4 | 90.1 | 90.6 | 87.2 | 86.8 | 86.2 | 85.6 | 89.8 | 90.9 | 90.6 | 91.0 | 91.0 | 91.7 | 91.3 | 92.2 |
| $C \rightarrow A$ | 85.8 | 86.5 | 89.9 | 90.3 | 88.5 | 89.3 | 87.5 | 88.4 | 89.5 | 89.9 | 90.4 | 90.9 | 92.1 | 92.5 | 92.0 | 92.5 |
| $C \rightarrow W$ | 67.5 | 67.8 | 91.6 | 90.6 | 89.4 | 87.8 | 87.7 | 86.4 | 91.2 | 89.0 | 91.8 | 90.8 | 93.3 | 93.5 | 92.7 | 93.1 |
| $D \rightarrow W$ | 95.4 | 95.1 | 97.9 | 98.0 | 96.8 | 95.7 | 97.0 | 96.5 | 97.5 | 97.5 | 98.2 | 97.8 | 98.7 | 98.3 | 98.7 | 98.5 |
| $A \rightarrow W$ | 70.5 | 71.6 | 90.4 | 91.1 | 87.2 | 88.1 | 89.5 | 88.6 | 90.3 | 87.8 | 92.2 | 91.5 | 92.1 | 92.9 | 92.3 | 93.0 |
| Average | 76.9 | 79.4 | 90.5 | 90.4 | 87.1 | 87.5 | 87.9 | 87.8 | 90.1 | 89.7 | 91.2 | 91.1 | 92.1 | 92.4 | 92.2 | 92.5 |

Table 1: Recognition accuracy (%) of different domain adaptation over 10 object categories on 4DA-CNN with deep feature representation



□ Experiments

| Tasks | SVM | TSL | RDALR [15] | LTS defense [27] | DTSL [31] | LSDT [37] | CRTL |
|---------------------|------|------|------------|------------------|-----------|-----------|-------------|
| $C1 \rightarrow C2$ | 82.7 | 80.0 | 80.7 | 75.4 | 84.6 | 81.7 | 87.0 |
| $C2 \rightarrow C1$ | 84.0 | 75.6 | 78.8 | 72.2 | 84.2 | 81.5 | 86.5 |
| Average | 83.3 | 77.8 | 79.7 | 73.8 | 84.4 | 81.6 | 86.8 |

Table 2: Recognition accuracy (%) of different domain adaptation on COIL-20

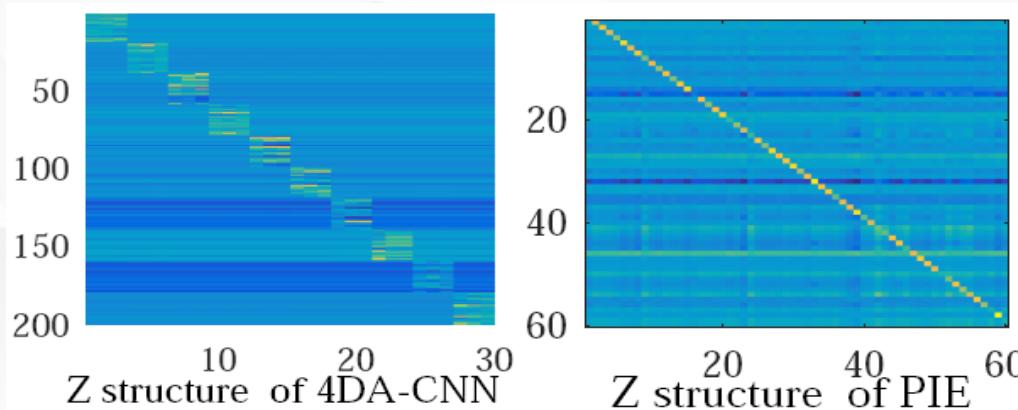
| Tasks | Naive Comb | A-SVM | SGF [10] | GFK [9] | SA [7] | LTS defense [27] | LSDT [37] | CRTL |
|--|------------|-------|----------|---------|--------|------------------|-----------|-------------|
| $S1 (0^\circ \rightarrow 60^\circ)$ | 61.0 | 57.0 | 53.7 | 56.0 | 51.3 | 56.0 | 59.7 | 65.7 |
| $S2 (0^\circ \rightarrow 60^\circ)$ | 62.7 | 62.7 | 55.0 | 58.7 | 62.7 | 60.7 | 63.3 | 69.0 |
| $S1+S2 (0^\circ \rightarrow 60^\circ)$ | 60.2 | 60.1 | 53.8 | 56.3 | 61.7 | 60.7 | 61.7 | 68.5 |
| $S1 \rightarrow S2$ | 93.6 | 94.3 | 92.5 | 96.7 | 98.3 | 96.7 | 95.8 | 98.7 |
| Average | 69.4 | 68.5 | 63.8 | 67.0 | 68.5 | 68.5 | 70.1 | 75.5 |

Table 3: Recognition accuracy (%) of different domain adaptation on face recognition across poses



☐ Experiments

| Tasks | Naive Comb | A-SVM | SGF [10] | GFK [9] | SA [7] | LTSI [27] | LSDT [37] | CRTL |
|-------------------|------------|-------|----------|---------|--------|-----------|-----------|-------------|
| $M \rightarrow U$ | 78.8 | 78.3 | 79.2 | 82.6 | 78.8 | 83.2 | 79.3 | 85.4 |
| $S \rightarrow U$ | 83.6 | 76.8 | 77.5 | 82.7 | 82.5 | 83.6 | 84.7 | 86.2 |
| $M \rightarrow S$ | 51.9 | 70.5 | 51.6 | 70.5 | 74.4 | 72.8 | 69.1 | 76.2 |
| $U \rightarrow S$ | 65.3 | 74.5 | 70.9 | 76.7 | 74.6 | 65.3 | 67.4 | 82.6 |
| $U \rightarrow M$ | 71.7 | 73.2 | 71.1 | 74.9 | 72.9 | 71.7 | 70.5 | 82.0 |
| $S \rightarrow M$ | 67.6 | 69.3 | 66.9 | 74.5 | 72.9 | 67.6 | 70.0 | 78.4 |
| Average | 69.8 | 73.8 | 69.5 | 77.0 | 76.0 | 74.0 | 73.5 | 81.8 |





Manifold Criterion Guided Transfer Learning (MCTL, TNNLS'18)

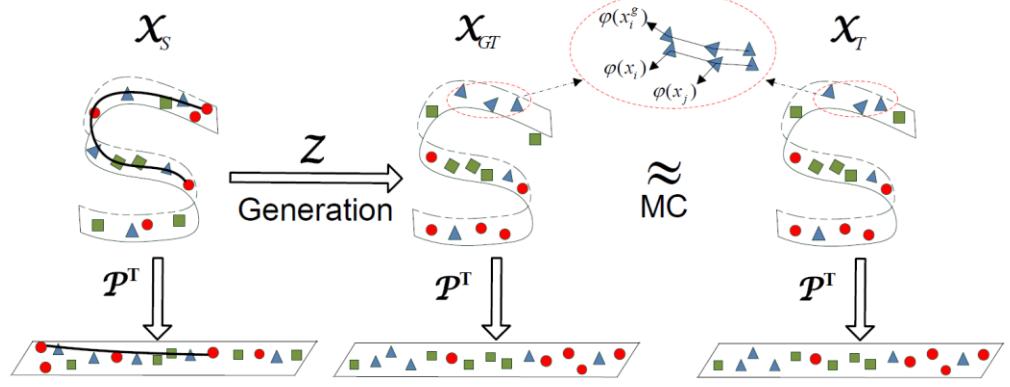
- A new manifold criterion for measuring domain match is proposed.
- Intermediate domain generation idea is proposed.

*Bridging the GAP between Transfer Learning and Semi-supervised Learning!!
Three Assumptions: Smooth, Cluster, Manifold*

Def. When manifold criterion is satisfied, domain distribution is matched.



Manifold Criterion Guided Transfer Learning (MCTL, TNNLS'18)



(a) Source domain (b) Intermediate domain

(c) Target domain

$$\varphi(\mathbf{X}_{GT}) = \varphi(\mathbf{X}_S)\mathcal{Z}$$

Local Generative Discrepancy Metric:

$$\begin{aligned} LGDM(D_{GT}, D_T) &= \sum_{p,q}^{n_T} W_{pq} \|\varphi(x_{GT}^p) - \varphi(x_T^q)\|_2^2 \\ &= Tr(\varphi(\mathbf{X}_{GT})\mathbf{D}(\varphi(\mathbf{X}_{GT})^T) \\ &\quad + Tr(\varphi(\mathbf{X}_T)\mathbf{D}(\varphi(\mathbf{X}_T)^T) \\ &\quad - 2Tr(\varphi(\mathbf{X}_{GT})\mathbf{W}(\varphi(\mathbf{X}_T)^T) \end{aligned}$$

Global Generative Discrepancy Metric:

$$\begin{aligned} GGDM(D_{GT}, D_T) &= \frac{1}{n_T} \sum_{i=1}^{n_T} \left\| \mathcal{P}^T (\varphi(\mathbf{X}_{GT}^i) - \varphi(\mathbf{X}_T^i)) \right\|_2^2 \\ &= \frac{1}{n_T} \left\| \mathcal{P}^T (\varphi(\mathbf{X}_S)\mathcal{Z} - \varphi(\mathbf{X}_T))\mathbf{1} \right\|_2^2 \end{aligned} \quad (4)$$

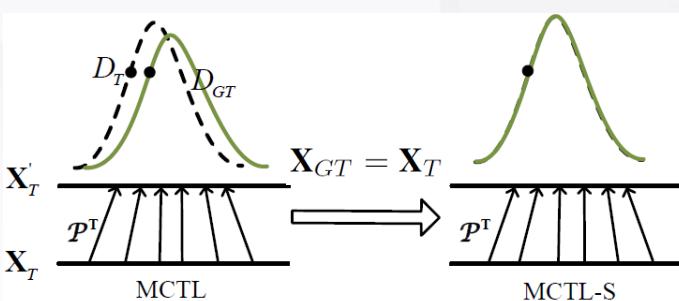
Let $\mathcal{P}^T = \Phi^T \varphi(\mathbf{X})^T$



Manifold Criterion Guided Transfer Learning (MCTL, TNNLS'18)

Derived MCTL model:

$$\begin{aligned} \min_{\Phi, \mathcal{Z}} & \frac{1}{(n_T)^2} \text{Tr}(\Phi^T \mathbf{K}_S \mathcal{Z} \mathbf{D} (\Phi^T \mathbf{K}_S \mathcal{Z})^T) \\ & + \frac{1}{(n_T)^2} \text{Tr}(\Phi^T \mathbf{K}_T \mathbf{D} (\Phi^T \mathbf{K}_T)^T) \\ & - \frac{2}{(n_T)^2} \text{Tr}(\Phi^T \mathbf{K}_S \mathcal{Z} \mathbf{W} (\Phi^T \mathbf{K}_T)^T) \\ & + \tau \frac{1}{n_T} \left\| \Phi^T (\mathbf{K}_S \mathcal{Z} - \mathbf{K}_T) \mathbf{1} \right\|_2^2 \\ & + \lambda_1 \|\mathcal{Z}\|_* \\ \text{s.t. } & \Phi^T \mathbf{K} \Phi = \mathbf{I} \end{aligned}$$



Simplified MCTL-s model:

$$\begin{aligned} \min_{\Phi, \mathcal{Z}} & \frac{2}{(n_T)^2} \text{Tr}(\Phi^T \mathbf{K}_S \mathcal{Z} \mathbf{L} (\Phi^T \mathbf{K}_S \mathcal{Z})^T) \\ & + \tau \left\| \frac{1}{n_T} \Phi^T (\mathbf{K}_S \mathcal{Z} - \mathbf{K}_T) \mathbf{1} \right\|_2^2 \\ & + \lambda_1 \|\mathcal{Z}\|_* \end{aligned}$$



Results

Face recognition on PIE across poses

| Tasks | Naive Comb | A-SVM | SGF [35] | GFK [34] | SA [60] | LTSI [40] | LSDT [41] | MCTL |
|--|------------|-------|----------|----------|---------|-----------|-----------|-------------|
| $S1 (0^\circ \rightarrow 60^\circ)$ | 61.0 | 57.0 | 53.7 | 61.0 | 51.3 | 56.0 | 59.7 | 65.3 |
| $S2 (0^\circ \rightarrow 60^\circ)$ | 62.7 | 62.7 | 55.0 | 58.7 | 62.7 | 62.7 | 63.3 | 70.0 |
| $S1 + S2 (0^\circ \rightarrow 60^\circ)$ | 60.2 | 60.1 | 53.8 | 56.3 | 61.7 | 60.2 | 61.7 | 68.3 |
| $S1 \rightarrow S2$ | 93.6 | 94.3 | 92.5 | 96.7 | 98.3 | 97.2 | 95.8 | 98.7 |
| <i>Average</i> | 69.4 | 68.5 | 63.8 | 67.0 | 68.5 | 70.3 | 70.1 | 75.6 |

Handwritten digits recognition on MNIST, USPS and SEMEION

| Tasks | Naive Comb | A-SVM | SGF [35] | GFK [34] | SA [60] | LTSI [40] | LSDT [41] | MCTL |
|-------------------|------------|-------|----------|----------|-------------|-----------|-----------|-------------|
| $M \rightarrow U$ | 78.8 | 78.3 | 79.2 | 82.6 | 78.8 | 83.2 | 79.3 | 87.8 |
| $S \rightarrow U$ | 83.6 | 76.8 | 77.5 | 82.7 | 82.5 | 83.6 | 84.7 | 84.8 |
| $M \rightarrow S$ | 51.9 | 70.5 | 51.6 | 70.5 | 74.4 | 72.8 | 69.1 | 74.0 |
| $U \rightarrow S$ | 65.3 | 74.5 | 70.9 | 76.7 | 74.6 | 65.3 | 67.4 | 83.0 |
| $U \rightarrow M$ | 71.7 | 73.2 | 71.1 | 74.9 | 72.9 | 71.7 | 70.5 | 81.2 |
| $S \rightarrow M$ | 67.6 | 69.3 | 66.9 | 74.5 | 72.9 | 67.6 | 70.0 | 74.0 |
| <i>Average</i> | 69.8 | 73.8 | 69.5 | 77.0 | 76.0 | 74.0 | 73.5 | 80.8 |



Results

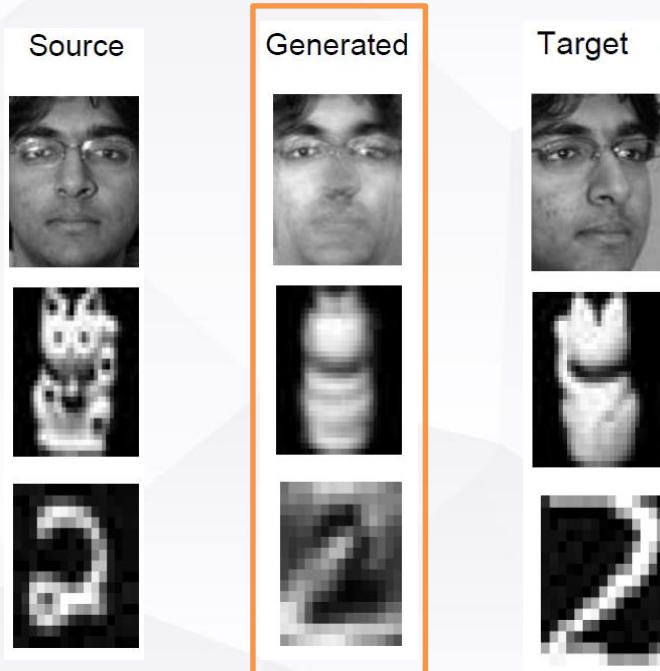




Table of Contents

Part I: Classifier-level Domain Adaptation

- [1] L. Zhang and D. Zhang, IEEE Trans. Image Processing, 2016.
- [2] L. Zhang and D. Zhang, IEEE Trans. Multimedia, 2016.

Part II: Feature-level Transfer Learning

- [3] L. Zhang, W. Zuo, and D. Zhang, IEEE Trans. Image Processing, 2016.
- [4] L. Zhang, J. Yang, and D. Zhang, Information Sciences, 2017.
- [5] S. Wang, L. Zhang, W. Zuo, ICCV W 2017.
- [6] L. Zhang, Y. Liu and P. Deng, IEEE Trans. Intru. Meas. 2017.
- [7] L. Zhang, S. Wang, G.B. Huang, W. Zuo, J. Yang, and D. Zhang, IEEE Trans. Neural Networks and Learning Systems, 2018.

Part III: Self-Adversarial Transfer Learning

- [8] Q. Duan, L. Zhang, W. Zuo, ACM MM, 2017.
- [9] L. Zhang, Q. Duan, W. Jia, D. Zhang, X. Wang, IEEE Trans. Cybernetics, 2018. in review

Part IV: Guide Learning (A try for TL/DA)

- [10] J. Fu, L. Zhang, B. Zhang, W. Jia, CCBR oral, 2018.
- [11] L. Zhang, J. Fu, S. Wang, D. Zhang, D.Y. Dong, C.L. Philip Chen, IEEE Trans. Neural Net. Learn. Syst. 2018. in review.



AdvNet (ACM MM'17):

Family and Kinship recognition
家庭和亲属关系识别



Sample family photos from F1W for 27 of the 1,000.



AdvNet:

- For over 1 million data, deep transfer learning is prior considered;
- MMD based Self-Adversarial (自我对抗) strategy is considered for discriminative feature adaptation;
- Residual net with Contrastive loss is used.



F-D

F-S

M-D

M-S

SIBS

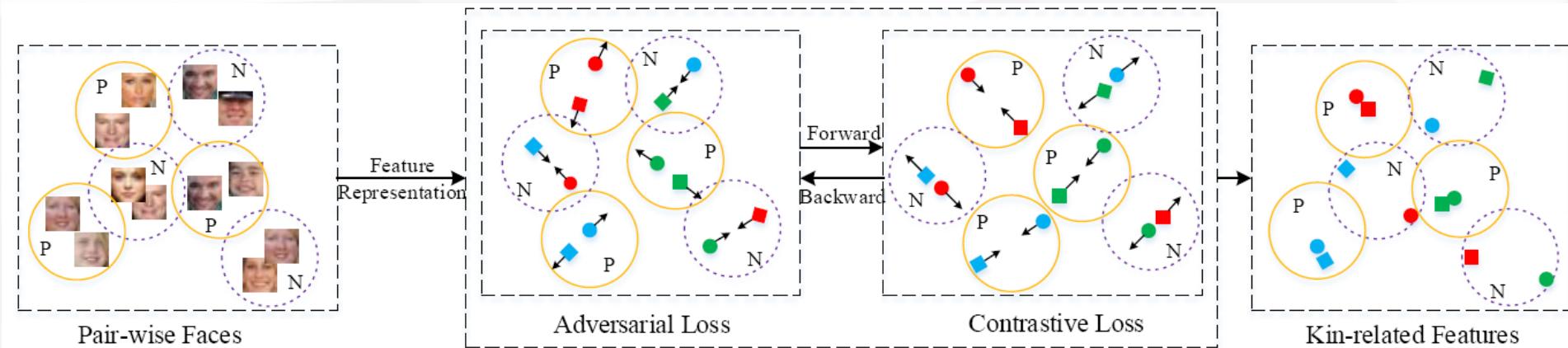
B-B

S-S

Challenge Competition on 7 Kinships



AdvNet:

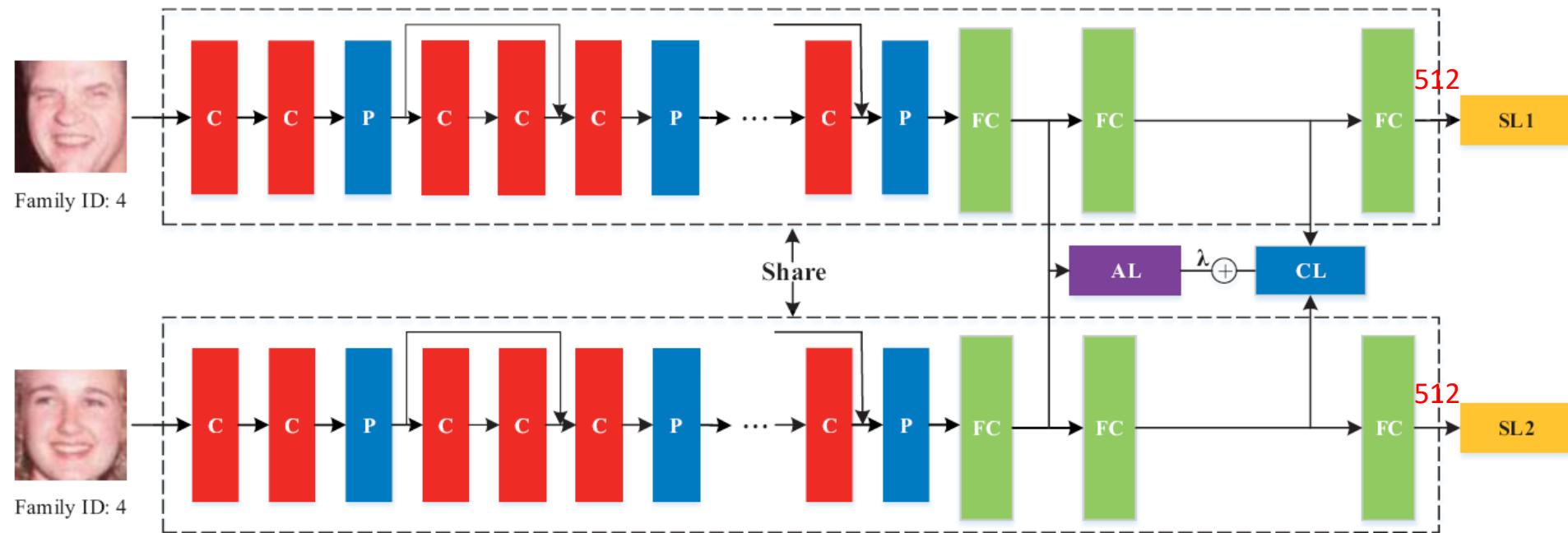


Learning discriminative kin-related features by adversarial loss and contrastive loss
通过模型自我对抗，实现有效特征学习

[8] Q. Duan and L. Zhang, "AdvNet: Adversarial Contrastive Residual Net for 1 Million Kinship Recognition," ACM MM, 2017



AdvNet:





Our proposed AdvNet (深度对抗网络)

$$L_C = \frac{1}{2N} \sum_{n=1}^N (\delta(y_n^1 = y_n^2)d^2 - \delta(y_n^1 \neq y_n^2) \max(\text{margin}-d, 0)^2) \quad (2)$$

where $\delta(\text{condition}) = 1$ if the condition is satisfied. y_n^1 and y_n^2 are the family IDs of x_n^1 and x_n^2 , respectively.

Family ID guided
Contrastive Loss

$$L = L_C + \lambda L_A + L_{S1} + L_{S2}$$

$$\begin{aligned} L_A = & -\frac{1}{2N} \sum_{n=1}^N (\delta(y_n^1 = y_n^2) \|\phi(\mathbf{x}_n^1) - \phi(\mathbf{x}_n^2)\|_\hbar^2 \\ & - \delta(y_n^1 \neq y_n^2) \|\phi(\mathbf{x}_n^1) - \phi(\mathbf{x}_n^2)\|_\hbar^2) \end{aligned} \quad (5)$$

MMD guided
Adversarial Loss



Experiments

- **Dataset:** Families in the Wild (FIW)
- **Size:** 12000 family photos of 1001 families
- **Input:** 644,000 pairs of 7 kinship relations
- The dataset is partitioned into 3 disjoint sets:
Train, Validation, Test (Test is blind)



Experiments

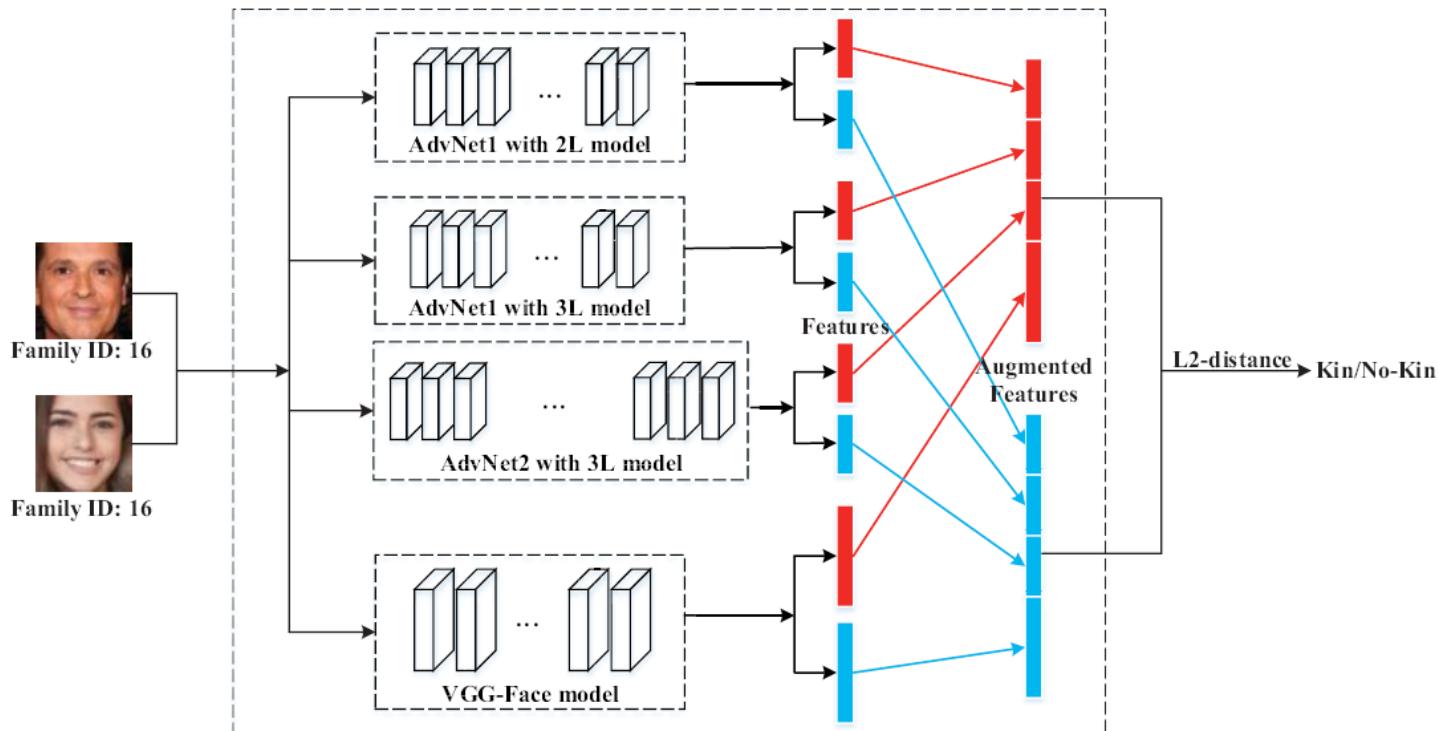
Table 3: Accuracy of AdvNet with different loss

| Loss | M-D | M-S | S-S | B-B | SIBS | F-S | F-D | Mean |
|------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| CL | 61.06 | 61.95 | 62.45 | 65.35 | 62.05 | 61.33 | 59.18 | 61.91 |
| 2L | 60.50 | 64.07 | 64.17 | 63.76 | 61.99 | 62.23 | 60.53 | 62.46 |
| 3L | 64.11 | 65.65 | 64.53 | 65.80 | 64.82 | 63.42 | 63.18 | 64.50 |

Performance is still not good?!



Feature Augmentation (Network Fusion: AdvNets+VGG-Face Net)





Feature Augmentation (Network Fusion: AdvNets+VGG-Face Net)

Table 5: Accuracy of different model, loss and feature augmentation

| Index | Loss | Model | M-D | M-S | S-S | B-B | SIBS | F-S | F-D | Mean |
|---------|-------|----------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| 1 | 2L | AdvNet1 | 60.50 | 64.07 | 64.17 | 63.76 | 61.99 | 62.23 | 60.53 | 62.46 |
| 2 | 3L | AdvNet1 | 64.11 | 65.65 | 64.53 | 65.80 | 64.82 | 63.42 | 63.18 | 64.50 |
| 3 | 3L | AdvNet2 | 63.56 | 66.80 | 65.48 | 65.77 | 65.35 | 64.14 | 63.59 | 64.97 |
| 4 | SL | VGG-Face | 65.99 | 58.88 | 74.59 | 71.99 | 64.69 | 64.71 | 62.87 | 66.25 |
| 1+2+3 | Joint | Feature Augmentation | 64.20 | 67.55 | 65.71 | 66.82 | 66.45 | 64.78 | 64.04 | 65.65 |
| 2+3+4 | Joint | Feature Augmentation | 70.07 | 65.60 | 77.52 | <u>71.88</u> | <u>69.72</u> | 68.79 | <u>67.56</u> | <u>70.16</u> |
| 1+2+3+4 | Joint | Feature Augmentation | <u>69.93</u> | <u>67.33</u> | <u>77.44</u> | 71.76 | 69.80 | <u>68.77</u> | 67.82 | 70.41 |



Feature Augmentation (Network Fusion: AdvNets+VGG-Face Net)

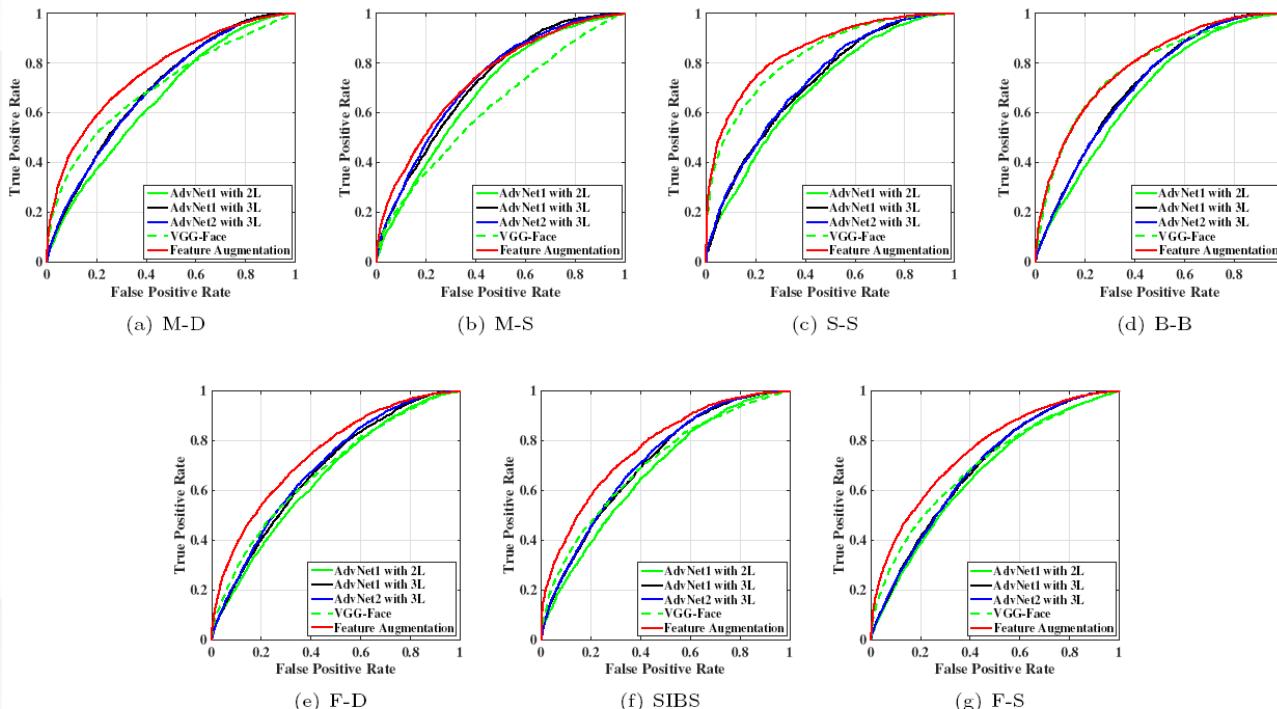




Table of Contents

Part I: Classifier-level Domain Adaptation

- [1] L. Zhang and D. Zhang, IEEE Trans. Image Processing, 2016.
- [2] L. Zhang and D. Zhang, IEEE Trans. Multimedia, 2016.

Part II: Feature-level Transfer Learning

- [3] L. Zhang, W. Zuo, and D. Zhang, IEEE Trans. Image Processing, 2016.
- [4] L. Zhang, J. Yang, and D. Zhang, Information Sciences, 2017.
- [5] S. Wang, L. Zhang, W. Zuo, ICCV W 2017.
- [6] L. Zhang, Y. Liu and P. Deng, IEEE Trans. Intru. Meas. 2017.
- [7] L. Zhang, S. Wang, G.B. Huang, W. Zuo, J. Yang, and D. Zhang, IEEE Trans. Neural Networks and Learning Systems, 2018.

Part III: Self-Adversarial Transfer Learning

- [8] Q. Duan, L. Zhang, W. Zuo, ACM MM, 2017.
- [9] L. Zhang, Q. Duan, W. Jia, D. Zhang, X. Wang, IEEE Trans. Cybernetics, 2018. in review

Part IV: Guide Learning (An ambition for TL/DA)

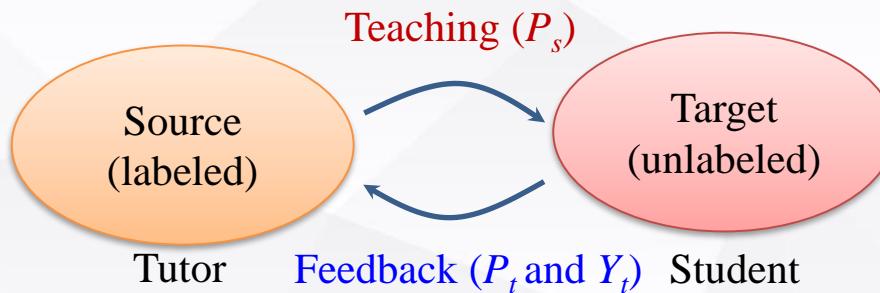
- [10] J. Fu, L. Zhang, B. Zhang, W. Jia, CCBR oral, 2018.
- [11] L. Zhang, J. Fu, S. Wang, D. Zhang, D.Y. Dong, C.L. Philip Chen, IEEE Trans. Neural Net. Learn. Syst. 2018. in review.



Guided Learning

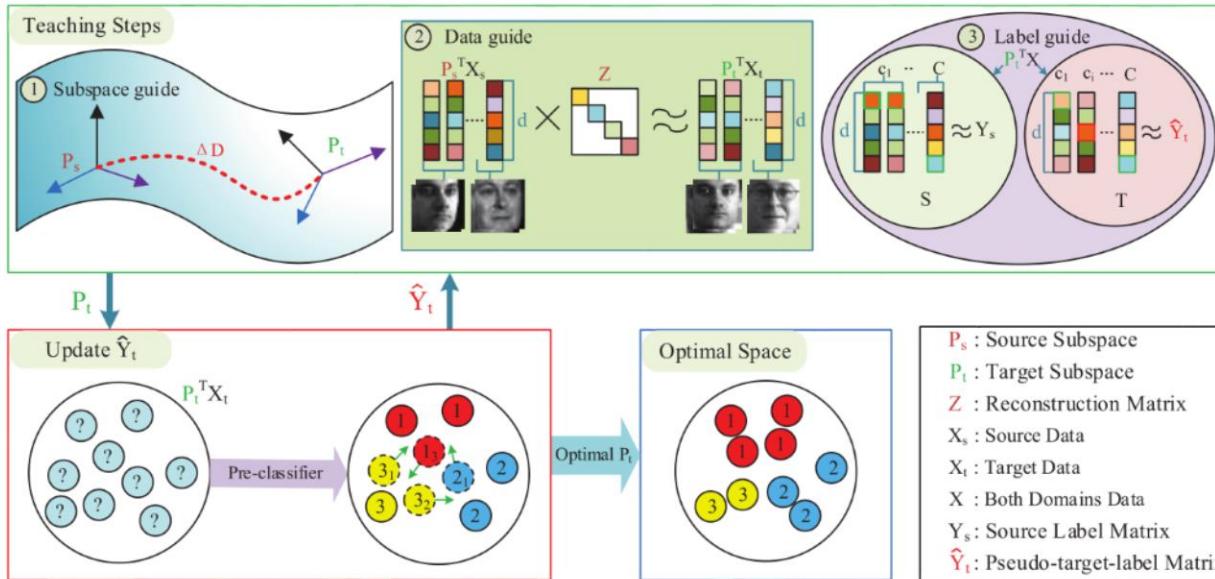
Guided Learning (GL) is a new, simple but effective paradigm, for domain disparity reduction through a progressive, guided, and multi-stage strategy, with the main idea of “**tutor guides student**” mode in human world.

Goal: “*The student surpasses the Master*” (青出于藍而勝于藍)





Guided Subspace Learning (GSL)



Three elements:

- ① Subspace guidance
- ② Data guidance-domain confusion
- ③ Label guidance-semantic confusion

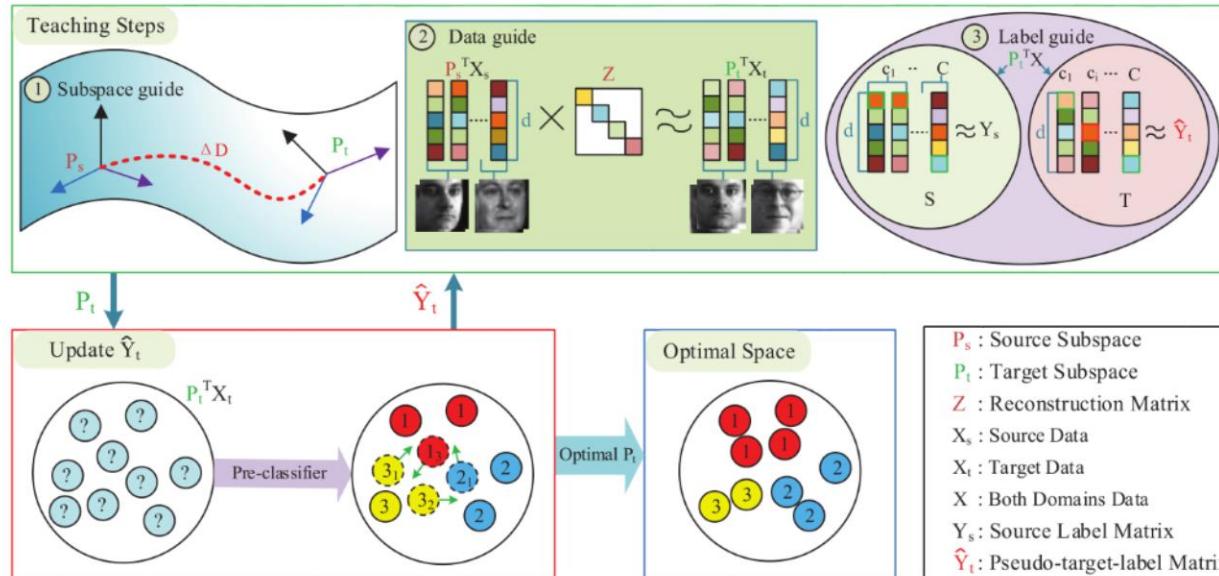
$$\begin{aligned} \min_{P_s, P_t, M, Z} \quad & \beta \|P_s - P_t\|_F^2 + \|P_t^T X_t - P_s^T X_s Z\|_F^2 + \alpha \|Z\|_* + \frac{1}{2} \|P_t^T X - Y \circ M\|_F^2 \\ \text{s.t. } & M \succ 0 \end{aligned}$$

$$X = [X_s, X_t]$$

$$Y = [Y_s, \hat{Y}_t]$$



Guided Subspace Learning (GSL)



Kernel construction

2018/9/29

$$\begin{aligned} \min_{\Phi_s, \Phi_t, M, Z} & \beta \|\phi(X)\Phi_s - \phi(X)\Phi_t\|_F^2 + \|\Phi_t^T K_t - \Phi_s^T K_s Z\|_F^2 \\ & + \alpha \|Z\|_* + \frac{1}{2} \|\Phi_t^T K - Y \circ M\|_F^2 \\ \text{s.t. } & M \succcurlyeq 0 \end{aligned}$$



Experiments on Benchmarks

| Data Set | Compared TL/DA Methods | | | | | | | | | | | | | |
|----------|------------------------|-------------|-------------|------|-------|-------------|-------------|------|-------------|-------|-------------|-------------|------------------------|---------------------|
| | NA | SA | JDA* | TSL | RDALR | LTSI | DTSL | GFK | JGSA* | CORAL | LDADA* | GSL | NGSL _{linear} | NGSL _{rbf} |
| C→A(1) | 50.1 | 54.4 | 59.8 | 52.3 | 52.5 | 24.1 | 53.3 | 56.6 | 55.1 | 45.9 | 54.8 | 56.6 | 58.7 | <u>59.3</u> |
| C→W(2) | 43.1 | 45.8 | 50.1 | 40.3 | 40.7 | 22.9 | 45.8 | 48.1 | 49.7 | 37.8 | <u>60.2</u> | 55.9 | 59.7 | 63.4 |
| C→D(3) | 47.8 | 40.9 | 44.1 | 49.0 | 45.2 | 14.6 | 51.0 | 42.9 | 46.0 | 31.8 | 41.5 | <u>49.7</u> | <u>49.7</u> | 49.0 |
| A→C(4) | 42.8 | 44.8 | 44.9 | 43.3 | 43.6 | 21.4 | 44.7 | 44.3 | 40.8 | 37.1 | 38.4 | 45.4 | 46.0 | <u>45.6</u> |
| A→W(5) | 37.0 | 44.1 | 47.0 | 34.6 | 35.9 | 18.2 | 38.3 | 42.7 | 59.0 | 37.9 | <u>49.3</u> | 41.7 | 44.1 | 45.1 |
| A→D(6) | 37.2 | 37.7 | 44.2 | 38.9 | 36.9 | 22.3 | 39.5 | 39.9 | 49.4 | 38.5 | 39.1 | 44.0 | <u>47.1</u> | <u>47.1</u> |
| W→C(7) | 29.5 | 32.3 | 29.8 | 31.4 | 28.1 | 34.6 | 30.3 | 32.0 | 29.7 | 32.5 | 31.7 | 35.3 | 37.9 | <u>37.8</u> |
| W→A(8) | 34.2 | 43.3 | 42.0 | 34.7 | 31.2 | 39.5 | 34.7 | 38.3 | 34.6 | 39.4 | 35.1 | 40.7 | 41.8 | <u>42.1</u> |
| W→D(9) | 80.6 | 70.3 | 86.3 | 79.6 | 83.4 | 72.6 | 82.8 | 78.7 | 78.5 | 80.9 | 74.6 | <u>88.5</u> | <u>88.5</u> | 89.8 |
| D→C(10) | 30.1 | 31.1 | 34.4 | 33.1 | 32.3 | <u>35.4</u> | 30.7 | 30.8 | 30.2 | 27.8 | 29.9 | 31.8 | 35.3 | 37.6 |
| D→A(11) | 32.1 | 40.8 | 44.6 | 32.6 | 33.7 | 39.4 | 33.2 | 40.4 | 39.0 | 31.9 | 40.6 | 34.8 | 40.6 | <u>43.7</u> |
| D→W(12) | 72.2 | 74.4 | 83.3 | 72.5 | 72.5 | 74.9 | 76.6 | 80.3 | 75.1 | 69.4 | 74.7 | 84.1 | <u>85.8</u> | 86.1 |
| Average | 44.7 | 46.7 | 50.9 | 45.2 | 44.7 | 35.0 | 46.7 | 47.9 | 48.9 | 42.6 | 47.5 | 50.7 | <u>52.9</u> | 53.9 |

Wang et al. ACM MM'18: MEDA 52.7% (The Best)



Experiments on Benchmarks

MSRC-VOC2007

| Data Set | Compared TL/DA Methods | | | | | | | | | | | | | |
|----------|------------------------|------|------|------|-------|------|------|------|------|-------|-------|------|------------------------|---------------------|
| | NA | SA | JDA | TSL | RDALR | LTSL | DTSL | GFK* | JGSA | LDADA | CORAL | GSL | NGSL _{linear} | NGSL _{rbf} |
| M→V(1) | 37.1 | 31.8 | 38.2 | 32.4 | 37.5 | 38.0 | 38.0 | 28.8 | 38.7 | 25.1 | 33.9 | 41.8 | 40.7 | 42.0 |
| V→M(2) | 55.5 | 46.0 | 59.3 | 43.2 | 62.3 | 67.1 | 56.4 | 48.9 | 49.3 | 43.2 | 54.1 | 66.4 | 64.7 | 68.2 |
| Average | 46.3 | 38.9 | 48.8 | 37.8 | 49.9 | 52.6 | 47.2 | 38.9 | 44.0 | 34.2 | 44.0 | 54.1 | 52.7 | 55.1 |

COIL-20

| Data Set | Compared TL/DA Methods | | | | | | | | | | | | | |
|----------|------------------------|------|------|------|-------|------|------|------|------|-------|-------|------|------------------------|---------------------|
| | NA | SA | JDA | TSL | RDALR | LTSL | DTSL | GFK* | JGSA | LDADA | CORAL | GSL | NGSL _{linear} | NGSL _{rbf} |
| C1→C2(1) | 82.7 | 86.7 | 88.7 | 80.0 | 80.7 | 75.4 | 84.6 | 72.5 | 85.1 | 77.9 | 84.9 | 88.8 | 92.9 | 92.1 |
| C2→C1(2) | 84.0 | 90.6 | 93.1 | 75.6 | 78.8 | 72.2 | 84.2 | 74.2 | 83.9 | 81.5 | 87.9 | 89.2 | 89.3 | 90.3 |
| Average | 83.3 | 88.7 | 90.9 | 77.8 | 79.7 | 73.8 | 84.4 | 73.3 | 84.5 | 79.7 | 86.4 | 89.0 | 91.1 | 91.2 |

Multi-PIE

| Data Set | Compared Transfer Learning Methods | | | | | | | | | | | | | |
|----------|------------------------------------|------|------|------|-------|------|------|------|------|-------|-------|------|------------------------|---------------------|
| | NA | SA | JDA* | TSL | RDALR | LTSL | DTSL | GFK* | JGSA | LDADA | CORAL | GSL | NGSL _{linear} | NGSL _{rbf} |
| P1→P4(1) | 51.8 | 42.8 | 84.5 | 46.7 | 41.7 | 20.0 | 81.3 | 31.2 | 76.1 | 35.6 | 26.0 | 84.8 | 83.7 | 75.1 |
| P4→P1(2) | 65.9 | 51.4 | 80.6 | 59.2 | 48.1 | 52.8 | 79.7 | 34.2 | 73.3 | 39.5 | 36.6 | 83.9 | 83.1 | 81.1 |
| P4→P5(3) | 52.0 | 47.9 | 54.6 | 45.2 | 48.8 | 47.0 | 71.0 | 37.4 | 55.3 | 26.9 | 40.8 | 71.8 | 65.2 | 67.0 |
| P5→P4(4) | 53.4 | 43.1 | 57.0 | 53.1 | 44.5 | 23.6 | 66.1 | 31.3 | 64.4 | 29.3 | 30.2 | 63.2 | 64.4 | 70.0 |
| Average | 55.8 | 53.8 | 69.2 | 51.1 | 45.8 | 35.9 | 74.5 | 33.5 | 67.3 | 32.8 | 33.4 | 75.9 | 74.1 | 73.3 |



References

- [1] L. Zhang and D. Zhang, "LSDT: Latent sparse domain transfer for visual adaptation," **IEEE Transactions on Image Processing**, 2016.
- [2] L. Zhang and D. Zhang, "Robust Visual Knowledge Transfer via EDA," **IEEE Transactions on Image Processing**, 2016.
- [3] L. Zhang and D. Zhang, "Cost-sensitive Discriminative Learning with application to Vision and Olfaction," **IEEE Transactions on Instrumentation and Measurement**, 2017.
- [4] L. Zhang, S. Wang, G.B. Huang, W. Zuo, J. Yang, and D. Zhang, "Manifold Criterion Guided Transfer Learning via Intermediate Domain Generation," **IEEE Transactions on Neural Networks and Learning Systems**, 2018.
- [5] L. Zhang, Y. Liu, and P. Deng, "Odor Recognition in Multiple E-nose Systems with Cross-domain Discriminative Subspace Learning," **IEEE Transactions on Instr. Meas.**, 2017.
- [6] L. Zhang and D. Zhang, "Visual Understanding via Multi-feature Shared Learning with Global Consistency," **IEEE Transactions on Multimedia**, 2016.
- [7] L. Zhang, J. Yang, D. Zhang, "Domain Class Consistency based Transfer Learning for Image Classification Across Domain," **Information Sciences**, 2017.
- [8] L. Zhang, Sunil Kr. Jha, T. Liu, "Discriminative Kernel Transfer Learning via $l_2,1$ -Norm Minimization," **IEEE International Joint Conference on Neural Networks**, 2016.
- [9] Q. Duan, L. Zhang, "AdvNet: Adversarial Contrastive Residual Net for 1 Million Kinship Recognition," **ACM MM**, 2017.
- [10] S. Wang, L. Zhang, "Class-specific Recognition Transfer Learning via Sparse Low-rank Constraint," **ICCV W**, 2017.



Resources/Codes can be found in

http://www.leizhang.tk/publications_and_codes.html

Learning Intelligence & Vision Essential (LiVE) Group

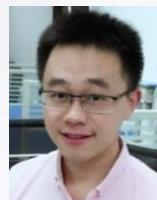
Ph.D Students



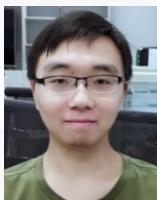
Qingyan Duan
Deep learning,
Face Recognition



Shanshan Wang
Transfer learning,
Image Recognition



Chao Yin
Deep learning,
Fine-grained Vision



Zhenwei He
Deep learning,
Object Detection



Zhipu Liu
Domain adaptation,
Person Re-ID

Master Students



Ji Liu
Hashing learning,
Image Retrieval



Fangyi Liu
Deep learning,
Person Re-ID



Jingru Fu
Transfer learning
Image Recognition,



Ni Xiao
Transfer learning,
Face Recognition



Fuxiang Huang
Hashing learning,
Computer Vision



Keyang Wang
Deep learning,
Video Detection



Yingguo Xu
Deep learning,
Machine Vision



Zhongzhou Zhang
Transfer learning,
Computer Vision



Yan Liu
Subspace learning,
Machine Olfaction



Pingling Deng
Sparse learning,
Machine Olfaction

THANK YOU!



+86-13629788369



leizhang@cqu.edu.cn



<http://www.leizhang.tk>



website