Conditional Bures Metric for Domain Adaptation

SUN YAT-SEN UNIVERSITY · 琶洲实验室 PAZHOU LAB

You-Wei Luo   Chuan-Xian Ren

CVPR VIRTUAL JUNE 19-25

Highlights

- **Problem:** we study the Unsupervised Domain Adaptation (UDA) for knowledge transfer under the conditional shift pre-condition.
- **Theory:** we propose the **Conditional Kernel Bures (CKB)** metric, which explores the Wasserstein-Bures geometry of conditional distributions.
- **Model:** we develop a conditional distributions matching network to mitigate the **negative transfer** and **mismatched distributions** in UDA.
- **Experiments:** the proposed model achieves the SOTA on UDA benchmarks while using a simple and fast network architecture.

Domain Adaptation

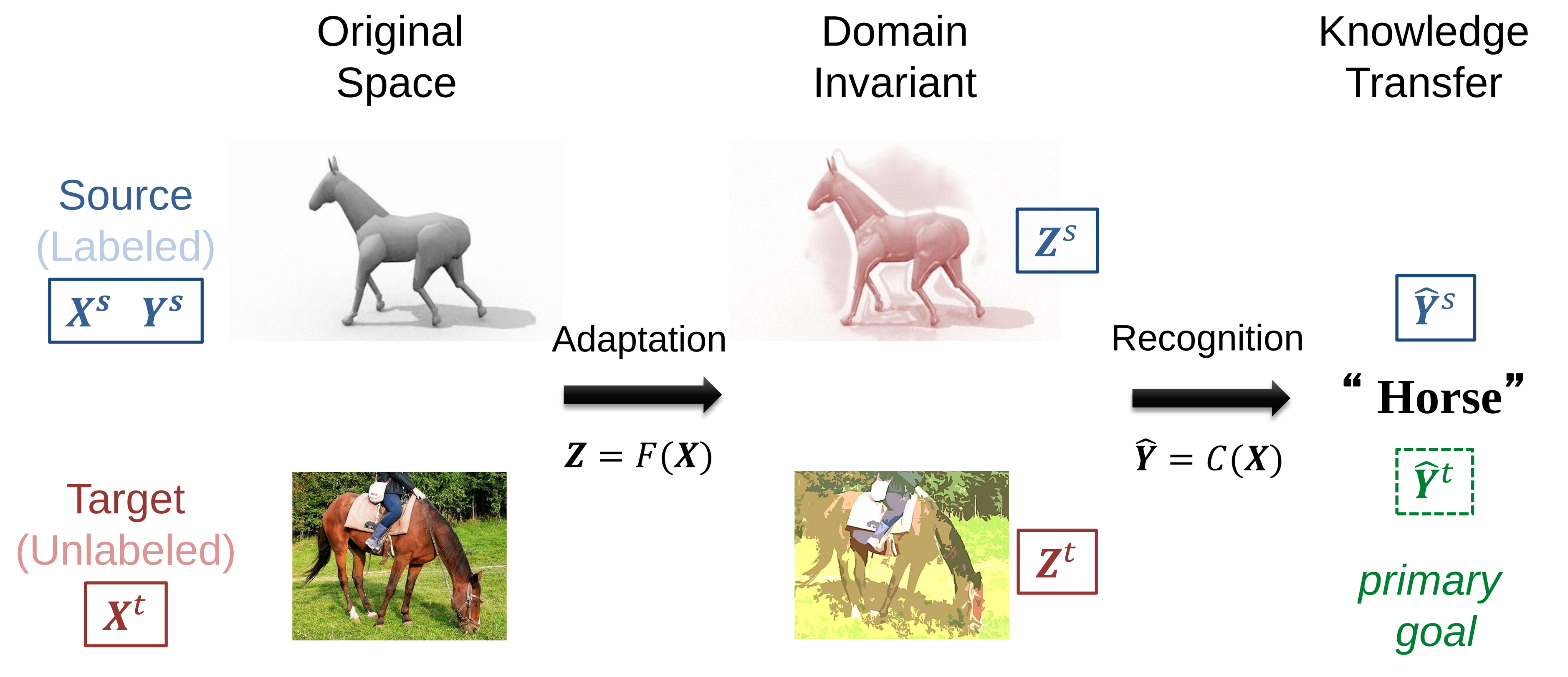
- **Problem Definition:** transfer the *task-related knowledge* from a *labeled source* domain to an *unlabeled target* domain.
- **Statistical Modeling:** the domains are defined as the joint distribution

Source $\mathcal{D}^s$: $(\boldsymbol{X}^s, \boldsymbol{Y}^s) \sim P^s_{XY}$ Target $\mathcal{D}^t$: $(\boldsymbol{X}^t, \boldsymbol{Y}^t) \sim P^t_{XY}$ *unknown*

- **Shift Assumption:** the shift of dataset is induced by conditional distributions

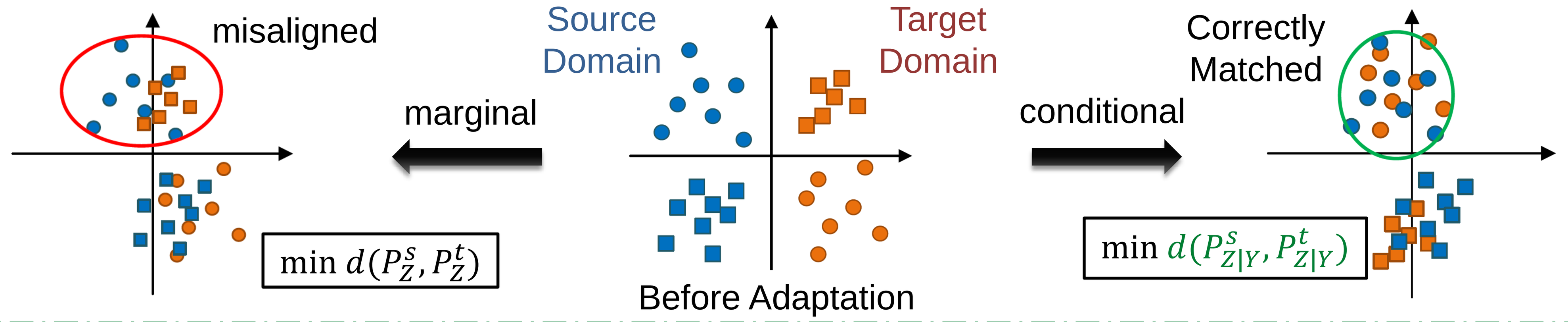
conditional shift $\quad P^s_{X|Y} \neq P^t_{X|Y} \quad \Rightarrow \quad P^s_{XY} \neq P^t_{XY}$

- **Pipeline:**

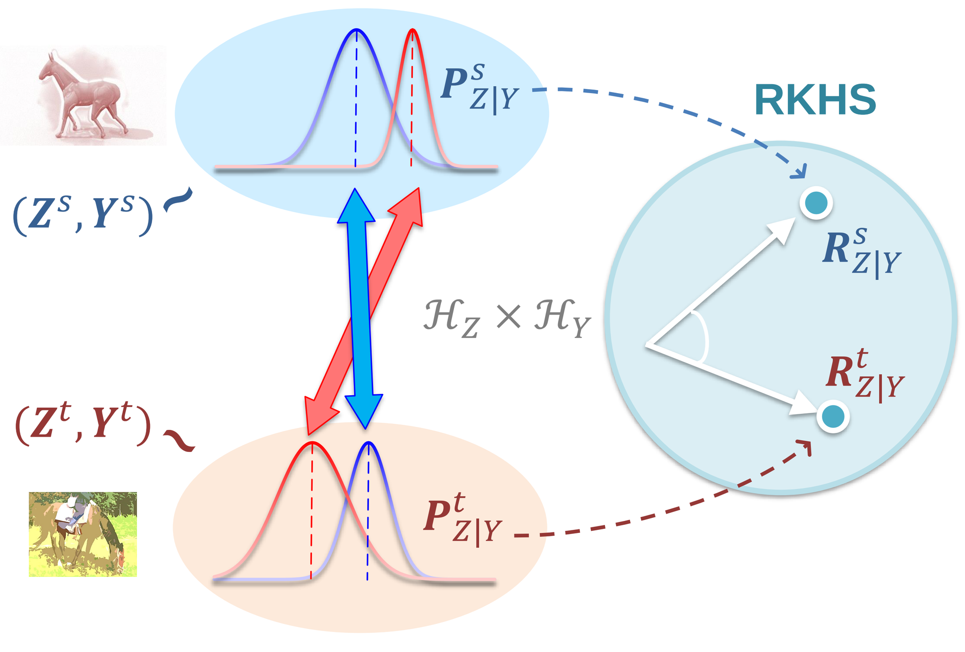


CKB for Distribution Matching

- **Insight:**
  - marginal adaptation of covariate $\boldsymbol{Z}$ may ignore the conditional distribution $P_{Z|Y}$;
  - conditional adaptation correctly matches $P_{Z|Y}$, but **well-defined metric** is **unexplored**.



- **Wasserstein-Bures Geometry:**
  - A close-form solution for conditional OT in Hilbert space



Conditional Distribution Embedding

$$\boxed{d^2_{\mathrm{CKB}} = \mathrm{tr}(\boldsymbol{R}^s_{ZZ|Y} + \boldsymbol{R}^t_{ZZ|Y} - 2\boldsymbol{R}^{st}_{ZZ|Y})}$$

- **Identifiability** of CKB metric

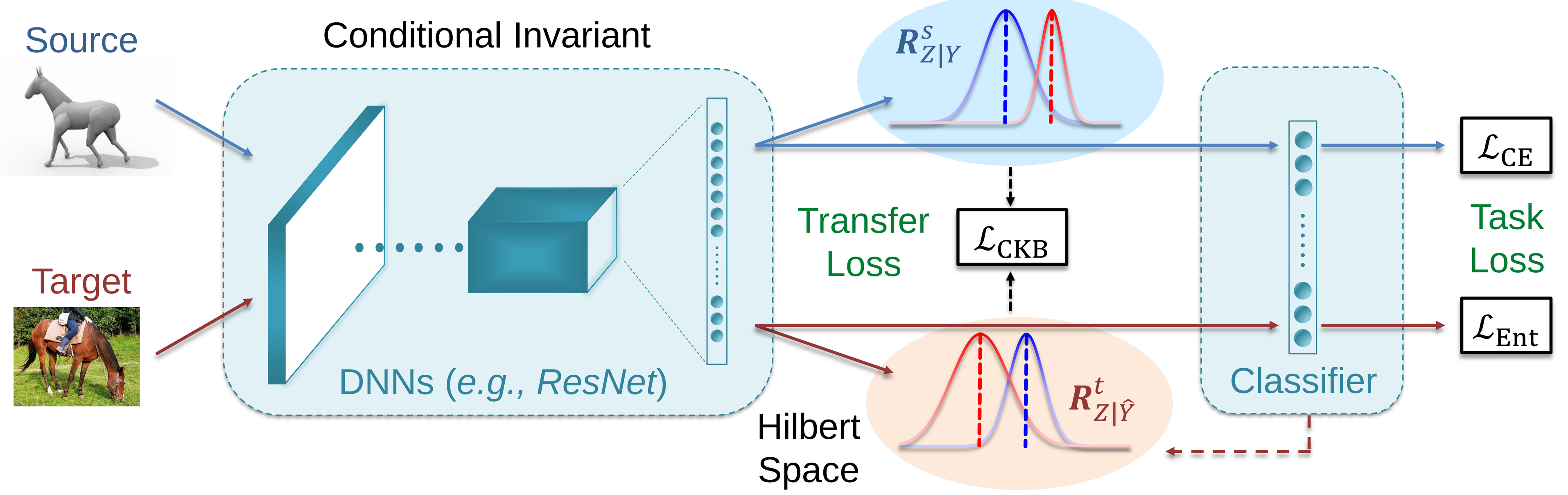
Thm. *If $\varphi\#P^s_{(X,Y)}$ and $\varphi\#P^t_{(X,Y)}$ are Gaussian, $k_\mathcal{X}$ and $k_\mathcal{Y}$ are c_0-universal , we have*

$$d_W^{\mathcal{H}_\mathcal{X}} = d_{\mathrm{CKB}} = 0 \quad \Longrightarrow \quad P^s_{Z|Y} = P^t_{Z|Y}$$

- CKB is **generally well-defined and converge** as $n \to \infty$
- The empirical CKB is **consistent** and **distribution-free**
- CKB can be **applied to any model straightforwardly**

Conditional Distribution Matching Network

- A simple and efficient model for learning conditional invariant representations.



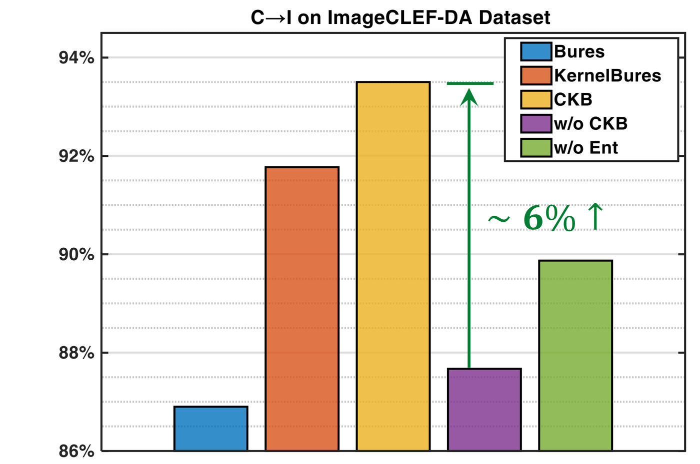
Experiments

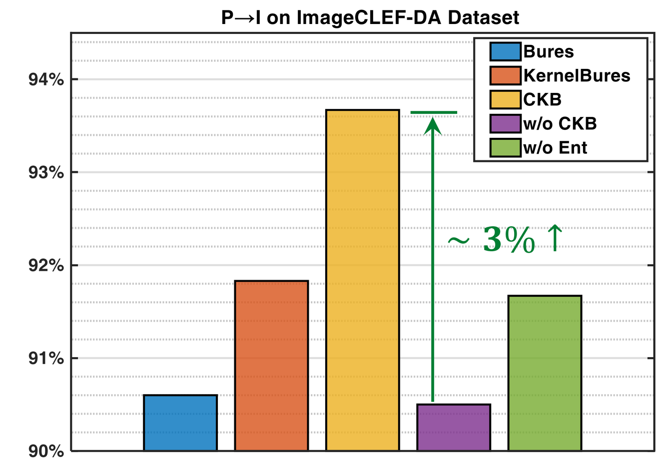
- **Comparison with SOTA methods**
  - Office-Home: 3 domains, 12 classes   Image-CLEF: 3 domains, 12 classes
  - Significant improvement compared with other OT models (e.g., KGOT and ETD)

Tab. *Classification Accuracy (%) on the Target Domain*

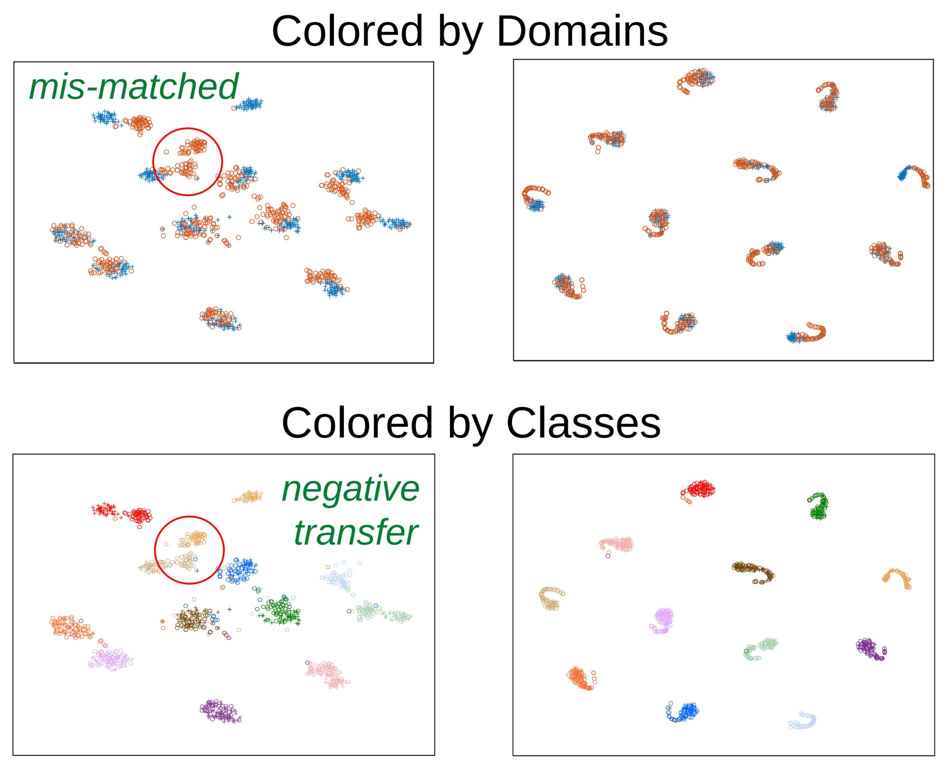
Datasets	ResNet *CVPR 16*	KGOT *TPAMI 20*	CDAN+E *NeurIPS 20*	ETD *CVPR 20*	DMP *TPAMI 20*	CKB
Office-Home	46.1%	54.7%	65.8%	67.3%	68.1%	**68.5%**
Image-CLEF	80.7%	84.1%	87.7%	89.7%	89.1%	**90.2%**

- **Ablation Study**
  - **Bures:** matches the covariances $\cancel{\mathcal{L}_{\mathrm{CKB}}} \to \mathcal{L}_{\mathrm{Bures}}$
  - **Kernel Bures:** matches P_Z $\cancel{\mathcal{L}_{\mathrm{CKB}}} \to \mathcal{L}_{\mathrm{KB}}$
  - **w/o $\mathcal{L}_{\mathrm{CKB}}$:** without CKB loss
  - **w/o $\mathcal{L}_{\mathrm{Ent}}$:** without target entropy





- **Feature Visualization**
  - 1st Col.: before adaptation
  - 2nd Col.: after adaptation



- **Hyper-parameter Sensitivity**
  - Surface is smooth