Multi-Source Domain Adaptation through Dataset Dictionary Learning in Wasserstein Space

Eduardo Montesuma [1] Fred Ngolè Mboula [1] Antoine Souloumiac [1]

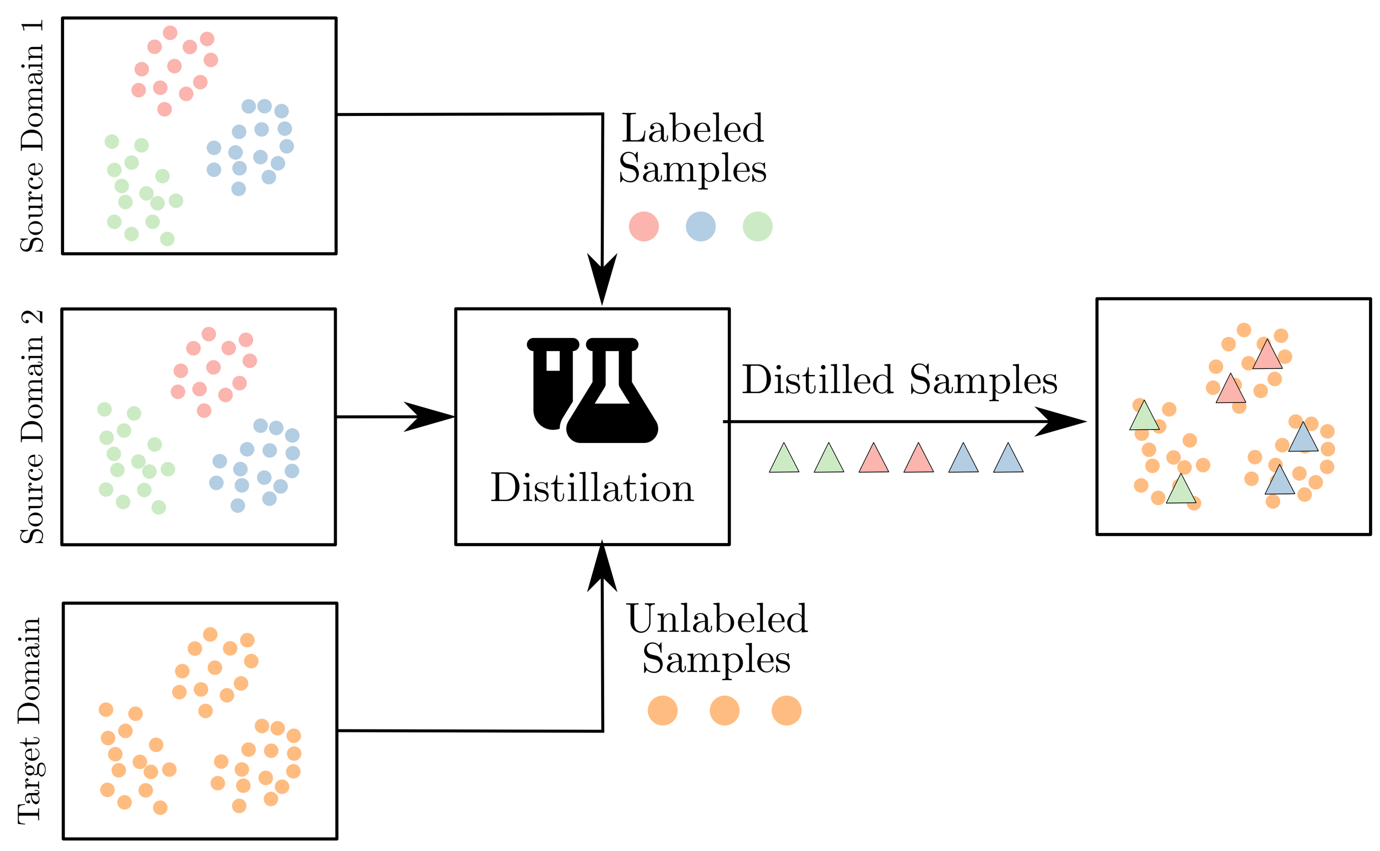
[1]Université Paris-Saclay, CEA, List, F-91120 Palaiseau France

Abstract

In this paper, we consider the intersection of two problems in machine learning: **Multi-Source Domain Adaptation (MSDA)** and **Dataset Distillation (DD)**. On the one hand, the first considers adapting multiple heterogeneous labeled source domains to an unlabeled target domain. On the other hand, the second attacks the problem of synthesizing a small summary containing all the information about the datasets. We thus consider a new problem called MSDA-DD. To solve it, we adapt previous works in the MSDA literature, such as Wasserstein Barycenter Transport and Dataset Dictionary Learning, as well as DD method Distribution Matching. We thoroughly experiment with this novel problem on four benchmarks (Caltech-Office 10, Tennessee-Eastman Process, Continuous Stirred Tank Reactor, and Case Western Reserve University), where we show that, even with as little as 1 sample per class, one achieves state-of-the-art adaptation performance.

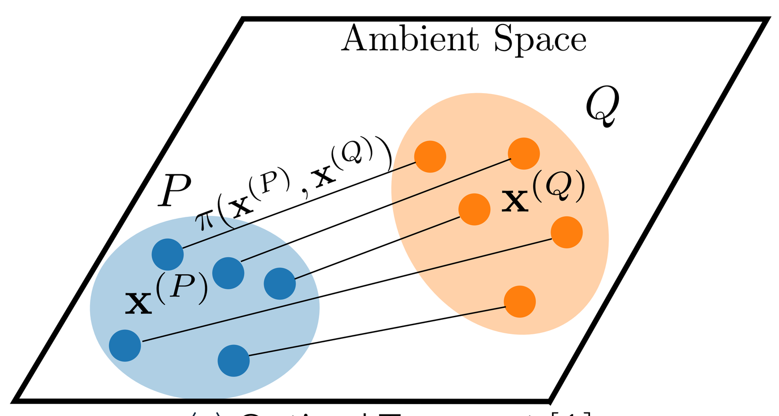
Methodology

Overview

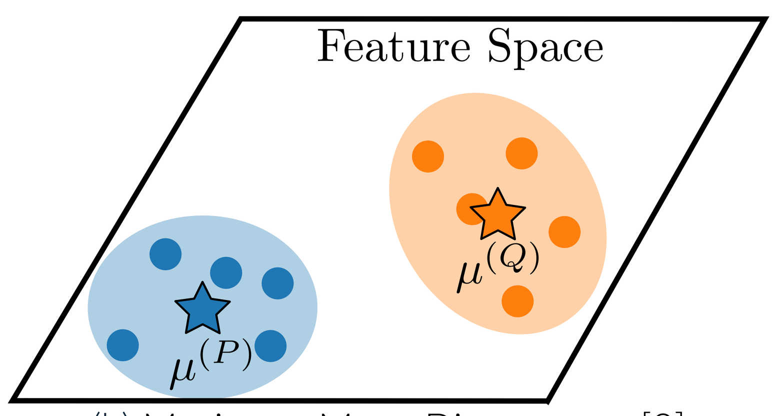


Multi-Source Domain Adaptation-Dataset Distillation. We search for a small, synthetic, labeled summary that is close **in distribution** to the target domain.

Probability Metrics



(a) Optimal Transport [1]



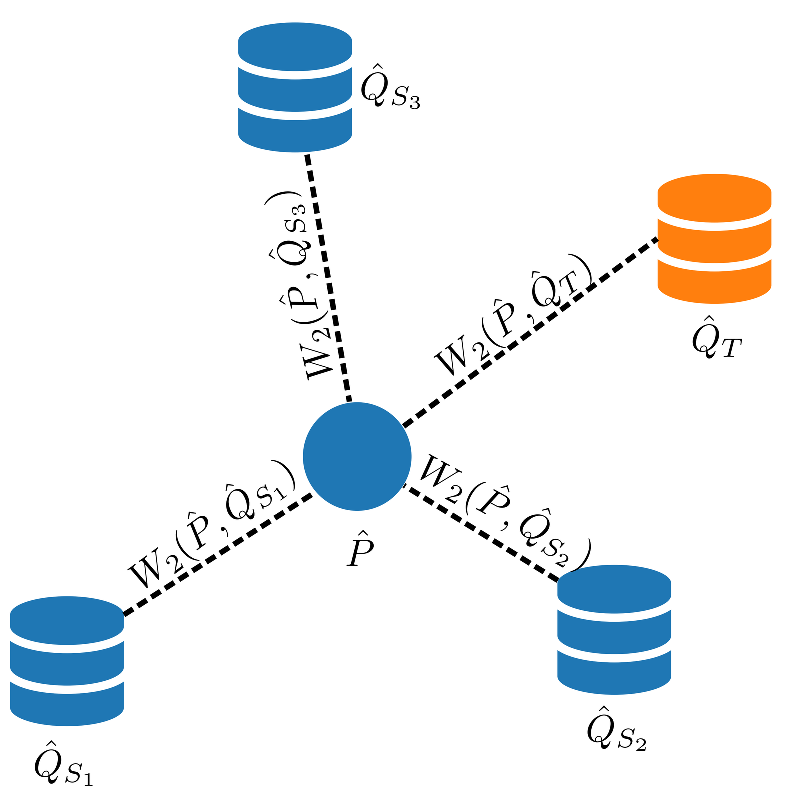
(b) Maximum Mean Discrepancy [2]

We use optimal transport and the maximum mean discrepancy for comparing probability distributions,

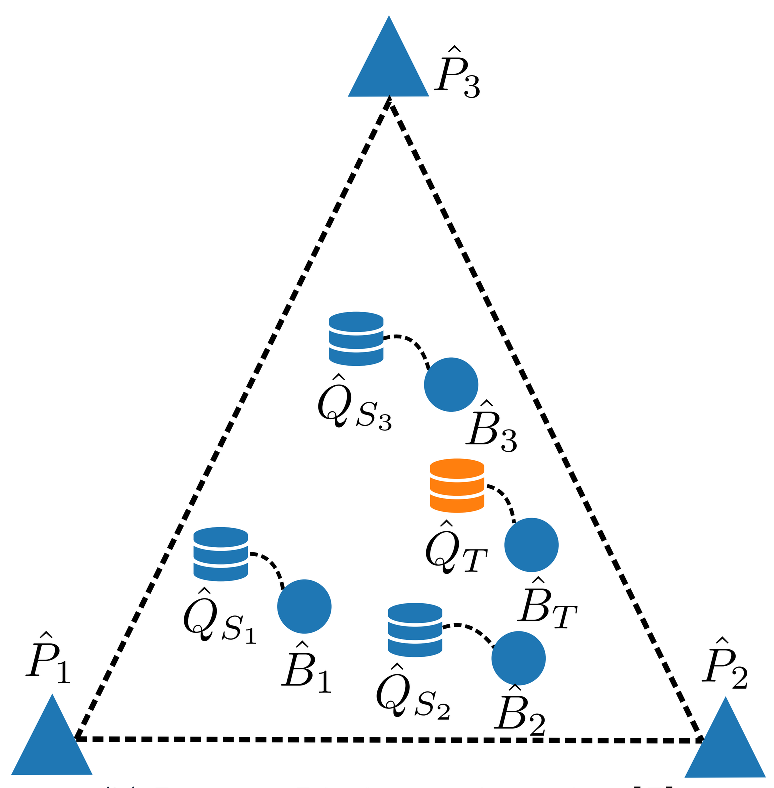
$$W_c(\hat{P},\hat{Q}) = \underset{\pi\in\Pi(P,Q)}{\text{argmin}} \sum_{i=1}^{n}\sum_{j=1}^{m}\pi_{ij}c((\mathbf{x}_i^{(P)},\mathbf{y}_i^{(P)}),(\mathbf{x}_j^{(Q)},\mathbf{y}_j^{(Q)})), \quad \text{(Optimal Transport)}$$

$$\text{MMD}_c(\hat{P},\hat{Q}) = \sum_{c=1}^{n_c}\|\mu_c^{(P)} - \mu_c^{(Q)}\|_2^2. \quad \text{(Maximum Mean Discrepancy)}$$

Proposed Methods



(a) Distribution Matching [3]



(b) Dataset Dictionary Learning [5]

$$\hat{P} = \underset{\{\mathbf{x}_i^{(P)},y_i^{(P)}\}_{i=1}^m}{\text{argmin}} D(\hat{P},\hat{Q}_T) + \sum_{\ell=1}^{N_S} D_c(\hat{P},\hat{Q}_{S_\ell}).$$

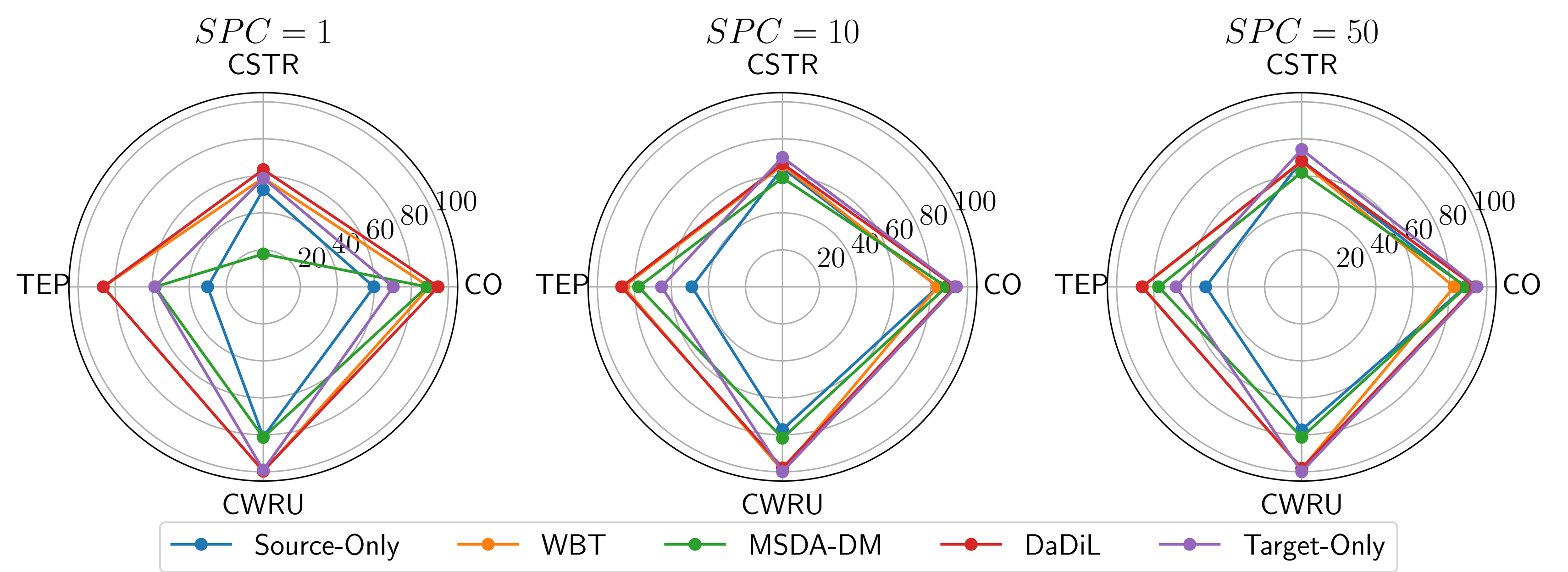
- Synthetic samples: $\{\mathbf{x}_i^{(P)}, y_i^{(P)}\}_{i=1}^m$
- $D = W_2$, $D_c = W_c \implies$ WBT [4]
- $D = \text{MMD}$, $D_c = \text{MMD}_c \implies$ MSDA-DM

$$(\mathcal{P}^\star,\mathcal{A}^\star) = \underset{\mathcal{P},\mathcal{A}}{\text{argmin}}\, W_2(\hat{Q}_T,\hat{B}_T) + \sum_{\ell=1}^{N_S} W_c(\hat{Q}_\ell,\hat{B}_\ell).$$

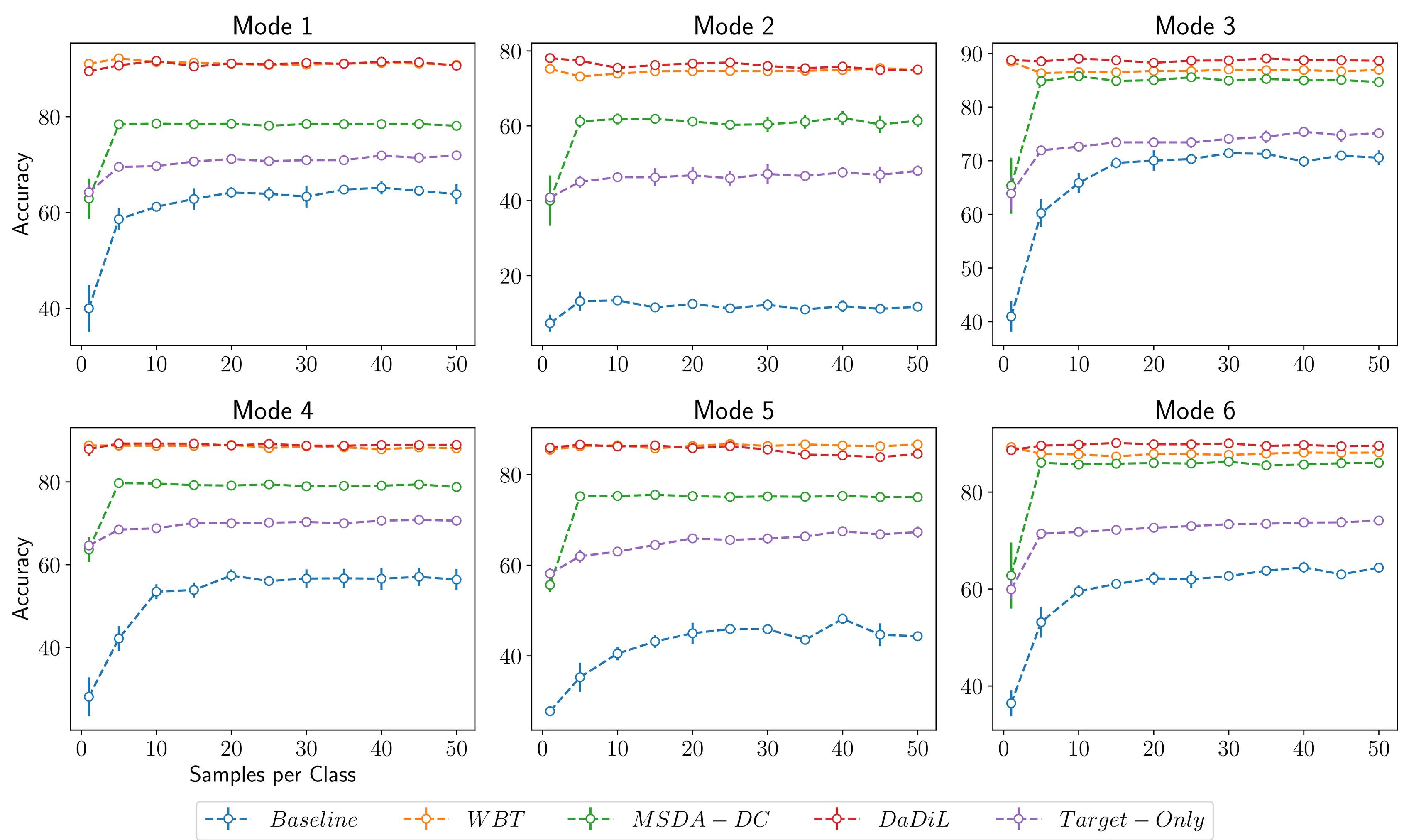
- Atoms: $\mathcal{P} = \{\hat{P}_k\}_{k=1}^K$,
- Barycentric coordinates $\mathcal{A} = \{\alpha_\ell\}_{\ell=1}^{N_S+1}$,
- $\hat{Q}_T$ is compressed via $\hat{B}_T = \mathcal{B}(\alpha_T;\mathcal{P})$.

Empirical Results

Overview

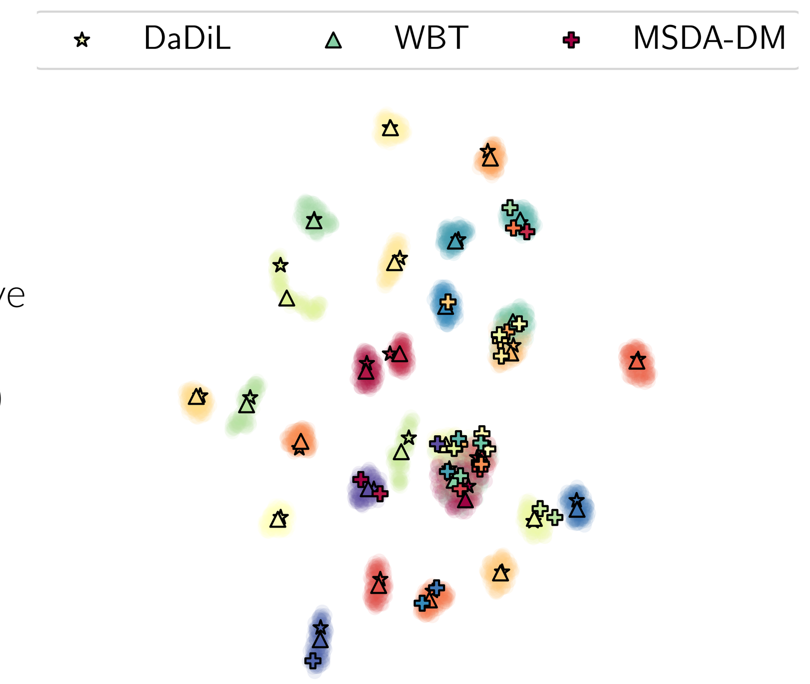


Performance vs. Dataset Size



Distribution Visualization

- DaDiL and WBT produce summaries that respect class boundaries, contrary to MSDA-DM.
- MSDA-DM based on the MMD is sensitive to initialization.
- MSDA-DM based on MMD (linear kernel) only aligns the first-order moments.
- The Wasserstein distance is a strong candidate for dataset distillation.



Conclusionk

- We perform **domain adaptation** and **data distillation** simultaneously.
- We **improve performance on taget domain** while **reducing the overall dataset size**.
- In the TEP benchmark we achieve state-of-the-art performance with only 0.16% of the total amount of samples.

Future Works

Our work opens a novel line of research on dataset distillation and domain adaptation. For future works we plan to use this framework for **incremental transfer learning**.

References

[1] Montesuma, E. F., Mboula, F. N., & Souloumiac, A. (2023). Recent advances in optimal transport for machine learning. arXiv preprint arXiv:2306.16156.

[2] Gretton, A., Borgwardt, K. M., Rasch, M. J., Schölkopf, B., & Smola, A. (2012). A kernel two-sample test. The Journal of Machine Learning Research, 13(1), 723-773.

[3] Zhao, B., & Bilen, H. (2023). Dataset condensation with distribution matching. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 6514-6523).

[4] Montesuma, E. F., & Mboula, F. N. M. (2021). Wasserstein barycenter for multi-source domain adaptation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 16785-16793).

[5] Montesuma, E. F., Mboula, F. N., & Souloumiac, A. (2023). Multi-source domain adaptation through dataset dictionary learning in wasserstein space. 26th European Conference on Artificial Intelligence

Learn more about DaDiL!

Main Paper LinkedIN Google Scholar

eduardo.fernandesmontesuma@cea.fr