MOST: Multi-Source Domain Adaptation via Optimal Transport for Student-Teacher Learning

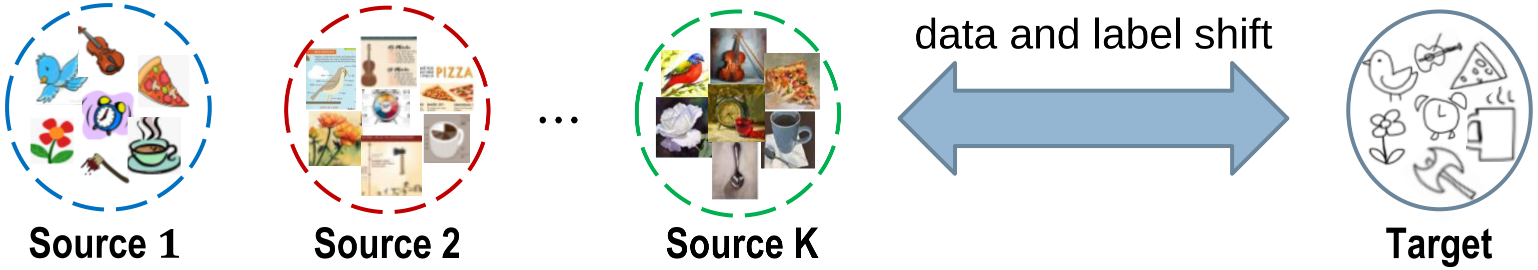
Tuan Nguyen[1], Trung Le[1], He Zhao[1], Quan Hung Tran[2], Truyen Nguyen[3], and Dinh Phung[1,4]
[1] Department of Data Science and AI, Monash University, Australia
[2] Adobe Research, San Jose, CA, USA
[3] University of Akron, USA
[4] VinAI Research, Vietnam

Tuan Nguyen
Email: tuan.ng@monash.edu

Introduction

Multi-source domain adaptation setting

- Multi-source domain adaptation (MSDA) is a difficult but practical problem, where labelled data are collected from multiple sources.
- Main problem: how to transfer knowledge from multiple source domains to a single unlabeled target domain.



Multi-source domain adaptation setting.

Introduction
Optimal Transport (OT)

- OT is a powerful mathematical theory.
- The underlying idea of OT is to seek an optimal plan of transforming one distribution of mass to another.

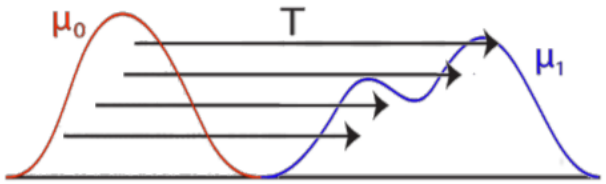
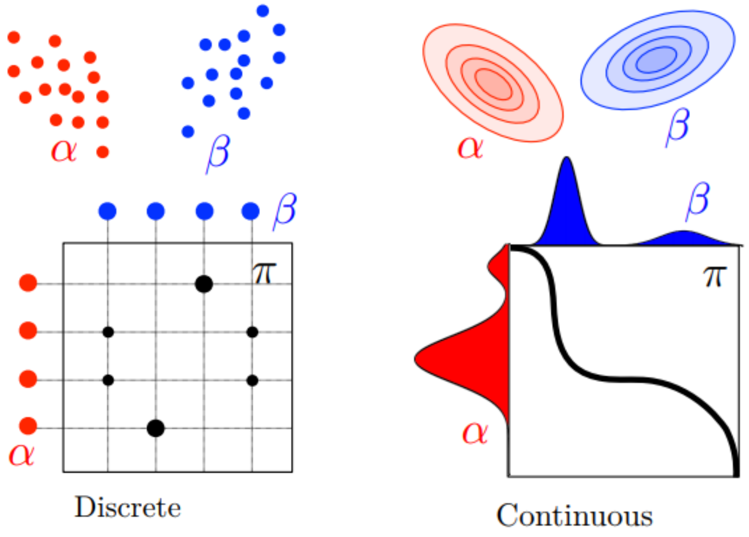


Illustration of the transport map from continuous density μ_0 to μ_1. [Papadakis, 2015]



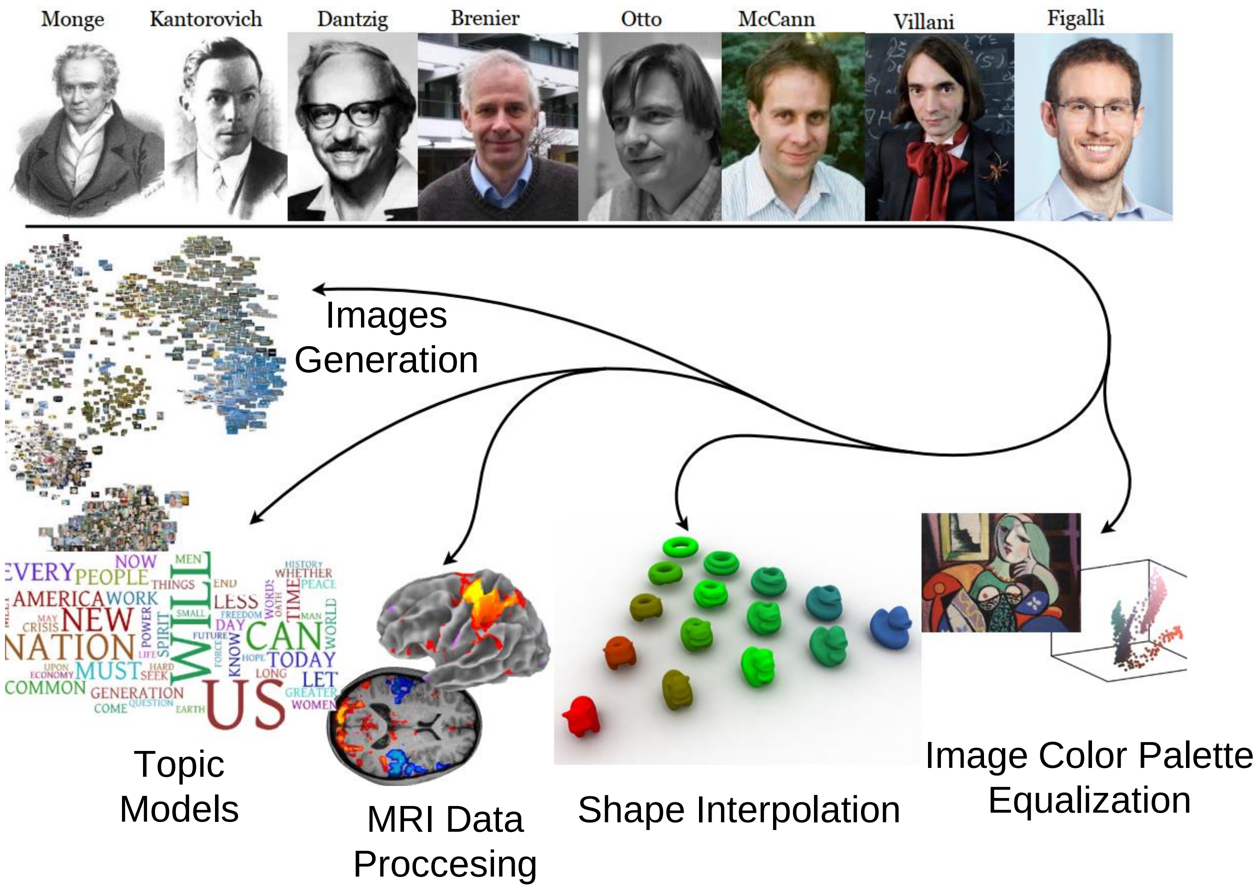
Discrete and continuous formulations of the OT problem. [Peyré and Cuturi, 2018]

N. Papadakis, Optimal Transport for Image Processing, Signal and Image Processing, 2015.

G. Peyré and M. Cuturi, Computational Optimal Transport, arxiv:1803.00567, 2018.

Introduction
Optimal Transport (OT)

- OT is a powerful mathematical theory.
- The underlying idea of OT is to seek an optimal plan of transforming one distribution of mass to another.
- OT has been applied to a diverse range of applications including images generation, topic models, shape interpolation and others.
- Clustering view of OT has been utilized to study a rich class of hierarchical and multilevel clustering problems.



Applications of Optimal Transport.

Image from slides of Optimal Transport course, Cuturi.

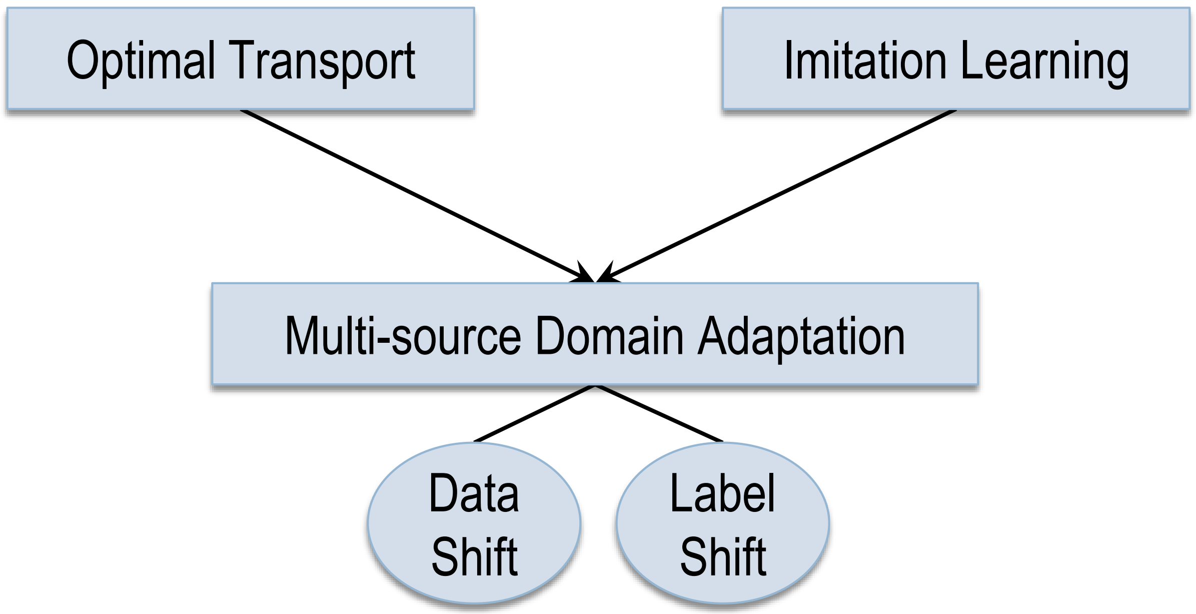
Introduction
Imitation Learning

- Two fundamental agents: an expert teacher and a student.
- The former agent knows how to do its job perfectly, whilst the latter learns a policy to mimic the teacher’s behavior.

Images from Stephane Ross.

Introduction

- Research question: How can we leverage the clustering view of optimal transport and the learning paradigm of imitation learning to tackle data shift and label shift in multi-source domain adaptation?



Proposed Method

MOST: Multi-Source Domain Adaptation via Optimal Transport for Student-Teacher Learning

- Optimal Transport based Imitation Learning
  - Consider two data domains $\mathcal{X}^A$ and $\mathcal{X}^B$ with two data distributions $\mathbb{P}^A$ and $\mathbb{P}^B$ respectively.
  - h^A is a classifier that predicts well on $\mathbf{x} \sim \mathbb{P}^A$.
  - Goal: learns the classifier h^B to predict accurately $\mathbf{x} \sim \mathbb{P}^B$.

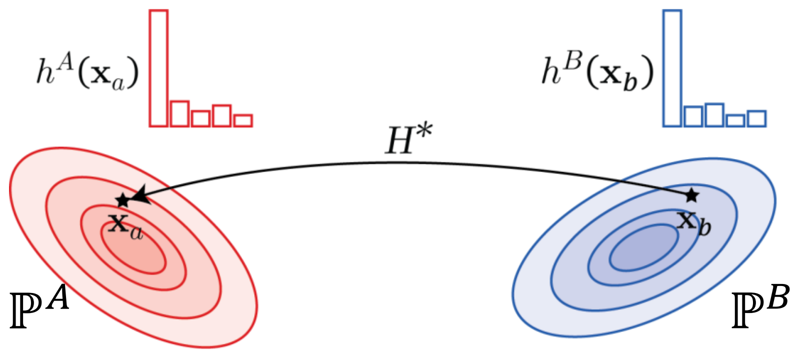


Figure 1: Imitation view explanation.

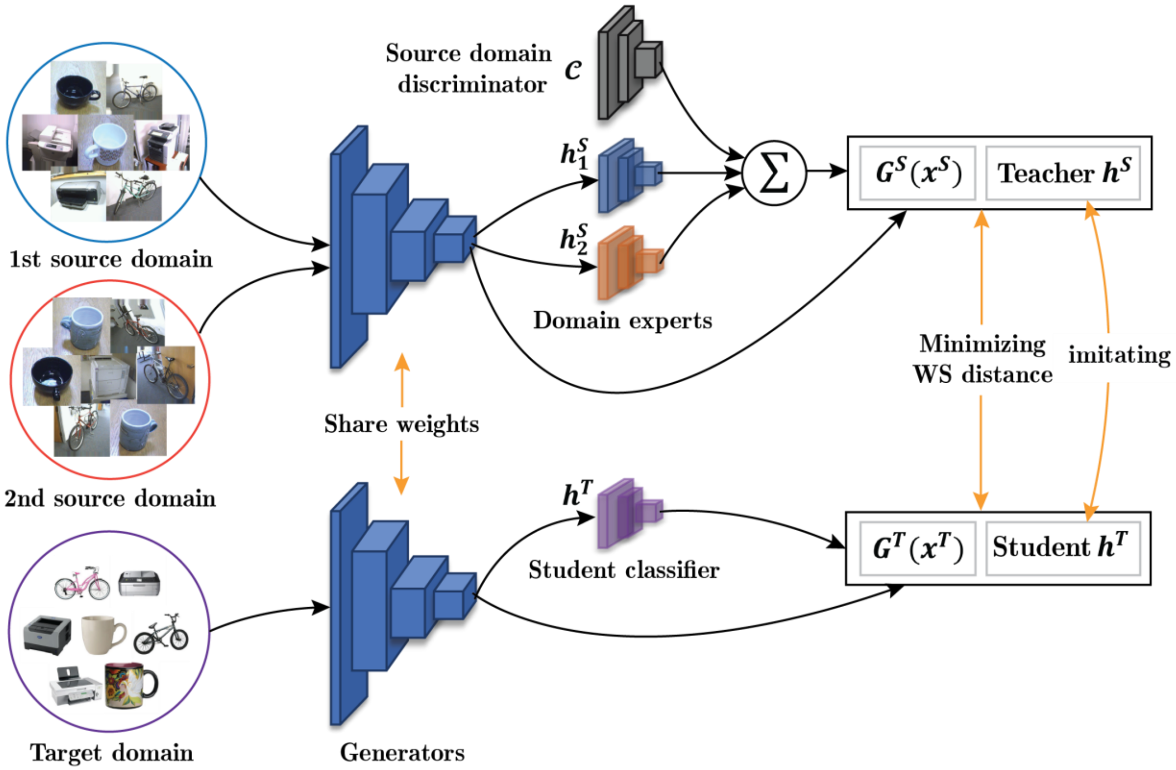
$$\boxed{\min_{h_B} W_d\left(\mathbb{P}_{A,h_A}, \mathbb{P}_{B,h_B}\right)}$$

Given $\mathbf{x}_b \sim \mathbb{P}^B$, the optimal transport H^* finds its closest counterpart in the space of data domain $\mathcal{X}^A$, which is $\mathbf{x}_a = H^*(\mathbf{x}_b)$ so that h^B can predict well on $\mathbf{x}_b$ by conveniently imitating the prediction of h^A on $\mathbf{x}_a$.

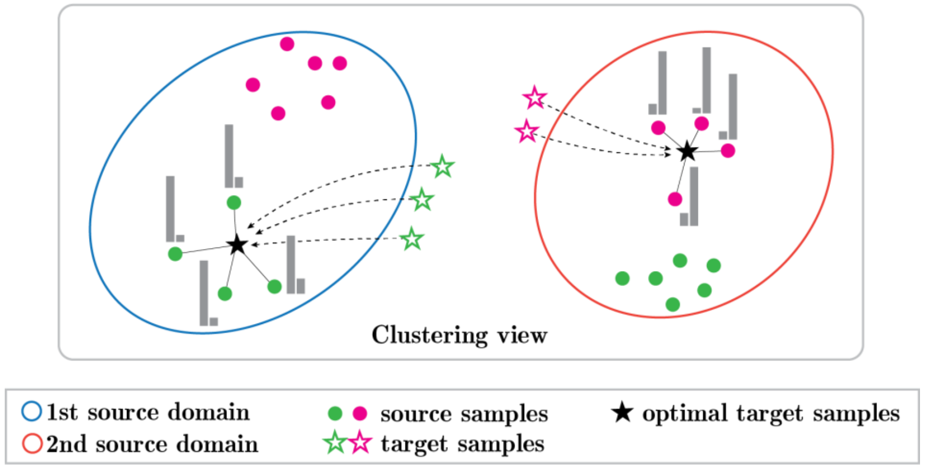
Proposed Method

MOST: Multi-Source Domain Adaptation via Optimal Transport for Student-Teacher Learning

- Main idea
  - Teacher classifier: a weighted combination of domain experts.
  - Student classifier: acts on the target domain and tries to imitate the teacher via the OT-based imitation learning.



The overall structure of our MOST.



Clustering view explanation on the joint space.

Experimental results

MOST: Multi-Source Domain Adaptation via Optimal Transport for Student-Teacher Learning

- MOST achieves state-of-the-art performance on benchmark datasets for multi-source domain adaptation, including Digits-five, Office-Caltech10, and Office-31.

Table 1: Classification results on Digits-five.

Methods	→**mm**	→**mt**	→**up**	→**sv**	→**sy**	Avg
LtC-MSDA [1]	85.6	99.0	98.3	83.2	93.0	91.8
MOST	**91.5**	**99.6**	**98.4**	**90.9**	**96.4**	**95.4**

Table 2: Classification results on Office-Caltech10

Methods	→**W**	→**D**	→**C**	→**A**	Avg
M^3SDA [2]	99.5	99.2	92.2	94.5	96.4
MOST	**100**	**100**	**96.0**	**96.4**	**98.1**

Table 3: Classification results on Office-31

Methods	→**D**	→**W**	→**A**	Avg
LtC-MSDA [1]	99.6	97.2	56.9	84.6
MOST	**100**	**98.7**	**60.6**	**86.4**

[1] H. Wang et al. Learning to combine: Knowledge aggregation for multi-source domain adaptation. In ECCV, 2020.

[2] X. Peng et al. Moment matching for multi-source domain adaptation. In ICCV, 2019.

Conclusion

- In this work, we propose a *rigorous OT-based theory* to leverage imitation learning into domain adaptation.
- Under imitation learning's perspective, we propose a novel model which utilizes two cooperative agents: teacher and student for multi-source domain adaptation.
- Comprehensive experiments are conducted with state-of-the-art performance on benchmark datasets.

Thanks for your attention
Q&A