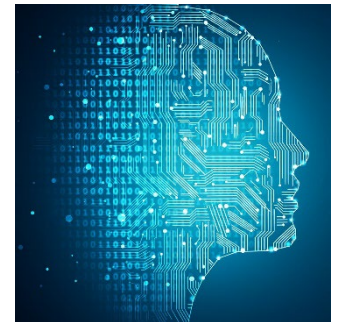


# Machine Learning Manifold Alignment



Kevin Moon (kevin.moon@usu.edu)  
STAT/CS 5810/6655



# Multimodal data



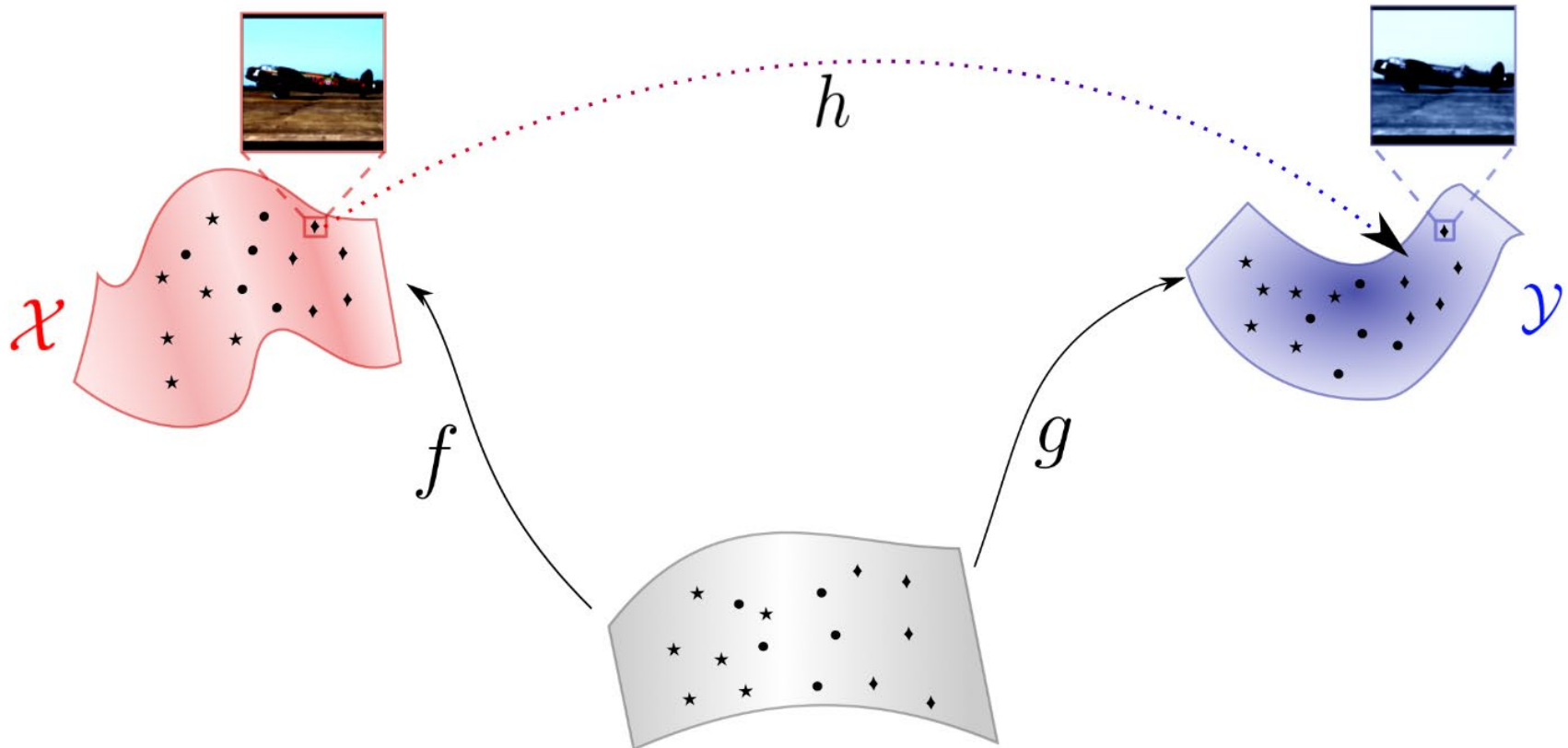
- Data may be collected of the same system under different conditions, protocols, or instruments
- **Examples**
  - Biological measurements from the same batch of cells
  - Text documents translated into different languages
  - Brain images from multiple neuroimaging techniques
  - Images captured from different views



# Manifold Alignment



- **Primary Goal:** find projections of the different domains onto a common latent space

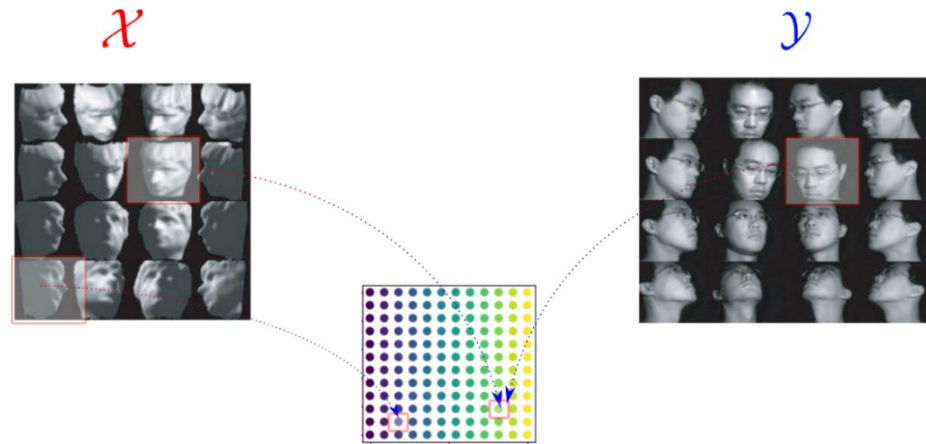
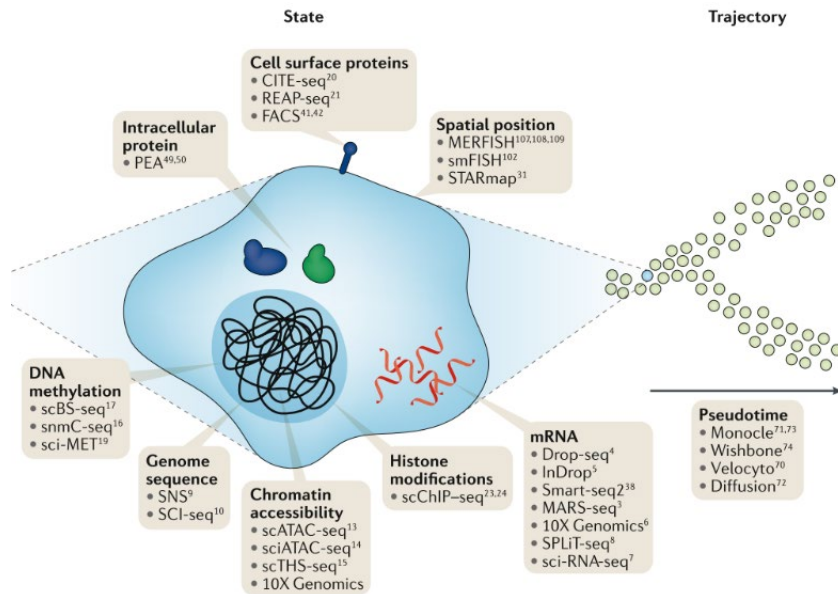


- **Secondary Goal:** learn correspondences between domains

# Manifold Alignment Benefits



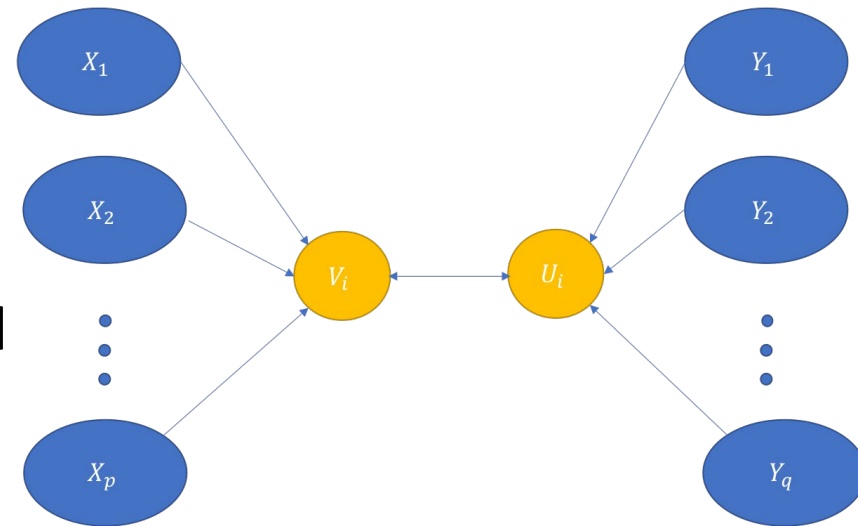
- Enhance understanding of the system
- Understand inter-domain relationships
- Batch effects correction
- Improve downstream analysis such as classification
  - Domain adaptation



# Supervised Manifold Alignment



- All data points in  $\mathcal{X}$  and  $\mathcal{Y}$  have known correspondences
- **Example:** Canonical Correlation Analysis (CCA)
  - Find directions  $V$  in  $\mathcal{X}$  and  $U$  in  $\mathcal{Y}$  that are maximally correlated
- **Example:** alternating diffusion and integrated diffusion
- Often too restrictive in practice
  - Data doesn't always come paired



<https://medium.com/analytics-vidhya/what-is-canonical-correlation-analysis-58ef4349c0b0>

# Unsupervised Manifold Alignment



- No known relationships between domains are assumed
- Typically relies on mostly rigid transformations
- Efficacy is limited

# Semi-Supervised Manifold Alignment

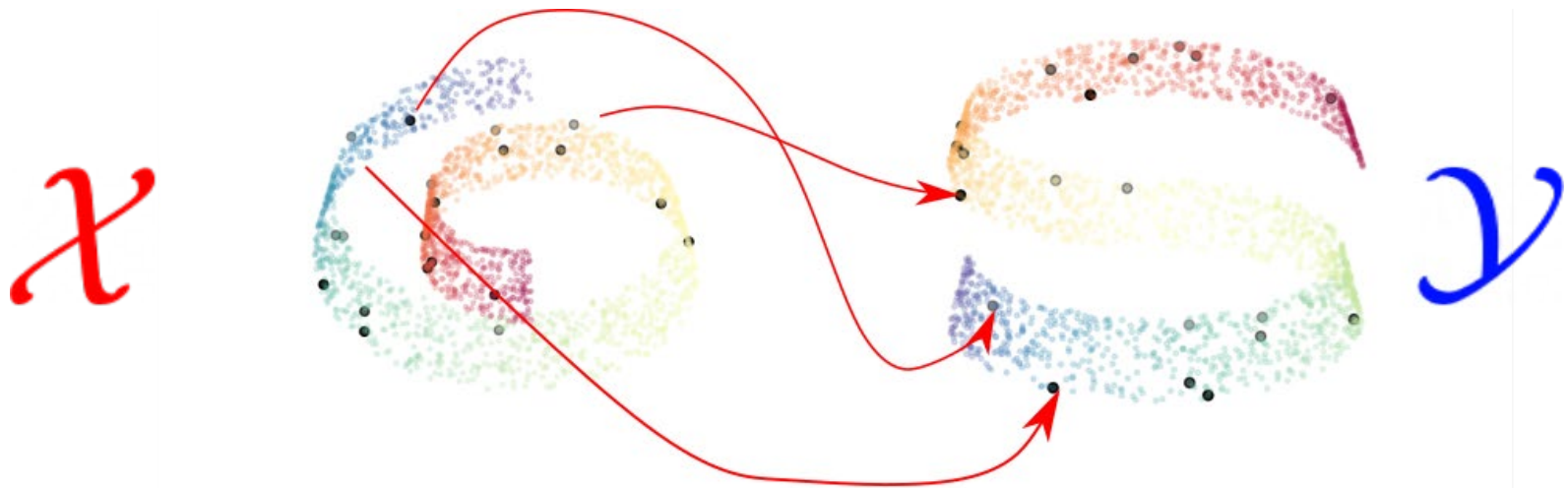


- In-between supervised and unsupervised
- Full correspondences are unknown, but some level of shared knowledge is assumed
- **Case 1:** A one-to-one correspondence between domains is known for only a few of the data points (DTA, IDA 2023)
- **Case 2:** Some coarse-grained information is available in both domains (e.g. class labels in MALI, SampTA 2023)

# Diffusion Transport Alignment (DTA)



- **Case 1:** A one-to-one correspondence between domains is known for only a few of the data points
- Let's use diffusion to learn correspondences between remaining points
- **Inherent assumption:** diffusion distances between points and known correspondences is similar in each domain



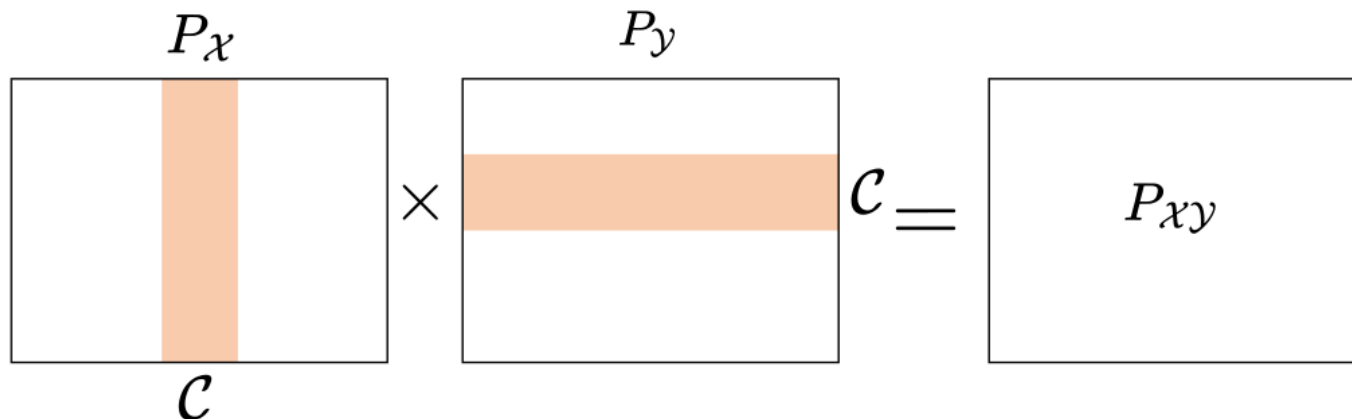


# DTA Steps



Let  $\mathcal{C}$  = set of points with known correspondences in  $\mathcal{X}$  and  $\mathcal{Y}$ .  
Then

1. Compute kernel (affinity) matrices  $K_{\mathcal{X}}$  and  $K_{\mathcal{Y}}$  in each domain
  - E.g.  $\alpha$ -decay kernel from PHATE
2. Normalize each row to obtain diffusion operators  $P_{\mathcal{X}}$  and  $P_{\mathcal{Y}}$
3. Extract the transition probabilities from every  $x \in \mathcal{X}$  to every  $c \in \mathcal{C}$  and from every  $c \in \mathcal{C}$  to every  $y \in \mathcal{Y}$
4. Compute  $P_{\mathcal{X}\mathcal{Y}}$ :

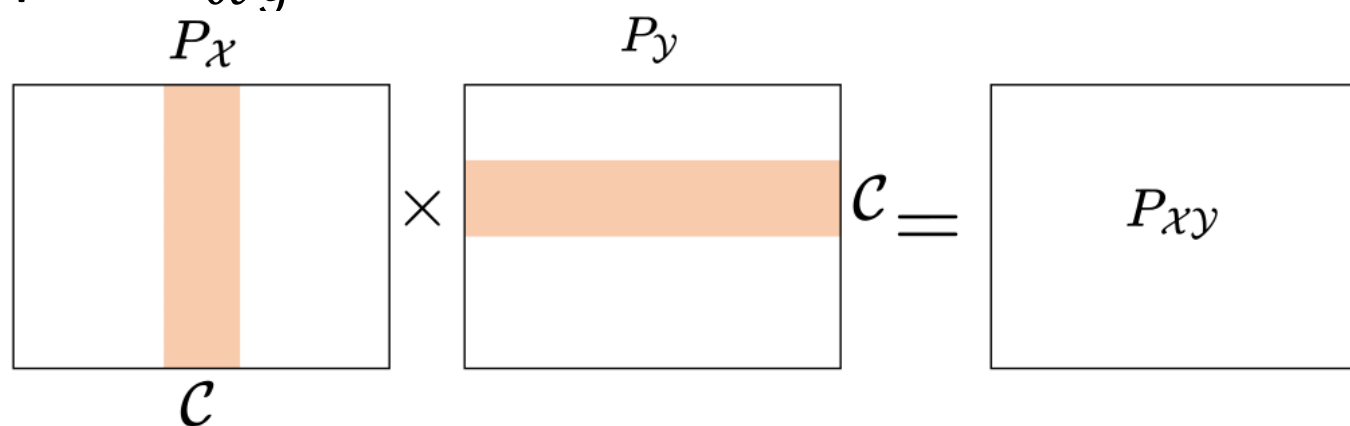


# DTA Steps



Let  $\mathcal{C}$  = set of points with known correspondences in  $\mathcal{X}$  and  $\mathcal{Y}$ . Then

4. Compute  $P_{\mathcal{X}\mathcal{Y}}$ :



5. Compute  $P_{\mathcal{Y}\mathcal{X}}$  similarly

- These represent transition probabilities from observations of one domain to the other
- Known correspondences form a link between domains



# DTA Steps

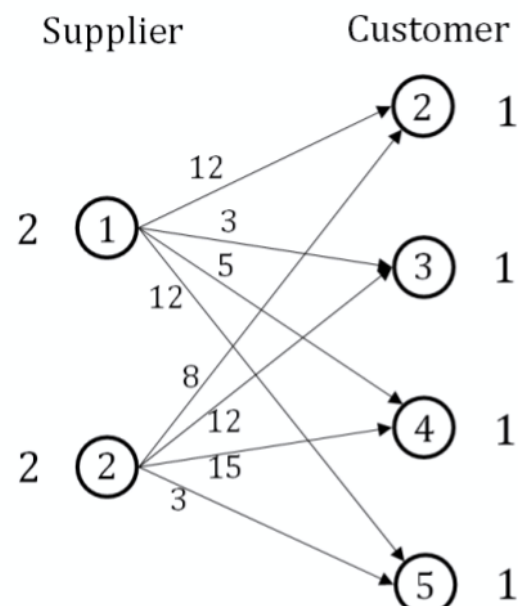


6. Compute inter-domain diffusion distances:

$$D_{ij} = \|P_{\mathcal{X}}(i, :) - P_{\mathcal{Y}\mathcal{X}}(j, :)\| + \|P_{\mathcal{Y}}(j, :) - P_{\mathcal{X}\mathcal{Y}}(i, :)\|$$

7. Assign observations from one domain to the other via an **optimal transport** problem with  $D_{ij}$  forming the “cost”:

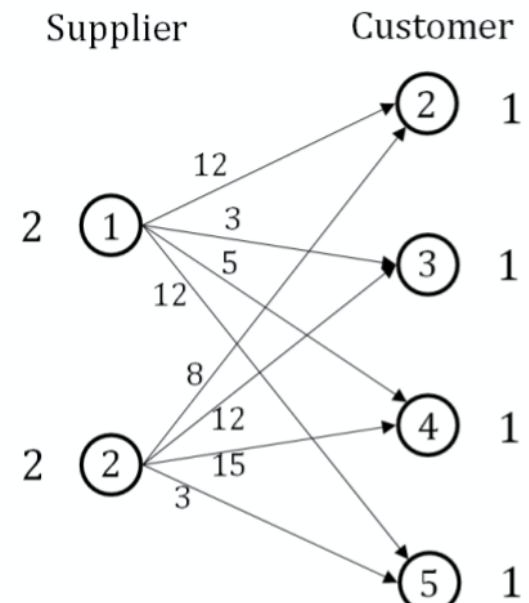
$$\begin{aligned}
& \min_T \sum_{i \in \mathcal{X}, j \in \mathcal{Y}} D_{ij} T_{ij} \\
& \text{s.t.} \quad \sum_{i \in \mathcal{X}} T_{ij} \leq q_j, \quad \forall j \in \mathcal{Y} \\
& \quad \sum_{j \in \mathcal{Y}} T_{ij} \leq v_i, \quad \forall i \in \mathcal{X} \\
& \quad \sum_{i,j} T_{ij} = m \leq \min\left(\sum_j q_j, \sum_i v_i\right) \\
& \quad T_{ij} \geq 0, \quad \forall i \in \mathcal{X}, \forall j \in \mathcal{Y}.
\end{aligned}$$



# DTA Steps



$$\begin{aligned}
 & \min_T \sum_{i \in \mathcal{X}, j \in \mathcal{Y}} D_{ij} T_{ij} \\
 & \text{s.t.} \quad \sum_{i \in \mathcal{X}} T_{ij} \leq q_j, \quad \forall j \in \mathcal{Y} \\
 & \quad \sum_{j \in \mathcal{Y}} T_{ij} \leq v_i, \quad \forall i \in \mathcal{X} \\
 & \quad \sum_{i,j} T_{ij} = m \leq \min\left(\sum_j q_j, \sum_i v_i\right) \\
 & \quad T_{ij} \geq 0, \quad \forall i \in \mathcal{X}, \forall j \in \mathcal{Y}.
 \end{aligned}$$



## 8. Find the correspondences or shared embedding

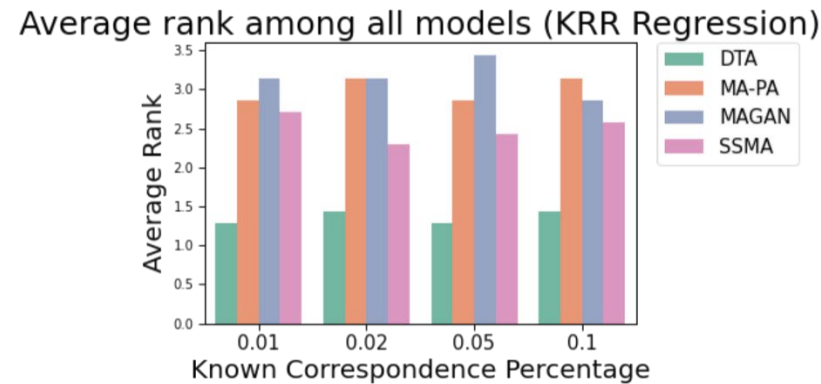
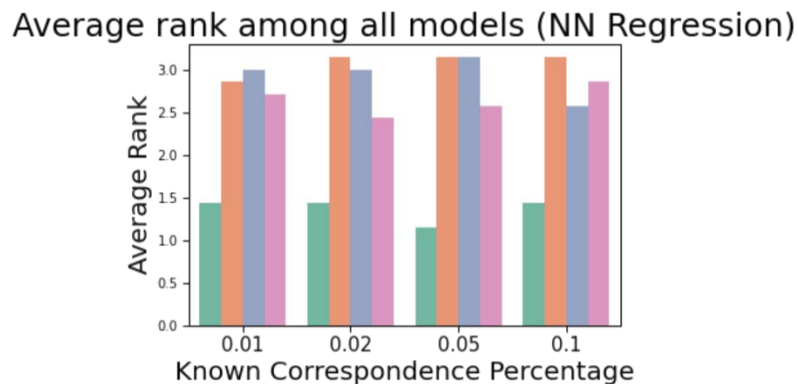
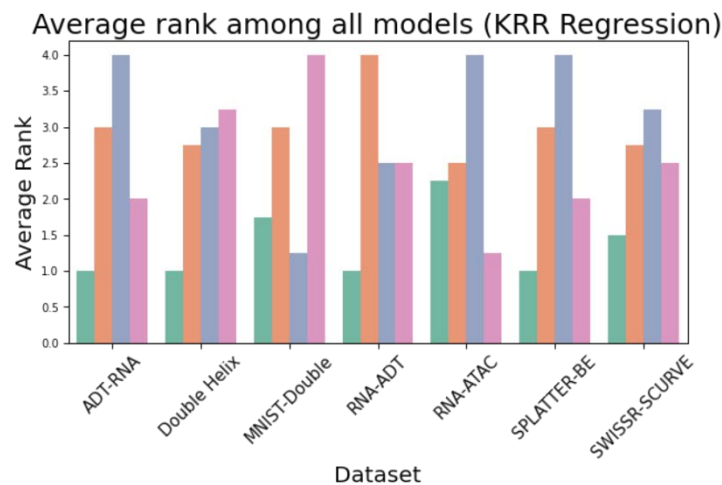
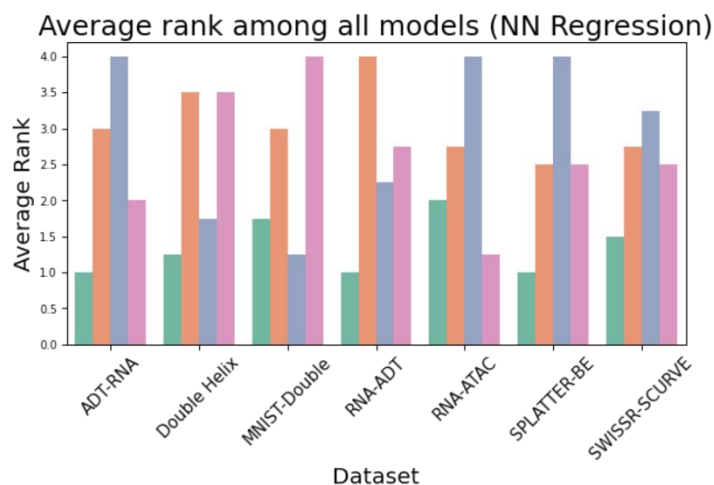
1. Map points from  $\mathcal{X}$  to  $\mathcal{Y}$  by barycentric projection  $x_i \rightarrow \sum_j \tilde{T}_{ij} y_j$  **OR**
2. Build a similarity matrix and apply a spectral embedding:

$$K = \begin{bmatrix} \mu K_{\mathcal{X}} & (1 - \mu)(K_{\mathcal{X}} T + T K_{\mathcal{Y}}) \\ (1 - \mu)(T^T K_{\mathcal{X}}^T + K_{\mathcal{Y}}^T T^T) & \mu K_{\mathcal{Y}} \end{bmatrix}$$

# DTA Regression Experiment



- Perform regression to map from one domain to the other
  - Use aligned data as labels
  - Compared with other methods



# DTA Domain Adaptation Experiment



- Train a classification model on  $\mathcal{Y}$  and test on  $\mathcal{X}$  after alignment
- 1-nearest neighbor accuracy:

Dataset	Model	1%	2%	5%
ADT-RNA	<b>DTA</b>	<b>0.672</b>	<b>0.712</b>	<b>0.719</b>
	MA-PA	0.539	0.606	0.622
	MAGAN	0.473	0.542	0.598
	SSMA	0.633	0.670	0.711
RNA-ADT	<b>DTA</b>	<b>0.681</b>	<b>0.700</b>	<b>0.727</b>
	MA-PA	0.563	0.625	0.660
	MAGAN	0.605	0.653	0.696
	SSMA	0.649	0.670	0.703

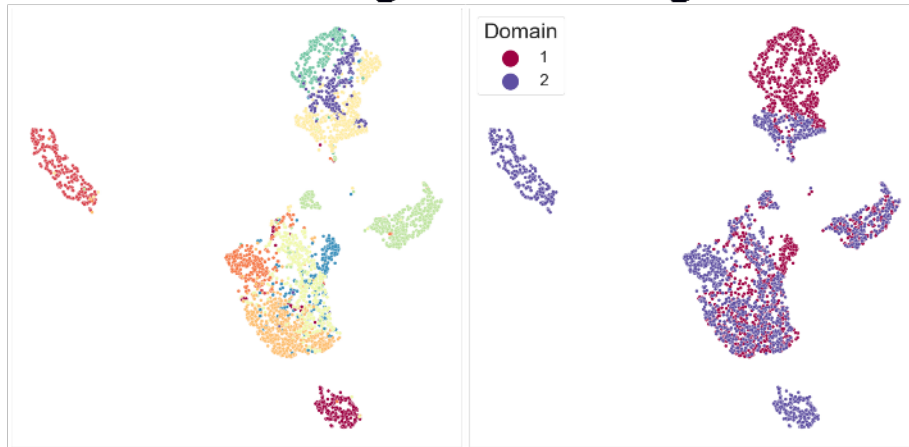
# DTA Partial Alignment



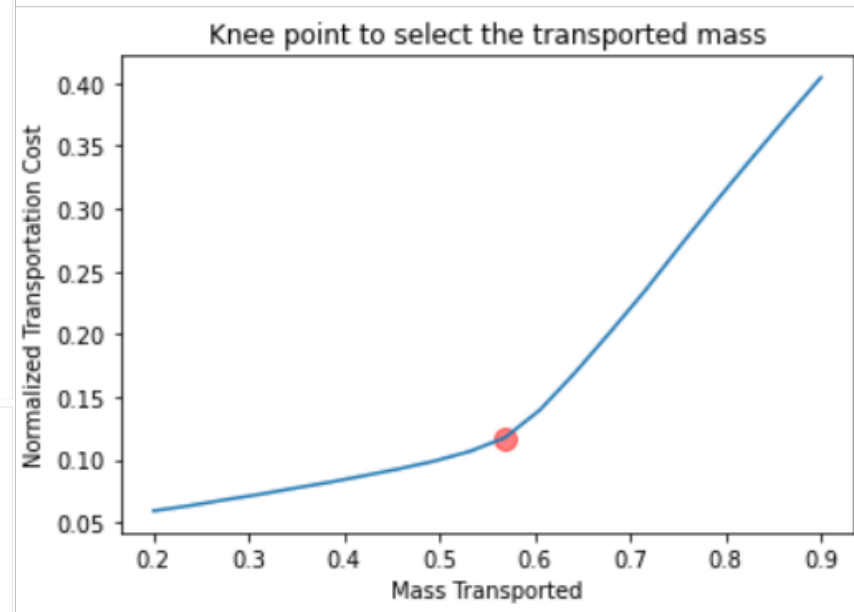
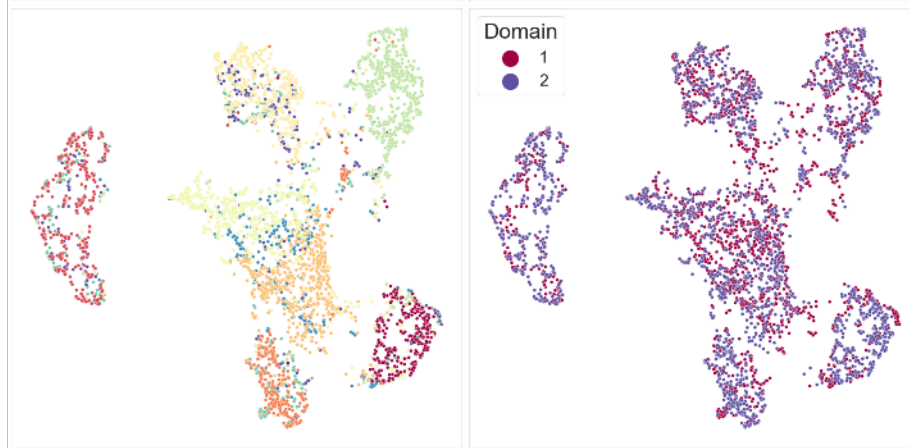
- Some data may be domain-specific
- Choose the amount of mass to transport via a knee plot

Embeddings after alignment

DTA



MAGAN



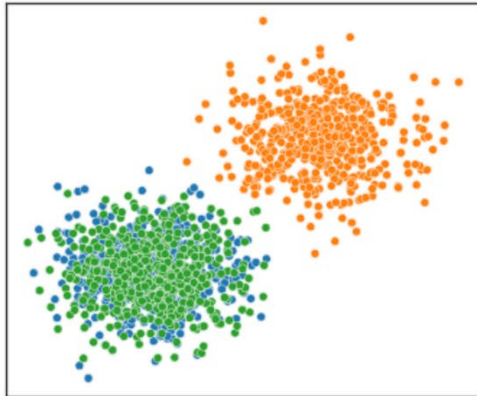
# DTA Alignment for Downstream Analysis



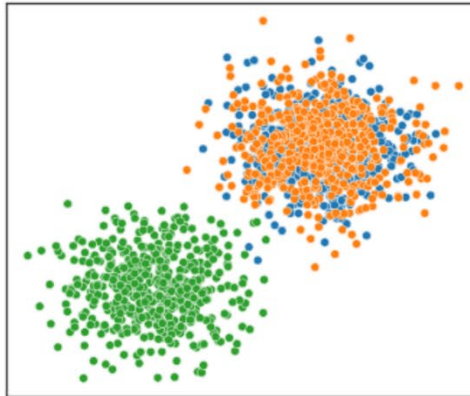
- Find an alignment that improves performance after concatenation based on learned correspondences

Overlapping  
Gaussians

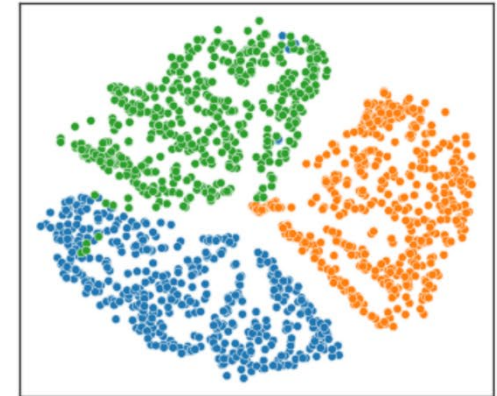
Domain 1  
classification acc.: 0.653



Domain 2  
classification acc.: 0.661

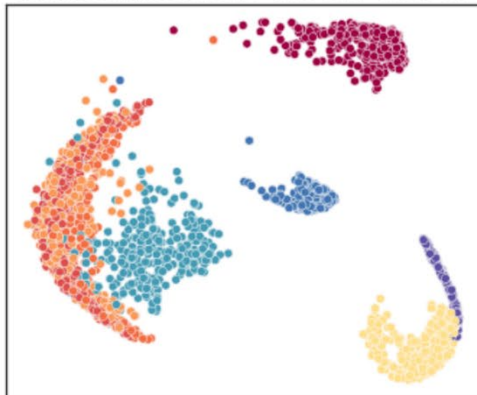


Concatenation after DTA  
classification acc.: 0.992

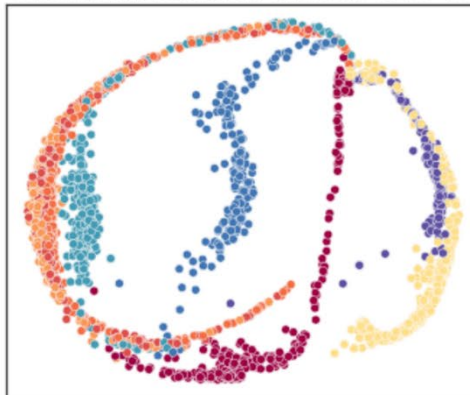


RNA-ATAC

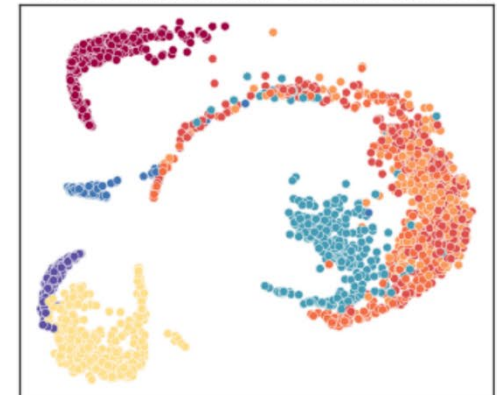
Domain 1  
classification acc.: 0.932



Domain 2  
classification acc.: 0.910



Concatenation after DTA  
classification acc.: 0.960

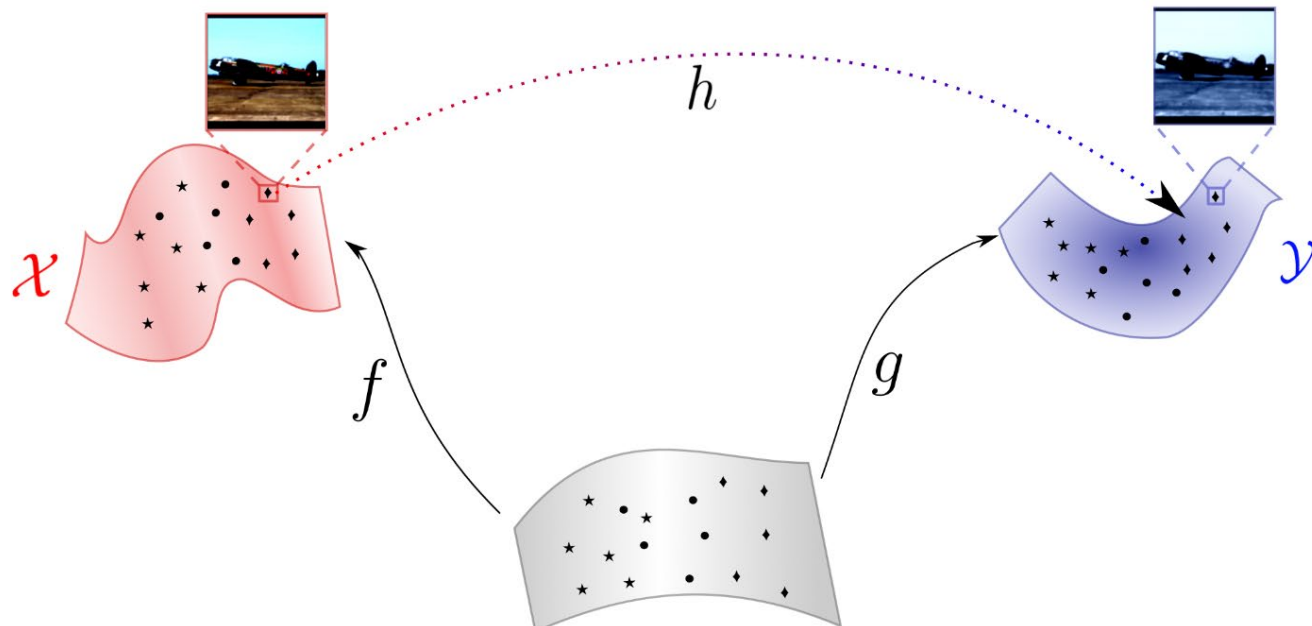




# Manifold Alignment with Label Information (MALI)



- In many cases, we can't obtain any correspondences
  - E.g. many biological measurements destroy the cells
- **Case 2:** Some coarse-grained information (e.g. class labels) is available in both domains
- **Inherent Assumption:** class relationships and points' relationships to their class are similar in both domains





## 1. Intra-domain similarities

- a) Construct domain-specific diffusion operators  $P_{\mathcal{X}}$  and  $P_{\mathcal{Y}}$
- b) Aggregate probabilities from every possible  $t$ -step random walk

$$M_{\mathcal{X}} = \sum_{t=1}^{\infty} (P_{\mathcal{X}} - \mathbf{1}\phi_0^T)^t = \left( I - (P_{\mathcal{X}} - \mathbf{1}\phi_0^T) \right)^{-1} - I$$

- $\phi_0$  is the stationary distribution of  $P_{\mathcal{X}}$
- $M_{\mathcal{Y}}$  is defined similarly
- Used to construct Diffusion pseudotime

## 2. Inter-domain distances

- a) Aggregate similarities within each class  $c$ :

$$M_{\mathcal{X}}^l(i, c) = \frac{1}{p_c^{\mathcal{X}}} \sum_{j \in I_c^{\mathcal{X}}} M_{\mathcal{X}}(i, j)$$

- $p_c^{\mathcal{X}}$  = # of points in  $\mathcal{X}$  in class  $c$
- $I_c^{\mathcal{X}}$  = set of points in  $\mathcal{X}$  in class  $c$



## 2. Inter-domain distances

a) Aggregate similarities within each class  $c$ :

$$M_{\mathcal{X}}^l(i, c) = \frac{1}{p_c^{\mathcal{X}}} \sum_{j \in I_c^{\mathcal{X}}} M_{\mathcal{X}}(i, j)$$

- $p_c^{\mathcal{X}}$  = # of points in  $\mathcal{X}$  in class  $c$
- $I_c^{\mathcal{X}}$  = set of points in  $\mathcal{X}$  in class  $c$
- Repeat for  $M_{\mathcal{Y}}^l(i, c)$

b) Compute inter-domain cosine distances between  $x_i \in \mathcal{X}$  and  $y_j \in \mathcal{Y}$ :

$$D_{ij} = \left( 1 - \frac{\langle M_{\mathcal{X}}^l(i, :), M_{\mathcal{Y}}^l(j, :) \rangle}{\|M_{\mathcal{X}}^l(i, :)\| \|M_{\mathcal{Y}}^l(j, :)\|} \right).$$

3. Match observations via optimal transport

4. Find common representation via barycentric projection or spectral embedding

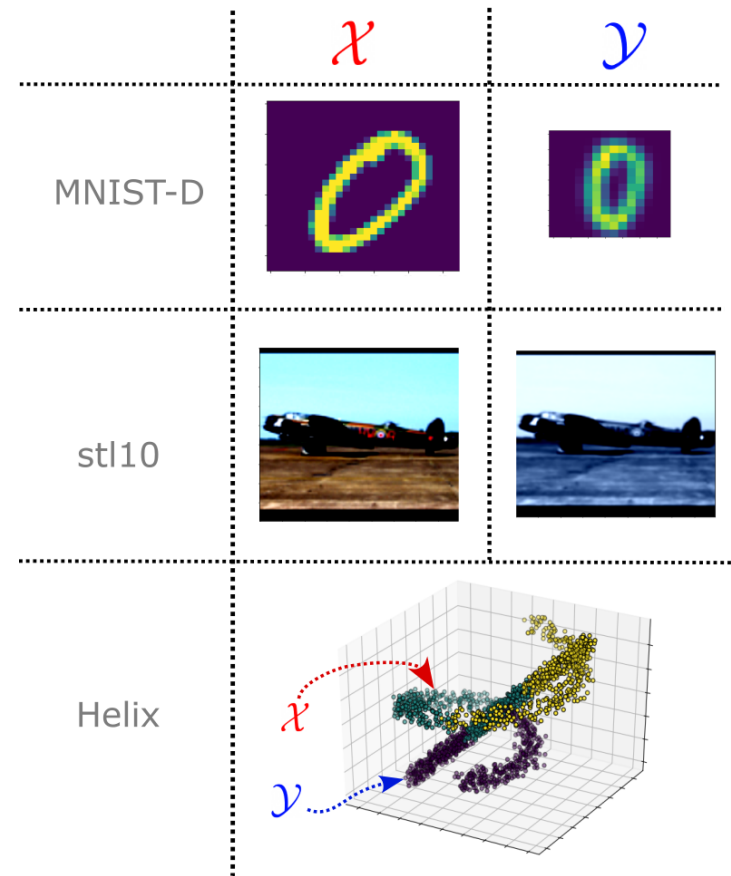
# Performance Metrics



- Good alignment should 1) map matching observations close to each other and 2) be useful for classifying unlabeled target observations (domain adaptation)

## Metrics

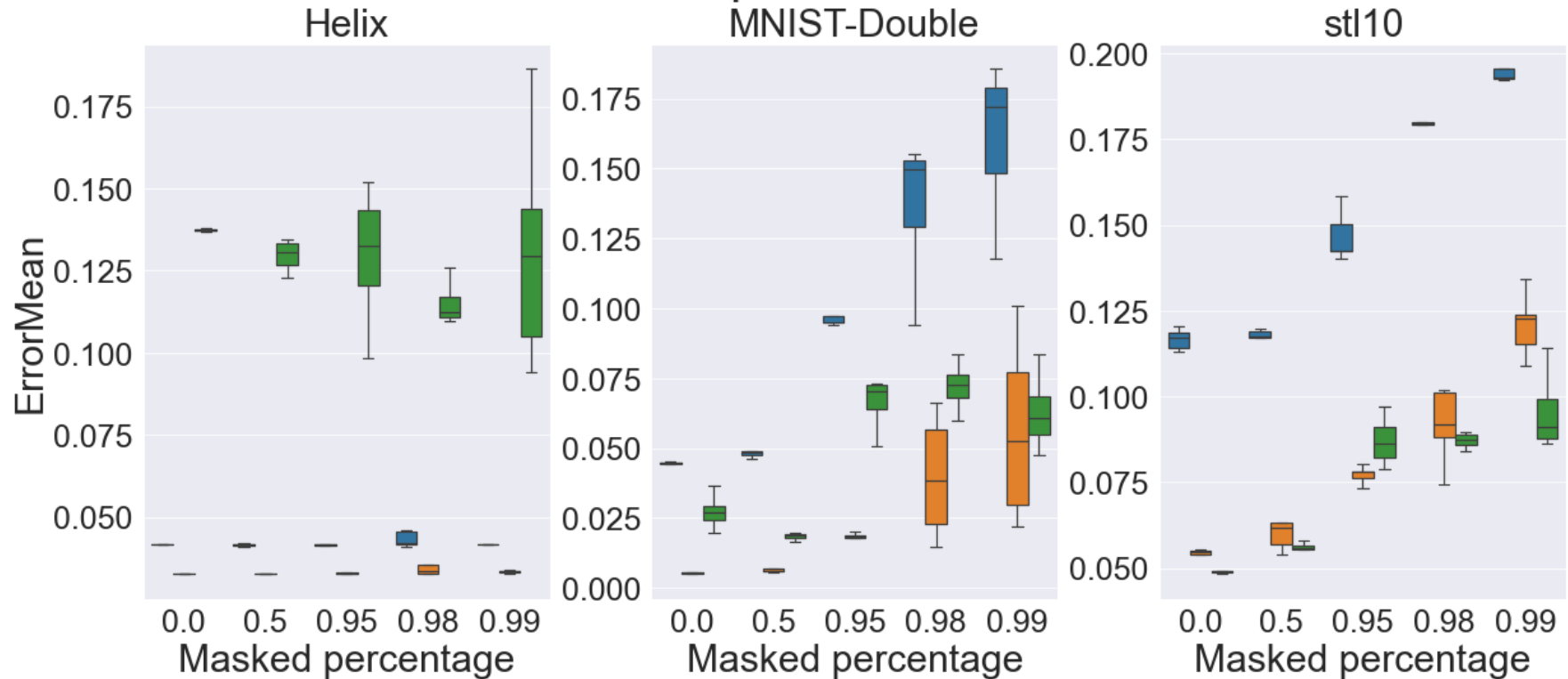
1. Fraction of samples close to the true match (FOSCTTM)
  - How many samples are close to the true match after alignment
  - Lower is better
2. Label Accuracy
  - Accuracy of a 1-NN classifier after alignment
  - Higher is better



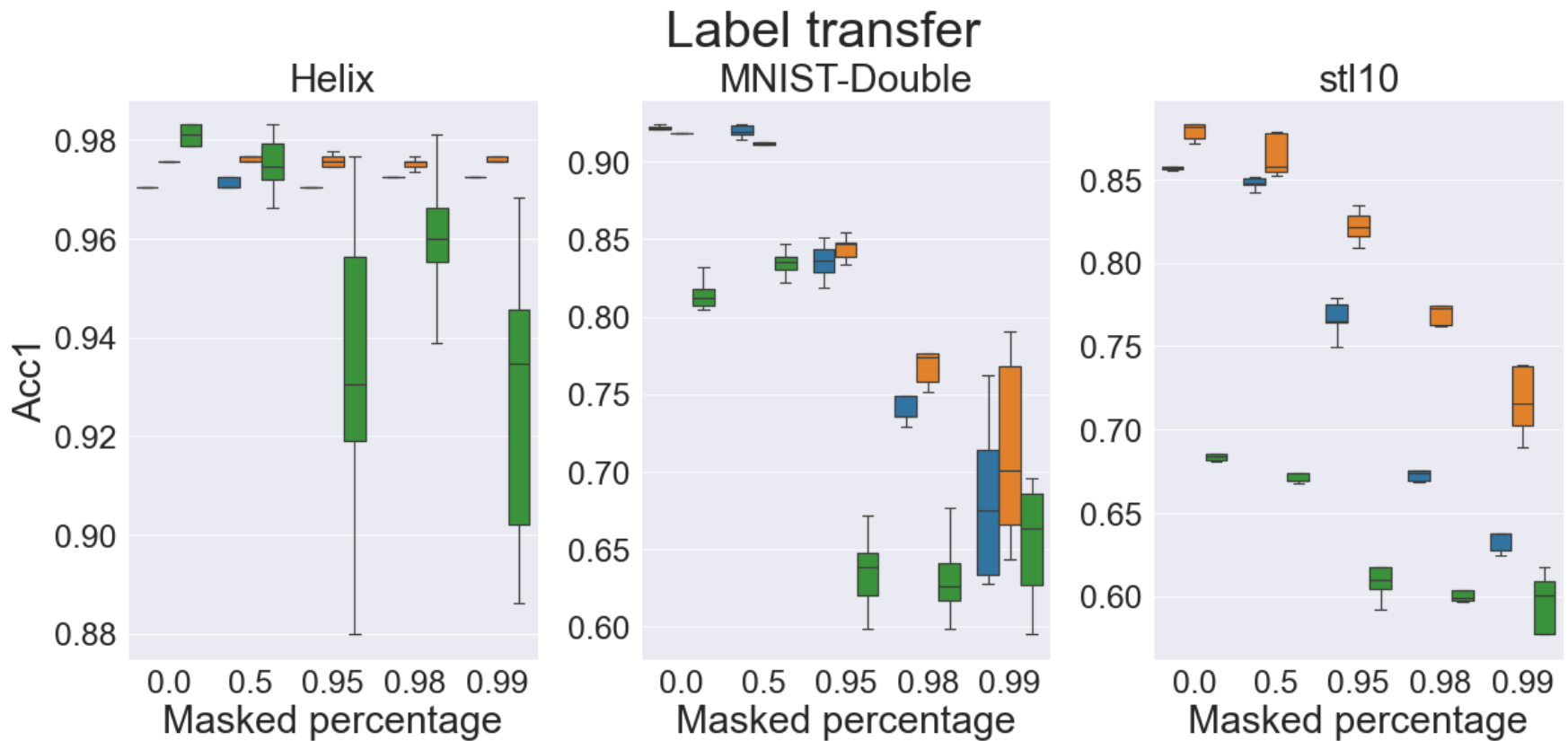
# MALI Experimental Results



Fraction of samples closer to true match



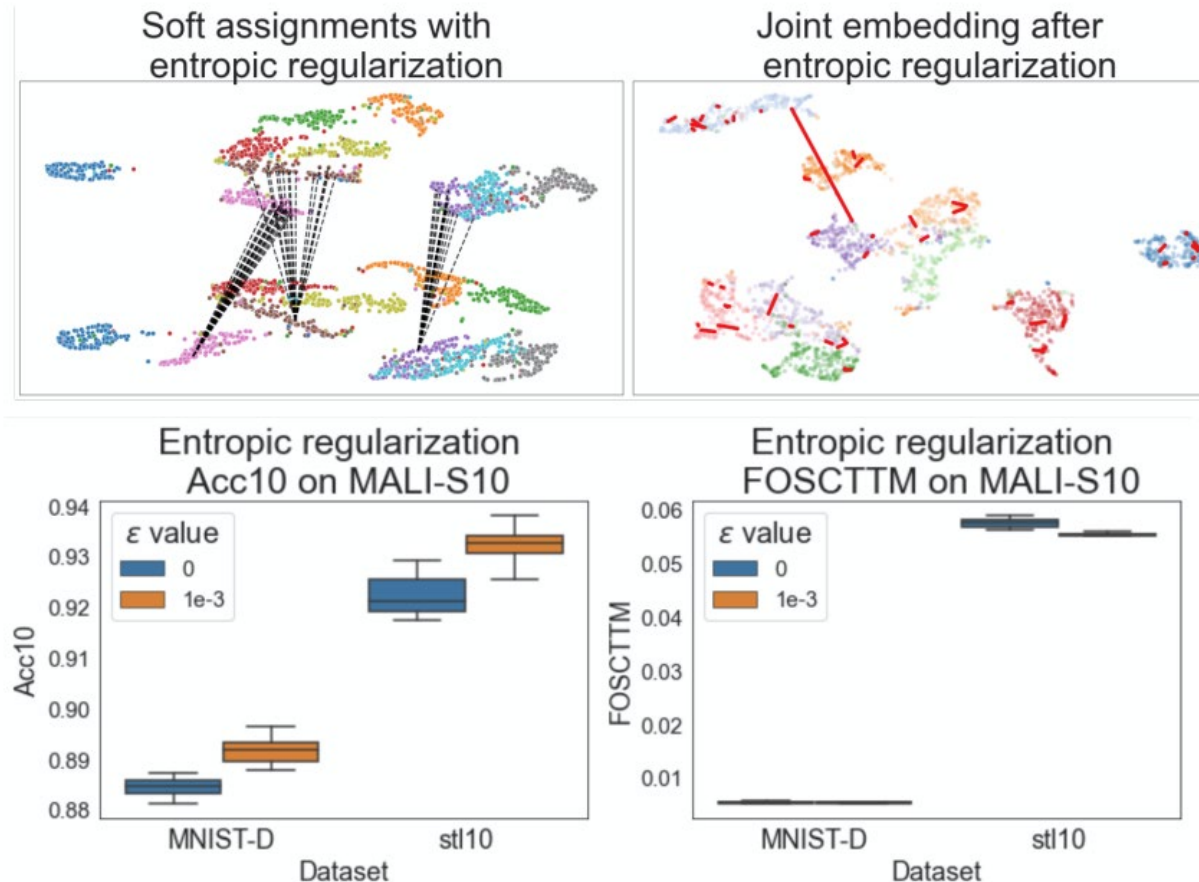
# MALI Experimental Results



# MALI Soft Assignments



- Different sampling density can make it difficult to align data
- Entropic





- Many data fusion/manifold alignment problems fall in the semi-supervised regime
- DTA and MALI are both effective in their respective problems

Future work:

- Explore how to use other side information besides class labels
  - E.g. graph structure, regression labels, etc.
- Combine DTA and MALI into a single framework
- Extend to neural data



# Further Reading



- DTA: [https://doi.org/10.1007/978-3-031-30047-9\\_10](https://doi.org/10.1007/978-3-031-30047-9_10)
- MALI:  
<https://doi.org/10.1109/SampTA59647.2023.10301369>

# Final Class Remarks



- We have scratched the surface of machine learning
- Topics we didn't cover much or at all:
  - Neural networks, reinforcement learning, semi-supervised learning, many other methods of dimensionality reduction, many other methods of clustering, graphical models, sparsity, very high dimensional problems, word embeddings, image processing, Gaussian processes, nonparametric Bayesian methods, statistical learning theory, matrix factorization, multitask learning, and others
- However, we've covered some of the most relevant tools and principles that should help you to understand most of these methods when studying them on your own
  - Thus you'll be less likely to get replaced when machine learning figures out how to do itself!

# Further reading in ML



- The ESL book and other books
- **Relevant Journals:** *Journal of Machine Learning Research* (JMLR), *IEEE Transactions on Pattern Analysis and Machine Intelligence* (TPAMI)
- **Top ML conferences:** NIPS (recently changed to NeurIPS), ICML, AISTATS; more recently ICLR and KDD
- **Top computer vision conferences** (you'll find a lot of ML there): CVPR and ECCV
- You'll also find ML papers in other journals and conferences in **statistics, signal processing/information theory, and computer science**