

Amortized Optimal Transport from Sliced Potentials

Minh-Phuc Truong[2], Khai Nguyen[1]

[1]Texas A&M University

[2]Hanoi Univeristy of Science and Technology





HUST

hust.edu.vn fb.com/dhbkhn

CONTENTS

01 Introduction

02 Motivation

03 Methodology

04 Experiments & Results

05 Conclusion

1. INTRODUCTION: Optimal Transport

- Optimal transport (OT) compares distributions through an optimal transportation plan.
- OT is widely used in generative modeling, domain adaptation, point clouds, and image processing.
- The transport plan describes how mass moves from one distribution to another.
- Computing this plan can be expensive, especially for large discrete problems.

1. INTRODUCTION: Repeated OT

- Many machine learning tasks require solving OT for many different pairs of distributions.
- Examples include mini-batch generative models, dataset comparison, and point-cloud learning.
- Solving each OT problem independently leads to repeated computational cost.
- Even with Sinkhorn, a new optimization must still be performed for every new pair.
- This motivates learning from previously solved OT problems.

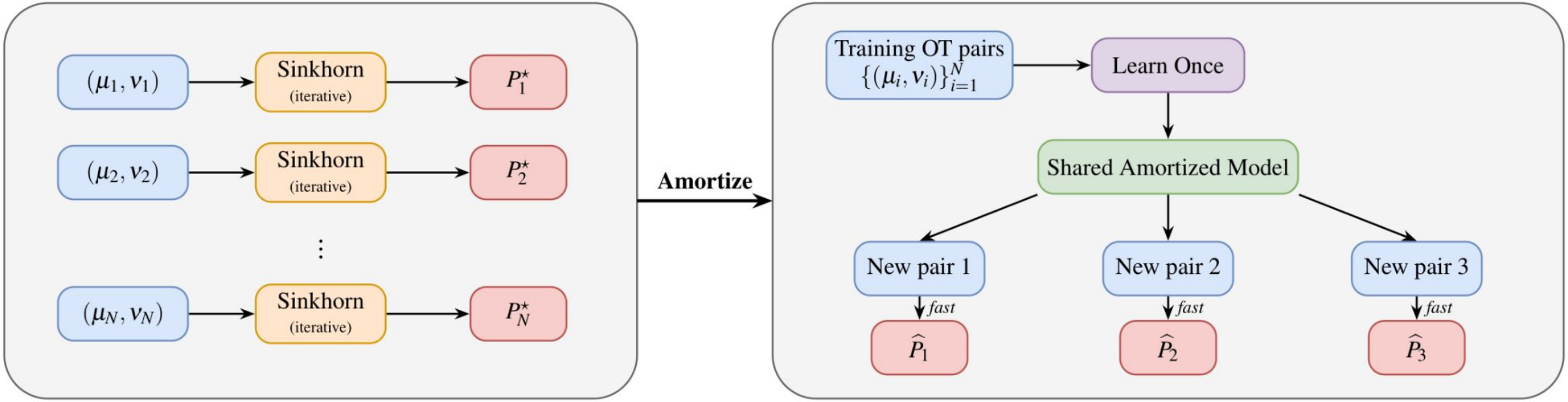
1. INTRODUCTION: Amortized OT

- Assume that OT problems are drawn from a meta-distribution D
- An amortized model learns from observed problems and predicts solutions for future ones.
- Instead of predicting the full plan, we predict one Kantorovich potential.
- The transport plan can then be recovered from the predicted potential.
- Meta-OT follows this idea using a neural amortized model.

Takeaway:

Amortizing Kantorovich potentials provides a practical way to predict OT plan

1. INTRODUCTION: Amortized OT



Repeated OT Solving

Solve a new OT problem for every new pair

Amortized OT

Reuse learned structure across future OT problems

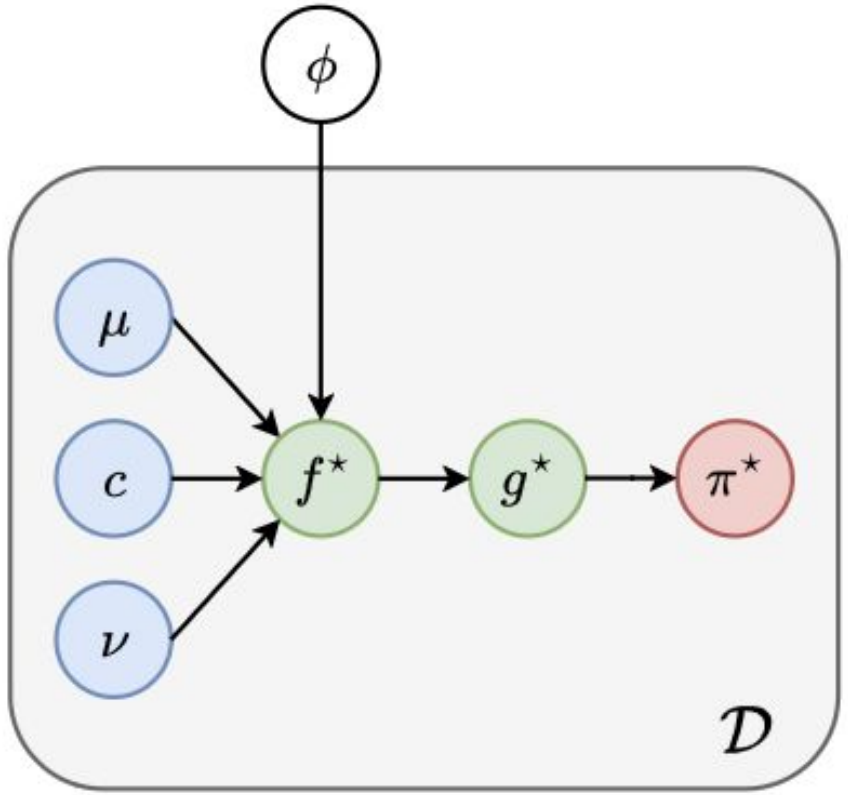
Input OT pairs
Sinkhorn solver
Learning step
Amortized model
Transport plan

2. MOTIVATION

2.1 Limitation Meta-OT

- Meta-OT predicts Kantorovich potentials directly from the raw input measures.
- Its neural model can require many trainable parameters.
- In the discrete setting, the architecture is tied to fixed input and output dimensions.
- This makes varying numbers of atoms difficult to handle with the same model.
- We therefore seek a smaller and more structured representation.

Can we replace raw measures with informative transport features?



Meta-OT

2. MOTIVATION

2.2 Sliced OT

- Sliced OT projects the original measures onto one-dimensional spaces.
- In one dimension, OT admits an efficient closed-form solution.

$$\pi^\star = (F_\mu^{-1}, F_\nu^{-1}) \sharp \mathcal{U}([0,1]),$$

- Each projection provides a corresponding sliced Kantorovich potential.
- Multiple projections capture transport information from different directions.
- These sliced potentials are inexpensive to compute and informative.

Takeaway:

Sliced potentials provide cheap and structured transport features.

2. MOTIVATION

2.3 Core Idea

Rethinking the Role of Sliced OT

- Sliced OT is traditionally used as a fast approximation to high-dimensional OT.
- Existing methods construct transport plans from projected 1D OT problems.
- However, they do not aim to recover the **original OT solution**.

Sliced Kantorovich Potentials as Basis Functions

- In amortized OT, measure pairs share structure through a common meta-distribution.
- Sliced Kantorovich potentials are cheap to compute and encode useful transport information.
- **Key idea:** use sliced potentials as basis functions to reconstruct the original Kantorovich potential.

2. MOTIVATION

2.3 Core Idea

From Approximation to Representation

- We view sliced OT as more than a computational shortcut.
- Instead, sliced potentials form a representation space for learning OT solutions.
- **Our approach is the first attempt to reconstruct the original OT plan from sliced OT.**

Takeaway:
Use sliced OT not to approximate each problem independently, but to learn a reusable representation of OT solutions.

3. METHODOLOGY

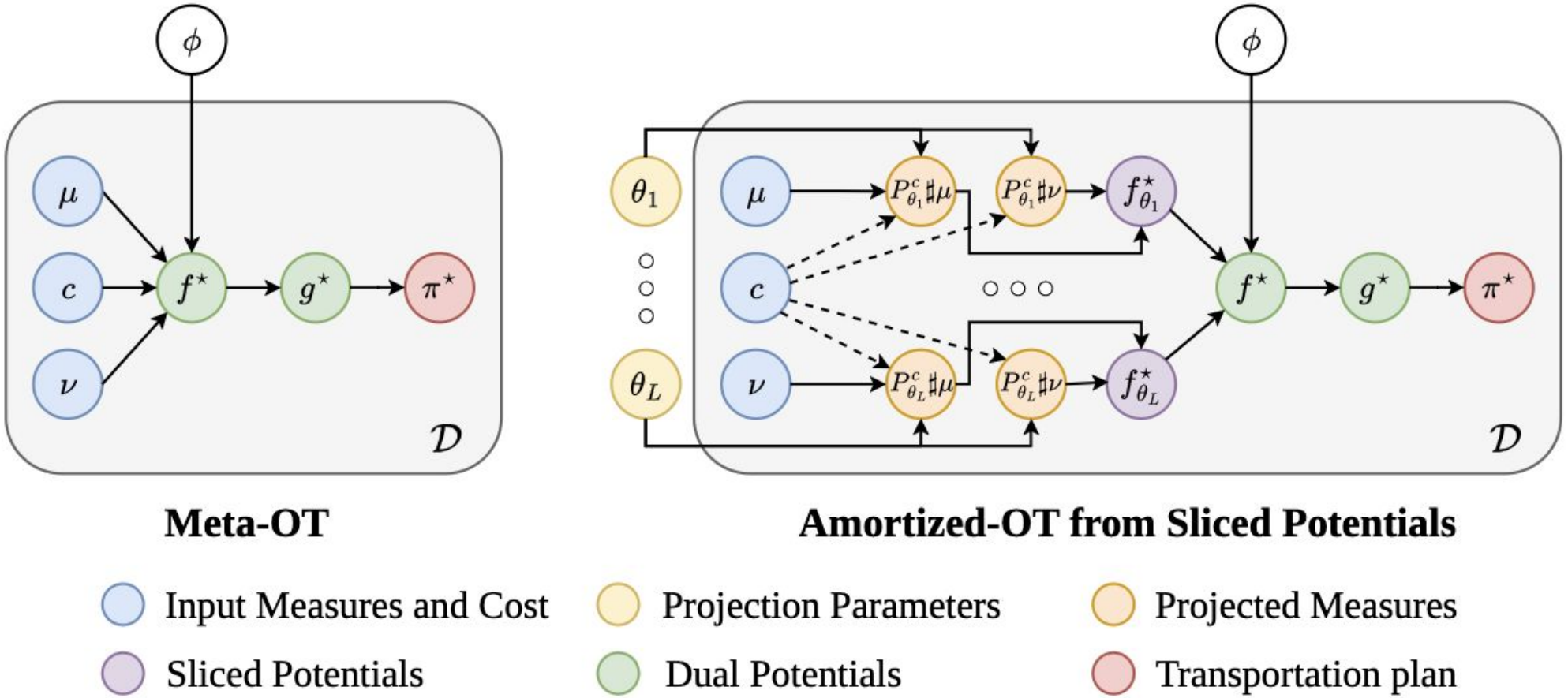


Figure 1: Amortized-OT uses sliced potentials to predict original dual potentials.

Amortized Model from Sliced Potentials

Common Philosophy

- Use sliced Kantorovich potentials as a compact representation of an OT problem.
- Each sliced potential captures transport information along one projection direction.
- Combine information from multiple slices to predict the original Kantorovich potential.

Amortized Model

- Given (L) projection directions, we first compute (L) sliced Kantorovich potentials.
- In general, these potentials can be combined by an operator.
- We use a simple linear combination as the amortized model.

$$\hat{f}_{\boldsymbol{\omega}}[\mu, \nu, c](x) = \sum_{l=1}^{L} \omega_l f^{\star}_{\theta_l}[\mu, \nu, c](P^{c}_{\theta_l}(x))$$

Amortized Model from Sliced Potentials

Why a Linear Model?

- The sliced potentials already provide informative transport features.
- The model therefore only needs to learn how to combine these features.
- Only (L) coefficients are learned, independent of the number of input atoms.
- Cheap 1D OT computation makes the overall model lightweight and computationally efficient.

Takeaway:

Learn a small set of coefficients instead of a large model directly from raw measures.

3. METHODOLOGY

RA-OT

- Treat the original Kantorovich potential as the regression target.
- Use sliced Kantorovich potentials as the regression features.
- Learn a shared coefficient vector across OT problems.

$$\min_{\boldsymbol{\omega}\in\mathbb{R}^L} \mathbb{E}_{(\mu,\nu,c)\sim\mathcal{D}}[\|\hat{f}_{\boldsymbol{\omega}}[\mu,\nu,c] - f^\star[\mu,\nu,c]\|_2^2],$$

Discrete Case

- For discrete measures, the functional regression becomes ordinary linear least squares.
- The optimal coefficients can be obtained from the corresponding normal equation.
- In practice, a standard linear solver can be used for numerical stability.

$$\boldsymbol{\omega}^\star = \left(\mathbb{E}[X_\Theta[\mu,\nu,c]^\top X_\Theta[\mu,\nu,c]]\right)^{-1} \mathbb{E}[X_\Theta[\mu,\nu,c]^\top Y[\mu,\nu,c]],$$

3. METHODOLOGY

RA-OT

- RA-OT is simple and does not require iterative neural-network training.
- Training requires ground-truth OT potentials from solved OT problems.
- Once trained, prediction only requires sliced potentials and a linear combination.

Takeaway:

RA-OT learns how to reconstruct OT potentials from sliced potentials through functional regression.

3. METHODOLOGY

OA-OT

- OA-OT uses the same sliced-potential representation as RA-OT.
- Instead of regressing onto ground-truth potentials, it optimizes the Kantorovich dual objective directly.
- The coefficients are learned using gradient-based optimization.

$$\min_{\omega \in \mathbb{R}^L} \mathbb{E}_{(\mu,\nu,c)\sim\mathcal{D}}[J(\hat{f}_\omega[\mu,\nu,c];\mu,\nu,c)],$$

Connection to Meta-OT

- OA-OT follows the same objective-based amortization principle as Meta-OT.
- The main difference lies in the representation used by the amortized model.
- Meta-OT learns from raw measures, while OA-OT learns from sliced OT potentials.
- As a result, OA-OT uses far fewer trainable parameters.

3. METHODOLOGY

OA-OT

- No regression onto ground-truth Kantorovich potentials is required.
- OA-OT is designed for the entropic OT setting.
- The learned model remains independent of the number of input atoms.

Takeaway:

OA-OT keeps Meta-OT's training objective but replaces its large raw-input model with a lightweight sliced-potential model.

From predicted potentials to OT plan

Recovering the Transport Plan

- Both RA-OT and OA-OT predict one Kantorovich potential.
- For entropic OT, the second potential can be obtained from the first using a Sinkhorn update.
- The transport plan is then reconstructed from the two potentials.

$$(\mathbf{f}^\star, \mathbf{g}^\star) \in \underset{\mathbf{f}\in\mathbb{R}^n, \mathbf{g}\in\mathbb{R}^m}{\arg\max} \langle \mathbf{f}, \boldsymbol{\alpha}\rangle + \langle \mathbf{g}, \boldsymbol{\beta}\rangle - \epsilon \sum_{i=1}^{n}\sum_{j=1}^{m} \exp\left(\frac{\mathbf{f}_i + \mathbf{g}_j - C_{ij}}{\epsilon}\right). \tag{6}$$

We can recover a discrete plan from any pairs $(\mathbf{f}, \mathbf{g})$ as follows:

$$P^\star_{ij} = \exp\left(\frac{\mathbf{f}^\star_i + \mathbf{g}^\star_j - C_{ij}}{\epsilon}\right). \tag{7}$$

The mapping between the dual can be rewritten as:

$$\mathbf{g} = \epsilon \log \boldsymbol{\alpha} - \epsilon \log\left(\exp(-C^\top/\epsilon)\exp(\mathbf{f}/\epsilon)\right), \tag{8}$$

$$\mathbf{f} = \epsilon \log \boldsymbol{\beta} - \epsilon \log\left(\exp(-C/\epsilon)\exp(\mathbf{g}/\epsilon)\right). \tag{9}$$

3. METHODOLOGY

Key Properties

- **Lightweight:** only (L) coefficients are learned in the linear model.
- **Flexible:** model parameters do not depend on the number of atoms.
- **Reusable:** the same coefficients are shared across different OT instances.
- **Structured:** sliced potentials provide transport-aware features before learning begins.
- **Generalizable architecture:** the linear operator can be replaced by a richer (\gamma_\phi) when needed.

Overall Pipeline

Input measures → 1D projections → Sliced potentials → RA-OT / OA-OT → Predicted potential → OT plan

Takeaway:
The learned model is small; the transport structure comes from sliced OT.

3. METHODOLOGY

Meta-OT

- Uses **raw measure representations** as the input to the amortized model.
- In the discrete setting, an MLP predicts a Kantorovich potential from the measure weights.
- The architecture depends on fixed input and output dimensions (requires a relatively large number of learnable parameters)
- Handling different numbers of atoms is not straightforward.

RA-OT / OA-OT

- Use **sliced Kantorovich potentials** as the input representation (Sliced potentials already encode useful transport information)
- A simple linear combination is sufficient for the amortized model (Only (L) coefficients need to be learned)
- The parameter dimension is independent of the number of atoms.

Meta-OT learns transport structure from raw measures, while our methods learn from structured sliced-OT representations.

Result: fewer parameters, simpler training, and support for varying measure sizes.

4. EXPERIMENTS & RESULTS

Experiment Setup

Evaluation Tasks

- We evaluate RA-OT and OA-OT on three representative discrete OT problems.
- The tasks cover Euclidean, non-Euclidean, and high-dimensional feature spaces.
- We additionally study amortized OT inside mini-batch conditional flow matching.

Standard OT Tasks

- **MNIST Grayscale Transport:** transport between digit images on a (28x28) grid.
- **Spherical Transport:** supply-demand matching on the unit sphere.
- **Color Transfer:** transport between image color distributions in RGB space.

4. EXPERIMENTS & RESULTS

Experiment Setup

Baselines

- **Meta-OT:** neural amortization from raw measures.
- **Min-STP:** learns an optimized projection direction.
- **min-SWGG:** training-free sliced transport baseline.
- **Sinkhorn:** converged entropic OT used as ground truth.

Evaluation

- Plan prediction quality is measured using **RMSE against converged Sinkhorn**.
- We also report **training time** and **per-pair inference time**.
- All sliced methods use (L=100) projections.
- Training sizes vary over (*M* in {10,20,50,200}).

4. EXPERIMENTS & RESULTS: MNIST Grayscale

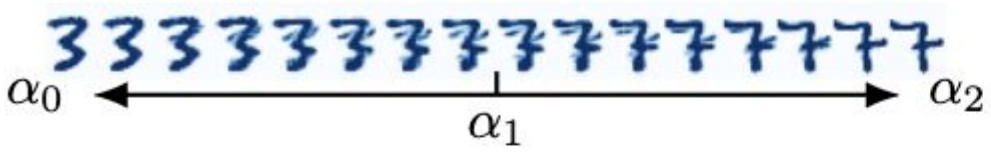
Setting

- Each MNIST image is represented as a discrete measure with (784) fixed grid locations.
- Pixel intensities define the probability weights.
- We use squared Euclidean transport cost with (epsilon=0.1).

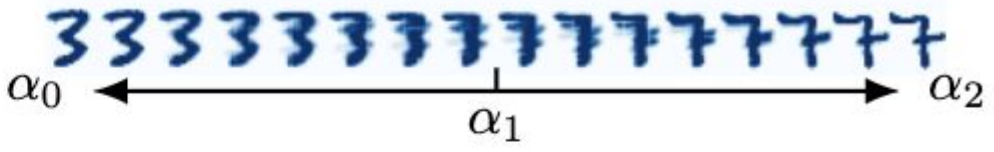
M	Method	RMSE ($\times 10^{-6}$, ↓)	Train (s)	Infer (ms,↓)	M	Method	RMSE ($\times 10^{-6}$, ↓)	Train (s)	Infer (ms,↓)
10	Meta-OT [5]	16.16 ± 4.90	38.05	$\mathbf{2.46 \pm 0.20}$	50	Meta-OT [5]	15.54 ± 4.74	37.11	$\mathbf{2.39 \pm 0.29}$
	Min-STP [40]	93.51 ± 10.18	199.52	6.18 ± 0.37		Min-STP [40]	93.50 ± 10.18	198.01	6.25 ± 0.48
	min-SWGG [42]	89.85 ± 8.64	no train	10.01 ± 0.73		min-SWGG [42]	89.85 ± 8.64	no train	9.72 ± 0.41
	RA-OT	8.23 ± 3.55	**1.36**	44.33 ± 2.91		**RA-OT**	7.77 ± 3.06	**3.03**	39.36 ± 3.61
	OA-OT	$\mathbf{6.16 \pm 2.53}$	16.65	42.54 ± 2.95		**OA-OT**	$\mathbf{6.02 \pm 2.52}$	15.78	38.92 ± 2.23
20	Meta-OT [5]	16.68 ± 4.98	37.27	$\mathbf{2.44 \pm 0.39}$	200	Meta-OT [5]	14.12 ± 4.70	37.22	$\mathbf{2.37 \pm 0.40}$
	Min-STP [40]	93.51 ± 10.18	198.56	6.21 ± 0.41		Min-STP [40]	93.51 ± 10.18	197.86	6.17 ± 0.49
	min-SWGG [42]	89.85 ± 8.64	no train	9.70 ± 0.47		min-SWGG [42]	89.85 ± 8.64	no train	9.73 ± 0.61
	RA-OT	7.49 ± 3.19	**1.44**	40.29 ± 2.74		**RA-OT**	7.74 ± 3.07	**11.12**	40.30 ± 4.39
	OA-OT	$\mathbf{6.04 \pm 2.50}$	15.60	40.65 ± 2.95		**OA-OT**	$\mathbf{6.11 \pm 2.54}$	16.39	40.48 ± 4.16

4. EXPERIMENTS & RESULTS: MNIST Grayscale

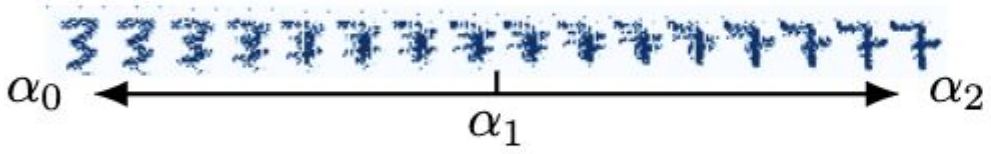
Sinkhorn (converged, ground-truth)



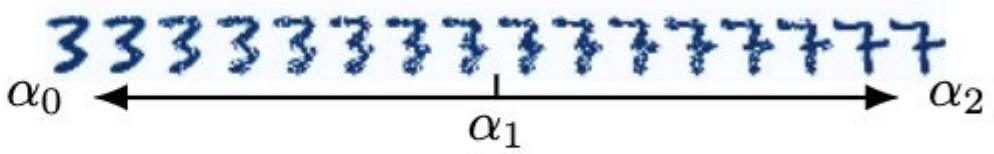
Meta-OT



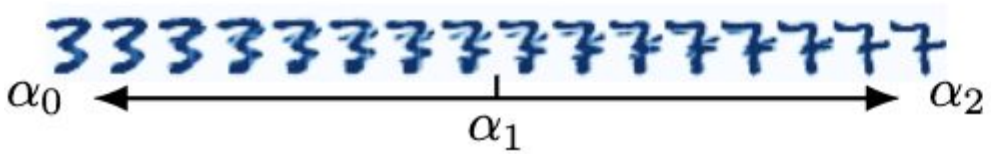
Min-STP



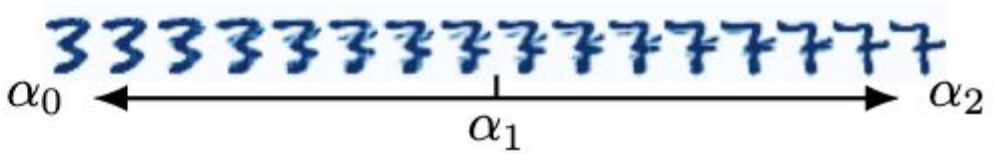
min-SWGG



RA-OT



OA-OT



4. EXPERIMENTS & RESULTS: Supply-Demand Transport

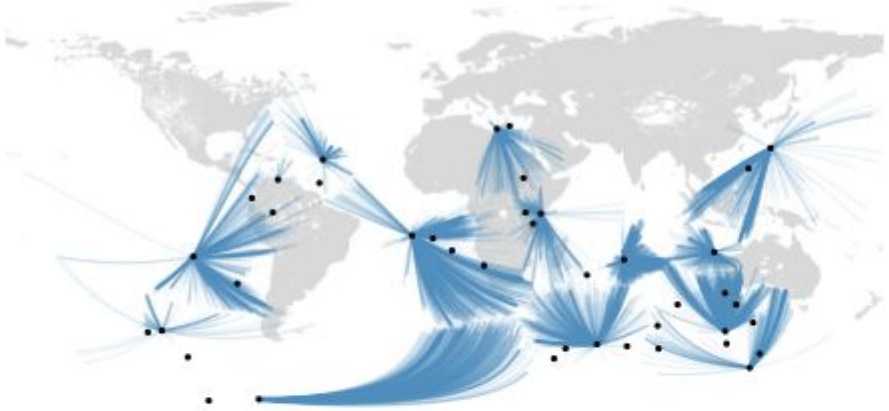
Setting

- Supply is distributed over (n=100) locations on Earth's landmass.Demand is defined over (m=10000) population locations.
- The ground cost is the geodesic distance on the sphere.
- We use stereographic projections to respect the spherical geometry.

M	Method	RMSE ($\times 10^{-7}$, ↓)	Train (s)	Infer (ms,↓)	M	Method	RMSE ($\times 10^{-7}$, ↓)	Train (s)	Infer (ms,↓)
10	Meta-OT [5]	6.83 ± 2.49	52.08	15.94 ± 2.14	50	Meta-OT [5]	4.42 ± 1.55	52.07	16.09 ± 2.17
	Min-STP [40]	115.07 ± 0.54	46.06	$\mathbf{7.51 \pm 7.23}$		Min-STP [40]	115.10 ± 0.54	46.38	$\mathbf{7.88 \pm 1.22}$
	min-SWGG [42]	112.15 ± 0.99	no train	34.62 ± 2.70		min-SWGG [42]	112.15 ± 0.99	no train	33.25 ± 1.71
	RA-OT	7.92 ± 2.77	**1.05**	41.16 ± 3.11		**RA-OT**	7.82 ± 1.89	**2.53**	41.96 ± 6.29
	OA-OT	$\mathbf{4.80 \pm 1.96}$	18.60	39.92 ± 3.12		**OA-OT**	$\mathbf{3.93 \pm 1.92}$	19.37	41.03 ± 2.76
20	Meta-OT [5]	6.03 ± 2.20	51.99	15.57 ± 1.22	200	Meta-OT [5]	$\mathbf{3.62 \pm 1.14}$	52.12	15.17 ± 2.54
	Min-STP [40]	115.06 ± 0.54	46.39	$\mathbf{7.95 \pm 1.41}$		Min-STP [40]	115.05 ± 0.54	46.41	$\mathbf{7.93 \pm 1.22}$
	min-SWGG [42]	112.15 ± 0.99	no train	33.85 ± 2.49		min-SWGG [42]	112.15 ± 0.99	no train	33.59 ± 1.99
	RA-OT	7.82 ± 2.25	**1.31**	40.70 ± 3.25		**RA-OT**	7.55 ± 1.88	**8.74**	39.40 ± 1.43
	OA-OT	$\mathbf{4.24 \pm 1.98}$	18.75	39.65 ± 3.26		**OA-OT**	3.66 ± 1.56	21.32	40.03 ± 3.83

4. EXPERIMENTS & RESULTS: Supply-Demand Transport

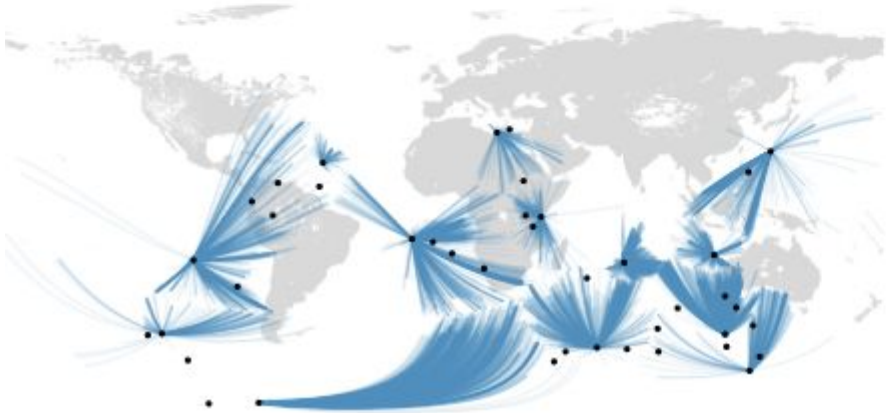
Sinkhorn (converged, ground-truth)



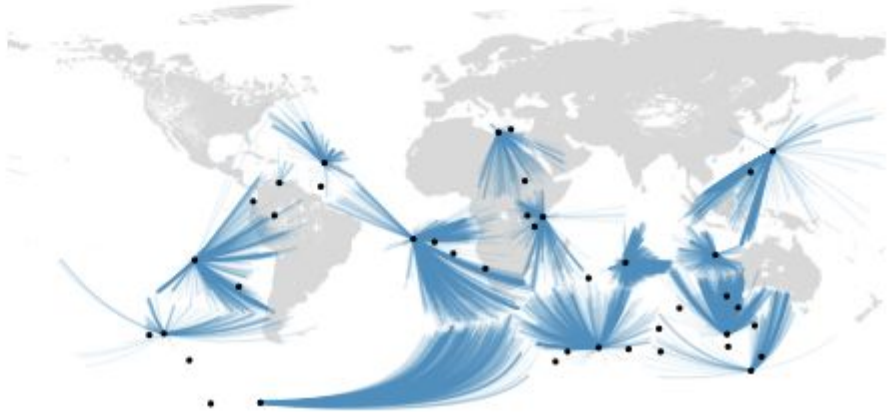
Meta-OT



RA-OT



OA-OT



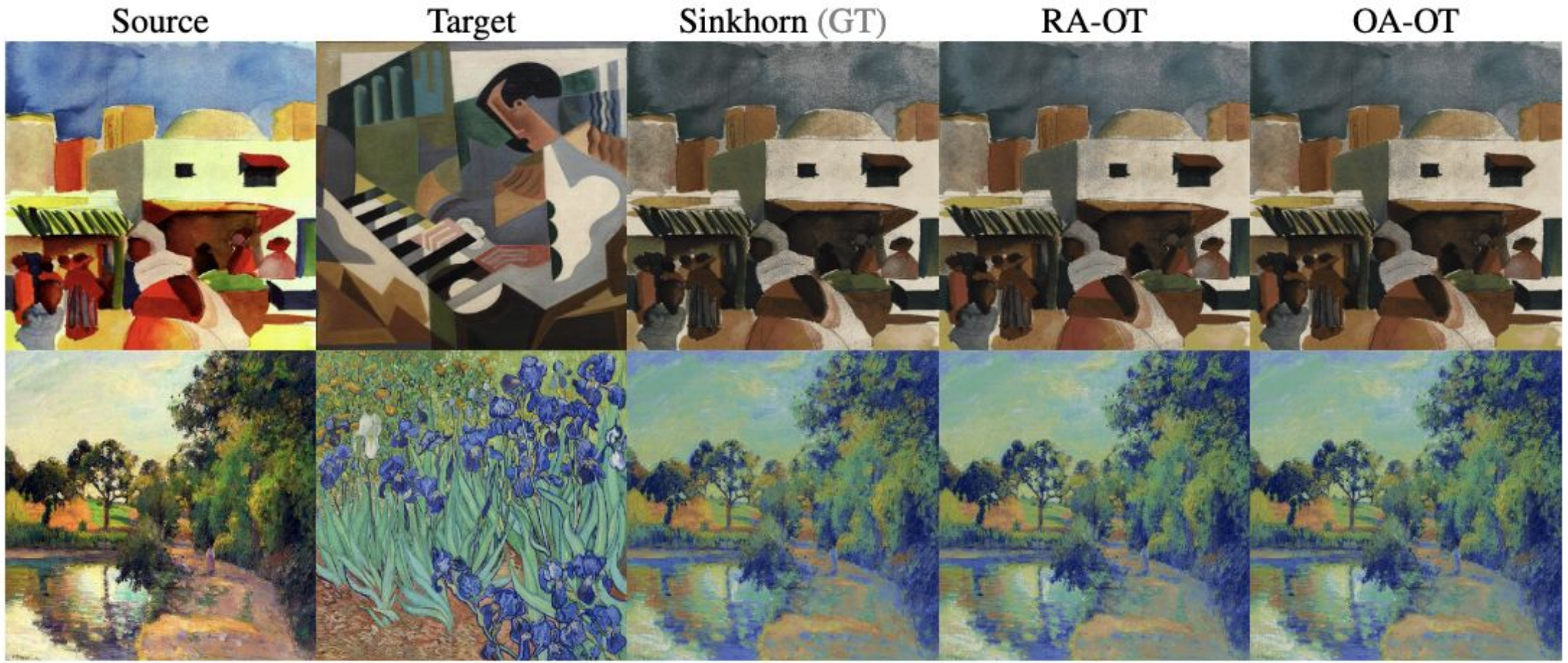
4. EXPERIMENTS & RESULTS: Color transfer

Setting

- Images are represented as weighted point clouds in normalized RGB space.
- Each image is quantized into (K=500) color clusters.
- The transport cost is squared Euclidean distance with (epsilon=0.005).

M	Method	RMSE ($\times 10^{-6}$, ↓)	Train (s)	Infer (ms,↓)	M	Method	RMSE ($\times 10^{-6}$, ↓)	Train (s)	Infer (ms,↓)
10	Meta-OT [5]	36.83 ± 10.97	33.32	16.83 ± 1.33	50	Meta-OT [5]	33.16 ± 10.45	32.91	15.16 ± 0.64
	Min-STP [40]	80.93 ± 3.24	49.51	$\mathbf{4.02 \pm 0.27}$		Min-STP [40]	81.04 ± 3.42	50.00	$\mathbf{4.10 \pm 0.20}$
	min-SWGG [42]	77.69 ± 3.19	no train	4.25 ± 0.21		min-SWGG [42]	77.69 ± 3.19	no train	4.34 ± 0.17
	RA-OT	10.73 ± 6.35	**1.95**	18.41 ± 1.21		**RA-OT**	9.99 ± 5.61	**7.40**	17.40 ± 0.83
	OA-OT	$\mathbf{9.78 \pm 5.78}$	17.52	18.77 ± 1.31		**OA-OT**	$\mathbf{9.00 \pm 5.02}$	18.13	17.76 ± 0.79
20	Meta-OT [5]	34.73 ± 10.94	34.63	15.03 ± 0.70	200	Meta-OT [5]	33.23 ± 10.89	33.11	15.44 ± 1.20
	Min-STP [40]	80.97 ± 3.33	49.92	$\mathbf{4.07 \pm 0.23}$		Min-STP [40]	81.10 ± 3.38	49.68	$\mathbf{4.08 \pm 0.22}$
	min-SWGG [42]	77.69 ± 3.19	no train	4.30 ± 0.17		min-SWGG [42]	77.69 ± 3.19	no train	4.27 ± 0.19
	RA-OT	9.74 ± 5.68	**3.11**	18.02 ± 1.34		**RA-OT**	10.21 ± 5.53	26.57	18.08 ± 1.40
	OA-OT	$\mathbf{9.31 \pm 5.24}$	17.92	18.45 ± 1.20		**OA-OT**	$\mathbf{9.00 \pm 5.05}$	**20.37**	18.25 ± 0.71

4. EXPERIMENTS & RESULTS: Color transfer



4. EXPERIMENTS & RESULTS: Mini-batch OT Flow

Motivation

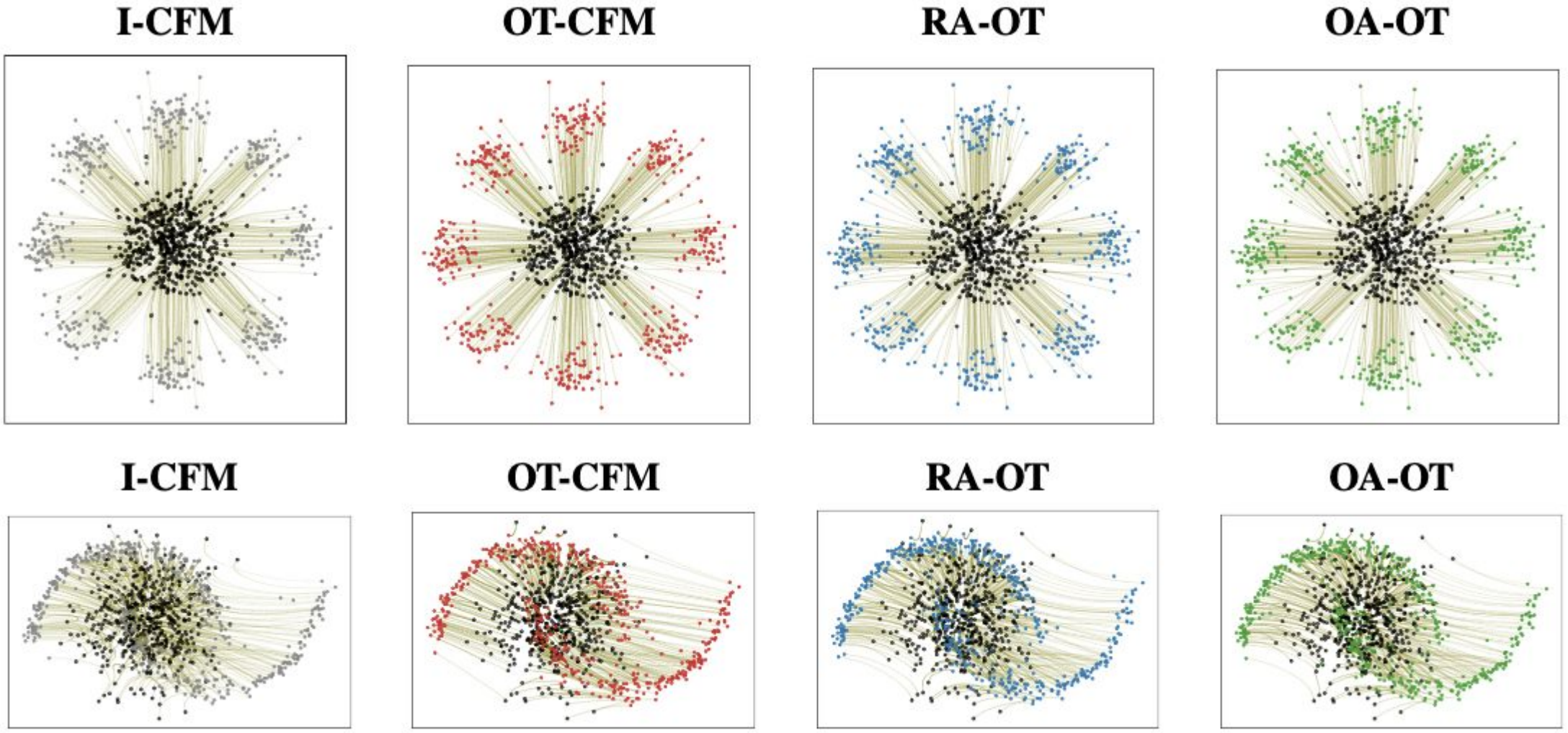
- OT-CFM uses optimal transport couplings to create straighter flow-matching trajectories.
- However, solving mini-batch OT at every training step creates a major computational bottleneck.
- We replace repeated exact OT solves with RA-OT or OA-OT coupling predictions.

Setup

- We consider Gaussian-to-**8-Gaussians**, **Moons**, and **S-Curve** transformations.
- Amortized models are pretrained using (M=50) batches and (L=100) projections.
- We evaluate trajectory straightness using **NPE** and final distribution quality using (W2^2).

	Gaussian → 8gaussians				Gaussian → moons				Gaussian → scurve			
Method	$W_2^2 \downarrow$	**NPE** ↓	**Train (s)**	**Pretrain**	$W_2^2 \downarrow$	**NPE** ↓	**Train (s)**	**Pretrain**	$W_2^2 \downarrow$	**NPE** ↓	**Train (s)**	**Pretrain**
I-CFM	0.2800	0.2590	**168.7**	0.0s	0.3901	0.5380	**45.6**	0.0s	0.1113	1.9230	**44.5**	0.0s
OT-CFM	**0.1956**	**0.0108**	1054.7	0.0s	**0.2668**	**0.0142**	1513.1	0.0s	**0.0991**	**0.0243**	992.0	0.0s
RA-OT	0.5072	0.0144	411.8	**2.6s**	0.5569	**0.0619**	331.7	**1.0s**	0.4675	0.0400	299.3	**6.3s**
OA-OT	0.5041	0.0142	413.8	14.0s	0.5459	0.0631	319.7	14.6s	0.4672	0.0425	294.0	14.4s

4. EXPERIMENTS & RESULTS: Mini-batch OT Flow



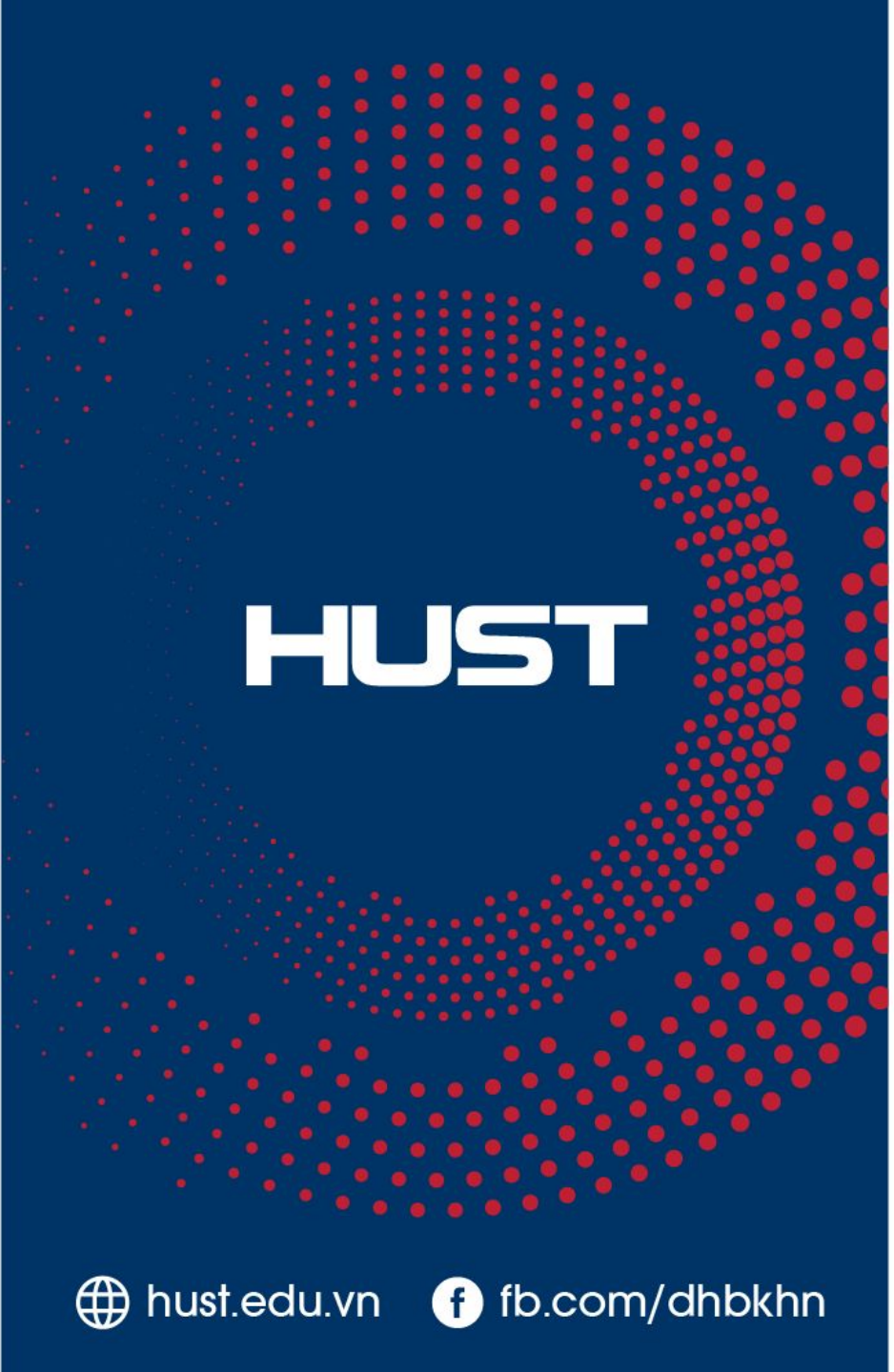
5. CONCLUSION

→ We introduce **RA-OT** and **OA-OT**, two amortized OT methods built from sliced Kantorovich potentials.

- Both methods use sliced OT as a structured representation rather than only as an approximation tool.
- **RA-OT** learns the mapping through functional regression, while **OA-OT** optimizes the Kantorovich dual objective directly.
- The resulting models are lightweight and their parameters are independent of the number of atoms.
- Experiments show strong performance on MNIST transport, spherical transport, color transfer, and conditional flow matching.
- The proposed methods can reduce training cost while maintaining accurate transport-plan predictions across repeated OT problems.

Final Takeaway

Sliced Kantorovich potentials provide a compact and reusable representation for learning optimal transport solutions.



THANK YOU !