Optimal Transport-Guided Conditional Score-Based Diffusion Model

Xiang Gu, Liwei Yang, Jian Sun and Zongben Xu

School of Mathematics and Statistics

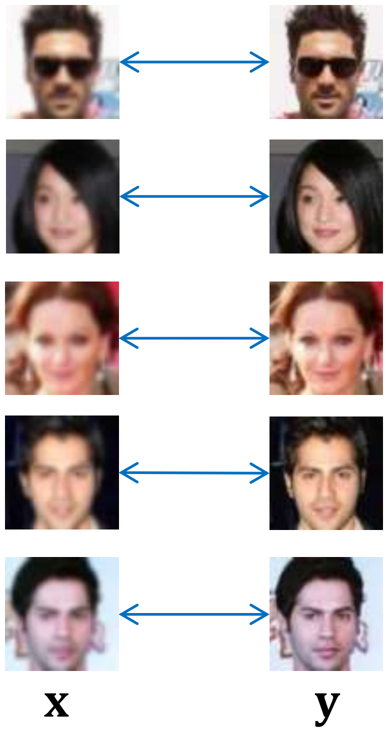
Xi'an Jiaotong University

xianggu@stu.xjtu.edu.cn,

jiansun@xjtu.edu.cn

1. Background

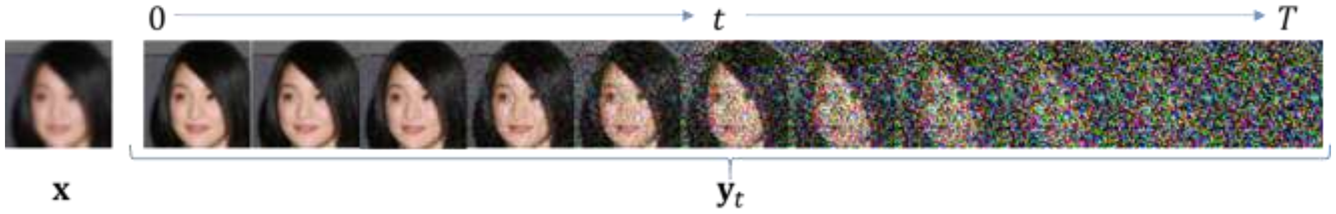
- **Conditional Score-Based Diffusion Model (Conditional SBDM) for Paired Data**



Paired condition data ($\mathbf{x}$) and target data ($\mathbf{y}$)

Forward Stochastic Differential Equation (SDE):

$$\mathrm{d}\mathbf{y}_t = f(\mathbf{y}_t, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}$$

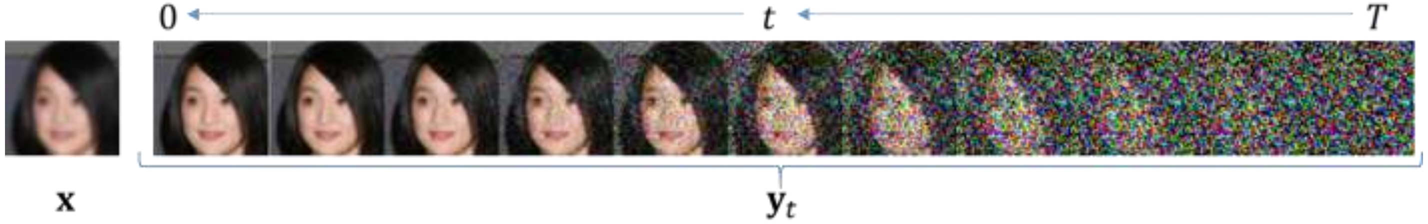


Denoising score matching:

$$\mathcal{J}_{\mathrm{DSM}}(\theta) = \mathbb{E}_t w_t \mathbb{E}_{\mathbf{y}_0 \sim q} \mathbb{E}_{\mathbf{y}_t \sim p_{t|0}(\mathbf{y}_t|\mathbf{y}_0)} \left\| s_\theta(\mathbf{y}_t; \mathbf{x}_{\mathrm{cond}}(\mathbf{y}_0), t) - \nabla_{\mathbf{y}_t} \log p_{t|0}(\mathbf{y}_t|\mathbf{y}_0) \right\|_2^2$$

Reverse SDE:

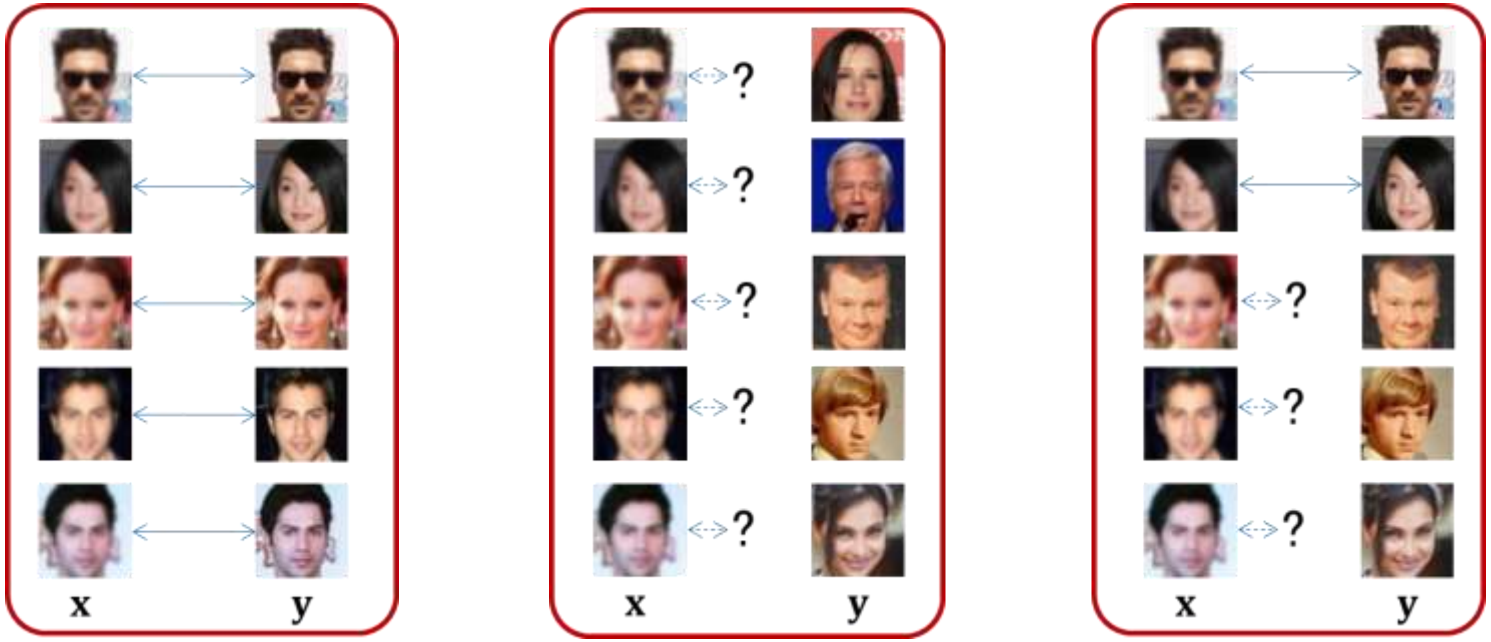
$$\mathrm{d}\mathbf{y}_t = [f(\mathbf{y}_t, t) - g^2(t) s_\theta(\mathbf{y}_t; \mathbf{x}, t)]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}}$$



2. Motivations

- **Why OT-Guided Conditional SBDM?**

  Insufficient paired data in practical applications



Paired Unpaired Partially paired

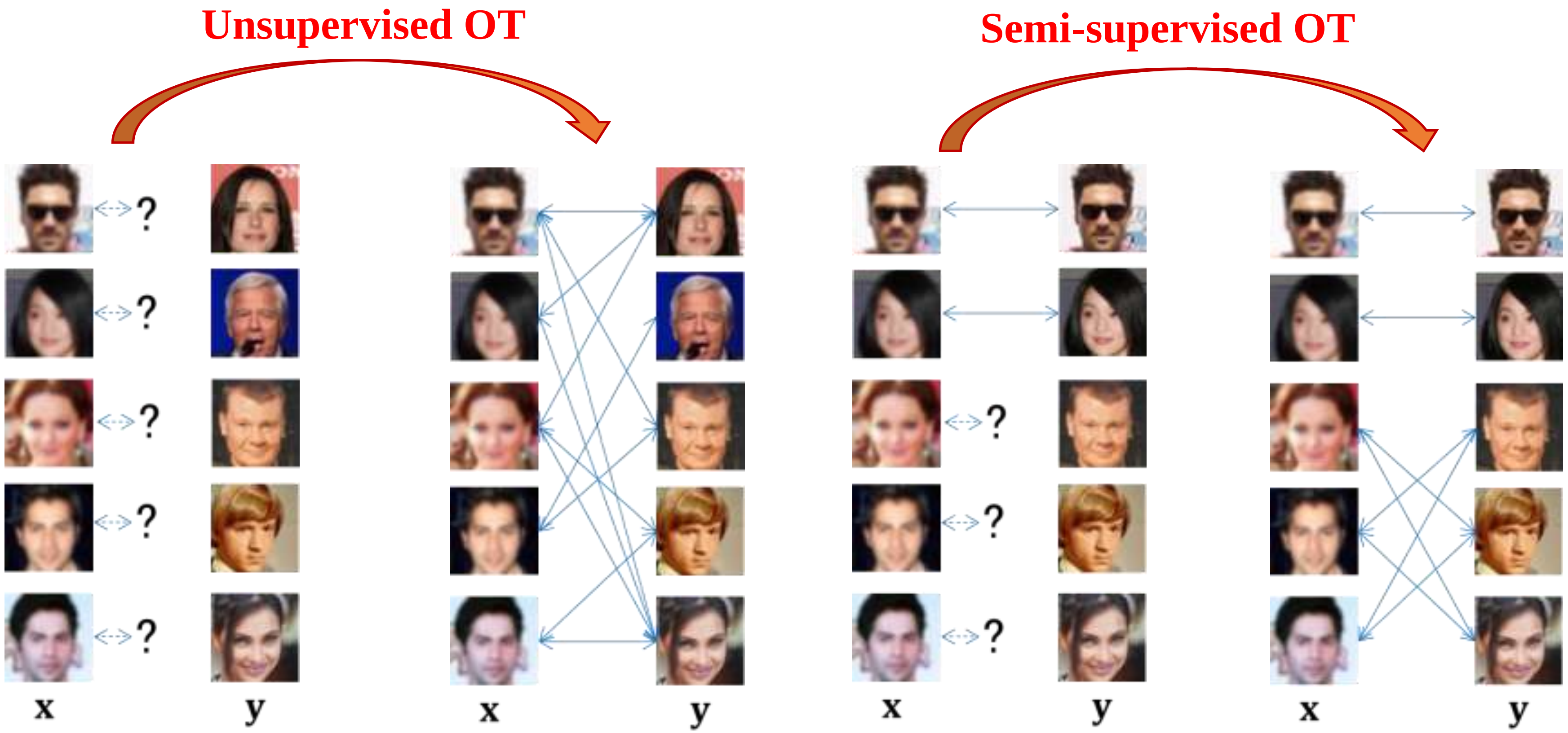
- **Main challenges:**
  - **Lacking coupling relationship of data**
  - **Unclear formulation for SBDM**

3. Contributions

- Optimal transport-guided Conditional Score-Based Diffusion Model (Conditional SBDM) for unpaired data.
- An approach to realize large-scale optimal transport based on diffusion model.
- Applications in unpaired super-resolution and semi-paired image-to-image translation.

4. OT-Guided Conditional SBDM

- For the first challenge, we build coupling relationship using unsupervised OT and semi-supervised OT for unpaired and partially paired settings, respectively.



Unpaired setting

Partially paired setting

4. OT-Guided Conditional SBDM

L_2-Regularized Large-Scale Optimal Transport

$$\min_{\pi \in \Gamma} \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim\pi} c(\mathbf{x},\mathbf{y}) + \epsilon\chi^2(\pi \| p \times q)$$

Unsupervised OT (Seguy et al. 18)

$$\min_{\tilde{\pi} \in \tilde{\Gamma}} \mathbb{E}_{(\mathbf{x},\mathbf{y})\sim m\otimes\tilde{\pi}} g(\mathbf{x},\mathbf{y}) + \epsilon\chi^2(m \otimes \tilde{\pi} \| p \times q)$$

Semi-supervised OT (Gu et al. 23)

Unified duality:

$$\max_{u,v} \mathbb{E}_{\mathbf{x}\sim p} u(\mathbf{x}) + \mathbb{E}_{\mathbf{y}\sim q} v(\mathbf{y}) - \frac{1}{4\epsilon}\mathbb{E}_{\mathbf{x}\sim p,\mathbf{y}\sim q} I(\mathbf{x},\mathbf{y})\left(u(\mathbf{x}) + v(\mathbf{y}) - \xi(\mathbf{x},\mathbf{y})\right)_+^2 \quad (6)$$

For unsupervised OT, $I(\mathbf{x},\mathbf{y}) = 1$ and $\xi(\mathbf{x},\mathbf{y}) = c(\mathbf{x},\mathbf{y})$

For semi-supervised OT, $I(\mathbf{x},\mathbf{y}) = m(\mathbf{x},\mathbf{y})$ and $\xi(\mathbf{x},\mathbf{y}) = g(\mathbf{x},\mathbf{y})$

Guiding function: $g(\mathbf{x},\mathbf{y}) = d(R^s_{\mathbf{x}}, R^t_{\mathbf{y}})$

$$R^s_{\mathbf{x},k} = \frac{\exp(-c(\mathbf{x},\mathbf{x}_k)/\tau)}{\sum_{l=1}^K \exp(-c(\mathbf{x},\mathbf{x}_l)/\tau)}, \quad R^t_{\mathbf{y},k} = \frac{\exp(-c(\mathbf{y},\mathbf{y}_k)/\tau)}{\sum_{l=1}^K \exp(-c(\mathbf{y},\mathbf{y}_l)/\tau)}$$

Solving OT :

u, v are represented by neural networks u_ω, v_ω with parameters ω that are trained by mini-batch-based stochastic optimization algorithms using the loss function (6)， using parameters $\hat{\omega}$ after training， the estimate of optimal transport plan is

$$\hat{\pi}(\mathbf{x},\mathbf{y}) = H(\mathbf{x},\mathbf{y})p(\mathbf{x})q(\mathbf{y}), \text{ where } H(\mathbf{x},\mathbf{y}) = \frac{1}{2\epsilon} I(\mathbf{x},\mathbf{y})\left(u_{\hat{\omega}}(\mathbf{x}) + v_{\hat{\omega}}(\mathbf{y}) - \xi(\mathbf{x},\mathbf{y})\right)_+ \quad (7)$$

Seguy V., et al., Large-Scale Optimal Transport and Mapping Estimation, ICLR, 2018.
Xiang Gu, Yucheng Yang, Wei Zeng, Jian Sun, Zongben Xu. Keypoint-Guided Optimal Transport. JMLR. 2023, under review.

4. OT-Guided Conditional SBDM

➢ For the second challenge, we first provide a reformulation of paired setting and extend it for unpaired and partially paired settings.

Reformulation of paired setting:

Proposition 1. *Let* $\mathcal{C}(\mathbf{x},\mathbf{y}) = \frac{1}{p(\mathbf{x})}\delta(\mathbf{x}-\mathbf{x}_{\text{cond}}(\mathbf{y}))$ *where* δ *is the Dirac delta function, then* $\mathcal{J}_{\text{DSM}}(\theta)$ *in Eq. (1) can be reformulated as*

$$\mathcal{J}_{\text{DSM}}(\theta) = \mathbb{E}_t w_t \mathbb{E}_{\mathbf{x}\sim p}\mathbb{E}_{\mathbf{y}\sim q}\mathcal{C}(\mathbf{x},\mathbf{y})\mathbb{E}_{\mathbf{y}_t\sim p_{t|0}(\mathbf{y}_t|\mathbf{y})}\left\|s_\theta(\mathbf{y}_t;\mathbf{x},t) - \nabla_{\mathbf{y}_t}\log p_{t|0}(\mathbf{y}_t|\mathbf{y})\right\|_2^2. \quad (8)$$

Furthermore, $\gamma(\mathbf{x},\mathbf{y}) = \mathcal{C}(\mathbf{x},\mathbf{y})p(\mathbf{x})q(\mathbf{y})$ *is a joint distribution for marginal distributions* p *and* q.

Observations from Proposition 1:

- The coupling relationship of $\mathbf{x},\mathbf{y}$ is explicitly modeled in $\mathcal{C}(\mathbf{x},\mathbf{y})$
- $\gamma(\mathbf{x},\mathbf{y})$ is closely related to the transport plan of L_2-regularized OT

Transport plan

$\hat{\pi}(\mathbf{x},\mathbf{y}) = H(\mathbf{x},\mathbf{y})p(\mathbf{x})q(\mathbf{y})$

Inspiring us to extend Eq. (8) to partially paired or unpaired settings by replacing $\mathcal{C}(\mathbf{x},\mathbf{y})$ ***with*** $H(\mathbf{x},\mathbf{y})$ ***!!!***

4. OT-Guided Conditional SBDM

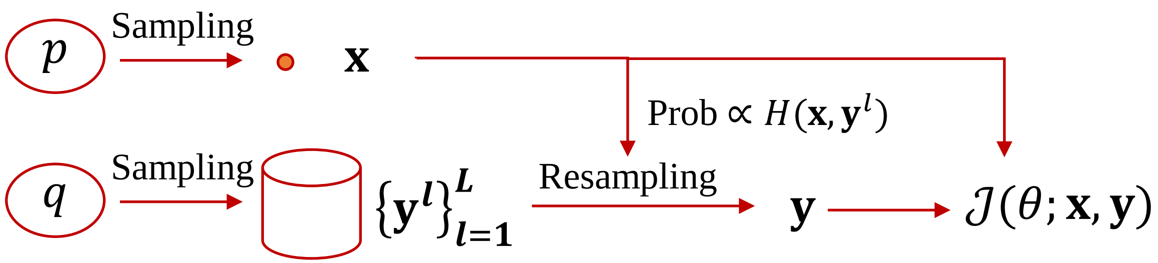
Formulation for unpaired and partially paired settings

$$\mathcal{J}_{\mathrm{CDSM}}(\theta) = \mathbb{E}_t w_t \mathbb{E}_{\mathbf{x}\sim p} \mathbb{E}_{\mathbf{y}\sim q} H(\mathbf{x},\mathbf{y}) \mathbb{E}_{\mathbf{y}_t \sim p_{t|0}(\mathbf{y}_t|\mathbf{y})} \left\| s_\theta(\mathbf{y}_t;\mathbf{x},t) - \nabla_{\mathbf{y}_t} \log p_{t|0}(\mathbf{y}_t|\mathbf{y}) \right\|_2^2 \quad (9)$$

Training the Conditional Score-based Model s_θ

- Standard minibatch-based training is less effective due to sparsity of $H(\mathbf{x},\mathbf{y})$.
- We present the following Resampling-by-compatibility strategy.

Resampling-by-compatibility:



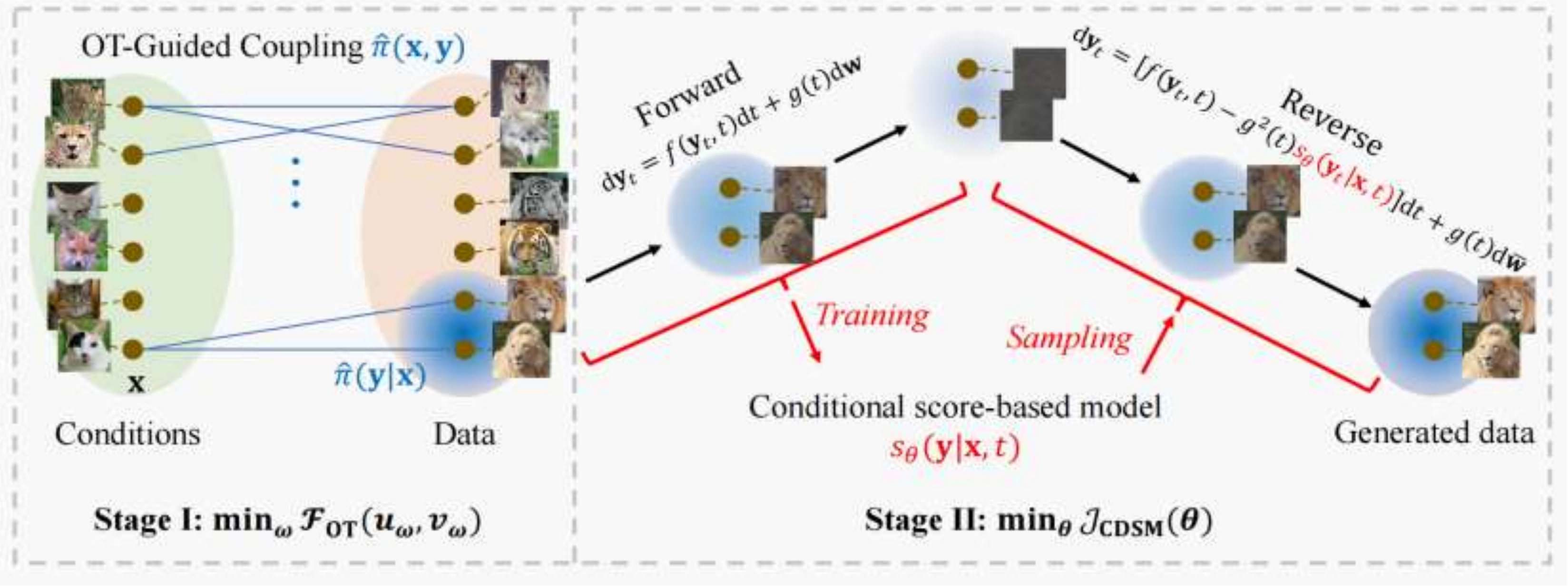
Sample generation by reverse SDE

$$\mathrm{d}\mathbf{y}_t = \left[f(\mathbf{y}_t,t) - g(t)^2 s_{\hat{\theta}}(\mathbf{y}_t;\mathbf{x},t)\right]\mathrm{d}t + g(t)\,\mathrm{d}\bar{\mathbf{w}},$$

4. OT-Guided Conditional SBDM

Understanding OT-Guided Conditional SBDM

Theorem 1. *For* $\mathbf{x} \sim p$*, we define the forward SDE* $\mathrm{d}\mathbf{y}_t = f(\mathbf{y}_t, t)\,\mathrm{d}t + g(t)\,\mathrm{d}\mathbf{w}$ *with* $\mathbf{y}_0 \sim \hat{\pi}(\cdot|\mathbf{x})$ *and* $t \in [0, T]$*, where* f, g, T *are given in Sect. 2.1. Let* $p_t(\mathbf{y}_t|\mathbf{x})$ *be the corresponding distribution of* $\mathbf{y}_t$ *and* $\mathcal{J}_{\mathrm{CSM}}(\theta) = \mathbb{E}_t w_t \mathbb{E}_{\mathbf{x}\sim p} \mathbb{E}_{\mathbf{y}_t \sim p_t(\mathbf{y}_t|\mathbf{x})} \|s_\theta(\mathbf{y}_t; \mathbf{x}, t) - \nabla_{\mathbf{y}_t} \log p_t(\mathbf{y}_t|\mathbf{x})\|_2^2$*, then we have* $\nabla_\theta \mathcal{J}_{\mathrm{CDSM}}(\theta) = \nabla_\theta \mathcal{J}_{\mathrm{CSM}}(\theta)$.

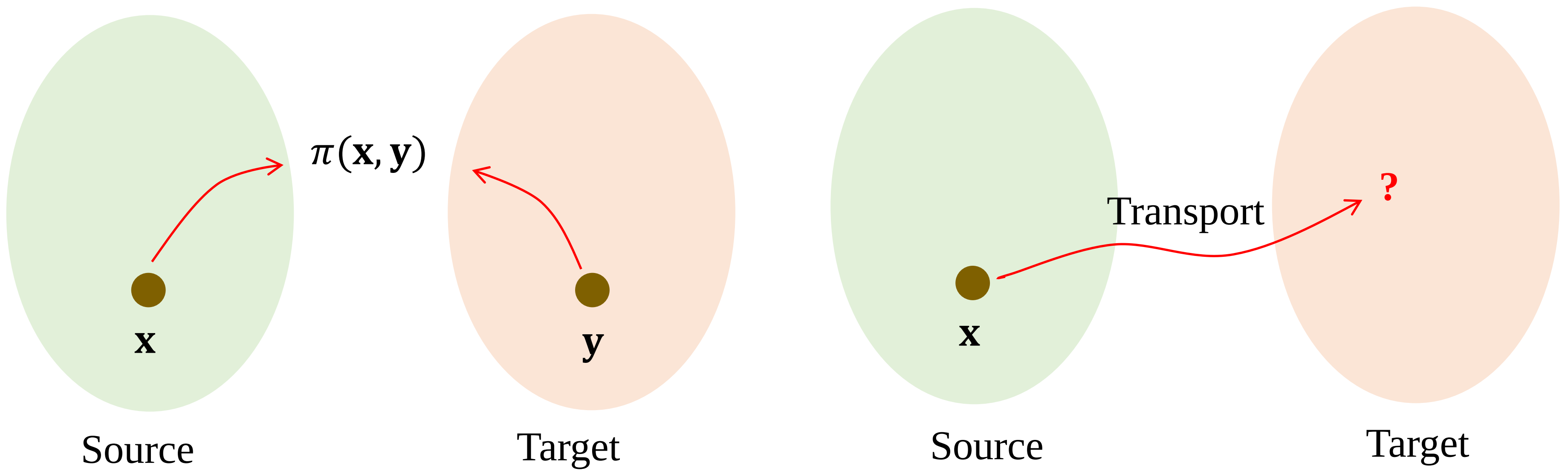


Workflow: (1) Building coupling $\hat{\pi}$ using OT; (2) Sampling clean sample $\mathbf{y}$ from $\hat{\pi}(\mathbf{y}|\mathbf{x})$; (3) Adding noise by forward SDE to train s_θ (4) Generating samples from $\hat{\pi}(\mathbf{y}|\mathbf{x})$ by reverse SDE.

4. OT-Guided Conditional SBDM

OTCS Realizing Data Transport for OT

- $\pi(\mathbf{x}, \mathbf{y})$ models the density function value of $\mathbf{x}, \mathbf{y}$ rather than the transported sample of $\mathbf{x}$
- How to transport source sample $\mathbf{x}$ to the target domain is known to be a challenge



Our proposed OTCS provides an diffusion-based approach to transport x to target domain by sampling from optimal conditional transport plan $\pi(\cdot\,|\mathrm{x})$, with a theoretical guarantee.

4. OT-Guided Conditional SBDM

OTCS Realizing Data Transport for OT

Notations:

$p^{sde}(\cdot\,|\mathbf{x})$: distribution of samples generated by OTCS

$\pi(\cdot\,|\mathbf{x})$: true conditional optimal transport plan of L_2-regularized OT

Theorem 2. *Suppose the assumptions in Appendix B hold, and $w_t = g(t)^2$, then we have*

$$\mathbb{E}_{\mathbf{x}\sim p} W_2\left(p^{\text{sde}}(\cdot|\mathbf{x}), \pi(\cdot|\mathbf{x})\right) \le C_1 \left\|\nabla_{\hat{\pi}} \mathcal{L}(\hat{\pi}, u_{\hat{\omega}}, v_{\hat{\omega}})\right\|_1 + \sqrt{C_2 \mathcal{J}_{\text{CSM}}(\hat{\theta})} + C_3 \mathbb{E}_{\mathbf{x}\sim p} W_2\left(p_T(\cdot|\mathbf{x}), p_{\text{prior}}\right), \quad (11)$$

where C_1, C_2, and C_3 are constants to $\hat{\omega}$ and $\hat{\theta}$ given in Appendix B.

Choose SDE to minimize it.

$u_{\hat{\omega}}, v_{\hat{\omega}}$ is near to the saddle point so that gradient norm is minimized.

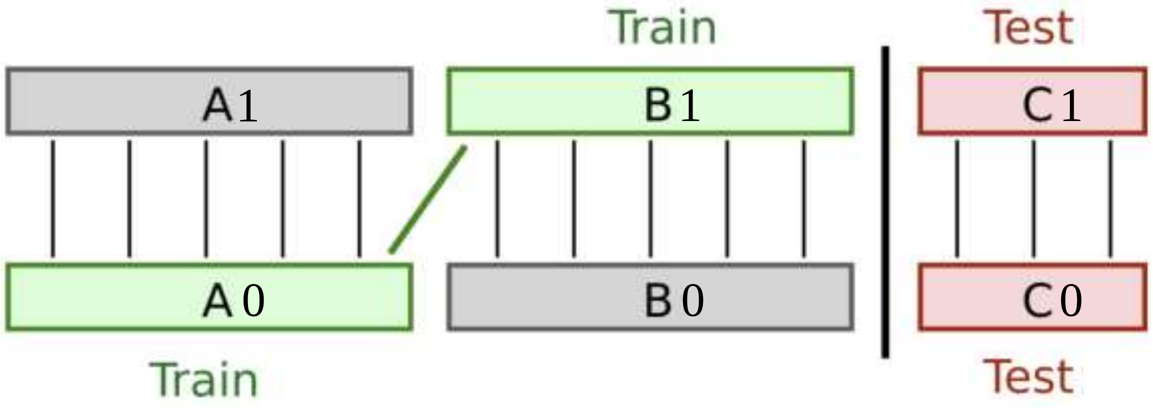
Training loss as in Theorem 1.

Theorem 2 shows that OTCS can generate samples from $\pi(\cdot\,|\mathbf{x})$.

4.1 Experiments on unpaired super-resolution

High-resolution

Degenerated



Results:

Degenerated images

Translated images

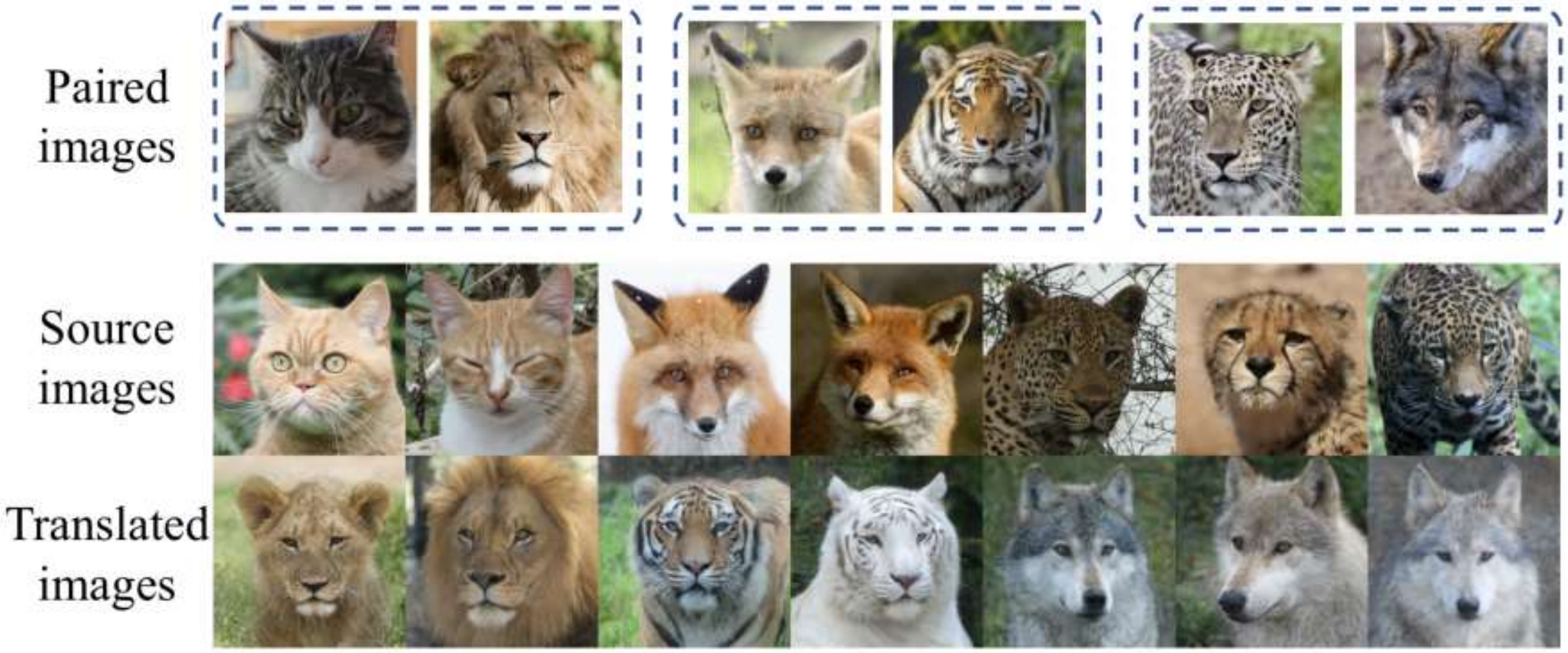
4.1 Application unpaired super-resolution

Table 1: Quantitative results for unpaired super-resolution on Celeba and semi-paired I2I on Animal images and Digits. The best and second best are respectively bolded and underlined.

Method	Method Type	Celeba		Animal images		Digits	
		FID ↓	SSIM ↑	FID ↓	Acc (%) ↑	FID ↓	Acc (%) ↑
W2GAN [33]	OT	48.83	0.7169	118.45	33.56	97.06	29.13
OT-ICNN [29]	OT	33.26	0.8904	148.29	38.44	50.33	10.84
OTM [30]	OT	22.93	0.8302	69.27	33.11	18.67	9.48
NOT [34]	OT	13.65	<u>0.9157</u>	156.07	28.44	23.90	15.72
KNOT [31]	OT	<u>5.95</u>	0.8887	118.26	27.33	**3.18**	9.25
ReFlow [36]	Flow	70.69	0.4544	56.04	29.33	138.59	11.57
EGSDE [17]	Diffusion	11.49	0.3835	52.11	29.33	34.72	11.78
DDIB [35]	Diffusion	11.35	0.1275	28.45	32.44	9.47	9.15
TCR [14]	–	–	–	34.61	<u>40.44</u>	6.90	<u>36.21</u>
SCONES [19]	OT, Diffusion	15.46	0.1042	<u>25.24</u>	35.33	6.68	10.37
OTCS (ours)	OT, Diffusion	**1.77**	**0.9313**	**13.68**	**96.44**	<u>5.12</u>	**67.42**

4.2 Experiments on semi-paired image-to-image translation

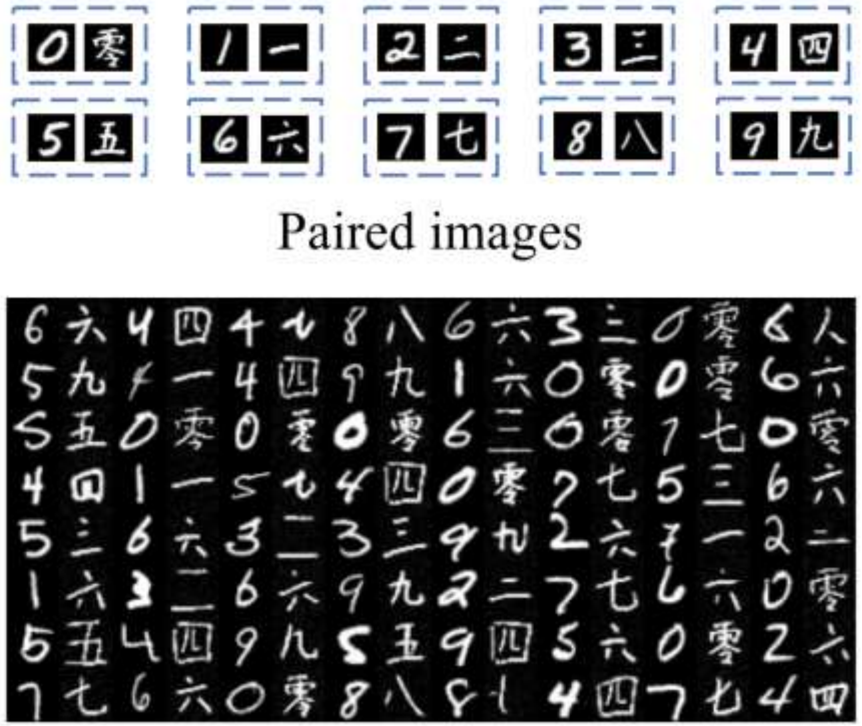
Task: Translate cat/fox/leopard images to lion/tiger/wolf using guidance of three image pairs, without class label



(a) Animal images (256×256)

4.2 Experiments on semi-paired image-to-image translation

Task: Translate MNIST to Chinese-MNIST using guidance of ten image pairs, without class label



Paired images

Source and translated images

(b) Digits (28×28)

4.2 Experiments on semi-paired image-to-image translation

Table 1: Quantitative results for unpaired super-resolution on Celeba and semi-paired I2I on Animal images and Digits. The best and second best are respectively bolded and underlined.

Method	Method Type	Celeba		Animal images		Digits	
		FID ↓	SSIM ↑	FID ↓	Acc (%) ↑	FID ↓	Acc (%) ↑
W2GAN [33]	OT	48.83	0.7169	118.45	33.56	97.06	29.13
OT-ICNN [29]	OT	33.26	0.8904	148.29	38.44	50.33	10.84
OTM [30]	OT	22.93	0.8302	69.27	33.11	18.67	9.48
NOT [34]	OT	13.65	<u>0.9157</u>	156.07	28.44	23.90	15.72
KNOT [31]	OT	<u>5.95</u>	0.8887	118.26	27.33	**3.18**	9.25
ReFlow [36]	Flow	70.69	0.4544	56.04	29.33	138.59	11.57
EGSDE [17]	Diffusion	11.49	0.3835	52.11	29.33	34.72	11.78
DDIB [35]	Diffusion	11.35	0.1275	28.45	32.44	9.47	9.15
TCR [14]	–	–	–	34.61	<u>40.44</u>	6.90	<u>36.21</u>
SCONES [19]	OT, Diffusion	15.46	0.1042	<u>25.24</u>	35.33	6.68	10.37
OTCS (ours)	OT, Diffusion	**1.77**	**0.9313**	**13.68**	**96.44**	<u>5.12</u>	**67.42**

Thanks for your attention!

Code: https://github.com/XJTU-XGU/OTCS