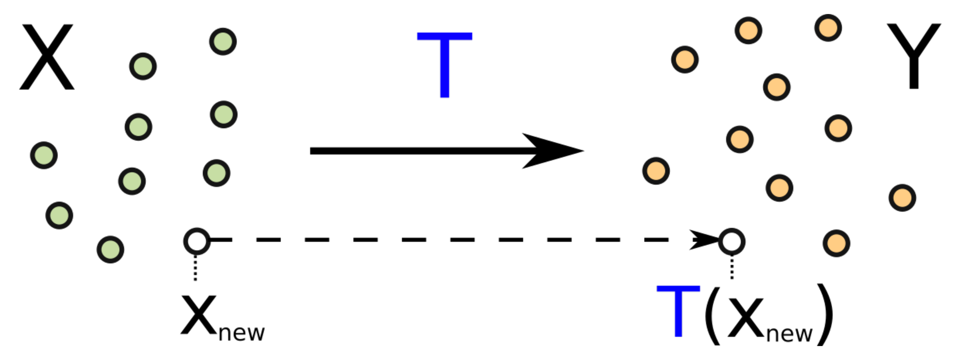
Extremal Domain Translation with Neural Optimal Transport

Skoltech[1] Skolkovo Institute of Science and Technology — AIRI[2]

We present a novel neural-networks-based algorithm to learn **incomplete optimal transport** maps.

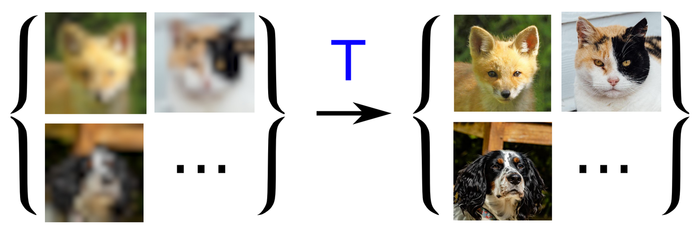
Milena Gazdieva[1] *Alexander Korotin*[1,2] *Daniil Selikhanovych*[1] *Evgeny Burnaev*[1,2]

I Motivation: *Unpaired* Domain Translation

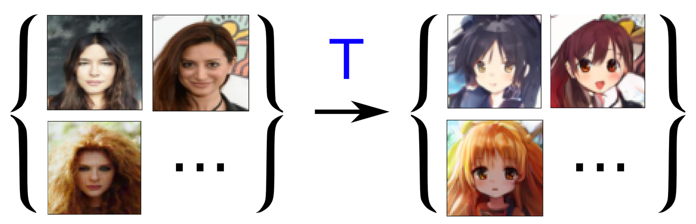


The (informal) task: given samples X, Y from two domains, construct a map T which can translate new samples from the input domain to the target domain.

Example I: image super-resolution

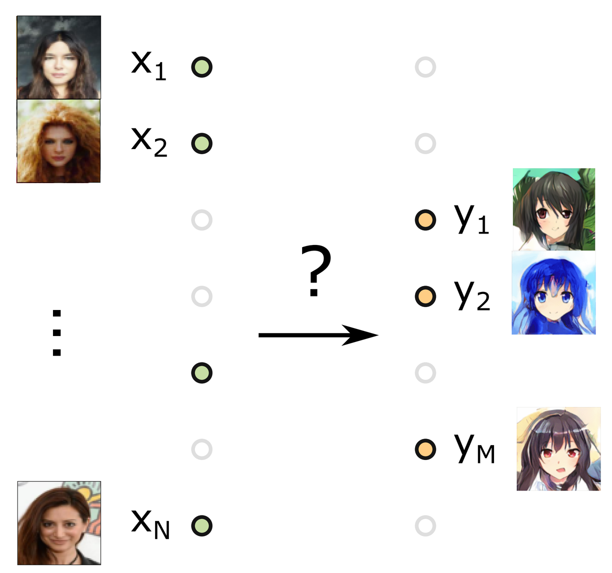


Example II: image style transfer



Unpaired setup

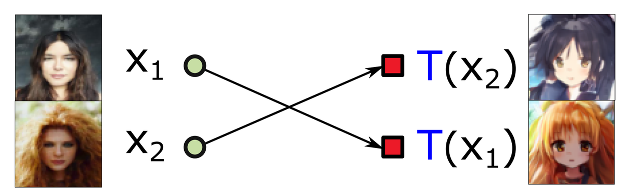
No paired training examples are available.



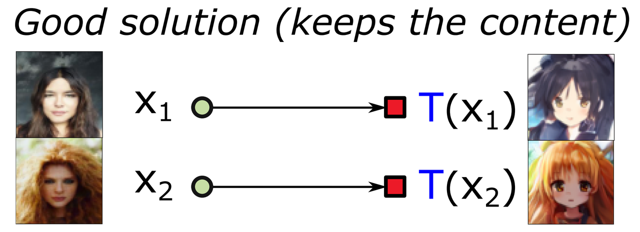
Issues of the unpaired setup

Ambiguity in translations

Bad solution (changes the content)

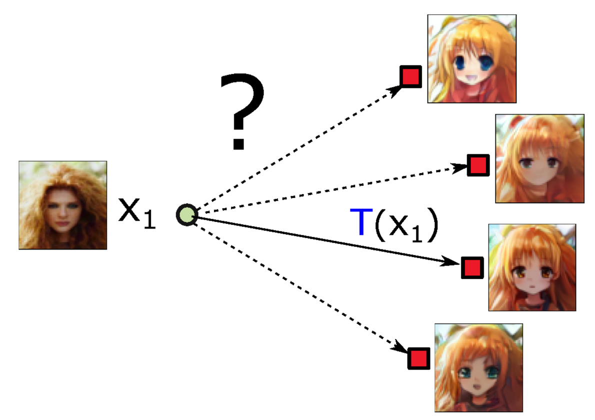


Good solution (keeps the content)



How to maximize the similarity?

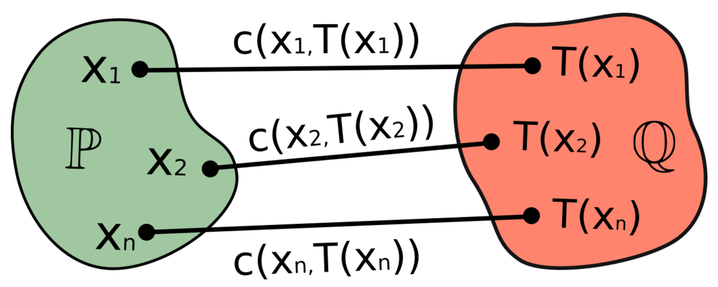
Plenty of translations keep the content. How to find the most similar?



II Background: Optimal Transport

Given two probability distributions $\mathbb{P}\in\mathcal{P}(\mathcal{X})$ $\mathbb{Q}\in\mathcal{P}(\mathcal{Y})$ on spaces $\mathcal{X}$, $\mathcal{Y}$, respectively,
(1) how to find the optimal way to transport the probability mass of $\mathbb{P}$ to $\mathbb{Q}$?
(2) how to find the way to transport the mass to the nearest points of $\mathbb{Q}$?

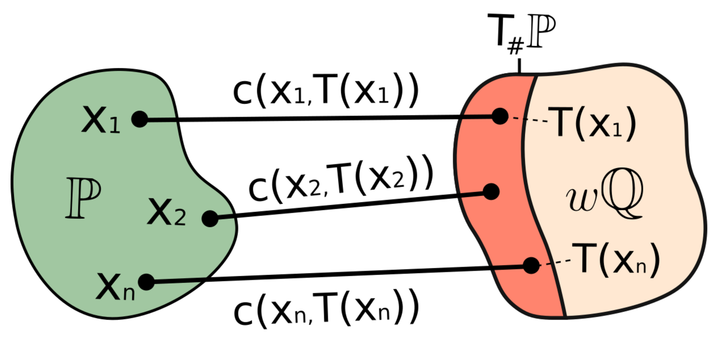
Classic OT



$$\inf_{T\sharp\mathbb{P}=\mathbb{Q}}\int_{\mathcal{X}} c(x,T(x))d\mathbb{P}(x)$$

The minimizer T^* is called the OT map.

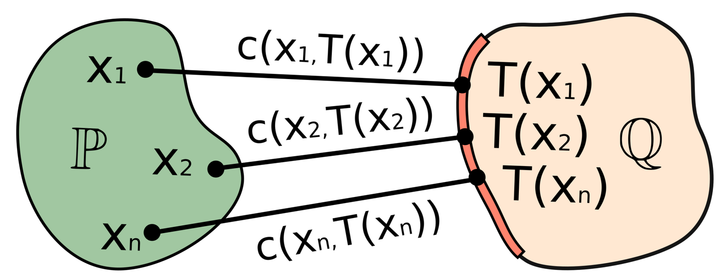
Incomplete OT (IT)



$$\underbrace{\inf_{T\sharp\mathbb{P}\leq w\mathbb{Q}}\int_{\mathcal{X}} c(x,T(x))d\mathbb{P}(x)}_{\text{Cost}_w(\mathbb{P},\mathbb{Q})}$$

The minimizer T^* is called the IT map.

Extremal OT (ET)



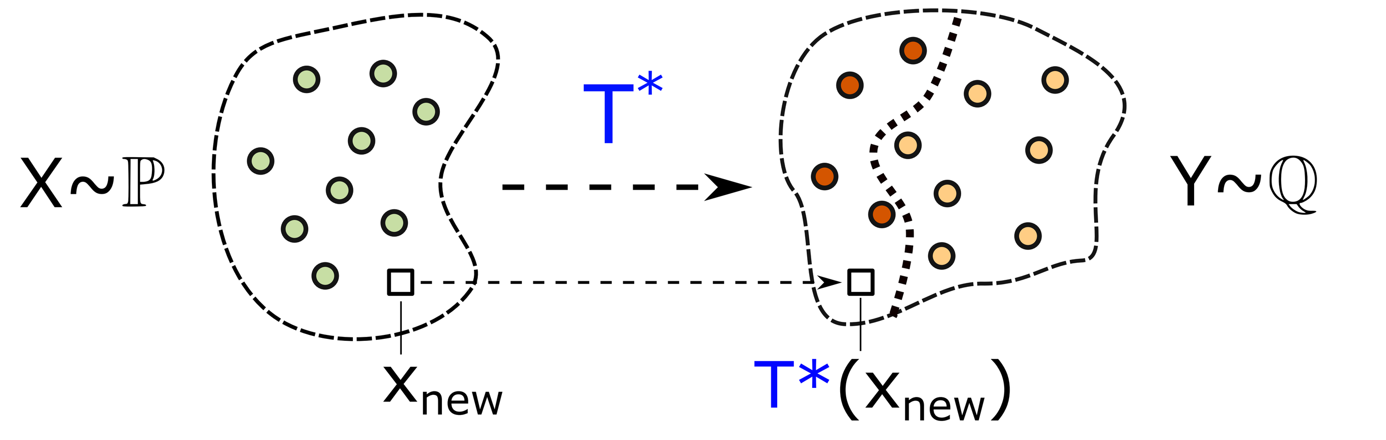
$$\underbrace{\inf_{\text{Supp}(T\sharp\mathbb{P})\subset\text{Supp}(\mathbb{Q})}\int_{\mathcal{X}} c(x,T(x))d\mathbb{P}(x)}_{\text{Cost}_\infty(\mathbb{P},\mathbb{Q})}$$

The minimizer T^* is called the ET map.

When $w\rightarrow\infty$ IT problem converges to ET.

III Proposed Algorithm: Incomplete Optimal Transport

Continuous IT problem



The (formal) task: given empirical samples $X\sim\mathbb{P}$, $Y\sim\mathbb{Q}$, reconstruct an IT map T^* between $\mathbb{P},\mathbb{Q}$.

Lemma (Minimax reformulation of the dual problem)

$$\text{Cost}_w(\mathbb{P},\mathbb{Q})=\sup_{f\leq 0}\inf_T \mathcal{L}(f,T),$$

where the functional $\mathcal{L}$ is defined by

$$\mathcal{L}(f,T)\stackrel{def}{=}\int_{\mathcal{X}} c(x,T(x))d\mathbb{P}(x)-\int_{\mathcal{X}} f(T(x))d\mathbb{P}(x)+\mathbf{w}\int_{\mathcal{Y}} f(y)d\mathbb{Q}(y).$$

Algorithm 1: Incomplete neural optimal transport (**IT**)

repeat
- Sample batches $X\sim\mathbb{P}$, $Y\sim\mathbb{Q}$;
- $\mathcal{L}_f\leftarrow w\cdot\frac{1}{|Y|}\sum_{y\in Y} f_\psi(y)-\frac{1}{|X|}\sum_{x\in X} f_\psi(T_\theta(x))$;
- Update ψ by using $\frac{\partial\mathcal{L}_f}{\partial\psi}$ to maximize $\mathcal{L}_f$;
- **for** $k_T=1,2,\dots,K_T$ **do**
  - Sample batch $X\sim\mathbb{P}$;
  - $\mathcal{L}_T\leftarrow\frac{1}{|X|}\sum_{x\in X}\big[c(x,T_\theta(x))-f_\psi(T_\theta(x))\big]$;
  - Update θ by using $\frac{\partial\mathcal{L}_T}{\partial\theta}$ to minimize $\mathcal{L}_T$;

until *not converged*;

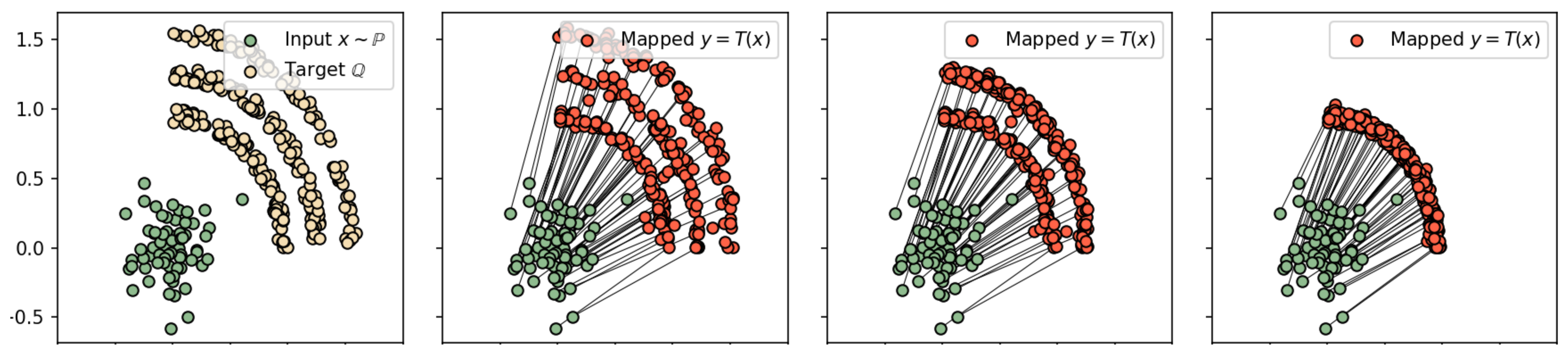
Lemma (IT maps solve the problem)

For any potential f^* which attains the optimal value of the problem, and for any IT map T^* which realizes some IT plan π^*:

$$T^*\in\arg\inf_T\mathcal{L}(f^*,T).$$

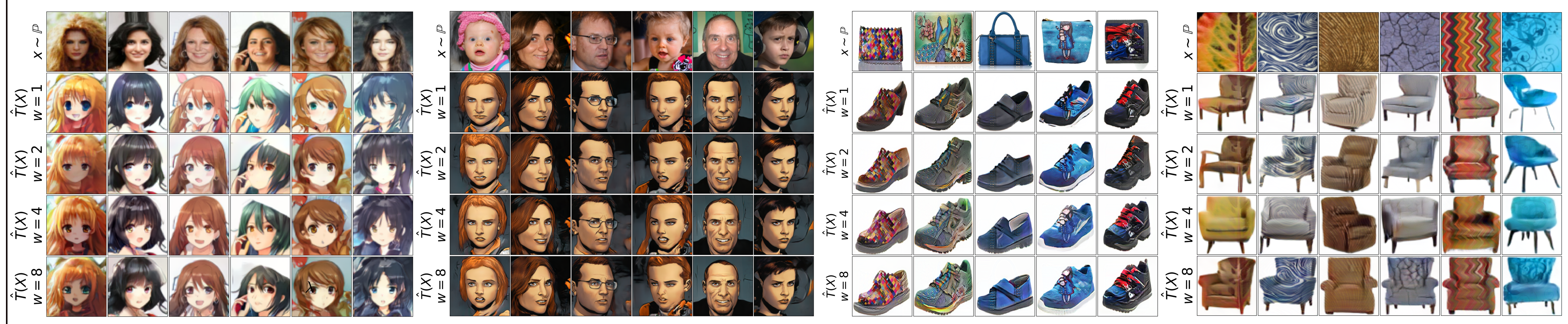
IV Toy Experiments: Learning Part of the Distribution

Example: gaussian to Wi-Fi



V Illustrative Experiments: Unpaired Translation

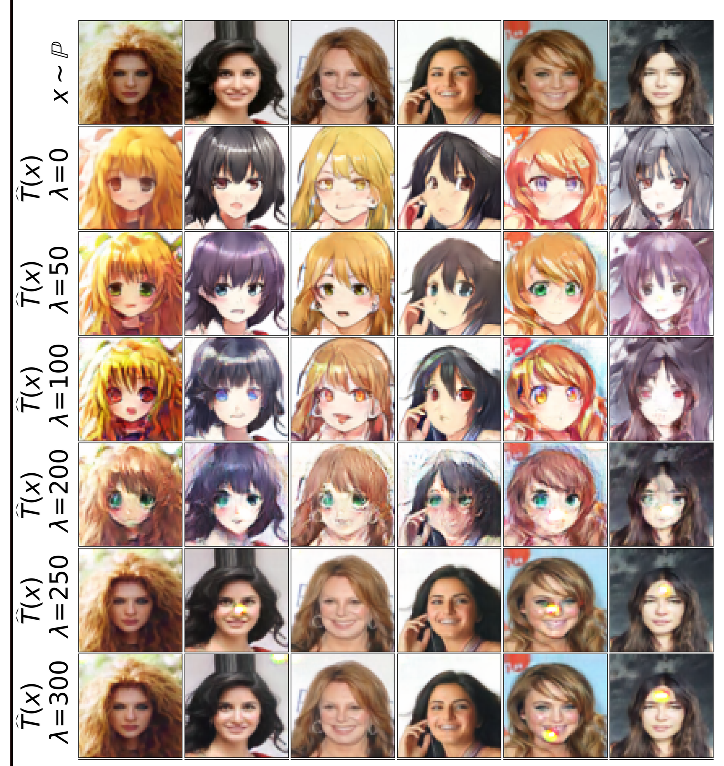
Proposed method



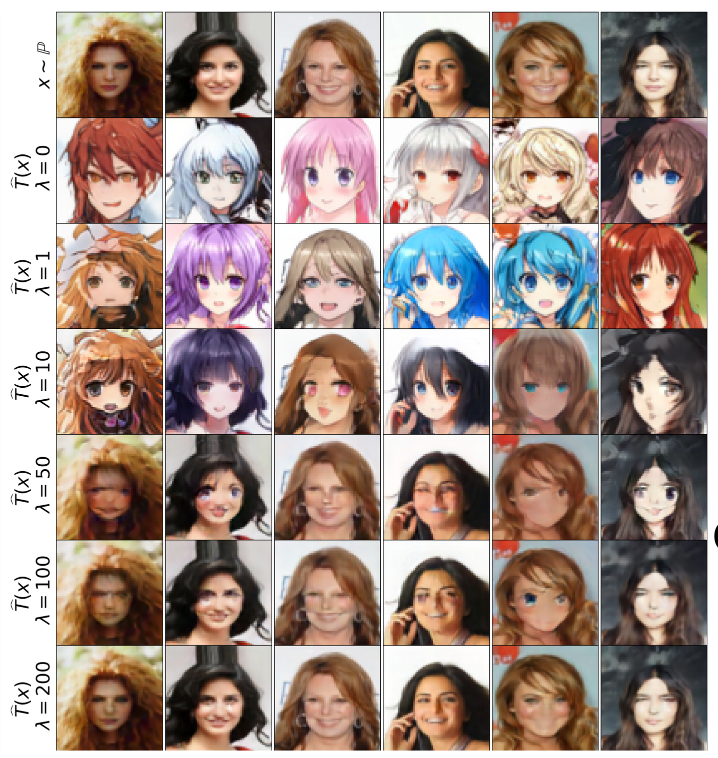
Celeba → anime FFHQ → Comics Handbag → shoes Texture → chairs

Results: IT method provides a way to control and maximize the similarity by varying the parameter w

Failures of existing methods



CycleGAN



StarGAN-v2

GANs

Typical objectives of GANs:

$$\underbrace{\mathcal{L}_{Dom}(T)}_{\text{Domain loss}}+\lambda\cdot\underbrace{\mathcal{L}_{Sim}(T)}_{\text{Similarity loss}}+[\text{other terms}]$$

Varying λ yields the nasty *realism-similarity* tradeoff

Discrete OT+Neural Nets

Common interpolation strategy is to learn the barycentric projection $\overline{T}(x)=\int_{\mathcal{Y}} y\,d\pi(y|x)$

Result: *blurry* images

Discrete OT