(Kernel) Neural Optimal Transport

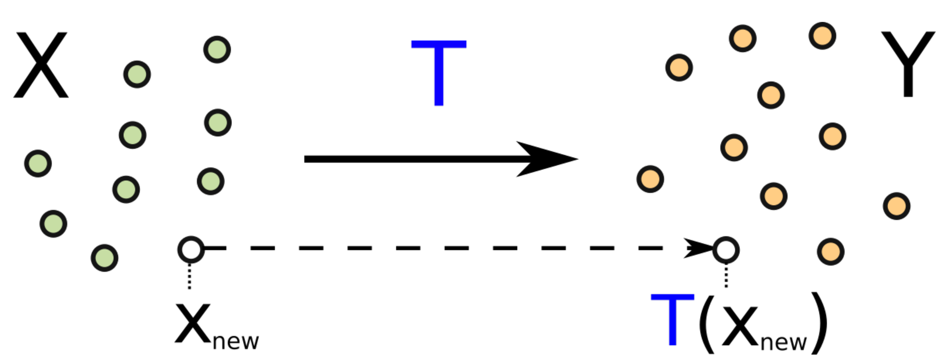
Skoltech[1] Skolkovo Institute of Science and Technology

AIRI[2]

We present a novel neural-networks-based algorithm to compute **optimal transport** maps and plans for **strong** and **weak** transport costs.

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

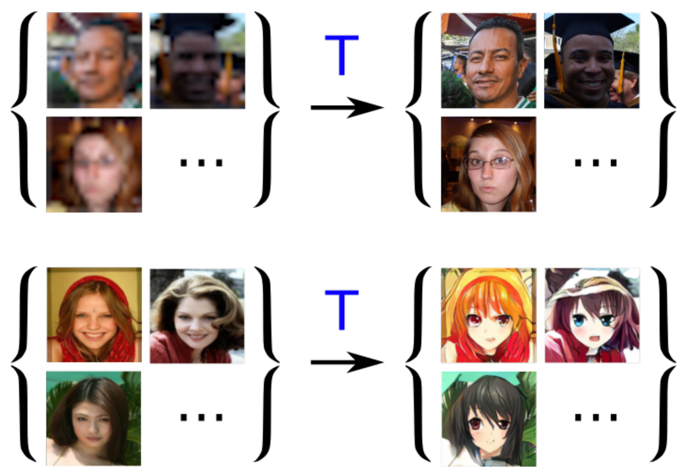
I Motivation: *Unpaired* Domain Translation



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.

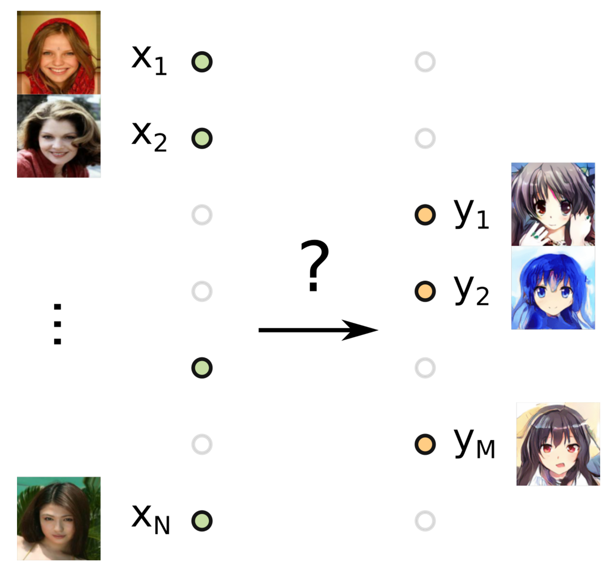
Example I: image super-resolution

Example II: image style transfer



Unpaired setup

No paired training examples are available.

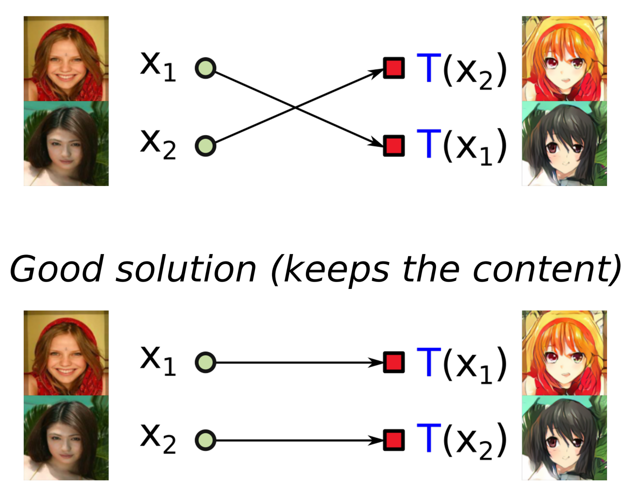


Issues of the unpaired setup

Ambiguity in translations

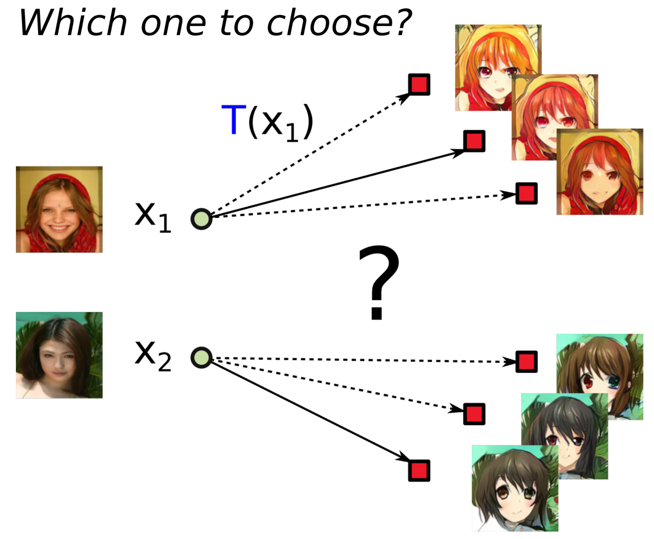
Bad solution (changes the content)

Good solution (keeps the content)



Plenty of possible translations

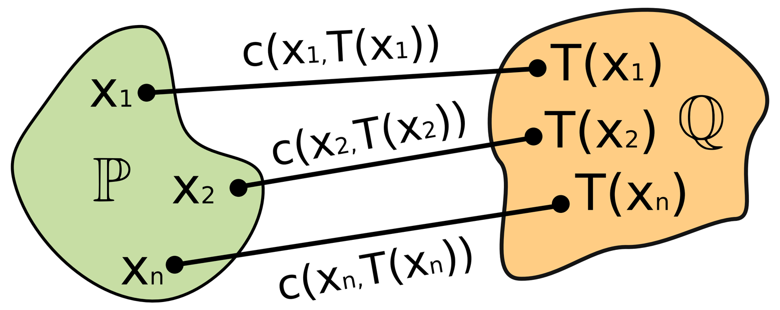
Which one to choose?



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, how to find the optimal way to transport the probability mass of $\mathbb{P}$ to $\mathbb{Q}$?

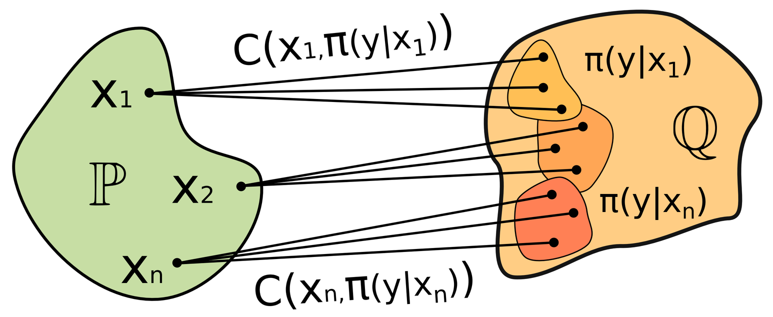
Monge (Strong) OT



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

The minimizer T^* is called the OT map.

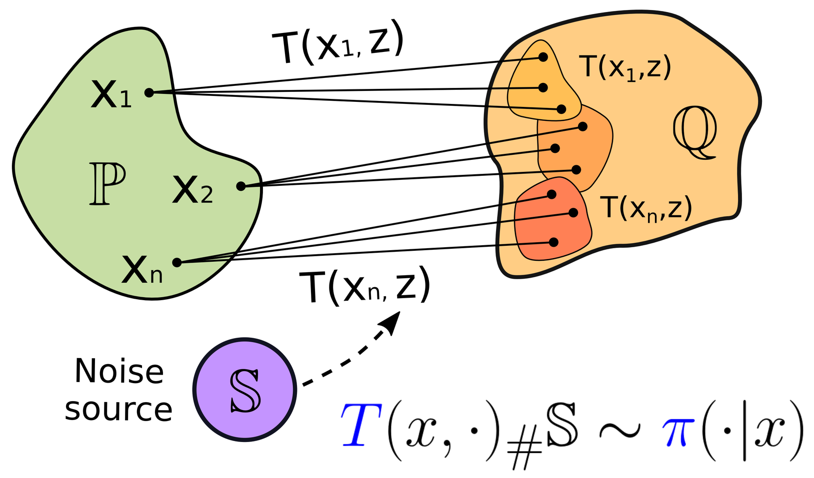
Weak OT



$$\inf_{\pi\in\Pi(\mathbb{P},\mathbb{Q})}\int_{\mathcal{X}} C(x,\pi(\cdot|x))d\mathbb{P}(x)$$

The minimizer π^* is called the OT plan.

Monge meets Weak OT



$$T(x,\cdot)_{\#}\mathbb{S}\sim\pi(\cdot|x)$$

If a stochastic function T^* represents some OT plan π^*, we say that it is an OT map.

Typical **weak** cost functions for $\mathcal{X}=\mathcal{Y}=\mathbb{R}^D$

$$C(x,\pi(\cdot|x))=\underbrace{\int_{\mathcal{Y}}c(x,y)d\pi(y|x)}_{\substack{\text{Dissimilarity}\\ \textit{(strong part)}}}-\gamma\underbrace{\mathcal{R}(\pi(\cdot|x))}_{\substack{\text{Diversity}\\ \textit{(weak part)}}}\qquad \gamma\geq 0$$

Similarity-diversity trade-off controlling parameter

Entropy-regularized cost

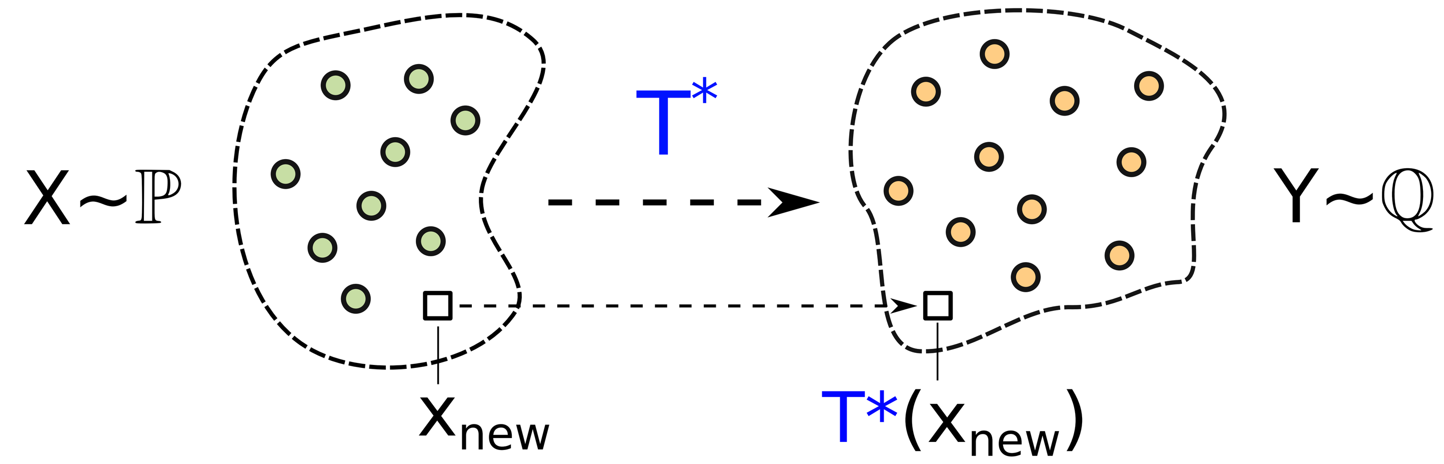
$$\int_{\mathcal{Y}}\frac{1}{2}\|x-y\|^2d\pi(y|x)-\frac{\gamma}{2}\text{Ent}(\pi(\cdot|x))$$

Weak quadratic cost

$$\int_{\mathcal{Y}}\frac{1}{2}\|x-y\|^2d\pi(y|x)-\frac{\gamma}{2}\text{Var}(\pi(\cdot|x))$$

III Proposed Algorithm: Neural Optimal Transport (NOT)

Continuous OT problem



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

Algorithm 1: Neural optimal transport (**NOT**)

repeat
- Sample batches $Y\sim\mathbb{Q}$, $X\sim\mathbb{P}$;
- For each $x\in X$ sample batch $Z_x\sim\mathbb{S}$;
- $\mathcal{L}_f\leftarrow\frac{1}{|X|}\sum_{x\in X}\frac{1}{|Z_x|}\sum_{z\in Z_x}f_\omega(T_\theta(x,z))-\frac{1}{|Y|}\sum_{y\in Y}f_\omega(y)$;
- Update ω by using $\frac{\partial\mathcal{L}_f}{\partial\omega}$;
- **for** $k_T=1,2,\dots,K_T$ **do**
  - Sample batches $X\sim\mathbb{P}$;
  - For each $x\in X$ sample batch $Z_x\sim\mathbb{S}$;
  - $\mathcal{L}_T\leftarrow\frac{1}{|X|}\sum_{x\in X}\big[\widehat{C}(x,T_\theta(x,Z_x))-\frac{1}{|Z_x|}\sum_{z\in Z_x}f_\omega(T_\theta(x,z))\big]$;
  - Update θ by using $\frac{\partial\mathcal{L}_T}{\partial\theta}$;

until *not converged*;

Lemma (Minimax reformulation of the dual problem)

$$\text{Cost}(\mathbb{P},\mathbb{Q})=\sup_f\inf_T\mathcal{L}(f,T),$$

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

$$\mathcal{L}(f,T)\stackrel{def}{=}\int_{\mathcal{Y}}f(y)d\mathbb{Q}(y)+\int_{\mathcal{X}}\Big(C(x,T(x,\cdot)_{\#}\mathbb{S})-\int_{\mathcal{Z}}f(T(x,z))d\mathbb{S}(z)\Big)d\mathbb{P}(x).$$

Lemma (Stochastic OT maps solve the problem)

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

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

IV Proposed Weak Kernel Cost Functions

Lemma (Solutions of the maximin problem are OT maps)

*Let $C(x,\mu)$ be a weak cost which is **strictly convex** in μ. Consider any optimal potential $f^*\in\arg\sup_f\inf_T\mathcal{L}(f,T)$. It holds that*

$$\hat{T}\in\arg\inf_T\mathcal{L}(f^*,T)\Leftrightarrow\hat{T}\text{ is a stochastic OT map.}$$

This lemma **does not** guarantee that all the solutions are OT maps. There may exist **fake solutions**.

This lemma **does** guarantee, but only for **strictly** convex transport costs.

(I) weak quadratic cost is **not** strictly convex; (II) entropic cost is **hard to estimate** from samples.

Consider a feature map $u:\mathcal{Y}\to\mathcal{H}$ and the respective PDS kernel $k(y,y')\stackrel{def}{=}\langle u(y),u(y')\rangle_{\mathcal{H}}$. Consider the cost:

$$C_{u,\gamma}(x,\pi(\cdot|x))\stackrel{def}{=}\underbrace{\frac{1}{2}\int_{\mathcal{Y}}\|u(x)-u(y)\|^2_{\mathcal{H}}d\pi(y|x)}_{\text{Feature (kernel) dissimilarity}}-\frac{\gamma}{2}\cdot\underbrace{\Big[\frac{1}{2}\int_{\mathcal{Y}\times\mathcal{Y}}\|u(y)-u(y')\|^2_{\mathcal{H}}\pi(y|x)\pi(y'|x)\Big]}_{\text{Feature (kernel) variance}},\qquad\gamma>0$$

For characteristic kernels, this cost is **strictly** convex and easy to estimate from samples via the **kernel trick**.

V Illustrative Experiments: Diverse Unpaired Translation

Celeba → anime

Outdoor → church   Handbag → shoe   Texture → shoe