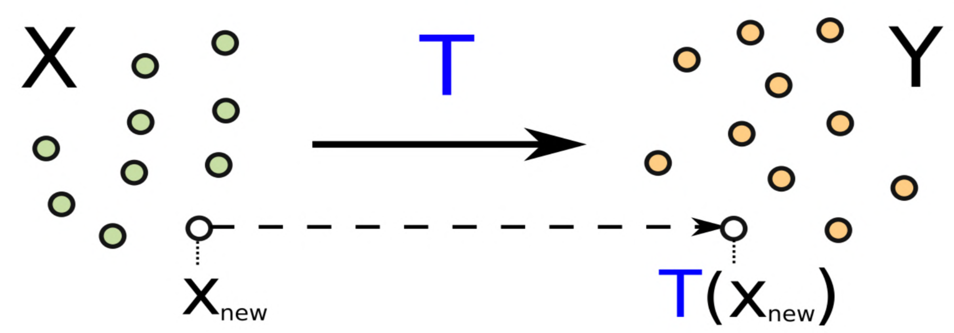
Neural Optimal Transport with General Cost Functionals

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

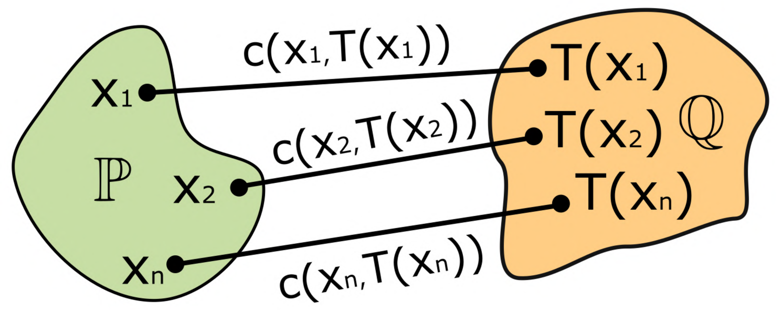
Arip Asadulaev[*1] *Alexander Korotin*[*2,1] *Vage Egiazarian*[3,4] *Petr Mokrov*[2] *Evgeny Burnaev*[2,1]

Yandex [3] HSE University [4]

Background: Classic Optimal Transport



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.



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

The minimizer T* is called the OT map.

Theorem (Maximin reformulation of the general OT)

*For separably *-increasing convex and lower semi-continuous functional $\mathcal{F}:\mathcal{M}(\mathcal{X}\times\mathcal{Y})\rightarrow\mathbb{R}\cup\{+\infty\}$ it holds (we identify $\widetilde{\mathcal{F}}(T)\stackrel{def}{=}\mathcal{F}(\pi_T)$):*

$$\sup_v \inf_T \mathcal{L}(v,T) = \sup_v \inf_T \Big\{\widetilde{\mathcal{F}}(T) - \int_{\mathcal{X}\times\mathcal{Z}} v(T(x,z))d\mathbb{P}(x)d\mathbb{S}(z) + \int_{\mathcal{Y}} v(y)d\mathbb{Q}(y)\Big\},$$

where the sup *is taken over potentials $v\in\mathcal{C}(\mathcal{Y})$ and* inf *– over measurable functions $T:\mathcal{X}\times\mathcal{Z}\rightarrow\mathcal{Y}$.*

Theorem (Optimal saddle points provide stochastic OT maps)

Let $v^*\in\arg\sup_v\inf_T\mathcal{L}(v,T)$ be any optimal potential. Then for every T^* it holds:

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

If $\mathcal{F}$ is strictly convex in π, then the objective above permits the unique OT plan π^*. In this case, $T^*\in\arg\inf_T\mathcal{L}(v^*,T)\Leftrightarrow T^*$ is a stochastic OT map.

Algorithm 1: Neural optimal transport for general cost functionals

Input : Distributions $\mathbb{P},\mathbb{Q},\mathbb{S}$ accessible by samples; mapping network $T_\theta:\mathbb{R}^P\times\mathbb{R}^S\rightarrow\mathbb{R}^Q$; potential network $v_\omega:\mathbb{R}^Q\rightarrow\mathbb{R}$; number of inner iterations K_T; empirical estimator $\widehat{\mathcal{F}}(X,T(X,Z))$ for cost $\widetilde{\mathcal{F}}(T)$;

Output: Learned stochastic map T_θ representing an OT plan between $\mathbb{P},\mathbb{Q}$;

repeat

Sample batches $Y\sim\mathbb{Q}$, $X\sim\mathbb{P}$ and for each $x\in X$ sample batch $Z[x]\sim\mathbb{S}$;

$\mathcal{L}_v\leftarrow\sum_{x\in X}\sum_{z\in Z[x]}\frac{v_\omega(T_\theta(x,z))}{|X|\cdot|Z[x]|}-\sum_{y\in Y}\frac{v_\omega(y)}{|Y|}$;

Update ω by using $\frac{\partial\mathcal{L}_v}{\partial\omega}$;

for $k_T=1,2,\dots,K_T$ **do**

Sample batch $X\sim\mathbb{P}$ and for each $x\in X$ sample batch $Z[x]\sim\mathbb{S}$;

$\mathcal{L}_T\leftarrow\widehat{\mathcal{F}}(X,T_\theta(X,Z))-\sum_{x\in X}\sum_{z\in Z[x]}\frac{v_\omega(T_\theta(x,z))}{|X|\cdot|Z[x]|}$;

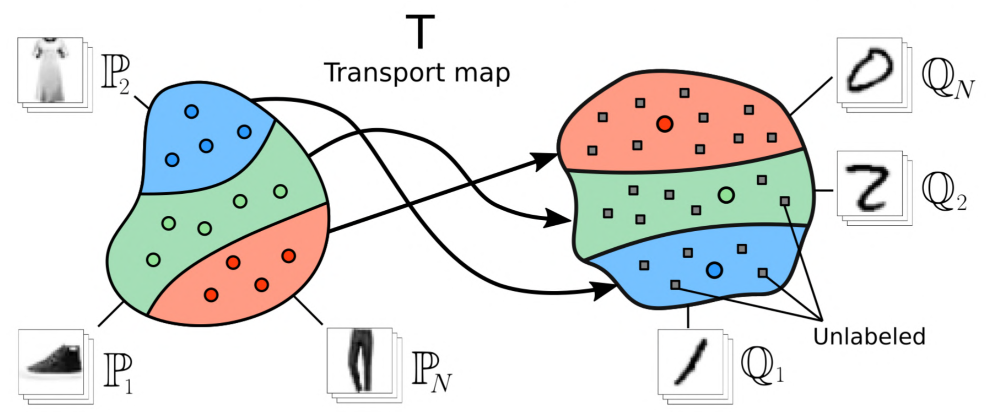
Update θ by using $\frac{\partial\mathcal{L}_T}{\partial\theta}$;

until *not converged*;

Goal: Problems with not Euqlidean Optimality

Case 1: Class-guided domain translation

Given distributions $\mathbb{P}=\sum_n\alpha_n\mathbb{P}_n$, and target $\mathbb{Q}=\sum_n\beta_n\mathbb{Q}_n$, which are mixtures of N classes, the goal is to learn (from samples) a class-preserving *optimal transport* map T between the two given data domains. **Important:** target data $\sim\mathbb{Q}$ only *partially labeled*.

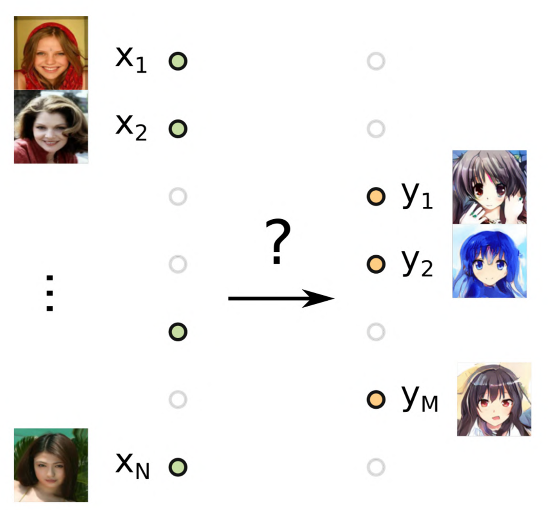


Case 2: Pair-guided domain translation

Unsupervised

Only *unpaired* train samples are given:

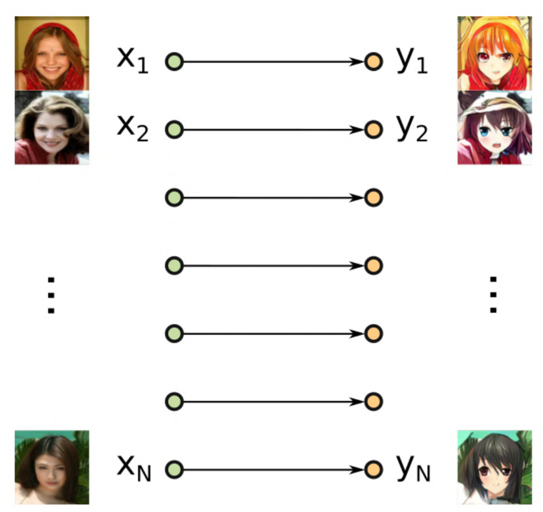
$\{x_1,\dots,x_N\},\{y_1,\dots,y_N\}$.



Semi-Supervised (our paper)

Some *paired* train samples are available:

$\{(x_1,y_1),\dots,(x_N,y_N)\}$.

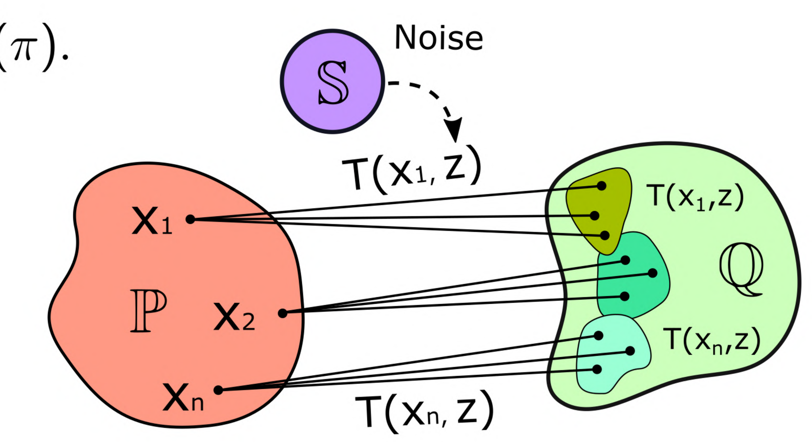


Method: General Neural Optimal Transport

To solve such a problem, we propose a **General Neural Optimal Transport (GNOT)**. Let $\mathcal{F}:\mathcal{M}(\mathcal{X}\times\mathcal{Y})\rightarrow\mathbb{R}\cup\{+\infty\}$ be a convex lower semi-continuous functional. Assume that there exists $\pi\in\Pi(\mathbb{P},\mathbb{Q})$ for which $\mathcal{F}(\pi)<\infty$. Let

$$\text{Cost}(\mathbb{P},\mathbb{Q})\stackrel{def}{=}\inf_{\pi\in\Pi(\mathbb{P},\mathbb{Q})}\mathcal{F}(\pi).$$

This problem is a *generalization* of classic, weak, and regularized optimal transport. To parameterize plan π, we utilize stochastic map $T:\mathcal{X}\times\mathcal{Z}\rightarrow\mathcal{Y}$, s.t. $\pi(\cdot|x)=T(x,\cdot)\sharp\mathbb{S}$, $\mathbb{S}$ is a simple latent distribution.



Class-guided OT (Case 1)

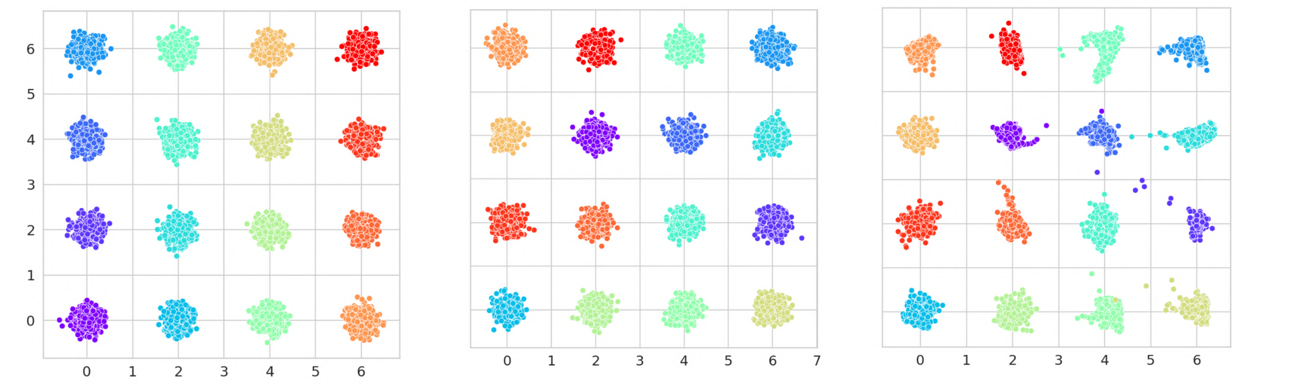
For **class-guided** optimal transport we propose the following functional:

$$\mathcal{F}_G(\pi)=\widetilde{\mathcal{F}}_G(T_\pi)=\sum_{n=1}^N\alpha_n\mathcal{E}^2(T_\pi\sharp(\mathbb{P}_n\times\mathbb{S}),\mathbb{Q}_n),$$

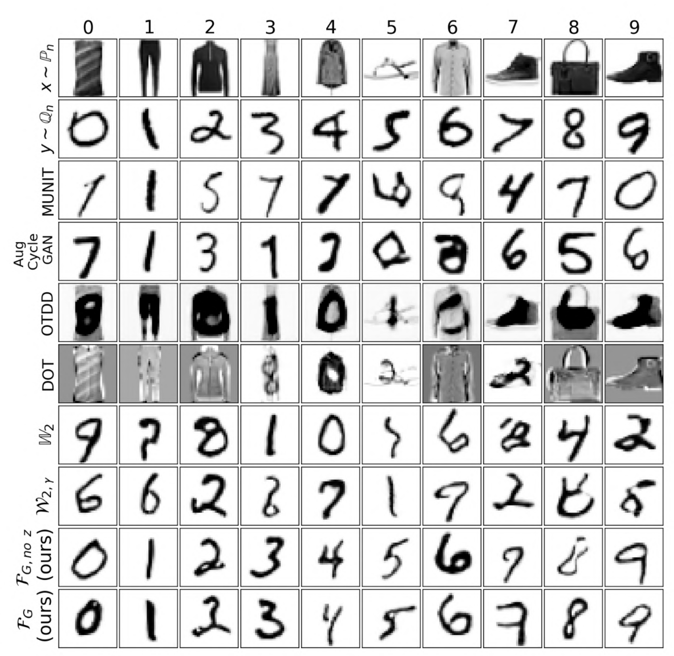
where $\mathcal{E}$ is the (square of) energy distance for $\mathbb{Q},\mathbb{Q}'\in\mathcal{P}(\mathcal{Y})$ with $\mathcal{Y}\subset\mathbb{R}^D$:

$$\mathcal{E}^2(\mathbb{Q},\mathbb{Q}')=\mathbb{E}\|Y_1-Y_2\|_2-\tfrac{1}{2}\mathbb{E}\|Y_1-Y_1'\|_2-\tfrac{1}{2}\mathbb{E}\|Y_2-Y_2'\|_2.$$

Note: we use $\approx$ 10 labelled samples per each class in our experiments.

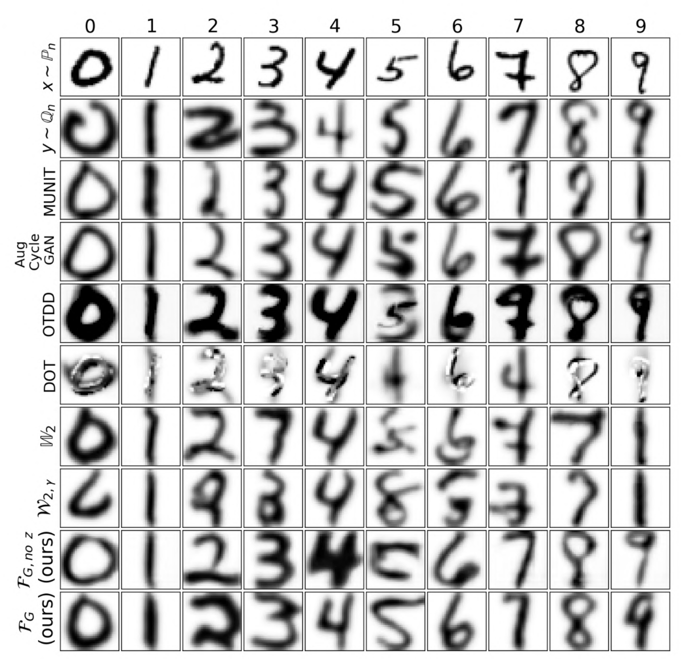


Two Gaussians mixtures mapping

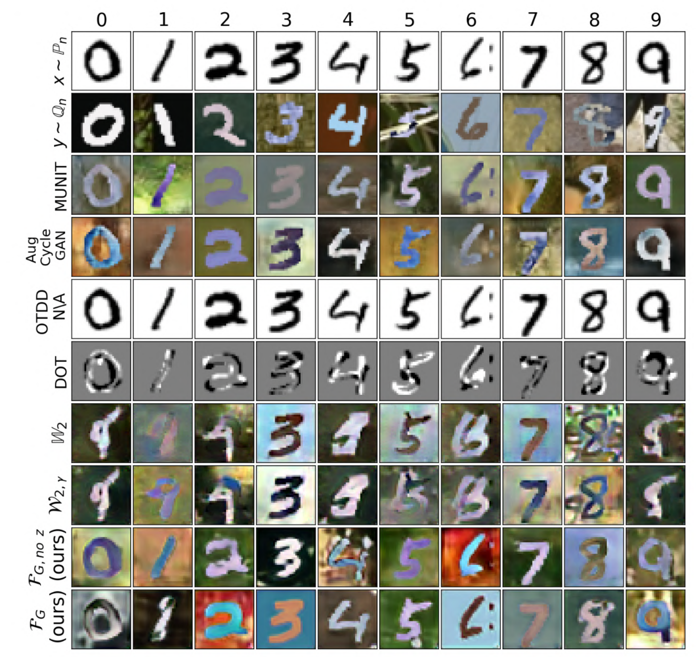


FashionMNIST to MNIST

MNIST to KMNIST



MNIST to USPS



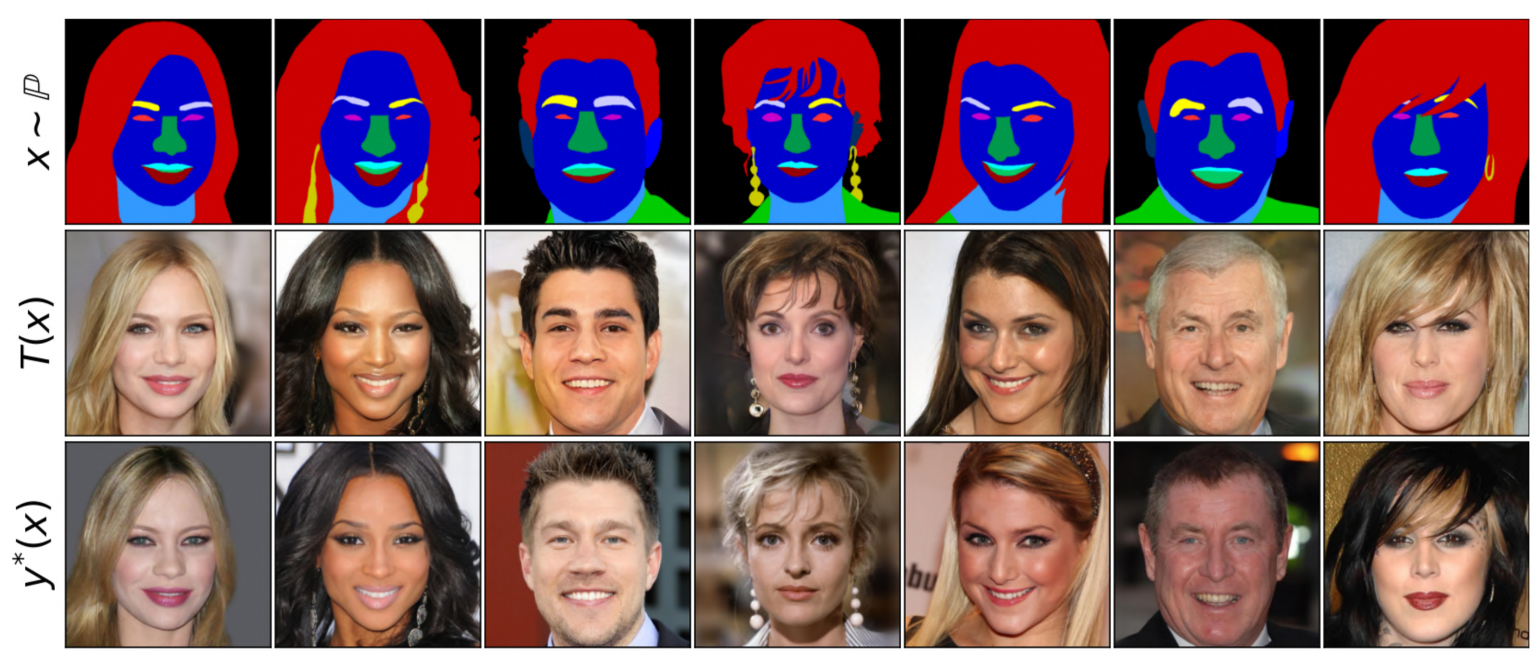
MNIST to MNIST-M

Pair-guided OT (Case 2)

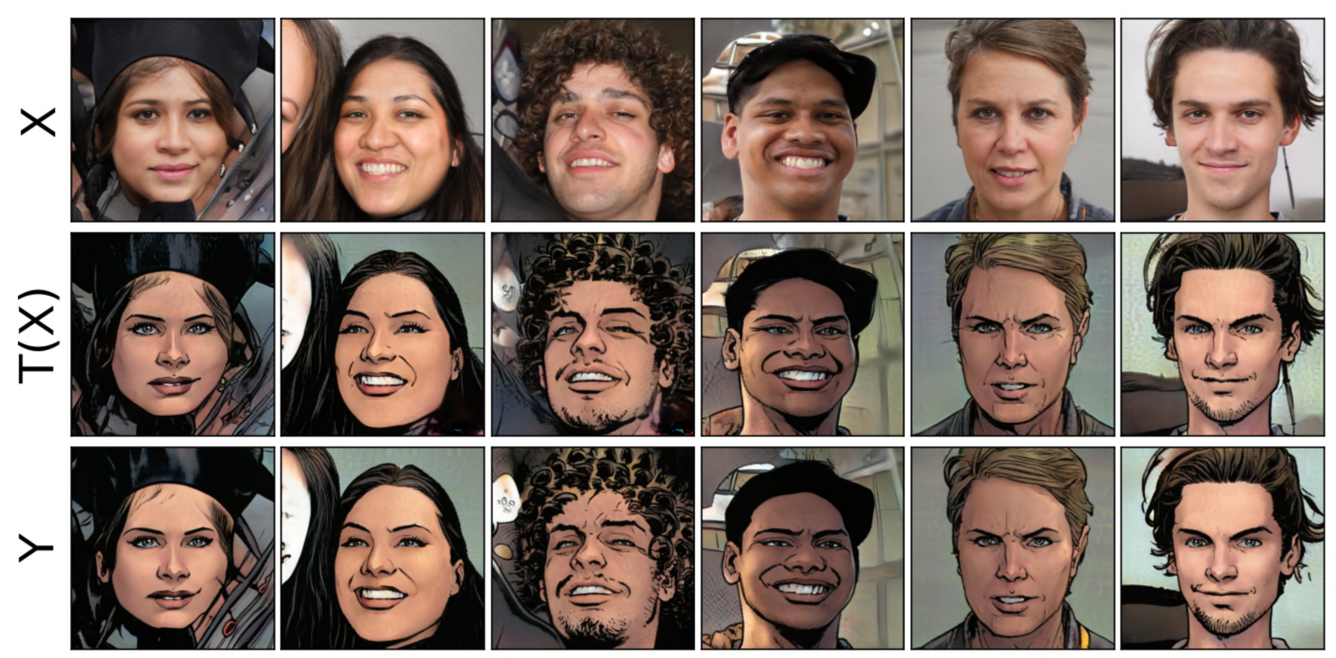
In **pair-guided** optimal transport for a given paired data set $(x_1,y^*(x_1)),\dots,(x_N,y^*(x_N))$ with samples $X_{1:N}=\{x_1,\dots x_N\}$ and $y^*(X_{1:N})=\{y^*(x_1),\dots,y^*(x_N)\}$ we introduce:

$$\mathcal{F}_S(\pi)=\int_{\mathcal{X}\times\mathcal{Y}}\ell(y,y^*(x))d\pi(x,y).$$

The function $\ell:\mathcal{X}\times\mathcal{Y}\rightarrow\mathbb{R}$ is an appropriate loss measuring the difference between samples.



CelebaMaskHQ 256 x 256



ComicFacesV1 512 x 512