Final project: “Stacking for improving neural optimal transport based style-transfer models”

Skoltech

Team:
Nikita Bogdanov
Daniil Panov
Anastasia Gavrish
Nikita Vasilev
Nikolay Kashin
TA: Nikita Gushchin

Unpaired image to image translation:

First image sample

Second image sample [1]

As input: two unpaired data sets
As output: some mapping function

How to find a mapping between input and output set of images?

[1] Korotin A., Selikhanovych D., Burnaev E. (2022). Neural Optimal Transport.

Generative models

First image sample

Mapping

Second image sample [1]

Approaches such as:

GAN

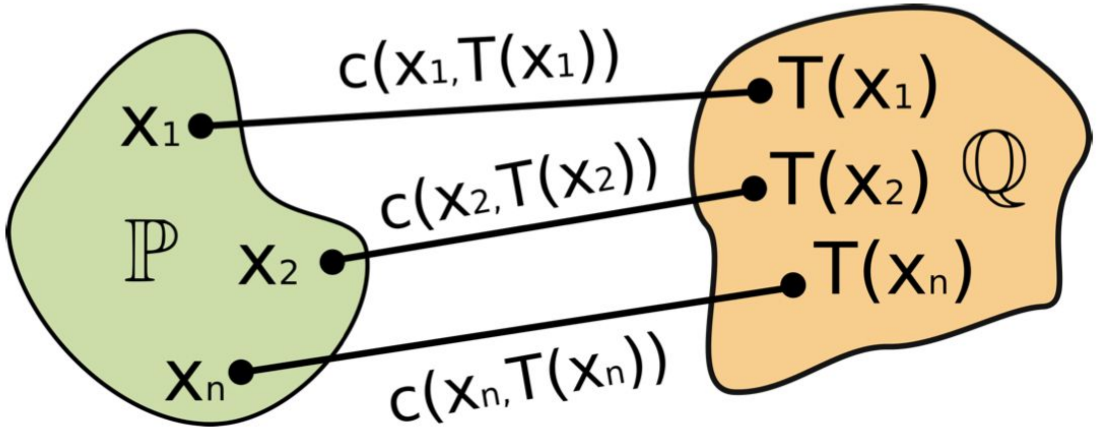
Diffusional models

VAE

Neural Optimal Transport

[1] Korotin A., Selikhanovych D., Burnaev E. (2022). Neural Optimal Transport.

Optimal transport Monge's formulation



Visualisation of Monge's OT formulation [1]

P∈P(X), Q∈P(Y) and a cost function c:X × Y→R, Monge's primal formulation of OT cost is

$$\mathrm{Cost}(\mathbb{P},\mathbb{Q}) \stackrel{\mathrm{def}}{=} \inf_{T_{\#}\mathbb{P}=\mathbb{Q}} \int_{\mathcal{X}} c(x, T(x))\, d\mathbb{P}(x),$$

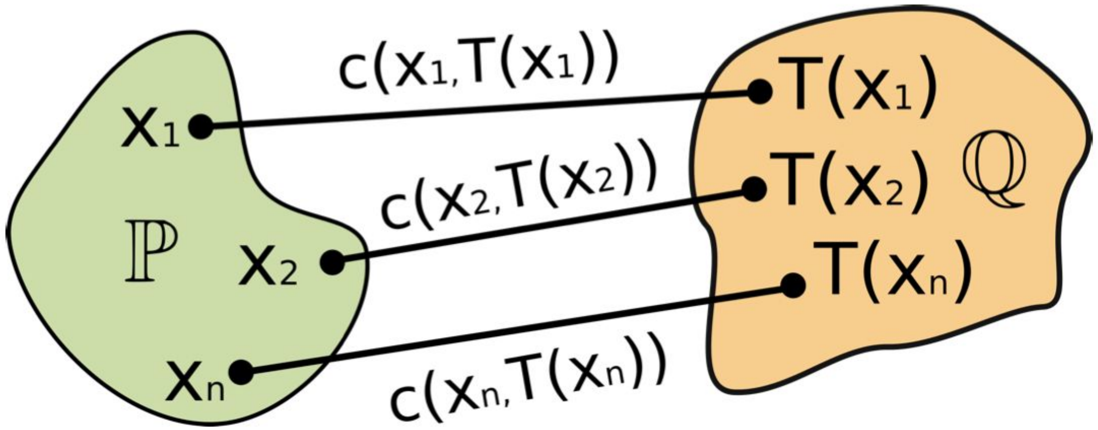
where the minimum is taken over measurable functions (transport maps) T : X → Y that map P to Q.

The optimal T* is called OT map

Skoltech

[1] Korotin A., Selikhanovych D., Burnaev E. (2022). Neural Optimal Transport.

Max min reformulation of OT problem



Visualisation of Monge's OT formulation [1]

Max min reformulation was applied for the problem

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

where

$$\mathcal{L}(f, T) = \int \frac{||x - T\ (x)||_2^2}{2} d\mathbb{P}(x) + \int f(y) d\mathbb{Y} - \int f(T\ (x)) d\mathbb{P}(x)$$

f is an upper-bounded continuous function

Skoltech

[1] Korotin A., Selikhanovych D., Burnaev E. (2022). Neural Optimal Transport.

Max min reformulation of OT problem on practice

$$\sup \inf \mathcal{L}(\omega, \theta) = \int \frac{||x - T_\theta(x)||_2^2}{2} d\mathbb{P}(x) + \int f_\omega(y) d\mathbb{Y} - \int f_\omega(T_\theta(x)) d\mathbb{P}(x)$$

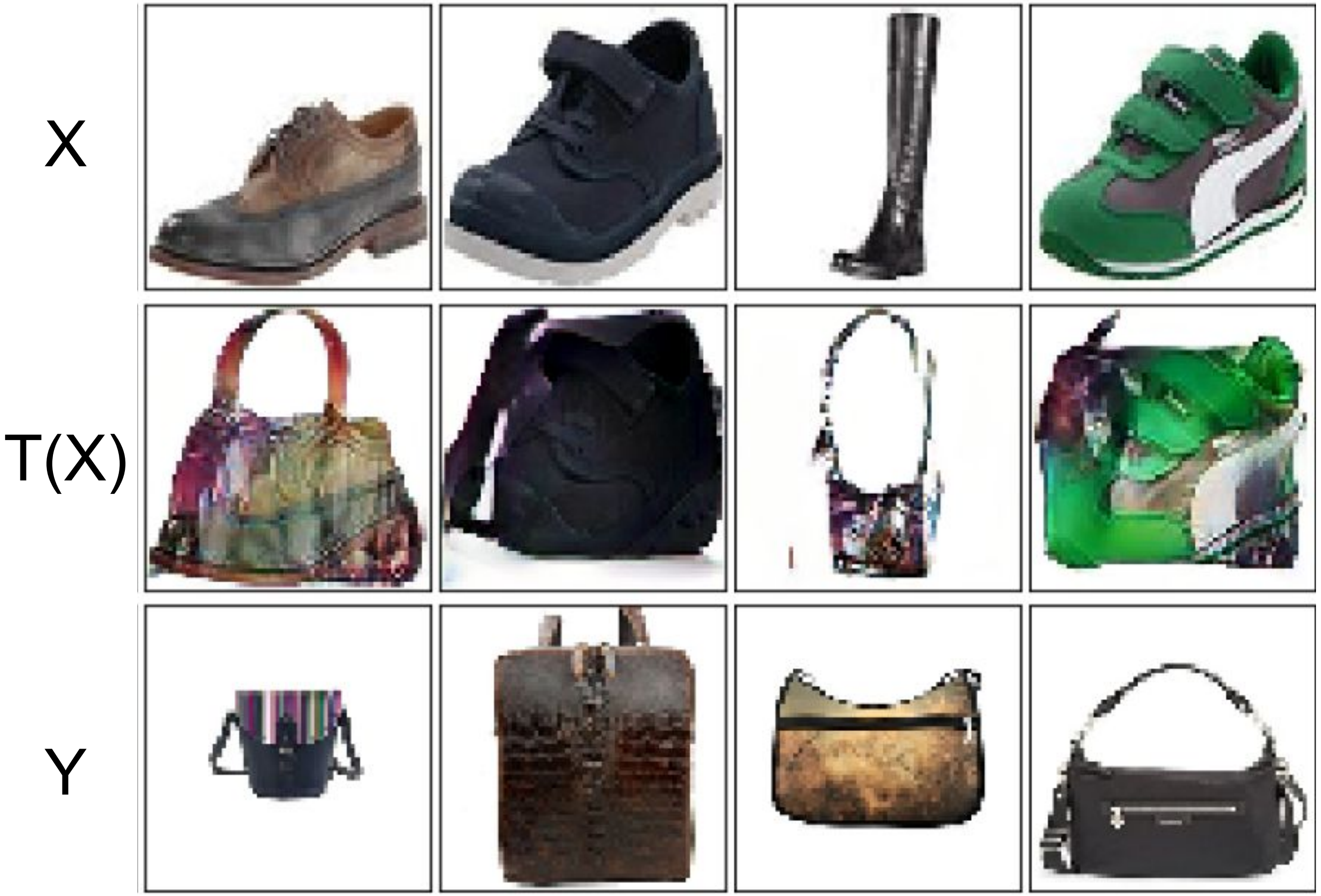
ResNet for kind of discrimination $f_\omega : \mathbb{R}^{3\times H\times W} \to \mathbb{R}$

UNet for kind of generation $T_\theta : \mathbb{R}^{3\times H\times W} \to \mathbb{R}^{3\times H\times W}$

Shoes (3x64x64) dataset mapping to Bags(3x64x64) dataset

Skoltech

Problem statement



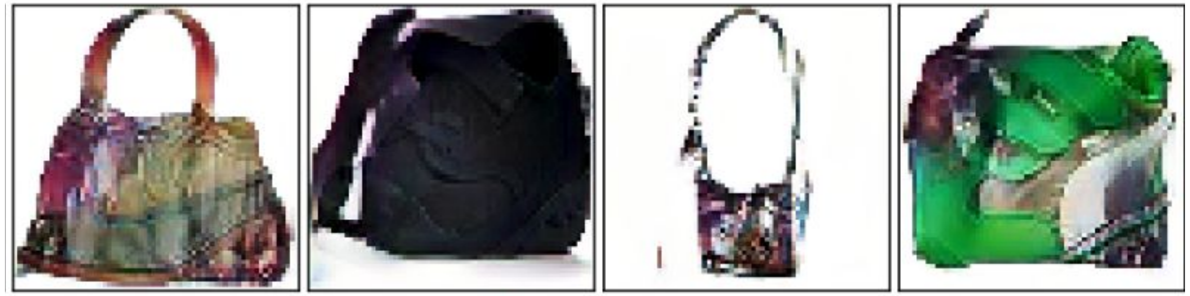
It is possible to find mapping function but the it is not ideal (high FID).

How to improve?

Problem statement and motivation

X

T(X)



T(T(X))

Mapping T(x) on Y

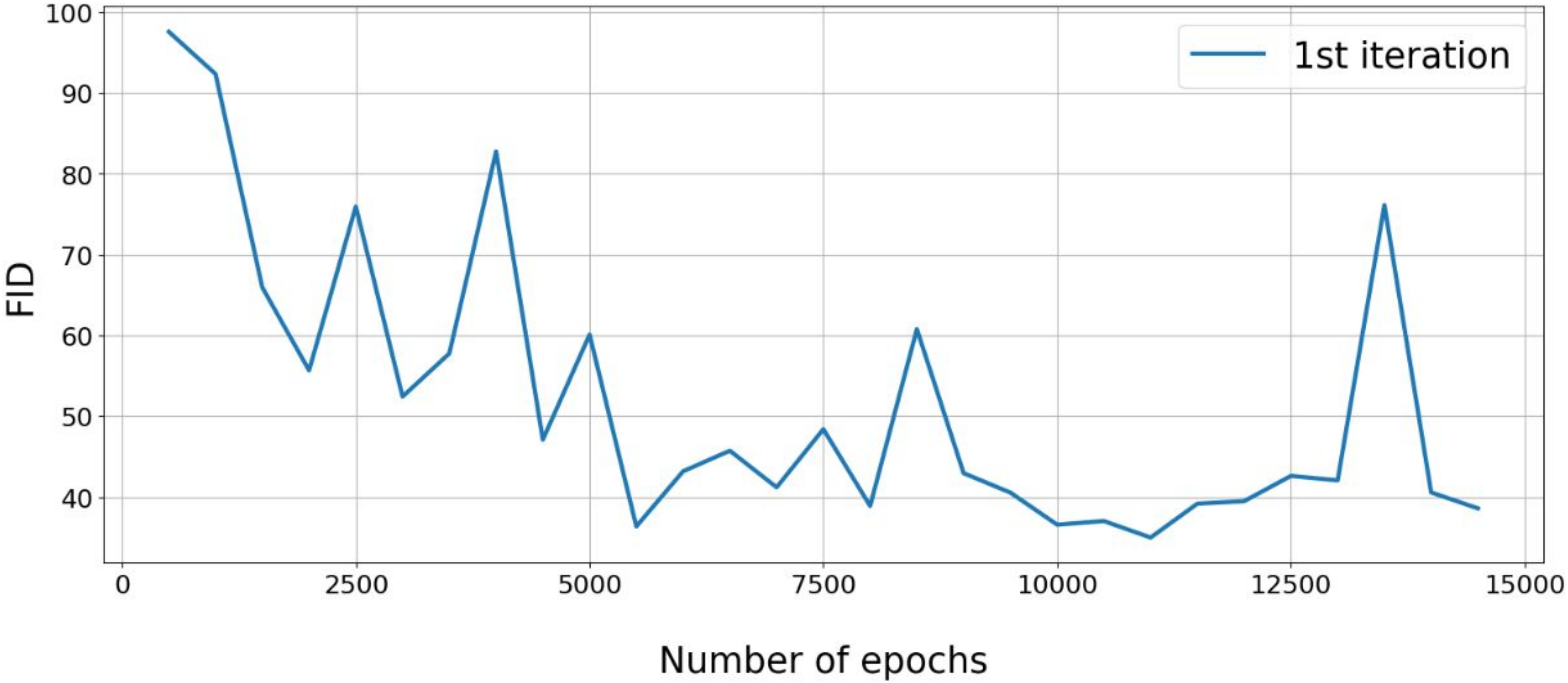
Y

Try to stack more models.

The hypothesis that next generation of models will improve on defects of the previous generation.

Estimated parameter: FID

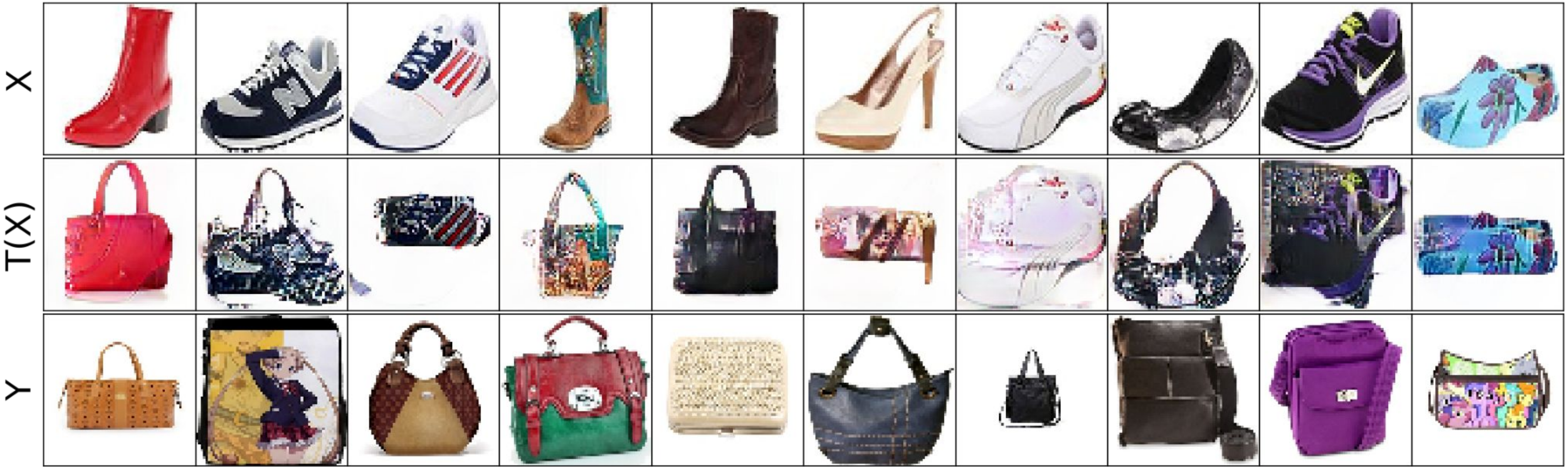
Results zero levels of stacking



Shoes to Bags minimal FID is 34.9.

Skoltech

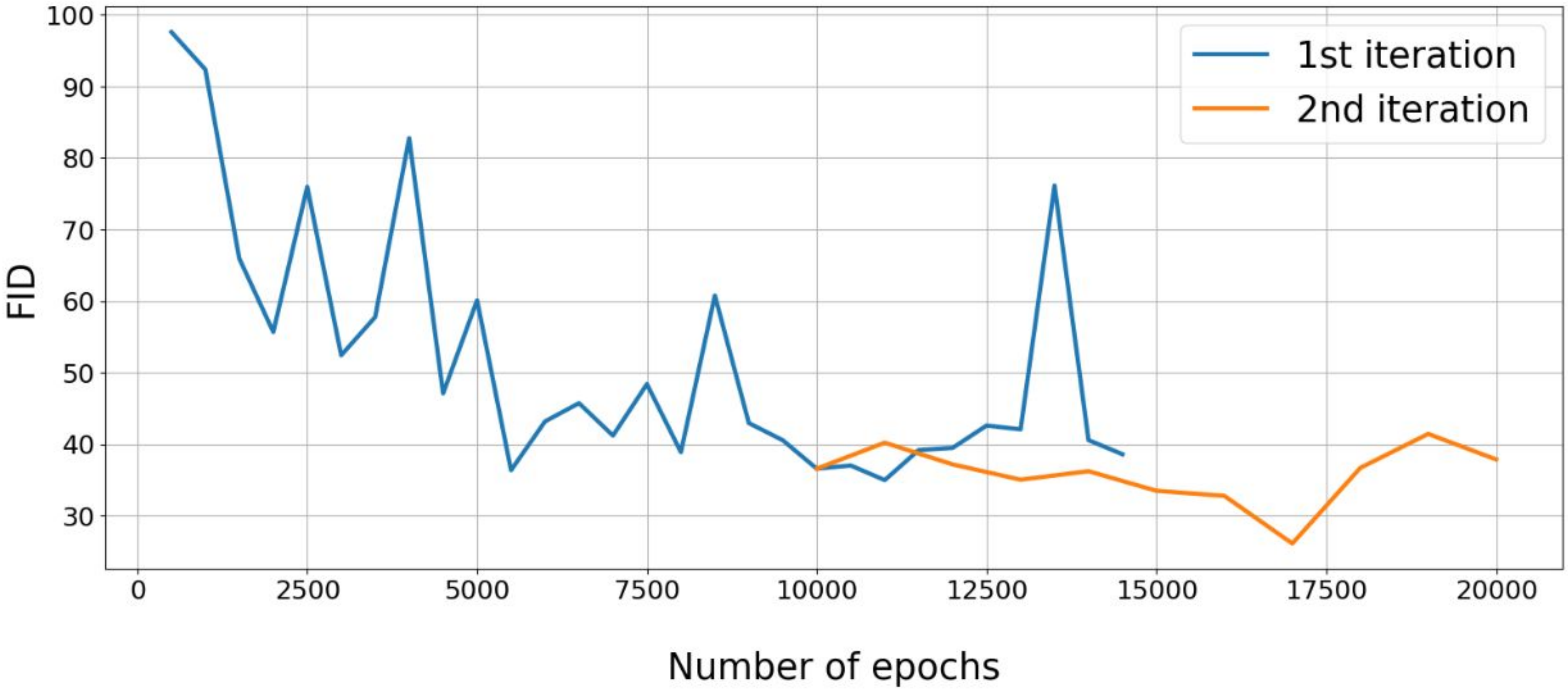
Results zero levels of stacking



Shoes to Bags min FID is 34.9

Skoltech

Results one level of stacking



Bags from shoes to Bags min FID is 26.2 (orange line)

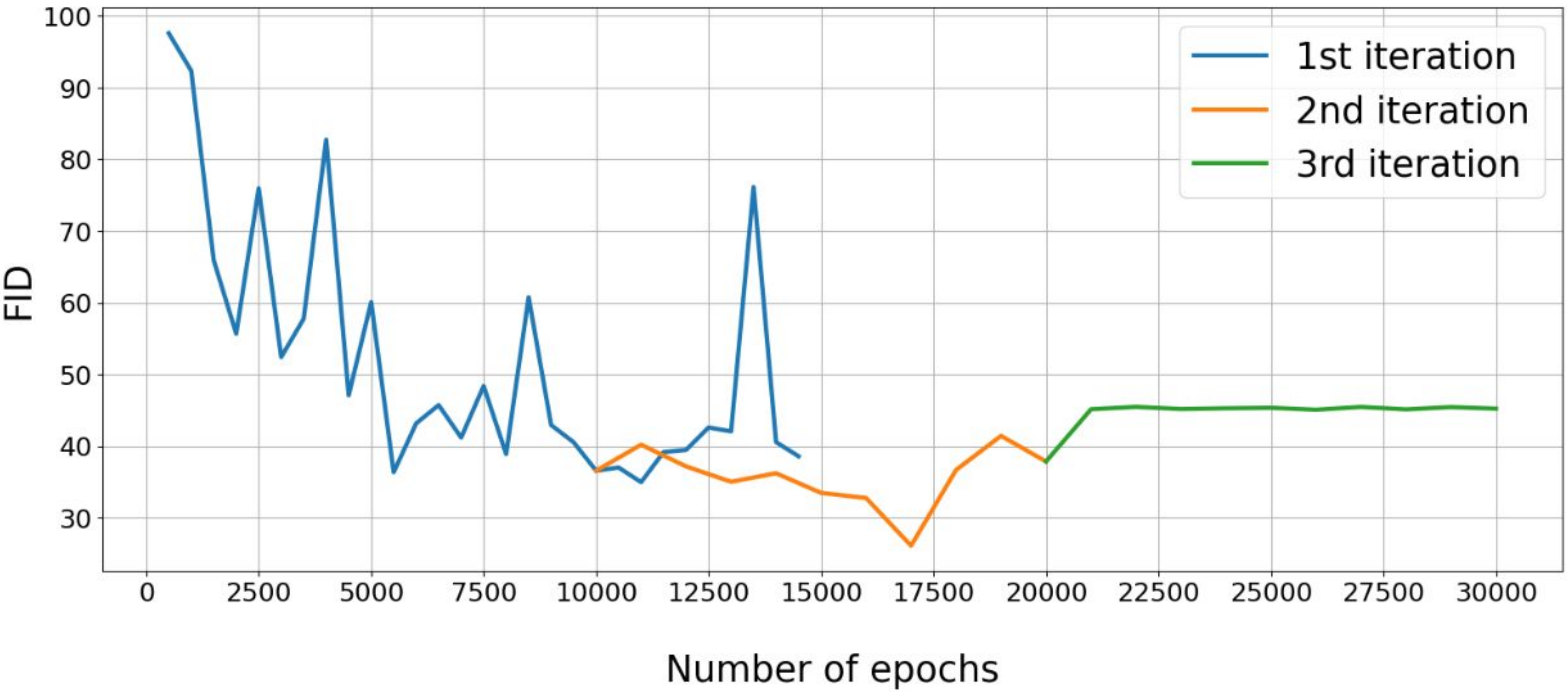
Skoltech

Results one level of stacking

Bags from shoes to Bags min FID is 26.2

Skoltech

Results two levels of stacking



Bags from shoes to Bags min FID is 38.9 (green line)

Skoltech

Results two levels of stacking

Bags from the first stack to Bags min FID is 38.9

Skoltech

Results of stacking 3 models

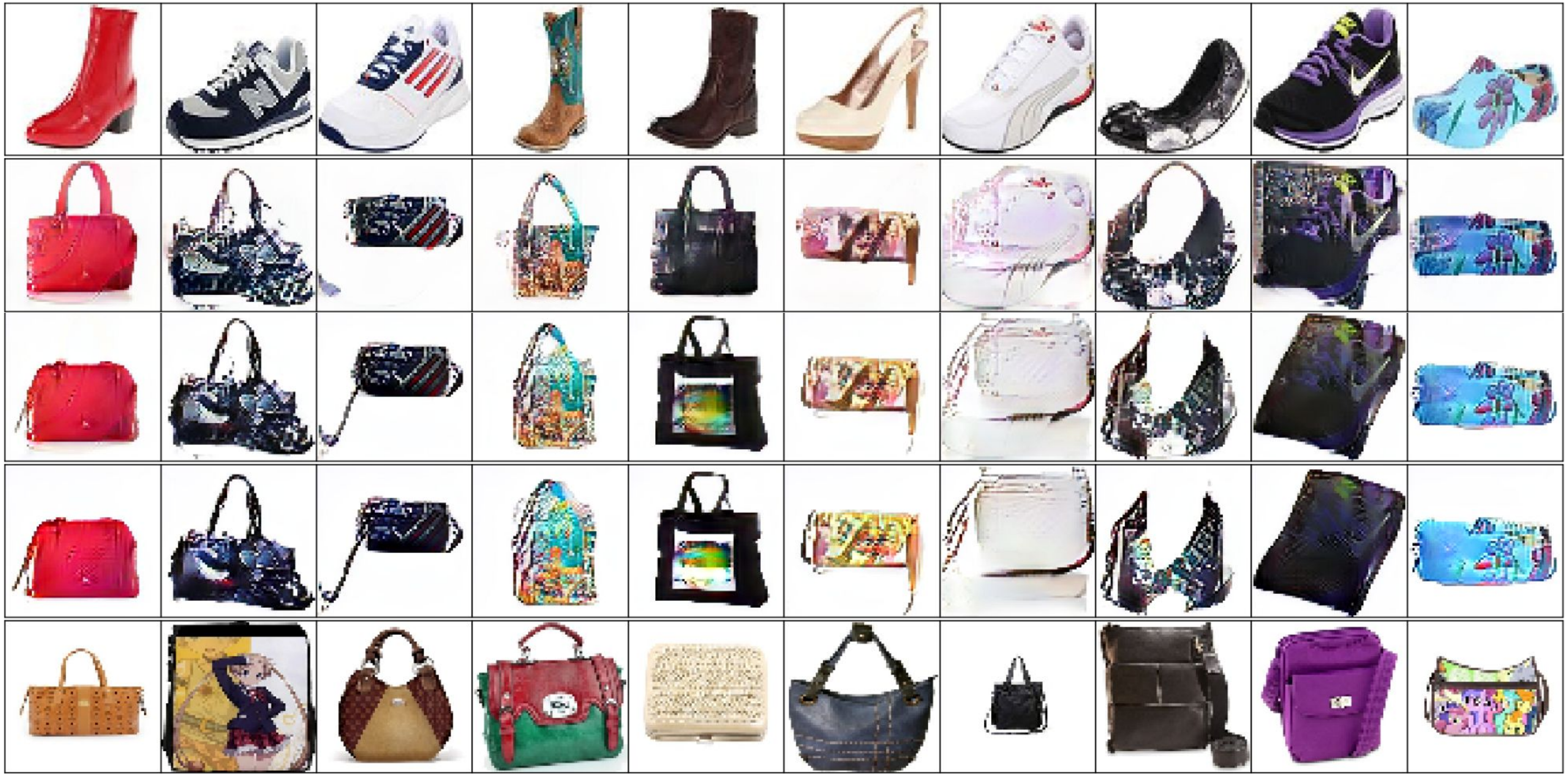
X

$T_1(X)$

$T_2(X)$

$T_3(X)$

Y



Conclusions

- Model stacking was applied to Neural Optimal Transport realisation from Korotin et al. [1] for shoes to handbag image to image translation problem
- One level of stacking helps to obtain minimal FID compare to initial model. However, the second level provides no improvement compare to both initial model and the first level of stacking.

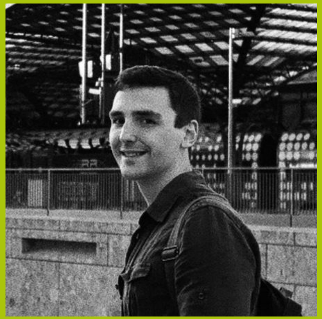
NOT Stacks number	Min FID
0 (no stacking)	34.9
1	26.2
2	38.9

Skoltech

[1] Korotin A., Selikhanovych D., Burnaev E. (2022). Neural Optimal Transport.

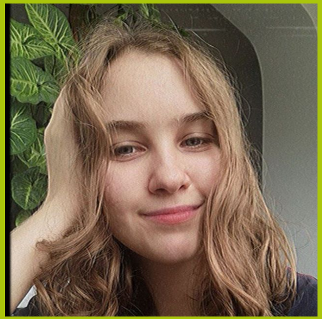
Our team

Our TA
Nikita Gushchin



Daniil Panov

Nikita Vasilev



Anastasia Gavrish

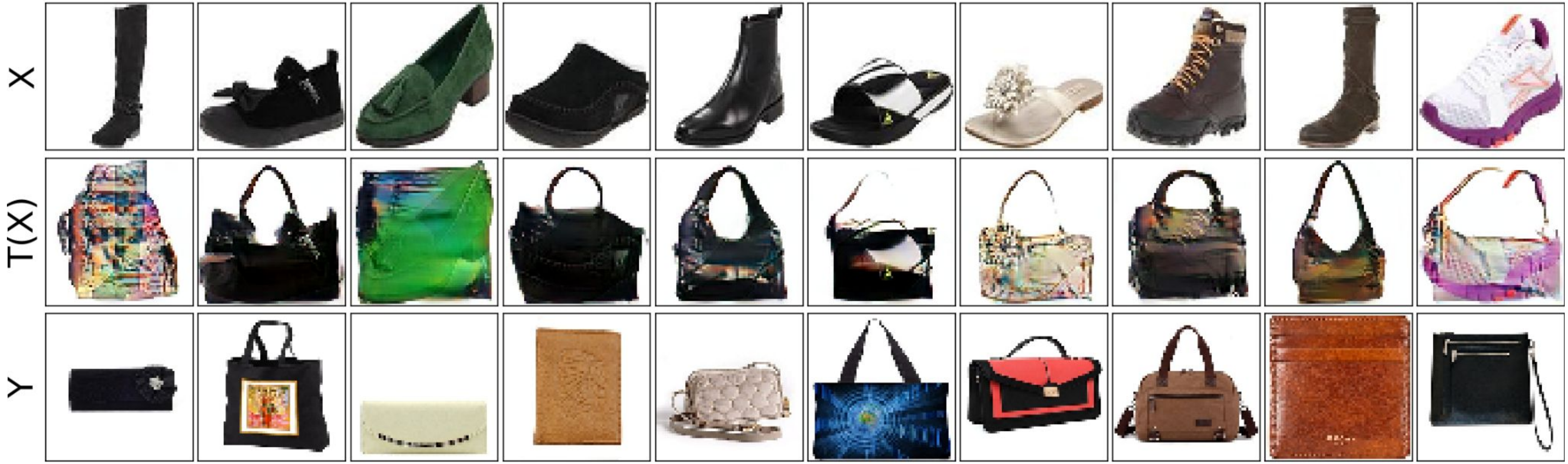
Nikolay Kashin

Nikita Bogdanov

thx.

Skoltech

Results zero levels of stacking



Shoes to Bags min FID is 36.

Results one level of stacking

X

T(X)

T(T(X))

Mapping T(x) on Y

MAKE SIMMILAR PICTURES THAT SHOW IMPROVMENT