Project in BMML: Team with bayesian flavor

A Preprint

Dmitriy Salnikov
Skolkovo Institute of Science and Technology
Moscow, Russia

Aliaksandr Nekrashevich
Skolkovo Institute of Science and Technology
Moscow, Russia

Nurlan Shagadatov
Skolkovo Institute of Science and Technology
Moscow, Russia

October 27, 2018

***K*eywords** GAN, deep learning, optimal transport, VAE, variational autoencoder, autoencoder, generative adversarial network

1 Introduction

Throughout recent years, there's been a lot of research devoted to Generative models, which find a lot of applications in today's applied statistics. But in order to correctly use them you need to know what you're doing. When evaluating their models, researchers often use latent space arithmetics to explore the latent vector space. For objects that are natural for humans(text, images) it's often argued that vectors in that space carry the semantic meaning. Thus, using such transforms in latent space leads to a natural transformation in the original space.

We study the paper [**?**] exploring latent space modifications for generative models. The most known generative models are Generative Adversarial Networks (GANs) and Variational Auto Encoders (VAEs). The problem is that latent space operations can create distributional mismatch, and the paper we study tries to resolve these mismatches with methods of optimal transport.

The rest of the report is organized as follows. In the Theoretical Explanation chapter we summarize the points presented in the paper. In the Workflow description, we describe what cases and extensions we study. In the Experimental Results chapter we present outcomes of our experiments. Later, in Team contribution, we explain what is done by each team member. Finally, references are given.

2 Theoretical Explanation

As briefly mentioned in the introduction part, we are looking to correct the distributional mismatch caused by the transformation applied to the sample. To do that we construct optimal transport map, which is the solution to the Monge transporation problem. The usual way to solve it is to consider the relaxation also known as Monge-Kantorovich problem, but for the case of our interest they coincide. For more information on the OT theory refer to [**?**], [**?**] for a comprehensive description.

Definition 1 (Monge-Kantorovich relaxation)**.** *For a given* $c : \mathcal{X} \times \mathcal{Y} \to R^+$

$$\inf_{p_x, p_y} \{E_{(x,y)} \sim p_{x,y} c(x,y) | (x,y) \sim p_{x,y}, x \sim p_x, y \sim p_y\}$$

Now, we will state theorems that to formulas we use. For proofs refer to [**?**]

Theorem 1 (Decomposition)**.** *If probability measures* p_x, p_y *have i.i.d components and cost function admits*

$$c(x,y) = \sum_{i=1}^{d} C(x^i, y^i)$$

Then, the solution of Monge-Kantorovich problem is component-wise i.e

$$p^{opt}_{x,y}(x,y) = \prod_{i=1}^{d} p^{i_{opt}}_{X,Y}$$

Theorem 2. *If c(x, y) = h(x - y), where $h : \mathbb{R} \to \mathbb{R}^+$, then the optimal transport map from p_x to p_y is given by the*

$$T_{X \to Y}(x) = F_Y^{-1}(F_X(x))$$

Where

- F_X *- CDF of X*
- $F_Y^{-1} = \inf\{t \in \mathbb{R} | F_Y y(t) \geq y\}$ *- pseudo-inverse CDF of Y*

Then, the distributional mismatch is usually addressed in the following steps. First, intuitive operator y is constructed. Then resulting distribution is computed. The following table contains operations y and corresponding matched $\overline{y}$ for $\mathcal{N}(0,1)$

2-point interpolation	$y = tz_1 + (1-t)z_2$	$\overline{y} = \frac{y}{\sqrt{(1-t)^2+t^2}}$
n-point interpolation	$y = \sum_{i=1}^{n} t_i z_i, \sum_{i=1} t_i = 1$	$\overline{y} = \frac{y}{\sqrt{\sum_{i=1}^{n} t_i^2}}$
vicinity sampling	$y = y_0 + \varepsilon u, u \sim N(0, I)$	$\overline{y} = \frac{y}{\sqrt{1+\varepsilon^2}}$
analogies	$y = z_3 + (z_2 - z_1)$	$\overline{y} = \frac{y}{\sqrt{3}}$

Table 1: Operations and distribution preserving maps for $\mathcal{N}(0,1)$

3 Workflow Desription

We have separated the project into three parts:

1. Reproducing the paper. This part is dedicated to GANs. Here we train and adopt DCGAN, compute inception score, evaluate results on different datasets.
2. Applying distribution matching for VAE. Here, we work only with interpolation models.
3. Correcting missing values for VAE. We take initial image, corrupt it in different way, and then reconstruct it using optimal transport distribution matching.

4 Experimental Setup

4.1 Data description

To evaluate and compare our models we use 3 freely available datasets used to train and assess generative models:

- LSUN-bedrooms - collection of images of bedrooms
- CelebA - collection of celebrity face photos
- LLD - collection of logo pictures

4.2 Quality assessment

We compare initial and matched distributions using two approaches

- Qualitative results: Since we obtain images by using the generator, we can just look at them and tell if they're looking good or natural to a human eye or not.
- Quantitative results: But as people doing the project, we may be biased. So as the metric that doesn't depend on our perception of the world, we use Inception Score.

For a given generator the inception score is just

$$\exp E_{p_x} KL(p(y|x)||p(y)), p_x \sim G \text{ where } p(y) = \int p(y|x)p_g(x)dx \approx \frac{1}{N}\sum_{i=1}^{N} p(y_i|x_i); \text{ where } x_i \sim G$$

4.3 DCGAN and VAE models

We use images of size 64x64 and latent space size 100 for the training of DCGAN model, 45x45 deep-funneled images and latent space of size 100 for training of VAE. There is a need to note that for the uniform prior we could not obtain the good model for GAN, so we focus more on Normal distribution.

5 Results

5.1 2 point interpolation

We begin with the classic example of 2-point interpolation: Figure 1 shows three examples per dataset for an interpolation between 2 points in latent space. Each example is first done via matched interpolation (upper raw), then with linear interpolation(lower raw). It is immediately obvious that linear interpolation produces inferior results with generally more blurry, less saturated and less detailed output images. Differences between the two types of interpolation methods for CelebA are much more subtle to the point that they are virtually indistinguishable when viewed side-by-side. This is not an inconsistency though: while distribution mismatch can cause large differences, it can also happen that the model generalizes well enough that it does not matter.

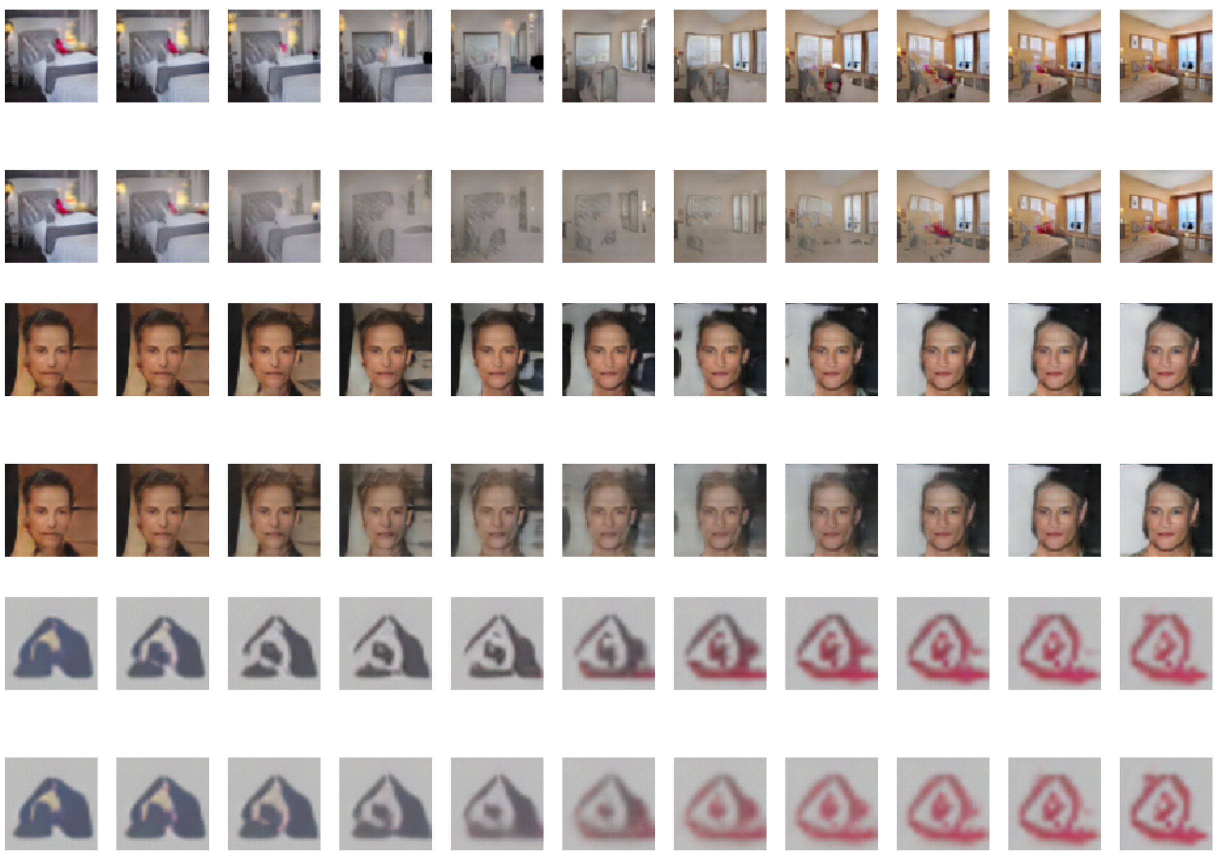

Figure 1: 2 point interpolation

5.2 4 point interpolation

An even stronger effect can be observed when we do 4-point interpolation, showcased in Figure 2 (LSUN) and Figure 4 (LLD icons). In our experiment we constructed interpolation between 4 sampled points which are corresponded to the corners of 2 squares of side 5 images on Figure 2,3, and 4. Left is interpolation using matched operation, right is the linear interpolation. The higher resolution of the LSUN output highlights the very apparent loss of detail and increasing prevalence of artifacts towards the midpoint in the linear version, compared to matched interpolation.

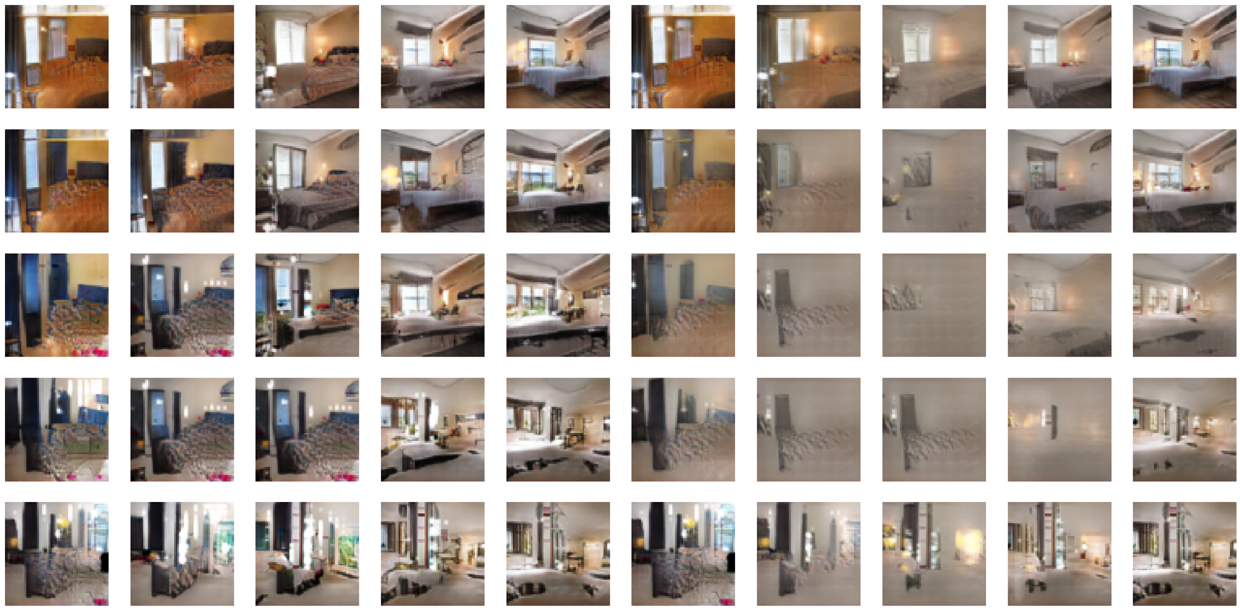

Figure 2: 4 point interpolation (LSUN); left: matched, right: linear

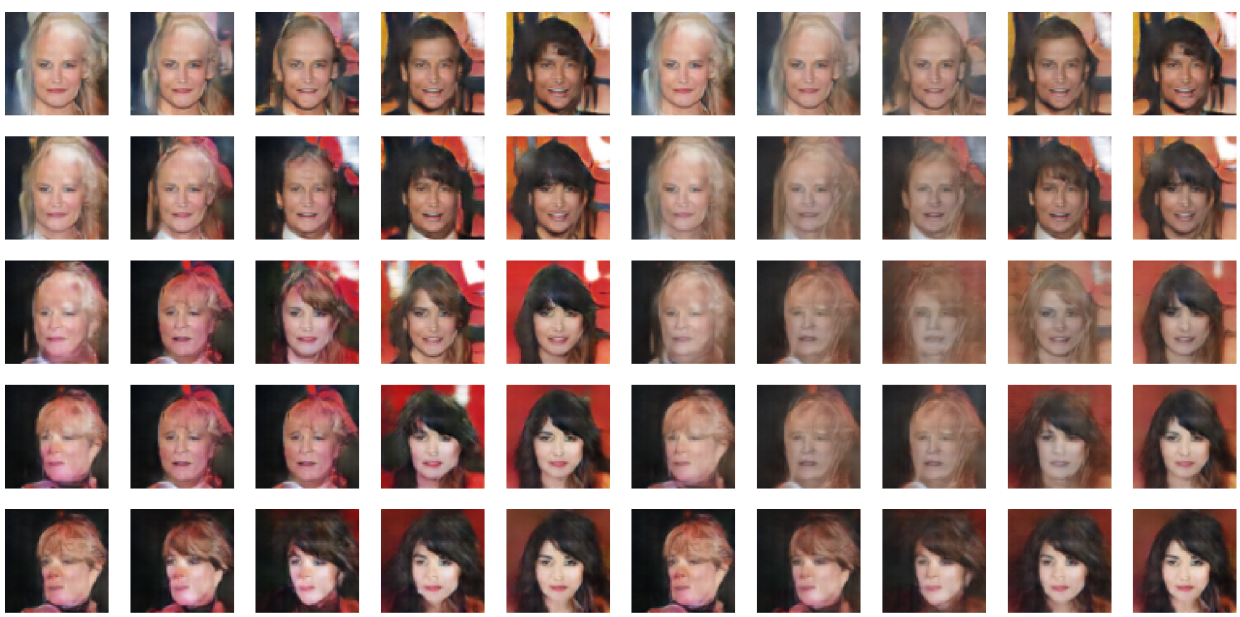

Figure 3: 4 point interpolation (CelebA); left: matched, right: linear

Figure 4: 4 point interpolation (LLD icon); left: matched, right: linear

	CelebA	LSUN	LLD
Usual 2-point	1.799	1.448	1.699
Matched 2-point	2.008	1.662	2.130
Usual 4-point	1.206	1.370	1.658
Matched 4-point	2.052	2.128	3.285

Table 2: 2 point interpolation inception score

5.3 Random walk (vicinity sampling)

An intuitive meaning of latent space can be obtain by a random walk through latent manifold, which leads to the interesting property of matched vicinity sampling. we start at a point $y_0 = z$ drawn from the prior, and then obtain point y_i by sampling a single point in the vicinity of y_{i-1} , using some fixed 'step size' ε. We show an example of such a walk in Figure 5 (left to the right), using ε = 0.5. As a result of the repeated application of the vicinity sampling operation, the divergence from the prior distribution in the non-matched case becomes stronger with each step, resulting in completely unrecognizable output images on the LSUN and LLD icon models. Even for the CelebA model where differences where minimal be- fore, this experiment indicates disadvantages of linear interpolation. The random walk thus perfectly illustrates the need for respecting the prior distribution when performing any operation in latent space, as the error can be cumulate through the repeated application of operators that do not correspond to the prior distribution.

5.4 Inception score

In order to quantitatively confirm the observations of the previous subsection, we computed the Inception score of our trained models (i.e. using random samples from the prior). More precisely, we sampled midpoints from 2-point interpolation and 4-point interpolation described above with 5000 samples and computed the mean value of Inception Score. Compared to the original scores of the trained models, matched operations are statistically indistinguishable (as expected) while the linear interpolation gives a significantly lower score in all settings.

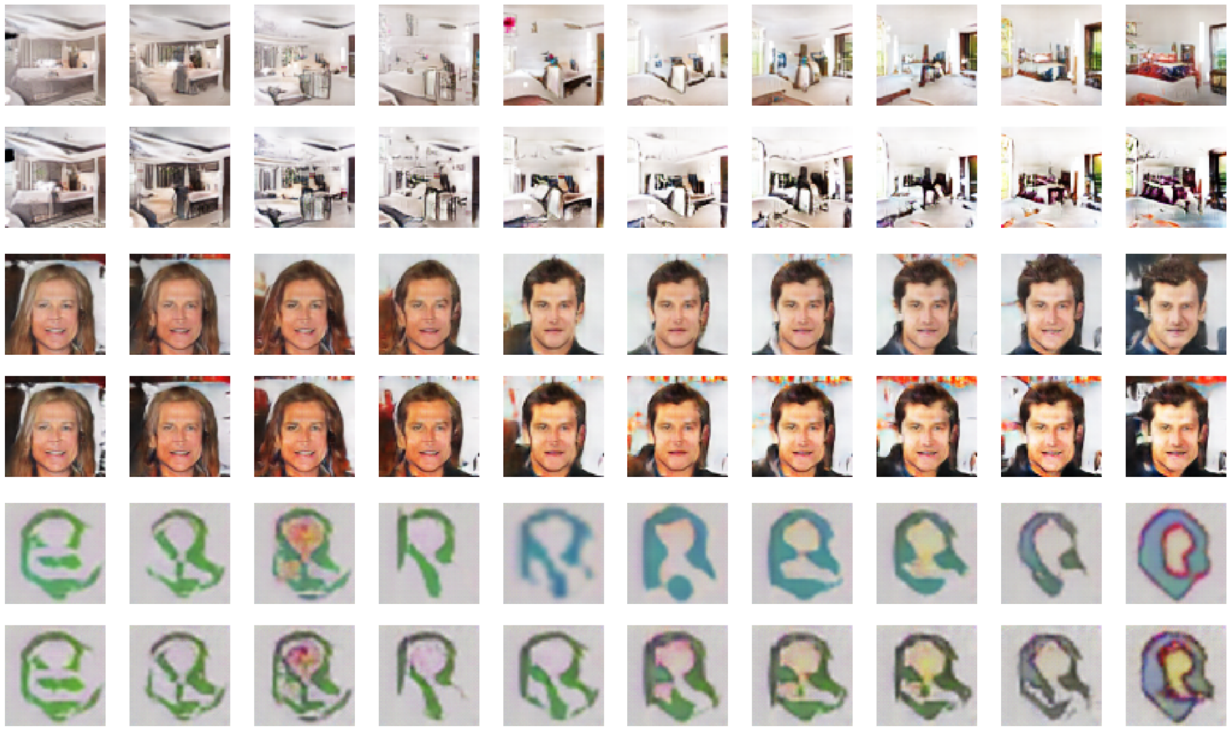

Figure 5: Random walks through latent space

5.5 VAE results

In addition, we conducted the same experiments for Variational Autoencoder trained on Labelled Faces in the Wild Dataset. Unfortunately, the difference between linear and matched interpolations is not obvious, although in some cases the exact interpolation generates clearer (under trained decoder conditions) images. Eventually, we attempted to reconstruct corrupted image using interpolations. Precisely, we took initial image from dataset and corrupted by changing canals of some pixels and reconstructed midpoint interpolation between two latent vectors corresponding to corrupted images. In this case also results are not satisfactory. Corresponding images can be found in Appendix.

6 Conclusions

We notice that it the impact that distribution mismatch has on the result depends on how well the generator is fitted. For the biggest dataset LSUN, we obtain a very good generator, which is very sensitive to the mismatch. If the generator is not that good, the mismatch does not introduce that big of a change, as we can see on examples of VAE or DCGAN on LLD dataset.

7 Team contribution

1. Dmitriy Salnikov: DCGAN experiments, distribution matching operations, presentation, team management, theoretical part of the paper
2. Aliaksandr Nekrashevich: VAE experiments, inception score for DCGANs, report
3. Nurlan Shagadatov: DCGAN experiments, VAE experiments, presentation