Image-to-Image style transfer problem with DDPM

Victor Zelenkovsky [1] **Nikita Manuylenko** [2] **Ivan Kogut**

Abstract

The main advantage of diffusion models is good quality of generating, however these models not usually preserves structure of image in translation tasks. OT and adversarial approaches although don't demonstrate remarkable results of generating, but they perfectly preserve structure of data. There are 2 approaches of DOOM that are used for translation tasks as DEBT and UNIT-DDPM.

This paper provides comparison of diffusion model as DEBT and UNIT DP for translation tasks with adversarial and OT approaches on large-scaled datasets.

Github repo: Image to image DDPM
Presentation: slides

1. Introduction

Transferring images from one domain to another while preserving the content representation is an important problem in computer vision, with wide applications that span style transfer and semantic segmentation. In tasks such as style transfer, it is usually difficult to obtain paired images of realistic scenes and their artistic renditions. Consequently, unpaired translation methods are particularly relevant, sense only the datasets, and not the one-to one correspondence between image translation pairs, are required.

Common methods on unpaired translation are based on generative adversarial networks or normalizing flows. Training such models typically involves minimizing an adversarial loss between a specific pair of source and target datasets.

This article discusses the main approaches to solving the problem of style transfer using diffusion models that allows to make Image-to-image translation in high resolution.

2. Style transfer methods

2.1. DDIB

DDIBs(Xuan Su, 2023) are an image-to-image translation method, based on probability flow ordinary differential equations (PF ODEs) of generative diffusion models. DDIBs enable independent model training, guarantee exact cycle-consistent translation, and produce impressive results on high-resolution image datasets.

Common image translation methods are based on generative adversarial networks (GANs) and normalizing flows. They learn to translate images by optimizing a loss that requires concurrent access to both the source and target datasets. What are potential problems of this training approach?

- Adaptability: the resulting models cannot be easily applied to other source-target pairs.
- Data Privacy: the training process requires simultaneous access to both datasets, which makes it impossible to separate the datasets and protect data privacy.

DDIBs solve both problems! DDIBs rely on diffusion models trained independently on the two domains.

In detail, DDIBs rely on so-called probability flow ordinary differential equations (PF ODEs), that are a specific, deterministic way of navigating the diffusion process. While there are many paths to generate samples from the latent space, defined via stochastic differential equations; among them, there is one, unique ODE that shares the same marginal densities across time as the SDEs. In graphical terms, the ODEs are the red trajectories below.

The DDIBs translation process is equivalent to flowing from the source space, to the latent space, and then to the target, via the red lines / bridges.

DDIBs guarantee exact cycle consistency.

A desirable property of image translation methods is the cycle consistency property: translating an image from the source to the target domain, and then back to the source domain, recovers the original image.

DDIBs guarantee cycle consistency. Indeed, if we were to travel through the red lines (see Fig.1), from left to right and

[1] Belarusian Sate University, Minsk, Belarus [2] Higher School of Economics, Moscow, Russia. Correspondence to: Victor Zelenkovsky <victor.zelenkovsky@gmail.com>, Nikita Manuylenko <nmanuylenko@hse.ru>.

Final Projects of the Machine Learning 2020 Course, Skoltech, Moscow, Russian Federation, 2020.

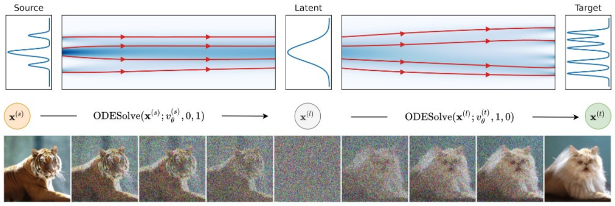

Source
Latent
Target
$\mathbf{x}^{(s)}$
ODESolve($\mathbf{x}^{(s)}; v_\theta^{(s)}, 0, 1$)
$\mathbf{x}^{(l)}$
ODESolve($\mathbf{x}^{(l)}; v_\theta^{(t)}, 1, 0$)
$\mathbf{x}^{(t)}$

Figure 1. DDIB translation process

then right to left, across the three spaces, we are certain to arrive at the original positions.

Exact cycle consistency of DDIBs is empirically validated by 2D synthetic translation experiments, shown in the following figures.

DDIBs are intrinsically optimal transport bridges.

Translation process of DDIBs is closely related to optimal transport: seeking the cost-minimizing method to transport masses from one distribution to another. In particular, DDIBs are two concatenated Schrödinger Bridges. These are a special type of optimal transport with entropy regularization.

To build the intuition, let's inspect the following visualizations, with the leftmost column being the source images and the rightmost being the targets. Clearly, DDIBs minimize pixel-wise distances between the image pairs, morphing the edges, hands, and shapes into tree branches and suitable parts of the dog body, that the diffusion models consider appropriate, and are closest in distances to the original shapes.

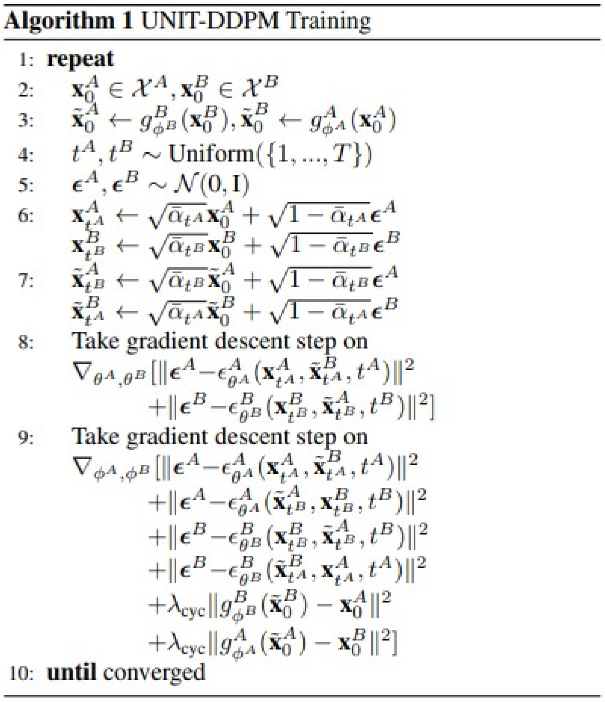
2.2. UNIT-DDPM

The next method id UNIT-DDPM(Hiroshi Sasaki, 2021), which is a modification of DDPM(Alex Nichol, 2021). It's aim is to develop I2I translation between different domains (unpaired). The method needs to learn the parameters of the models from a given dataset of source and target domains via empirical risk minimisation and subsequently be able to infer the target domain images from the corresponding source domain images.

Model Training Assuming a source domain ${x_0}^A \in X^A$ and a target domain ${x_0}^B \in X^B$, UNIT-DDPM interatively optimise the reverse process in each domain ${p^A}_{\theta^A}$, ${p^B}_{\theta^B}$ (neural networks). For more details see Fig.2.

Model Inference

Using the trained θ^A, θ^B, the input images are translated from source to target domain. The domain translation functions are no longer used in inference. Instead, the target domain images are progressively synthesised from Gaussian noise and the noisy source domain images. During sampling, the generative process is conditioned on the input source domain images that are perturbed by the forward process from $t = T$ until an arbitrary timtestep $t_r \in [1, T]$. For more details see Fig.3.



```
Algorithm 1 UNIT-DDPM Training
repeat
    $\mathbf{x}_0^A \in \mathcal{X}^A, \mathbf{x}_0^B \in \mathcal{X}^B$
    $\tilde{\mathbf{x}}_0^A \leftarrow g_{\phi^B}^B(\mathbf{x}_0^B), \tilde{\mathbf{x}}_0^B \leftarrow g_{\phi^A}^A(\mathbf{x}_0^A)$
    $t^A, t^B \sim \text{Uniform}(\{1, ..., T\})$
    $\boldsymbol{\epsilon}^A, \boldsymbol{\epsilon}^B \sim \mathcal{N}(0, \text{I})$
    $\mathbf{x}_{t^A}^A \leftarrow \sqrt{\bar{\alpha}_{t^A}}\mathbf{x}_0^A + \sqrt{1 - \bar{\alpha}_{t^A}}\boldsymbol{\epsilon}^A$
    $\mathbf{x}_{t^B}^B \leftarrow \sqrt{\bar{\alpha}_{t^B}}\mathbf{x}_0^B + \sqrt{1 - \bar{\alpha}_{t^B}}\boldsymbol{\epsilon}^B$
    $\tilde{\mathbf{x}}_{t^B}^A \leftarrow \sqrt{\bar{\alpha}_{t^B}}\tilde{\mathbf{x}}_0^A + \sqrt{1 - \bar{\alpha}_{t^B}}\boldsymbol{\epsilon}^A$
    $\tilde{\mathbf{x}}_{t^A}^B \leftarrow \sqrt{\bar{\alpha}_{t^A}}\tilde{\mathbf{x}}_0^B + \sqrt{1 - \bar{\alpha}_{t^A}}\boldsymbol{\epsilon}^B$
    Take gradient descent step on
    $\nabla_{\theta^A, \theta^B}[\|\boldsymbol{\epsilon}^A - \epsilon_{\theta^A}^A(\mathbf{x}_{t^A}^A, \tilde{\mathbf{x}}_{t^A}^B, t^A)\|^2$
            $+\|\boldsymbol{\epsilon}^B - \epsilon_{\theta^B}^B(\mathbf{x}_{t^B}^B, \tilde{\mathbf{x}}_{t^B}^A, t^B)\|^2]$
    Take gradient descent step on
    $\nabla_{\phi^A, \phi^B}[\|\boldsymbol{\epsilon}^A - \epsilon_{\theta^A}^A(\mathbf{x}_{t^A}^A, \tilde{\mathbf{x}}_{t^A}^B, t^A)\|^2$
            $+\|\boldsymbol{\epsilon}^A - \epsilon_{\theta^A}^A(\tilde{\mathbf{x}}_{t^B}^A, \mathbf{x}_{t^B}^B, t^B)\|^2$
            $+\|\boldsymbol{\epsilon}^B - \epsilon_{\theta^B}^B(\mathbf{x}_{t^B}^B, \tilde{\mathbf{x}}_{t^B}^A, t^B)\|^2$
            $+\|\boldsymbol{\epsilon}^B - \epsilon_{\theta^B}^B(\tilde{\mathbf{x}}_{t^A}^B, \mathbf{x}_{t^A}^A, t^A)\|^2$
            $+\lambda_{\text{cyc}}\|g_{\phi^B}^B(\tilde{\mathbf{x}}_0^B) - \mathbf{x}_0^A\|^2$
            $+\lambda_{\text{cyc}}\|g_{\phi^A}^A(\tilde{\mathbf{x}}_0^A) - \mathbf{x}_0^B\|^2]$
until converged
```

Figure 2. UNIT-DDPM training process

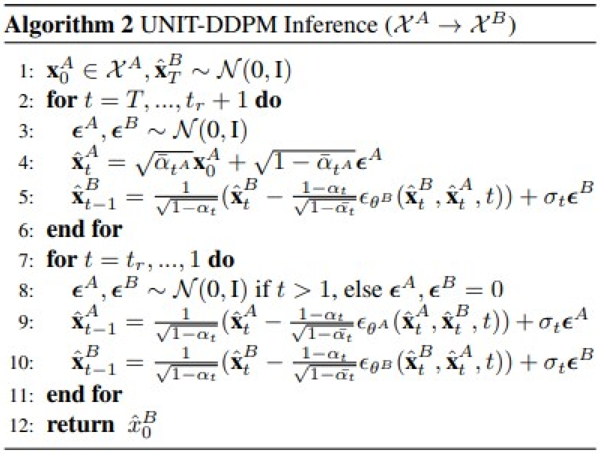
2.3. CycleGAN

(Jun-Yan Zhu, 2017) Let X and Y be two different domains. CycleGAN model contains two mapping functions $G : X \rightarrow Y$ and $F : Y \rightarrow X$, and associated adversarial discriminators D_Y and D_X. D_Y encourages G to translate X into outputs indistinguishable from domain Y, and vice versa for D_X and F. To further regularize the mappings, two cycle consistency losses are introduced, that capture the intuition that if we translate from one domain to the other and back again we should arrive at where we started: forward cycle-consistency loss: $x \rightarrow G(x) \rightarrow F(G(x))$ should approximately be equivalent to x, and backward cycle-consistency loss: $y \rightarrow F(y) \rightarrow G(F(y))$ should approximately be equivalent to y.

3. Experiments and results

Github repo contains all acquired, albeit limited, results.

- DDIB: No notable results were achieved at the moment,



Algorithm 2 UNIT-DDPM Inference ($\mathcal{X}^A \rightarrow \mathcal{X}^B$)

```
$\mathbf{x}_0^A \in \mathcal{X}^A, \hat{\mathbf{x}}_T^B \sim \mathcal{N}(0, \mathrm{I})$
**for** $t = T, ..., t_r + 1$ **do**
    $\boldsymbol{\epsilon}^A, \boldsymbol{\epsilon}^B \sim \mathcal{N}(0, \mathrm{I})$
    $\hat{\mathbf{x}}_t^A = \sqrt{\bar{\alpha}_{t^A}}\mathbf{x}_0^A + \sqrt{1 - \bar{\alpha}_{t^A}}\boldsymbol{\epsilon}^A$
    $\hat{\mathbf{x}}_{t-1}^B = \frac{1}{\sqrt{1-\alpha_t}}(\hat{\mathbf{x}}_t^B - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_{\theta^B}(\hat{\mathbf{x}}_t^B, \hat{\mathbf{x}}_t^A, t)) + \sigma_t\boldsymbol{\epsilon}^B$
**end for**
**for** $t = t_r, ..., 1$ **do**
    $\boldsymbol{\epsilon}^A, \boldsymbol{\epsilon}^B \sim \mathcal{N}(0, \mathrm{I})$ if $t > 1$, else $\boldsymbol{\epsilon}^A, \boldsymbol{\epsilon}^B = 0$
    $\hat{\mathbf{x}}_{t-1}^A = \frac{1}{\sqrt{1-\alpha_t}}(\hat{\mathbf{x}}_t^A - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_{\theta^A}(\hat{\mathbf{x}}_t^A, \hat{\mathbf{x}}_t^B, t)) + \sigma_t\boldsymbol{\epsilon}^A$
    $\hat{\mathbf{x}}_{t-1}^B = \frac{1}{\sqrt{1-\alpha_t}}(\hat{\mathbf{x}}_t^B - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_{\theta^B}(\hat{\mathbf{x}}_t^B, \hat{\mathbf{x}}_t^A, t)) + \sigma_t\boldsymbol{\epsilon}^B$
**end for**
**return** $\hat{x}_0^B$
```

Figure 3. UNIT-DDPM inference process

generated image from one domain still had its previous shape, although coloration was subject to scrambling due to the noising during transfer.

- UNIT-DDPM: Main result from running the model is just a noise even when running it through a decoder, although on simple datasets, such as color-mnist, for example, there can be seen some patterns, slightly resembling the shape of the image from the desired domain.

- CycleGAN: Main result is a slight change in the images of domain during transfer to other domains. No full gender changes are observable at that level of training.

4. Conclusion

The article considered such methods of solving the style transfer problem as DDIB, UNIT-DDPM and CycleGAN. We have analyzed the ideas and model architectures of each of these methods. Next, we conducted experiments reflecting their work.

References

Alex Nichol, P. D. Improved denoising diffusion probabilistic models. *https://arxiv.org/abs/2102.09672*, 2021.

Hiroshi Sasaki, Chris G. Willcocks, T. P. B. Unit-ddpm: Unpaired image translation with denoising diffusion probabilistic models. *https://arxiv.org/pdf/2104.05358.pdf*, 2021.

Jun-Yan Zhu, Taesung Park, P. I. A. A. E. Unpaired image-to-image translation using cycle-consistent adversarial networks. *https://arxiv.org/pdf/1703.10593.pdf*, 2017.

Xuan Su, Jiaming Song, C. M. S. E. Dual diffusion implicit bridges for image-to-image translation. *https://openreview.net/forum?id=5HLoTvVGDe*, 2023.

A. Team member's contributions

Explicitly stated contributions of each team member to the final project.

Victor Zelenkovsky (35% of work)

- Reviewing literate on the topics of DDPM and UNIT-DDPM(2 papers)
- Analysis of UNIT-DDPM code implemenation
- Running of UNIT-DDPM on colorized MNIST and CelebA datasets
- Preparing the Sections 1, 2.1, 4 of this report

Nikita Manuylenko (35% of work)

- Reviewing literate mainly on the topics of CycleGAN and NOT. (2 papers)
- Running CycleGAN on celebA dataset for male-to-female image transfer.
- Attempted to run NOT on img_align_celeba dataset. (unsuccessfully).
- Preparing the GitHub Repo
- Preparing the Section 2.3 of this report

Ivan Kogut (30% of work)

- Running of DDIB on colorized MNIST