Department of Mathematics
香港城市大學
City University of Hong Kong

MA4530 Project 2023/24
Image Denoising via Advanced Diffusion Generative Modeling
基於先進擴散生成模型的圖像去噪

LI Xingyao (56641676) | Supervisor: Prof. ZHOU Xiang

Introduction

Recent advances in deep generative models have significantly improved the generation of high-quality samples. Techniques like GANs, VAEs, and Normalizing Flows model data distributions in unique ways. **Denoising Diffusion Probabilistic Models (DDPMs)** are particularly powerful, using a Markov chain to reverse noise added to data, creating realistic samples.

Score-based models, including Score-Matching Langevin Dynamics (SMLD) and DDPMs, estimate data distribution scores for sample generation. **Stochastic Differential Equations (SDEs)** enhance these methods with new sampling techniques, diffusing data into noise and reversing it to recreate the data.

We explore the principles of DDPMs and SDEs, review recent developments, and present experimental results on datasets like MNIST, Swiss Roll and FFHQ, showcasing the effectiveness and potential of these models.

Mathematical Principles

Diffusion models are latent variable models characterized by:

$$p_\theta(x_0) = \int p_\theta(x_{0:T})dx_{1:T},$$

where $x_1, \ldots, x_T$ are latent variables with the same dimensionality as the data x_0, sampled from $q(x_0)$. The joint distribution $p_\theta(x_{0:T})$, referred to as the reverse process, follows a **Markov chain** with Gaussian transitions:

$$p_\theta(x_{0:T}) = p(x_T)\prod_{t=1}^{T} p_\theta(x_{t-1}|x_t),$$

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)).$$

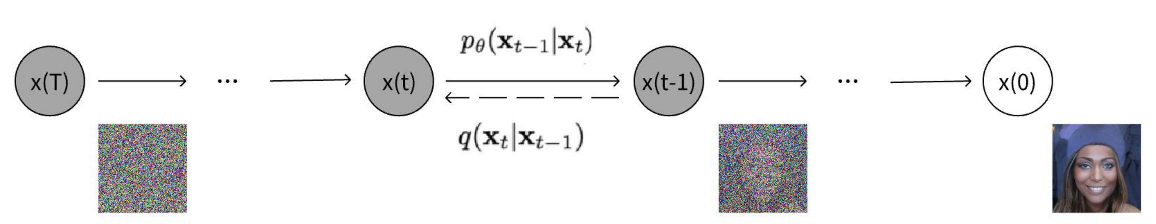

Figure 1. Diagram of the forward and reverse processes of the DDPM model.

Algorithm 1 Training Algorithm

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \mathrm{Uniform}(\{1, \ldots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take gradient descent step on
6: $\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta \left(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t \right) \right\|^2$
7: **until** converged

Algorithm 2 Sampling Algorithm

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T$ to 1 **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

In line with Song, we aim to formulate a **diffusion process** $x(t)_{t=0}^{T}$ that obeys the Itˆo stochastic differential equation:

$$\mathrm{d}x = f(x, t)\mathrm{d}t + g(t)\mathrm{d}w$$

The **reverse of a diffusion process** also constitutes a diffusion process, characterized by:

$$dx = \left[f(x, t) - g^2(t)\nabla_{x_t} \log p(x_t) \right] dt + g(t)d\omega$$

Flexible probability distributions are expressed in the form:

$$p_\theta(x) = \frac{1}{Z_\theta} e^{-f_\theta(x)},$$

where $f_\theta(x)$ is a parameterizable function, often a neural network, and Z_θ is a normalizing constant. Computing Z_θ is challenging in maximum likelihood estimation, so we use a **neural network $s_\theta(x)$** to estimate the **score function $\nabla_x \log p(x)$**.

$$\begin{aligned} \nabla_x \log p_\theta(x) &= \nabla_x \log \left(\frac{1}{Z_\theta} e^{-f_\theta(x)} \right) \\ &= \nabla_x \log \frac{1}{Z_\theta} + \nabla_x \log e^{-f_\theta(x)} \\ &= -\nabla_x f_\theta(x) \\ &\approx s_\theta(x). \end{aligned}$$

Optimize the score model by minimizing the Fisher Divergence with the ground truth score function:

$$\mathbb{E}_{p(x)} \left[\| s_\theta(x) - \nabla \log p(x) \|_2^2 \right].$$

Ultimately, this entails minimizing the following objective function:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathbb{E}_t \Big\{ \lambda(t) \mathbb{E}_{\mathbf{x}(0)} \mathbb{E}_{\mathbf{x}(t)|\mathbf{x}(0)} \Big[\left\| \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}(t), t) - \nabla_{\mathbf{x}(t)} \log p_{0t}(\mathbf{x}(t) \mid \mathbf{x}(0)) \right\|_2^2 \Big] \Big\}.$$

Experimental Methodology

Implementation Details:
We conducted initial experiments on MNIST and Swiss Roll datasets. Using **TensorFlow 2.4.0** and **PyTorch 1.7.0** on Tesla V100 GPUs, we then extensively explored the Flickr-Faces-HQ (FFHQ) dataset, preprocessing the images by **resizing to 256×256 pixels** and applying data augmentations such as random horizontal flips and rotations. We trained our models with a **batch size of 32**, using the **Adam optimizer** and applying a **linear learning rate warm-up** for the first 5,000 iterations.

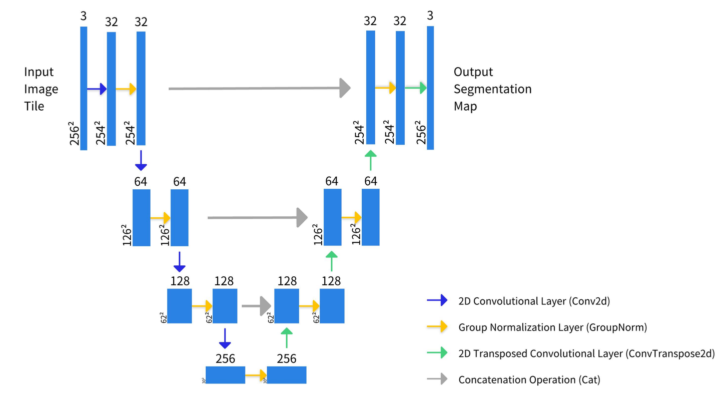

Figure 2. Temporal U-Net: Gaussian Embedding Enhanced Score-based Model.

Evaluation Metrics:

(1) loss function

$$\mathrm{loss}_{\mathrm{SDE}} = \mathrm{mean}\left(\sum_i \left((\mathrm{score}_i \cdot \sigma_i + \mathbf{z}_i)^2 \right) \right)$$

(2) Fréchet Inception Distance (FID)

$$\mathrm{FID} = \|\mu_r - \mu_g\|^2 + \mathrm{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2})$$

Results Analysis

Generation Process:

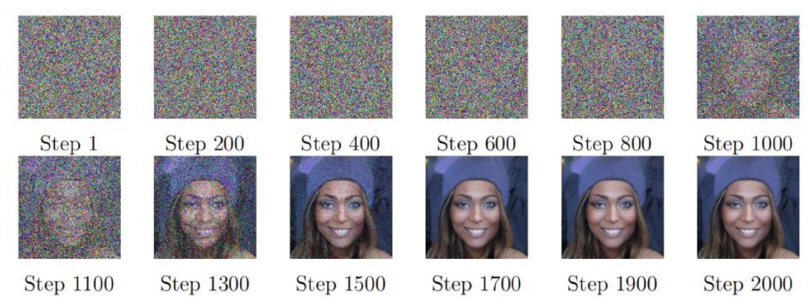

Figure 3. Denoising Process of FFHQ Dataset with NCSNPP (Continuous).

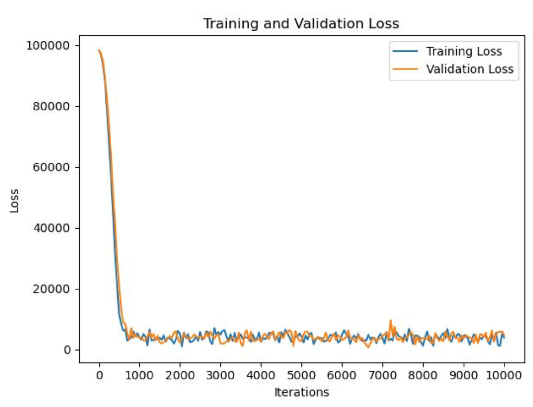

Figure 4. Figure of Training and Validation Loss.

Table 1. FID Values for FFHQ and CIFAR-10 Datasets.

Dataset	Method	FID
FFHQ (Ours)	DDPM	4.96
CIFAR-10 (Song Yang's)	DDPM	3.37

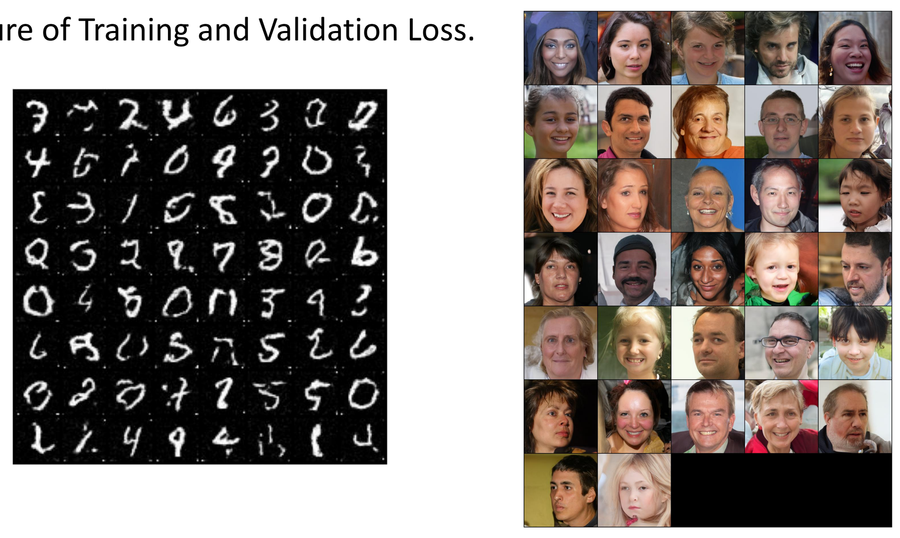

Figure 5. Sample Generation Results.

Conclusions

In conclusion, this study offers a comprehensive analysis of Denoising Diffusion Probabilistic Models (DDPMs) and Score-based models employing Stochastic Differential Equations (SDEs). We have elucidated the fundamental principles underlying DDPMs and SDEs, highlighting their significance in image generation. Our experiments have demonstrated the stability of score-based models in sample generation, indicating their practical feasibility.

Looking forward, while our methodologies have yielded promising results, there are opportunities for further exploration. Specifically, balancing the stability of score-based models with the efficient sampling capabilities of implicit models like GANs remains an area of interest. Additionally, enhancing the speed of sampling processes is crucial and warrants attention in future research endeavors.

References

[1] Yang Song et al. "Score-Based Generative Modeling through Stochastic Differential Equations". In: International Conference on Learning Representations. 2021. URL: https://openreview.net/forum?id=PxTIG12RRHS.
[2] Jonathan Ho, Ajay Jain, and Pieter Abbeel. "Denoising Diffusion Probabilistic Models". In: arXiv preprint arXiv:2006.11239 (2020). [Submitted on 19 Jun 2020 (v1), last revised 16 Dec 2020 (this version, v2)]. DOI: 10.48550/arXiv.2006.11239. arXiv: 2006.11239 [cs.LG]. URL: https://arxiv.org/abs/2006.11239.
[3] Yang Song. "Generative Modeling by Estimating Gradients of the Data Distribution | Yang Song". 2021. URL: https://yang-song.net/blog/2021/score/
[4] Stephen Marsland. "Machine Learning: An Algorithmic Perspective". https://homepages.ecs.vuw.ac.nz/~marslast/Code/Ch6/lle.py. 2014.
[5] Tero Karras and Janne Hellsten. "NVlabs/Ffhq-Dataset". https://github.com/NVlabs/ffhq-dataset. 2021.
[6] Chin-Wei Huang, Jae Hyun Lim, and Aaron C. Courville. "A Variational Perspective on Diffusion-Based Generative Models and Score Matching". In: CoRR abs/2106.02808 (2021). arXiv: 2106.02808. URL: https://arxiv.org/abs/2106.02808.
[7] Tero Karras et al. "Elucidating the Design Space of Diffusion-Based Generative Models". 2022. arXiv: 2206.00364 [cs.CV].
[8] Brian D.O. Anderson. "Reverse-time diffusion equation models". In: Journal of Mathematical Analysis and Applications 87.2 (1982). Received November 13, 1979; Revised June 12, 1980; Available online April 2, 2002., pp. 592–602. DOI: 10.1016/0304-4149(82)90051-5. URL: https://doi.org/10.1016/0304-4149(82)90051-5.