Denoising Diffusion Probabilistic Model and Image Inpainting

In this project, We worked on Denoising Diffusion Probabilistic Models (DDPMs) implementation, as well as performing several application such as sample generation and image inpainting. The model proposed extends score-based modeling by incorporating a time-dependent noise schedule and a reverse denoising process to generate high-quality samples.

Modeling

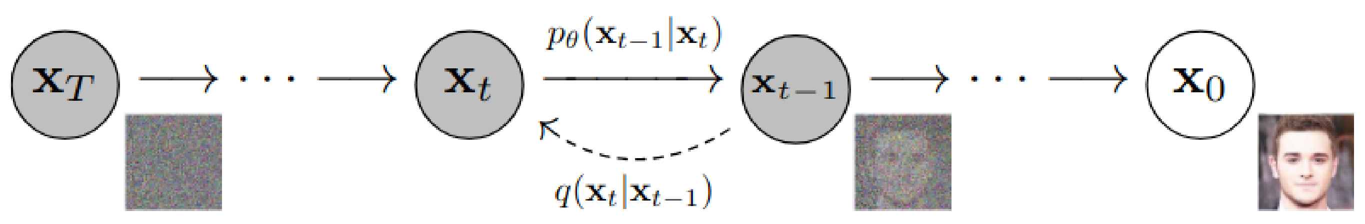

Figure : Directed graphical diffusion model introduced in [1]

- **Forward process :** Add Gaussian noise to the input data step-by-step until it becomes pure noise

$$q(x_t \mid x_{t-1}) := \mathcal{N}(x_t; (1-\beta_t)x_{t-1}, \beta_t I)$$

- **Reverse process :** Reverse the noising process to recover the original data

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

where β_t represents the noise scheduler

- **Variational Objective / ELBO :**

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{t\sim[1,T]}\mathbb{E}_{x_0\sim q(x_0)}\mathbb{E}_{\epsilon_t\sim\mathcal{N}(0,I)}\left[\|\epsilon_t - \epsilon_\theta(x_t, t)\|^2\right]$$

where ϵ_t represents the true noise

Experiments

- **Selecting schedulers**

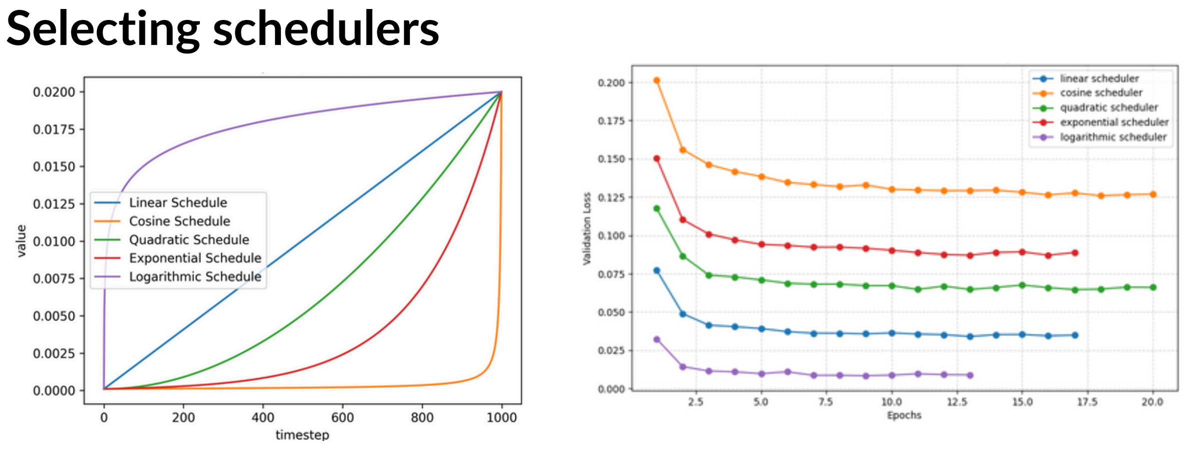

Figure : Shape of the five test schedulers on left, MSE validation on the right.

- **Hyperparameter Tuning using Bayesian Optimization**

Dataset	T	beta_min	beta_max	Best Loss
Fashion	1757	0.0007645481	0.0328679383	0.01454419
CIFAR	1132	0.0004998979	0.0387503176	0.02055065
MNIST	1722	0.0000252098	0.0479183811	0.01064030

Table : Results of hyperparameter tuning (diffusion using Linear Scheduler).

Dataset	T	beta_min	beta_max	Best Loss
Fashion	1927	1.1146056 − 05	0.041009289	0.00310464
CIFAR	1931	6.4127417 − 05	0.042472445	0.00290811
MNIST	1931	0.0003399997	0.039108512	0.00201147

Table : Results of hyperparameter tuning (diffusion using Log Scheduler).

- **Neural Network Architecture:**
  - UNET architecture containing Convolutional layers, residual blocks, DownSampling and UpSampling blocks.
  - Same UNET with additional attention blocks after the DownSampling and UpSampling blocks

Sampling

The reverse process is used :

- sample random noise $x_T \sim \mathcal{N}(0, I)$
- remove the noise using the reverse process

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{1-\beta_t}}\left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z}$$

- After T iterations, the final output is the denoised data sample

Contributors

Ibrahim RIDENE
Malek BOUHADIDA

Master MVA,
ENS Paris-Saclay

RePainting (Linear vs Log scheduler)

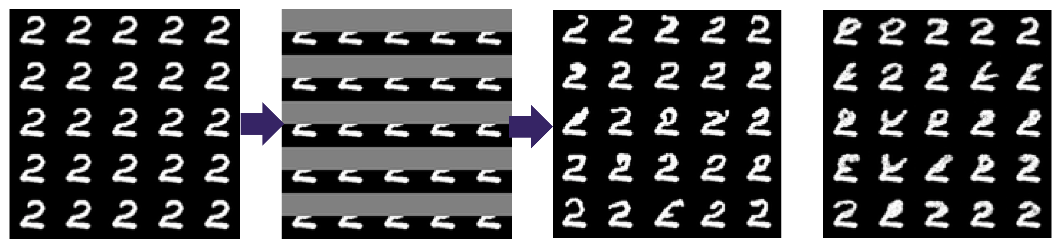

Figure : MNIST samples: original, masked, repainted (Linear scheduler), repainted (Log scheduler) in order.

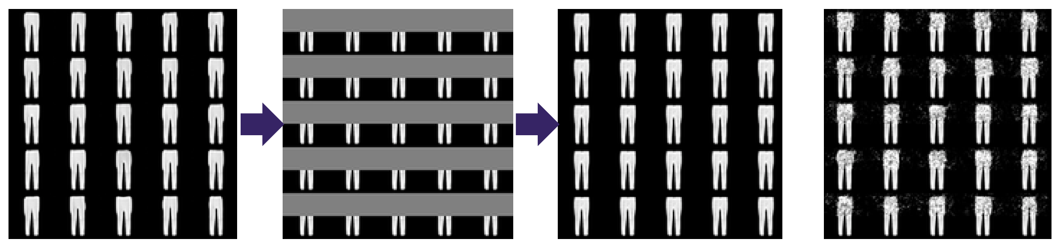

Figure : Fashion-MNIST samples: original, masked, repainted (Linear scheduler), repainted (Log scheduler) in order.

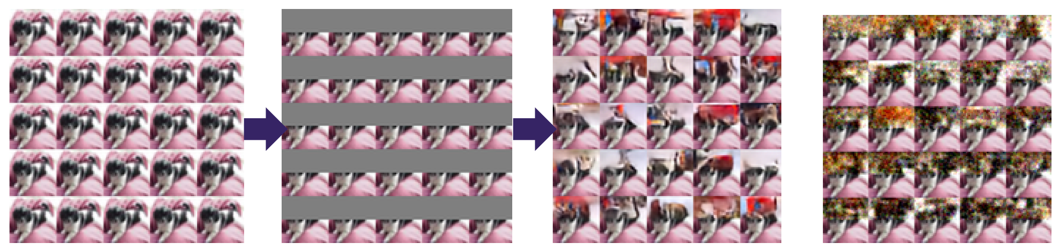

Figure : CIFAR samples: original, masked, repainted (Linear scheduler), repainted (Log scheduler) in order.

RePainting (U-Net w/wo Attention)

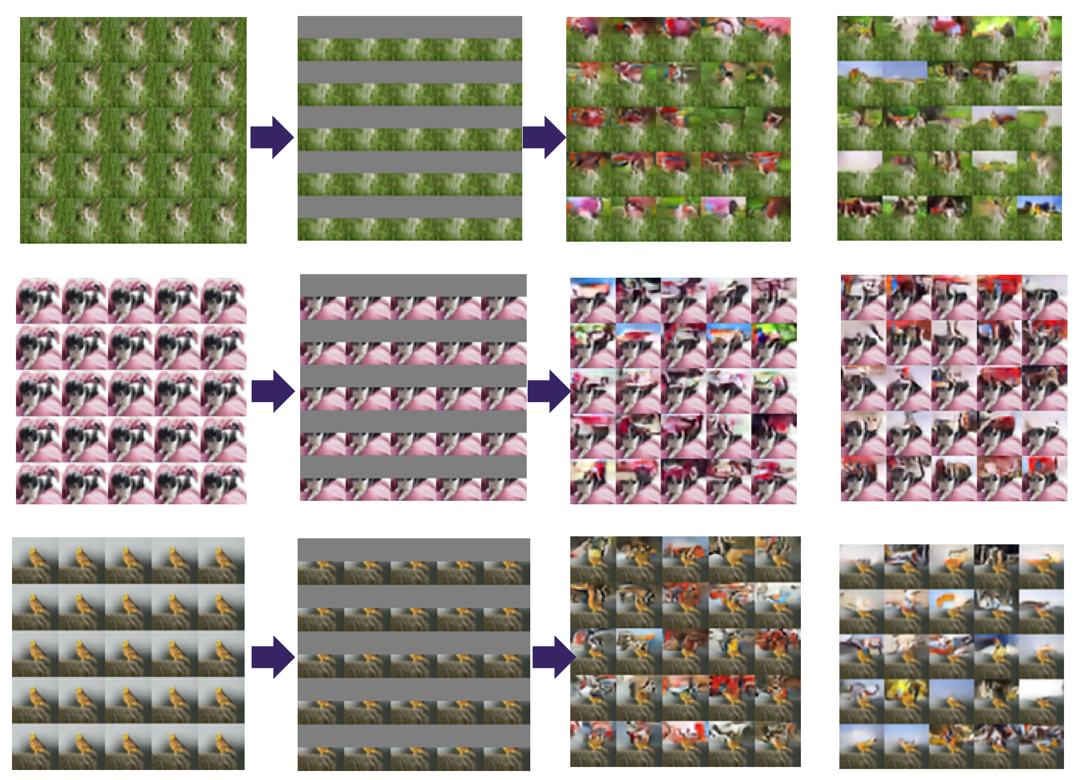

Figure : CIFAR samples: original, masked, simple UNET, UNET with Attention

Discussion

- The linear and Logarithmic schedulers perform better than the other schedulers based on validation data.
- The logarithmic scheduler introduces more noise in the generated samples as well as the repainted samples, compared to the linear scheduler.
- The attention-based UNet generates inpainted regions that are more consistent with the global distribution of the image, compared to the basic UNet model.
- Including the FID score during training will provide a better evaluation of the generated images' quality with respect to the global distribution of the data.
- Including class labels will improve model performance.

References

[1] **Ho Jonathan, Ajay Jain, and Pieter Abbeel.** "Denoising diffusion probabilistic models.", NeurIPS 2020

[2] **A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool.** "RePaint: Inpainting using Denoising Diffusion Probabilistic Models.", IEEE/CVF 2022