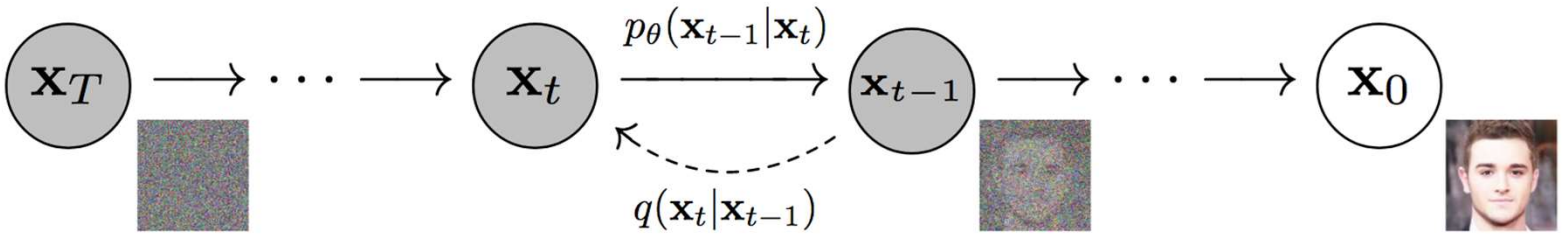
Project 3: Diffusion models

TA: Qian Wang
Sunday, October 8th, 2023

Diffusion models

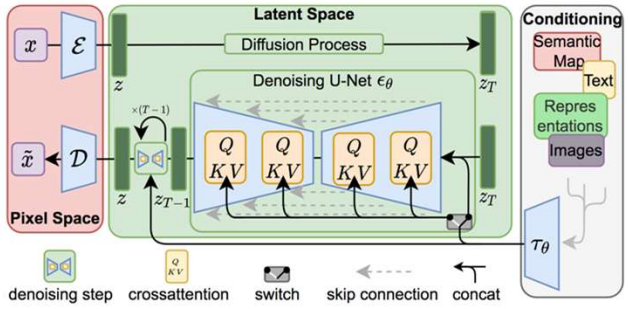
- One of the most popular generative models nowadays
- Can be used for generating image, video, audio, 3D ...
- What you need to understand:



Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in neural information processing systems* 33 (2020): 6840-6851.

Stable Diffusion

- A large text-to-image diffusion models.
- Quickly try it out by yourself: https://huggingface.co/spaces/stabilityai/stable-diffusion
- Build on top on the Latent Diffusion Models:



Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." *CVPR* 2022.

Diffusers library

- Why using diffusers library?
  - High-level implementation of diffusion pipelines
  - Easy access of pre-trained diffusion models
- Only basic usage in this project.
- More interesting pipelines? Training speedup? Less memory consumption? Go to the official documentation: https://huggingface.co/docs/diffusers/index

Project 3-1: Inversion and sampling

- Method: DDIM (Denoising Diffusion Implicit Models)
- The same formula for inversion and sampling, but use it in different orders.
- Inversion: sample x_{t+1} from x_t;
- Sampling: sample x_{t-1} from x_t.
- We will implement the method based on the formula (12) in the DDIM paper.

Song, Jiaming, Chenlin Meng, and Stefano Ermon. "Denoising diffusion implicit models." *arXiv preprint arXiv:2010.02502* (2020).

Project 3-2: Dreambooth

- Personalization of a new concept given a few reference images as input.
- We will build it on top of the Stable Diffusion models.
- It includes learning an embedding for the newly added concept and finetuning the unet model.
- You need to find a good setting of hyperparameters: not underfitting or overfitting.

Ruiz, Nataniel, et al. "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.

Project goals

- Main goal: have a high-level understanding of how diffusion works and how to implement it, but not how to derive formulas.
- However, you should still know the basic notations, such as x_0, x_T and ε.
- You should also know the components in Stable Diffusion: UNet, VAE, Text Encoder, Tokenizer, Noise scheduler ...
- GPU is needed, preferably a single A100 or V100. Multi-GPU is not supported in this project.

Others

- Due date: Wednesday, October 25th
- Please contact TA (Qian Wang) if you have any questions about the project, or you find any mistakes in the notebook ☺
- Training time wouldn't be long, but good to start earlier. Good luck!