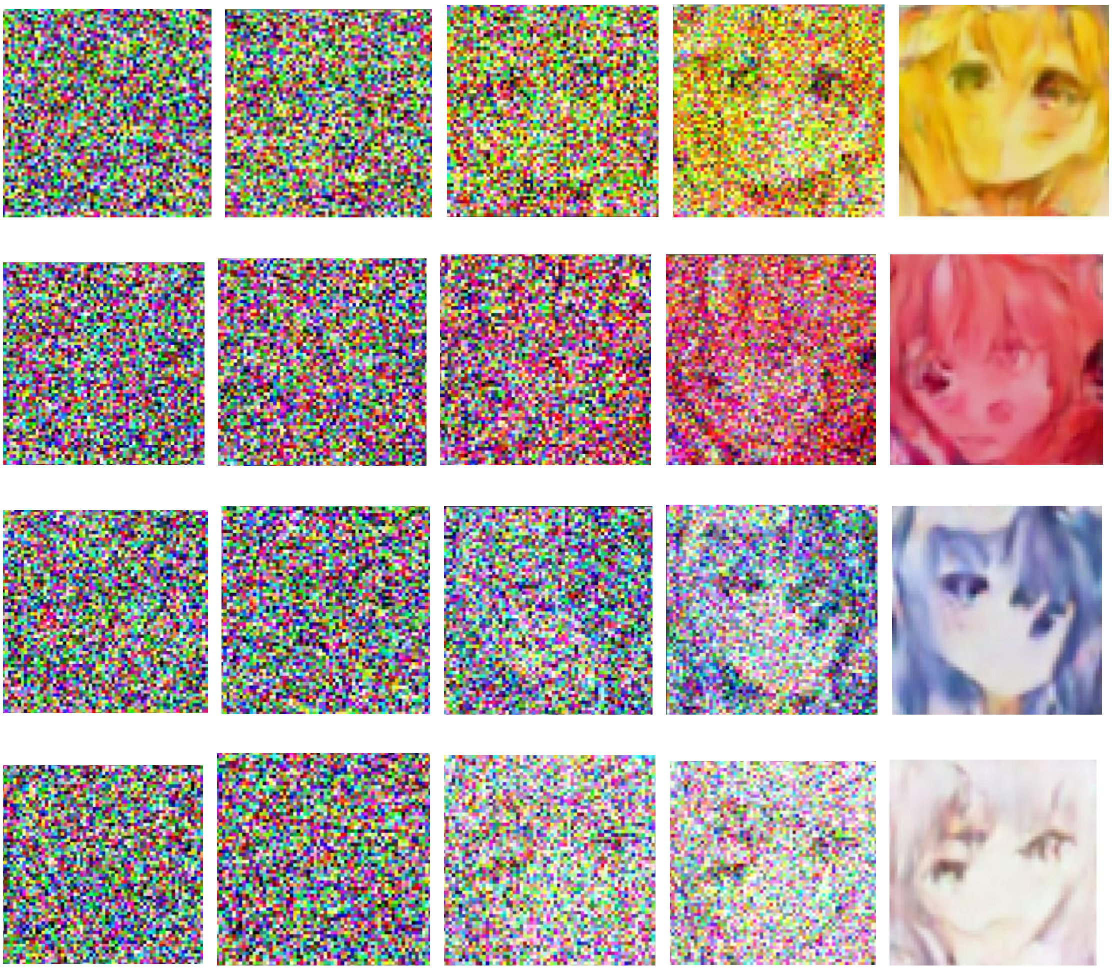
Gradescope

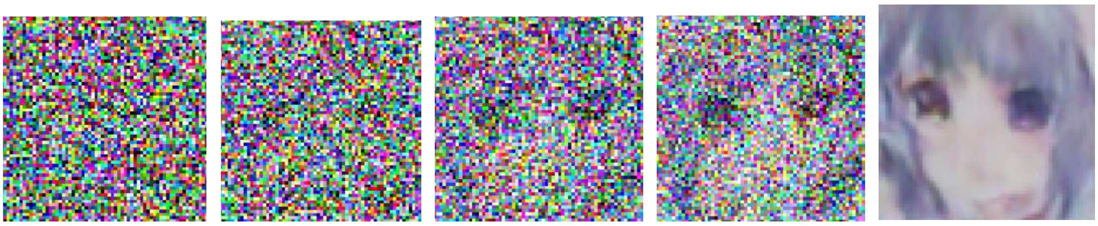
t11902210 張一凡

1. Sample 5 images and show the progressive generation. Then, briefly describe their differences in different time steps.

Answer:

Note in advance: The model I am using is the secret weapon stylegan2 provided by HW6, but the internal packaging of this stylegan2 is too good to generate images during the process. Therefore, I followed the tutorials provided by the teaching assistant and teacher to make modifications to the original code, that is, output the image part. However, I did not spend too much experience improving the original model, so the accuracy of the final image is not high, But basically, you can see the appearance of the portrait, which meets the requirements of the teaching assistant in the discussion area. Here, I use these pictures to answer question one, hoping to gain the understanding of the teachers. Of course, the code I submitted on Judge and ntu cool is related to the code and grades of stylegan2.





In the diffusion model, the image generation process is carried out through a series of diffusion steps, each step making the image clearer and less noisy. The following is a brief description of the differences in images generated at different time steps:

In the early stages of training, the images are very blurry and there is a lot of noise. For example, the first photo of each group I showed is completely indistinguishable from the cartoon face, only showing many mosaic-like things. This is because, at this stage, noise is distributed throughout the entire image and has not yet been concentrated in any specific area.

In the mid-term process, as the diffusion step progresses, noise begins to become more concentrated in specific areas of the image, leading to the appearance of patterns and structures. For example, as shown in Figures 2, 3, and 4 of each group, although there are obvious defects in the eyes, nose, and mouth, as the training progresses, it can be seen that this is a face. The image generated at this stage is still very blurry, but some basic shapes and features begin to become visible.

In the later stages of the diffusion process, noise becomes more concentrated and begins to display finer details and textures in the image. The images generated in this stage are clearer and easier to recognize than those generated earlier, such as the last image in each group I presented. Although the final effect of using the model answering Q1 is not as good as the model I obtained using stylegan2, it can also be seen that this is a cartoon portrait with major defects and noise that is invisible. If the training steps of the model can be improved or the model can be replaced, better results can be obtained.

Overall, the gradual generation of images in diffusion models demonstrates the power of iterative processes in generating high-quality images from noisy and incomplete data.

2. Canonical diffusion model (DDPM) is slow during inference, Denoising The diffusion Implicit Model (DDIM) is at least 10 times faster than DDPM during inference and preserves the qualities. Please describe the differences in training, inference process, and the generated images of the two models respectively. Briefly explain why DDIM is faster.

Answer:

Both DDPM and DDIM are generative models based on diffusion processes, and they show different behaviors based on similar diffusion processes, but there are some differences in definition and model application. The specific comparison is shown below.

For the complete training process, the process includes sampling from the diffusion process and minimizing the possibility of estimation and the reverse KL deviation between the real data distribution. The two models use the same loss function and optimization algorithm. However, DDIM uses a different architecture from DDPM, which has residual blocks and modified attention mechanisms, resulting in faster convergence during the training process.

For reference processes, DDPM requires MCMC sampling to generate images, while DDIM can directly generate samples by solving differential equations using ODE solvers. This makes DDIM much faster in the inference process than DDPM. On the other hand, DDPM requires multiple forward and reverse diffusion processes to generate samples, thus requiring high computational costs. DDIM only needs one forward transmission through the model to generate samples, which makes it much faster than DDPM.

For the image generation process, both models can generate high-quality samples with clear details and diverse structures, achieving the desired results. However, DDIM uses a denoising fractional matching method, where models are trained to denoise noisy images rather than directly estimating their likelihood, so it can generate samples with higher fidelity and better visual quality.

Finally, in terms of specific application methods, the reason why DDIM is faster is that it avoids the need for MCMC sampling in the inference process by using a differential equation solver, which is extremely time-consuming, resulting in significant acceleration without sacrificing the quality of generated samples.

Reference Paper:
Denoising Diffusion Probabilistic Models
Denoising Diffusion Implicit Models

Reference link:
https://xduwq.blog.csdn.net/article/details/124839018?spm=1001.2101.3001.6661.1&utm_medium=distribute.pc_relevant_t0.none-task-blog-2%7Edefault%7ECTRLIST%7EPayColumn-1-124839018-blog-128265411.235%5Ev29%5Epc_relevant_default_base3&depth_1-utm_source=distribute.pc_relevant_t0.none-task-blog-2%7Edefault%7ECTRLIST%7EPayColumn-1-124839018-blog-128265411.235%5Ev29%5Epc_relevant_default_base3&utm_relevant_index=1&ydreferer=aHR0cHM6Ly9ibG9nLmNzZG4ubmV0L3dhbmd5dW5zZW5nMzMvYXJ0aWNsZS9kZXRhaWxzLzEyODI2NTQxMQ%3D%3D