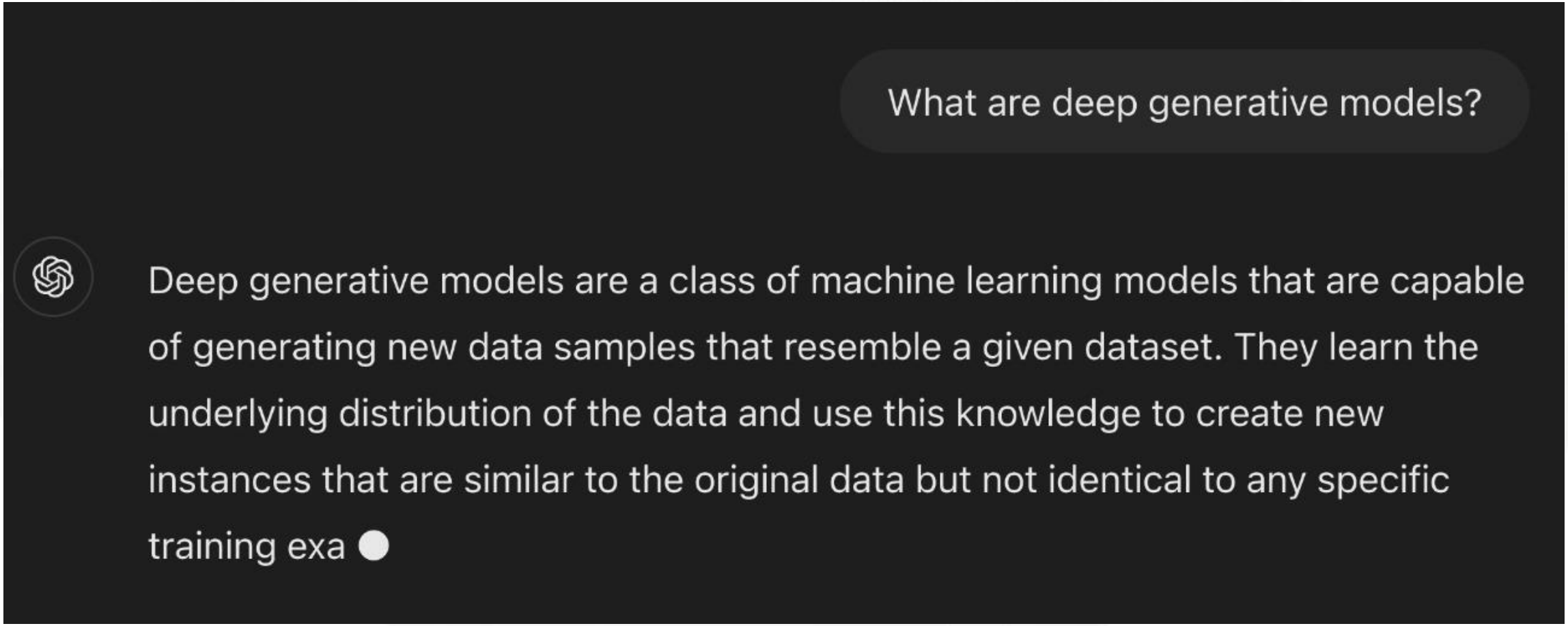
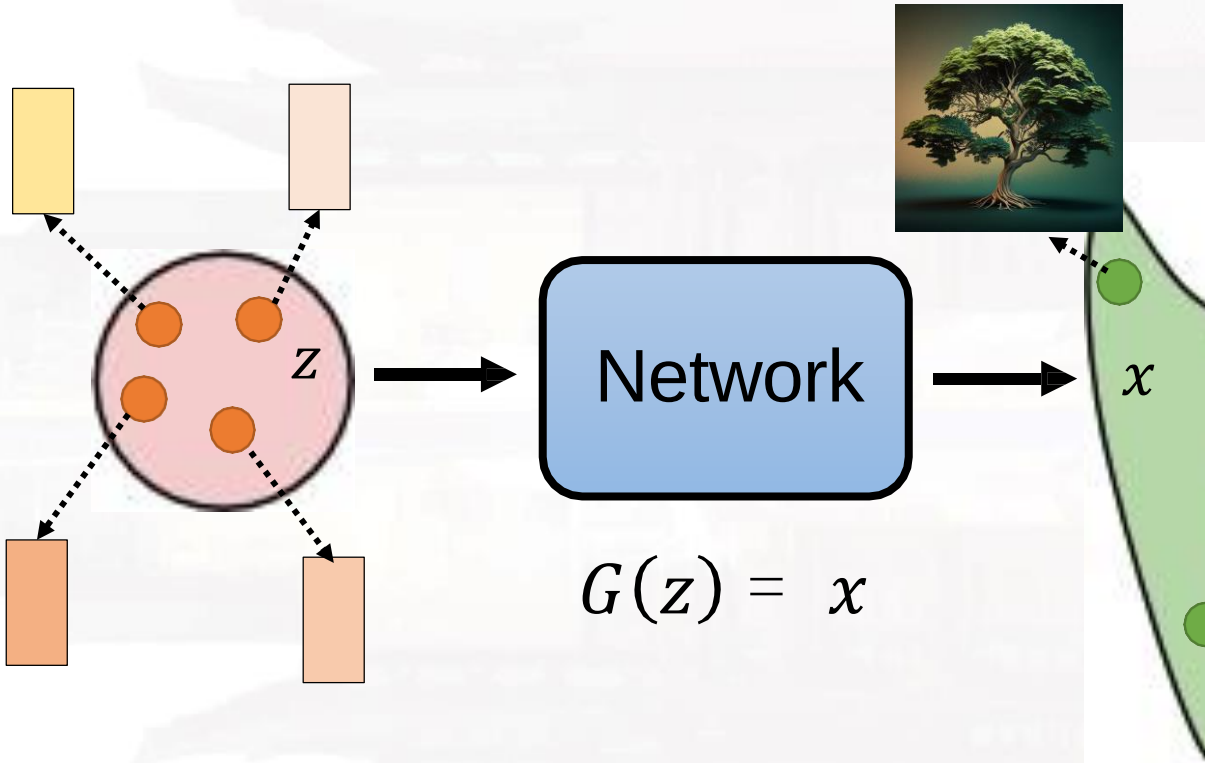
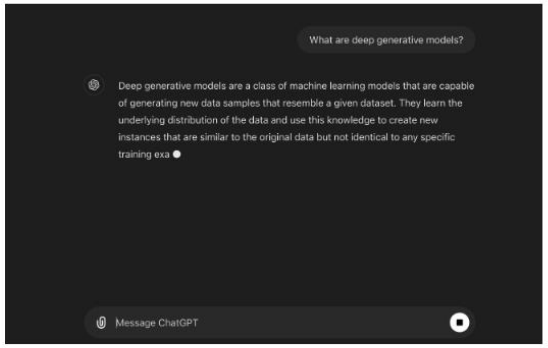
16. Generative Models

Generative Model



Generative Models

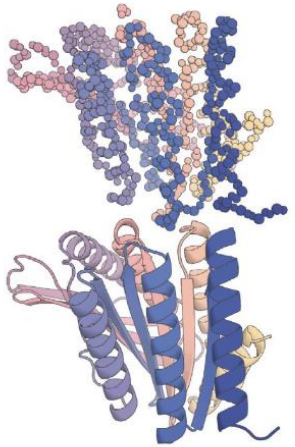




Chatbot

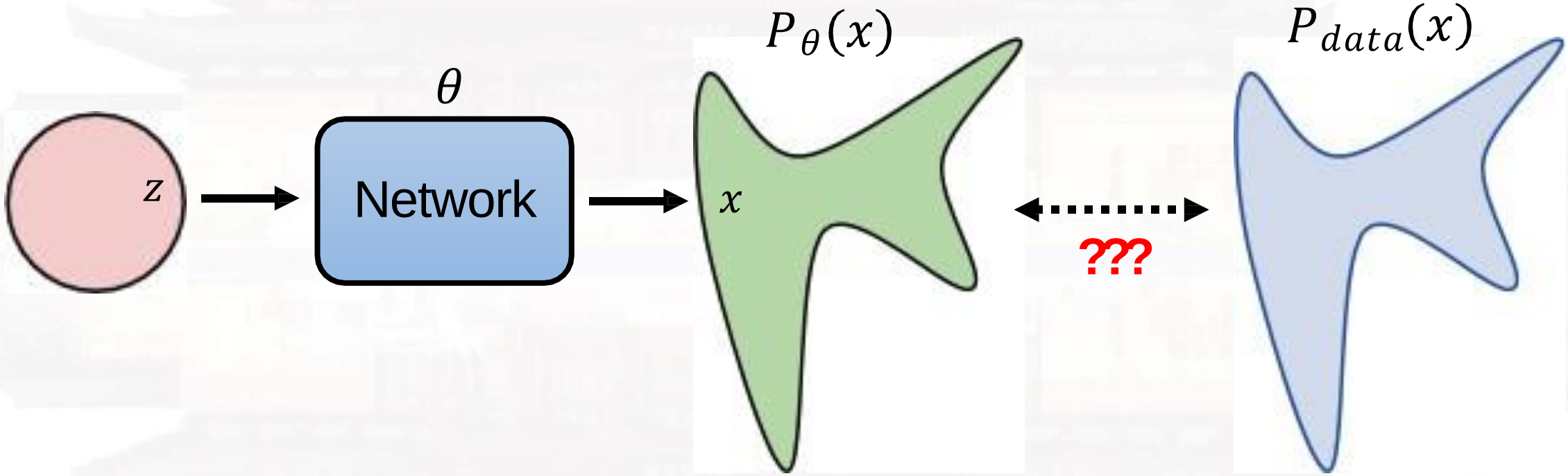
Image generation

Video generation

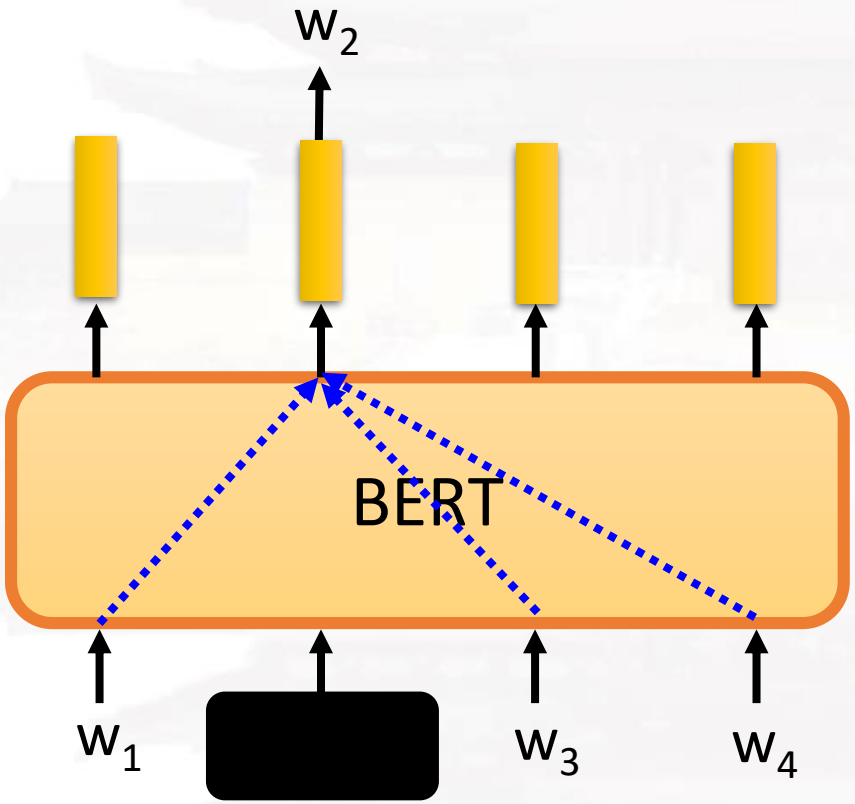


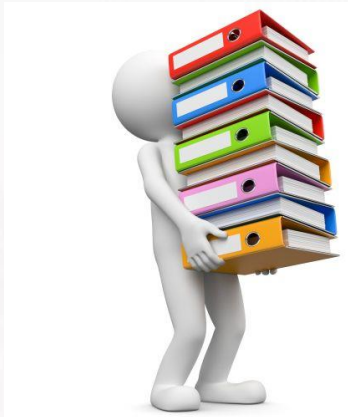
Protein generation

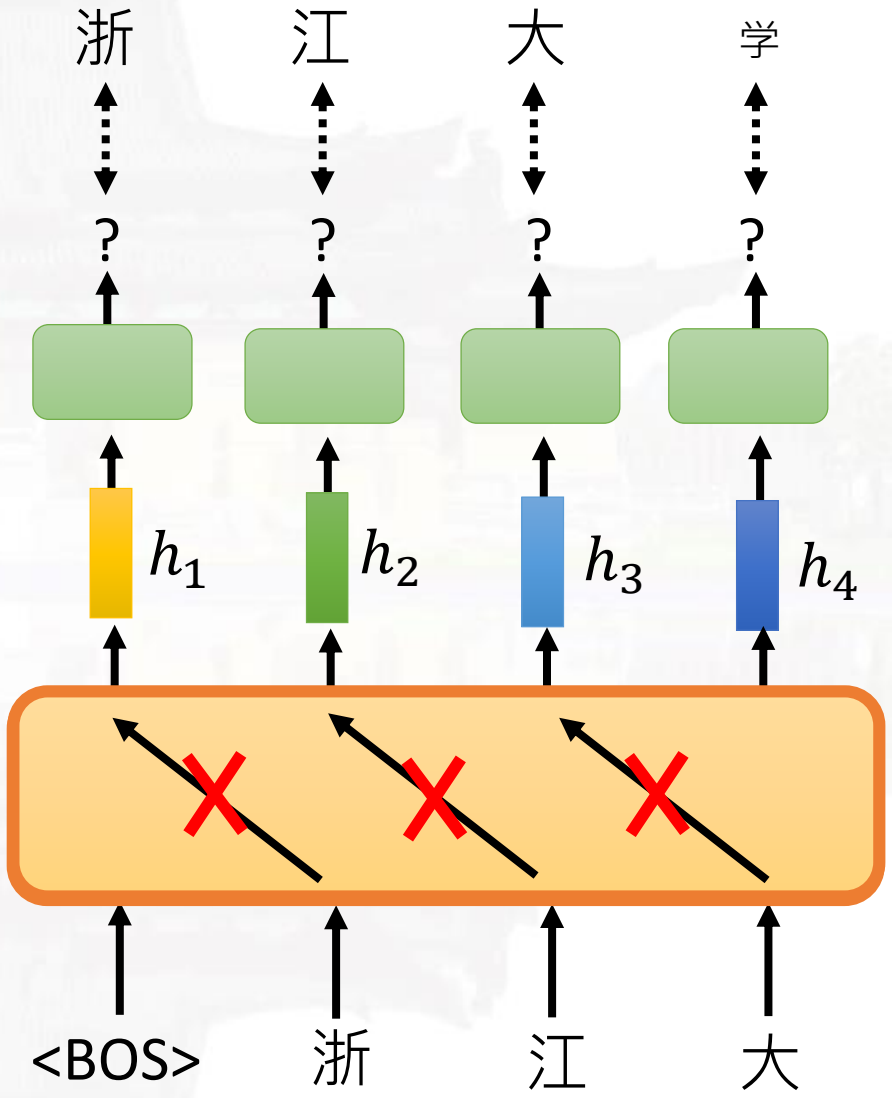
Generative Models



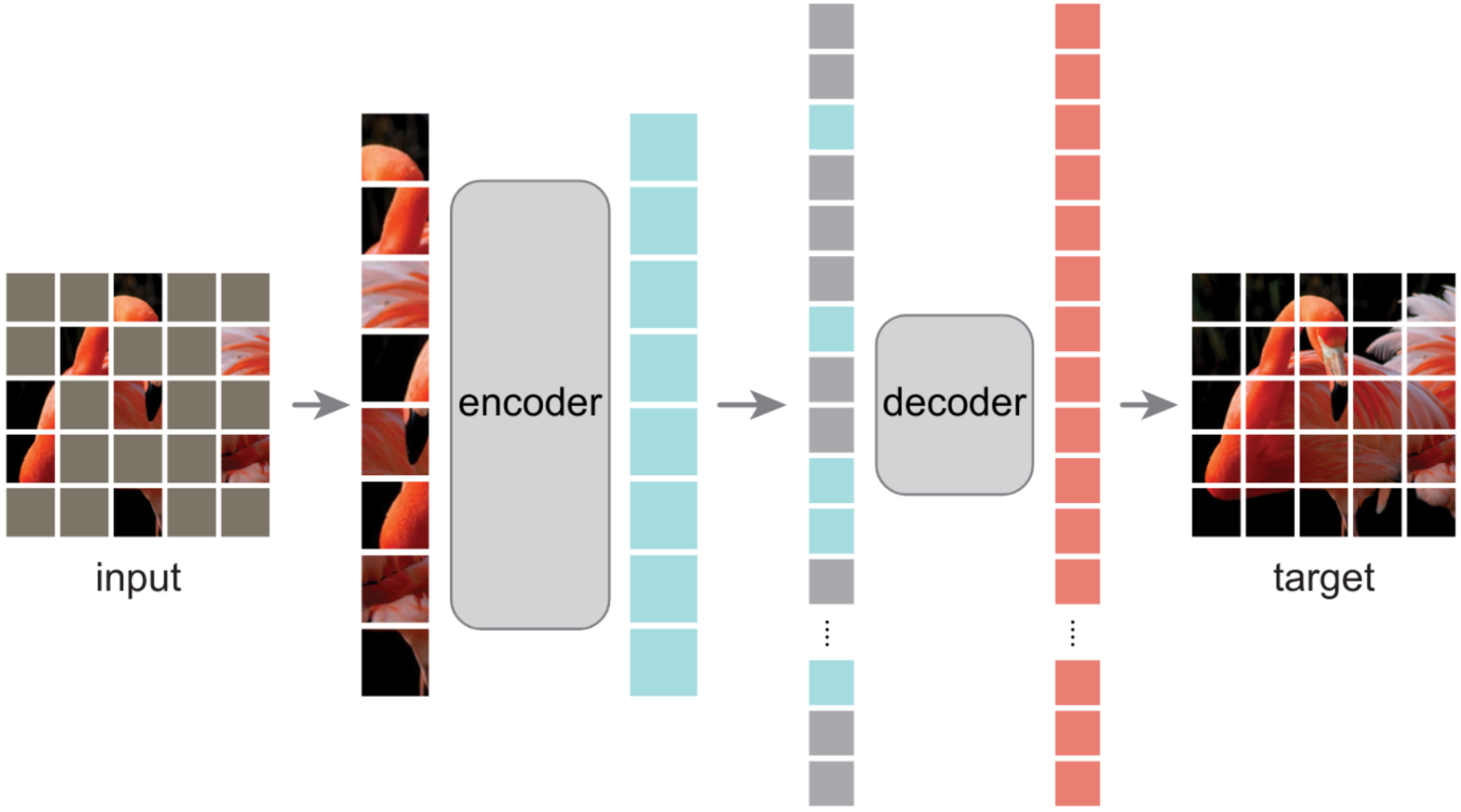
How: Self-supervised learning



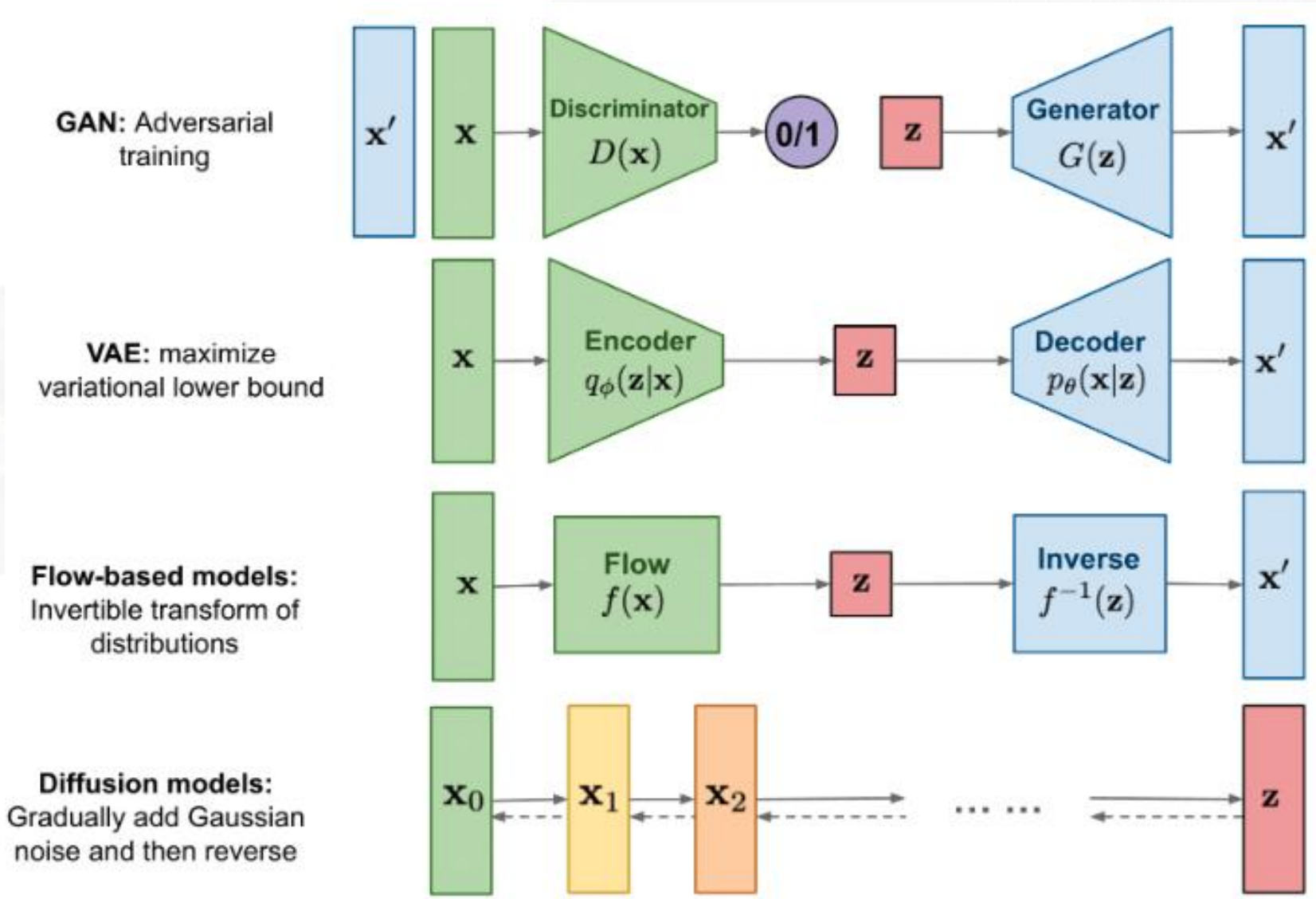


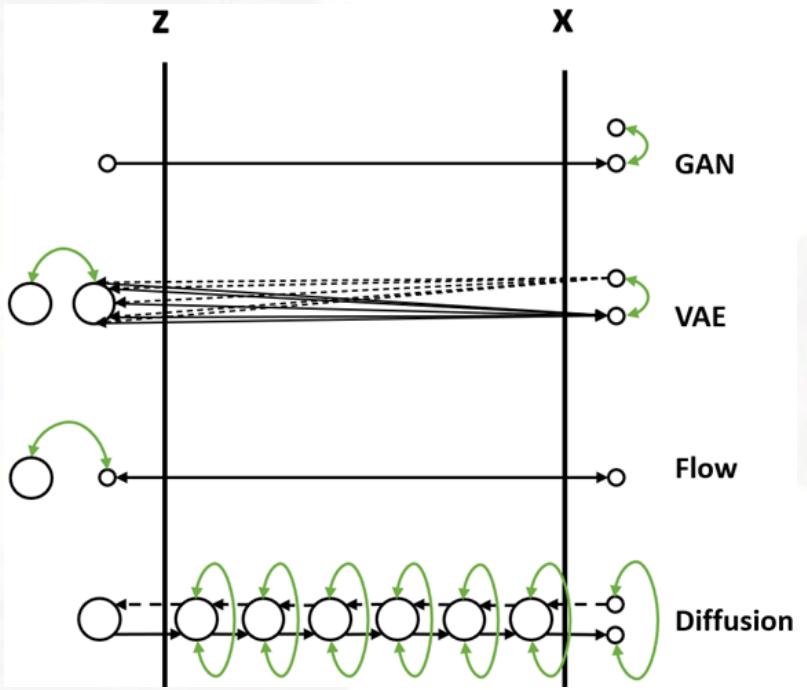


How: Self-supervised learning

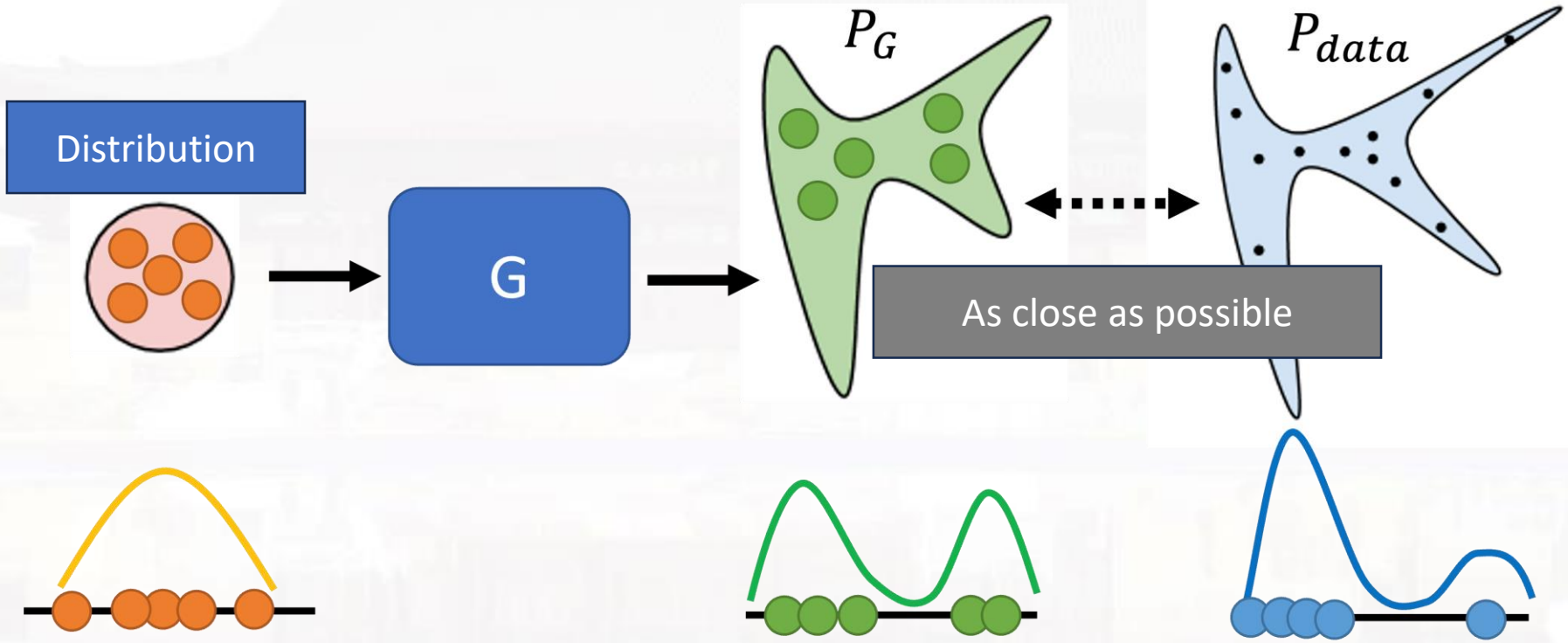


Generative Models





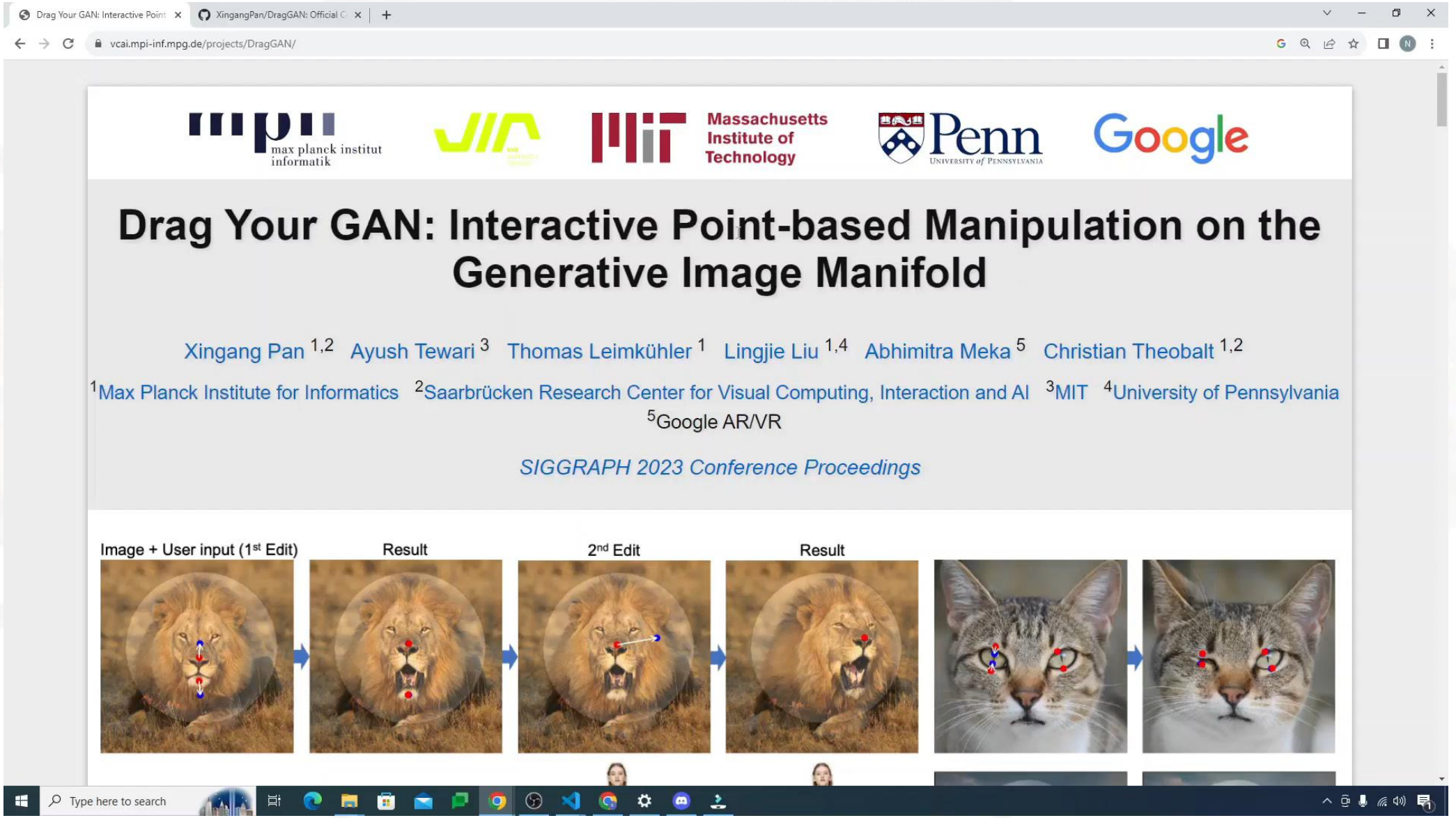
GAN



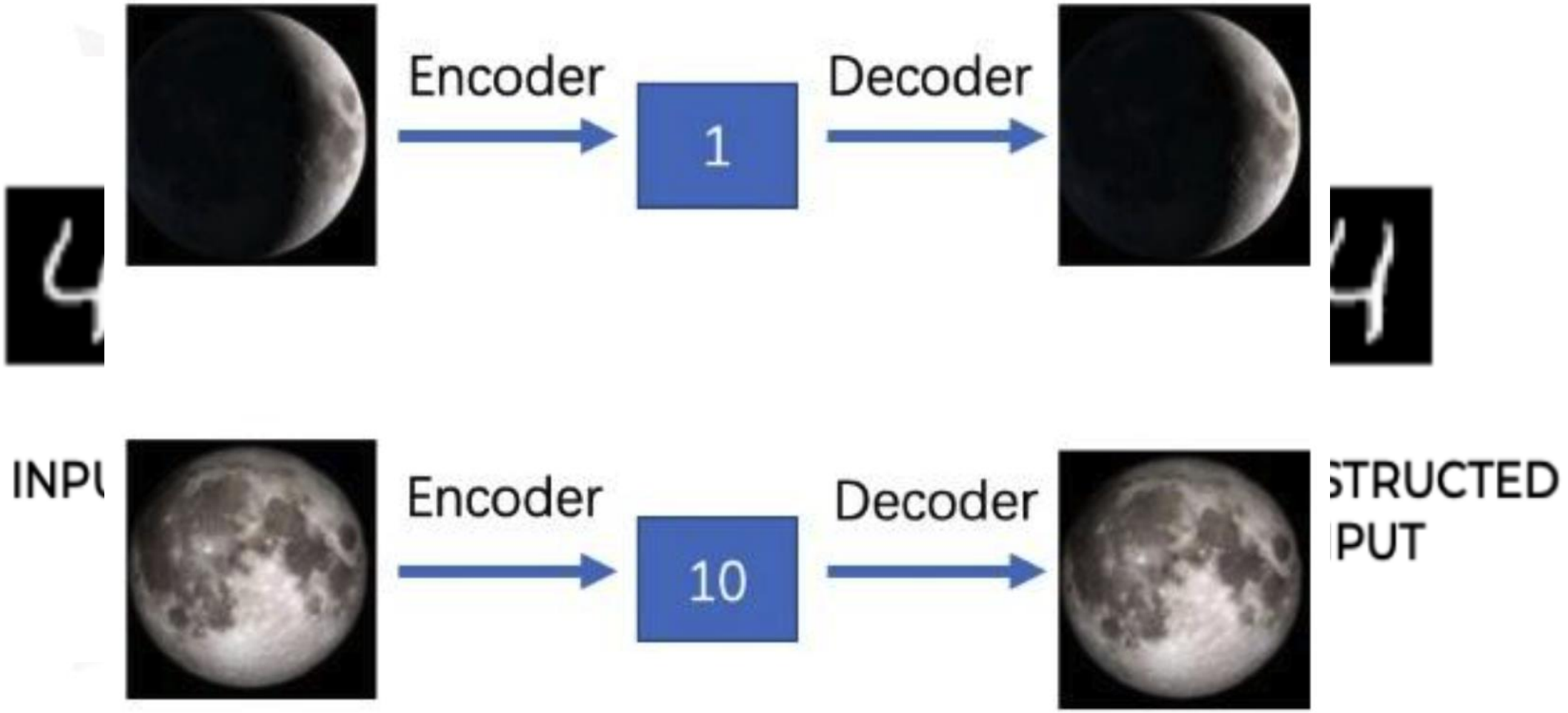
$$G^* = arg \min_G Div(P_G, P_{data})$$

Difference between P_G and P_{data}

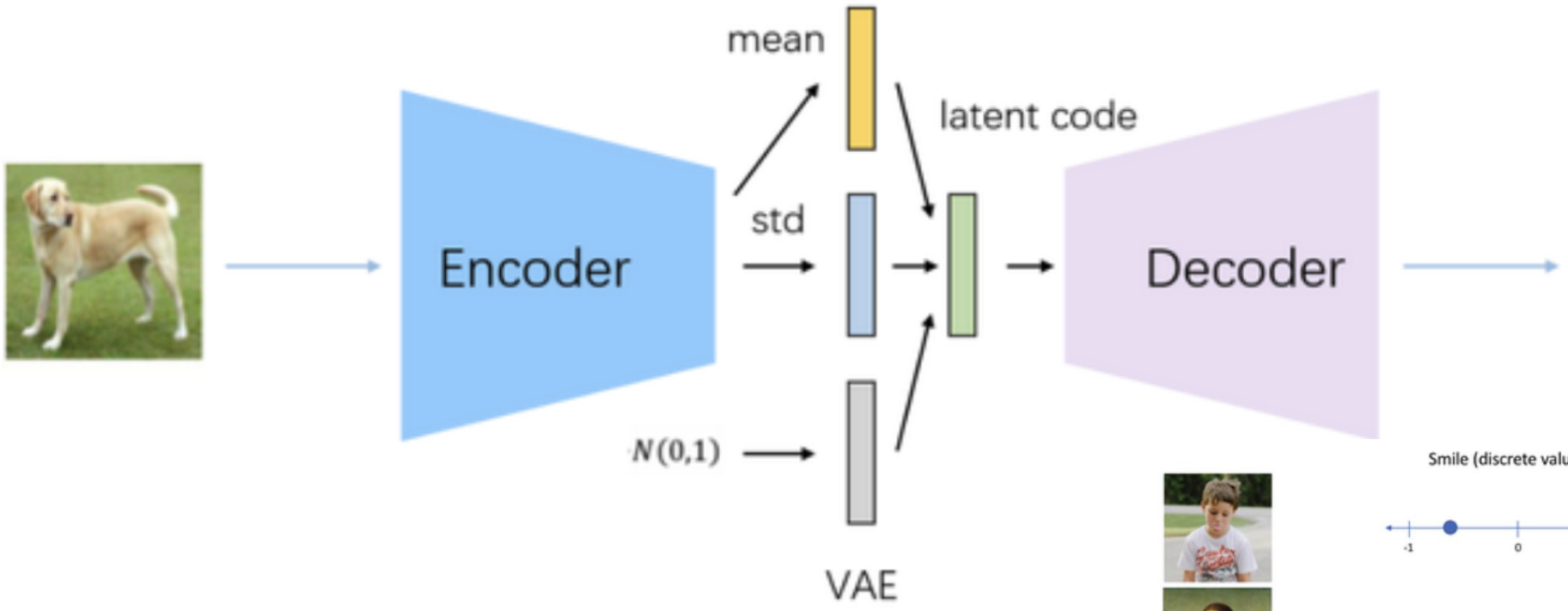
GAN

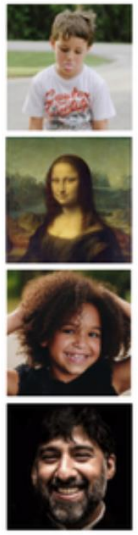

Drag Your GAN: Interactive Point
XingangPan/DragGAN: Official
vcai.mpi-inf.mpg.de/projects/DragGAN/
max planck institut informatik
Massachusetts Institute of Technology
Penn
University of Pennsylvania
Google
Drag Your GAN: Interactive Point-based Manipulation on the Generative Image Manifold
Xingang Pan [1,2] Ayush Tewari [3] Thomas Leimkühler [1] Lingjie Liu [1,4] Abhimitra Meka [5] Christian Theobalt [1,2]
[1]Max Planck Institute for Informatics [2]Saarbrücken Research Center for Visual Computing, Interaction and AI [3]MIT [4]University of Pennsylvania
[5]Google AR/VR
SIGGRAPH 2023 Conference Proceedings
Image + User input (1st Edit)
Result
2nd Edit
Result
Type here to search

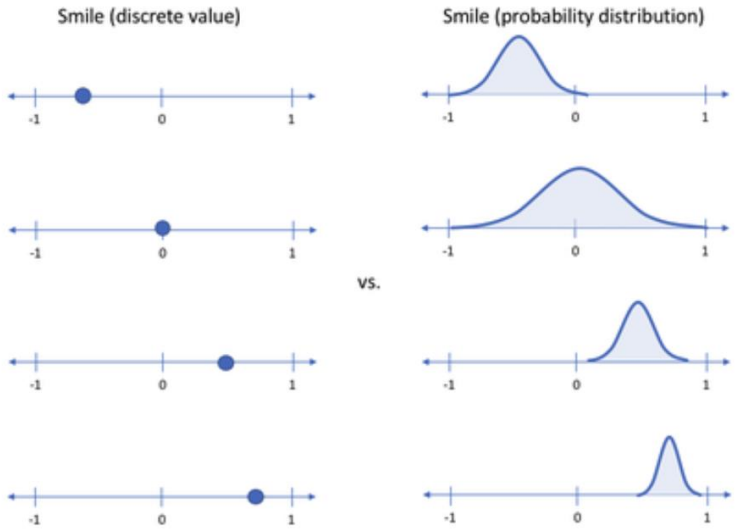
VAE



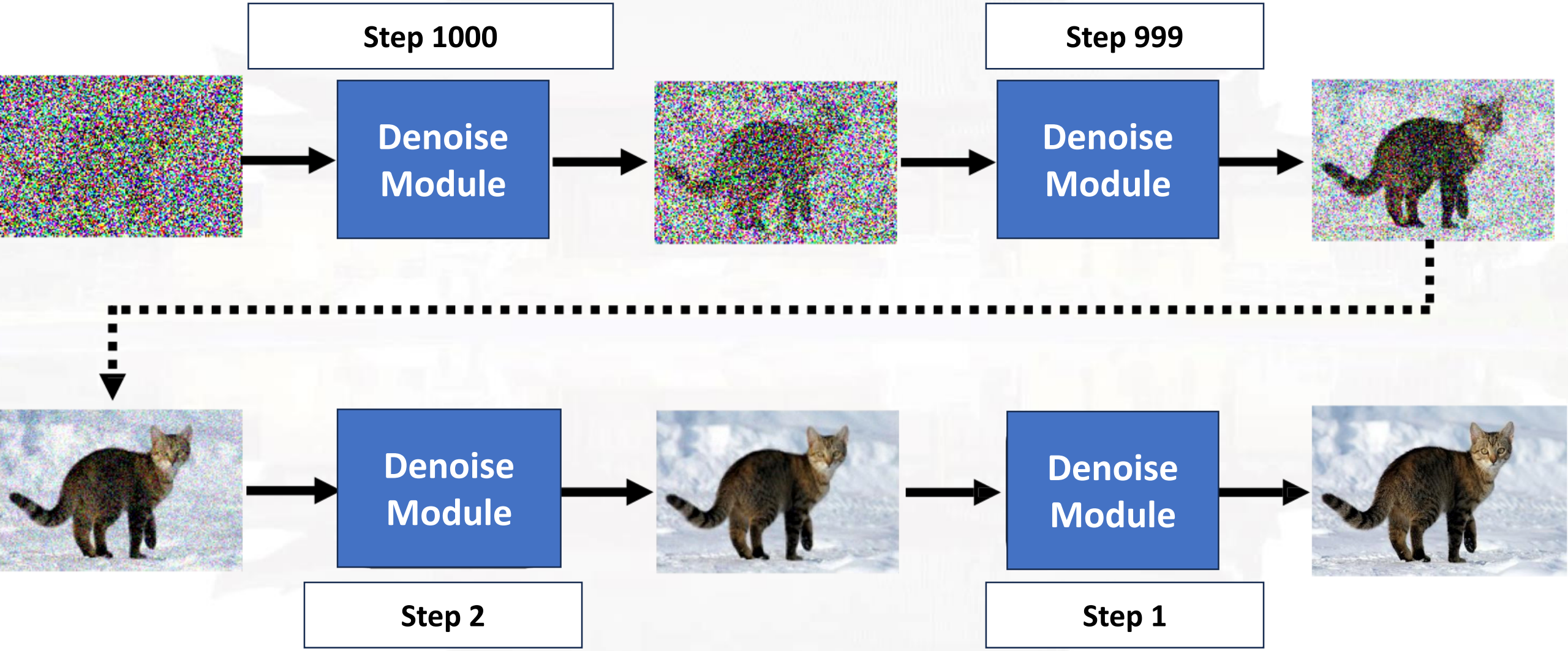
VAE


mean
latent code
std
Encoder
Decoder
$N(0,1)$
VAE

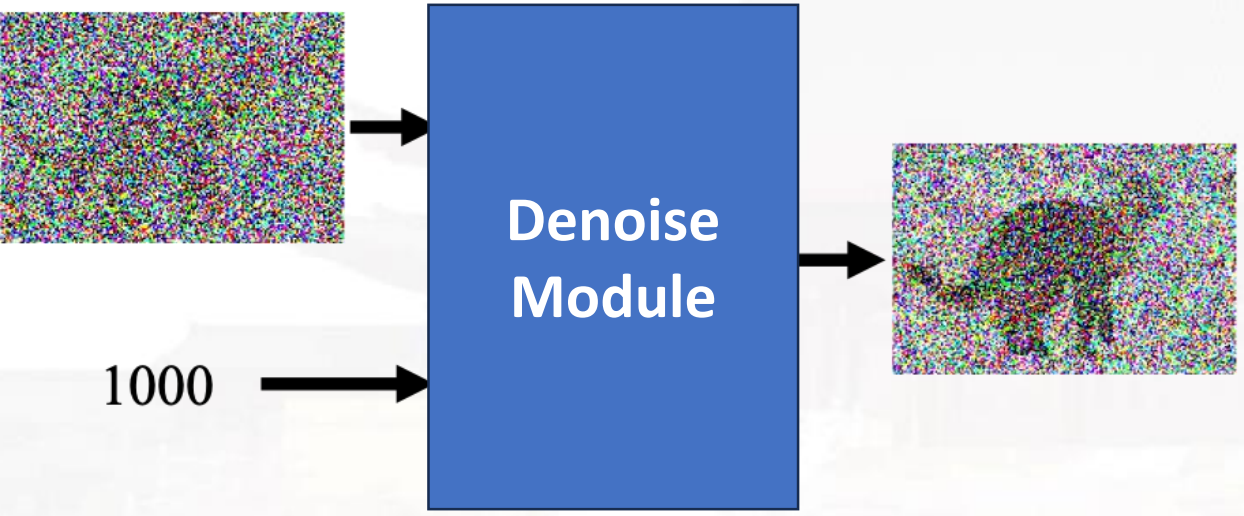


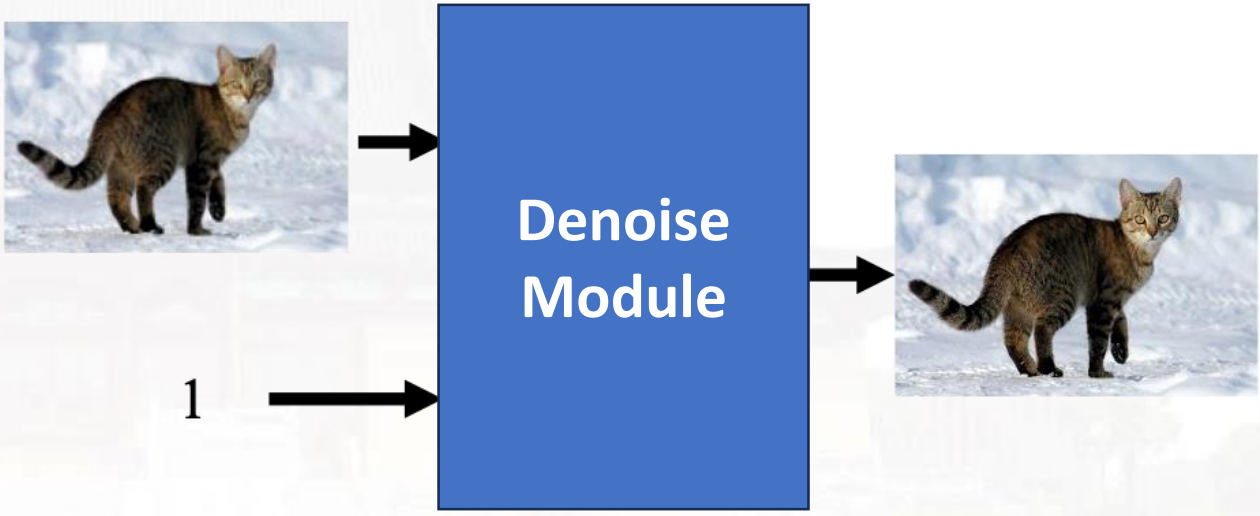

Smile (discrete value)
Smile (probability distribution)
vs.
-1
0
1

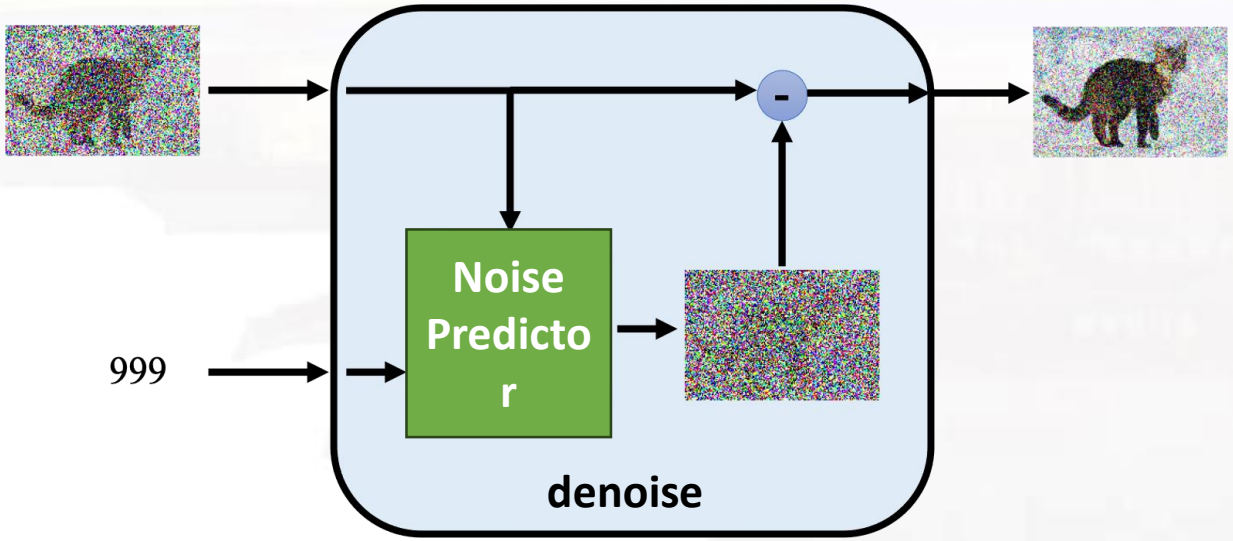
Diffusion Model



Diffusion Model



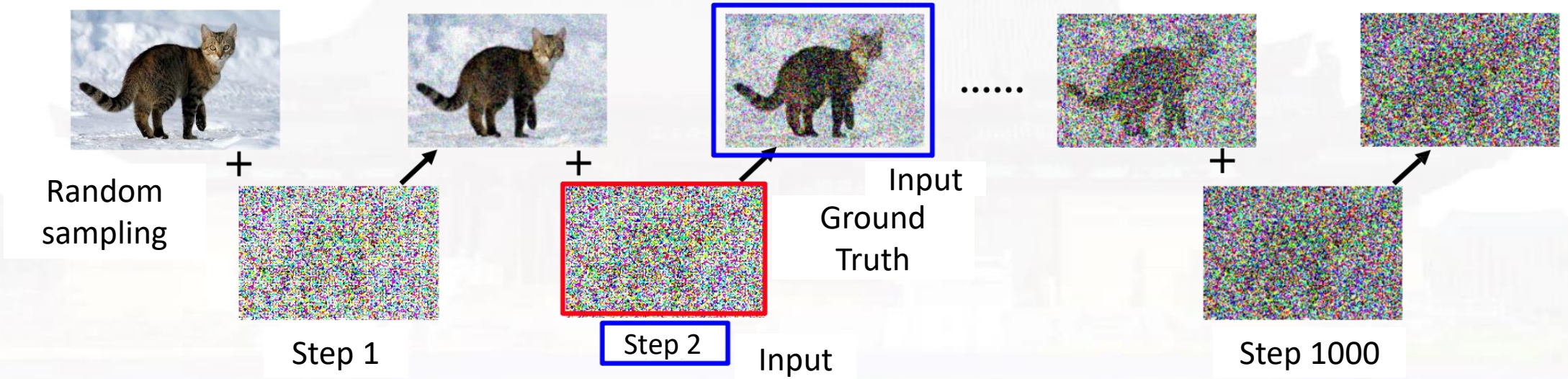


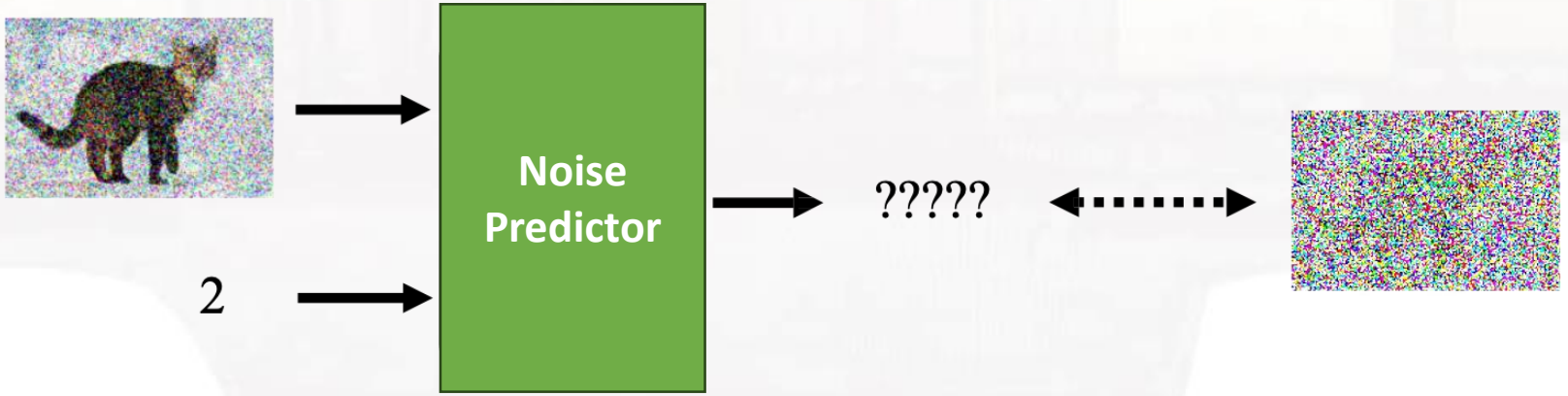


Q: Why not directly use an end-to-end model, where the input is the image to be denoised and the output is directly the denoised result?

A: It is possible to do so. However, most current papers still choose to use a noise predictor because the difficulty of generating an image is different from that of generating noise. If a denoising model can generate a noisy cat, it can almost be said that it already knows how to draw a cat. Therefore, the difficulty of generating a noisy cat is different from generating the noise contained within an image. Thus, directly using a noise predictor might be relatively simpler, whereas using an end-to-end model to directly produce the denoised result is relatively difficult.

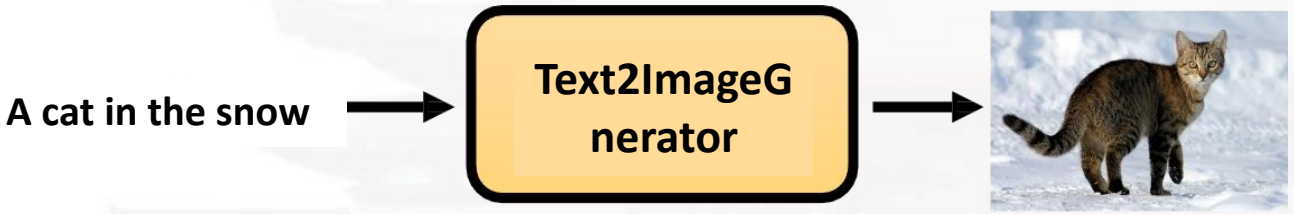
Diffusion Model

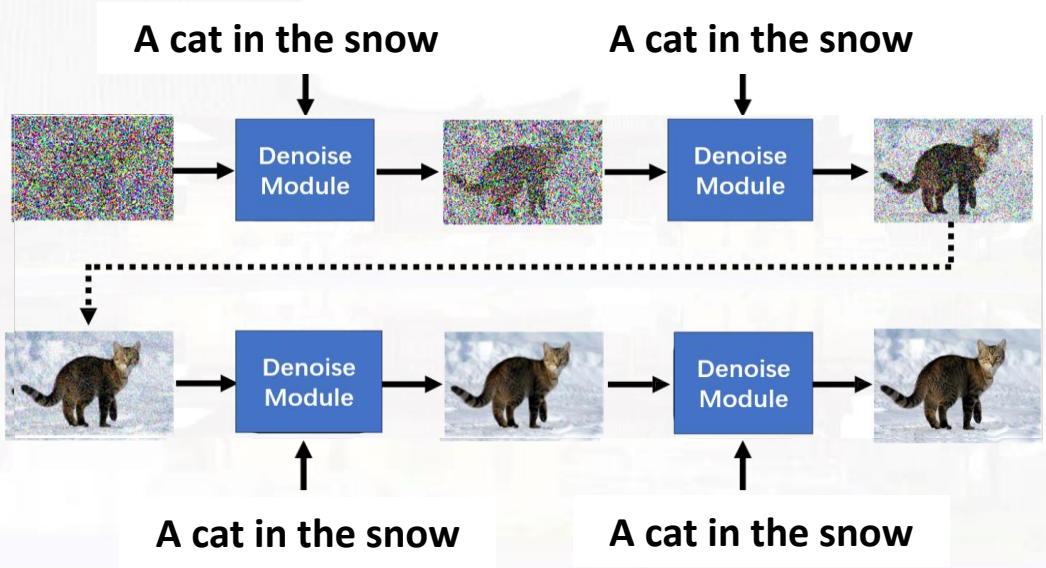


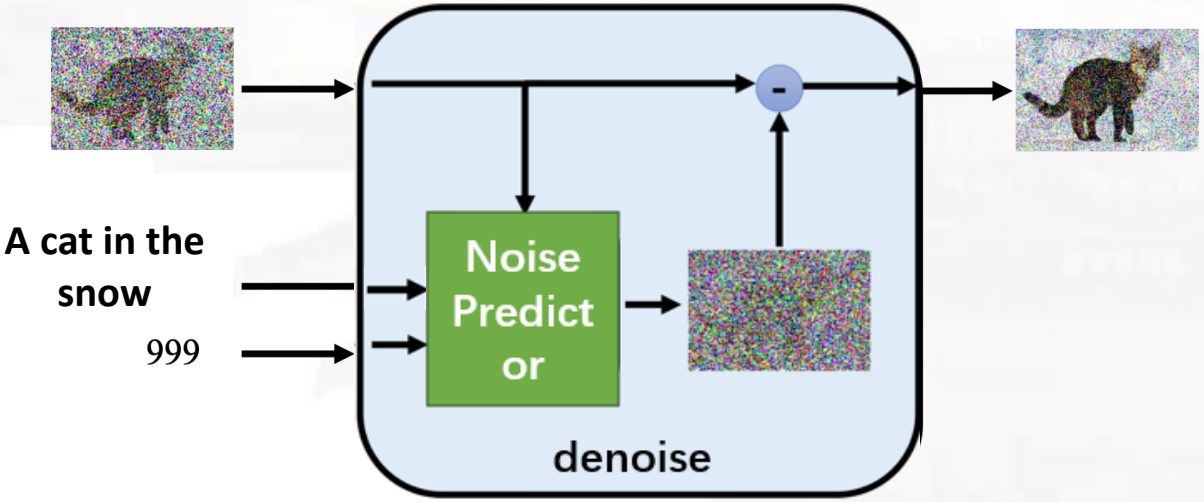


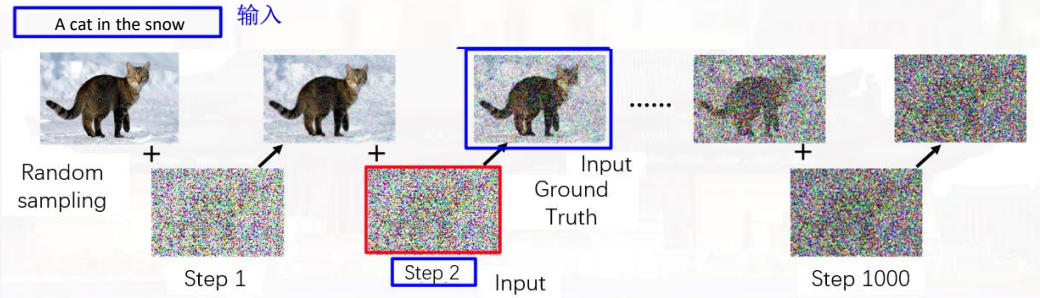
Diffusion Model

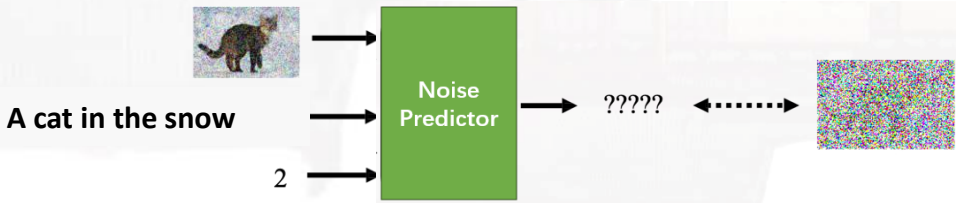
Text to Image





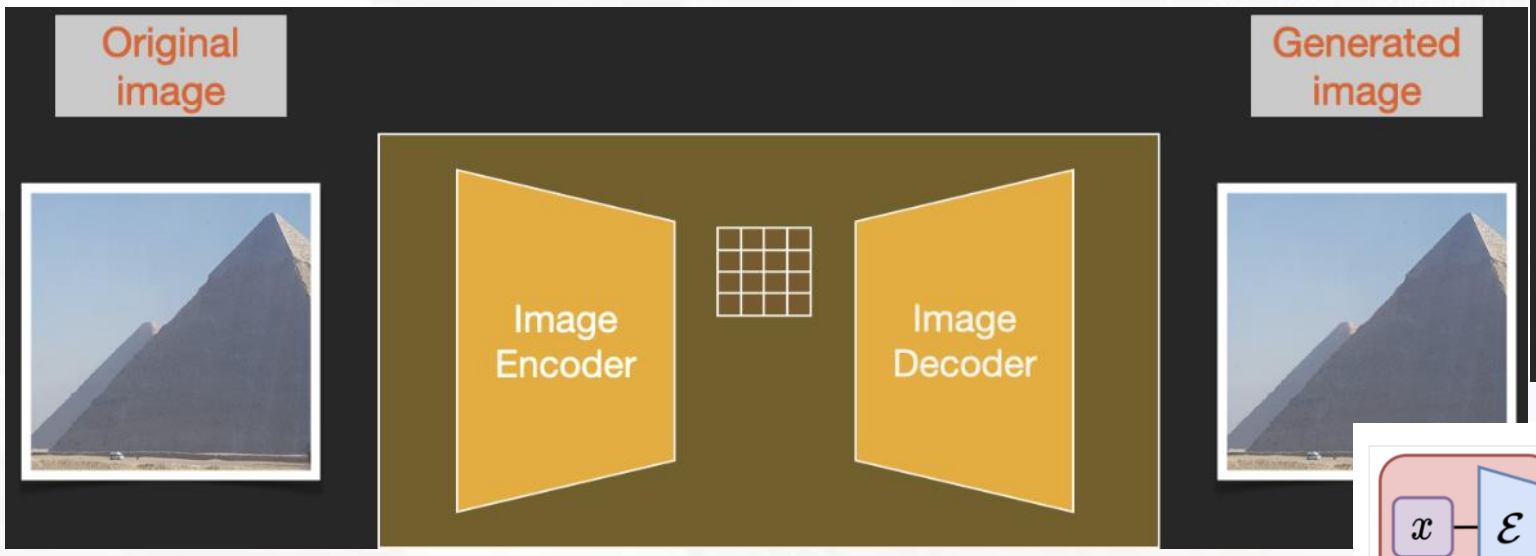


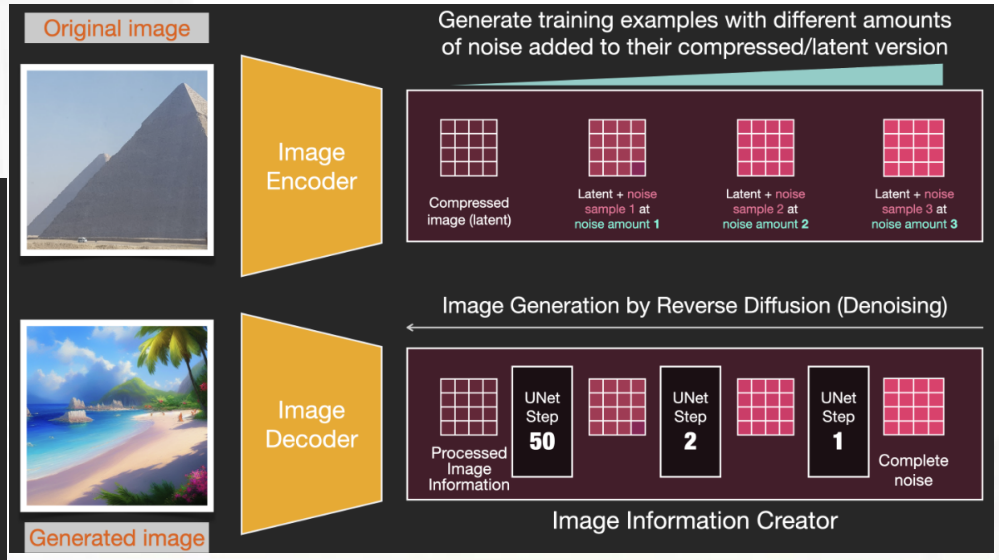


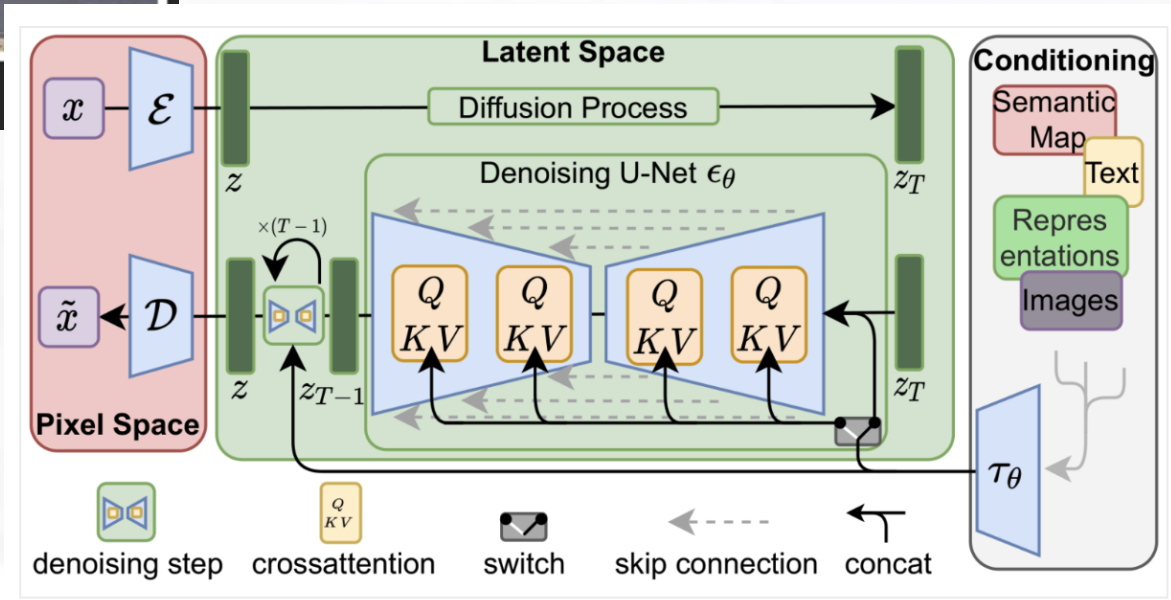


Diffusion Model

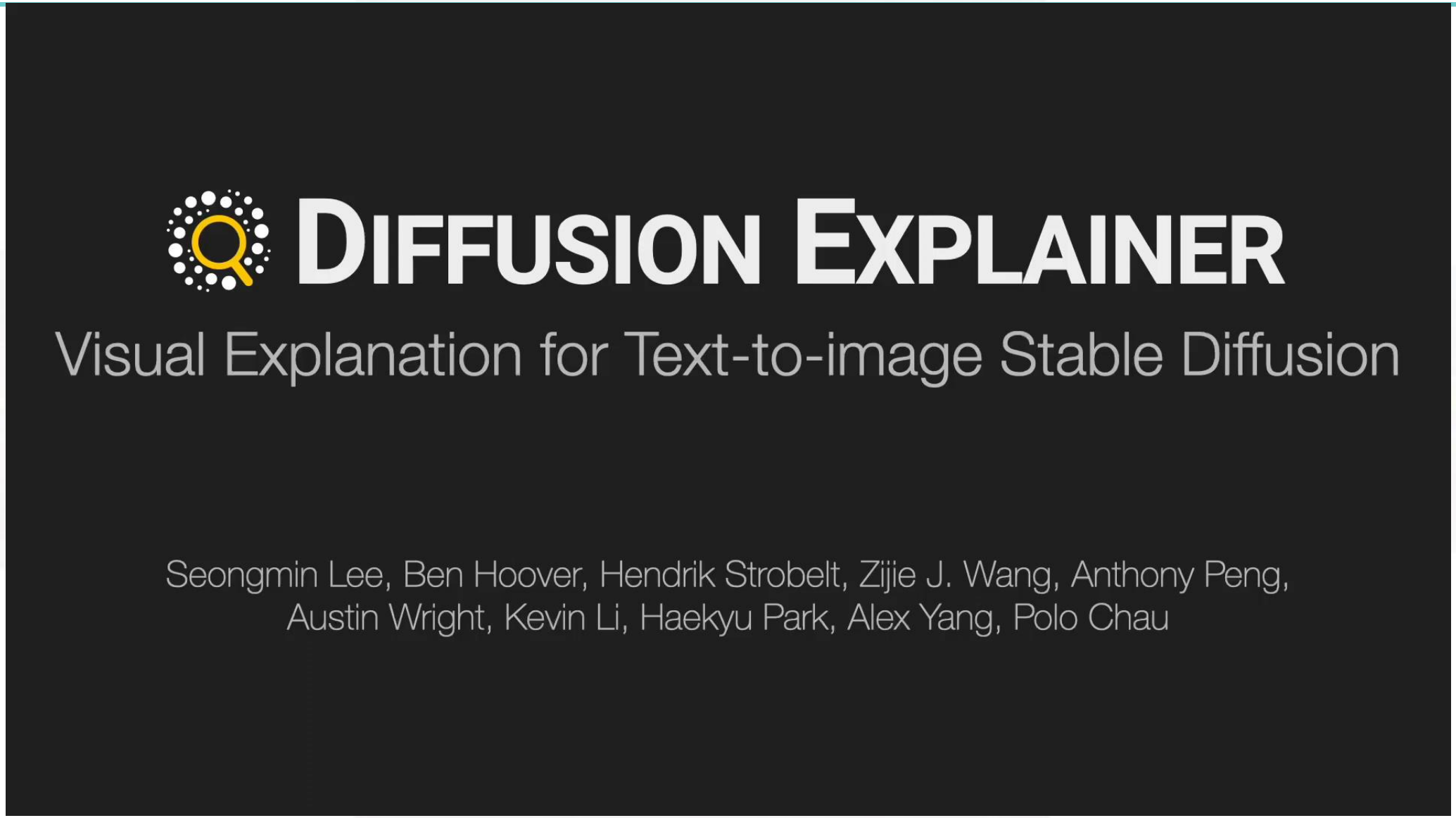
- **LDM**







Diffusion Model

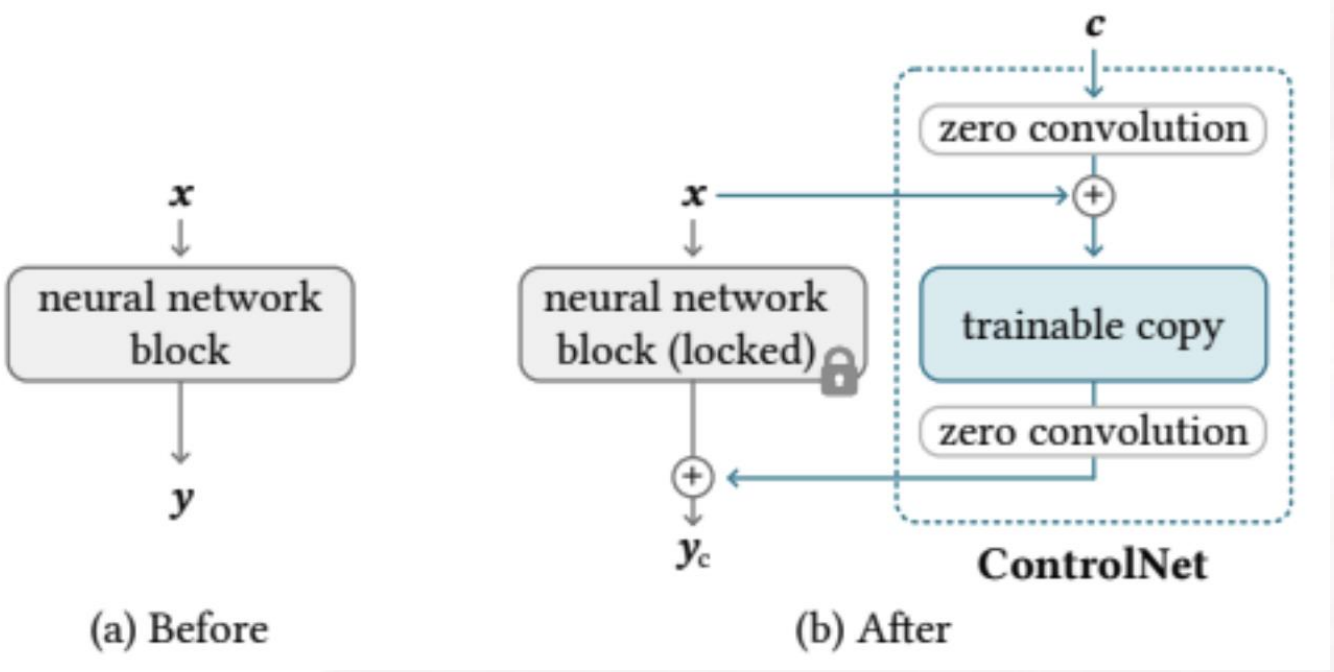


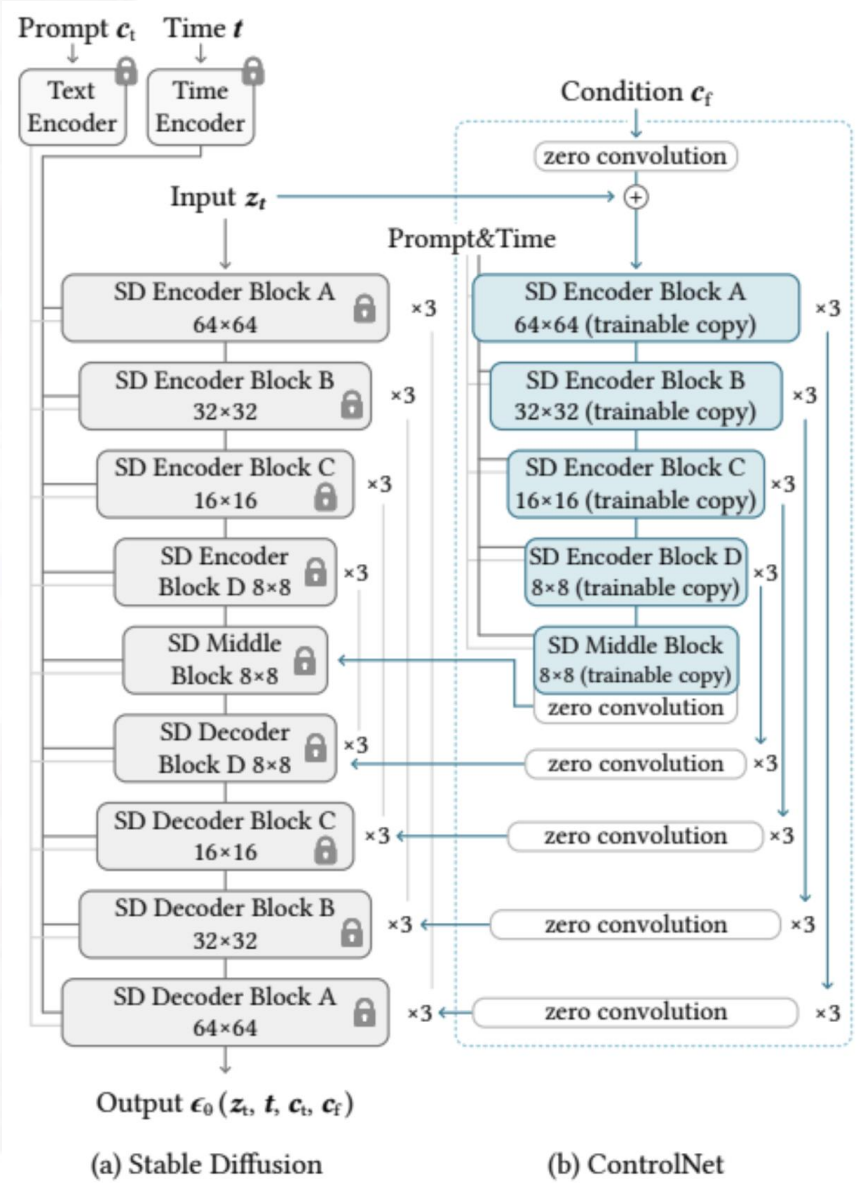
Diffusion Model

ControlNet

Diffusion Model

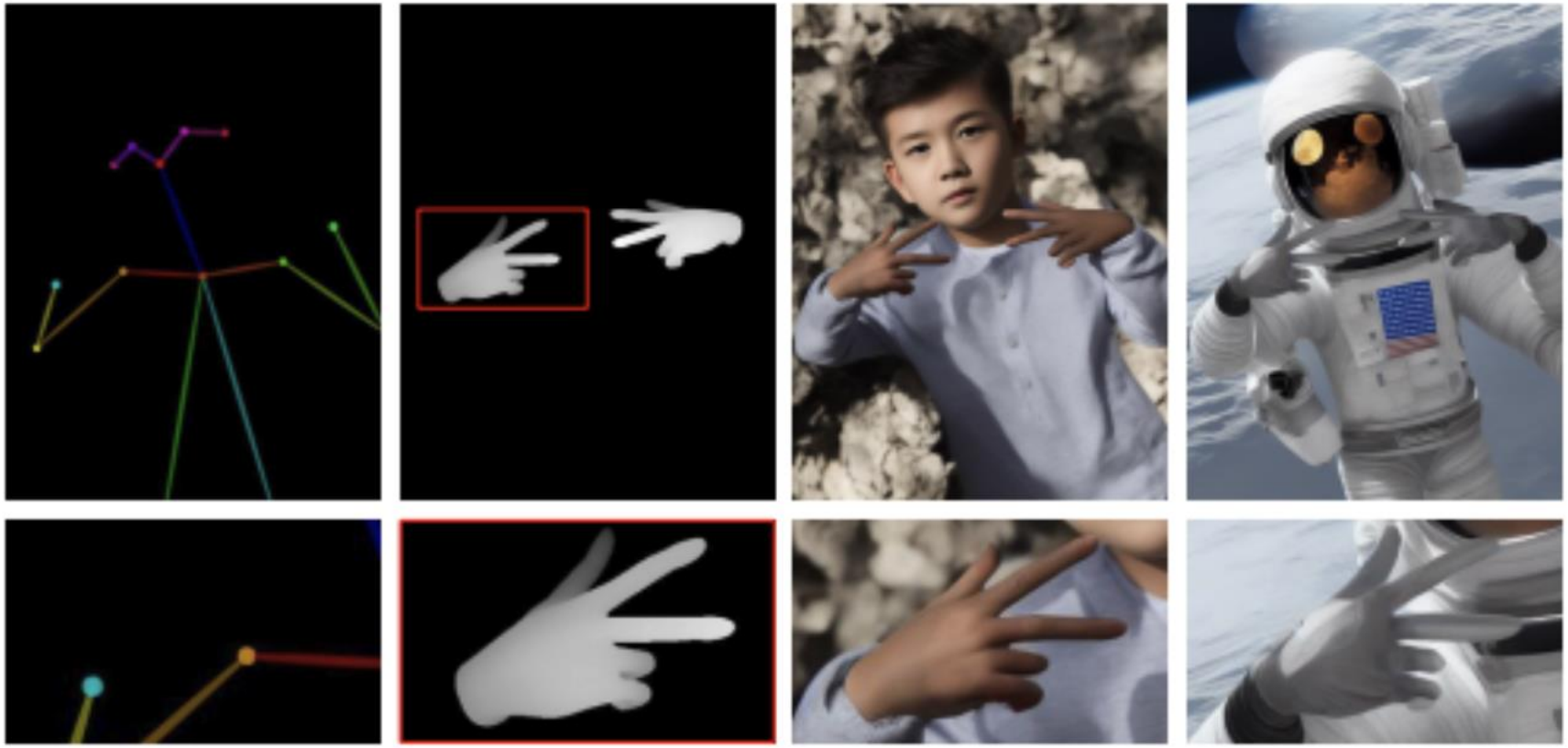
ControlNet





Diffusion Model

- **ControlNet**



Diffusion Model

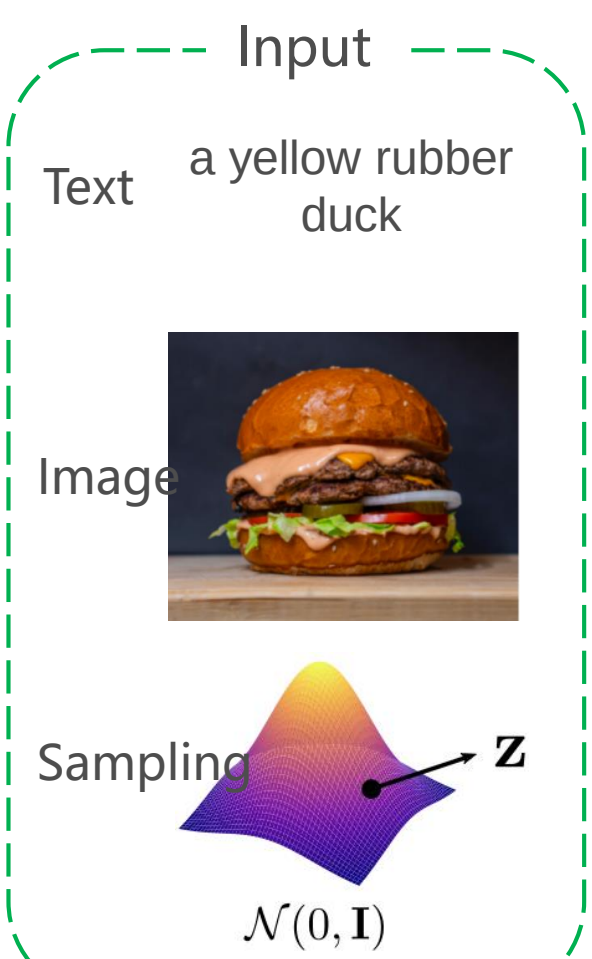
- **ControlNet**

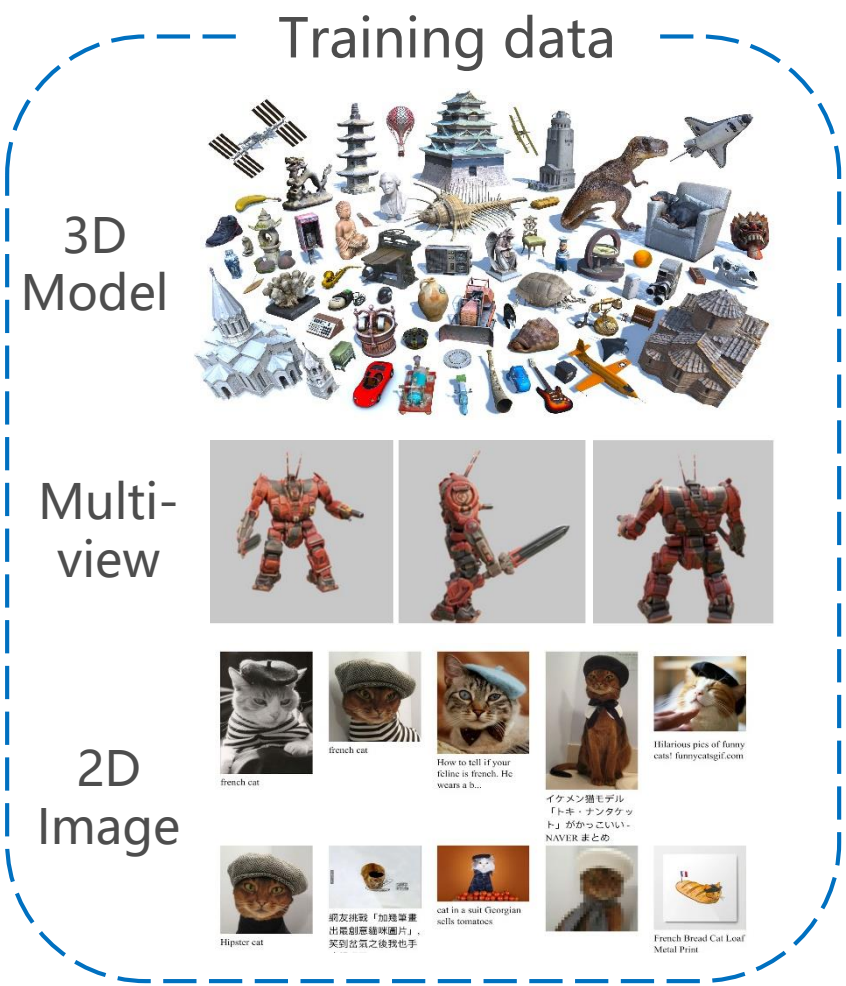
Diffusion Model

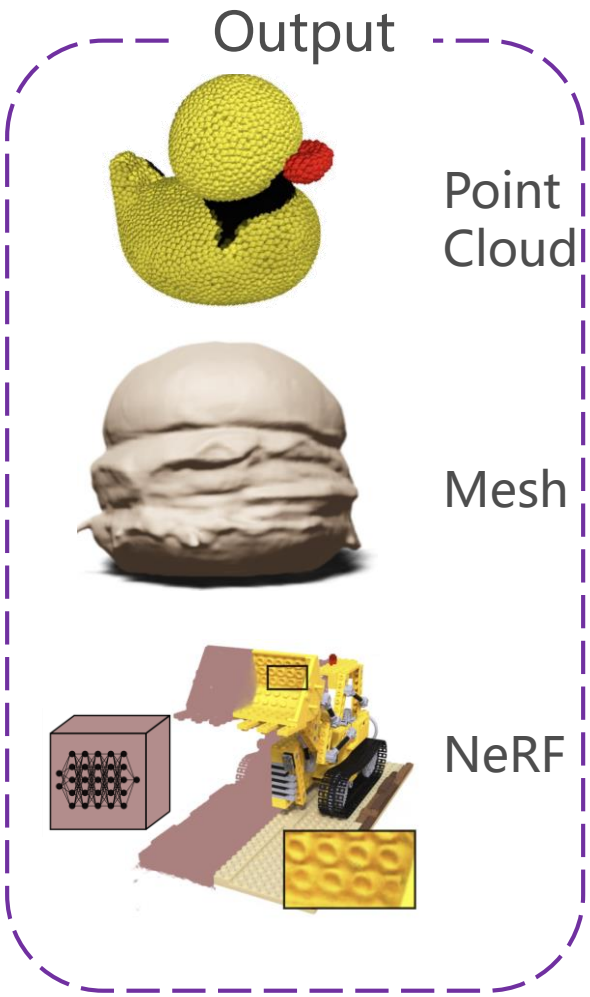
- **ControlNet**

Input "a high-quality and extremely detailed image"

3D AIGC

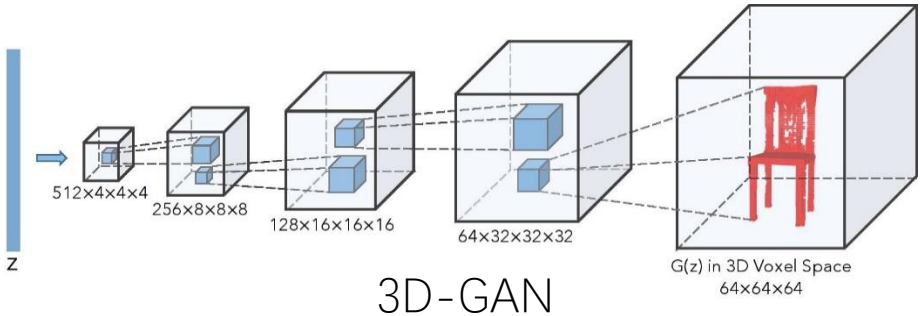

Input
Text
a yellow rubber duck
Image
Sampling
$\mathbf{z}$
$\mathcal{N}(0, \mathbf{I})$


Training data
3D Model
Multi-view
2D Image
french cat
How to tell if your feline is french. He wears a b...
Hilarious pics of funny cats! funnycatsgif.com
Hipster cat
cat in a suit Georgian sells tomatoes
French Bread Cat Loaf Metal Print

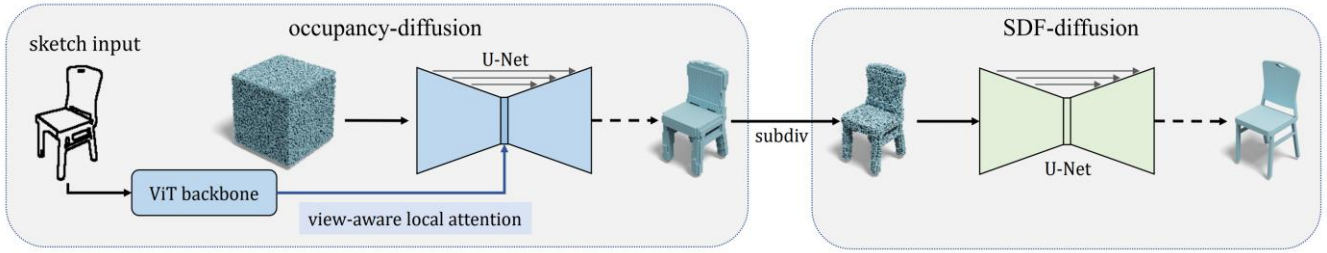

Output
Point Cloud
Mesh
NeRF

3D AIGC

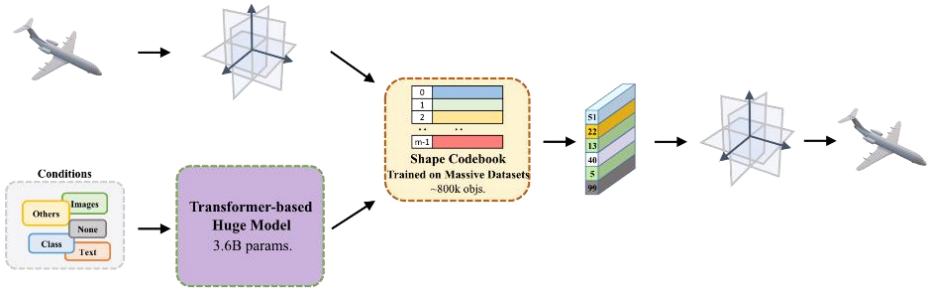
- 3D generation from 3D data
  - **GAN based**: 3D-GAN, Get3D
  - **Diffusion + Volume**: Rodin, LAS-diffusion…
  - **Autoregression**: Argus-3D, PolyGen
  - **LLM + procedural modeling**: 3D-GPT



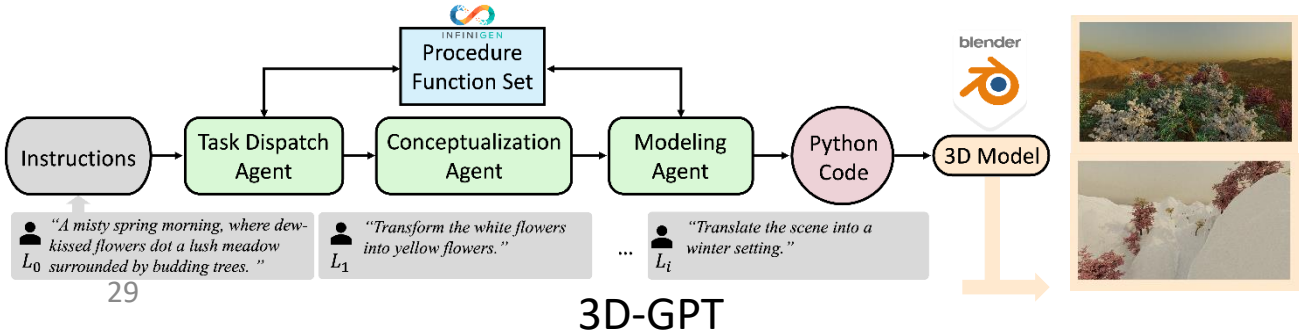
3D-GAN



LAS-diffusion



Argus-3D



3D-GPT

3D AIGC

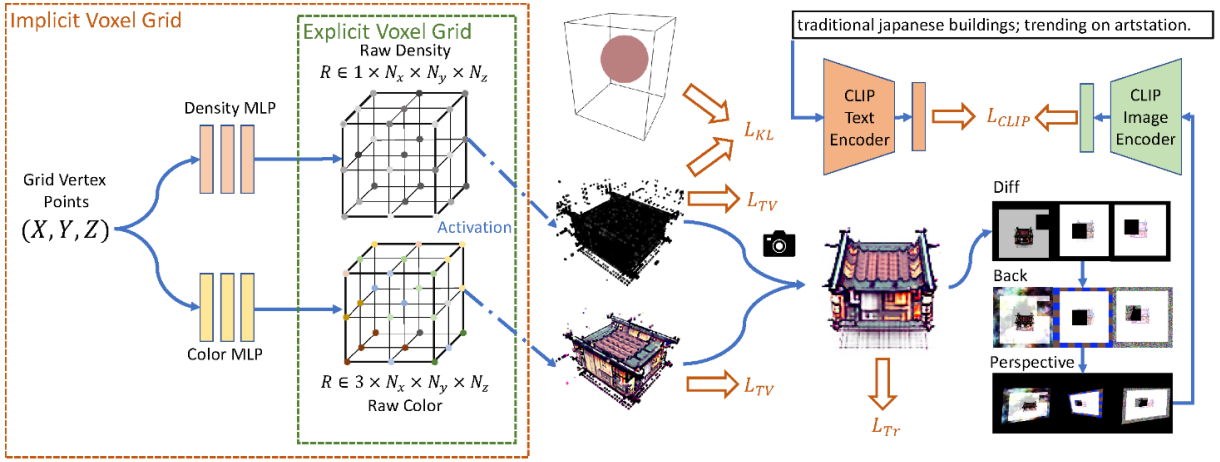
- 3D generation from 3D data

LAION-5B: 5.85 Billion of Images with Text

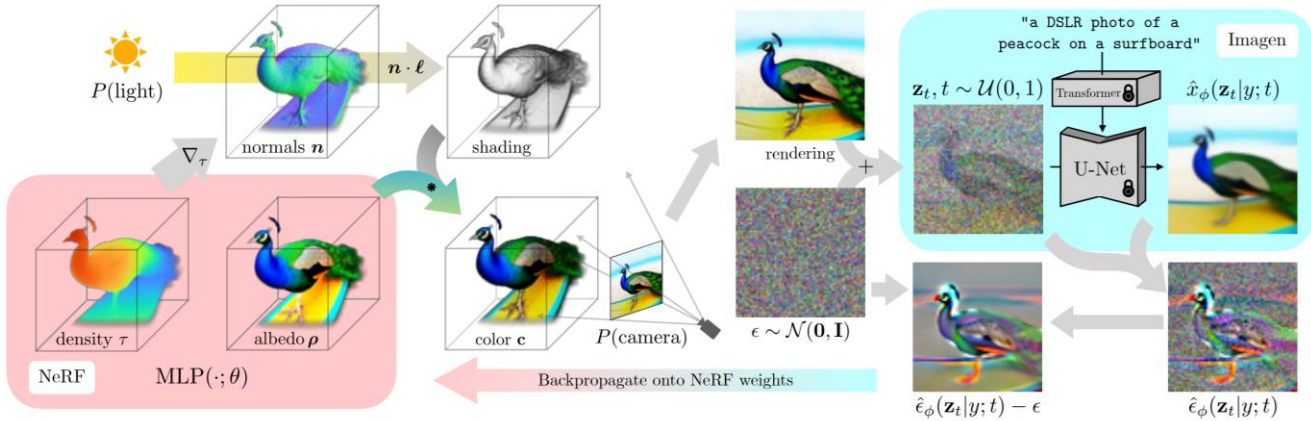
Objaverse-XL: 10 Million of 3D Objects

3D AIGC

- 3D Generation from 2D Data
  - **CLIP-guided 3D optimization**: PureCLIPNeRF, CLIP-mesh, Dream3D…
  - **Diffusion-based 3D optimization**: DreamFusion, Magic3D, Fantasia3D, ProlificDreamer…



PureCLIPNeRF



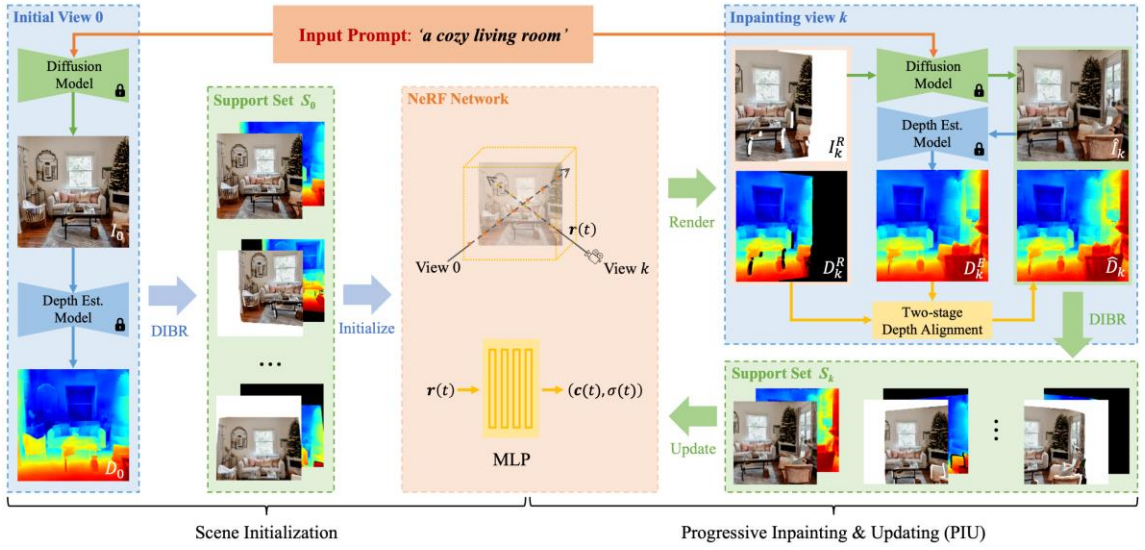
DreamFusion

3D AIGC

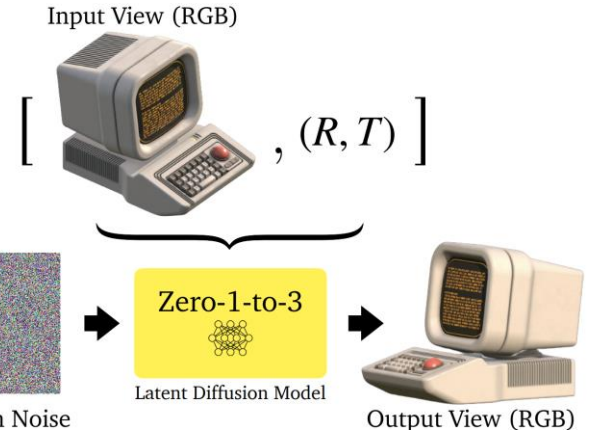
- 3D Generation from 2D Data
  - **CLIP-guided 3D optimization**: PureCLIPNeRF, CLIP-mesh, Dream3D…
  - **Diffusion-based 3D optimization**: DreamFusion, Magic3D, Fantasia3D, ProlificDreamer…

3D AIGC

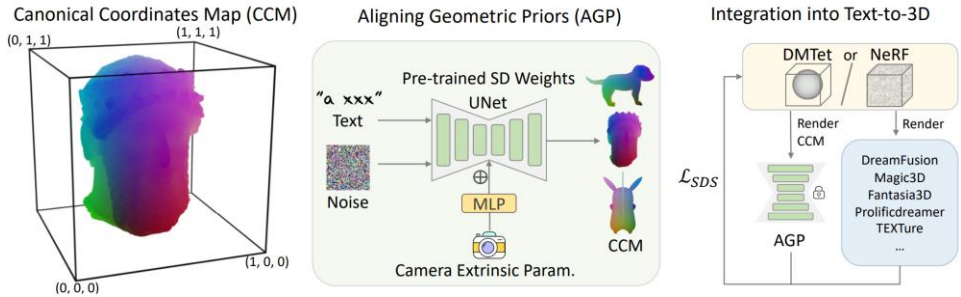
- 3D generation from hybrid data
  - **Progressive Generation**: Text2NeRF, MVDiffusion…
  - **Viewpoint-conditioned Diffusion**: Zero-1-2-3, MVDreamer, SyncDreamer, Wonder3D…
  - **3D-enhanced 2D Diffusion**: SweetDreamer



Text2NeRF



Zero-1-2-3



SweetDreamer