Image Generation

Bo Kang, Thomas Demeester, Tijl De Bie

Outline

- Introduction
- A Brief History
- Stable Diffusion Walk Through
- Demo: Train Your Own LoRA Model
- References

Outline

- Introduction
- A Brief History
- Stable Diffusion Walk Through
- Demo: Train Your Own LoRA Model
- References

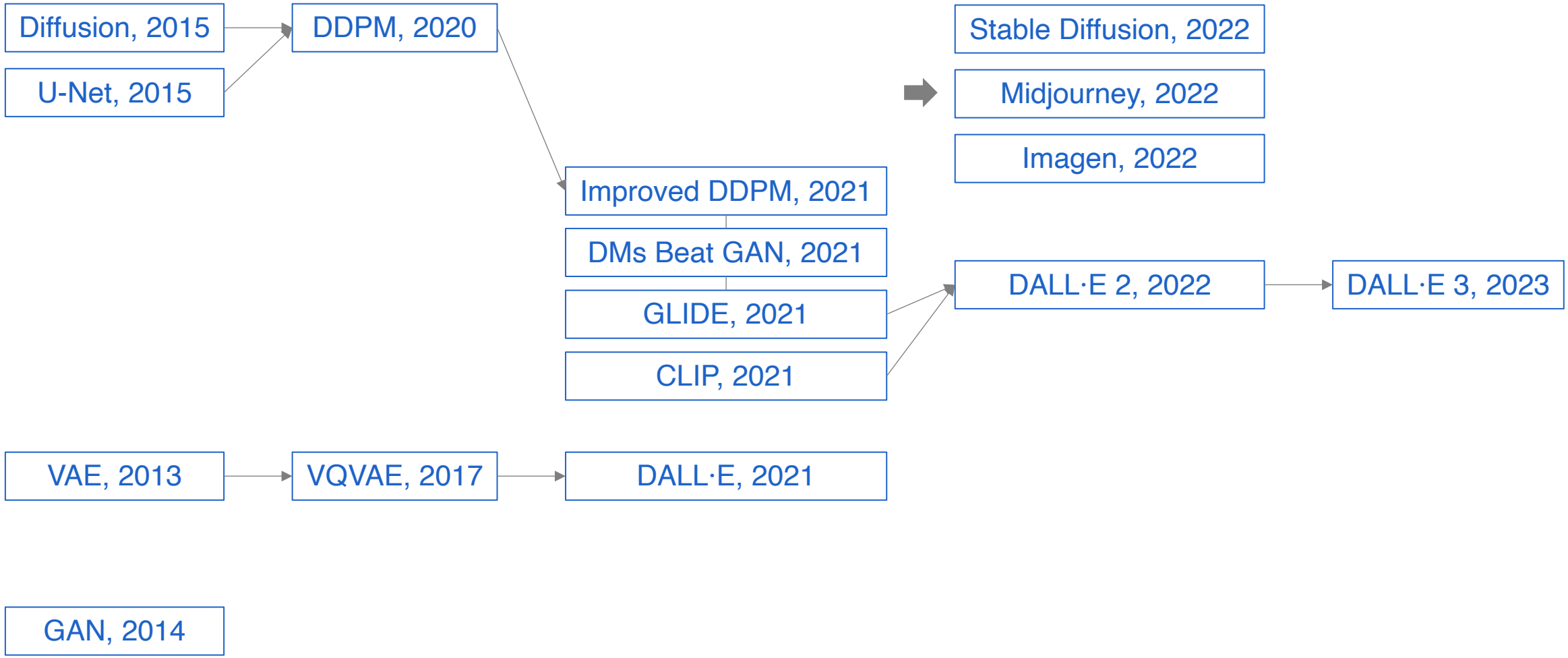
Introduction

- Generate or manipulate images with neural network models
- Applications
  - Unconditional generation
  - Text to image
  - Image to image
  - Inpainting
  - Many more…
- Tools
  - Commercial: OpenAI[1], Midjourney[2]
  - Open source: HF diffusers[3], SD Webui[4], ComfyUI[5], Kohya SS[6]

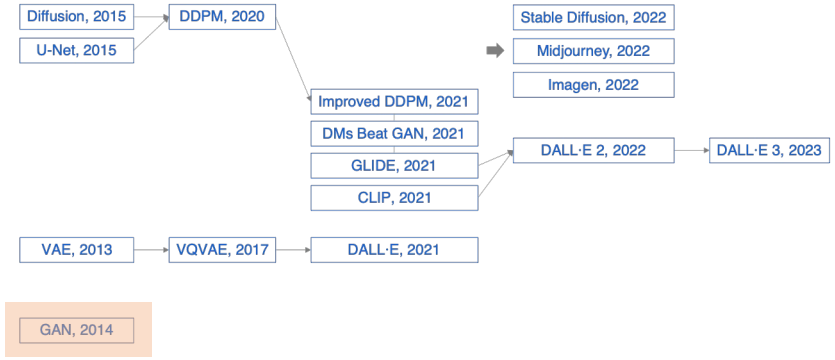
Outline

- Introduction
- A Brief History
- Stable Diffusion Walk Through
- Demo: Train Your Own LoRA Model
- References

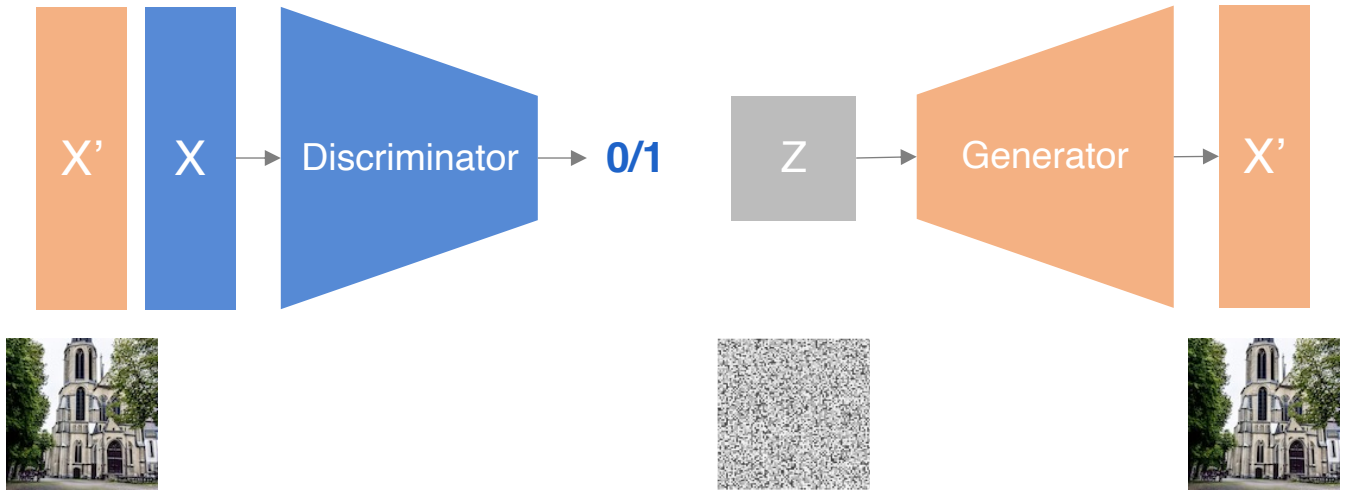
A Brief History



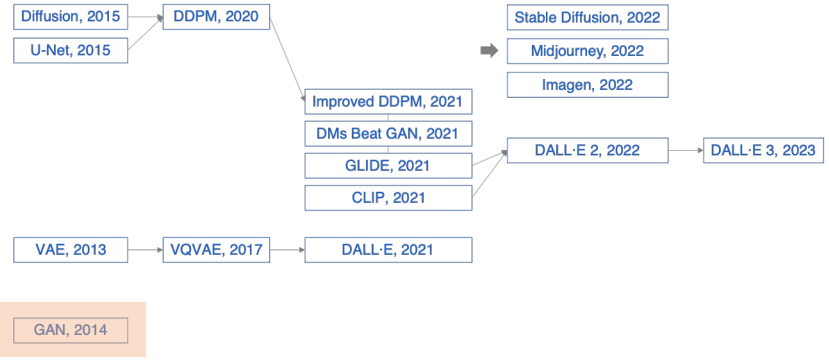
Generative Adversarial Network (GAN)



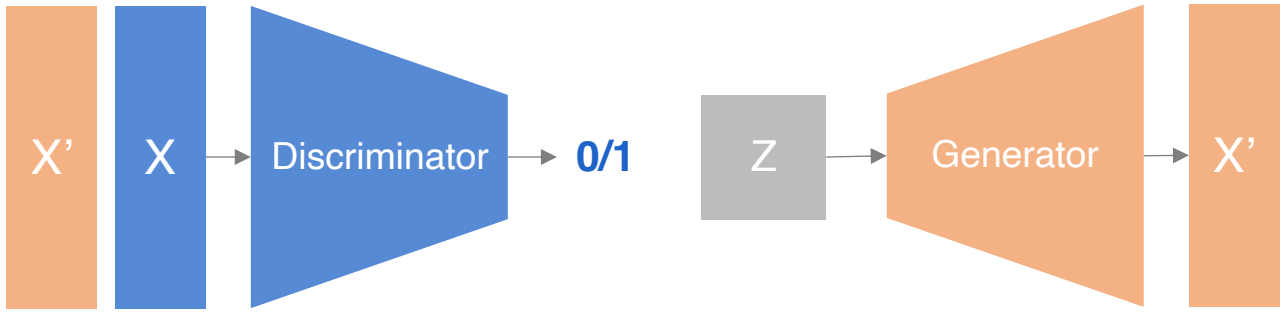
- Idea: train an image generator against a discriminator



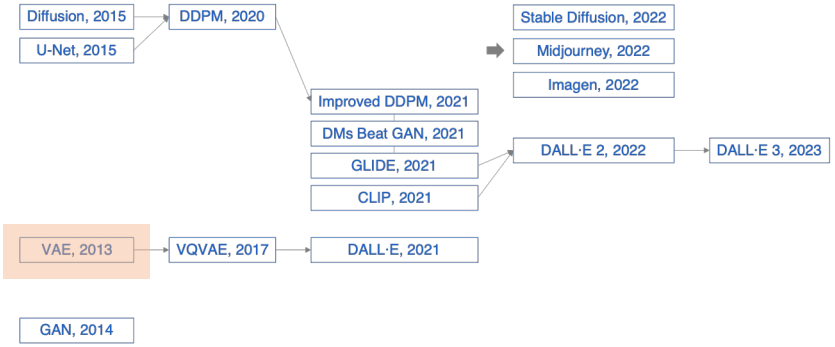
Generative Adversarial Network (GAN)



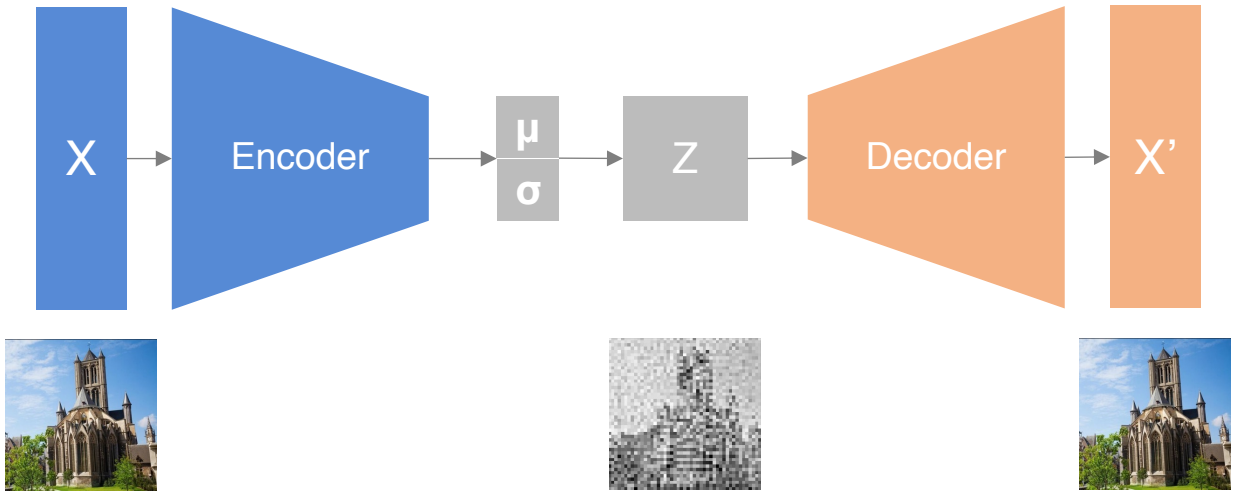
- Idea: train an image generator against a discriminator
- Pros
  - High fidelity
  - Many years of improving, easy to use
- Cons
  - Difficult to train
  - Low diversity
  - Mathematically less elegant



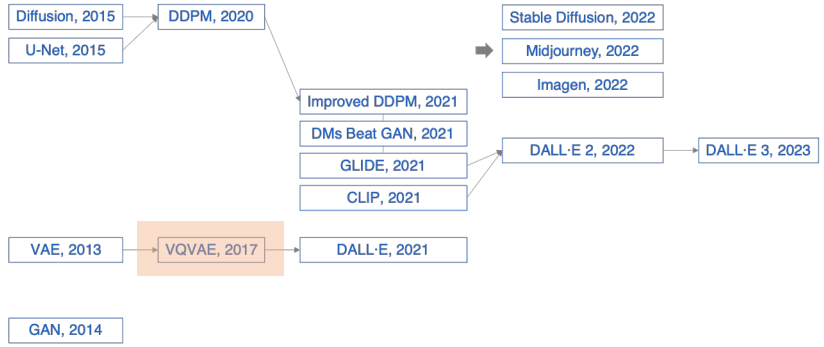
Variational Auto Encoder (VAE)



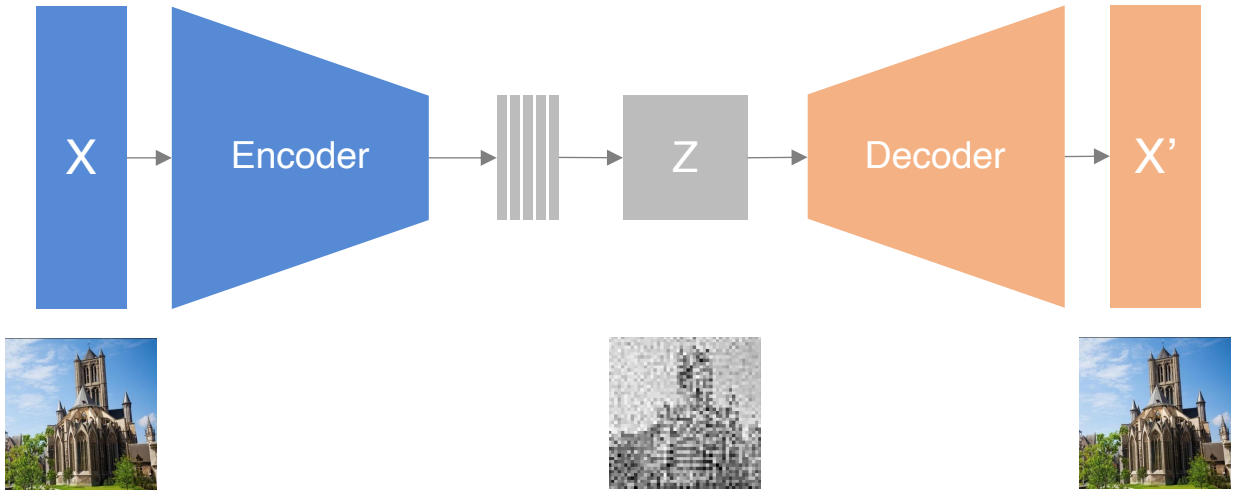
- Idea: learn a distribution over a latent space
- Pros
  - Better diversity
  - Principled probabilistic modeling
- Cons
  - Blurry outputs
  - Still difficult to train



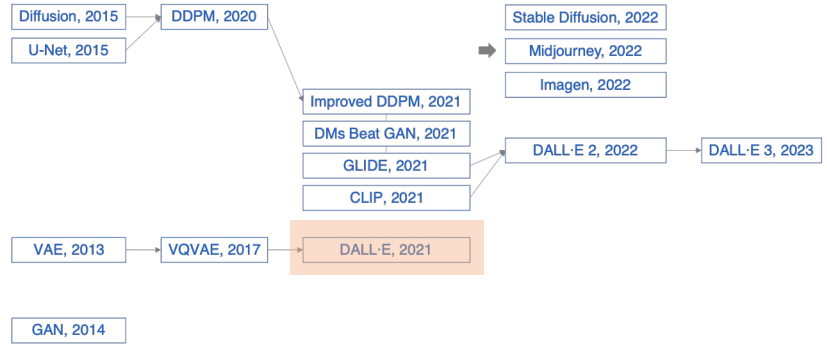
Vector Quantized VAE (VQVAE)



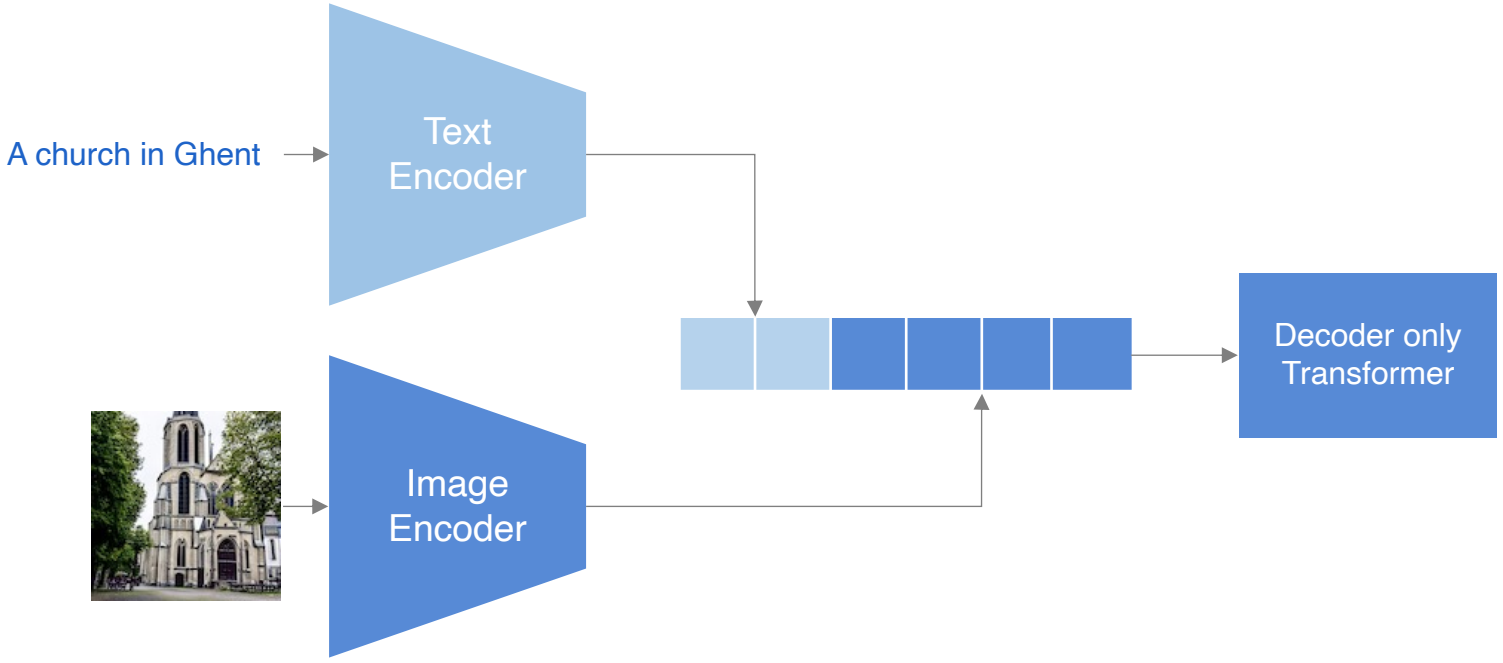
- Idea: learns discrete latent space using vector quantization; learns an autoregressive model for generation
- Pros:
  - Better sample quality
  - More efficient representation
- Cons:
  - Generation needs extra model
  - Training unstableness



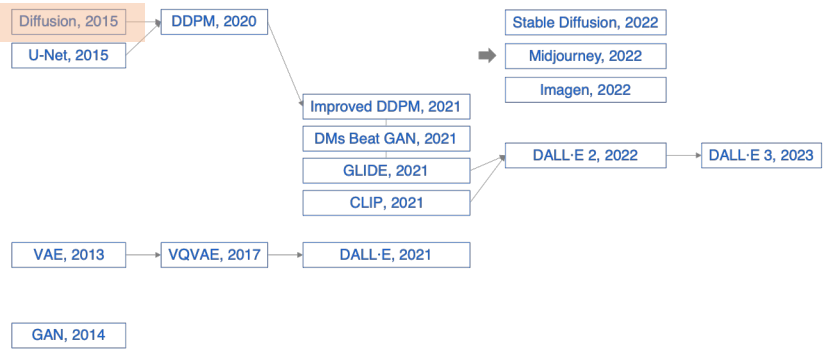
DALL·E



- Idea: VQVAE with text guidance and GPT style autoregressive latent representation generation



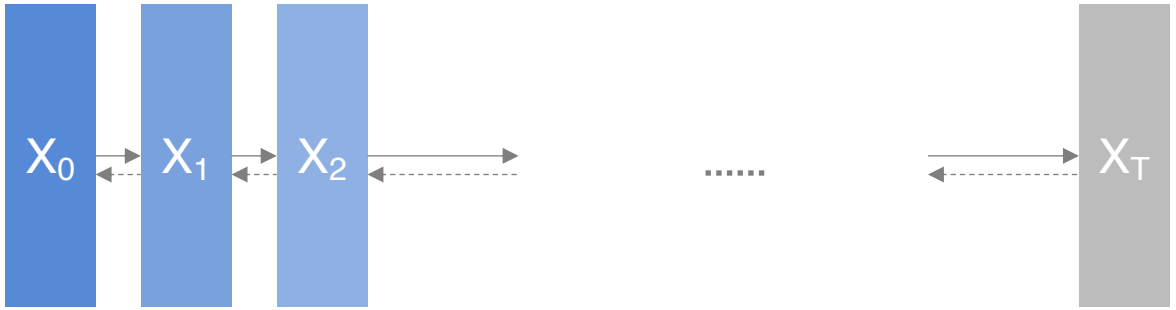
Diffusion model



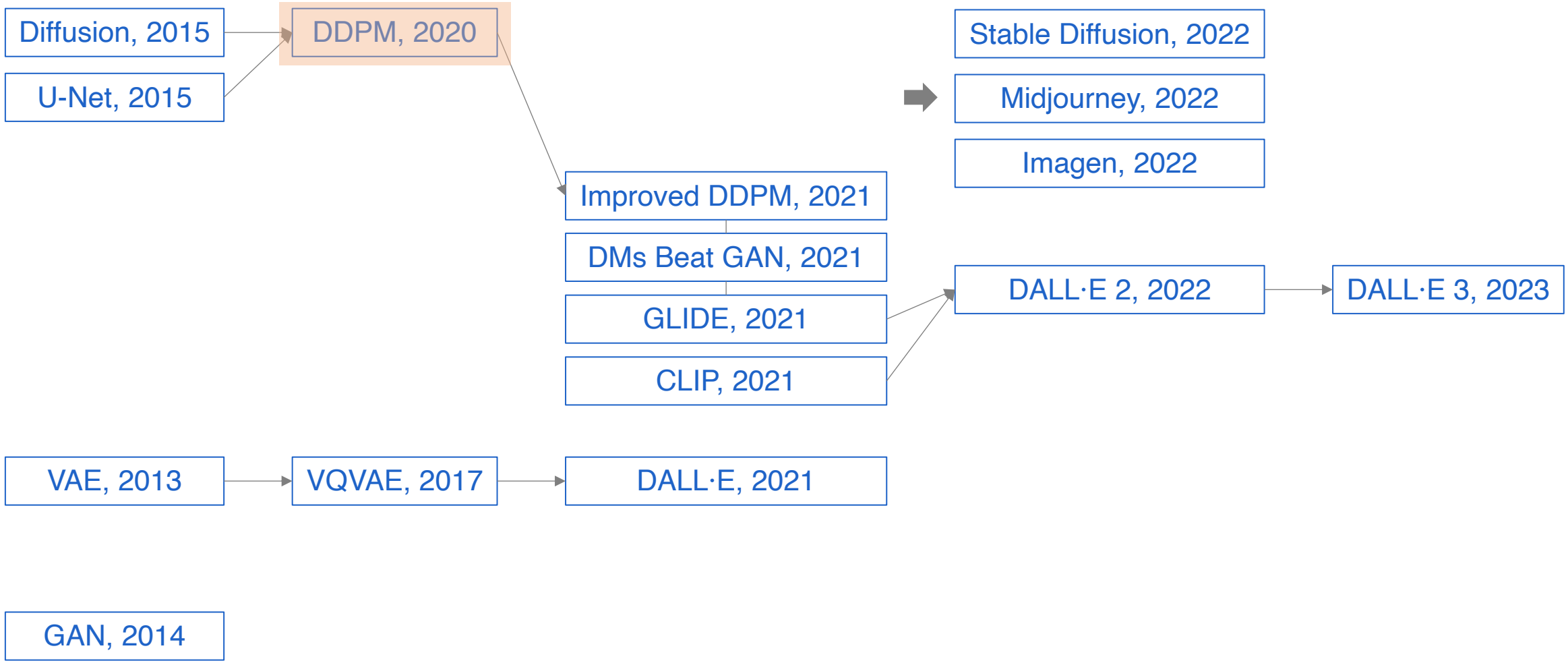
- Idea: gradually transform a distribution of random noise into a complex image data distribution through a reverse diffusion process

- Pros:
  - New paradigm
  - Mathematically principled

- Cons:
  - Generation is still not good enough
  - Slow

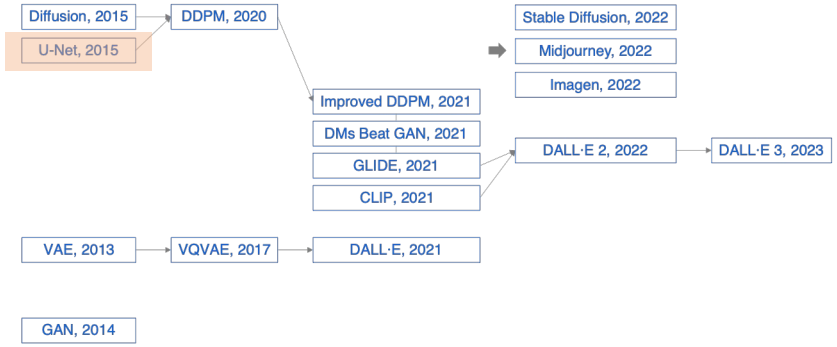


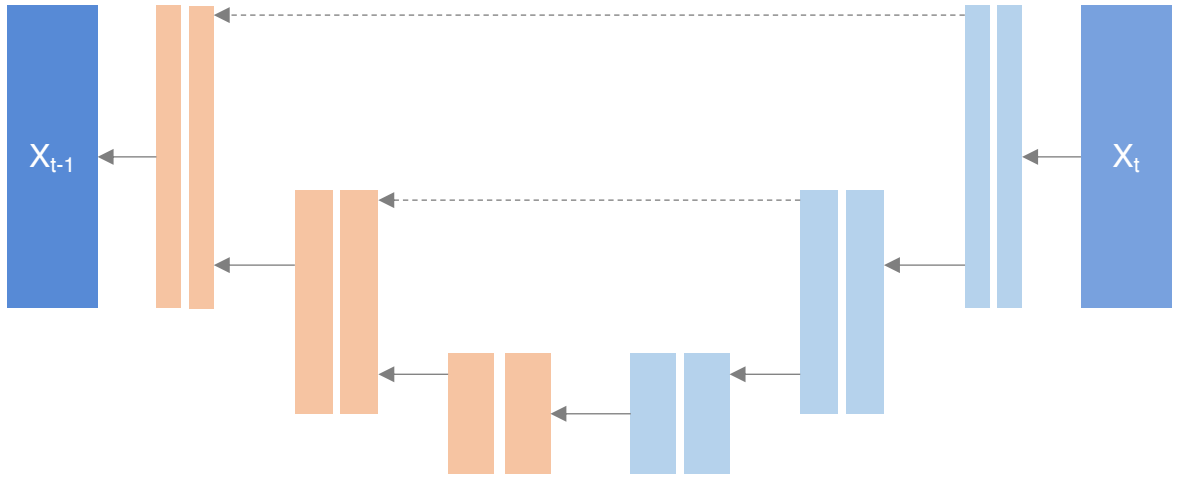
Diffusion model Improvements



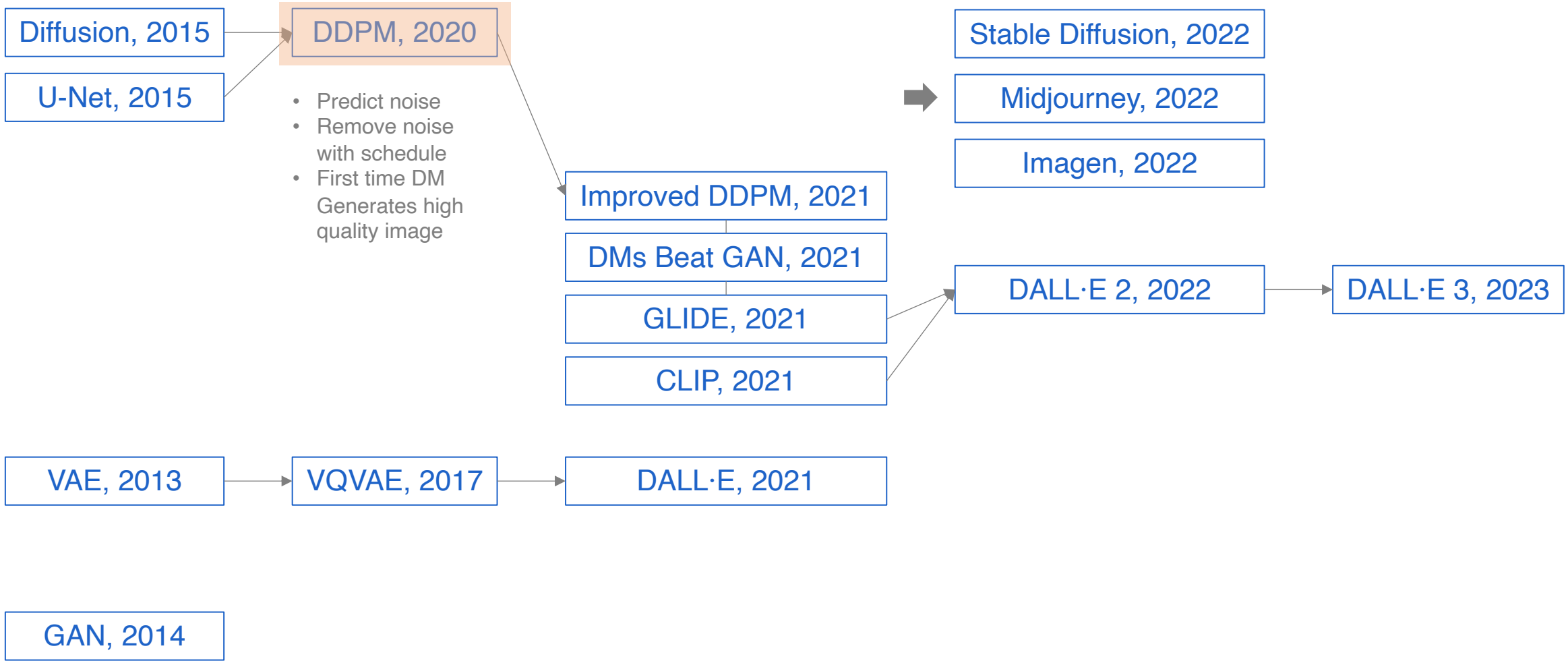
U-Net

- Idea:
  - CNN based encoder decoder architecture
  - Originally used to predict segmentation of an image
  - here predicts noise in the reverse diffusion process

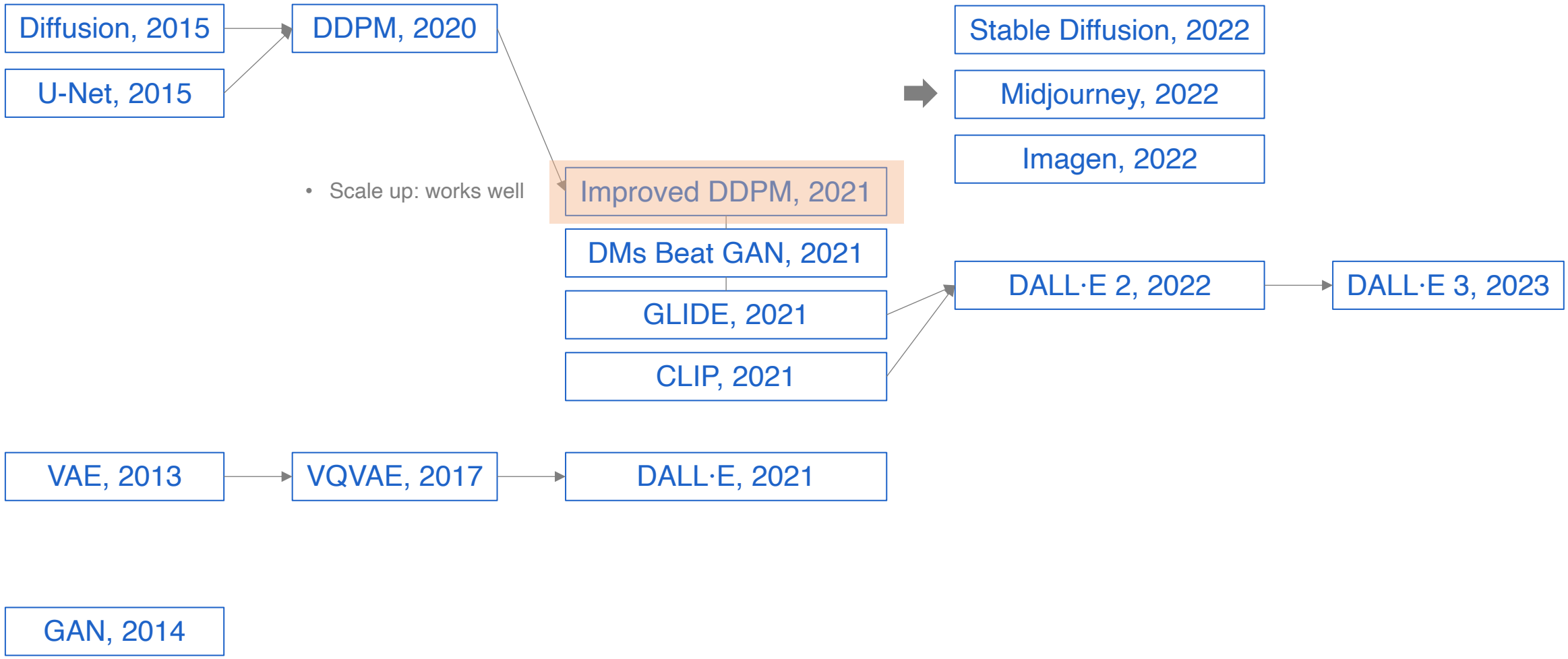




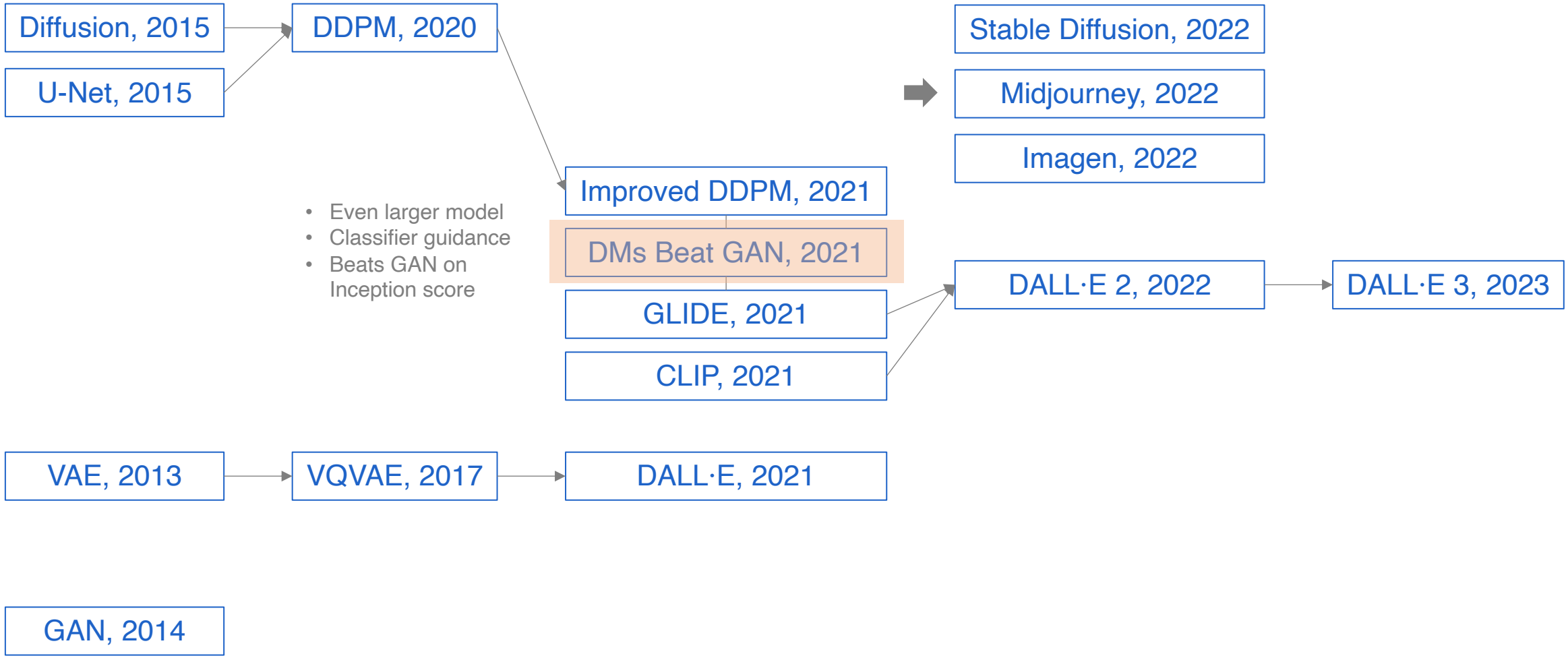
Diffusion model Improvements



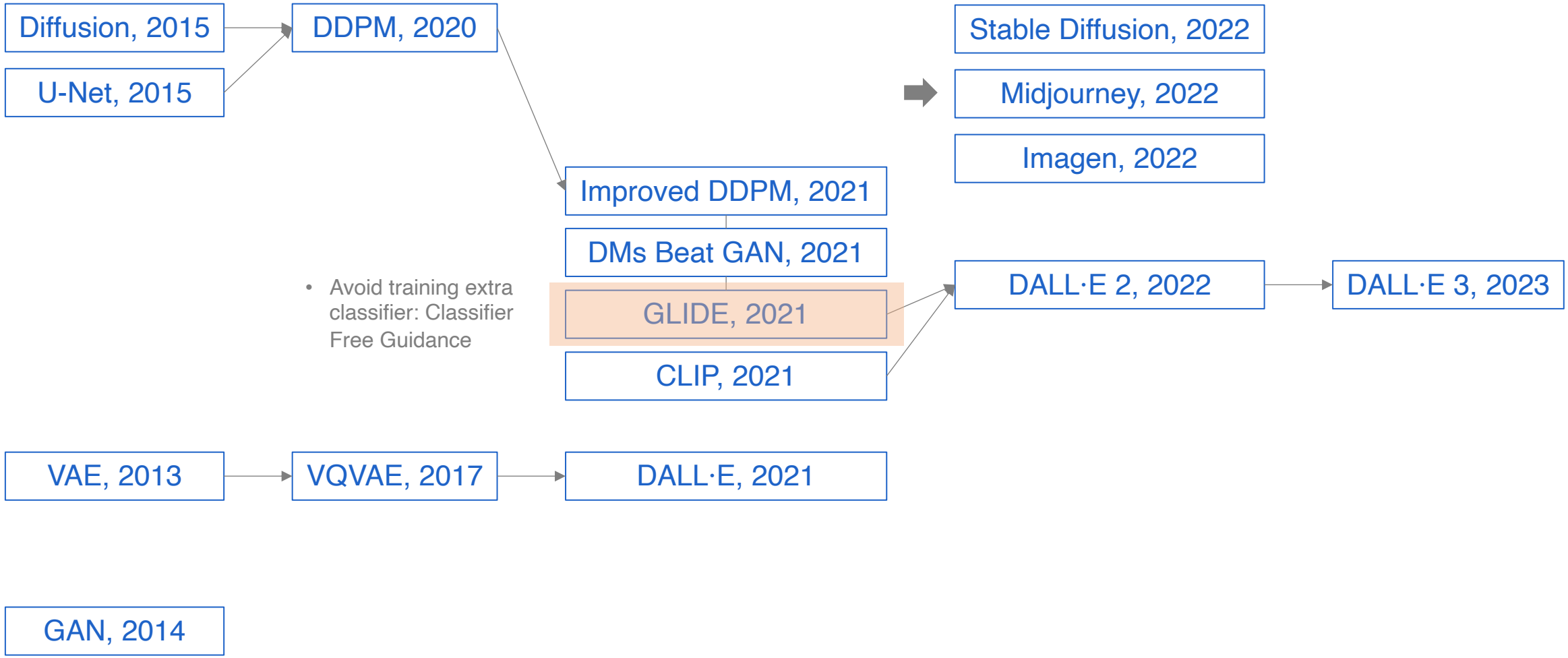
Diffusion model Improvements



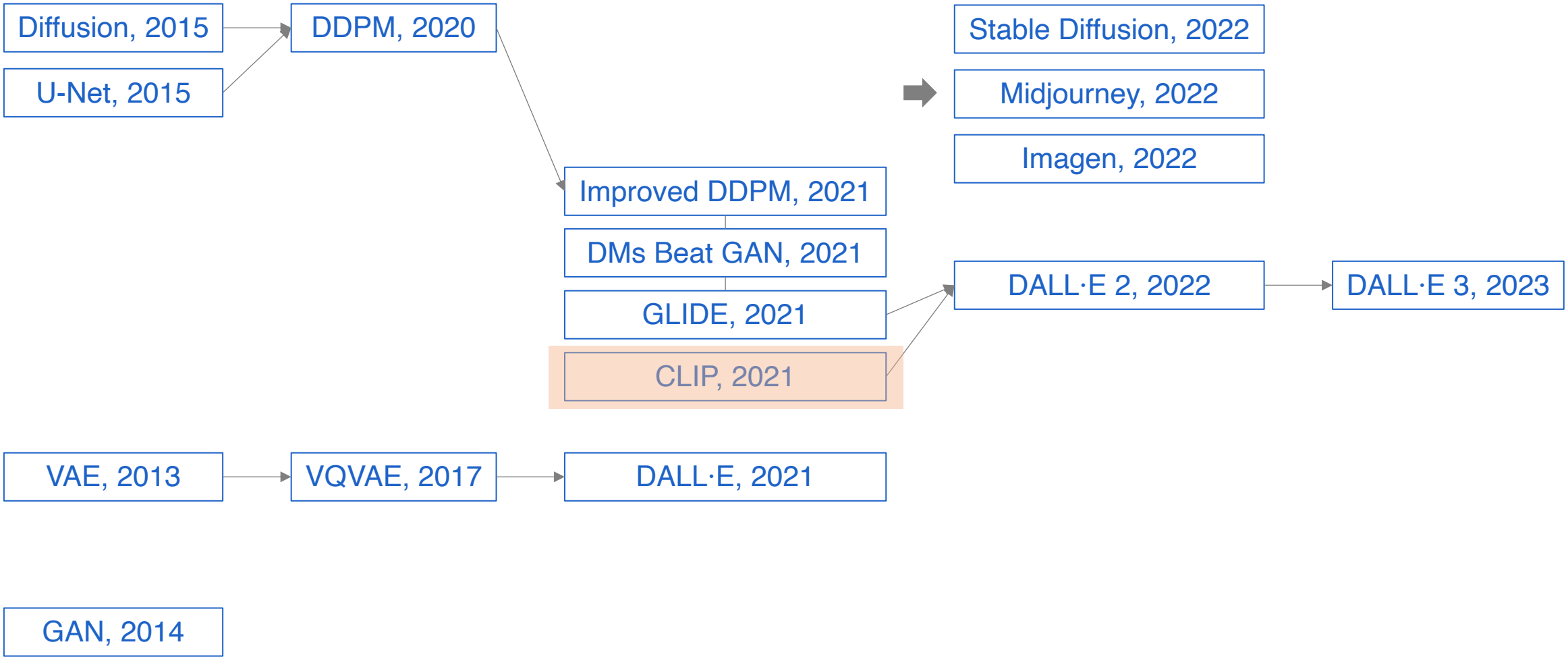
Diffusion model Improvements



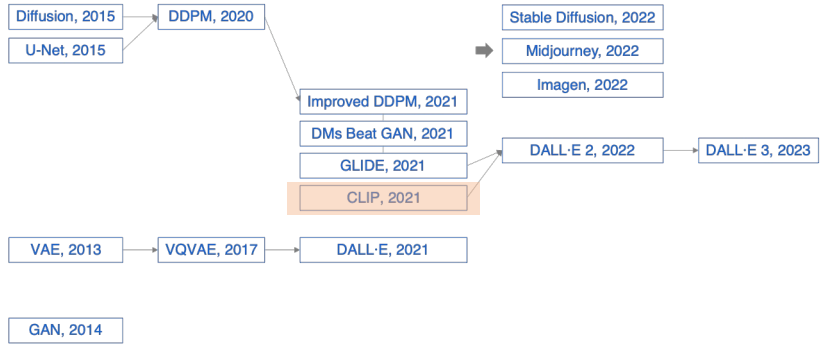
Diffusion model Improvements



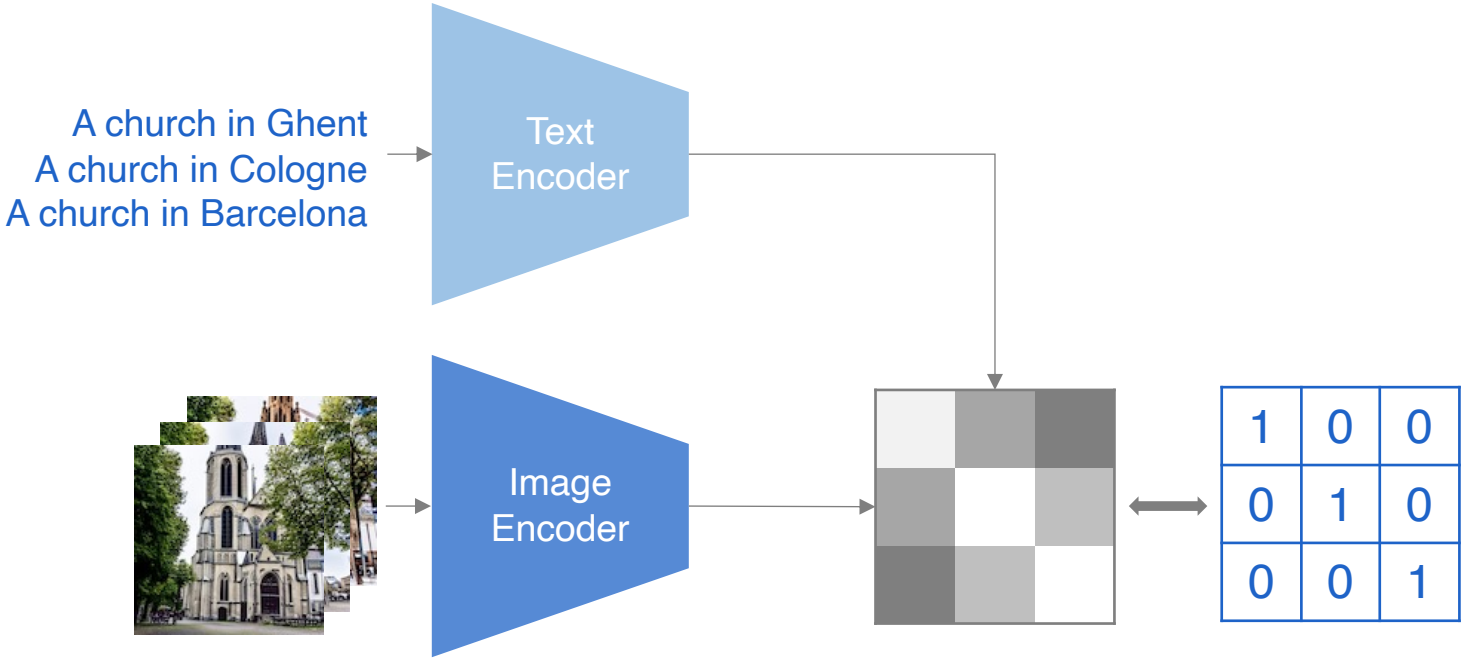
Diffusion model Improvements



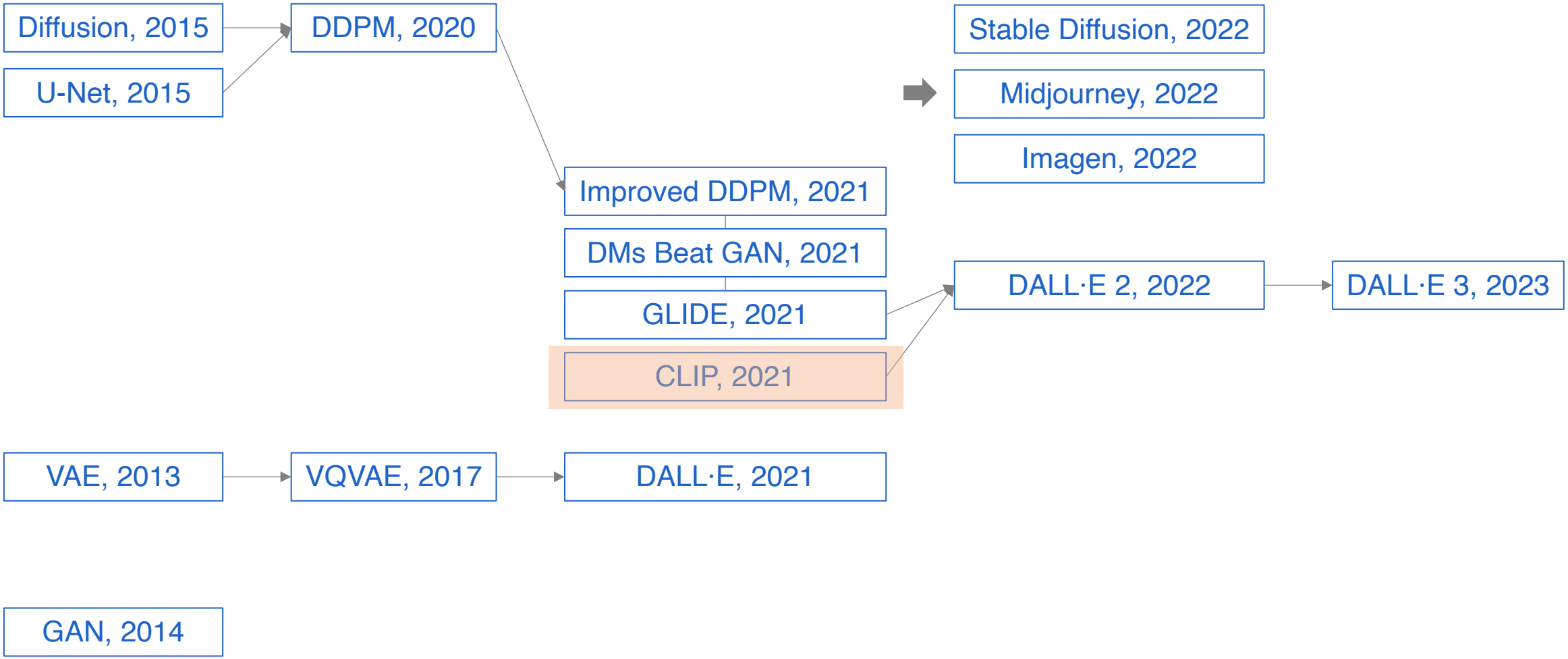
CLIP: Contrastive Language-Image Pre-training



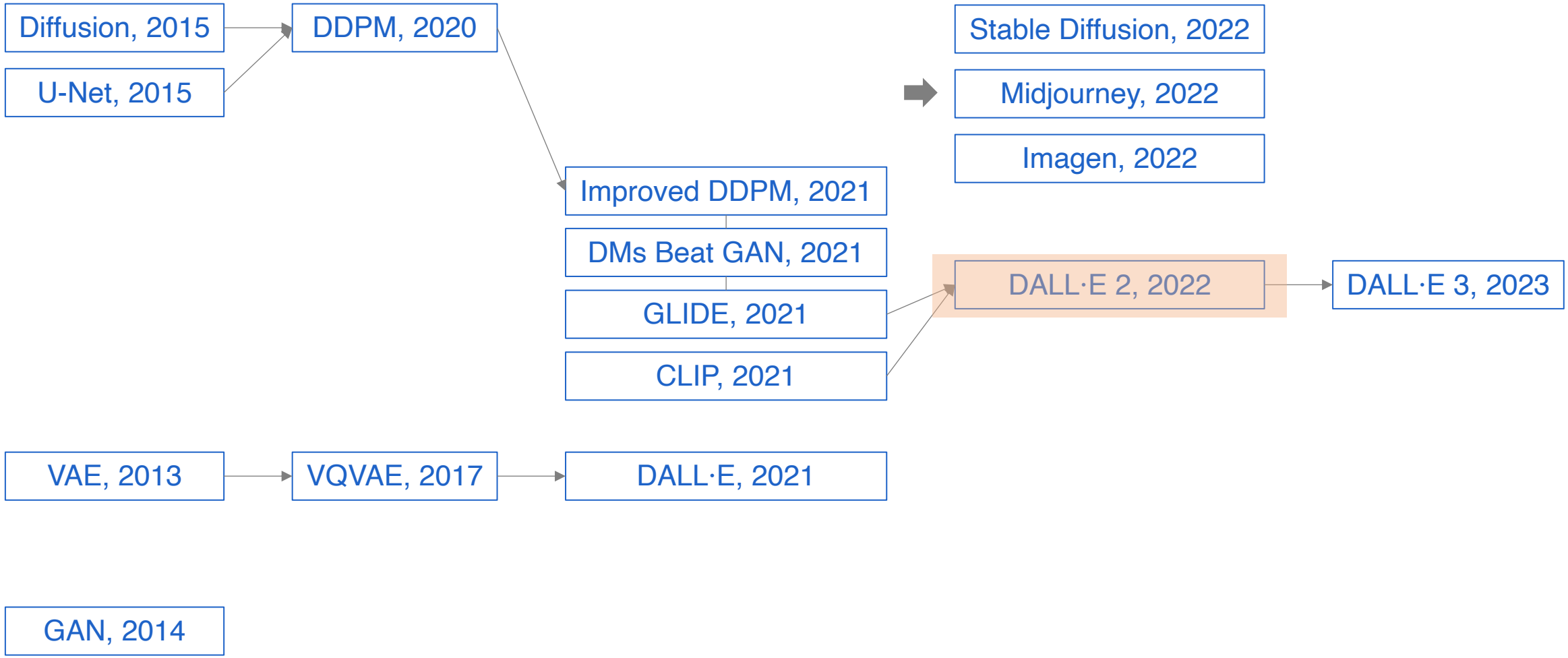
- Idea: learns image embeddings that matches relevant text embeddings



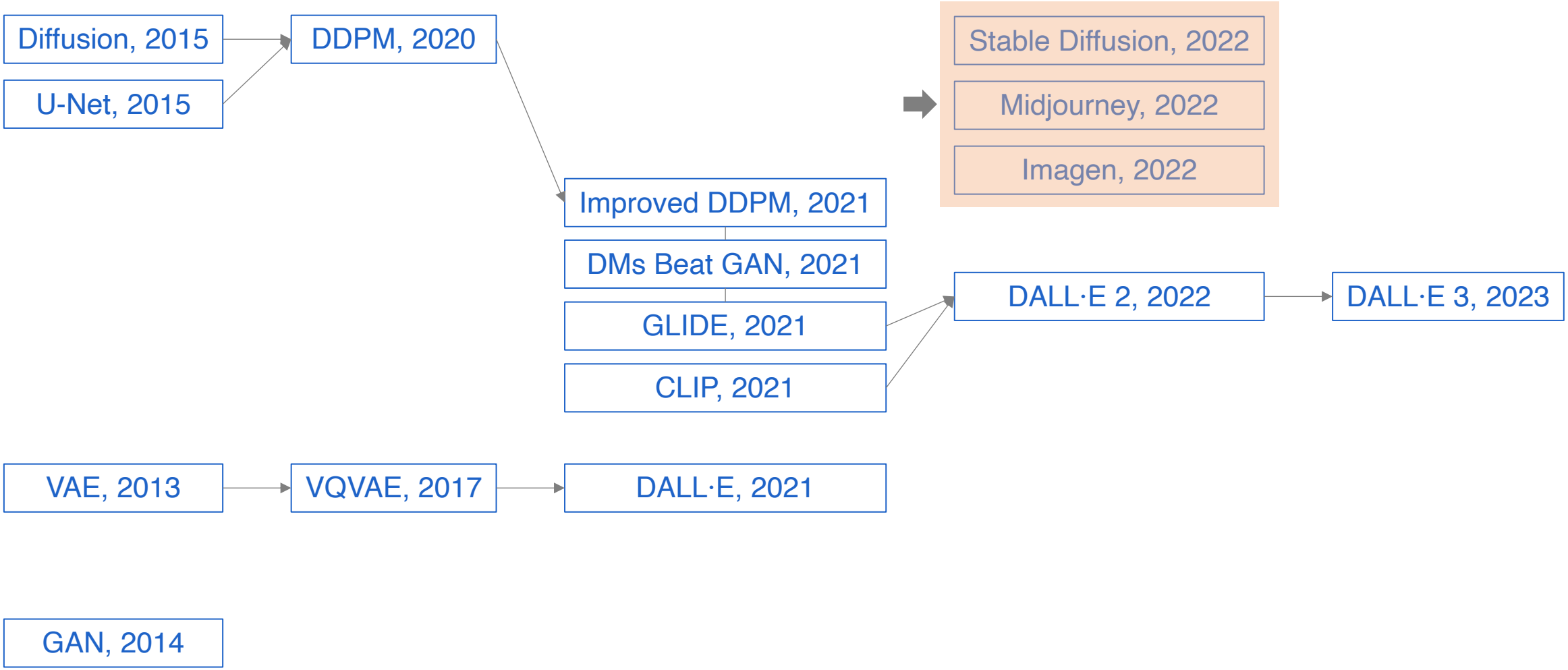
Diffusion model Improvements



Diffusion model Improvements

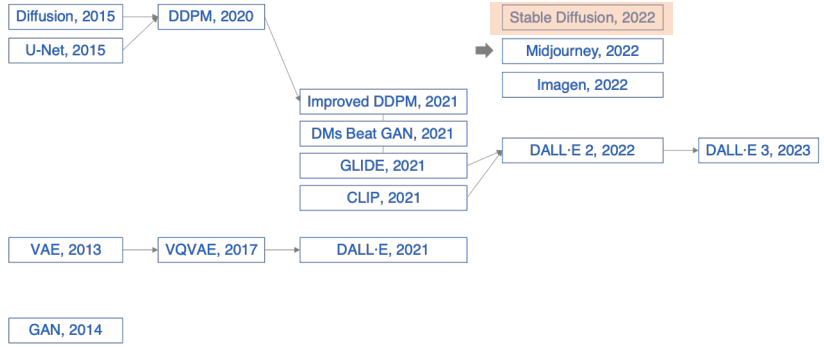


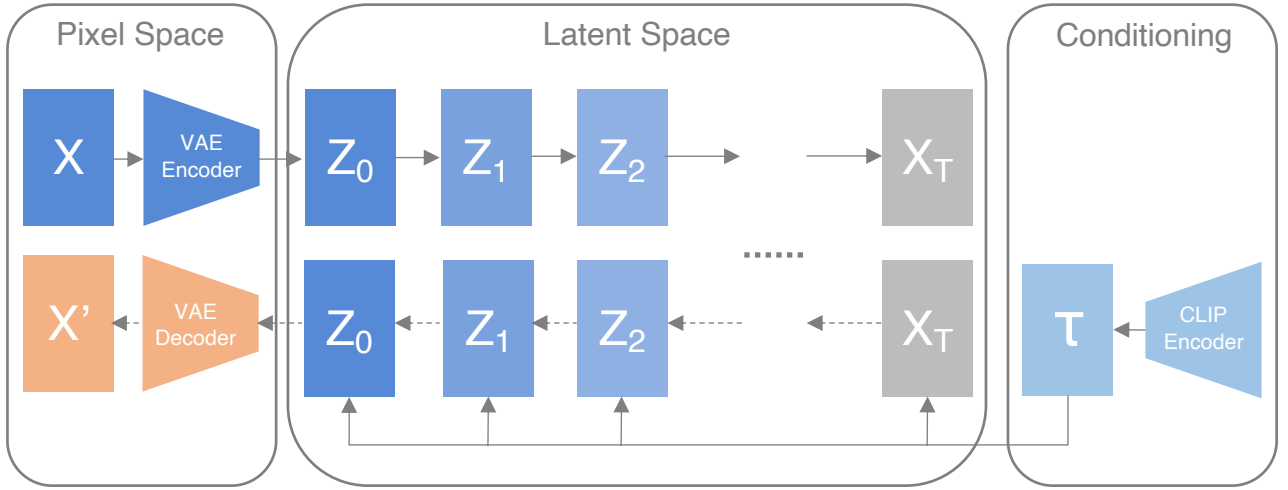
Diffusion model Improvements



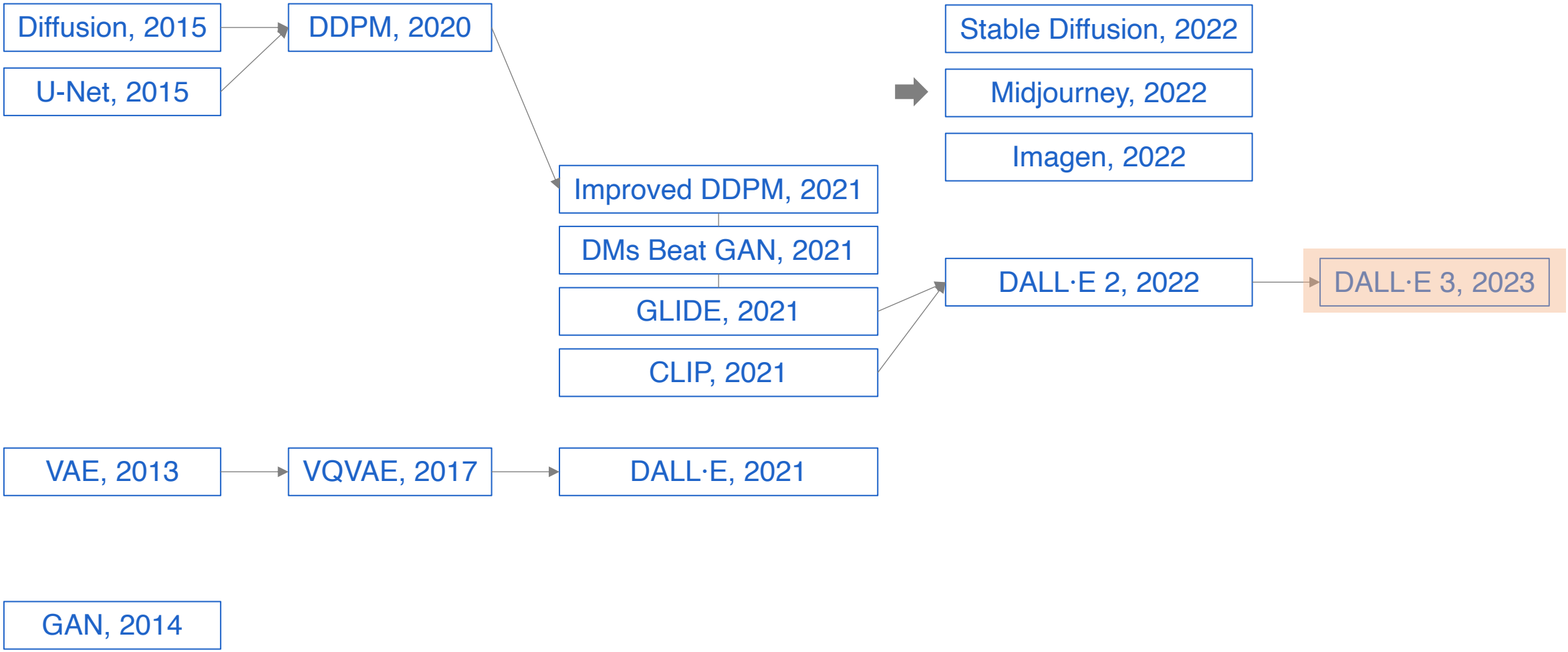
Stable Diffusion

- Idea: apply diffusion process in latent space





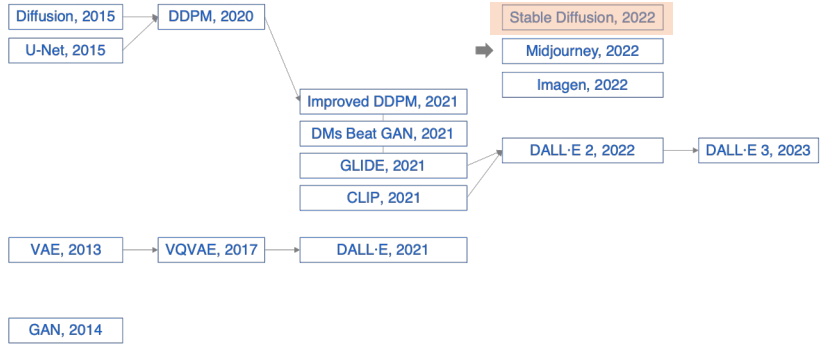
Diffusion model Improvements



Outline

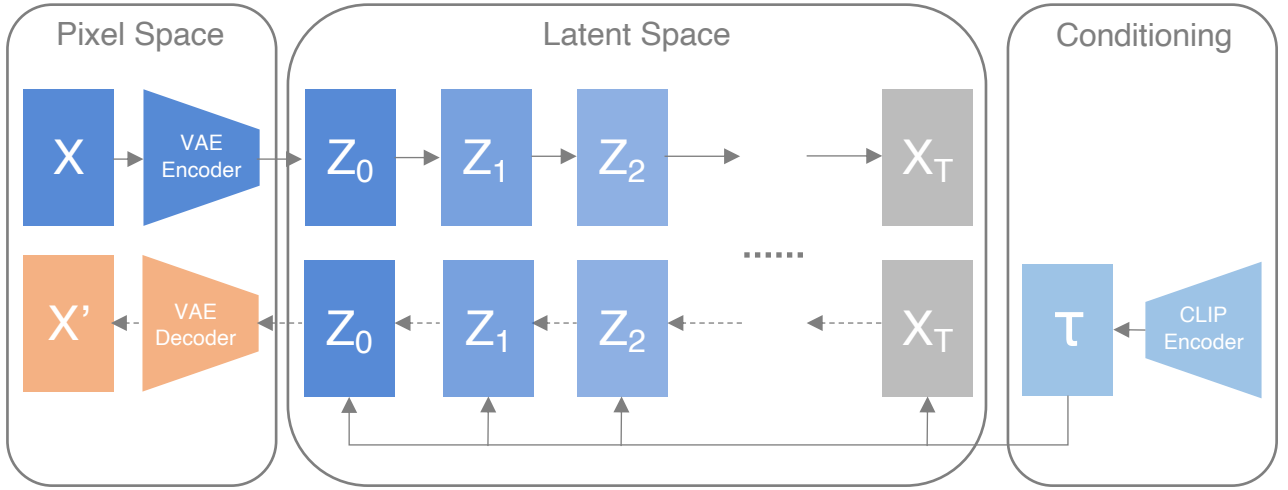
- Introduction
- A Brief History
- Stable Diffusion Walk Through
- Demo: Train Your Own LoRA Model
- References

Stable Diffusion Walk Through



- Idea: apply diffusion process in latent space

- Sub modules
  - Encoder
  - CLIP
  - Scheduler
  - U-Net

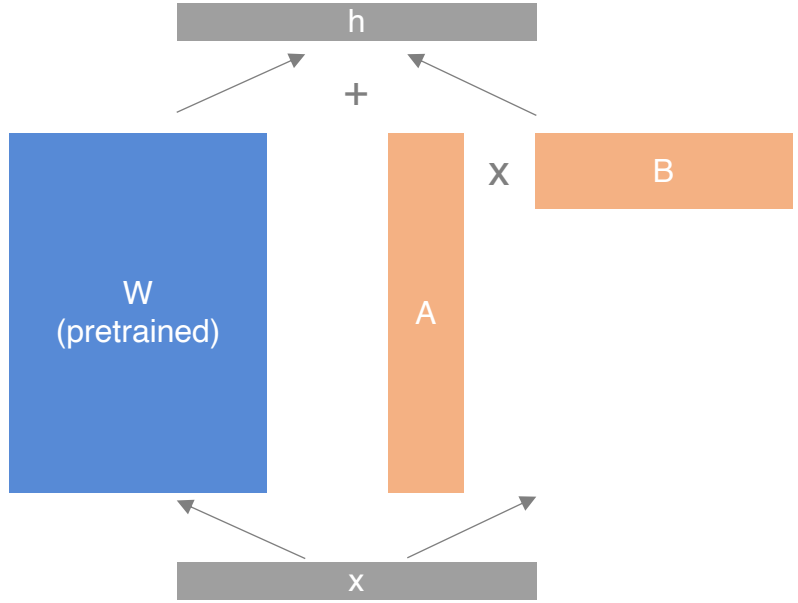


Outline

- Introduction
- A Brief History
- Stable Diffusion Walk Through
- Demo: Train Your Own LoRA Model
- References

Demo: LoRA Training

- Idea: finetune stable diffusion model by adapt the model weights using extra low rank parameter matrices
- Tools: Huggingface, Kohya_SS
- Training data preparation
- Train
- Usage



Outline

- Introduction
- A Brief History
- Stable Diffusion Walk Through
- Demo: Train Your Own LoRA Model
- References

References

1. https://openai.com/dall-e-3
2. https://www.midjourney.com/home
3. https://huggingface.co/docs/diffusers/index
4. https://github.com/AUTOMATIC1111/stable-diffusion-webui
5. https://github.com/comfyanonymous/ComfyUI
6. https://github.com/bmaltais/kohya_ss
7. [VAE] Kingma & Willing, Auto-Encoding Variational Bayes, 2013
8. [GAN] Goodfellow et al., Generative Adversarial Networks, 2014
9. [Diffusion Model] Sohl-Dickstein et al., Deep Unsupervised Learning using Nonequilibrium Thermodynamics, 2015
10. [U-Net] Ronneberger et al., U-Net: Convolutional Networks for Biomedical Image Segmentation, 2015
11. [VQVAE] Van den Oord et al., Neural Discrete Representation Learning, 2017
12. [DDPM] Ho et al., Denoising Diffusion Probabilistic Models, 2020
13. [Improved DDPM] Nichol & Dhariwal, Improved Denoising Diffusion Probabilistic Models, 2021
14. [DMs Beat GAN] Dhariwal & Nichol, Diffusion Models Beat GANs on Image Synthesis, 2021]
15. [GLIDE] Nichol et al., GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models, 2021
16. [CLIP] Radford et al., Learning Transferable Visual Models From Natural Language Supervision, 2021
17. [DALL·E] Ramesh et al., Zero-Shot Text-to-Image Generation, 2021
18. [DALL·E 2] Ramesh et al., Hierarchical Text-Conditional Image Generation with CLIP Latents, 2022
19. [Stable Diffusion] Rombach et al., High-Resolution Image Synthesis with Latent Diffusion Models, 2022
20. [Imagen] Saharia et al., Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding, 2022
21. [DALL·E 3] Betker et al., Improving Image Generation with Better Captions, 2023
22. [LoRA] Hu et al., LoRA: Low-Rank Adaptation of Large Language Models, 2021