

How does Stable Diffusion work?

Michael Pieler

Stable Diffusion

- How does it work?
- Applications

What is Stable Diffusion?

<https://www.reddit.com/r/StableDiffusion/>

What is Stable Diffusion?



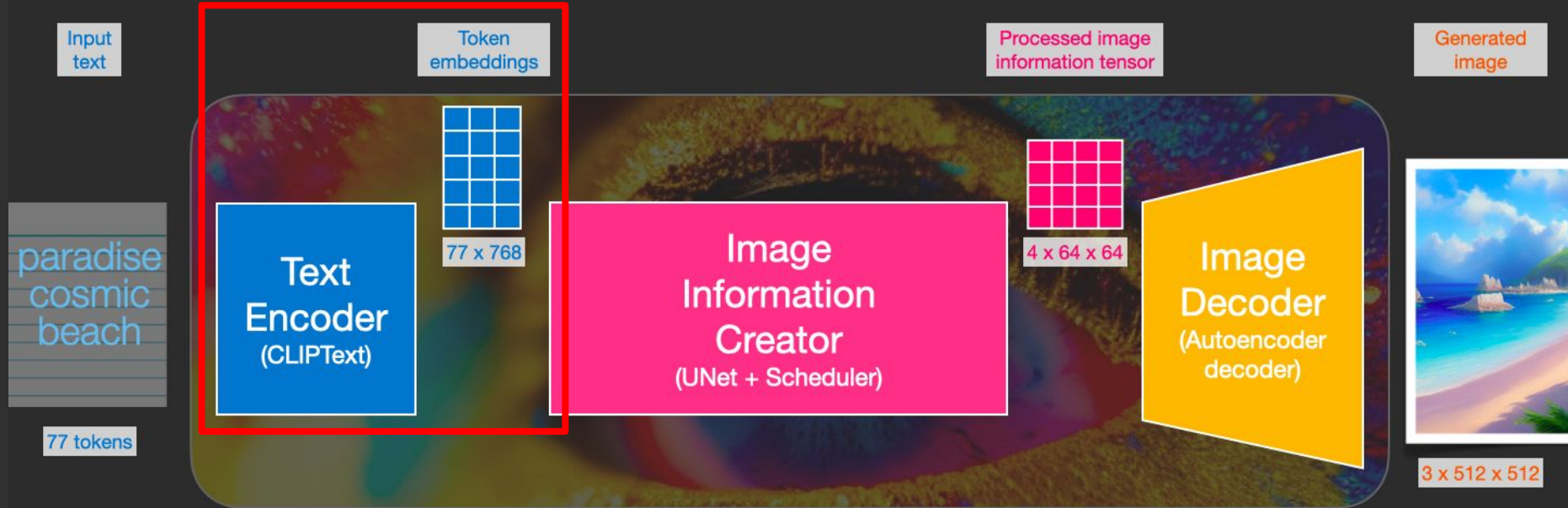
- Stable Diffusion is a latent text-to-image diffusion model.
- Trained on 512x512 images from a subset of the [LAION-5B](#) dataset.
- Compute donation from [Stability AI](#) and support from [LAION](#).

This slide deck is heavily based on [The Illustrated Stable Diffusion by Jay Alammar](#) so please check out his blog post! :-)

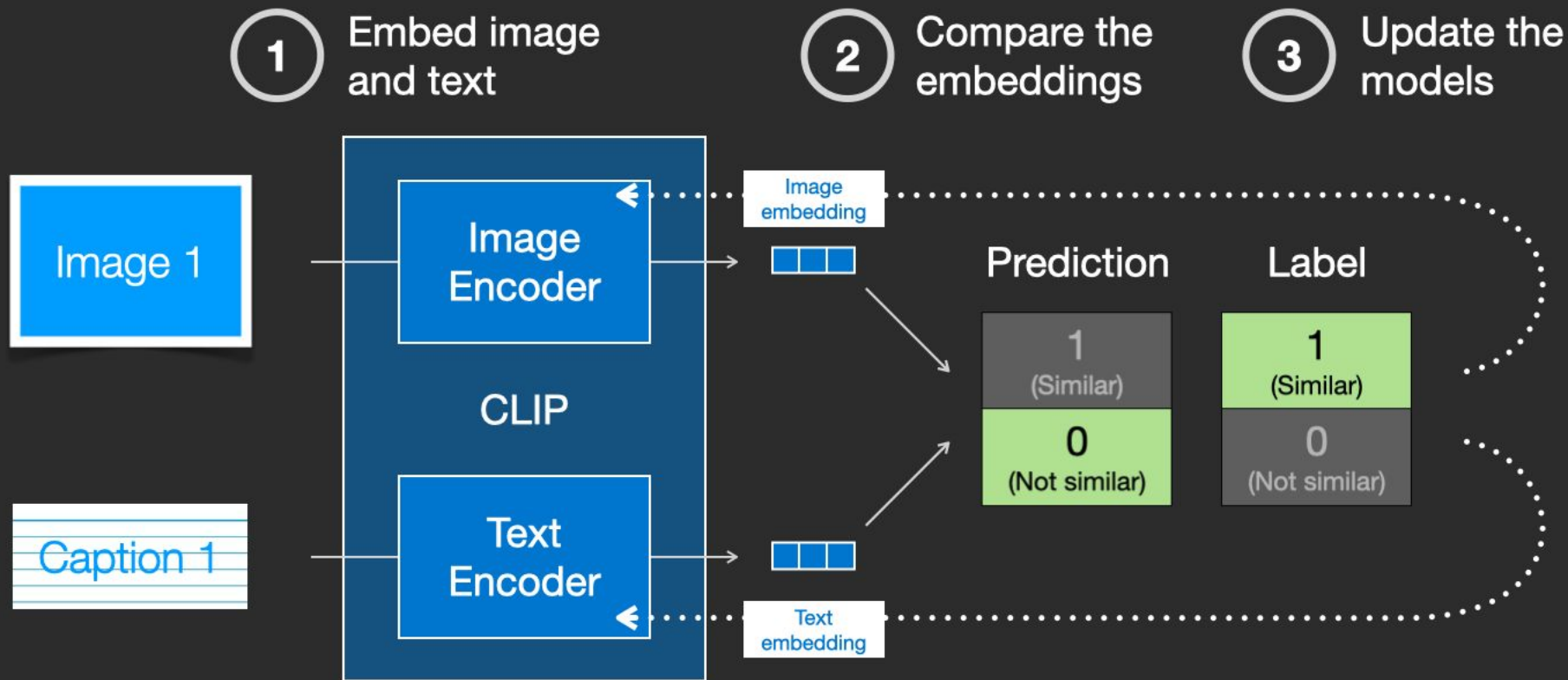
Components

- Text encoder (CLIP)
- Diffusion model (UNet)
- Autoencoder (VAE)

The Components of Stable Diffusion



The text encoder from CLIP



Components

- Text encoder (CLIP)
- Diffusion model (UNet)
- Autoencoder (VAE)

Stable Diffusion

paradise
cosmic
beach

77 tokens

**Text
Encoder**
(CLIPText)



Token
embeddings

1

2



Random image
information tensor

Image Information Creator
(UNet + Scheduler)

Diffusion

3



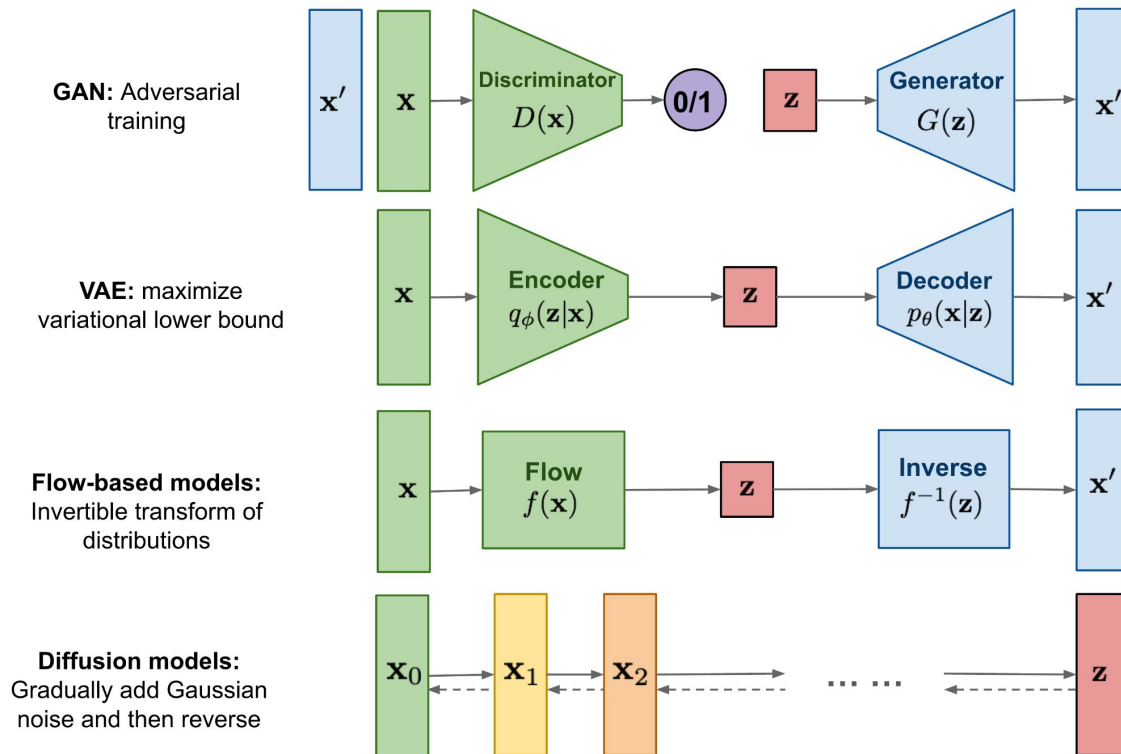
Processed image
information tensor

**Image
Decoder**
(Autoencoder
decoder)

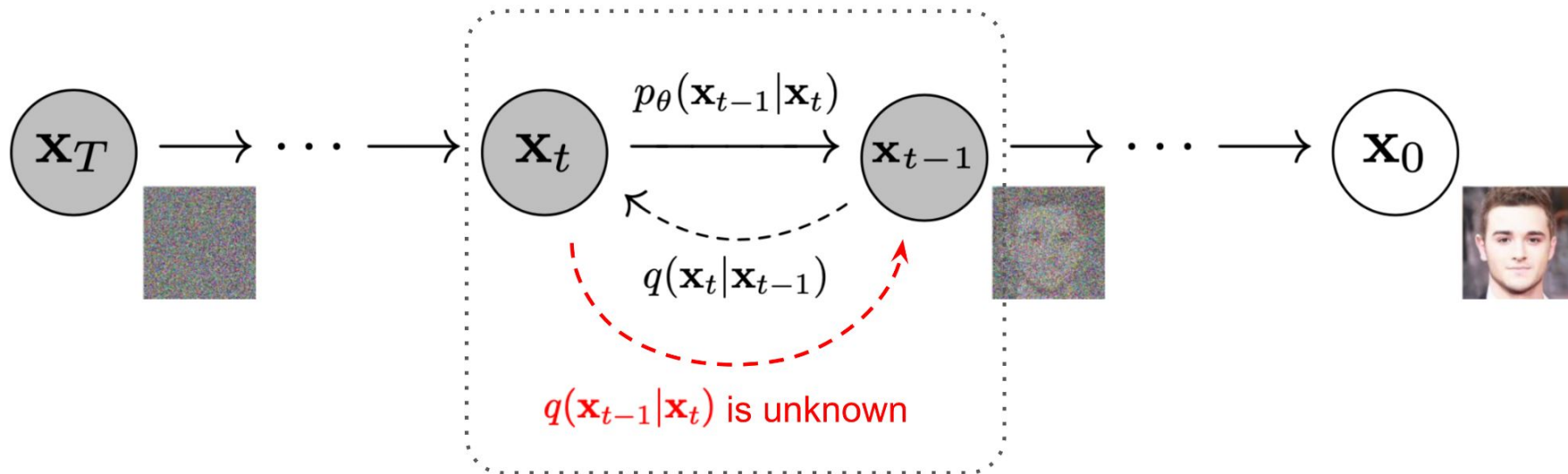
Generated
image



What is Diffusion?



What is Diffusion?

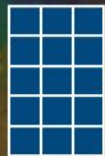


Stable Diffusion

paradise
cosmic
beach

77 tokens

**Text
Encoder**
(CLIPText)



Token
embeddings

Image Information Creator
(UNet + Scheduler)



**Random image
information tensor**

UNet
Step
1



UNet
Step
2



...

UNet
Step
50



**Processed image
information tensor**

**Image
Decoder**
(Autoencoder
decoder)

Diffusion

**Generated
image**



Diffusion



Image Information Creator

Image
Decoder
(Autoencoder
decoder)



How diffusion works

Step 1



+

Step 2



noise
slice 1

=



+

Step 3



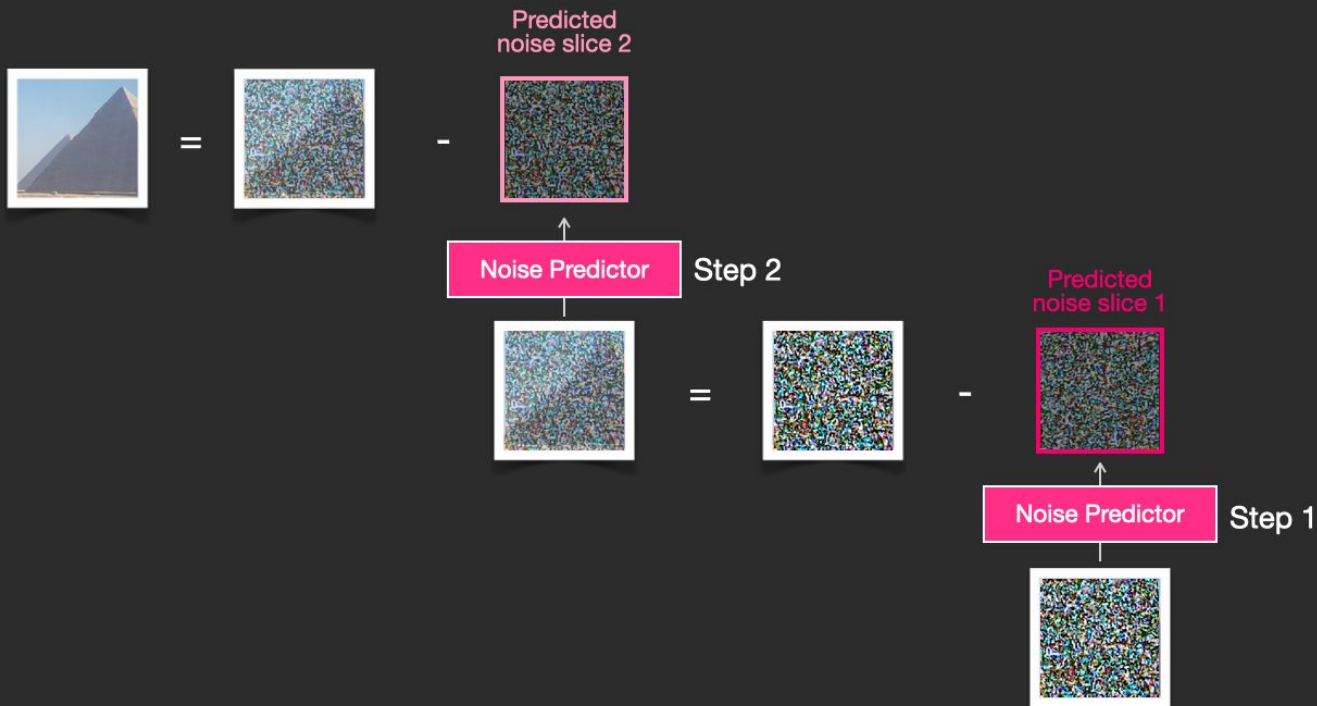
noise
slice 2

=



Painting images by removing noise

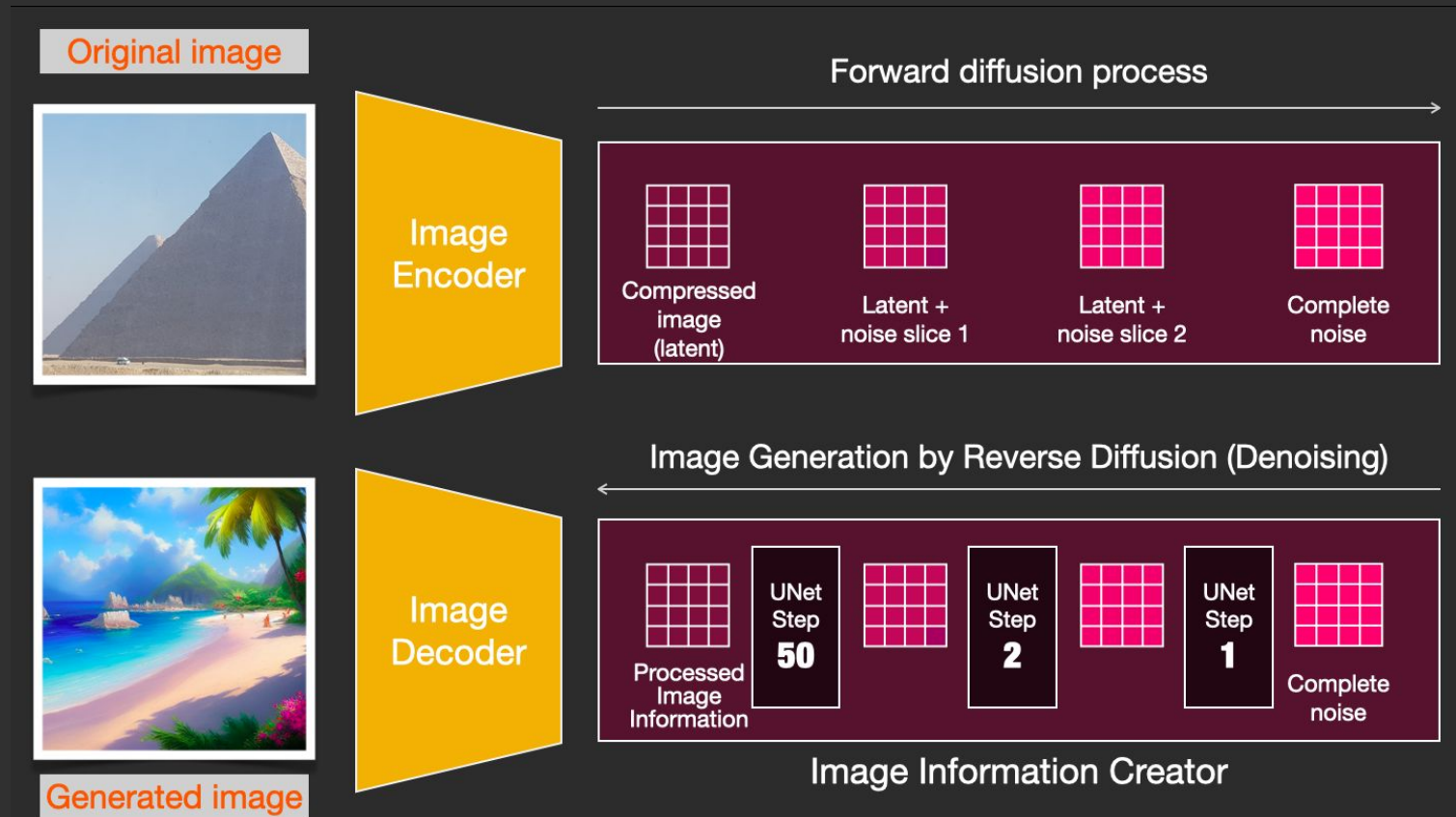
Image Generation by Reverse Diffusion (Denoising)



Components

- Text encoder (CLIP)
- Diffusion model (UNet)
- Autoencoder (VAE)

Diffusion on compressed (latent) data instead of the pixel images



Stable Diffusion

paradise
cosmic
beach

77 tokens

Text
Encoder
(CLIPText)



1



2



Random image
information tensor

Image Information Creator
(UNet + Scheduler)

Diffusion

3



Processed image
information tensor

Image
Decoder
(Autoencoder
decoder)

Generated
image



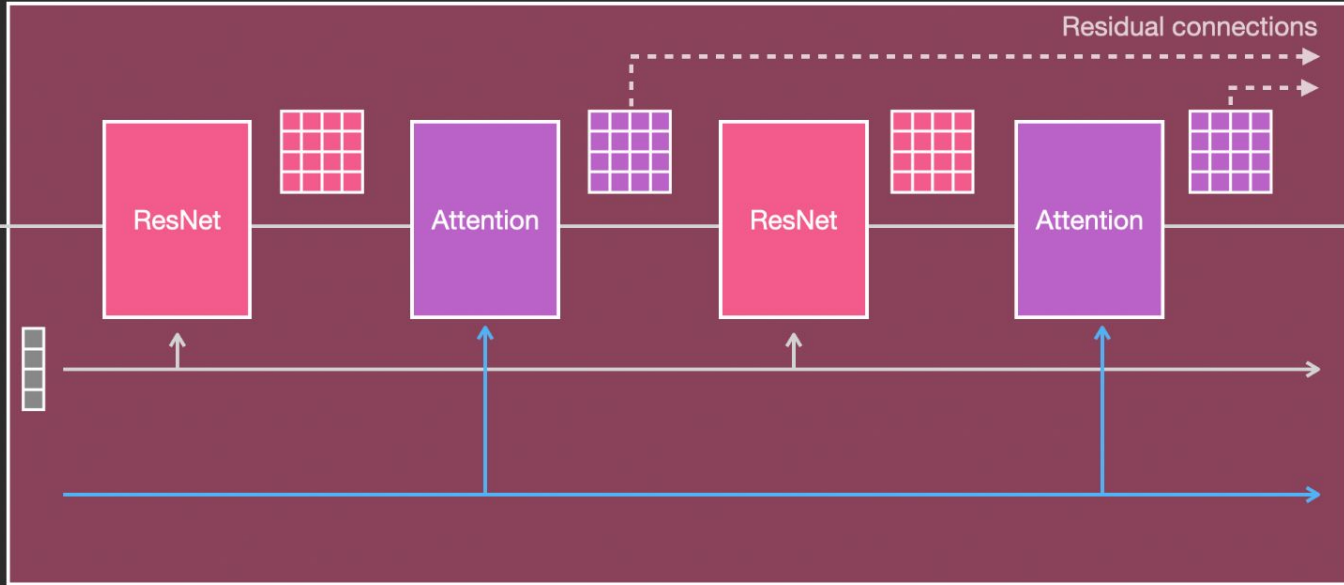
Feeding text information into the image generation process

Noise Predictor with Text Conditioning (UNet with attention)

Noisy image
information
(latents)



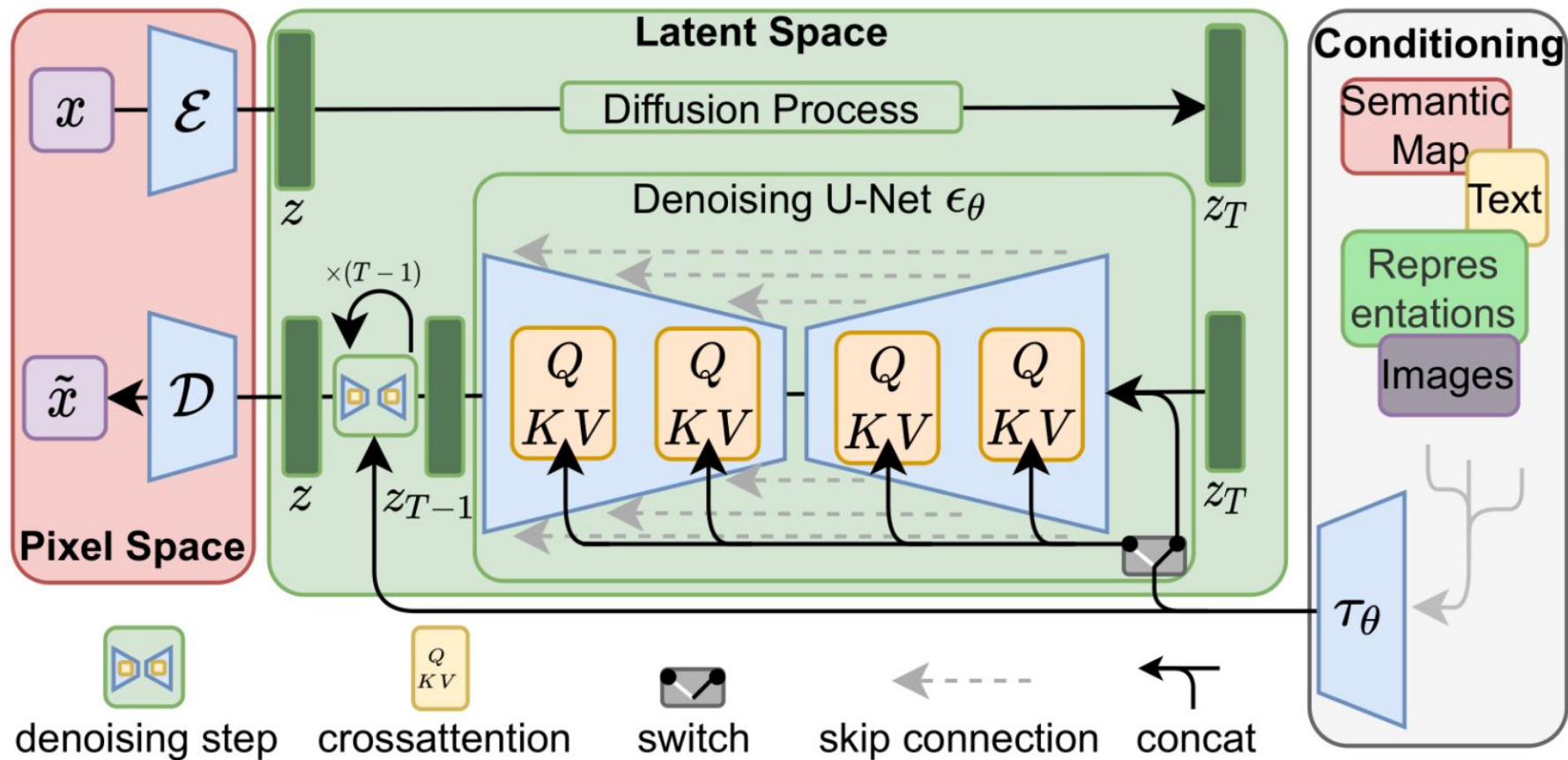
Step 3



Predicted
noise slice



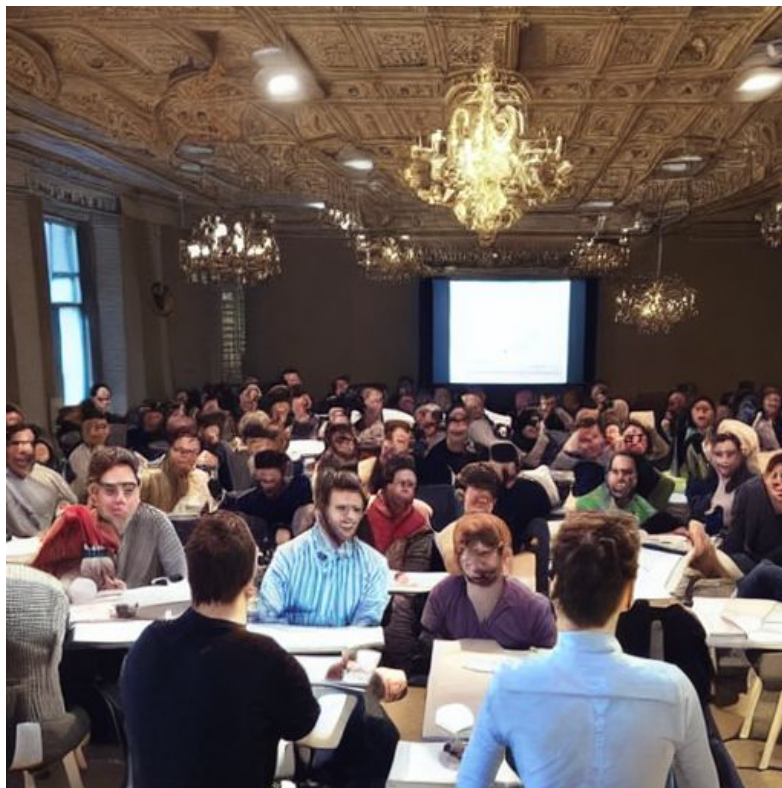
Text information
(token embeddings)



Stable Diffusion

- How does it work?
- Applications

Text to image



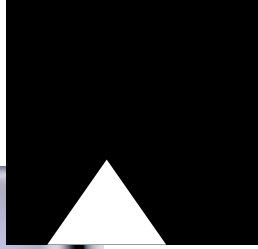
Prompt: "Vienna deep learning meetup in October"

Image to image



Inpainting

+ Mask



+ Prompt: "Cinematic movie still of
Elon Musik in 2001 space Odyssey"

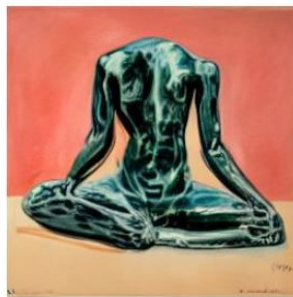
Fine tune Stable Diffusion



Textual inversion



Input samples $\xrightarrow{\text{invert}}$ “ S_* ”



“An oil painting of S_* ”



“App icon of S_* ”



“Elmo sitting in the same pose as S_* ”



“Crochet S_* ”



Input samples $\xrightarrow{\text{invert}}$ “ S_* ”



“Painting of two S_* fishing on a boat”



“A S_* backpack”

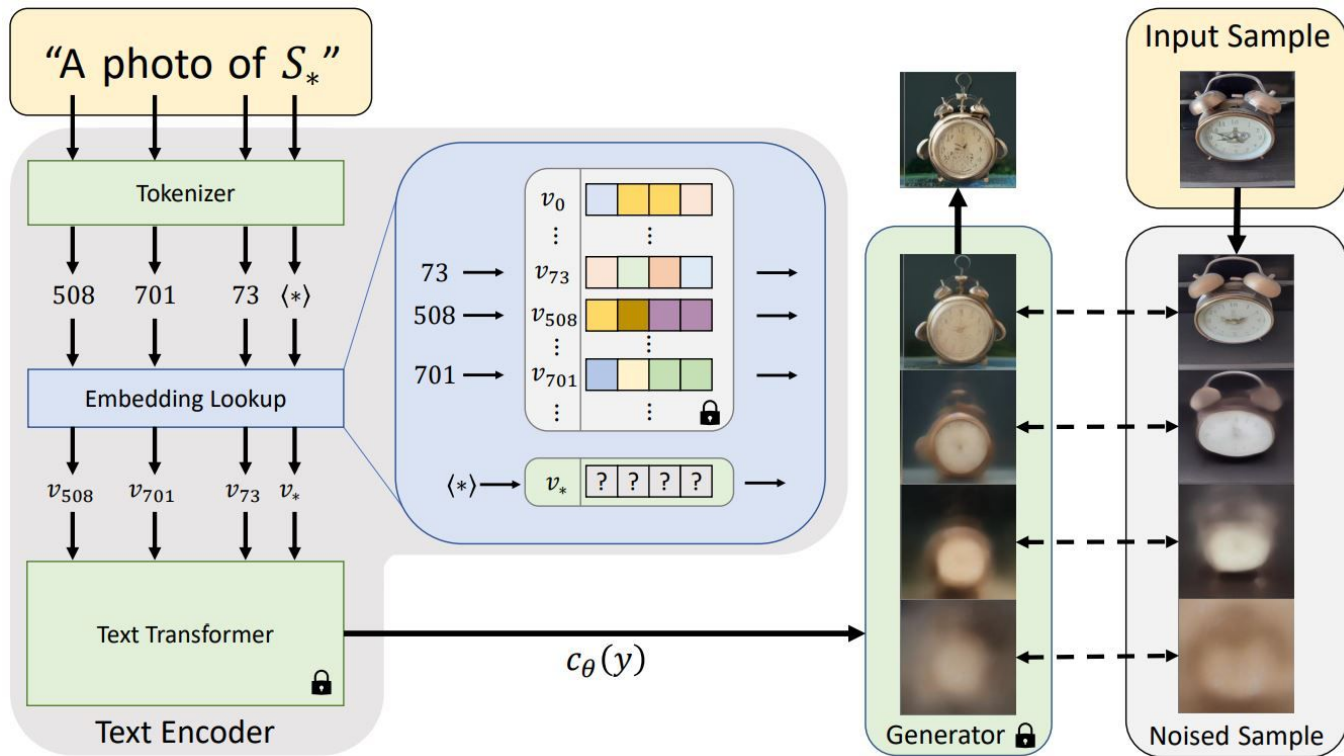


“Banksy art of S_* ”



“A S_* themed lunchbox”

Textual inversion setup



Dreambooth



Input images



in the Acropolis



swimming



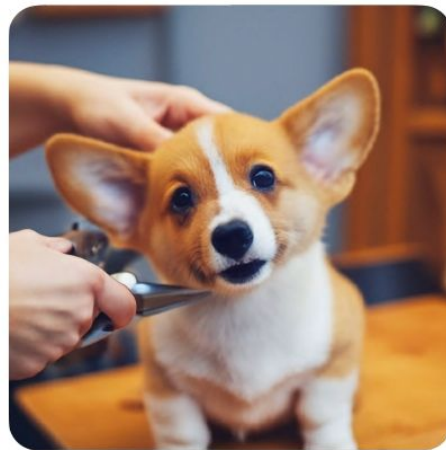
sleeping



in a doghouse

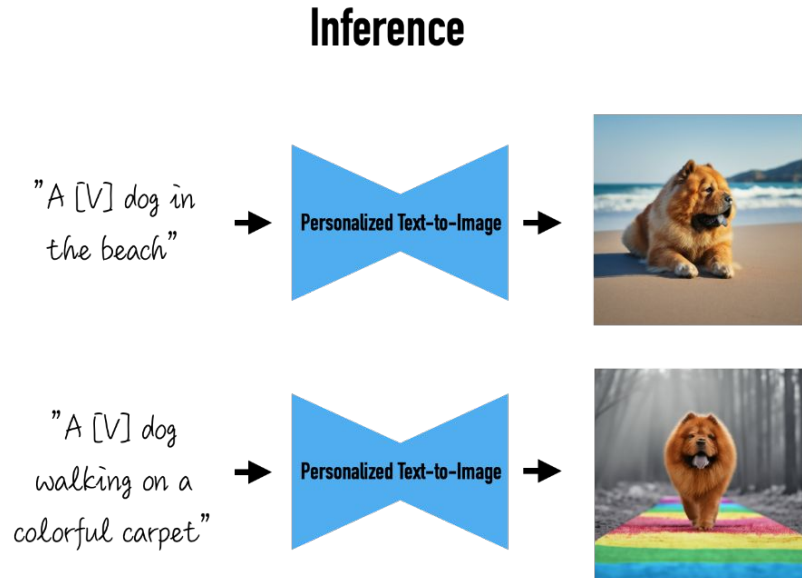
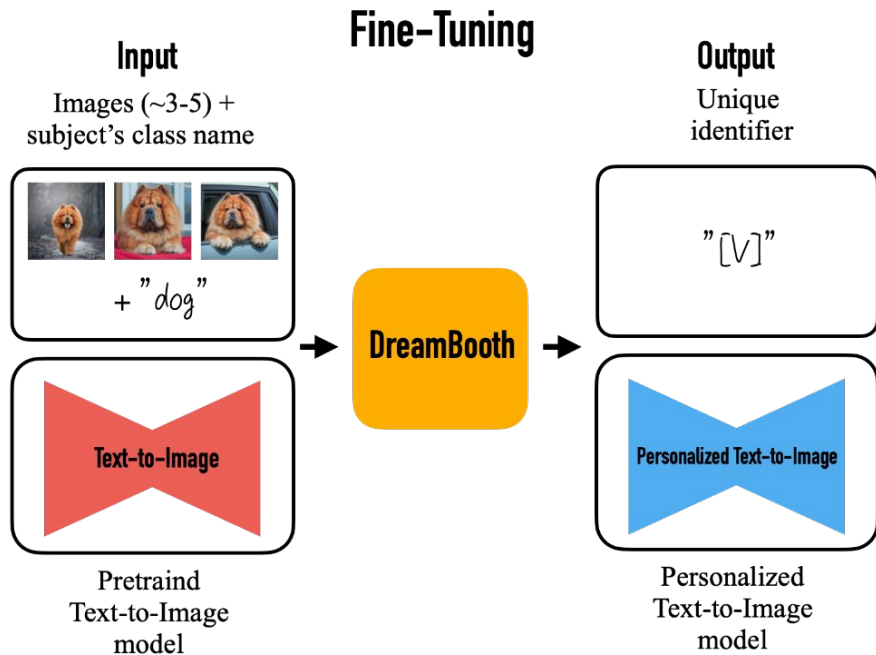


in a bucket

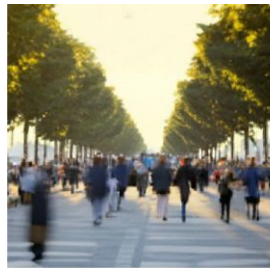


getting a haircut

Dreambooth setup



Prompt-to-prompt image editing



“The boulevards are crowded today.”



“My fluffy bunny doll.”



“Photo of a cat riding on a bicycle.”

~~car~~



“a cake with decorations.”

jelly beans



“Landscape with a house near a river

and a rainbow in the background?”



“Children drawing of a castle next to a river.”

Thank you for your attention!

Additional resources

- [Stable Diffusion Public Release - Stability.AI](#)
- [The Annotated Diffusion Model](#)
- [Stable Diffusion with !\[\]\(95b42f0077faf7439a26242a54e021ec_img.jpg\) Diffusers, Diffusers notebooks](#)
- [How diffusion models work: the math from scratch | AI Summer](#)
- [What are Diffusion Models? | Lil'Log](#)
- [Generative Modeling by Estimating Gradients of the Data Distribution | Yang Song](#)
- [Diffusion Models: A Comprehensive Survey of Methods and Applications](#)
- [How DALL-E 2, Imagen and Parti Architectures Differ](#)
- <https://github.com/CompVis/stable-diffusion>
- <https://github.com/lucidrains/denoising-diffusion-pytorch>
- New fastai part 2 course 2022 will be released soon and has a focus on Stable Diffusion.