

Latent Stable Diffusion (LDM)

Latent Stable Diffusion

Es un modelo de deep learning para pasar texto a imagen creado en 2022. Fue creado por Stability AI, CompVis LMU y Runway.

Aplicaciones:

- Convertir texto en imagen (es el más usado)
- Inpainting
- Outpainting
- Imagen a imagen

Texto en imagen

```
▶ prompt = "astronaut riding a horse"  
image = pipe(prompt).images[0]  
image.save(f"astronaut_rides_horse.png")  
image
```

100%  50/50 [00:09<00:00, 5.32it/s]



Input: texto
Output: imagen

```
prompt = "people eating pizza"  
image = pipe(prompt).images[0]  
image
```

100%  50/50 [00:10<00:00, 4.97it/s]



Inpainting

Input: imagen

Output: imagen

Original Image



Inpainted Image



Outpainting



Input: imagen
Output: imagen



Imagen a imagen

Input: imagen
Output: imagen

INPUT

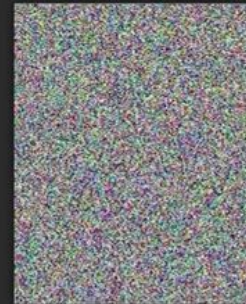


OUTPUT



Idea general de stable diffusion

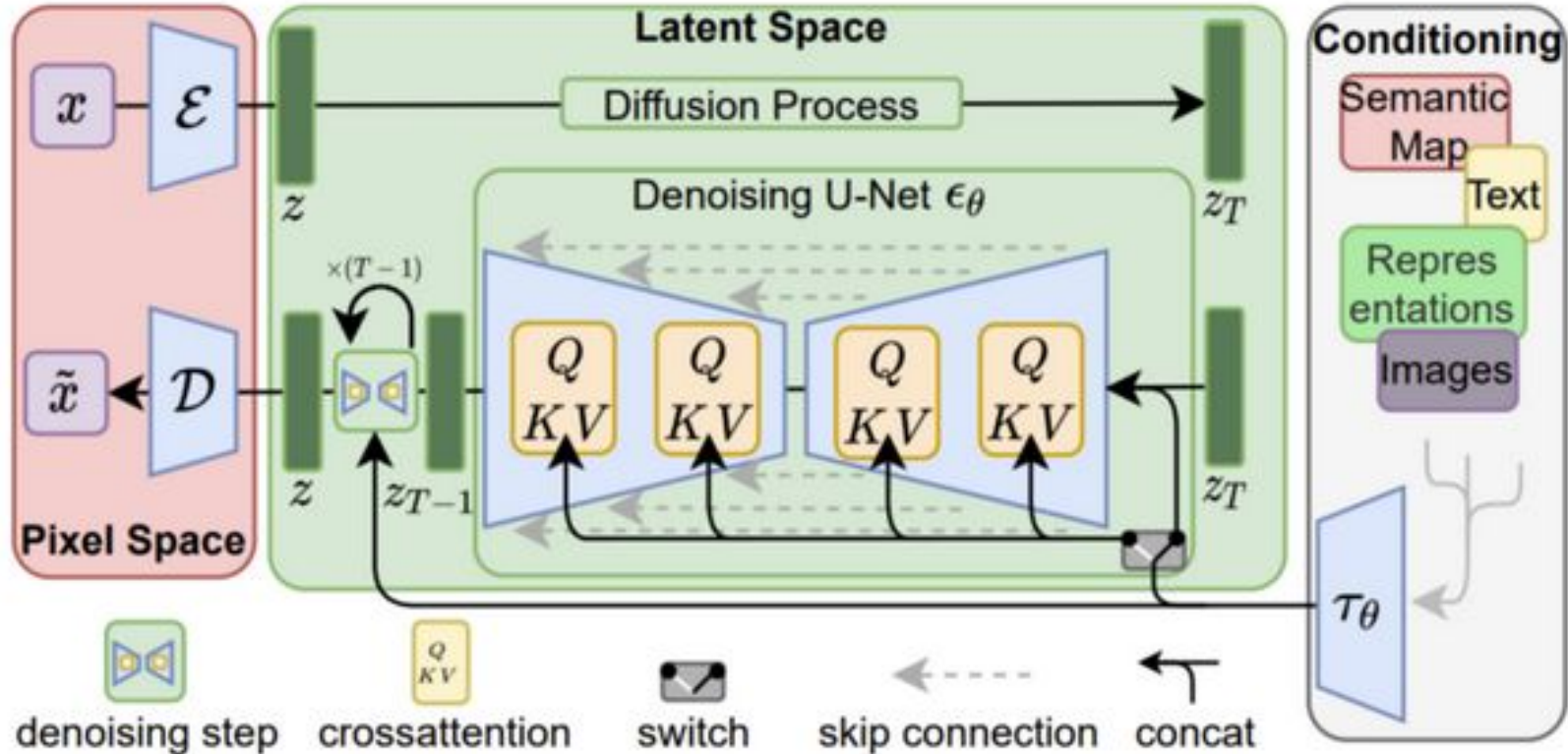
FORWARD DIFFUSION PROCESS



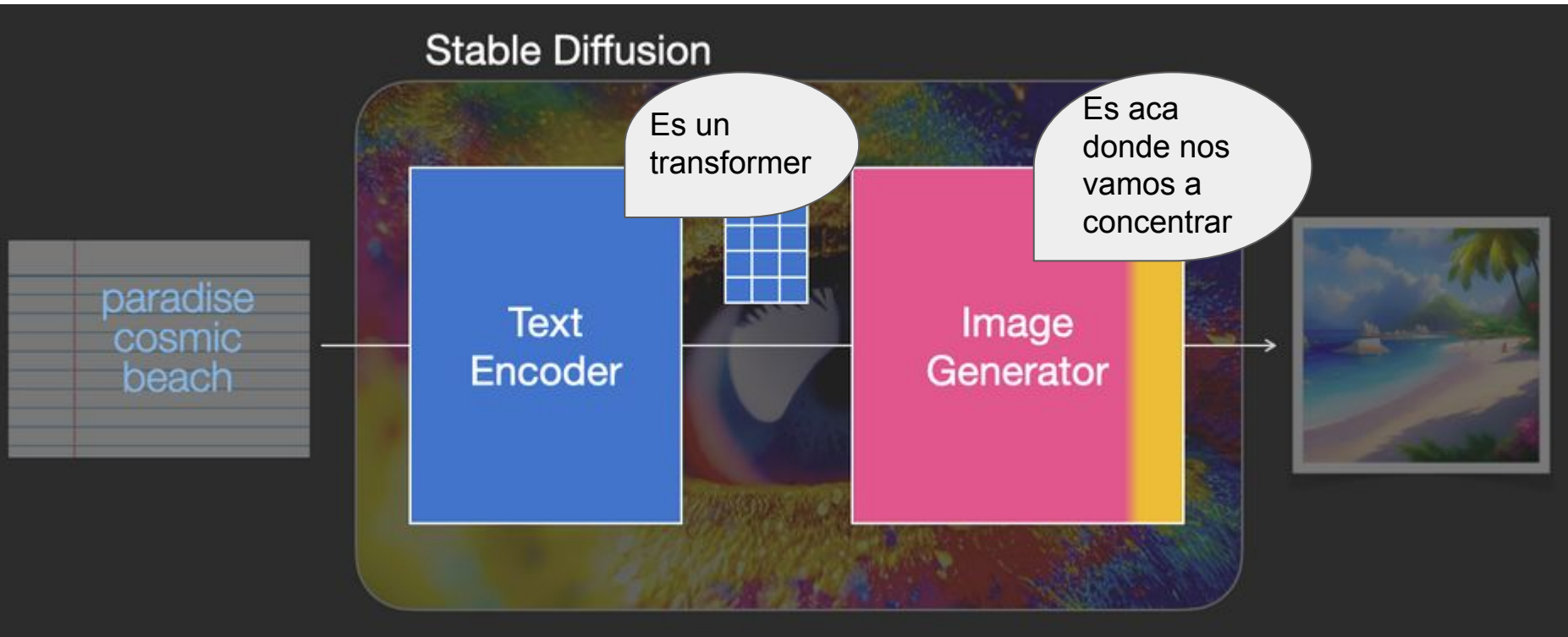
Se puede dividir en 3 partes:

1. Autoencoder (VAE) (parte roja)
2. U-Net (Verde abajo, denoising U-net)
3. Text-encoder, e.g. CLIP's Text Encoder. (conditioning if text, derecha)

Modelo



Vamos por partes como dice jack:



El text encoder toma el texto como input y devuelve una lista de números por palabra/token

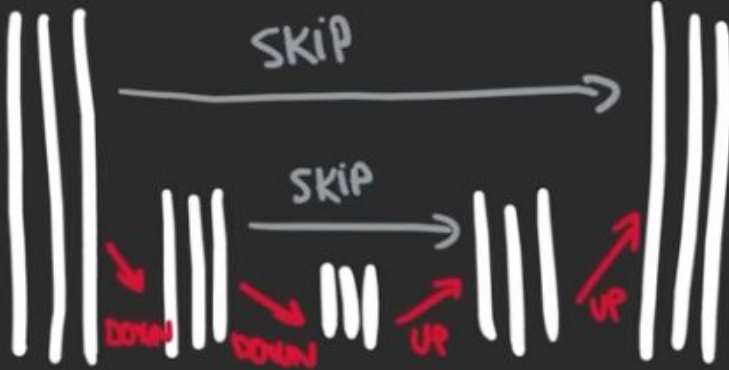
The U-Net part

Objetivo de la U-Net es detectar el ruido. Output ruido de la imagen



Neural Net (UNet)

t
IN

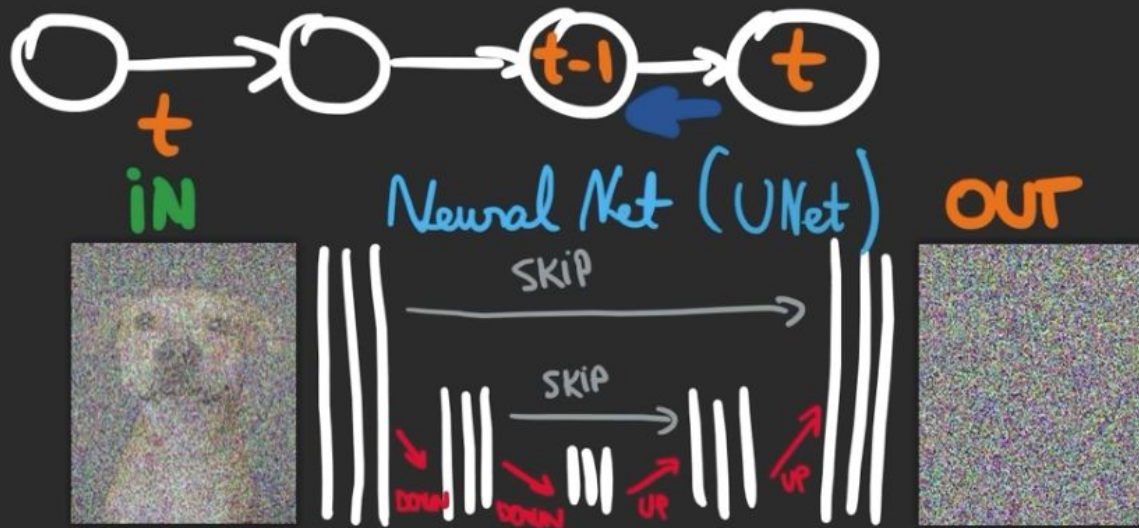


OUT



$t-1$

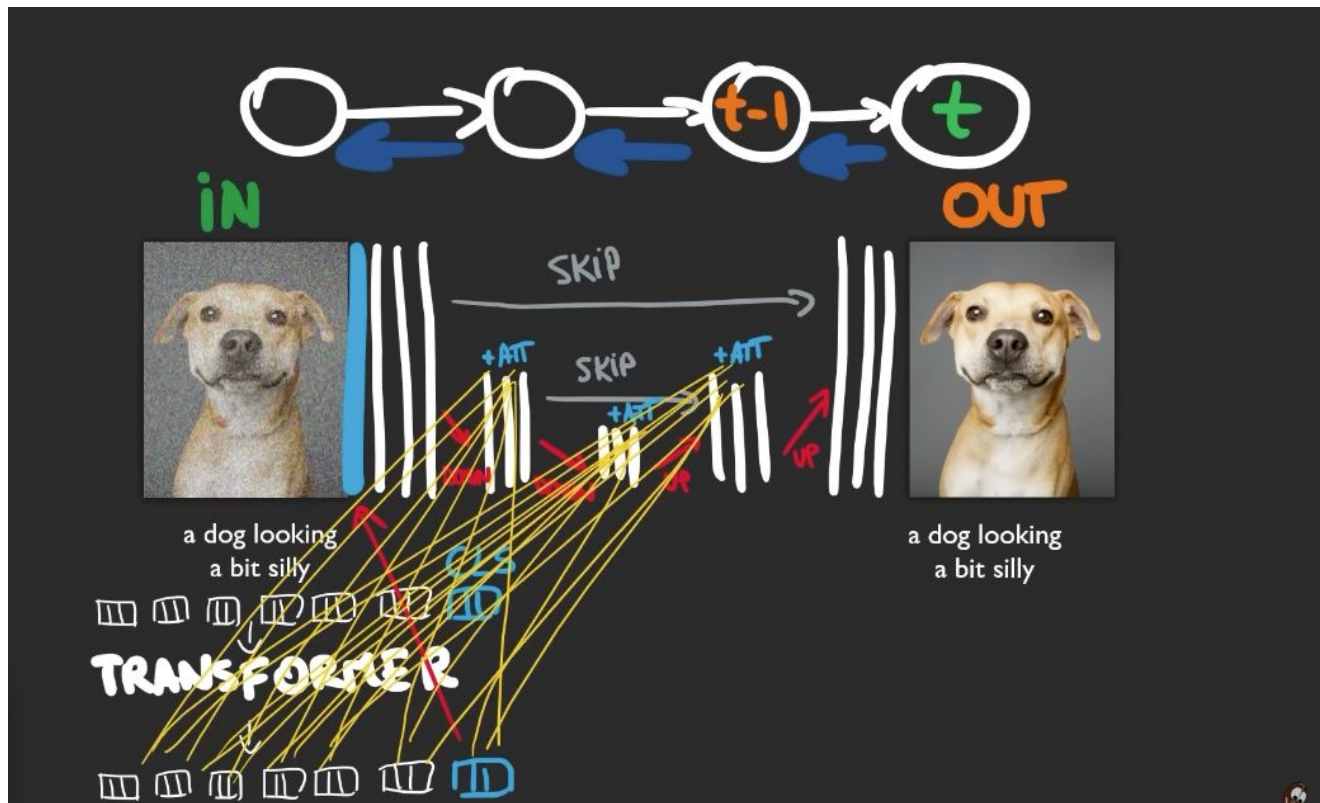




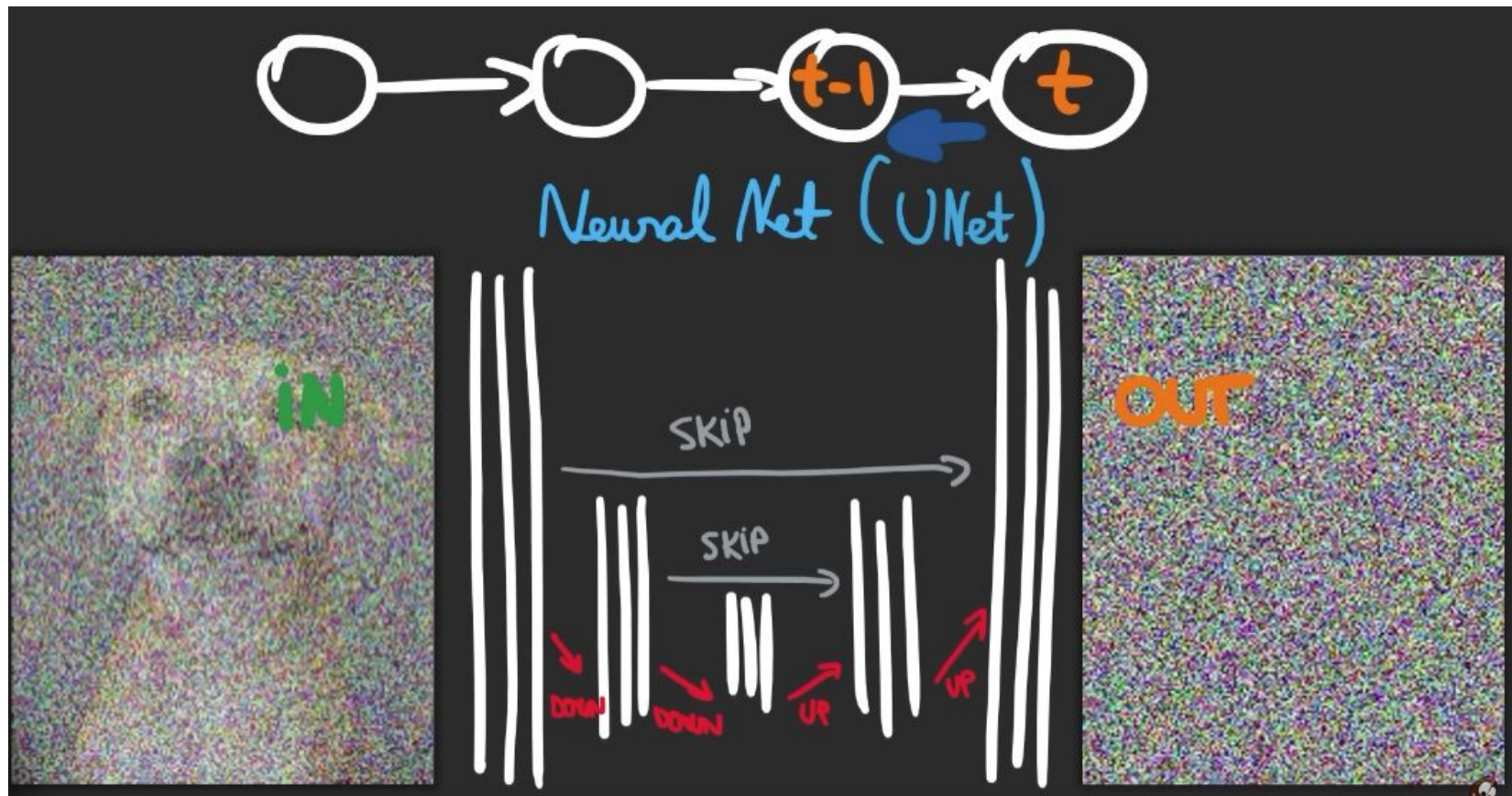
$$t \text{ (noisy dog image)} - \varepsilon \cdot \text{(noisy image)} = t-1 \text{ (clear dog image)}$$

En realidad no se sustrae todo el ruido sino el ruido por epsilon (entre 0 - 1)

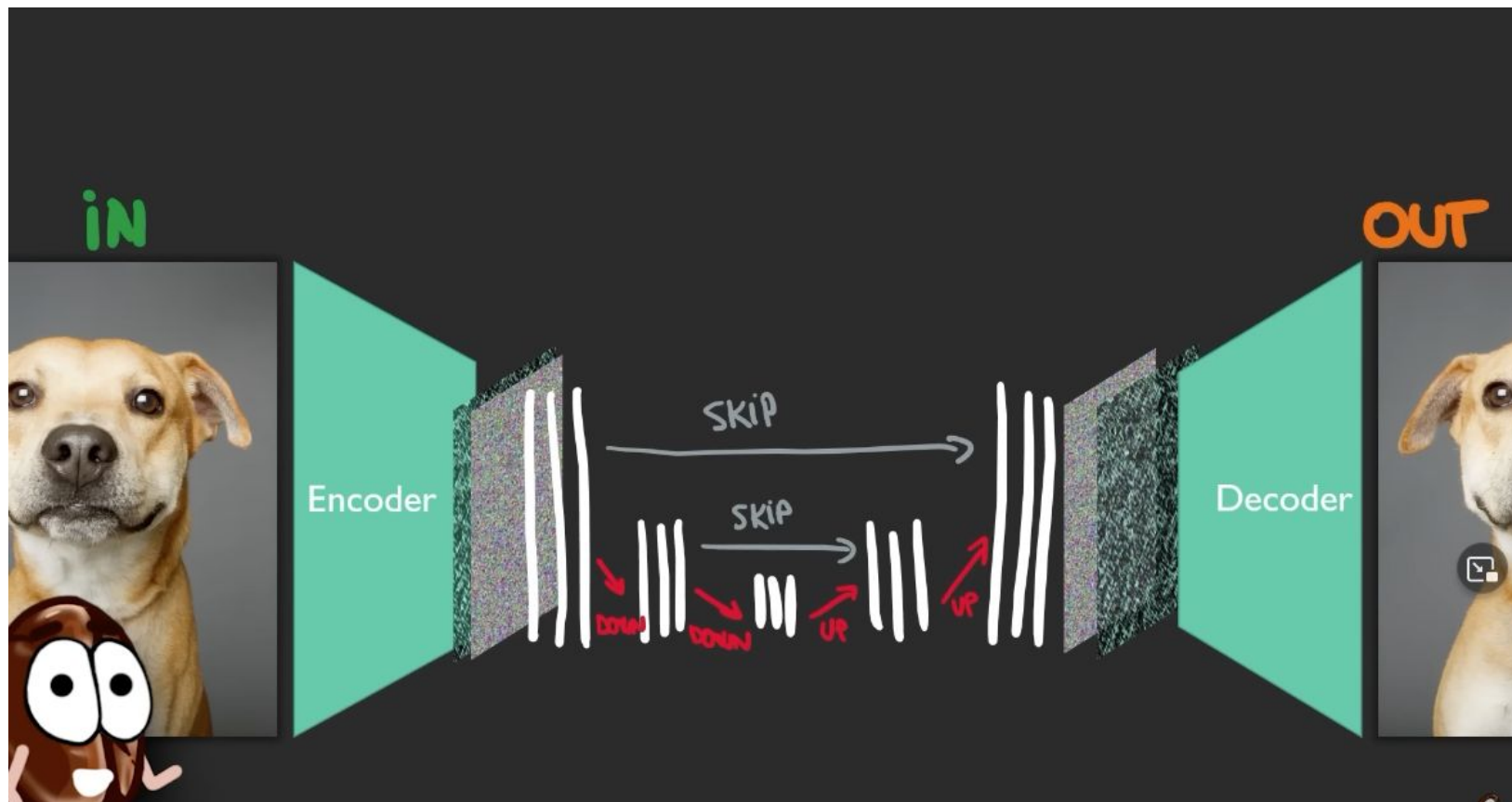
Como se agrega el texto:



Problema: Muy costoso para imágenes grandes



Solución: AutoEncoder



Recapitulando:

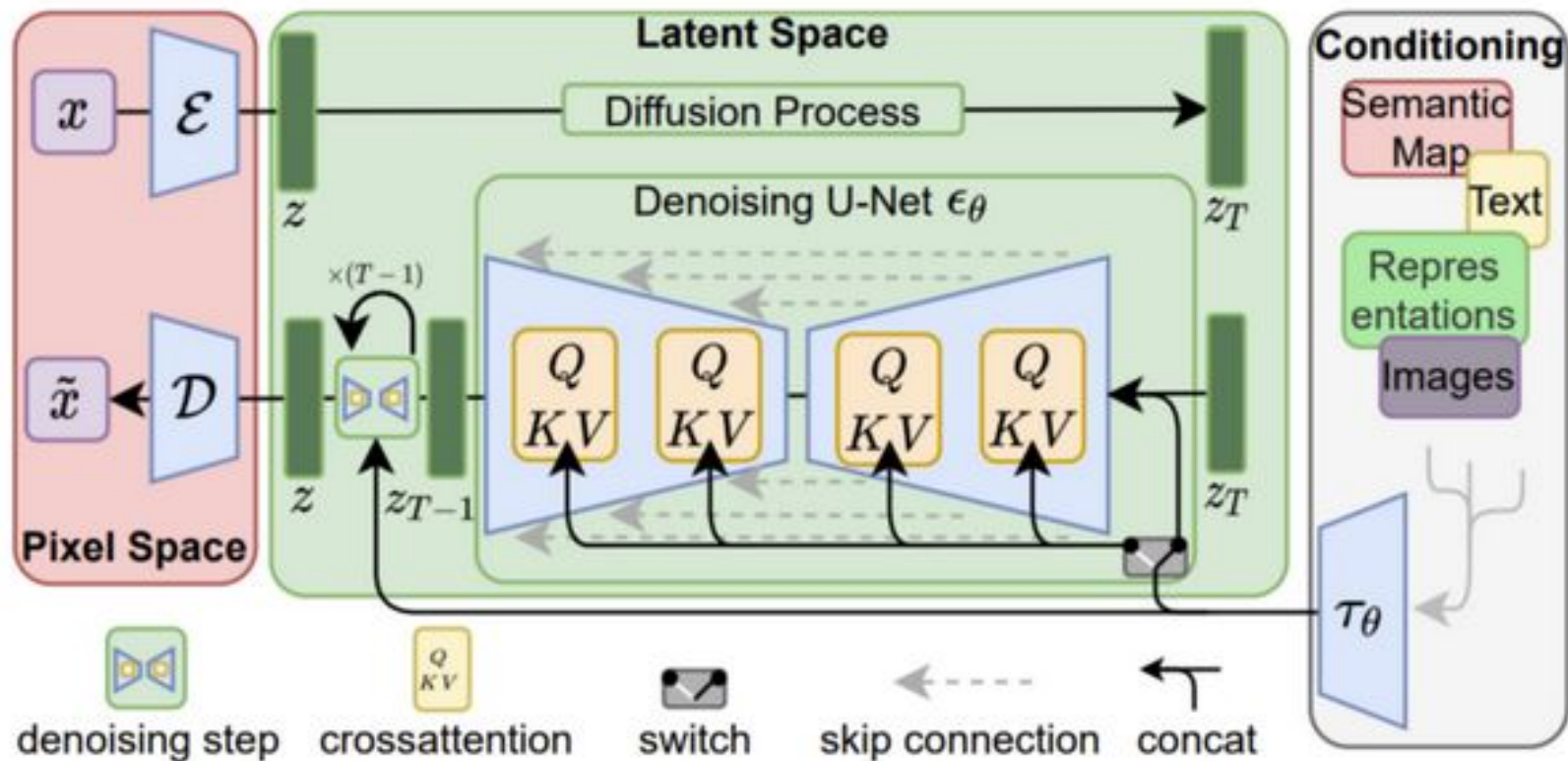


Imagen generada por la U-Net cada 10 iteraciones



Dibujando a Artica!





Resources:

<https://www.youtube.com/watch?v=J87hffSMB60>

https://colab.research.google.com/github/huggingface/notebooks/blob/main/diffusers/stable_diffusion.ipynb#scrollTo=b2i37Salfyce