¿Reemplazará la IA a los artistas? Modelos de difusión y sus implicaciones

Imágenes generadas por IA

UNIVERSIDAD DE LA RIOJA

Índice

1. ¿Qué es?
2. ¿Cómo funciona?
3. Cuestiones éticas/legales
4. ¿Qué más se puede hacer?

Índice

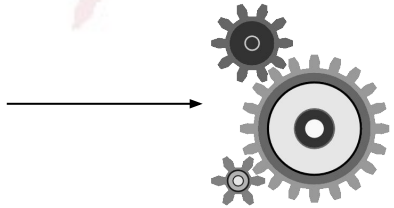
1. **¿Qué es?**
2. ¿Cómo funciona?
3. Cuestiones éticas/legales
4. ¿Qué más se puede hacer?

Generación de imágenes

Un agaporni al estilo de Pixar

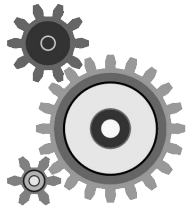
Generación de imágenes

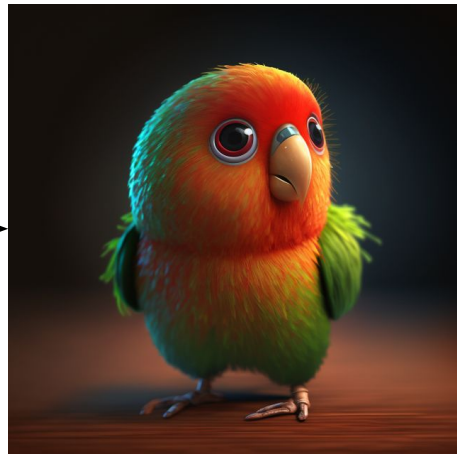
Un agaporni al estilo de Pixar



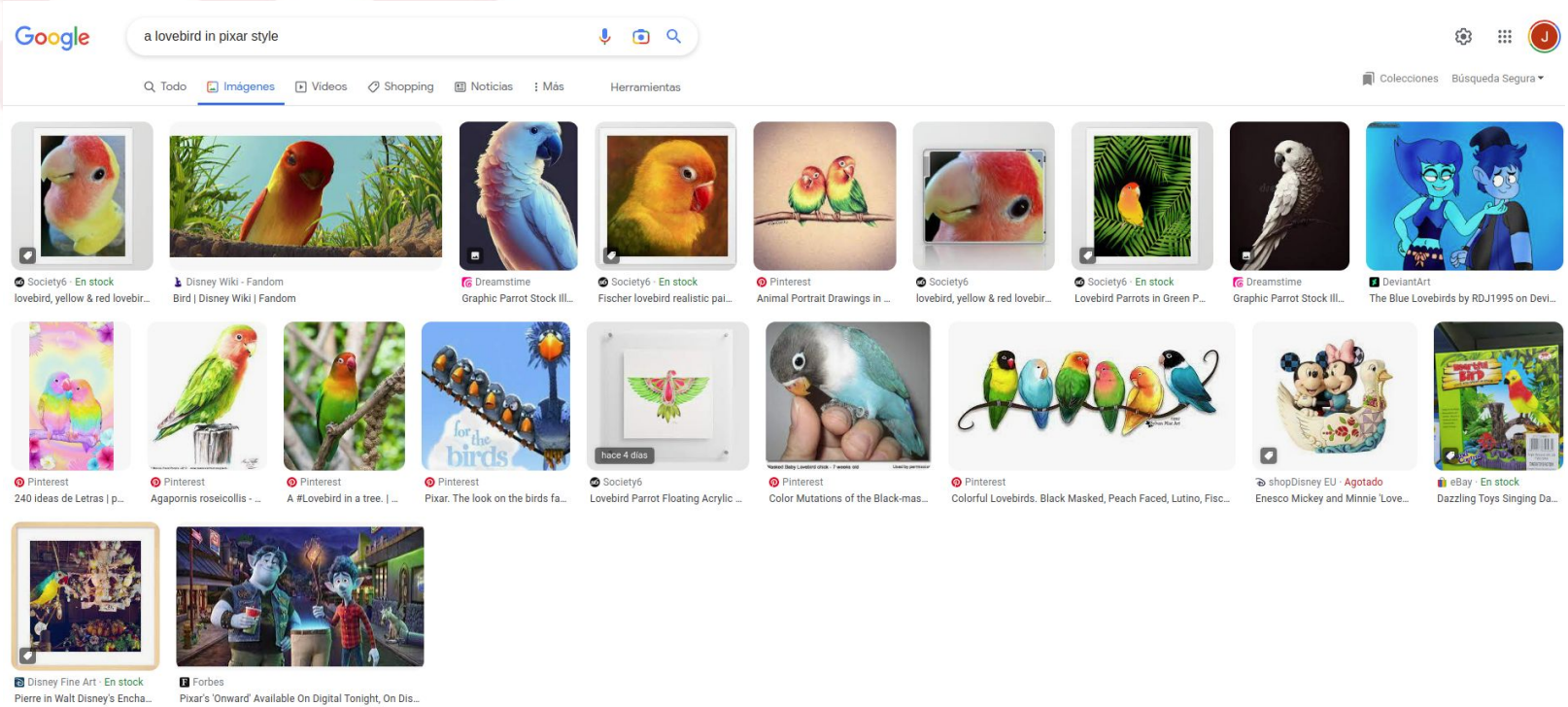
Generación de imágenes

Un agaporni al estilo de Pixar





¿Es una búsqueda?



UNIVERSIDAD DE LA RIOJA

Herramientas

- Dall-E
- Midjourney
- Lexica
- Stable Diffusion
- Karlo
- …

Índice

1. ¿Qué es?
2. **¿Cómo funciona?**
3. Cuestiones éticas/legales
4. ¿Qué más se puede hacer?

Modelos de difusión

High-Resolution Image Synthesis with Latent Diffusion Models

(A.K.A. LDM & Stable Diffusion)

Robin Rombach[1,2], Andreas Blattmann[1,2], Dominik Lorenz[1,2], Patrick Esser[3],
Björn Ommer[1,2]

[1]LMU Munich, [2]IWR, Heidelberg University, [3]Runway

CVPR 2022 (ORAL)

[Submitted on 20 Dec 2021 (v1), last revised 13 Apr 2022 (this version, v2)]

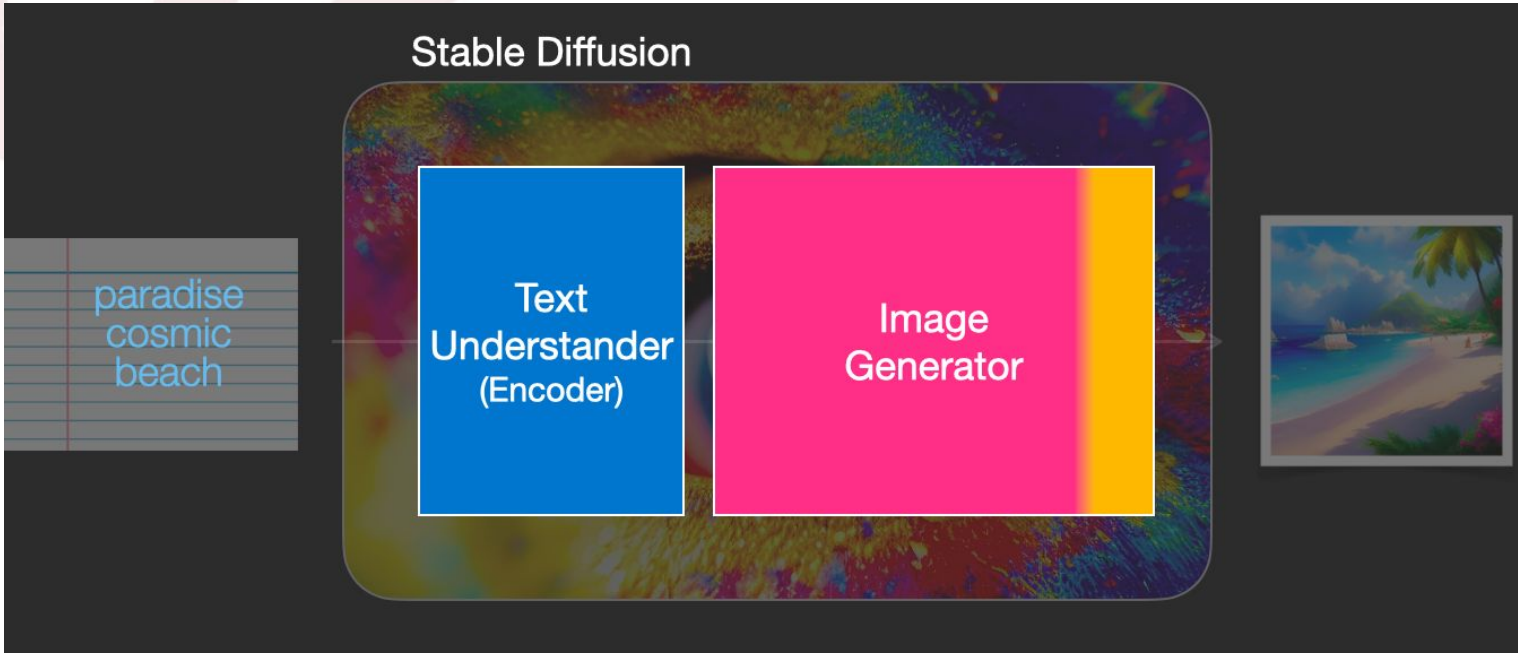
High-Resolution Image Synthesis with Latent Diffusion Models

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer

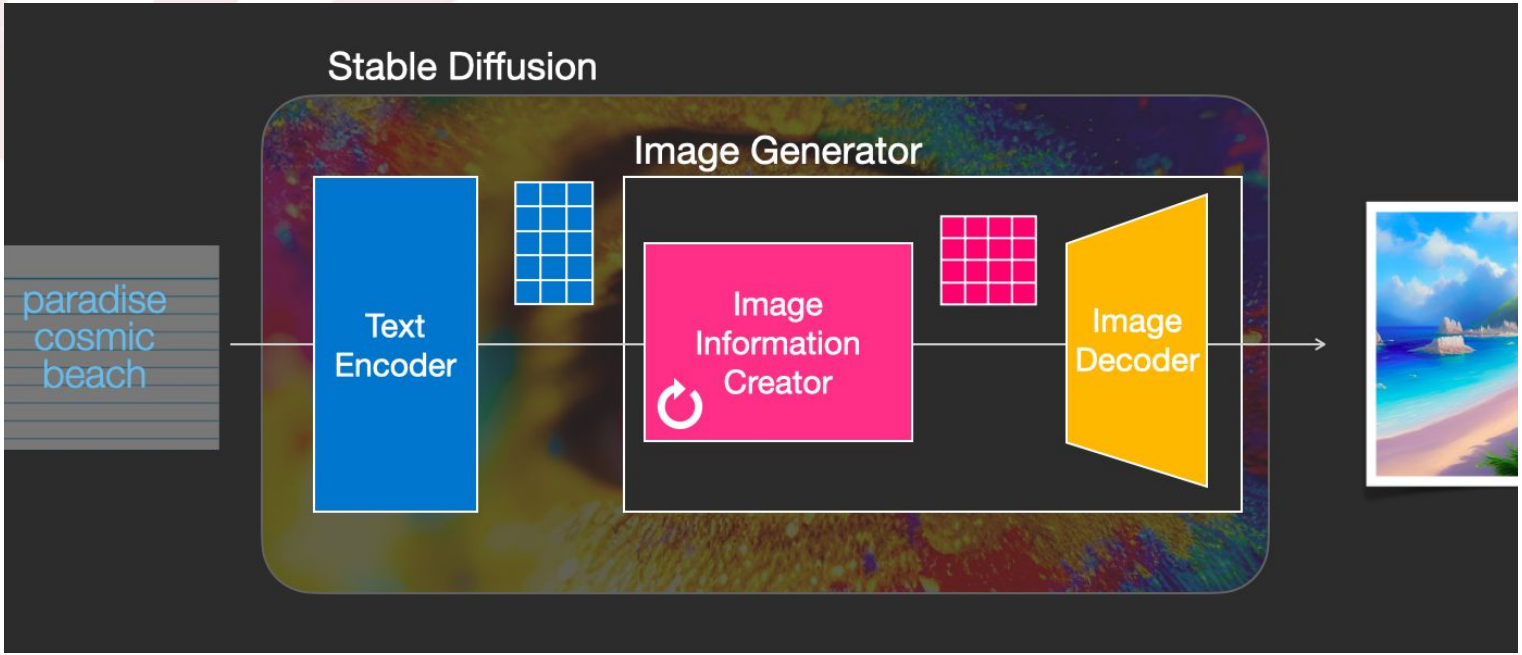
Introducción a Stable Diffusion

Explicación basada en https://jalammar.github.io/illustrated-stable-diffusion/

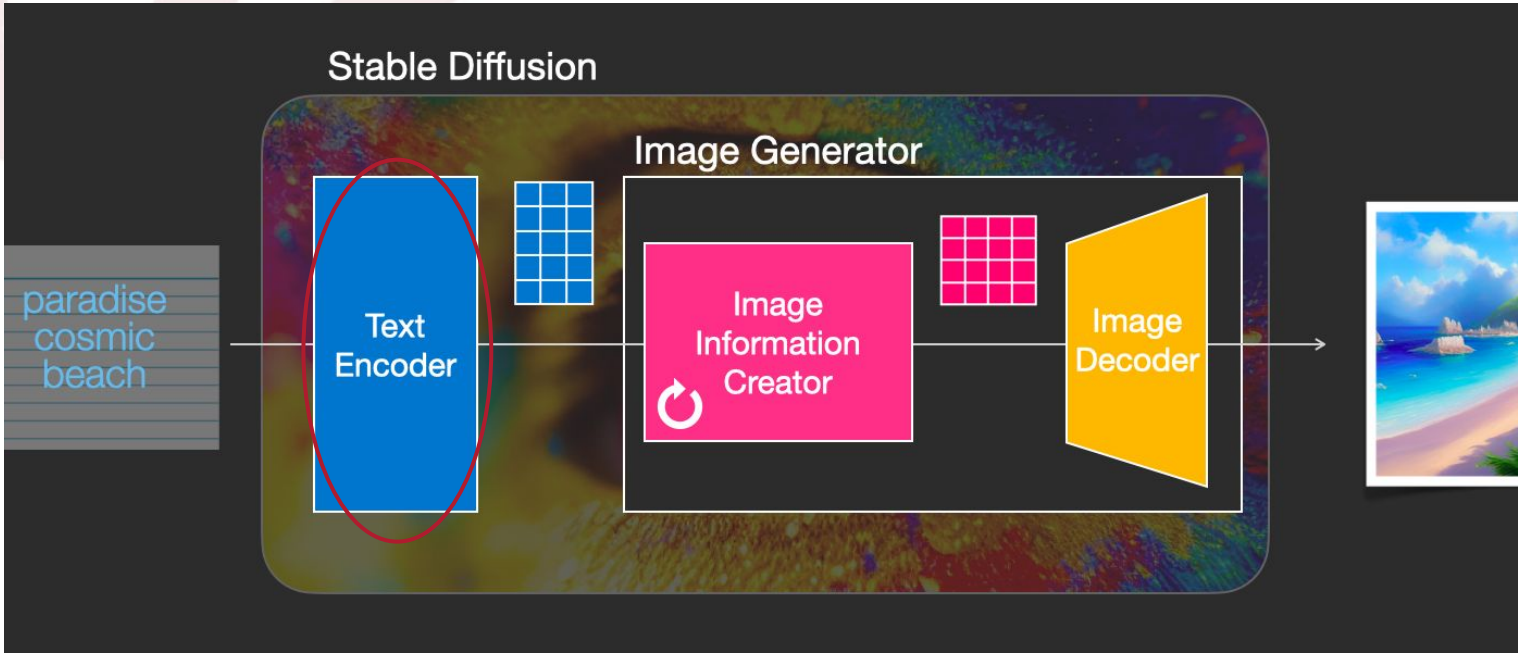
Componentes



Componentes



Componentes



Text encoder

CLIP modelo entrenado en

- Imágenes (modelo de clasificación)
- Descripciones (modelo de lenguaje)

Image encoder

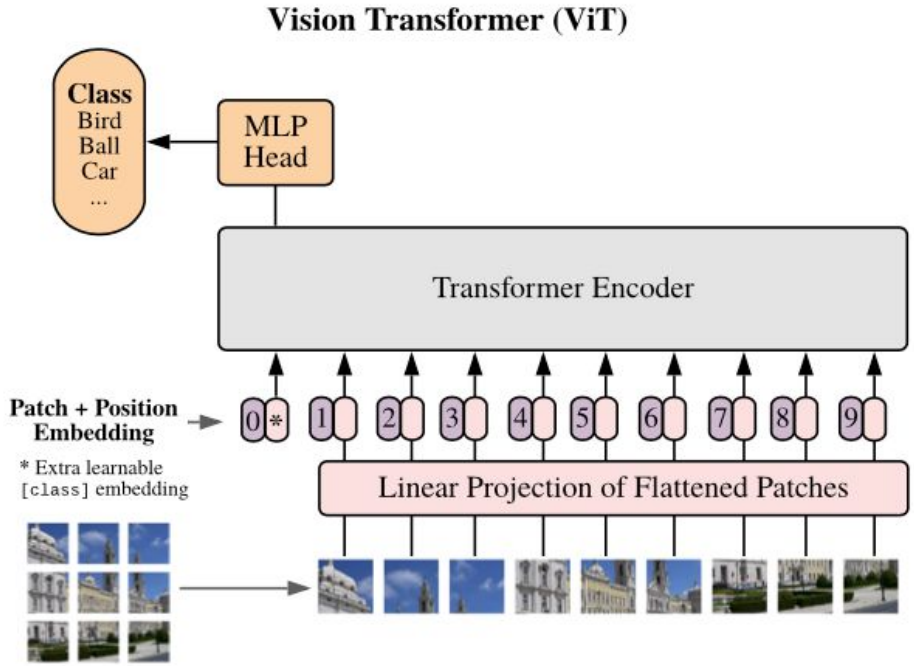
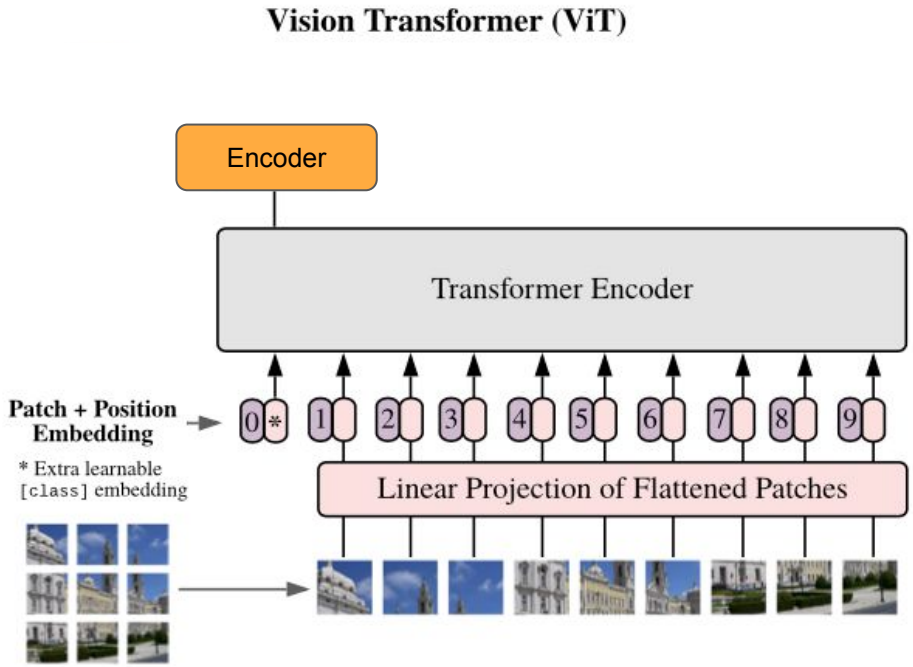
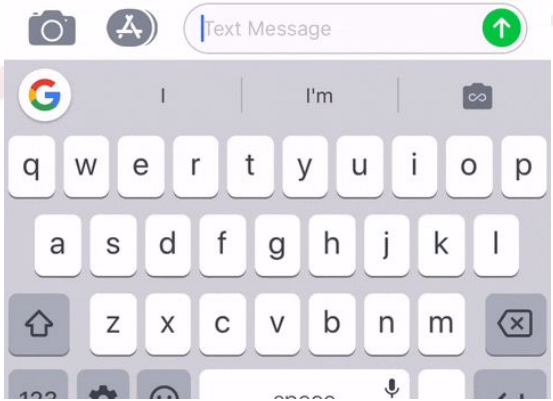


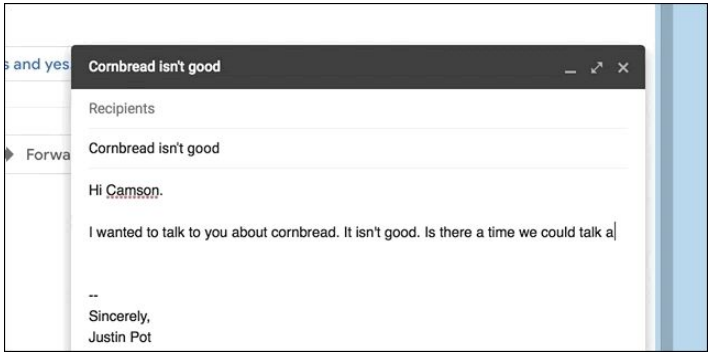
Image encoder

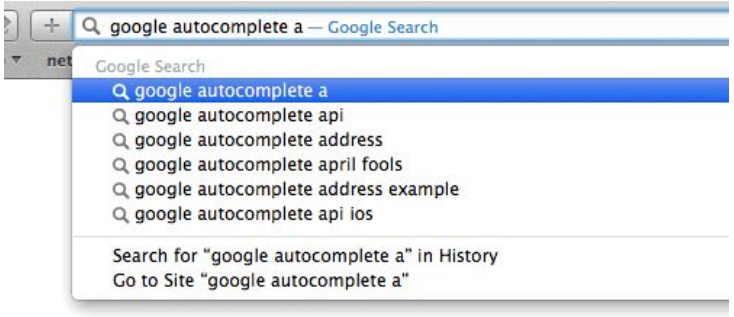


Text encoder

Modelo de lenguaje (BERT)



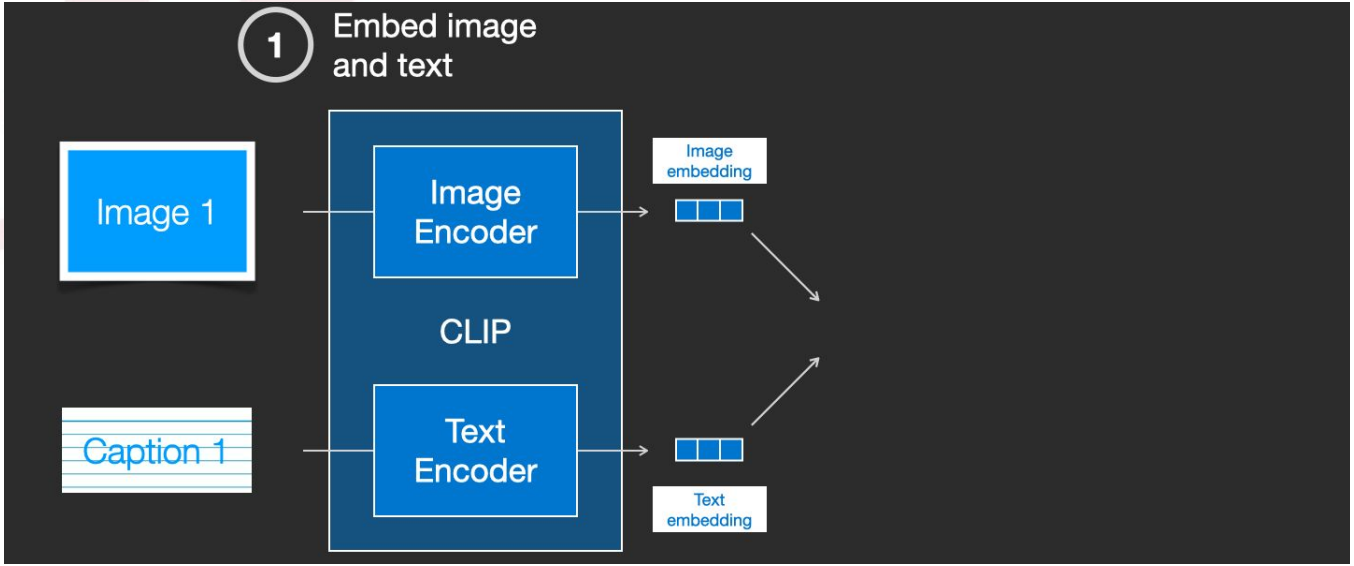




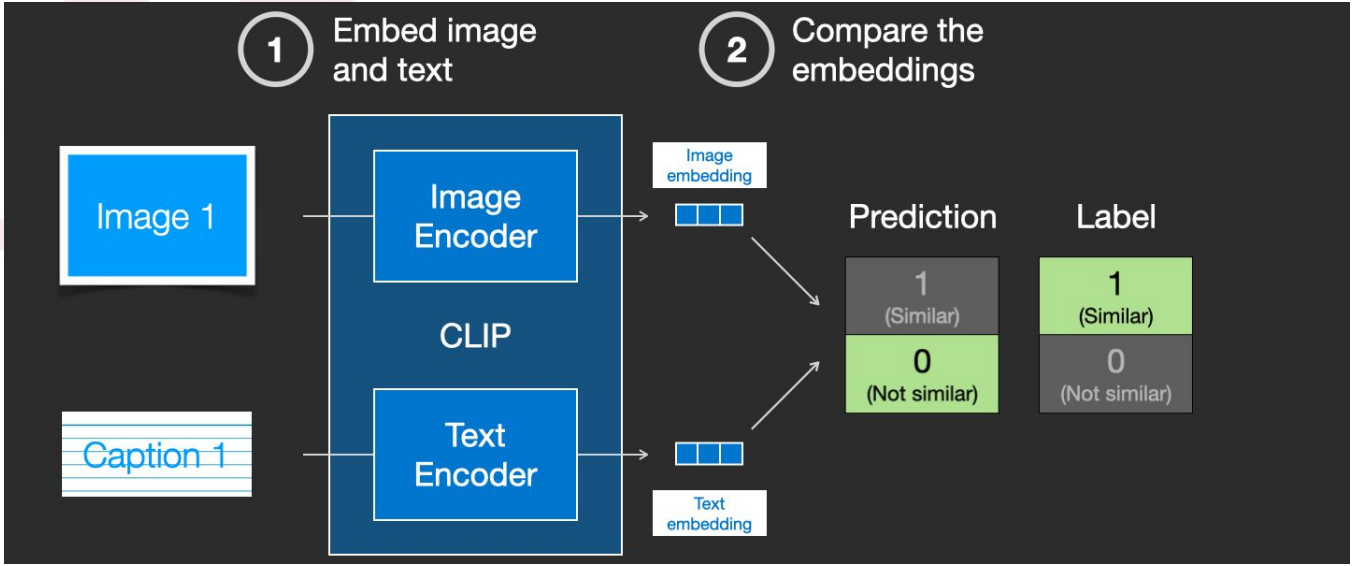
Entrenamiento CLIP

IMAGE			
CAPTION	Photo pour Japanese pagoda and old house in Kyoto at twilight - image libre de droit	Soaring by Peter Eades	far cry 4 concept art is the reason why it 39 s a beautiful game vg247. Black Bedroom Furniture Sets. Home Design Ideas

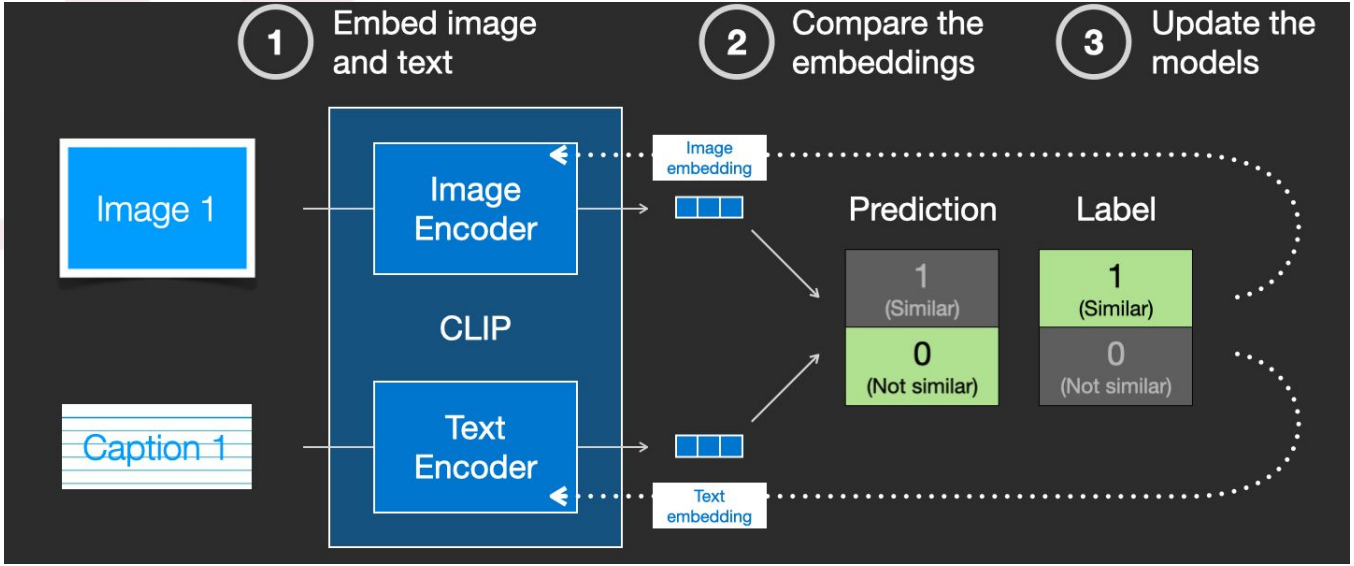
Entrenamiento CLIP



Entrenamiento CLIP



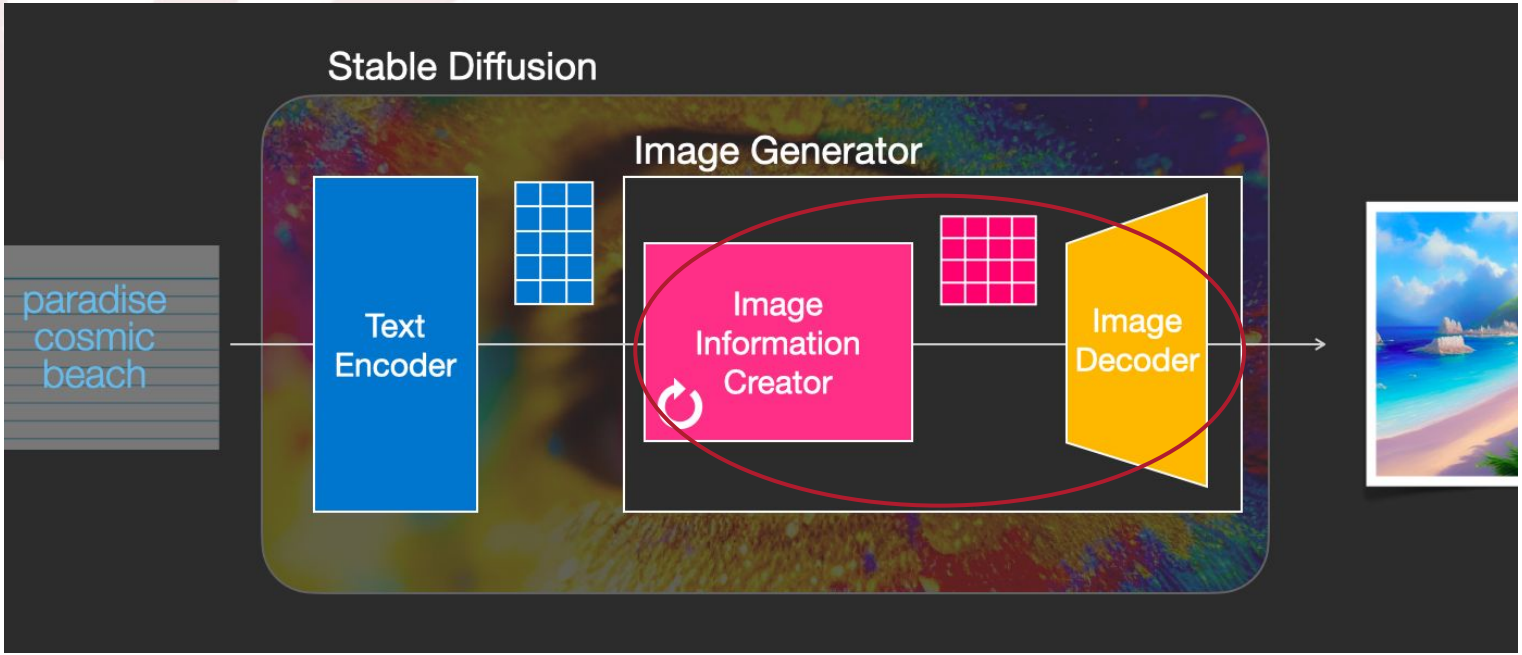
Entrenamiento CLIP



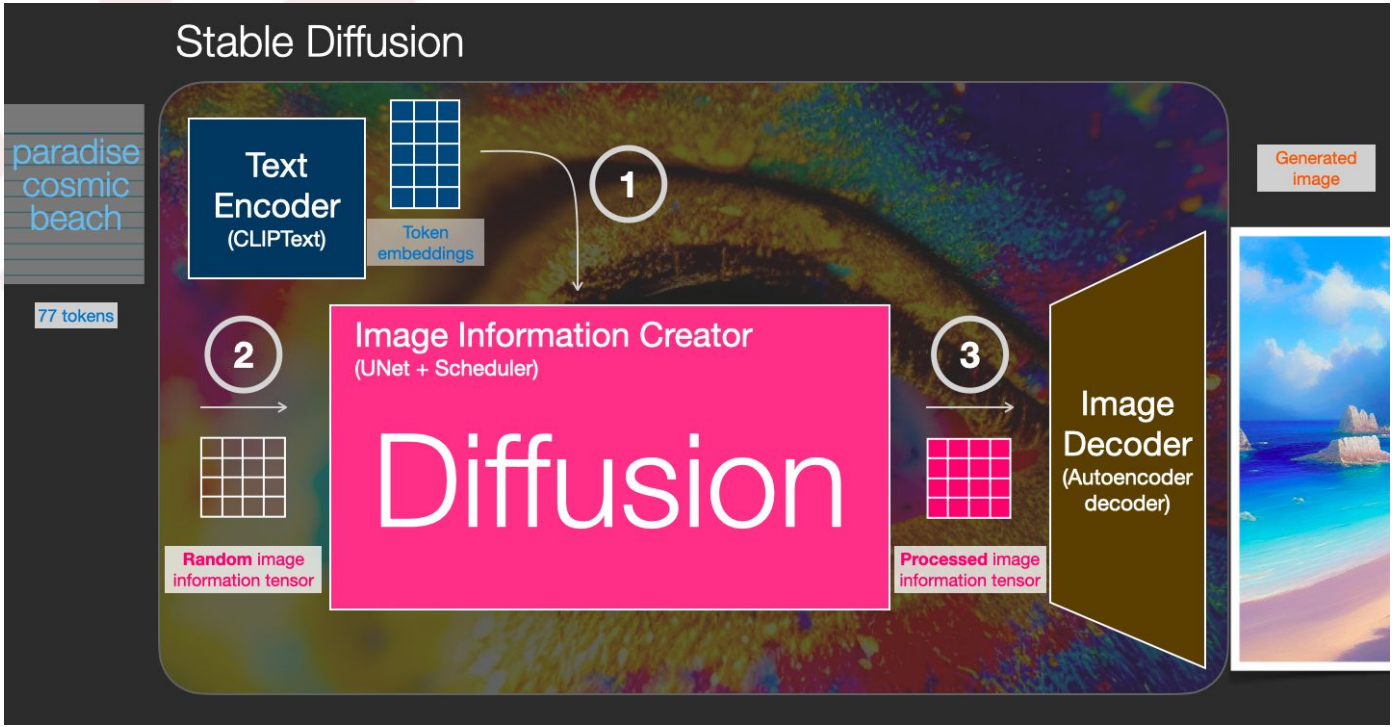
¿Cómo obtener prompts?

CLIP Interrogator

Componentes

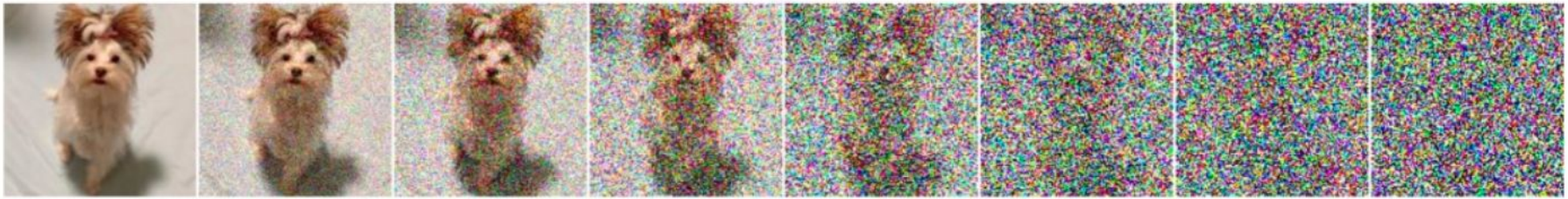


Componentes



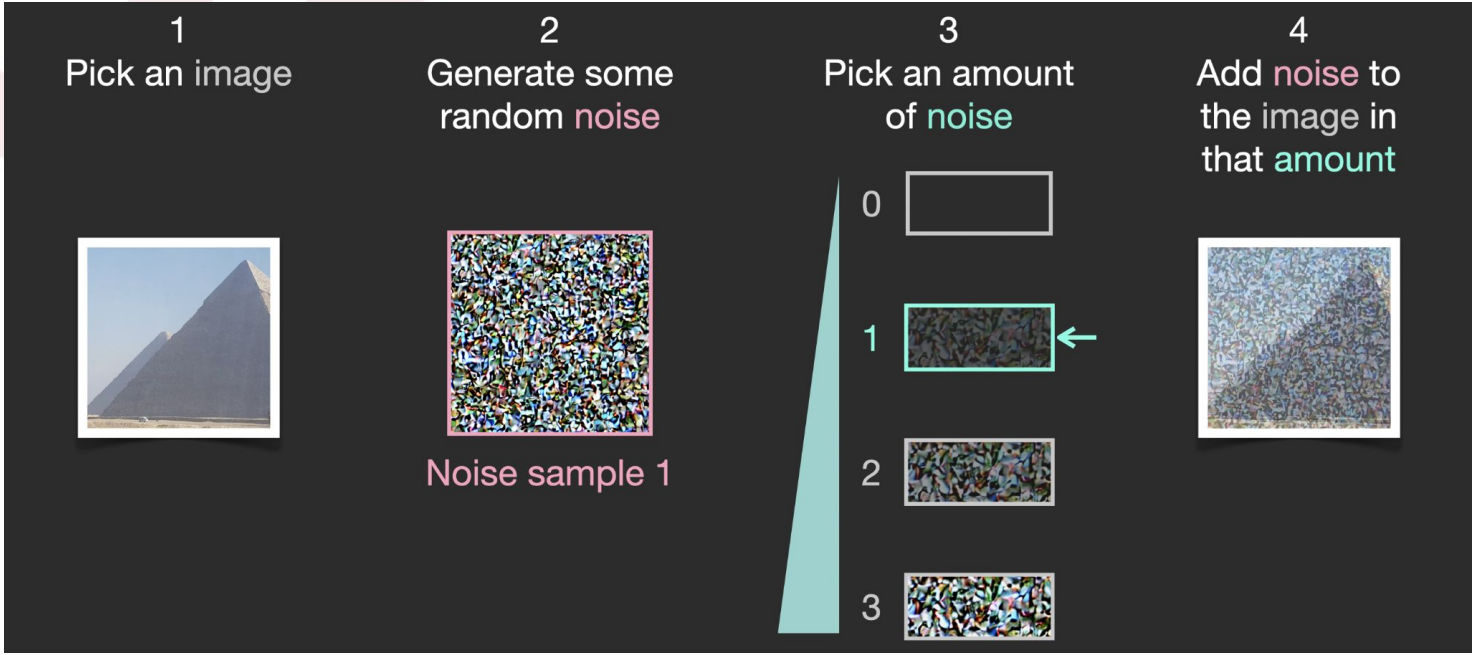
¿Cómo funciona la difusión?

Forward pass

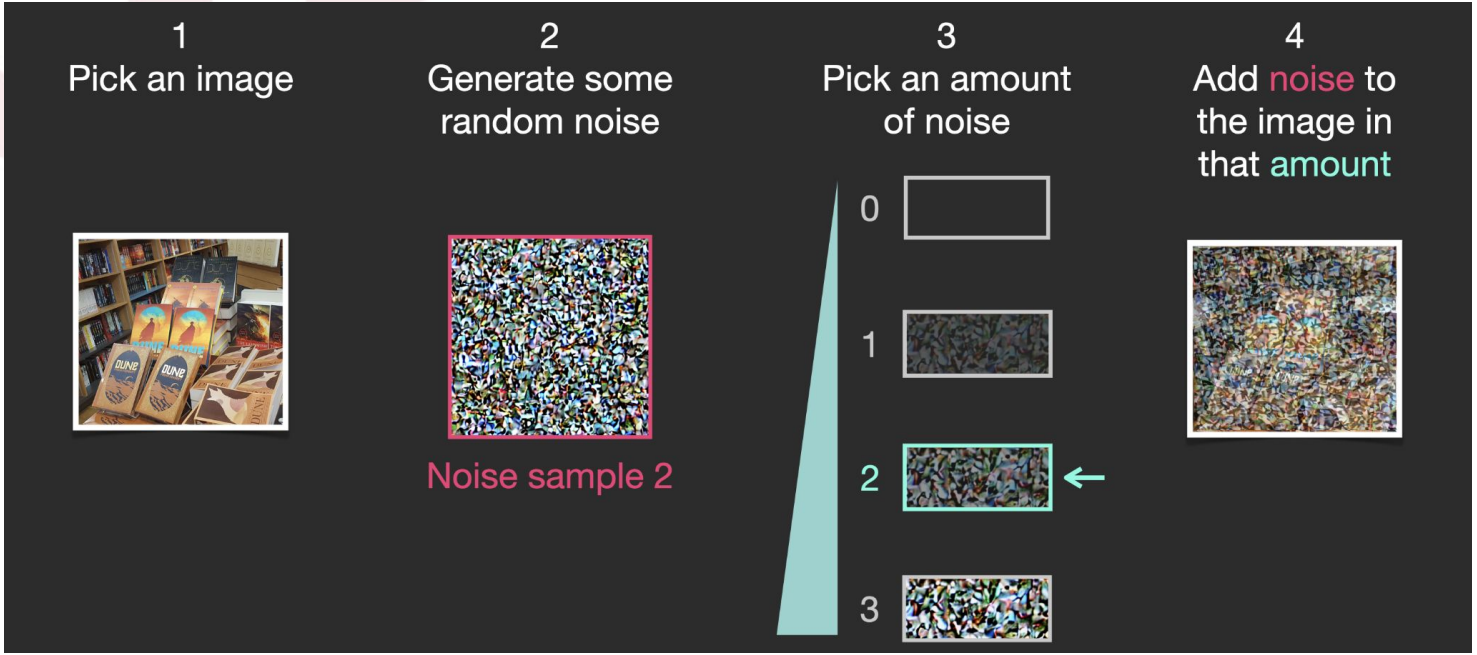


Backward pass

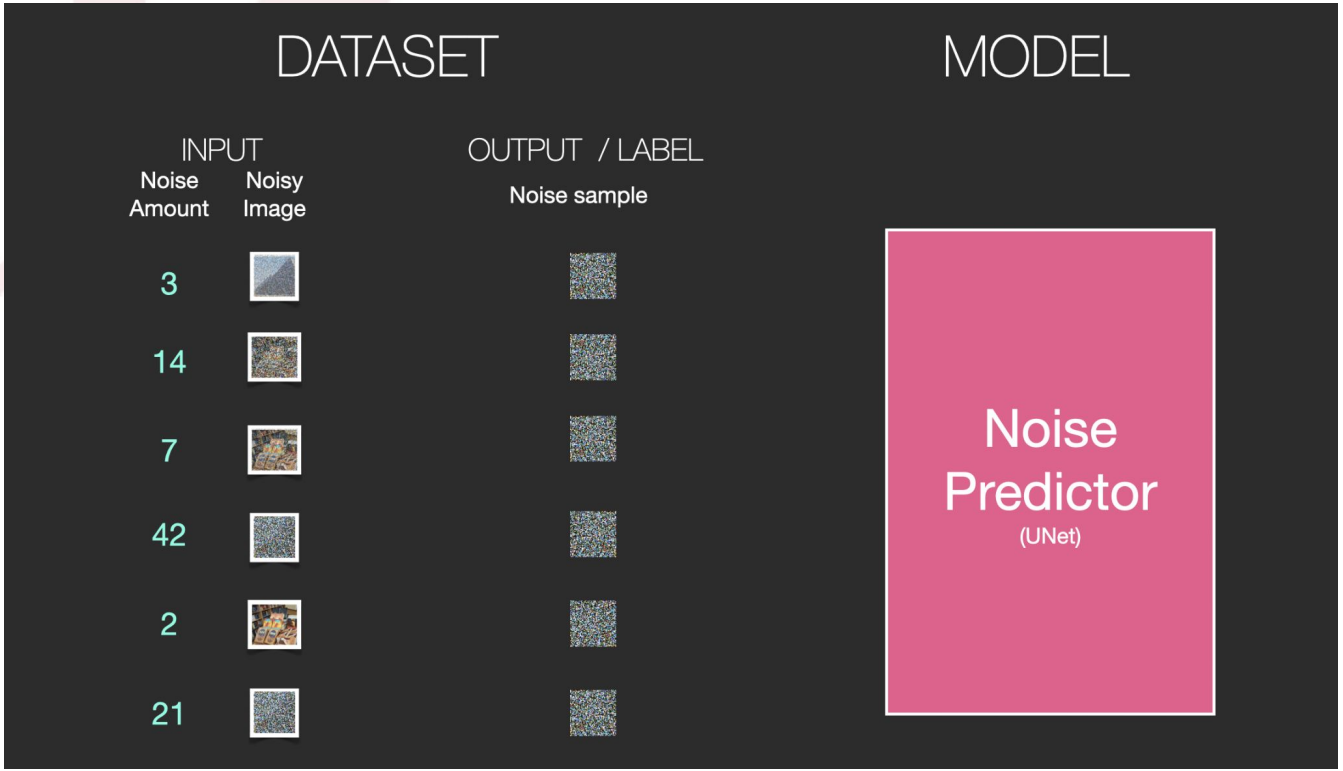
Generación dataset



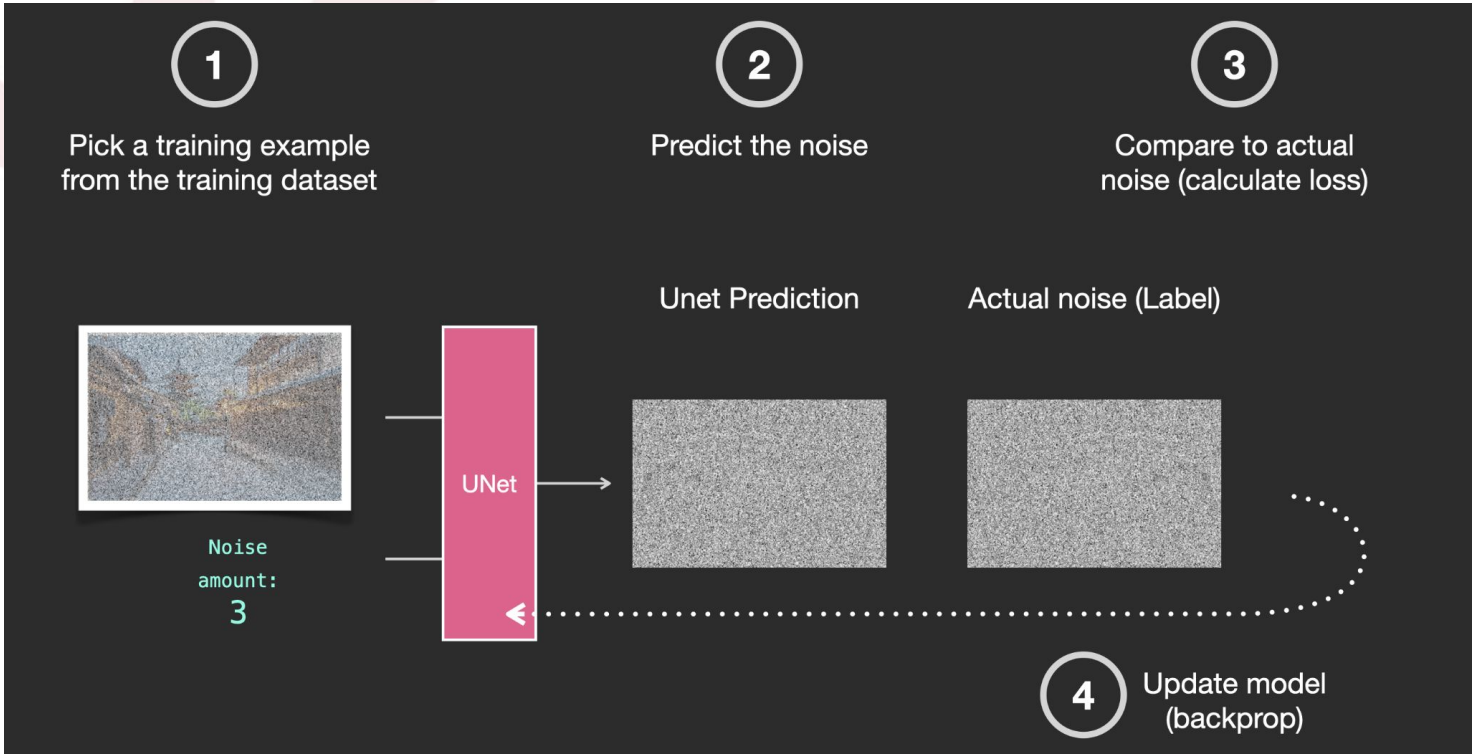
Generación dataset



Generación dataset

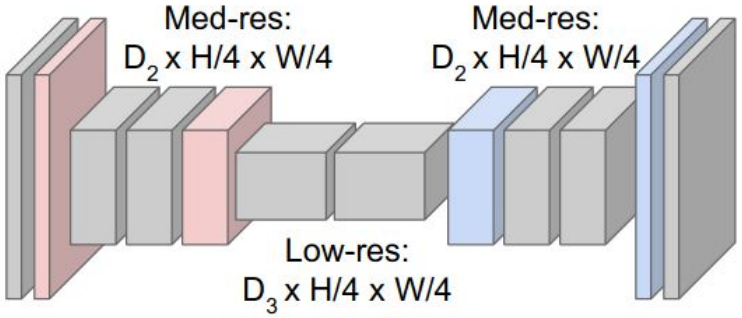


Entrenamiento modelo



U-Net

Input:
3 x H x W

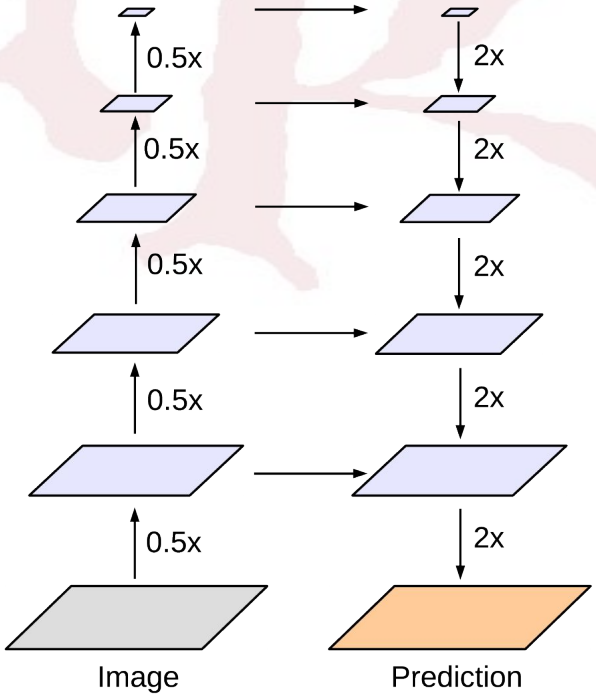


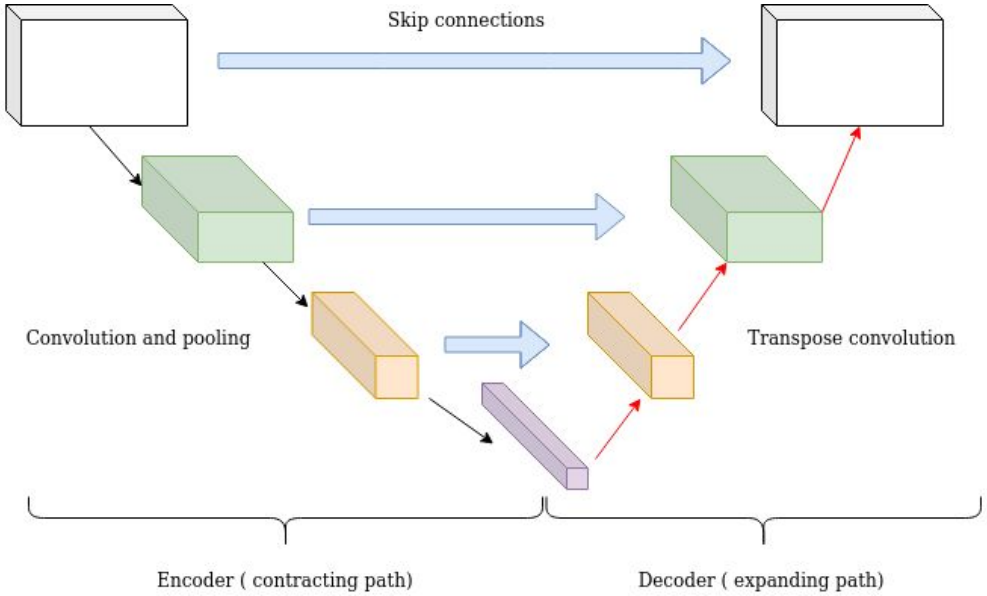
High-res:
D_1 x H/2 x W/2

High-res:
D_1 x H/2 x W/2

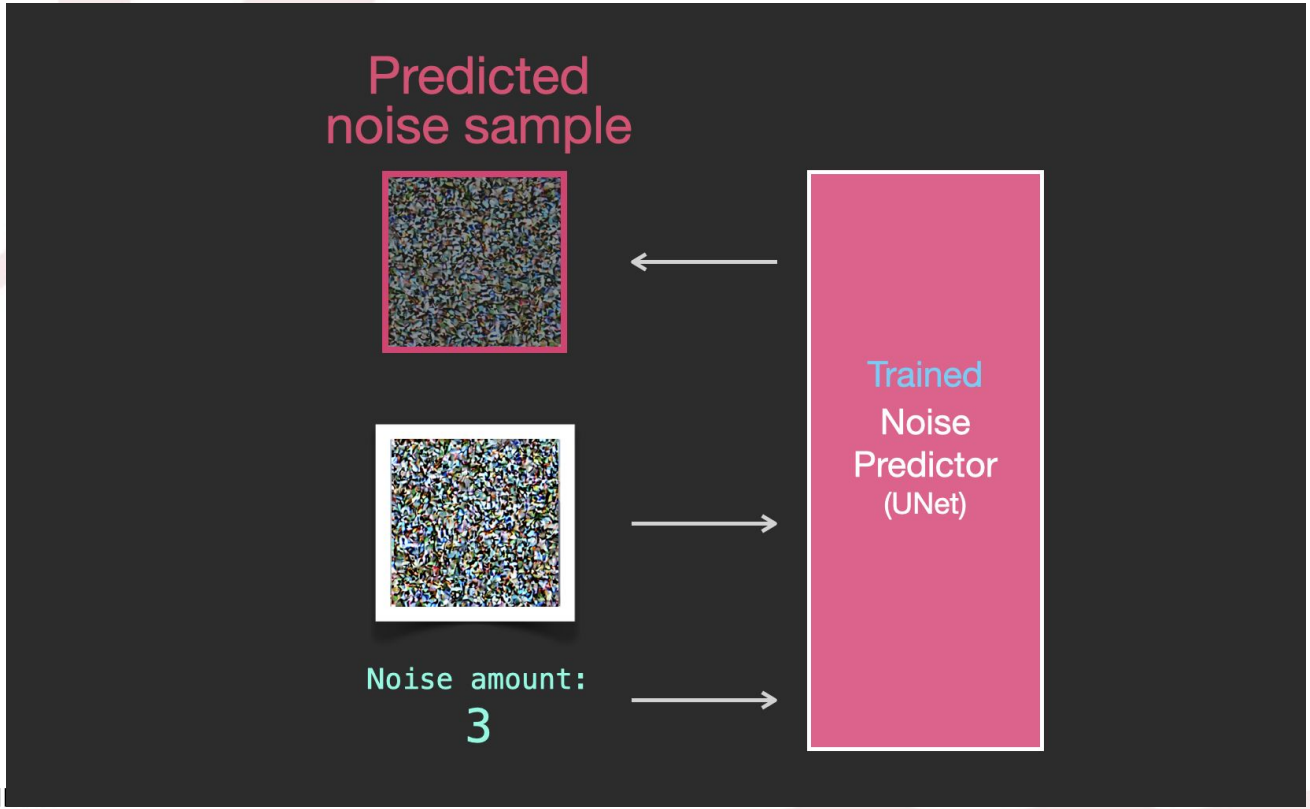
Predictions:
H x W

U-Net

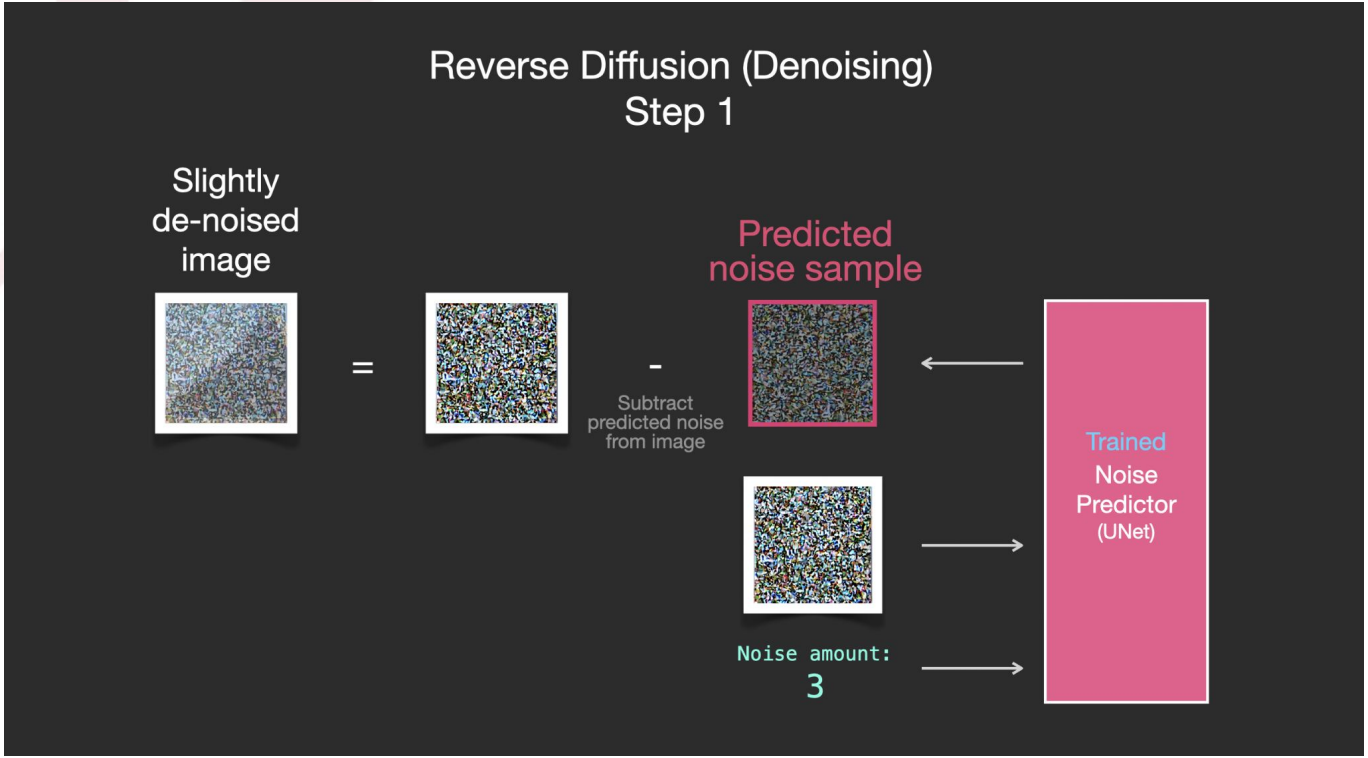

0.5x
0.5x
0.5x
0.5x
0.5x
2x
2x
2x
2x
2x
Image
Prediction


Skip connections
Convolution and pooling
Transpose convolution
Encoder (contracting path)
Decoder (expanding path)

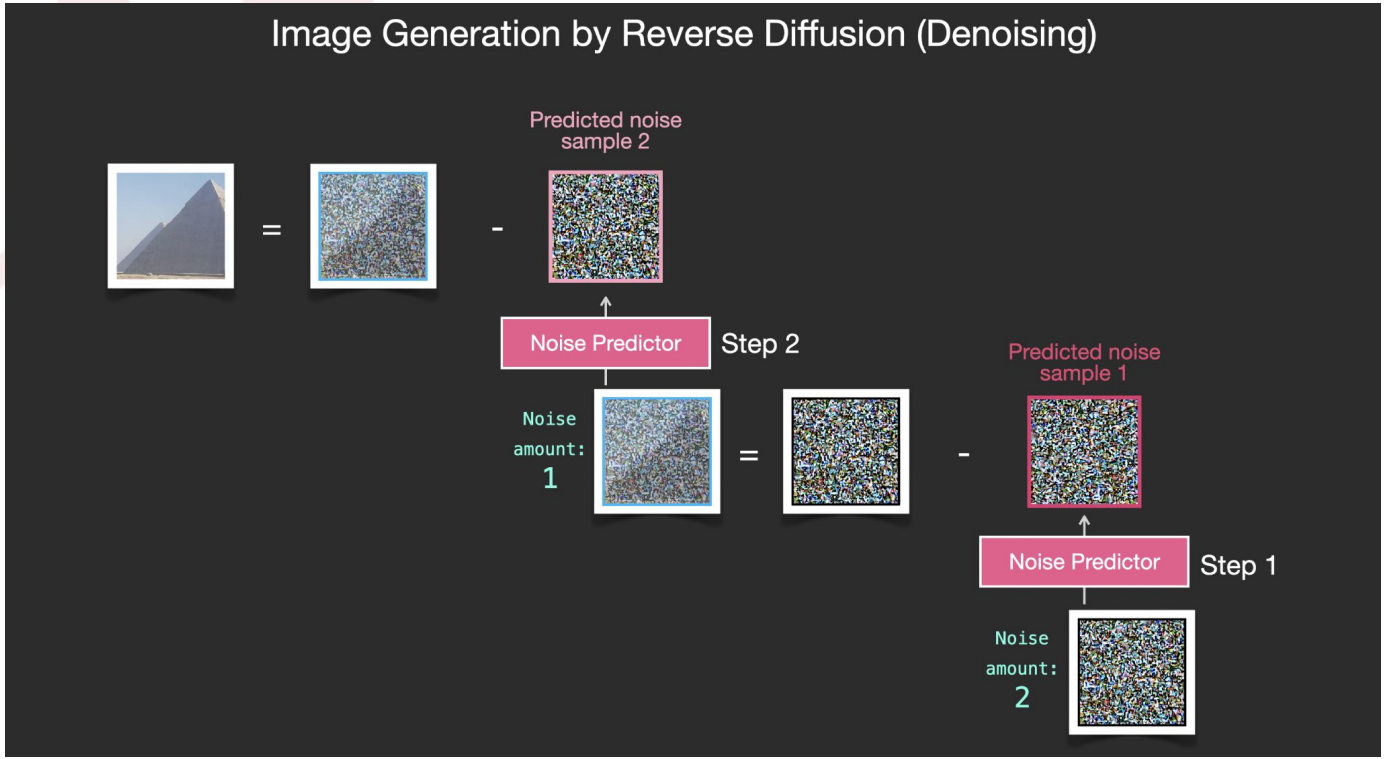
Generando imágenes a partir del ruido



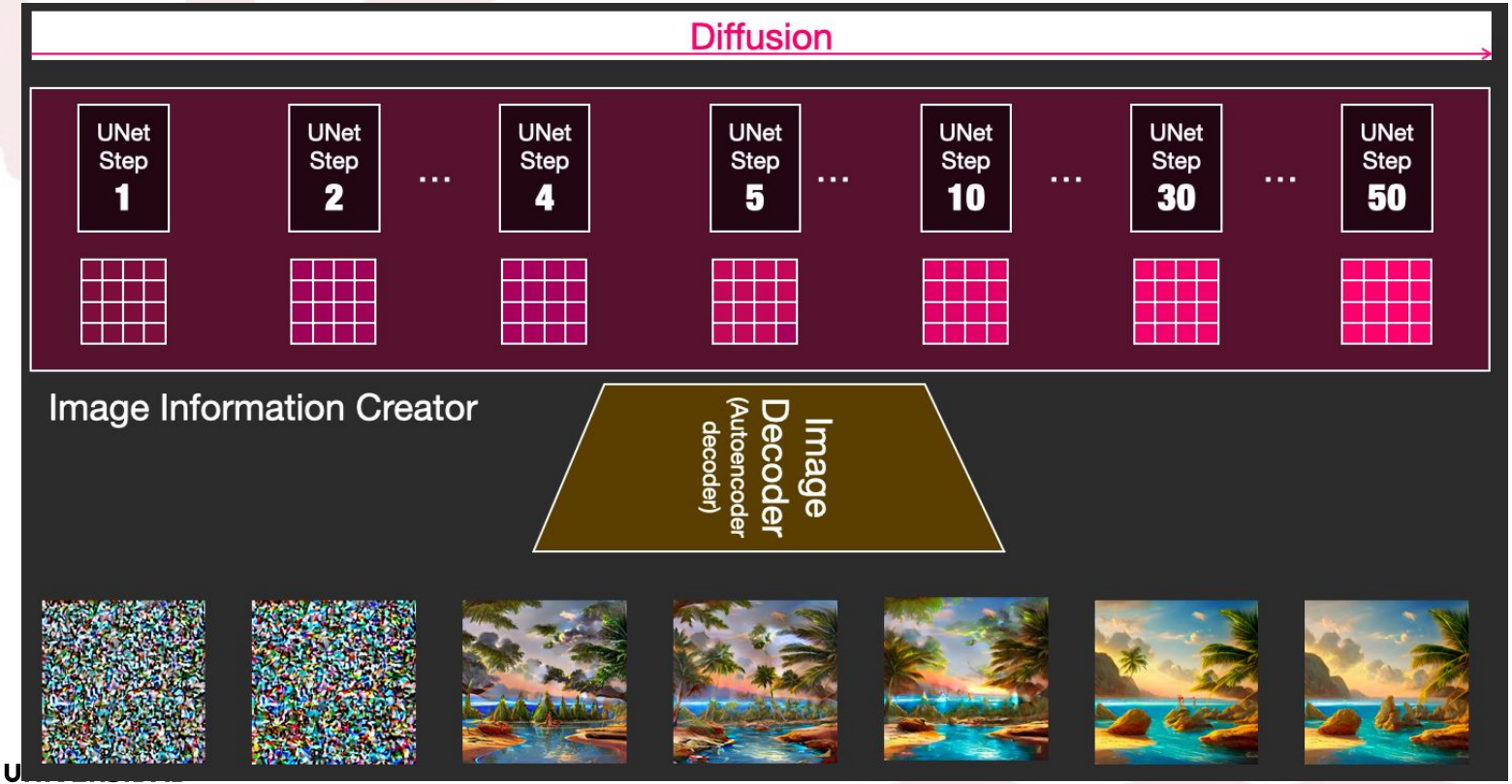
Generando imágenes a partir del ruido



Generando imágenes a partir del ruido

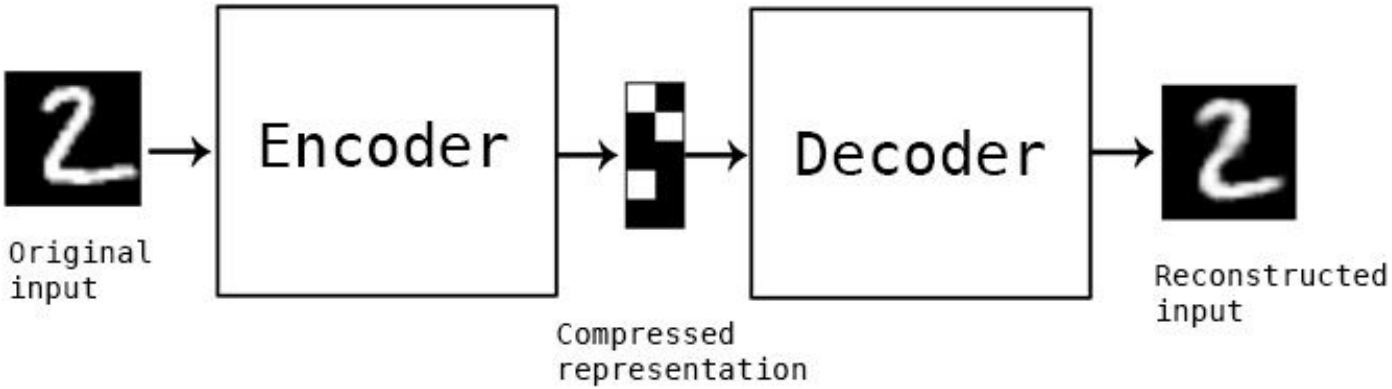


Generando imágenes a partir del ruido

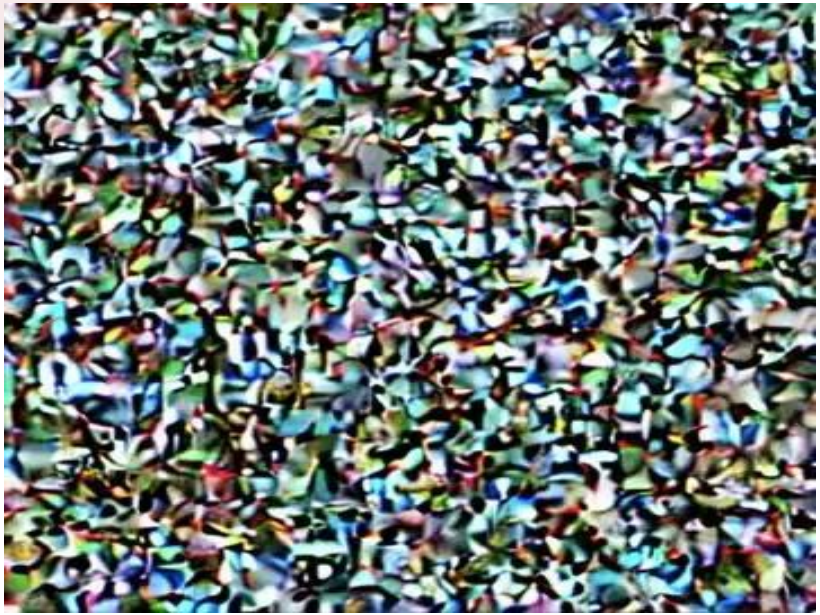


Autoencoders

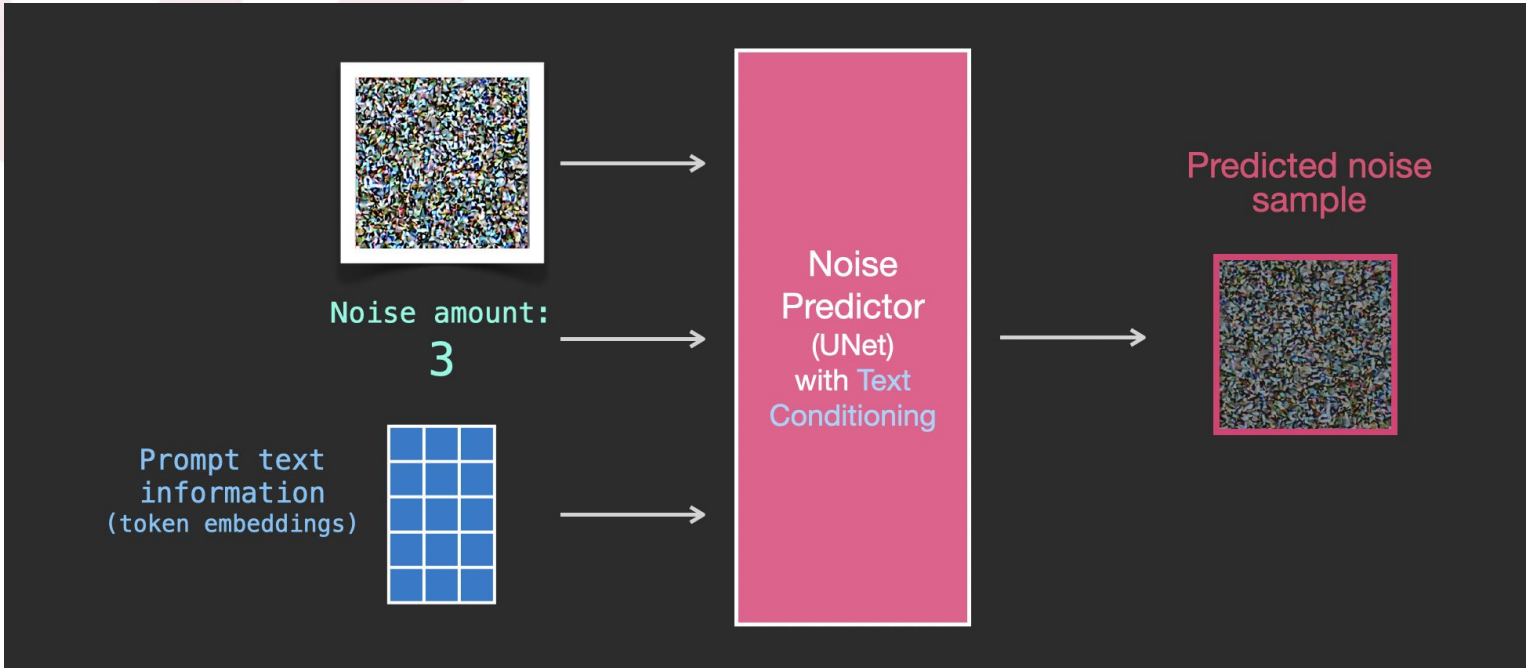
Compresión de datos



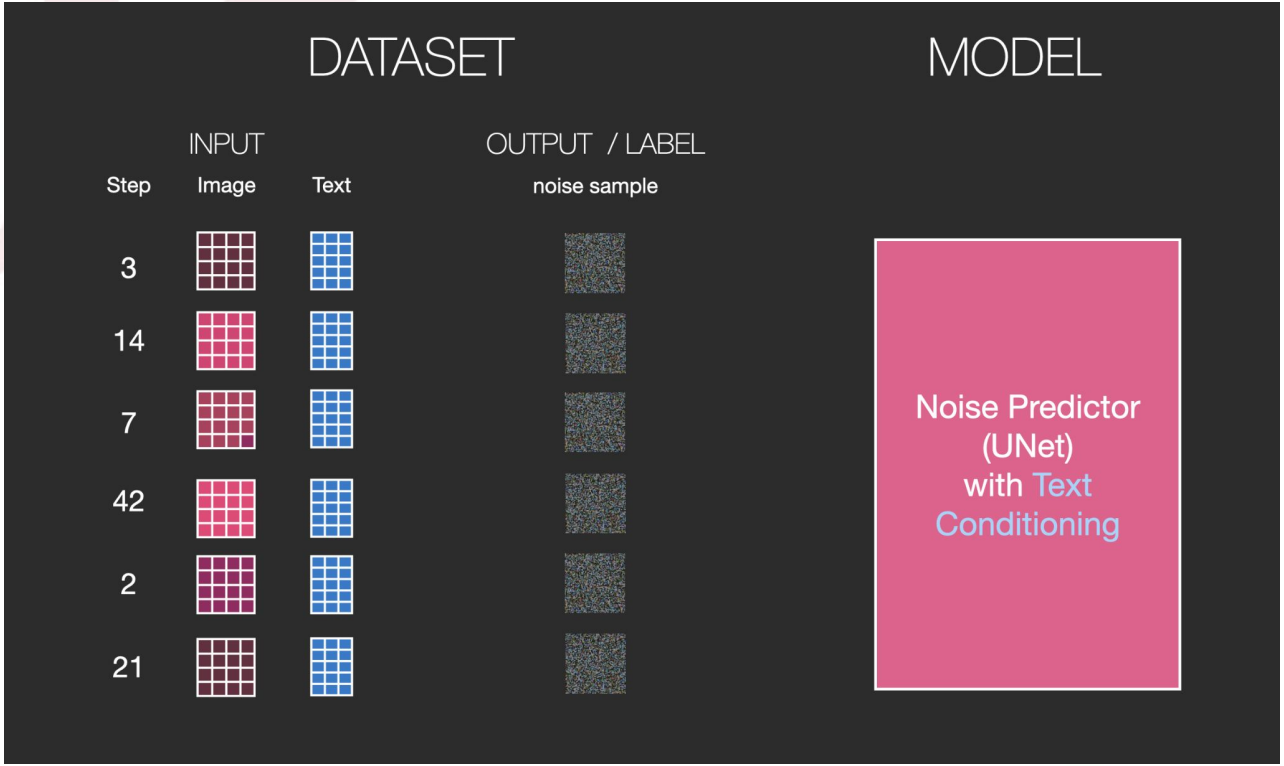
Generando imágenes a partir de ruido



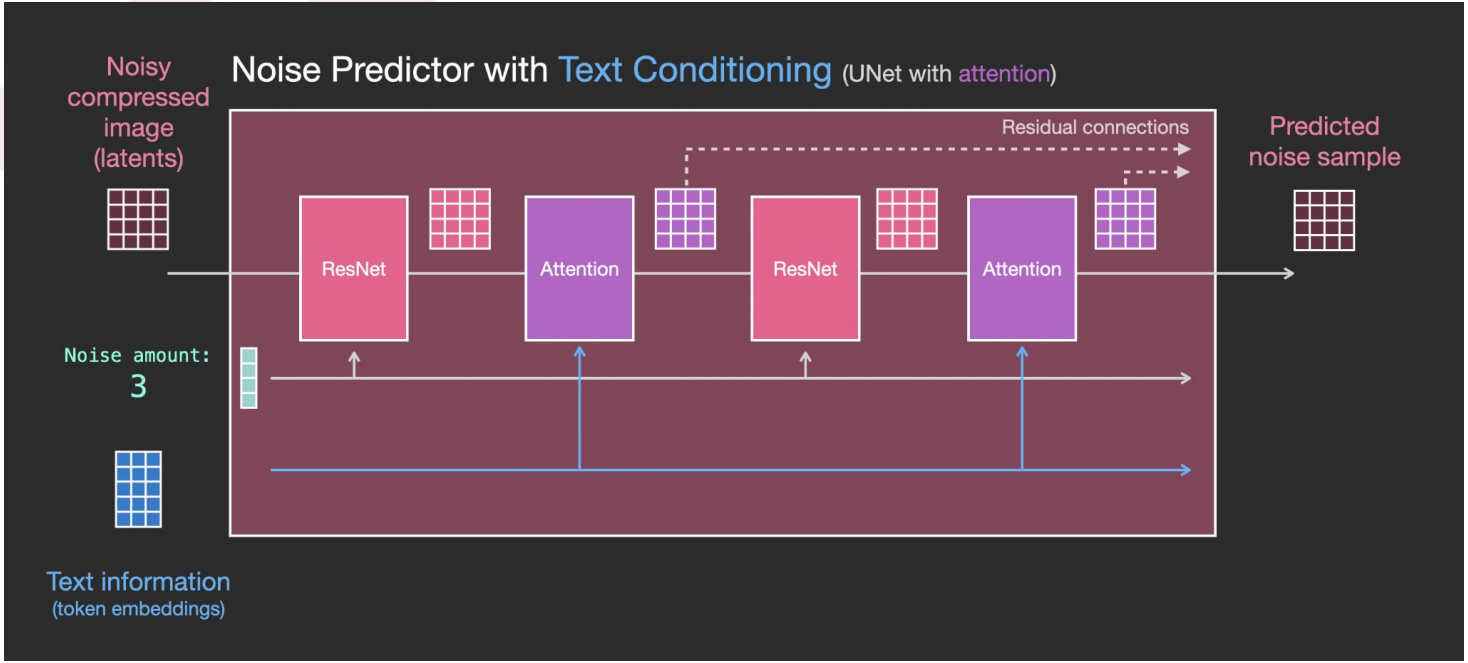
Proporcionando información textual



Proporcionando información textual



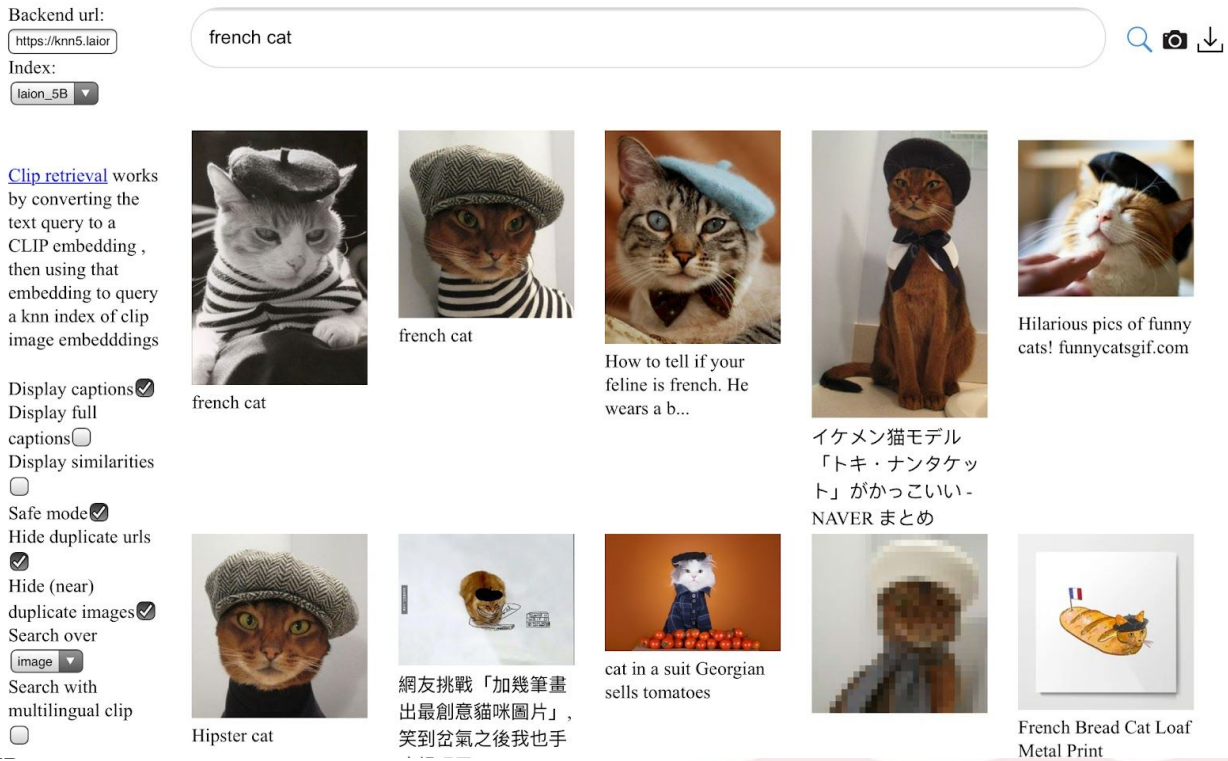
Proporcionando información textual



Índice

1. ¿Qué es?
2. ¿Cómo funciona?
3. **Cuestiones éticas/legales**
4. ¿Qué más se puede hacer?

Datasets entrenamiento



Copyright

Do AI images violate copyright? A lawyer explains the Stable Diffusion lawsuit

MARK FRAUENFELDER / 9:39 AM MON JAN 23, 2023

¿A quién pertenecen las imágenes?

- En general copyright pertenece a quien crea las imágenes
  - StableDiffusion, MidJourney and Dalle-2 asignan autoría a creador
- En USA no hay copyright para trabajos creados por máquinas…
- … pero sí si el creador prueba estar involucrado activamente en la creación

Imitación de estilos

- ¿Puedo crear un cómic de Super Mario y venderlo?
  - Personajes como Mickey Mouse o Spiderman protegidos por propiedad intelectual
  - Puedes dibujarlos de la manera que quieras pero no venderlo
- ¿Puedes imitar estilos?
  - Los estilos no tienen copyright, así que sí

Sesgos

A computer scientist

A nurse

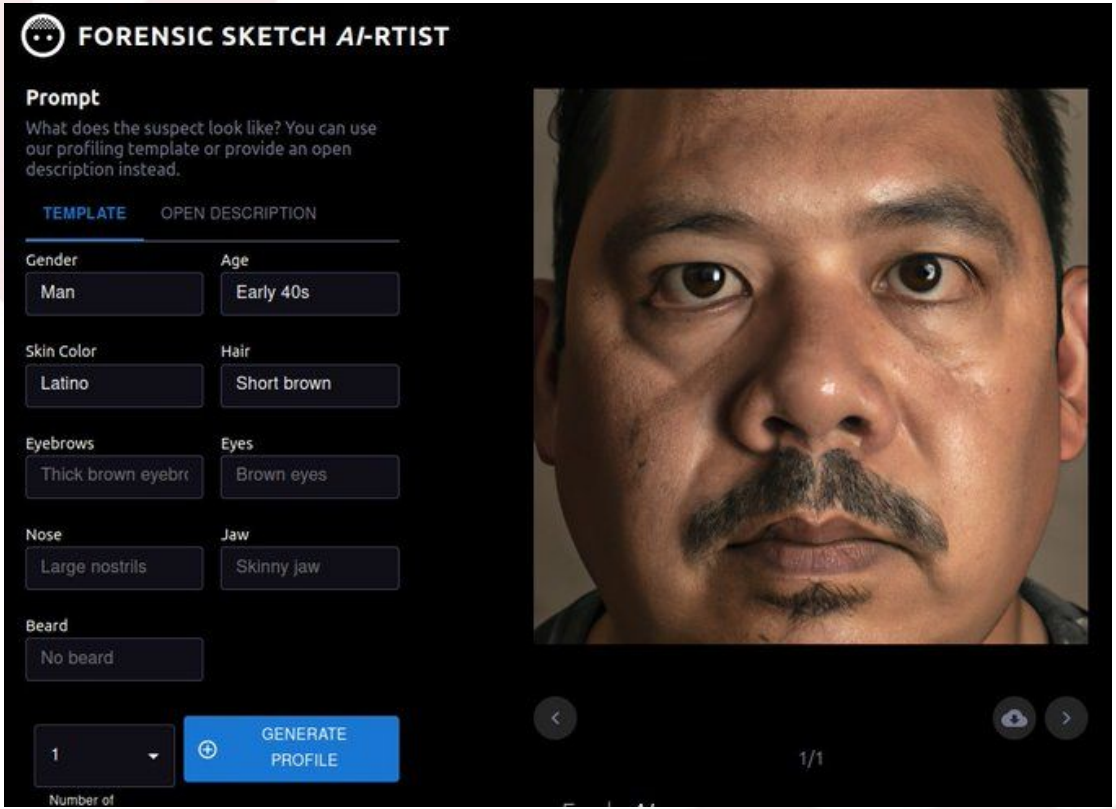
A mathematician

Sesgos

Imágenes NSFW

- Imágenes NSFW de individuos
- Imágenes violentas
- Pornografía infantil
- …

Forensic sketches



Índice

1. ¿Qué es?
2. ¿Cómo funciona?
3. Cuestiones éticas/legales
4. **¿Qué más se puede hacer?**

¿Qué más?

- Edición de imágenes
- Generación de música
- Generación 3D
- Generación de vídeos
- …

Edición de imágenes

"zebras roaming in the field"

"a girl hugging a corgi on a pedestal"

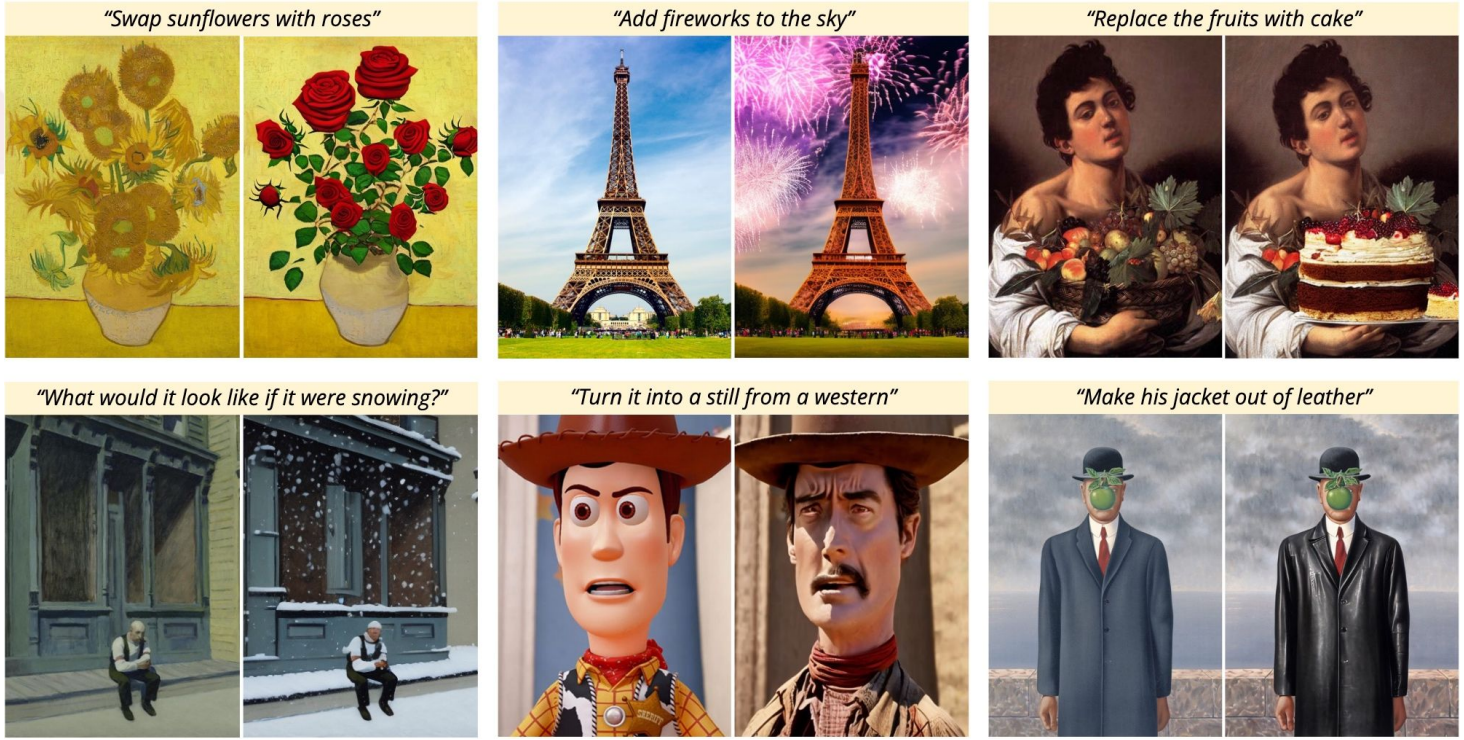
"a man with red hair"

"a vase of flowers"

"an old car in a snowy forest"

"a man wearing a white hat"

Edición de imágenes



Generación de música

MusicLM: Generating Music From Text

| paper | dataset |

Andrea Agostinelli, Timo I. Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, Christian Frank
Google Research

Abstract We introduce MusicLM, a model generating high-fidelity music from text descriptions such as *"a calming violin melody backed by a distorted guitar riff"*. MusicLM casts the process of conditional music generation as a hierarchical sequence-to-sequence modeling task, and it generates music at 24 kHz that remains consistent over several minutes. Our experiments show that MusicLM outperforms previous systems both in audio quality and adherence to the text description. Moreover, we demonstrate that MusicLM can be conditioned on both text and a melody in that it can transform whistled and hummed melodies according to the style described in a text caption. To support future research, we publicly release MusicCaps, a dataset composed of 5.5k music-text pairs, with rich text descriptions provided by human experts.

Generación 3D

Generación de vídeos

Edición de vídeo

Gracias

UNIVERSIDAD DE LA RIOJA

a lovebird in pixar style that is really happy with a computer science background