

FINETUNER UN MODÈLE DE GÉNÉRATION D'IMAGE POUR REPRODUIRE UNE IDENTITÉ VISUELLE

Matthieu Grosselin & Dmitry Kuzovkin
17/04/2024



Matthieu Grosselin

Co-founder & co-CEO @Seelab.ai

Working in tech since 2015 in
Product / UX

Former Chief Product Officer @

kewego

viadeo

GIROPTIC

HiPay



Dmitry Kuzovkin

Head of AI @Seelab.ai

PHD in Computer vision from
Rennes University

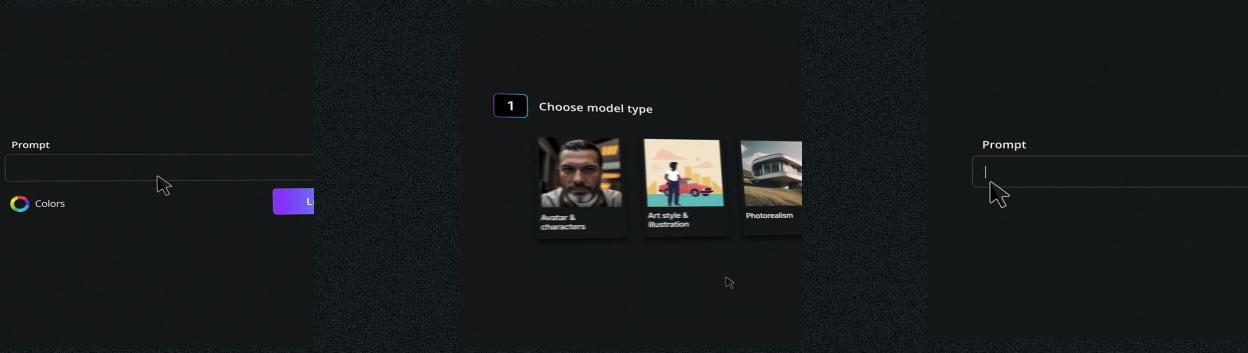
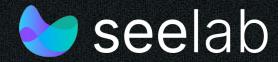
Former R&D Team Lead @ Meero

technicolor

meero



Seelab.ai in a few words...



Simple AI tool
made for
marketing &
creative teams

Create your own
image AI based
on your brand
style assets

Produce
consistent content
you can share with
your team and
clients

Founded in 2023

10 employees

French , built in
Brittany

Model agnostic

All your data stays
private

Full IP on images
you produce with
your models

Create your own image AI



Usecase:

Training on an illustration style for Gandi.net

User uploads his dataset (10 images here)



Gandi's
model is
created

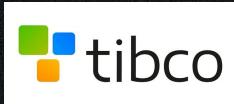
after 40 mn of
automated work



User generates images with simple prompts in a few seconds



Create your own image AI



Usecase:
Training on an illustration style for Tibco

User uploads his dataset (12 images here)



Tibco's
model is
created

after 45 mn of
automated work



User generates images with simple prompts in a few seconds



Create your own image AI



Usecase:

Training an avatar for France Pronos

User uploads his dataset (8 images here)



France
Pronos
model is
created

after 40 mn of
automated work



User generates images with simple prompts in a few seconds



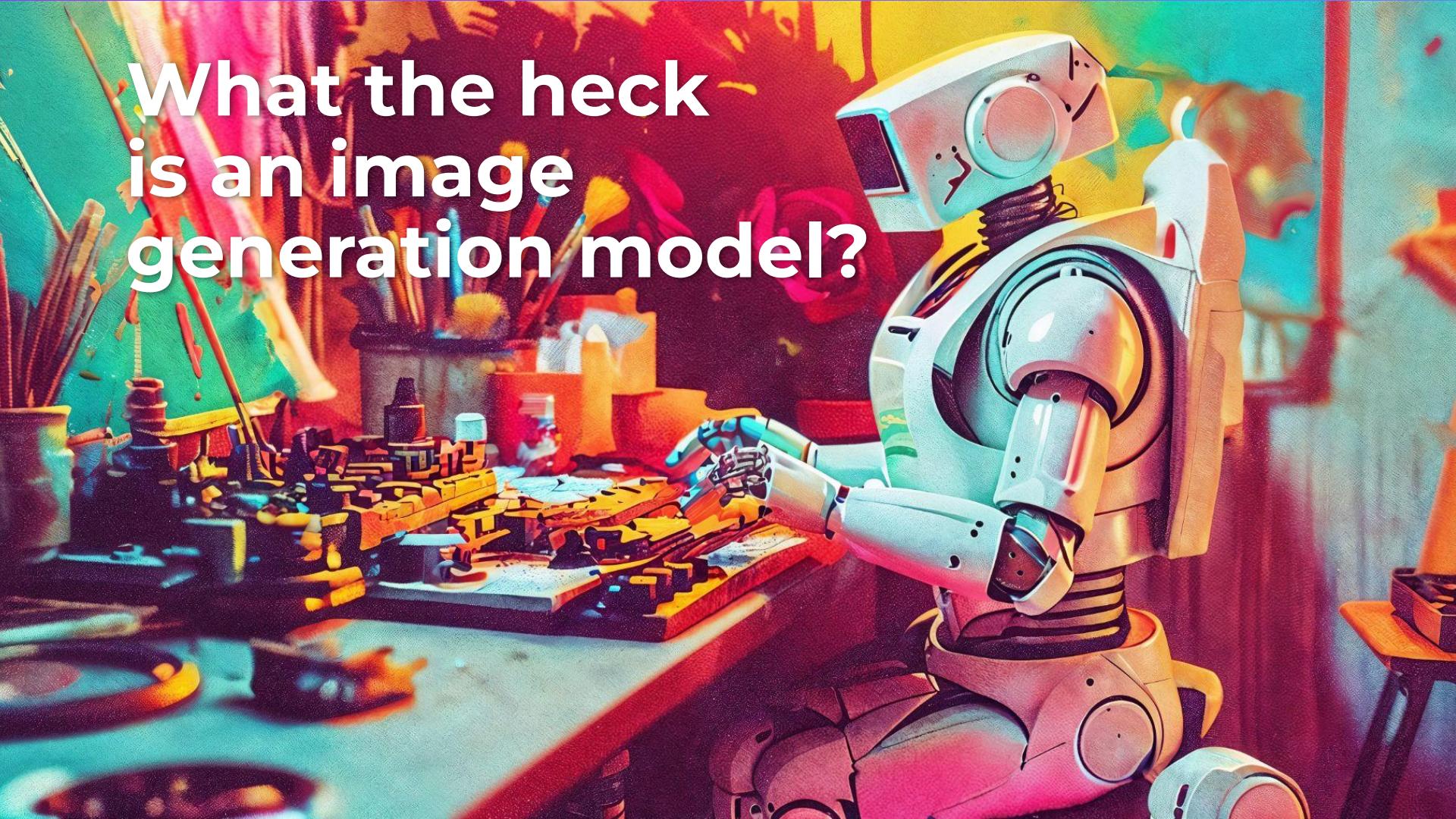
Remember “Shift Hackaton”?

SHIFT
nantes

LE HACKATHON GEN AI

Du **31 Mai au 2 Juin**, tu as exactement 48h pour créer le futur en intégrant de l'IA Générative dans un produit tech !
Tout ça au **Palace**, à Nantes.





What the heck
is an image
generation model?

Text-to-image

National geographic award winning breathtaking stunning photography of a unicorn



Dall-e 3



Midjourney V6



Seelab/SD XL

} Prompt



"Human hand" Midjourney v1 Mars 2022



“Human hand” Midjourney v5.1 Mars 2023

Models evolve freaking quickly



"National geographic award winning breathtaking stunning photography of a unicorn"



1



3



4



5



6

Midjourney V1
feb 2022

Midjourney V3
july 2022

Midjourney V4
nov 2022

Midjourney V5
nov 2022

Midjourney V6
dec 2023



Rapid quality and inference speed growth in a few months

source: [algoartist](#)

Main AI image models

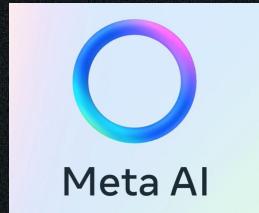


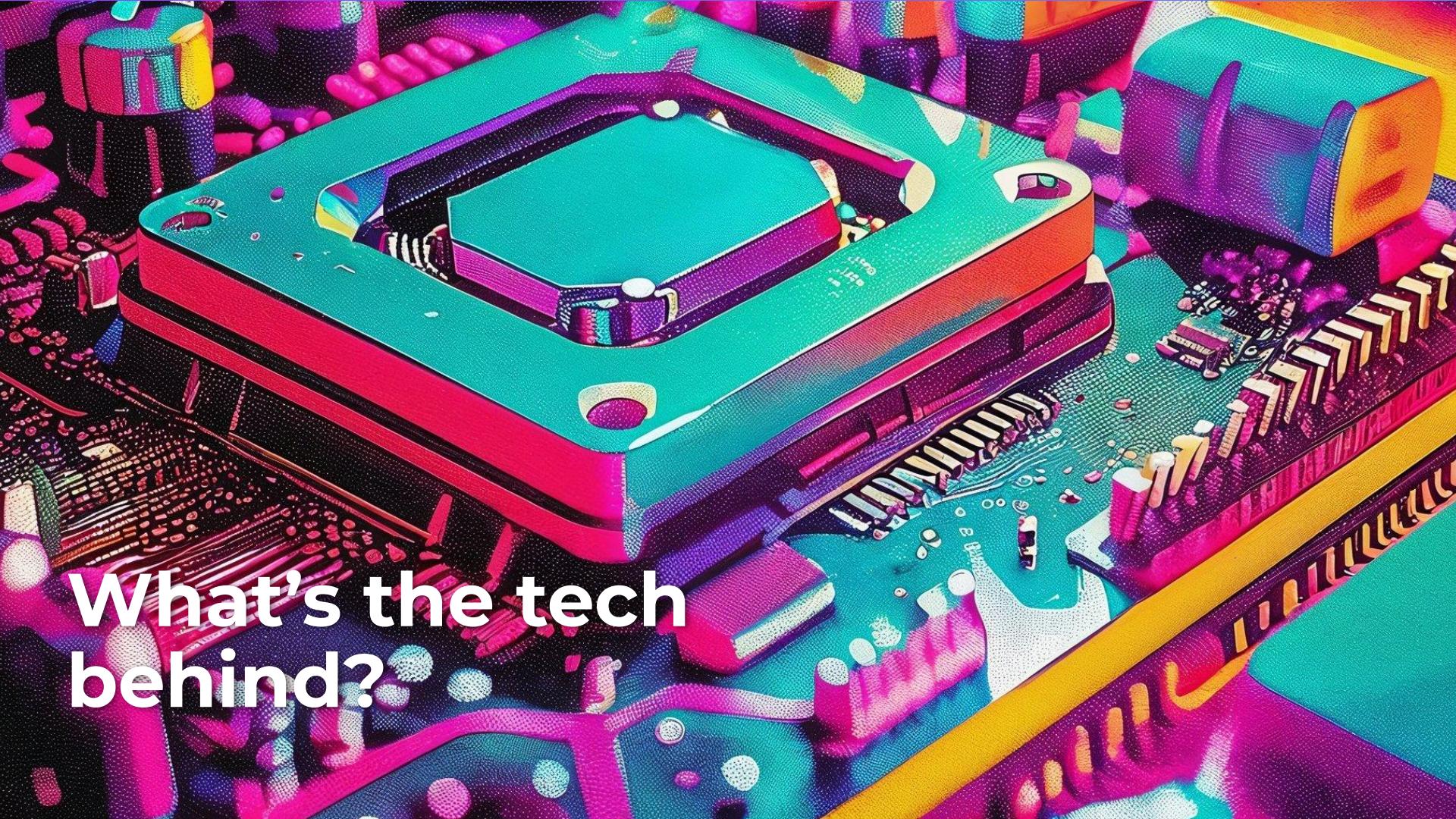
Open source

stability.ai

Stable Diffusion

Proprietary





What's the tech
behind?

Origins of Diffusion for Image Generation



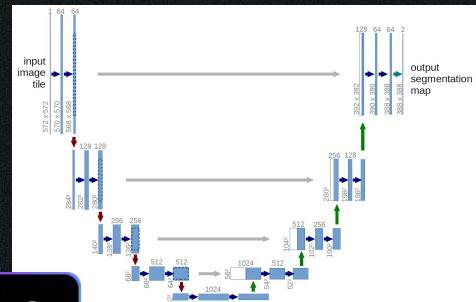
3 factors behind diffusion-based generation:



1

Diffusion models

from thermodynamics
physics - help to deal with
complex distributions that
are hard to model and
sample - just like image data



2

U-Net network

proposed for Biomedical Image
Segmentation - which used for many
Computer Vision applications



3

Transformers: Text Encoding + Attention Mechanism

from text processing (e.g.
translation): help with context
understanding. It inspired the
cross-attention mechanism, where
images and text are encoded in a
shared vectorial representation

Origins of Diffusion for Image Generation



1

Diffusion Models

Forward diffusion process (data → noise)



Backward diffusion process (noise → data)

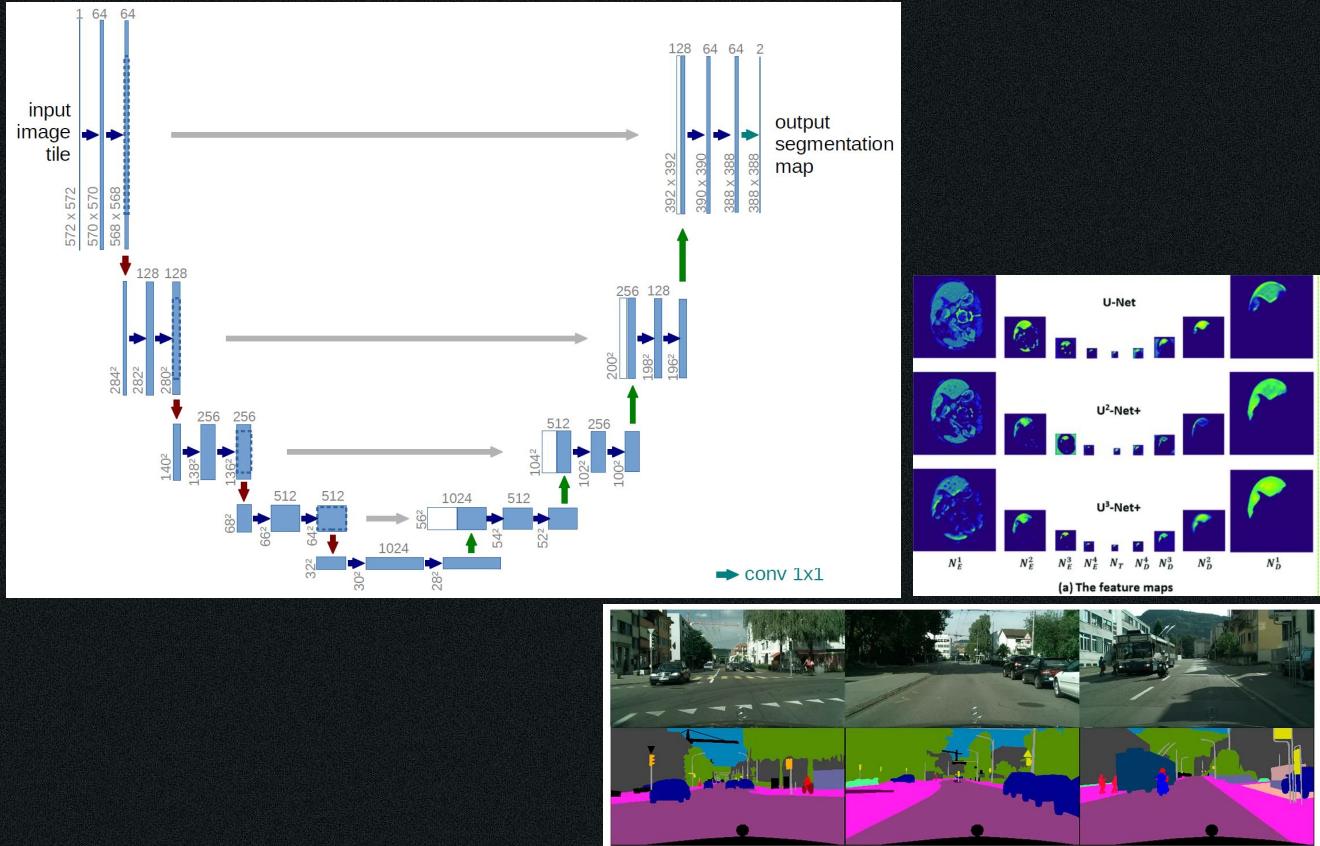
Origins of Diffusion for Image Generation



2 U-Net

Most used application - predicting segmentation maps using multi-level context.

Its convolution features and connections between them are good to learn pixel relations.



Origins of Diffusion for Image Generation

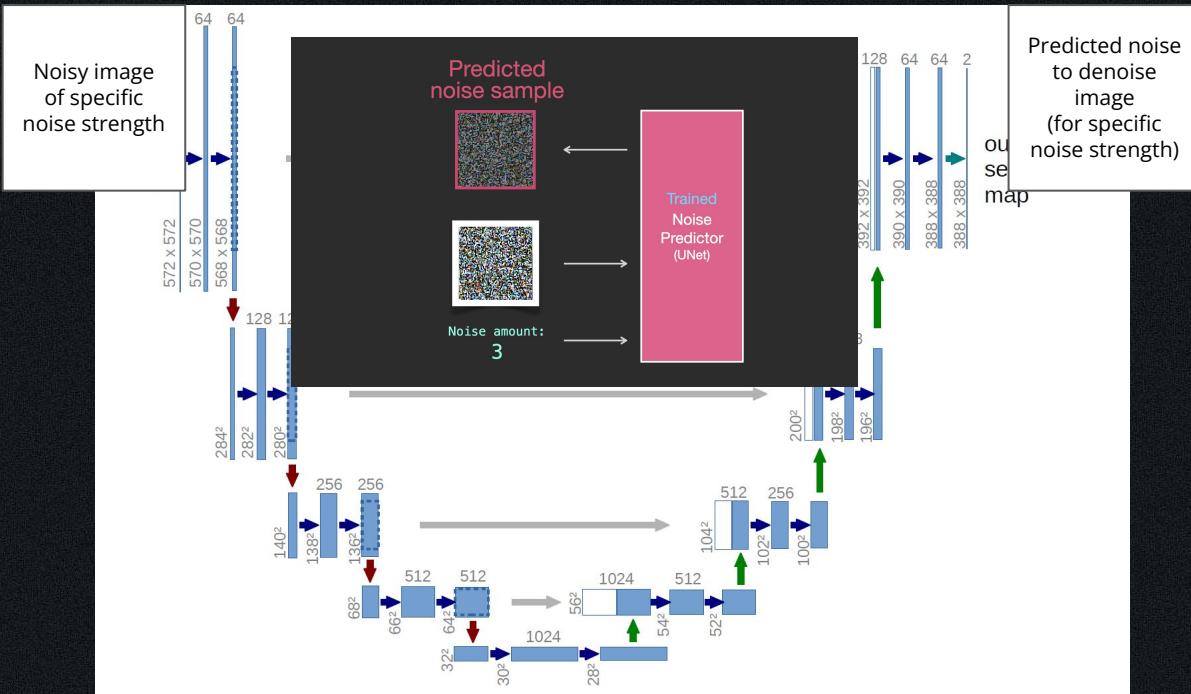
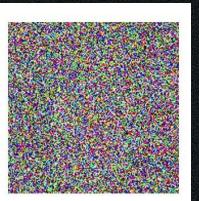


1 Diffusion Models

+ | 2 U-Net as Noise Predictor



Noisy image
of specific
noise strength



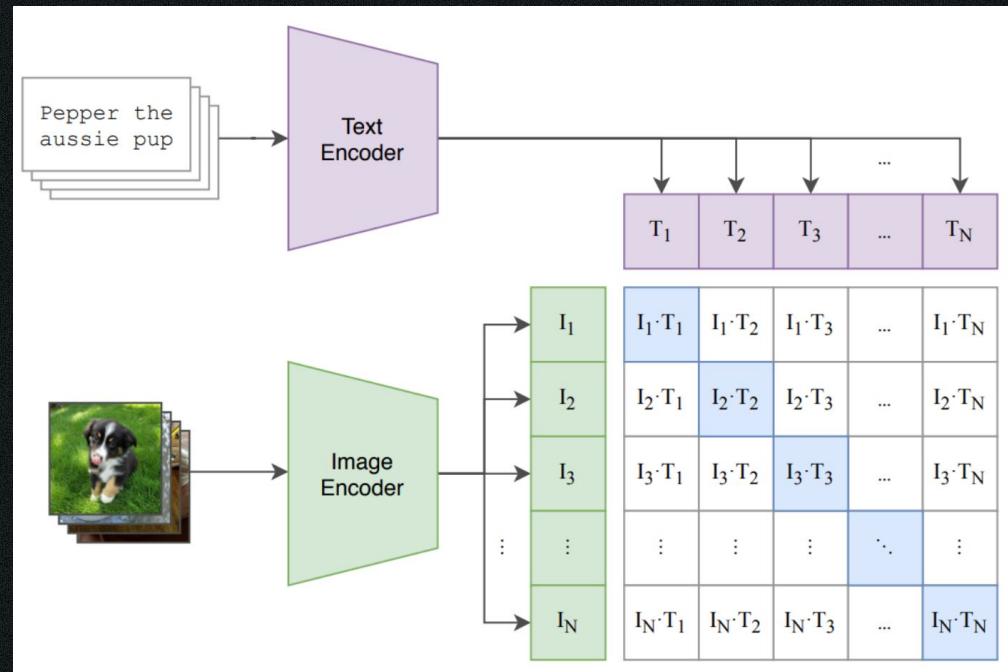
Origins of Diffusion for Image Generation



3

Encoding with cross-attention

Encode text and image together, so that the influence between text and image can be measured and used in the training → proposed by OpenAI with CLIP technique: Contrastive Language-Image Pre-training (CLIP)



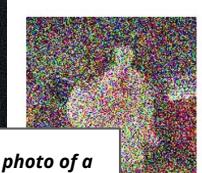
Origins of Diffusion for Image Generation



1 Diffusion Models



a photo of a
cat chilling



a photo of a
cat chilling



a photo of a
cat chilling

+

2 U-Net as Noise Predictor

Noisy image
of specific
noise strength

572²
570²
568²
568²

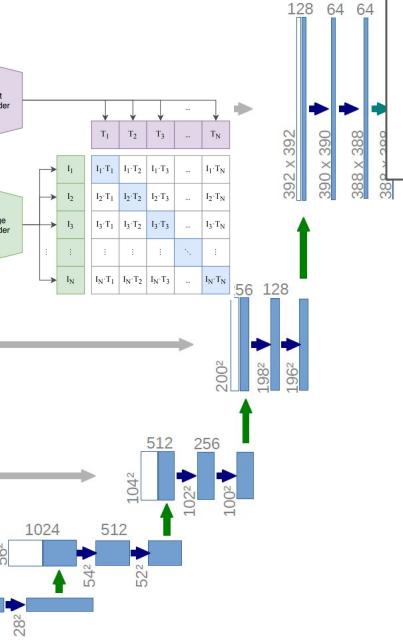
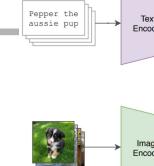
284²
282²
280²

140²
138²
136²
256²
256²

68²
66²
512²
512²

32²
30²
1024²
1024²
56²
54²
512²
104²
102²
256²
100²

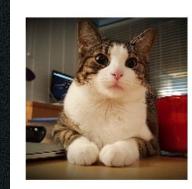
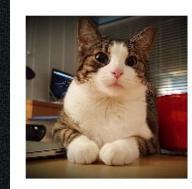
U-Net as Noise Predictor



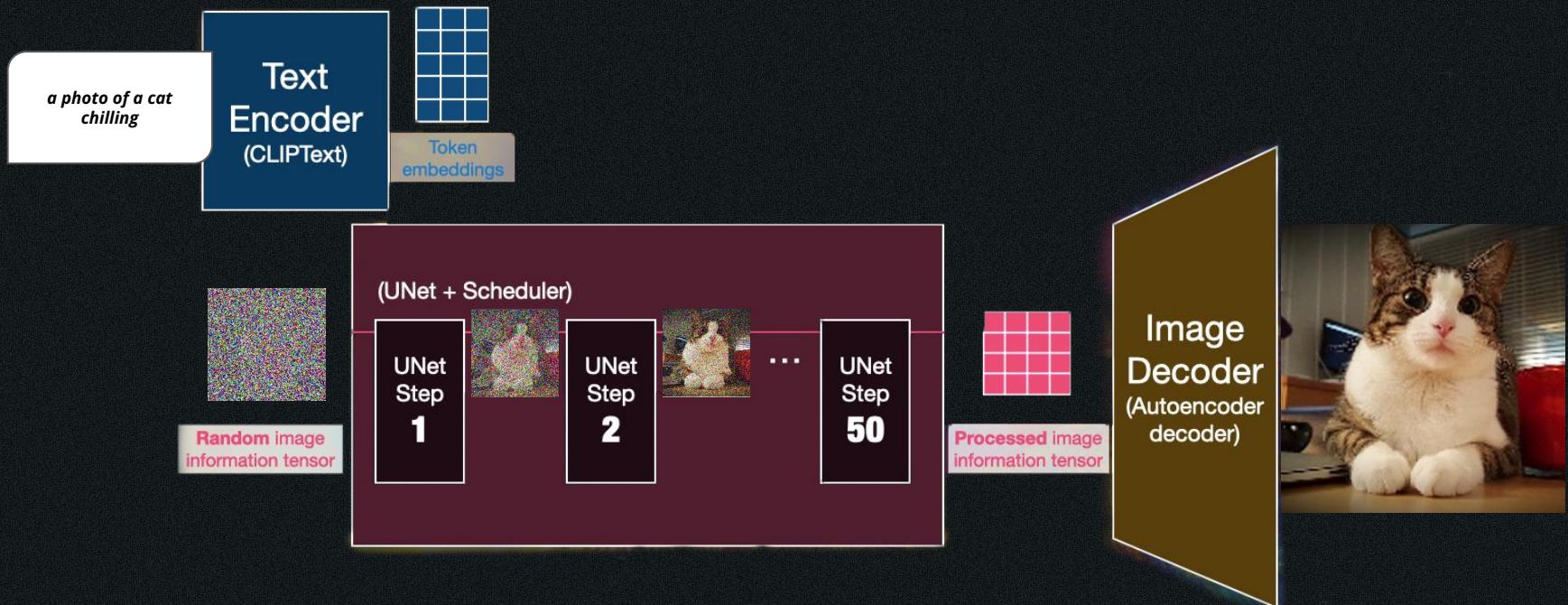
+

3 Encoding with cross-attention

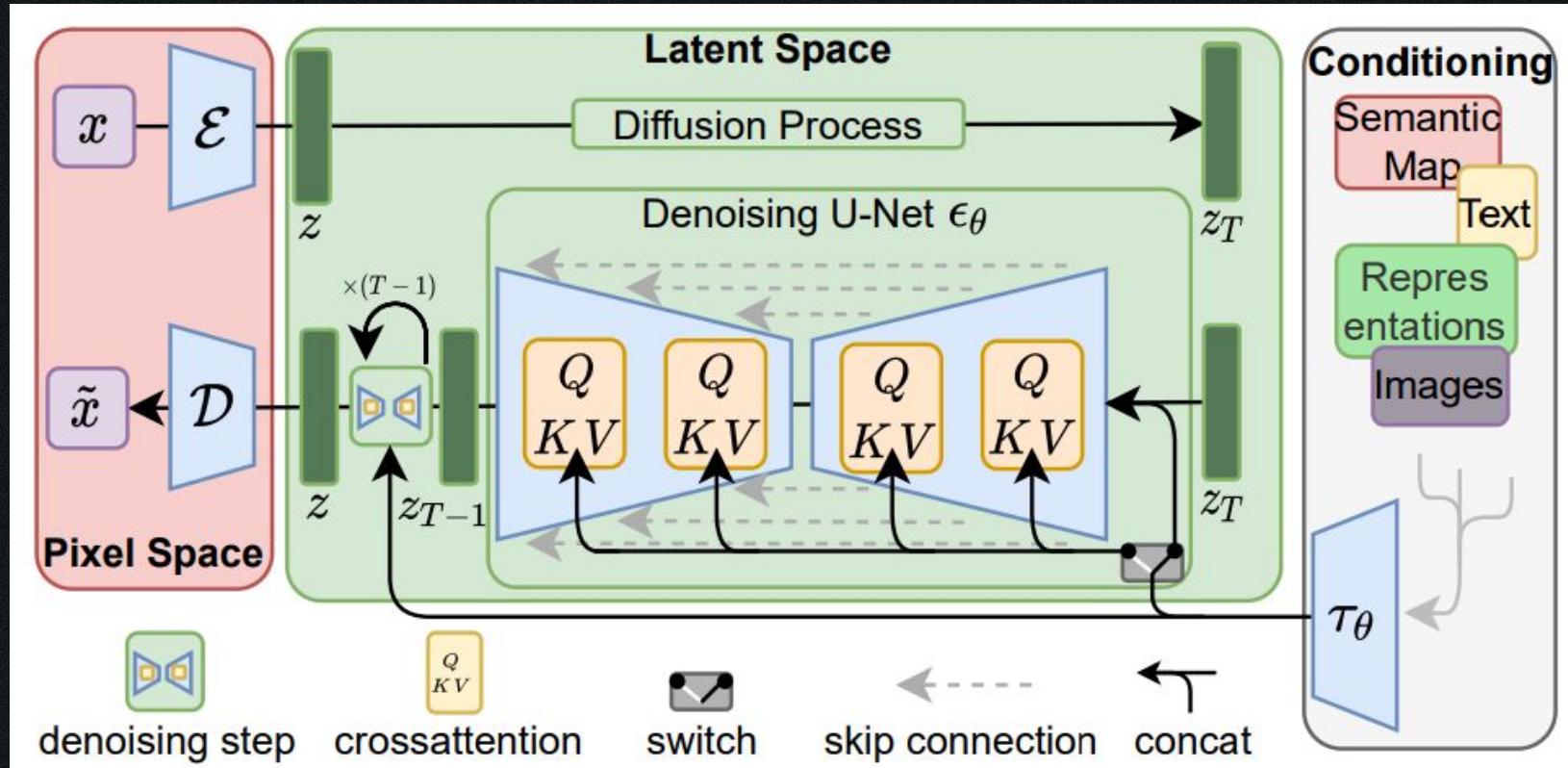
Predicted noise
to denoise
image
(for specific
noise strength)



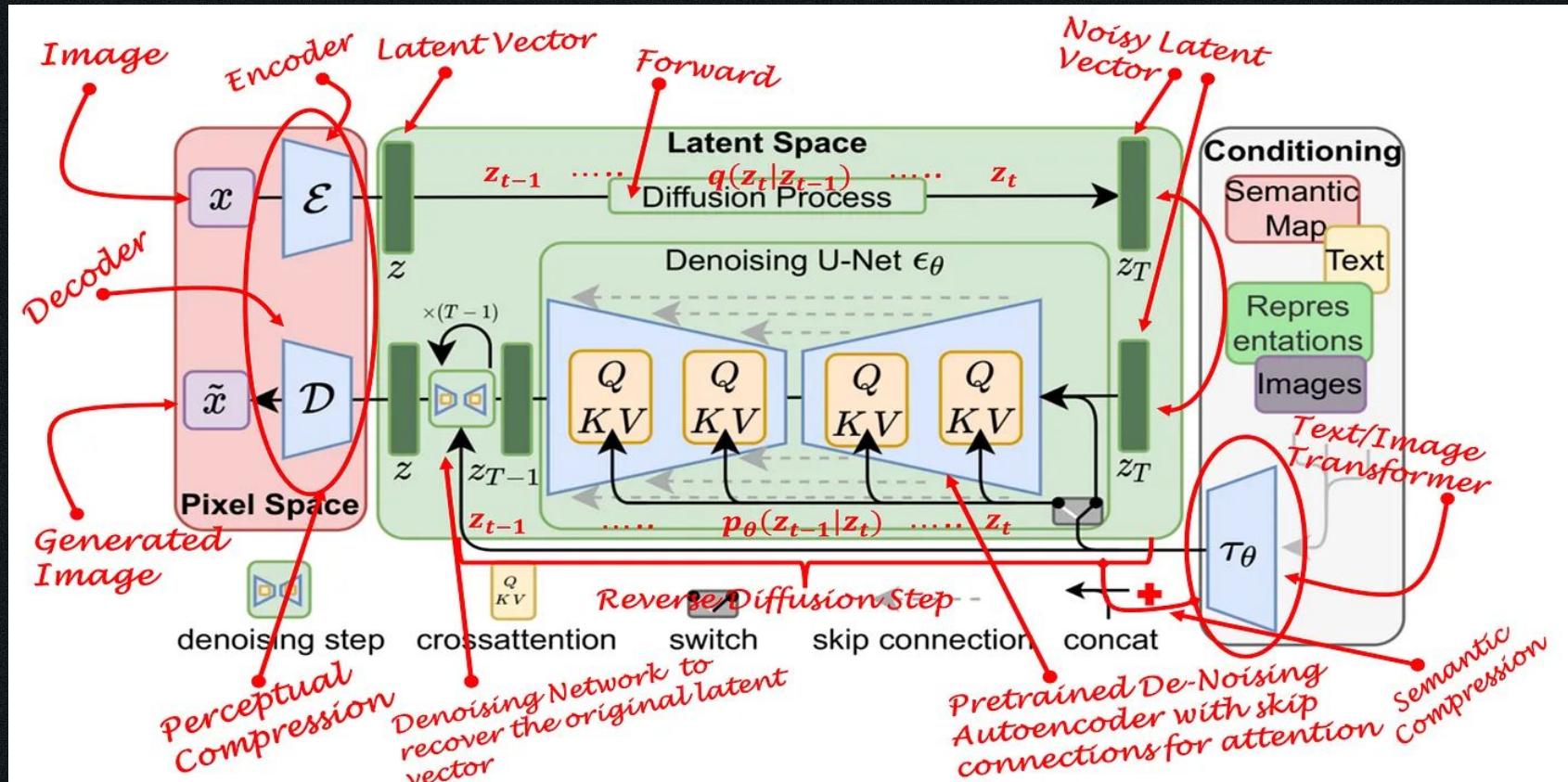
Origins of Diffusion for Image Generation



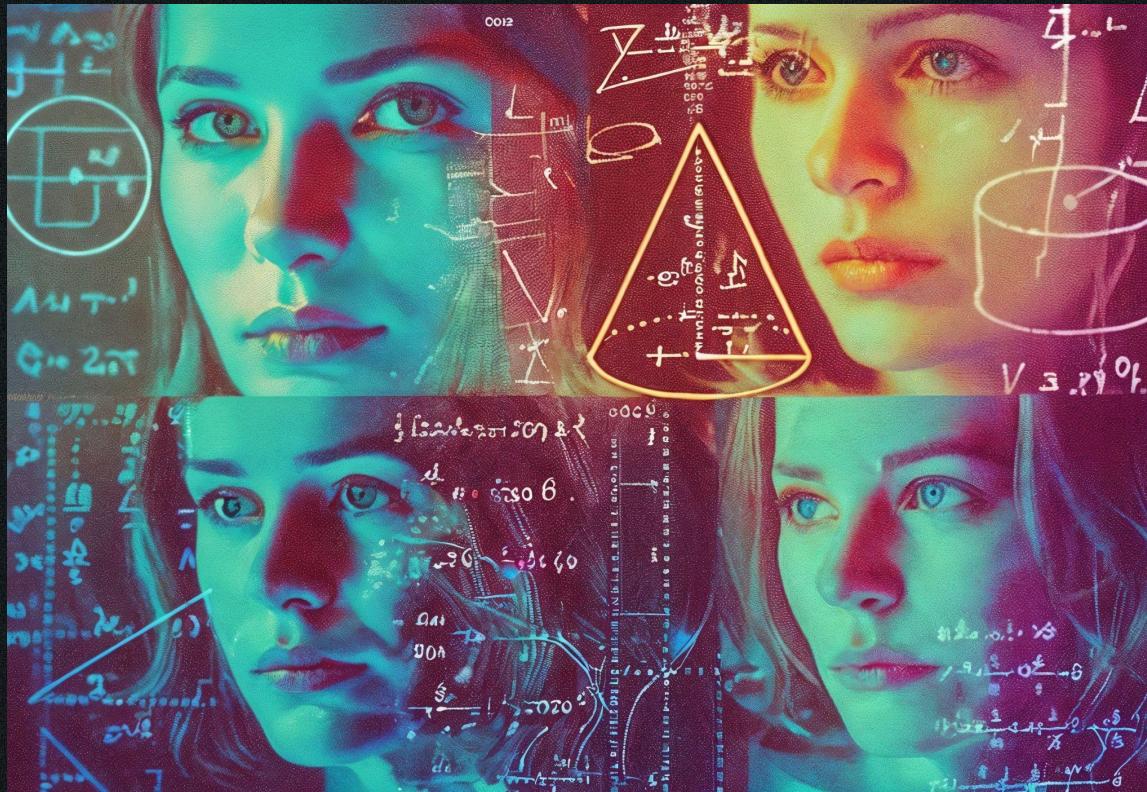
Latent Diffusion: approach behind Stable Diffusion



Latent Diffusion: approach behind Stable Diffusion



Latent Diffusion: approach behind Stable Diffusion



Data behind



- LAION-5B: dataset behind SD1.5 and SD2
- 5,85 billion CLIP-filtered image-text pairs
- SD XL and later versions of Stability AI Stable Diffusion models used internal datasets for training

Backend url:
<https://knn5.laion>

Index:
laion_5B

french cat

[Clip retrieval](#) works by converting the text query to a CLIP embedding , then using that embedding to query a knn index of clip image embedddings

Display captions Display full captions Display similarities
Safe mode Hide duplicate urls
Hide (near) duplicate images Search over [image](#)
Search with multilingual clip

french cat

french cat

french cat

french cat

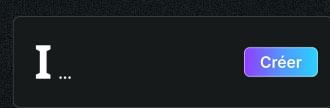
How to tell if your feline is french. He wears a b...

Hipster cat

cat in a suit Georgian sells tomatoes

イケメン猫モデル「トキ・ナントケット」がかっこいい--NAVERまとめ

Different ways of generating images



Text to image

Image to image

Inpainting

Outpainting

Upscaling & enhancing

Image to image



The screenshot shows the Seelab AI platform interface. At the top left is the logo "seelab BETA". The top navigation bar includes links for Home, Image creation (which is the active tab), Collections, Editor, and Model builder. On the far right, there's a user icon and a "M.C." button.

Image input

Image input

Drag and drop an image or select a generated image

Image output

Image count: 3

Image size: 1024x1024

W: [Slider]

H: [Slider]

Styles

content type, lens, season, daytime, color

San AndrlA

National geographic award winning breathtaking stunning photography of a unicorn

Colors Negative prompt

Let's create

Generated on April 17th, 2024
une belle cuisine moderne, colorful

Three generated images of a modern kitchen with vibrant colors (cyan, magenta, yellow) and tropical elements like palm trees and flowers.

Generated on April 17th, 2024
une belle cuisine moderne, colorful

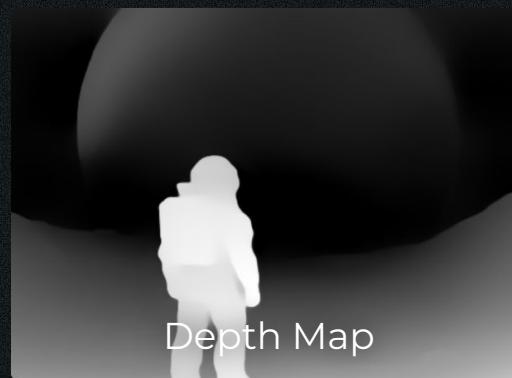
Three generated images of a modern kitchen with vibrant colors (cyan, magenta, yellow) and tropical elements like palm trees and flowers.

A circular navigation button with a speech bubble icon is located at the bottom right.

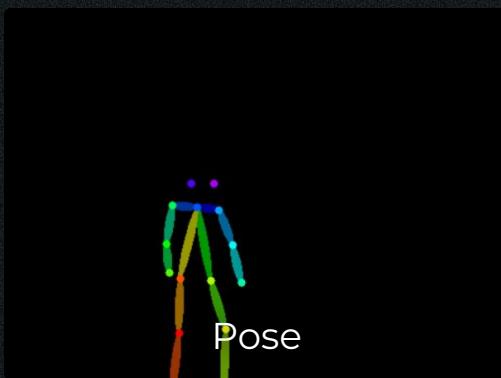
Image to image with controlNet Conditioning



Original



Depth Map



Scribble

A scientist in a high-tech suit, on a glowing volcanic Venus landscape, black hole visible in the sky.

ControlNet Conditioning examples



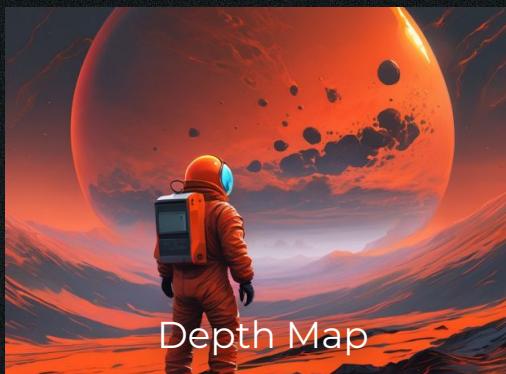
Original



Canny Edge



Line Art



Depth Map



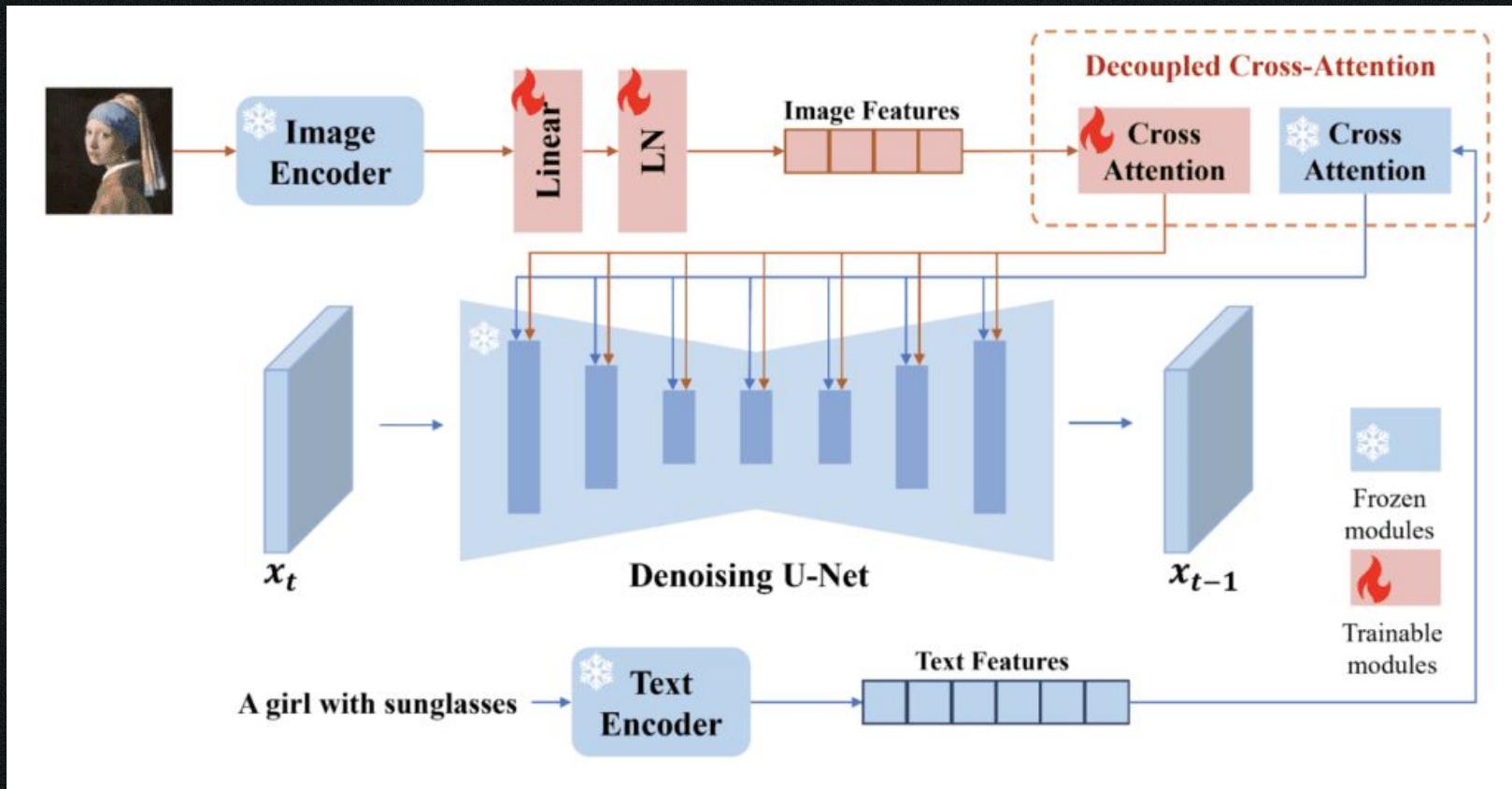
Pose



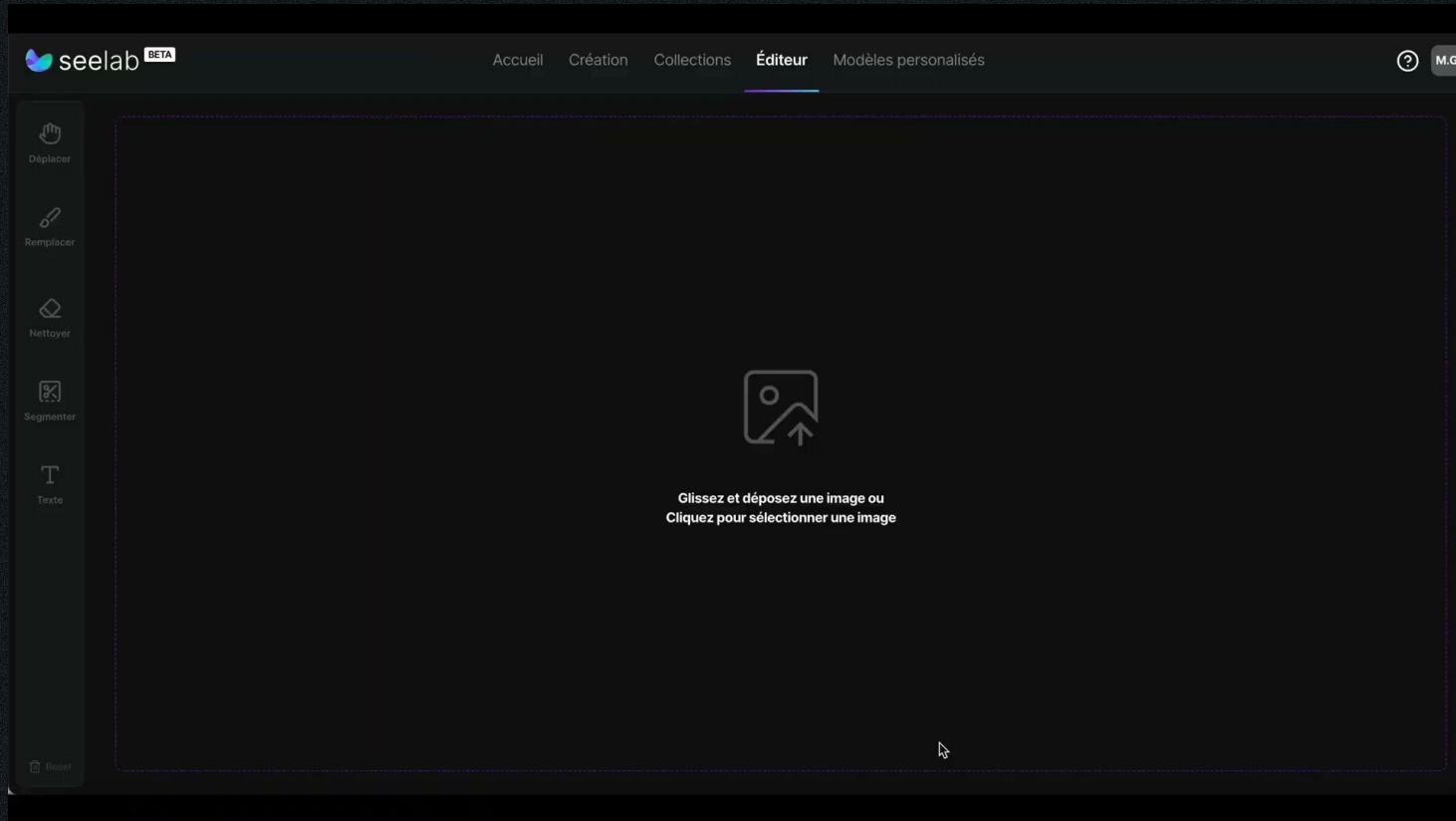
Scribble

A scientist in a high-tech suit, on a glowing volcanic Venus landscape, black hole visible in the sky.

Image to image with controlNet Conditioning



Inpainting



A screenshot of the seelab Inpainting interface. The top navigation bar includes the seelab logo (BETA), Accueil, Création, Collections, Éditeur (which is underlined in blue), and Modèles personnalisés. On the right, there's a user profile icon with "M.G". The left sidebar contains icons for Déplacer, Remplacer, Nettoyer, Segmenter, Texte, and a Reset button at the bottom. The main workspace features a large dashed purple rectangular area for inpainting. In the center is a placeholder icon with a camera and an upward arrow, accompanied by the text "Glissez et déposez une image ou Cliquez pour sélectionner une image". A cursor arrow is visible at the bottom center of the workspace.

Outpainting



"Girl with a pearl earring", Vermeer, 1665

Upscaling & SD-based enhancement



A screenshot of the MAGNIFIC AI upscaling interface. On the left, there's a sidebar with a "Prompt" input field containing "Write your prompt here...", five sliders for "Creativity" (-3), "HDR" (-3), "Resemblance" (8), "Fractality" (0), and an "Engine" dropdown set to "Automatic". Below these is a large orange "Upscale" button. At the bottom, it says "Final size: 2432 x 1664" and "This will cost ⚡ 5". The main area shows a black and white photograph of a horse from the chest up, standing in a field. To the left of the horse is a "Before" button, and to the right is an "After" button. A zoom control at the bottom right indicates "Zoom = [Z] + [wheel]." The background of the interface is dark.

The secret tool of all AI artists :) Soon on Seelab

How does image fine-tuning work?

What is a fine-tuned image model?



*Our unicorn generated with a finetuned
"Shift hackathon" model with Seelab*

"A fine-tuned image model is a deep learning model optimized for specific tasks by retraining on a narrower dataset after broad initial training.

This process sharpens its performance on targeted tasks like image classification, object detection, and style transfer, leveraging existing knowledge efficiently "

by chat GPT ;)

Fine Tuned model examples



“A chocolate labrador with sunglasses in an office setting”

While transferring the style, a fine tuned model will also learn from other component of the dataset, such as:

- pose
- environment
- view points etc...

Different approaches of image model personalization



Textual Inversion



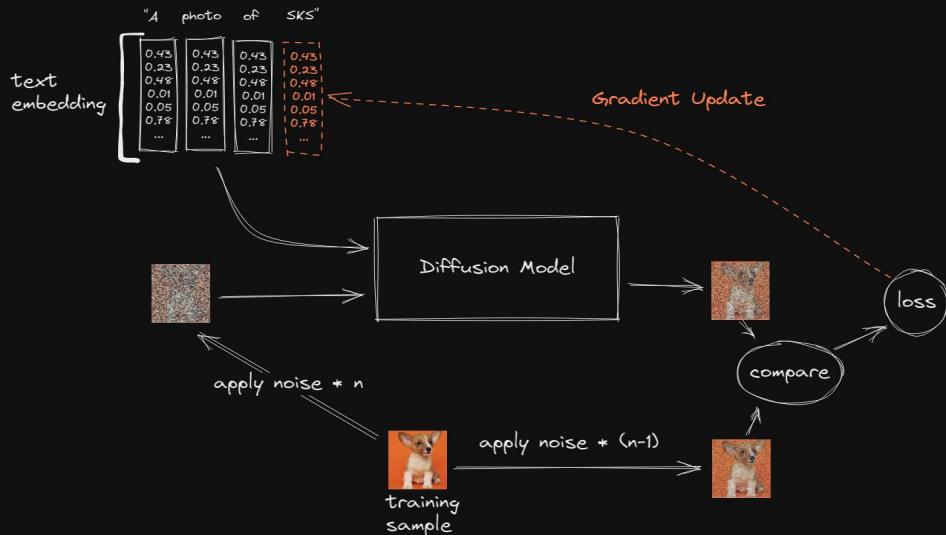
Dreambooth



LoRA



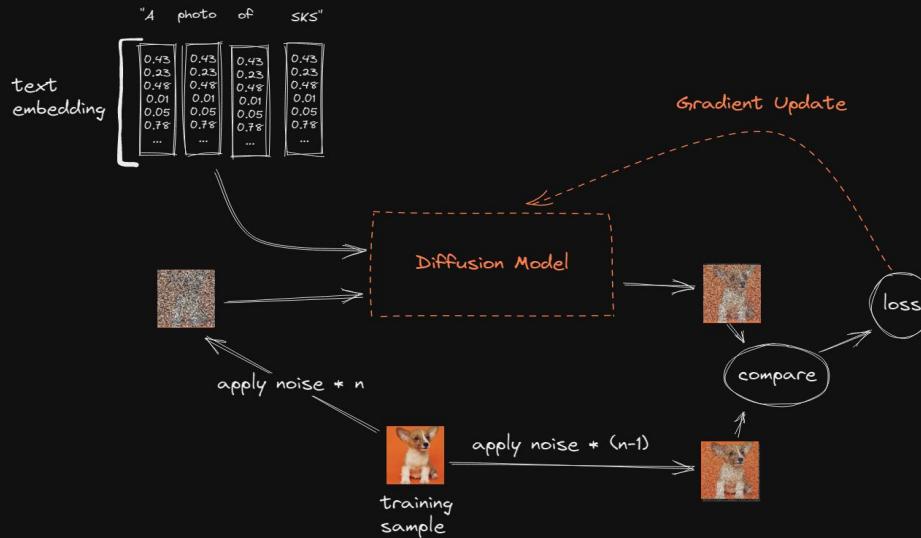
Textual Inversion



Create a special word embedding which captures the new concept

- + Output is a tiny embedding
- Quality is usually not the best

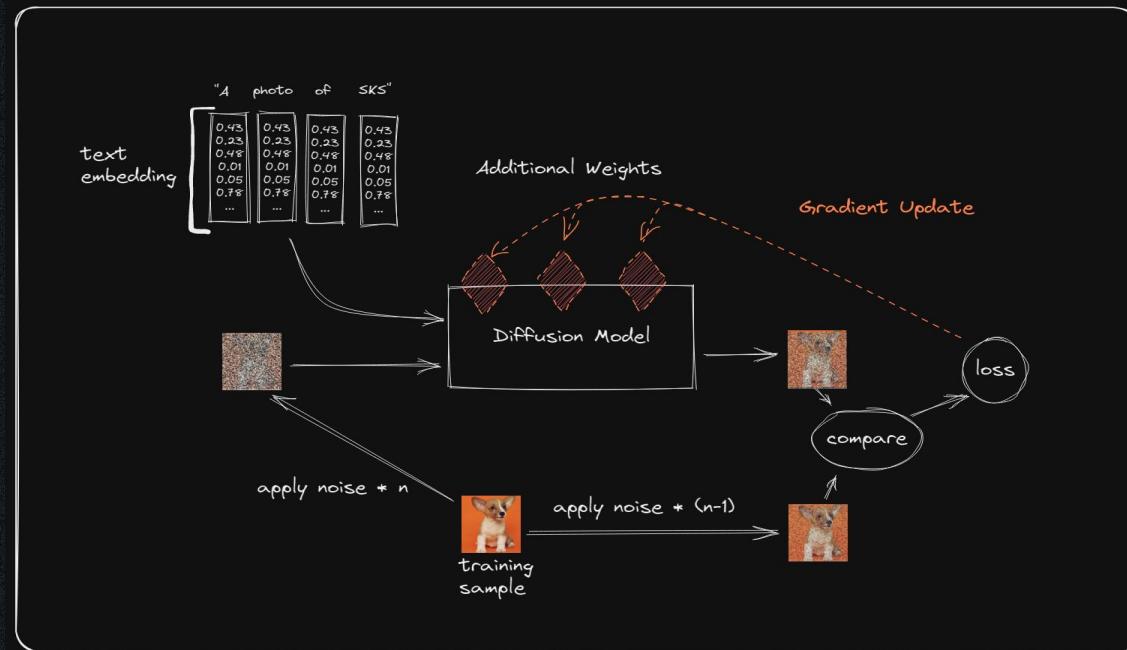
Dreambooth



Fine-tune the model itself until it understands the new concept → creating a new model checkpoint

- + The most effective
- Inefficient in storage
(and cost)

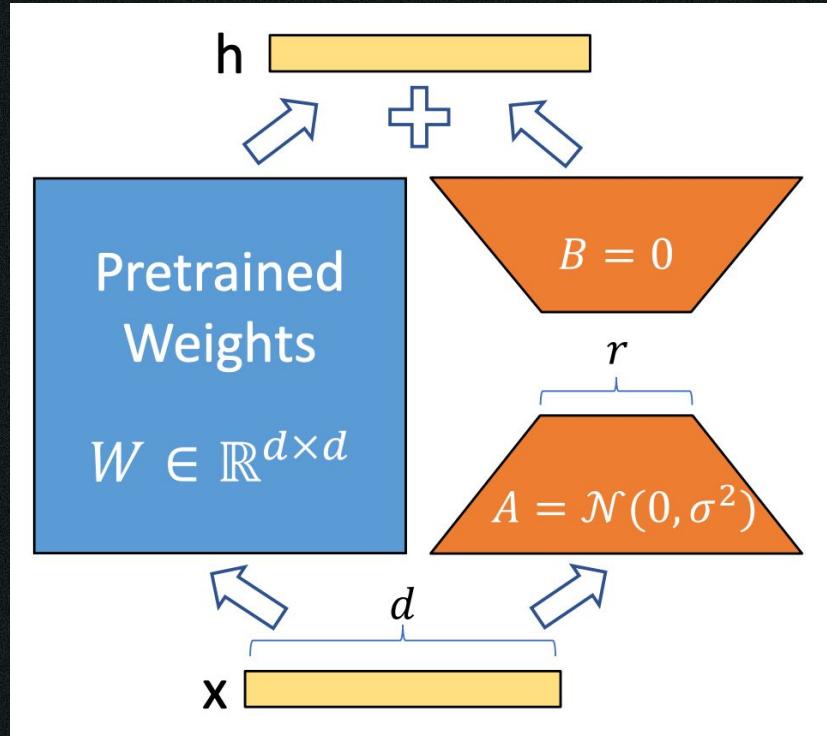
LoRa: Low Rank Adaptation



Add small update weights on top of an existing diffusion model and train only those weights to make the model understand the concept

- + Quick to train
- + Cheap
- + Good quality

LoRA: Low Rank Adaptation

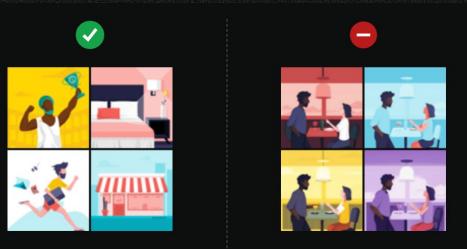


- Approach coming from LLM domain
- Training only decomposition update matrices instead of full weights
- Less parameters to train → less training resources needed
- Base model frozen → trained models can be easily swapped on top
- Smallest LoRA can be just a few Mb. Optimal size for SD XL models: 150Mb+

Garbage in, garbage out



Data set quality is - of course - key for a good fine tuning.



Make sure to include your most common characters or style elements, but keep diversity:

- > No duplicate images
- > Include various people/animal/objects representations you have



No text or text-like elements on images that is not essential to your style.

Try to avoid images where the important subjects are too small.



Prefer simple compositions with not too many elements or interactions.

Customer side



Screenshot of the Seelab AI model builder interface, showing the "Modèles personnalisés" (Custom Models) section.

The interface includes a navigation bar with links: accueil, création, collections, éditeur, and modèles personnalisés (which is underlined). There are also icons for a globe and user profile.

The main area displays a step-by-step process:

- 1 Informations sur le modèle** (Step 1: Model information)
- Étape suivante >** (Next step)

The "Nom du modèle" (Model name) field is active, with a placeholder "Entrez le nom de votre modèle".

A section titled "Choisissez le type de modèle que vous souhaitez générer" (Choose the type of model you want to generate) shows five preview cards:

- A portrait of a man with a beard.
- A person standing next to a red car against a colorful, abstract background.
- A modern, curved white building set in a landscape.
- A green bottle of shampoo labeled "Bientôt dispo" (Coming soon).
- A circular graphic featuring a sunset over mountains, also labeled "Bientôt dispo".

“Style” training example: Shift



Dataset



Captioning example

Digital psychedelic photography. Rendered on a high-resolution screen, a person in profile sits at a vintage computer setup, focusing intently on the screen while wearing a yellow cap and glasses, bathed in vibrant reds and blues. Their environment is a brightly colored room with large windows, casting dynamic shadows. The vibe is nostalgic and contemplative, with a vivid palette that gives a dreamy, retro ambiance.

Prompt token & decoration

Digital psychedelic photography. \${prompt}

Big thanks to Simon Timssale-Bourrioux, AI artist and trainer who created this model

<https://www.linkedin.com/in/simon-timssale-bourrioux-746a1aa5/>

“Style” training example: Shift



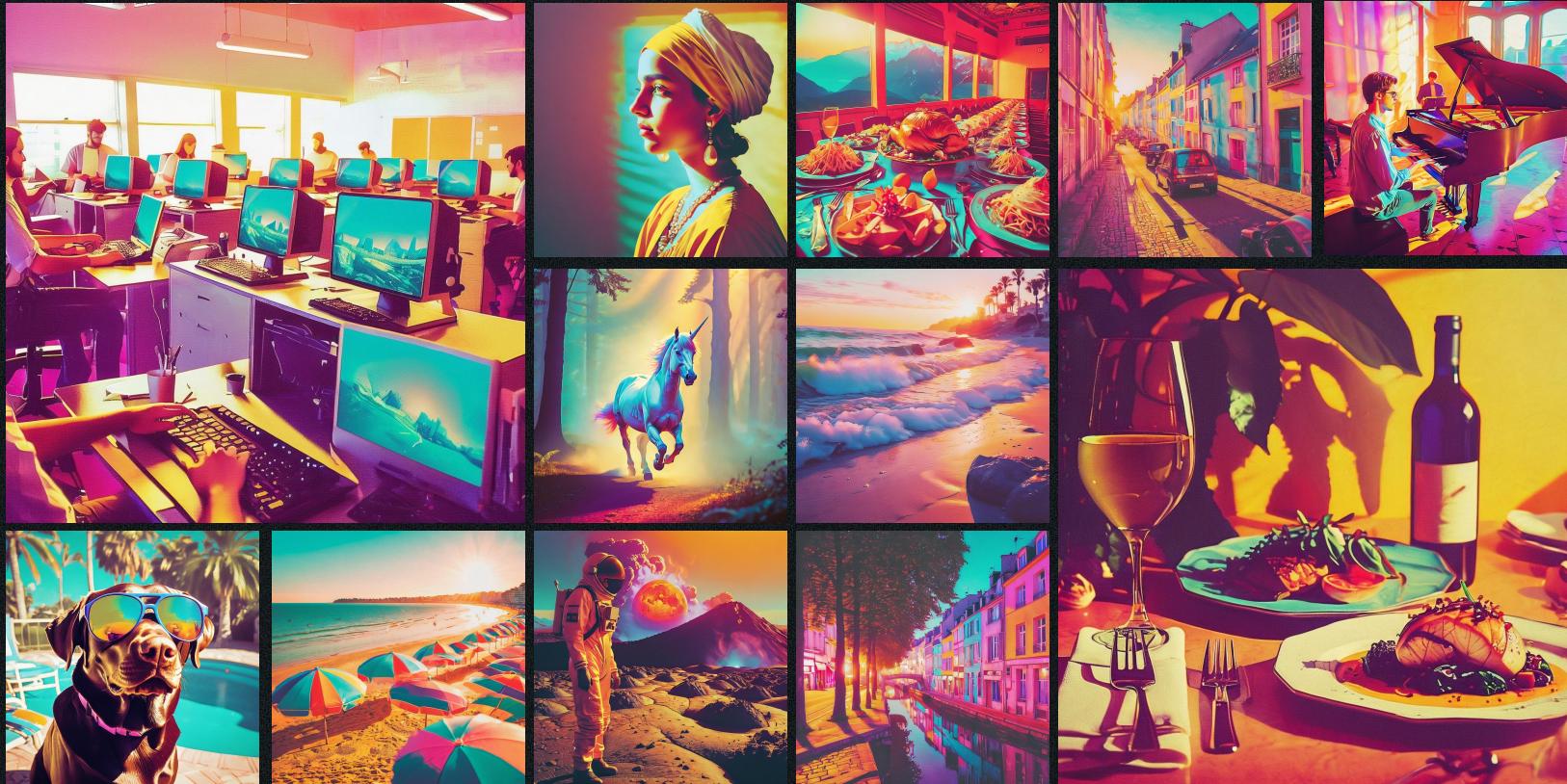
Dataset



Generated images with prompts



“Style” training example: Shift



“Avatar” training example: Anne Kerdi



Dataset



Captioning example

A medium-format camera captures Anne Kerdi, elbows on table, head resting on hands, wearing a casual white t-shirt. She is viewed from the front. The outdoor café setting basks in sunlight, casting soft shadows and creating a serene mood. The image is vivid with earthy and warm tones, showcasing crisp detail and high resolution. No text is present.

Prompt token & decoration

Anne Kerdi, \${prompt}

Big thanks to Sébastien Keranvran, creator of Anne Kerdi, the first Bretonne AI influencer

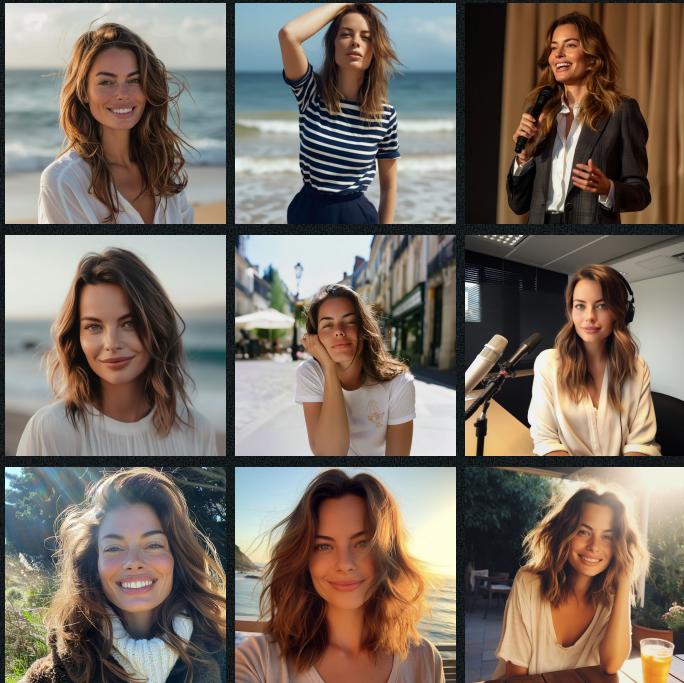


<https://www.instagram.com/annekerdi/>

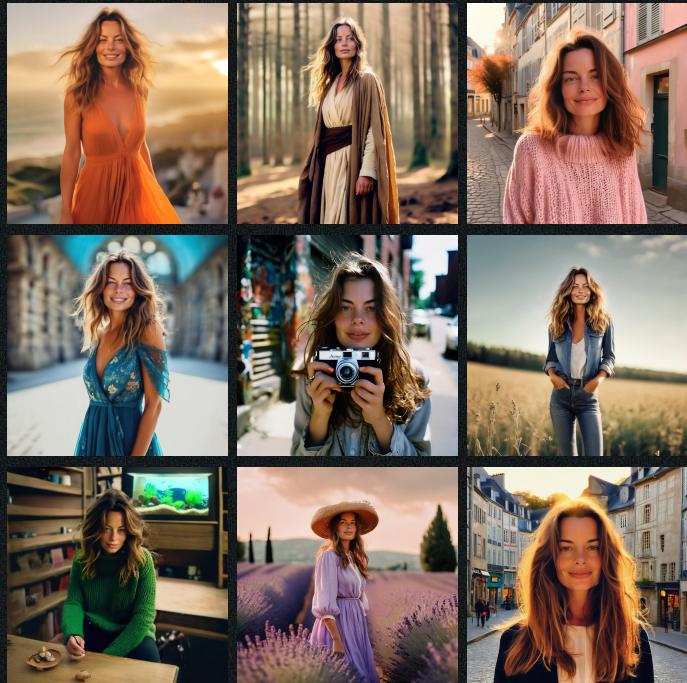
“Avatar” training example: Anne Kerdi



Dataset



Generated images with prompts



“Avatar” training example: Anne Kerdi



A vibrant, colorful photograph of a desert landscape under a clear blue sky. A two-lane road with yellow center and edge lines curves through the scene. The ground is covered in small, rounded, reddish-orange shrubs. In the background, rolling hills and mountains are visible, with the colors transitioning from red and orange in the foreground to blue and purple in the distance.

WHAT'S NEXT?

Future base models



Stable Cascade

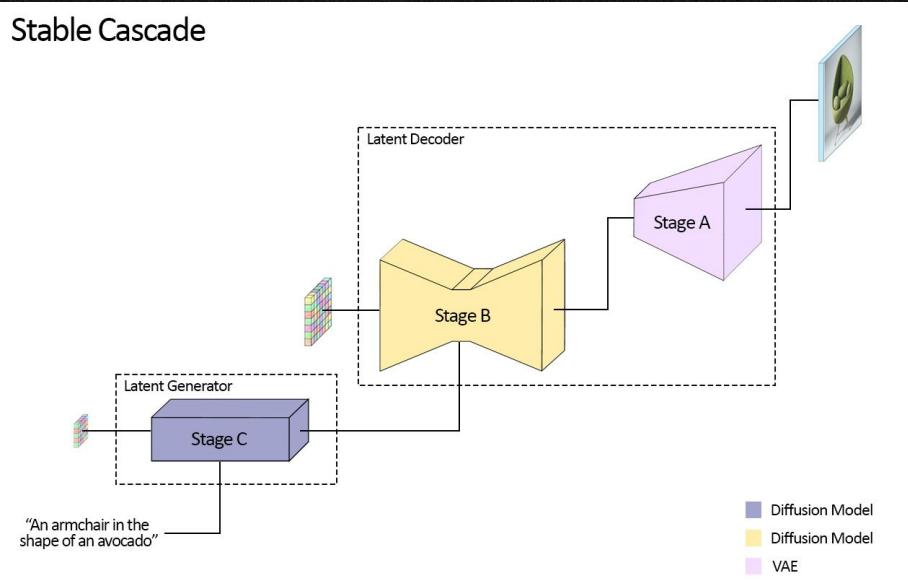


Stable Diffusion 3

Future Models

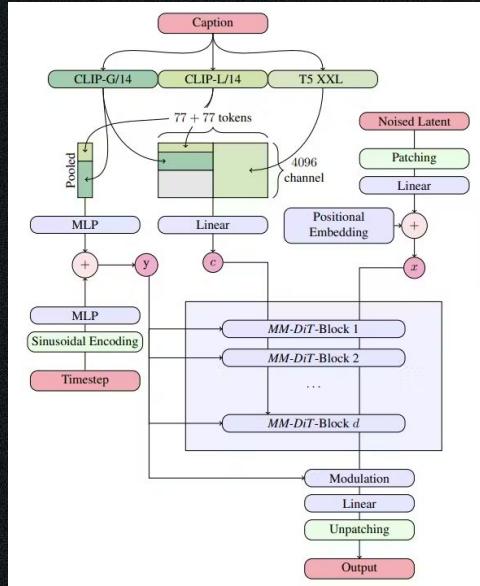


Stable Cascade



Stable Cascade

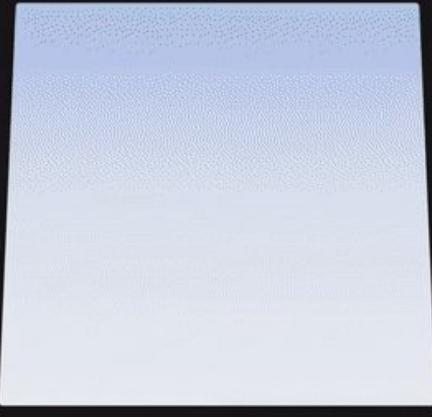
Splitting the network into three stages, to make trainings and fine-tunings more efficient, only using lighter stages



Stable Diffusion 3

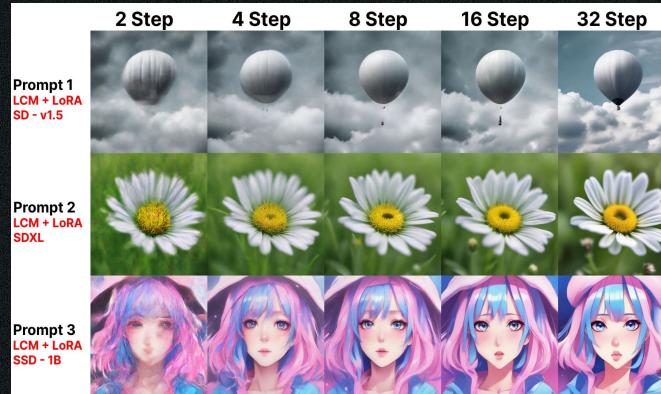
Diffusion Transformer with latent patches instead U-Net, different flow-based denoising paradigm

Fast Generation Models



a white persian cat

SD XL Turbo

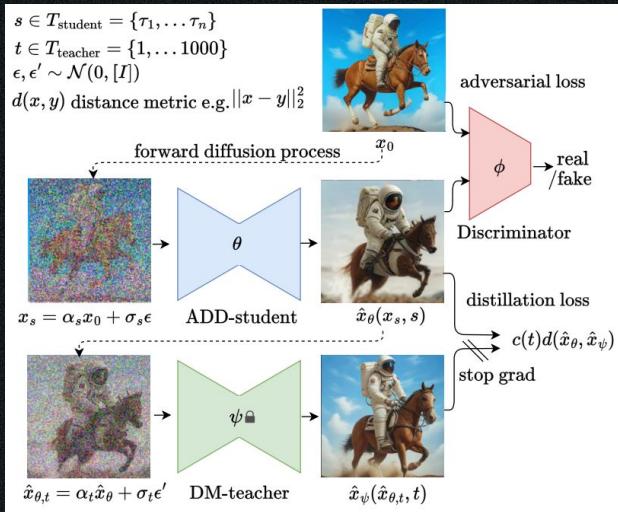


LCM: Latent
Consistency Model



SD XL Lightning

Fast Generation Models



- Fast models based on knowledge distillation approach: how to transfer knowledge of a large model to a much smaller one
- SD XL Turbo: Adversarial Distillation - a competitive approach similar to GANs
- SD XL Lightning: Progressive Adversarial Distillation - smoother iterative approach of transferring knowledge

SD XL Turbo
Adversarial Diffusion Distillation

LoRA merge



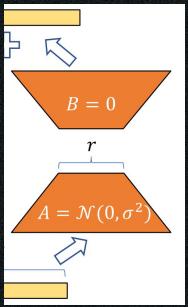
+



<Lora:anne_kerdi:1.0>

<Lora:shift:1.0>

LoRA Merging



$$A_{merged} = concat(weight_1 * scaling_1 * A_1, weight_2 * scaling_2 * A_2, dim = 0)$$

$$B_{merged} = concat(B_1, B_2, dim = 1)$$

$$B_{merged}A_{merged} = weight_1 * scaling_1 * B_1A_1 + weight_2 * scaling_2 * B_2A_2$$

Straightforward method:
weighted concatenation

Alternative methods:

- Task Arithmetic: more complex weighting of separate deltas
- SVD: singular value decomposition of the merged delta weights
- TIES: smarter merging with pre-processing and calculating agreement of weights
- DARE: another smarter merging with active pruning of weights

Merge de Lora



Anne Kerdi + Shift model merge



Anne Kerdi 90%
Shift Hackathon 10%



Anne Kerdi 50%
Shift Hackathon 50%



Anne Kerdi 20%
Shift Hackathon 80%

<Lora:anne_kerdi:0.6><Lora:shift:0.4>

Anne Kerdi on a beach in Brittany, at sunset

Lora Merge



Clay Fun 80%
Cel Shading 20%



Clay Fun 20%
Cel Shading 80%

Lora Merge

Mixing Models gives
you the opportunity
to create your

Own Unique Style !

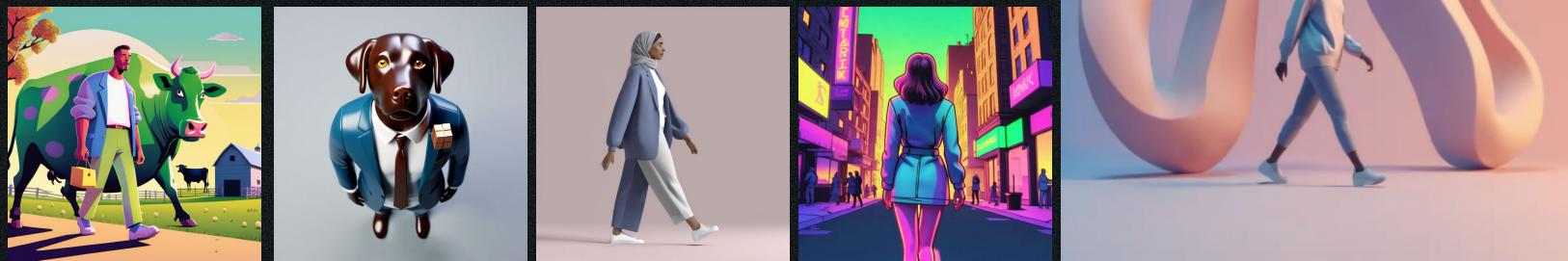
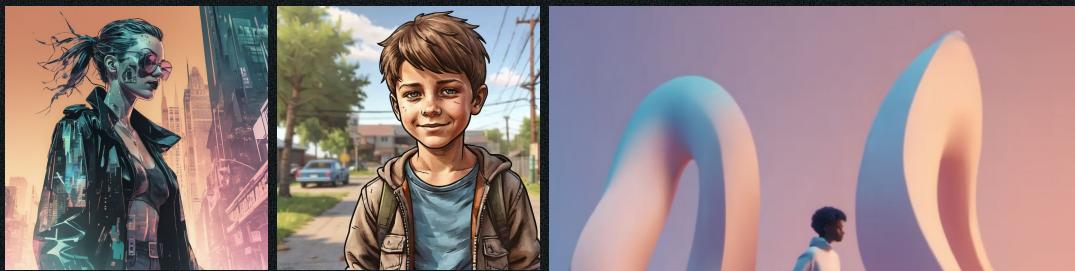
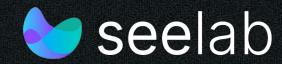


Image to video



> > >

A vertical sequence of three small white right-pointing chevrons, indicating a process or transformation.



CONCLUSION



Traceability
Transparency
Ethics

QUESTIONS... AND THEN BEER TIME!

Let's talk:

matthieu@seelab.ai

dmitry@seelab.ai

