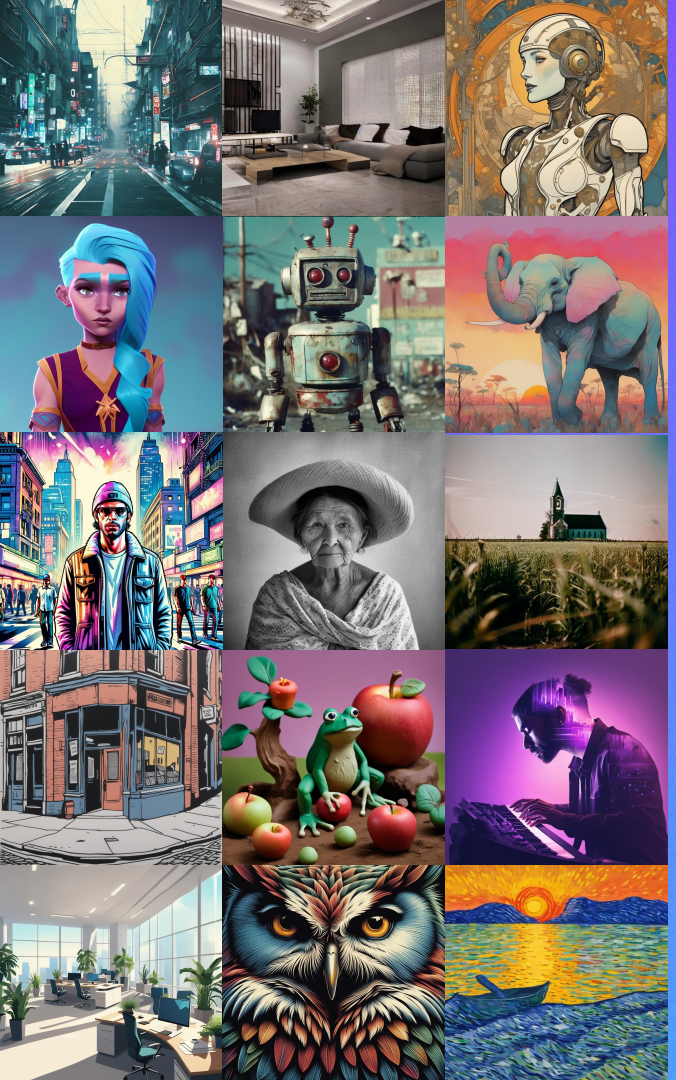


seelab

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

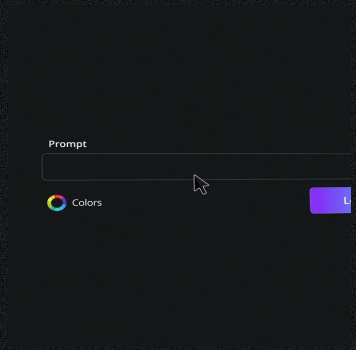
PHD in Computer vision from Rennes University

Former R&D Team Lead @ Meero

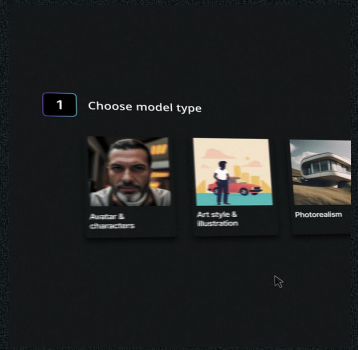
technicolor meero

Seelab.ai in a few words...

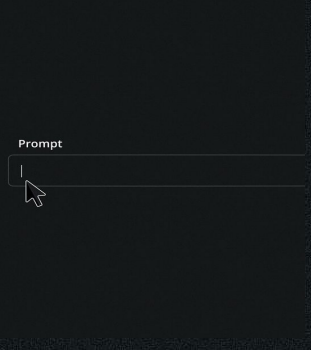
seelab



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

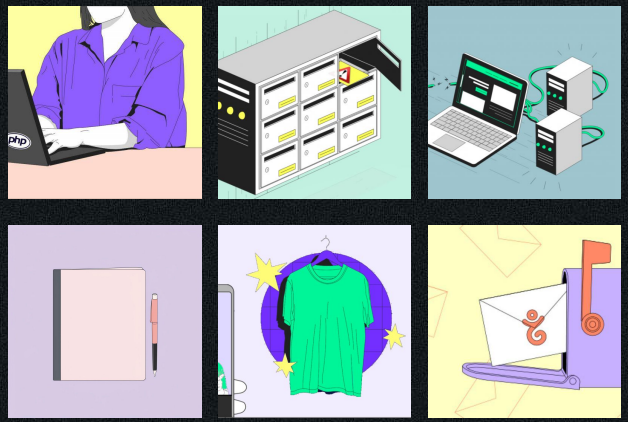
Create your own image AI

seelab

gandi.net

Usecase:
Training on an illustration style for Gandi.net

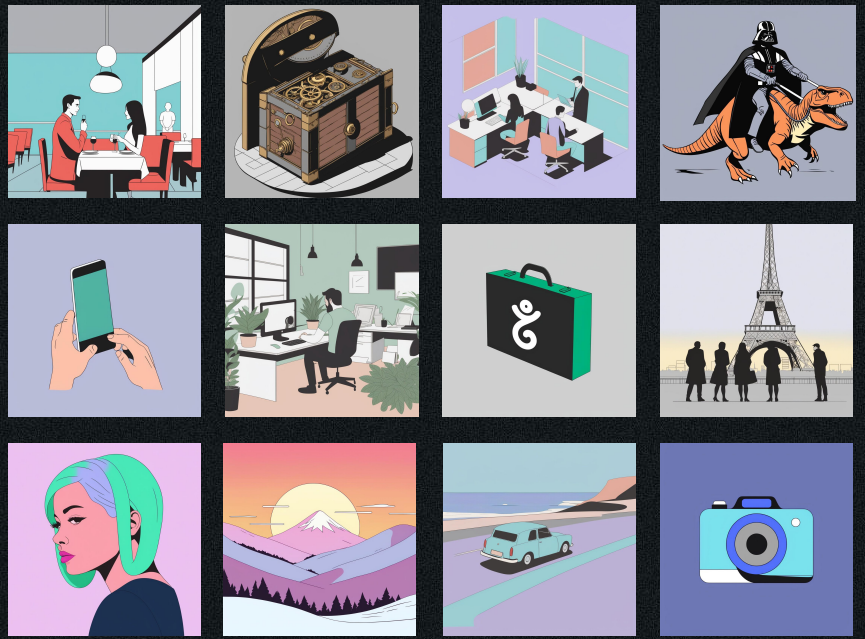
User uploads his dataset (10 images here)



Gandi 's model is created

after 40 mn of automated work

User generates images with simple prompts in a few seconds



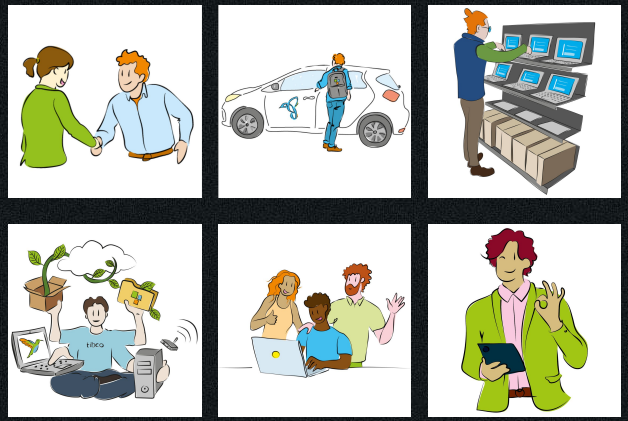
Create your own image AI

seelab

tibco

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

seelab

Usecase:
Training an avatar for France Pronos

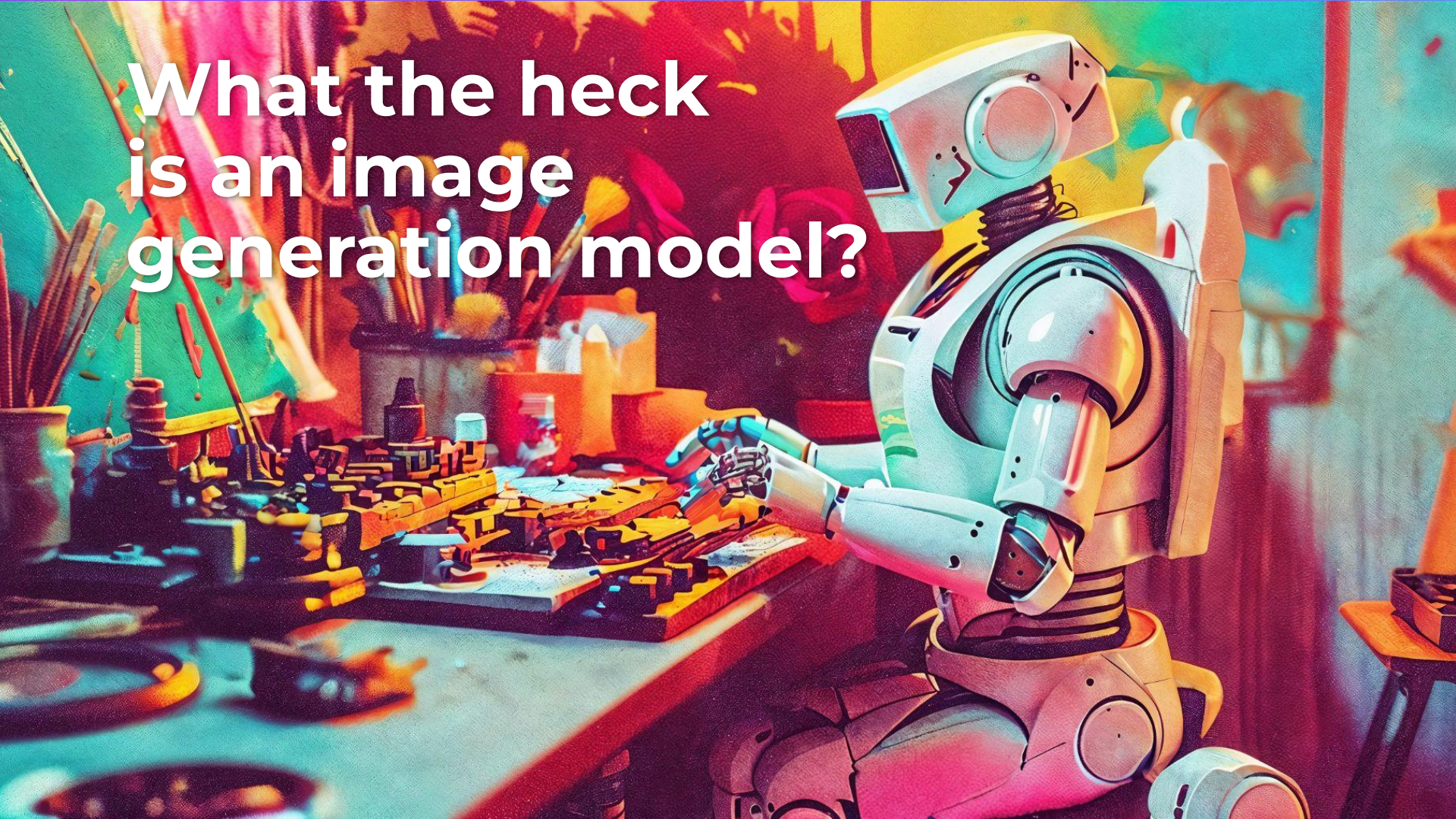
User uploads his dataset (8 images here)

France Pronos model is created

after 40 mn of automated work

User generates images with simple prompts in a few seconds

Remember “Shift Hackaton”?


What the heck
is an image
generation model?

Text-to-image

seelab

National geographic award winning breathtaking stunning photography of a unicorn

Prompt

Dall-e 3

Midjourney V6

Seelab/SD XL

“Human hand” Midjourney v1 Mars 2022

“Human hand” Midjourney v5.1 Mars 2023

Models evolve freaking quickly

seelab

“National geographic award winning breathtaking stunning photography of a unicorn”

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

seelab

Open source

stability.ai

Stable Diffusion

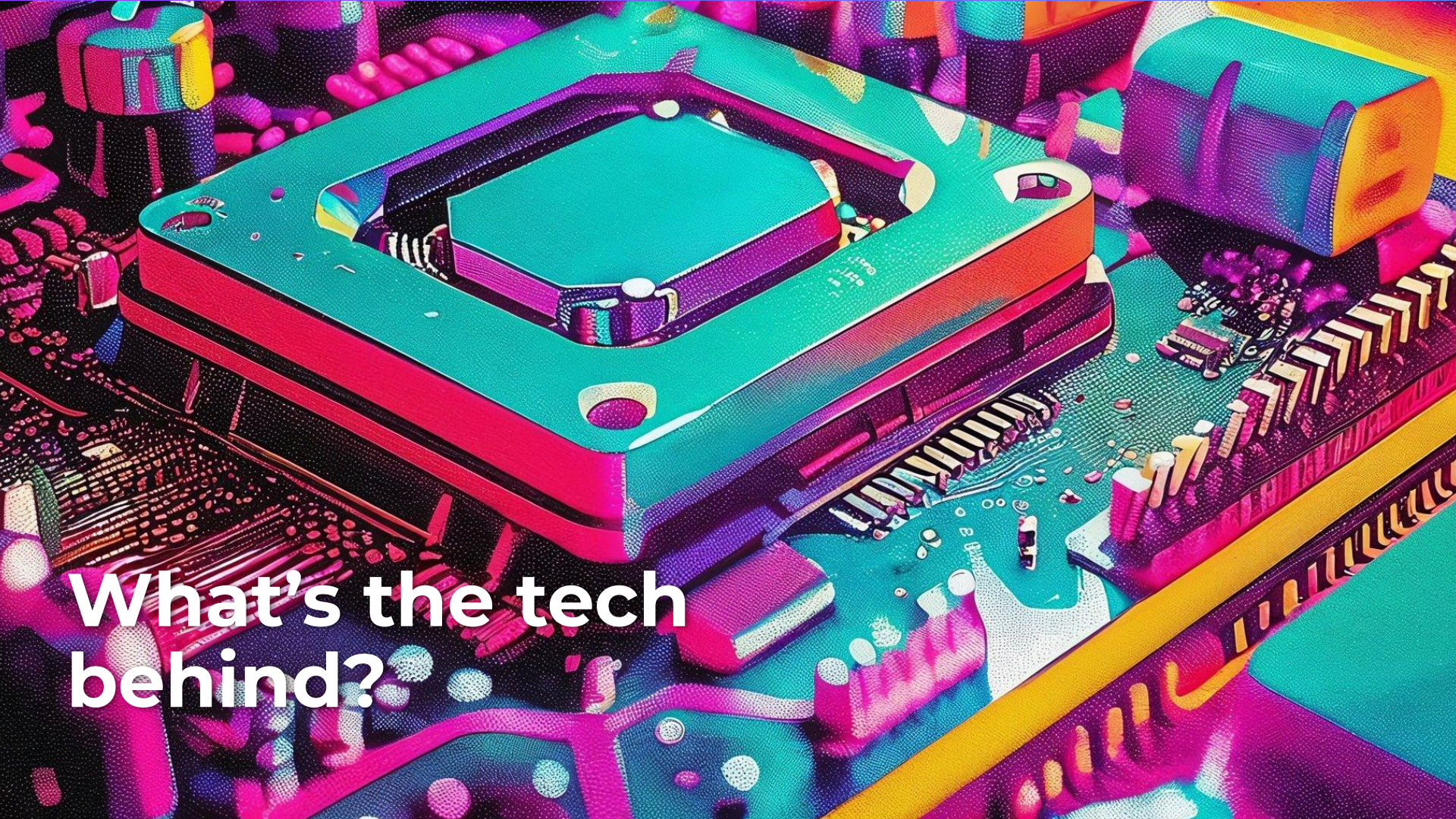
Proprietary

Midjourney

DALL·E

Meta AI

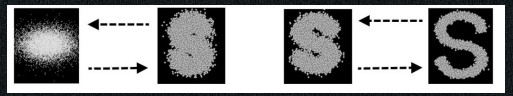
Google AI


What’s the tech behind?

Origins of Diffusion for Image Generation

seelab

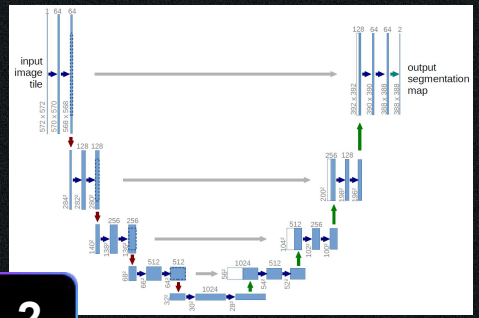
3 factors behind diffusion-based generation:



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

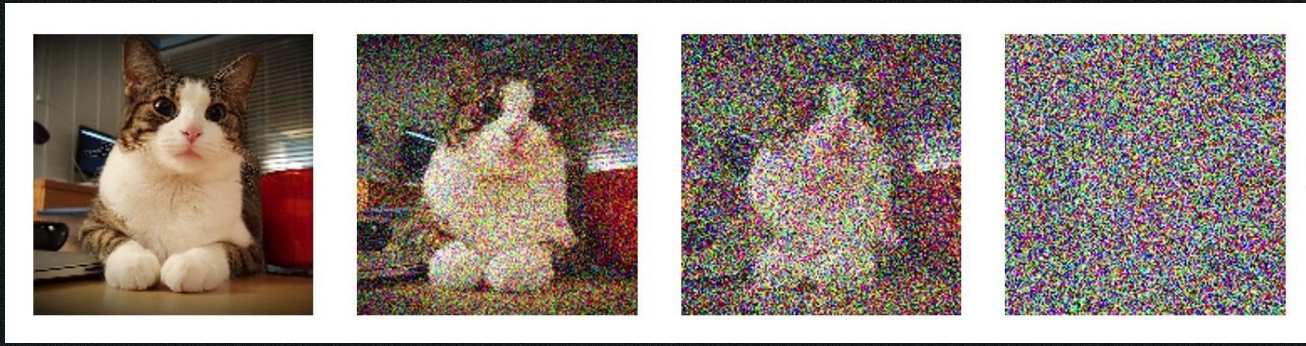
from text processing (e.g. translation): help with context understanding. It inspired the **cross-attention mechanism**, where images and text are encoded in a shared vectorial representation

Origins of Diffusion for Image Generation

seelab

1 Diffusion Models

Forward diffusion process (data → noise)



Backward diffusion process (noise → data)

https://jalammar.github.io/illustrated-stable-diffusion/

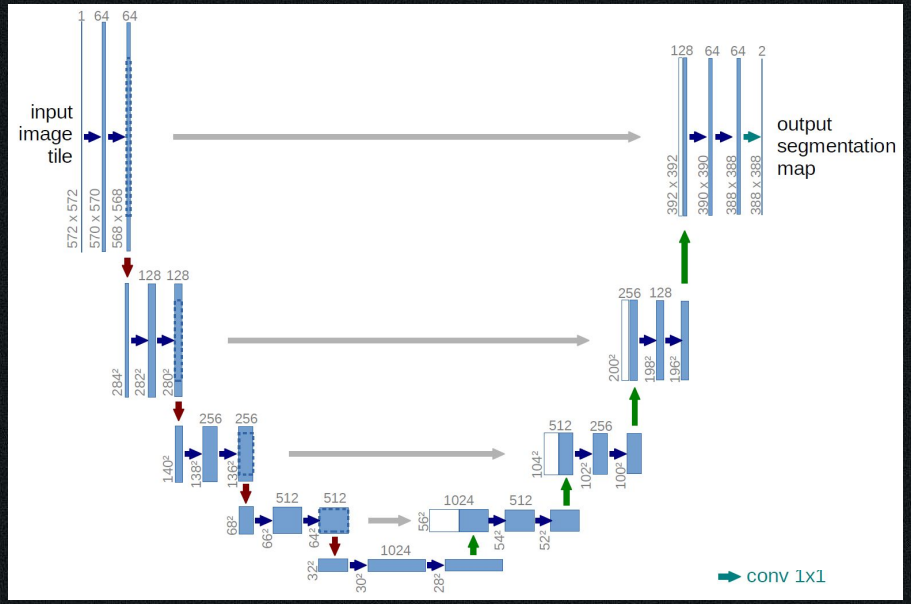
Origins of Diffusion for Image Generation

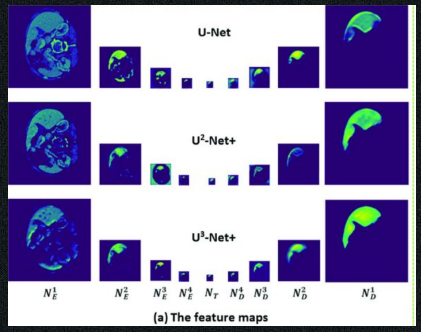
seelab

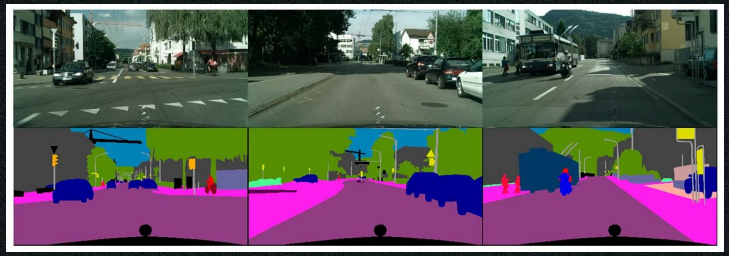
2 U-Net

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

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



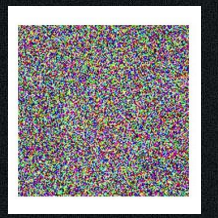




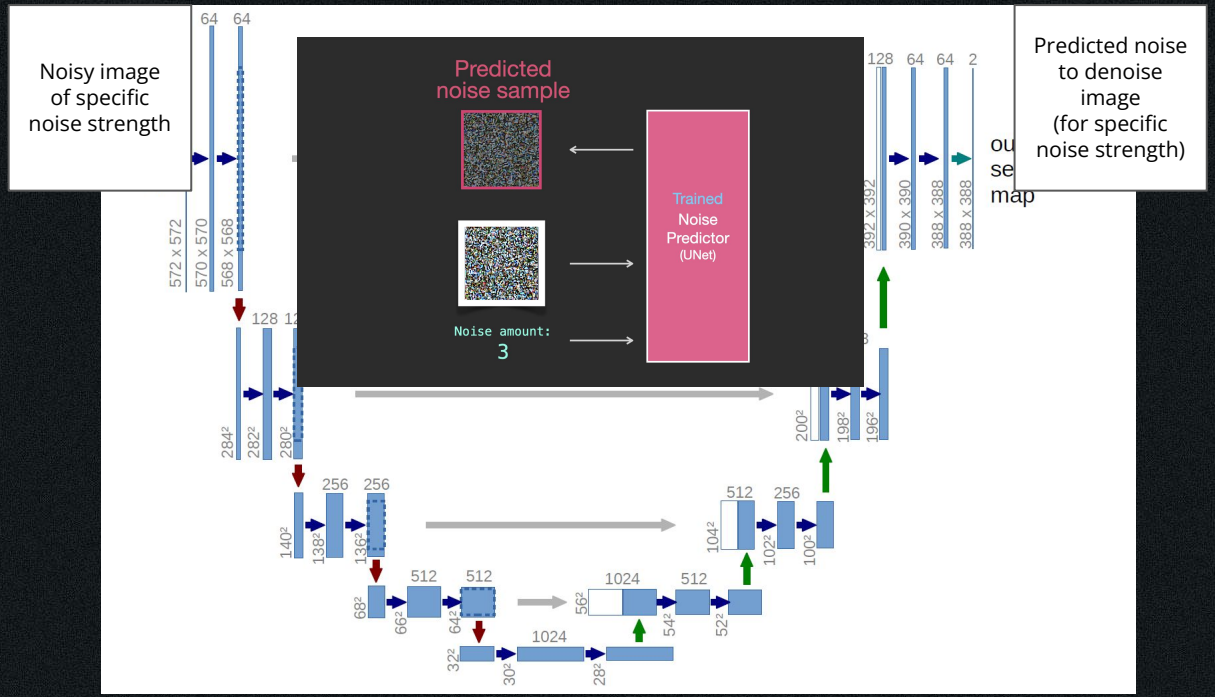
Origins of Diffusion for Image Generation

seelab

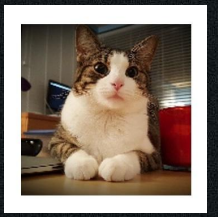
1 Diffusion Models **+** **2** U-Net as Noise Predictor

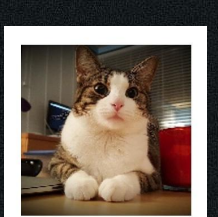


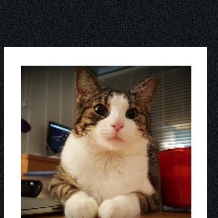
Noisy image of specific noise strength



Predicted noise to denoise image (for specific noise strength)







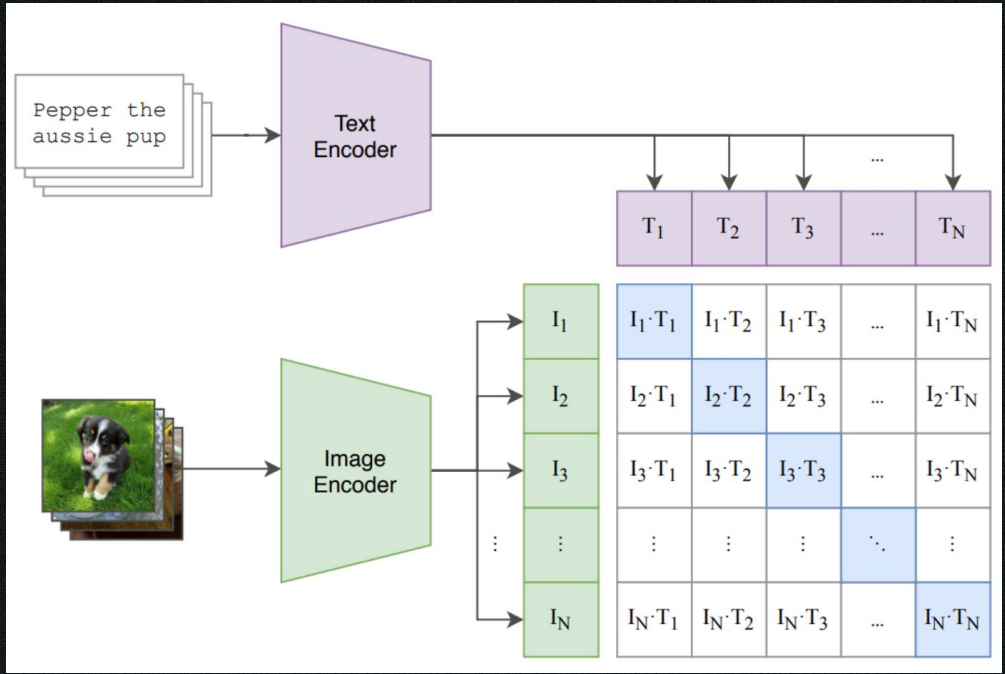
https://jalammar.github.io/il usion/

Origins of Diffusion for Image Generation

seelab

3 **Encoding with cross-attention**

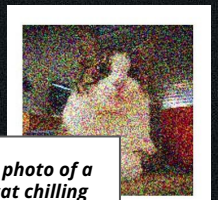
Encode text and image together, so that the influence between text and image can be measured and used in the training → proposed by OpenAI with CLIP technique: Contrastive Language-Image Pre-training (CLIP)



Origins of Diffusion for Image Generation

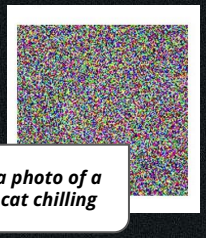
seelab

1 Diffusion Models + **2** U-Net as Noise Predictor + **3** Encoding with cross-attention



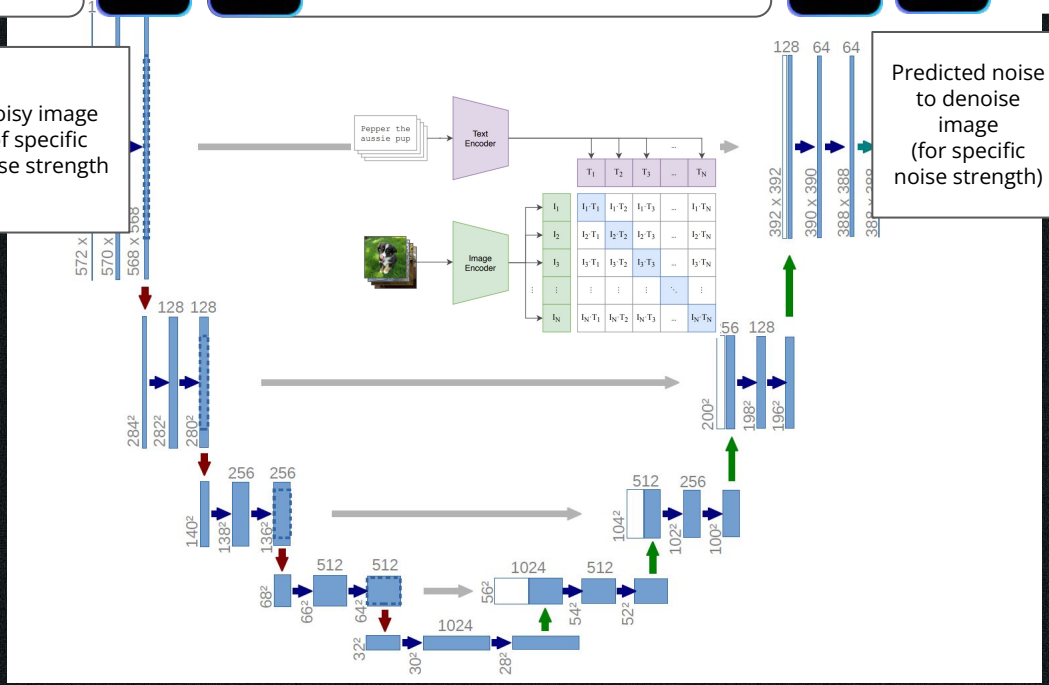
a photo of a cat chilling

a photo of a cat chilling

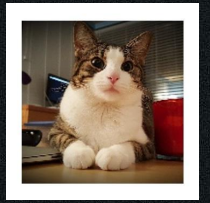


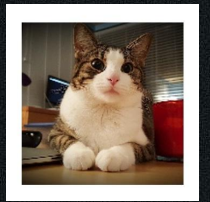
a photo of a cat chilling

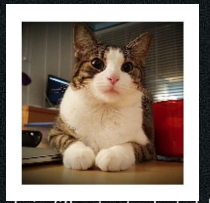
Noisy image of specific noise strength



Predicted noise to denoise image (for specific noise strength)



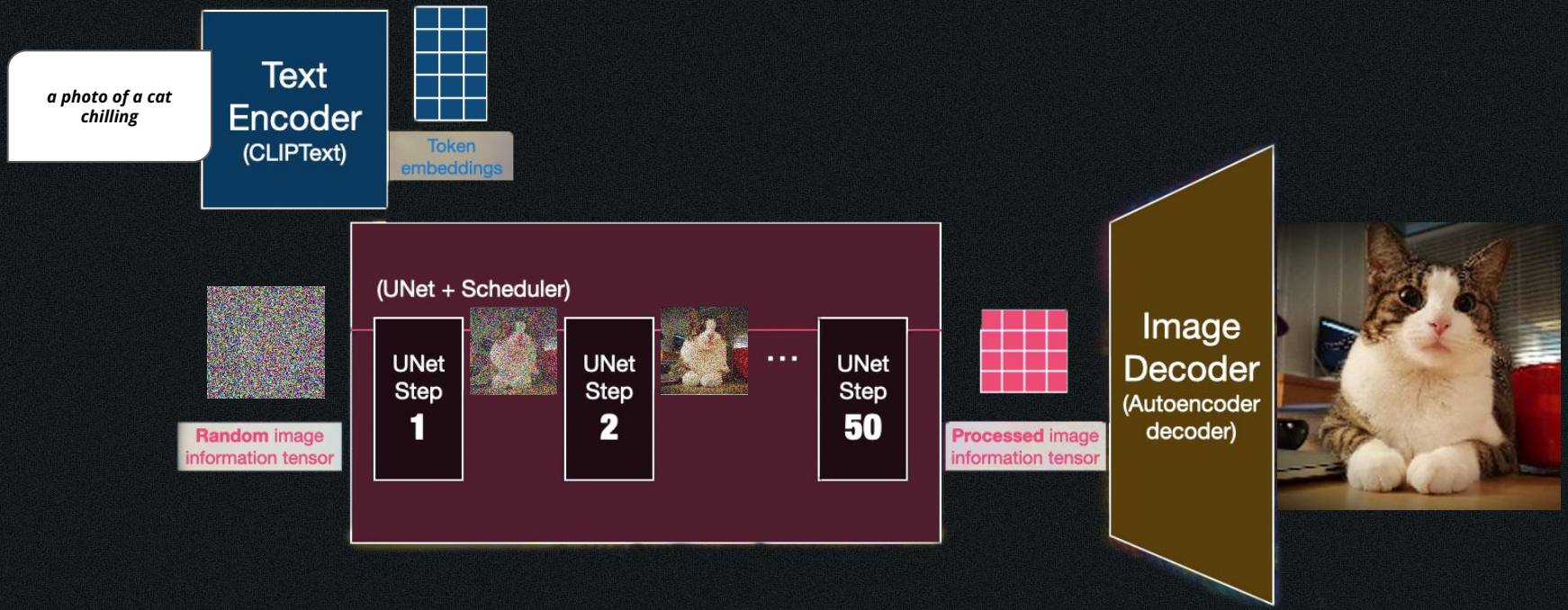




https://jalammar.github.io/illustrated-stable-diffusion/

Origins of Diffusion for Image Generation

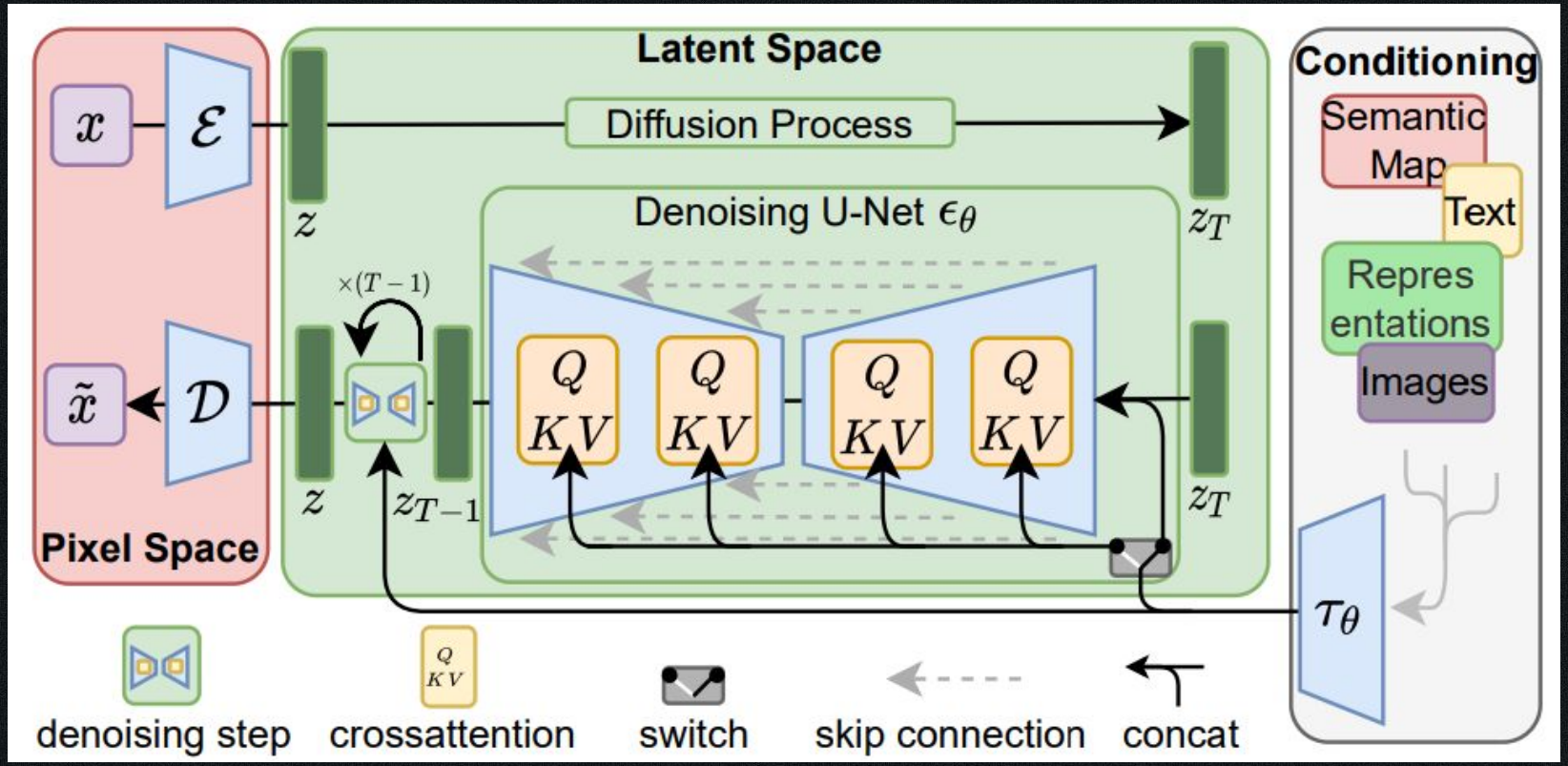
seelab



https://jalammar.github.io/illustrated-stable-diffusion/

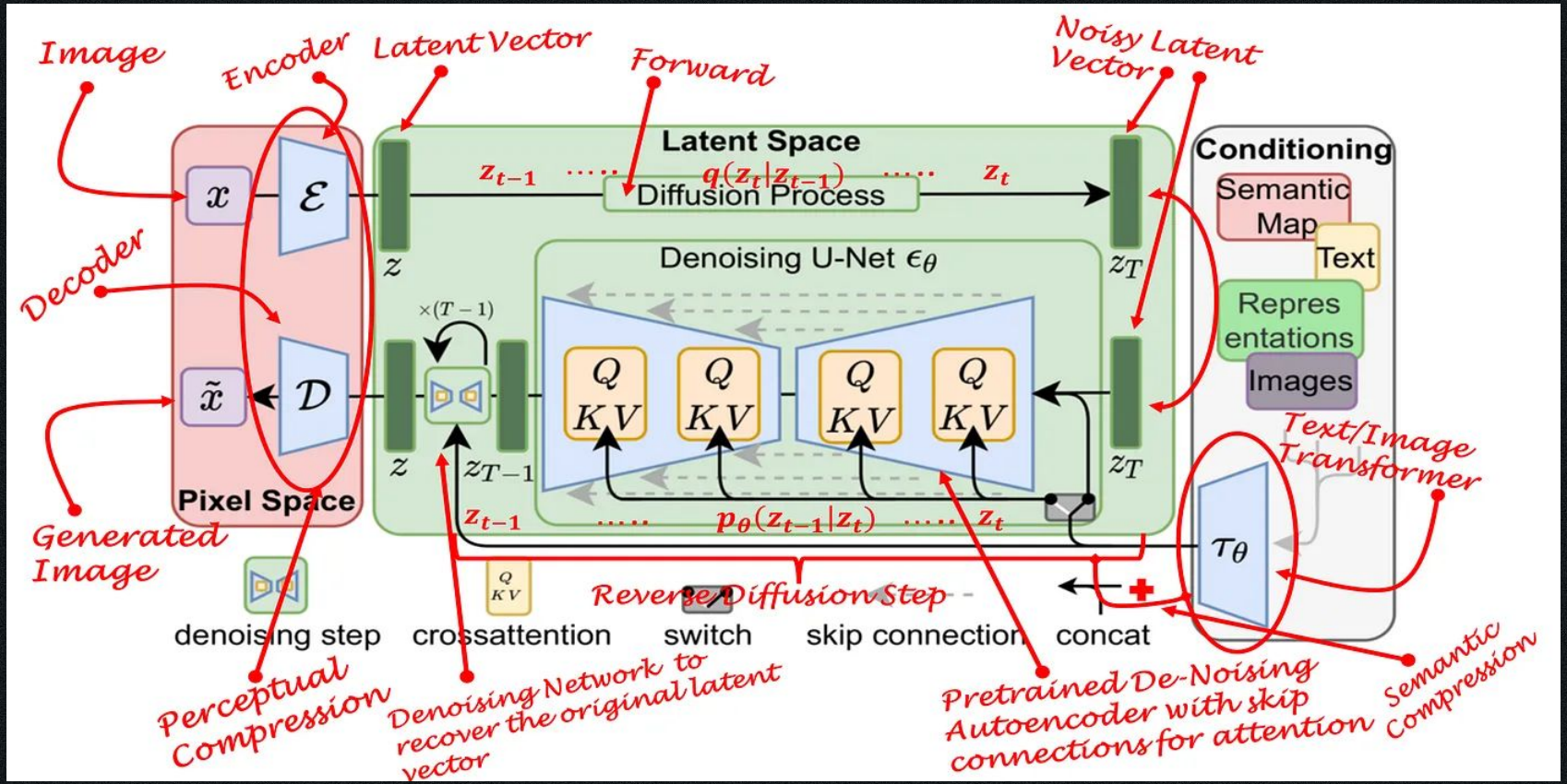
Latent Diffusion: approach behind Stable Diffusion

seelab



Latent Diffusion: approach behind Stable Diffusion

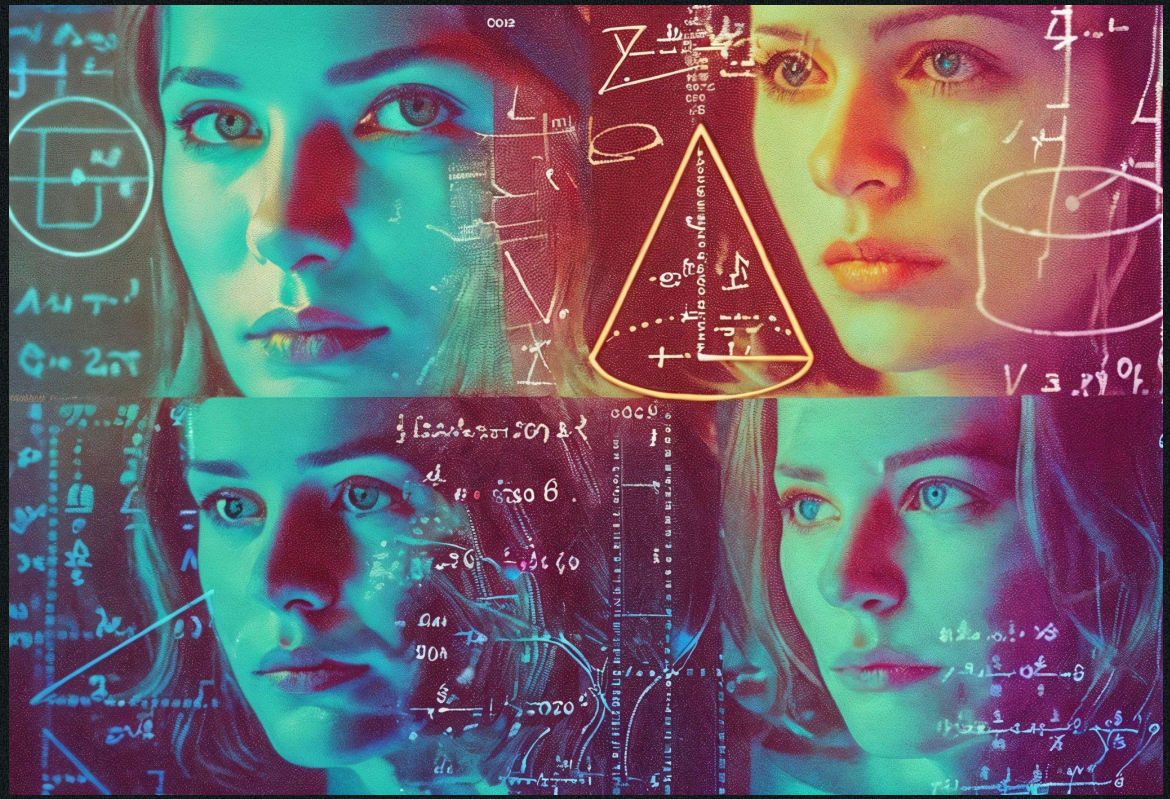
seelab



https://towardsdatascience.com/what-are-stable-diffusion-models-and-why-are-they-a-step-forward-for-image-generation-aa1182801d46

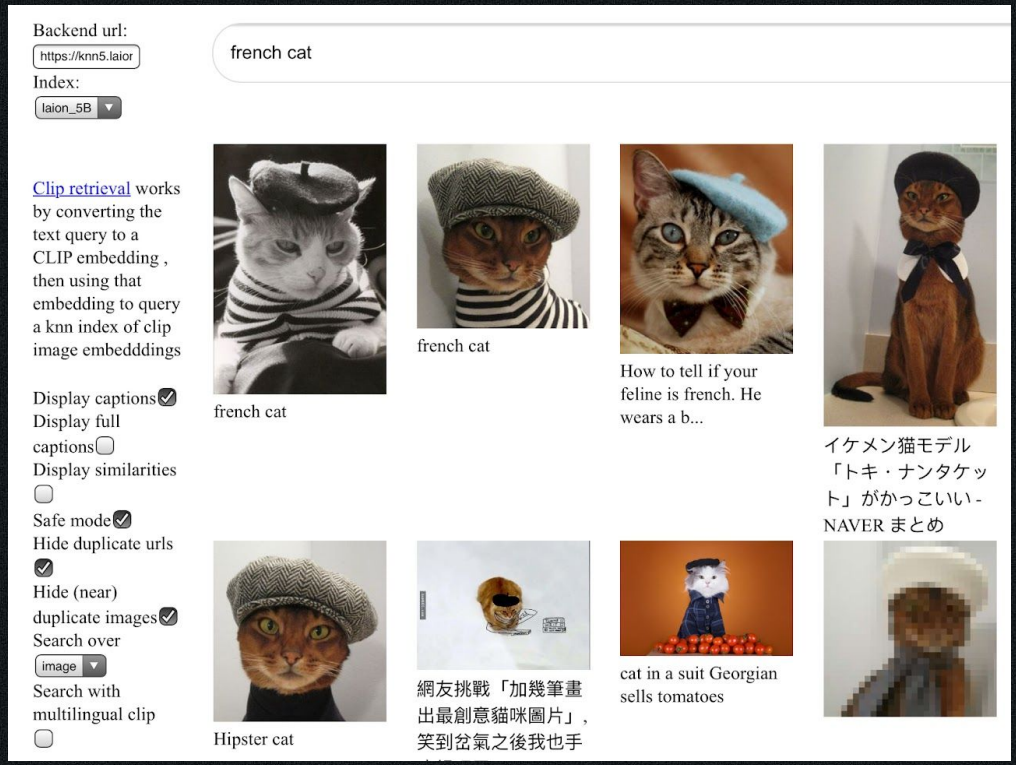
Latent Diffusion: approach behind Stable Diffusion

seelab



Data behind

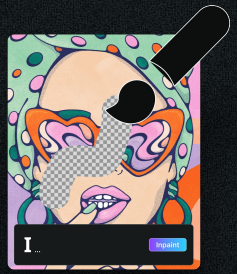
- LAION-5B: dataset behind SD1.5 and SD2
- 5,85 billion CLIP-filtered image-text pairs
- SD XL and later versions of Stability AI Stable Diffusion models used internal datasets for training

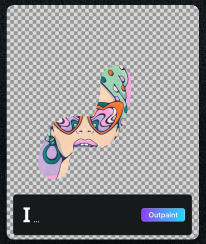


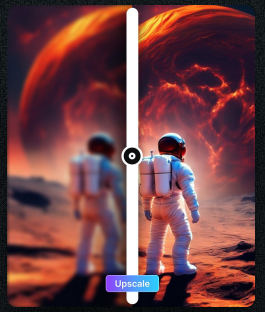
https://laion.ai/blog/laion-5b/

Different ways of generating images

seelab







Text to image

Image to image

Inpainting

Outpainting

Upscaling & enhancing

Image to image

seelab

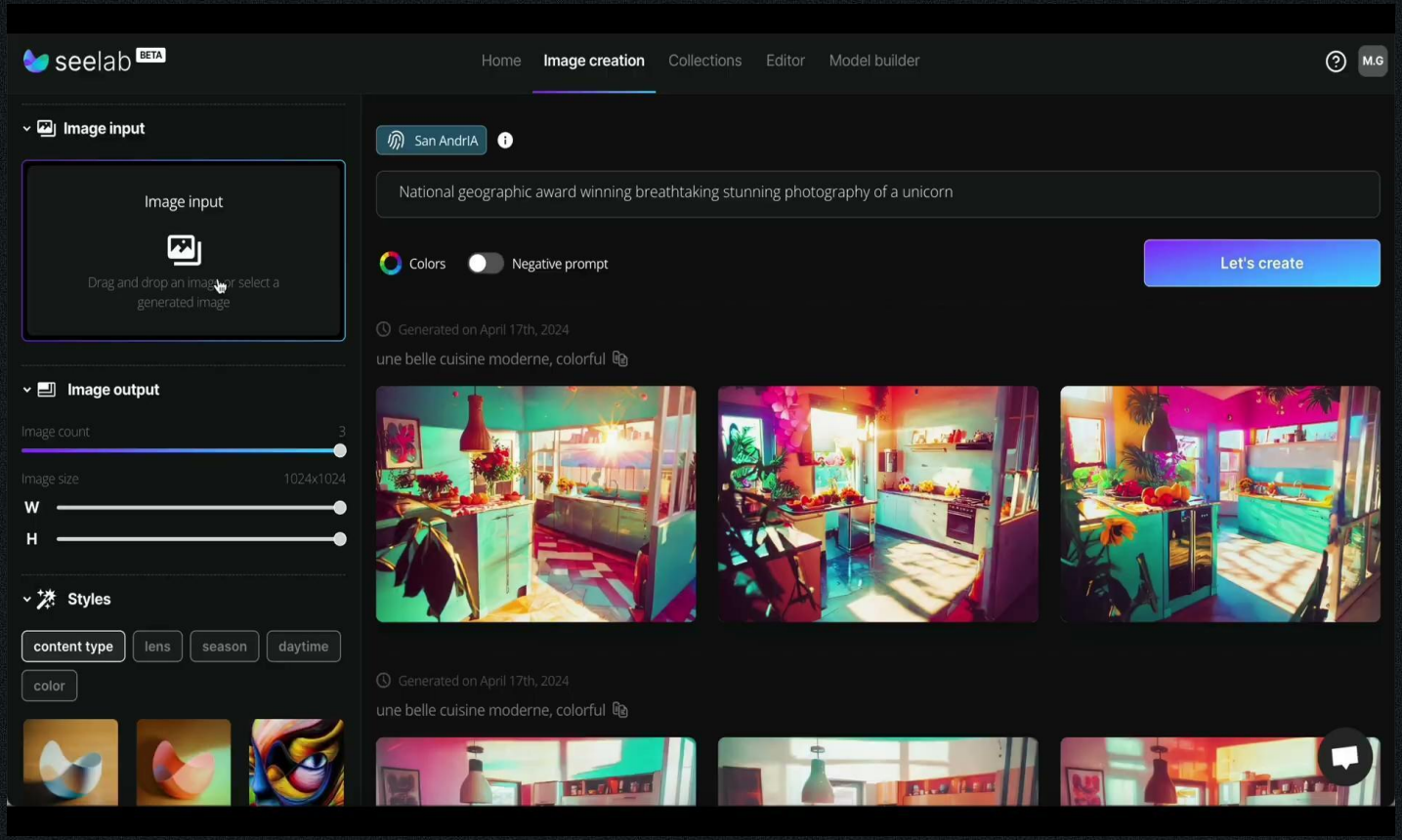
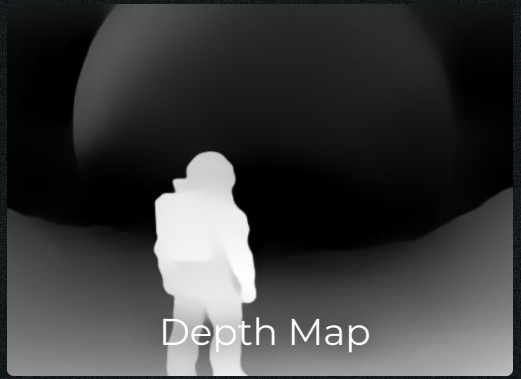
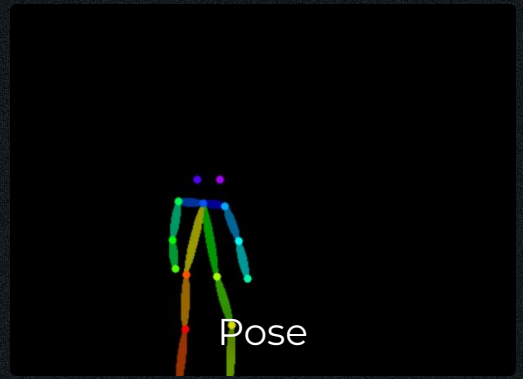


Image to image with controlNet Conditioning

seelab

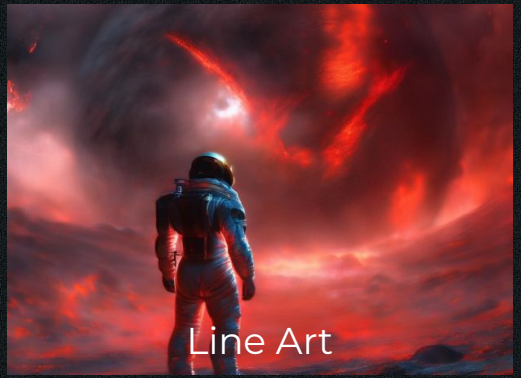




A scientist in a high-tech suit, on a glowing volcanic Venus landscape, black hole visible in the sky.

ControlNet Conditioning examples

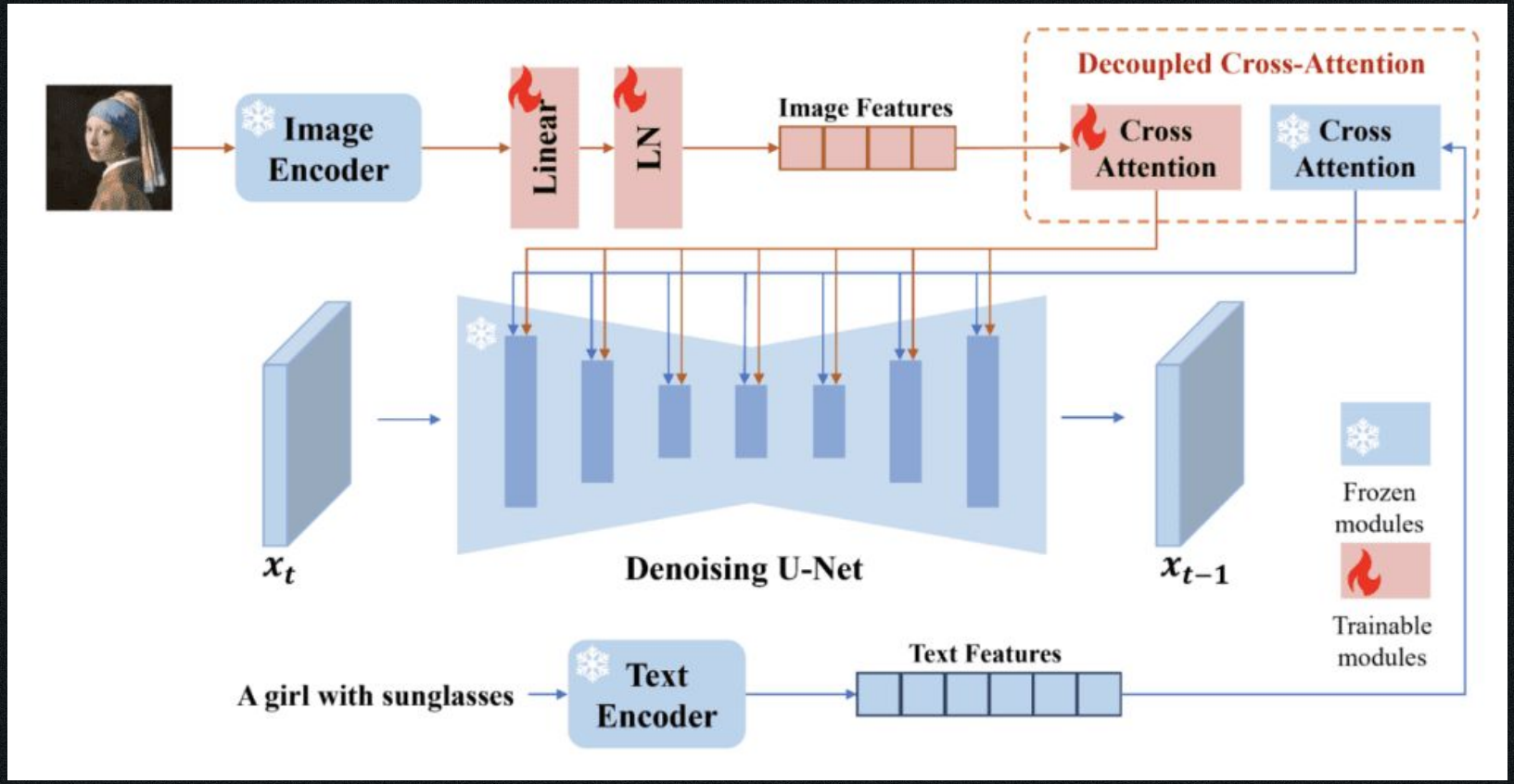
seelab



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

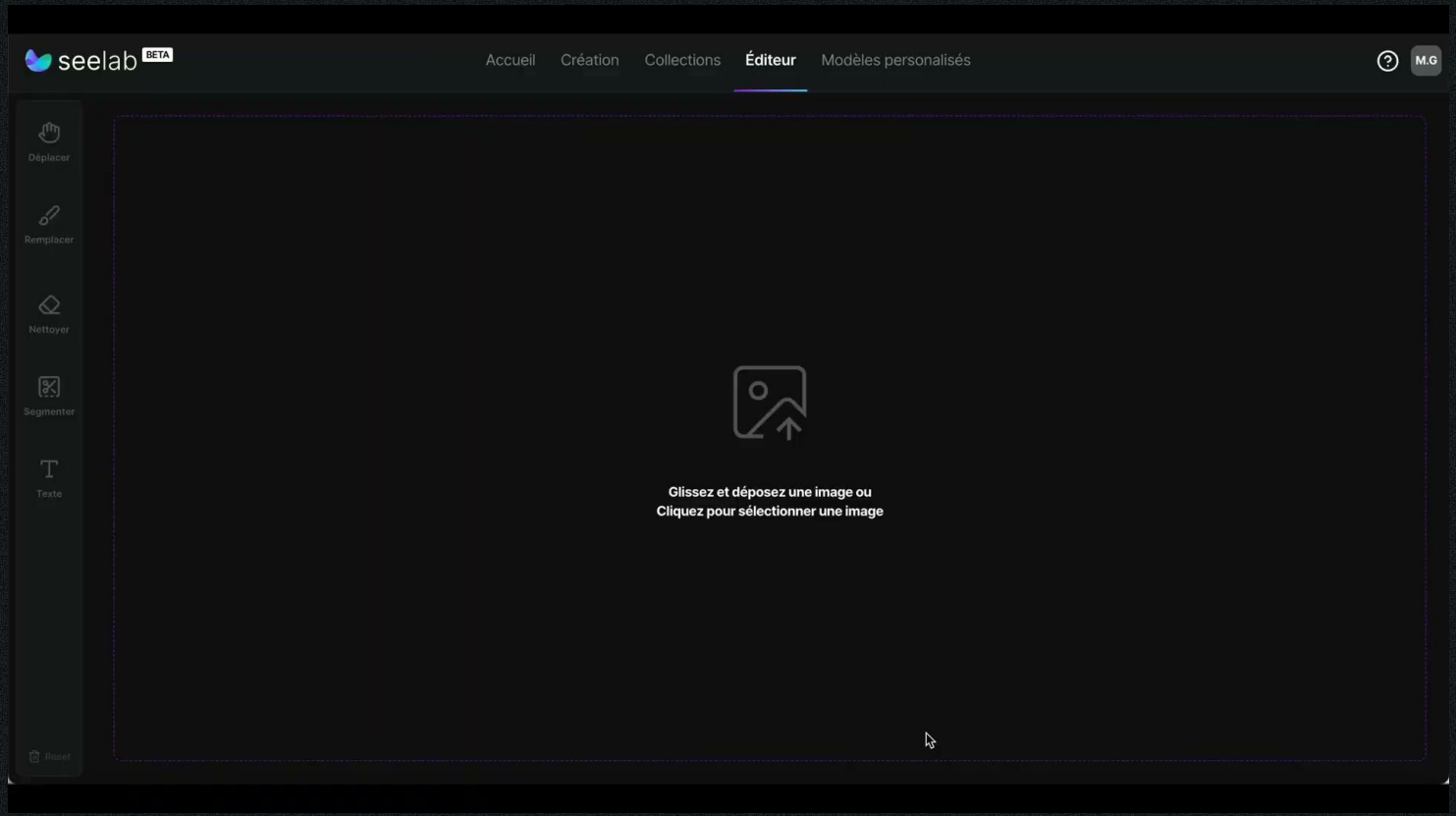
Image to image with controlNet Conditioning

seelab



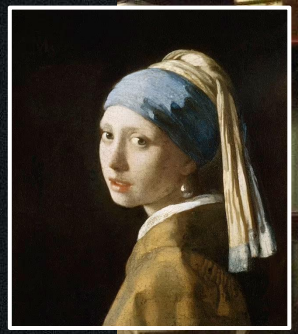
Inpainting

seelab



Outpainting

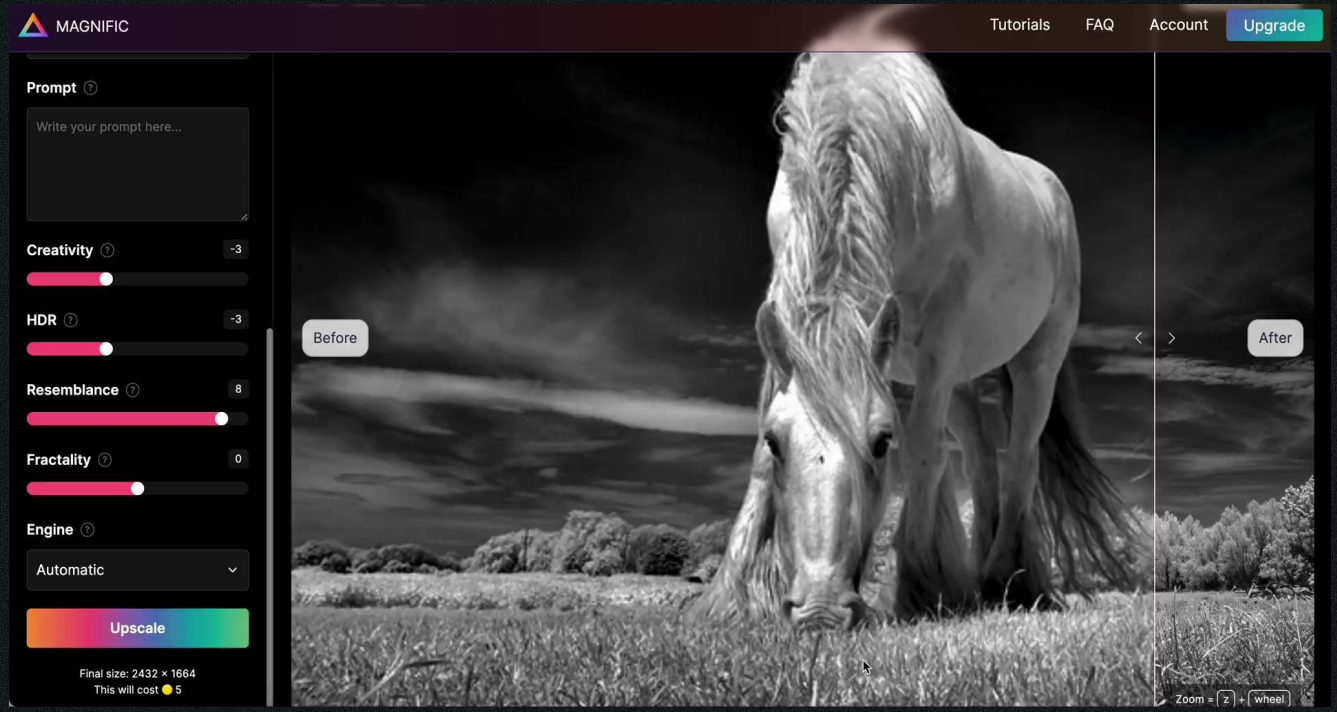
seelab



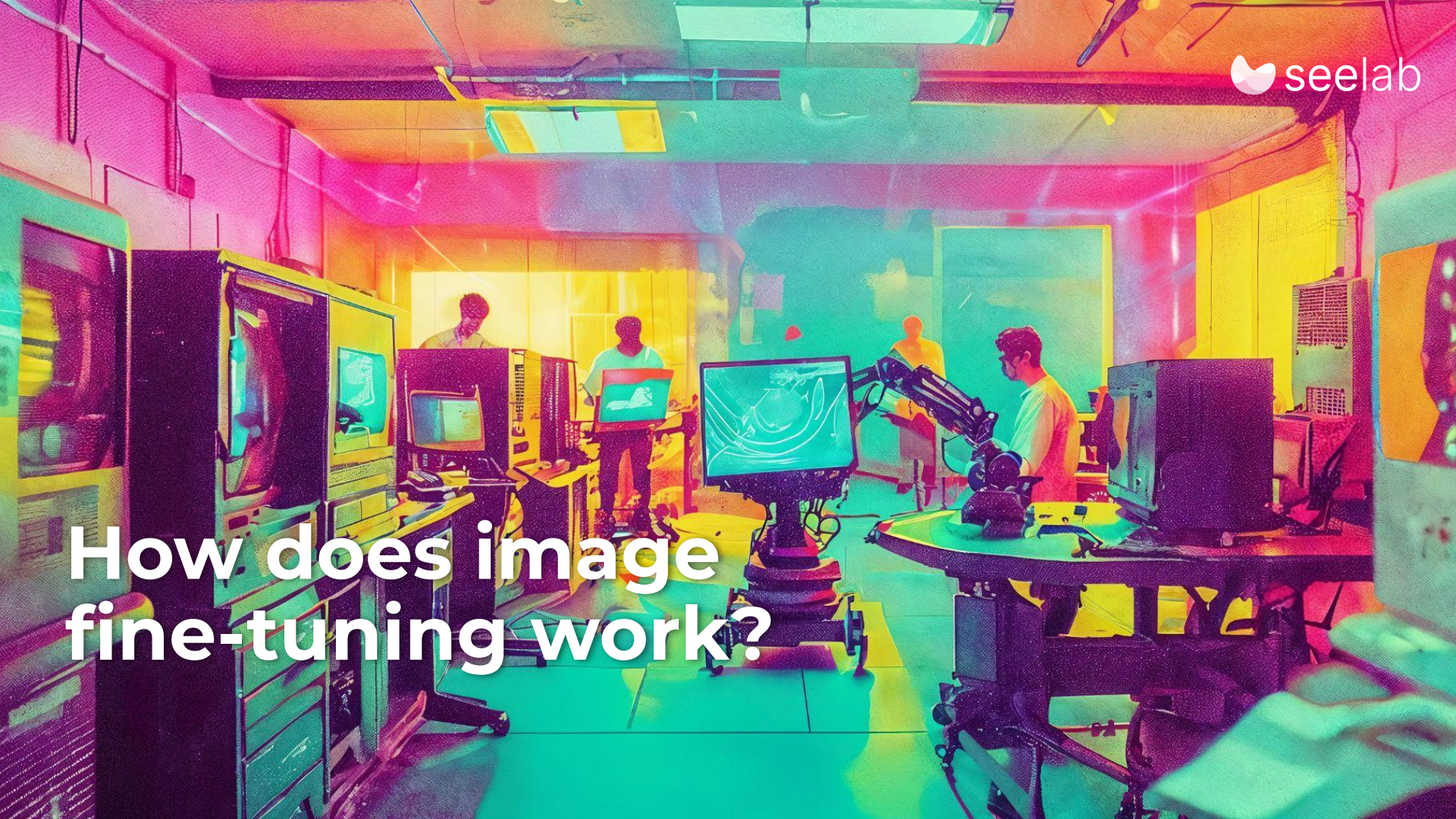
“Girl with a pearl earring”, Vermeer, 1665

Upscaling & SD-based enhancement

seelab



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


seelab
How does image fine-tuning work?

What is a fine-tuned image model?

seelab

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

seelab

“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...

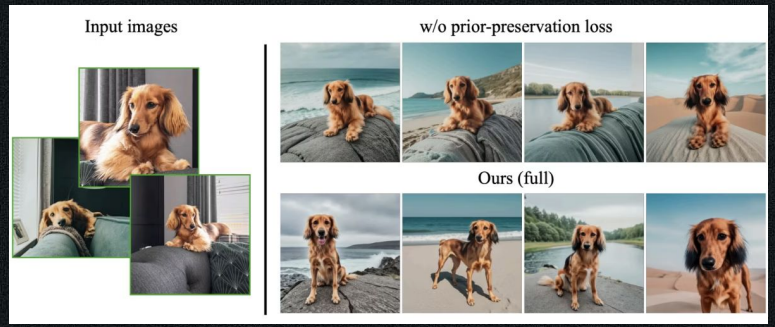
Different approaches of image model personalization

seelab

Textual Inversion

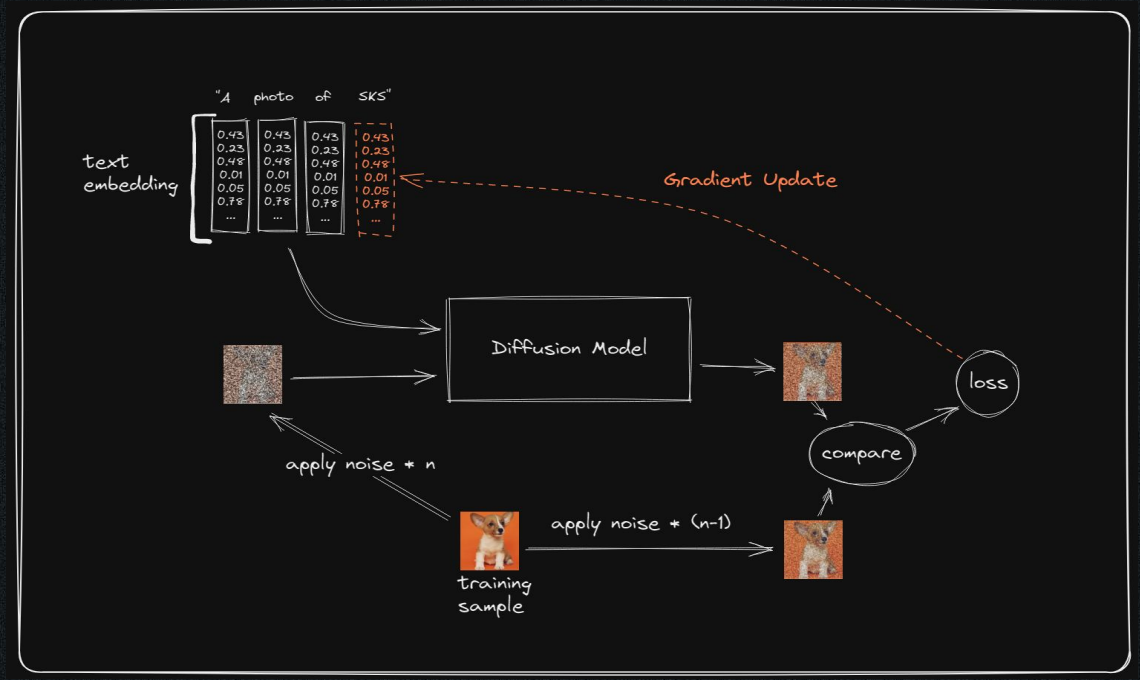
Dreambooth

LoRA



Textual Inversion

seelab

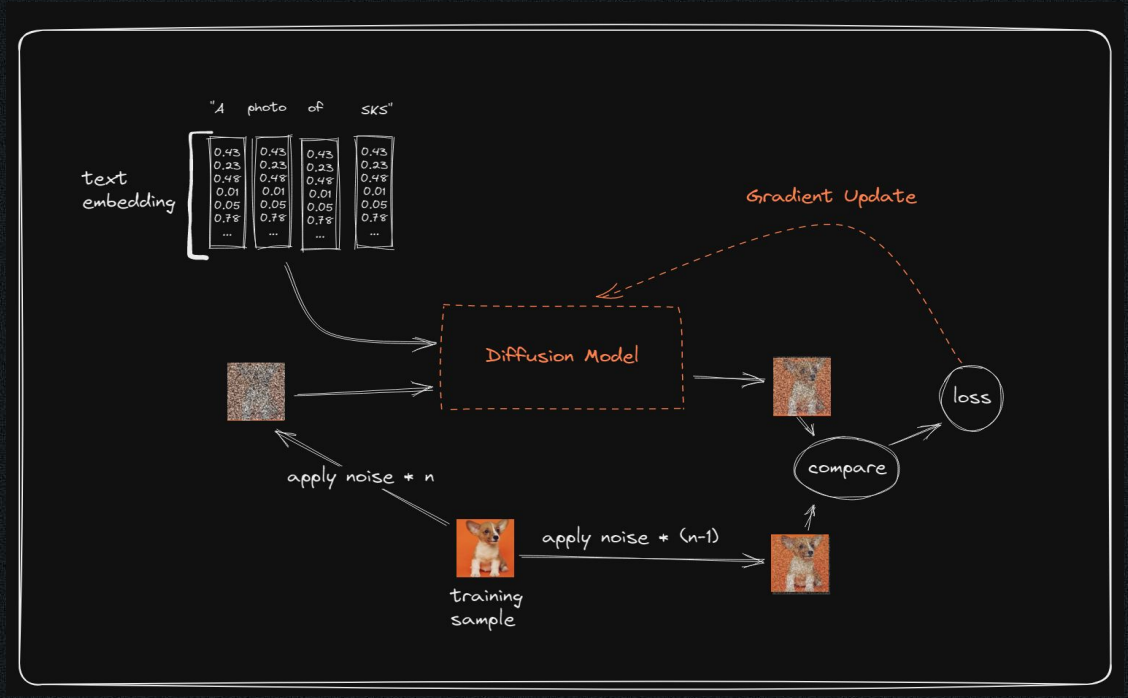


Create a special word embedding which captures the new concept

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

Dreambooth

seelab



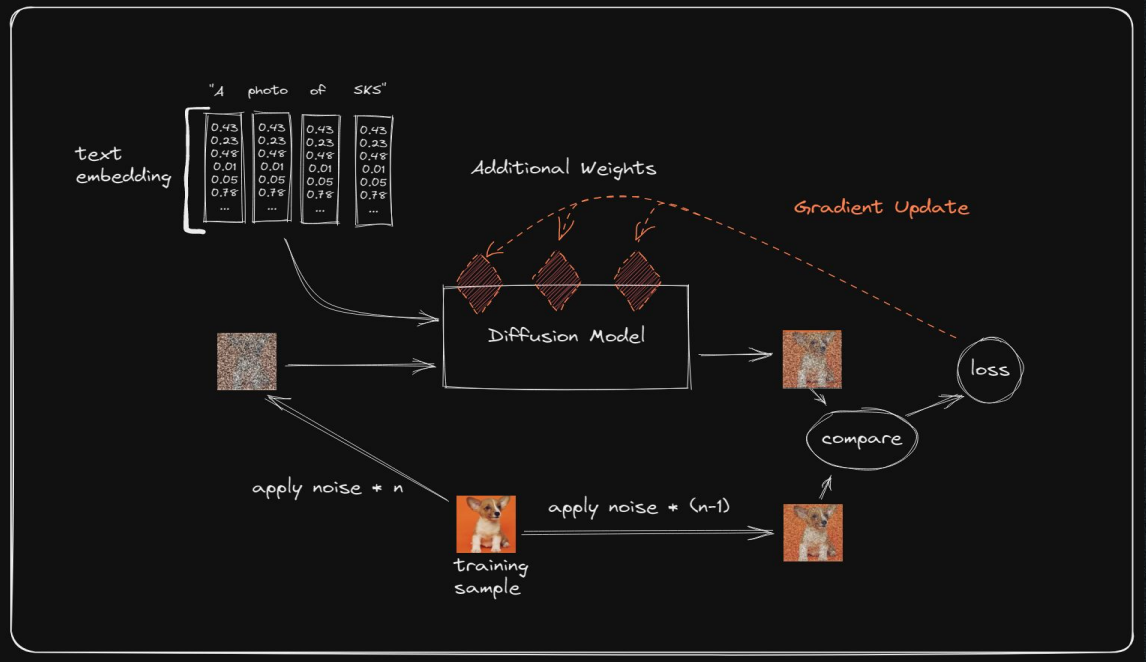
Fine-tune the model itself until it understands the new concept → creating a new model checkpoint

\+ The most effective

\- Inefficient in storage (and cost)

LoRa: Low Rank Adaptation

seelab

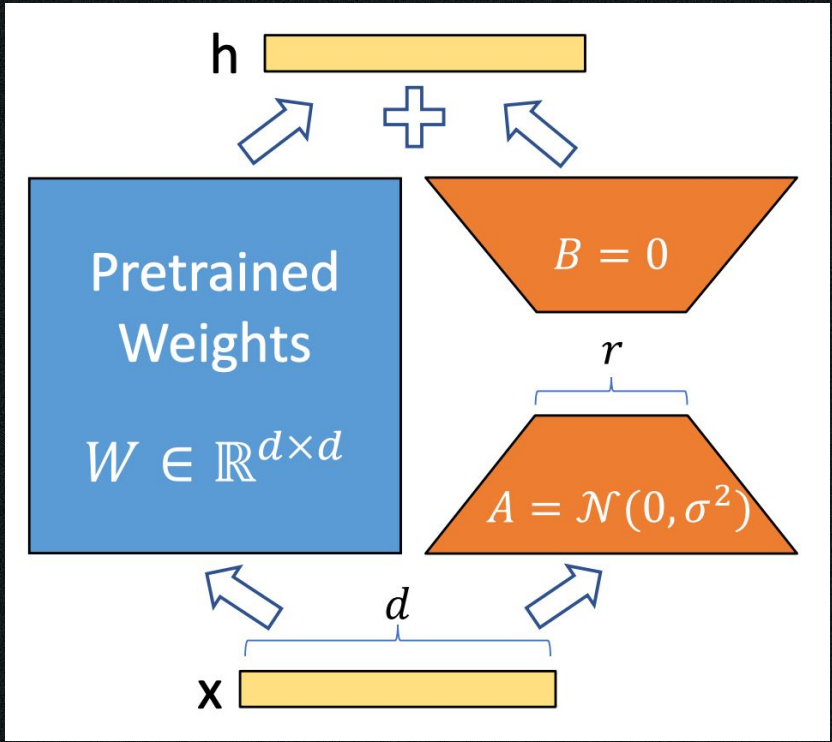


Add small update weights on top of an existing diffusion model and train only those weights to make the model understand the concept

+ Quick to train
+ Cheap
+ Good quality

LoRa: Low Rank Adaptation

seelab

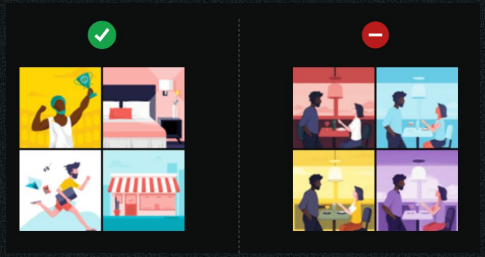


- Approach coming from LLM domain
- Training only decomposition update matrices instead of full weights
- Less parameters to train → less training resources needed
- Base model frozen → trained models can be easily swapped on top
- Smallest LoRA can be just a few Mb. Optimal size for SD XL models: 150Mb+

Garbage in, garbage out

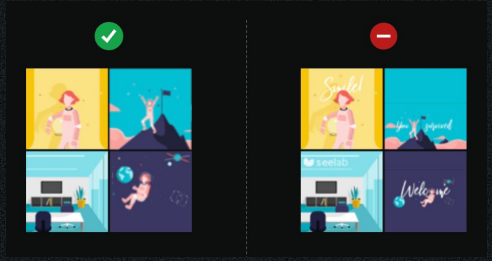
seelab

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



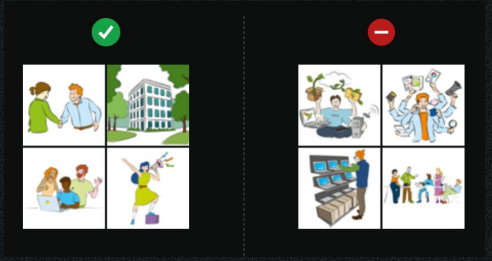
Make sure to include your most common characters or style elements, but keep diversity:

> 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.

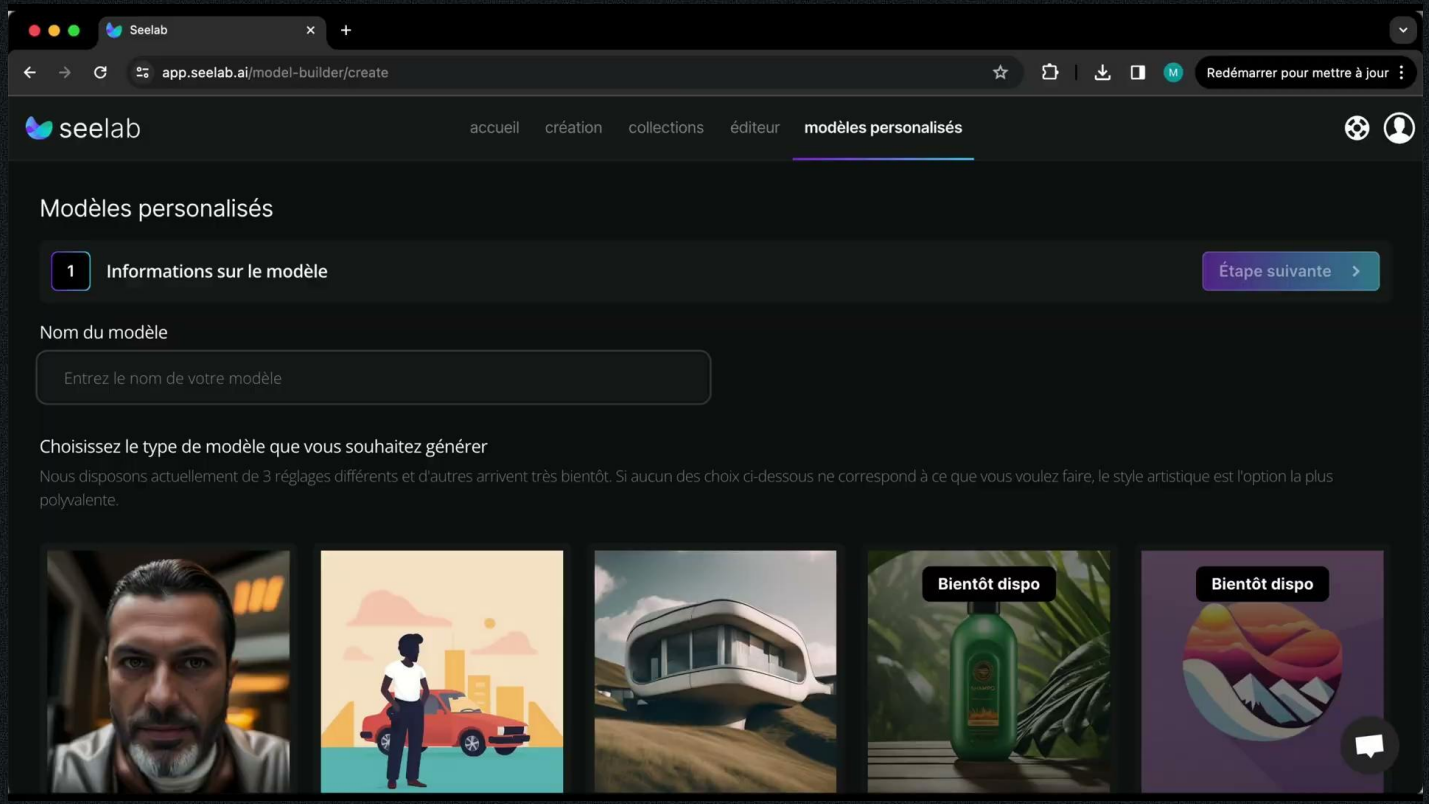
Try to avoid images where the important subjects are too small.



Prefer simple compositions with not too many elements or interactions.

Customer side

seelab



“Style” training example: Shift

seelab

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

seelab

Dataset

Generated images with prompts

“Style” training example: Shift

seelab

“Avatar” training example: Anne Kerdi

seelab

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

seelab

Dataset

Generated images with prompts

“Avatar” training example: Anne Kerdi

seelab


WHAT'S NEXT?

Future base models

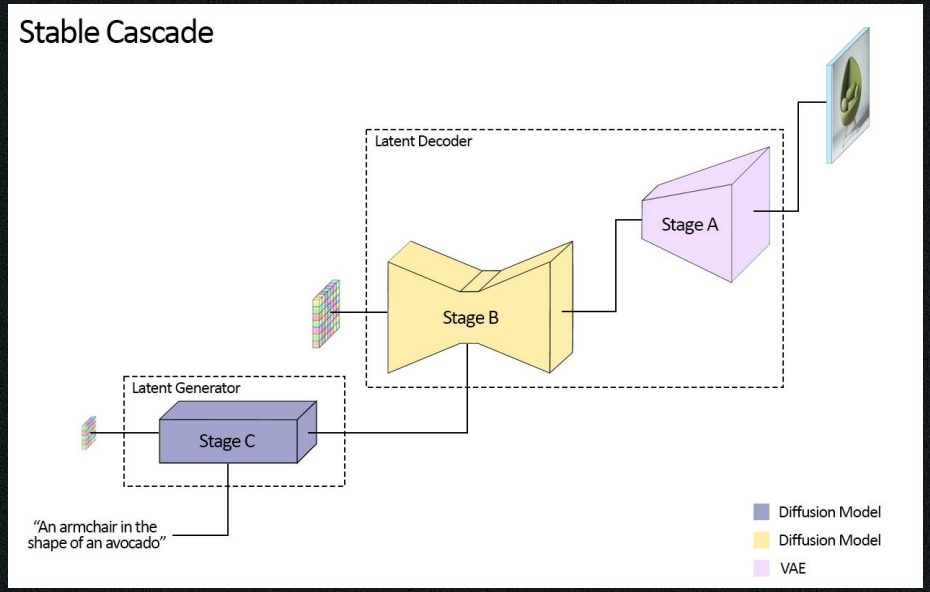
seelab

Stable Cascade

Stable Diffusion 3

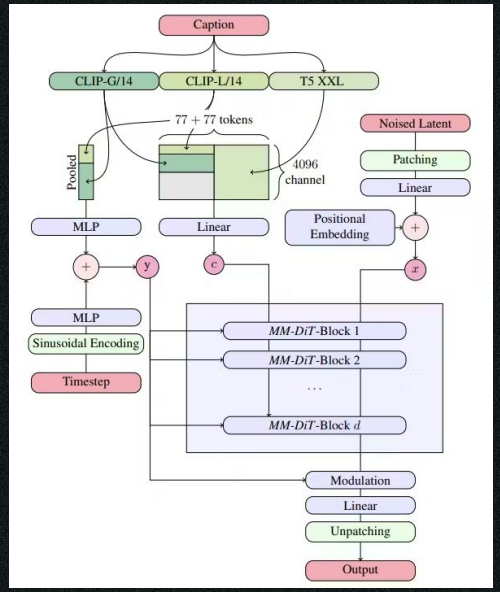
Future Models

seelab



Stable Cascade

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

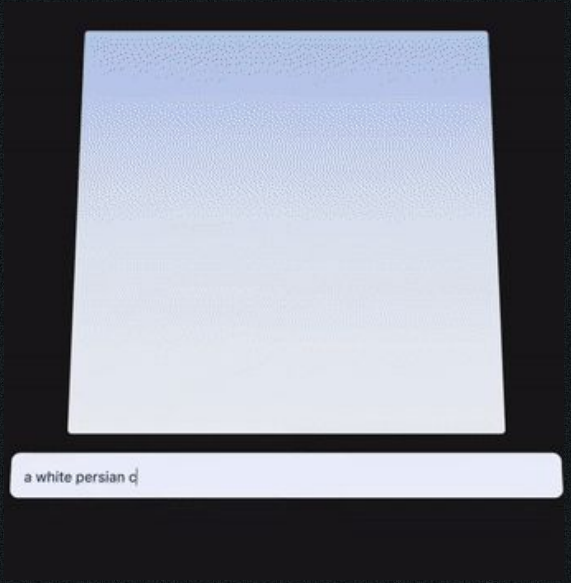


Stable Diffusion 3

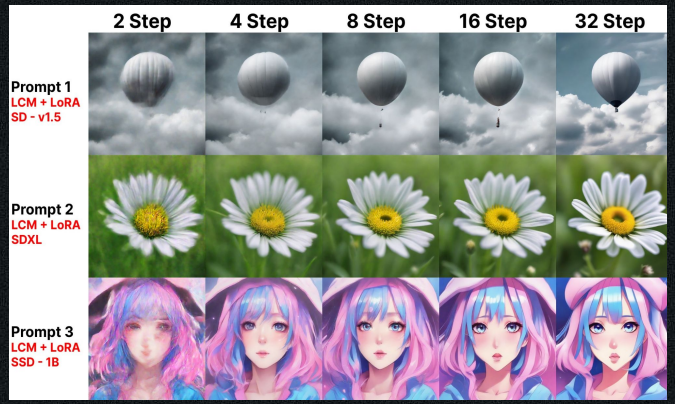
Diffusion Transformer with latent patches instead U-Net, different flow-based denoising paradigm

Fast Generation Models

seelab



SD XL Turbo



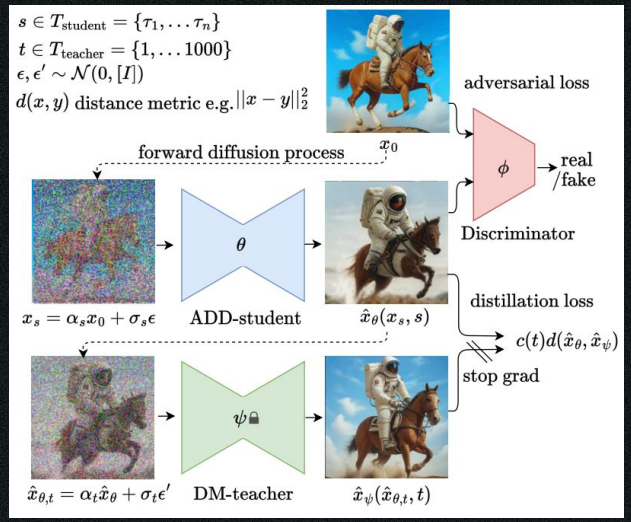
LCM: Latent Consistency Model

SD XL Lightning

https://laion.ai/blog/laion-5b/

Fast Generation Models

seelab



SD XL Turbo

Adversarial Diffusion Distillation

- 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

LoRA merge

seelab

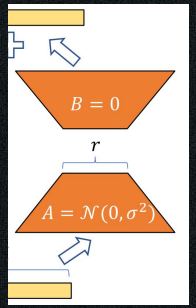
<Lora:anne_kerdi:1.0>

+

<Lora:shift:1.0>

LoRA Merging

seelab



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

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

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

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

seelab

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

seelab

Clay Fun 80%
Cel Shading 20%





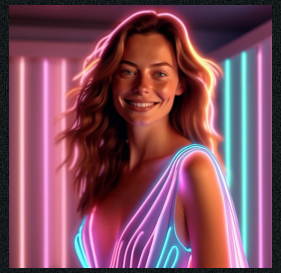

Clay Fun 20%
Cel Shading 80%

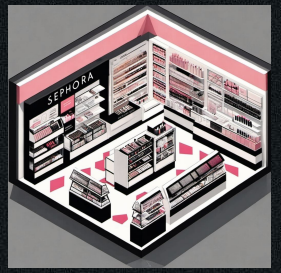
Lora Merge

seelab

Mixing Models gives you the opportunity to create your

Own Unique Style !







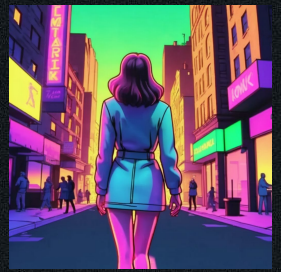



Image to video

seelab

>
>
>


CONCLUSION


Traceability
Transparency
Ethics

QUESTIONS...
AND THEN BEER TIME!

Let's talk:

matthieu@seelab.ai
dmitry@seelab.ai