

Project Idea for Learning from Images

Applications of Stable Diffusion:

- DreamBooth
- Prompt-to-Prompt
- Conceptualizer

28.11.2022

Dennis Fast, Amin Suaad, Manuel Freistein

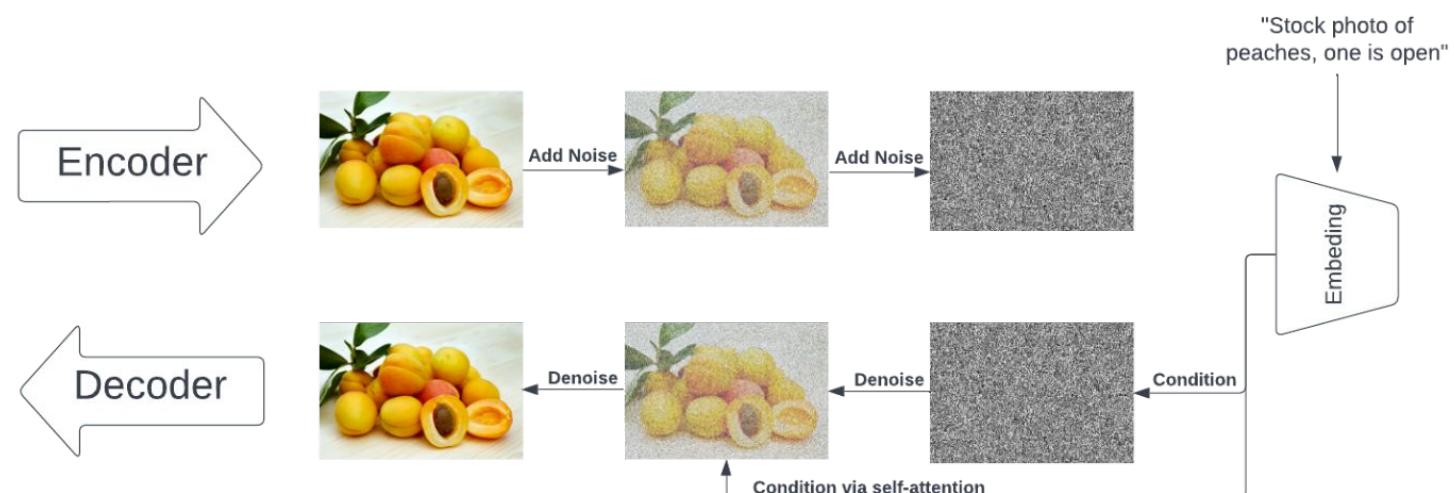
Outline

1. Short Intro into Stable Diffusion
2. DreamBooth
3. Prompt-to-Prompt
4. Stable Diffusion Conceptualizer
 - a. Examples
 - b. Data and Metrics

Short Intro into Stable Diffusion (1)

- text-to-image latent diffusion model
- created by the researchers and engineers from CompVis, Stability AI and LAION
- trained on 512x512 images from a subset of the LAION-5B database (the largest, freely accessible multi-modal dataset at the moment)

Pre-training process:

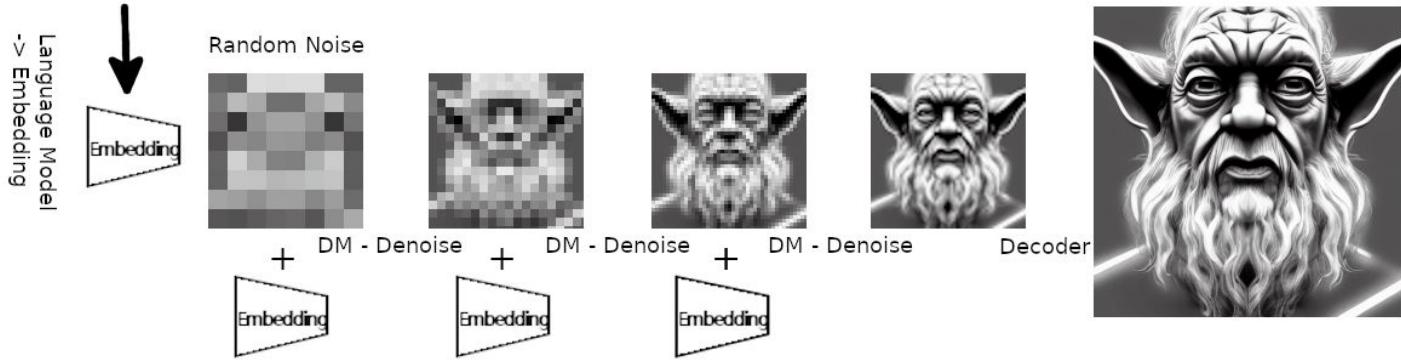


https://huggingface.co/blog/stable_diffusion

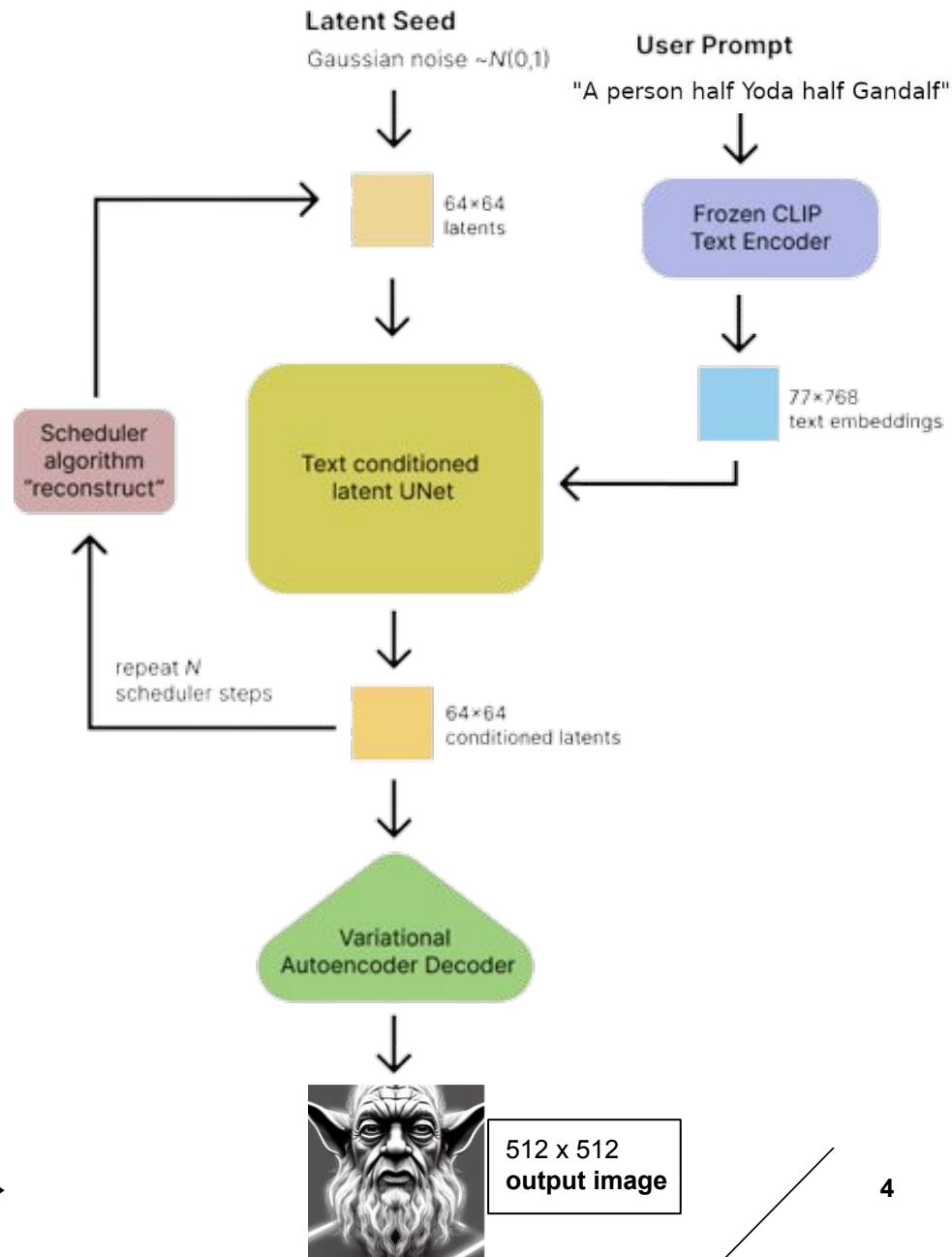
Short Intro into Stable Diffusion (2)

Inference process:

"A person half Yoda half Gandalf"



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



Evaluation metrics

- image quality (e.g. inception score)
- text relevance (e.g. R-precision)
- object accuracy (e.g. Semantic Object Accuracy)
- Type of a metric: objective (automatic) or subjective (survey among Data Science students)

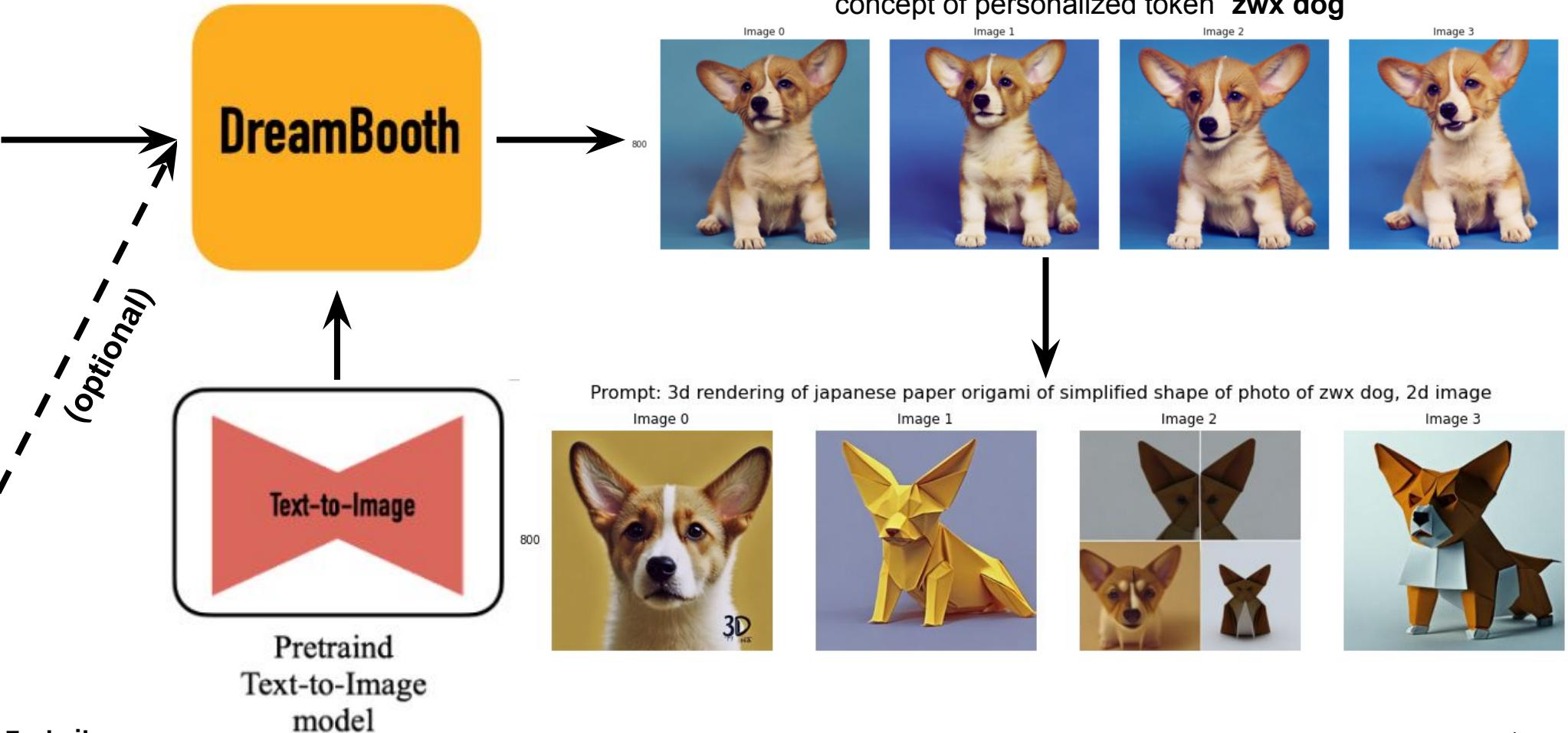
TISE: Bag of metrics https://www.ecva.net/papers/eccv_2022/papers_ECCV/html/5685_ECCV_2022_paper.php

DreamBooth: Fine-tuning to a personalized token

“picture of zwx dog”



200 x “picture of a dog”



DreamBooth: Creating of visual avatar

“picture of **dfast** person”



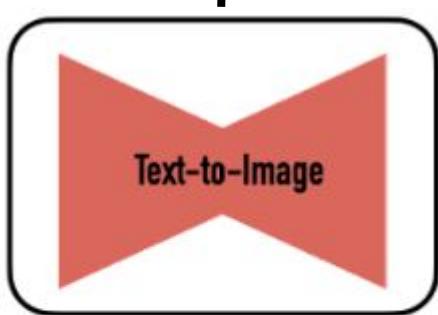
concept of personalized token
“**dfast person**”



???

Variation:

- clothes
- body pose
- facial expression
- age
- environment
- background
- light situation



Pretrain
Text-to-Image
model

Prompt-to-Prompt

- Editing is challenging for generative models.
- Often small change in the text can lead to a massive change in the image.
- Cross attention layers are key to establish the relation between the image and each word of the prompt.
- Localized editing by replacing a word, global editing by adding a specification, and even controlling the extent to which a word is reflected in the image.



Prompt-to-Prompt example (Changing one particular entity)

```
personprompt = "A boy playing in the field"  
boy = stablediffusion(personprompt, seed=seed)  
girl = stablediffusion(personprompt.replace("boy", "girl"), seed=seed)  
display(*[boy, girl])
```



Prompt-to-Prompt example (Changing style)

```
personprompt = "A boy playing in the field"  
boy = stablediffusion(personprompt, seed=seed)  
painting_boy = stablediffusion(personprompt, "a watercolor painting of " + personprompt, seed=seed)  
display(*[boy, painting_boy])
```



Prompt-to-Prompt example (Changing weight)

```
prompt = "A fantasy landscape with a pine tree in the ground and sunset in the distance"  
seed = 2483964025  
base = stablediffusion(prompt, seed=seed)  
lesspine = stablediffusion(prompt, prompt_edit_token_weights=[(6, -3)], seed=seed) # make the tree less "pine"  
imgs = [base, lesspine]  
display(*imgs)
```

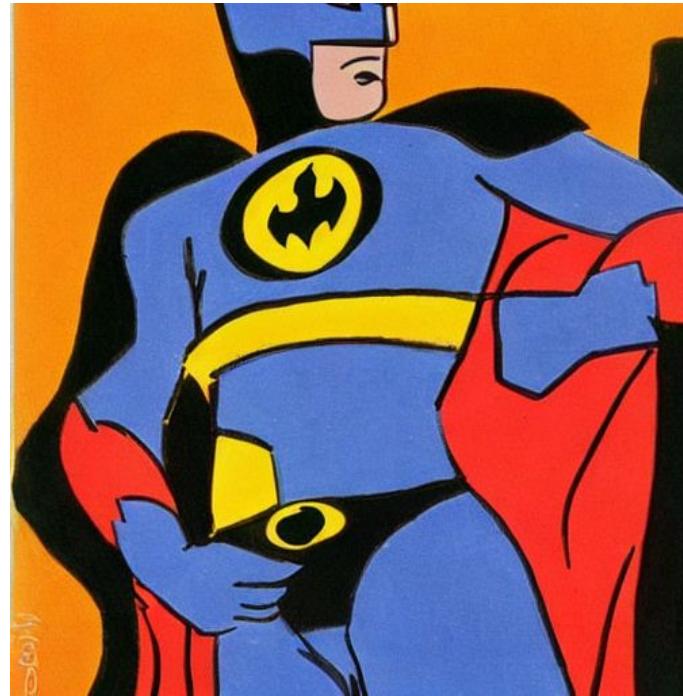


Stable Diffusion Conceptualizer: Examples

<Batman> in <George Grosz> style



<Batman> in <Matisse> style



<Batman> in <Caravaggio> style



Stable Diffusion Conceptualizer: Data and Metrics

Conceptualizer:

- uses style tags to generate images with a particular art style

Data

- Public Domain Images
- Training with small datasets

Fréchet Inception Distance (FID)

- FID is a performance metric that calculates the distance between the feature vectors of real images and the feature vectors of fake images
- Measures if the similarity between real and generated images is close.
- Lower the FID, the better the quality.

Sources

Stable Diffusion:

- <https://github.com/CompVis/stable-diffusion>
- <https://arxiv.org/abs/2112.10752>

DreamBooth:

- <https://dreambooth.github.io/>
- <https://arxiv.org/pdf/2208.12242.pdf>

Prompt-to-Prompt:

- <https://github.com/google/prompt-to-prompt>
- https://prompt-to-prompt.github.io/ptp_files/Prompt-to-Prompt_preprint.pdf

Stable Diffusion Conceptualizer:

- <https://huggingface.co/spaces/sd-concepts-library/stable-diffusion-conceptualizer>
- Rinon Gal, Yuval Alaluf, Yuval Atzmon et al.: An Image is Worth One Word: Personalizing Text-to-Image Generation using Textual Inversion, 2022 online: <https://arxiv.org/abs/2208.01618> (retrieved: 27 November)
- Ali Borji: Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2, 2022, online: <https://arxiv.org/abs/2210.00586> (retrieved: 27 November)
- Muhammad Ferjad Naeem, Seong Joon Oh, Youngjung Uh et al.: Reliable Fidelity and Diversity Metrics for Generative Models, 2020, online: <https://arxiv.org/abs/2002.09797> (retrieved: 27 November)

Thank you for your attention!