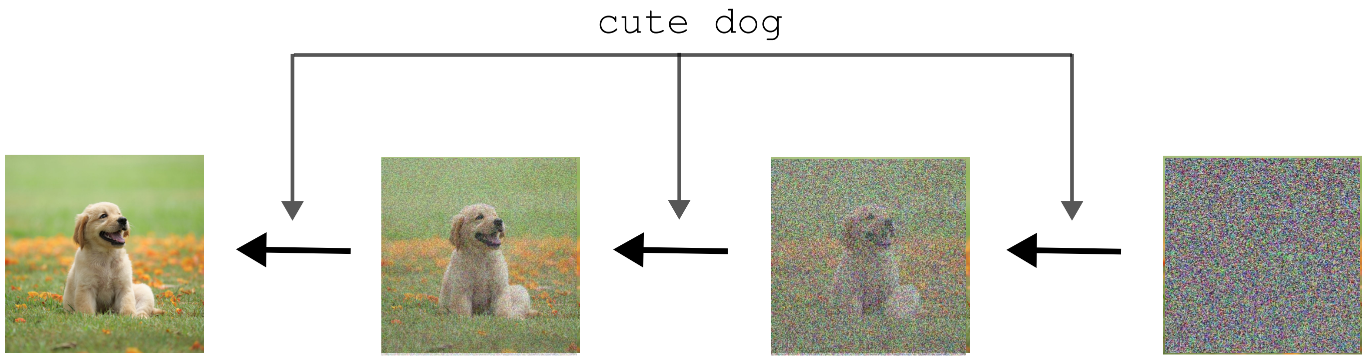


From Noise to Art:
User-Controlled Image Generation Beyond Prompt Engineering

Niklas Deckers

UNIVERSITÄT LEIPZIG

ScaDS.AI DRESDEN LEIPZIG

Recent Advancements in Image Generation

- Allow to generate images based on a given text
- Surprisingly high quality
- Accessible to a wide audience

`mountain with a sunset and a river`

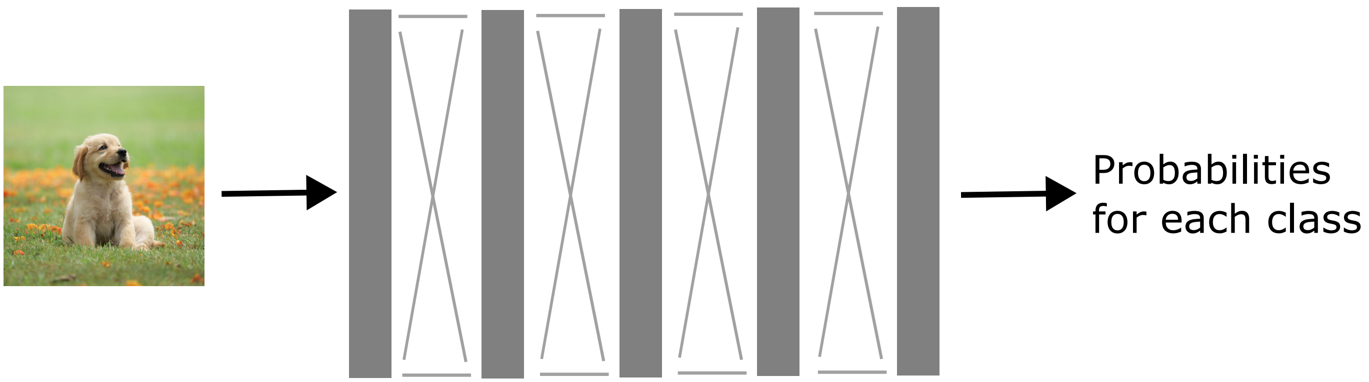
Public event at ScaDS.AI

1

Generative Text-To-Image Models

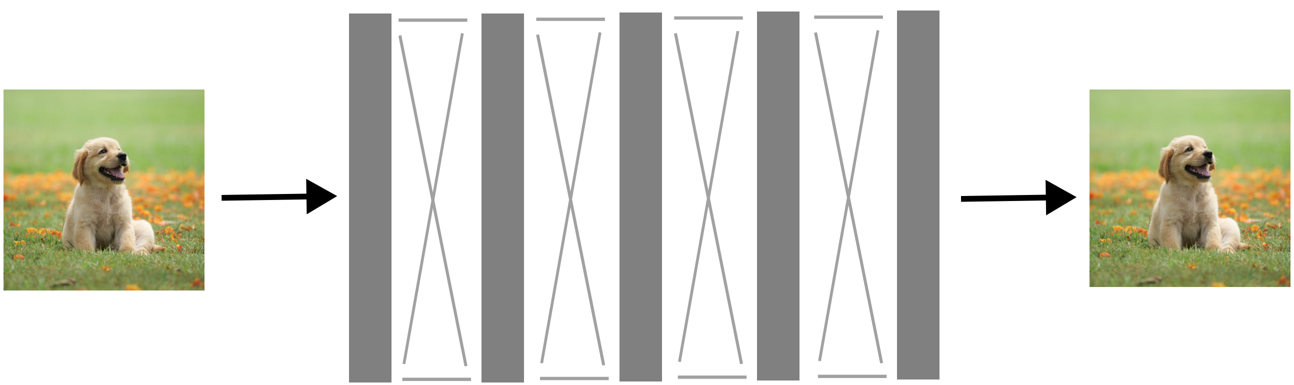
Basic Neural Networks

- Used for classification tasks
- Input: Single example (image)
- Output: Probability for each class
- Trained on labeled data, e.g., from web crawls



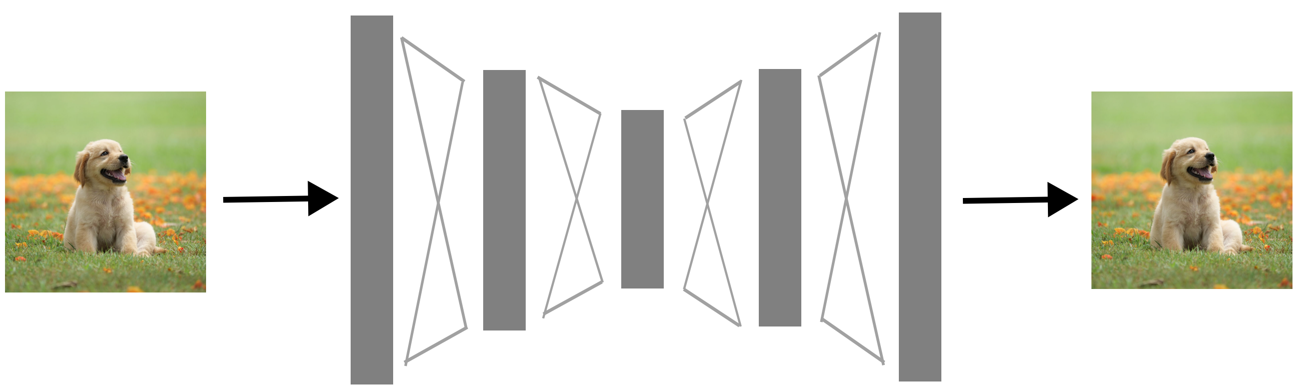
Generative Networks: Autoencoders

- Autoencoders are trained on re-generating the input image
- Forced to pass information through an information bottleneck
  - The space of possible images is larger than the space of desirable images
  - Exclude irrelevant information by reducing the amount of information that is passed through
- Changing the latent value at the bottleneck to an unseen value generates unseen image
- Problem: Choosing this value is not easy



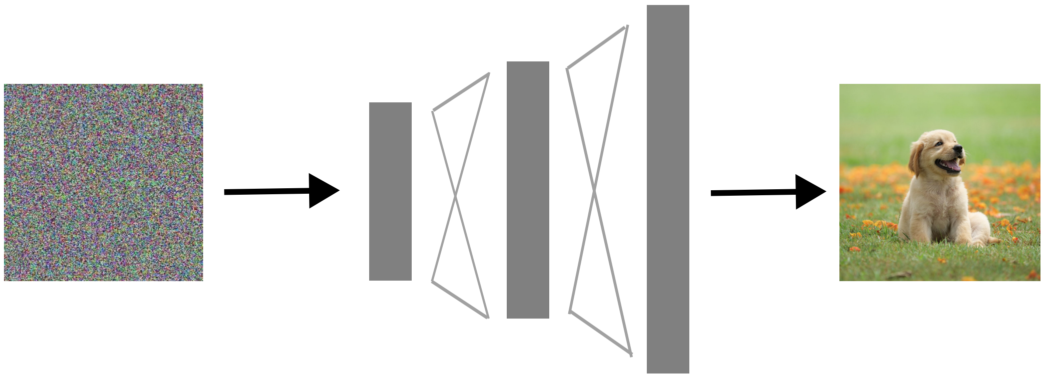
Generative Networks: Autoencoders

- Autoencoders are trained on re-generating the input image
- Forced to pass information through an information bottleneck
  - The space of possible images is larger than the space of desirable images
  - Exclude irrelevant information by reducing the amount of information that is passed through
- Changing the latent value at the bottleneck to an unseen value generates unseen image
- Problem: Choosing this value is not easy



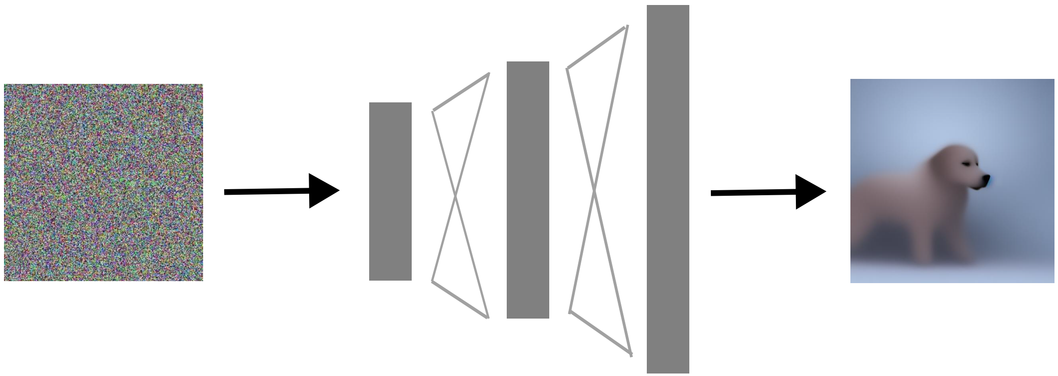
Generative Networks: Autoencoders

- Autoencoders are trained on re-generating the input image
- Forced to pass information through an information bottleneck
  - The space of possible images is larger than the space of desirable images
  - Exclude irrelevant information by reducing the amount of information that is passed through
- Changing the latent value at the bottleneck to an unseen value generates unseen image
- Problem: Choosing this value is not easy



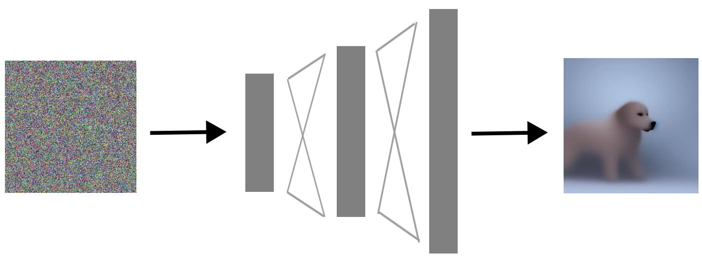
Generative Networks: Autoencoders

- Autoencoders are trained on re-generating the input image
- Forced to pass information through an information bottleneck
  - The space of possible images is larger than the space of desirable images
  - Exclude irrelevant information by reducing the amount of information that is passed through
- Changing the latent value at the bottleneck to an unseen value generates unseen image
- Problem: Choosing this value is not easy



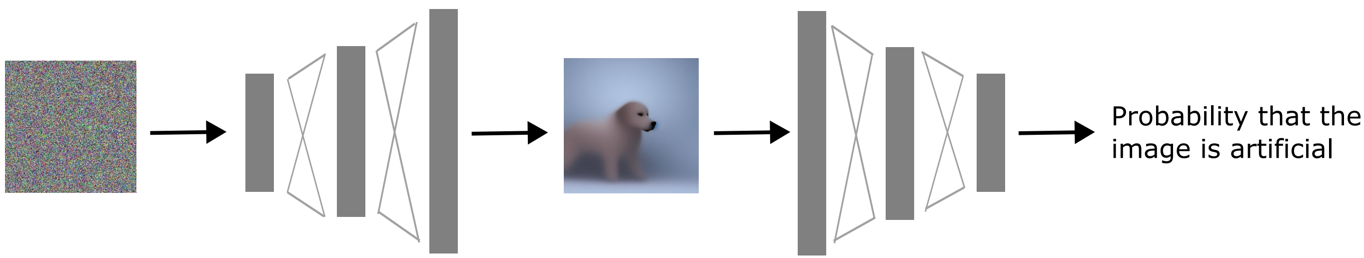
Generative Adversarial Networks (GAN)

- Idea: Optimize the decoder to generate good images even if using arbitrary noise at the latent (Generator)
- Use a second agent to tell whether the generated image is artificial or not (Discriminator)
- Train both agents concurrently



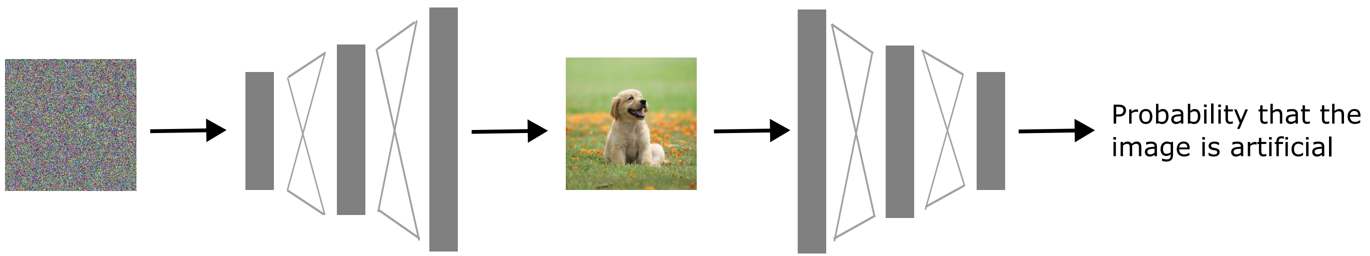
Generative Adversarial Networks (GAN)

- Idea: Optimize the decoder to generate good images even if using arbitrary noise at the latent (Generator)
- Use a second agent to tell whether the generated image is artificial or not (Discriminator)
- Train both agents concurrently



Generative Adversarial Networks (GAN)

- Idea: Optimize the decoder to generate good images even if using arbitrary noise at the latent (Generator)
- Use a second agent to tell whether the generated image is artificial or not (Discriminator)
- Train both agents concurrently

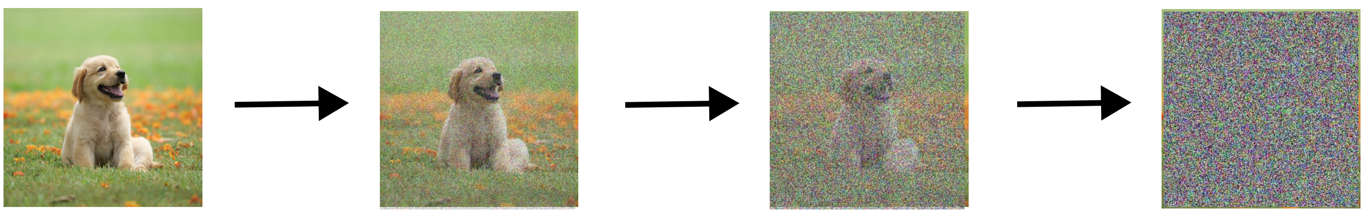


Generative Adversarial Networks (GAN)

- First example of noise to art
- Application example: *This person does not exist*
- Lack of control over the generated images

Diffusion Models

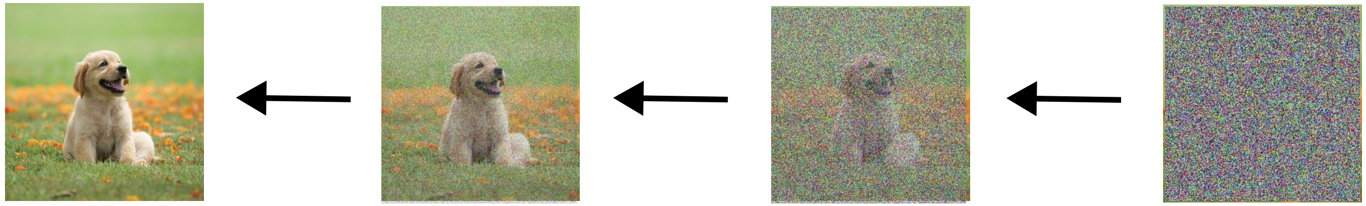
- A diffusion process turns information into random noise (information loss)
- Background in thermodynamics: Movement of particles in a system



- Can this process be reverted?
  - Not always
  - However, the space of images that we want to be able to generate is limited

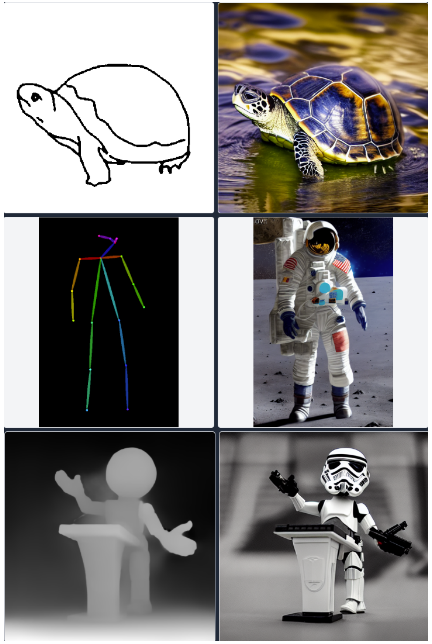
Diffusion Models

- Training a network to perform the backward process of diffusion
- Taking smaller steps makes this easier
- These steps are now deterministic



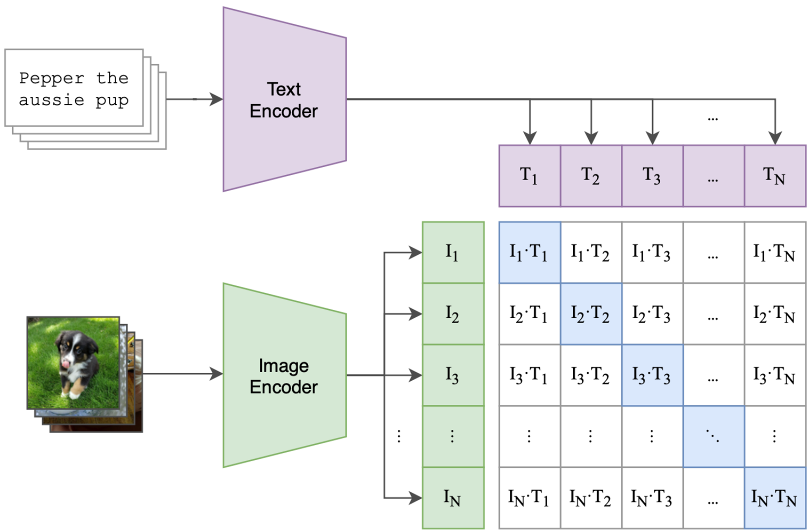
Conditioning Diffusion Models

- To be able to control the image generation, we will introduce additional information
- Consider relation of the image with information from another modality



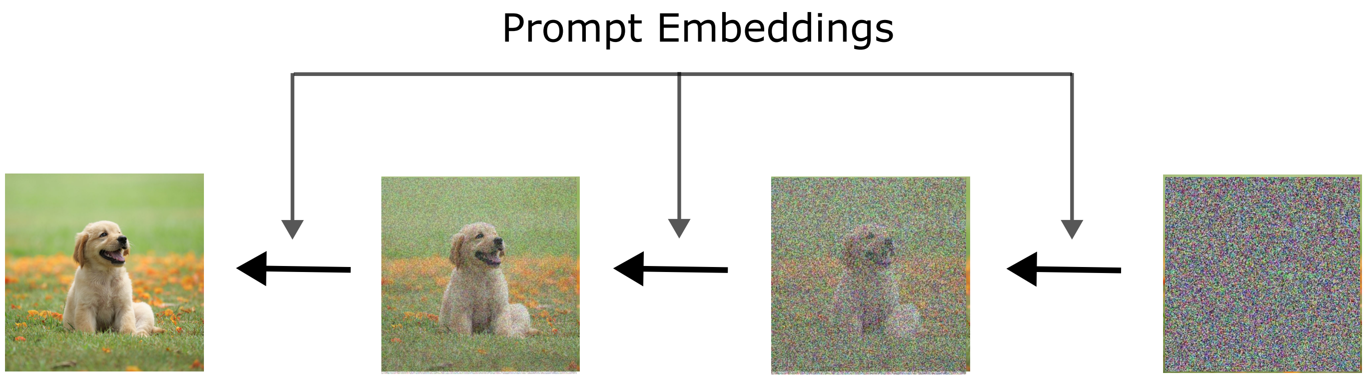
Conditioning Diffusion Models

- Idea: By operating a “simple” modality, the user gets control over a “difficult” modality (images)
- We use text as the input modality
- Need to convert text into a numerical representation

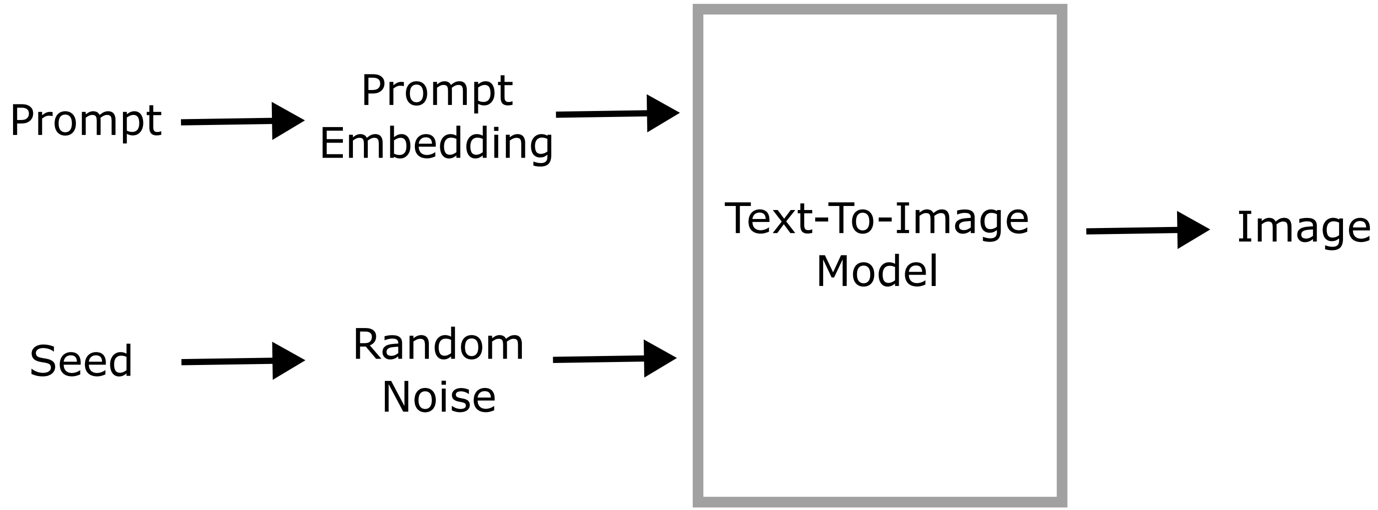


Conditioning Diffusion Models

- Provide text embedding information to the denoiser in each step
- Training on text-image pairs from web datasets (`alt` attributes)

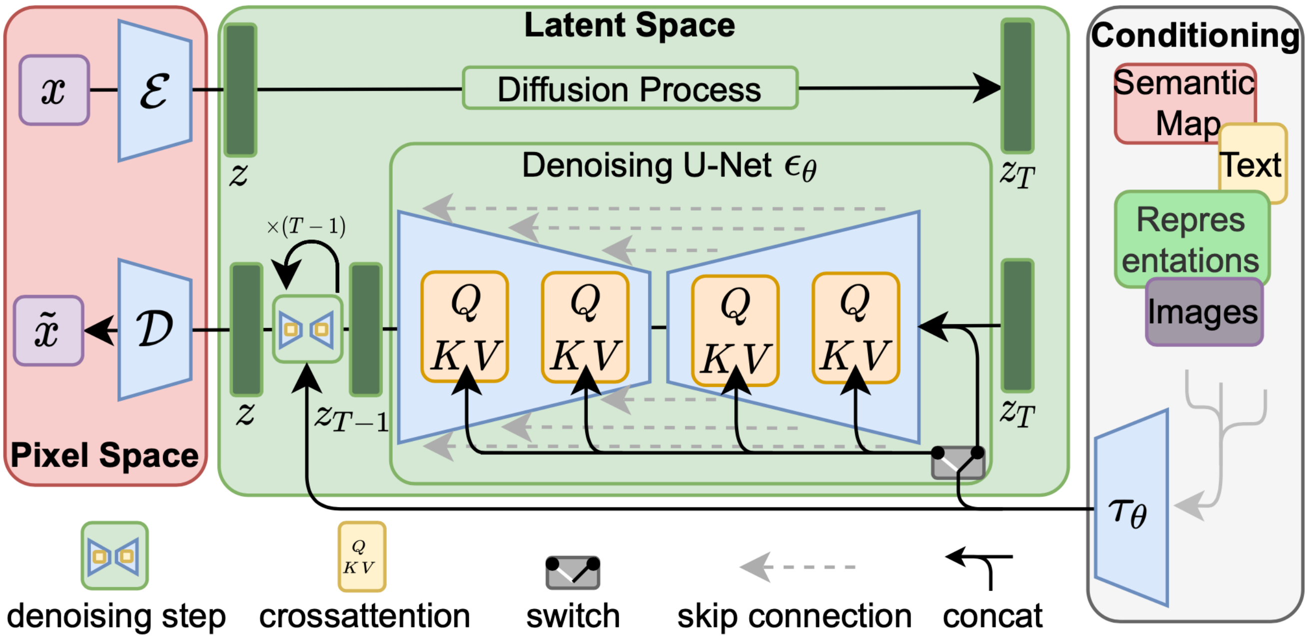


Generative Text-To-Image Models



Stable Diffusion

- Idea: Diffusion approach to generate latents of an autoencoder
- Provides good image quality, good control, good efficiency

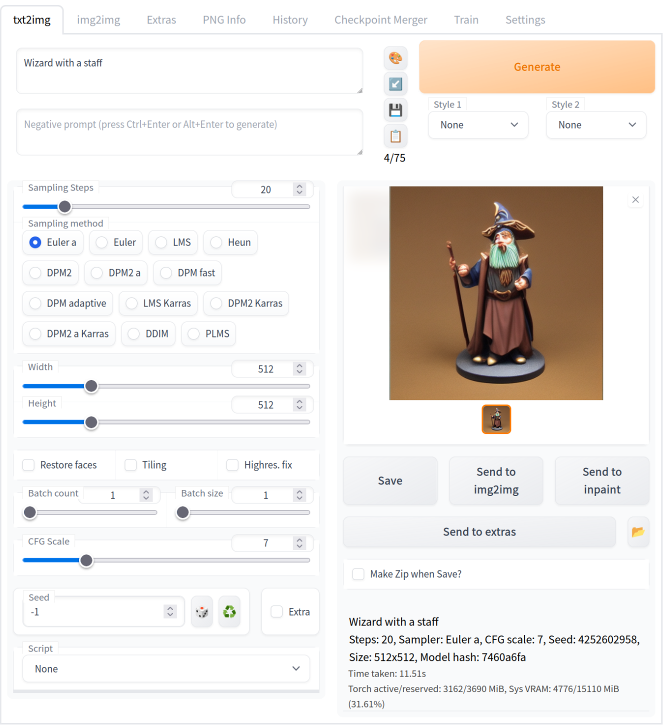


Stable Diffusion

- First large-scale model with publicly released weights (2022)
- Model weights can be used to
  - Use the model at home on own computer systems using GPU hardware
  - Build more advanced models without having to train a model from scratch (finetuning)
  - Develop approaches to add more functionalities

Stable Diffusion

- Building a community around the model
- Users develop and share plugins, modifications, best practices

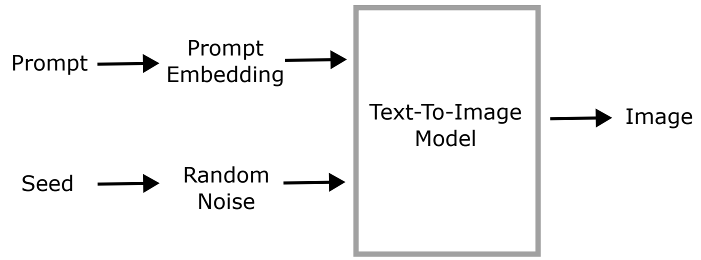


2

Prompt Engineering and the Infinite Index

Prompt Engineering

- Users only have directed control over the image generation by modifying the prompt and modifying the random seed



- Prompt contains all creative aspects
- Problem: Not obvious how prompts should be formulated to achieve the desired output
- Prompt engineering involves iteratively reformulating the prompt
- Usage of prompt modifiers like `4k high resolution award-winning image`

Prompt Engineering

User Behavior

- Two use cases:
- Descriptive approach:
- Creative approach:

User Behavior

- Two use cases:

- Descriptive approach:
  - The user has an idea of a fixed target image
  - Generates image that approximates their ideas

- Creative approach:

User Behavior

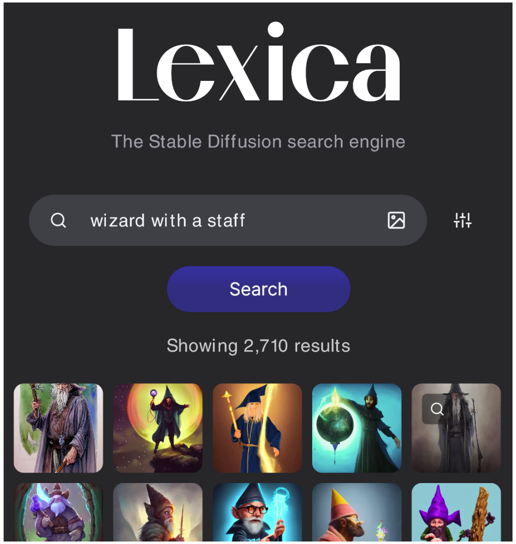
- Two use cases:

- Descriptive approach:
  - The user has an idea of a fixed target image
  - Generates image that approximates their ideas

- Creative approach:
  - The user has no clear vision or goal, but a set of constraints
  - Iteratively refines their prompt in a feedback-loop with random elements introduced by the system

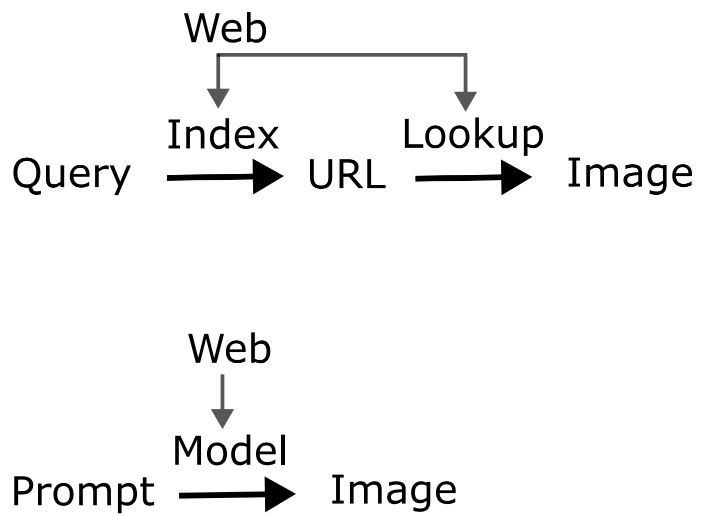
User Behavior

- This leads to two objectives:
  - The user needs fine-grained control over the input
  - The user wants to explore different aspects and get inspiration
- `lexica.art` as a search engine and an image feed



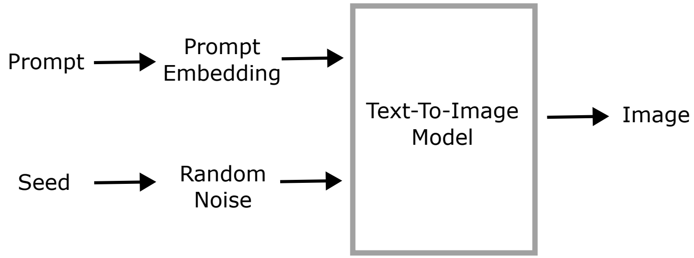
The Infinite Index

See image generation with a prompt as
image retrieval with a query, but on an infinite index



The Infinite Index

- Allows interpolation and extrapolation: Generate images that have never been indexed before



The Infinite Index

- Directly modify the numerical representation of the prompt in arbitrarily small steps

The Infinite Index

- Directly modify the numerical representation of the prompt in arbitrarily small steps

Prompt A
Seed 1

Prompt B
Seed 1

`beautiful mountain landscape, lake, snow, oil painting 8 k hd`
interpolated to
`a beautiful and highly detailed matte painting of the epic mountains of avalon, intricate details, epic scale, insanely complex, 8 k, sharp focus, hyperrealism, very realistic, by caspar friedrich, albert bierstadt, james gurney, brian froud,`

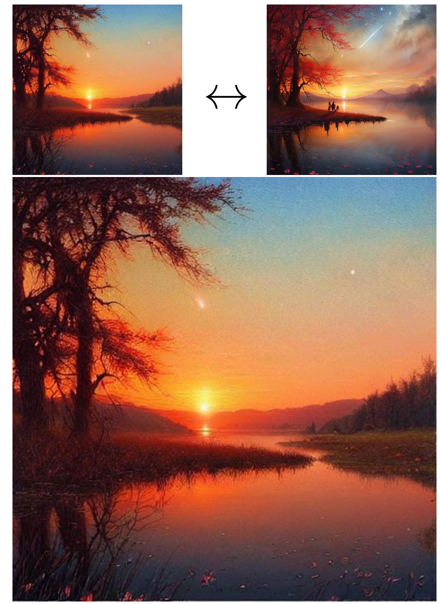
Prompt C
Seed 1

Prompt D
Seed 1

`Cute small squirrel sitting in a movie theater eating popcorn watching a movie ,unreal engine, cozy indoor lighting, artstation, detailed, digital painting,cinematic,character design by mark ryden and pixar and hayao miyazaki, unreal 5, daz, hyperrealistic, octane render`
interpolated to
`Cute small fox sitting [...]`

Prompt E
Seed 1

Prompt E
(modified)
Seed 1

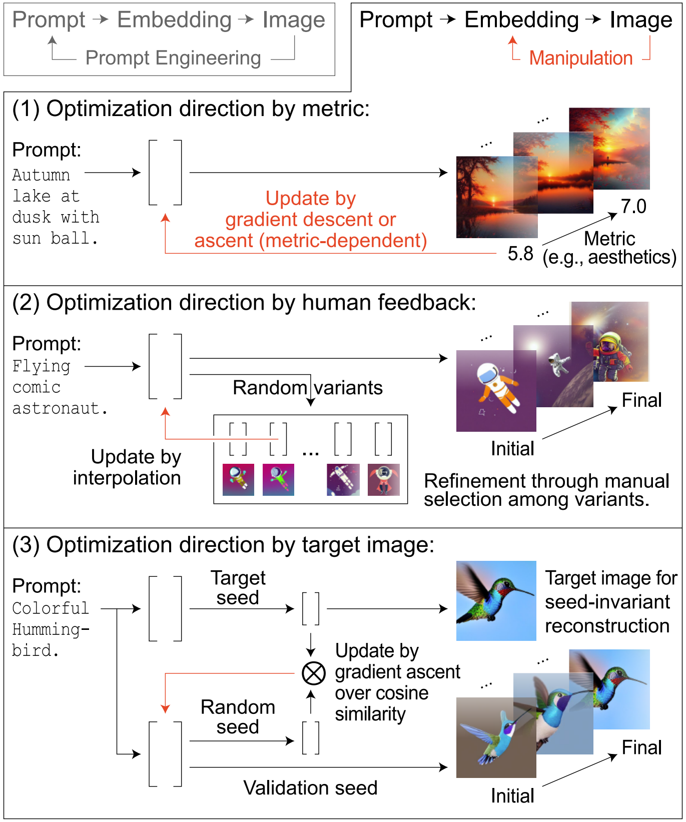


`a beautiful painting of a peaceful lake in the Land of the Dreams, full of grass, sunset, red horizon, starry-night!!!!!!!!!!!!!!!!!!!!, Greg Rutkowski, Moebius, Mohrbacher, peaceful, colorful`
and a gradient ascent on an aesthetic score

3

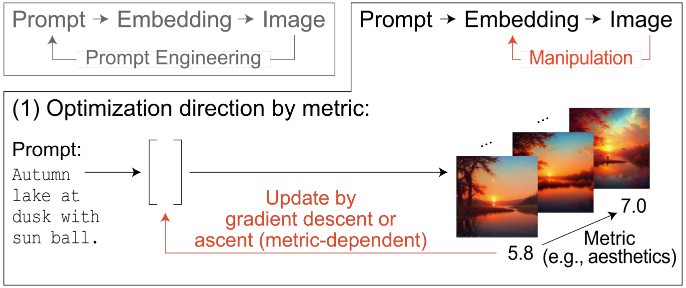
User-Centered Methods for Manipulating the Generated Images

Approaches for More User Control

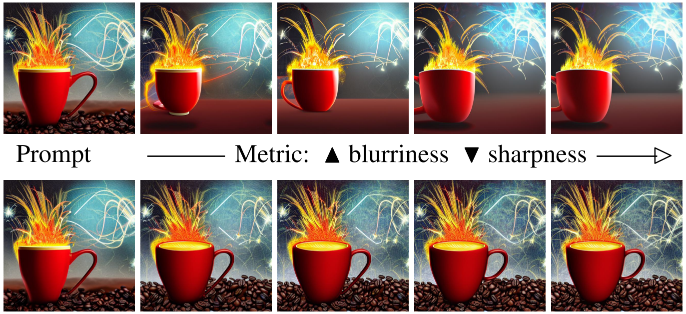


Metric-Based Optimization

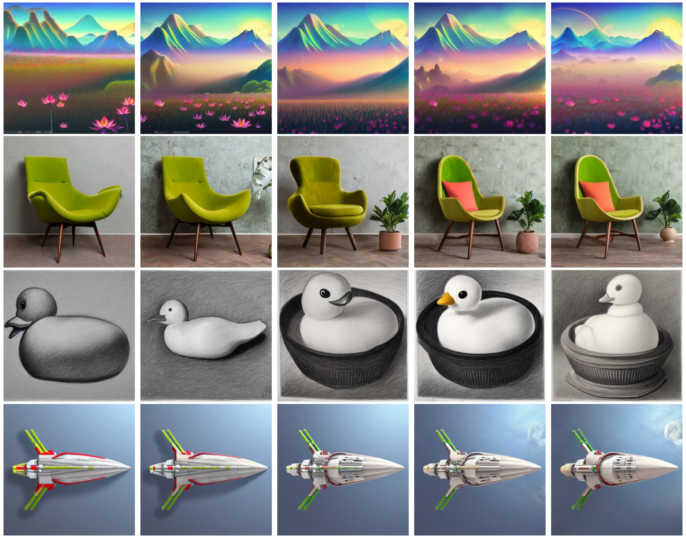
- Users apply arbitrary prompt modifiers like `4k high resolution award-winning image`
- Approach: Use a metric on the output for optimization



Metric-Based Optimization



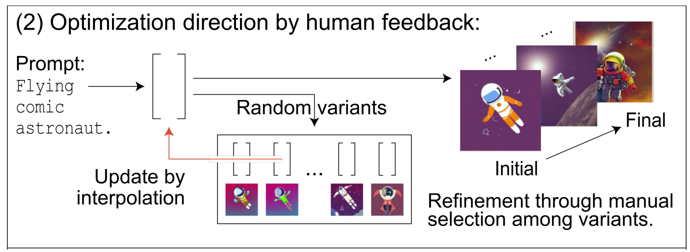
Metric-Based Optimization

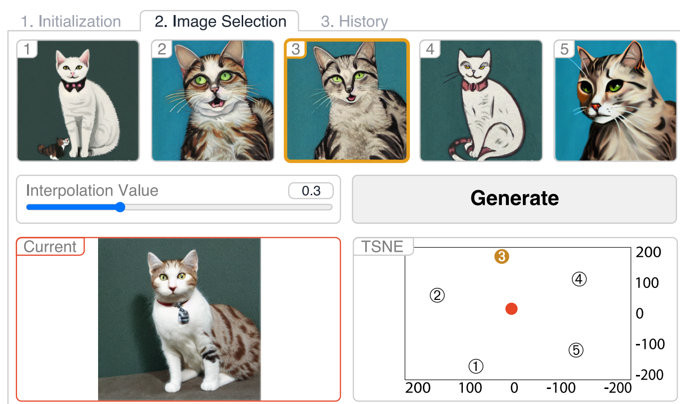


Prompts ——— Metric: aesthetics ———▷

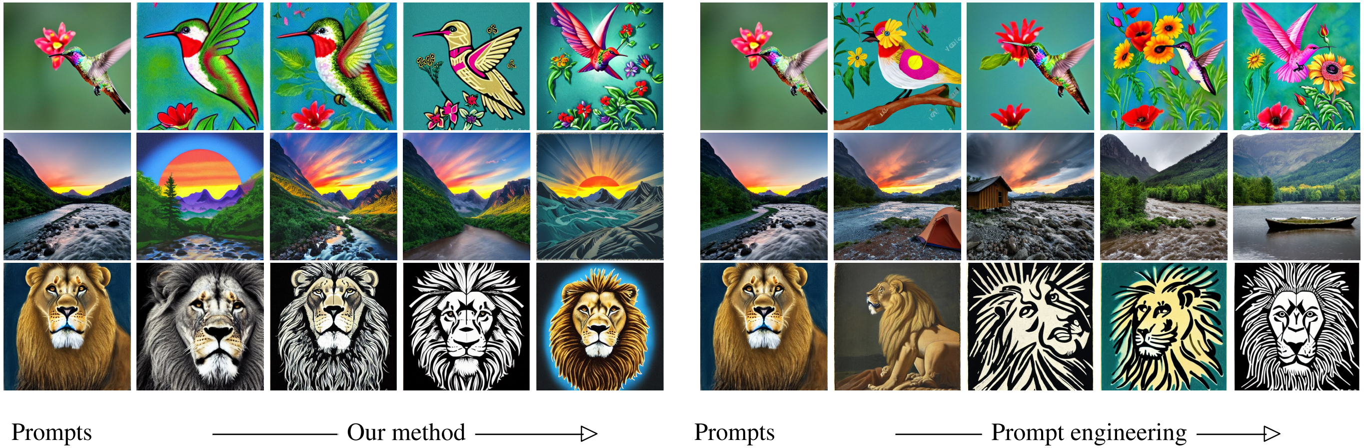
Iterative Human Feedback

- Users iteratively refine their prompt
- Approach: Allow movement through the prompt embedding space with random variants



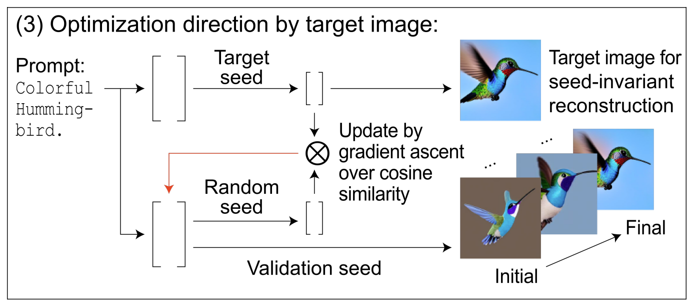


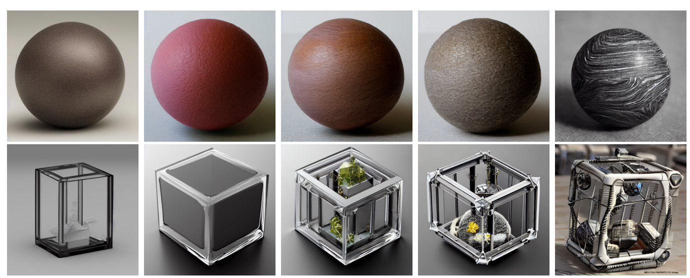
Iterative Human Feedback



Seed-Invariant Prompt Embeddings

- The used seed has a large effect on the generated image
- Approach: Use an automatic mechanism to transfer information from the image to the prompt embedding





Prompts (Target seed) —— Optimization ⟶ (Validation seed) (Another validation seed)

4

Generative Models and Feed-Based Platforms

Stock Images

- Symbolic images are used to describe an abstract scenario instead of a specific situation
- Similar to generated images
- Difference between literal description and hidden meaning of stock images
- Attempts to extract hidden meaning from given images
- Might be used to automatically illustrate sites like Medium.com

Generated Social Media

- Generative approaches can be used to generate image-based feeds like TikTok or Instagram
- Implications on perceived relevance?
  Do users care as long as they feel entertained?
- Algorithms can iteratively optimize feeds based on user feedback
- On the other hand: Algorithms can introduce serendipity to otherwise monothematic and highly optimized feeds (Similarly to lexica.art)

Generative Models for Advertisement

- Models might be used to generated individual advertisements based on user interests and feed context
- Modifying the infinite index directly: Prompts like `delicious food` can be directed towards specific brands

Generative Models as Custom Tools

- Generative models might support users to express themselves using custom images or other high-level modalities (GIFs, videos, sounds)
- Might include symbolic images in blogs or chats
- Help to express emotions
- Need for tools to generate images beyond prompt formulation

Future Research: Retrievability

- We noticed varying speed in the image space when interpolating two images
- Some images are “easier” to get than others
- Images with different probabilities
- Implications on relevance?

5

Backup

The Infinite Index

- Every query yields (some) results
- For each query, there exist infinitely many possible results
  with different random seeds

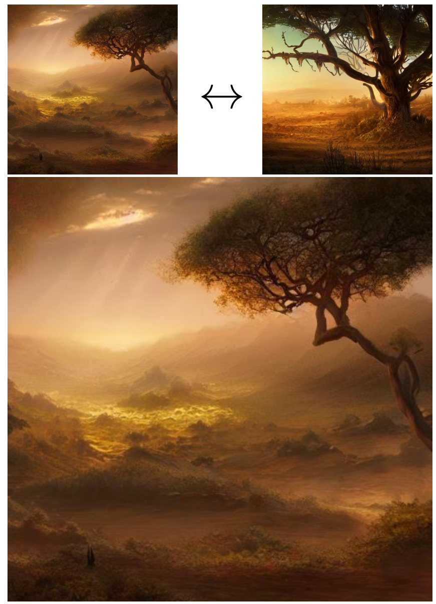
The Infinite Index

- Every query yields (some) results
- For each query, there exist infinitely many possible results
  with different random seeds

Seed 1 Seed 2

Golden treehouse in lush forest, better homes and hardens magazine, big glass windows, intricate woodworking, polaroid

Seed 1 Seed 2



ethiopian landscape, highly detailed, digital painting, concept art, sharp focus, cinematic lighting, diffuse lighting, fantasy, intricate, elegant, lifelike, photorealistic, illustration, smooth

Seed 1 Seed 2

a cybernetic samoyed and beagle, concept art, detailed face and body, detailled decor, fantasy, highly detailed, cinematic lighting, digital art painting, winter, nature, running

Backup: Interpolating Seeds