Stable Diffusion
Recap: Large Language Models
SpeakUp

Join by URL: https://web.speakup.info/room/join/80740

Join by key: web.speakup.info → use key 80740

Join by saved session: web.speakup.info

Join by QR:
• Some background on the diffusion concept

• Common applications of stable diffusion

• Stable diffusion intuition
Main Generative Neural Networks

Image
• Stable diffusion (uses a variational autoencoder VAE)
• Generative adversarial networks (comprises a generator and a descriminator)
• Transformers (e.g., OpenAI’s DALL·E 2)

Text (e.g., summaries, code, answers)
• Transformers (large language models, e.g., ChatGPT or FauxPilot)

Audio (e.g., speech and music)
• WaveNet (based on CNNs)
• Generative adversarial networks
• Transformers (Microsoft’s VALL-E or Bark)
• Stable diffusion (via spectrograms)
• Variational Autoencoders (VAEs)

A Viral AI-Generated Drake Song by ‘Ghostwriter’ Has Millions of Listens

An astronaut riding a horse in photorealistic style. – DALL·E 2

https://beta.elevenlabs.io/
Stable Diffusion Tools

• **AIVA** (music generation)
• **Astria** (fine-tuning and generation)
• **Civitai** (showcase and hosting)
• **Hugging Face** (fine-tuning, generation, and hosting)
• **DreamStudio** (generation, hosting, and editing)
• **Lexica** (prompt engineering and generation)

• You can find more here: [https://pharmapsychotic.com/tools.html](https://pharmapsychotic.com/tools.html)
Applications

Typically, Text-to-Image (and more recently *-to-Audio)

• Input a text prompt and output an image

An astronaut riding a horse.
Applications

Typically, Text-to-Image (and more recently *-to-Audio)

• Input a text prompt and output an image
• Input a text prompt and an image and output an image

"Oil painting of wolf howling at the moon by Van Gogh"
Applications

Typically, Text-to-Image (and more recently *-to-Audio)

• Input a text prompt and output an image
• Input a text prompt and an image and output an image
• Generate new variants based on a few examples (text inversion)
Applications

Typically, Text-to-Image (and more recently *-to-Audio)

• Input a text prompt and output an image
• Input a text prompt and an image and output an image
• Generate new variants based on a few examples (text inversion)
• Many more applications... it seems we’re finding new applications every day
Generation Terms/Applications

- **Text-to-image:** generate an image corresponding to the given text
- **Image-to-image:** seed text-to-image with an initial drawing or photo
- **Inpainting:** replace sections of image with generated pixels
- **Outpainting:** generate pixels surrounding original image (outside frame)
- **Restoration:** remove creases and add colors
- **Upscaling:** generate a higher resolution image from the original
- **Tiling:** generate tiled image sets that align with one another
- **Textual inversion:** customize model output for a specific subject or style
Quick Note on GANs
Stable Diffusion

Images from “High-Resolution Image Synthesis with Latent Diffusion Models” by Robin Rombach et al.
A central problem in machine learning involves modeling complex data-sets using highly flexible families of probability distributions in which learning, sampling, inference, and evaluation are still analytically or computationally tractable.

Here, we develop an approach that simultaneously achieves both flexibility and tractability.

The essential idea, inspired by non-equilibrium statistical physics, is to systematically and slowly destroy structure in a data distribution through an iterative forward diffusion process.

We then learn a reverse diffusion process that restores structure in data, yielding a highly flexible and tractable generative model of the data.

This approach allows us to rapidly learn, sample from, and evaluate probabilities in deep generative models with thousands of layers or time steps, as well as to compute conditional and posterior probabilities under the learned model.

We additionally release an open source reference implementation of the algorithm.
A central problem in machine learning involves modeling complex data-sets using highly flexible families of probability distributions in which learning, sampling, inference, and evaluation are still analytically or computationally tractable. Here, we develop an approach that simultaneously achieves both flexibility and tractability.

The essential idea, inspired by non-equilibrium statistical physics, is to systematically and slowly destroy structure in a data distribution through an iterative forward diffusion process. We then learn a reverse diffusion process that restores structure in data, yielding a highly flexible and tractable generative model of the data. This approach allows us to rapidly learn, sample from, and evaluate probabilities in deep generative models with thousands of layers or time steps, as well as to compute conditional and posterior probabilities under the learned model.

We additionally release an open source reference implementation of the algorithm.
A central problem in machine learning involves modeling complex data-sets using highly flexible families of probability distributions in which learning, sampling, inference, and evaluation are still analytically or computationally tractable.

Here, we develop an approach that simultaneously achieves both flexibility and tractability. The essential idea, inspired by non-equilibrium statistical physics, is to systematically and slowly destroy structure in a data distribution through an iterative forward diffusion process. We then learn a reverse diffusion process that restores structure in data, yielding a highly flexible and tractable generative model of the data.

This approach allows us to rapidly learn, sample from, and evaluate probabilities in deep generative models with thousands of layers or time steps, as well as to compute conditional and posterior probabilities under the learned model.

We additionally release an open source reference implementation of the algorithm.

Idea: Turning a Classifier Into a Generator
Learn to Extract the Added Noise
Learn to Extract the Added Noise

Variational Autoencoder
Stable Diffusion Training Diagram
Inference

Input: text prompt

1. Tokenize input text to produce and index for each word
2. Process indices with an encoder to produce embeddings
3. Generate an initial image-like matrix (latents) using the normal distribution
4. Create a noise scheduler (how many times do we update the output)
5. Generate output image-like matrix iteratively
6. Generate the output by passing the image-like matrix to the VAE
Resources

- https://github.com/huggingface/diffusion-models-class
- https://github.com/lllyasviel/ControlNet
- https://github.com/Stability-AI/stablediffusion
- https://pharmapsychotic.com/tools.html
- https://lexica.art/
- https://deepfloyd.ai/
- https://invoke-ai.github.io/InvokeAI/

- High-Resolution Image Synthesis with Latent Diffusion Models