CS492D: Diffusion Models and Their Applications

Course Introduction

LECTURE 1
MINHYUK SUNG

Fall 2024
KAIST

Before we get started…

- Please make sure that you're in the right (virtual) classroom.
  - This is CS492(D): Diffusion Models and Their Applications course.
- Please check whether you name is correct in the Zoom chat.
- Please turn on your camera but turn off the mic.


air head · Made by shy kids with Sora


OpenAI Sora

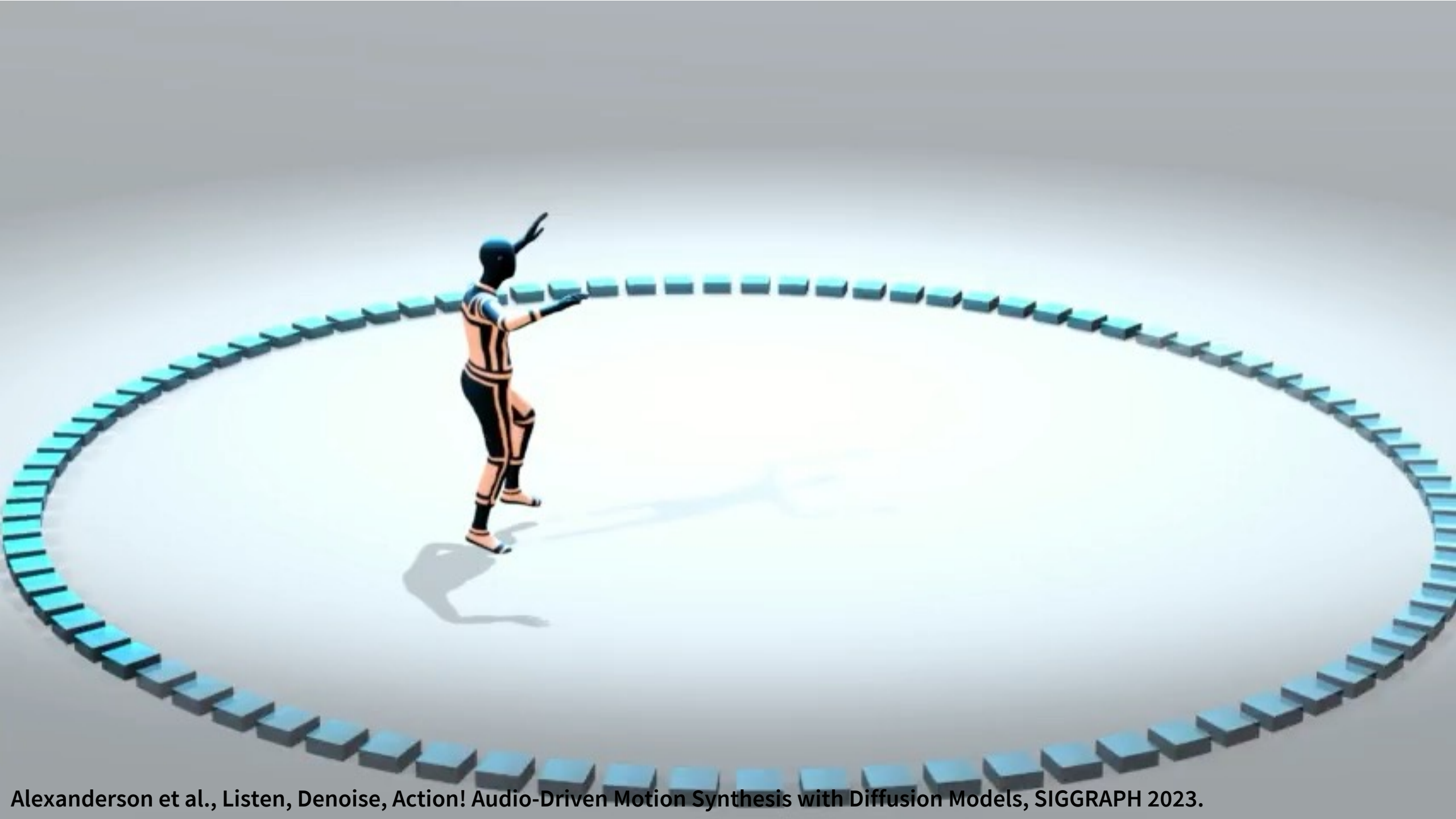

Luma Dream Machine

Text input: A traditional Irish fiddle playing a lively reel.

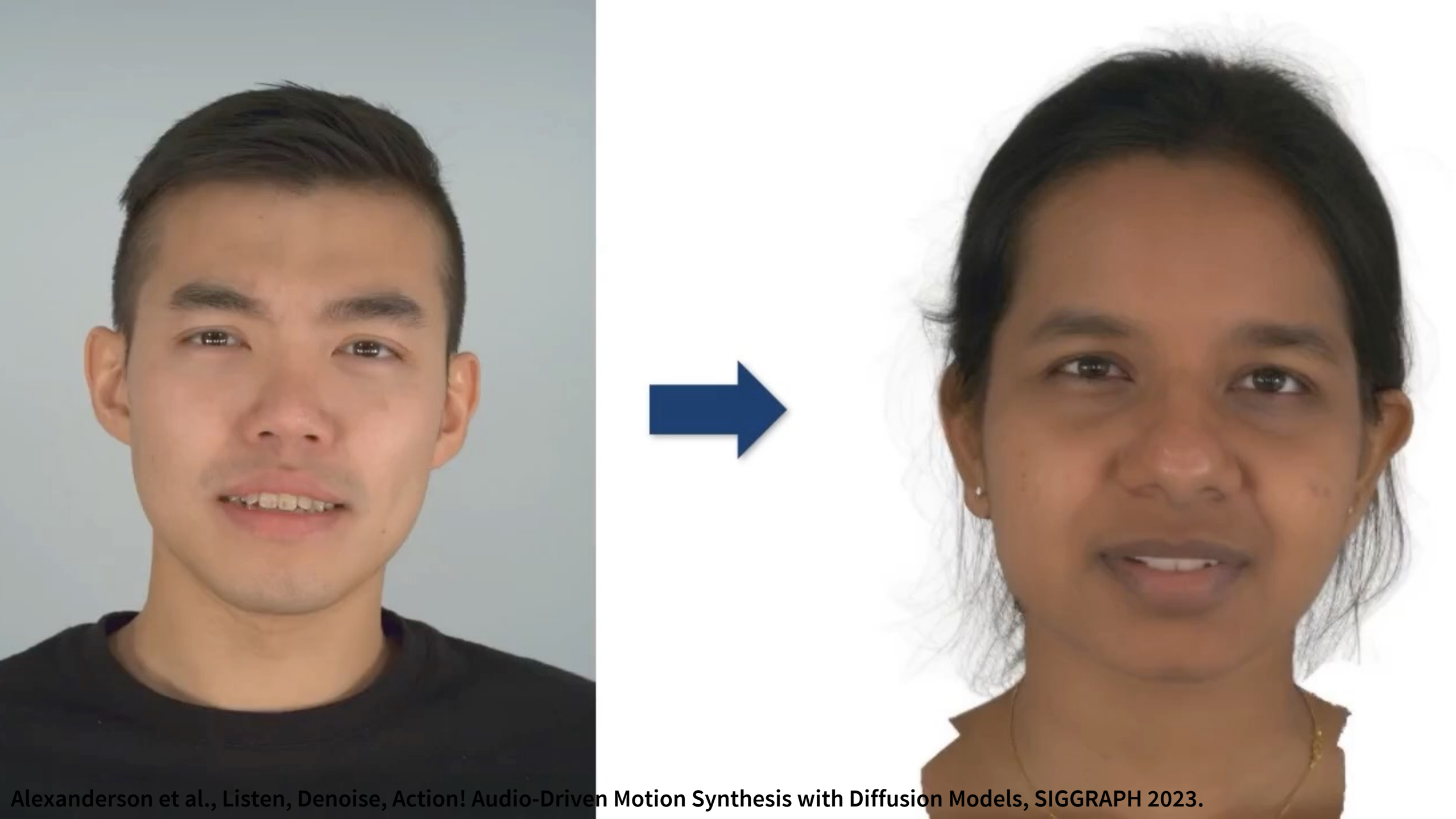
Up Next: The sound of a light saber

Meta Audio Box

AI research by Meta



Alexanderson et al., Listen, Denoise, Action! Audio-Driven Motion Synthesis with Diffusion Models, SIGGRAPH 2023.



Alexanderson et al., Listen, Denoise, Action! Audio-Driven Motion Synthesis with Diffusion Models, SIGGRAPH 2023.

Black Forest Labs Flux

Image Generation

https://huggingface.co/spaces/black-forest-labs/FLUX.1-schnell

Generate any image you want (with any prompt) and post it on the #lecture-01 channel on Slack.

Image Generation Example

A classroom for "Diffusion Models and Their Applications" with many students in the fall semester. On the blackboard, it is written as "Diffusion Models and Their Applications."

Adobe Firefly

Jason Allen won the digital-art competition at the Colorado State Fair last year for his piece “Théâtre D’opéra Spatial” that he created using the AI software Midjourney. Recently, the US Copyright Office refused to grant him a copyright for his piece, writing, “We have decided that we cannot register this copyright claim because the deposit does not contain any human authorship.” He plans to appeal.


demonflyingfox

Wang et al., ProlificDreamer: High-Fidelity and Diverse Text-to-3D Generation with Variational Score Distillation, arXiv 2023.

Tang et al., LGM: Large Multi-View Gaussian Model for High-Resolution 3D Content Creation, ECCV 2024.

3D Generation

Tang et al., LGM: Large Multi-View Gaussian Model for High-Resolution 3D Content Creation, ECCV 2024.

3D Generation

“motorcycle”

“mech suit”

“ghost lantern”

“furry fox head”

“dresser”

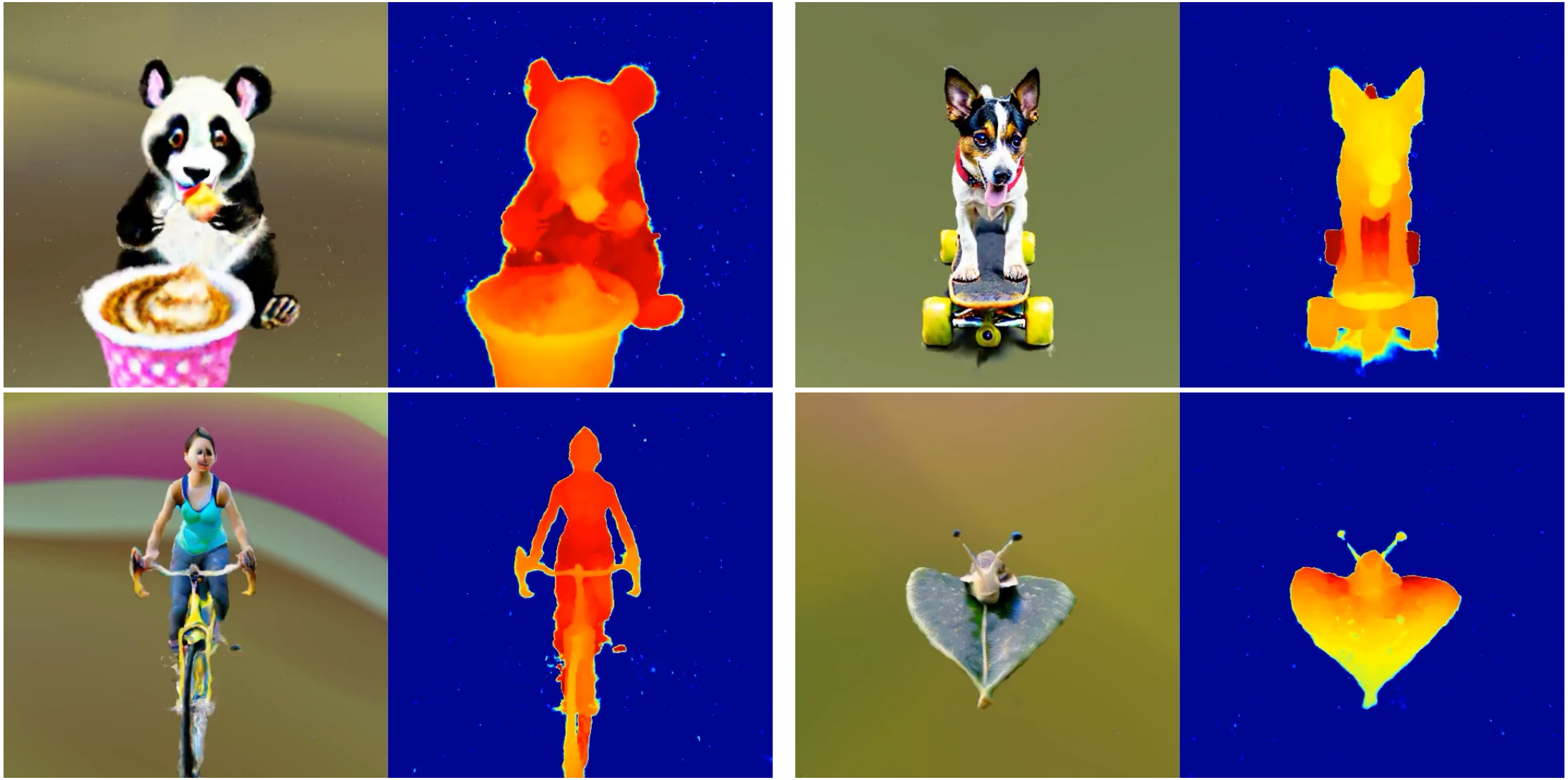
“swivel chair”

“astronaut”

“mushroom house”

Zheng et al., Dream-in-4D: A Unified Approach for Text- and Image-guided 4D Scene Generation, CVPR 2024.

4D Generation



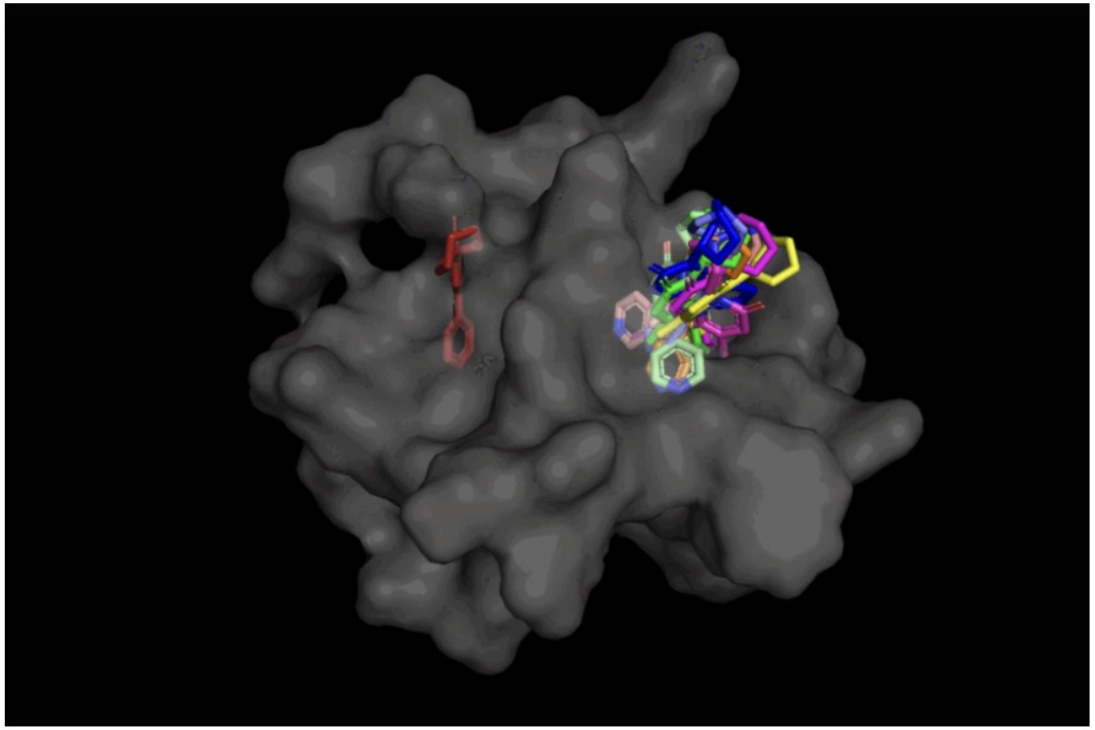
https://news.mit.edu/2023/speeding-drug-discovery-with-diffusion-generative-models-diffdock-0331

Drug Discovery

Speeding up drug discovery with diffusion generative models

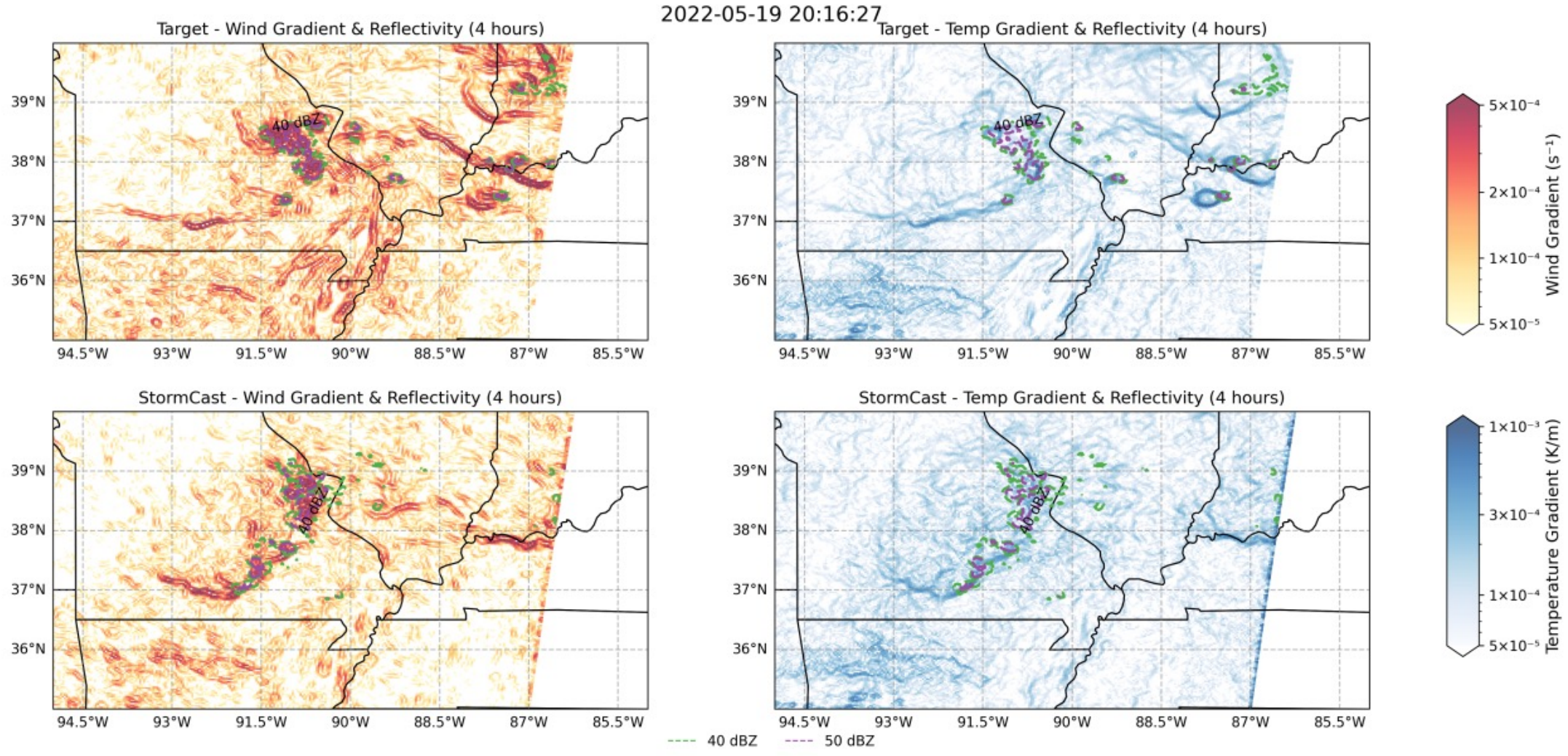
MIT researchers built DiffDock, a model that may one day be able to find new drugs faster than traditional methods and reduce the potential for adverse side effects.

Alex Ouyang | Abdul Latif Jameel Clinic for Machine Learning in Health
March 31, 2023



https://research.nvidia.com/publication/2024-08_kilometer-scale-convection-allowing-model-emulation-using-generative-diffusion

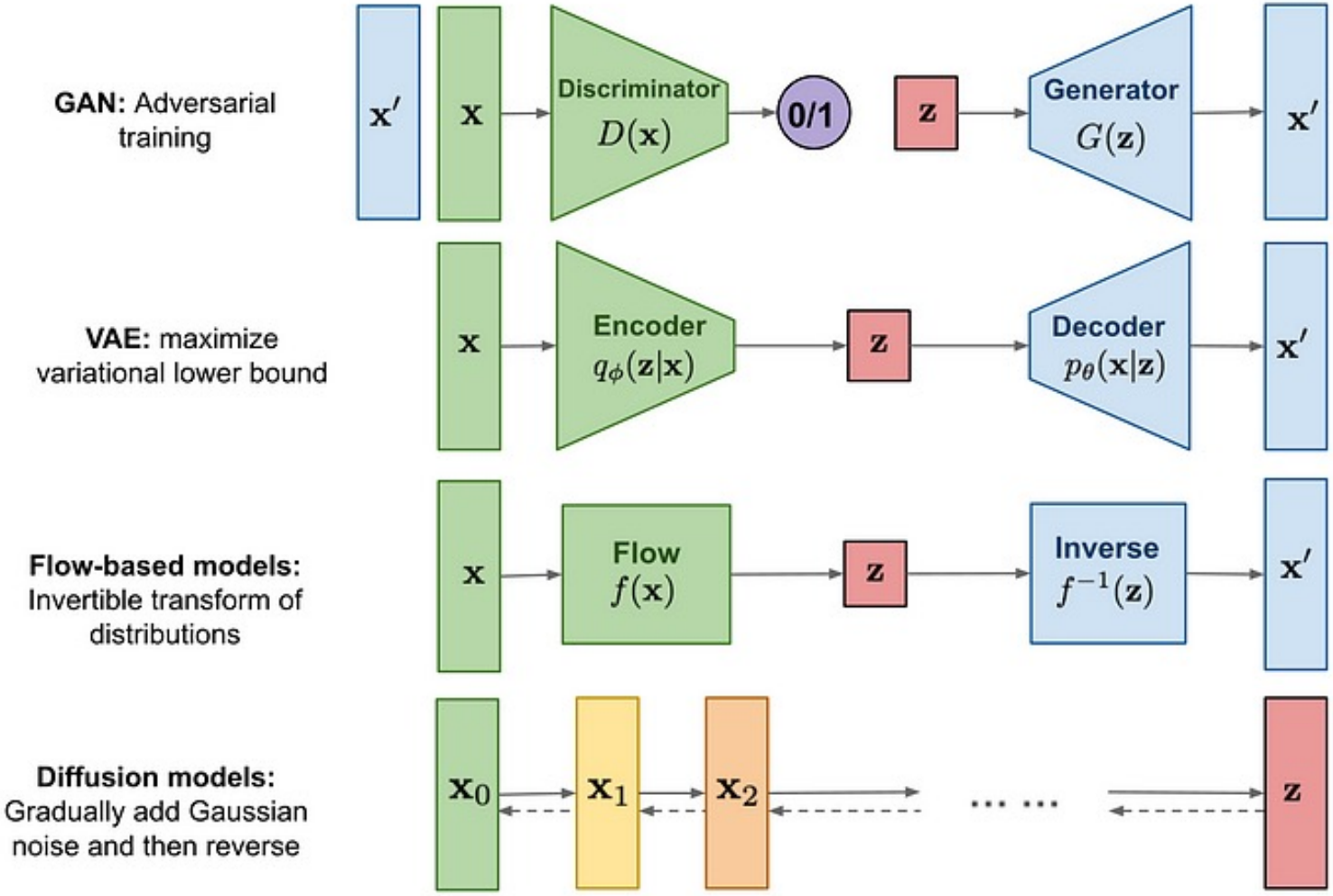
Weather Forecasting



A Brief Overview of the Course

https://lilianweng.github.io/posts/2021-07-11-diffusion-models/

Generative Models

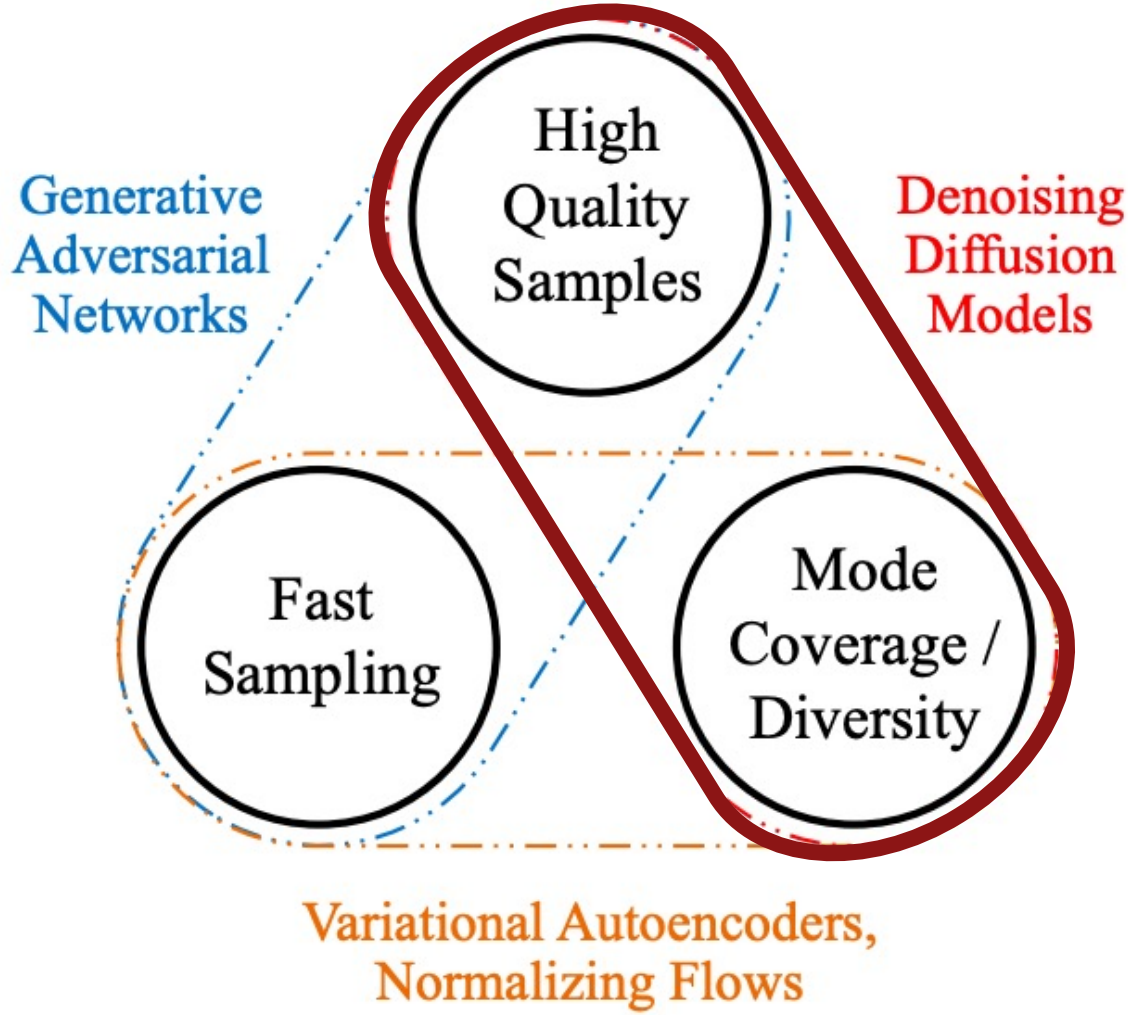


Kerras et al., Analyzing and Improving the Image Quality of StyleGAN.

StyleGAN2

https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/

Generative Models – Comparison



Higher Quality & Diversity

Diffusion Models Beat GANs on Image Synthesis

Prafulla Dhariwal*
OpenAI
prafulla@openai.com

Alex Nichol*
OpenAI
alex@openai.com

Abstract

We show that diffusion models can achieve image sample quality superior to the current state-of-the-art generative models. We achieve this on unconditional image synthesis by finding a better architecture through a series of ablations. For conditional image synthesis, we further improve sample quality with classifier guidance: a simple, compute-efficient method for trading off diversity for fidelity using gradients from a classifier. We achieve an FID of 2.97 on ImageNet 128×128, 4.59 on ImageNet 256×256, and 7.72 on ImageNet 512×512, and we match BigGAN-deep even with as few as 25 forward passes per sample, all while maintaining better coverage of the distribution. Finally, we find that classifier guidance combines well with upsampling diffusion models, further improving FID to 3.94 on ImageNet 256×256 and 3.85 on ImageNet 512×512. We release our code at https://github.com/openai/guided-diffusion.

Diffusion Models

- (+) High quality
- (+) Diversity
- (–) Slow

Diffusion Models

The generation process of a diffusion model is a denoising process.

Song et al., Score-Based Generative Modeling through Stochastic Differential Equations, ICLR 2021.

Score-Based Model / DDPM / DDIM

Forward SDE (data → noise)

$\mathbf{x}(0)$ — $\mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w}$ → $\mathbf{x}(T)$

score function

$\mathbf{x}(0)$ ← $\mathrm{d}\mathbf{x} = \left[\mathbf{f}(\mathbf{x}, t) - g^2(t)\nabla_{\mathbf{x}} \log p_t(\mathbf{x})\right] \mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}}$ — $\mathbf{x}(T)$

Reverse SDE (noise → data)

Diffusion Models

- (+) High quality
- (+) Diversity
- (–) Slow

- (+)(Relatively) easy to implement and train
- (+) Easy to convert a conditional model
- (+) Easy to personalize
- (+) Easy to distill knowledge
- …

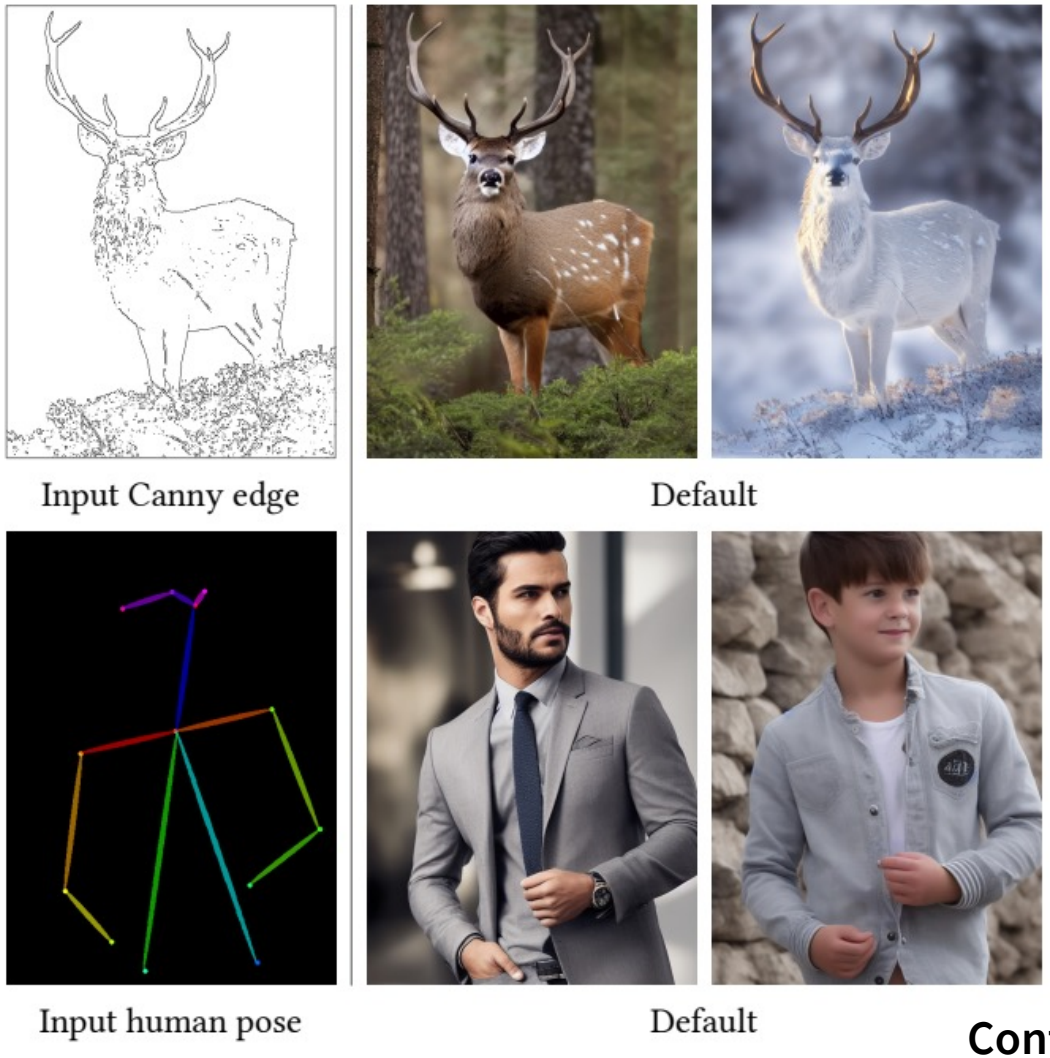
Ho and Salimans, Classifier-Free Diffusion Guidance, NeurIPS 2021 Workshop.

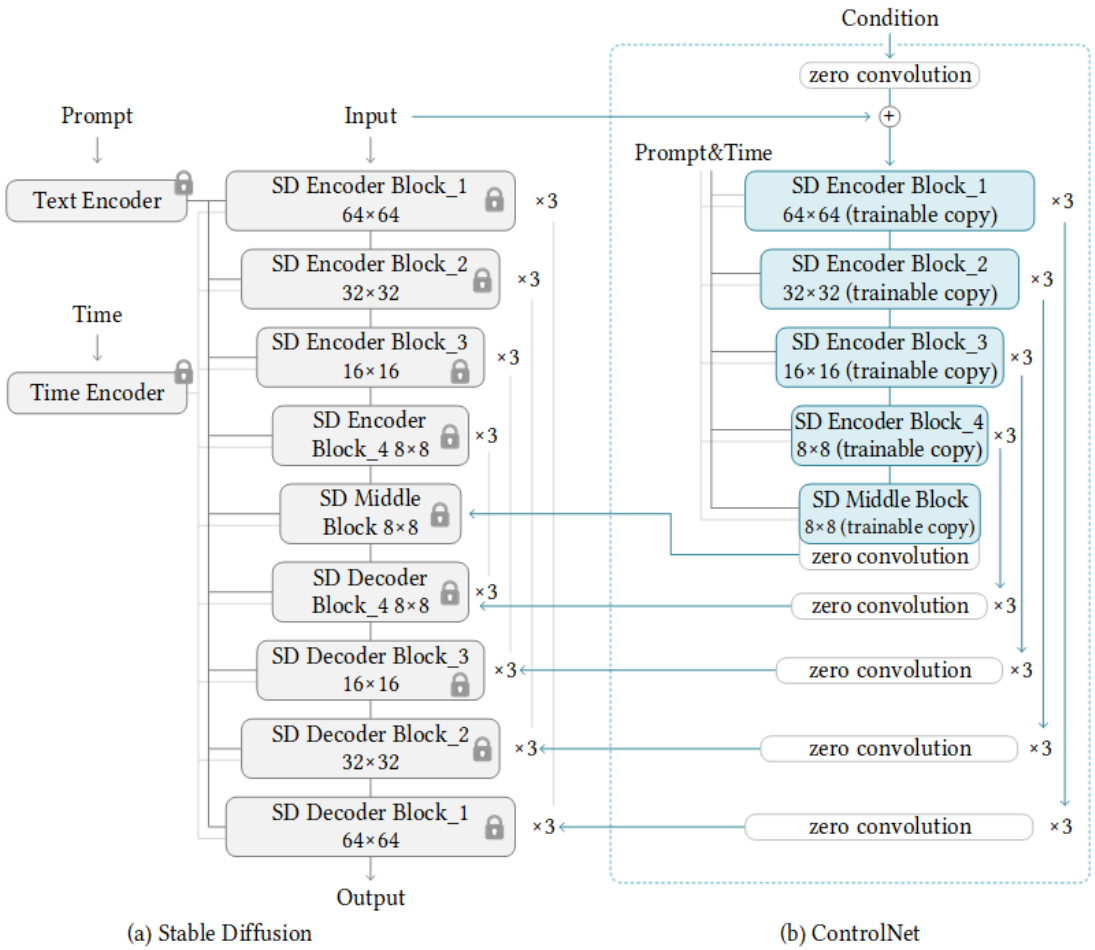
Leveraging Class Labels or Prompts

Classifer Guidance / Classifer-Free Guidance

Zhang et al., Adding Conditional Control to Text-to-Image Diffusion Models, ICCV 2023.

Conditional Generation





ControlNet

https://www.shruggingface.com/blog/self-portraits-with-stable-diffusion-and-lora

Stylization

LoRA

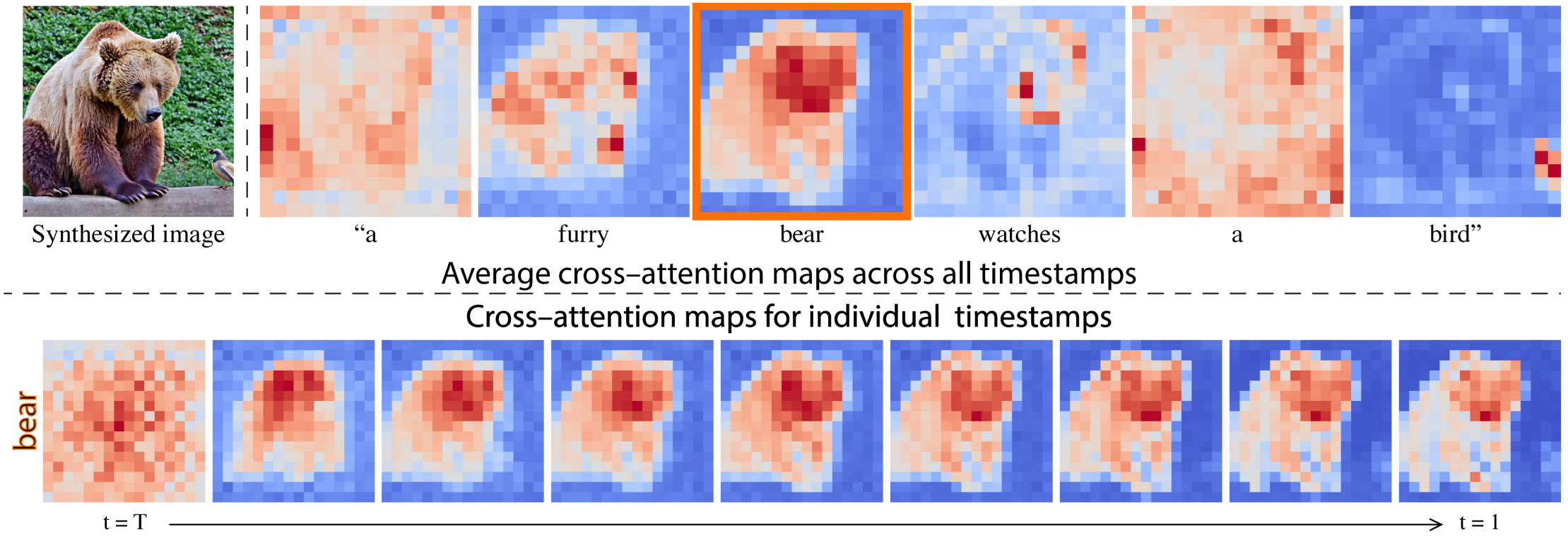
Ruiz et al., DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation, CVPR 2023.

Personalization

DreamBooth

Hertz et al., Prompt-to-Prompt Image Editing with Cross-Attention Control, 2022.

The Power of Attention-Layer

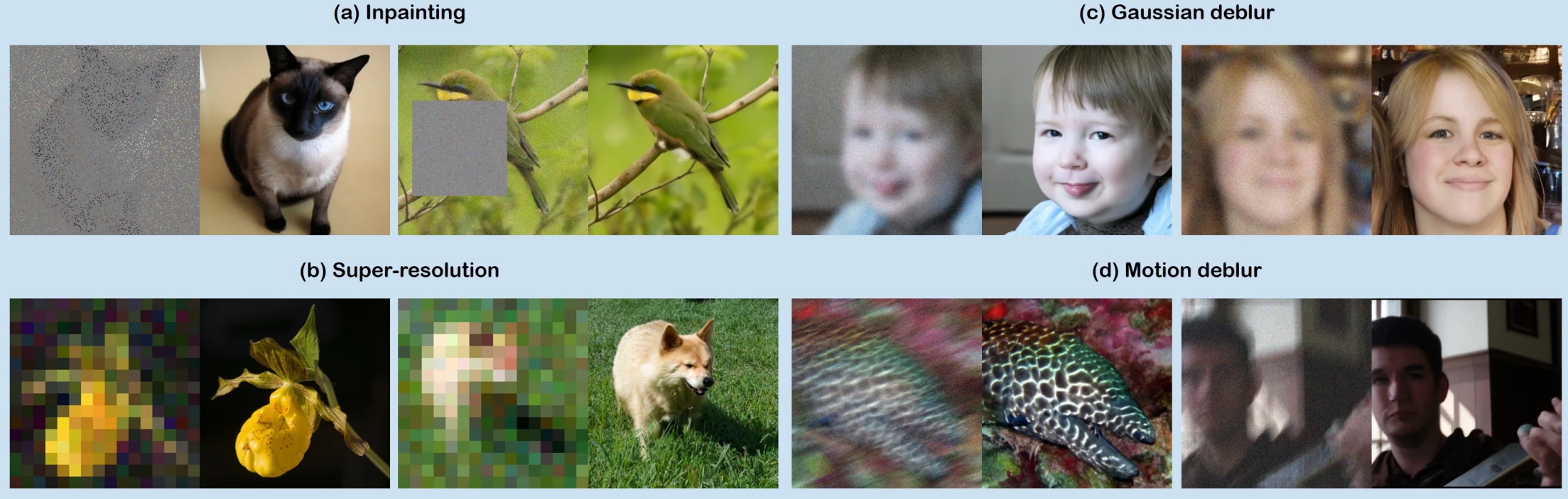


Hertz et al., Prompt-to-Prompt Image Editing with Cross-Attention Control, 2022.

The Power of Attention-Layer

Chung et al., Diffusion Posterior Sampling for General Noisy Inverse Problems, ICLR 2023.

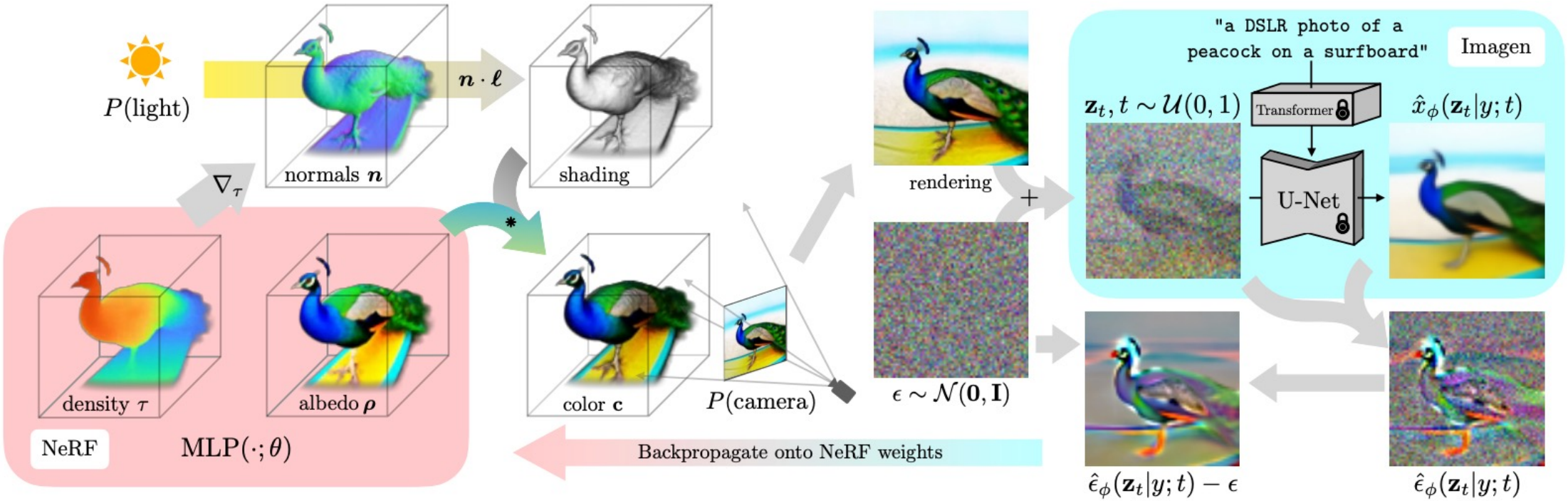
Inverse Problems



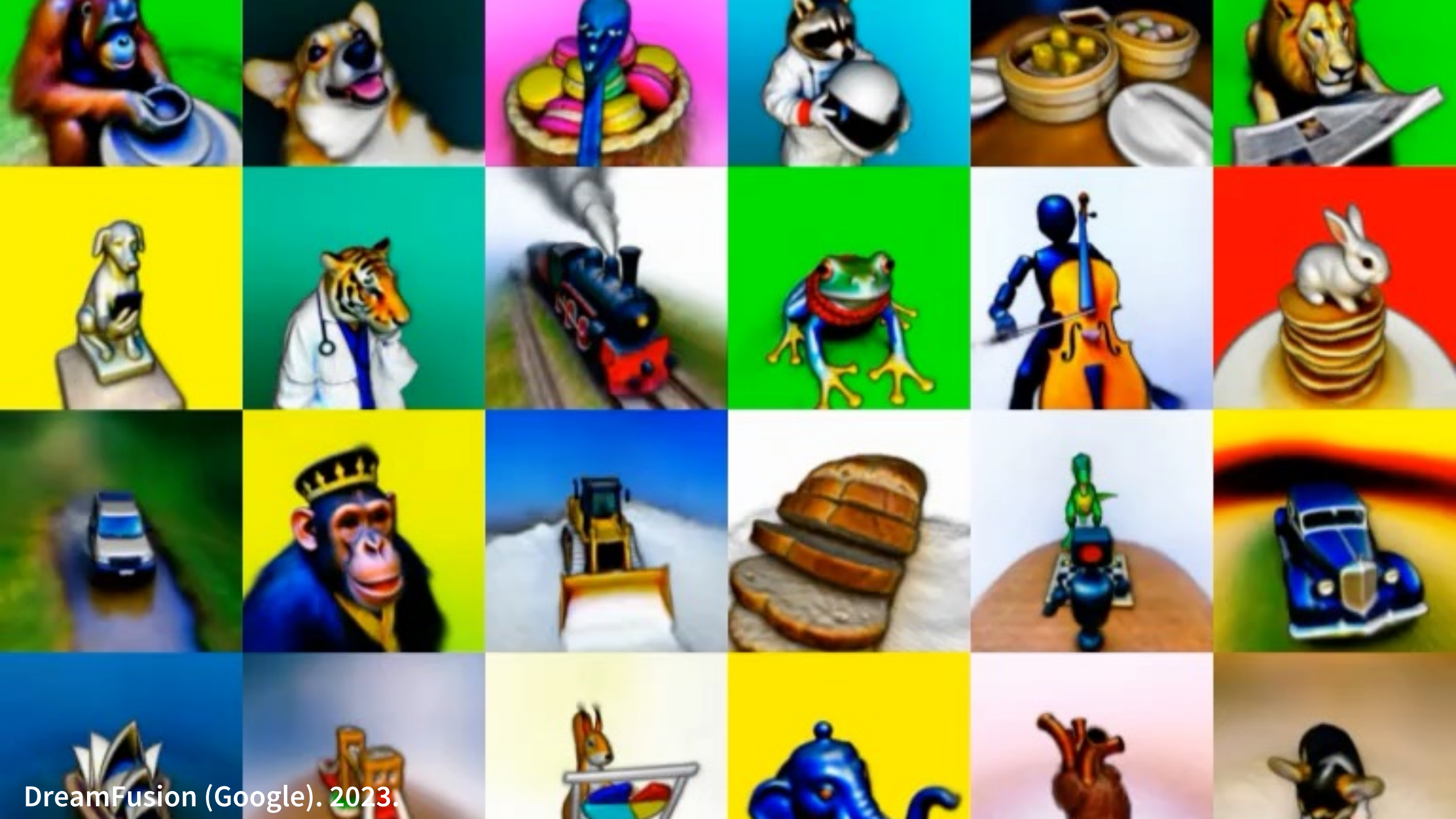
Diffusion Posterior Sampling

Poole et al., DreamFusion: Text-to-3D using 2D Diffusion, ICLR 2023.

Knowledge Distillation



Score Distillation Sampling



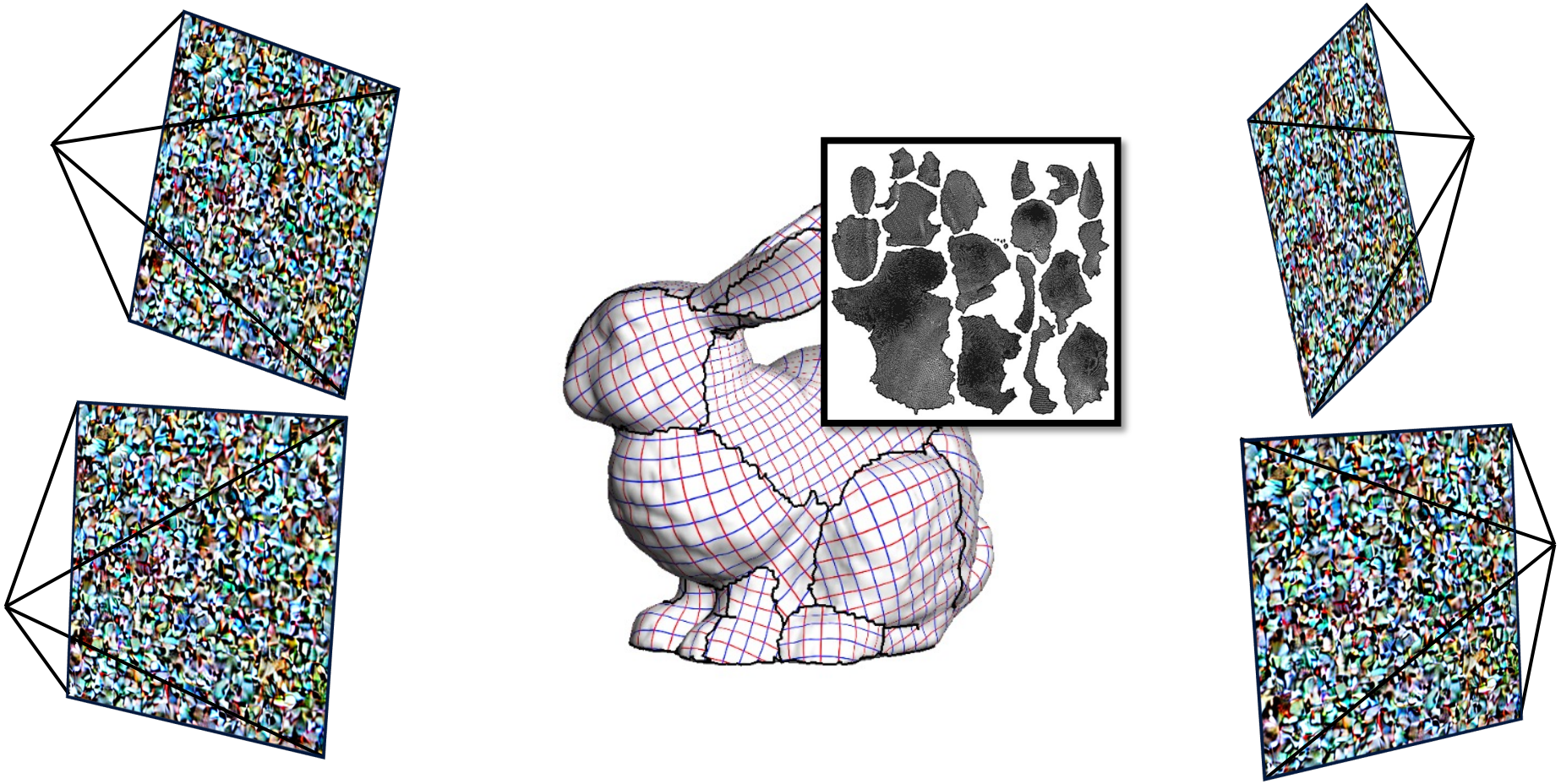
DreamFusion (Google). 2023.

Koo et al., Posterior Distillation Sampling, CVPR 2024.

Knowledge Distillation

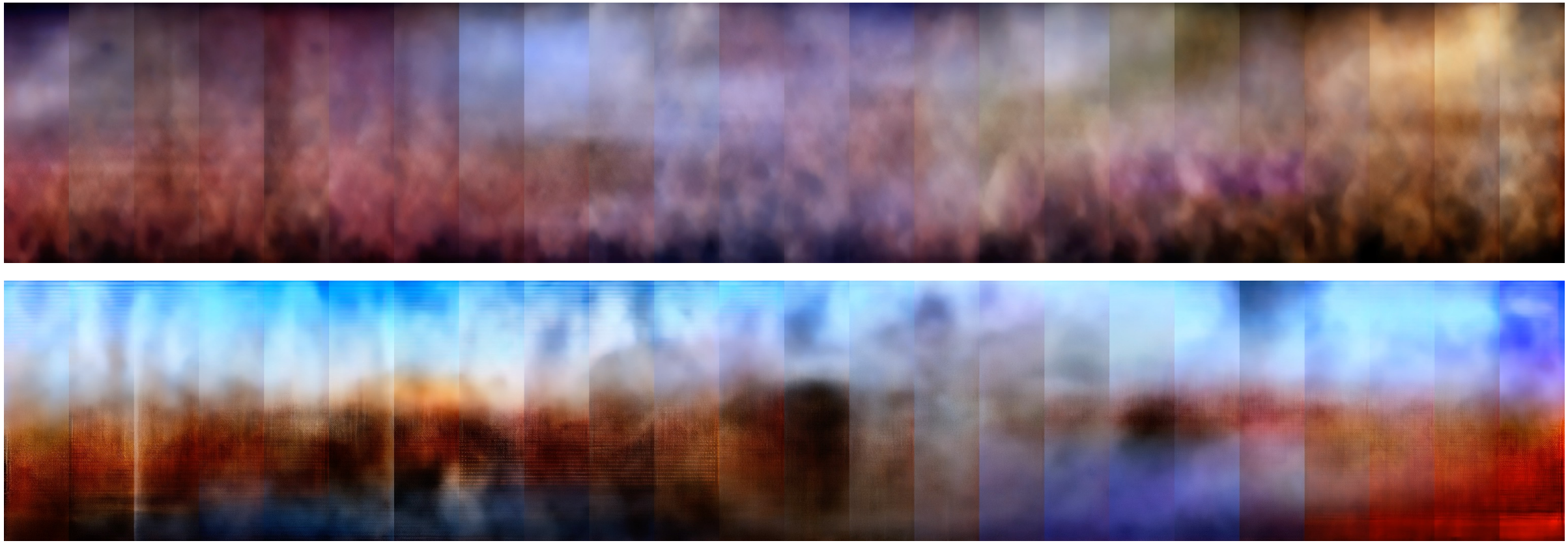
Posterior Distillation Sampling

Joint Denoising



Lee et al., SyncDiffusion: Coherent Montage via Synchronized Joint Diffusions, NeurIPS 2023.

Joint Denoising



SyncDiffusion

Kim et al., SyncTweedies: A General Generative Framework Based on Synchronized Diffusions, 2024.

Joint Denoising

"A hand carved wood turtle"

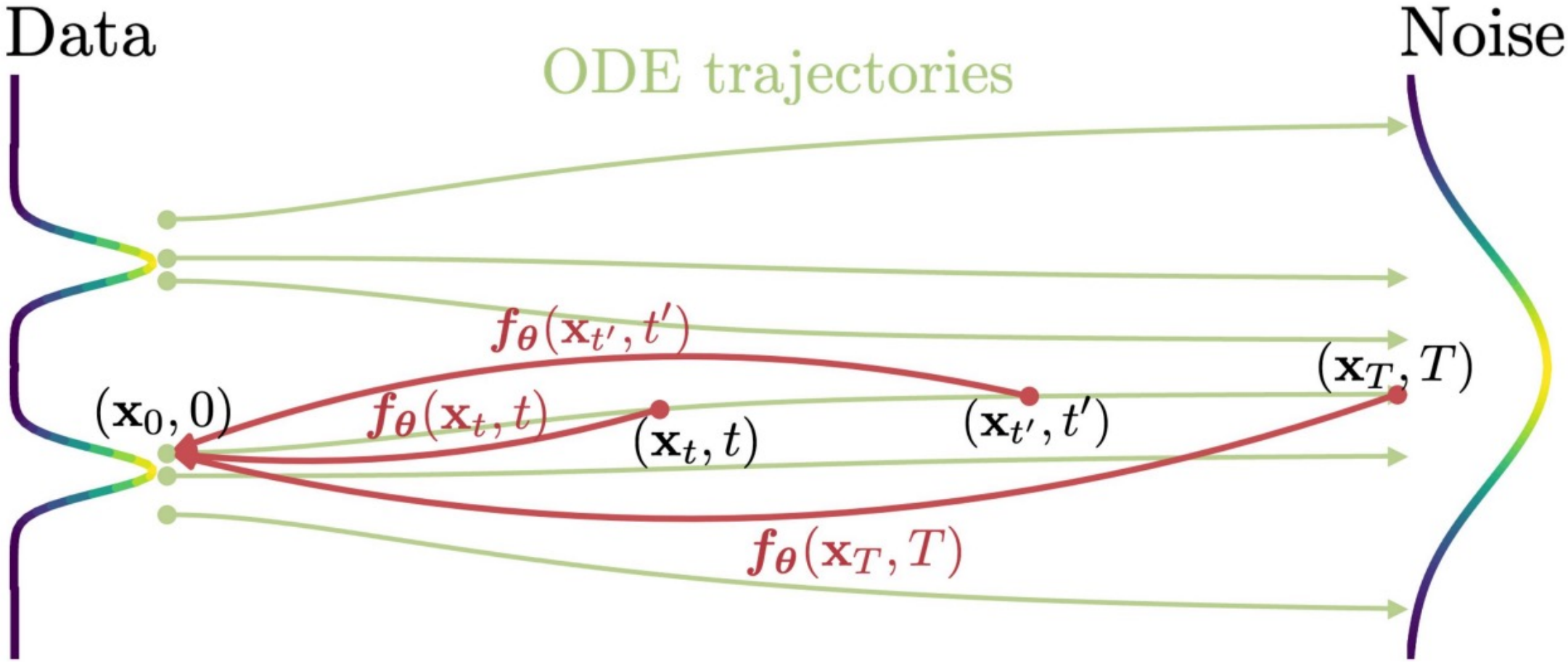
"A dumpster"

"A Chinese style lantern"

"A car with graffiti"

SyncTweedies

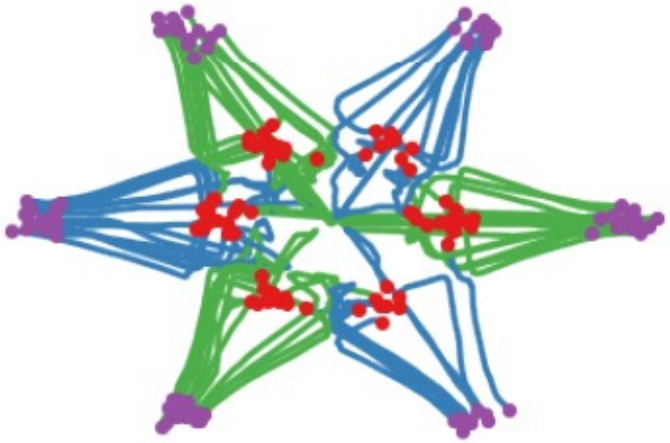
Few-Step Generation



Consistency Models

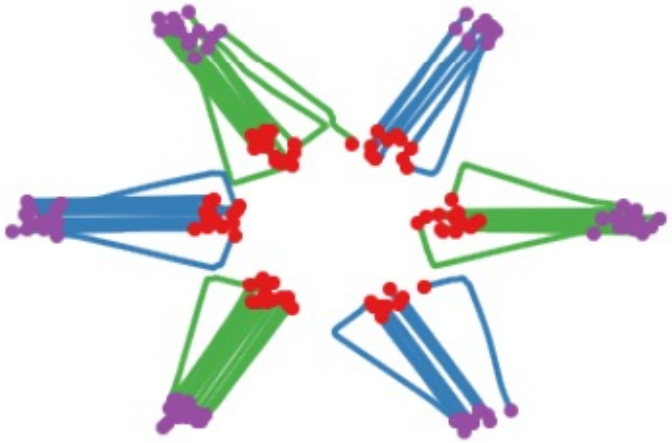
Liu et al., Flow straight and fast: Learning to generate and transfer data with rectified flow, ICLR 2023.

Rectified Flow



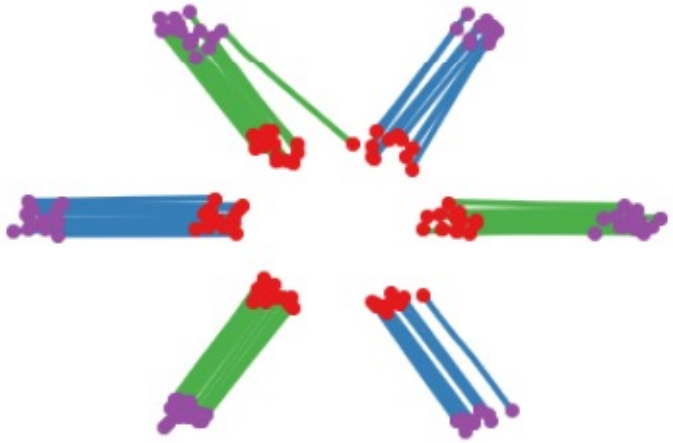
(a)The 1st rectified flow $\boldsymbol{Z}^1$

$\boldsymbol{Z}^1 = \texttt{RectFlow}((X_0, X_1))$



(b) Reflow $\boldsymbol{Z}^2$

$\boldsymbol{Z}^2 = \texttt{RectFlow}((Z_0^1, Z_1^1))$



(c) Reflow $\boldsymbol{Z}^3$

$\boldsymbol{Z}^3 = \texttt{RectFlow}((Z_0^2, Z_1^2))$

Reflow

Topics

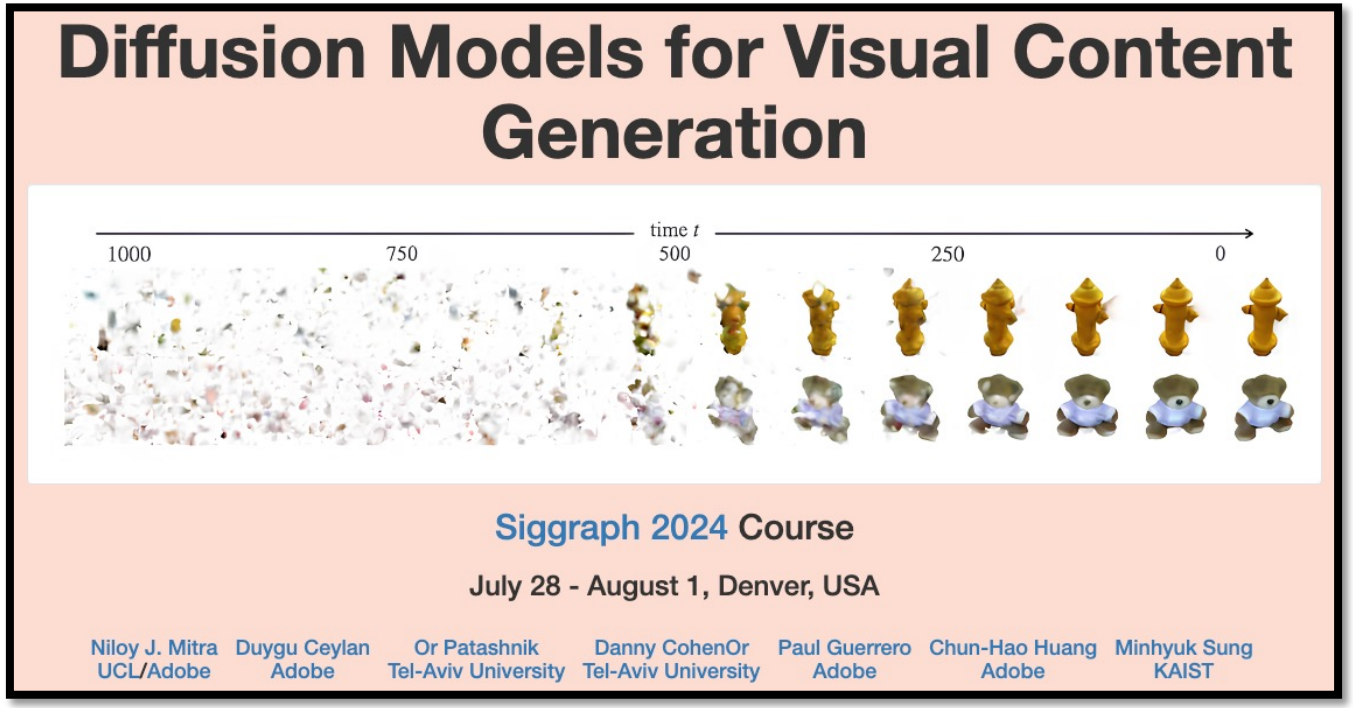
- Background of Generative Models
- DDPM / DDIM / Score-Based Models
- CFG / Latent Diffusion
- Conditional Generation
- Stylization / Personalization
- Inverse Problem
- Knowledge Distillation
- Diffusion Synchronization
- SDE/ODE Solvers
- Consistency Models / Flow-Based Models
- DiT / Applications / Future of Generative Models

Resources

SIGGRAPH 2024 Course:

Diffusion Models for Visual Content Generation

https://geometry.cs.ucl.ac.uk/courses/diffusion4ContentCreation_sigg24/



Resources

- CVPR 2023 Tutorial: Denoising Diffusion Models: A Generative Learning Big Bang.
- "Generative Modeling by Estimating Gradients of the Data Distribution", Yang Song.
- "What are Diffusion Models?", Lilian Weng.
- "Understanding Diffusion Models: A Unified Perspective". Calvin Luo.
- "Tutorial on Diffusion Models for Imaging and Vision". Stanley H. Chan.
- "Step-by-Step Diffusion: An Elementary Tutorial". Preetum Nakkiran, Arwen Bradley, Hattie Zhou, and Madhu Advani.
- Check out the course webpage for more information.

In this course…

We will discuss diffusion models, covering both their theoretical foundations and practical applications.

Prerequisite

- Background in machine learning/deep learning. We'll specifically focus on diffusion models (while briefly discussing the background of generative models).

- Experience with neural network implementation. There will be programming assignments and a term project.

Tentative Schedule (1/2)

Week	Mon	Topic	Wed	Topic
01	Sep 02	Course Introduction	Sep 04	Introduction to Generative Models
02	Sep 09	DDPM / Score-Based Models	Sep 11	Assignment 1 Session
03	Sep 16	No Class (Chuseok Holiday)	Sep 18	No Class (Chuseok Holiday)
04	Sep 23	DDIM / CFG / Latent Diffusion	Sep 25	Assignment 2 Session
05	Sep 30	ControlNet & Conditional Generation / LoRA & Personalization	Oct 02	No Class (Substitution of Hangul Day)
06	Oct 07	Assignment 3 Session	~~Oct 09~~ Oct 10	Guest Lecture 1 Or Patashnik
07	Oct 14	Inverse Problem / Knowledge Distillation	Oct 16	Assignment 4 Session
08	Oct 21	No Class (Midterm Week)	Oct 23	No Class (Midterm Week)

Tentative Schedule (2/2)

Week	Mon	Topic	Wed	Topic
09	Oct 28	Project Introduction	Oct 30	Inverse Problem / Diffusion Synchronization
10	Nov 04	Assignment 5 Session	Nov 06	SDE/ODE Solvers
11	Nov 11	Assignment 6 Session	Nov 13	No Class (Break)
12	Nov 18	Consistency Model / Flow-Based Models	Nov 20	Assignment 7 Session
13	Nov 25	DiT / More Applications / Future of Generative Models	Nov 27	Guest Lecture 2 Jiaming Song
14	Dec 02	Final Project Presentations 1	Dec 04	Final Project Presentations 2
15	Dec 09	No Class (Conference Trip)	Dec 11	No Class (Conference Trip)
16	Dec 16	Final Week (No Class)	Dec 18	Final Week (No Class)

Lectures

- All the lectures will be ONLINE.
  Recordings will be available on the course webpage.

- There will be hybrid (offline/online) sessions for assignments.

Holidays / Breaks / Special Session

- Holidays and breaks
  - Sep 16-18: Chuseok
  - Oct 2: Substitution of hangul day
  - Nov 13, Dec 9-11: Break
- Special sessions
  - Oct 10 (Thu), Nov 27: Guest lectures
  - Oct 28: Project introduction
  - Dec 2-4: Project presentations

Guest Lectures

Or Patashnik

Ph.D. Student at Tel-Aviv University

Oct 10 (Thu) 4:00 am KST

Topics: Editing using pretrained diffusion models

Jiaming Song

Chief Scientist at Luma AI

Nov 27 (Wed) 1:00 pm KST

Topics: Diffusion models from Luma / Future of generative models

Guest Lectures

- The guest lectures will be open to the public.

  (Feel free to invite the others.)

- The attendance will be checked.

More Information

Webpage: https://mhsung.github.io/kaist-cs492d-fall-2024

Course Logistics

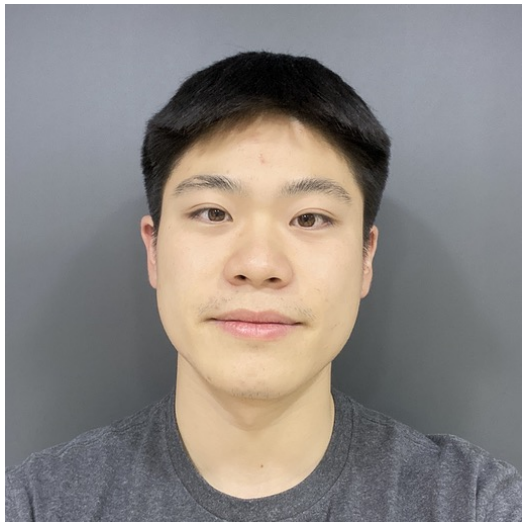
Minhyuk Sung (Instructor)

- Assistant Professor in School of Computing.
- Homepage: https://mhsung.github.io
- Office: N1, Rm 607.

Course Assistants

Seungwoo Yoo
dreamy1534@kaist.ac.kr

Yuseung Lee
phillip0701@kaist.ac.kr



Jaihoon Kim
jh27kim@kaist.ac.kr

Juil Koo
63days@kaist.ac.kr

Zoom Policy

- Turn on your camera, turn off your mic.
  You'll get a penalty in the participation score if don't turn on the camera.

- Please set your name on Zoom like this:
  (Student ID)_(Full name)
  Set your name in English and match the full name in the KLMS.

We'll use multiple tools.

- **Zoom**: Lectures

- **Slack**: Quizzes, Q&A, communication.

  Check #announcements channel in slack regularly.

- **GradeScope**: Programming assignment / project submission.

Quizzes

There will be quizzes during the lectures.

Please answer the questions in the #lecture-01 channel on **Slack**.

Let's try now!

Q. What motivated you to enroll in this course?

Evaluation

- In-Class Participation: 10%
- Programming Assignment: 45%
- Project: 45%

In-Class Participation (10%)

Education 4.0

https://academy.kaist.ac.kr/pages/sub/sub02_02

In-Class Participation – Quizzes

- You'll take short quizzes during the lectures.

- Each lecture may start with quizzes about the previous lecture – **Be prepared!**

In-Class Participation – Quizzes

- Sometimes quizzes will be about solving (easy) math problems.

- **Prepare a pencil and paper in lectures.**
  You may need to take a photo of your answer and post it on Slack.

In-Class Participation – Quizzes

- TA will randomly choose one quiz each lecture for the evaluation.
- TA may or may not check the correctness of your answer. But, the answer still needs to be reasonable.
- The goal of quizzes is to engage your learning and participation in classes, not to evaluate your performance.

In-Class Participation – Quizzes

A good question or answer posted on Slack will make up for one missing quiz! Good questions or answers in Slack will receive ❤️ from the instructor or TA.

In-Class Participation – Evaluation

- For the randomly sampled quizzes,
  you'll receive a 100% score if you answered 80% of them or more. Otherwise, your score will be prorated.

- Example:
  - 16 / 20 answered → Your participation score: 100%
  - 15 / 20 answered → Your participation score: 15 / 20 = 75%!
- **The participation score can affect your grade!**

Programming Assignment (45%)

Programming Assignments

- The programming assignment has already been released.

- Please visit the ‘Programming Assignment’ tab on the course web page and check out the GitHub repository of the assignment.

Programming Assignments

- Assignment1 (DDPM)
- Assignment2 (DDIM / CFG)
- Assignment3 (ControlNet / LoRA)
- Assignment4 (Distillation)
- Assignment5 (Synchronization)
- Assignment6 (DPMSolver)
- Assignment7 (Flow)

Prerequisite

- You'll need basic programming skills in Python and PyTorch to complete the programming assignments.

- **Start the programming assignments as early as possible!**

Programming Assignments

- We'll have a hybrid (online/offline) TA session for each assignment. Please check the schedule on the course webpage.

- **Offline classroom: N1, Room 201.**

Programming Assignments

- You are strongly encouraged to attend the assignment session in person so you can ask questions directly to the instructor and TAs. You can still join the session online if needed.

- We will also have quizzes during the assignment sessions, so your participation will count towards your participation score.

Programming Assignments

- Each assignment will have two tasks, with each task worth 10 points (a total of 20 points per assignment).

- Some assignments will have bonus credits that can compensate for missing points in other assignments.

Submission

- Each programming assignment is due **two weeks after** the assignment session.
- Note that the assignment periods may overlap. You will need to attend an assignment session before the due date of the previous assignment. **Start the assignment as early as possible.**
- Submit your solutions on Gradescope.

Submission

- There will be a 20% penalty for each day late.

- Exemptions can be made only when special arrangements have been made with the instructor BEFORE the due date.

Computing Resource

- We will provide a KCLOUD account with a single GPU to each undergraduate student who has requested (we won't have enough GPUs to allocate to the graduate students).

- Check out the #stack-overflow channel in Slack.

AI Coding Assistant Tool Policy

- You are allowed to use AI coding assistant tools, such as ChatGPT, Copilot, Codex, and Code Intelligence, for your programming assignments and projects.

- However, it is still strictly prohibited to directly copy code from the Internet or from someone else. Such actions will result in a score of zero and a report to the university.

Plagiarism

- Any cases identified by Moss as copies of
  - the original code or
  - as being taken from someone else's code found on the Internet, from classmates, or from any other sources,

  will be classified as plagiarism.

- Moss: https://theory.stanford.edu/~aiken/moss/

Project (45%)

Project

- **Create your own diffusion model!**

- Choose one of the candidate benchmarks (dataset and evaluation metrics) and create your own diffusion model based on it.

- You'll need to form a team with classmates and work on the project throughout the semester.

Project Proposal

Due: October 19 (Saturday)

- Submit a write-up (up to three pages) that includes the following:
  - Names and student IDs of all team members
  - Chosen benchmark
  - Basic ideas for the diffusion model implementation
  - Timeline for implementation
  - Plans for task allocation to team members

Project Interim Report

Due: November 9 (Saturday)

- Based on the timeline in the proposal, **color-code** each task in the timeline as Completed, In progress (as scheduled), Delayed, and Have issues.
- For the cases marked as Delayed or Have issues, provide a brief explanation of the problems and propose solutions.
- Submit a write-up (up to three pages).

Project Submission

Due: November 30 (Saturday)

- Submit the followings:
  - A poster
  - A report
  - Source code and data
- More details will be provided in the “Project” tab on the course webpage soon.

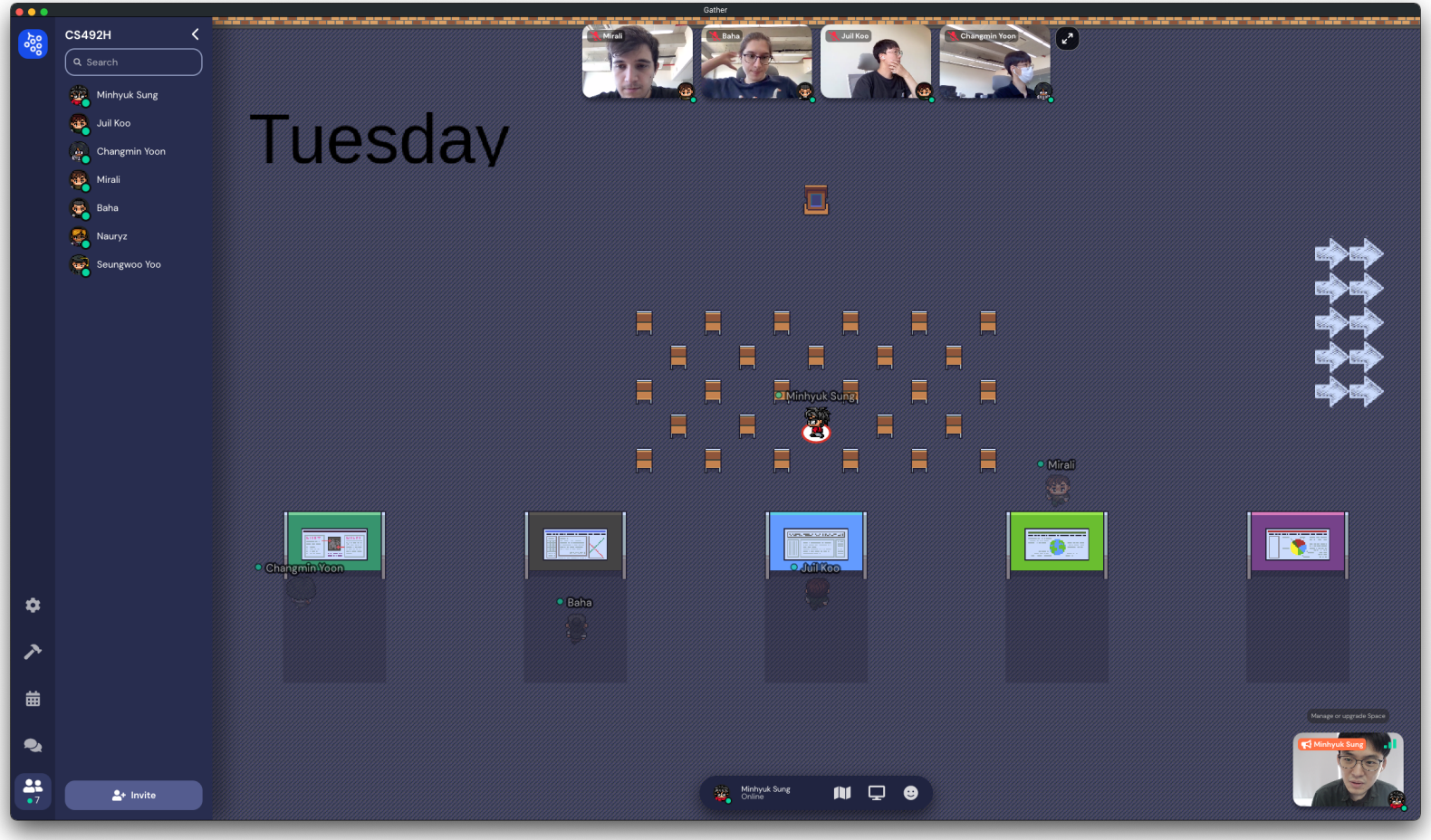
Policy

- You are allowed to use any existing code, pretrained models, and additional data, *provided* you properly cite these resources in your write-up and code.

- Missing references to any code, models, or data will be considered plagiarism.

Gather Town Poster Sessions

December 2 (Mon) and 4 (Wed)



Next Time

Introduction to Generative Models

GAN / VAE