Generative AI for 3D Data

December 2023

Seungwoo Yoo

KAIST Geometric AI Lab

The Era of Generative AIs

ChatGPT, OpenAI

The Era of Generative AIs

Gemini, Google DeepMind

The Era of Generative AIs

Midjourney

The Era of Generative AIs

Adobe Photoshop Generative Fill, Adobe

The Era of Generative AIs

Pika 1.0, Pika Labs

Speaker

Seungwoo Yoo

- M.S. Student, KAIST Geometric AI Lab (2023 -)
- B.S. in Computer Science, KAIST (2019 – 2023)
- Undergraduate Intern, KAIST Geometric AI Lab (2021 – 2023)
- Jeonbuk Science High School (2017 – 2019)

Research Interest: 3D Machine Learning, Computer Graphics

How do generative AIs work?

How can we build generative AIs for 3D data?

Generative AI Crash Course

Generative AI Crash Course

Q. How many RGB images of size 512 x 512 exist?

Assume that each pixel intensity is quantized to 0-255.

Generative AI Crash Course

Q. How many RGB images of size 512 x 512 exist?

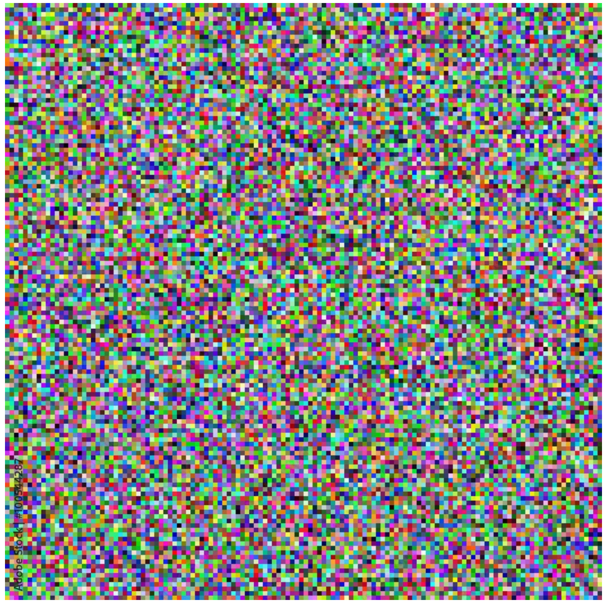
Assume that each pixel intensity is quantized to 0-255.

$$256^{512 \times 512 \times 3}$$

Generative AI Crash Course

Q. How many **interesting** RGB images of size 512 x 512 exist?

FFHQ Dataset, Karras et al., 2018



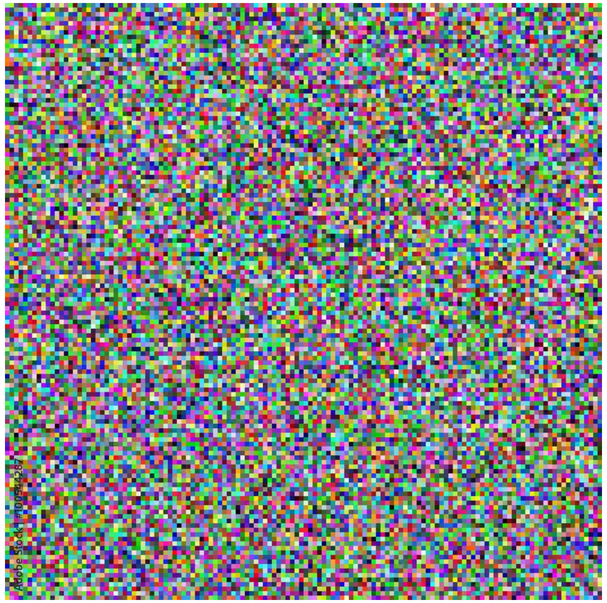
Gaussian Noise

Matterhorn, Wikipedia

Generative AI Crash Course

We assume that *real* images follow certain probability distributions defined over the image space.

FFHQ Dataset, Karras et al., 2018

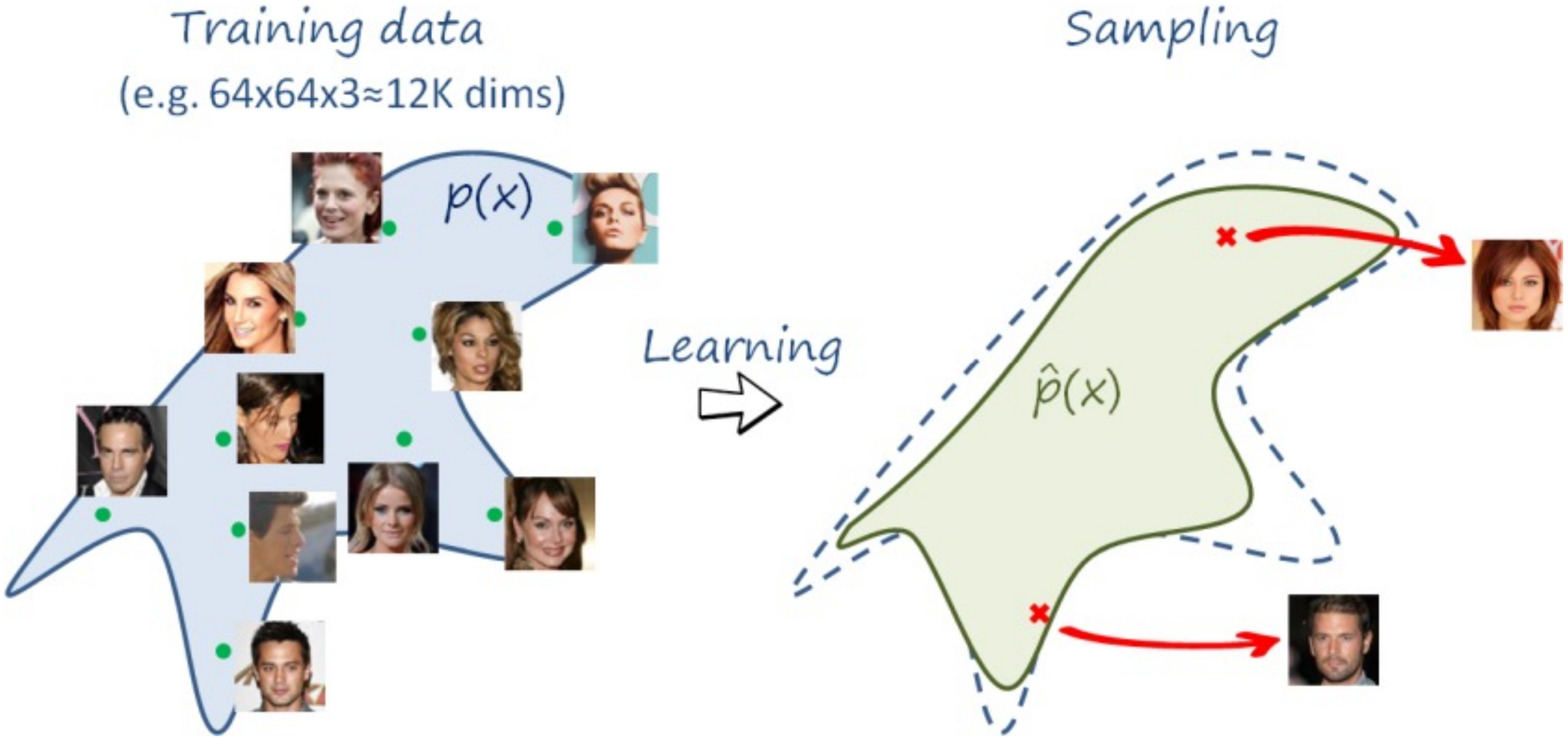


Gaussian Noise

Matterhorn, Wikipedia

Generative AI Crash Course

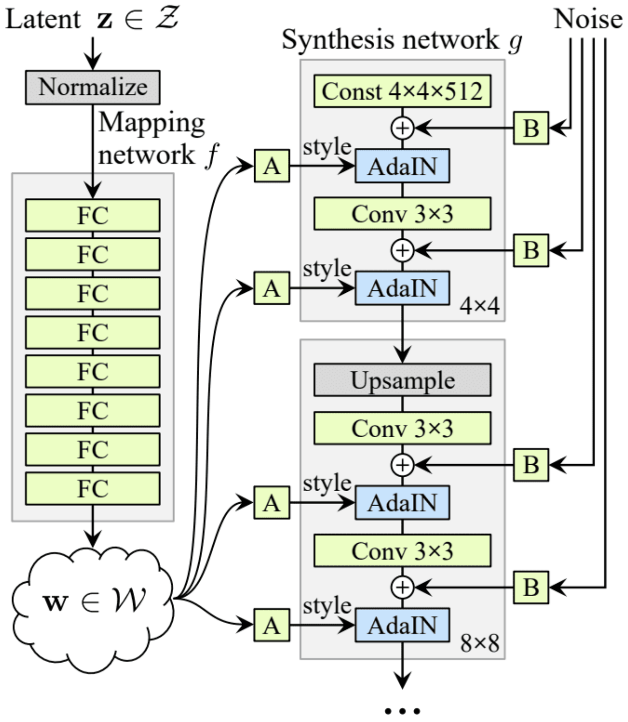
We *model* the underlying distributions of real images or data and expect to *draw* unseen samples from them.



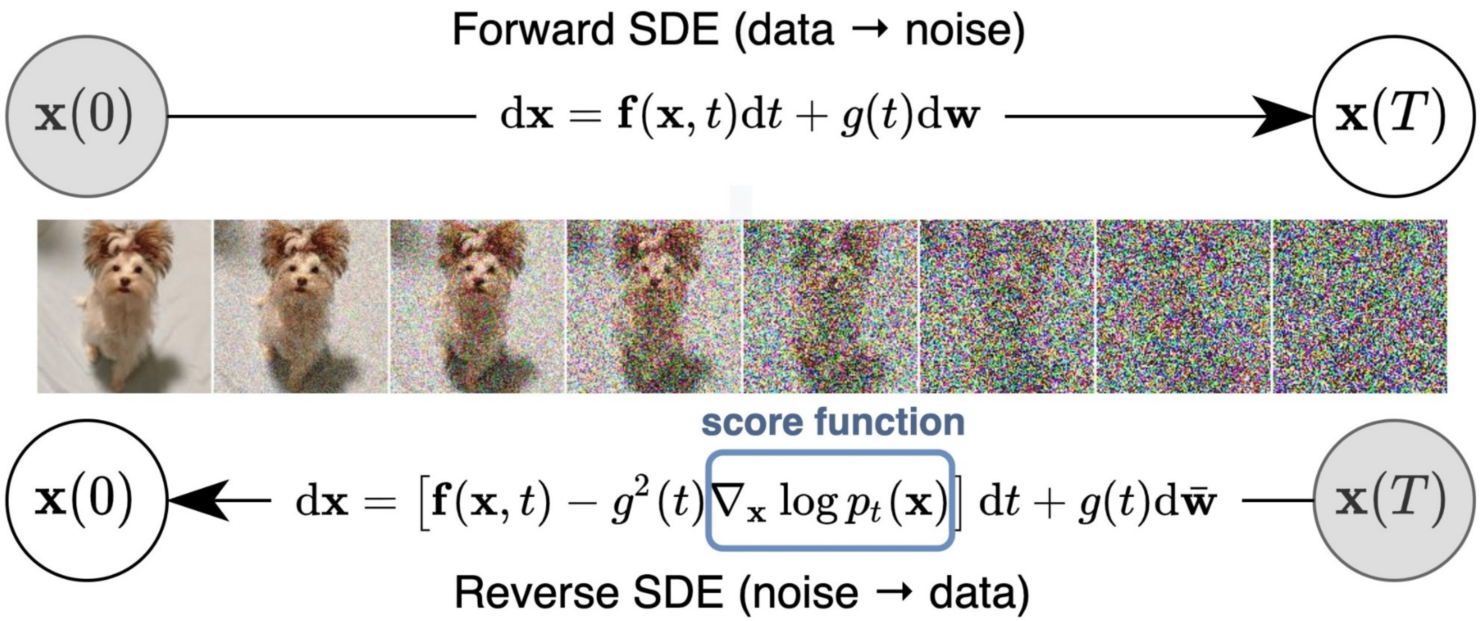
Generative Modeling, Luis Herranz

Generative AI Crash Course

Key Idea: Learn mappings from simple probabilistic distributions to complex real-world data distributions.



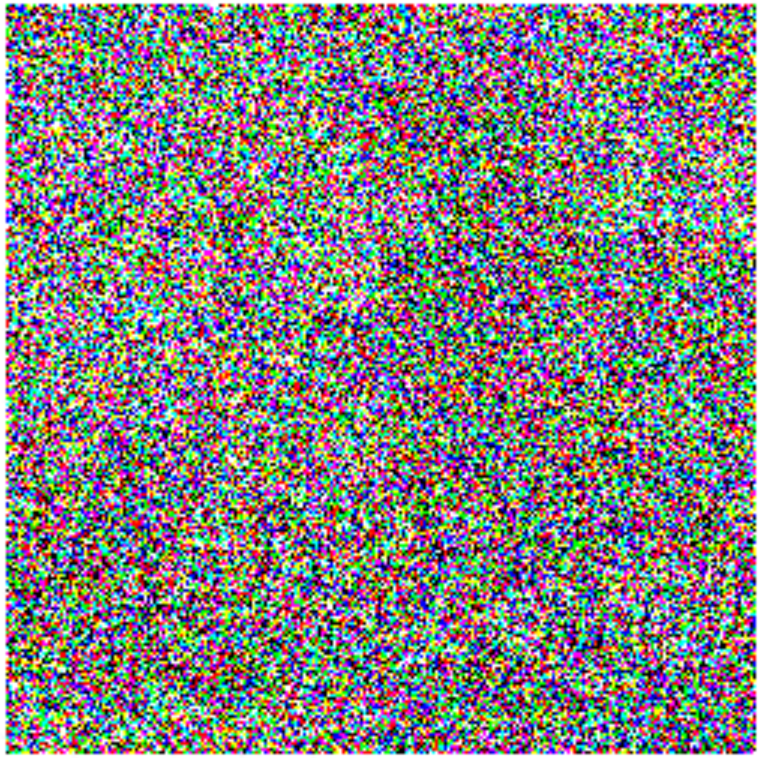
Karras et al., 2018

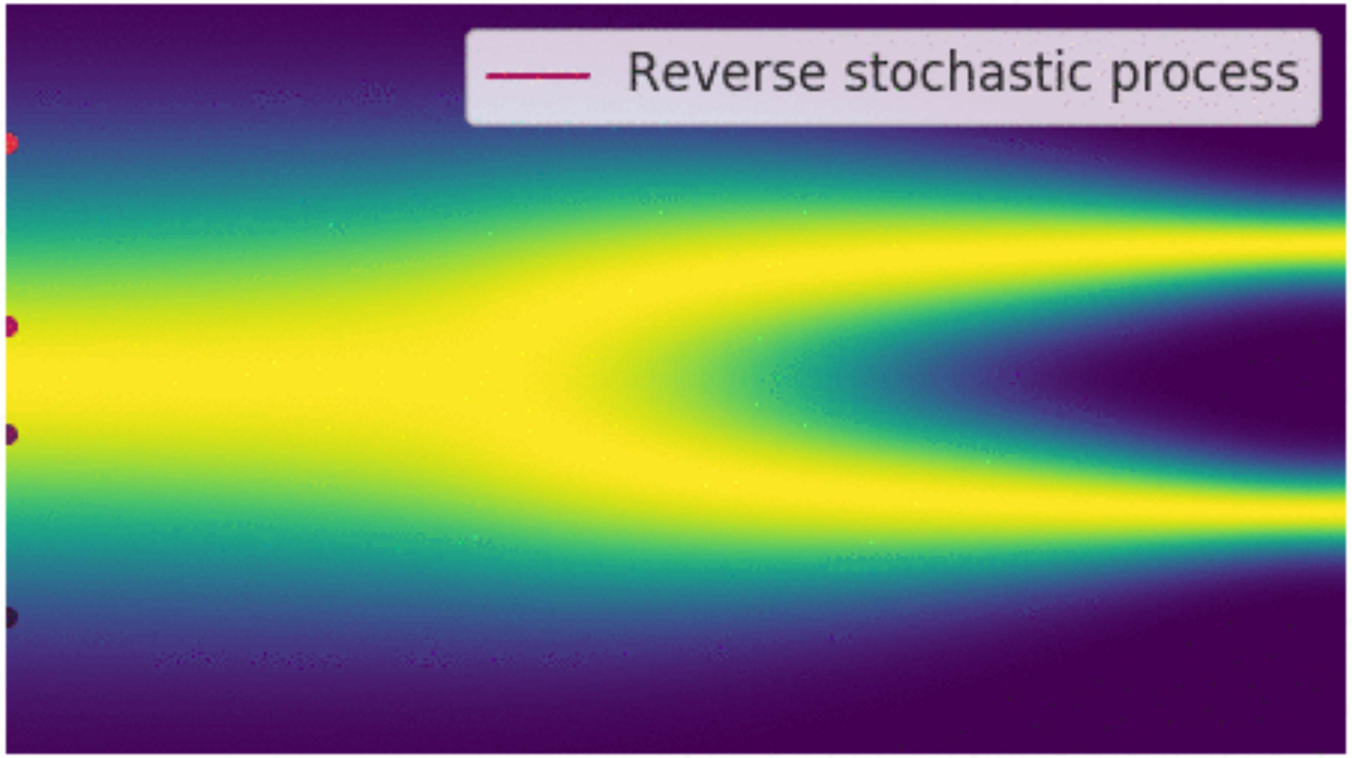


Song et al., 2021

Generative AI Crash Course

Key Idea: Learn mappings from simple probabilistic distributions to complex real-world data distributions.





Song et al., 2021

Generative AI Crash Course

Conditional generation (e.g., text, mask) is often achieved by training models using example pairs.

Zhang et al., 2023

Generative Modeling for 3D Data

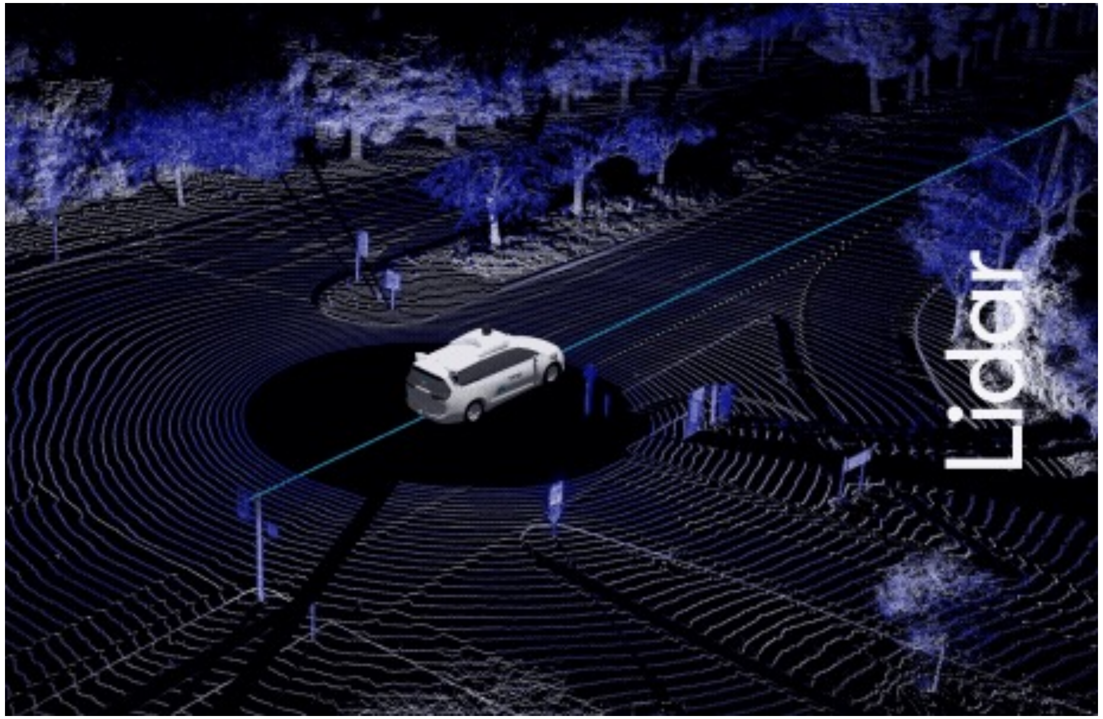
3D Data

3D shapes can be represented in various forms such as

- Point Clouds;
- Voxels;
- Polygon Meshes;
- Etc.

3D Data: Point Clouds

- Unordered sets of 3D coordinates: (X, Y, Z);
- Raw outputs of 3D scanning devices such as RADAR and LiDAR.

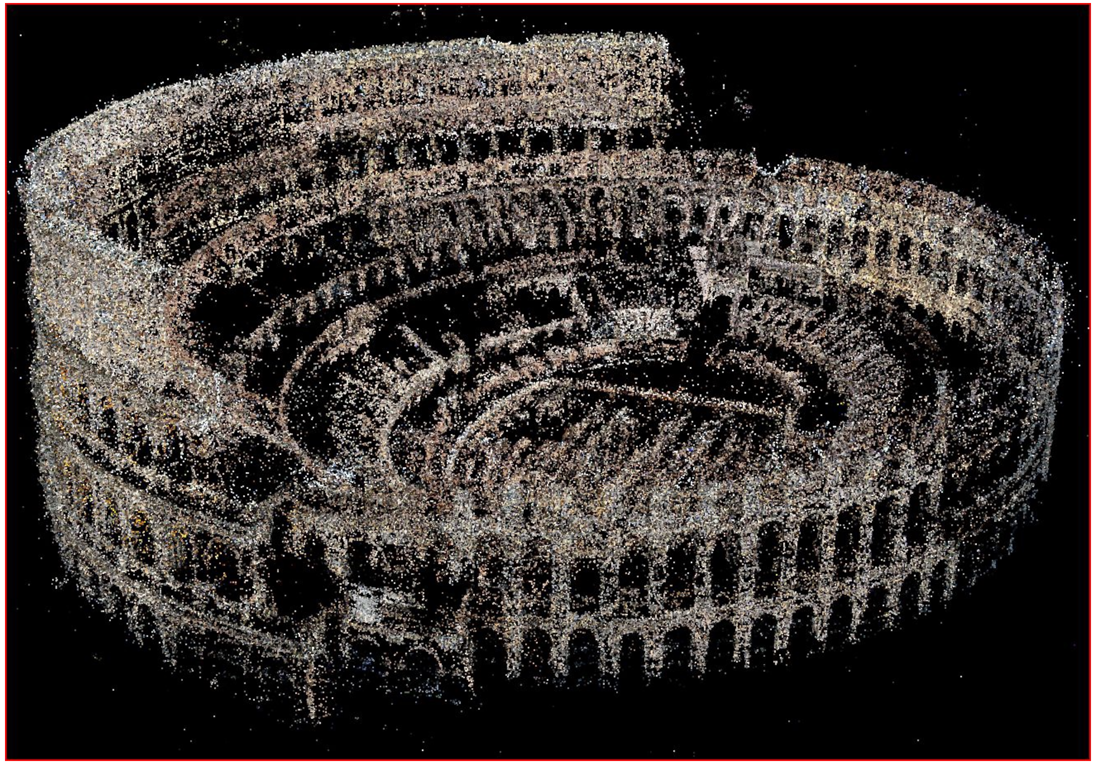


Waymo

3D Data: Point Clouds

(+) Simple, easy to process.

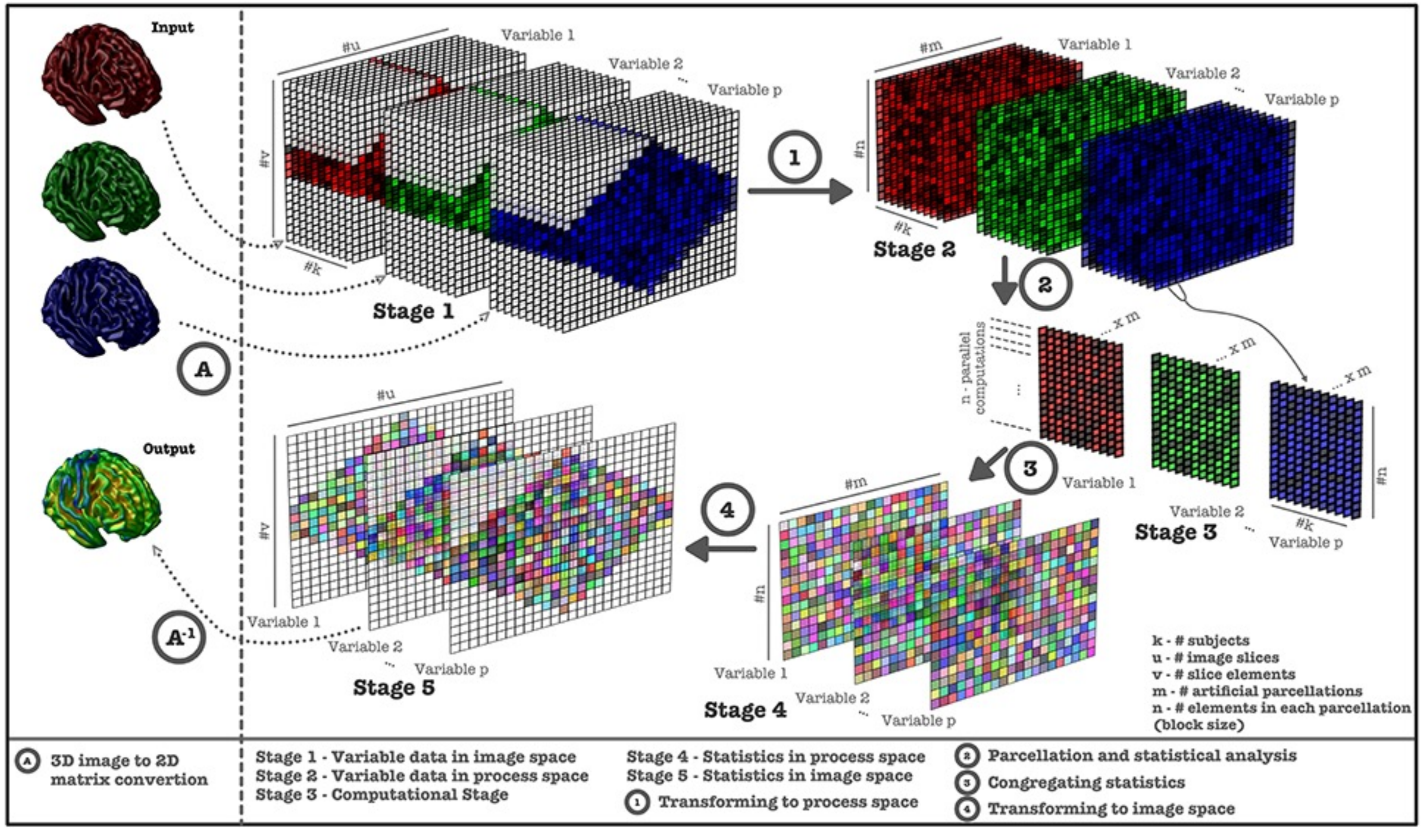
(-) No surface/topology information.



Agarwal et al., 2009

3D Data: Voxels

- Regular grids dividing 3D volumes into smaller cells;
- Widely adapted in medical imaging (e.g., MRI).

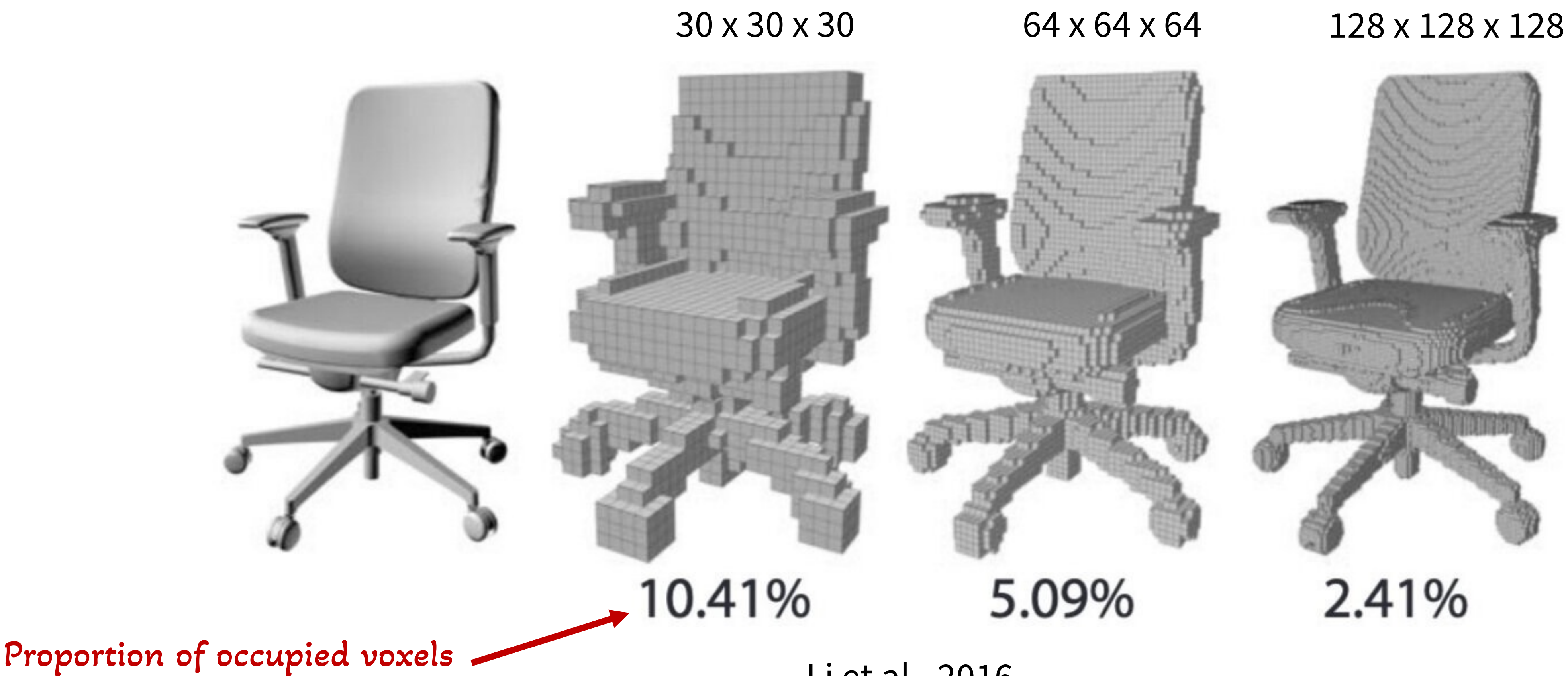


Mathotaarachchi et al., 2016

3D Data: Voxels

(+) Simple, 2D convolution naturally extends to 3D voxels.

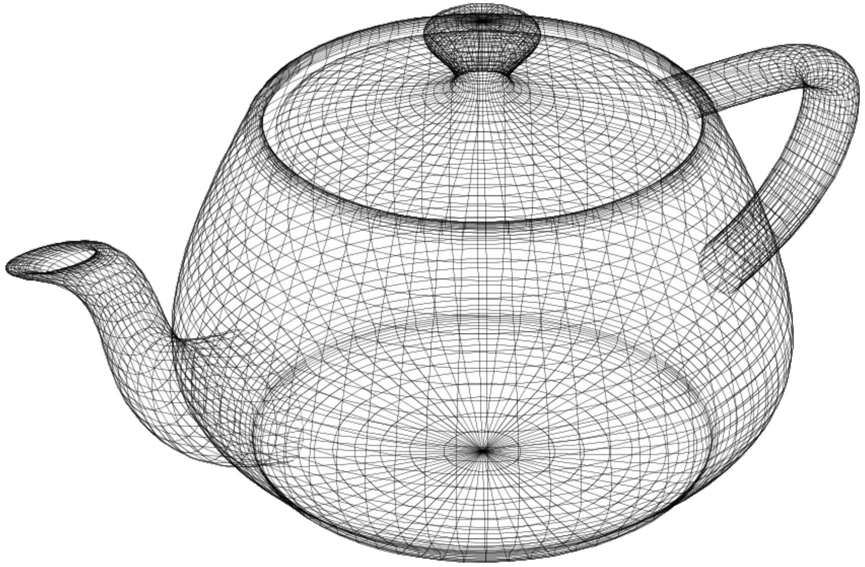
(-) Huge computation costs and waste of memory.



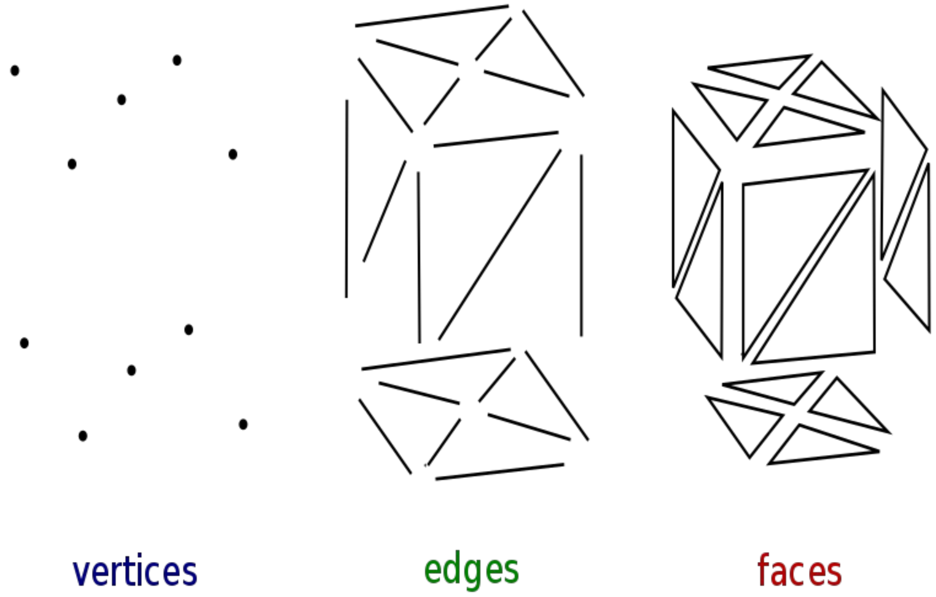
Li et al., 2016

3D Data: Polygon Meshes

- Collections of vertices, edges, and faces;
- Widely used in rendering, texturing, deformation, etc.



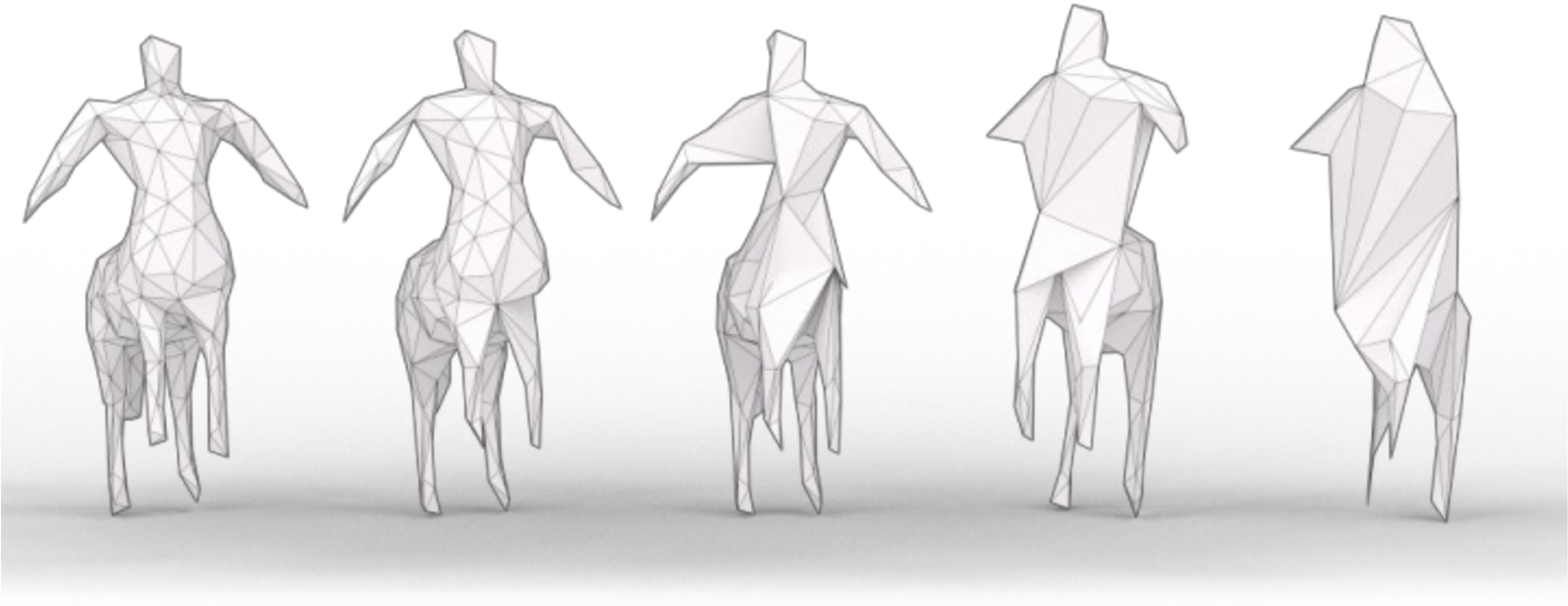
Utah Teapot, 1975



3D Data: Polygon Meshes

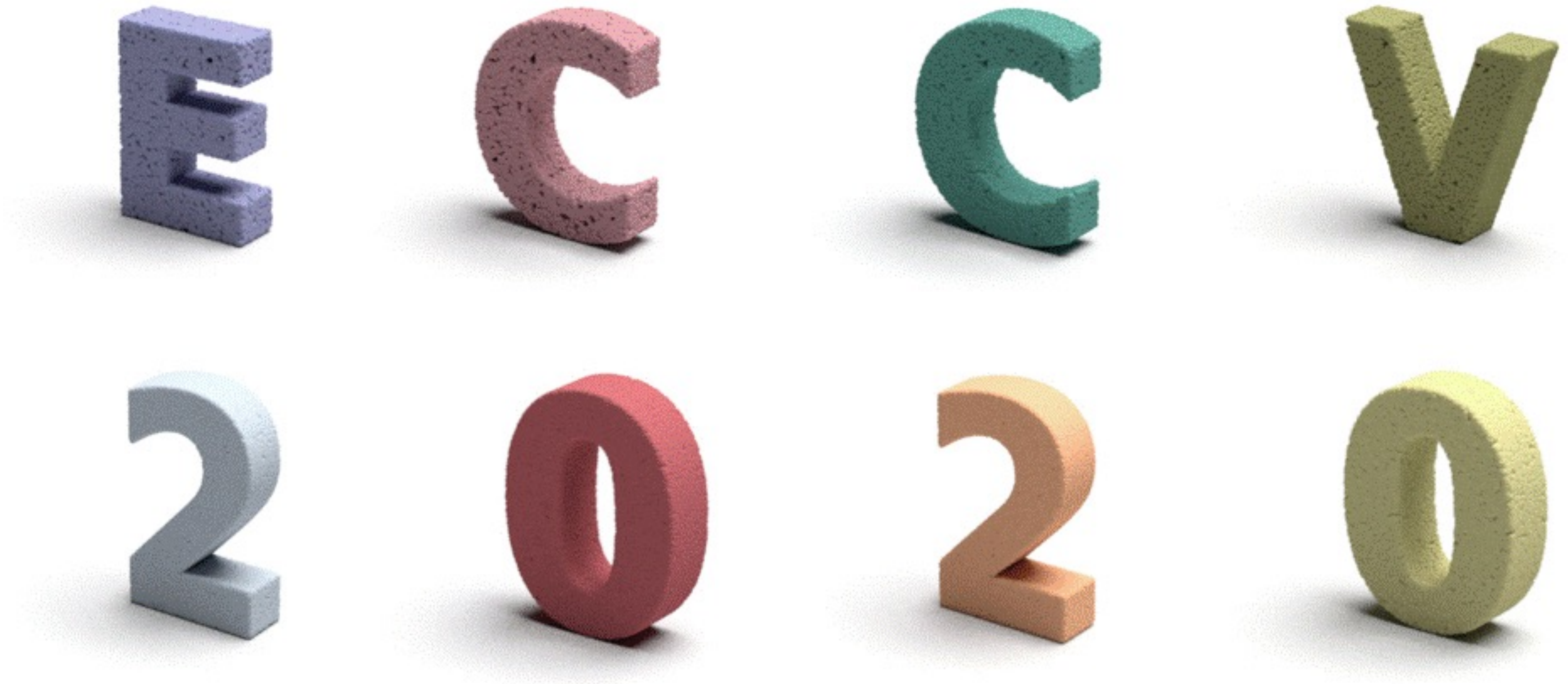
(+) Explicit surface/topology information, compact, easy-to-edit.

(-) Irregular, discrete connectivity is inappropriate for deep learning.



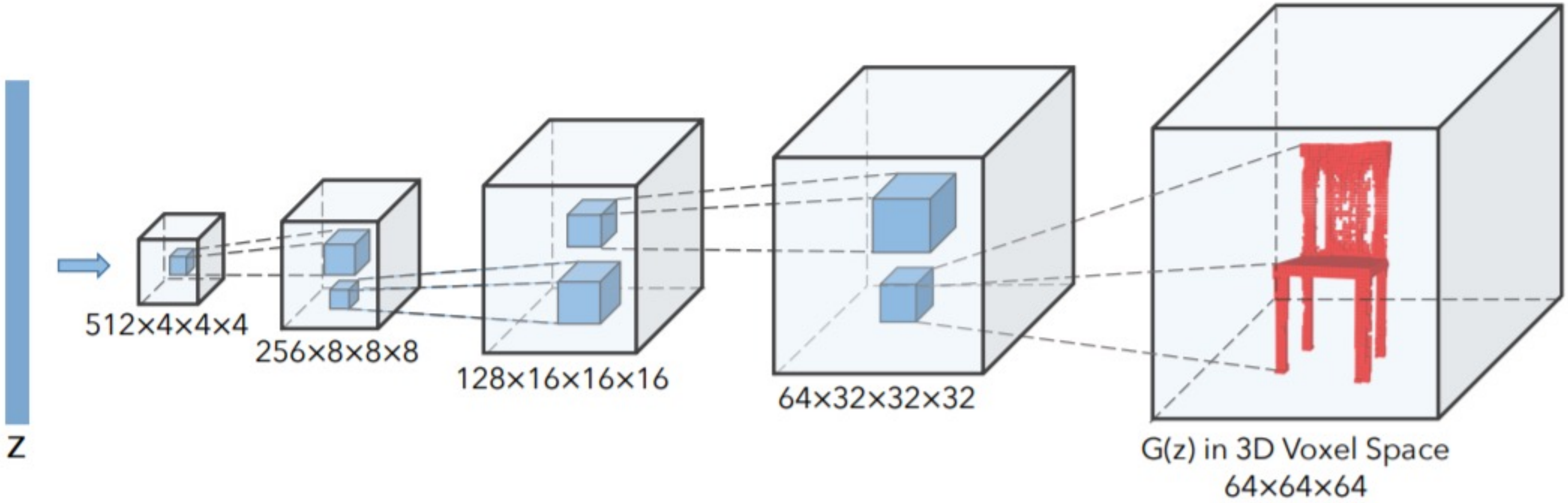
Hanocka et al., 2019

Generative Models for 3D Data



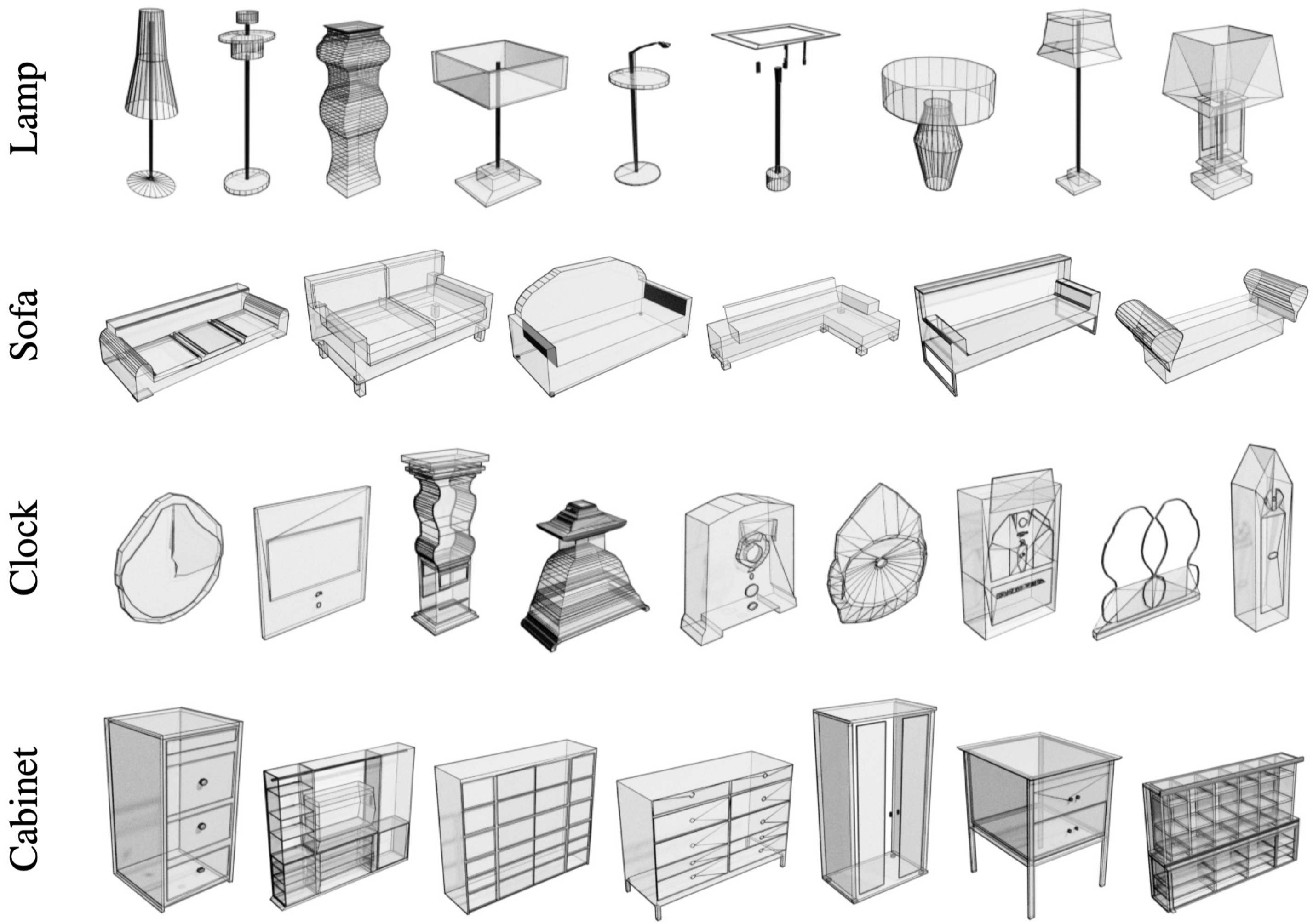
Cai et al., 2020

Generative Models for 3D Data



Wu et al., 2016

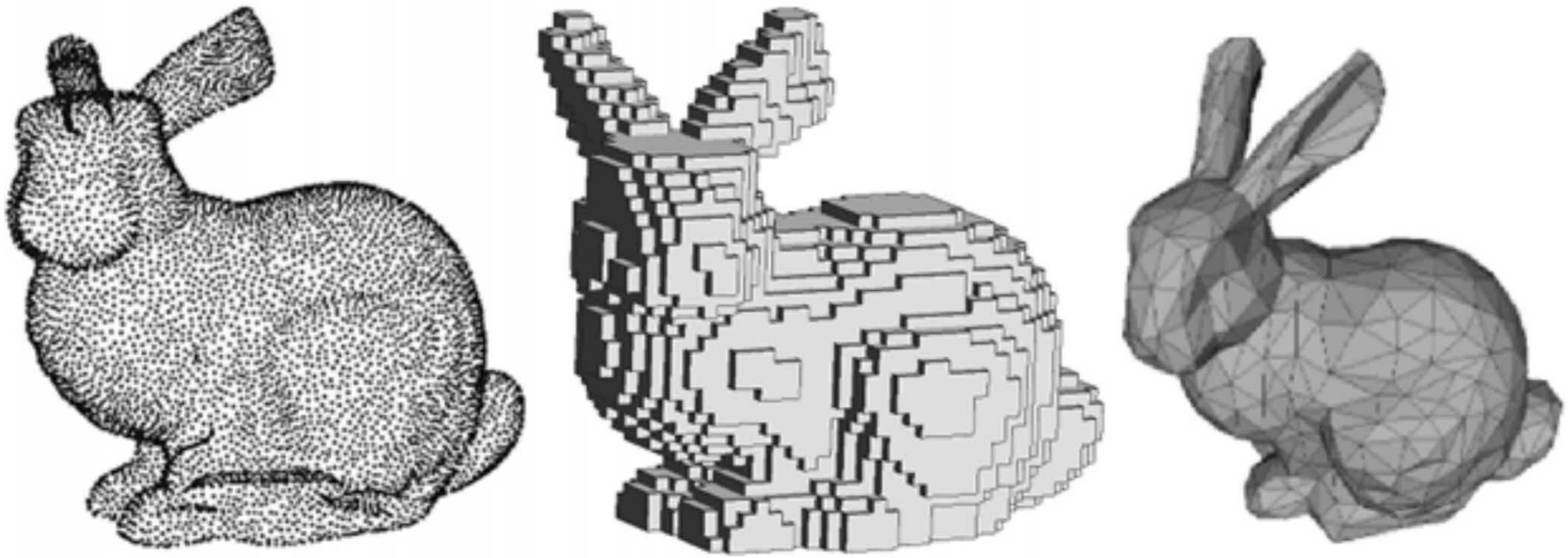
Generative Models for 3D Data



Nash et al., 2020

Challenges

The irregularity of 3D data poses challenges in designing neural network architectures for learning-based methods.



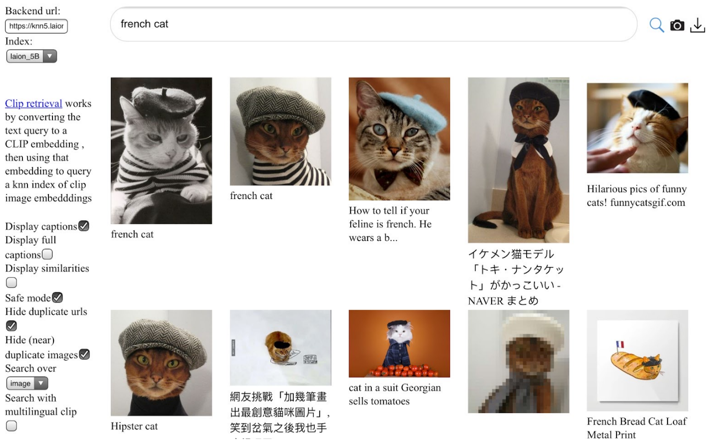
Hoang et al., 2019

Challenges

It is hard to populate high-quality, large-scale datasets of 3D shapes.

Dataset	Quantity (tokens)
Common Crawl (filtered)	410 billion
WebText2	19 billion
Books1	12 billion
Books2	55 billion
Wikipedia	3 billion

GPT-3 Training Data (400B)

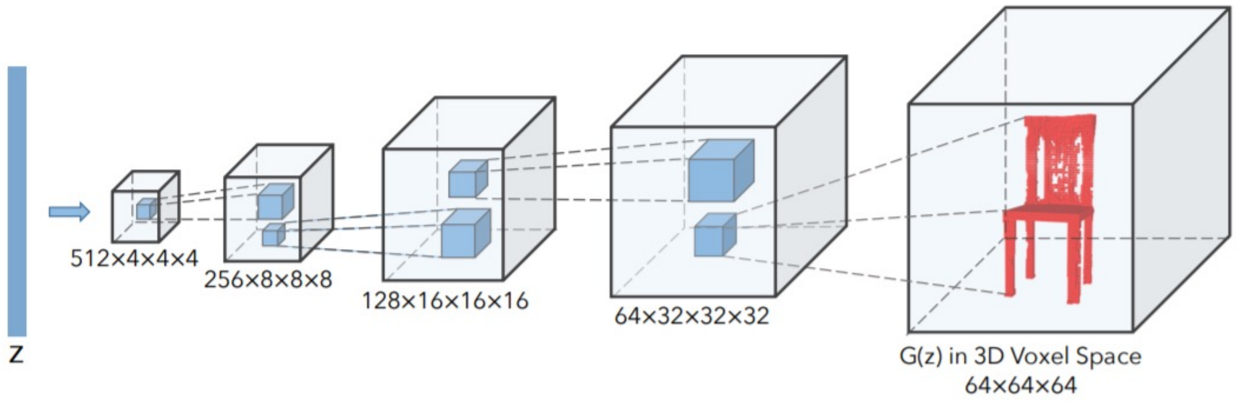


LAION-5B Dataset (5B)

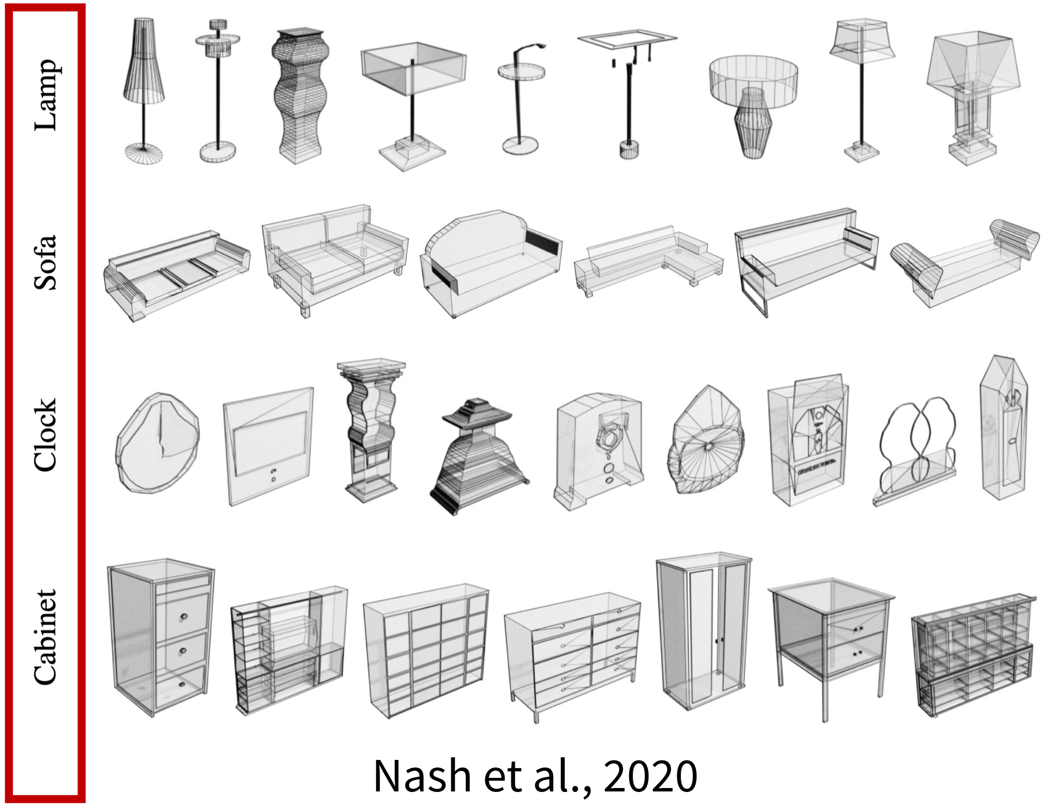
Objaverse-XL Dataset (10M)

Challenges

Limited generalization capability since models are trained on small set of shape categories at a time.



Wu et al., 2016



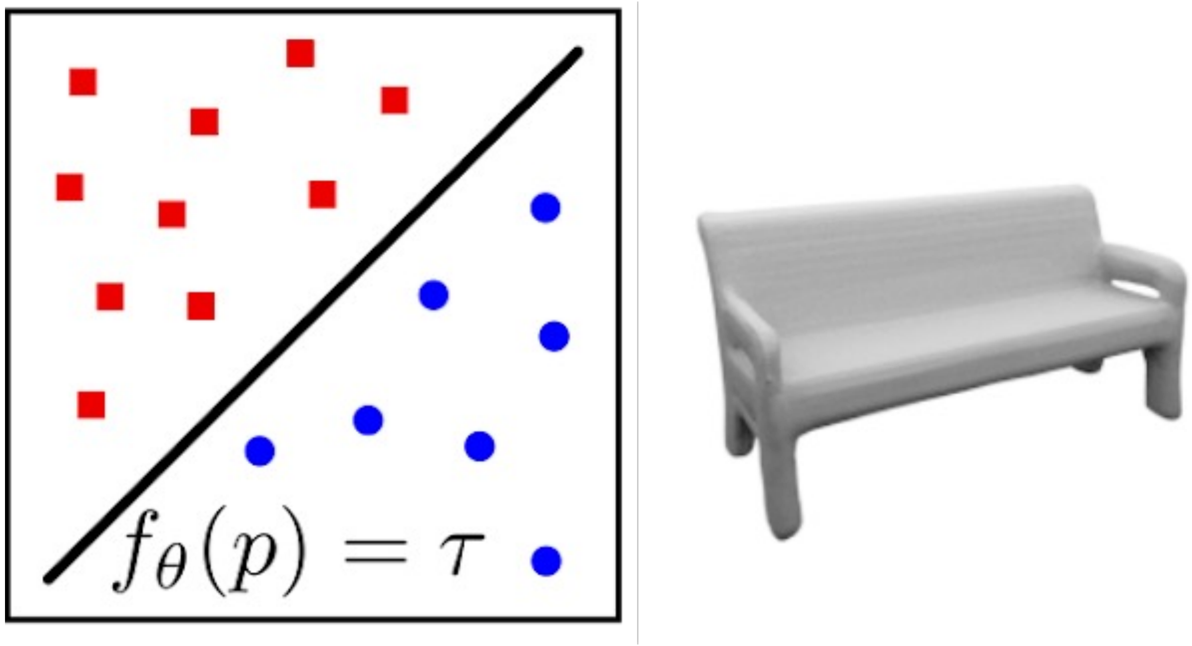
Nash et al., 2020

Requires lots of shapes from specific categories!

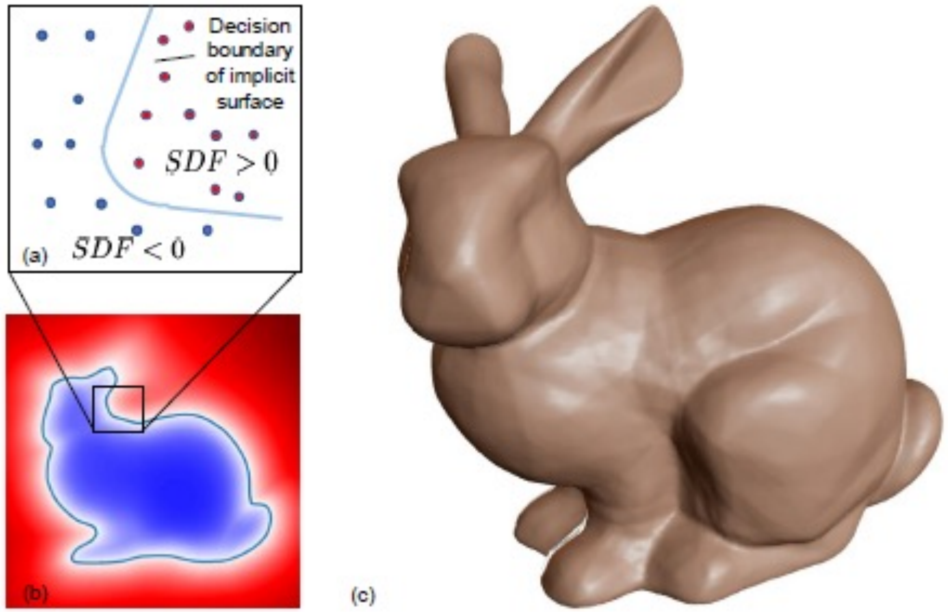
Challenges

1. Irregularity of 3D data → Use implicit functions!

2. Lack of high-quality, large-scale 3D datasets.



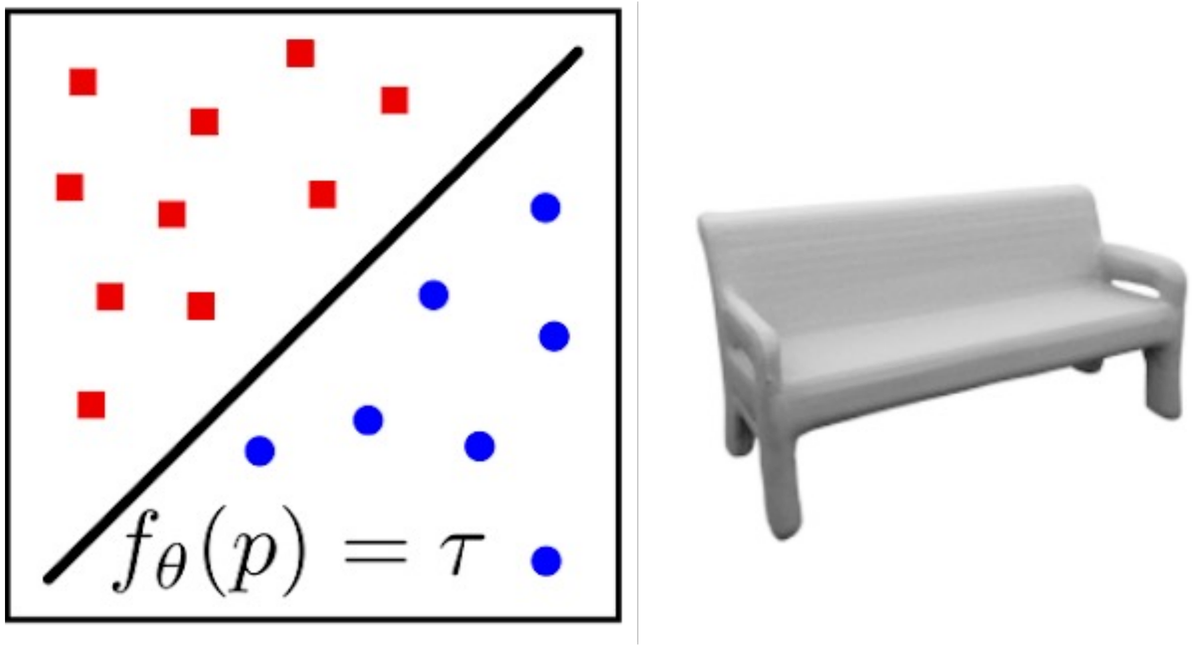
Mescheder et al., 2019



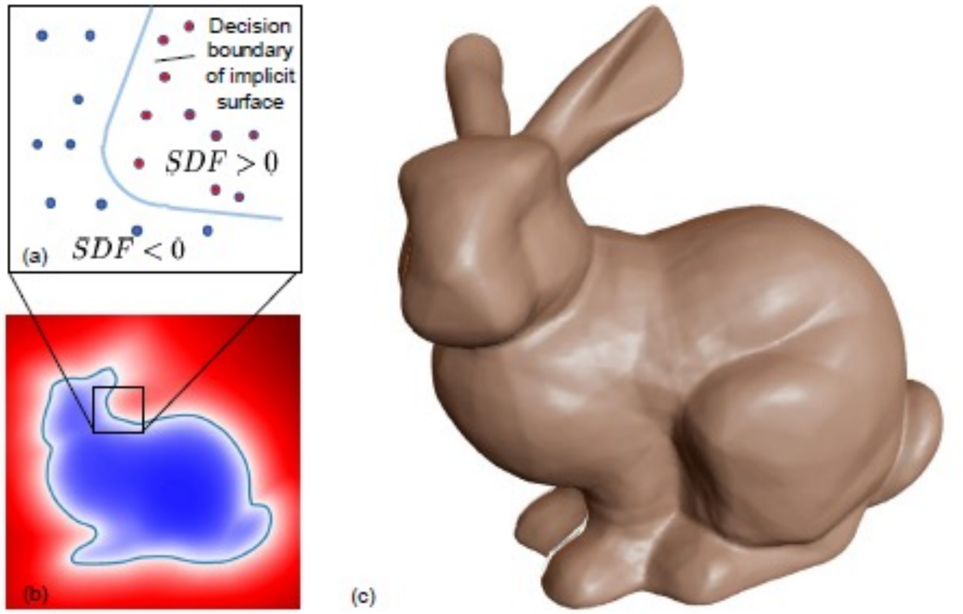
Park et al., 2019

Implicit Functions

- Functions that take (3D) coordinates as inputs can represent shapes;
- In/outside of shapes are indicated by occupancy or signed distance.



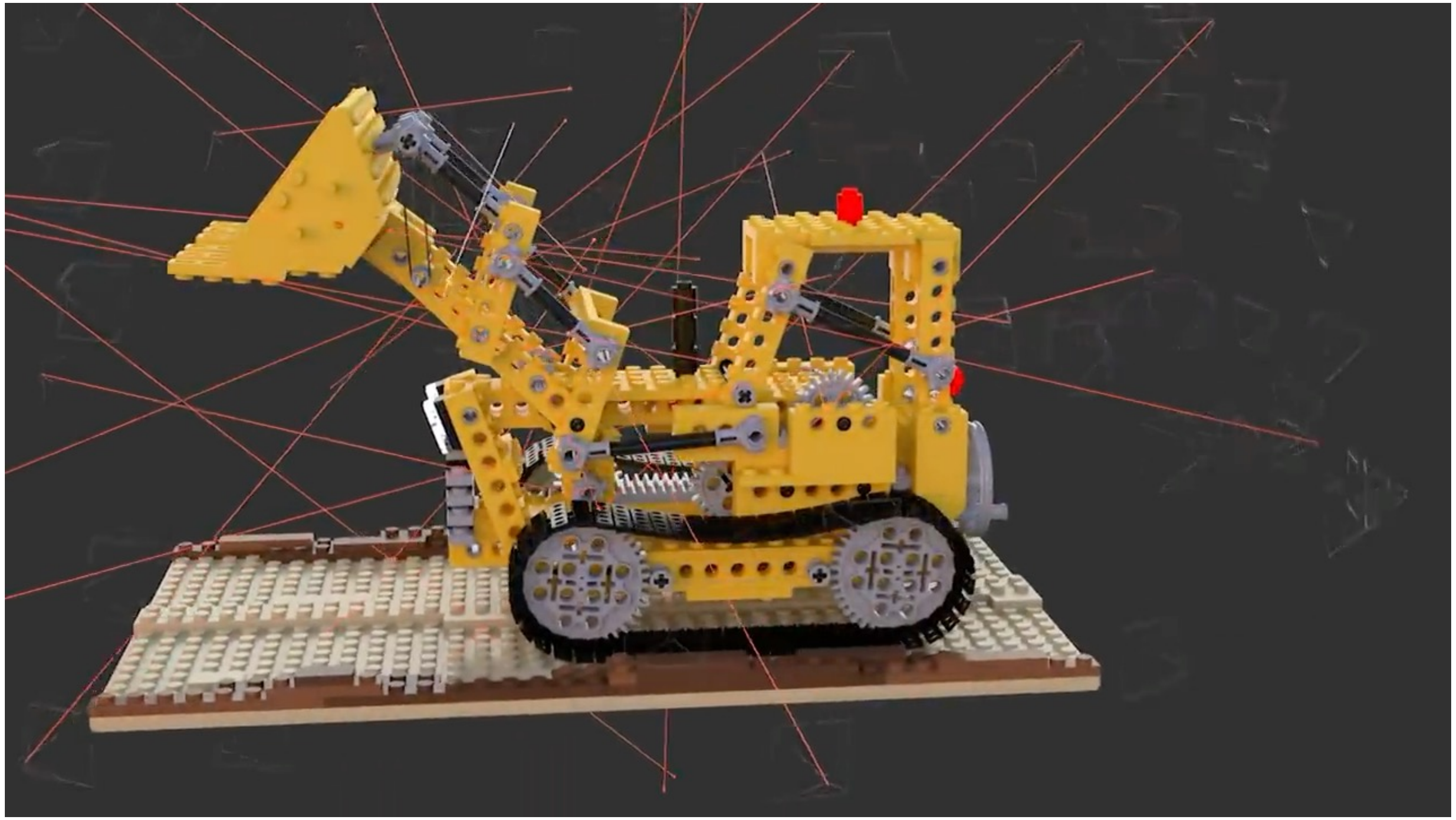
Mescheder et al., 2019



Park et al., 2019

Implicit Functions - NeRFs

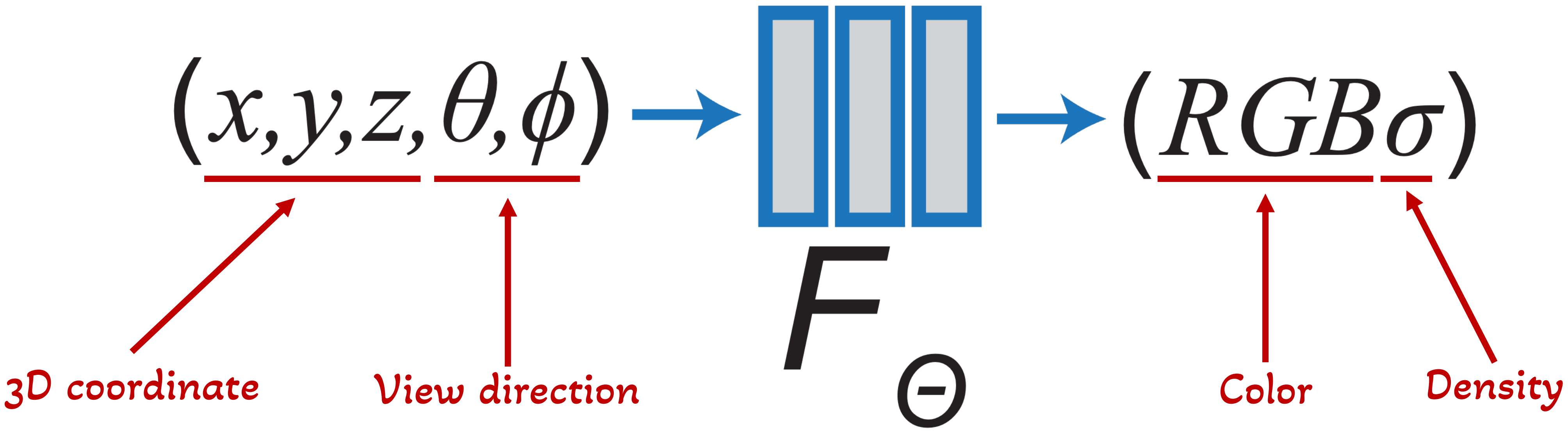
Recent 3D reconstruction techniques like NeRF produce implicit functions, taking advantage of their *flexibility* and *expressiveness*.



Mildenhall et al., 2020

Implicit Functions - NeRFs

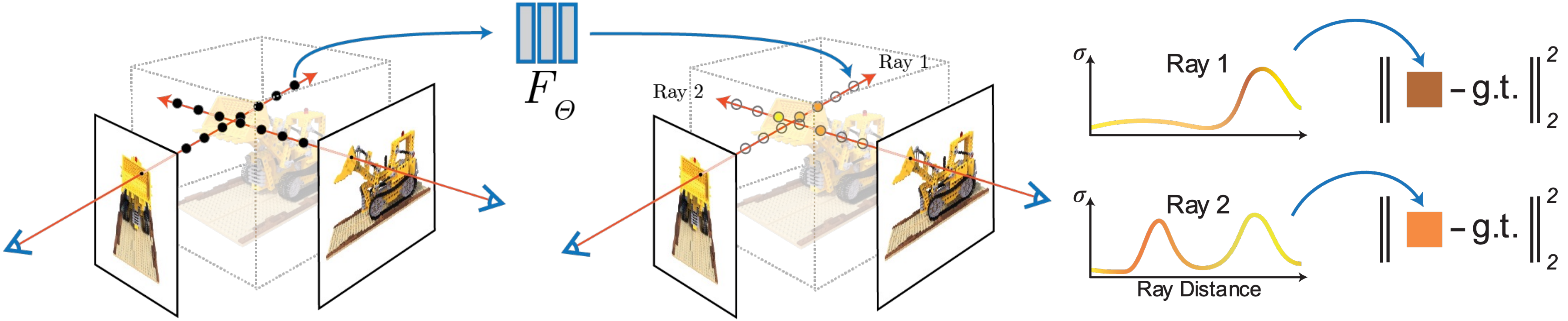
NeRF represents 3D scenes as vector-valued functions parameterized by neural networks.



Mildenhall et al., 2020

Implicit Functions - NeRFs

The networks are trained by comparing images produced via *volume rendering* and their ground truth in the dataset.



Mildenhall et al., 2020

Challenges

1. Irregularity of 3D data;

2. Lack of high-quality, large-scale 3D datasets.

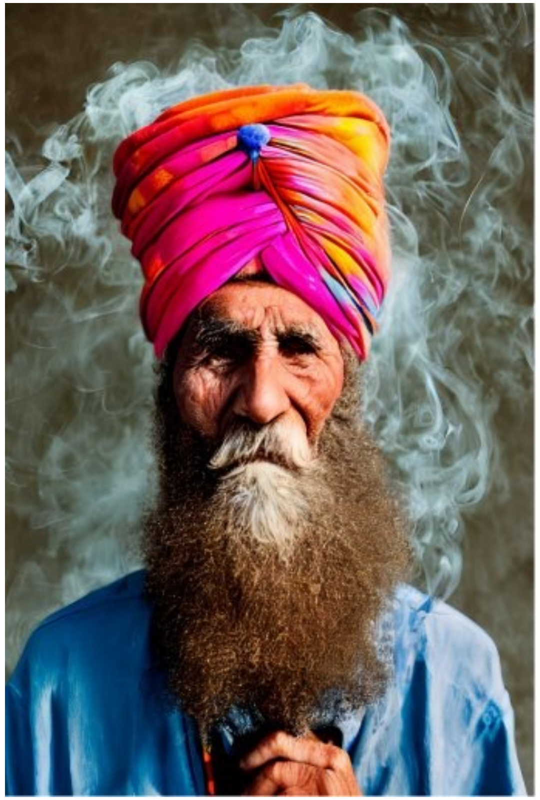
Objaverse-XL Dataset

ShapeNet Dataset

Challenges

Recent 2D image models are achieving impressive performance.

Behind their success is the availability of large-scale image datasets.



Stable Diffusion

Challenges

Interestingly, we perceive the world through its 2D projections.

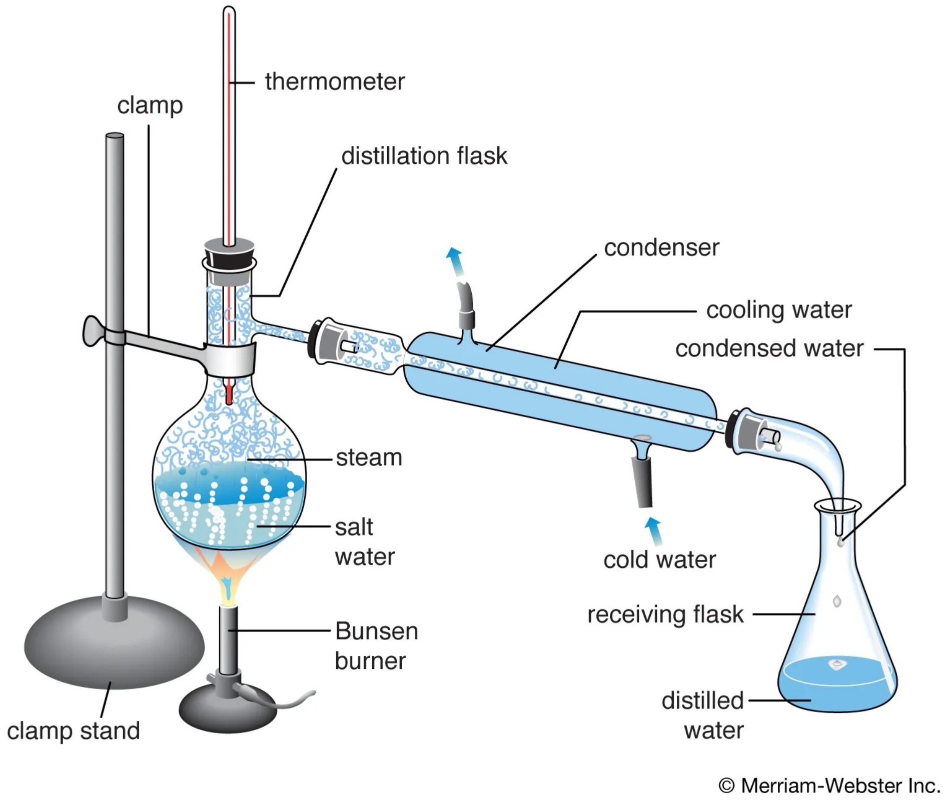
Human artists manipulate virtual objects based on 2D perception.

Blender

Can we train 3D generative models using 2D supervision?

Distilling 2D Generative Models

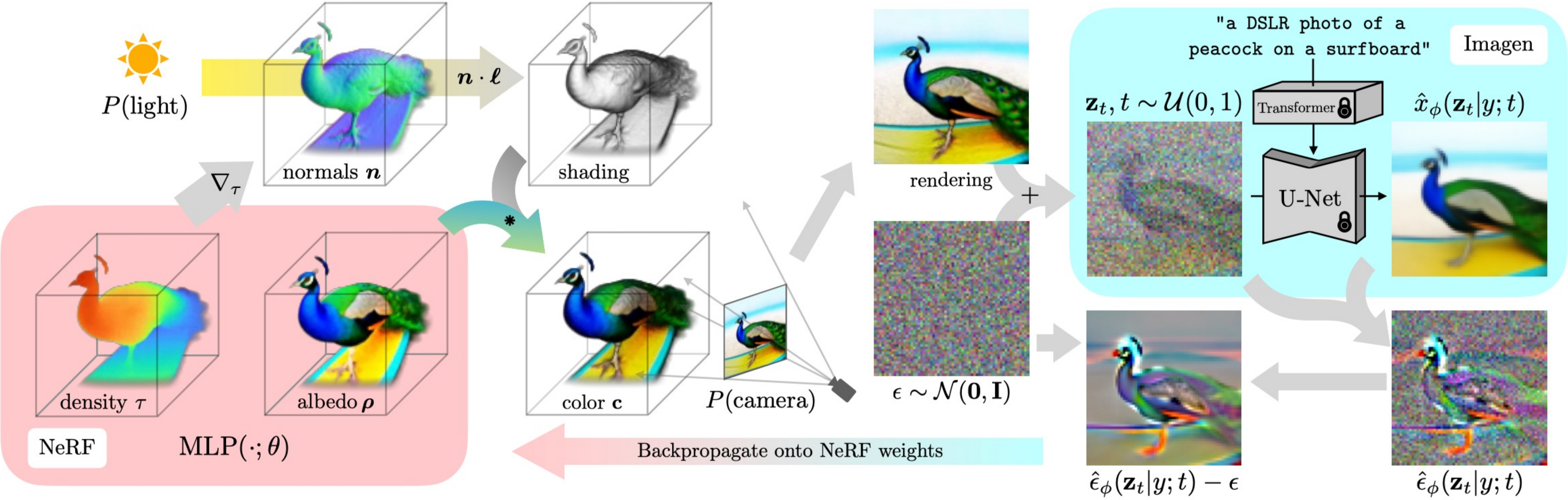
Distillation: The process of separating the components or substances from a liquid mixture by using selective boiling and condensation.



Distillation, Britannica

Distilling 2D Generative Models

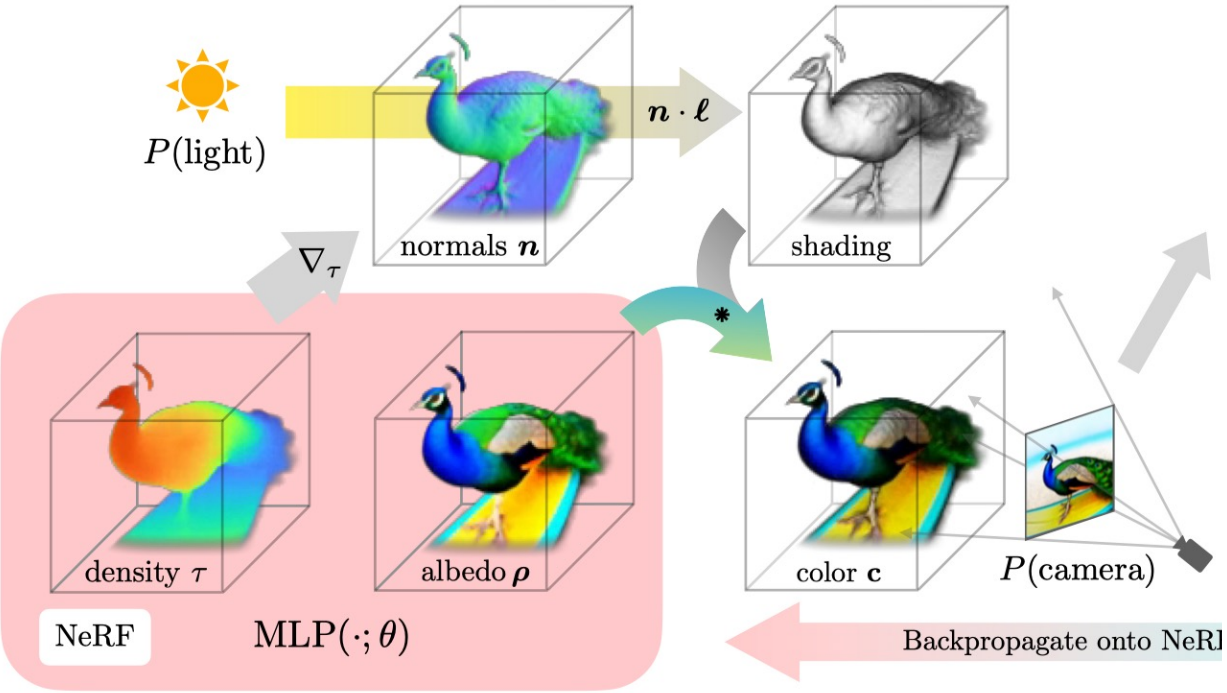
Key Idea: Leverage 2D image generative models as critique that evaluate how *realistic* the given images are.



Poole et al., 2023

Distilling 2D Generative Models

Optimize the parameters (e.g., density) of a scene represented by NeRF instead of other representations (e.g., point clouds, meshes).

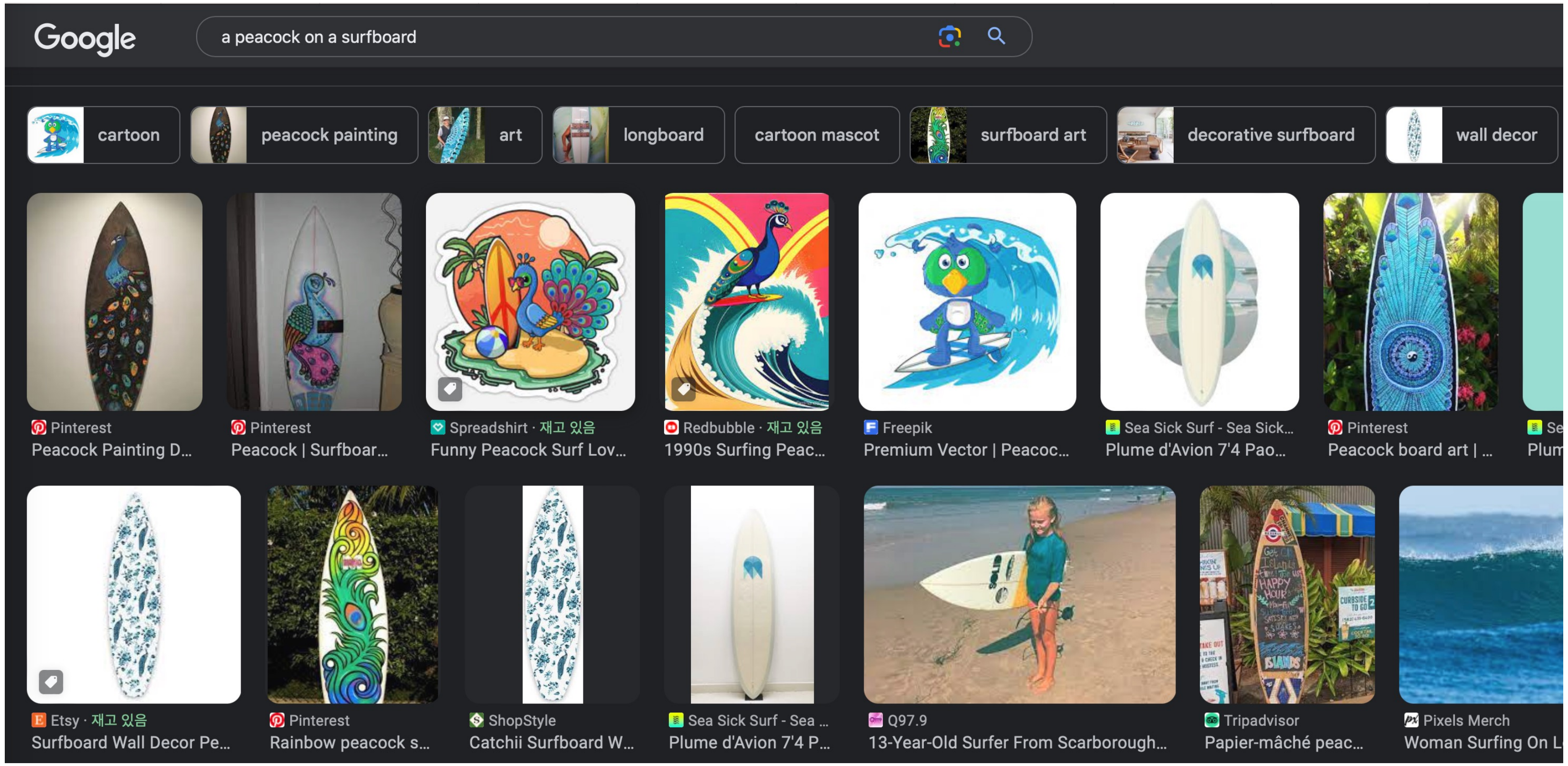


Prompt: "A peacock on a surfboard"

Distilling 2D Generative Models

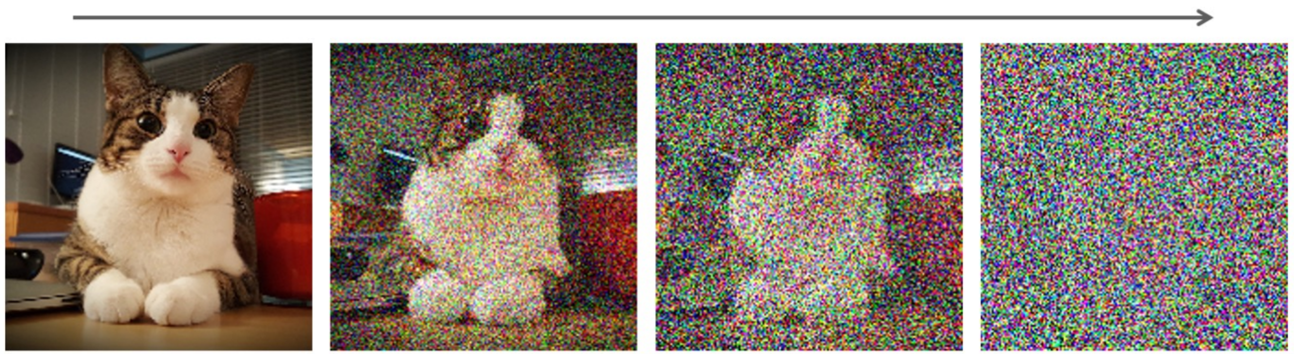
NeRF reconstructs 3D scenes from 2D images.

Do we have photos of "*a peacock on a surfboard*"?

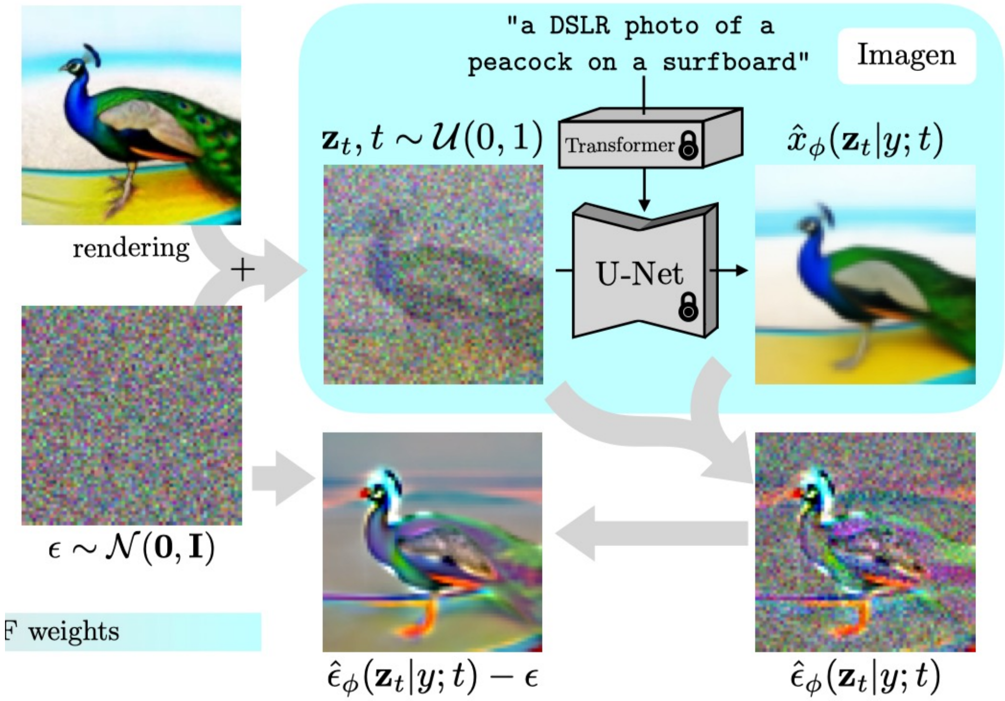


Distilling 2D Generative Models

Parameter updates are driven by guidance signals from a pre-trained text-to-image model.



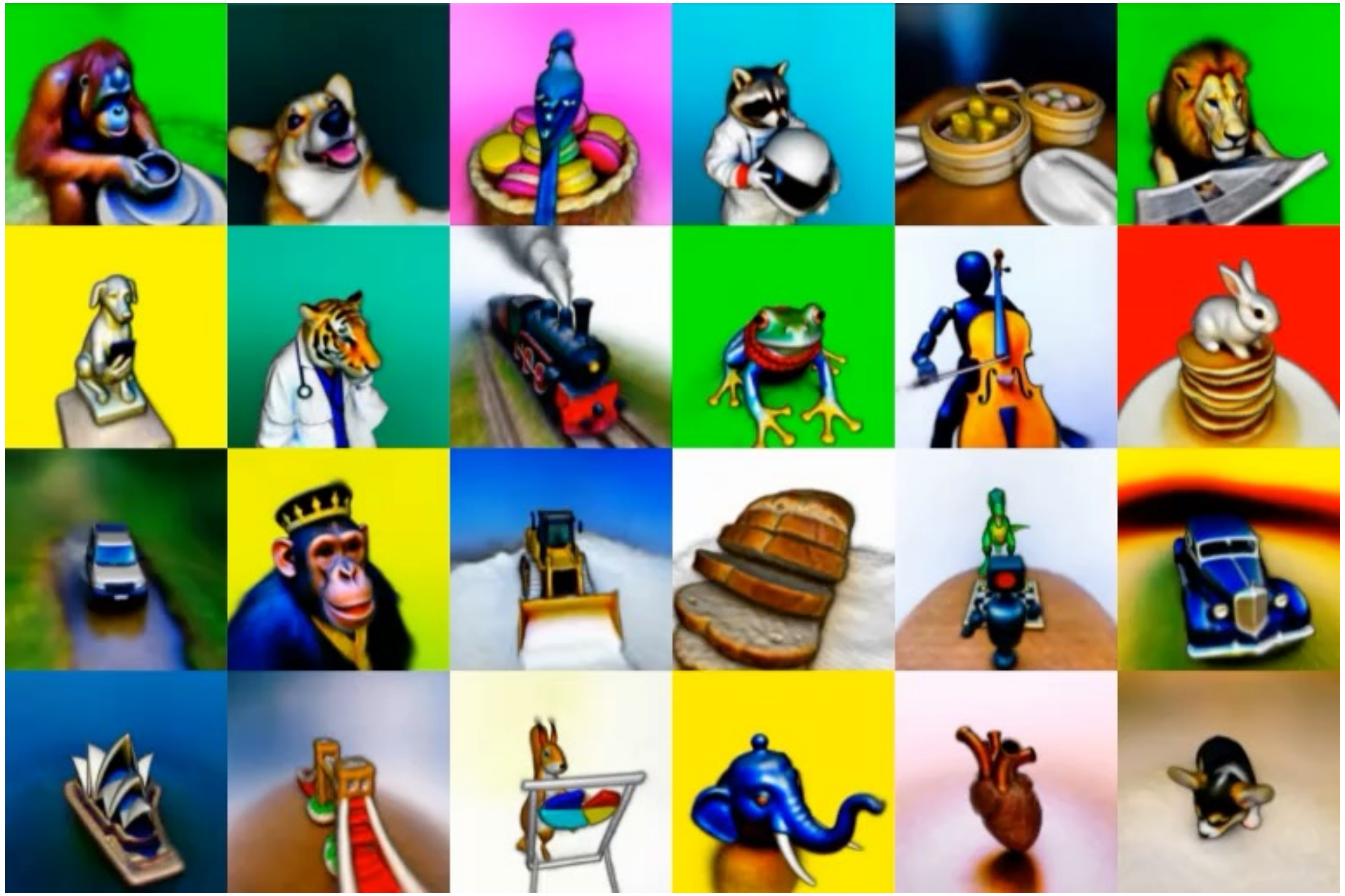
“How much noise should be removed to make the given image realistic?”



Recent Progress

First introduced in Sept. 2022. by Poole et al.

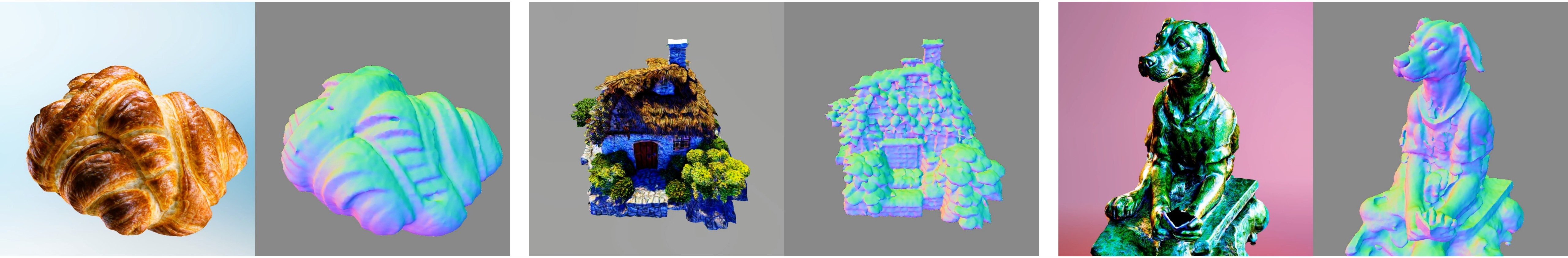
Results lack photorealism due to oversaturation.

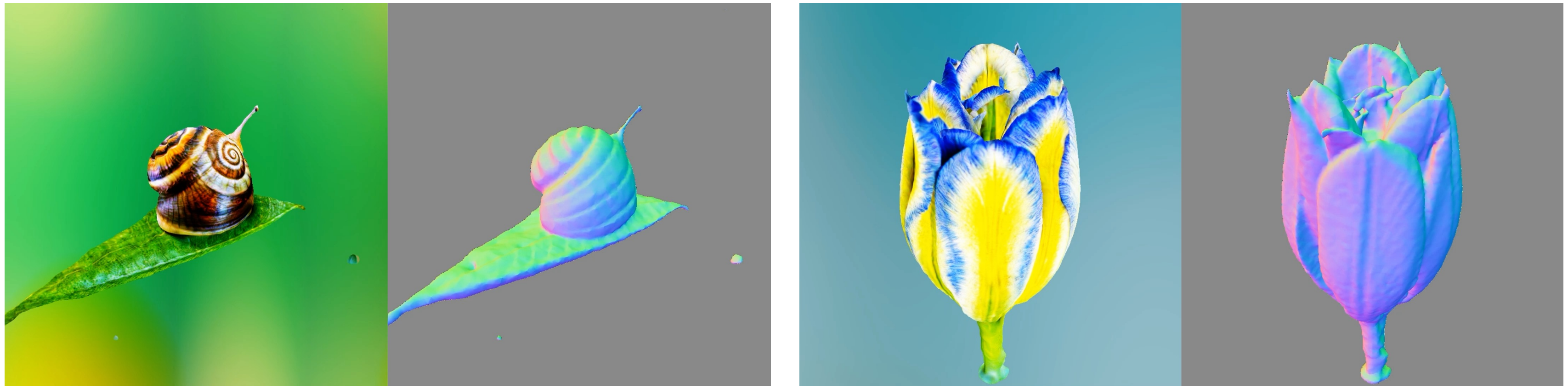


Recent Progress

Wang et al. improved the quality by incorporating variational sampling.

Compared to the predecessor, outputs are more photorealistic.





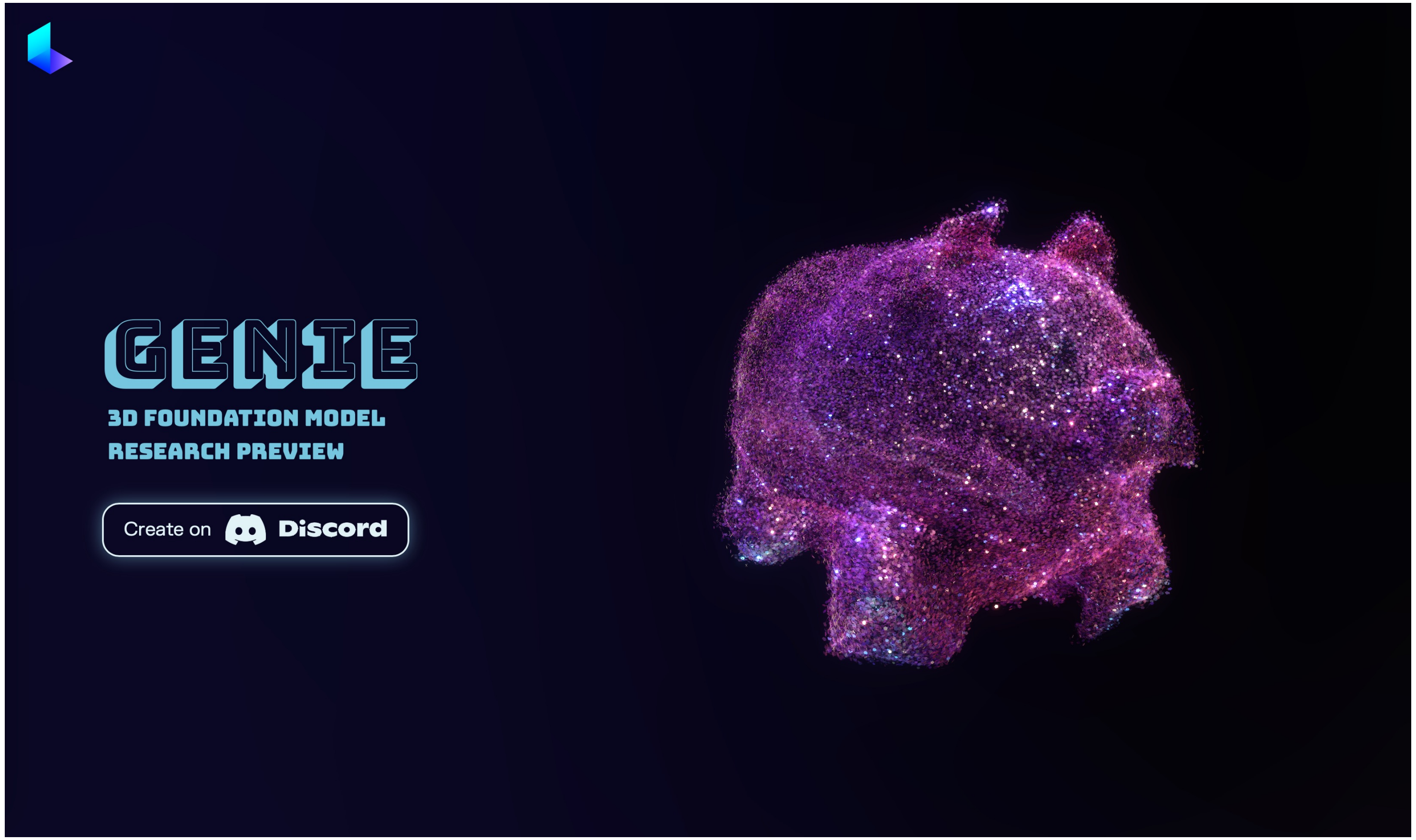
Recent Progress

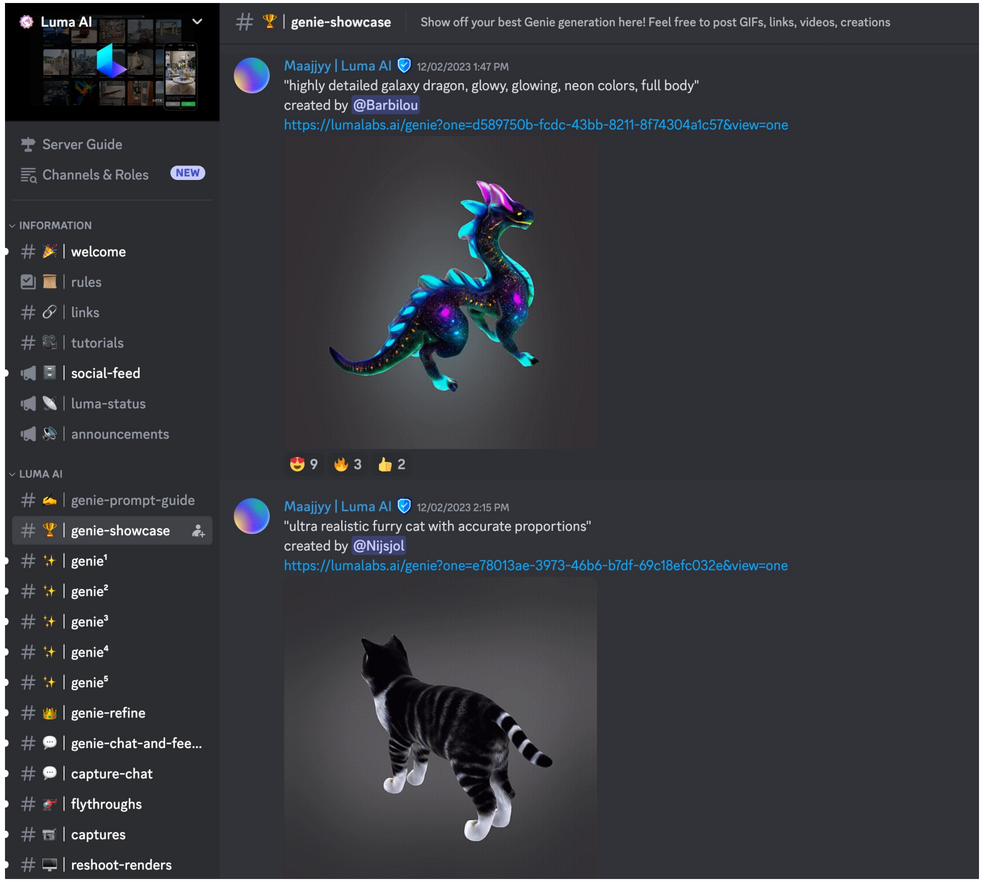
Today, text-to-3D has become one of the most rapidly growing fields in computer graphics and vision.

Liang et al., 2023

Recent Progress

The technology is already being shipped to by tech startups.





Genie, Luma AI

References

- Karras et al., A Style-Based Generator Architecture for Generative Adversarial Networks, CVPR 2019
- Song et al., Score-Based Generative Modeling through Stochastic Differential Equations, ICLR 2021
- Zhang et al., Adding Conditional Control to Text-to-Image Diffusion Models, ICCV 2023
- Agarwal et al., Building Rome in a Day, ICCV 2009
- Mathotaarachchi et al., VoxelStats: A MATLAB Package for Multi-Modal Voxel-Wise Brain Image Analysis, Neuroinform 2016
- Li et al., FPNN: Field Probing Neural Networks for 3D Data, NeurIPS 2016
- Hanocka et al., MeshCNN: A Network with an Edge, SIGGRAPH 2019
- Wu et al., Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling, NeurIPS 2016
- Cai et al., Learning Gradient Fields for Shape Generation, ECCV 2020
- Hoang et al., A Deep Learning Method for 3D Object Classification Using the Wave Kernel Signature and A Center Point of the 3D-Triangle Mesh Electronics 2019

References

- Mescheder et al., Occupancy Networks: Learning 3D Reconstruction in Function Space, CVPR 2019
- Park et al., DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation, CVPR 2019
- Mildenhall et al., NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, ECCV 2020
- Poole et al., DreamFusion: Text-to-3D using 2D Diffusion, ICLR 2023
- Wang et al., ProlificDreamer: High-Fidelity and Diverse Text-to-3D Generation with Variational Score Distillation, NeurIPS 2023
- Liang et al., LucidDreamer: Towards High-Fidelity Text-to-3D Generation via Interval Score Matching, arXiv 2023

Generative AI for 3D Data

"A photo of a snowy town in a Christmas mood."
by DALL-E 3

Seungwoo Yoo

KAIST Geometric AI Lab

https://dvelopery0115.github.io/