SDF Fields meet 3D shape generation

郑欣阳
清华大学 & 微软亚洲研究院

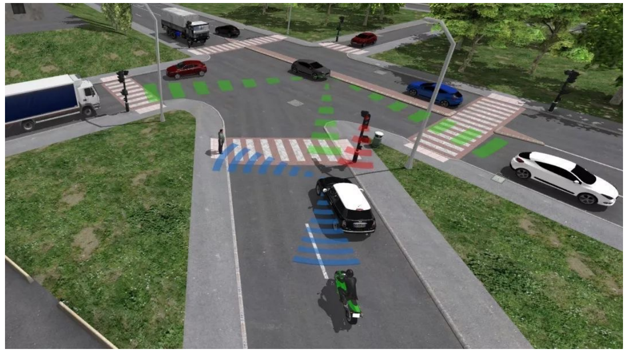
2023/9/7

Introduction

Virtual Reality

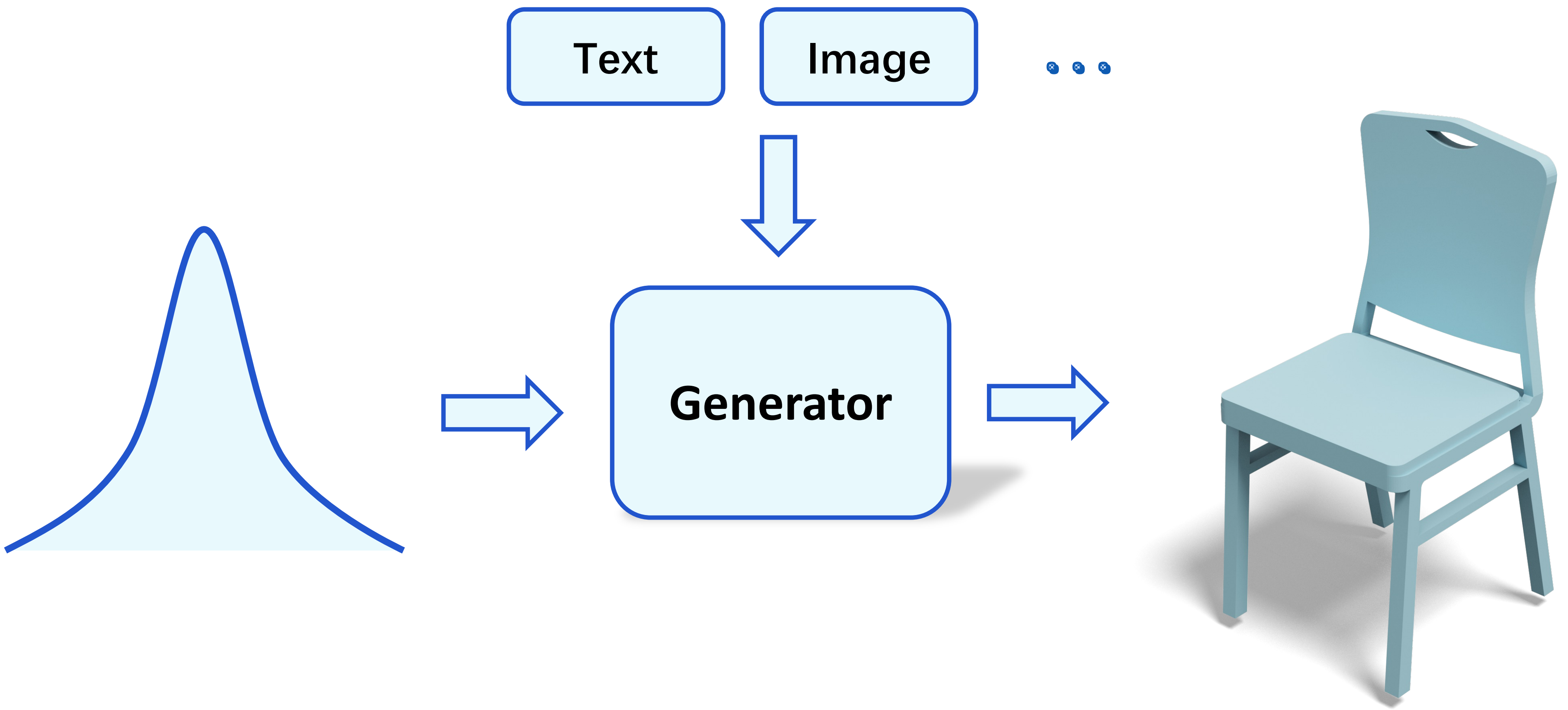
Movie

Game



Simulation

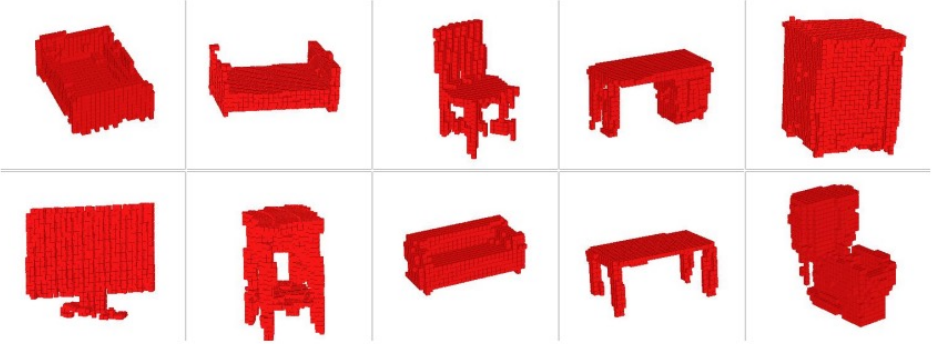
Introduction



Related Works

- 3D representations
  - Voxel

3D-GAN
[Wu et al. 2016]

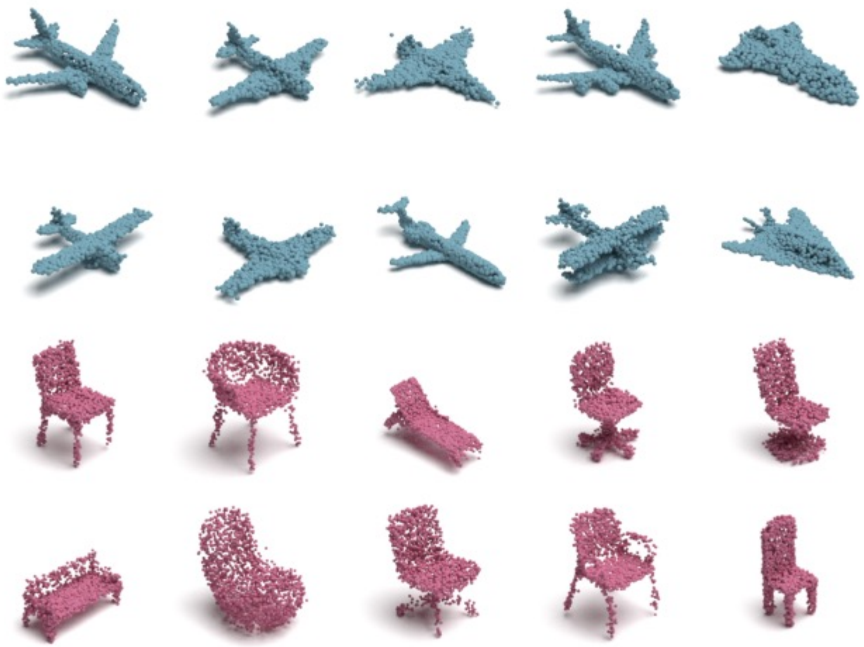


3D-IWGAN
[Smith and Meger 2017]

Related Works

- 3D representations
  - Voxel
  - Point Cloud

SP-GAN
[Li et al. 2021]



PC-Diffusion
[Luo and Hu 2021]

Related Works

- 3D representations
  - Voxel
  - Point Cloud
  - Mesh

PolyGen
[Nash et al. 2020]

MeshDiffusion
[Liu et al. 2023]

Related Works

- 3D representations
  - Voxel
  - Point Cloud
  - Mesh
  - Implicit Function
  - …

ShapeGAN
[Kleineberg et al. 2020]

ImplicitGrid
[Ibing et al. 2021]

Related Works

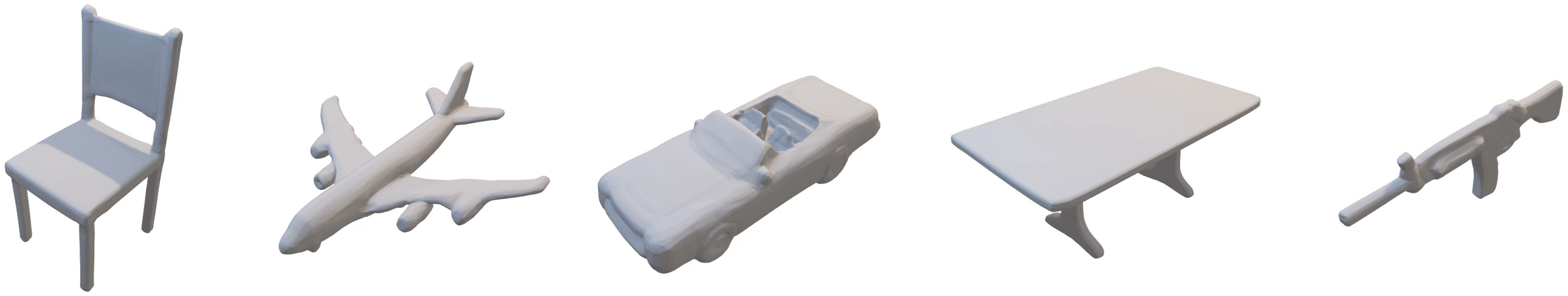
- GAN based method
  - 3DGAN [Wu et al. 2016], L-GAN [Achlioptas et al. 2018], SurfGen [Luo et al. 2021], IM-GAN [Chen and Zhang 2019], Implicit-Grid [Ibing et al. 2021]
- Diffusion based method
  - ShapeGF [Cai et al. 2020], PVD [Zhou et al. 2021], LION [Zeng et al. 2022], MeshDiffusion [Liu et al. 2023], Wavelet Diffusion [Hui et al. 2022], NFD [Shue et al. 2023], SDFusion [Cheng et al. 2023], DIffusionSDF [Chou et al. 2022]

1. **Geometry quality determined by the chosen 3D representation and the capability of generative models**
2. **Shape diversity affected by generative models**

SDF-StyleGAN: Implicit SDF-Based StyleGAN for 3D Shape Generation

Xin-Yang Zheng[1], Yang Liu[2], Peng-Shuai Wang[2], Xin Tong[2]

[1]Tsinghua University, [2]Microsoft Research Asia



Our Observation

- Existing 3D generators are weaker than SOTA 2D generators

2D Generators

BigGAN

StyleGAN

⋮

3D Generators

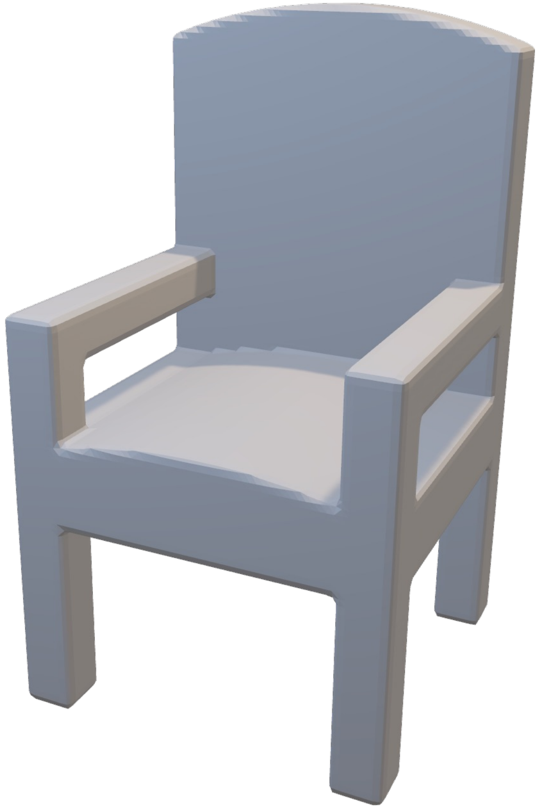
IM-Net

DeepSDF

⋮

Our Observation

- Shape normal affects visual perception



0.5% noise level on SDF value

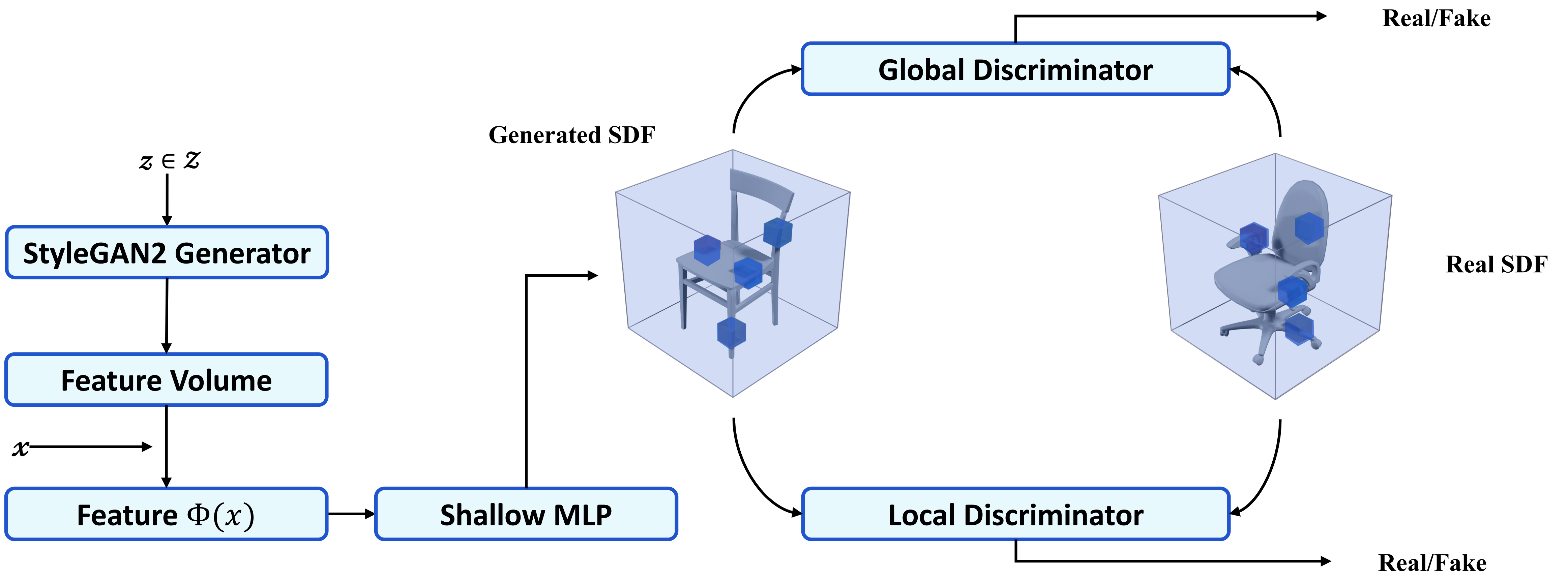
Our Solution

Representation → Feature Volume based Implicit SDF

Generator Design → StyleGAN based Generator

Discriminator Design → SDF Gradients for Discriminator
Global and Local Discriminators

Network Overview



Experiment and Analysis

- Dataset
  - Five categories from ShapeNet
    - chair (6778), airplane (4045), car (7497), table (8509), and rifle (2372)
  - SDF field resolution: 128^3

- Experiment Setup
  - 8 Tesla V100 GPUs (16 GB memory)
  - 2 days for training on average
  - Marching cube resolution: 128^3

Metrics

- Previous Metrics
  - Distance based (CD, EMD, LFD [Chen et al. 2003])
    - COV [Achliopts et al. 2018]
    - MMD [Achliopts et al. 2018]
    - 1-NNA [Yang et al. 2019]
    - ECD [Ibing et al. 2021]
  - Learning based
    - FPD [Shu et al. 2019]
      - Mimic FID [Heuse et al. 2017]

- **Visual appearance**

FID based on Shading Images

20 Views

Shading Images

$$FID = \frac{1}{20}\left[\sum_{i=1}^{20} \|\mu_g^i - \mu_r^i\|^2 + \mathrm{Tr}\left(\Sigma_g^i + \Sigma_r^i - 2\left(\Sigma_r^i \Sigma_g^i\right)^{1/2}\right)\right]$$

Quantitative Comparison

Microsoft Research 微软亚洲研究院

Training Configurations

	chair table	airplane car rifle
IMGAN [Chen and Zhang 2019]	64	128
Implicit-Grid [Ibing et al. 2021]	256	
ShapeGAN [Kleineberg et al. 2020]	128	
ours	128	

Data	Method	COV(%) ↑	MMD ↓	1-NNA↓	ECD↓	FID↓	FPD↓
Chair	IMGAN	72.57	**3326**	0.7042	1998	63.42	1.093
	Implicit-Grid	**82.23**	3447	**0.6655**	**1231**	119.5	1.456
	ShapeGAN	65.19	3726	0.7896	4171	126.7	1.177
	SDF-StyleGAN	75.07	3465	0.6690	1394	**36.48**	**1.040**
Airplane	IMGAN	76.89	**4557**	**0.7932**	**2222**	74.57	1.207
	Implicit-Grid	**81.71**	5504	0.8509	4254	145.4	2.341
	ShapeGAN	60.94	5306	0.8807	6769	162.4	2.235
	SDF-StyleGAN	74.17	4989	0.8430	3438	**65.77**	**0.942**
Car	IMGAN	54.13	2543	0.8970	12675	141.2	1.391
	Implicit-Grid	**75.13**	2549	0.8637	8670	209.3	1.416
	ShapeGAN	57.40	2625	0.9168	14400	225.2	0.787
	SDF-StyleGAN	73.60	**2517**	**0.8438**	**6653**	**97.99**	**0.767**
Table	IMGAN	83.43	**3012**	**0.6236**	**907**	51.70	1.022
	Implicit-Grid	**85.66**	3082	0.6318	1089	87.69	1.516
	ShapeGAN	76.26	3236	0.7069	1913	103.1	**0.934**
	SDF-StyleGAN	69.80	3119	0.6692	1729	**39.03**	1.061
Rifle	IMGAN	71.16	**5834**	0.6911	701	103.3	2.102
	Implicit-Grid	77.89	5921	**0.6648**	**357**	125.4	1.904
	ShapeGAN	46.74	6450	0.8446	3115	182.3	1.249
	SDF-StyleGAN	**80.63**	6091	0.7180	510	**64.86**	**0.978**

Qualitative Results

Microsoft Research 微软亚洲研究院

Qualitative comparison

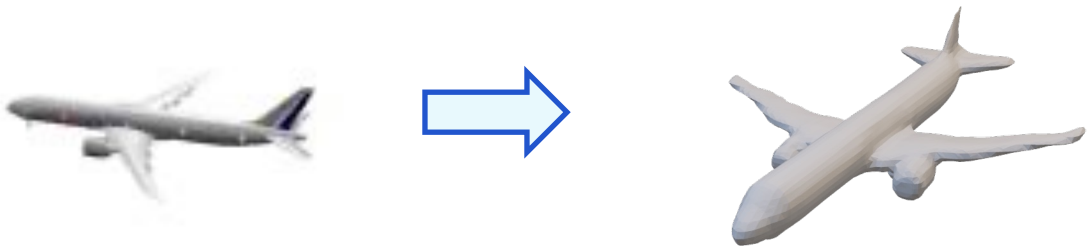
IMGAN [CZ19]

ShapeGAN [KFW20]

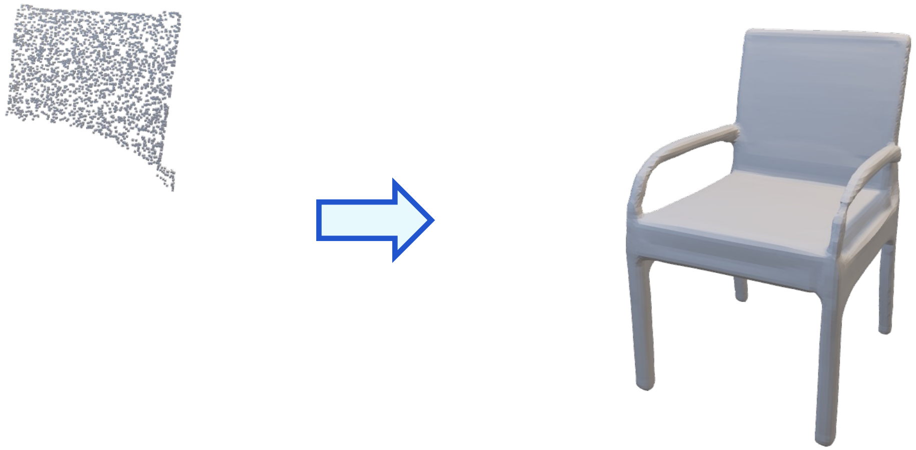
Implicit-Grid [ILK21]

SDF-StyleGAN

Applications



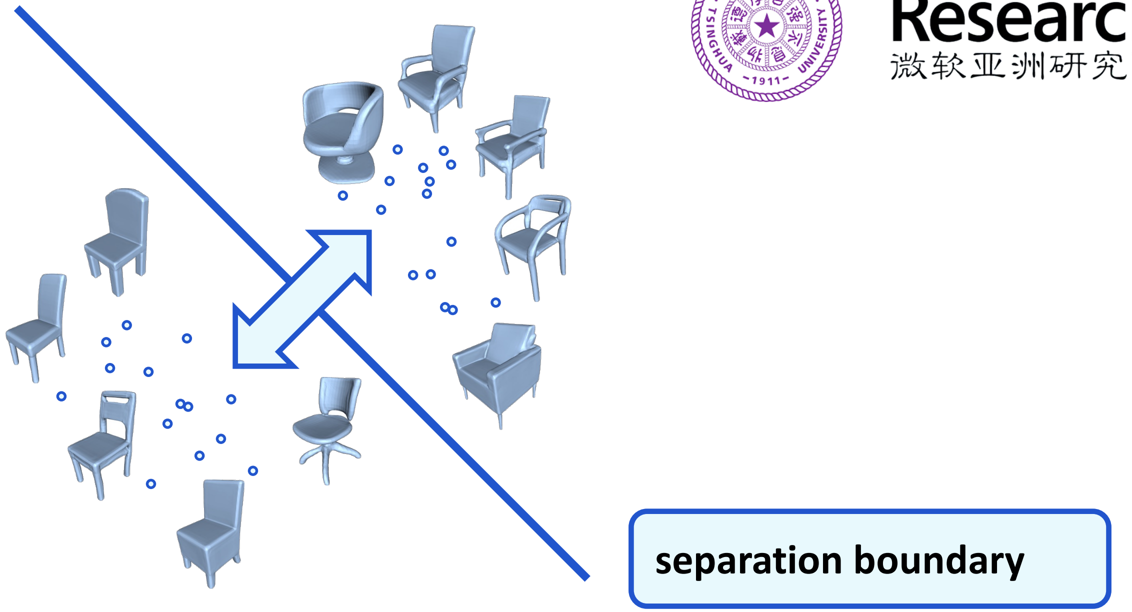
Single Image Based Reconstruction

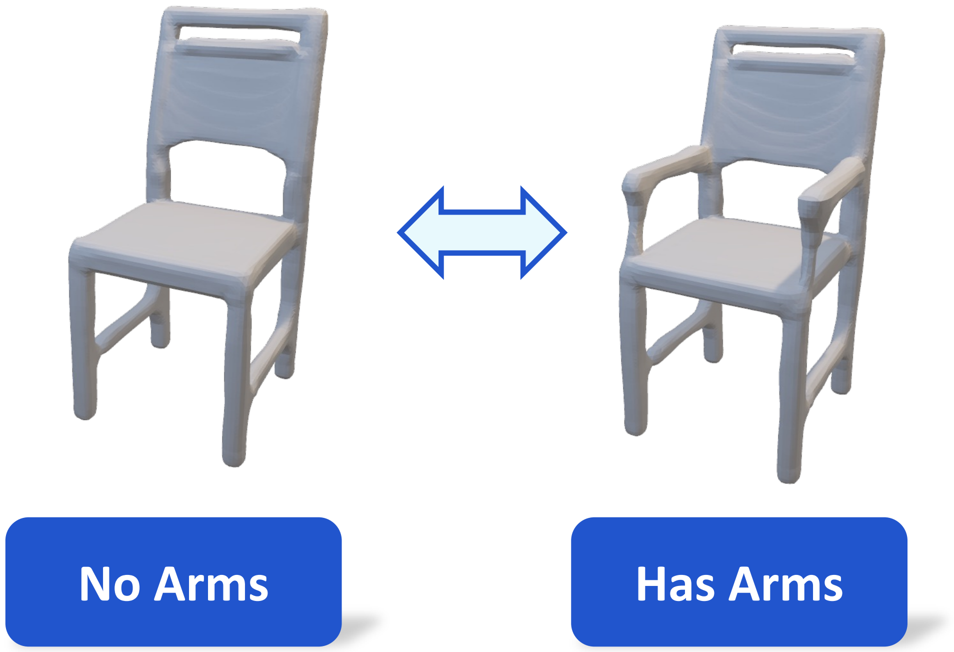


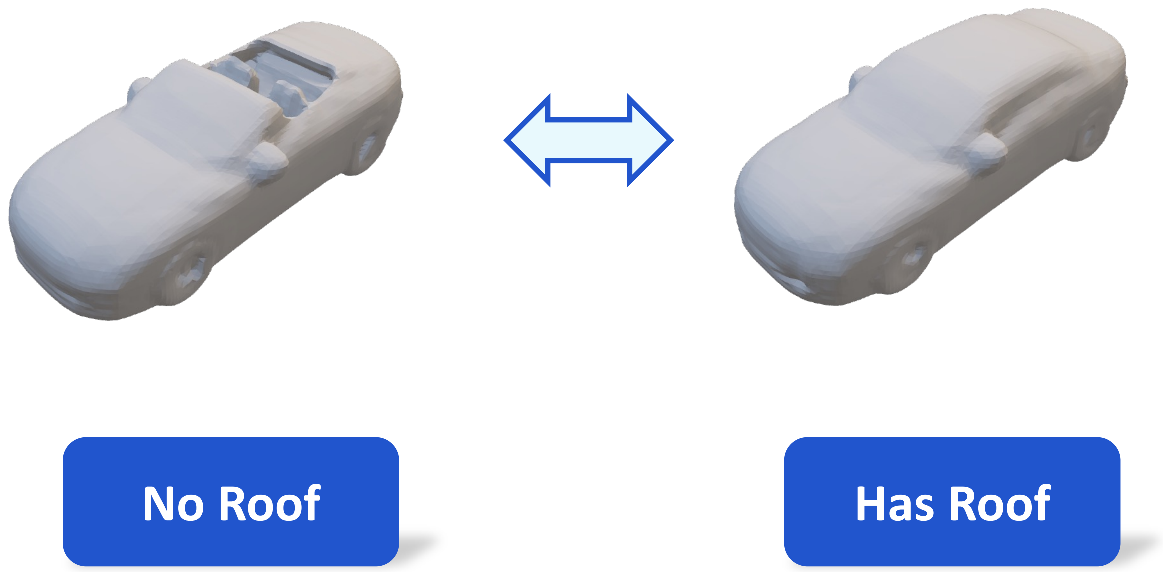
Shape Completion

Shape Style Edit

- InterfaceGAN [SYTZ20]
- Data label: from [MKC18]







Take-Away Messages

- SDF-StyleGAN: A GAN based method for 3D shape generation
  - Use strong generator backbone
  - Global and local discriminators
  - SDF gradients do matter
  - Better evaluation metric: calculate FID on shading images

- Welcome to visit our project page
  - https://zhengxinyang.github.io/projects/SDF_stylegan.html
  - For paper, code, model, and data

SIGGRAPH 2023
LOS ANGELES+ 6-10 AUG

THE PREMIER CONFERENCE & EXHIBITION ON
COMPUTER GRAPHICS & INTERACTIVE TECHNIQUES

LOCALLY ATTENTIONAL SDF DIFFUSION FOR CONTROLLABLE 3D SHAPE GENERATION

XIN-YANG ZHENG[1], HAO PAN[2], PENG-SHUAI WANG[3], XIN TONG[2], YANG LIU[2], HEUNG-YEUNG SHUM[1,4]

[1]TSINGHUA UNIVERSITY, [2]MICROSOFT RESEARCH ASIA,
[3]PEKING UNIVERSITY, [4]INTERNATIONAL DIGITAL ECONOMY ACADEMY

Controlled 3D shape Generation

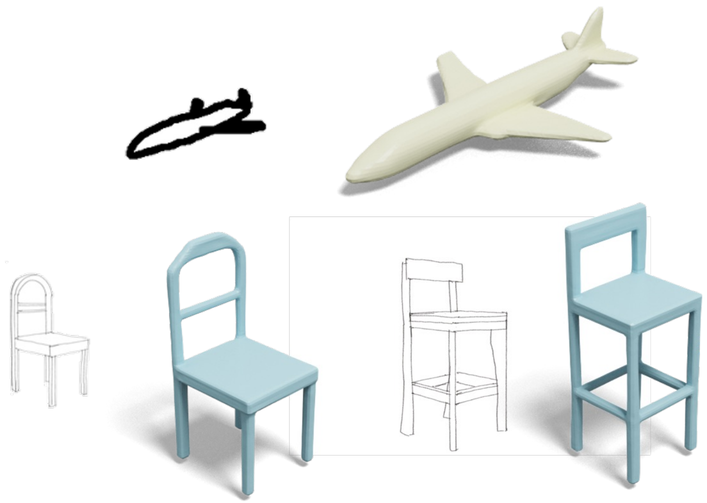
- Conditioned 3D generation
  - Text-based methods: CLIP-Forge [Sanghi et al. 2022], CLIP-Sculptor [Sanghi et al. 2023]
  - Image-based methods: AutoSDF [Mittal et al. 2022], 3DILG [Zhang et a. 2022], DiffusionSDF [Chou et al. 2022]
  - Lack of local controllability
- Sketch-based 3D reconstruction
  - Sketch2Model [Zhang et al. 2021] ,Sketch2Mesh [Guillard et al. 2021]…
  - Fail to handle one-to-many mapping between sketch and 3D shapes and result in low quality results

Our Contributions

- LAS-Diffusion: a 3D diffusion model for generating 3D shapes
  - High fidelity with rich geometric features

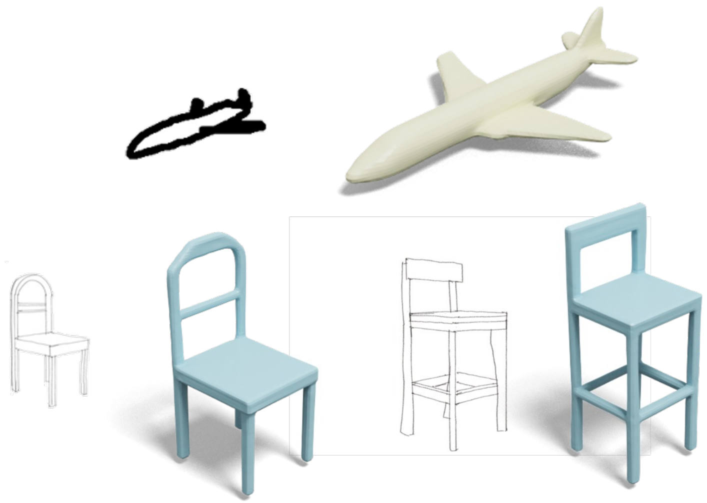
Our Contributions

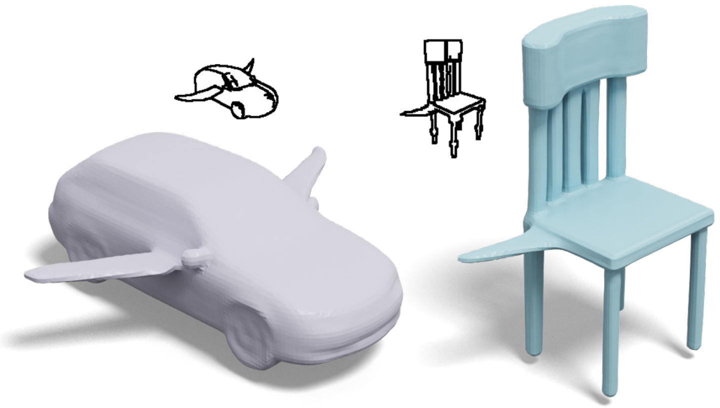
- LAS-Diffusion: a 3D diffusion model for generating 3D shapes
  - High fidelity with rich geometric features
  - Efficient 2D sketch control for 3D shape generation



Our Contributions

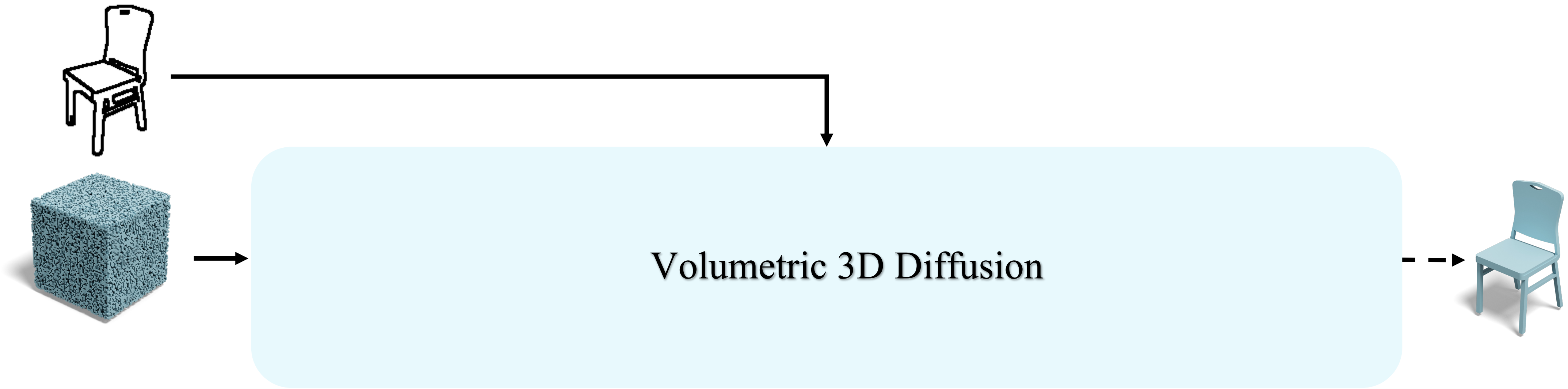
- LAS-Diffusion: a 3D diffusion model for generating 3D shapes
  - High fidelity with rich geometry features
  - Efficient 2D sketch control for 3D shape generation
  - Capability for generating 3D shapes with unseen geometric features





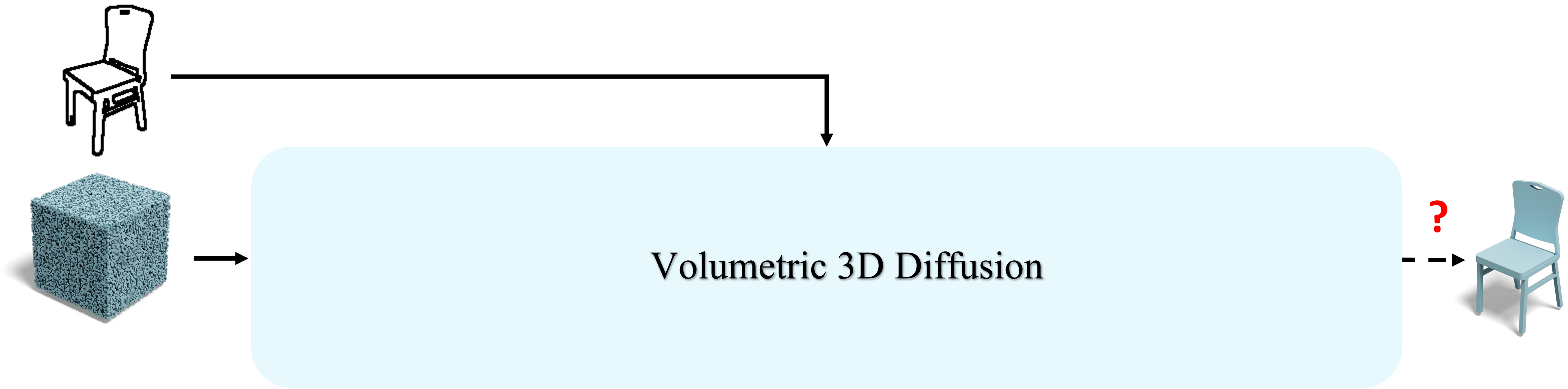
Our Method

- 3D shapes are modeled by signed distance functions (SDF)
- A volumetric 3D diffusion model for 3D SDF generation
- 2D sketch image as the condition of the diffusion model



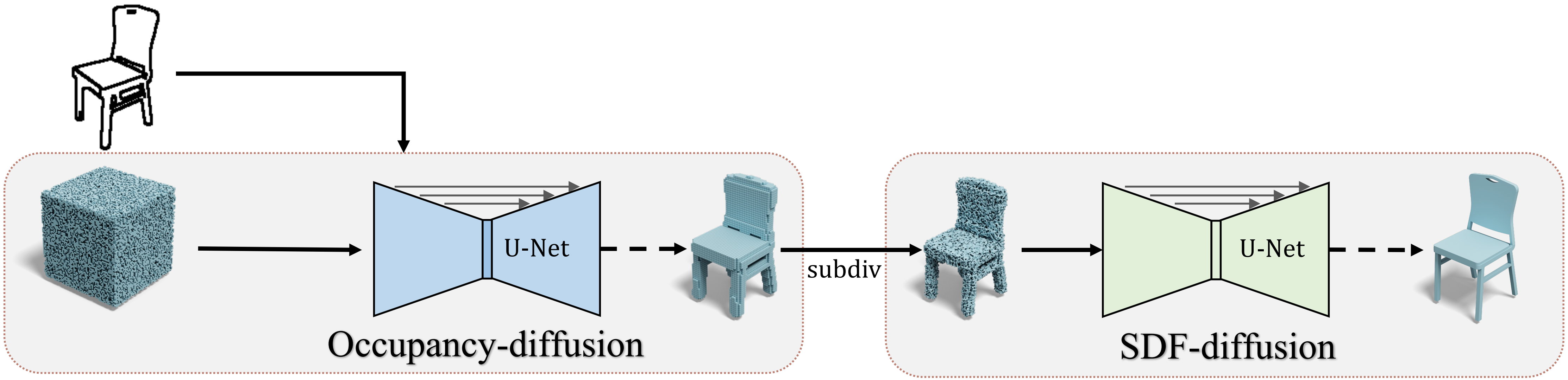
Key Challenge I

- How to generate high fidelity 3D shapes with various geometric features?



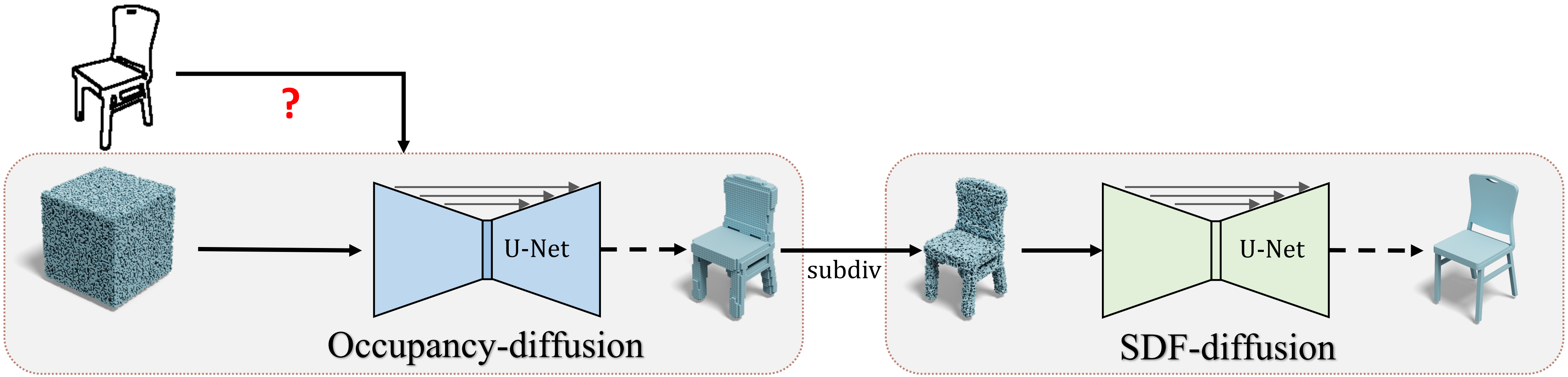
Our Key Idea I

- Two-stage diffusions
  - An occupancy diffusion for generating low-res occupancy field of 3D shapes
  - Subdividing the occupied voxels in high-res volume for SDF diffusion
  - An SDF diffusion for generating detailed shapes within the occupancy voxels



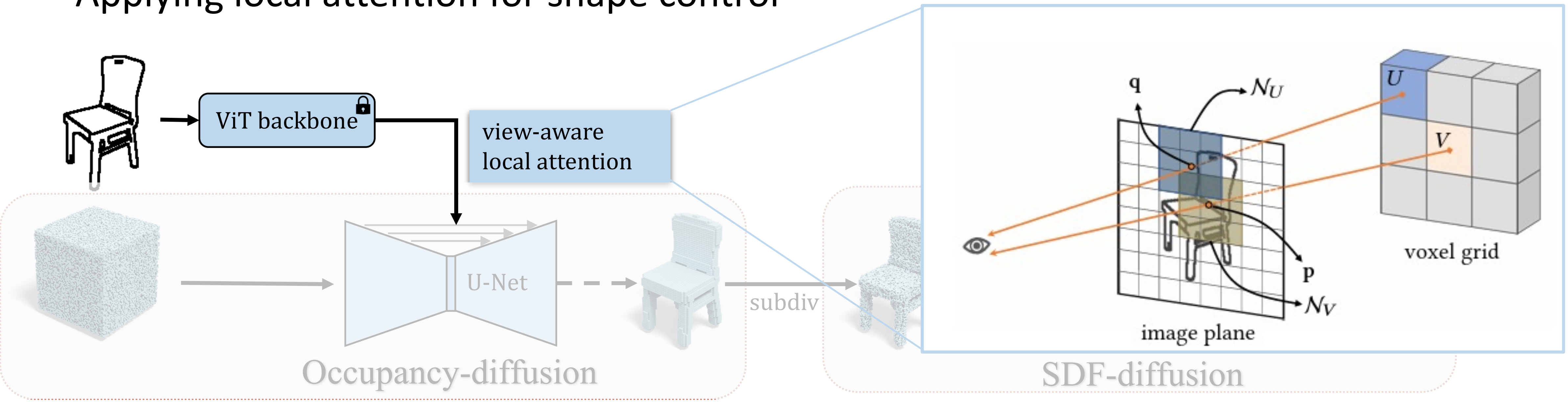
Key Challenge II

- How to control local geometric features of 3D shapes via 2D sketch?



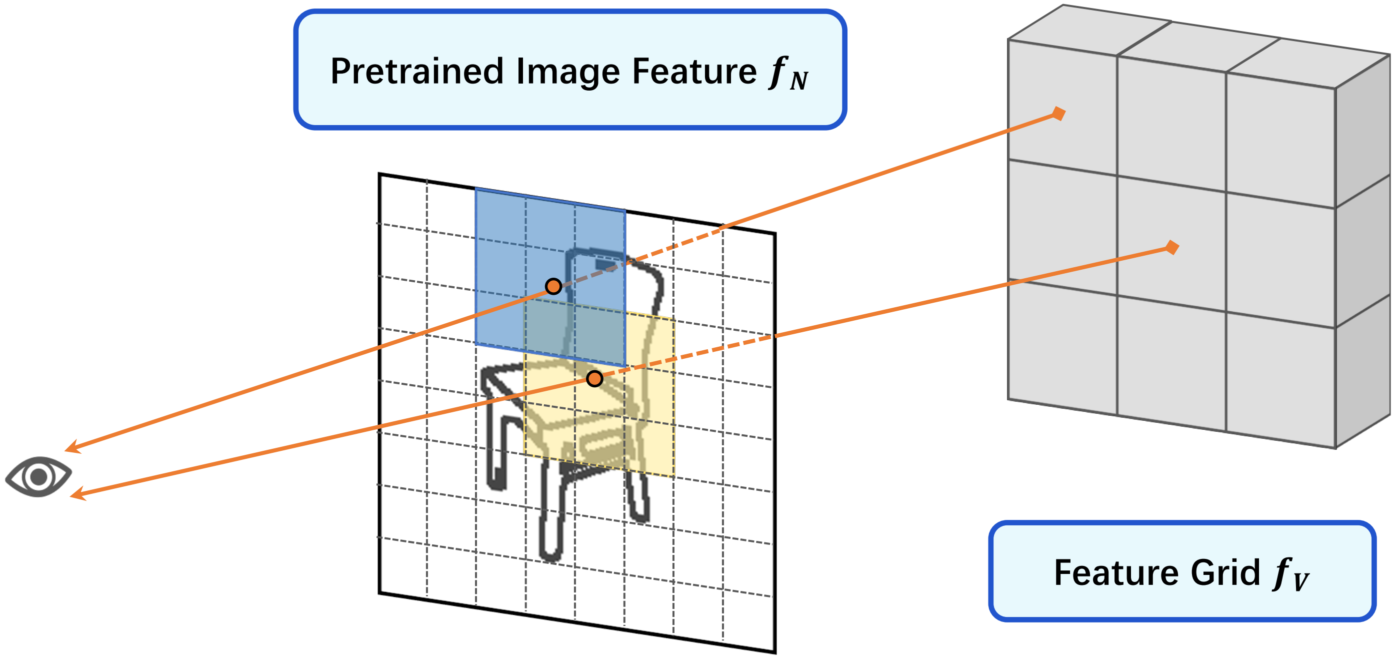
Our Key Idea II

- View-aware local attention scheme
  - Encoding sketch image into patch features
  - Projecting the diffusion feature volume into the image plane according to view direction of the sketch
  - Applying local attention for shape control



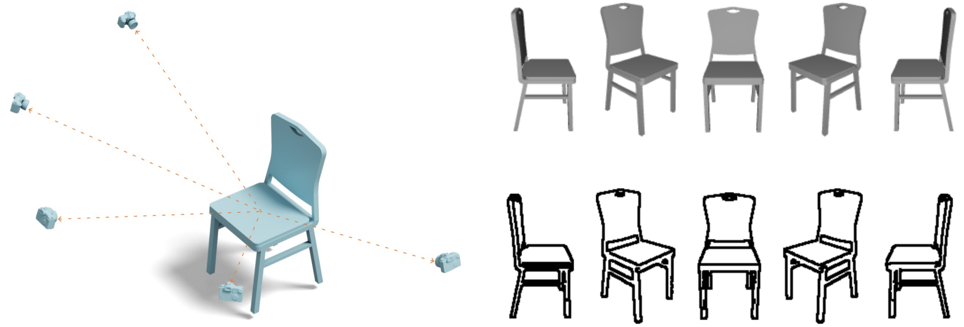
Our Key Idea II

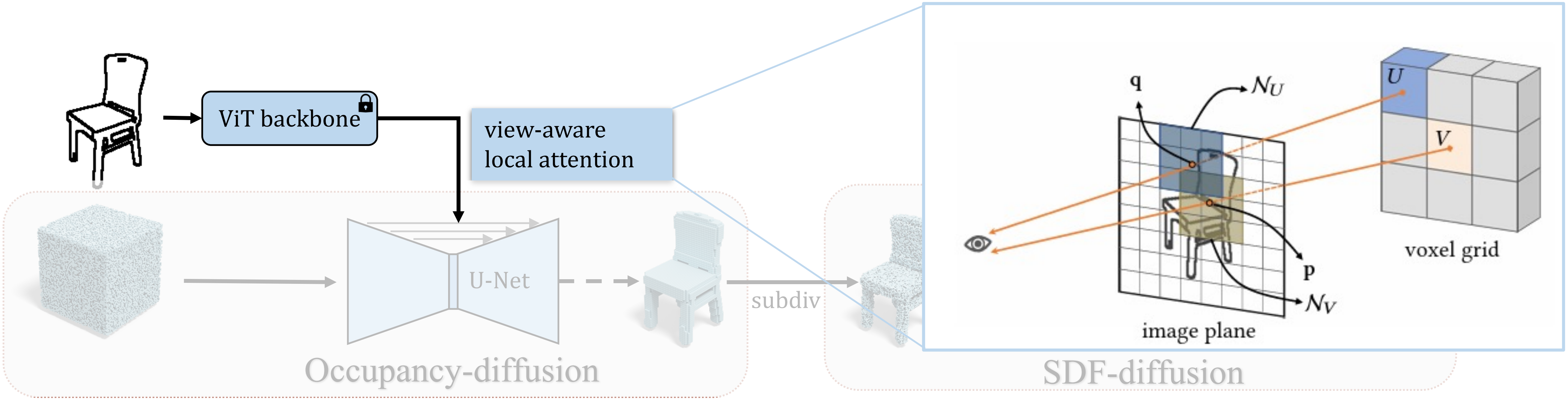
- View-aware local attention scheme



Our Key Idea II

- View-aware local attention scheme
  - Manually specify one of five approximate view directions for input sketch
    - Left, front left, front, front right, right





Our Implementation

- Dataset
  - ShapeNet: chair, table, car, airplane, rifle (29K 3D shapes)
  - Sketches: edge image by Canny detector on the rendered images
- Representation
  - 64^3 for occupancy field
  - 128^3 for SDF
- Network design
  - UNet for occupancy diffusion
  - Octree-based UNet for SDF diffusion
- Training and inference
  - 3 days for training on 8 V100 Nvidia GPUs (category agnostic)
  - 10 sec for inference a 3D shape on Nvidia Geforce 1080 Ti
  - Two stages are trained separately

Unconditional Generation Results

Quantitative evaluation

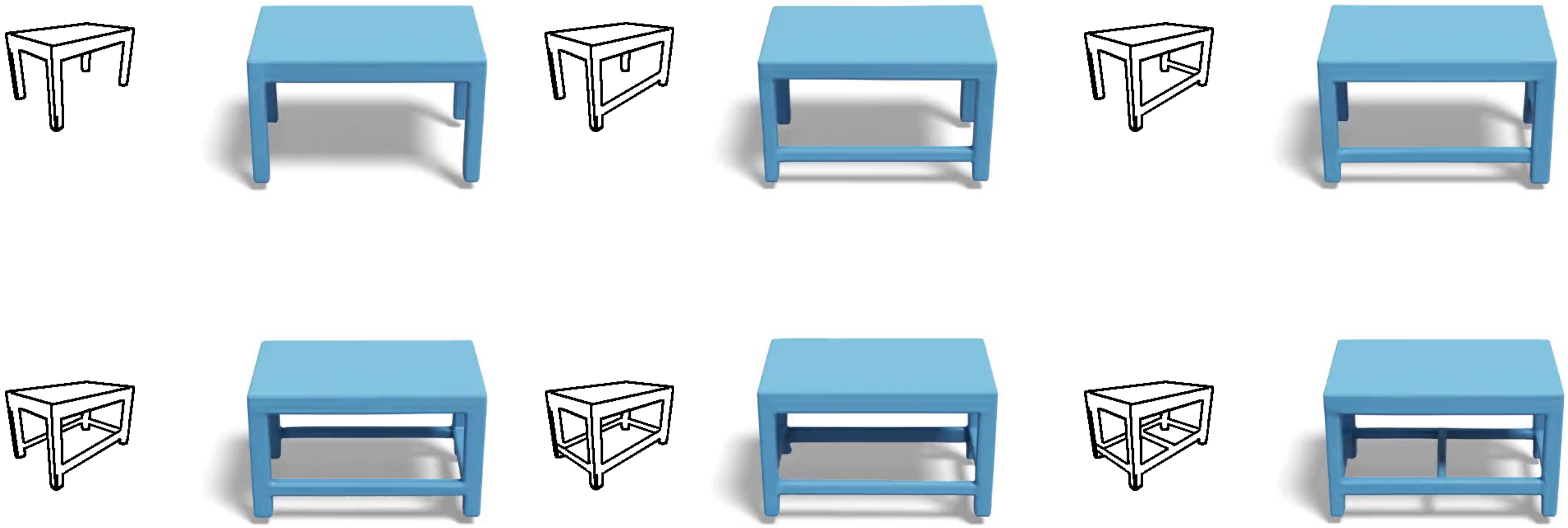
- Shading image based FID [Zheng et al. 2022]

$$FID = \frac{1}{20}\left[\sum_{i=1}^{20} \|\mu_g^i - \mu_r^i\|^2 + \mathrm{Tr}\left(\Sigma_g^i + \Sigma_r^i - 2\left(\Sigma_r^i\Sigma_g^i\right)^{1/2}\right)\right]$$

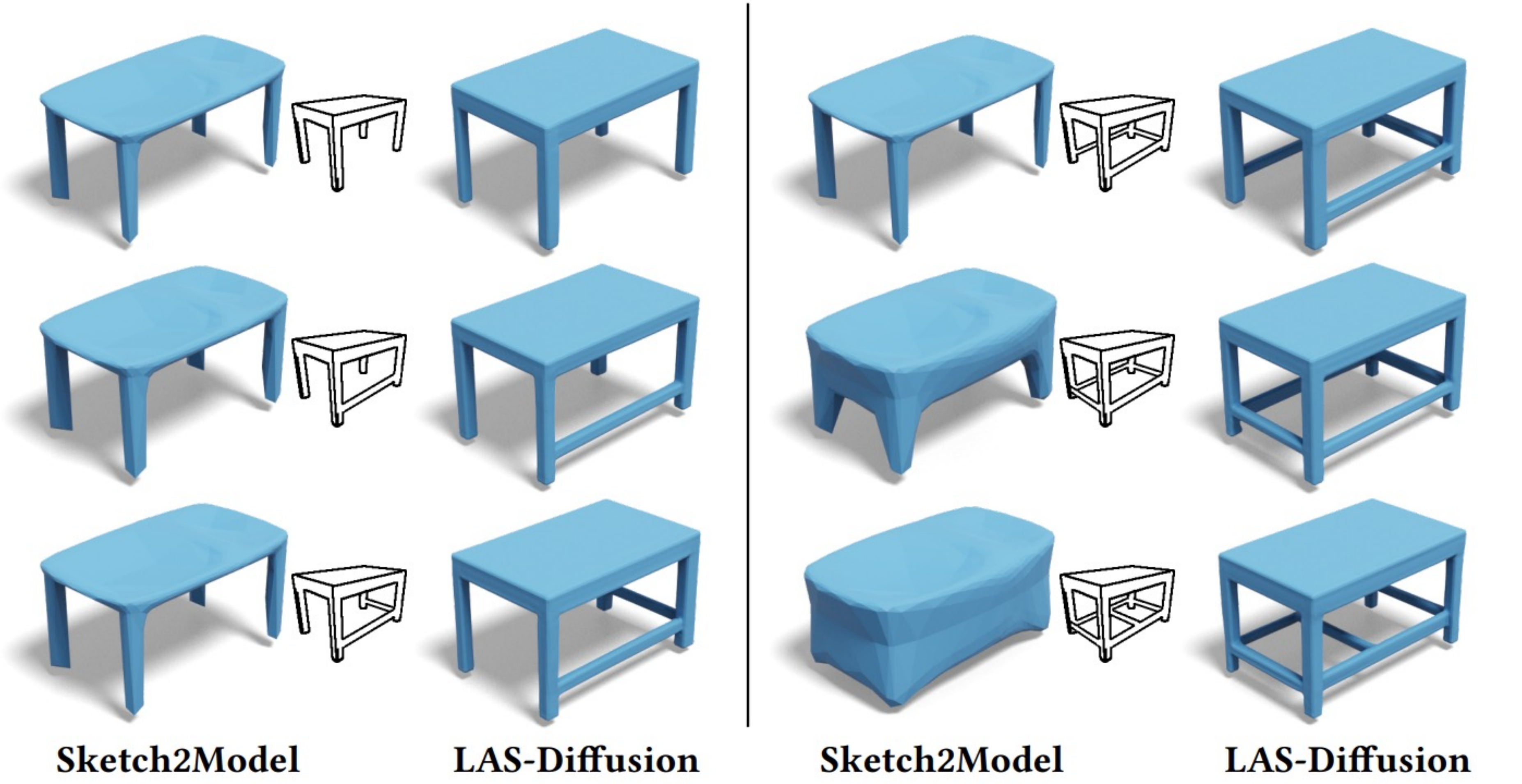
	Chair	Airplane	Car	Table	Rifle
IM-GAN	63.42	74.57	141.2	51.70	103.3
SDF-StyleGAN	36.48	65.77	97.99	39.03	64.86
Wavelet-Diffusion	28.64	35.05	N/A	30.27	N/A
Ours (unconditioned)	**20.45**	**32.71**	**80.55**	**17.25**	**44.93**
3DILG	31.64	54.38	164.15	54.13	77.74
Ours (conditioned)	21.55	43.08	86.34	17.41	70.39

- *Refer to our paper for comparisons under other metrics (e.g. Cov, MMD)*

Sketch Controlled Generation



Local Controllability



Model Robustness

Shape with Unseen Local Features

Take-Away Messages

- LAS-Diffusion: A diffusion-based method for 3D shape generation
  - Two-stage diffusion for generating diverse shapes with rich geometric features
  - View-aware local attention for better control and generalizability

- Welcome to visit our project page
  - https://zhengxinyang.github.io/projects/LAS-Diffusion.html
  - For paper, code, model, and data

Future Works

- Supporting shape appearance
- Scale to large and diverse data
  - Objaverse

Thank you!