

基于数据驱动的 三维几何生成

刘洋

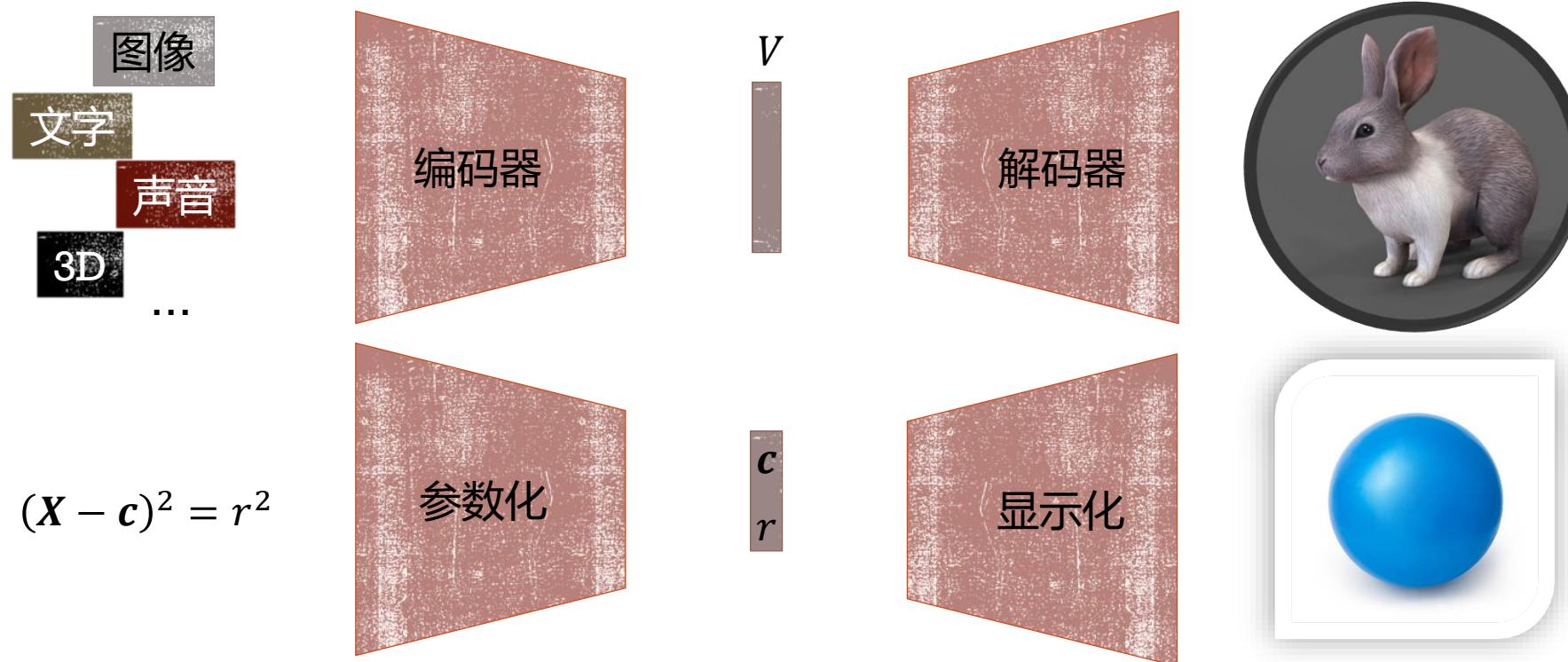
微软亚洲研究院

基于数据驱动三维生成的目的

- 降低三维建模的难度
- 提高三维建模的效率
- 增加三维几何的多样性
- 理解三维形状空间
- ...

生成问题的输入和输出

- 输入：三维内容的隐表达，如一个向量 $V \in \mathbb{R}^K$
- 输出：以某种三维表达形式呈现的三维内容



生成的手段

- 三维解码器 $f^{-1}: V \rightarrow 3D\ content$

- 参数化求逆
- 机器学习
 - 神经网络
 - 卷积神经网络 (CNN)
 - 循环神经网络 (RNN)

- ...

- 强化学习

- ...

三维生成质量的评判标准

- 形似
 - 几何外观与细节、纹理、表观

- 神似
 - 结构、功能

- 多样性 (针对编码器和生成器)

- 效率
 - 时间开销、内存开销

三维表达

三维表示

连续表示（精确表示）

- 参数曲面
- 隐式曲面
- CSG
- BREP
- 分片曲面
- ...

离散表示（近似表示）

- 点云
- 体素
- 网格
- 多视图
- ...

形状分析：识别、分割、对应...

三维表示

低层次表达

- 离散表达
- 连续表达



中低层次表达

- 型元表达
- 结构化表达



中高层表达

- 功能型表达
- 物理属性表达
- 布局表达（物体之间）



低层次表示



规则体素 (REGULAR GRID, VOXELS)

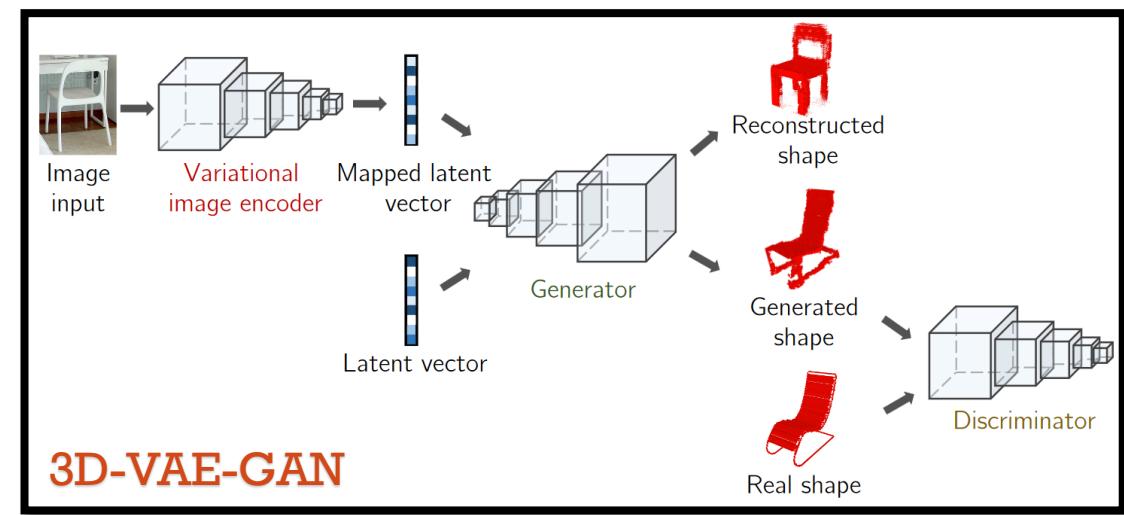
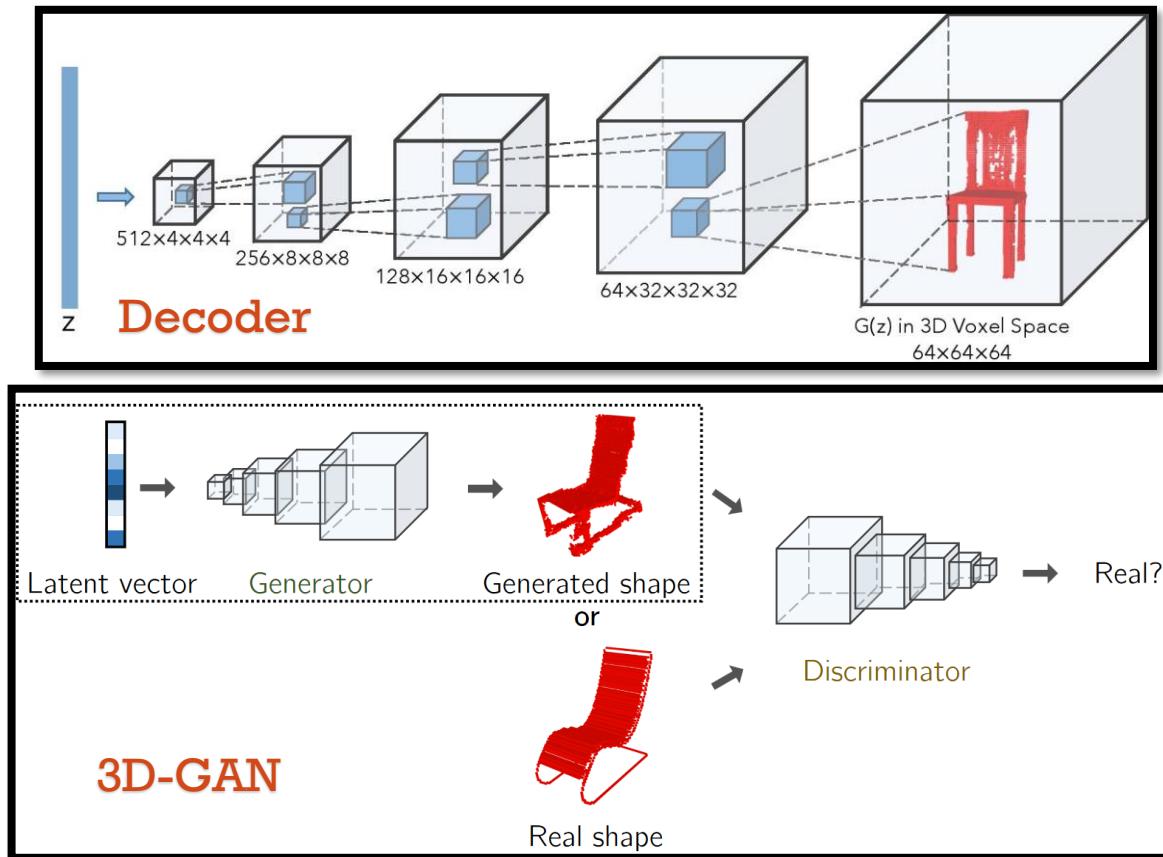
- 表达：空间规则划分 N^3 个体素，体素信号 (0: 空; 1: 实)



- 已有工作: 3D-GAN [Wu et al. 2016] (VAE&GAN), 3D-R2N2 [Choy et al. 2016] (RNN), 3D-IWGAN[Smith et al. 2017](VAE&WGAN),....

规则体素 (REGULAR GRID, VOXELS)

- 代表工作: 3D-GAN [Wu et al. 2016]



规则体素 (REGULAR GRID, VOXELS)

- 代表工作：3D-GAN [Wu et al. 2016]



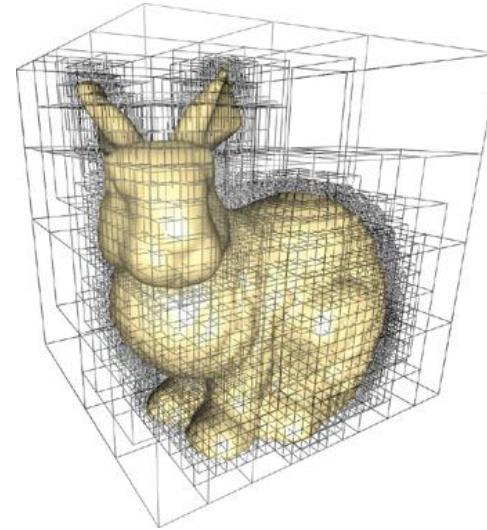
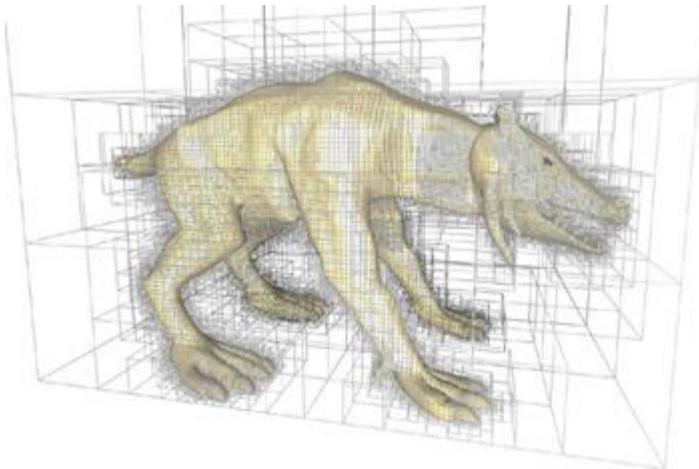
Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling, NIPS 2016

规则体素 (REGULAR GRID, VOXELS)

- 优点：2D 卷积直接推广到3D，图像域的工作容易借鉴。
- 缺点： $O(N^3)$ 空间复杂度 -- 内存开销大，训练时间长。常见网络受限于 $32^3, 64^3$ 分辨率下。生成物体精细度不够。

稀疏体素 (SPARSE VOXEL)

表达：用少量的体素表达三维物体。体素标记：外部、内部、边界。

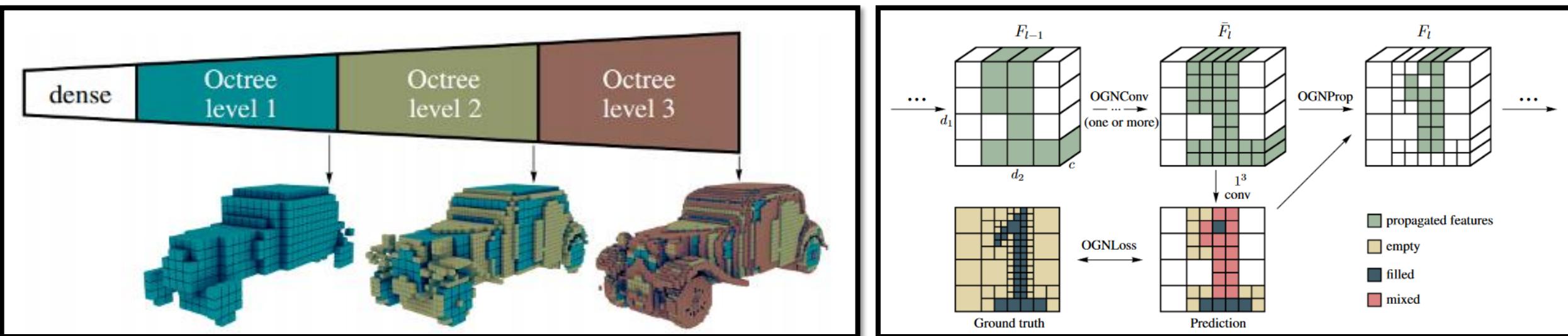


已有工作：OctGen [Tatarchenko et al. 2017] (八叉树); OctFusion [Riegler et al. 2017] (八叉树); HSP [Hane et al. 2017] (八叉树); AOCNN [Wang et al. 2018] (自适应八叉树)。

稀疏体素 (SPARSE VOXEL)

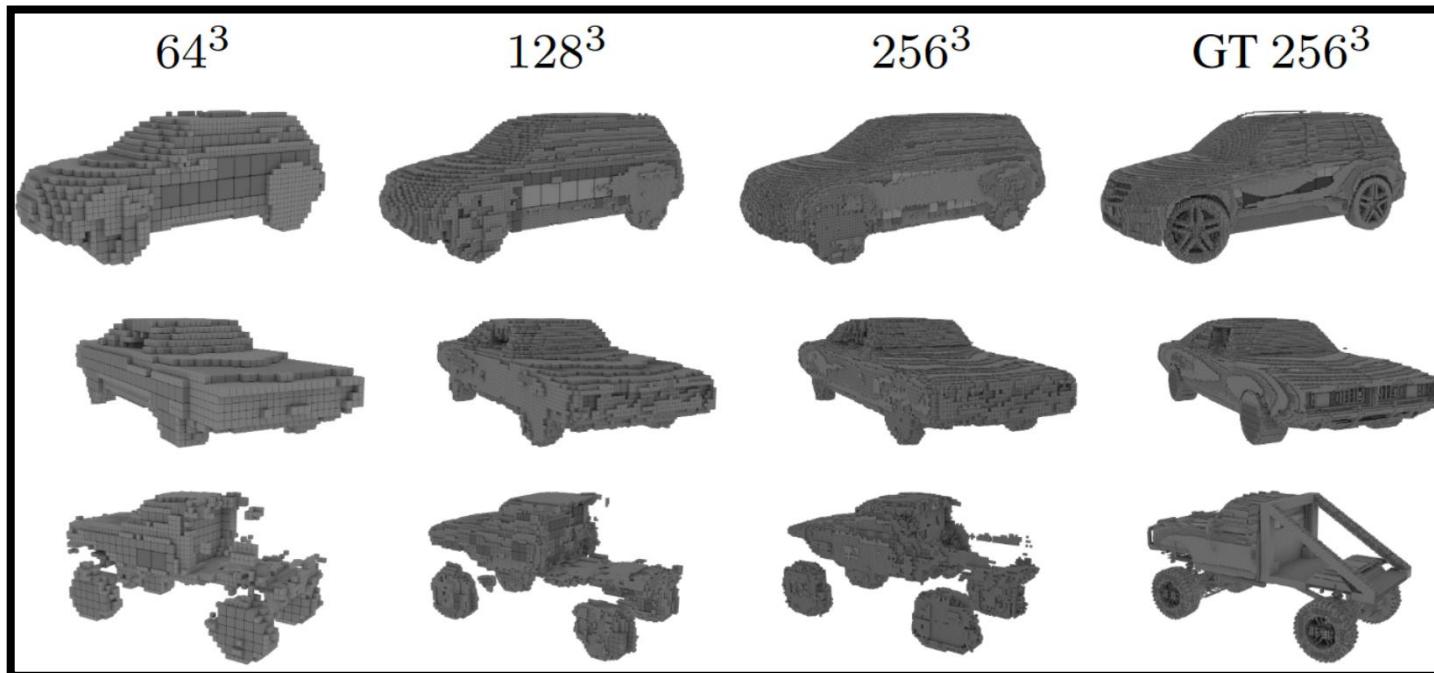
代表工作：OctGen [Tatarchenko et al. 2017]

网络：从低分辨率结点开始，预测体素标记：如果预测为物体表面，继续剖分该体素。



稀疏体素 (SPARSE VOXEL)

代表工作: OctGen [Tatarchenko et al. 2017]



Category	R2N2 [6]	OGN	Dense
Plane	0.513	0.587	0.570
Bench	0.421	0.481	0.481
Cabinet	0.716	0.729	0.747
Car	0.798	0.816	0.828
Chair	0.466	0.483	0.481
Monitor	0.468	0.502	0.509
Lamp	0.381	0.398	0.371
Speaker	0.662	0.637	0.650
Firearm	0.544	0.593	0.576
Couch	0.628	0.646	0.668
Table	0.513	0.536	0.545
Cellphone	0.661	0.702	0.698
Watercraft	0.513	0.632	0.550
Mean	0.560	0.596	0.590

单视图三维重建 32^3 , IOU 比较图

Octree generating networks: efficient convolutional architectures for high-resolution 3D outputs, ICCV 2017.

稀疏体素 (REGULAR GRID, VOXELS)

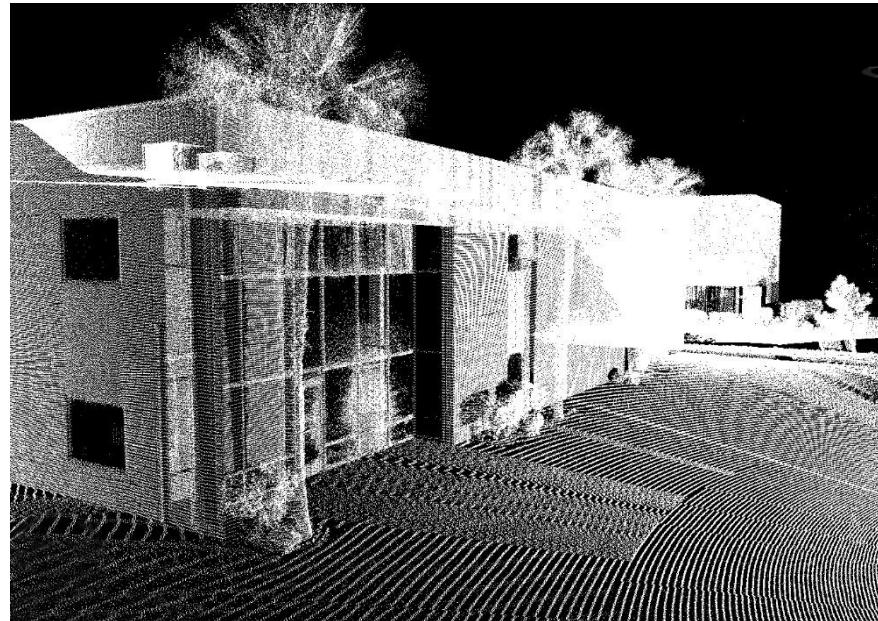
- 优点: $O(N^2)$ 复杂度, 高效省内存。可适用 $128^3, 256^3, 512^3$ 高分辨率。
可恢复更多细节。



Hierarchical Surface Prediction for 3D Object Reconstruction, 3DV 2017.

点云 (POINT CLOUD)

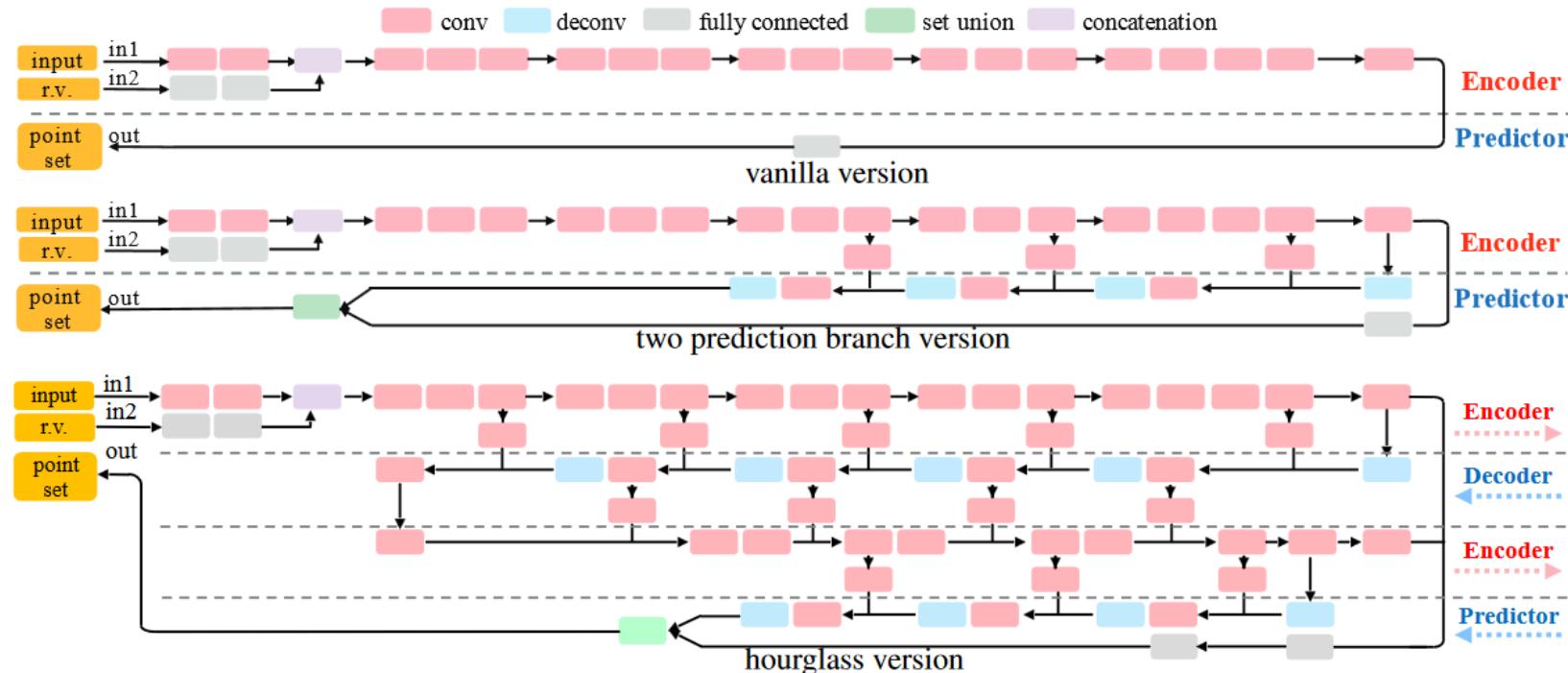
- 表达：离散点集表示三维形体。



- 已有工作：PSG [Fan et al. 2017], [Achlioptas et al. 2017]

点云 (POINT CLOUD)

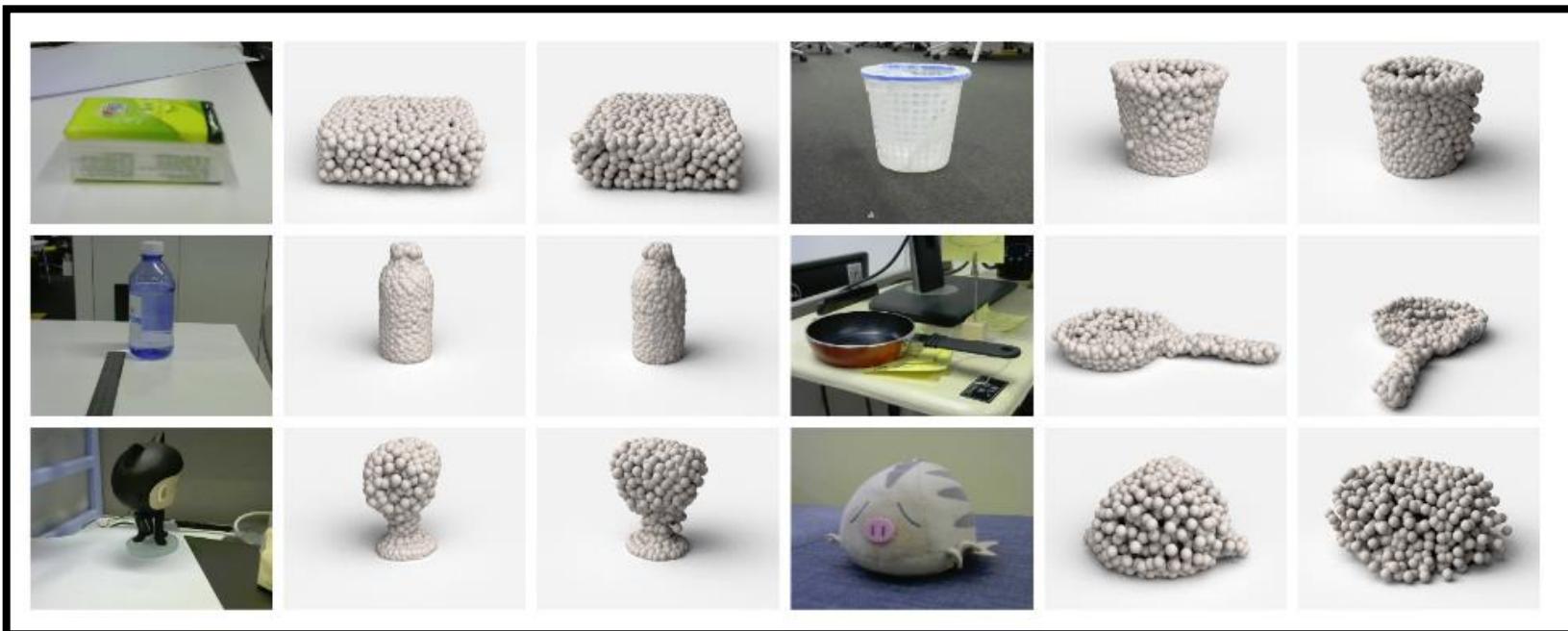
- 代表工作: PSG [Fan et al. 2017]
- 网络: 给定点数, 预测每个点的空间位置。



A point set generation network for 3D object reconstruction from a single image, CVPR 2017

点云 (POINT CLOUD)

- 代表工作: PSG [Fan et al. 2017]



category	Ours	3D-R2N2		
	1 view	1 view	3 views	5 views
plane	0.601	0.513	0.549	0.561
bench	0.550	0.421	0.502	0.527
cabinet	0.771	0.716	0.763	0.772
car	0.831	0.798	0.829	0.836
chair	0.544	0.466	0.533	0.550
monitor	0.552	0.468	0.545	0.565
lamp	0.462	0.381	0.415	0.421
speaker	0.737	0.662	0.708	0.717
firearm	0.604	0.544	0.593	0.600
couch	0.708	0.628	0.690	0.706
table	0.606	0.513	0.564	0.580
cellphone	0.749	0.661	0.732	0.754
watercraft	0.611	0.513	0.596	0.610
mean	0.640	0.560	0.617	0.631

A point set generation network for 3D object reconstruction from a single image, CVPR 2017

点云 (POINT CLOUD)

- 代表工作: [Achlioptas et al. 2017] (Latent GAN + GMM)



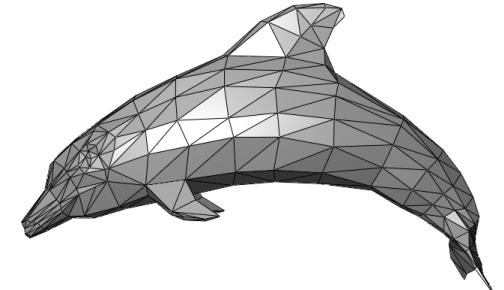
Learning Representations and Generative Models for 3D Point Clouds, ICLR 2018.

点云 (POINT CLOUD)

- 优点：表达灵活、也适用稀疏点集
- 缺点：缺乏点之间关联性、需后处理转换成连续表达

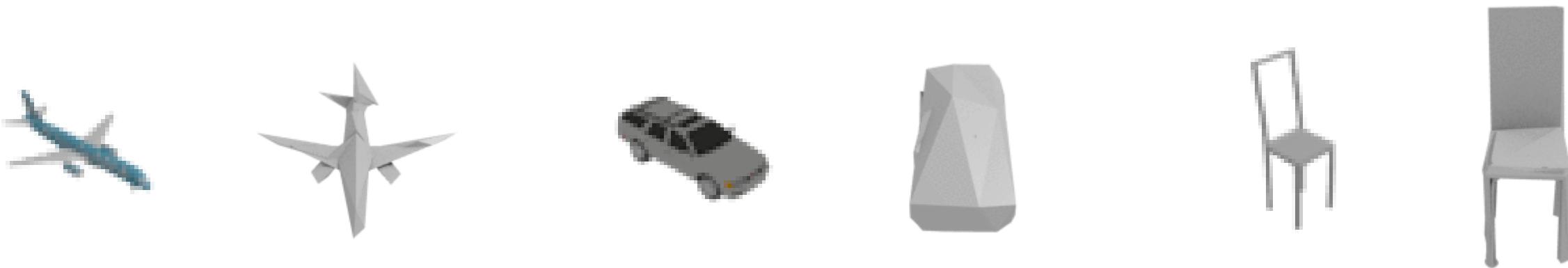
网格 (MESH)

- 表达: 一片或多片三角形网格表达三维形体
- 网络: 预测template网格的变形映射, 以构成新的形体。
- 已有工作
 - 单片网格:
 - SurfNet: Generating 3D shape surfaces using deep residual networks, CVPR 2017.
 - Pixel2Mesh: Generating 3D Mesh Models, from Single RGB Images, CVPR 2018.
 - Neural 3D Mesh Renderer, CVPR 2018.
 - FoldingNet: Point cloud auto-encoder via deep grid deformation, CVPR 2018
 - 多片网格:
 - AtlasNet: A Papier-Mache approach to learning 3D surface generation, CVPR 2018



网格 (MESH)

- 代表工作： Neural 3D Mesh Renderer [Kato et al. 2017]
- 网络： 预测给定球面网格顶点的偏移量，偏移量没有真实值做指导。



网格 (MESH)

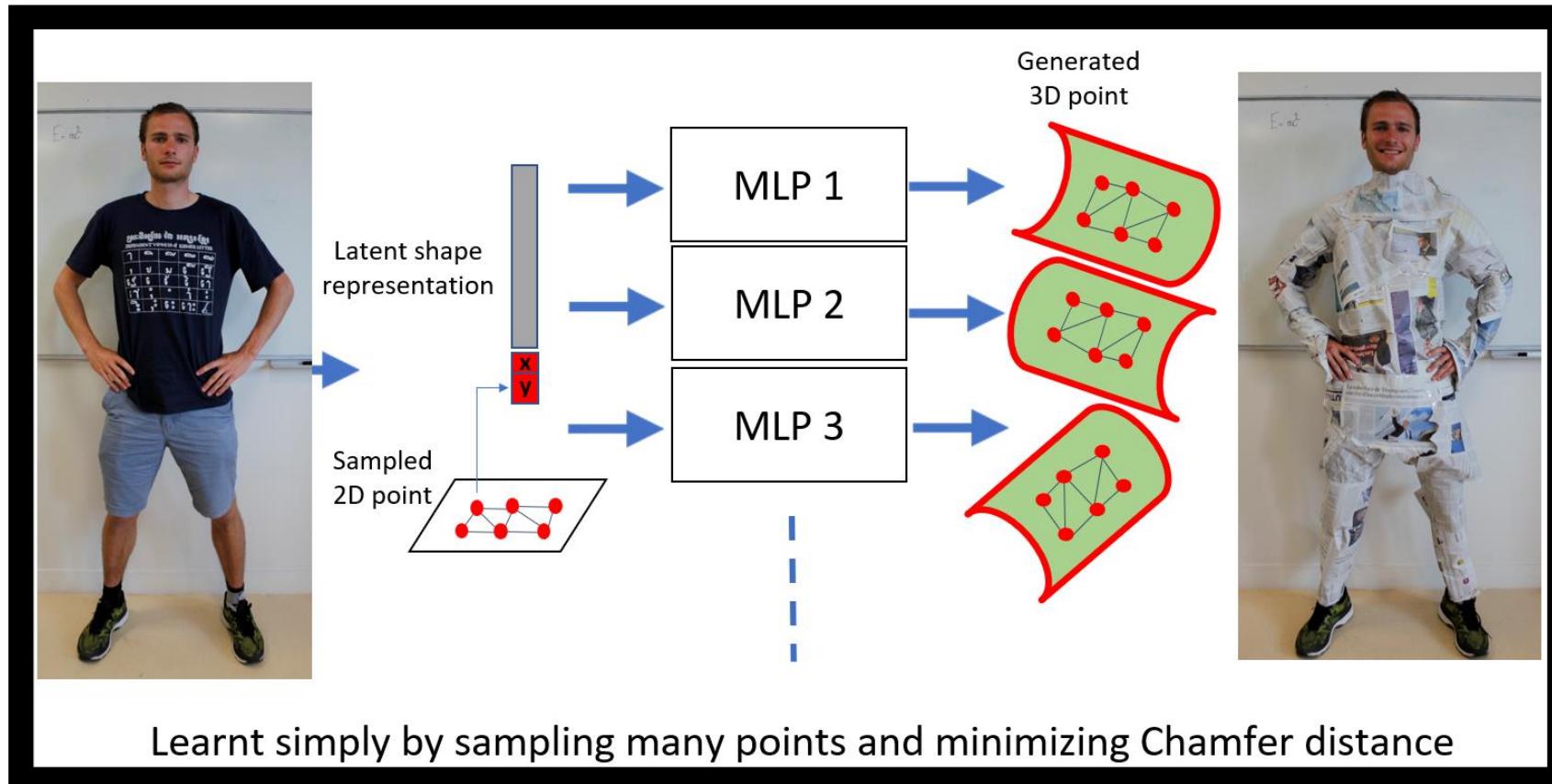
- 代表工作: Neural 3D Mesh Renderer [Kato et al. 2017]
- 网络: 预测给定球面网格顶点的偏移量, 偏移量没有真实值做指导。

	airplane	bench	dresser	car	chair	display	lamp
Retrieval [36]	0.5564	0.4875	0.5713	0.6519	0.3512	0.3958	0.2905
Voxel-based [36]	0.5556	0.4924	0.6823	0.7123	0.4494	0.5395	0.4223
Mesh-based (ours)	0.6172	0.4998	0.7143	0.7095	0.4990	0.5831	0.4126
	loudspeaker	rifle	sofa	table	telephone	vessel	mean
Retrieval [36]	0.4600	0.5133	0.5314	0.3097	0.6696	0.4078	0.4766
Voxel-based [36]	0.5868	0.5987	0.6221	0.4938	0.7504	0.5507	0.5736
Mesh-based (ours)	0.6536	0.6322	0.6735	0.4829	0.7777	0.5645	0.6016

重建质量对比 (IOU metric)

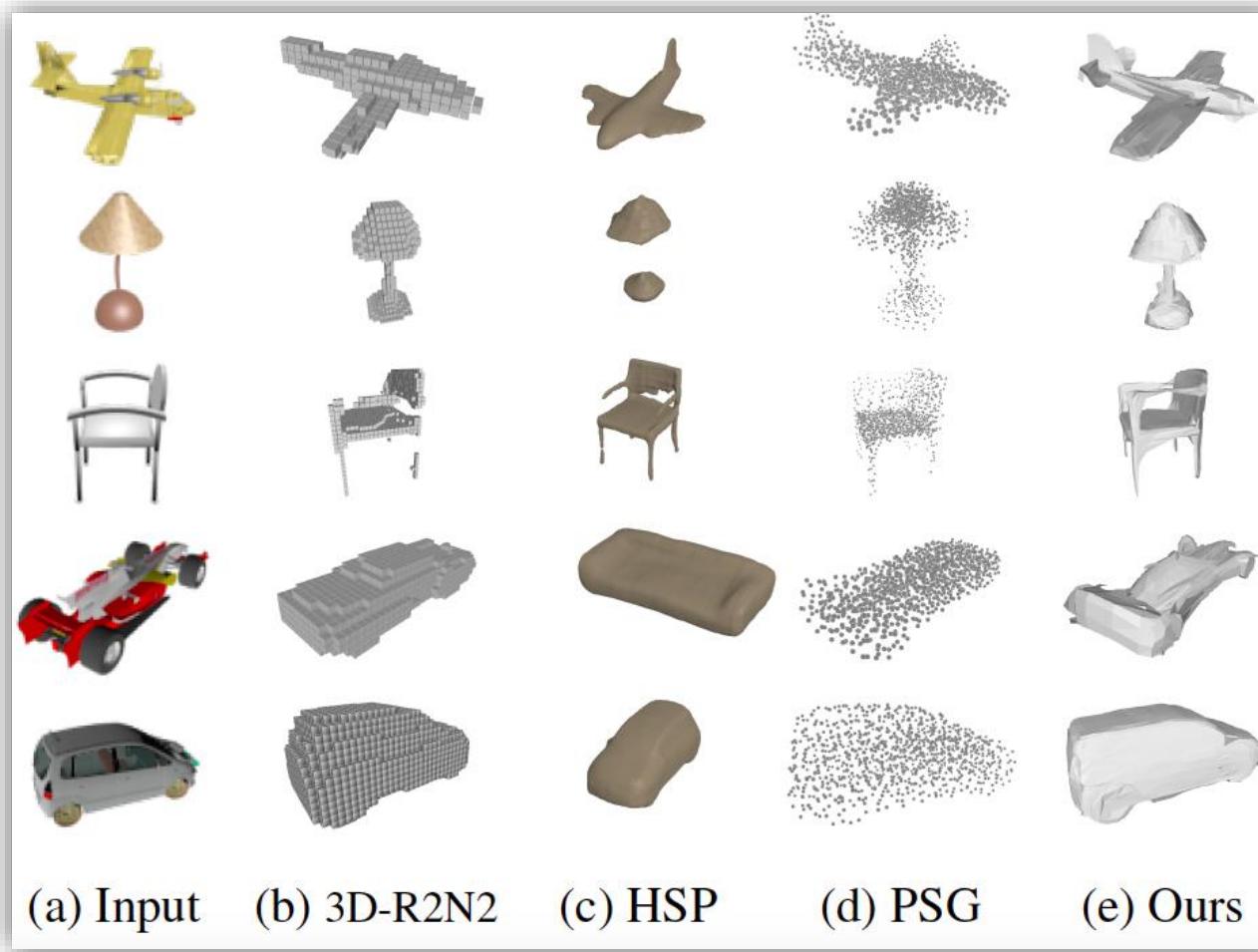
网格 (MESH)

- 代表工作：AtlasNet[Groueix et al. 2018]
- 网络：给定指定片数的二维网格，预测每片到三维空间的映射



网格 (MESH)

- 代表工作：AtlasNet[Groueix et al. 2018]

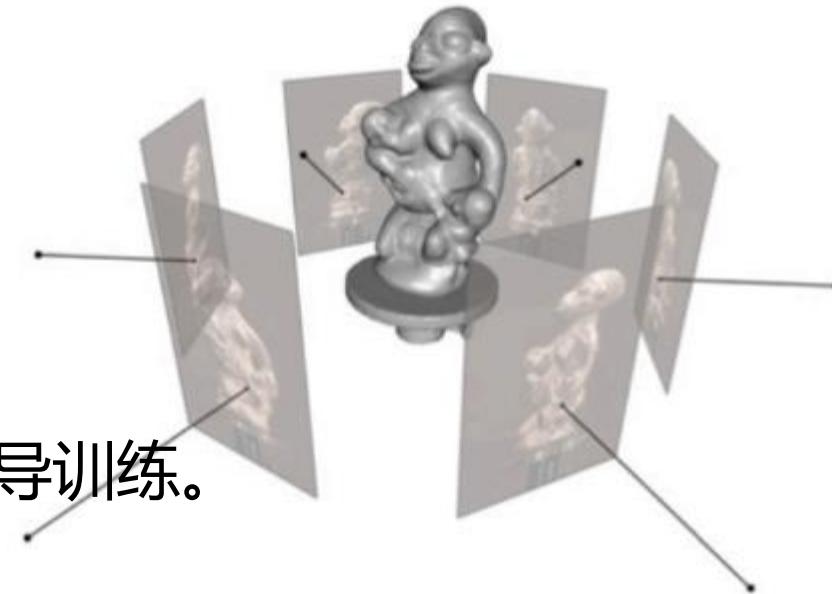


网格 (MESH)

- **优点:** 输出质量相对高、方法简单
- **缺点:** 映射扭曲程度没有控制。单片受限于给定拓扑；由多片构造完整曲面还存在困难。

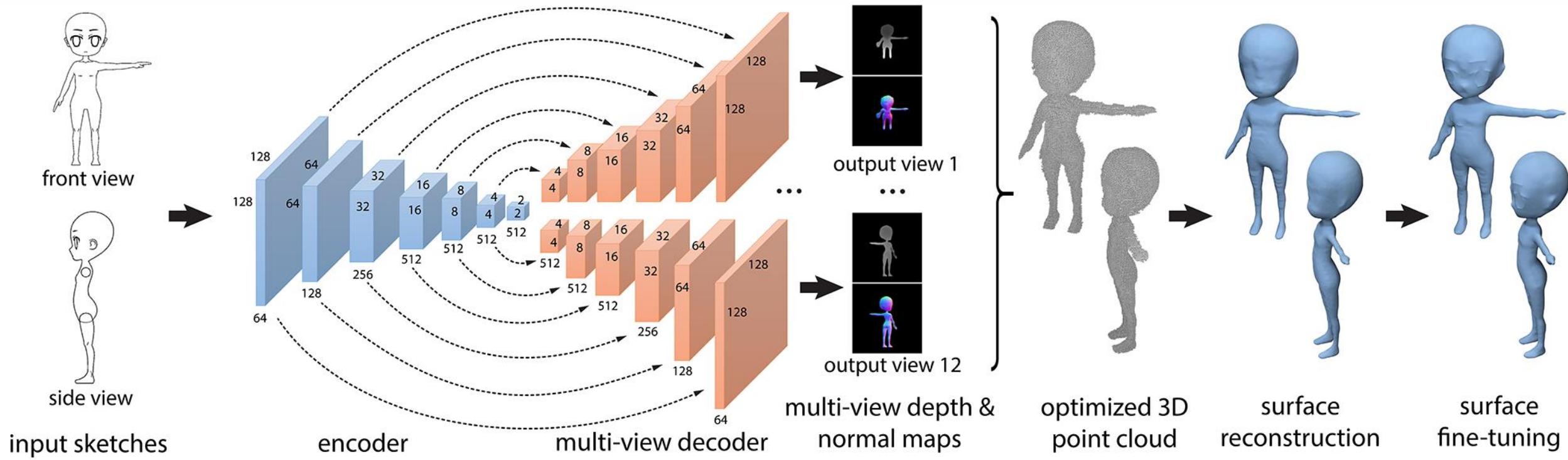
多视图 (MULTIVIEW)

- 表达：多个(深度)视图形成三维形体。
- 网络：预测若干个指定视角的视图，用投影误差指导训练。
- 已有工作：
 - Lun et al., 3D Shape Reconstruction from Sketches via Multi-view Convolutional Networks, 3DV 2017.
 - Lin et al. Learning Efficient Point Cloud Generation for Dense 3D Object Reconstruction, AAAI 2018.



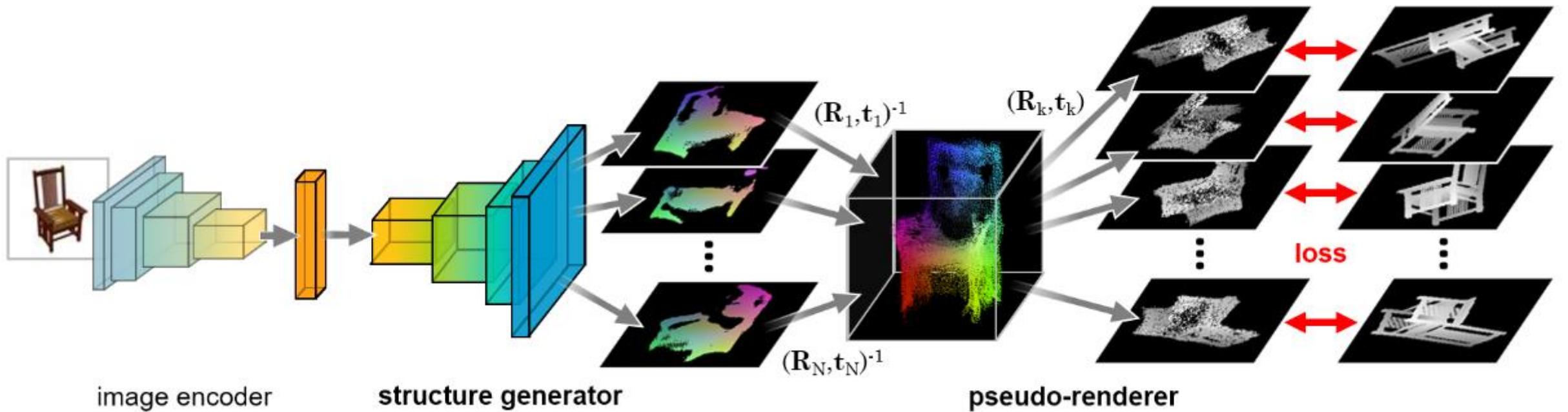
多视图 (MULTIVIEW)

- 表达：多个(深度)视图形成三维形体。



多视图 (MULTIVIEW)

- 表达：多个(深度)视图形成三维形体。



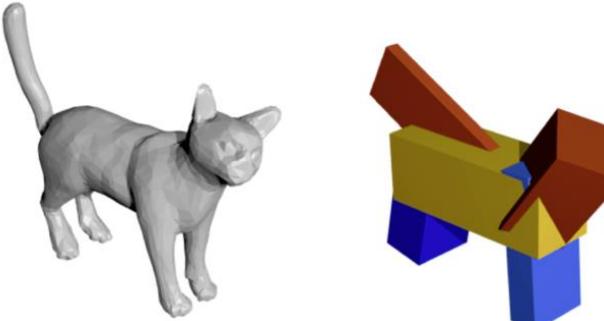
Learning Efficient Point Cloud Generation for Dense 3D Object Reconstruction

型元(PRIMITIVE)与结 构(STRUCTURE)表达



体元集合

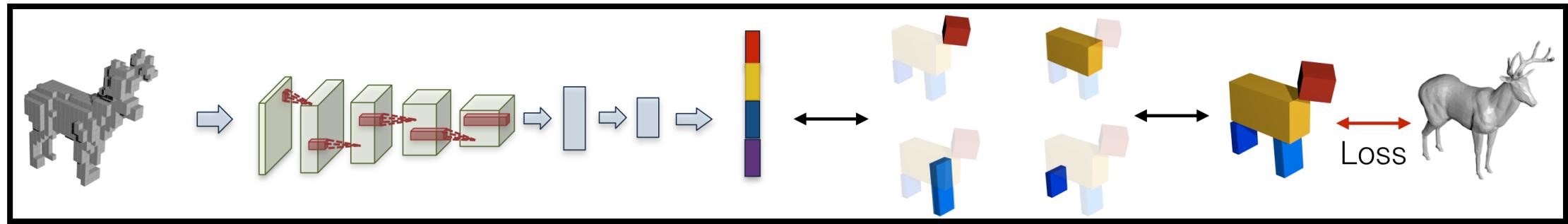
- 表达：用若干个简单的实体如长方体(cuboid)来近似物体。



- 已有工作：
 - Learning shape abstractions by assembling volumetric primitives, CVPR 2017.
 - 3D-PRNN: Generating shape primitives with recurrent neural networks, ICCV 2017.

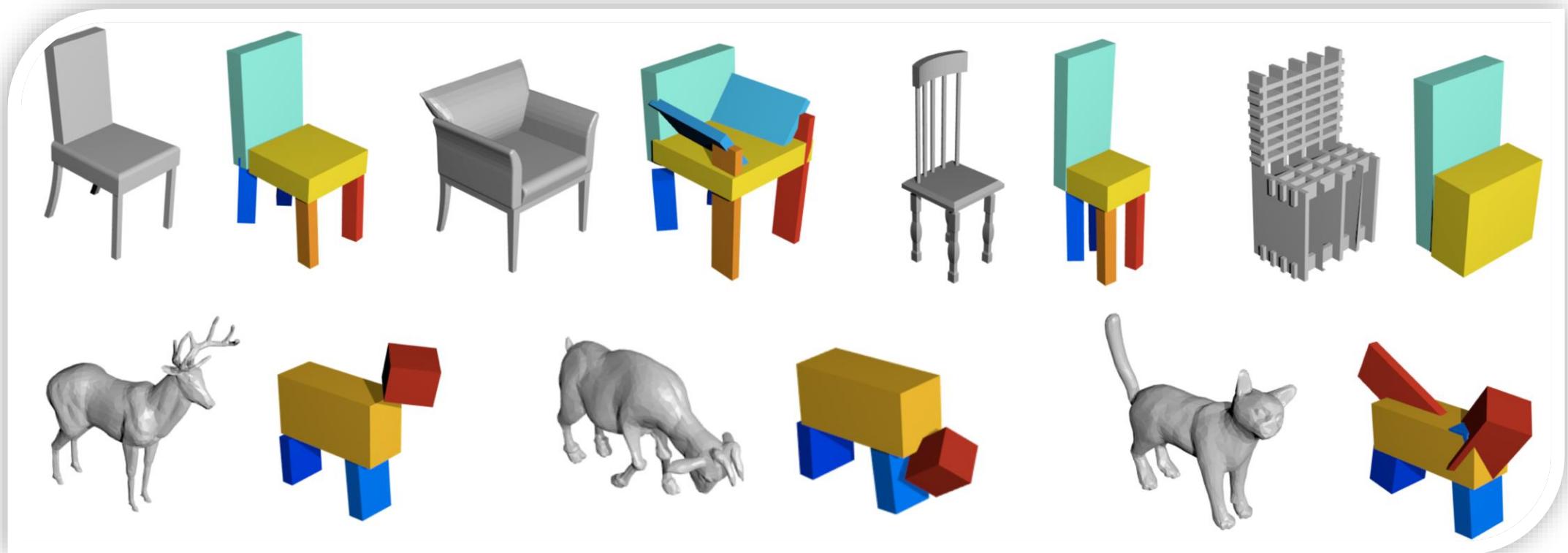
体元集合

- 代表工作: Learning shape abstractions by assembling volumetric primitives
- 网络: 预测指定数目长方体的大小、朝向、位置。 (无监督学习)



体元集合

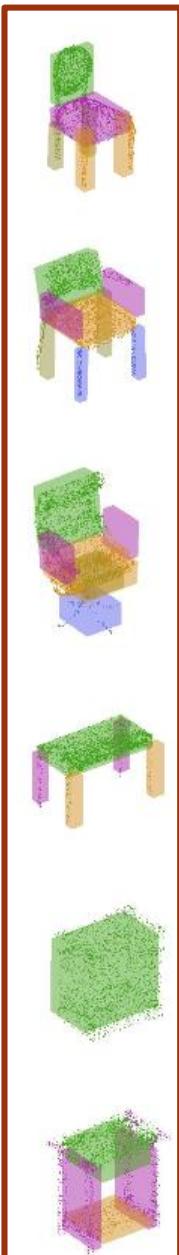
- 代表工作: Learning shape abstractions by assembling volumetric primitives
- 网络: 预测指定数目长方体的大小、朝向、位置。 (无监督学习)



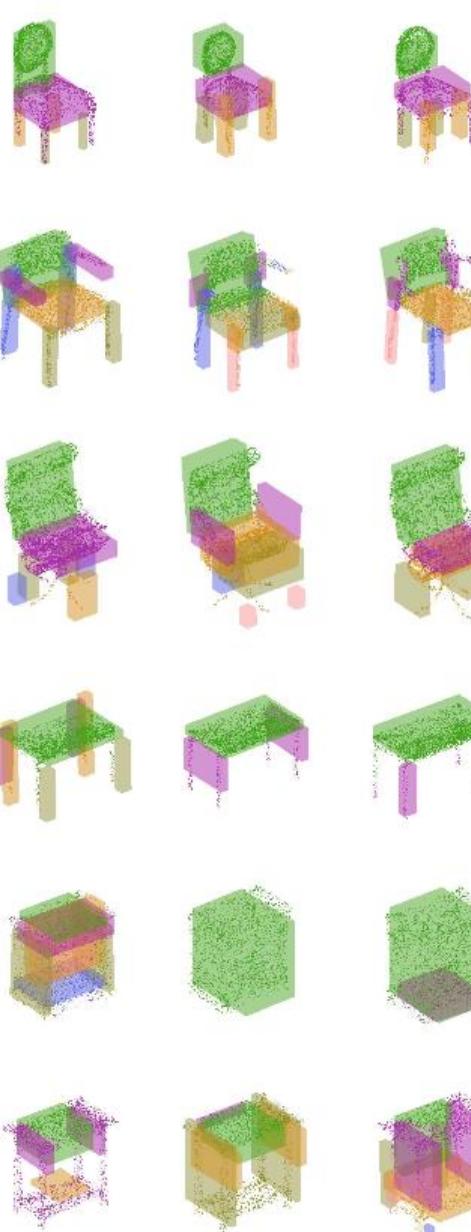
Input Depth Map



Most probable



Random sampling



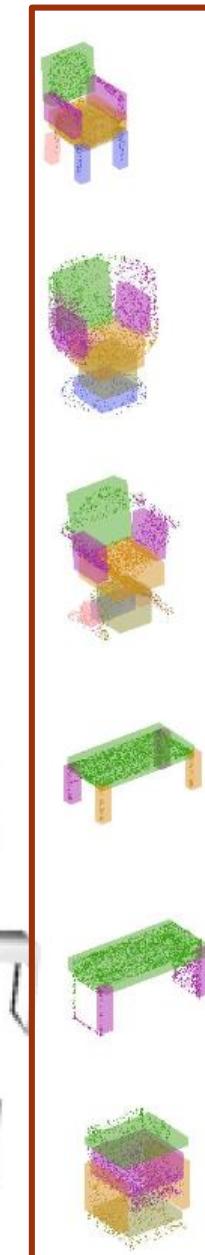
GT



Input Depth Map



Most probable



Random sampling



GT

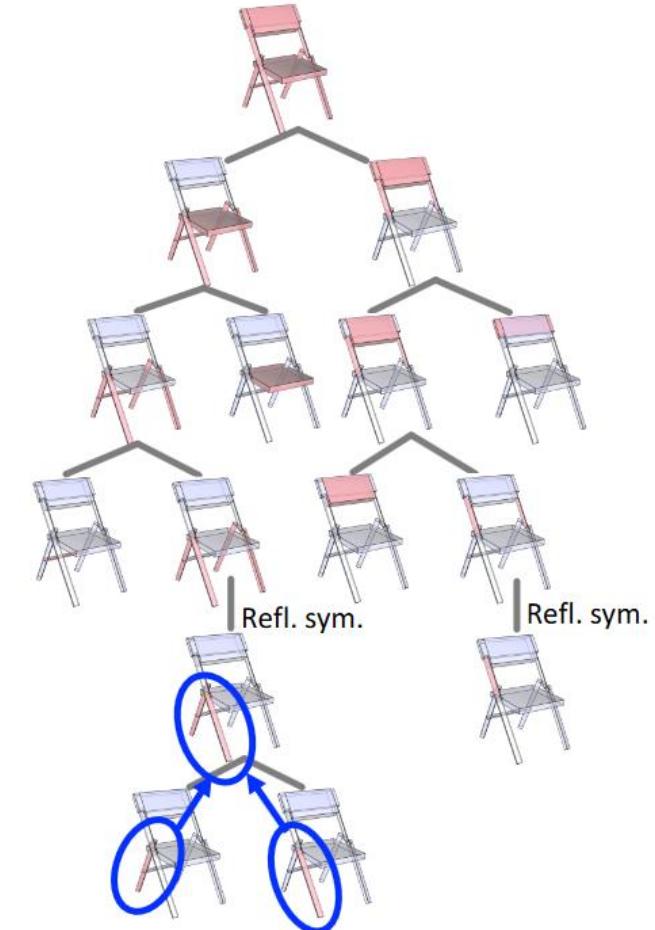


体元集合

- 优点：利用体元抽象出物体的共同结构，适合构建物体之间对应、描述物体相似性等。
- 缺点：体元之间关系没有定义；多类型体元的尚待扩展。

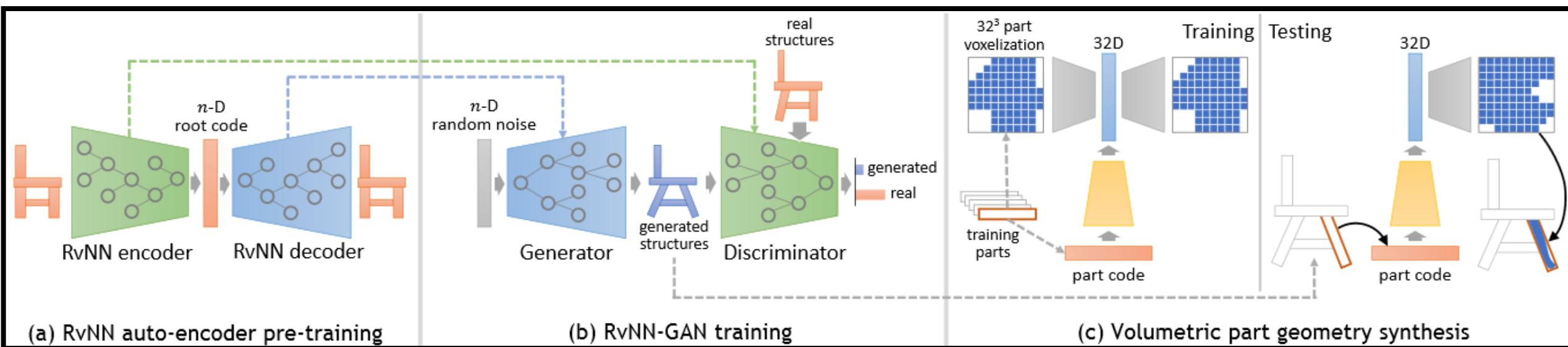
层次生成结构

- 表达：将三维内容看做一个层次结构（如树状结构），叶节点是内容的一部分，如物体的一个组件/场景中的一个物体等。
- 网络：递归地生成层次结构并预测节点上的定义的关系和叶节点中的内容。
- 已有工作：
 - GRASS: Generative recursive autoencoders for shape structures, SIGGRAPH 2017
 - Im2Struct: Recovering 3D Shape Structure from a Single RGB Image, CVPR 2018
 - GRAINS: Generative Recursive Autoencoders for INdoor Scenes, arXiv 2018.
 - The shape variational autoencoder: A deep generative model of part-segmented 3D objects, Computer Graphics Forum (SGP), 2017



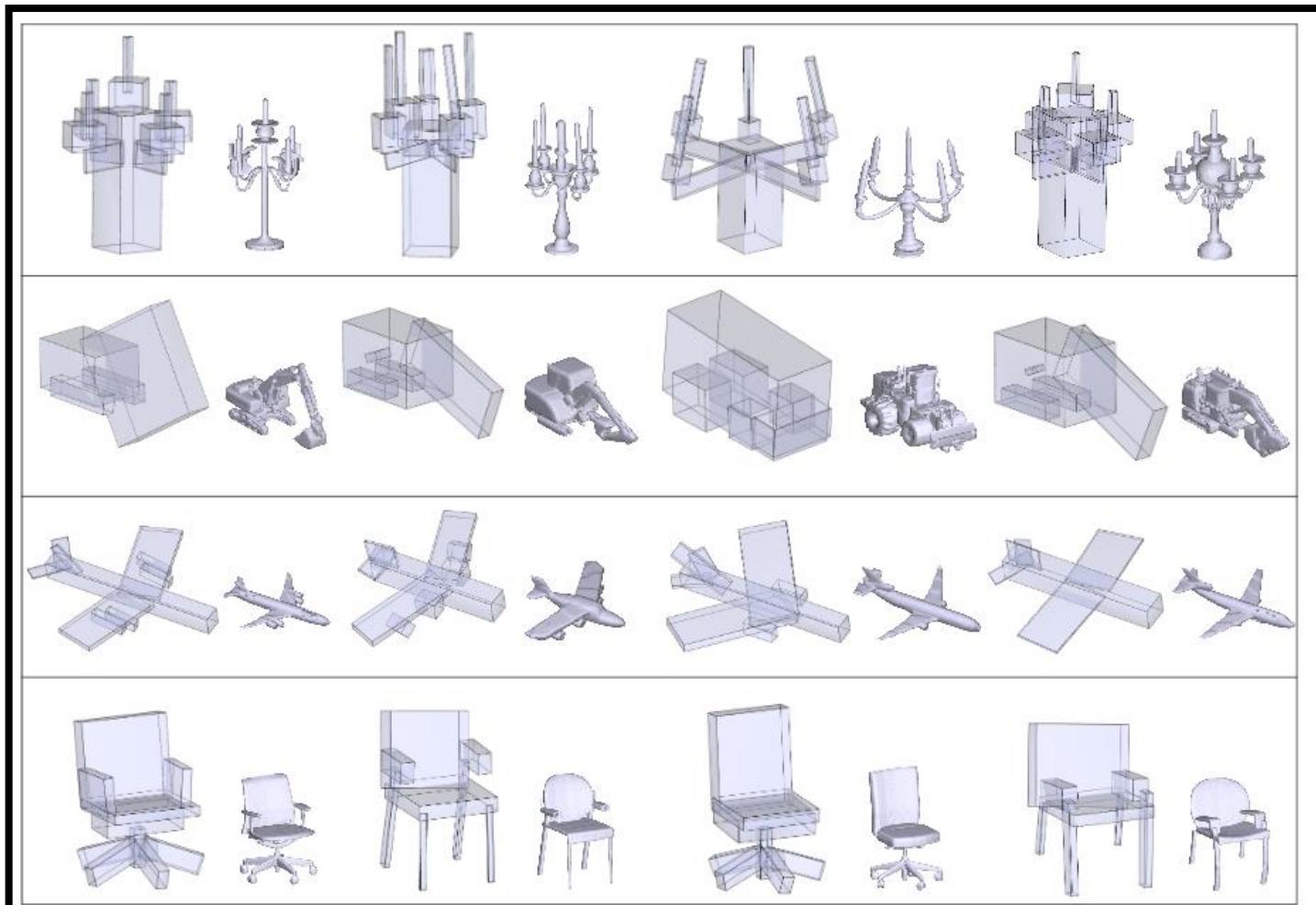
树型生成结构

- 代表工作： GRASS: Generative recursive autoencoders for shape structures, SIGGRAPH 2017



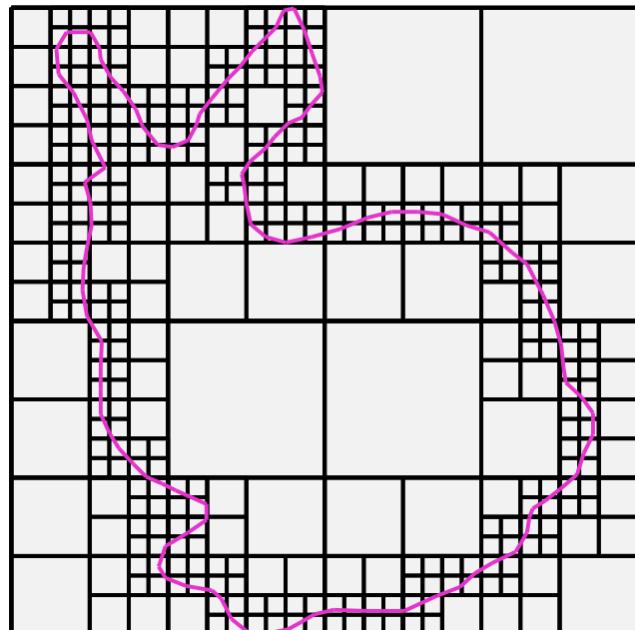
树型生成结构

- 代表工作： GRASS: Generative recursive autoencoders for shape structures, SIGGRAPH 2017

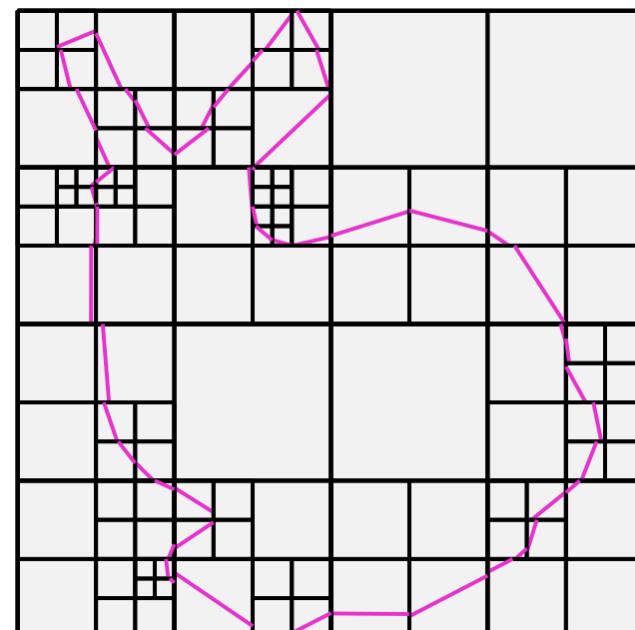


基于型元的自适应八叉树

- 表达：自适应八叉树结构，叶节点储存型元（如平面片）。各层上所有叶节点内的平面片构成三维形体。
- 工作： Adaptive O-CNN: A Patch-based Deep Representation of 3D Shapes, conditionally accepted to SiggraphAsia 2018.



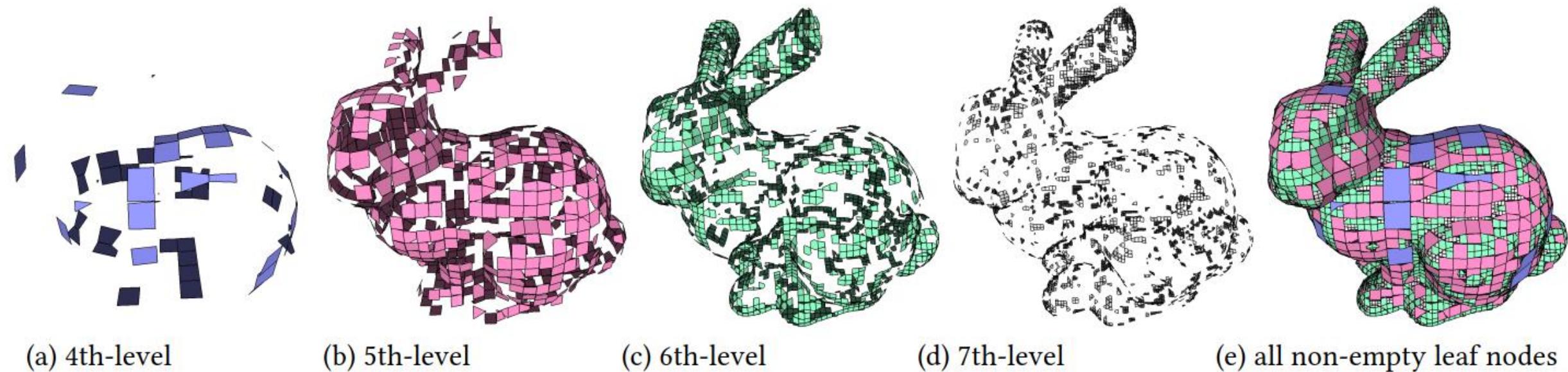
四叉树



基于线段的自适应四叉树

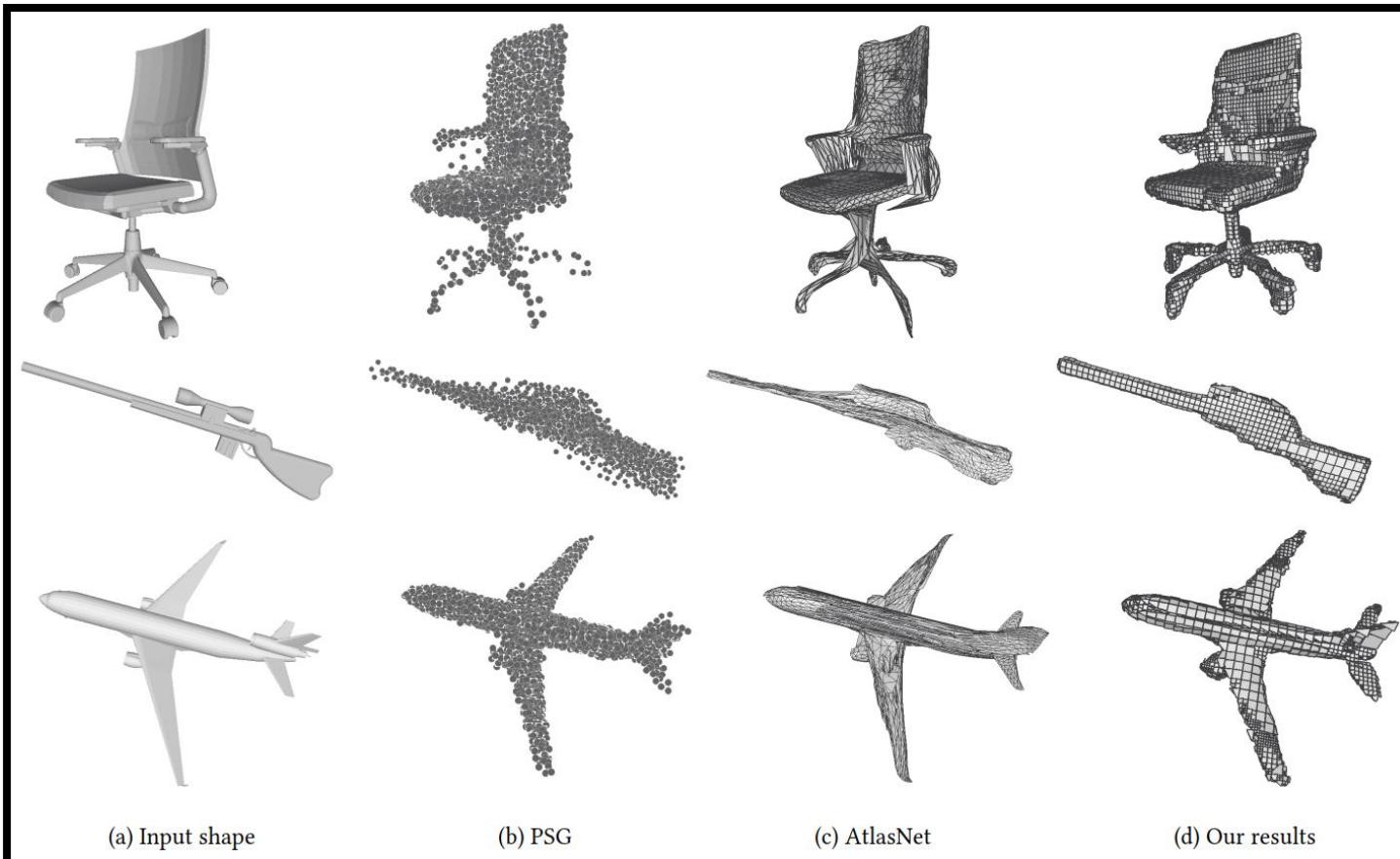
基于型元的自适应八叉树

- 表达：自适应八叉树结构，叶节点储存型元（如平面片）。各层上所有叶节点内的平面片构成三维形体。
- 工作： Adaptive O-CNN: A Patch-based Deep Representation of 3D Shapes, conditionally accepted to SiggraphAsia 2018.



基于型元的自适应八叉树

- 网络：预测八叉树结构，每个节点上判断三种状态之一：(1) 空（不含三维形体表面）；
(2) 所含形状表面可以被型元**很好的**近似；(3) 所含形状表面不能被型元**很好的**近似。对于(3)，八叉树继续剖分；对于(2)，回归型元参数。



基于型元的自适应八叉树

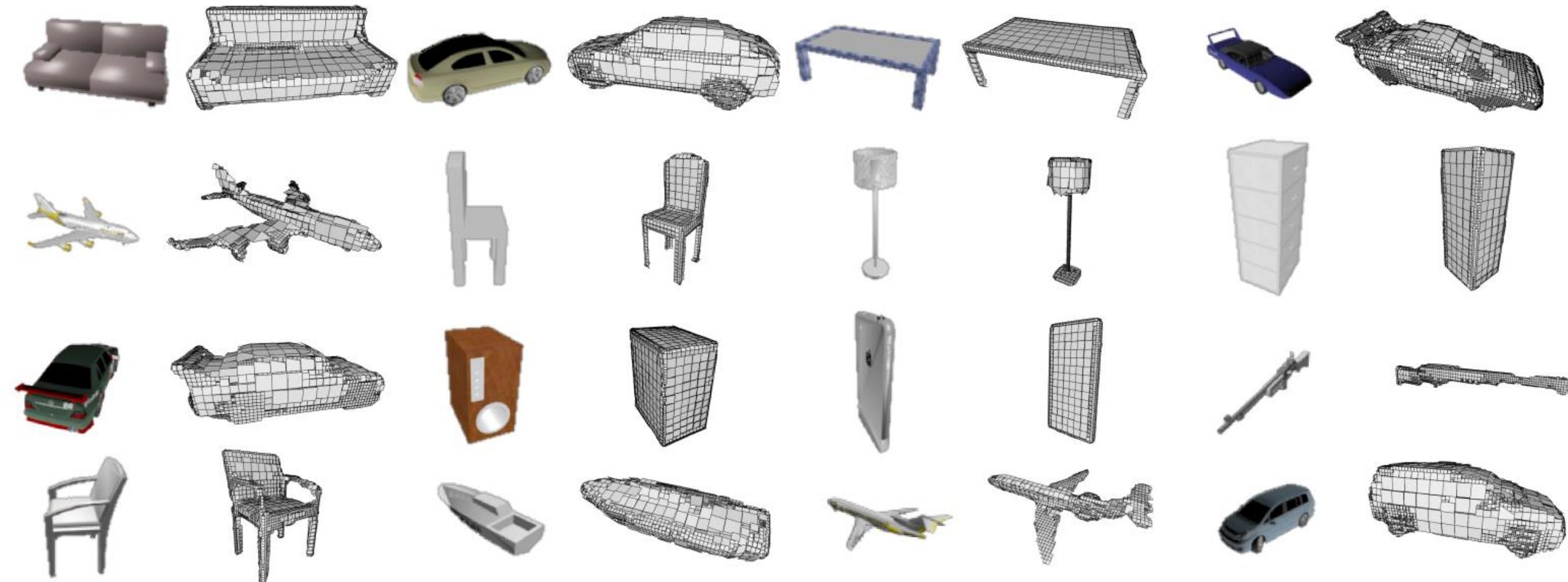
- 图像到3D任务上比较 Chamfer distance metric

Method	mean	pla.	ben.	cab.	car	cha.	mon.	lam.	spe.	fir.	cou.	tab.	cel.	wat.
PSG	6.41	3.36	4.31	8.51	8.63	6.35	6.47	7.66	15.9	1.58	6.92	3.93	3.76	5.94
AtlasNet(25)	5.11	2.54	3.91	5.39	4.18	6.77	6.71	7.24	8.18	1.63	6.76	4.35	3.91	4.91
AO-CNN	4.63	2.45	2.69	2.67	1.80	6.13	6.27	10.92	9.43	1.68	4.42	4.19	2.51	5.04

- 内存和计算时间开销:
ModelNet40 分类任务
batch size = 32

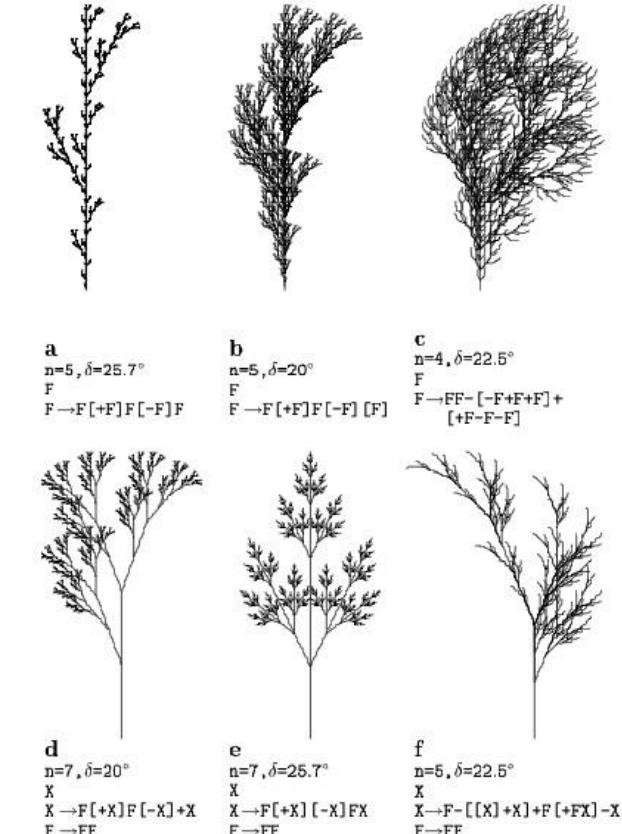
	Method	32^3	64^3	128^3	256^3
Memory	Voxel	0.71 GB	3.7 GB	–	–
	O-CNN	0.58 GB	1.1 GB	2.7 GB	6.4 GB
	AO-CNN	0.51 GB	0.95 GB	1.5 GB	1.7 GB
Time	Voxel	425 ms	1648 ms	–	–
	O-CNN	41 ms	117 ms	334 ms	1393 ms
	AO-CNN	34 ms	63 ms	112 ms	307 ms
Accuracy	O-CNN	90.4%	90.6%	90.1%	90.2%
	AO-CNN	90.5%	90.4%	90.0%	90.2%

基于型元的自适应八叉树



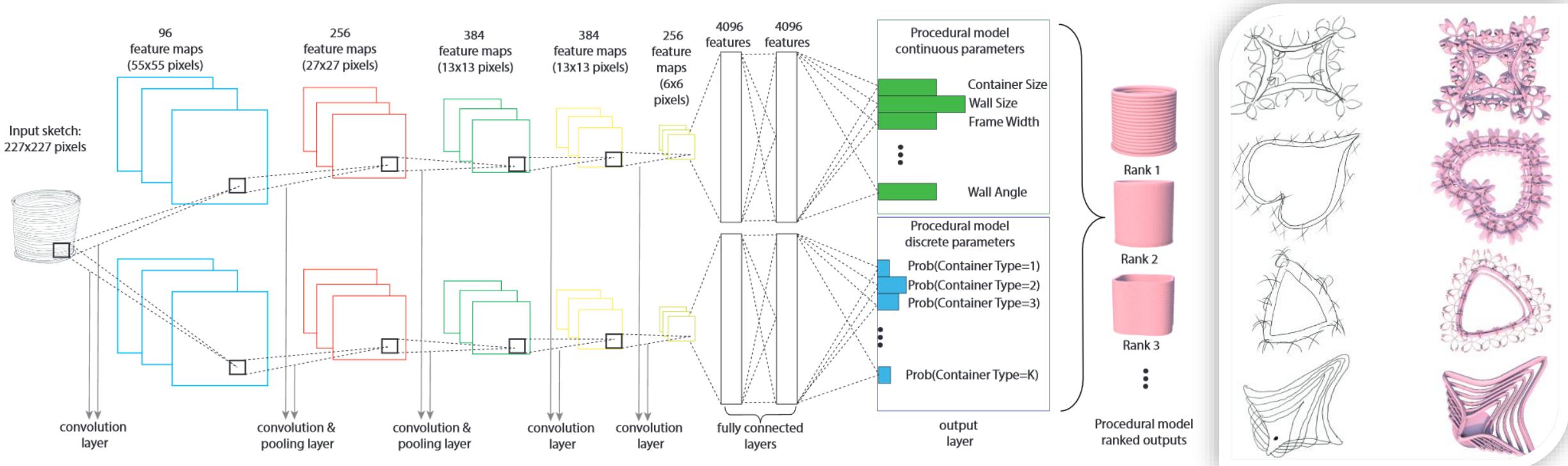
程序化建模 (PROCEDURAL MODELING)

- 表达：一系列规则定义模型，包括L-system, 分型, GML, CSG等。
- 网络：学习规则的参数
- 已有工作：
 - Shape Synthesis from Sketches via Procedural Models and Convolutional Networks, TVCG 2017.
 - CSGNet: Neural shape parser for constructive solid geometry, CVPR 2018.



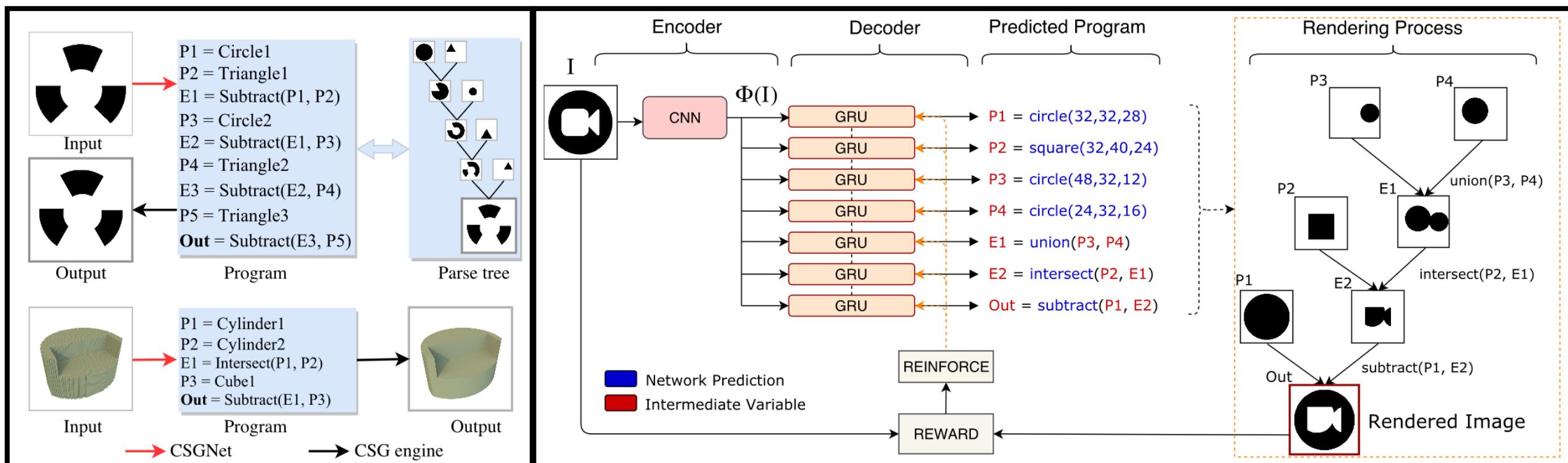
程序化建模 (PROCEDURAL MODELING)

- 代表工作: Shape Synthesis from Sketches via Procedural Models and Convolutional Networks, TVCG 2017.



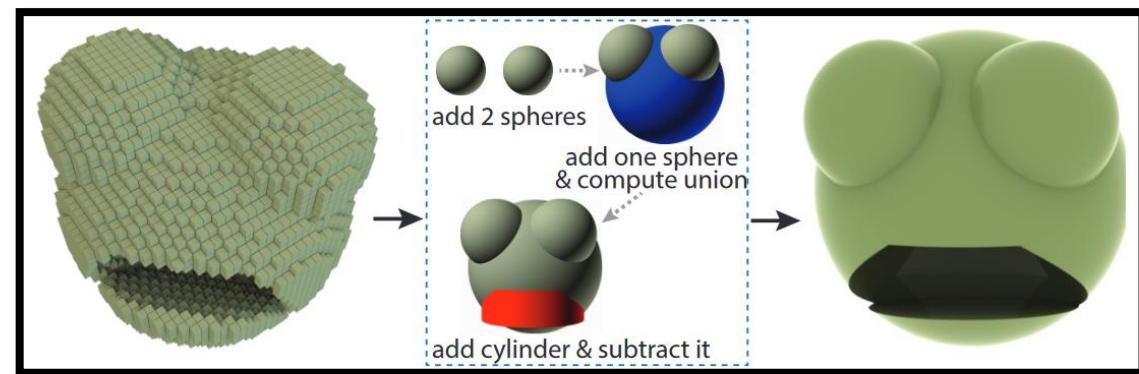
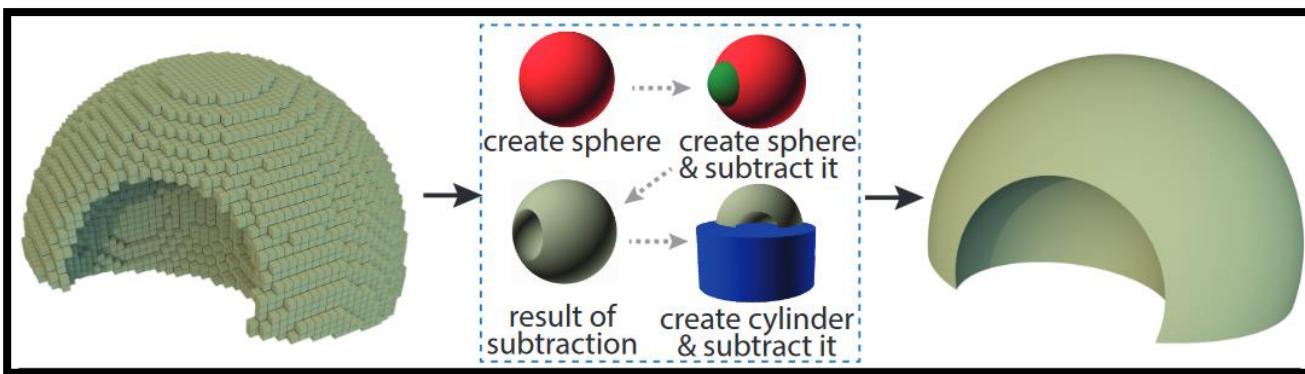
CSG 树

- 代表工作： CSGNet: Neural shape parser for constructive solid geometry, CVPR 2018.
- 网络：预测实体(类型和个数)和操作。



CSG 树

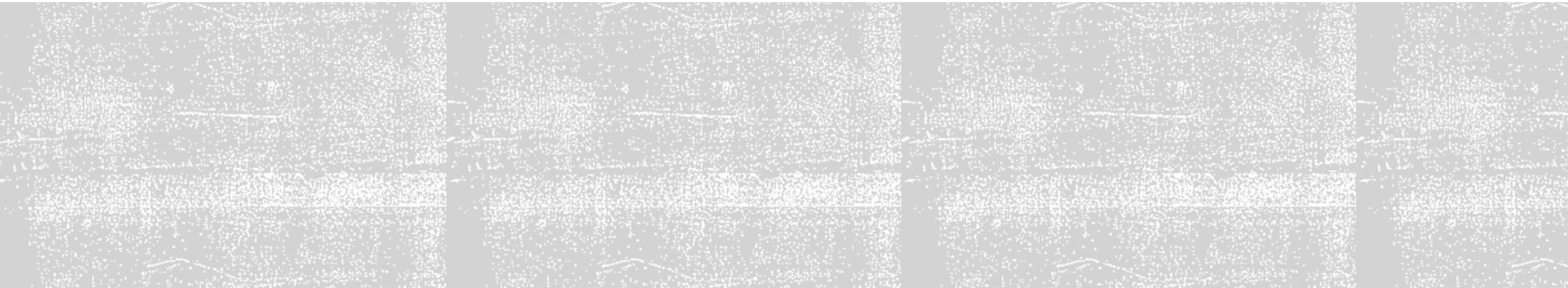
- 代表工作： CSGNet: Neural shape parser for constructive solid geometry, CVPR 2018.
- 网络：预测实体(类型和个数)和操作。



程序化建模 (PROCEDURAL MODELING)

- 优点：用简单规则构建复杂模型、有建模步骤适合编辑与理解。
- 缺点：训练复杂，需要强化学习技术，目前研究处于起步阶段。

应用

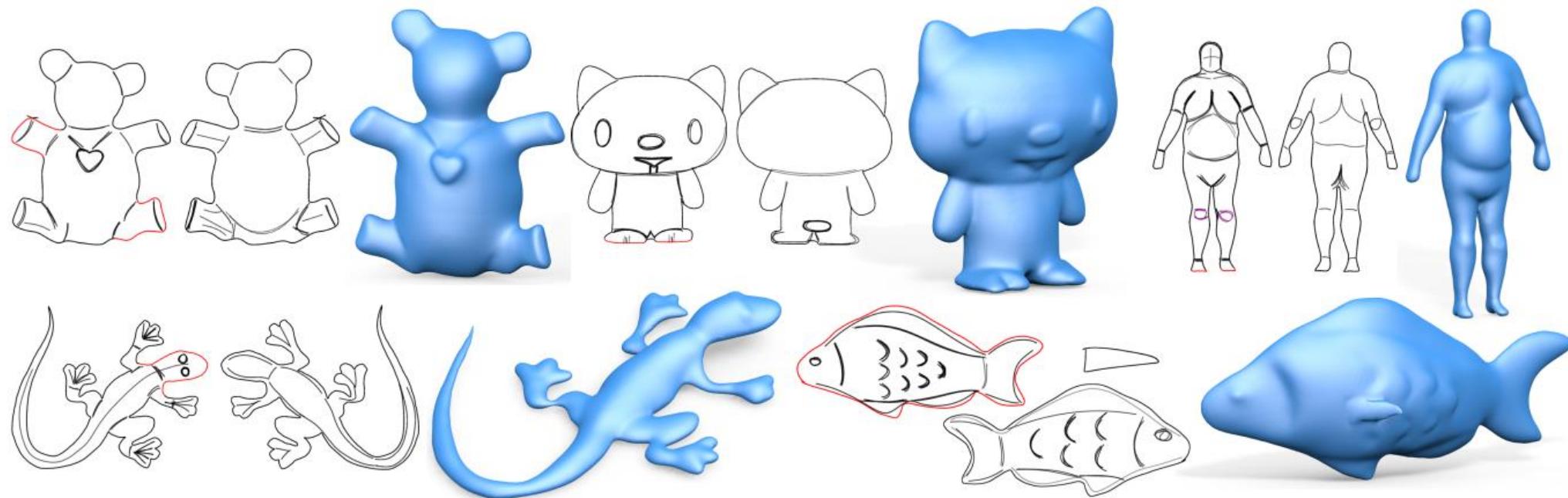


三维生成应用

- 基于图像的三维生成
- 基于草图的三维生成
- 基于文字的三维生成
- 学习形状空间: 采样、插值、对应...
- 三维补全与增强

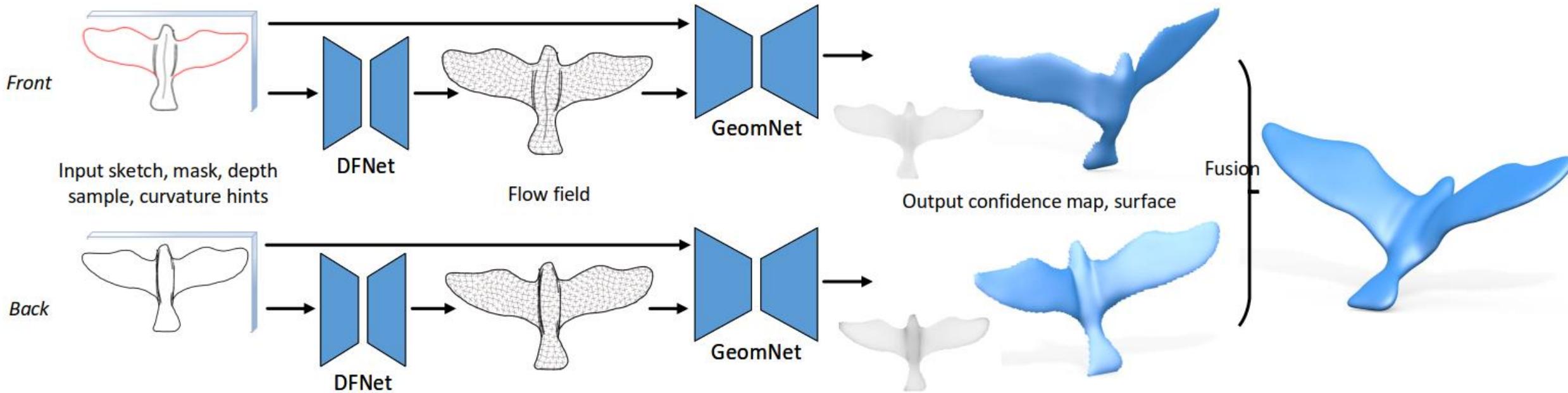
基于草图的三维生成

- 我们的工作： Robust Flow-Guided Neural Prediction for Sketch-Based Freeform Surface Modeling, conditionally accepted to Siggraph Asia 2018.
- 优点：无物体类别的限制；预测confidence map以降低二义性；更好的曲面质量；支持sharp feature等类型。



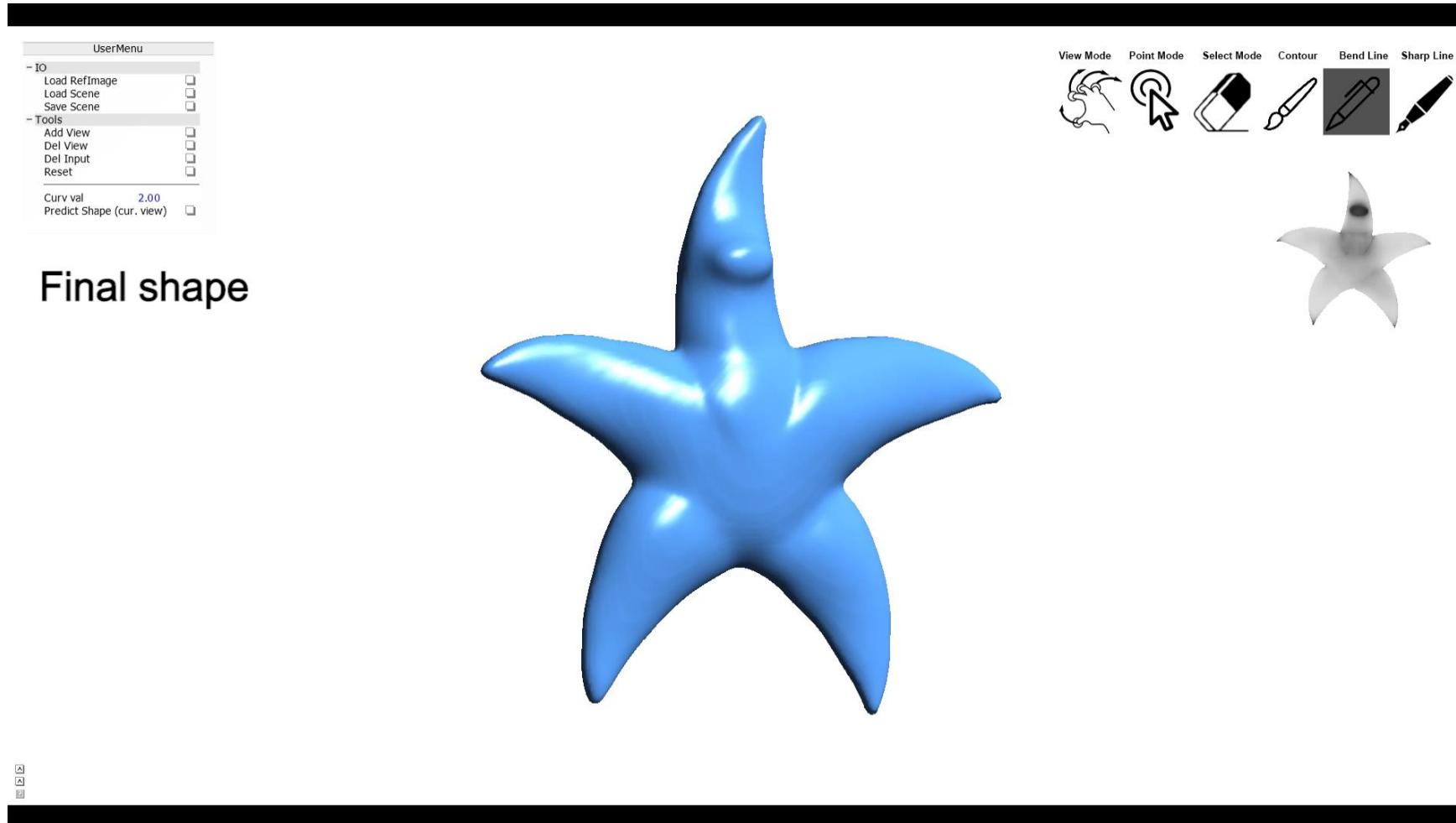
基于草图的三维生成

- 网络：单视图下先预测曲面的曲率场在2D上投影，再预测像素点的高度、法向以及confidence map；多视图之间利用重叠区域的信息做融合。



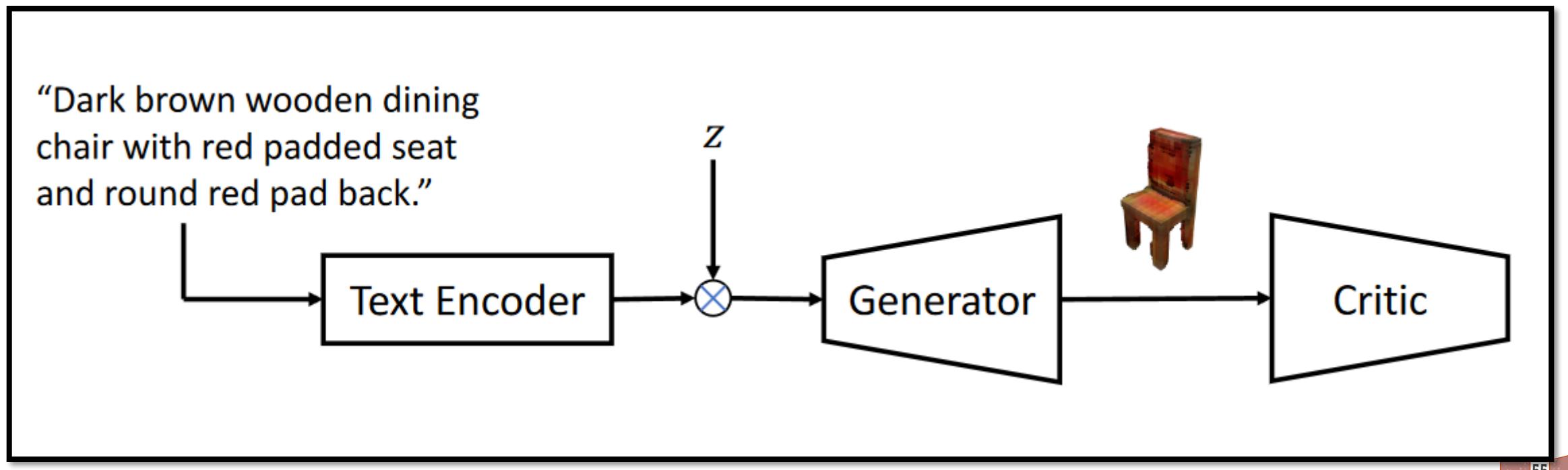
Robust Flow-Guided Neural Prediction for Sketch-Based Freeform Surface Modeling

基于草图的三维生成



基于文字的三维生成

- 已有工作：Text2Shape: Generating Shapes from Natural Language by Learning Joint Embeddings, arXiv 2018
- 核心：1. 学习 Text&3D joint embedding；2. Text-3D 生成网络



基于文字的三维生成

Input Text	GAN-INT-CLS [10]	Ours CGAN	Ours CWGAN	GT
1) modern dining table chair with nice and sturdy for wooden legs. looks comfortable				
2) Chair with wave back. Seat is light grey and legs and armrests are curved wood.				
3) brown wooden chair with fabric on the seat and the backrest				

目前三维生成工作的不足

- 外观完整度、精细度与传统三维建模的差距；
- 对于结构和功能的理解刚刚起步；
- 多归纳、少推理

展望

- 形 (shape) 神 (functional) 兼备
- 其他表达: Graph
- 形状空间的构建, 生成的多样性
- 不同媒体空间的联合嵌入
- 表观空间、物理属性空间
- 数据集的体量与质量
- 专注特定类别的应用: 如人脸、人体、室内家具、服装等。

- 近期MSRA-IG组三维深度学习的工作
- Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Chun-Yu Sun, Xin Tong: **O-CNN: Octree-based Convolutional Neural Networks for 3D Shape Analysis**, *SIGGRAPH* 2017.
- Peng-Shuai Wang, Yang Liu, Chun-Yu Sun, Xin Tong. **Adaptive O-CNN: A Patch-based Deep Representation of 3D Shapes**, conditionally accepted to *SIGGRAPH Asia* 2018.
- Changjian Li, Hao Pan, Yang Liu, Xin Tong, Alla Sheffer, Wenping Wang: **Robust Flow-Guided Neural Prediction for Sketch-Based Freeform Surface Modeling**, conditionally accepted to *SIGGRAPH Asia* 2018.
- Hao Pan, Shi-Lin Liu, Yang Liu, Xin Tong: **Convolutional Neural Networks on 3D Surfaces Using Parallel Frames**, arXiv:1808.04952, 2018. (未发表)
- 近期MSRA-IG组表观建模深度学习的工作
- Jian Shi, Yue Dong, Hao Su, Stella X. Yu, **Learning Non-Lambertian Object Intrinsics across ShapeNet Categories**, *CVPR* 2017.
- Xiao Li, Yue Dong, Pieter Peers, Xin Tong, **Modeling Surface Appearance from a Single Photograph using Self-augmented Convolutional Neural Networks**, *SIGGRAPH* 2017.

生成

表达

谢谢！

欢迎批评指正！