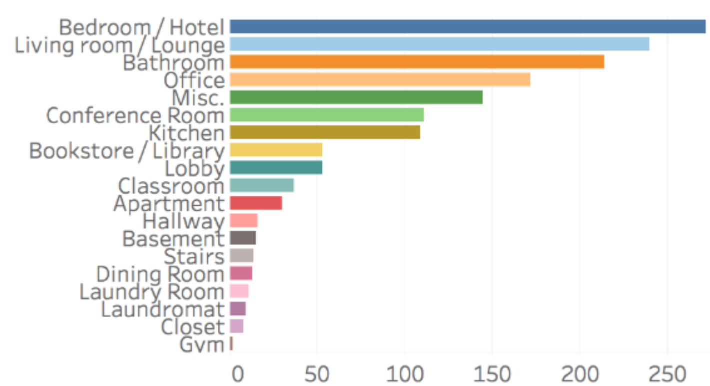
Datasets for Indoor 3D Scenes

Scanned Real Scenes: ScanNet

- 2.5M Views in 1,500 RGBD scans
- 3D camera poses
- Surface reconstructions
- Instance-level semantic segmentations

Most recently:

- ARKitScenes,
- ScanNet++ (with DSLR images)

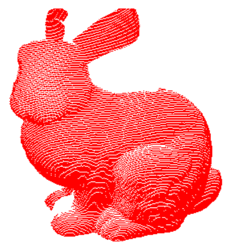
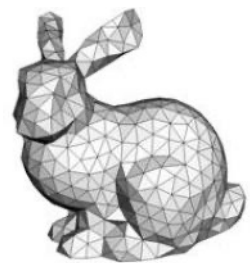
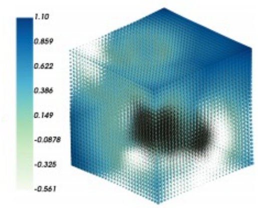
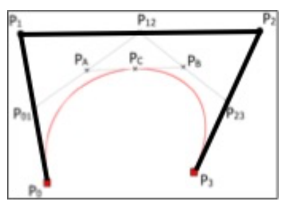
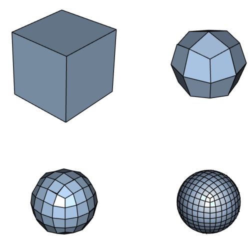
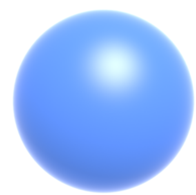
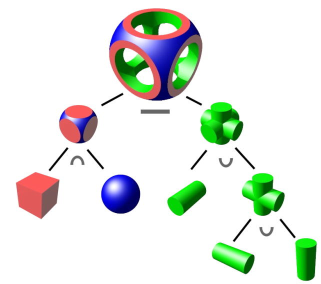


Dai et al. CVPR 2017

AI + Geometry: Tasks

- $P(S)$ or $P(S|c)$ --- Generative models
  - Learning (conditional) shape priors
  - Shape generation, completion, & geometry data processing
- $P(c|S)$ --- Discriminative models
  - Learning shape descriptors
  - Shape classification, segmentation, view estimation, etc.
- Joint modeling of 3D and 2D data
  - Large-scale 2D datasets & very good pretrained models
  - Differentiable projection/back-projection & differentiable/neural rendering
- Joint modeling of multi-modal data beyond visual (e.g., text)

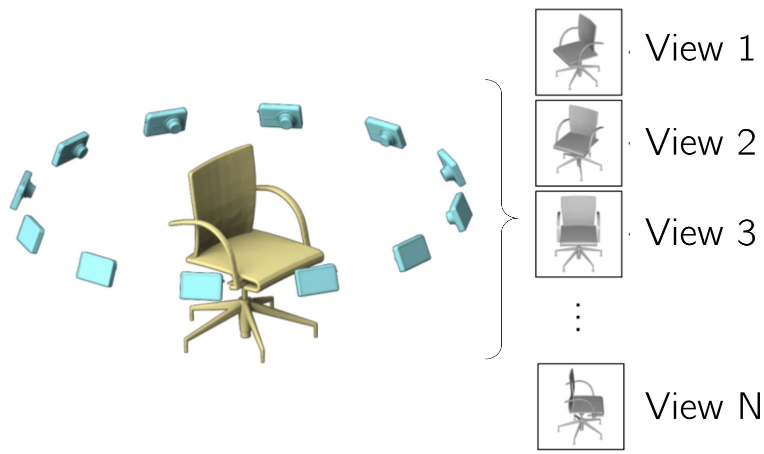
AI + Geometry: Which Representation?

	Explicit		Implicit (Eulerian)	
Non-parametric	 Points	 Meshes	 Voxels	 Level Sets
Parametric	 Splines	 Subdivision Surfaces	 $x^2+y^2+z^2=1$ Algebraic Surfaces	 Constructive Solid Geometry

Multi-View CNN

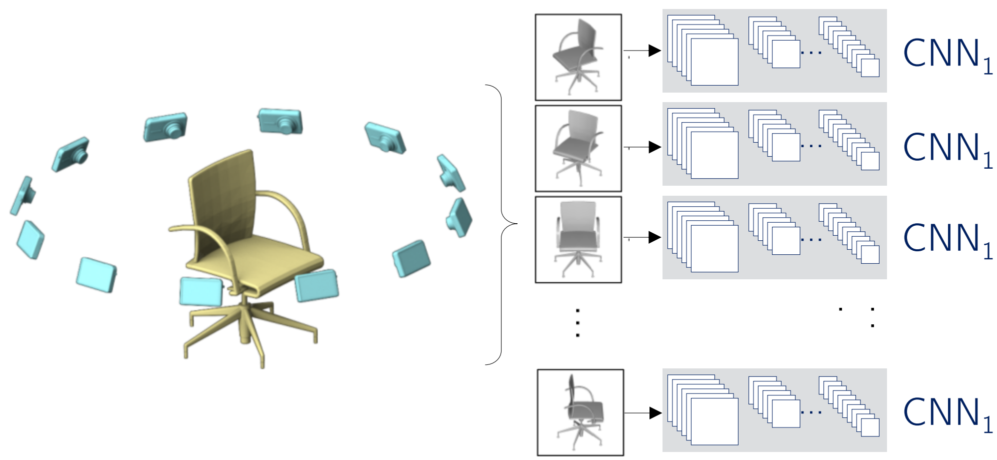
Su et al. ICCV 2015

Multi-View CNN



Su et al. ICCV 2015

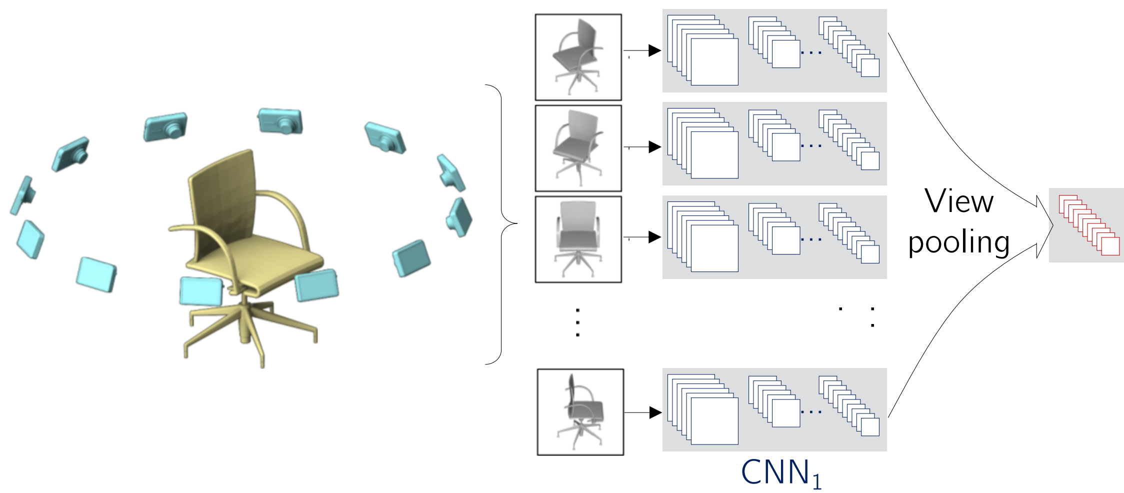
Multi-View CNN



CNN_1: a ConvNet extracting image features

Su et al. ICCV 2015

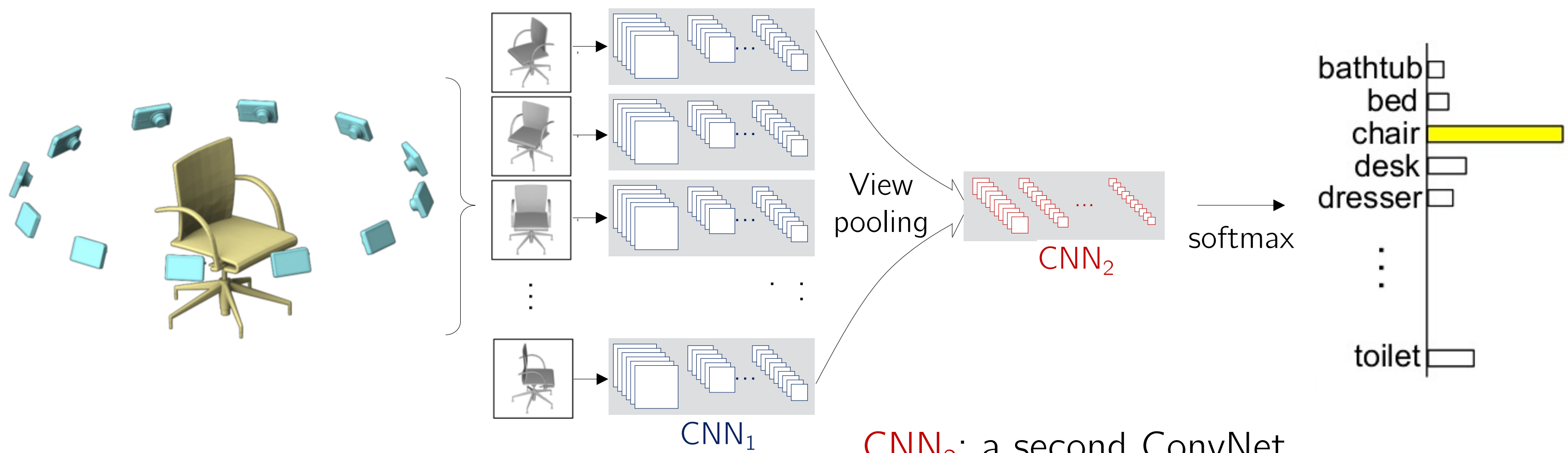
Multi-View CNN



View pooling: element-wise
max-pooling across all views

Su et al. ICCV 2015

Multi-View CNN



CNN$_2$: a second ConvNet producing shape descriptors

Su et al. ICCV 2015

Experiments – Classification & Retrieval

	Method	Classification (Accuracy)	Retrieval (mAP)
Non-deep	SPH	68.2%	33.3%
	LFD	75.5%	40.9%
	3D ShapeNets	77.3%	49.2%
	FV, 12 views	84.8%	43.9%
	CNN, 12 views	88.6%	62.8%
	MVCNN, 12 views	**89.9%**	70.1%
	MVCNN+metric, 12 views	89.5%	**80.2%**
	MVCNN, 80 views	90.1%	70.4%
	MVCNN+metric, 80 views	**90.1%**	**79.5%**

On ModelNet 40

Su et al. ICCV 2015

Multi-View Representations

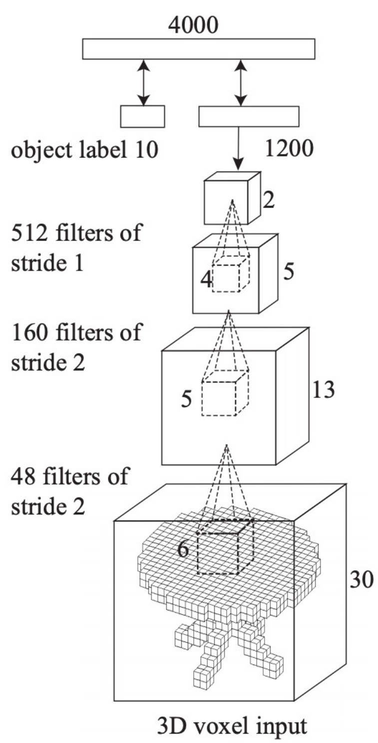
- Indeed gives good performance
- Can leverage vast literature of image classification
- Can use pretrained features

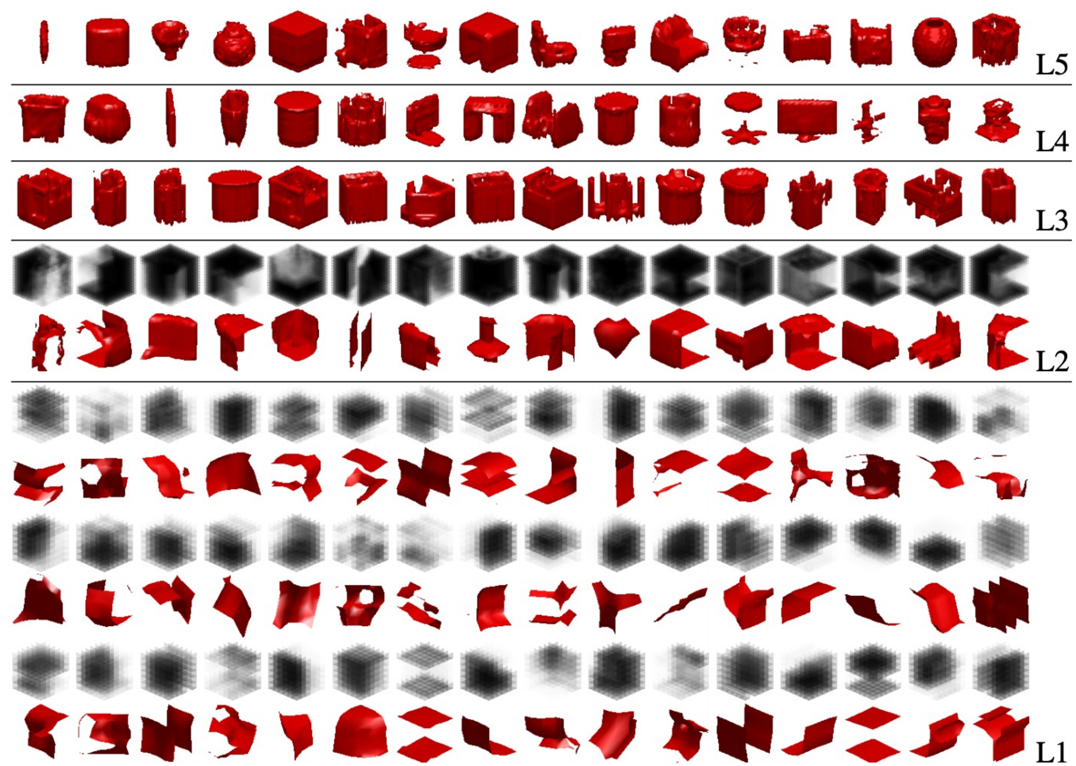
- Need projection
- What if the input is noisy and/or incomplete? e.g., point cloud

Slide Credit: Hao Su

Pixels -> Voxels

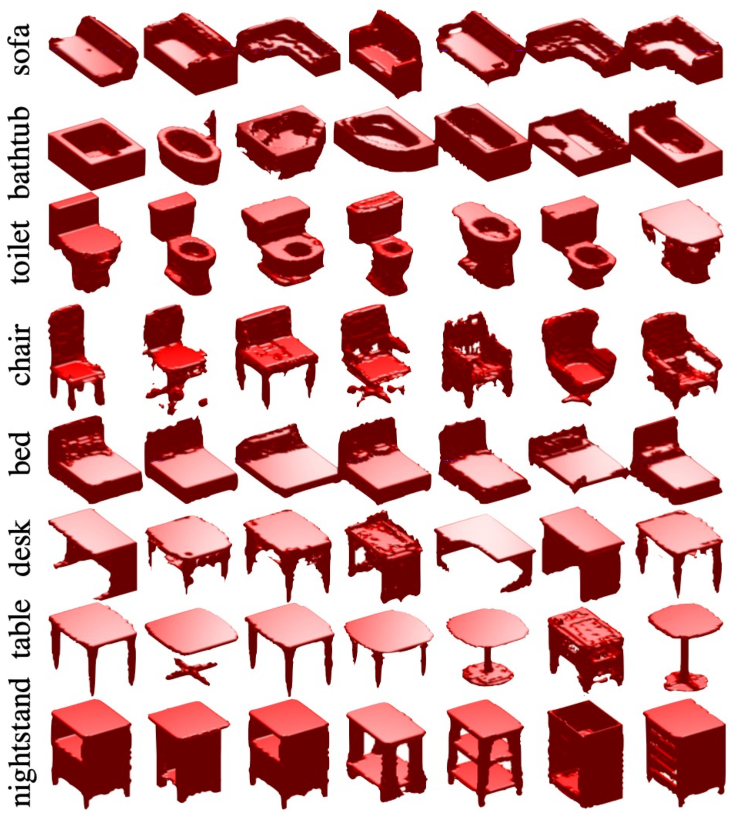
- 3D Conv Deep Belief Networks (CDBN)





Wu et al. CVPR 2015

Generative Modeling

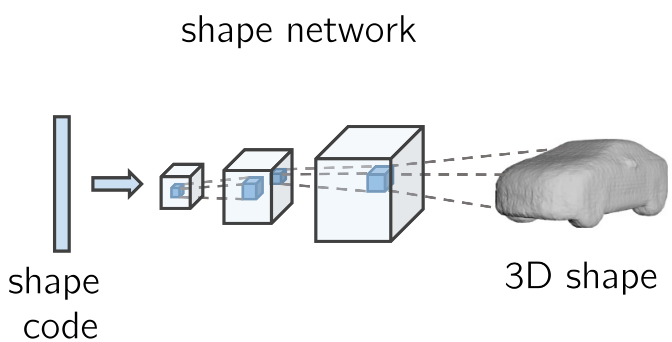


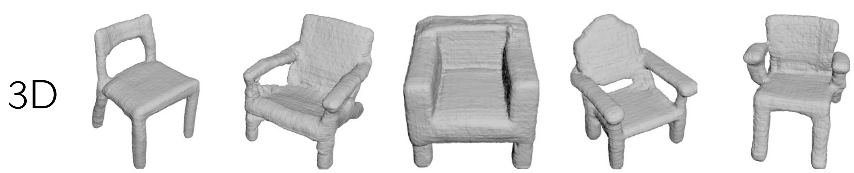
10 classes	SPH [18]	LFD [8]	Ours
classification	79.79 %	79.87 %	**83.54%**
retrieval AUC	45.97%	51.70%	**69.28%**
retrieval MAP	44.05%	49.82%	**68.26%**
40 classes	SPH [18]	LFD [8]	Ours
classification	68.23%	75.47%	**77.32%**
retrieval AUC	34.47%	42.04%	**49.94%**
retrieval MAP	33.26%	40.91%	**49.23%**

Table 1: **Shape Classification and Retrieval Results.**

Wu et al. CVPR 2015

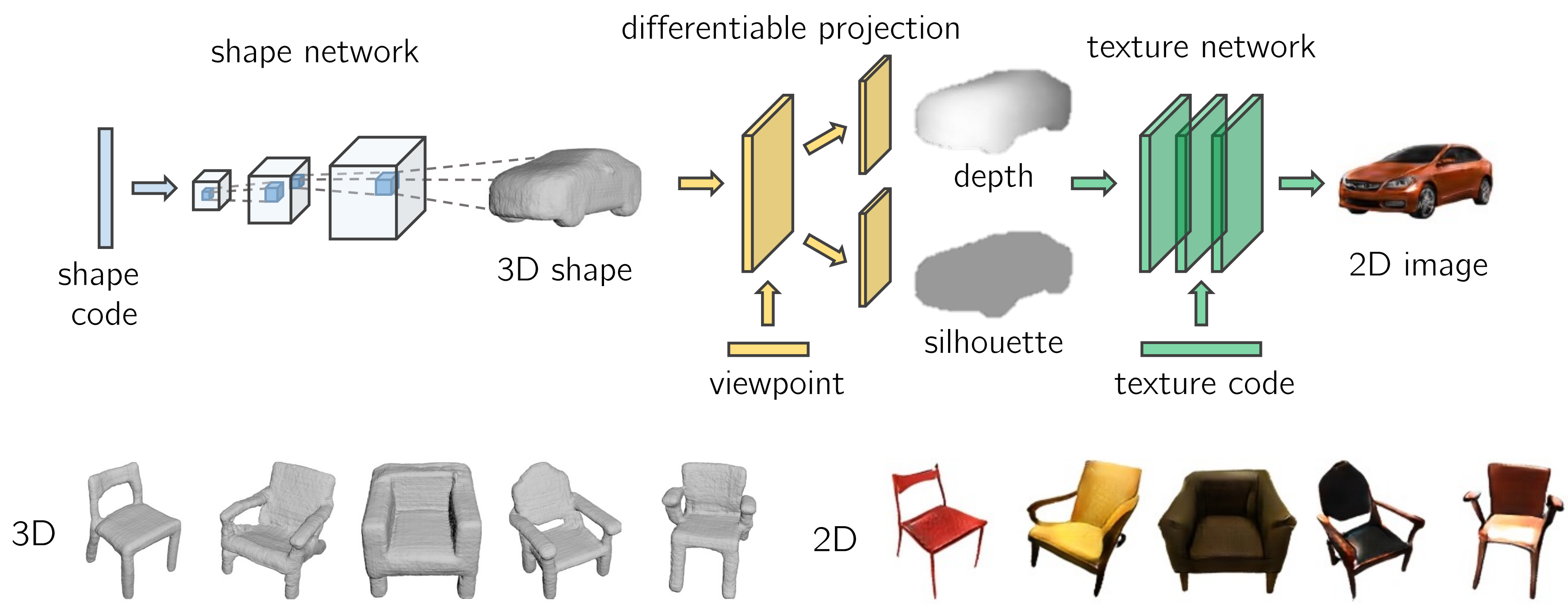
3D-GANs





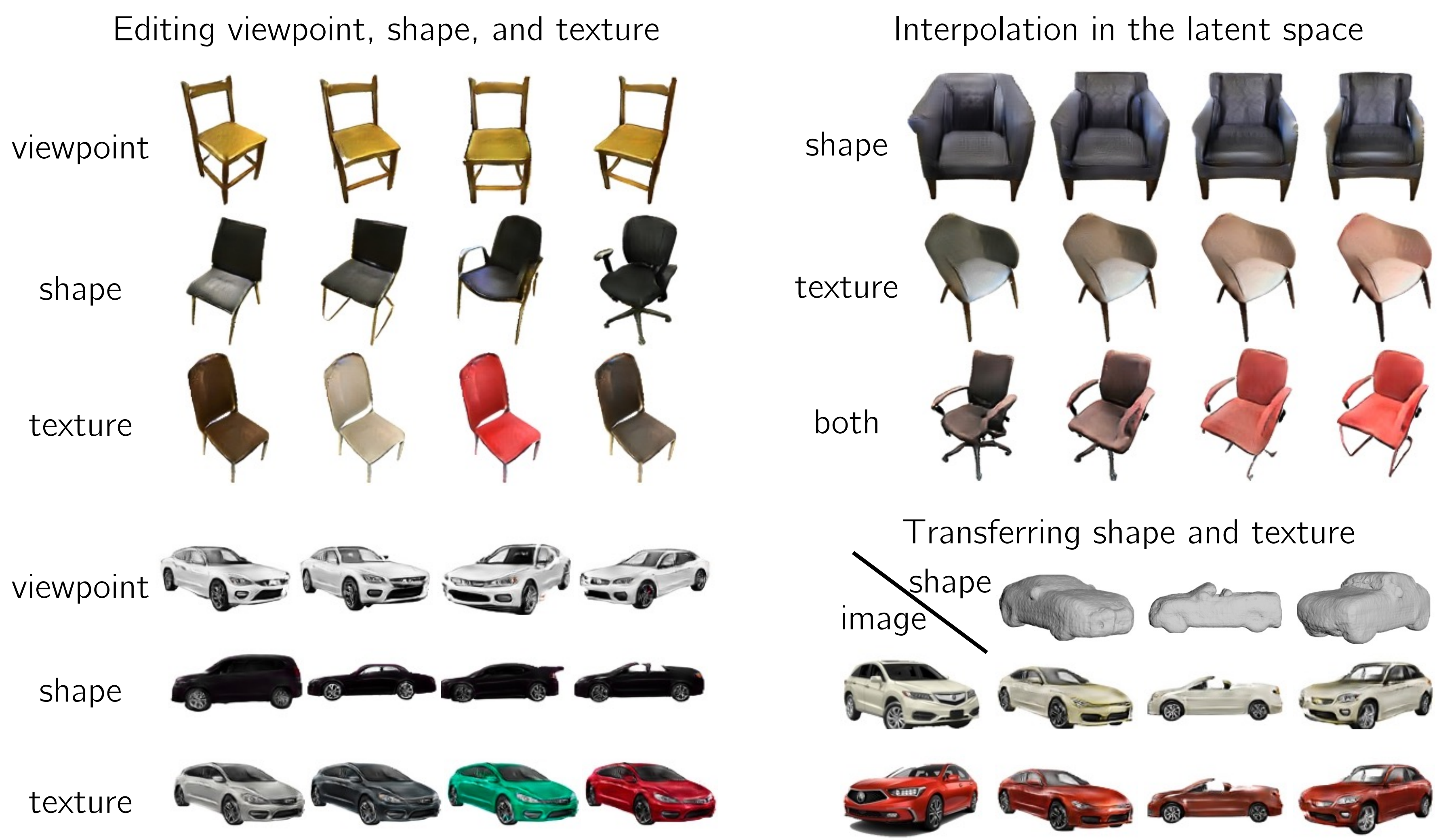
Wu et al. NeurIPS 2016

Visual Object Networks (Geometry + Rendering)



Wu et al. NeurIPS 2016

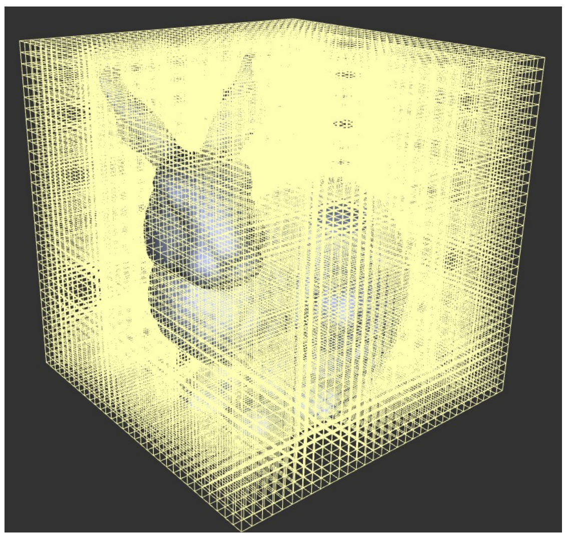
Zhu et al. NeurIPS 2018

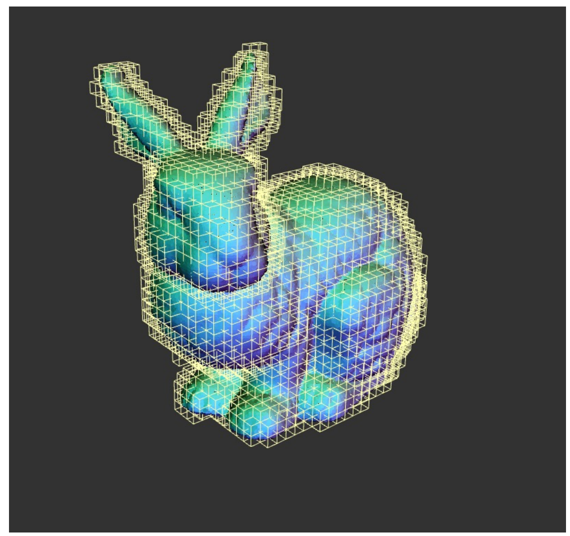


Zhu et al. NeurIPS 2018

Octave Tree Representations

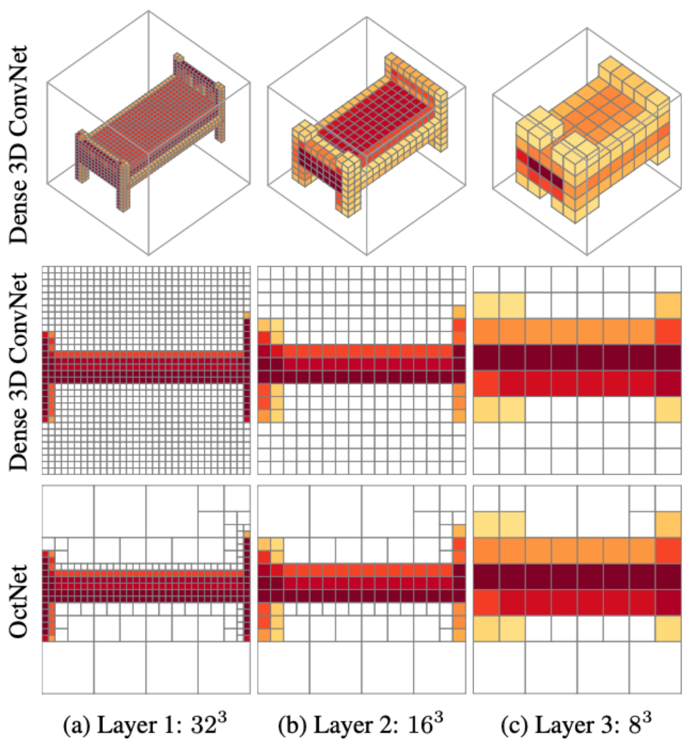
- Store the sparse surface signals
- Constrain the computation near the surface



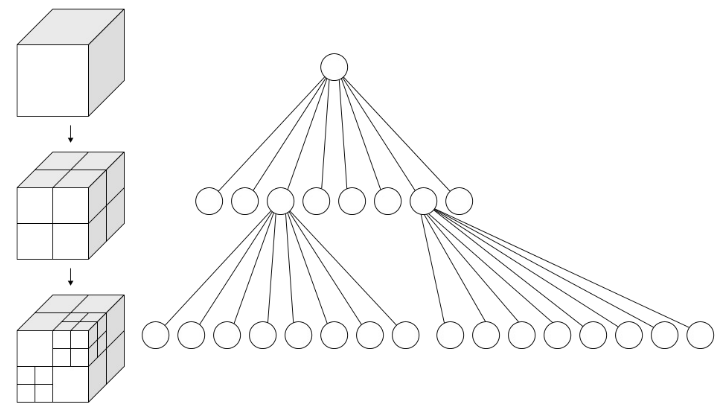


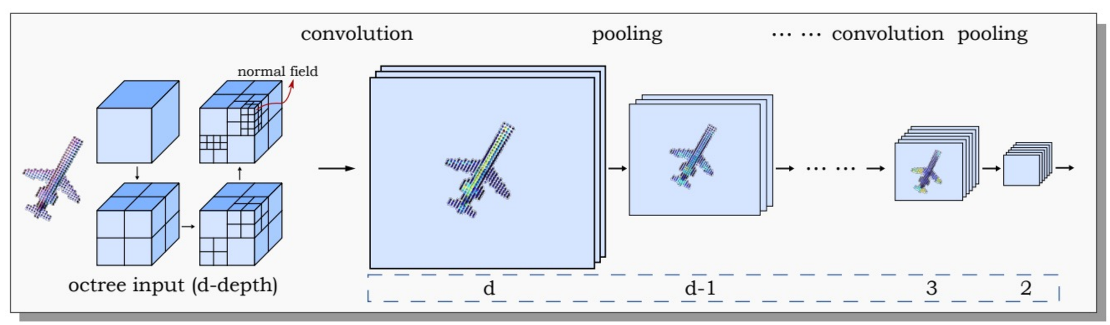
Slide Credit: Hao Su

Octree: Recursively Partition the Space



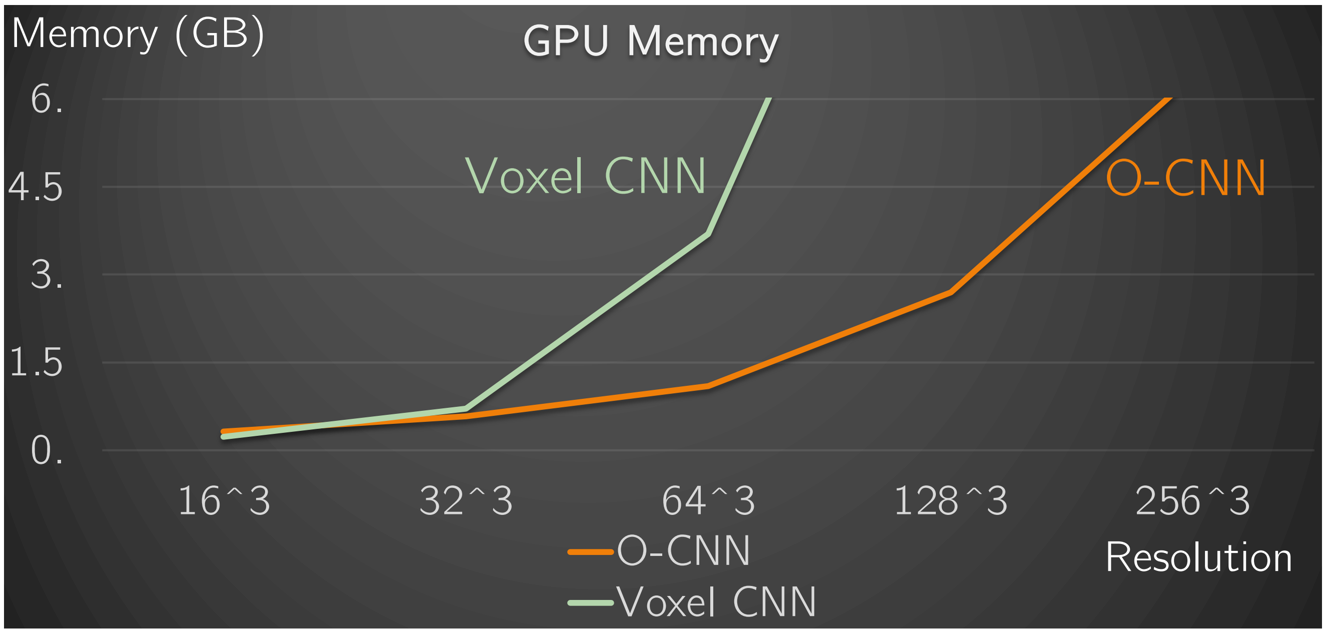
Riegler et al. OctNet. CVPR 2017





Wang et al. O-CNN. SIGGRAPH 2017

Memory Efficiency

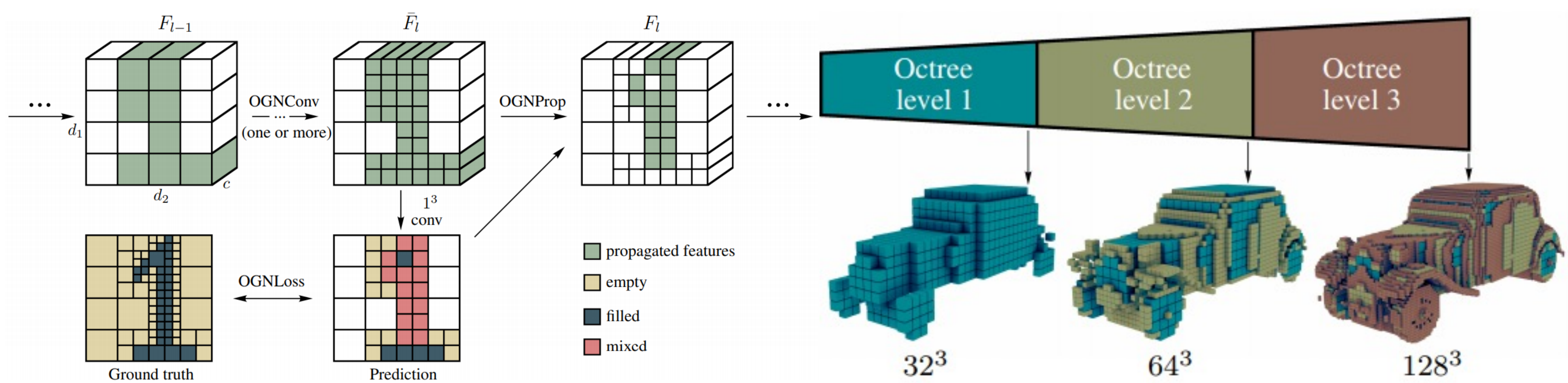


Slide Credit: Hao Su

Octree Generating Networks

Avoid $\mathcal{O}(n^3)$ reconstruction

- Octree representation of shapes
- Generate the octree layer by layer

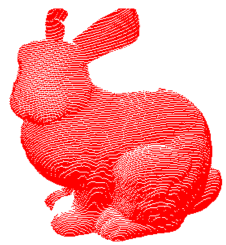


Tatarchenko et al. ICCV 2017. Slide Credit: Hao Su

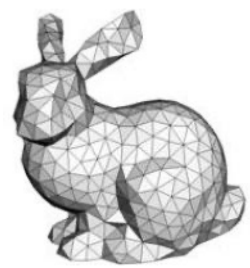
Eulerian -> Lagrangian

Explicit | Implicit

Non-parametric

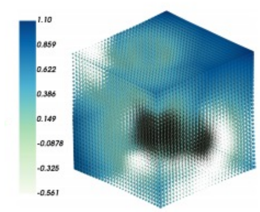


Points



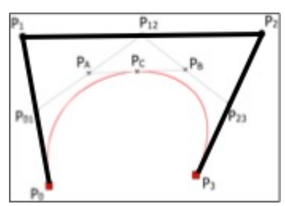
Meshes

Voxels

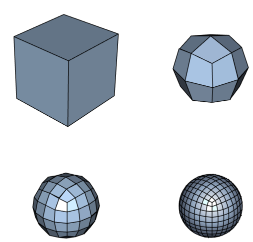


Level Sets

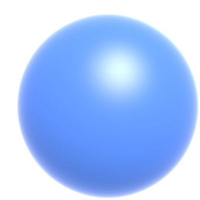
Parametric



Splines

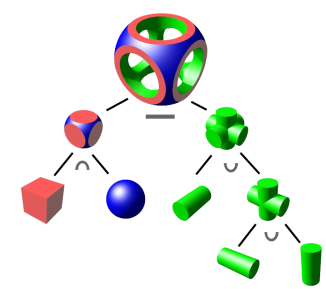


Subdivision Surfaces



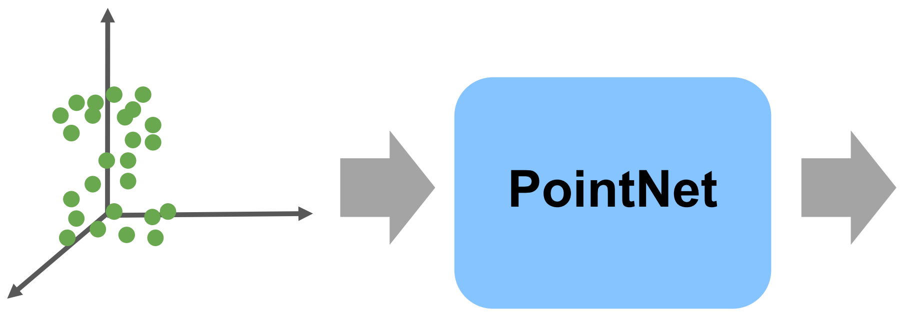
$x^2+y^2+z^2=1$

Algebraic Surfaces



Constructive Solid Geometry

PointNet: Learning on Point Clouds



Object Classification

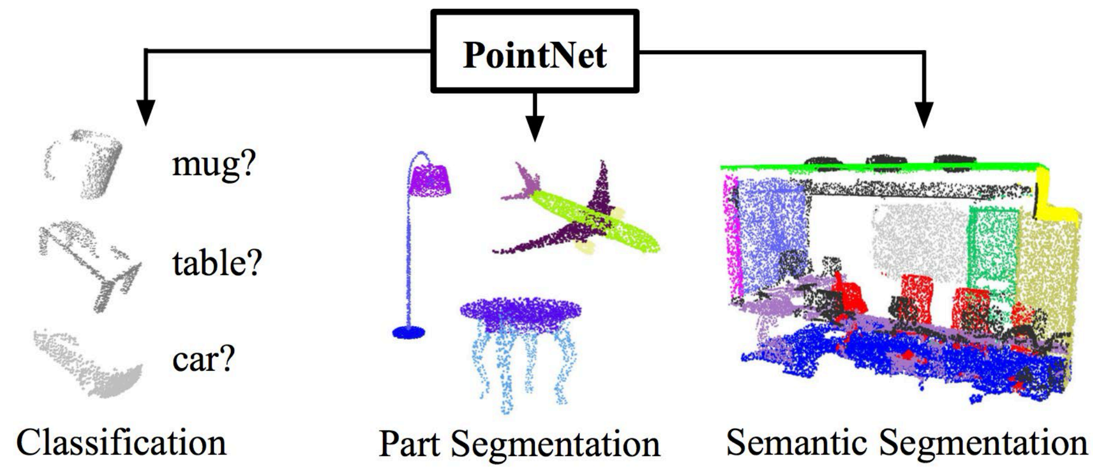
Object Part Segmentation

Semantic Scene Parsing

...

End-to-end learning for irregular point data

Unified framework for various tasks



Charles R. Qi, Hao Su, Kaichun Mo, Leonidas J. Guibas. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. (CVPR'17)

Slide credit: He Wang

Invariances

The model has to respect key desiderata for point clouds:

Point Permutation Invariance

Point cloud is a set of unordered points

Sampling Invariance

Output a function of the underlying geometry and not the sampling

Slide credit: He Wang

Permutation Invariance: Symmetric Functions

$$f(x_1, x_2, \ldots, x_n) \equiv f(x_{\pi_1}, x_{\pi_2}, \ldots, x_{\pi_n}), \quad x_i \in \mathbb{R}^D$$

Examples:

$$f(x_1, x_2, \ldots, x_n) = \max\{x_1, x_2, \ldots, x_n\}$$

$$f(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \ldots + x_n$$

$\ldots$

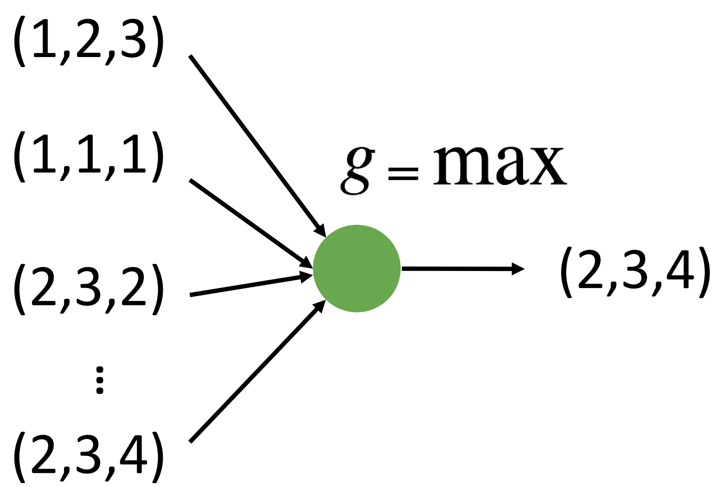
How can we construct a universal family of symmetric functions by neural networks?

Slide credit: He Wang

Construct Symmetric Functions by NNs

Simplest form: directly aggregate all points with a symmetric operator g

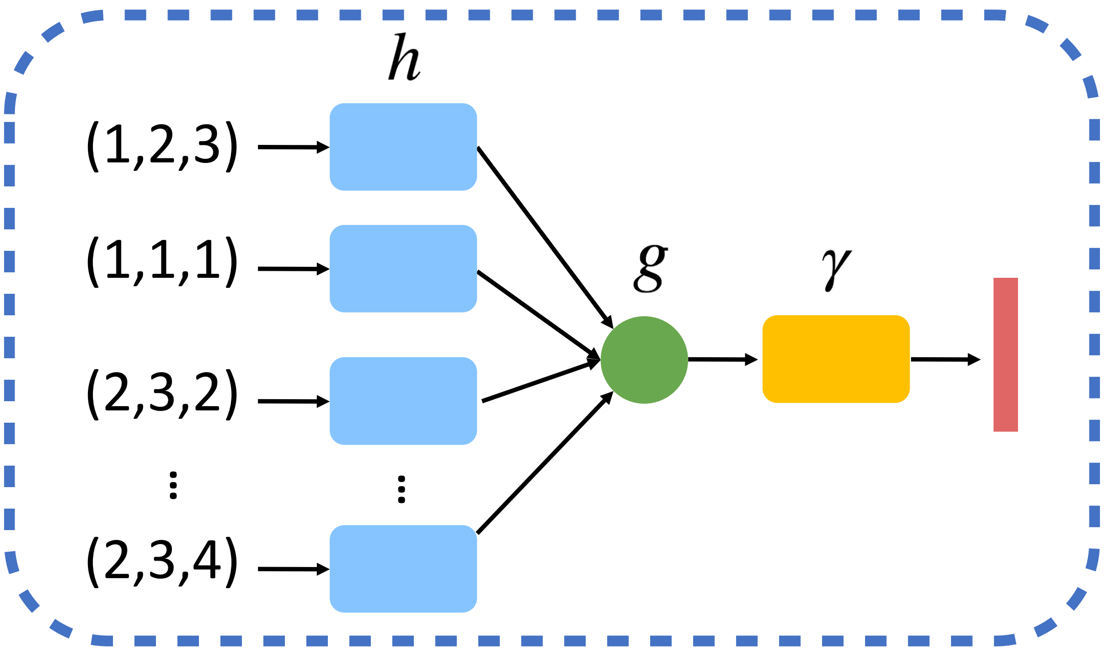
Just discovers simple extreme/aggregate properties of the geometry.



Slide credit: He Wang

Construct Symmetric Functions by NNs

$f(x_1, x_2, \ldots, x_n) = \gamma \circ g(h(x_1), \ldots, h(x_n))$ is symmetric if g is symmetric

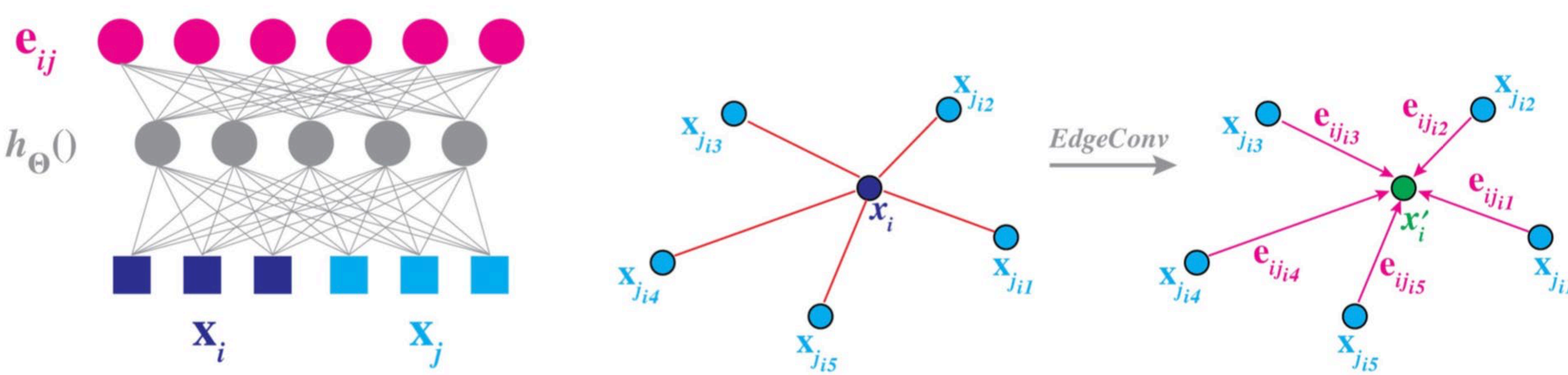


PointNet (vanilla)

Slide credit: He Wang

Graph NNs on Point Clouds

- Points -> Nodes
- Neighborhood -> Edges
- Graph NNs for point cloud processing

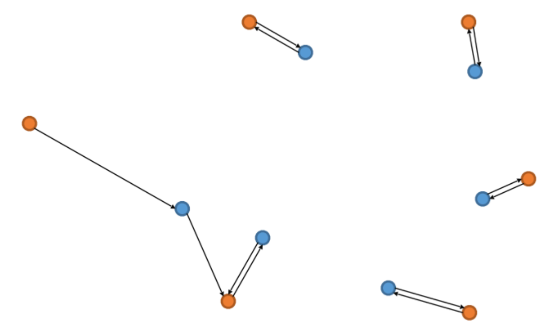


Wang et al. TOG 2019

Distance Metrics for Point Clouds

Chamfer distance We define the Chamfer distance between $S_1, S_2 \subseteq \mathbb{R}^3$ as:

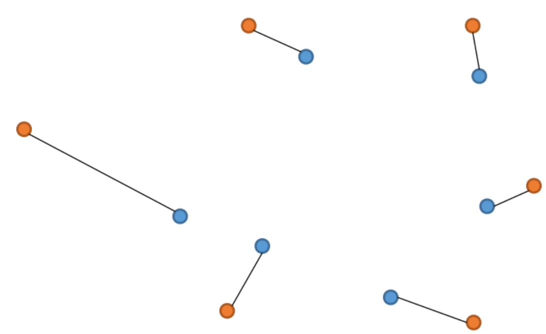
$$d_{CD}(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2 + \sum_{y \in S_2} \min_{x \in S_1} \|x - y\|_2$$



Earth Mover's distance Consider $S_1, S_2 \subseteq \mathbb{R}^3$ of equal size $s = |S_1| = |S_2|$. The EMD between A and B is defined as:

$$d_{EMD}(S_1, S_2) = \min_{\phi: S_1 \to S_2} \sum_{x \in S_1} \|x - \phi(x)\|_2$$

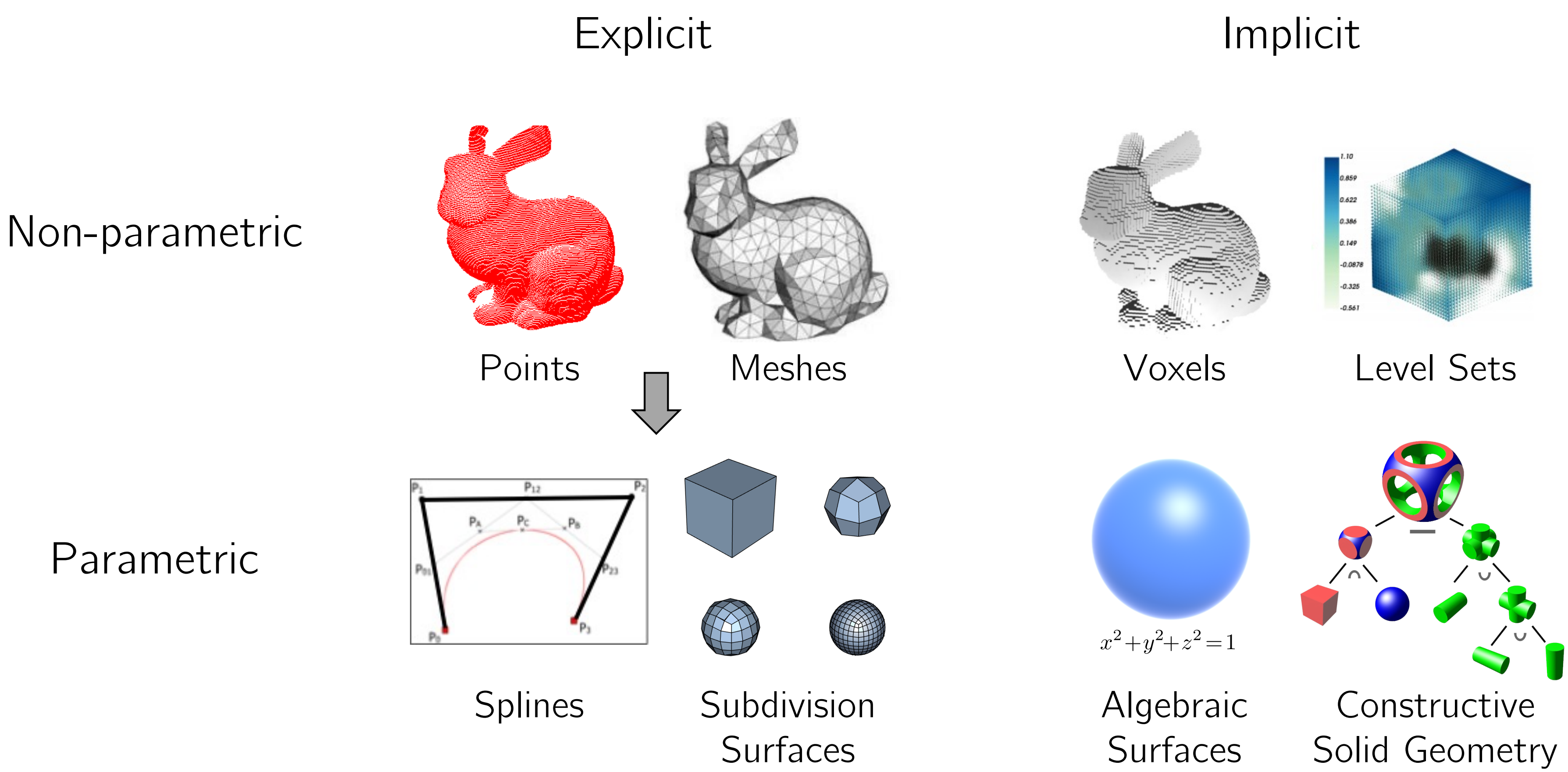
where $\phi : S_1 \to S_2$ is a bijection.



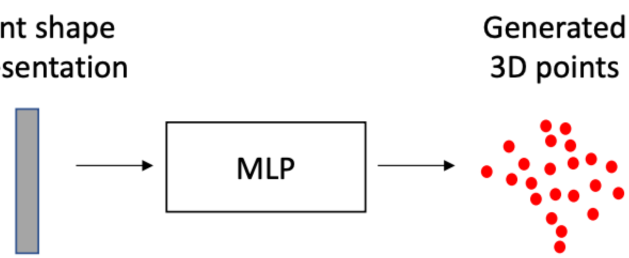
A Point Set Generation Network for 3D Object Reconstruction from a Single Image, CVPR 2016

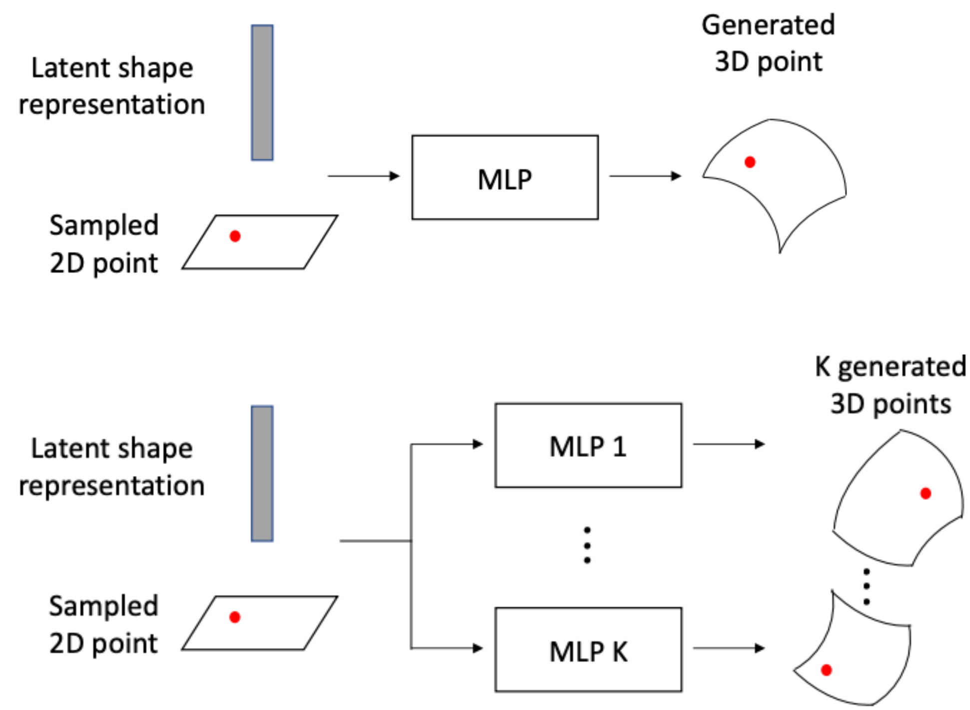
Slide credit: He Wang

Non-Parametric -> Parametric



Parametric Decoder: AtlasNet



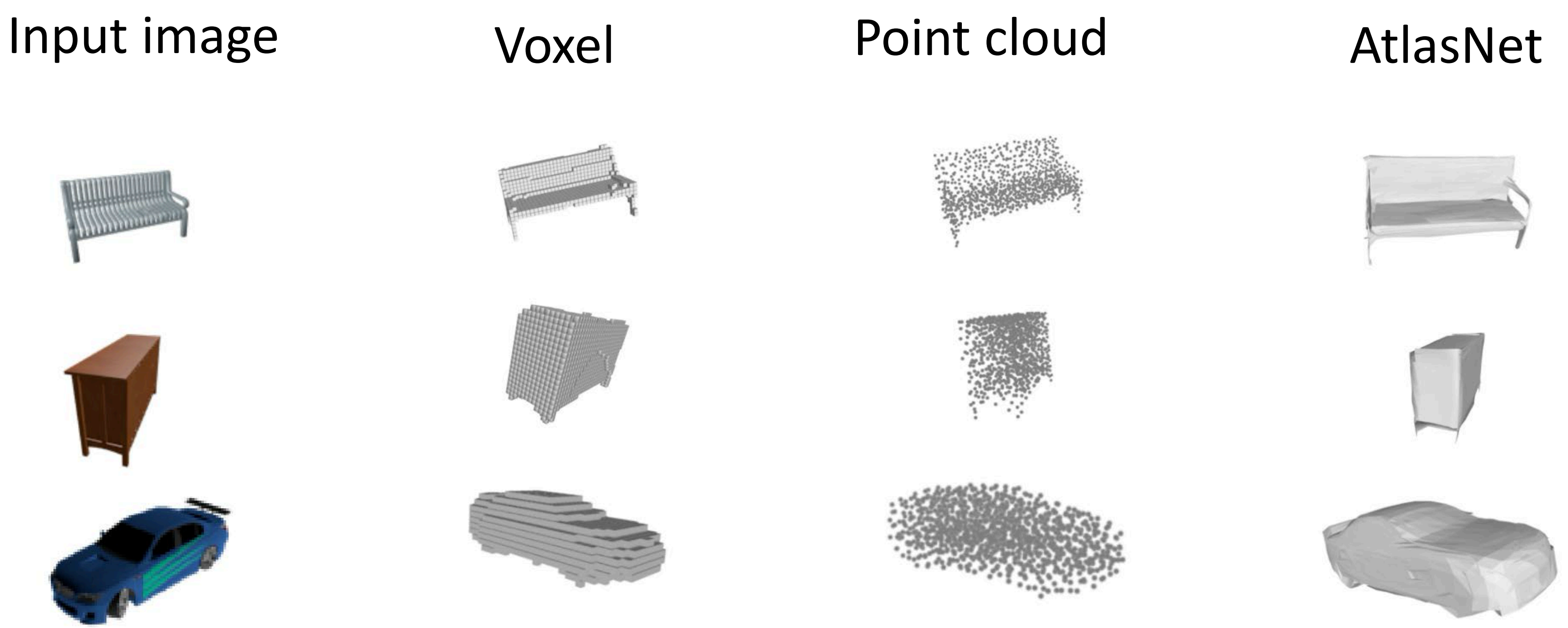


Given the output points form a smooth surface, enforce such a parametrization as input.

MLP(z, u, v) -> point

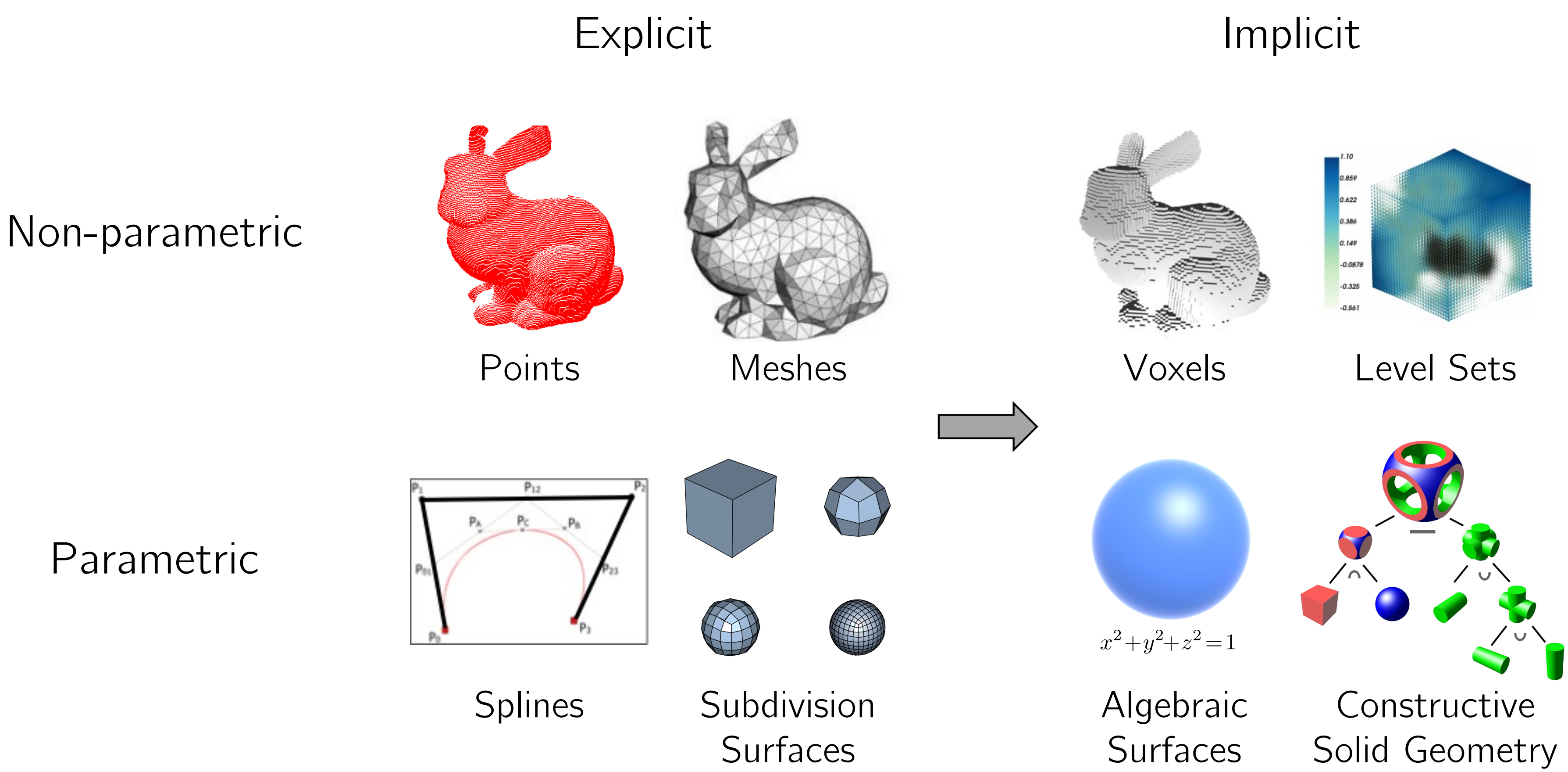
Groueix et al. CVPR 2018

Results

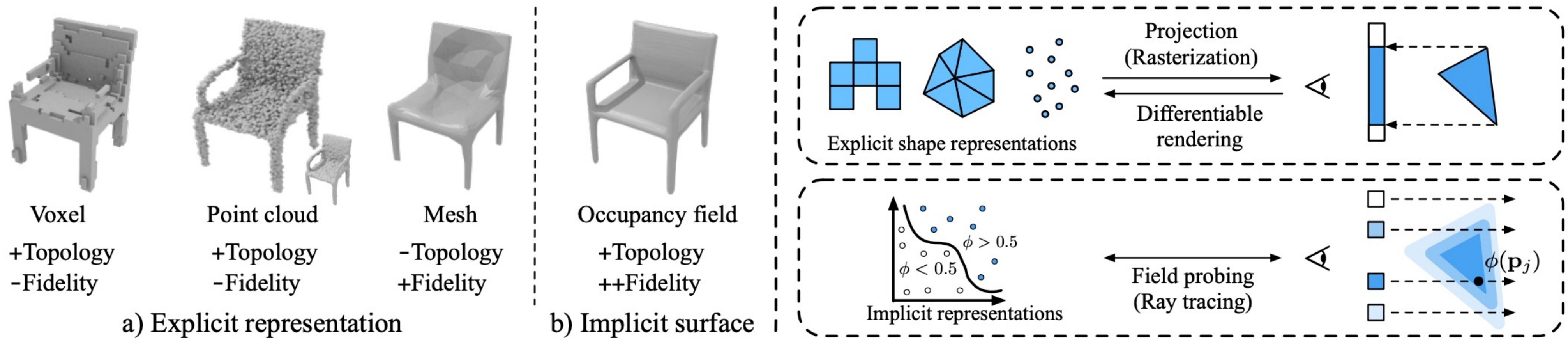


Groueix et al. CVPR 2018

Explicit -> Implicit



Deep Implicit Functions



Liu et al. Learning to Infer Implicit Surfaces without 3D Supervision. NeurIPS 2019

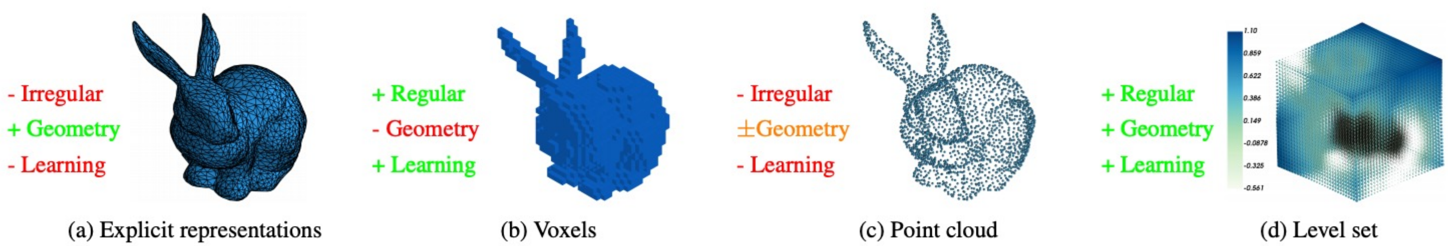
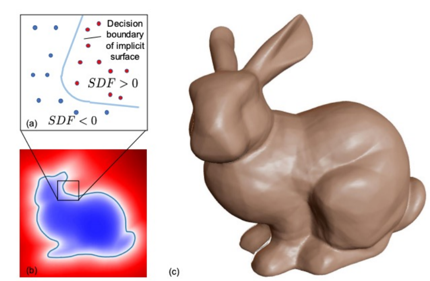
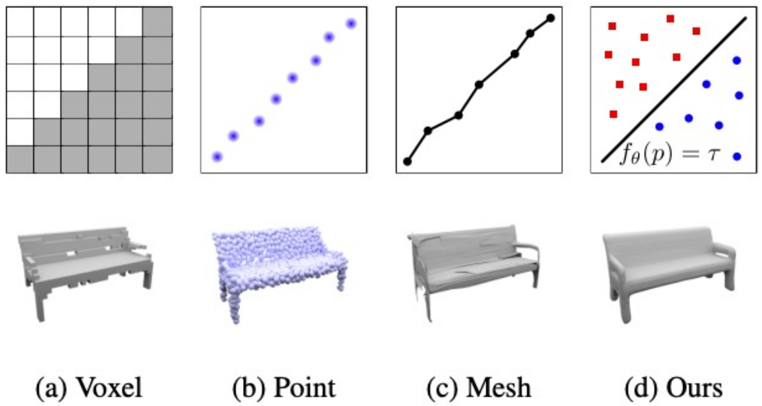


Figure 2. Four common representations of 3D shape along with their advantages and disadvantages.

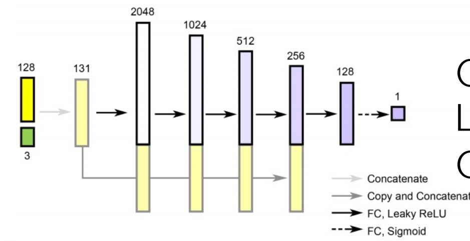
Deep Level Sets: Implicit Surface Representations for 3D Shape Inference. 2019



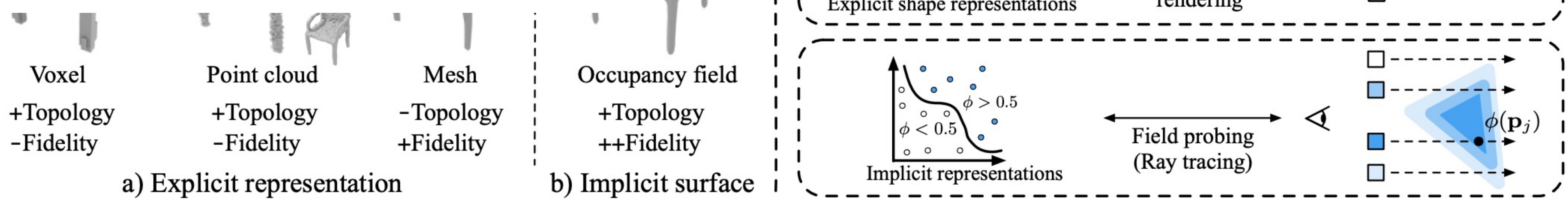
DeepSDF. CVPR 2019



Occupancy Networks
CVPR 2019



Chen and Zhang.
Learning Implicit Fields
CVPR 2019



Liu et al. Learning to Infer Implicit Surfaces without 3D Supervision. NeurIPS 2019

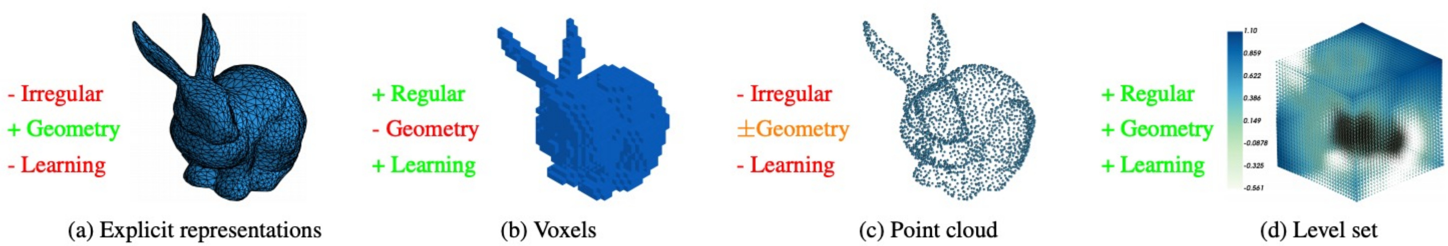
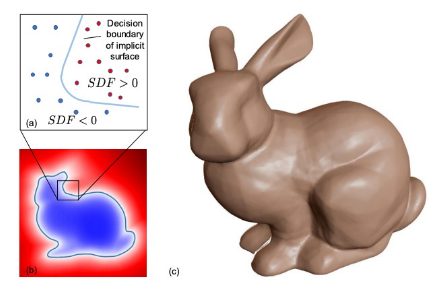


Figure 2. Four common representations of 3D shape along with their advantages and disadvantages.

Deep Level Sets: Implicit Surface Representations for 3D Shape Inference. 2019



DeepSDF. CVPR 2019

Collection of Implicit Functions

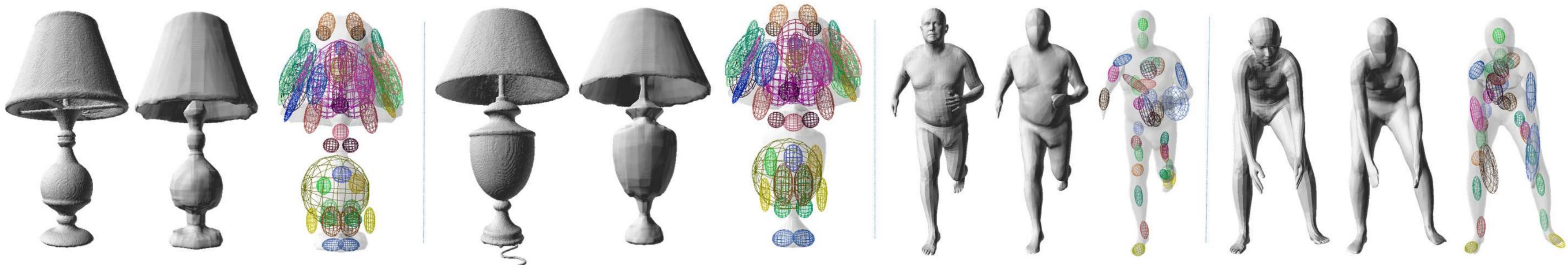
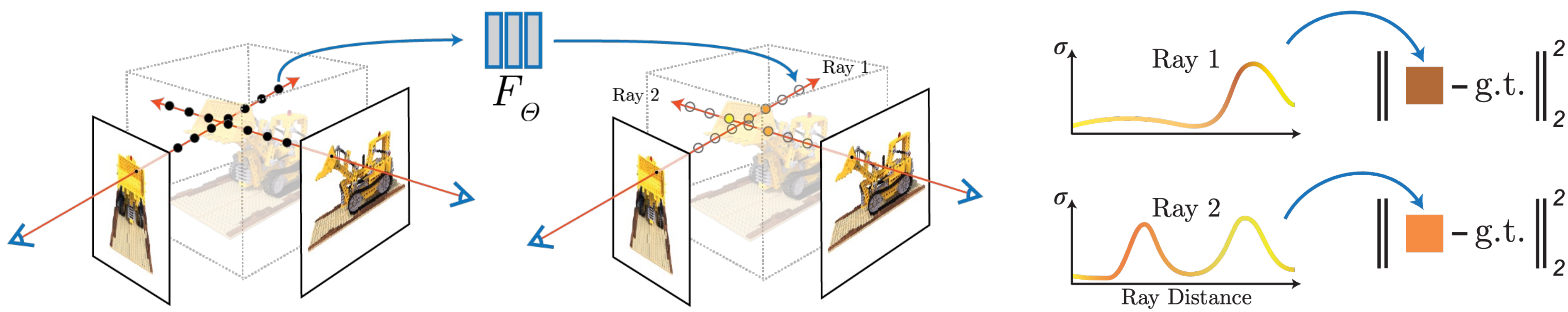


Figure 1. This paper introduces Local Deep Implicit Functions, a 3D shape representation that decomposes an input shape (mesh on left in every triplet) into a structured set of shape elements (colored ellipses on right) whose contributions to an implicit surface reconstruction (middle) are represented by latent vectors decoded by a deep network. Project video and website at ldif.cs.princeton.edu.

Genova et al. CVPR 2020

Implicit Functions for Geometry + Rendering

$$(x,y,z,\theta,\phi) \rightarrow F_\Theta \rightarrow (RGB\sigma)$$



NeRF. Mildenhall*, Srinivasan*, et al. ECCV 2020

Volume rendering is trivially differentiable.

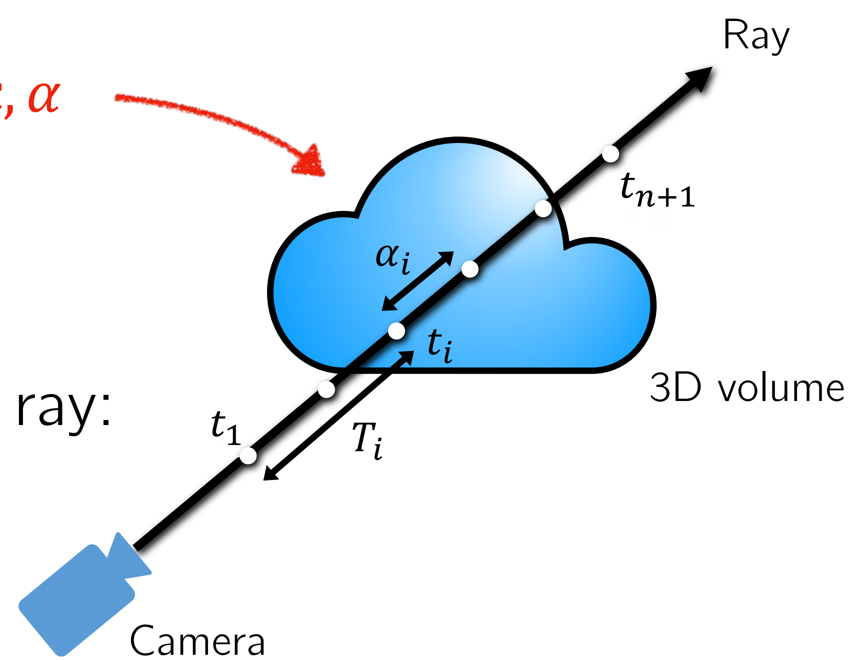
Rendering model for ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$:

$$\boxed{\mathbf{c} \approx \sum_{i=1}^{n} \underbrace{T_i \alpha_i}_{\text{weights}} \underbrace{\mathbf{c}_i}_{\text{colors}}}$$

differentiable w.r.t. $\mathbf{c}, \alpha$

How much light is blocked earlier along ray:

$$T_i = \prod_{j=1}^{i-1} (1 - \alpha_j)$$



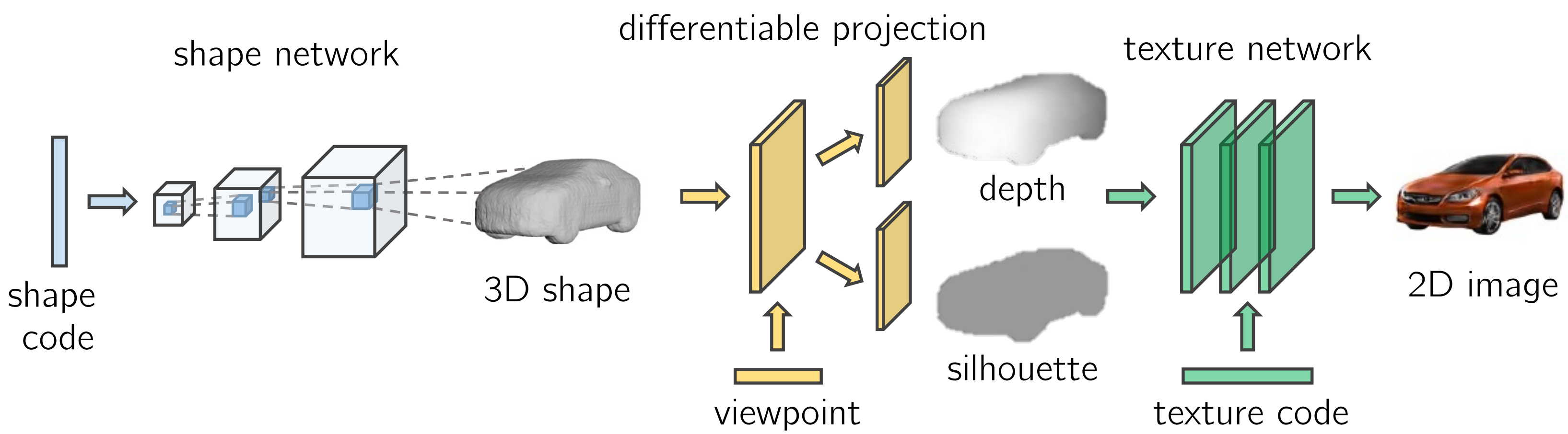
How much light is contributed by ray segment i: α_i

Slide credit: Ben Mildenhall

Reconstruction & Novel View Synthesis with NeRF

Mildenhall*, Srinivasan*, et al. ECCV 2020

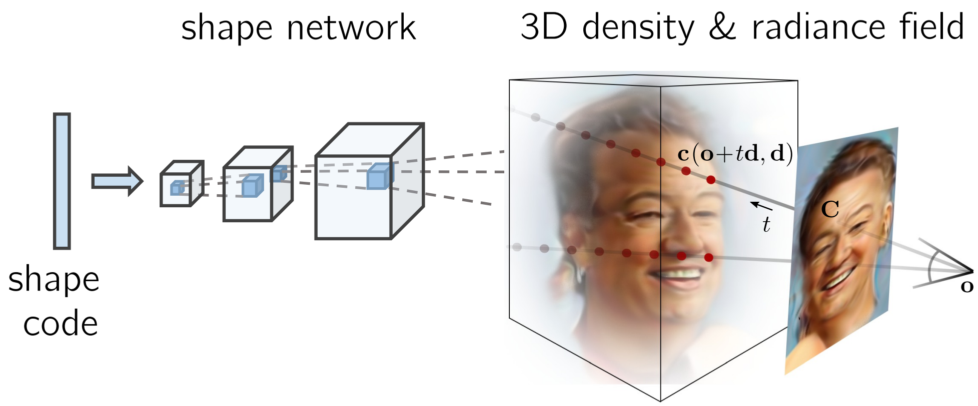
Generative Modeling with Implicit Geometry + Rendering

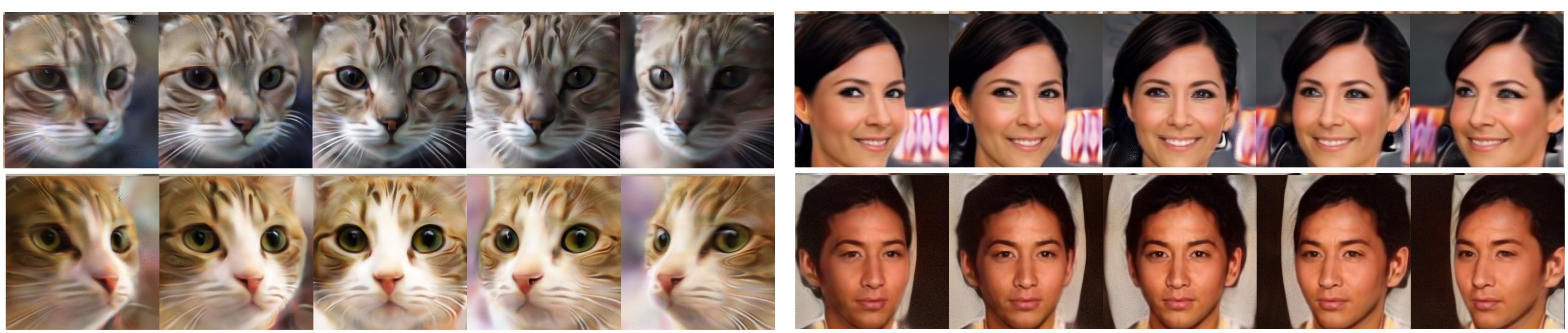


Wu et al. NeurIPS 2016

Zhu et al. NeurIPS 2018

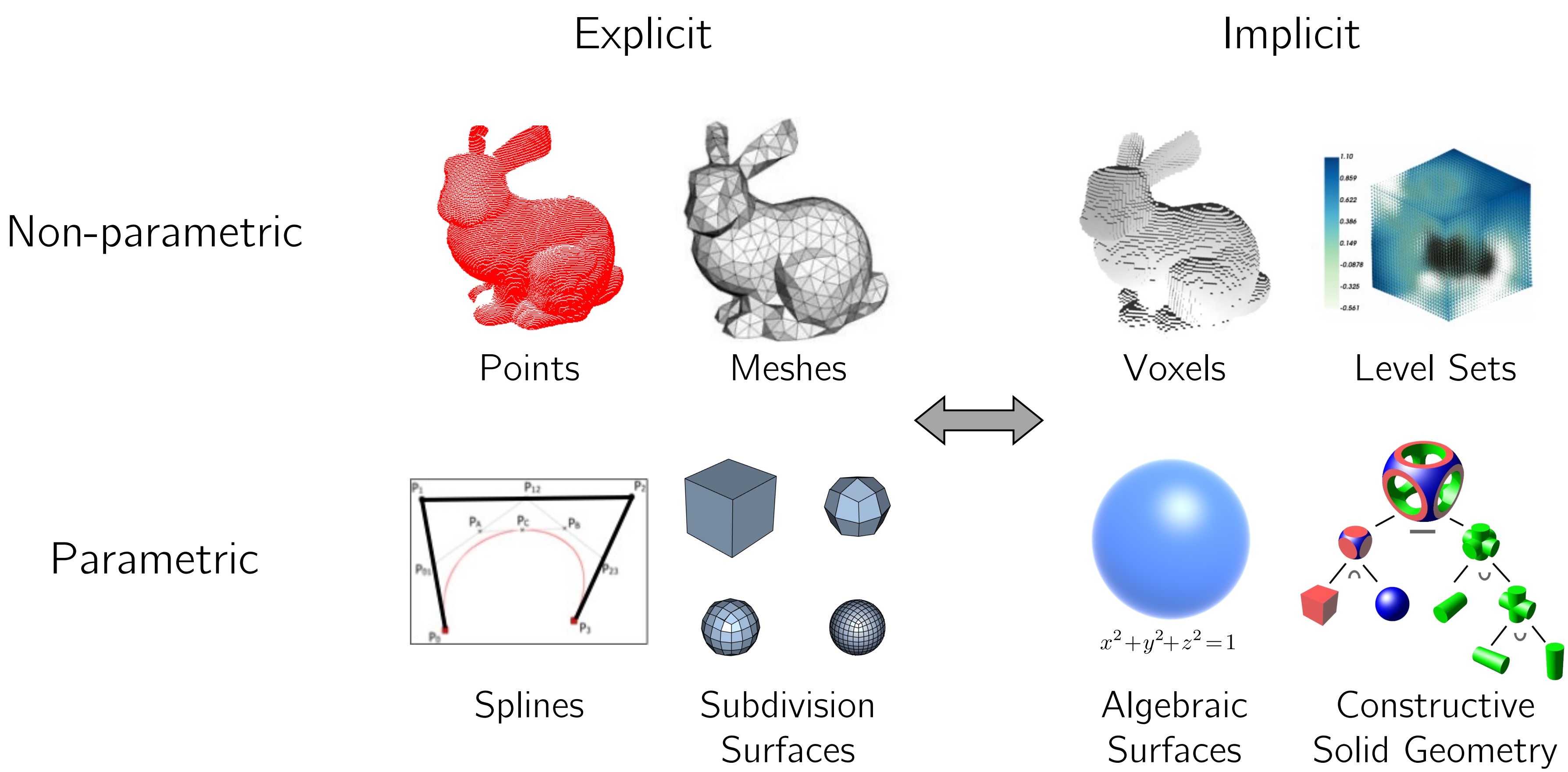
Generative Modeling with Implicit Geometry + Rendering





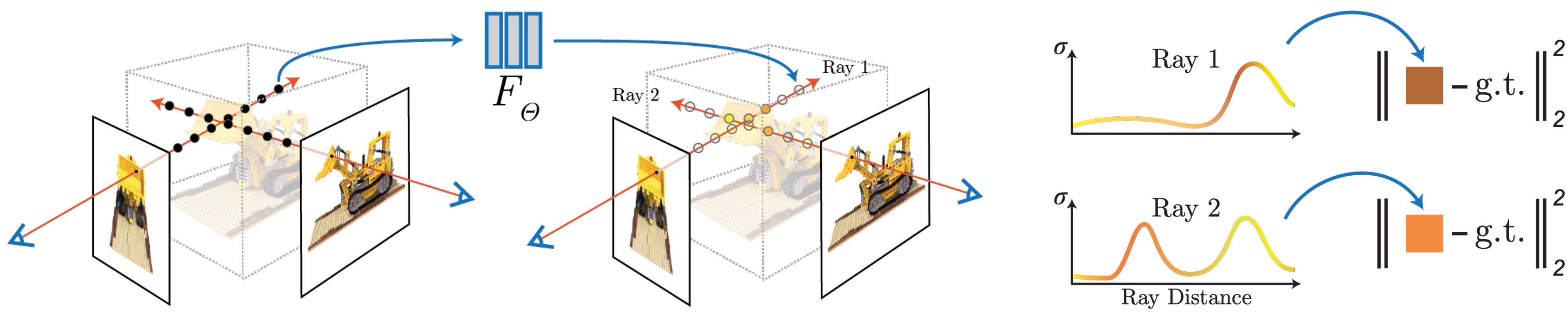
Kerbl*

Explicit <-> Implicit



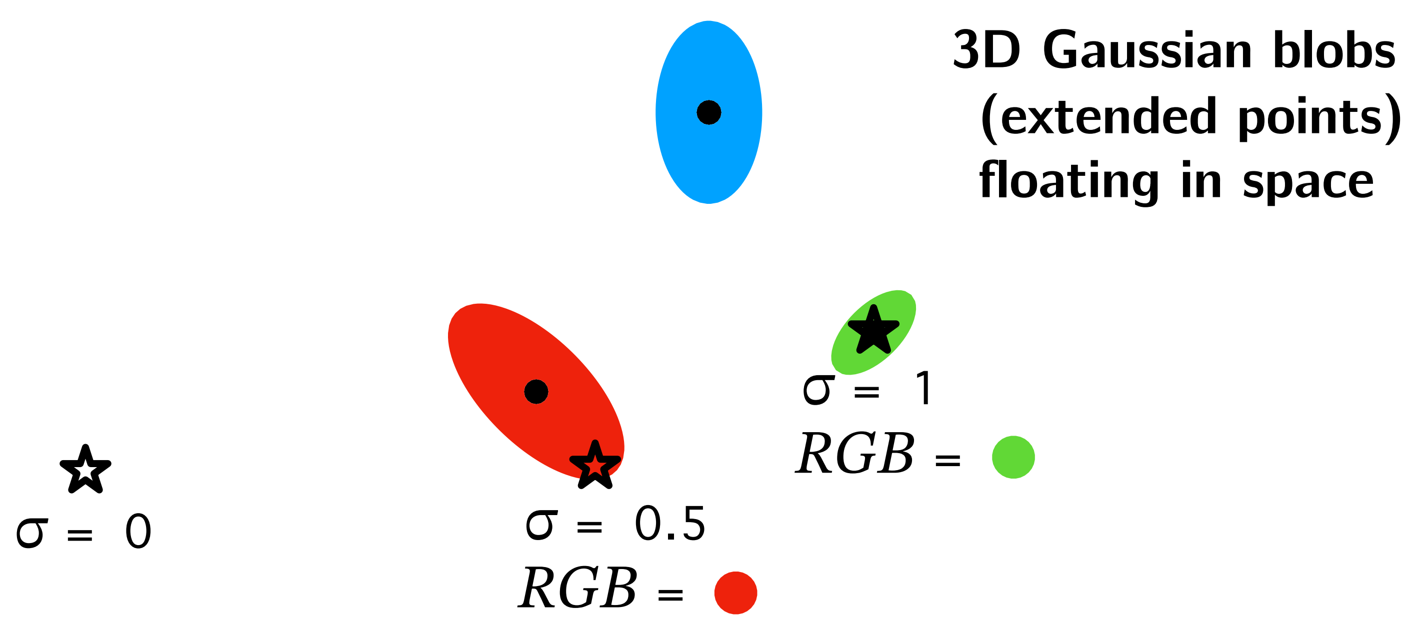
NeRF parameterizes scenes densely, at every point in space.

$$(x,y,z,\theta,\phi) \rightarrow F_\Theta \rightarrow (RGB\sigma)$$



Slide credit: Vincent Sitzmann

Gaussian splatting parameterizes the scene sparsely, only where density is nonzero.

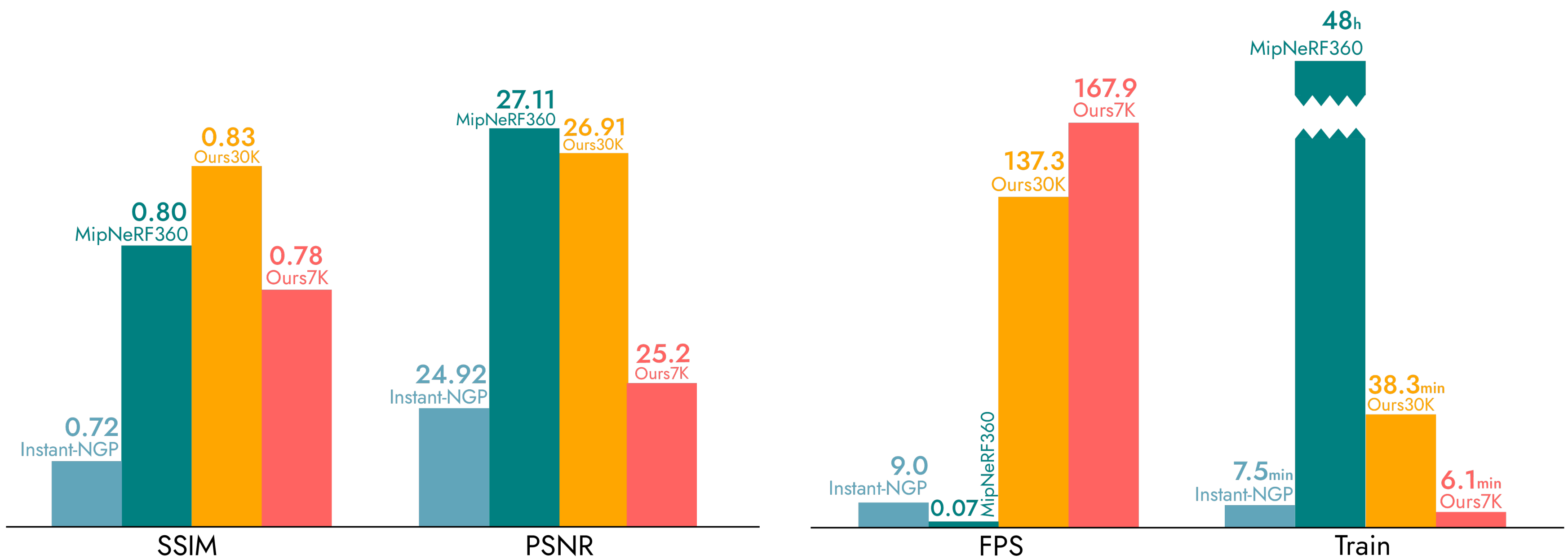


Slide credit: Vincent Sitzmann

Reconstruction Using 3DGS

Kerbl et al. ACM TOG (SIGGRAPH) 2023

Quality & Efficiency



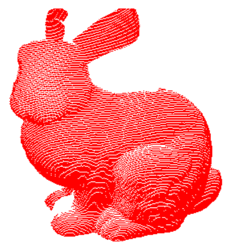
Kerbl et al. ACM TOG (SIGGRAPH) 2023

Shape Representations

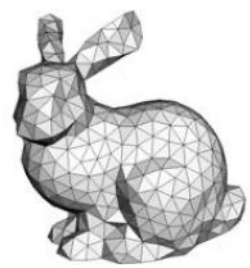
Explicit

Implicit

Non-parametric

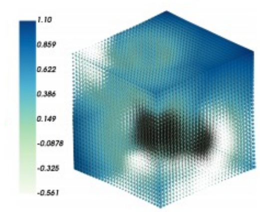


Points



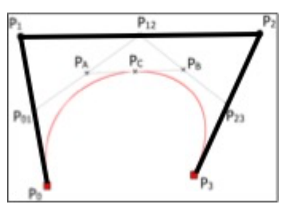
Meshes

Voxels

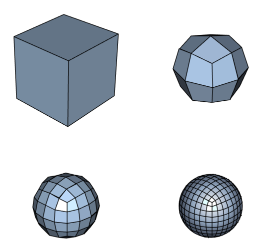


Level Sets

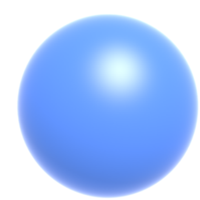
Parametric



Splines

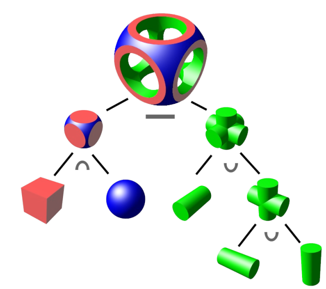


Subdivision Surfaces



$x^2+y^2+z^2=1$

Algebraic Surfaces

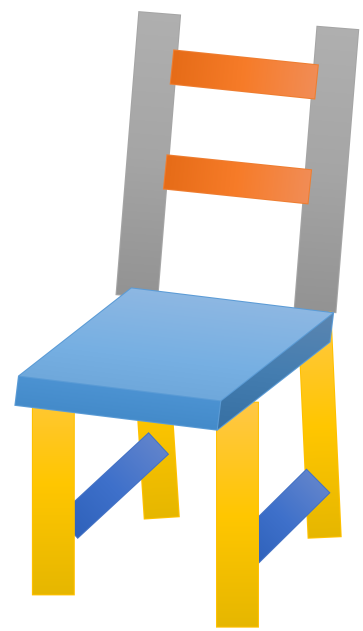


Constructive Solid Geometry

Anatomy of a Structure-Aware Representation

=

Element Structure

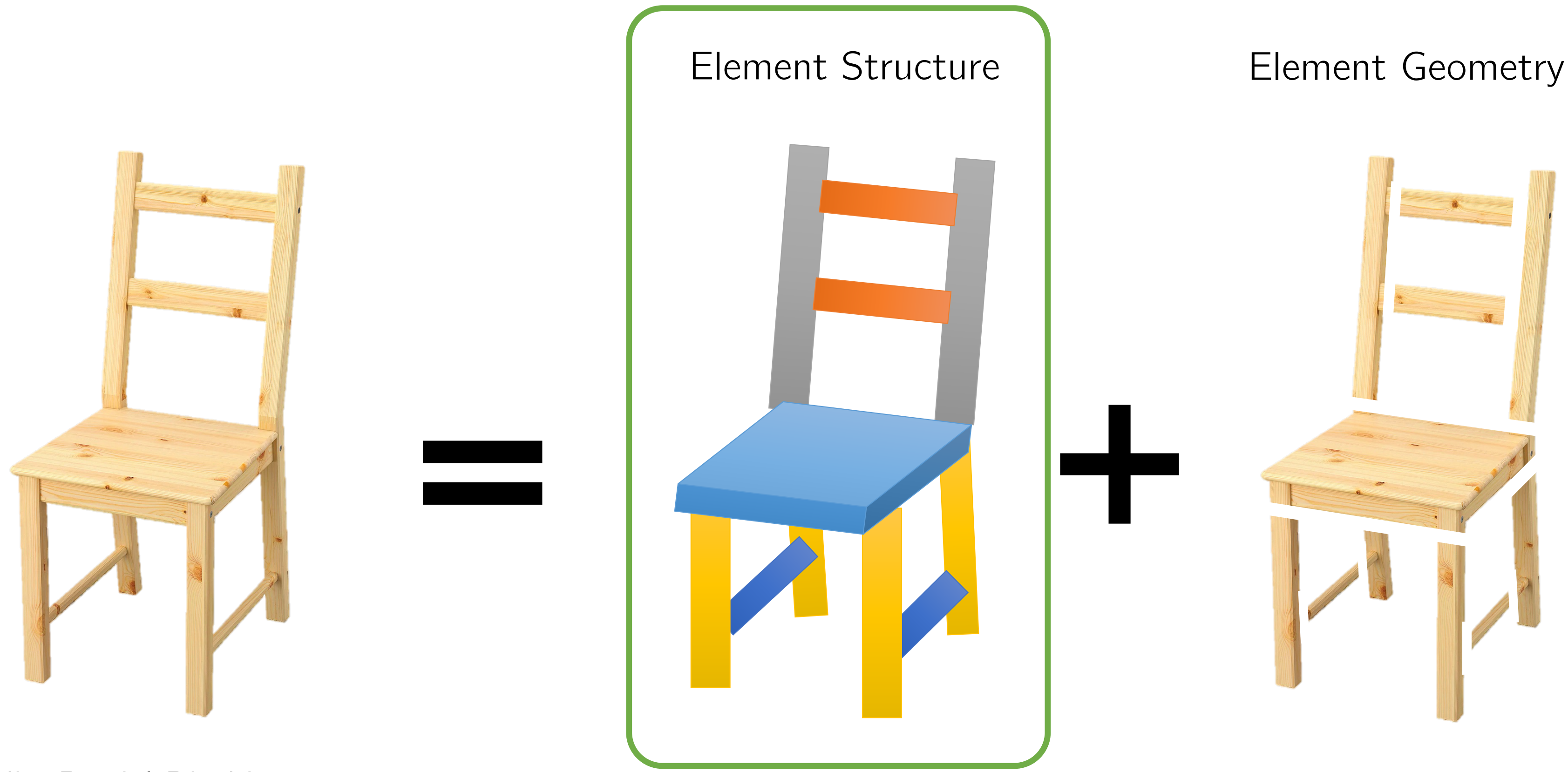


+

Element Geometry

Slide credit: Daniel Ritchie

Anatomy of a Structure-Aware Representation



Slide credit: Daniel Ritchie

Representing Element Structure

- **Segmented Geometry**

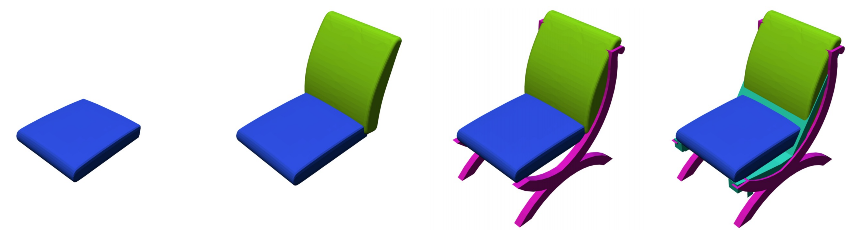
- Simple to construct
- Re-use models for unstructured geometry
- Integrity of atomic elements not guaranteed by construction (generative model must learn to output coherent segments)

[WSH∗18]
[WSCR18]

Slide credit: Daniel Ritchie

Representing Element Structure

- Segmented Geometry
- **Part Sets**

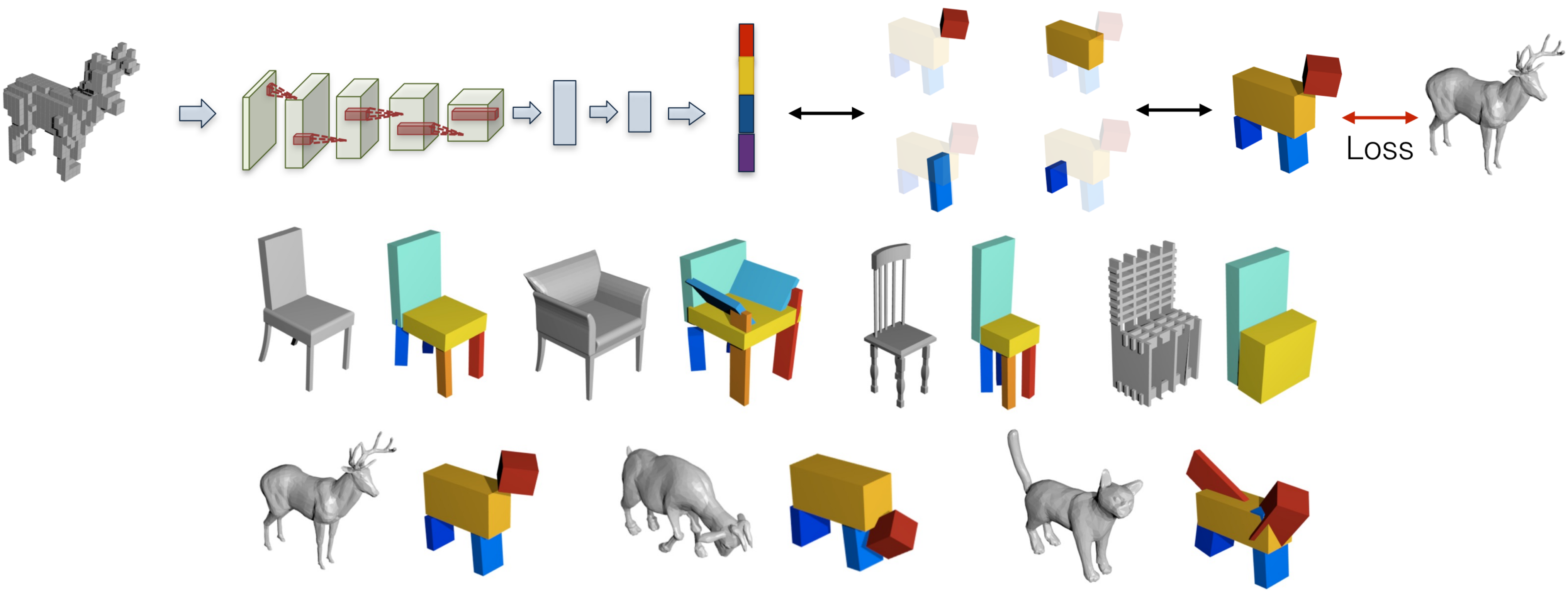


- Part integrity guaranteed
- No relationships between parts (e.g. nothing to prevent parts from "floating")

Slide credit: Daniel Ritchie

[SSK*17]

Sets of Volumetric Primitives



Tulsiani et al. CVPR 2017

Sets of Implicit Functions

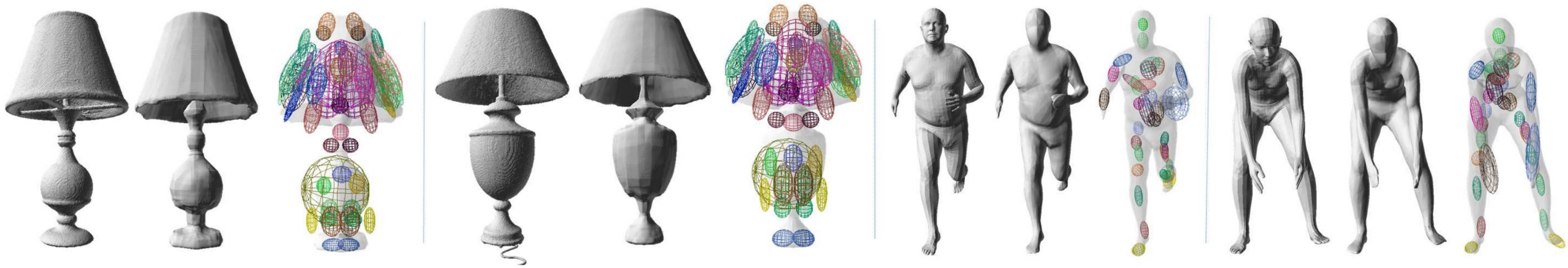
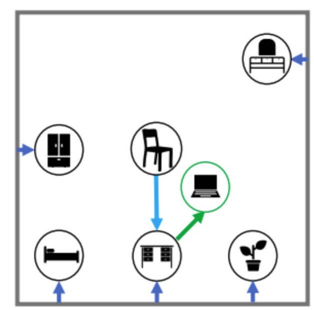


Figure 1. This paper introduces Local Deep Implicit Functions, a 3D shape representation that decomposes an input shape (mesh on left in every triplet) into a structured set of shape elements (colored ellipses on right) whose contributions to an implicit surface reconstruction (middle) are represented by latent vectors decoded by a deep network. Project video and website at ldif.cs.princeton.edu.

Genova et al. CVPR 2020

Representing Element Structure

- Segmented Geometry
- Part Sets
- **Relationship Graphs**



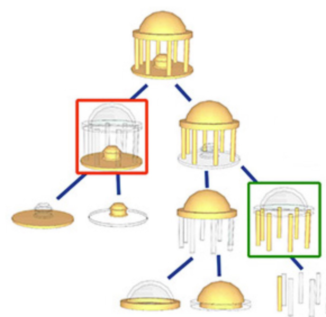
- Can enforce important relationships (e.g. connectivity)
- In general, machine learning models for graph generation still an open problem

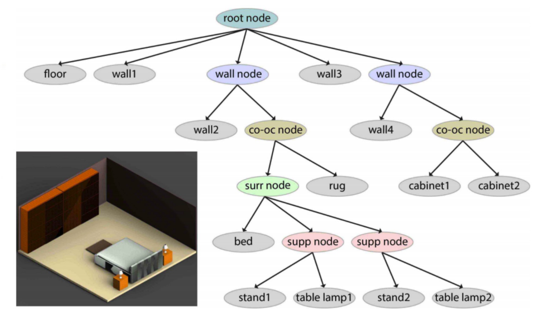
Slide credit: Daniel Ritchie

[WLW*19]

Representing Element Structure

- Segmented Geometry
- Part Sets
- Relationship Graphs
- **Hierarchies**





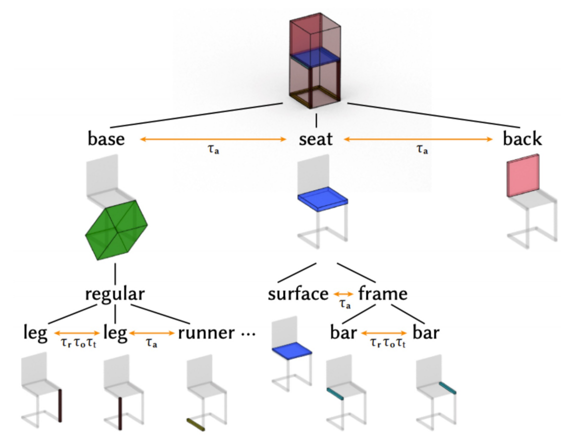
- Tree generative models better understood than graph generative models
- Not all structures of interest can be (naturally) expressed as trees

[WXL∗11]
[LPX*10]

Slide credit: Daniel Ritchie

Representing Element Structure

- Segmented Geometry
- Part Sets
- Relationship Graphs
- Hierarchies
- **Hierarchical Graphs**

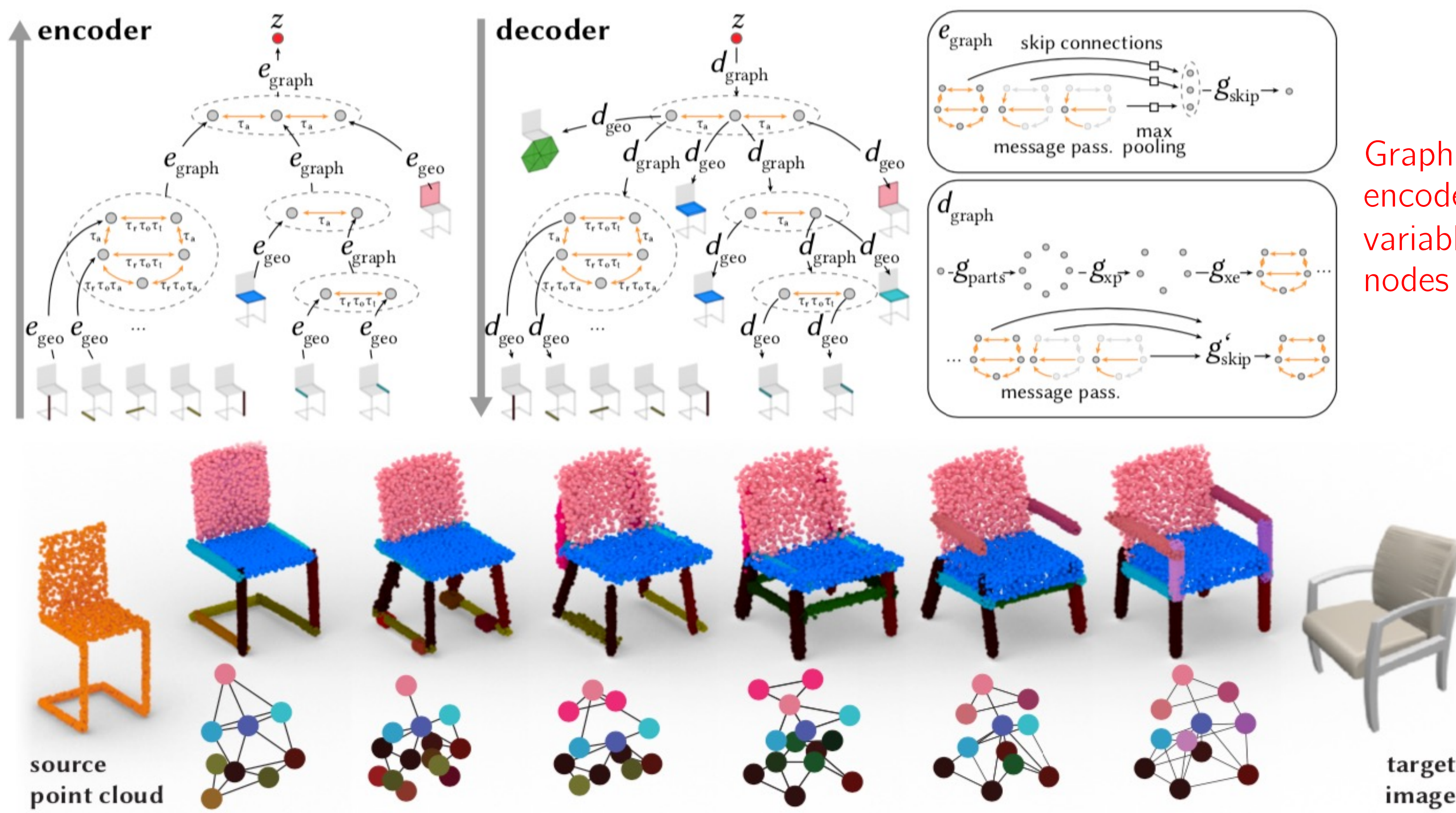


- Models both naturally hierarchical structure as well as naturally lateral relationships
- Graphs per level are simpler → easier to generate than large, general-purpose graphs
- Difficult to obtain / expensive to annotate data in this format

[MGY*19a]

Slide credit: Daniel Ritchie

Hierarchical Graph of Shape Primitives

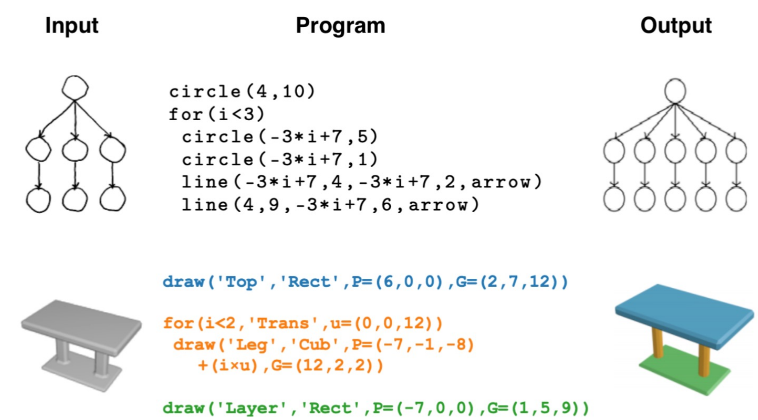


Graph convnets encode/decode variable-degree nodes

StructureNet. Mo et al. 2019

Representing Element Structure

- Segmented Geometry
- Part Sets
- Relationship Graphs
- Hierarchies
- Hierarchical Graphs
- **Programs**



- Subsumes all other representations (programs can generate any of them)
- Express natural degrees of freedom via free parameters
- Even more difficult to get data in this format

[ERSLT18]
[TLS*19]

Slide credit: Daniel Ritchie