Learning to generate 3D shapes

Subhransu Maji

College of Information and Computer Sciences

University of Massachusetts, Amherst

http://people.cs.umass.edu/smaji

August 10, 2018 @ Caltech

Creating 3D shapes is not easy

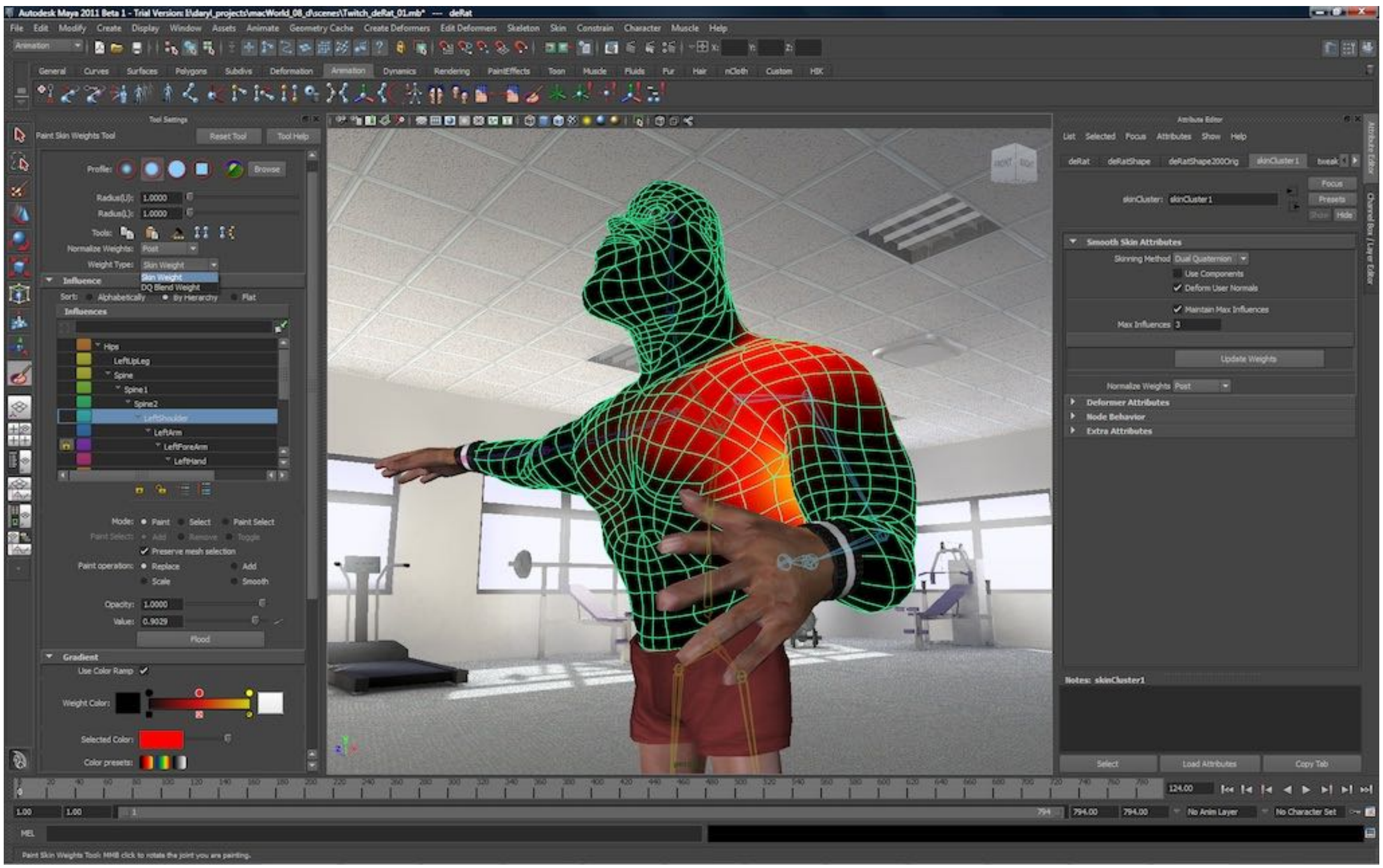
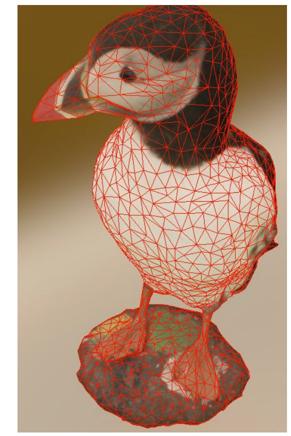


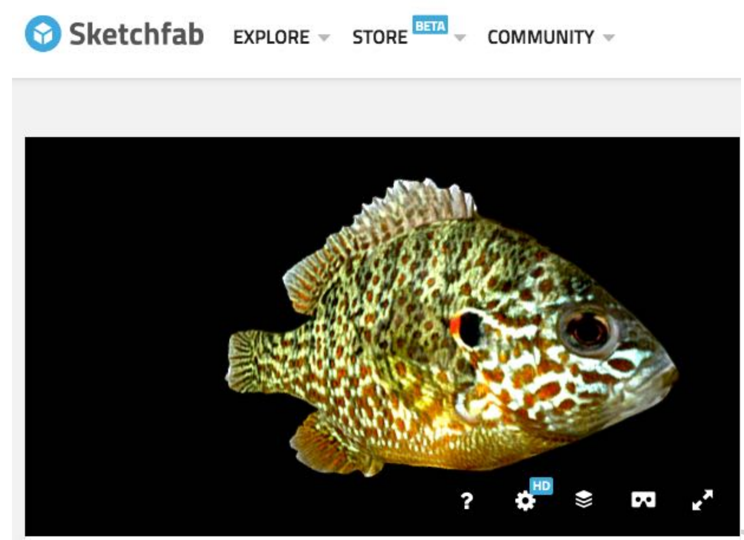
Image from Autodesk 3D Maya

Inferring 3D shapes from images

What shapes are puffins?



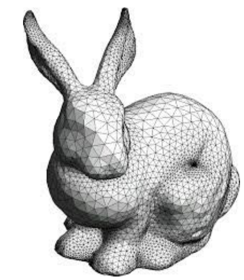
What shapes are pumpkinseed fish?

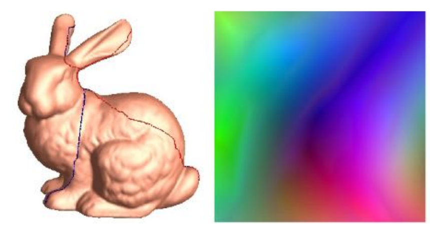


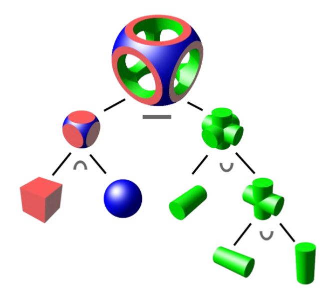
Creating 3D shapes is not easy

- Many techniques for recognizing 3D data, but relatively few techniques for generating them

- Representations for generation?
  - Voxels
  - Multiview
  - Geometry images
  - Shape basis
  - Set-based (points, triangles, etc.)
  - Procedural, e.g., constructive solid geometry







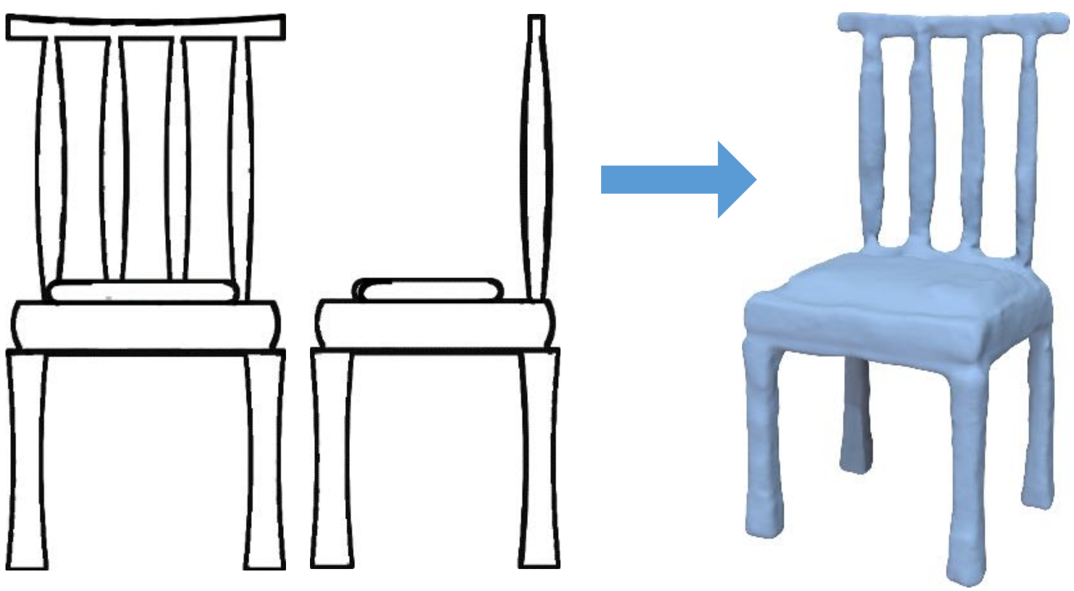
Talk overview

- Generative models for 3D shapes and applications
  - Multiview [3DV’17]
  - Multiresolution tree networks [ECCV’18]
  - Constructive solid geometry [CVPR’18]

- Learning 3D shapes with weak supervision [3DV’17]

Part 1

3D Shape Reconstruction from Sketches via Multi-view Convolutional Networks



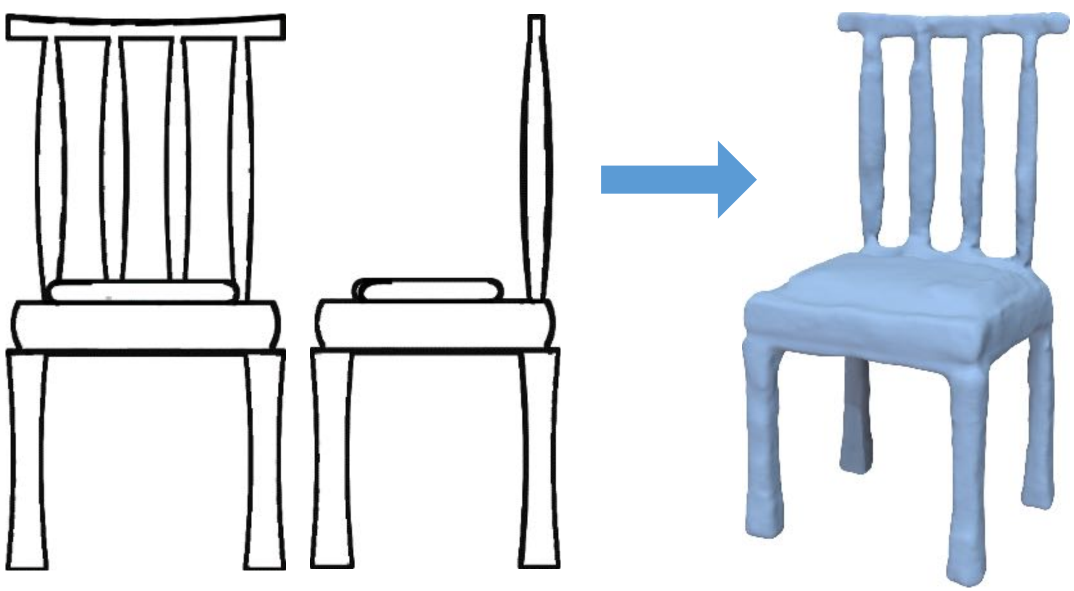
Zhaoliang Lun
Matheus Gadelha
Evangelos Kalogerakis
Subhransu Maji
Rui Wang

3DV 2017

Part 1

3D Shape Reconstruction from Sketches via Multi-view Convolutional Networks

Zhaoliang Lun
Matheus Gadelha
Evangelos Kalogerakis
Subhransu Maji
Rui Wang



3DV 2017

What Does the Occluding Contour Tell Us about Solid Shape?

Jan J Koenderink

First Published June 1, 1984 | Research Article

Why line drawings? Simple & intuitive medium to convey shape!

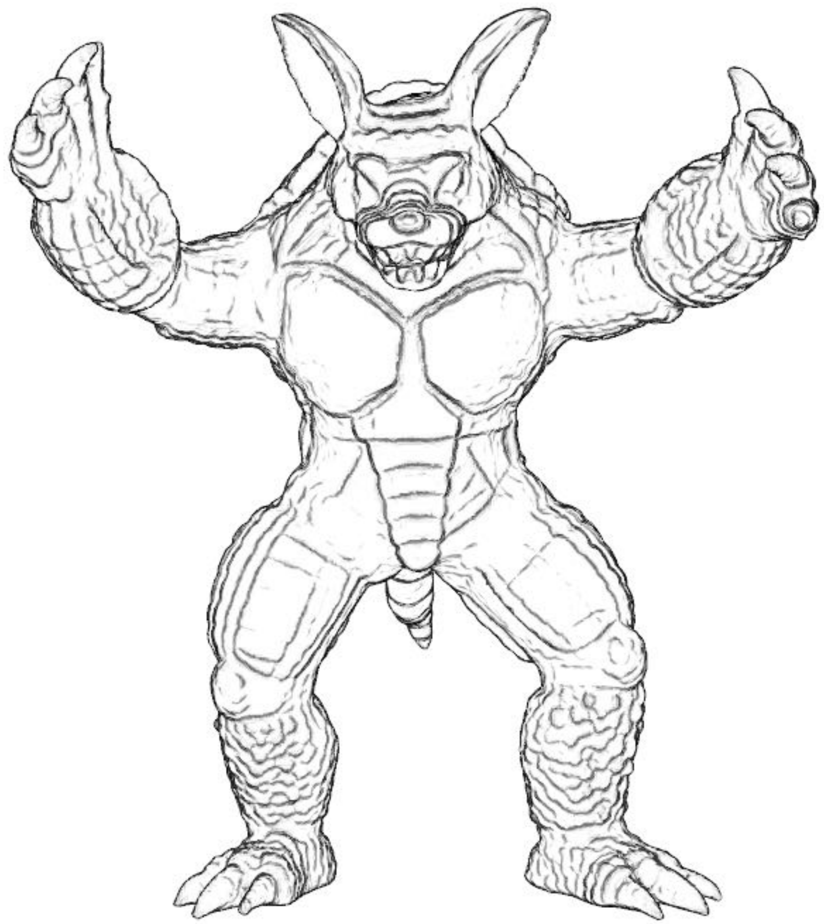
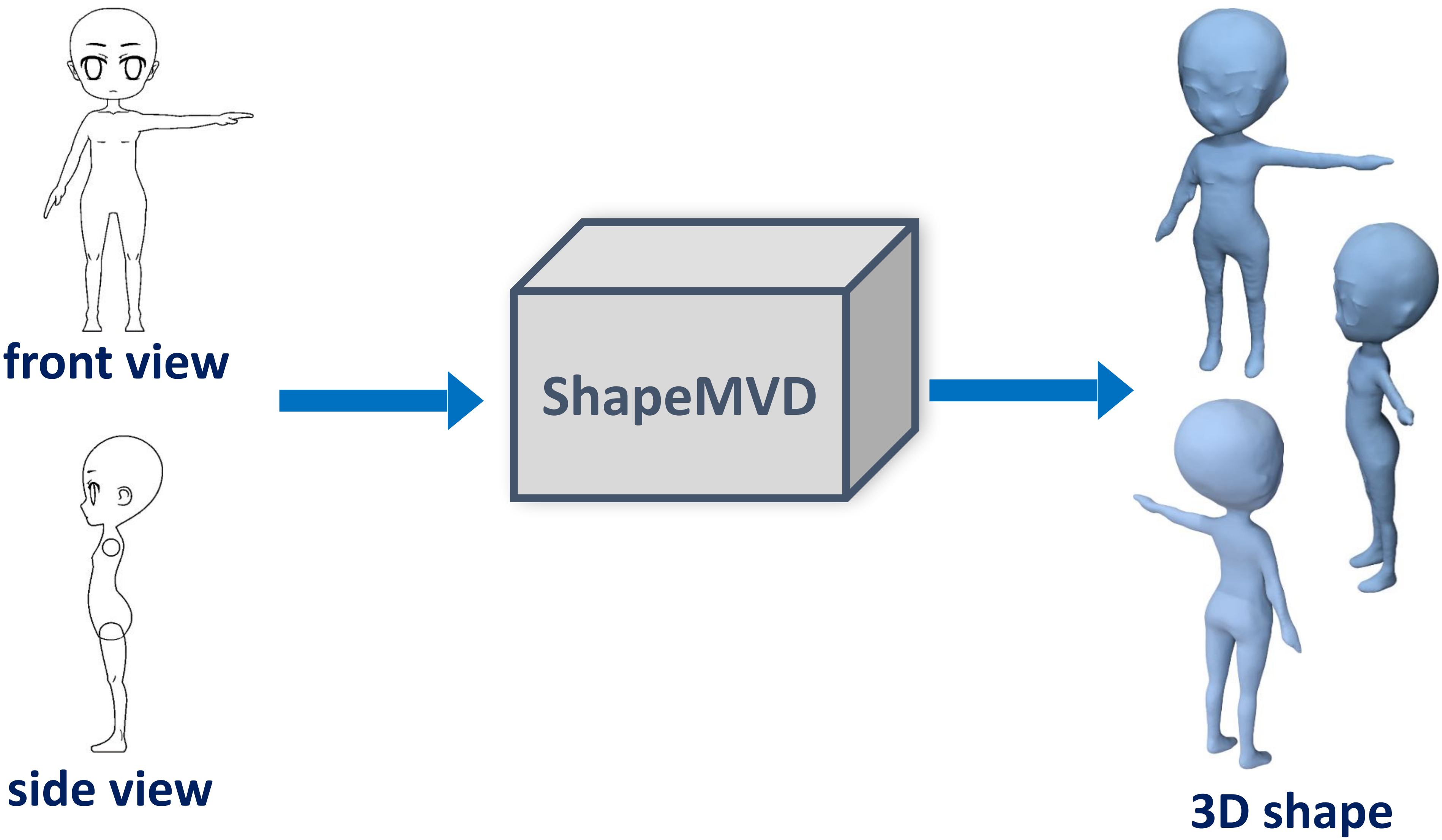


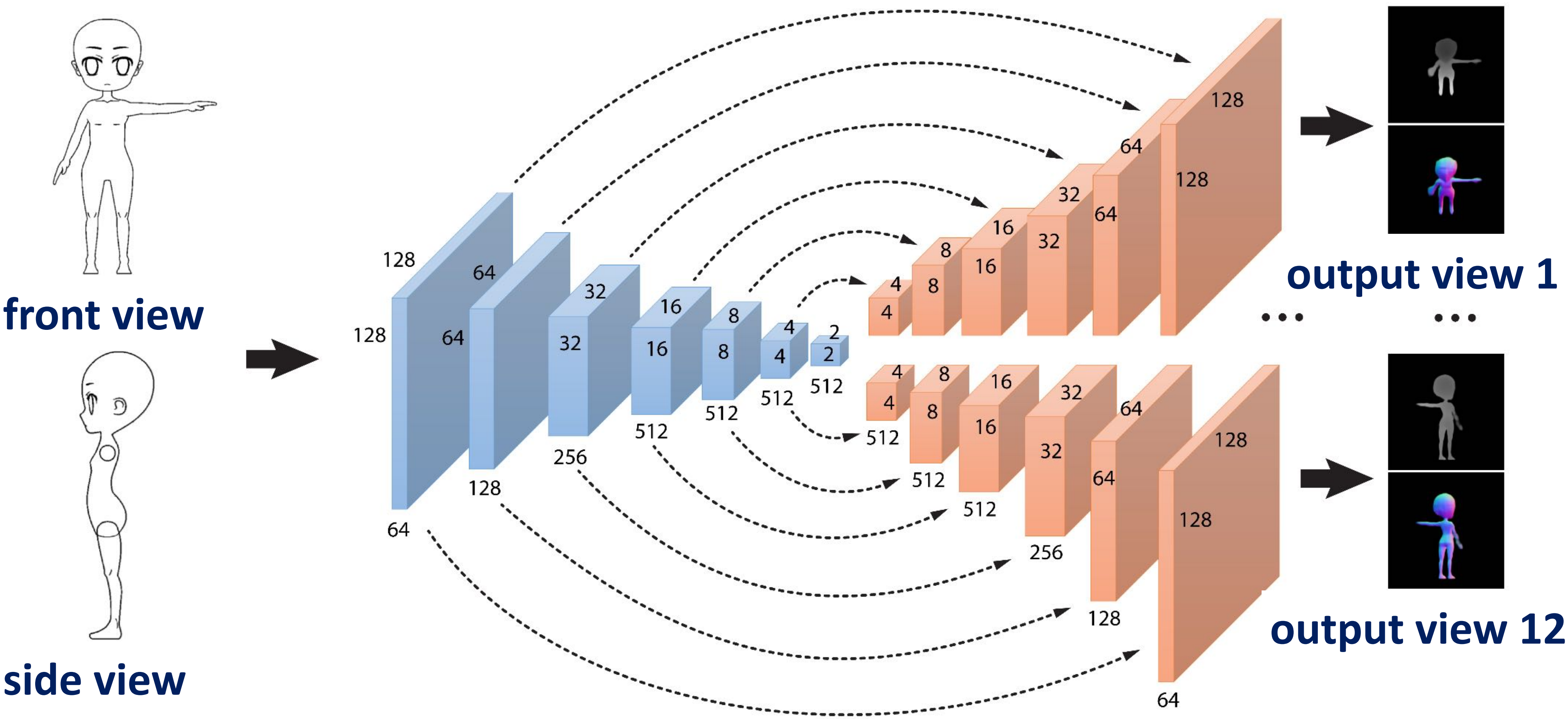
Image from Suggestive Contour Gallery, DeCarlo et al. 2003

Goal: 2D line drawings in, 3D shapes out!



Deep net architecture: U-net structure

Feature representations in the decoder depend on **previous layer & encoder's corresponding layer**



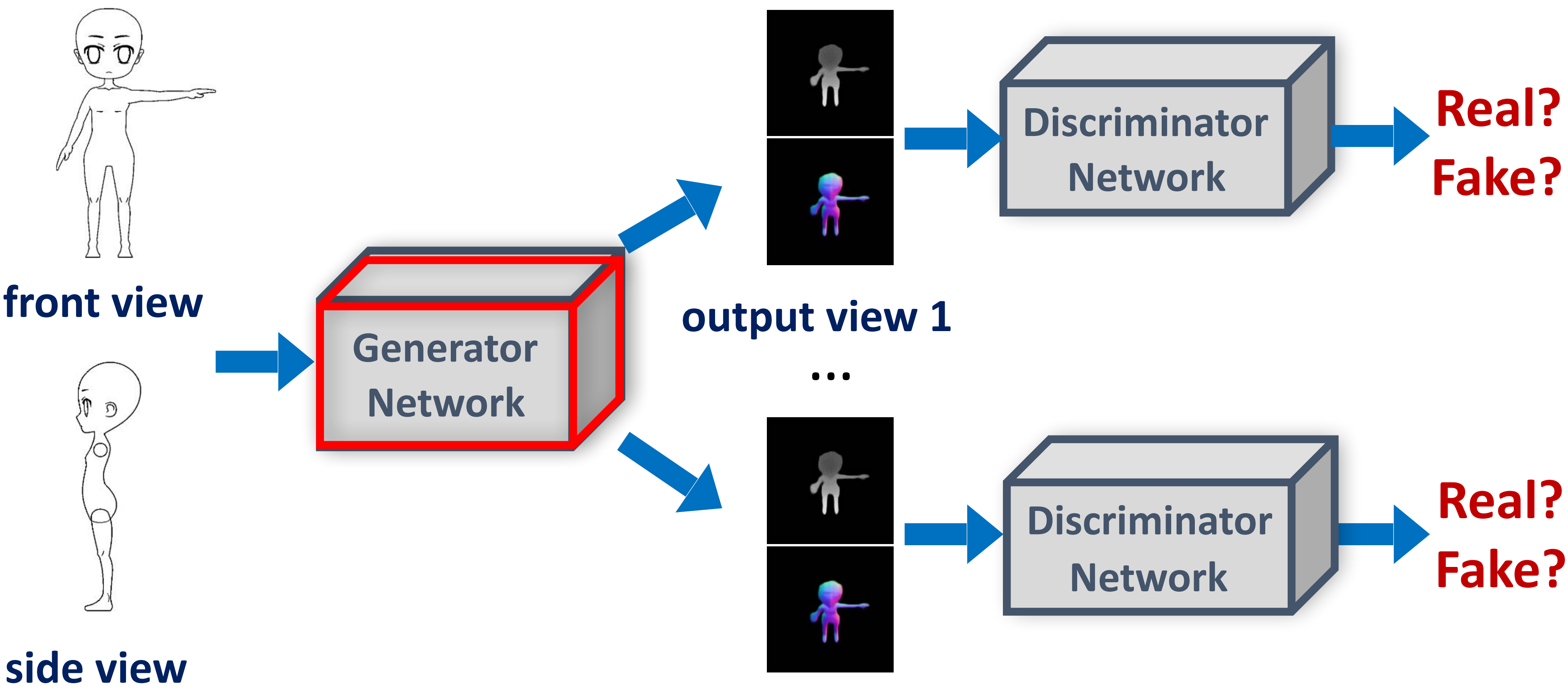
U-net: Ronneberger et al. 2015,
Isola et al. 2016

Training: full loss

Penalize per-pixel depth reconstruction error: $\sum_{pixels} | d_{pred} - d_{gt} |$

& per-pixel normal reconstruction error: $\sum_{pixels} (1 - n_{pred} \cdot n_{gt})$

& "unreal" outputs: $-\log P(real)$



cGAN: Isola et al. 2016

Training data

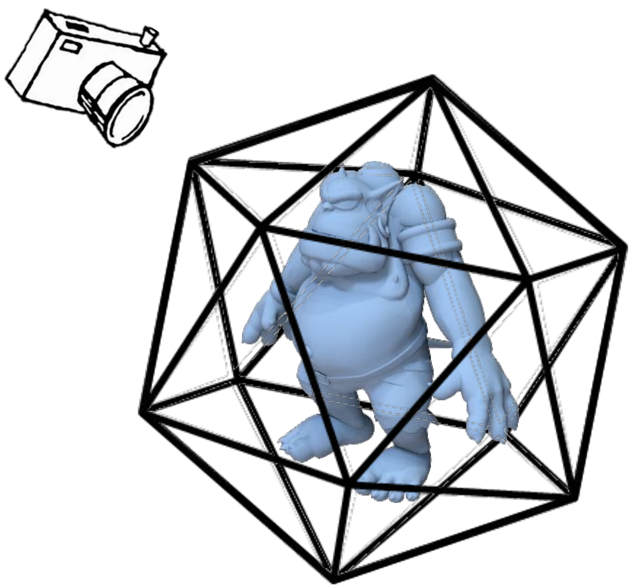
Character
10K models

Chair
10K models

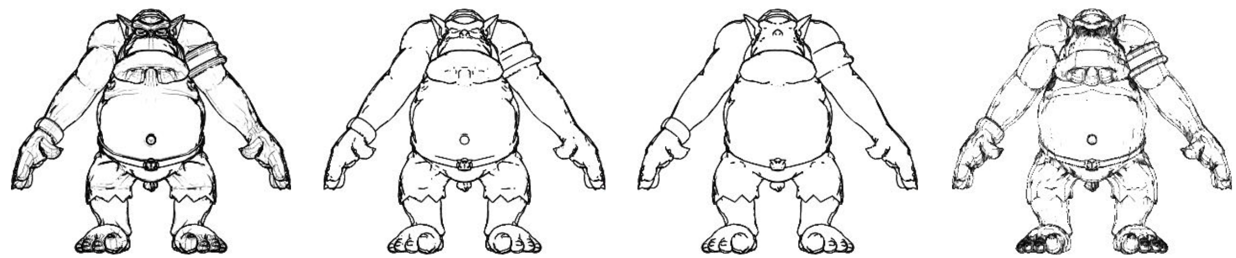
Airplane
3K models

Models from “The Models Resource” &
3D Warehouse

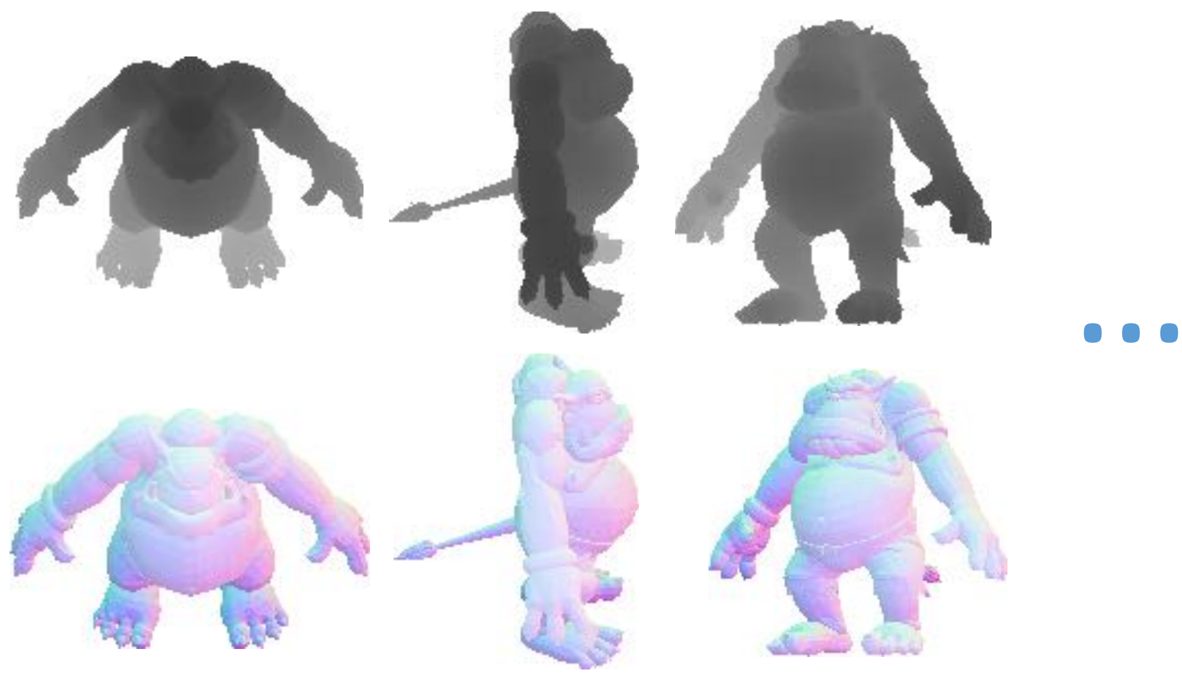
Training data



12 views



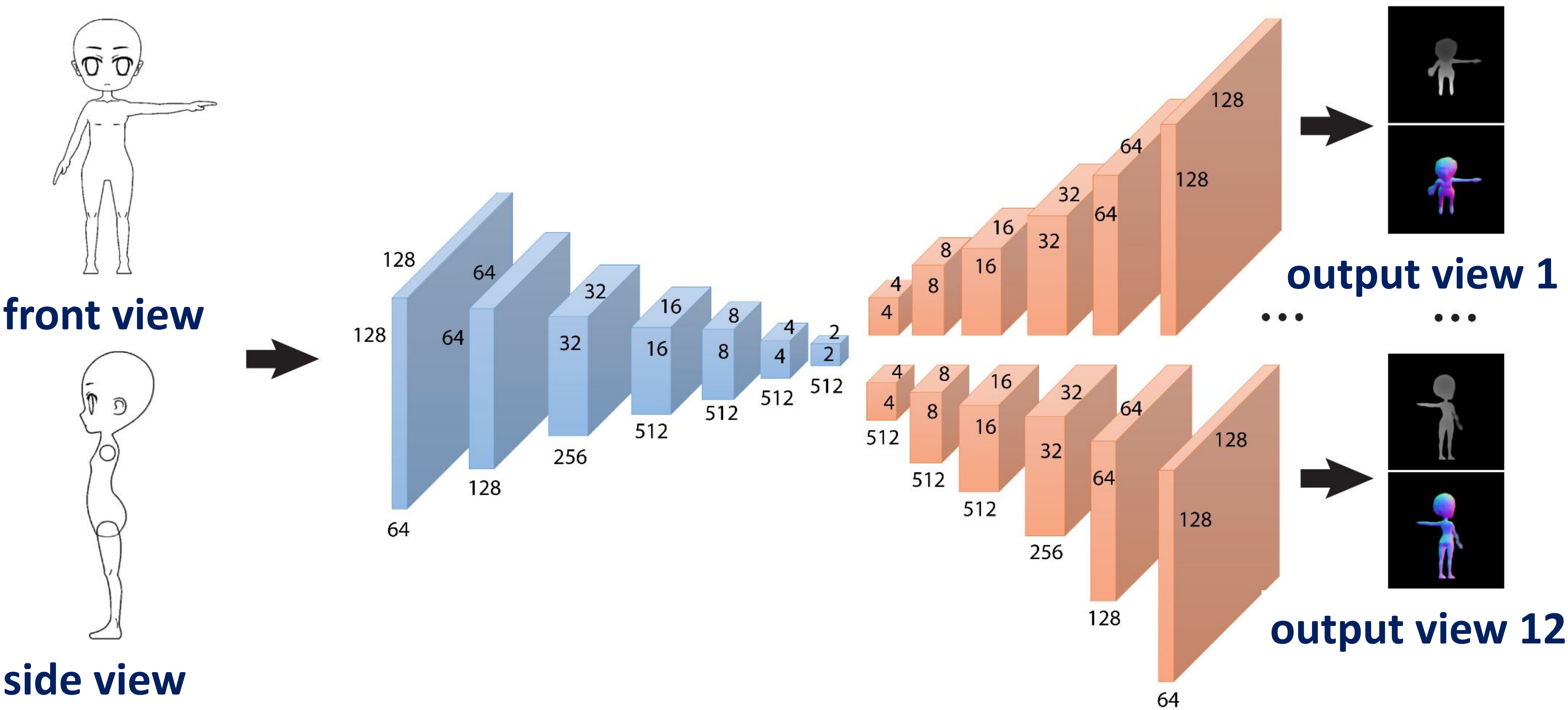
Synthetic line drawings



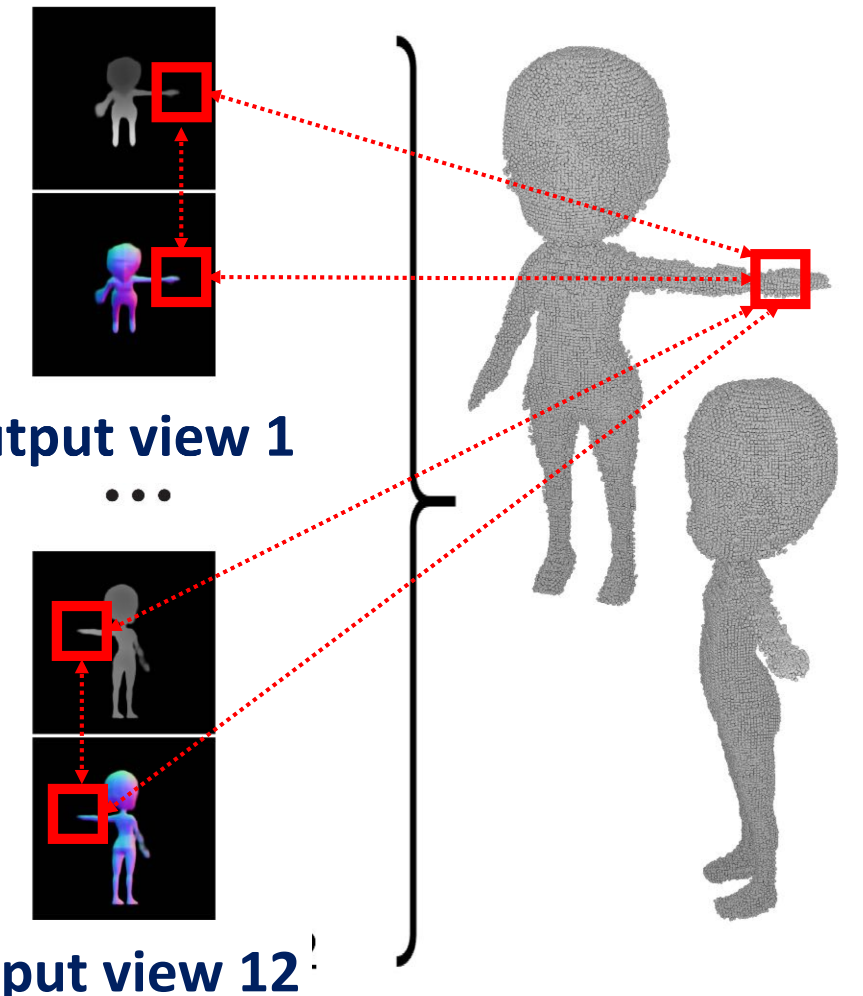
Training depth and normal maps

Test time

Predict multi-view depth and normal maps!



Multi-view depth & normal map fusion



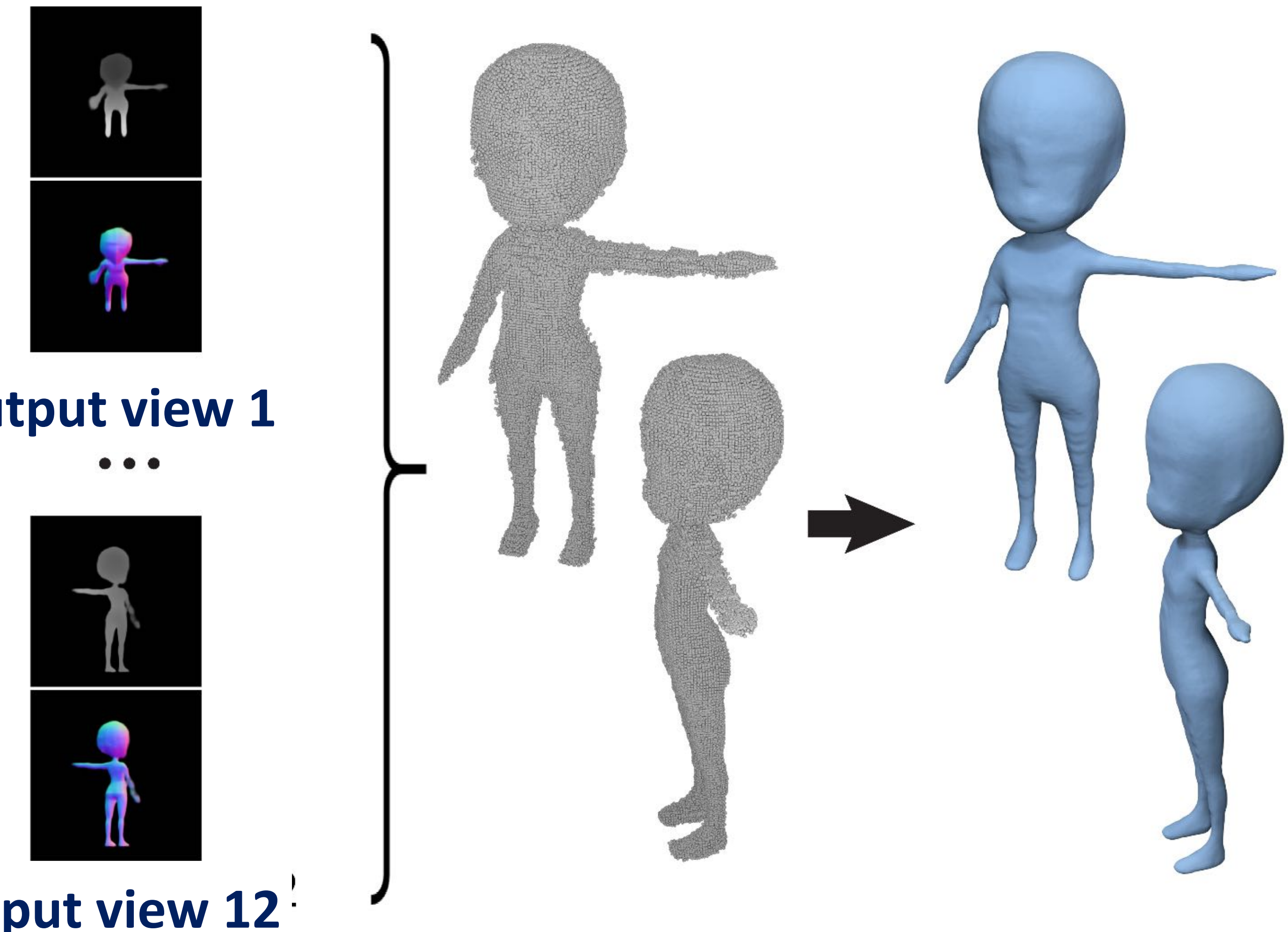
Multi-view depth & normal maps

Consolidated point cloud

Optimization problem

- Depth derivatives should be consistent with normals
- Corresponding depths and normals across different views should agree

Surface reconstruction

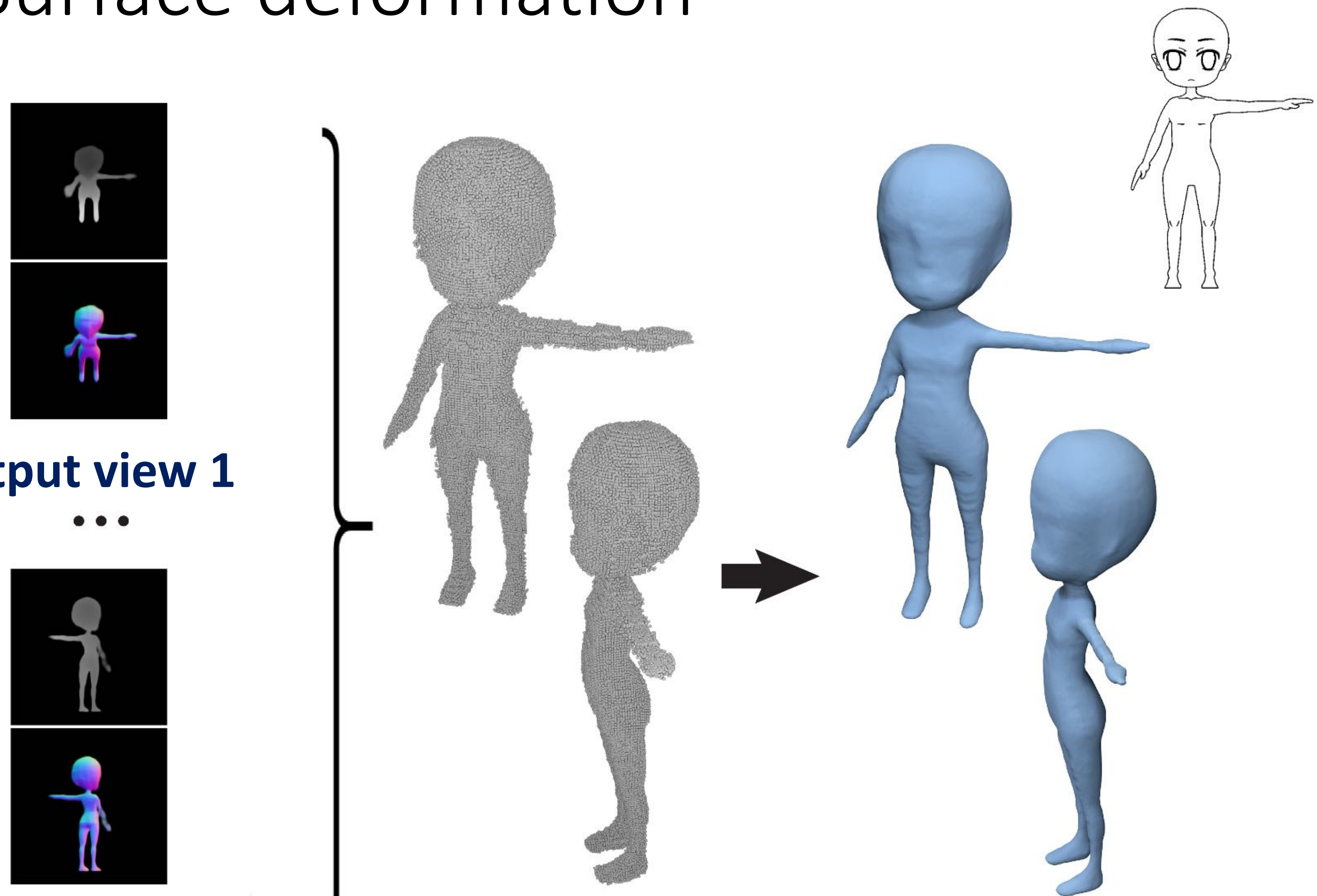


Multi-view depth & normal maps

Consolidated point cloud

Surface reconstruction
[Kazhdan et al. 2013]

Surface deformation

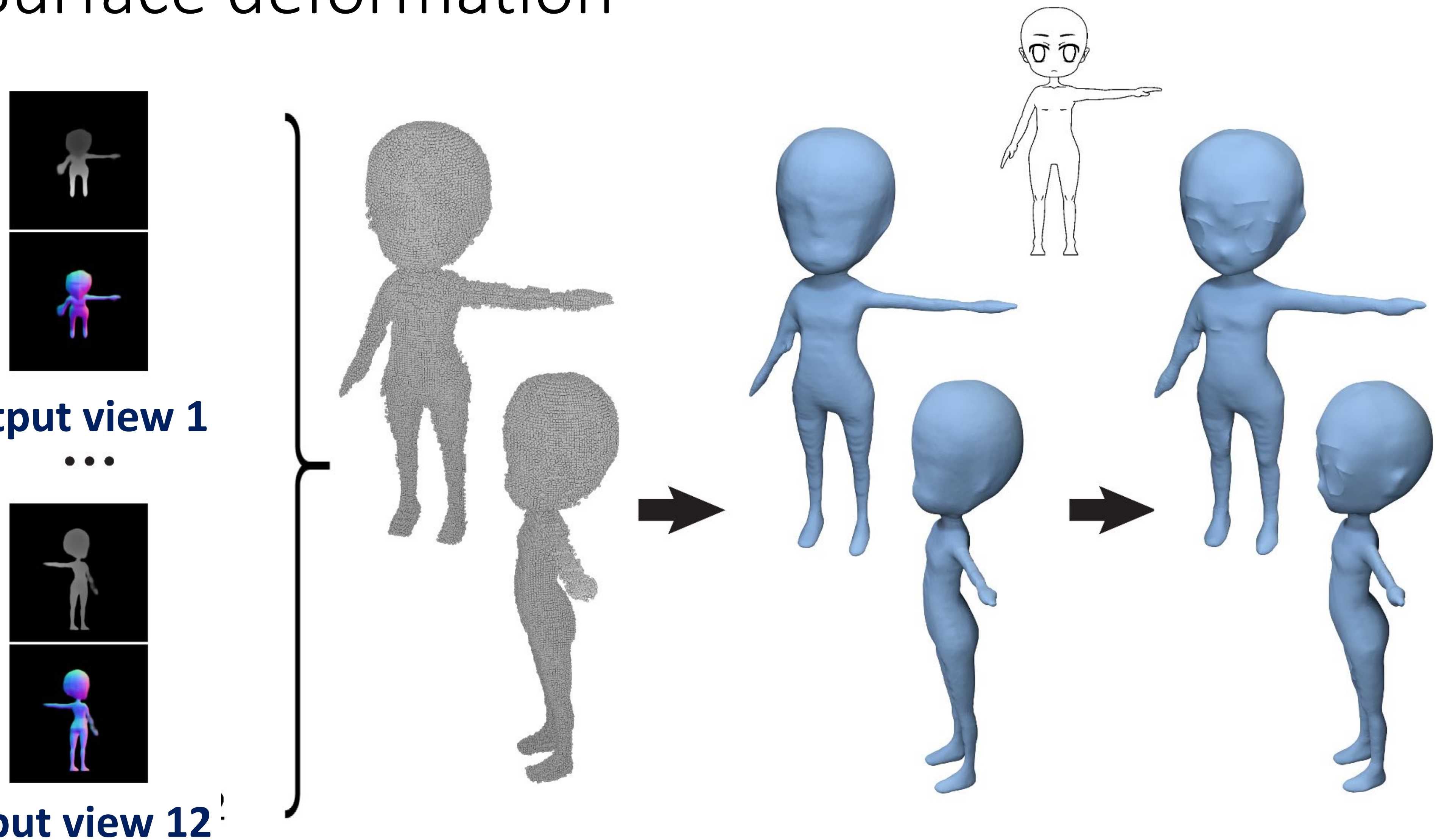


Multi-view depth & normal maps

Consolidated point cloud

Surface reconstruction
[Kazhdan et al. 2013]

Surface deformation



Multi-view depth & normal maps

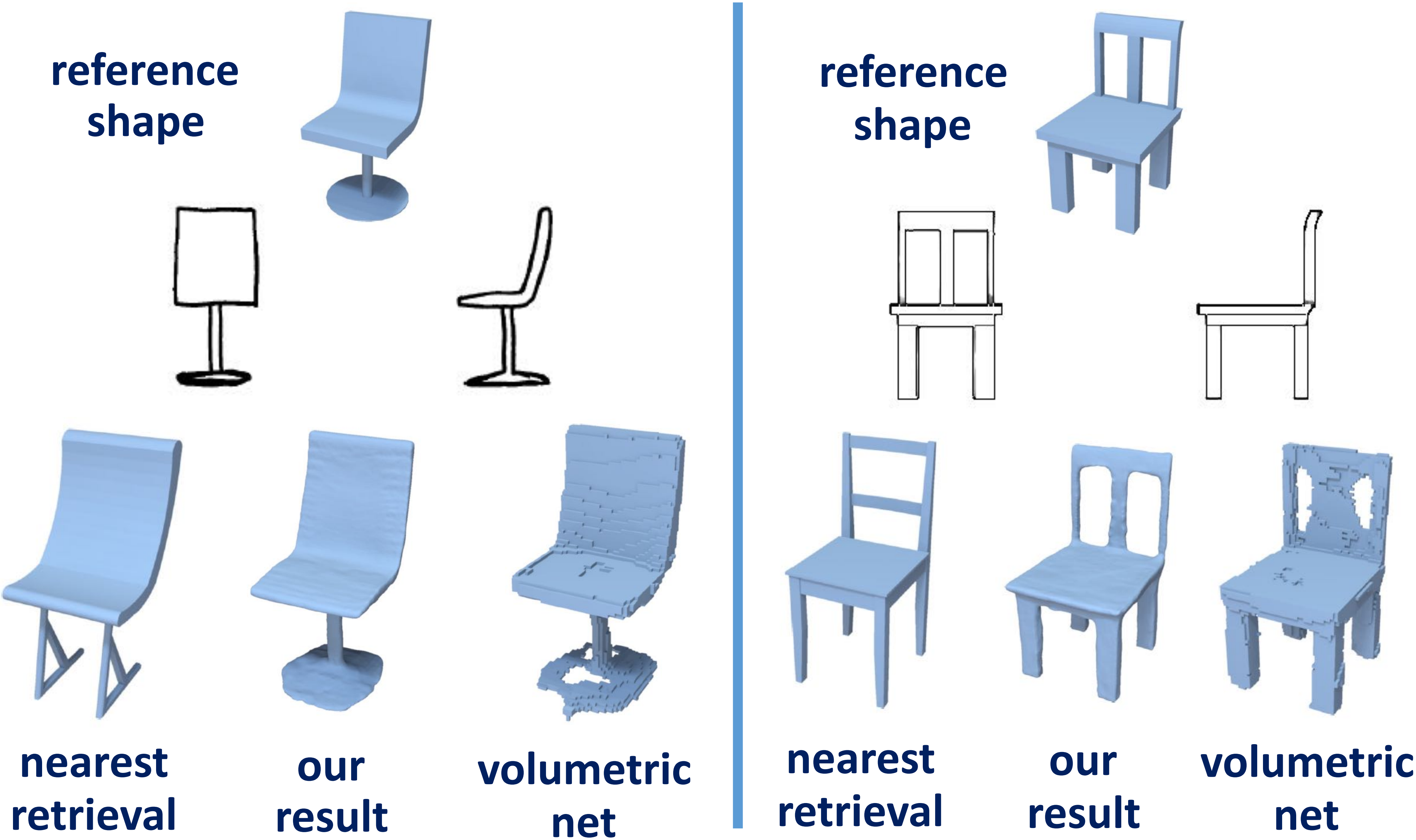
Consolidated point cloud

Surface reconstruction [Kazhdan et al. 2013]

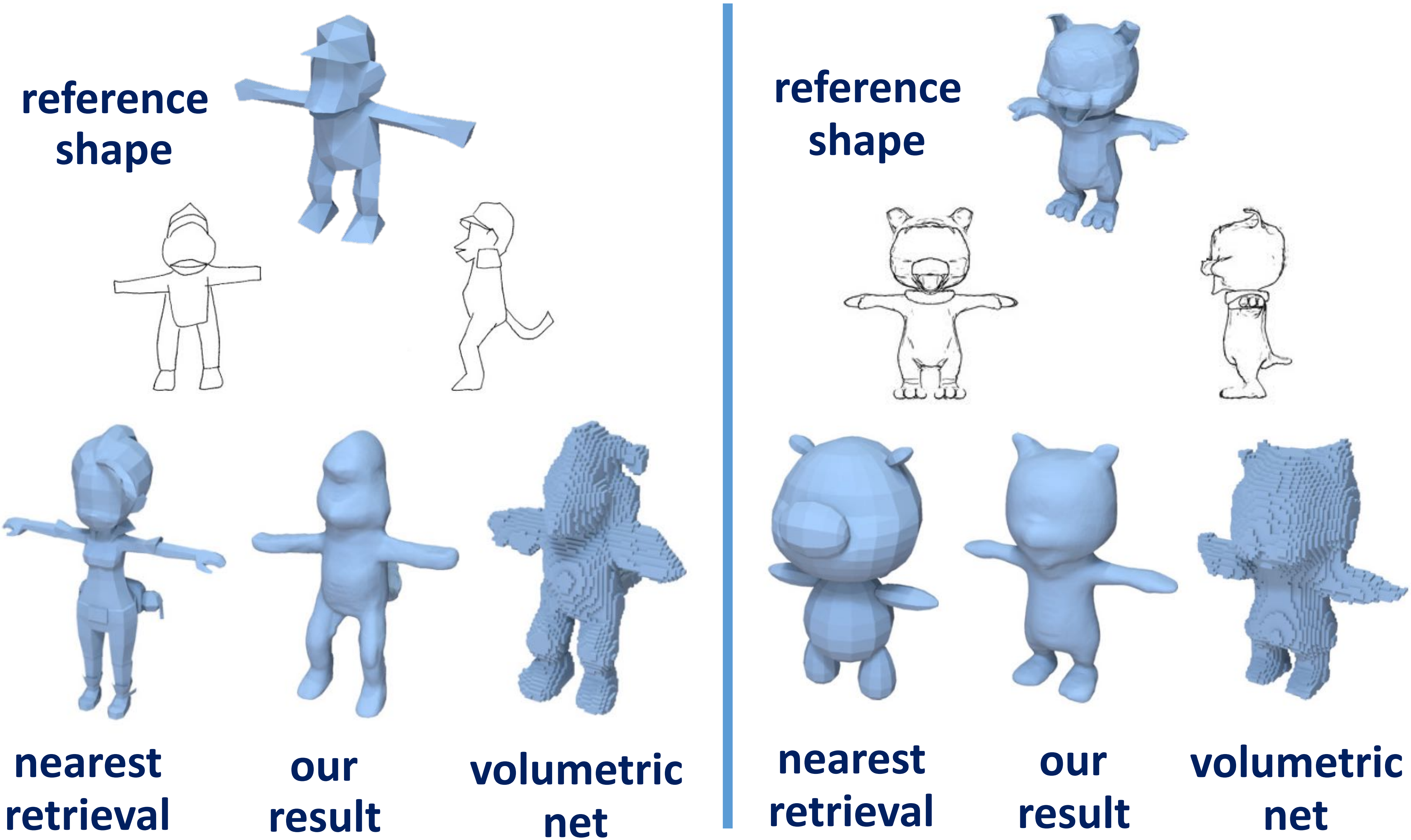
Surface "fine-tuning" [Nealen et al. 2005]

Experiments

Qualitative Results



Qualitative Results



Quantitative Results

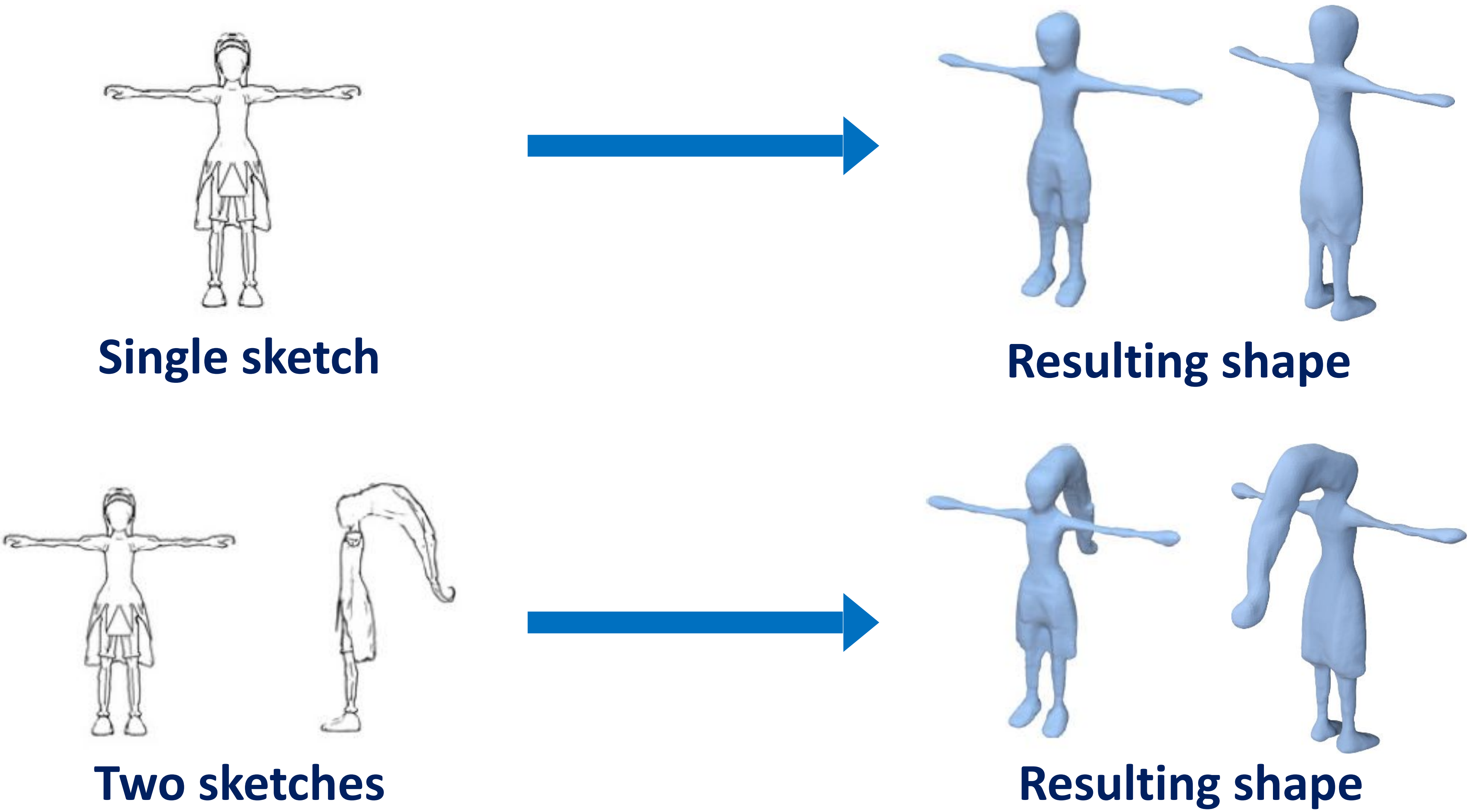
Character (human drawing)

Metric	Our method	Volumetric	NN
Hausdorff distance	**0.120**	0.638	0.242
Chamfer distance	**0.023**	0.052	0.045
normal distance	**34.27**	56.97	47.94
volumetric distance	**0.309**	0.497	0.550

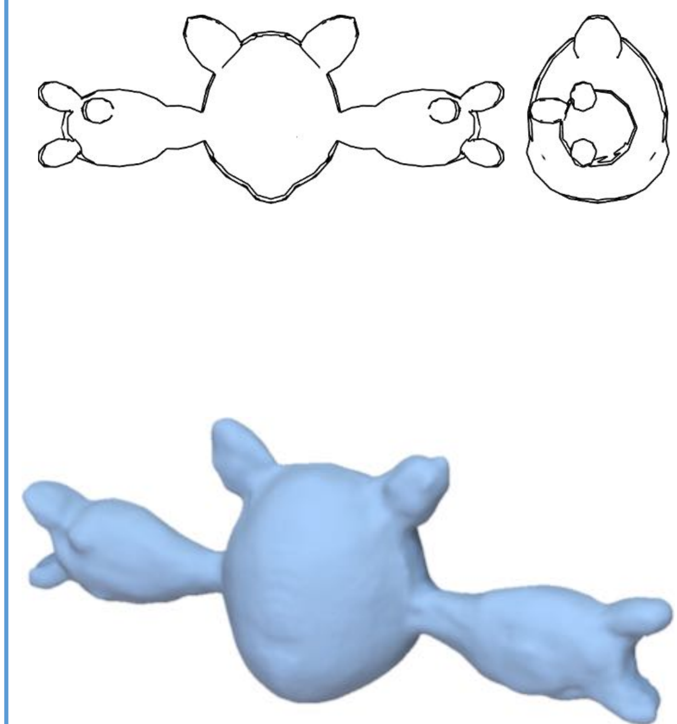
Man-made (human drawing)

Hausdorff distance	**0.171**	0.211	0.228
Chamfer distance	**0.028**	0.032	0.038
normal distance	**34.19**	48.81	43.75
volumetric distance	**0.439**	0.530	0.560

Single vs two input line drawings



More results



Part 2

Multiresolution Tree Networks for 3D Point Cloud Processing

Matheus Gadelha
Subhransu Maji
Rui Wang

ECCV 18

Part 2

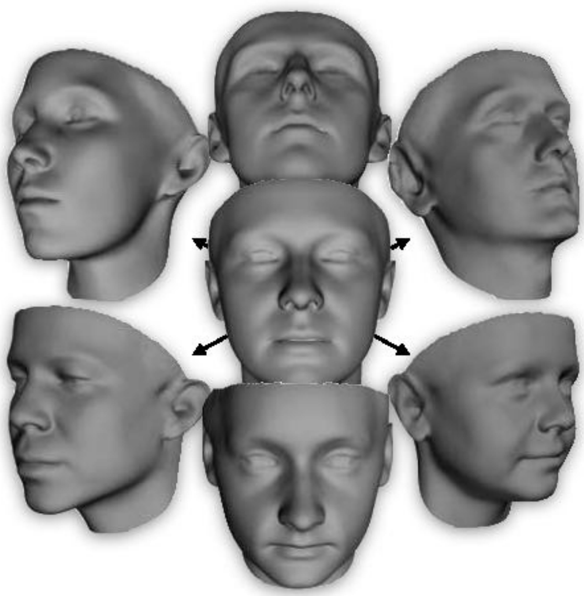
Multiresolution Tree Networks for 3D Point Cloud Processing

Matheus Gadelha
Subhransu Maji
Rui Wang

ECCV 18

Point-cloud decoders

- Global shape basis or fully-connected decoders



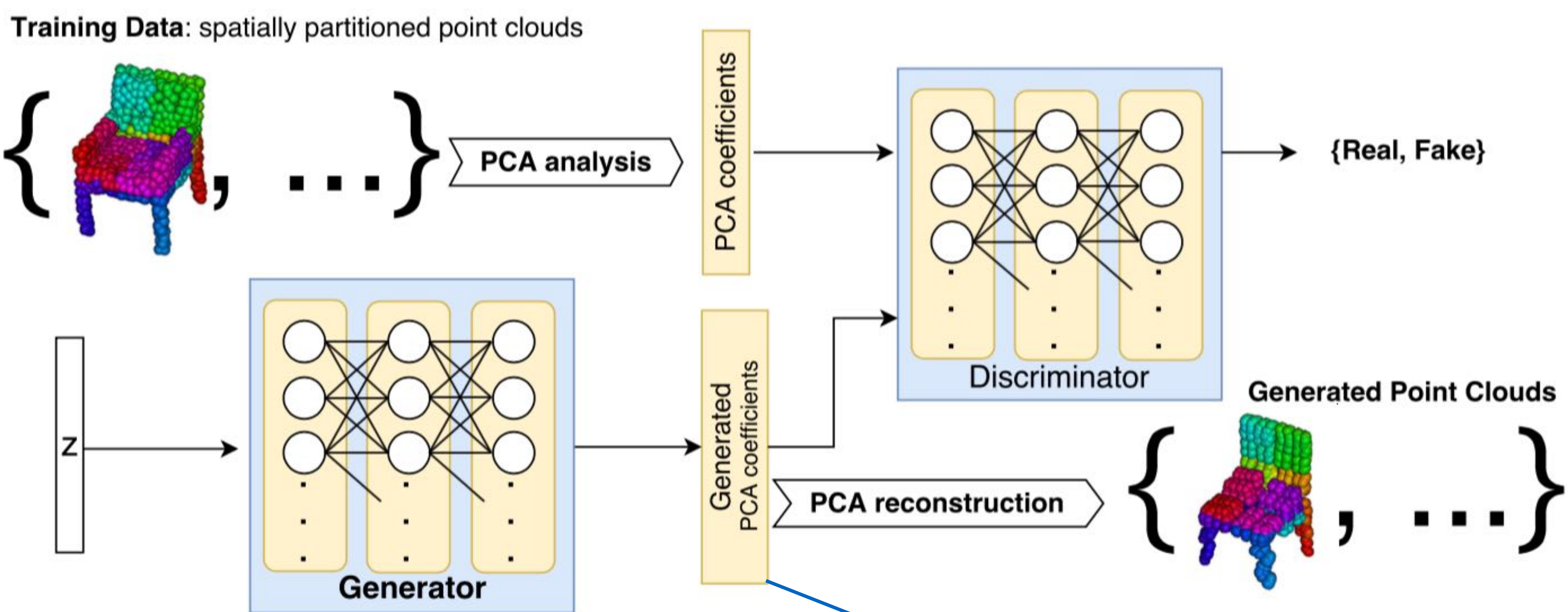
Morphable models
[Figure from Booth et al., 16]

$$\mathbf{X}^* = \boldsymbol{M} + \sum_{i=1}^{d} \alpha_i \boldsymbol{U}_i = \boldsymbol{M} + \mathbf{U}\boldsymbol{\alpha}$$

Requires perfect correspondence

Gadelha et al., BMVC 17

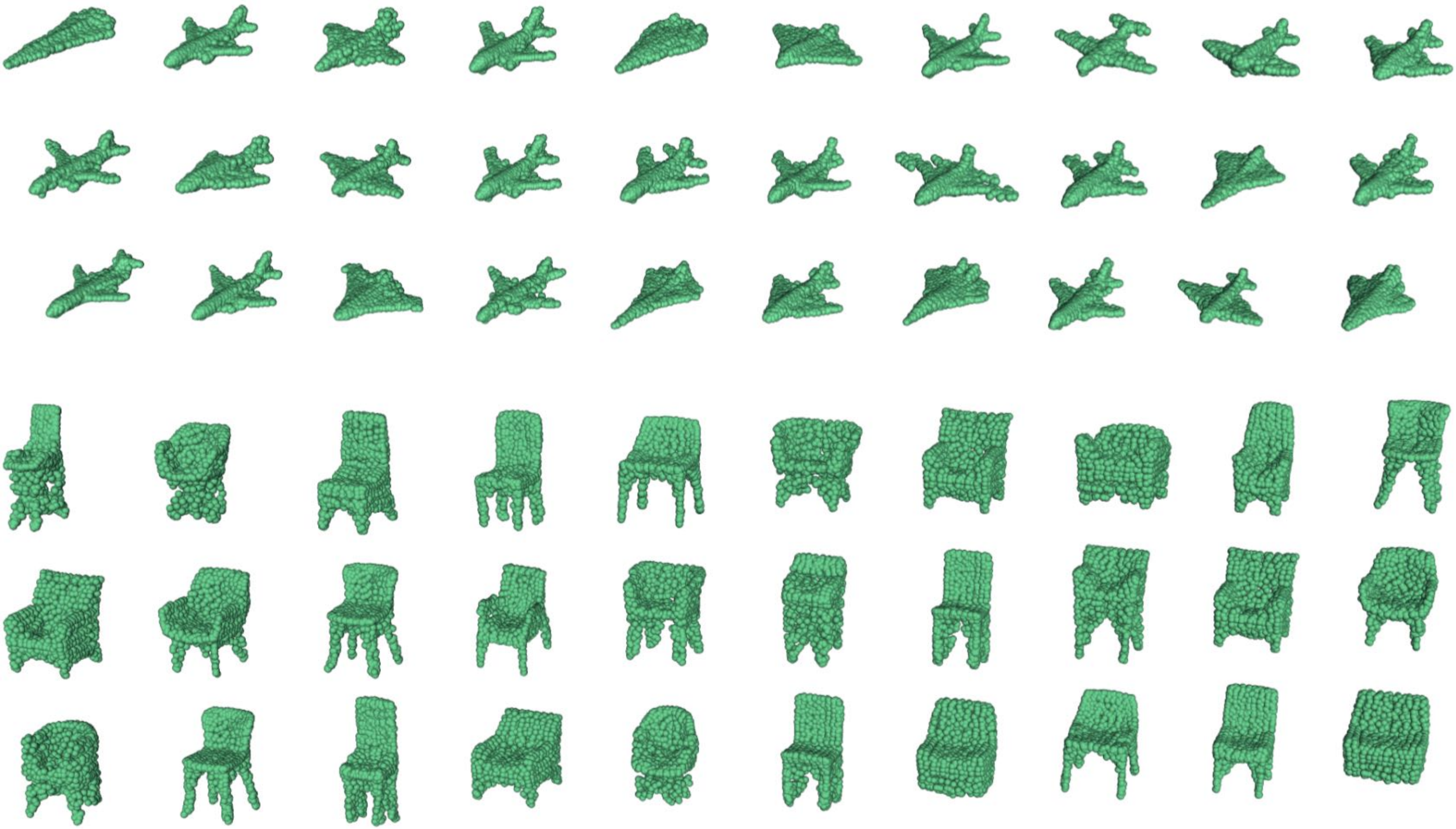
How important is correspondence?



$$\mathbf{X}^* = \boldsymbol{M} + \sum_{i=1}^{d} \alpha_i \boldsymbol{U}_i = \boldsymbol{M} + \mathbf{U}\boldsymbol{\alpha}$$

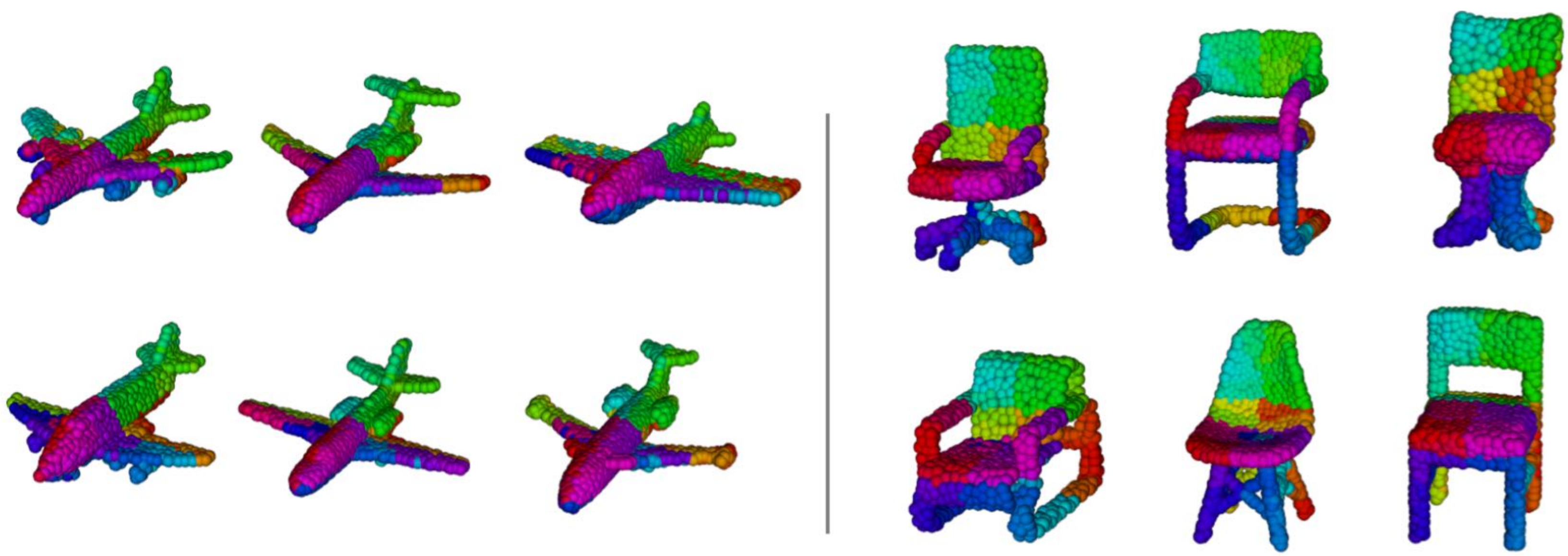
Related work: Fan et al., CVPR 2017

How important is correspondence?



Multiresolution tree networks

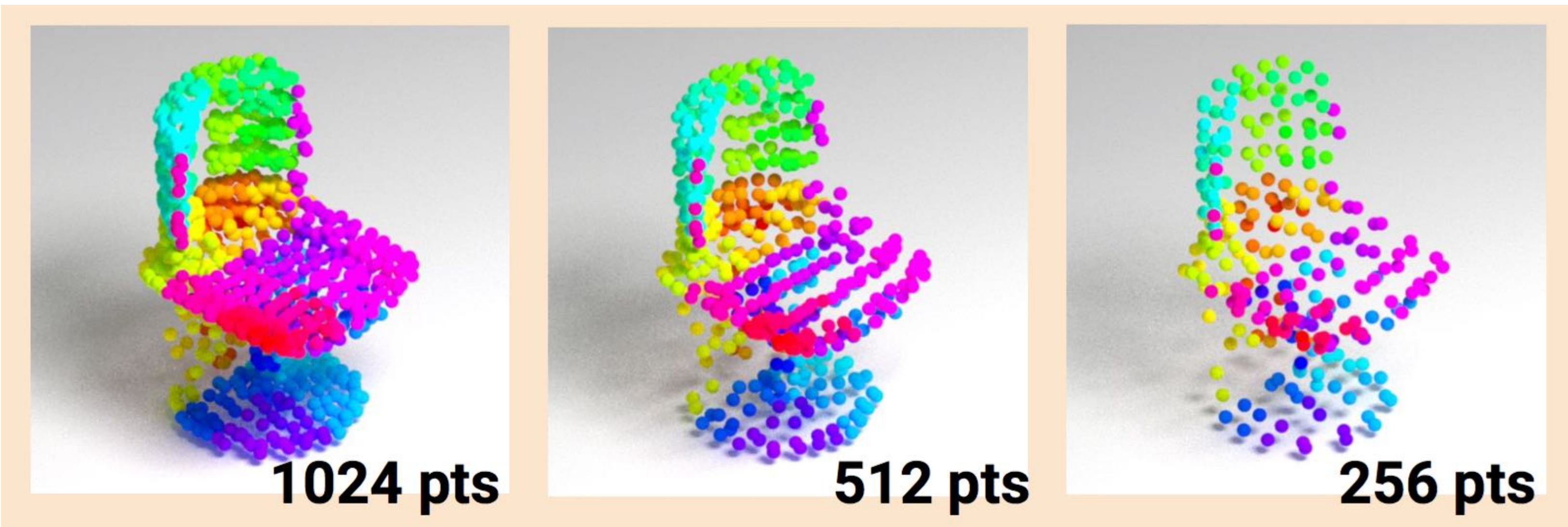
- Addresses the lack of
  - convolutional structure, and
  - coarse to fine reasoning
- Basic idea: linearize 3D points and use 1D convolutions



Points colored with kdtree sort index

Multiresolution tree networks

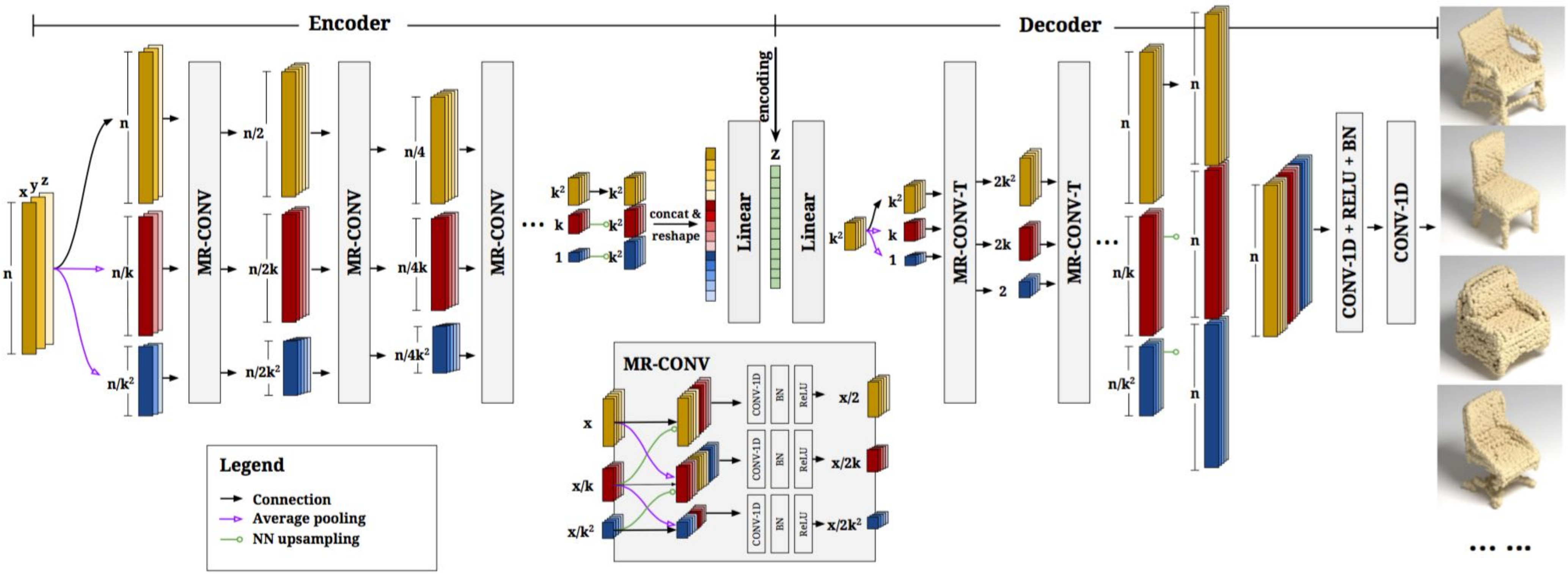
- Addresses the lack of
  - convolutional structure, and
  - coarse to fine reasoning
- Basic idea: linearize 3D points and use 1D convolutions



Implicit mulitresolution structure

Multiresolution tree networks

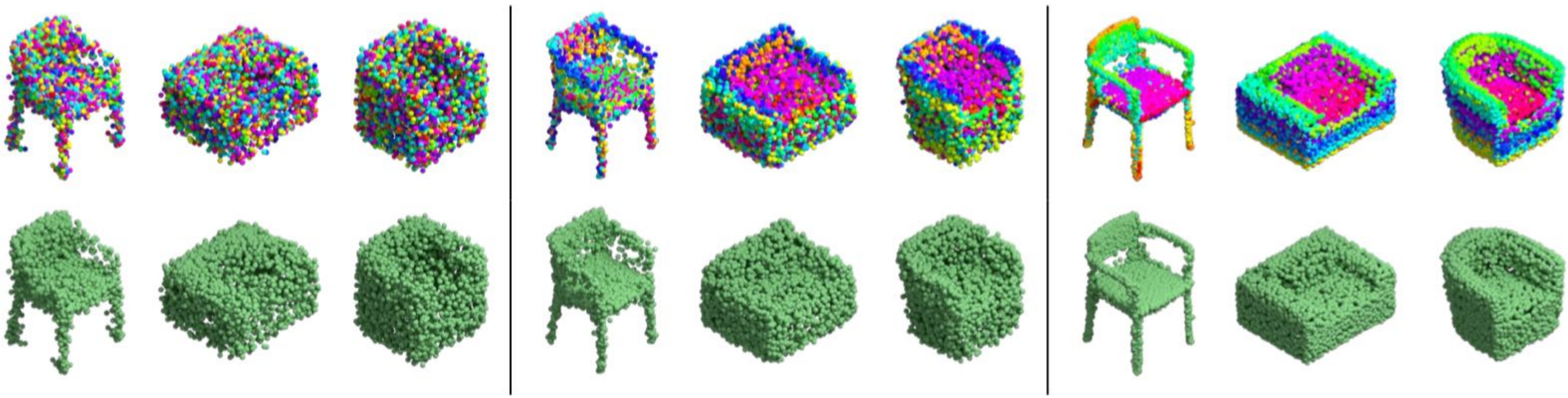
Architecture for encoding and decoding



Multiresolution convolution block

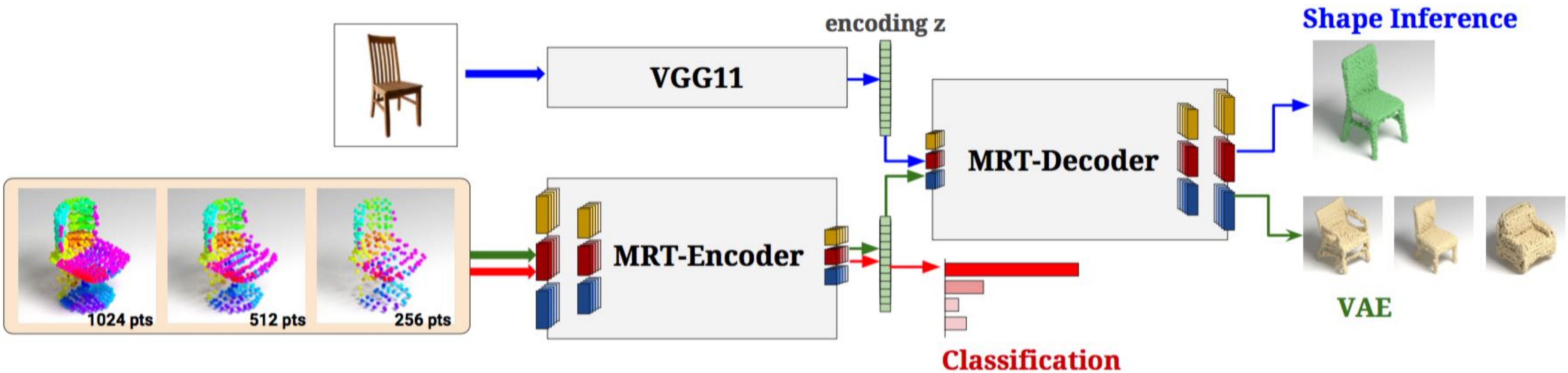
Does multiresolution analysis help?

fully-connected single-resolution multi-resolution

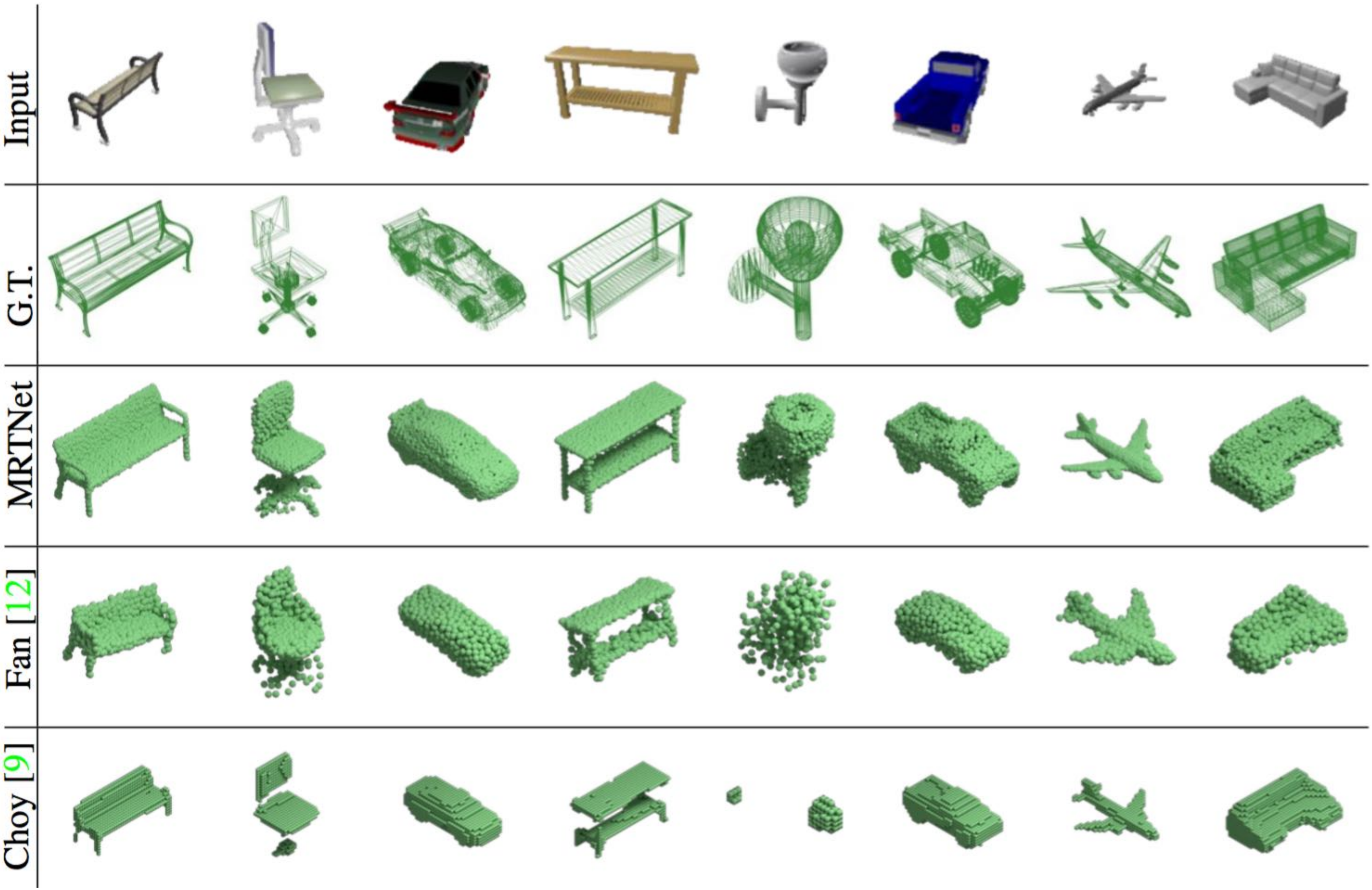


Color indicates the position in the list

Other shape tasks with MRTNet



Single image shape reconstruction



Quantitative evaluation: ShapeNet dataset

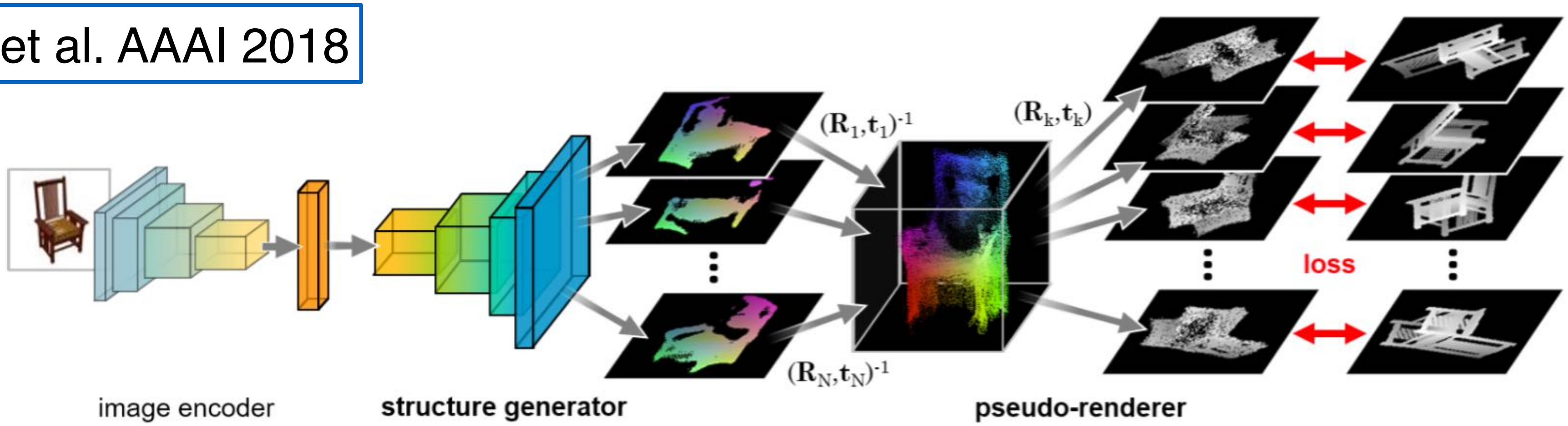
Chamfer distance: pred → GT / GT → pred

voxel-based | fully-conn. | multiview

Category	3D-R2N2 [9] 1 view	3D-R2N2 [9] 3 views	3D-R2N2 [9] 5 views	Fan et al. [12] (1 view)	Lin et al. [26] (1 view)	MRTNet (1 view)
mean	3.345 / 4.102	2.702 / 3.465	2.588 / 3.342	1.982 / 2.146	1.846 / 1.701	**1.559 / 1.529**

$$Ch(\mathbf{x},\mathbf{y}) = \frac{1}{|\mathbf{x}|}\sum_{x\in\mathbf{x}}\min_{y\in\mathbf{y}}\|x-y\|_2 + \frac{1}{|\mathbf{y}|}\sum_{y\in\mathbf{y}}\min_{x\in\mathbf{x}}\|x-y\|_2$$

Lin et al. AAAI 2018



Quantitative evaluation: ShapeNet dataset

Chamfer distance: pred → GT / GT → pred

voxel-based **fully-conn.** **multiview**

Category	3D-R2N2 [9] 1 view	3D-R2N2 [9] 3 views	3D-R2N2 [9] 5 views	Fan et al. [12] (1 view)	Lin et al. [26] (1 view)	MRTNet (1 view)
airplane	3.207 / 2.879	2.521 / 2.468	2.399 / 2.391	1.301 / 1.488	1.294 / 1.541	**0.976 / 0.920**
bench	3.350 / 3.697	2.465 / 2.746	2.323 / 2.603	1.814 / 1.983	1.757 / 1.487	**1.438 / 1.326**
cabinet	1.636 / 2.817	1.445 / 2.626	**1.420** / 2.619	2.463 / 2.444	1.814 / **1.072**	1.774 / 1.602
car	1.808 / 3.238	1.685 / 3.151	1.664 / 3.146	1.800 / 2.053	1.446 / **1.061**	**1.395** / 1.303
chair	2.759 / 4.207	1.960 / 3.238	1.854 / 3.080	1.887 / 2.355	1.886 / 2.041	**1.650 / 1.603**
display	3.235 / 4.283	2.262 / 3.151	2.088 / 2.953	1.919 / 2.334	2.142 / **1.440**	**1.815** / 1.901
lamp	8.400 / 9.722	6.001 / 7.755	5.698 / 7.331	2.347 / 2.212	2.635 / 4.459	**1.944 / 2.089**
speaker	2.652 / 4.335	2.577 / 4.302	2.487 / 4.203	3.215 / 2.788	2.371 / **1.706**	**2.165** / 2.121
rifle	4.798 / 2.996	4.307 / 2.546	4.193 / 2.447	1.316 / 1.358	1.289 / 1.510	**1.029 / 1.028**
sofa	2.725 / 3.628	2.371 / 3.252	2.306 / 3.196	2.592 / 2.784	1.917 / **1.423**	**1.768** / 1.756
table	3.118 / 4.208	2.268 / 3.277	2.128 / 3.134	1.874 / 2.229	1.689 / 1.620	**1.570 / 1.405**
telephone	2.202 / 3.314	1.969 / 2.834	1.874 / 2.734	1.516 / 1.989	1.939 / **1.198**	**1.346** / 1.332
watercraft	3.592 / 4.007	3.299 / 3.698	3.210 / 3.614	1.715 / 1.877	1.813 / 1.550	**1.394 / 1.490**
mean	3.345 / 4.102	2.702 / 3.465	2.588 / 3.342	1.982 / 2.146	1.846 / 1.701	**1.559 / 1.529**

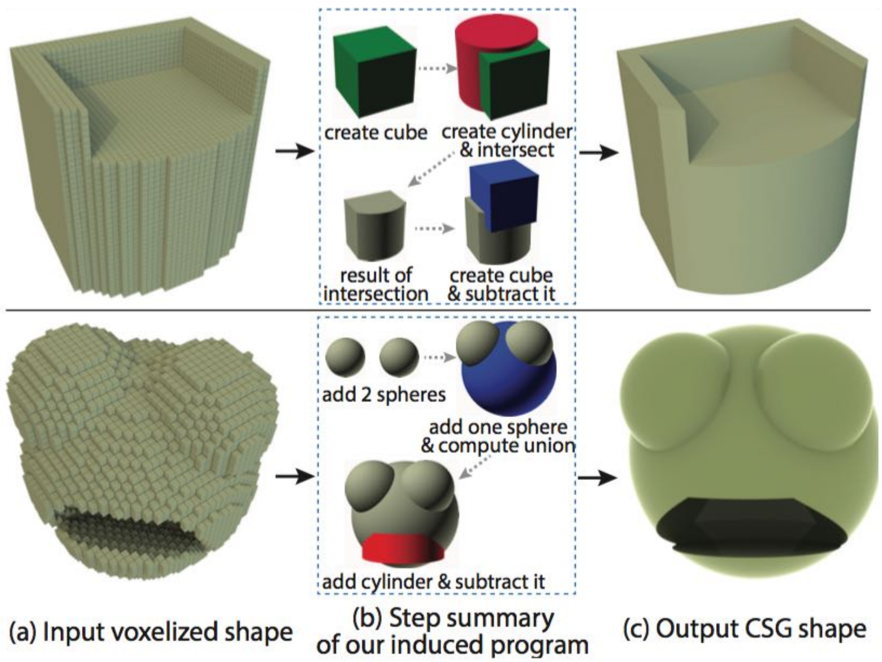
MRTNet summary

- A generic architecture for
  - Point cloud classification (**91.7%** on ModelNet40)
  - Semantic segmentation (see results in the paper)
  - Generation

- Project page: http://mgadelha.me/mrt/index.html

Part 3

CSGNet: Neural Shape Parser for Constructive Solid Geometry

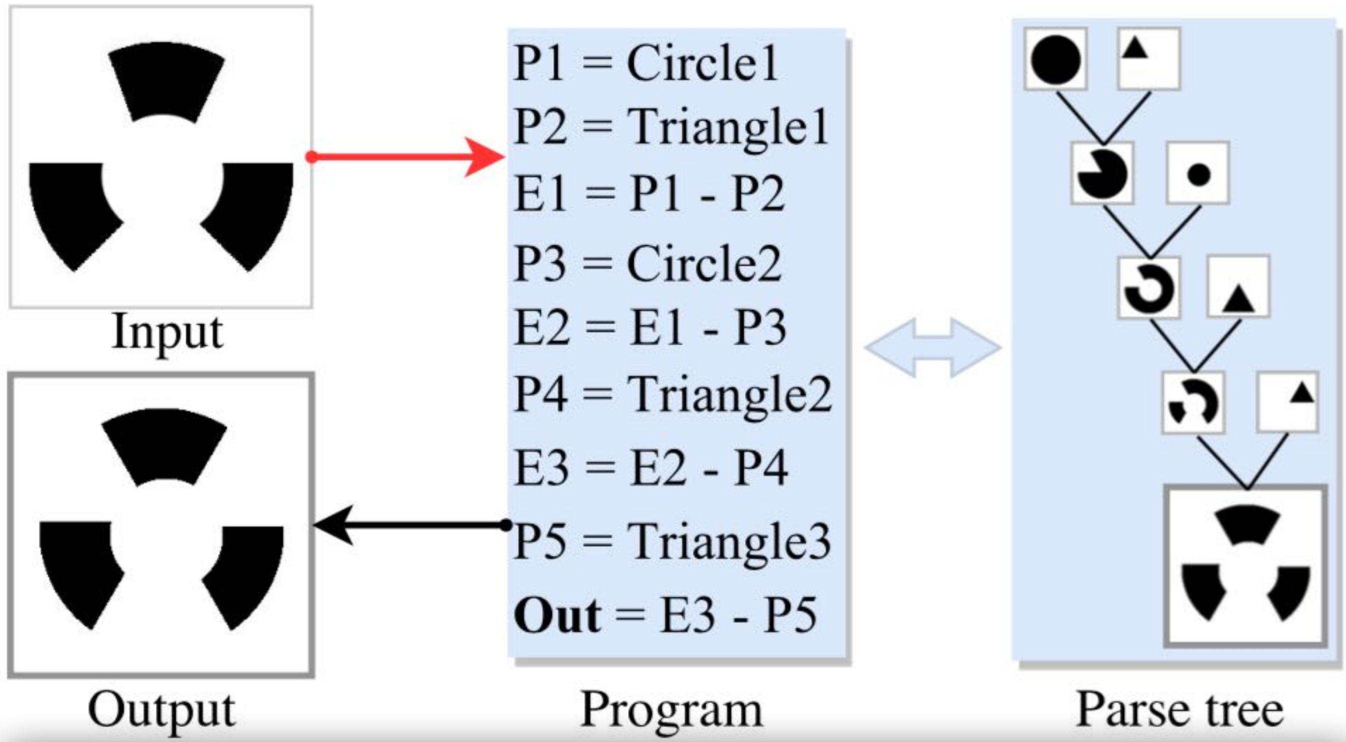


Gopal Sharma
Rishabh Goyal
Difan Liu
Evangelos Kalogerakis
Subhransu Maji

CVPR 18

interpretable and editable

Constructive 2D geometry



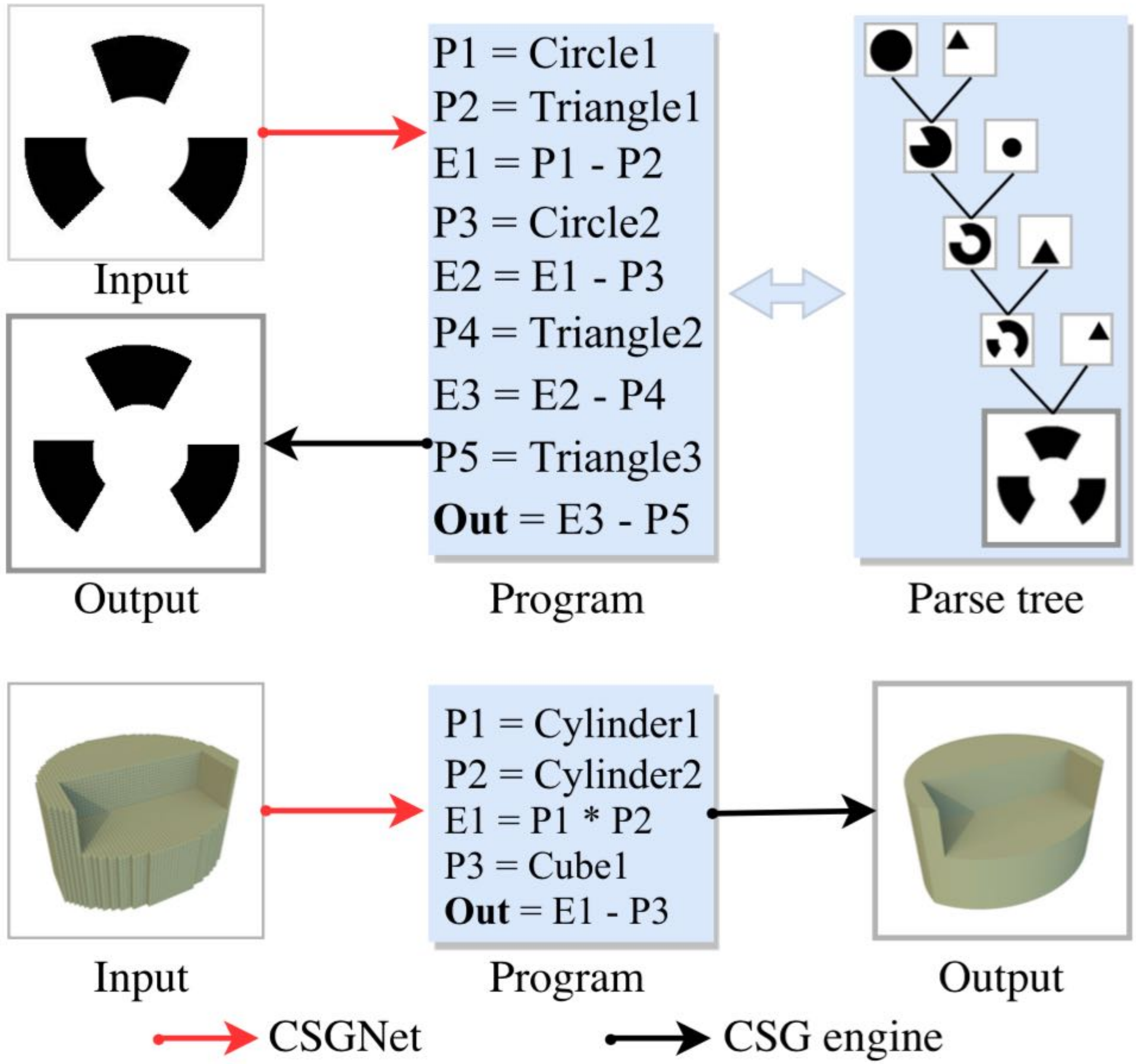
$$S \to E$$

$$E \to E\ E\ T \mid P$$

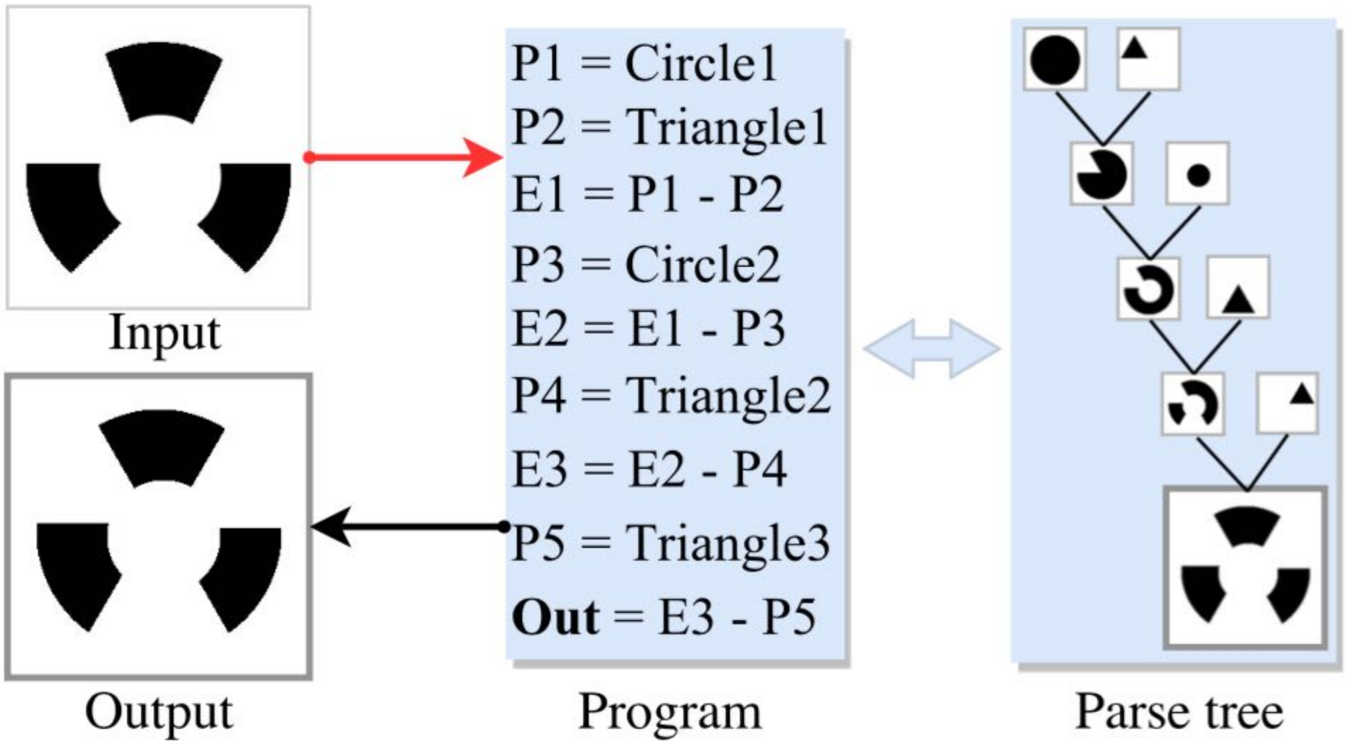
$$T \to \mathtt{OP}_1 | \mathtt{OP}_2 | \ldots | \mathtt{OP}_m$$

$$P \to \mathtt{SHAPE}_1 | \mathtt{SHAPE}_2 | \ldots | \mathtt{SHAPE}_n$$

Constructive solid geometry

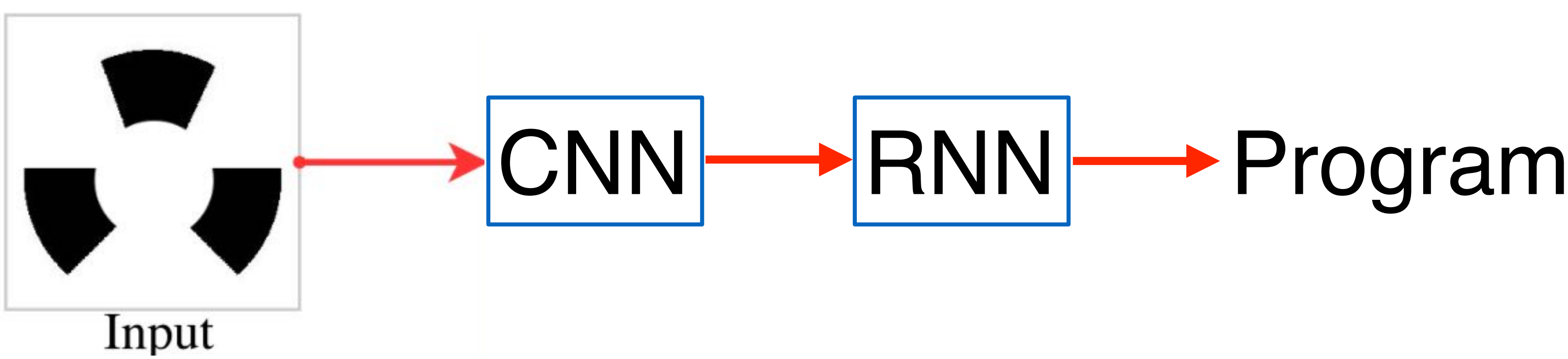


Constructive solid geometry



Circle1
Triangle1
-
Circle2
-
Triangle2
-
Triangle3
-
[STOP]

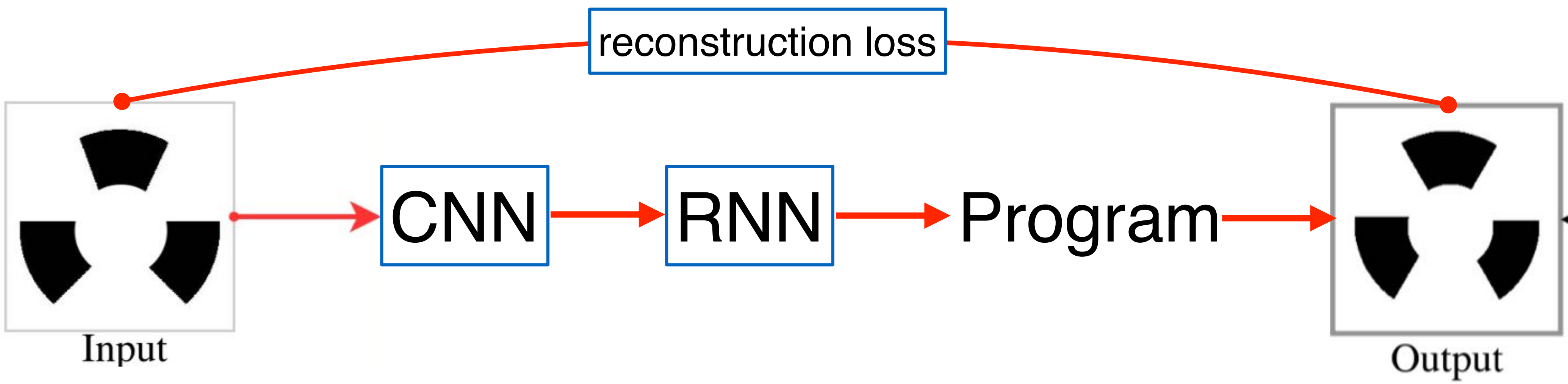
Postfix notation



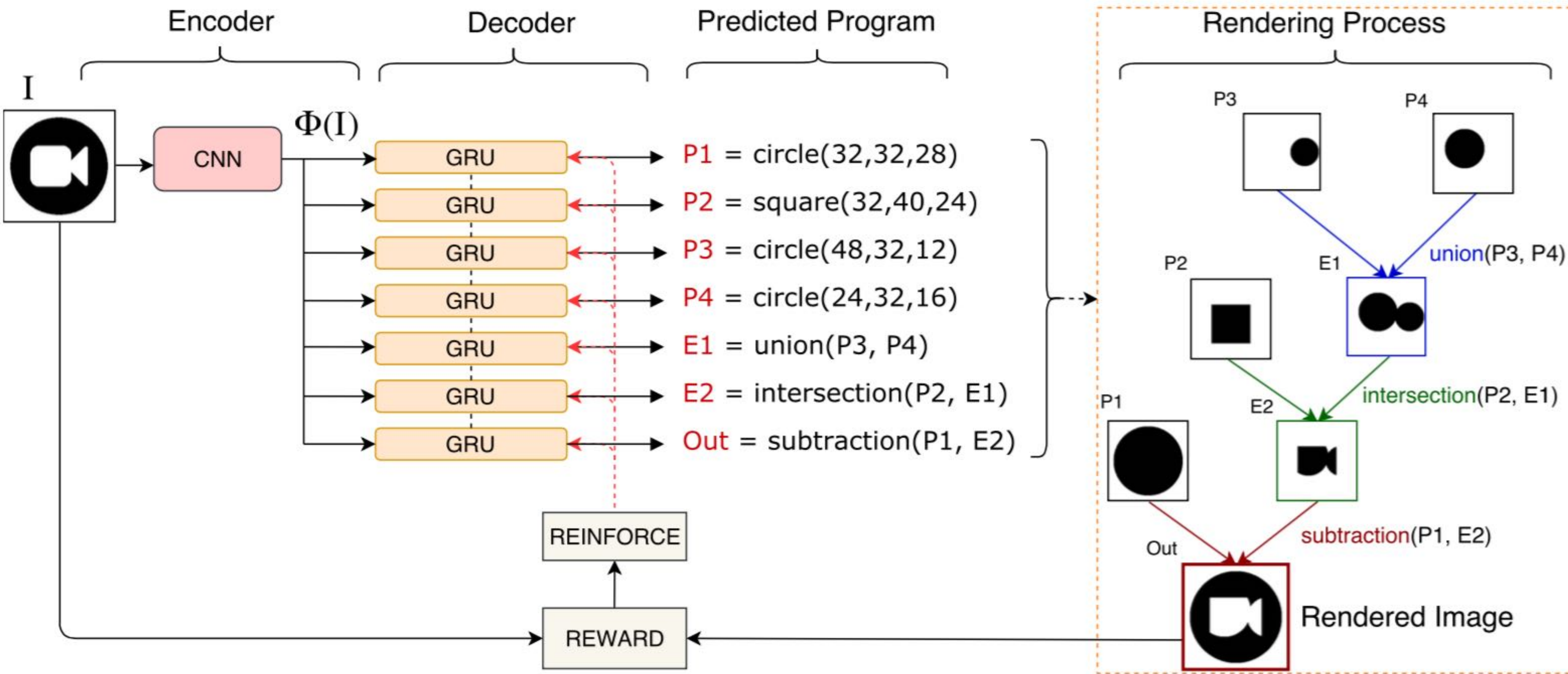
Learning

- **Supervised setting:** learn to predict programs directly

- **Unsupervised setting:** No ground-truth programs.
  - Learn parameters to minimize a reconstruction error through policy gradients [REINFORCE, Willams 1992]

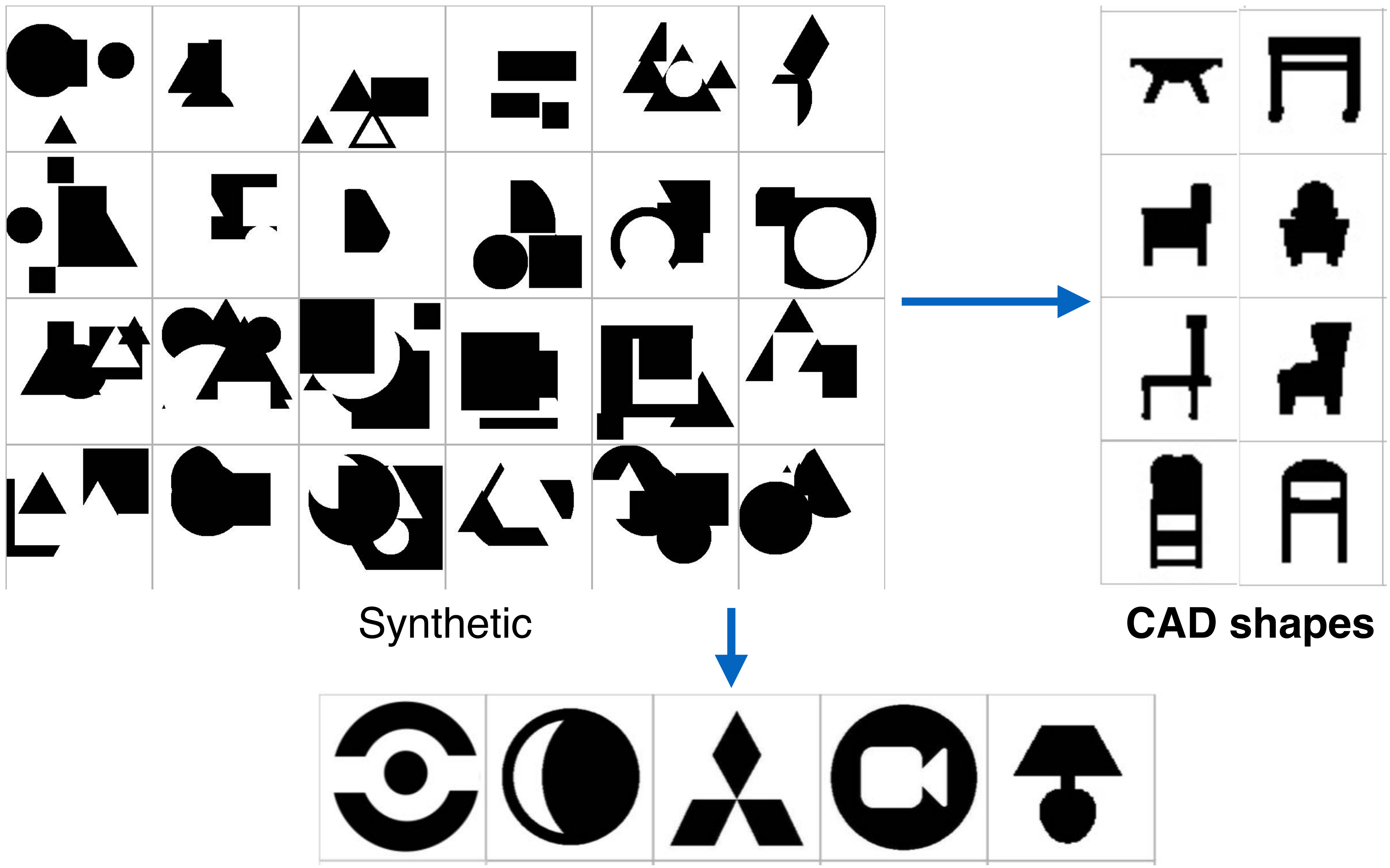


CSGNet: 2D/3D ⇨ programs



CSGNet: 2D/3D ⇨ programs

Train on <u>synthetic data</u> and adapt to <u>new domains</u> using policy gradients

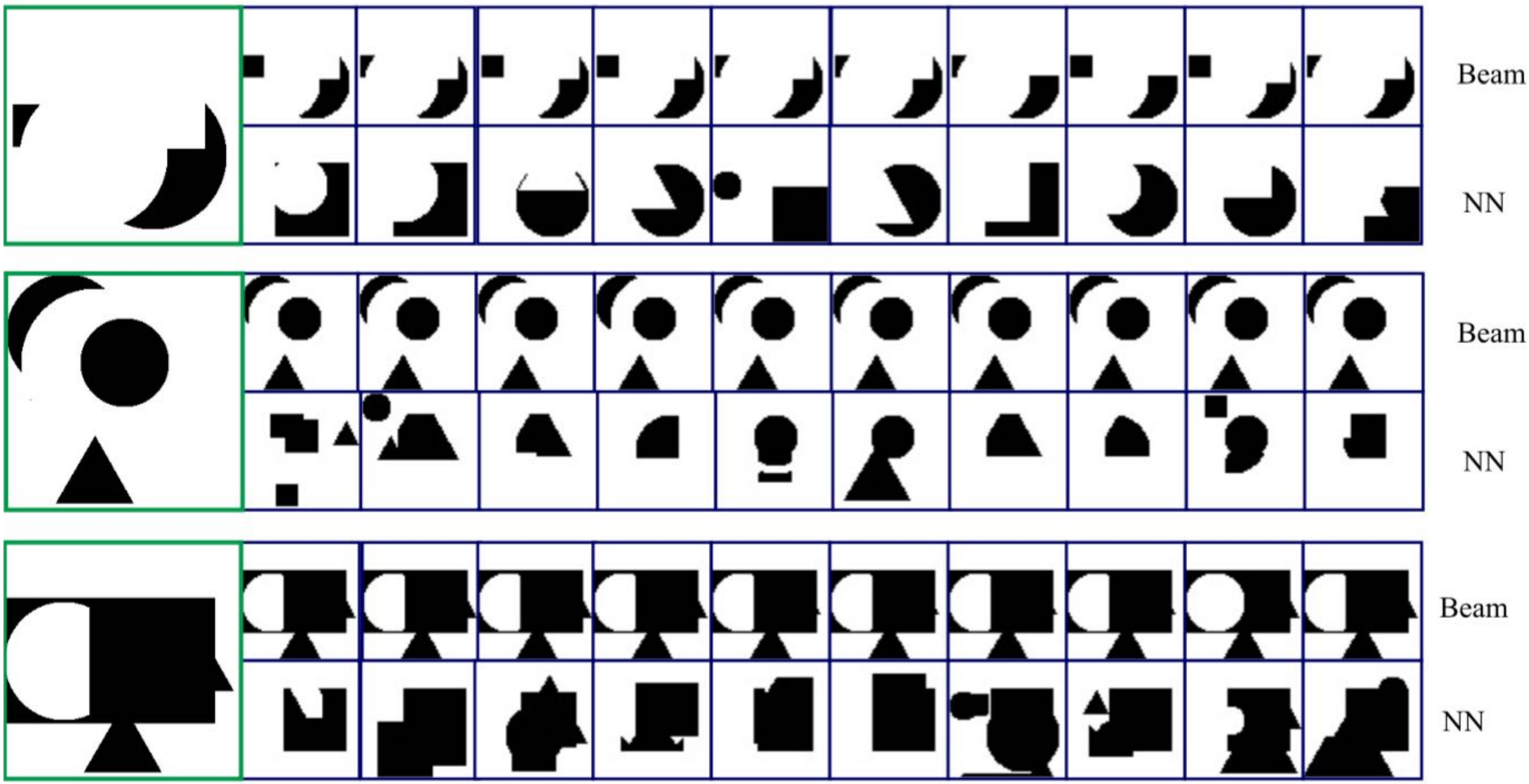


Logos

CSGNet: 2D/3D ⇨ programs

How well does the nearest neighbor perform?

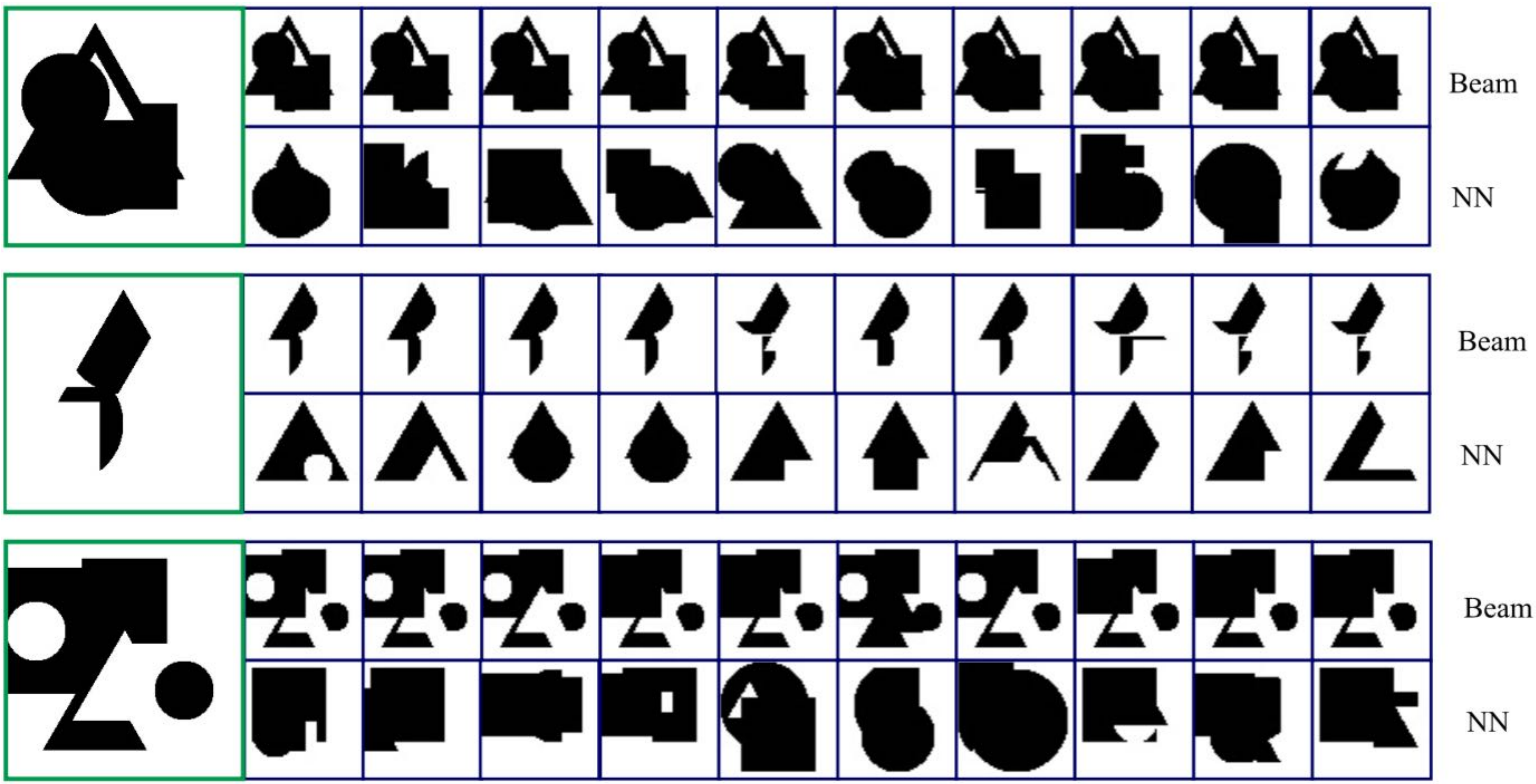
Chamfer distance **1.88** (NN), **1.36** (CSGNet) with 675K training examples



CSGNet: 2D/3D ⇨ programs

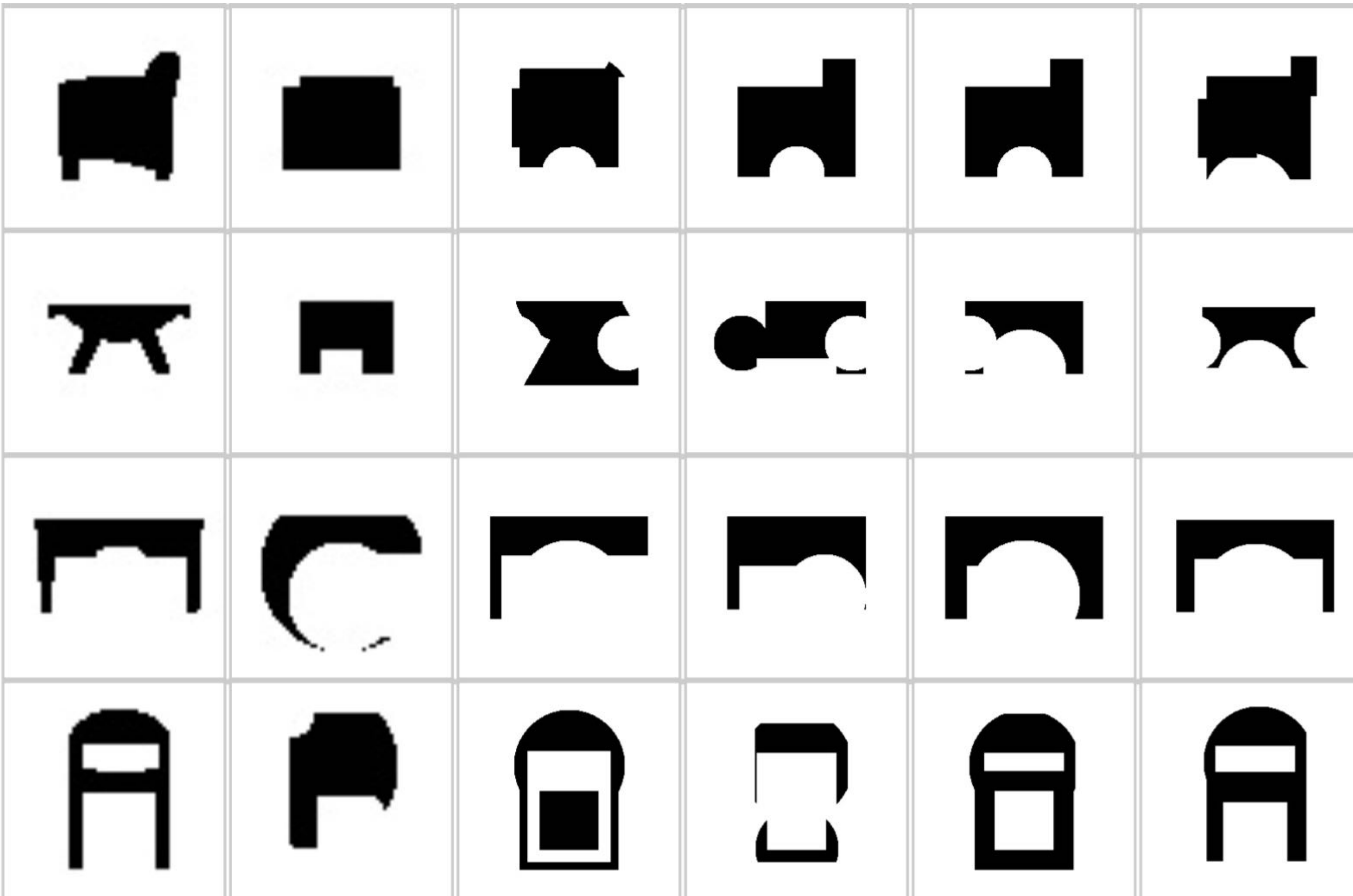
How well does the <u>nearest neighbor</u> perform?

Chamfer distance **1.88** (NN), **1.36** (CSGNet) with 675K training examples



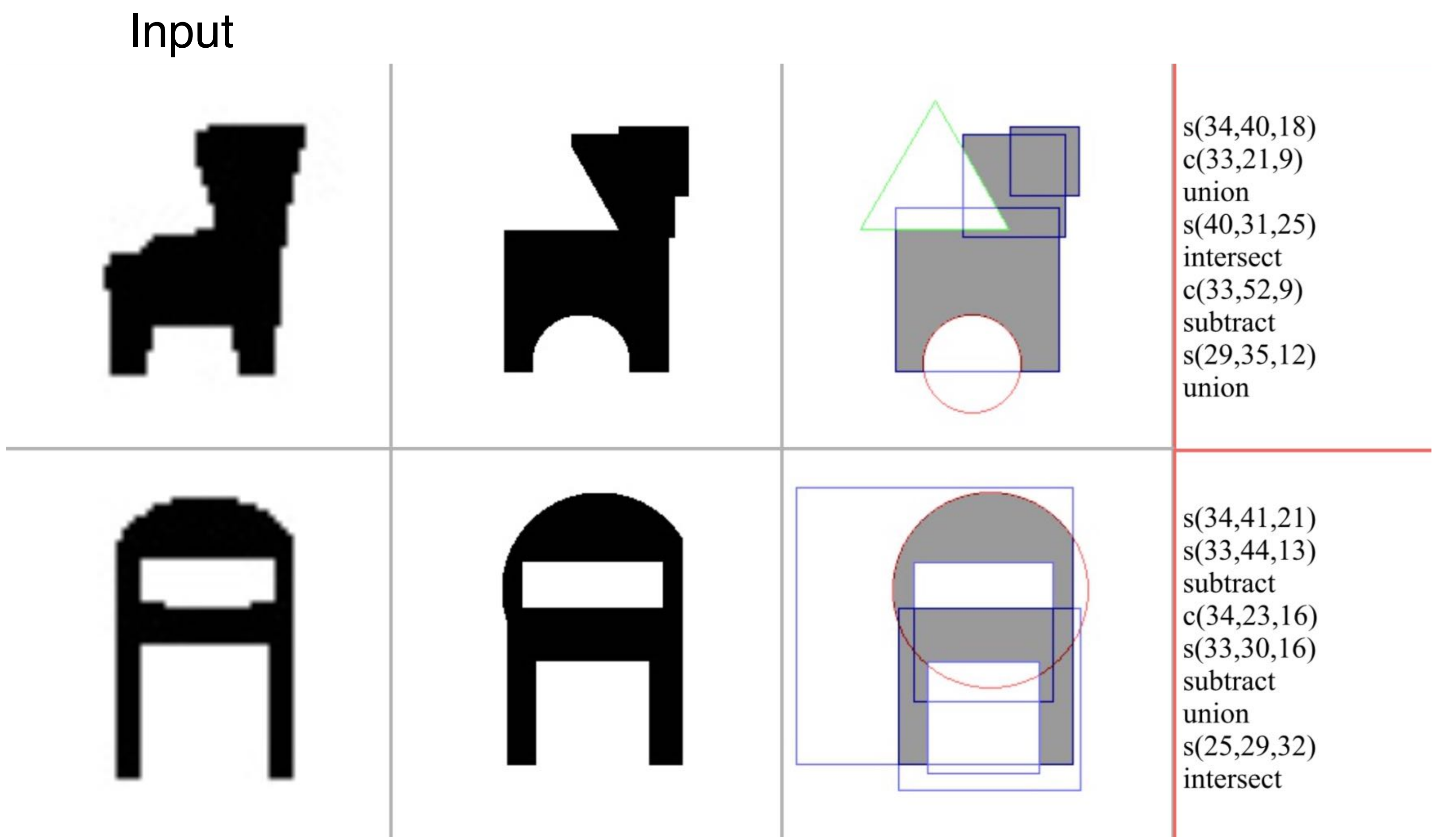
CSGNet: 2D/3D ⇨ programs

CAD shapes dataset: Chamfer distance **1.94** (NN), **0.51** (CSGNet)



a) Target b) NN c) Supervised d) RL e) RL + beam f) Full

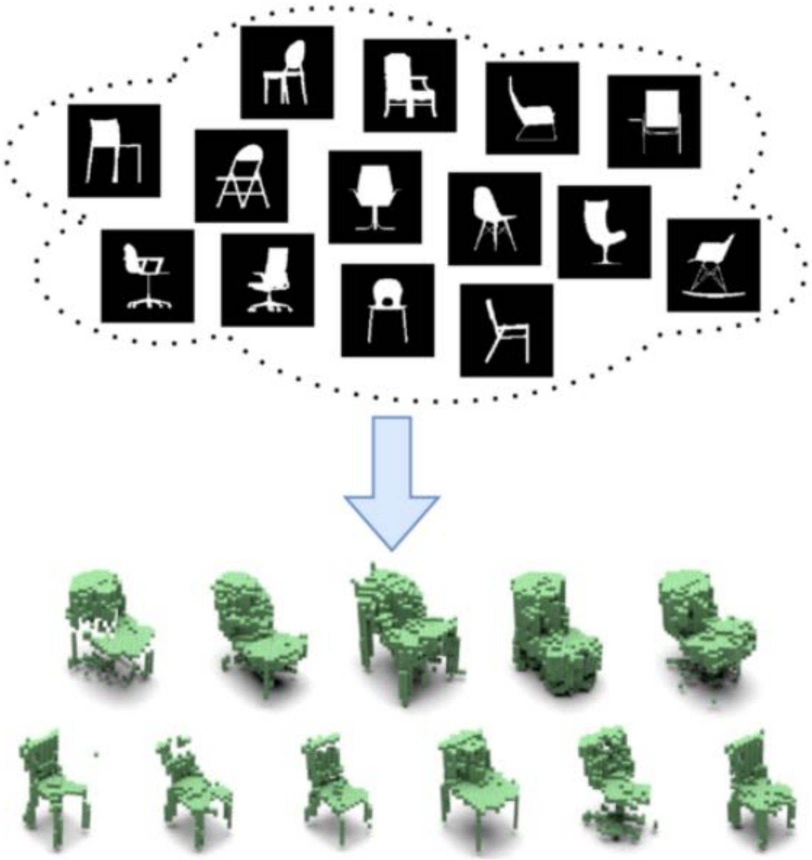
CSGNet: 2D/3D ➪ programs



- More results in the paper: reward shaping, comparison to Faster R-CNN for primitive detection, results on 3D, etc.
- Preprint available: https://arxiv.org/abs/1712.08290

Part 4

Learning 3D Shape Representations with Weak Supervision



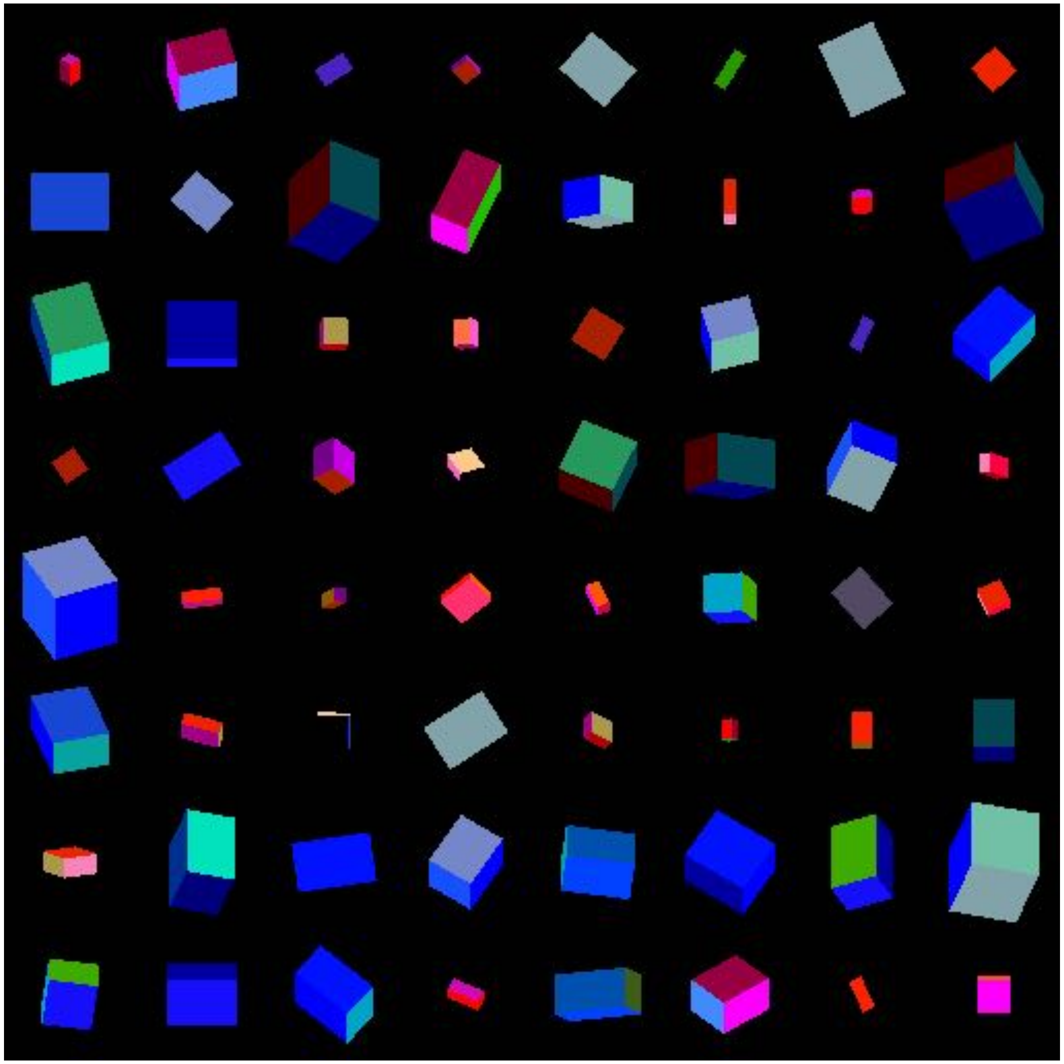
Matheus Gadelha
Subhransu Maji
Rui Wang

3DV 2017

Related Work

- 3D shape from collection of images
  - Visual hull — same instance, known viewpoints
  - Photometric stereo — same instance, known lighting, simple reflectance
  - Structure from motion — same instance (or 3D)
  - Non-rigid structure from motion — known shape family (e.g., faces)
  - **Our work** — unknown shape family, unknown viewpoints
- 3D shape from single image
  - Optimization-based approaches;
  - Recognition-based approaches;

A motivating example

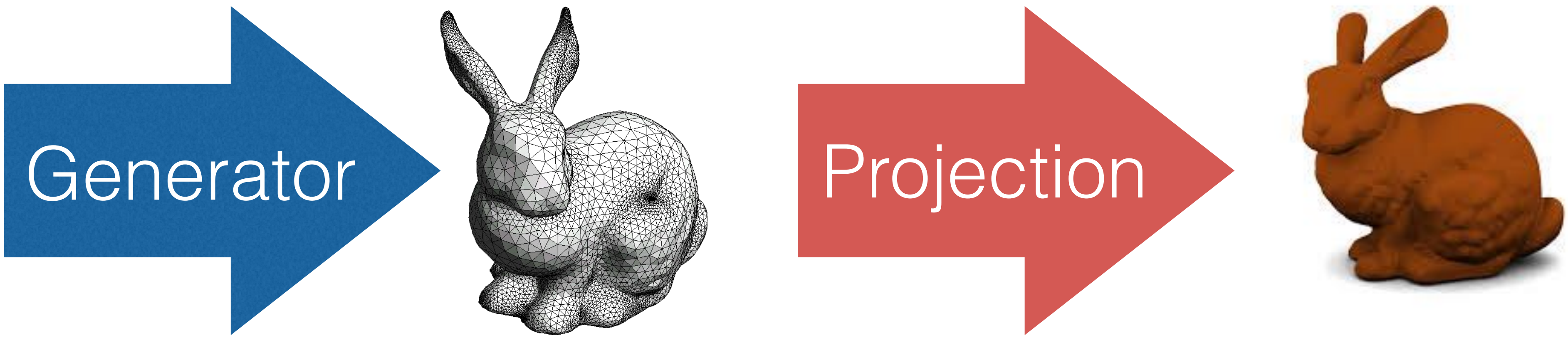


Small cubes are reddish
Big cubes are bluish

Hypothesis: It is easier to generate these images by reasoning in 3D

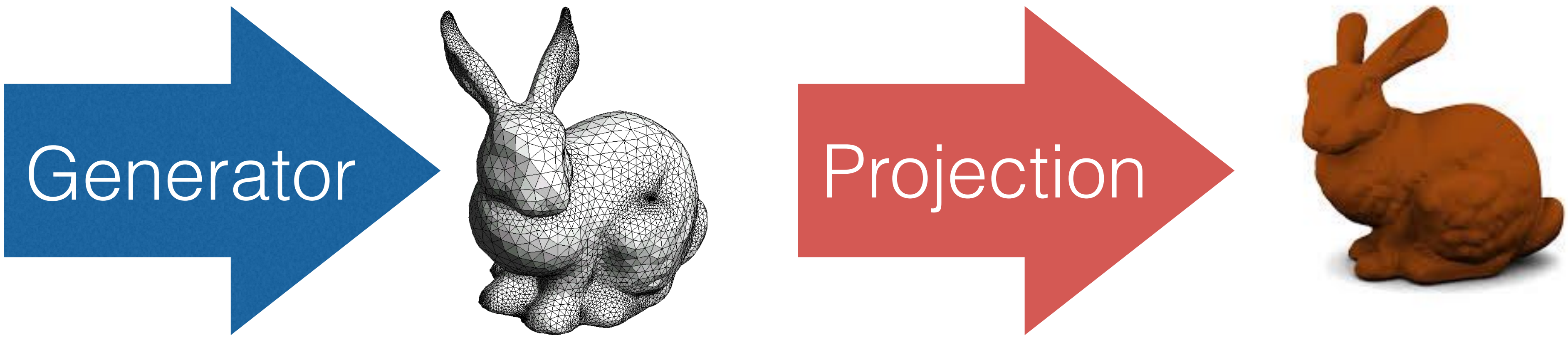
Approach

- Our goal is to learn a 3D shape generator whose projections match the provided set of the views



Approach

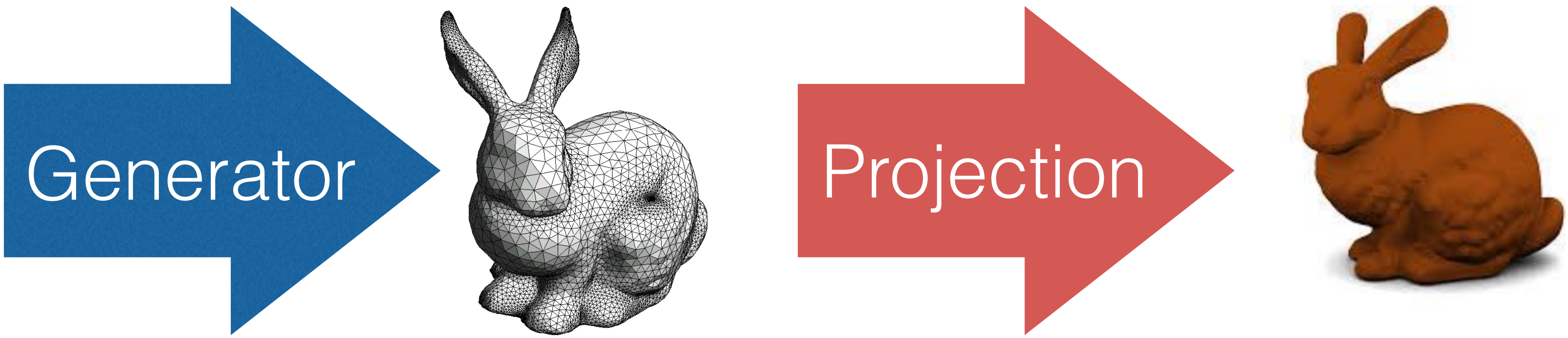
- Our goal is to learn a 3D shape generator whose projections match the provided set of the views



How do we match distributions?

Approach

- Our goal is to learn a 3D shape generator whose projections match the provided set of the views



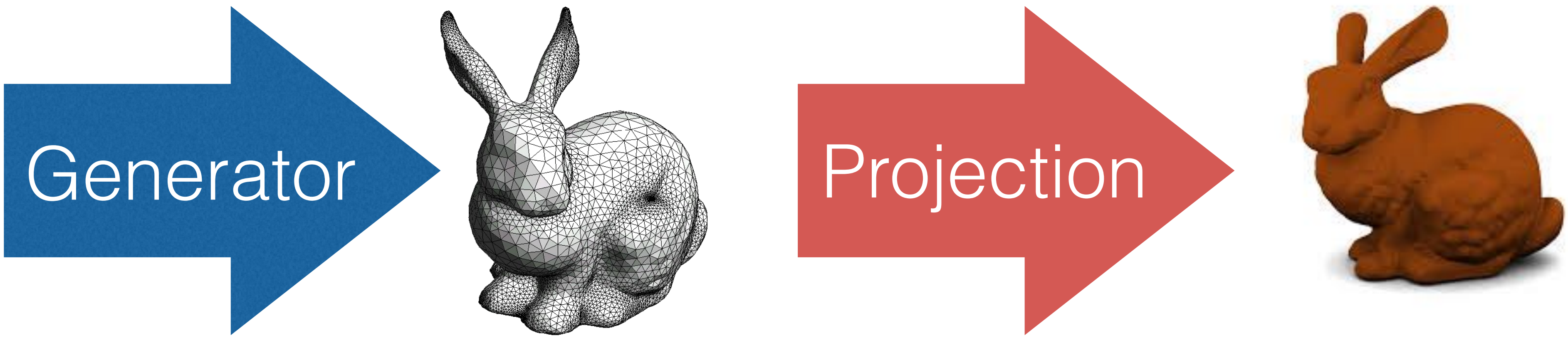
How do we match distributions?

generated true

$$\min_{G} D_{\mathrm{KL}}(G||D) = \min_{z \sim G} \mathbb{E}\left[\log \frac{G(z)}{D(z)}\right]$$

Approach

- Our goal is to learn a 3D shape generator whose projections match the provided set of the views



How do we match distributions?

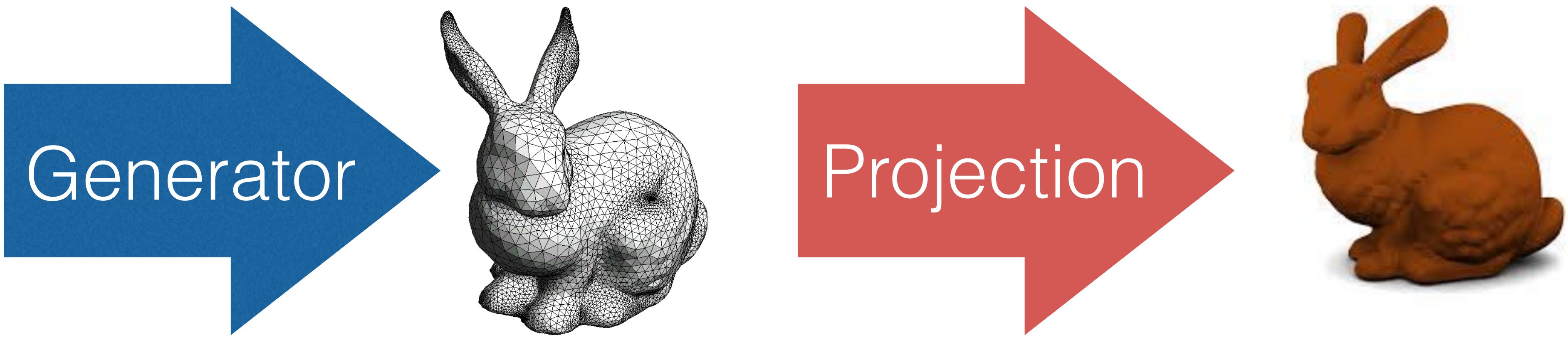
generated true

$$\min_G D_{\mathrm{KL}}(G||D) = \min_{z\sim G} \mathbb{E}\left[\log \frac{G(z)}{D(z)}\right]$$

estimate using logistic regression

Approach

- Our goal is to learn a 3D shape generator whose projections match the provided set of the views



How do we match distributions?

generated true

$$\min_G D_{\mathrm{KL}}(G||D) = \min_{z\sim G} \mathbb{E}\left[\log \frac{G(z)}{D(z)}\right]$$

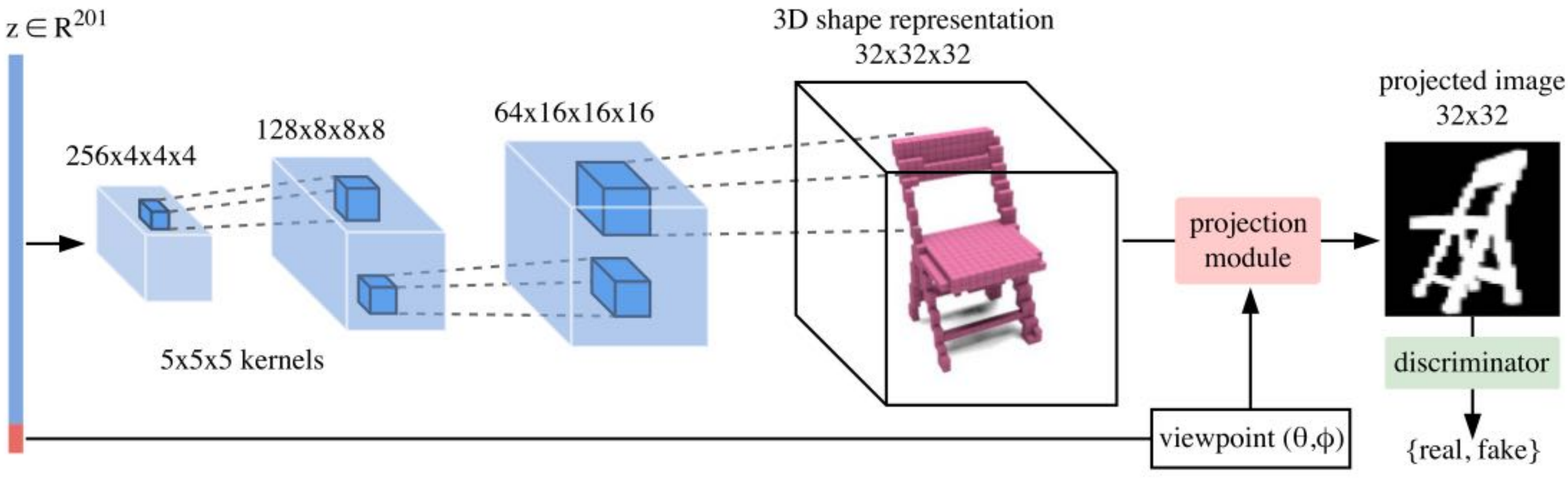
estimate using logistic regression

$$\min_G \max_d \mathbb{E}_{x\sim D}[\log d(x)] + \mathbb{E}_{z\sim G}[\log(1 - d(z))]$$

Generative adversarial networks [Goodfellow et al.]

PrGAN

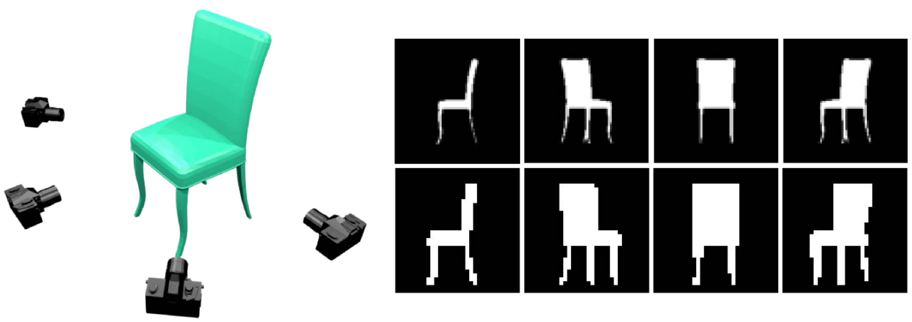
Generator maps z to a voxel occupancy grid and a viewpoint

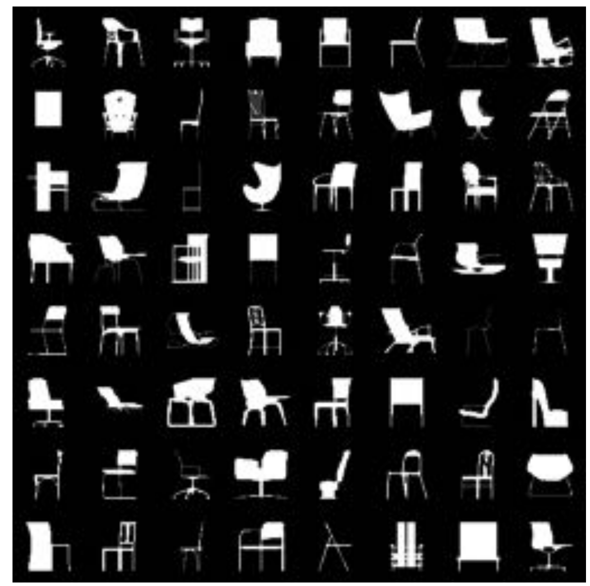


Projection using line integration along the view direction

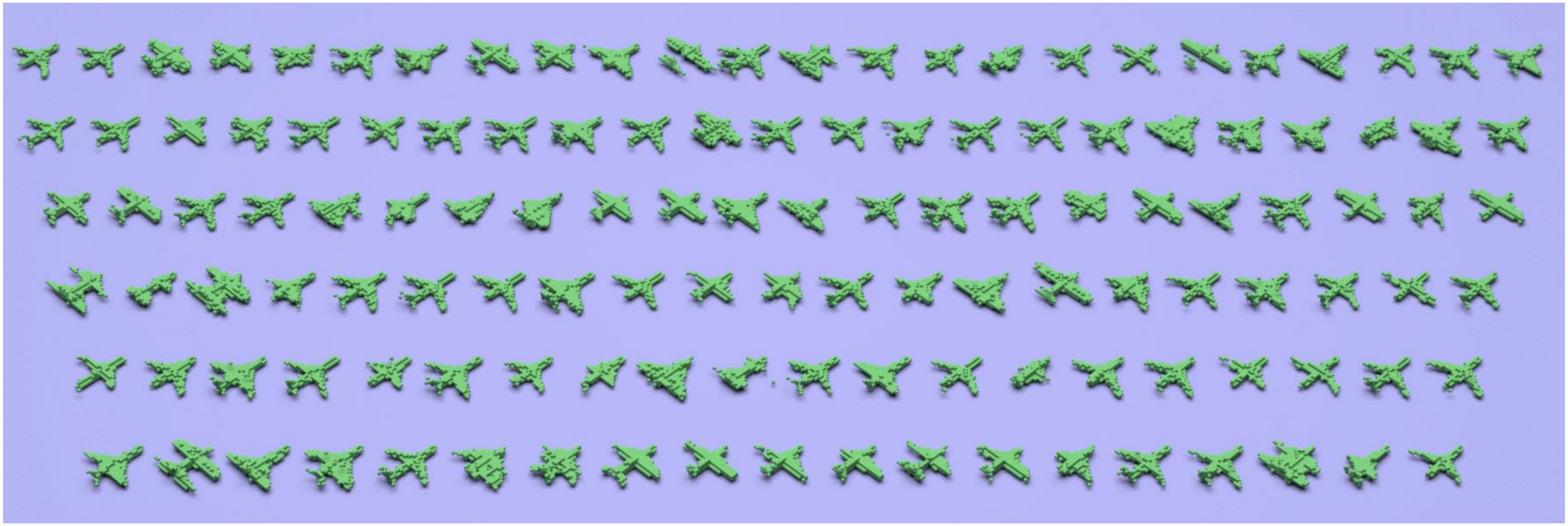
$$I(\mathbf{x}) = 1 - \exp\left(-\int_0^\infty V(\mathbf{x}+\mathbf{r})dr\right)$$

Dataset generation

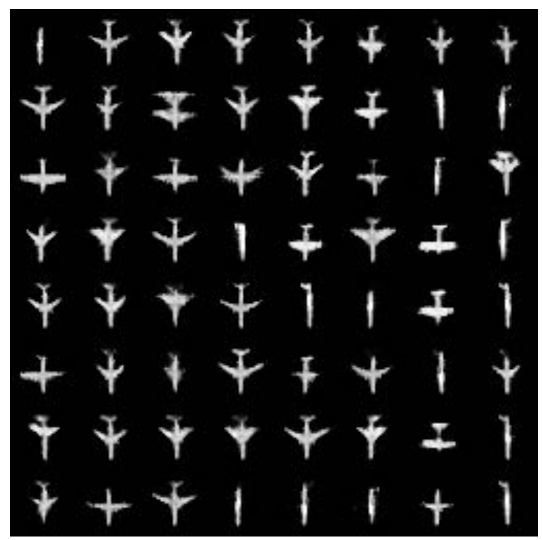




Airplanes

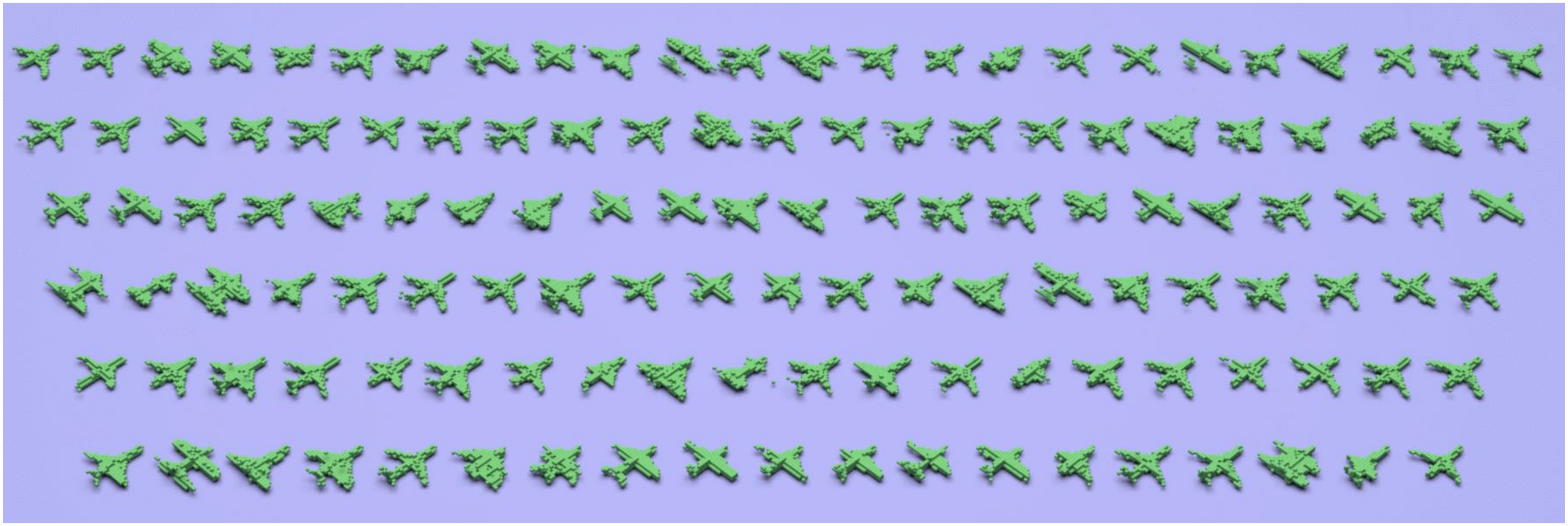


input

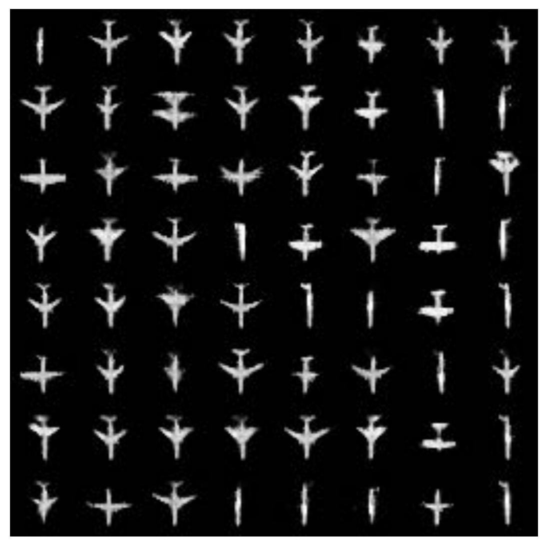


generated

Airplanes



input



generated

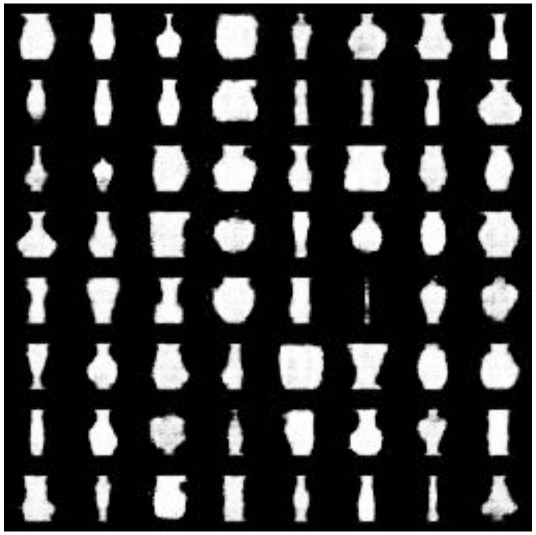
Vases

input

generated

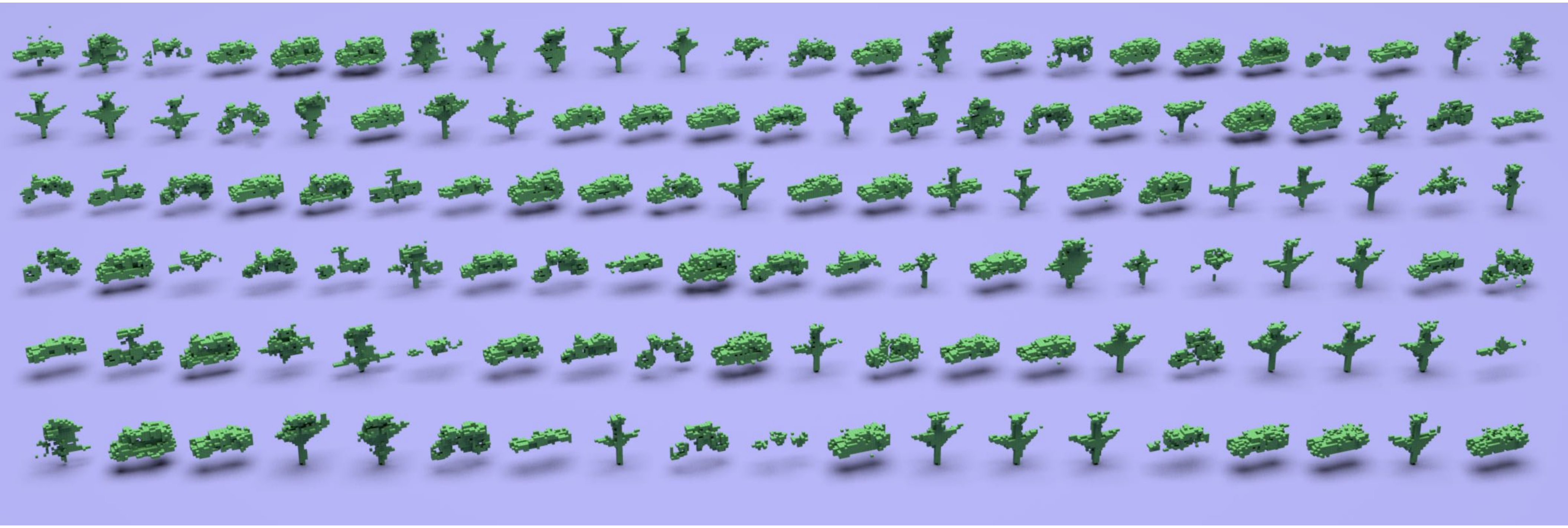
Vases

input



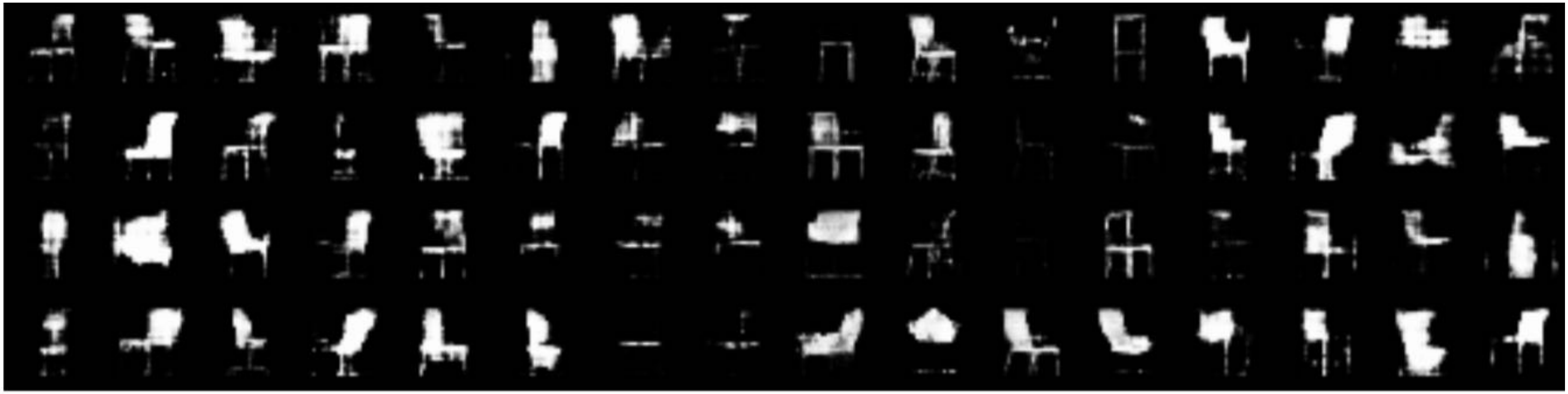
generated

Mixed categories



input

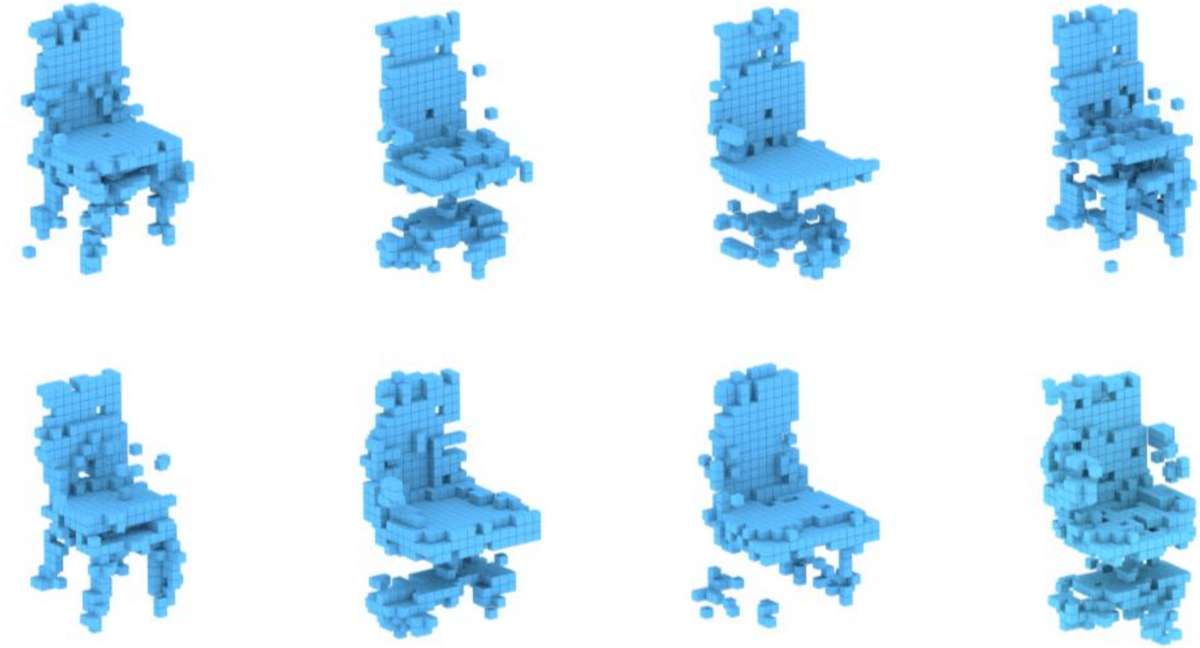
generated



(a) Results from 2D-GAN.

(a) Results from PrGAN.

MMD metric: 2D-GAN 90.1, PrGAN 88.3



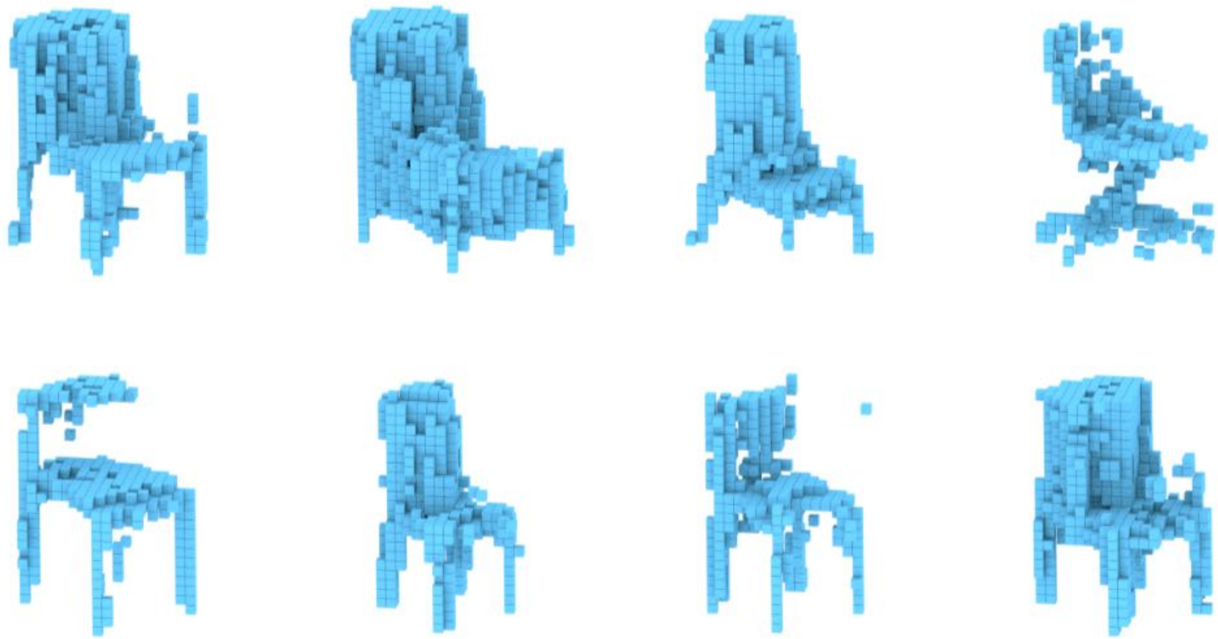
(a) Results from 3D-GAN.

Trained with 3D supervision

MMD metric

3D-GAN 347.5

PrGAN 442.9



(a) Results from PrGAN.

Projection GAN

- The model is able to recover the coarse 3D structure

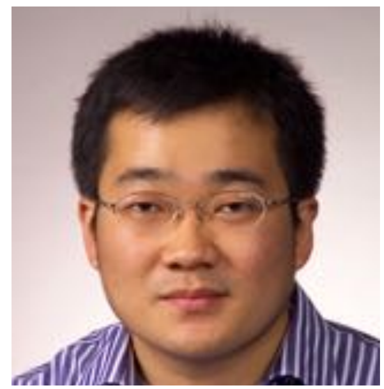
- But should use side information when available
  - Viewpoint
  - Landmarks / part labels
  - Pose estimates

- Iterative: bootstrap 3D to estimate pose & viewpoint

Thank you!

- Collaborators: Matheus Gadelha, Zhaoliang Lun, Gopal Sharma, Rui Wang, Evangelos Kalogerakis




- Funding from NSF, NVIDIA, Facebook
- https://people.cs.umass.edu/smaji/projects.html