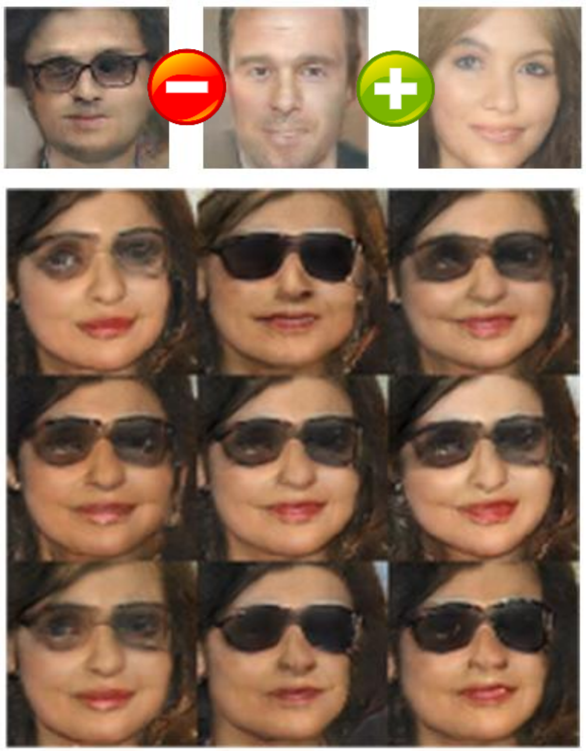
Learning Deep Generative Models for 3D Shape Structures

徐 凯

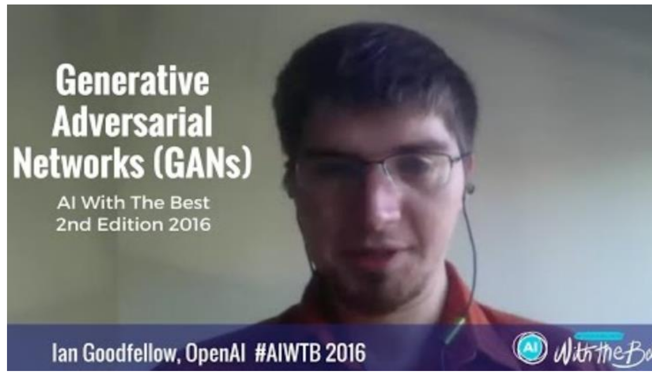
国防科技大学计算机学院

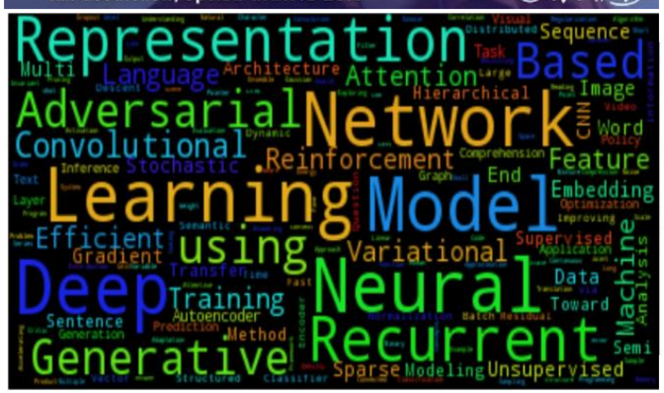
GAMES Web Seminar • 2017.6.29

Deep Generative Models recently …



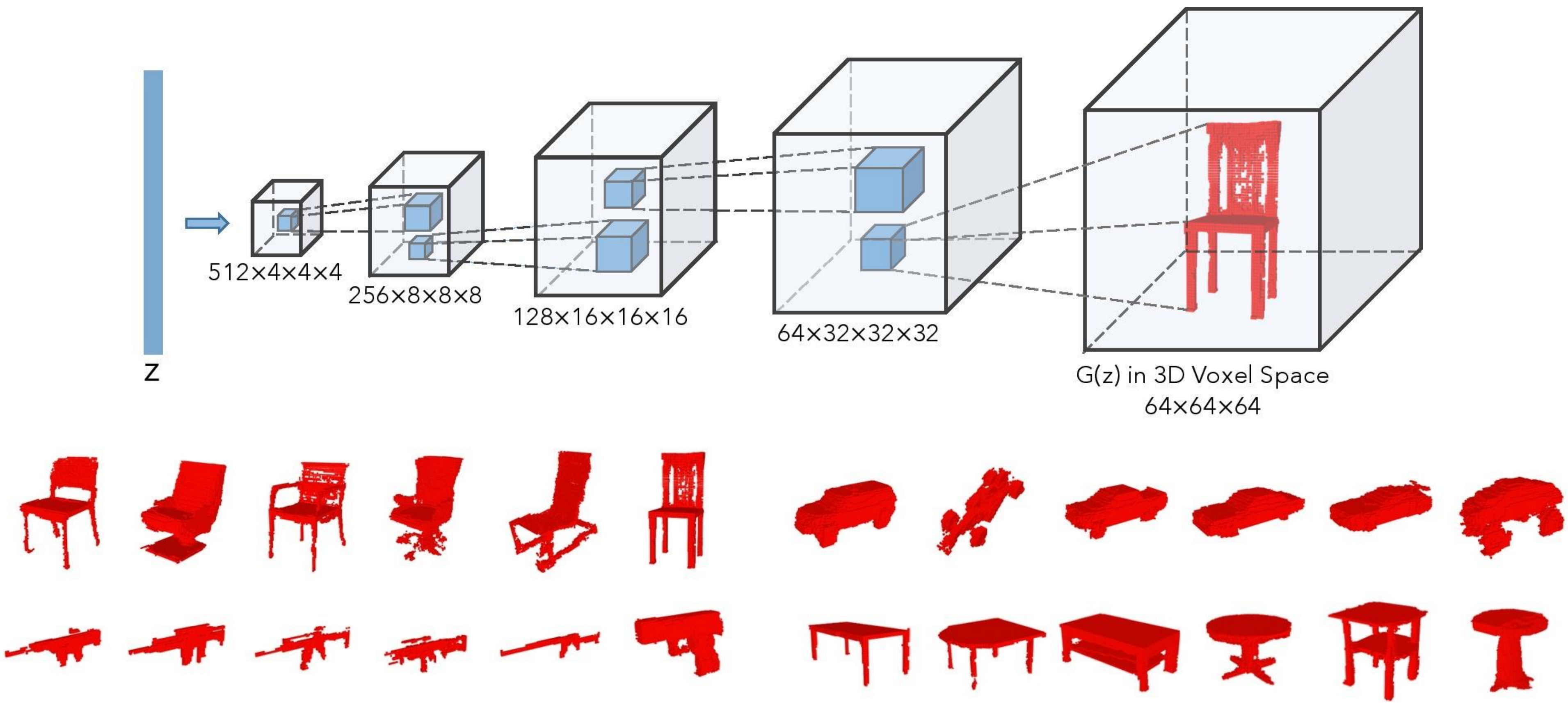
[Radford et al. 2015]





[Isola et al. 2017]

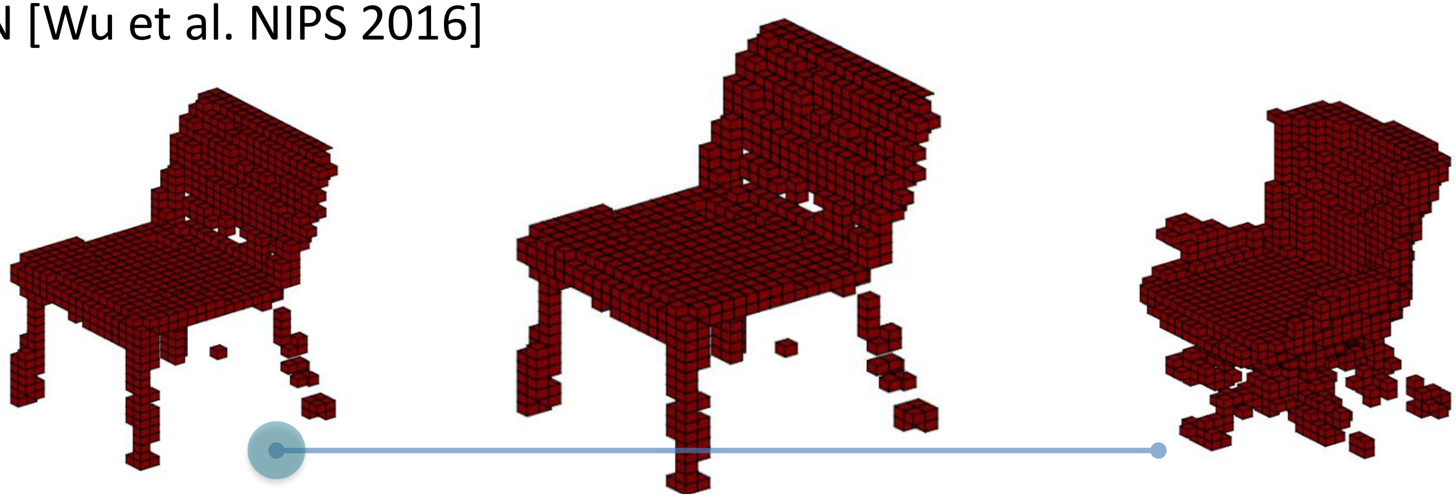
Even for 3D shapes …



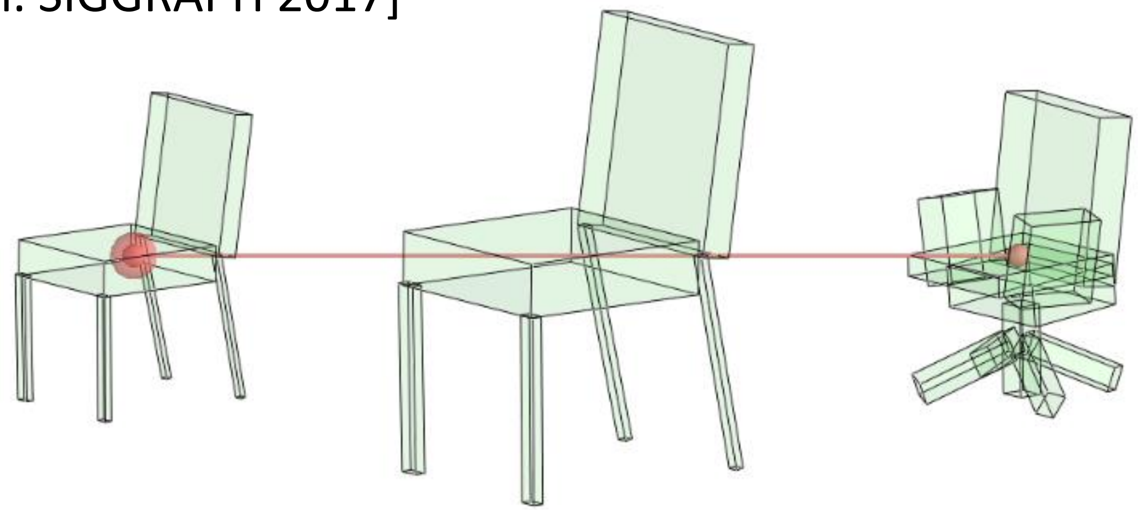
3D-GAN [Wu et al. NIPS 2016]

Generating 3D shapes

3D-GAN [Wu et al. NIPS 2016]



GRASS [Li et al. SIGGRAPH 2017]



Today

- What is special in generating 3D shapes?
- What is a good generative model for 3D shapes?
- What do we learn from building such a model?

Generative Models

both for analysis and synthesis

Generative models

- For analysis:
  - To form a conditional probability density function
  - Example: Maximum a Posterior (MAP)

Parameters Observation Likelihood

$$\text{argmax}_\theta \quad p(\boxed{\theta} \mid \boxed{x}) = \frac{\boxed{p(x|\theta)} \cdot p(\theta)}{C}$$

Analysis-by-Synthesis

Generative models

- For synthesis
  - To model data distribution directly
  - Example: Variational Bayesian method

Latent variable

$$\max.\ P(X) = \int \boxed{P(X|z,\theta)}\, P(\boxed{z})\mathrm{d}z$$

Likelihood

Sample from the likelihood

Generate new for ?

Generate natural image
[Zhang et al. 2016]

Level	Method	1-billion-word
Word	LSTM	An opposition was growing in China . This is undergoing operation a year . It has his everyone on a blame . Everyone shares that Miller seems converted President as Democrat . Which is actually the best of his children . Who has The eventual policy and weak ?
	CNN	Companies I upheld , respectively patented saga and Ambac. Independence Unit have any will MRI in these Lights It is a wrap for the annually of Morocco The town has Registration matched with unk and the citizens
Character	CNN	To holl is now my Hubby , The gry timers was faller After they work is jith a But in a linter a revent

Generate natural language
[Rajeswar et al. 2017]

- Curiosity about unsupervised learning?
- Better understanding / interpretation?
- Applications?

Generate new for 3D Content Creation

- In Computer Graphics, 3D Content Creation has been a **hard-core challenge** for long.

Generate new for creativity support !

- 3D content creation calls for creativity
  - Automatic, diverse & creative

Cars

Monsters University

Generate new for creativity support !

- 3D content creation calls for creativity
  - Automatic, diverse & creative

3D Content Creation

3D Content Creation

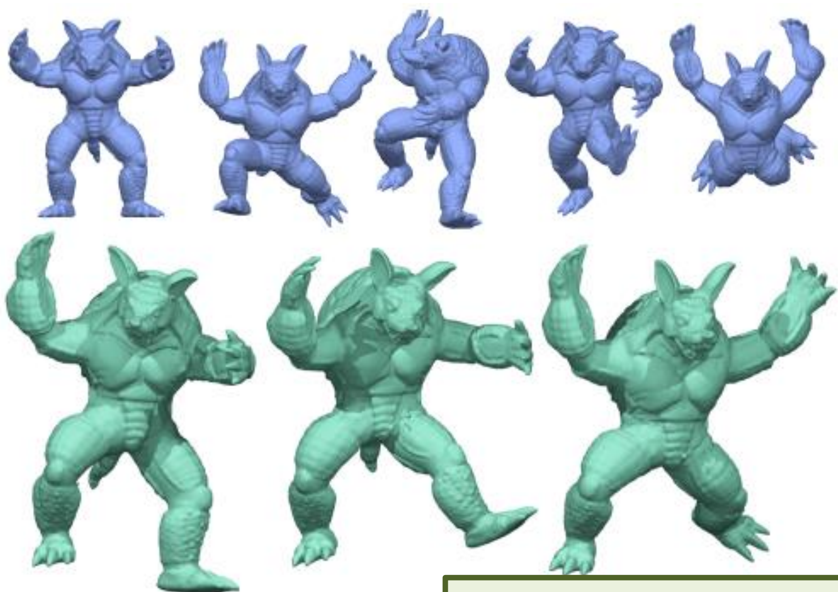
… finding powerful means to create 3D shapes is the key to make graphics as ubiquitous as we had wanted it to be.

Jim Kajiya

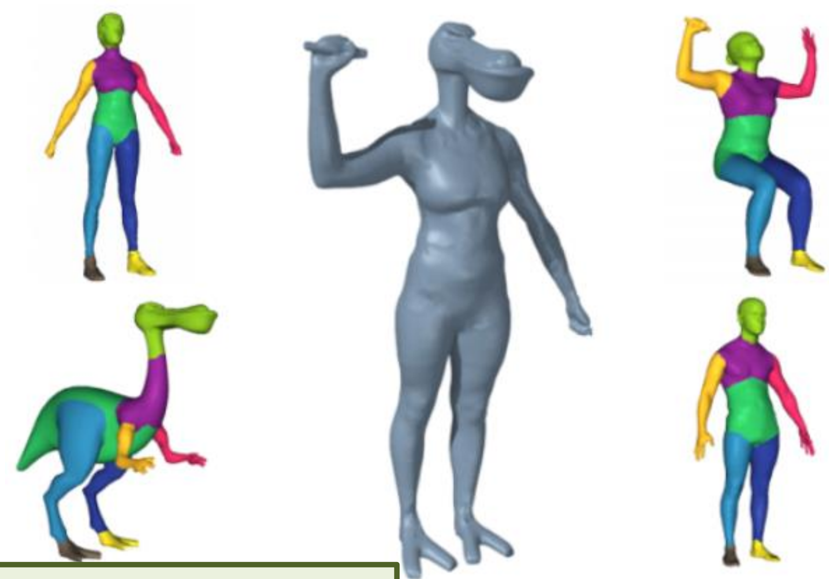
Recipient of Steven Anson Coons Award

Generating 3D shape variations

- Two basic approaches



Defor



mbling

Preserving local details

Easier

Automatic 3D shape synthesis

- Two basic approaches

Defo

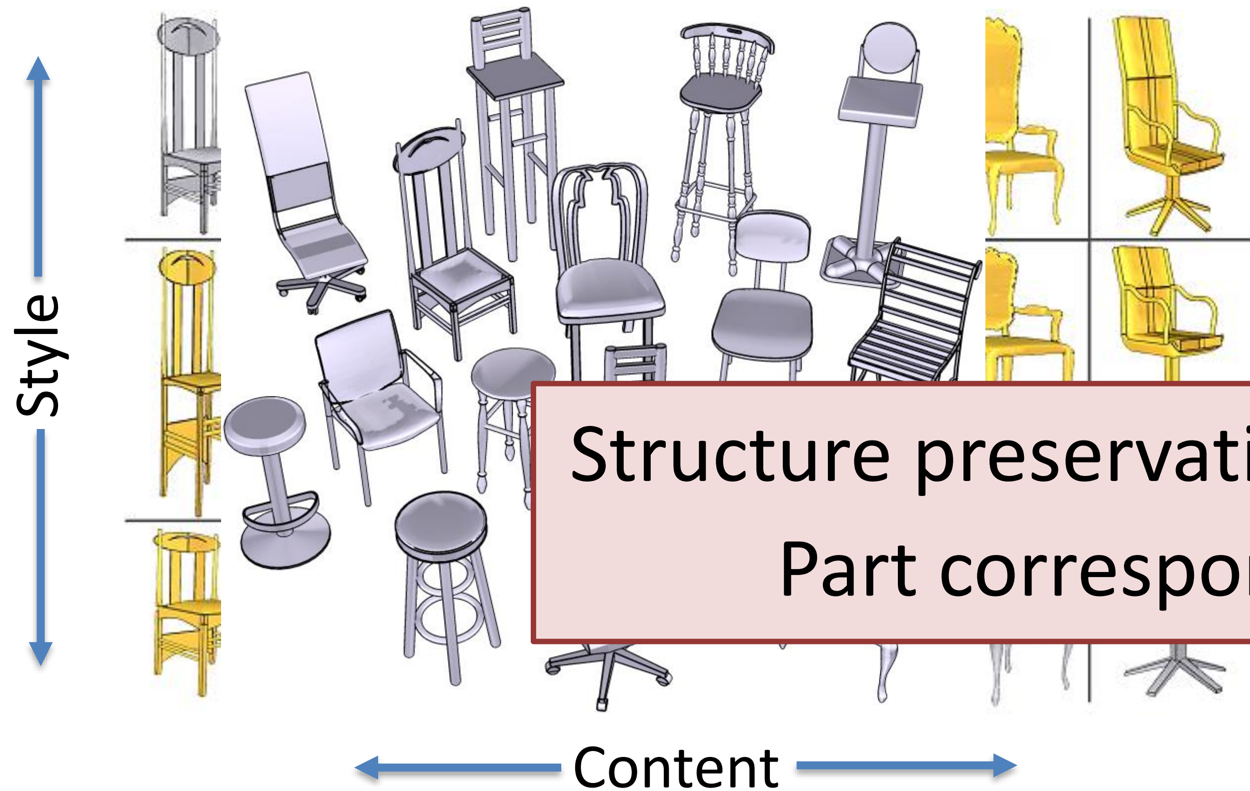
bling

Preserving global structures

Much harder

Our attempts over the past years

- Style-content separation [Xu et al. SIGGRAPH Asia 2010]



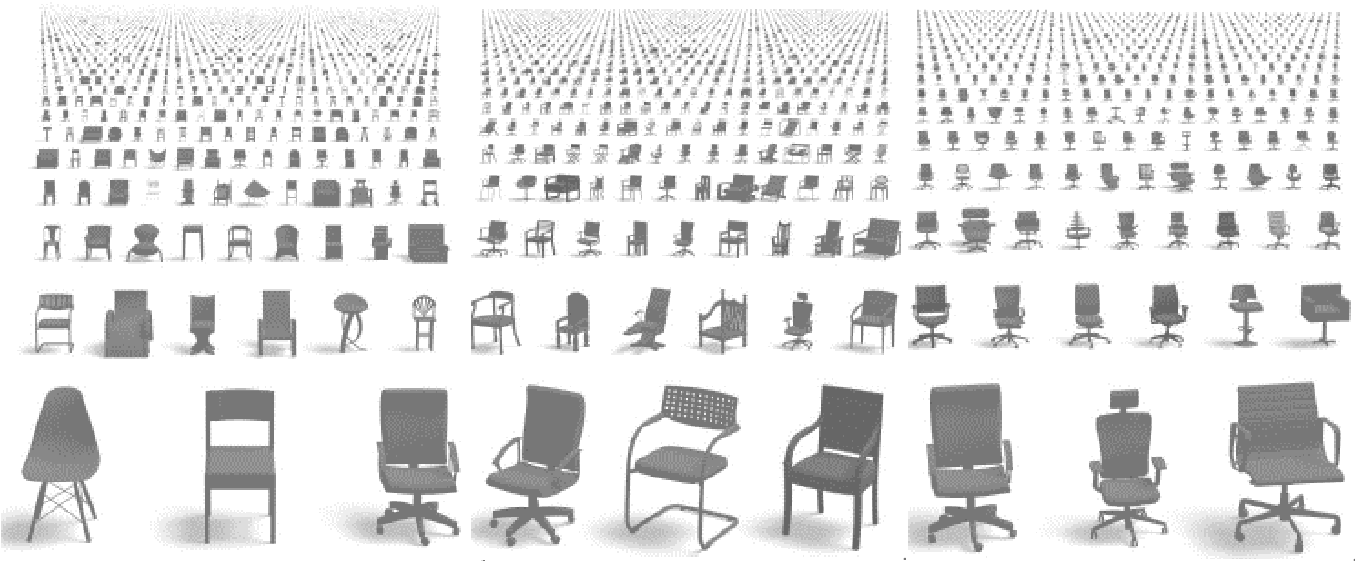
Structure preservation:
Part correspondence

Our attempts over the past years

- Structure-preserving variation driven by photos

[Xu et al. SIGGRAPH 2011]

Photograph



Shape database

Our attempts over the past years

- Structure-preserving variation driven by photos

[Xu et al. SIGGRAPH 2011]

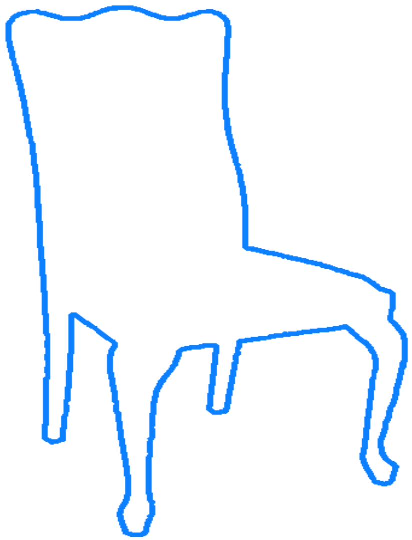
Photograph

Controllers
Shape

Our attempts over the past years

- Structure-preserving variation driven by photos

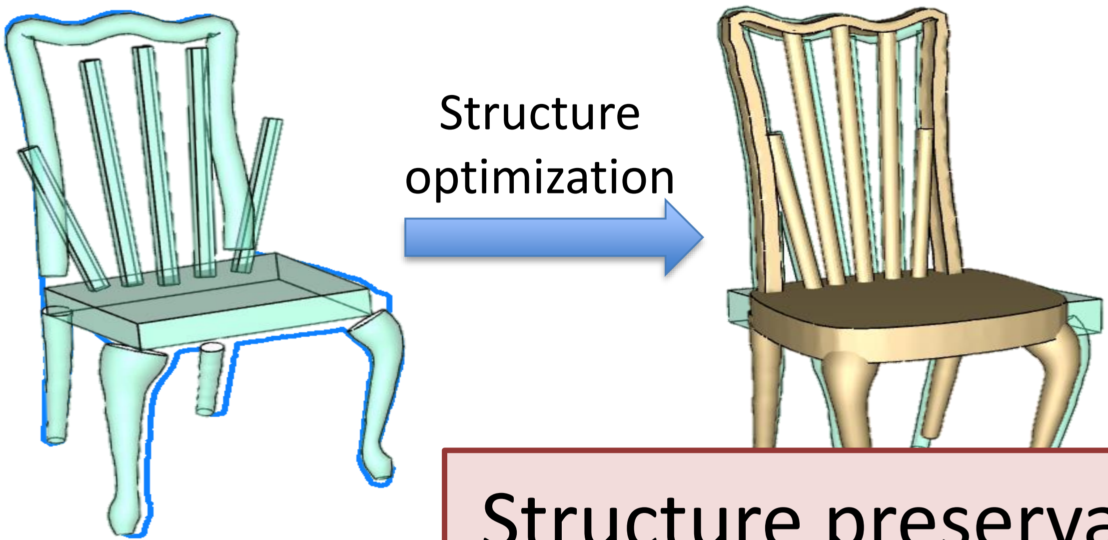
[Xu et al. SIGGRAPH 2011]



Our attempts over the past years

- Structure-preserving variation driven by photos

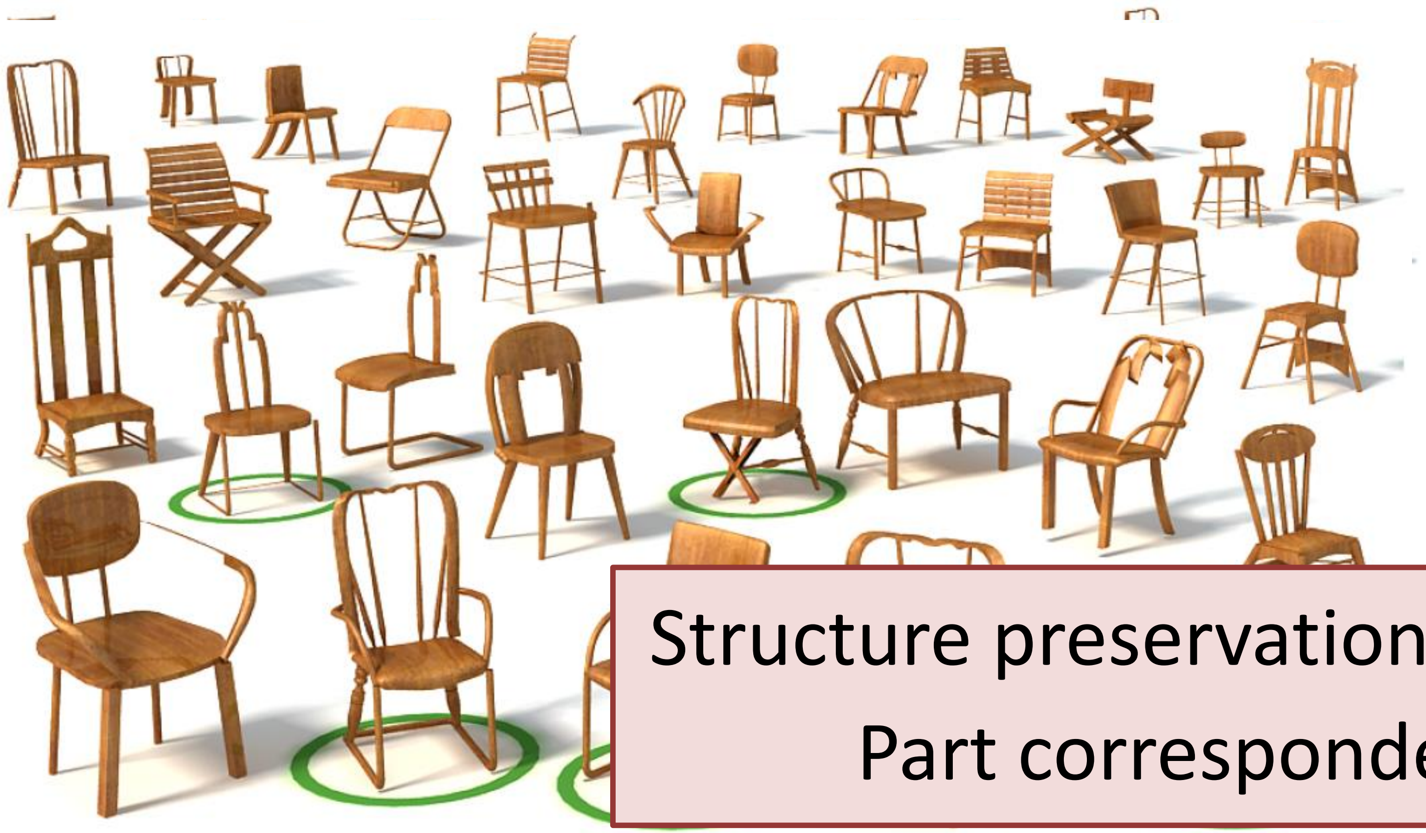
[Xu et al. SIGGRAPH 2011]



Structure preservation:
Hard-coded rules

Our attempts over the past years

- Shape set evolution [Xu et al. SIGGRAPH 2012]

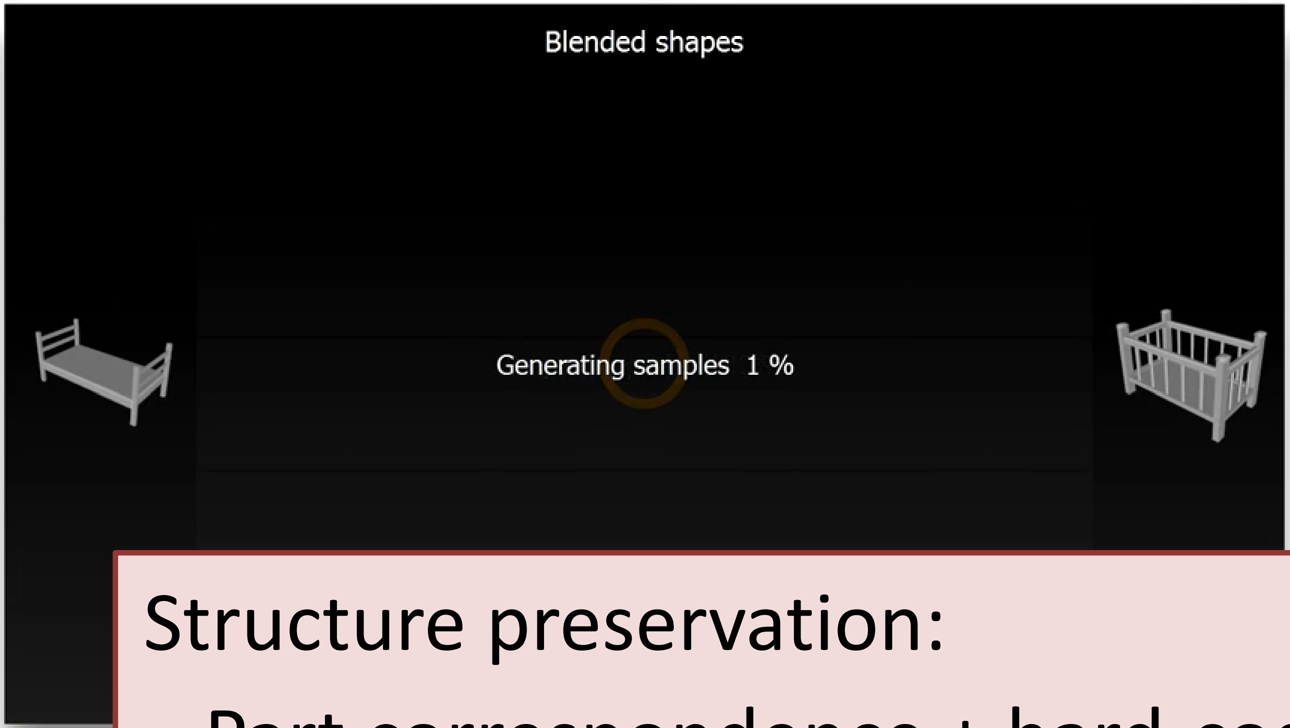


Structure preservation:
Part correspondence

The 5th generation
The 15th generation

Our attempts over the past years

- Structure blending [Alhashim et al. SIGGRAPH 2014]

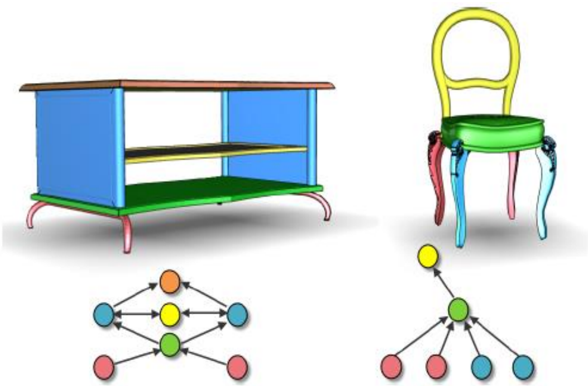


Structure preservation:

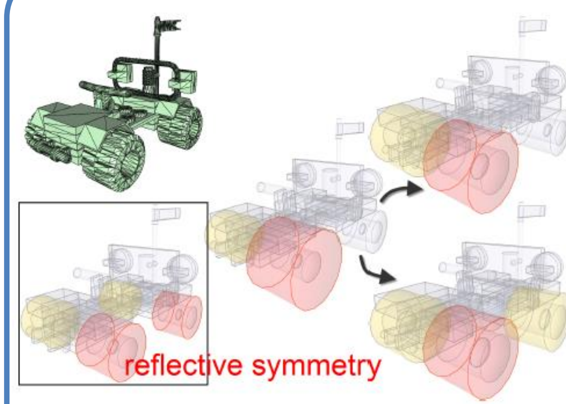
Part correspondence + hard-code rules

Our new goal

- Generating 3D shape variations



Structure preserving



No hard-coded rule

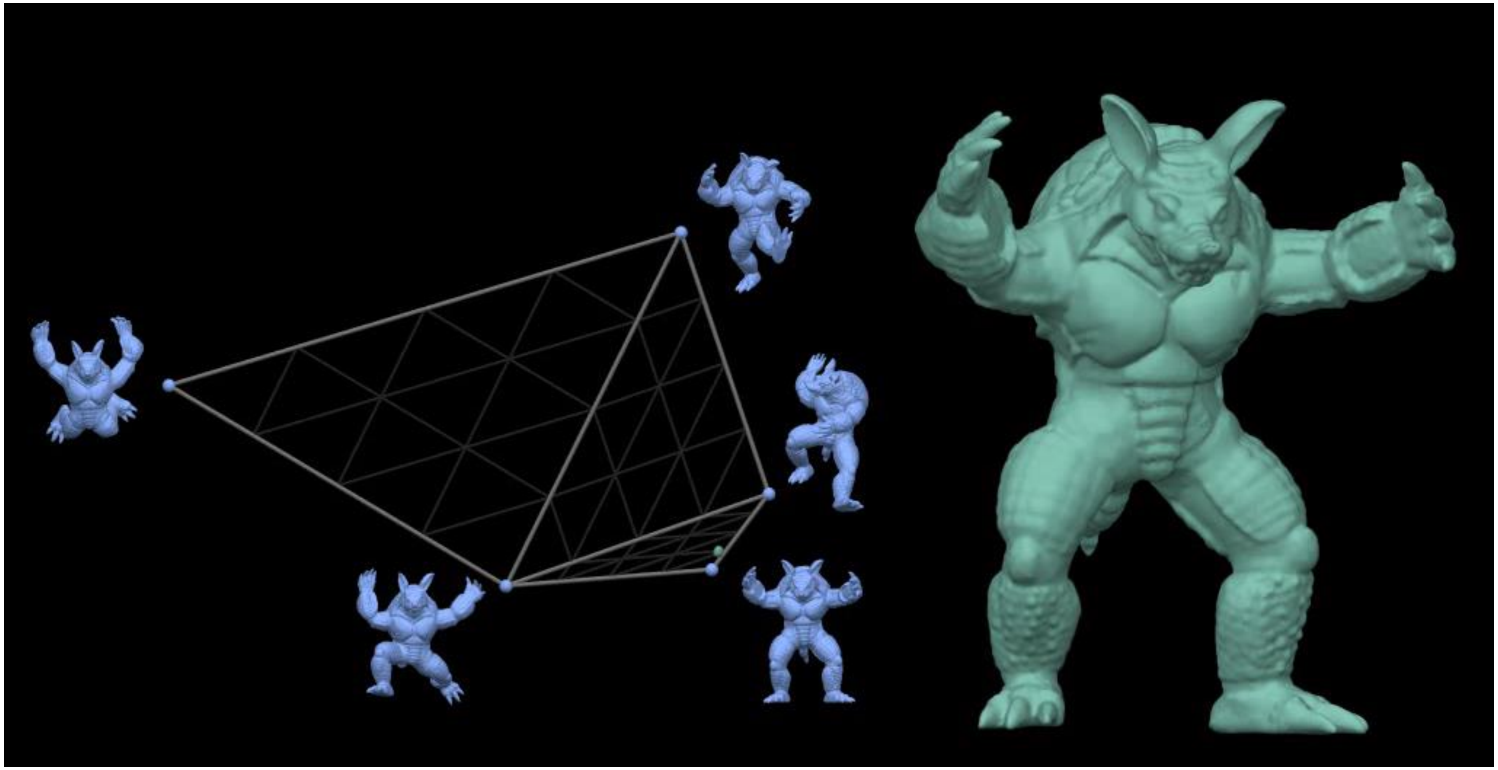
Correspondence-free

Completely data-driven

Completely unsupervised

Our new goal

- More generally, modelling the **structure space** of 3D shapes ?



Geometry space of 3D shapes [Kilian et al. 2007]

Our new goal

- More generally, modelling the **structure space** of 3D shapes ?

How about 3D shape structures?

✓ Structure preserving
✓ Completely data-driven
✓ Correspondence-free

Modeling 3D Structure Space

Our approach

- Basic idea:
  - Learn a **distribution** that approximates the data distribution of true 3D structures

$$P(X) \approx P_{gt}(X)$$

  - Marginalize over a latent variable

$$\text{maximize } P(X) = \int \underbrace{\boxed{P(X|\boxed{z};\boxed{\theta})}}_{\text{Likelihood}} P(z)\,dz$$

Likelihood

Parameters

Variational Bayesian formulation

maximize $P(X) = \int P(X|z;\theta)P(z)dz$

maximize $\boxed{E_{z\sim Q}\left[\log P(X|z)\right]} - \boxed{\mathcal{D}\left[Q(z|X)\|P(z)\right]}$

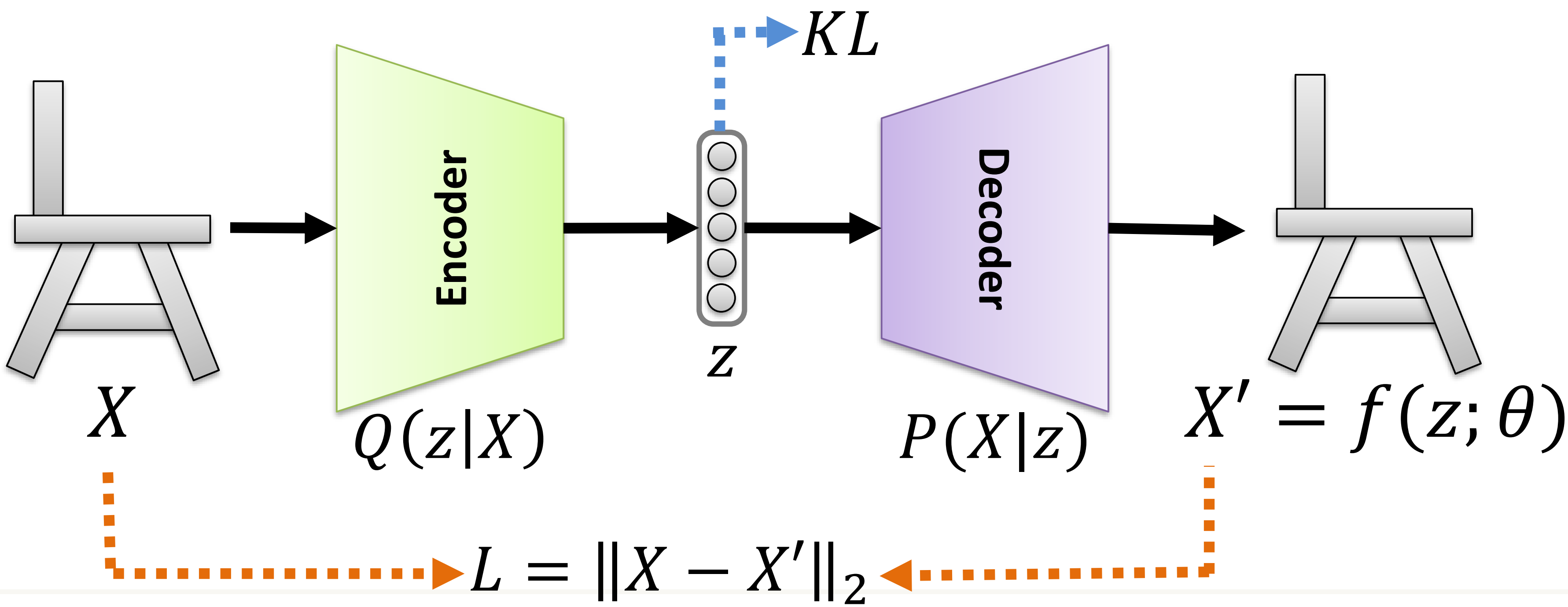
z should reconstruct X, given that it was drawn from $Q(z|X)$

Assuming z's follow a normal distribution

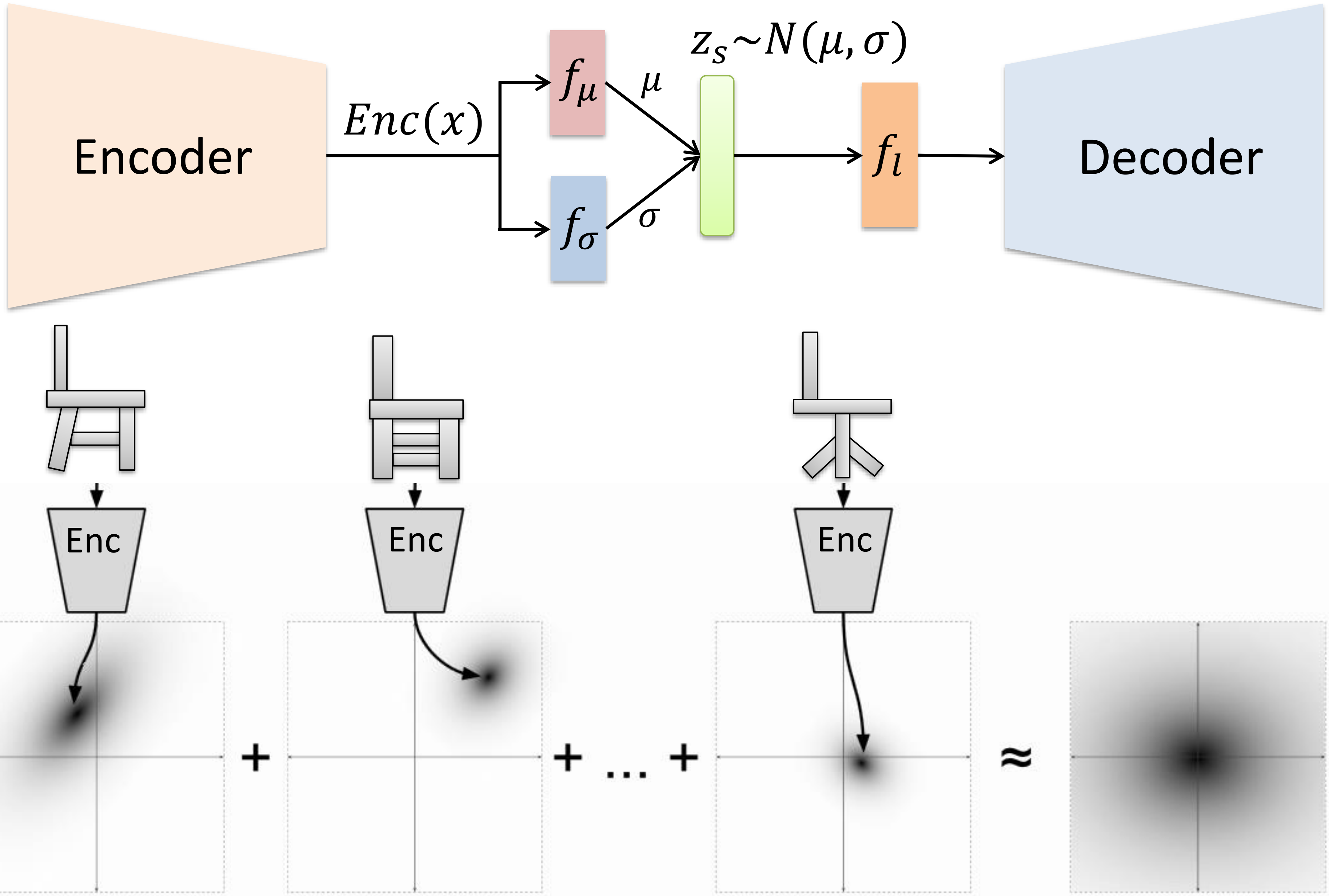
Variational auto-encoder

maximize $\boxed{E_{z\sim Q}[\log P(X|z)]} - \boxed{\mathcal{D}[Q(z|X)\|P(z)]}$

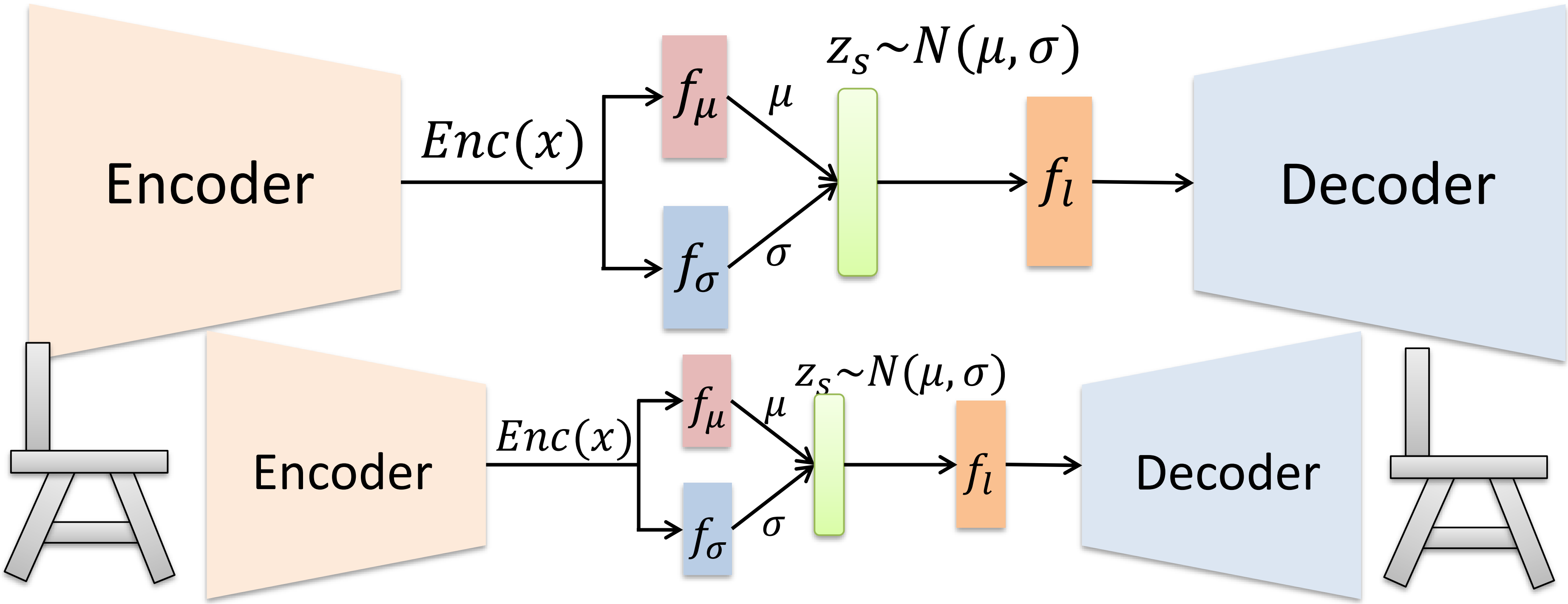
Reconstruction loss | KL divergence loss



Variational auto-encoder



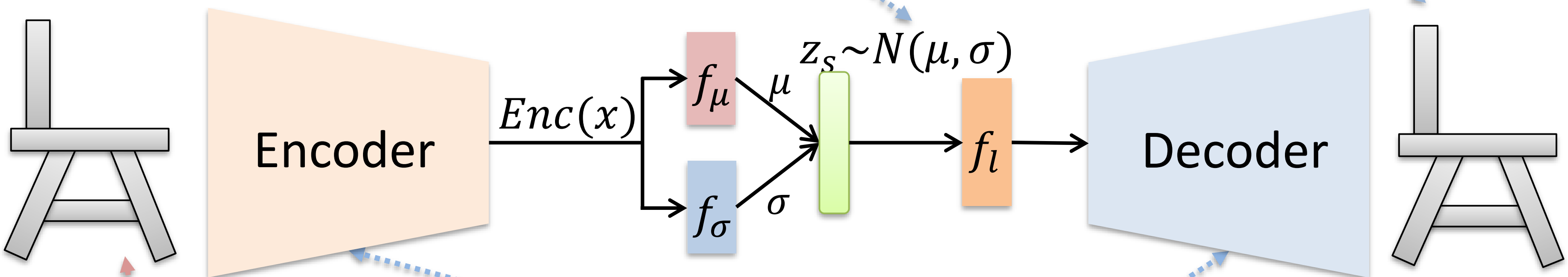
Variational auto-encoder



Remaining issues

How to deal with sampling far away from μ?

How to measure the quality of generated structures?



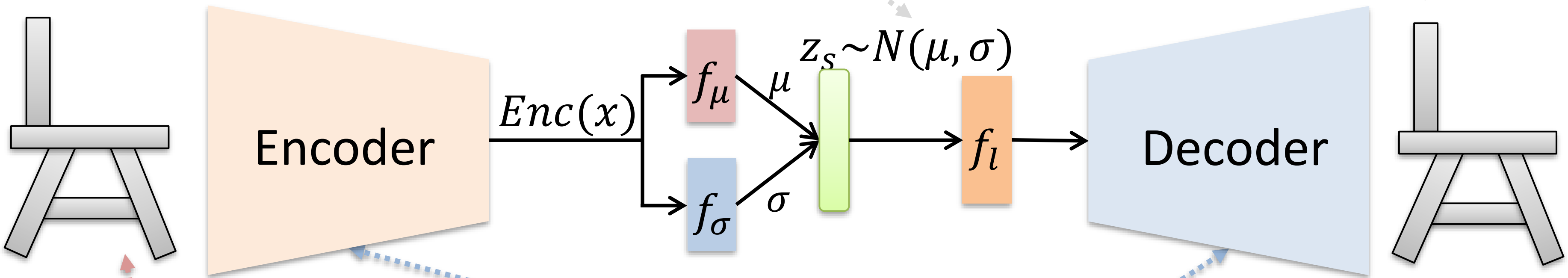
What would be a good representation for input structures?

How to encode / decode 3D shape structures?

Remaining issues

How to deal with sampling far away from μ?

How to measure the quality of generated structures?

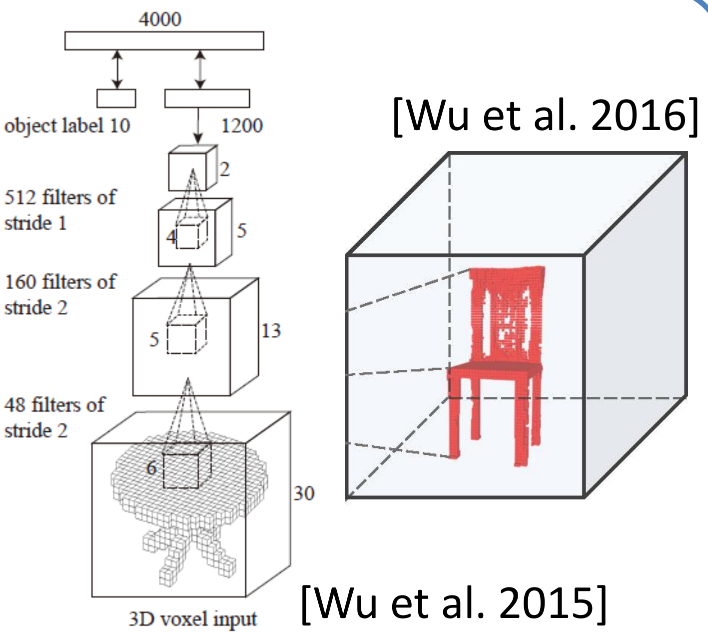


What would be a good representation for input structures?

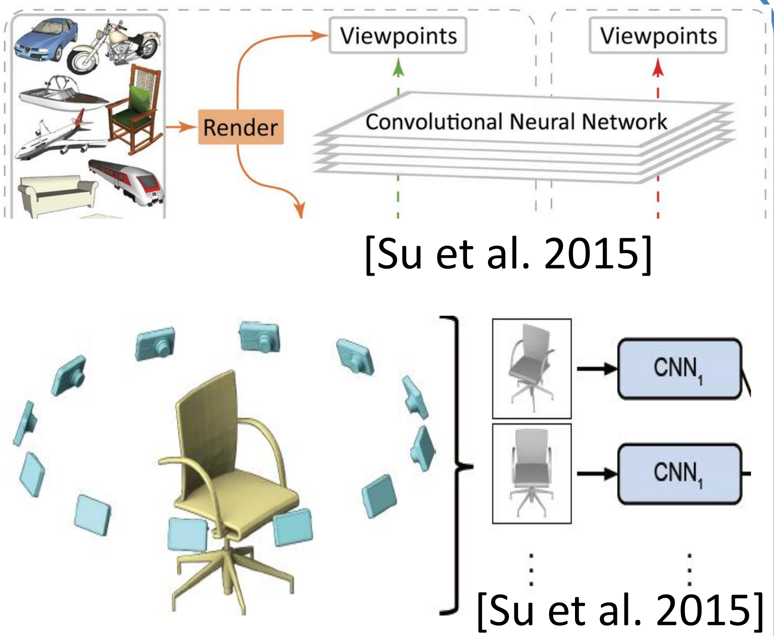
How to encode / decode 3D shape structures?

3D shape representation for DL

- Typically two methods



Volumetric representation

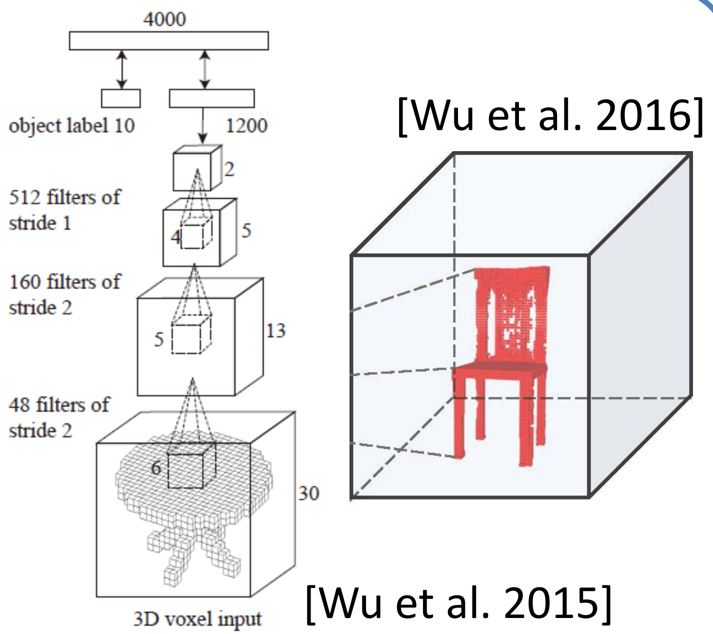


Multi-view rendered images

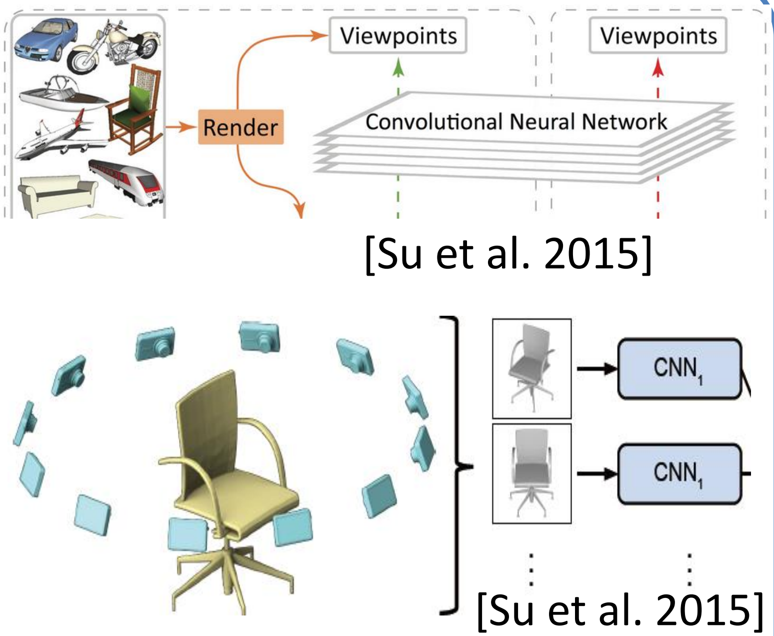
Good for visual classification & recognition

3D shape representation for DL

- Typically two methods



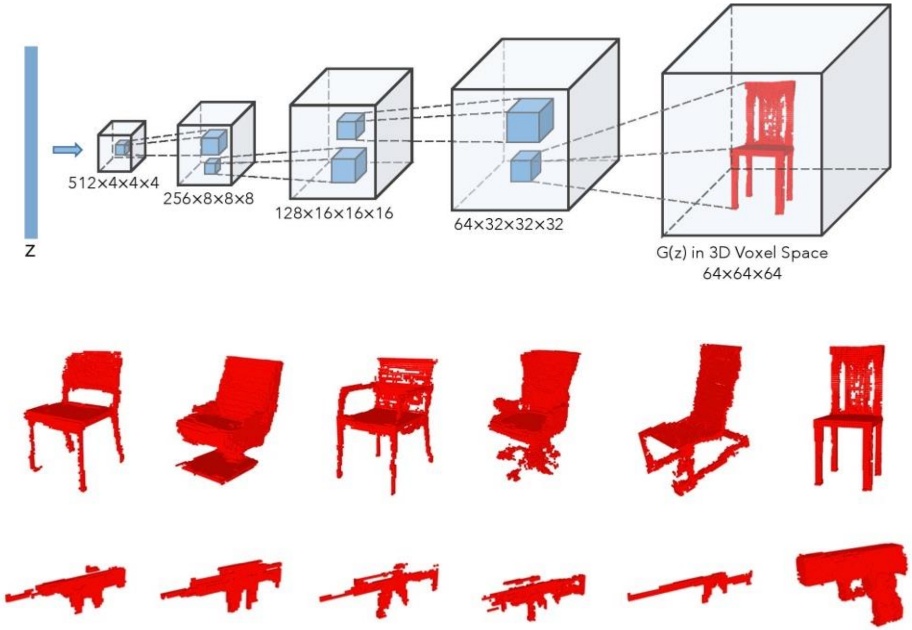
Volumetric representation

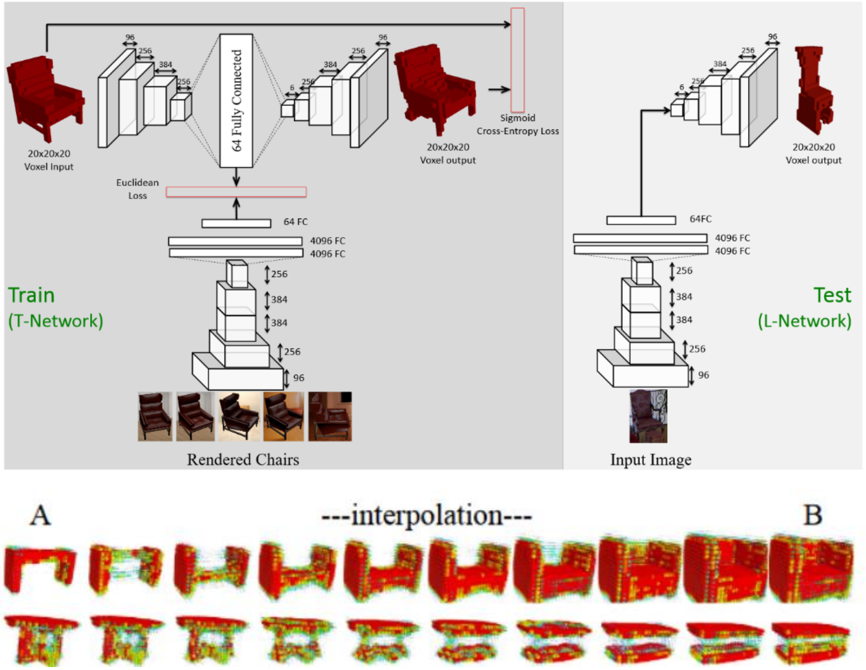


Multi-view rendered images

Limitation: Oblivious to structure!

Structure-oblivious 3D shape generation

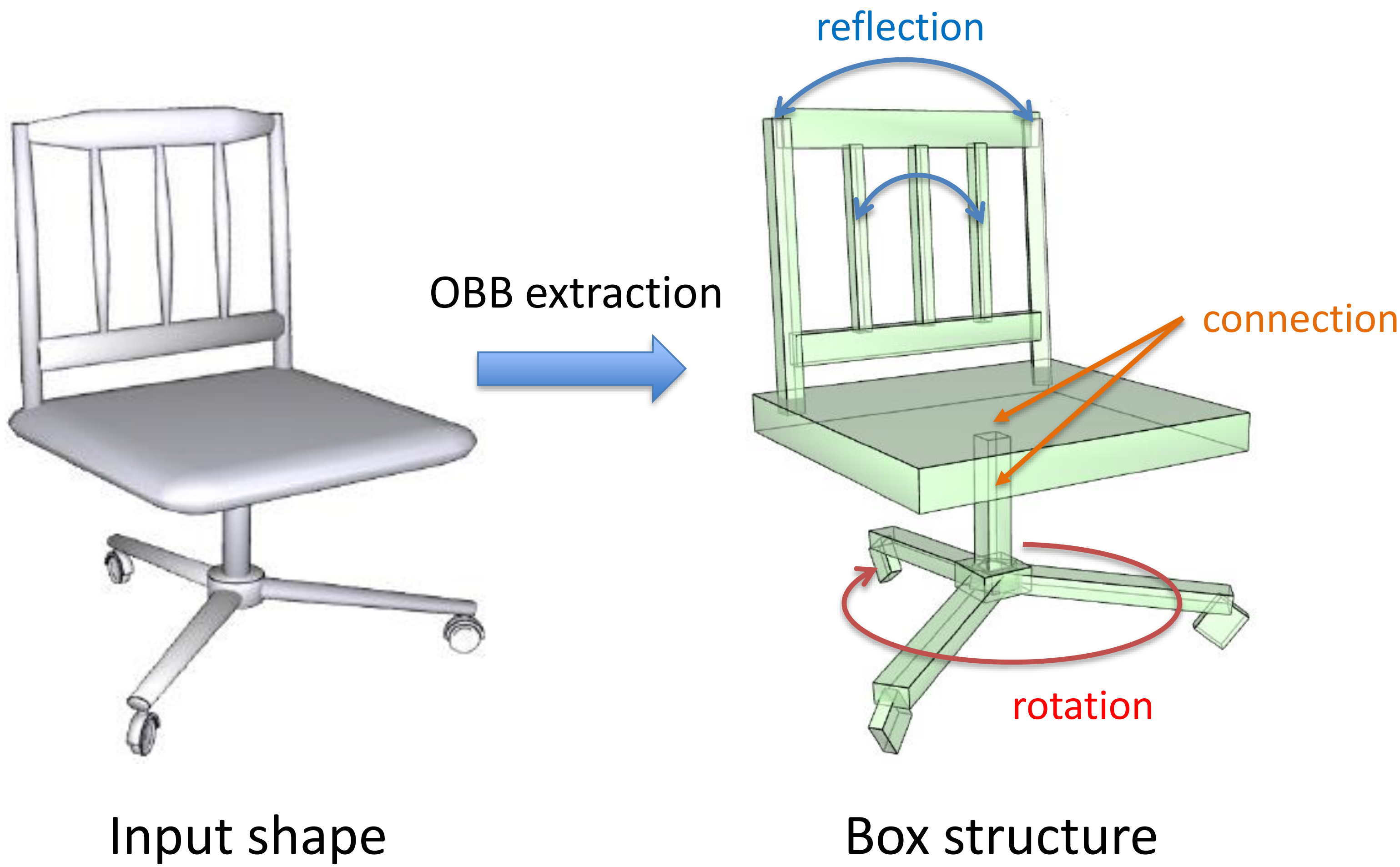




[Wu et al. 2016] by MIT

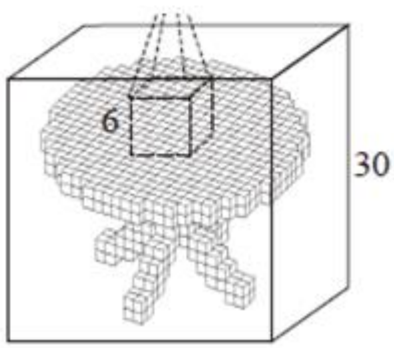
[Girdhar et al. 2016] by CMU

Structure-aware representation

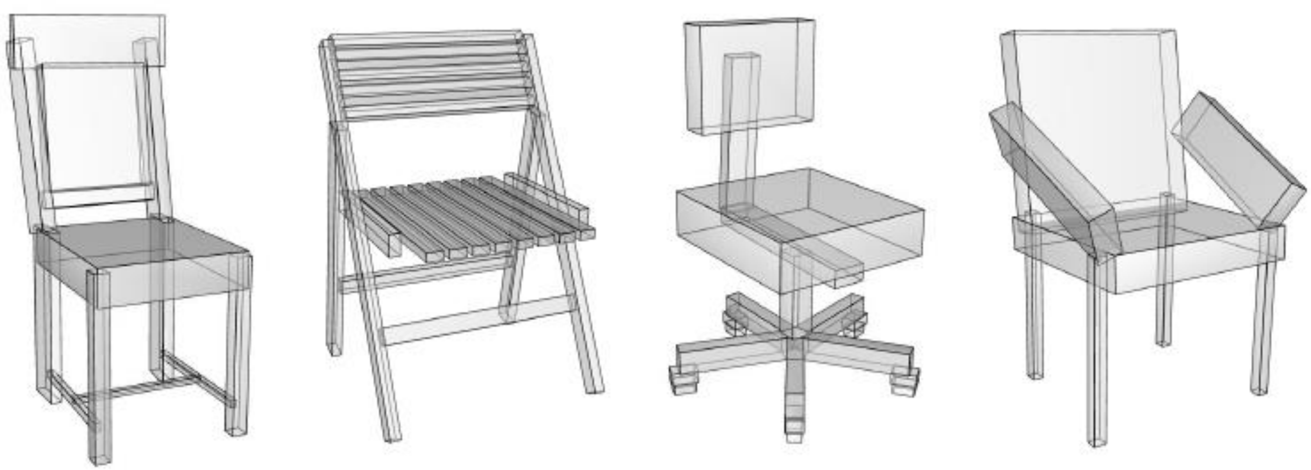


Structure-aware representation

- Problems with box structure representation
  - Number of boxes varies from shape to shape
    - ☹ Not neural networks friendly



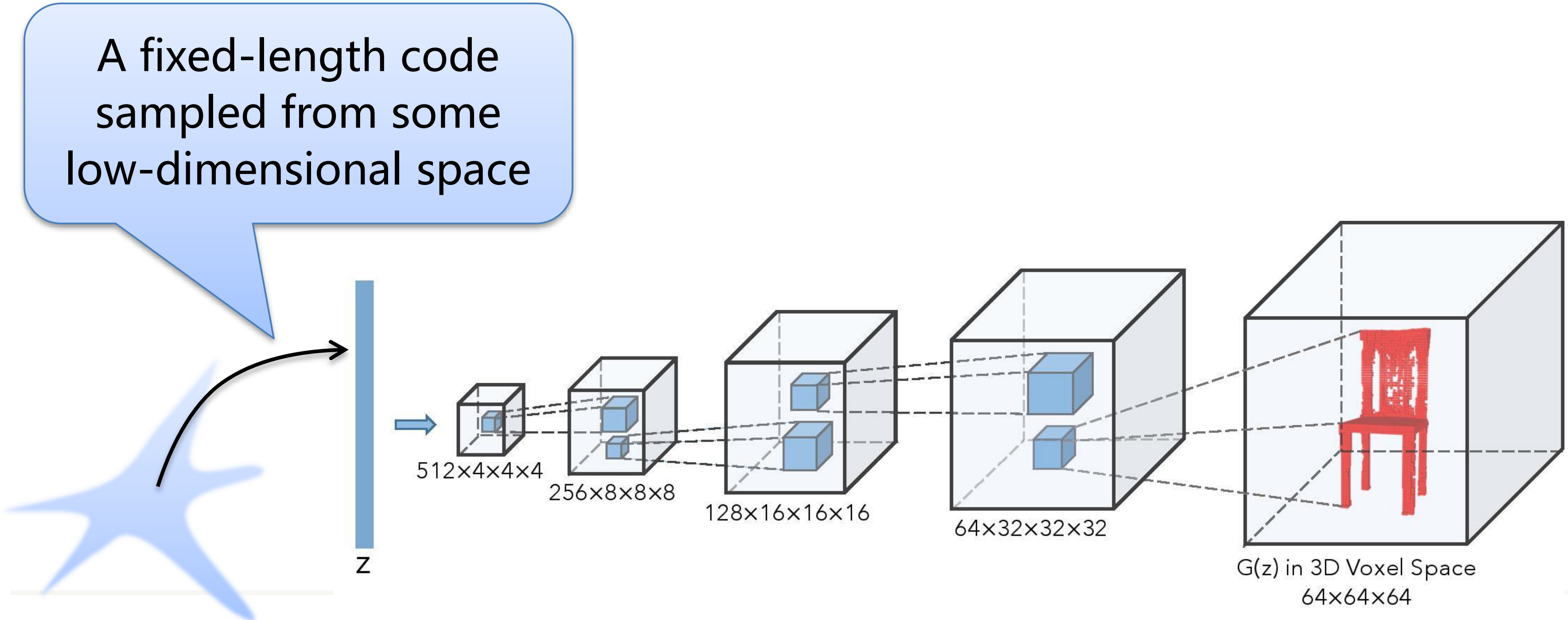
Volume representation



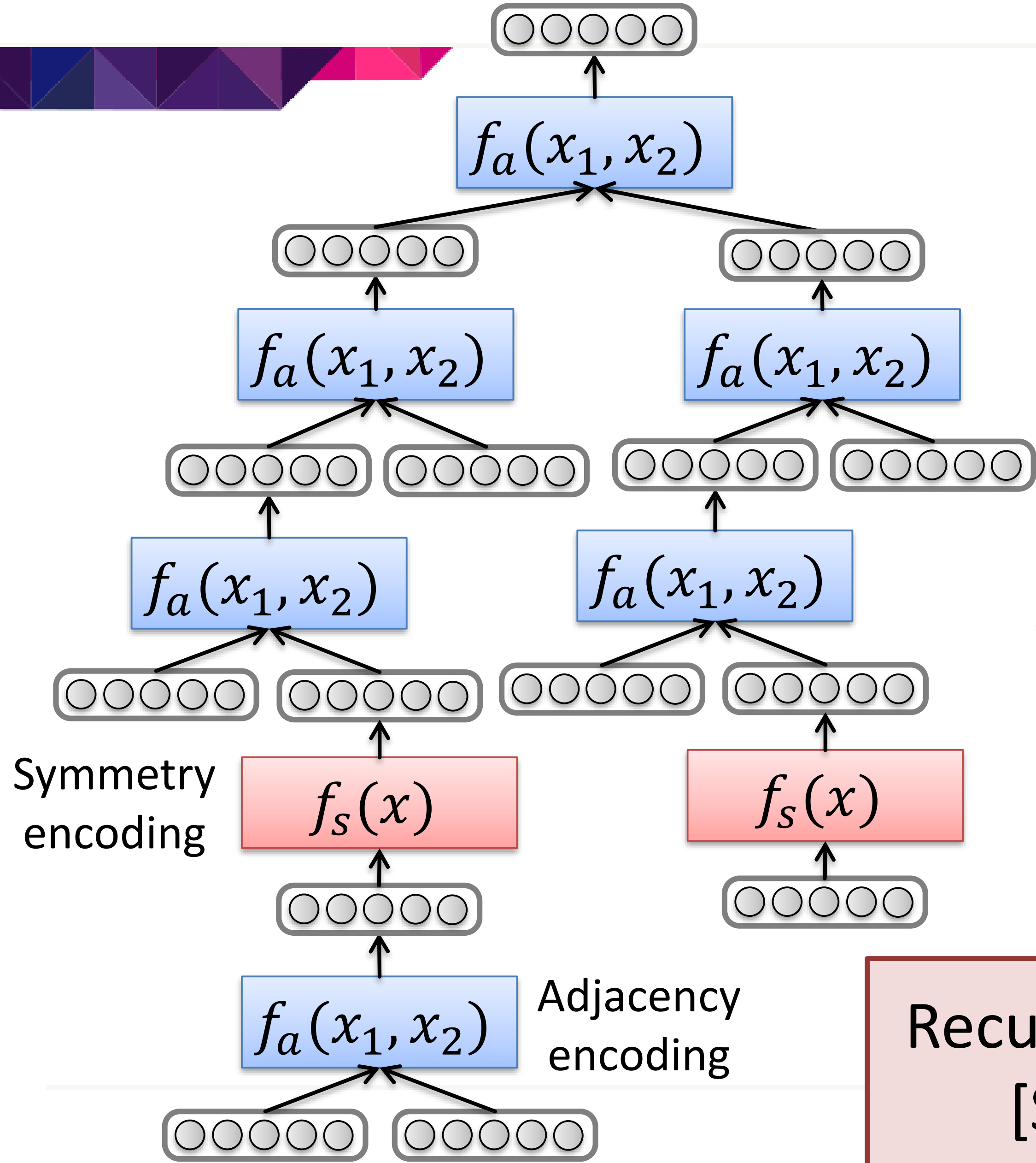
Box representation

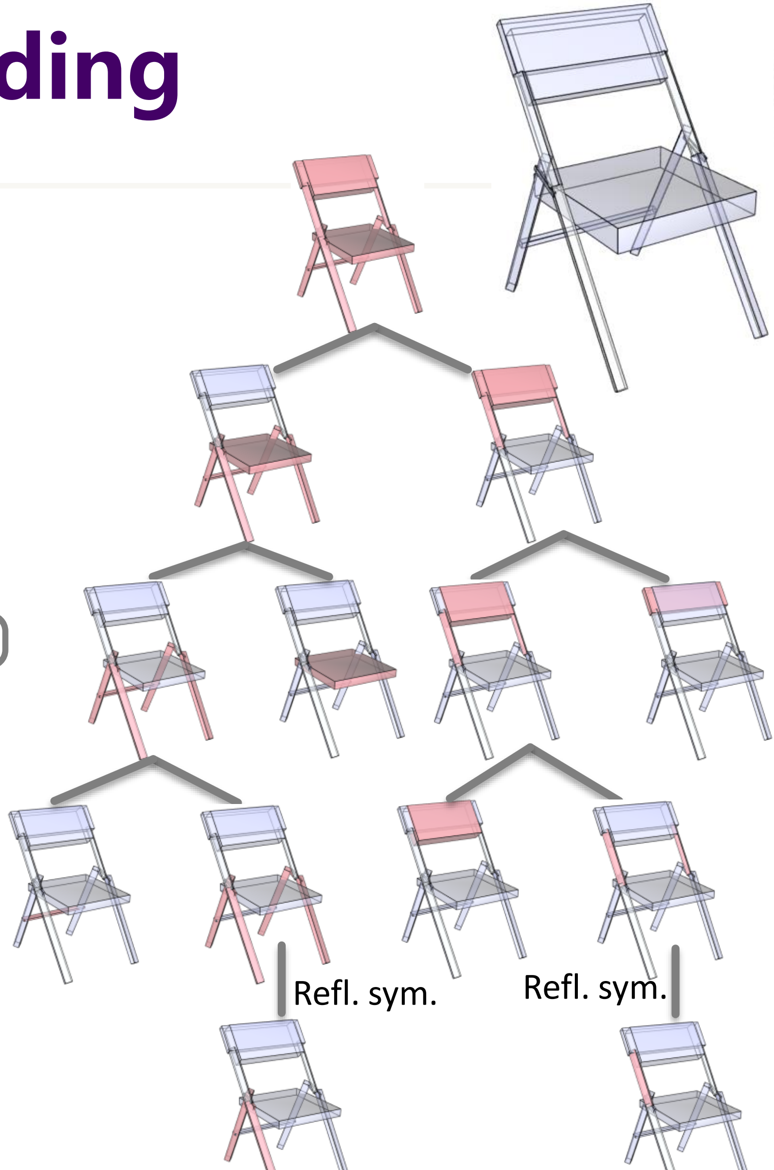
Structure-aware representation

- Problems with box structure representation
  - Number of boxes varies from shape to shape
    - ☹ Encode the whole structure into a fixed length code?



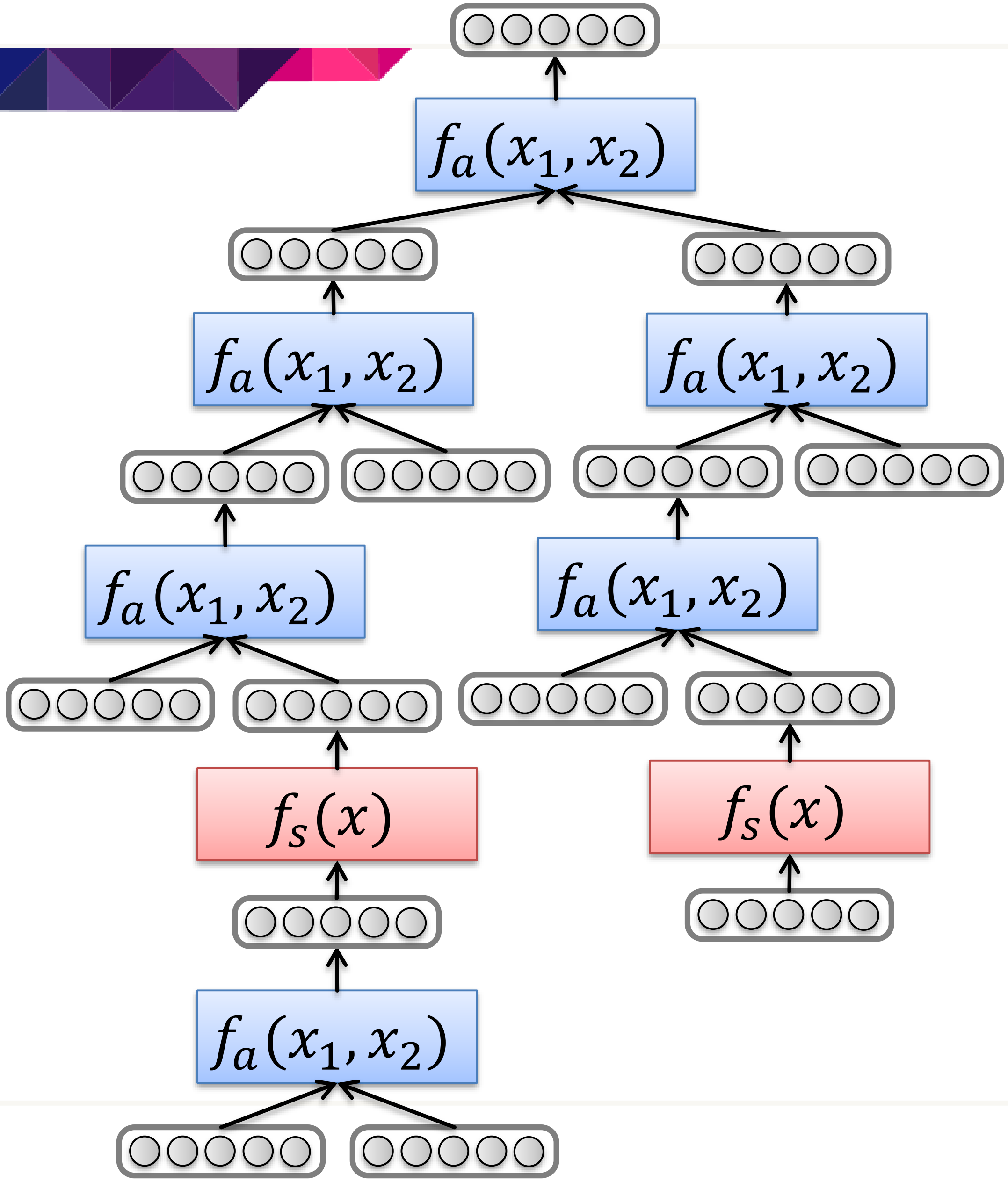
Structure encoding / decoding





Recursive Neural Networks
[Socher et al. 2011]

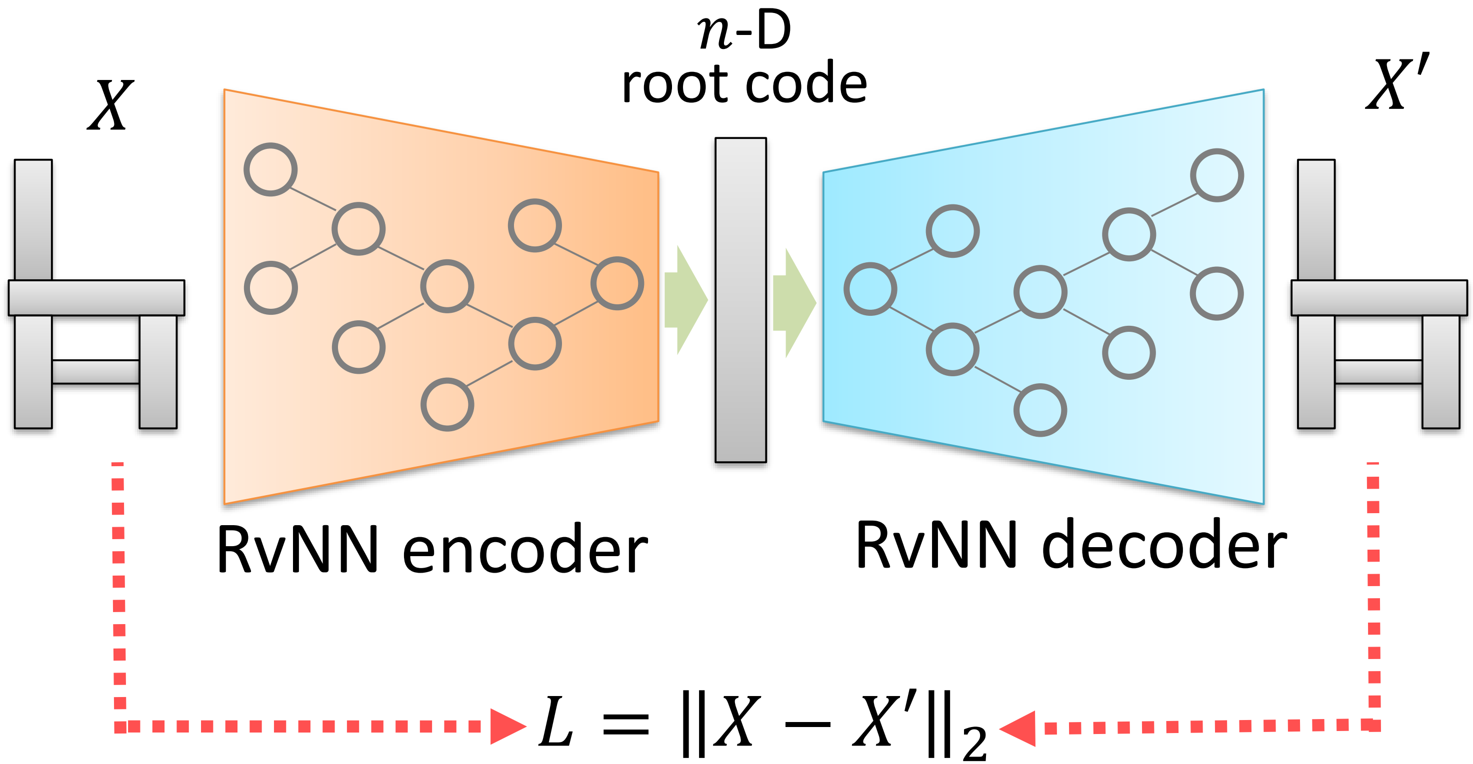
Recursive structure encoding



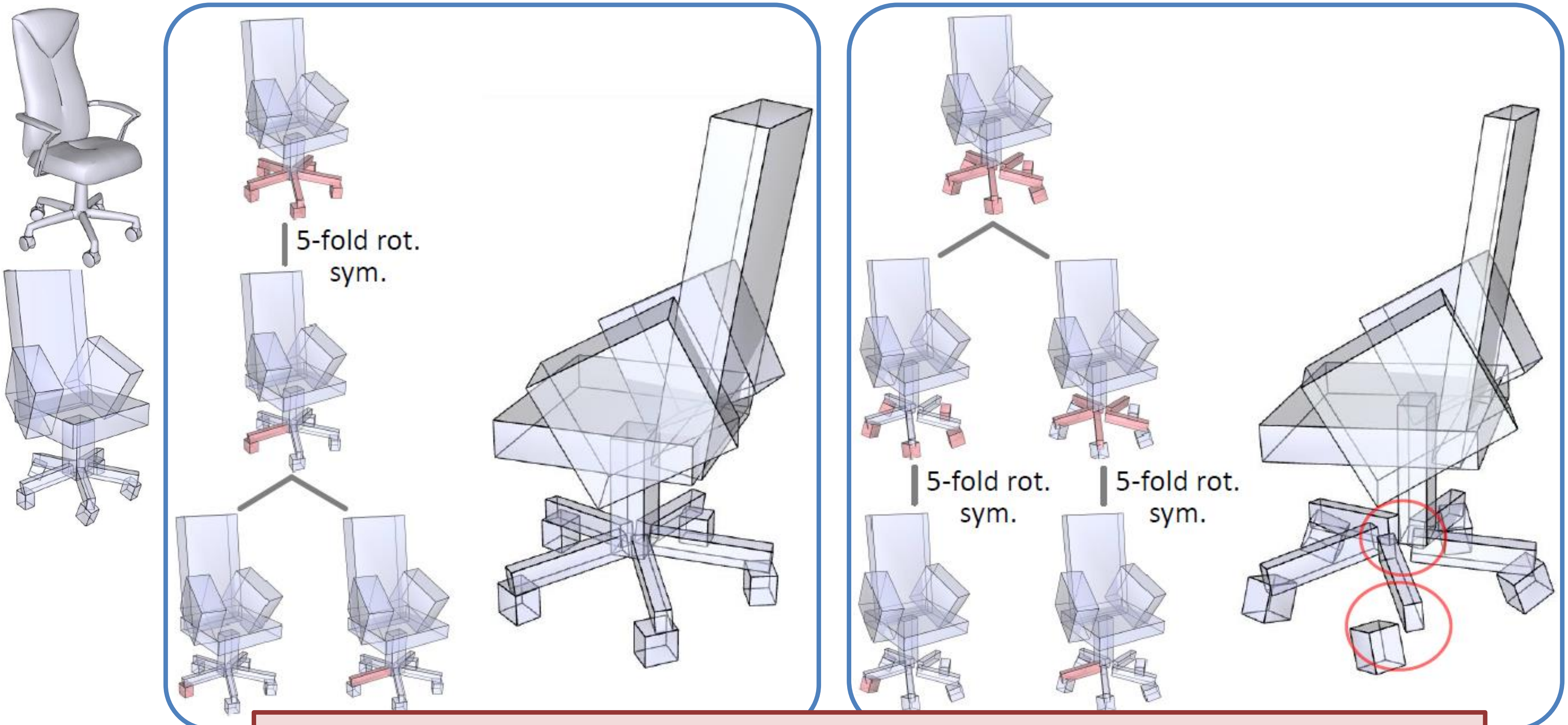
- Shared weights across different levels
- A fixed-length root code encodes the whole structure
- Different networks for adjacency and symmetry
- How to determine the grouping order?

Recursive auto-encoder

- Self-reconstruction



Recursive auto-encoder

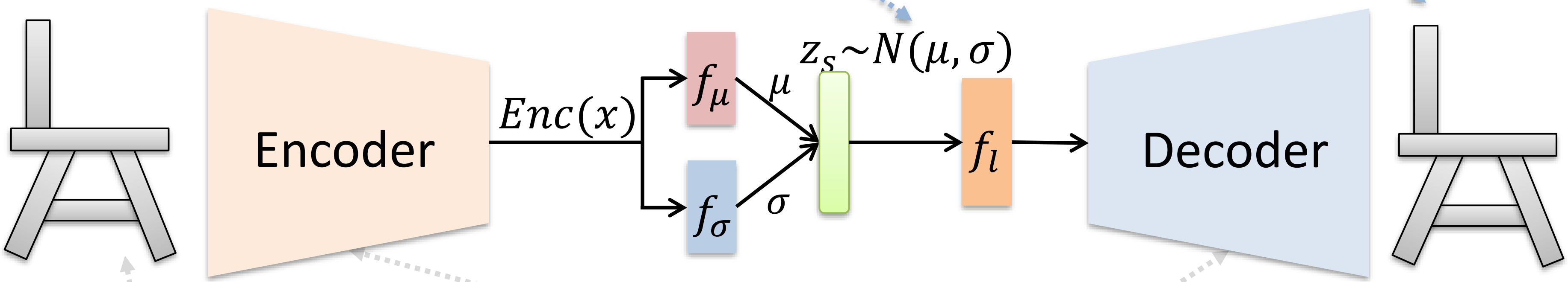


Perceptually meaningful grouping leads to low self-reconstruction error!

Remaining issues

How to deal with sampling far away from μ?

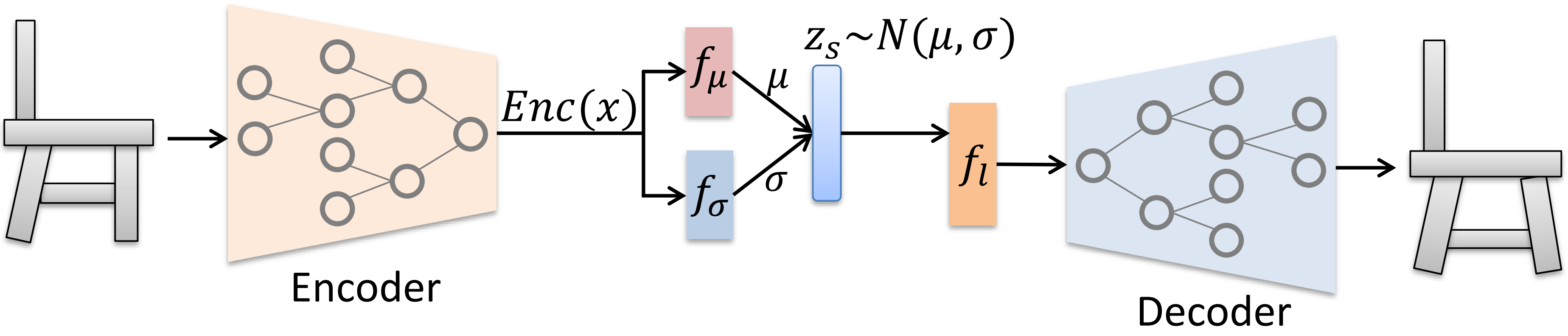
How to measure the quality of generated structures?

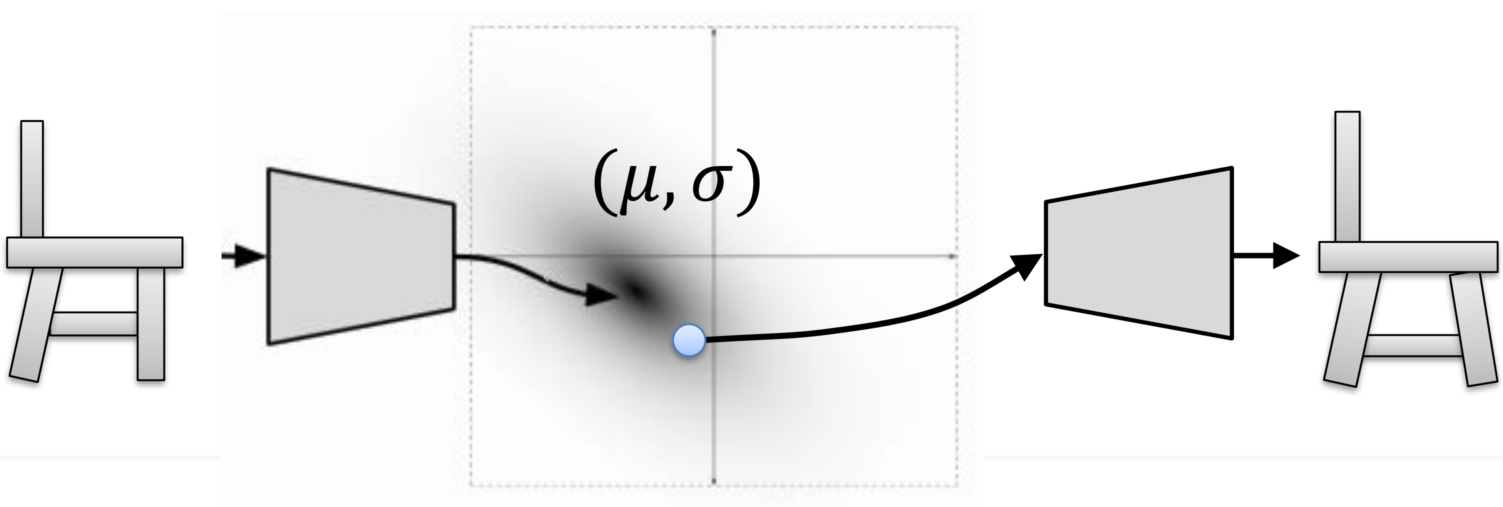


What would be a good representation for input structures?

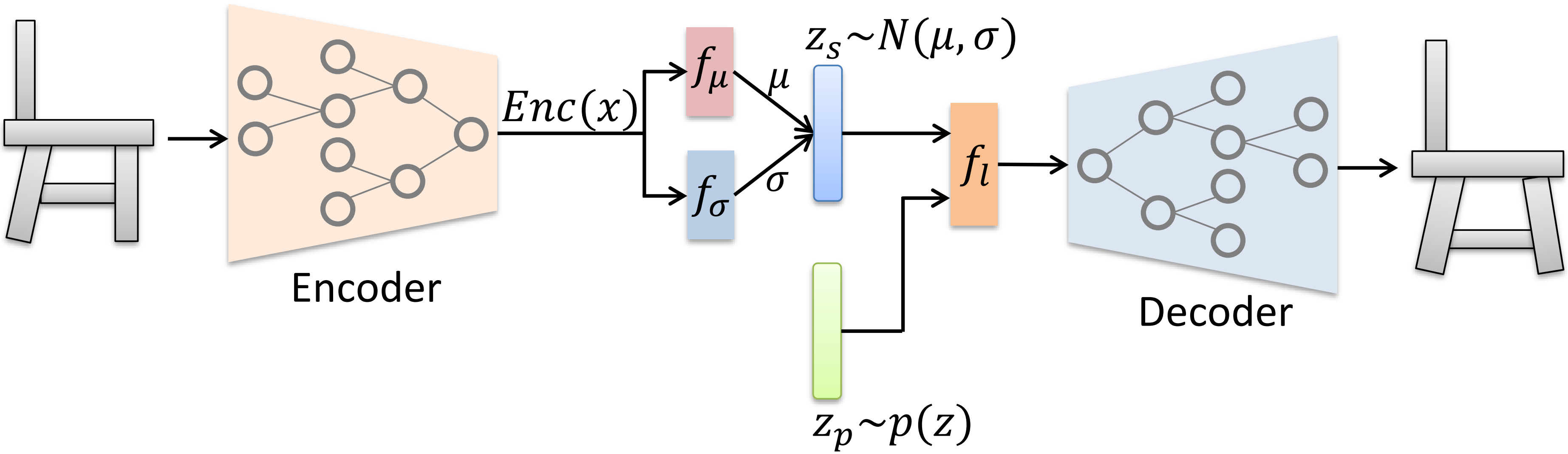
How to encode / decode 3D shape structures?

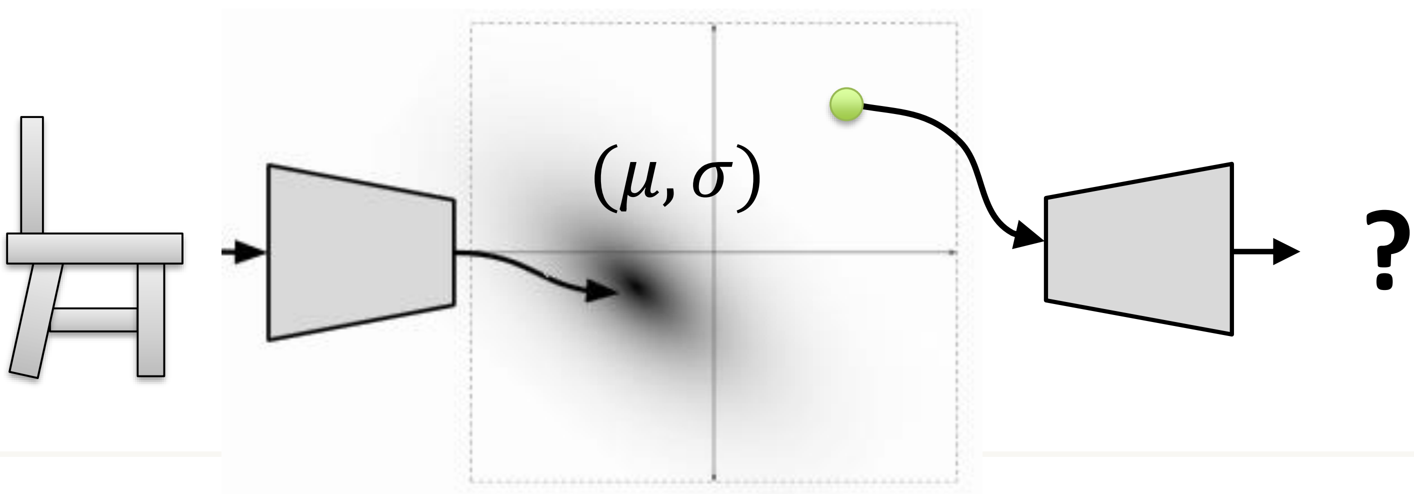
Sampling far away from μ





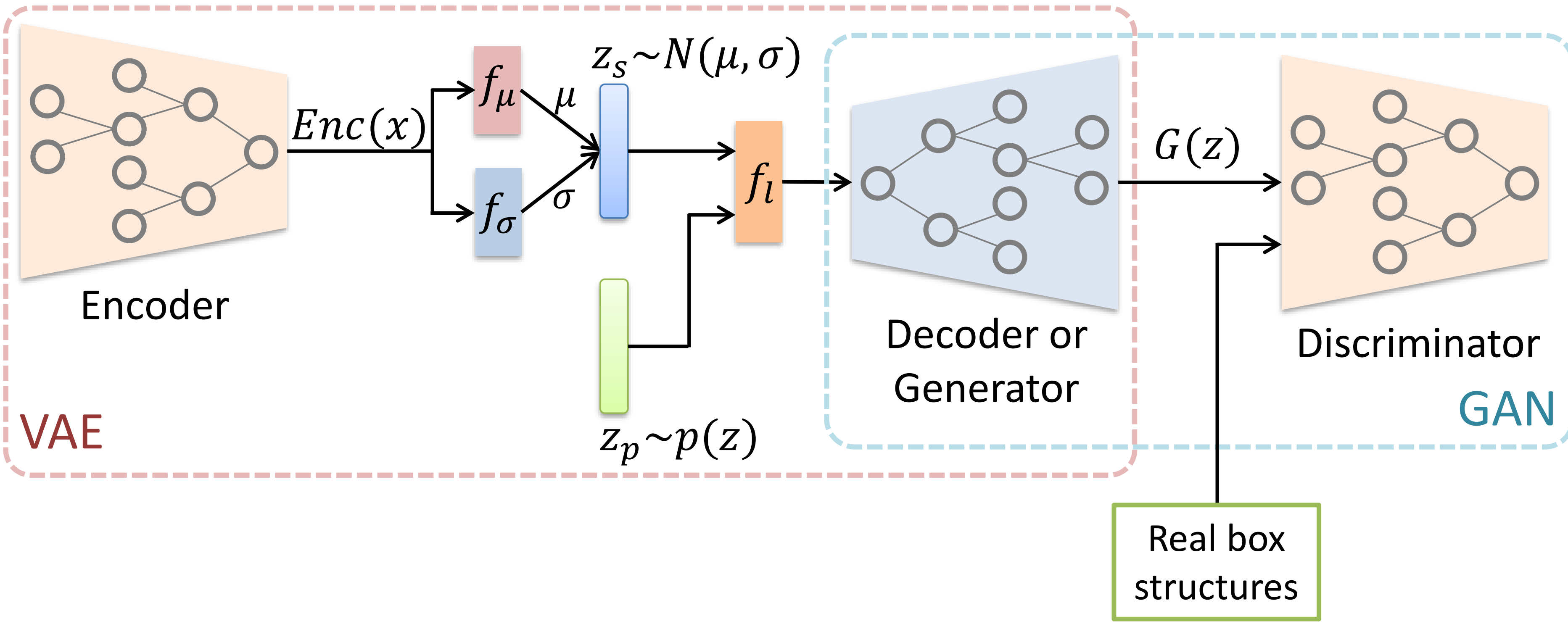
Sampling far away from μ





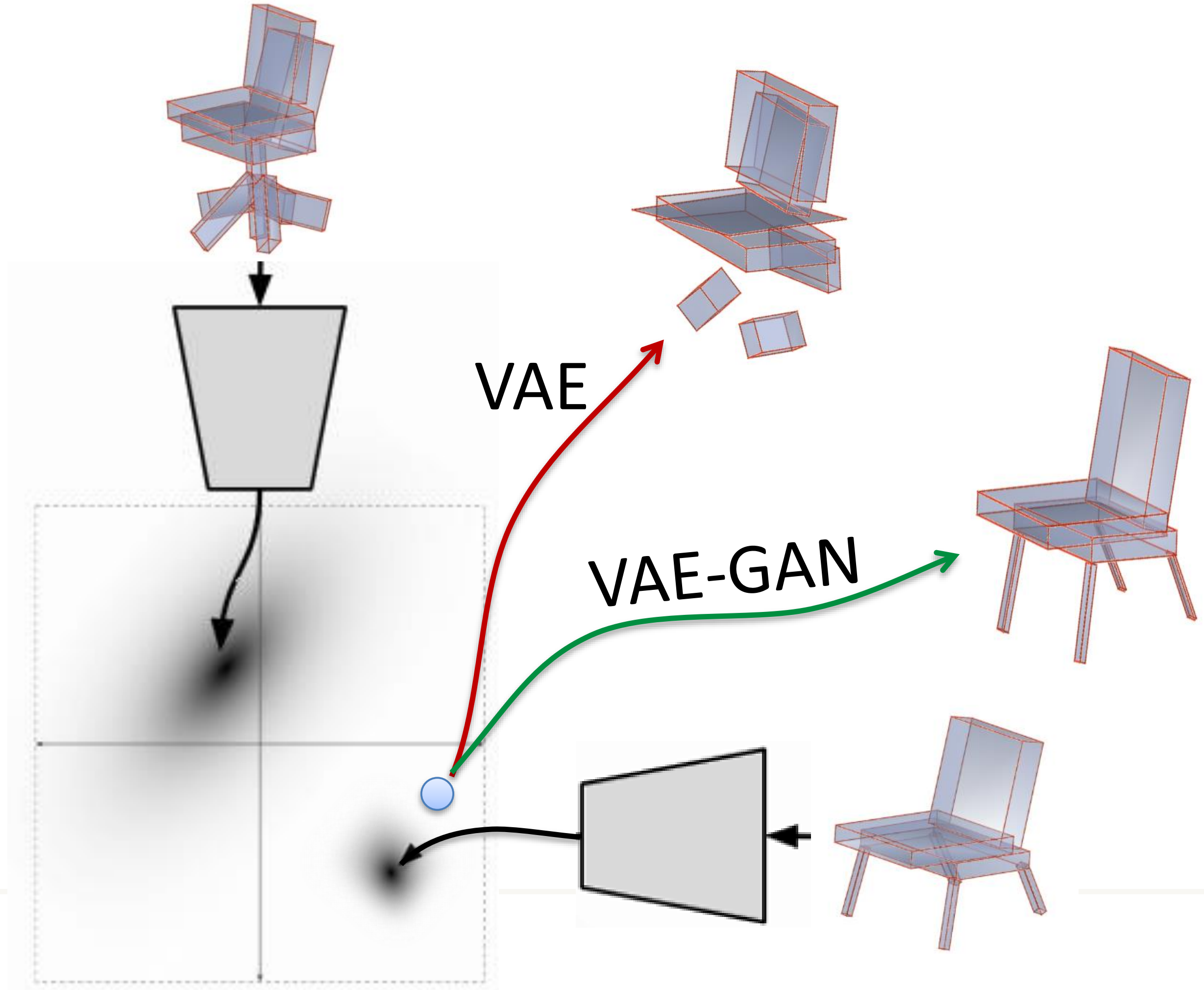
Adversarial training

- VAE-**GAN** (Generative Adversarial Network) architecture

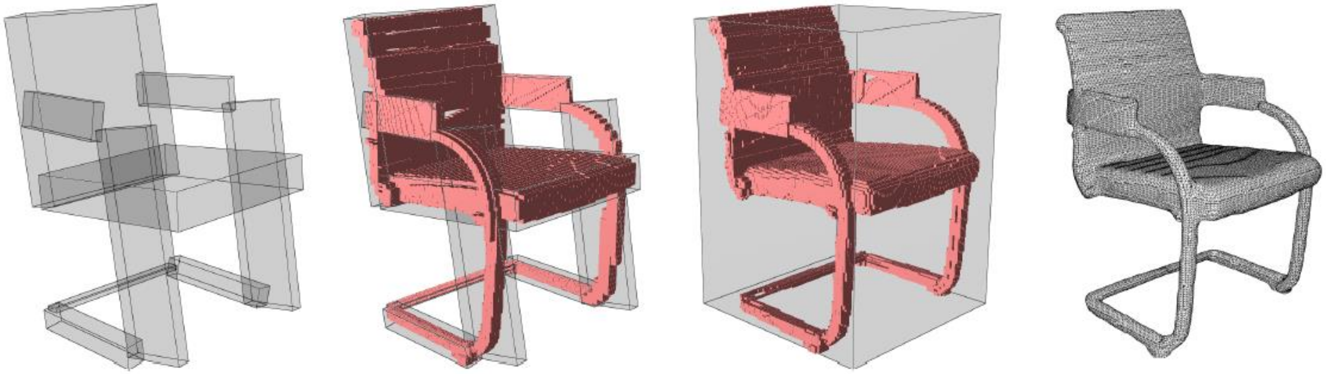


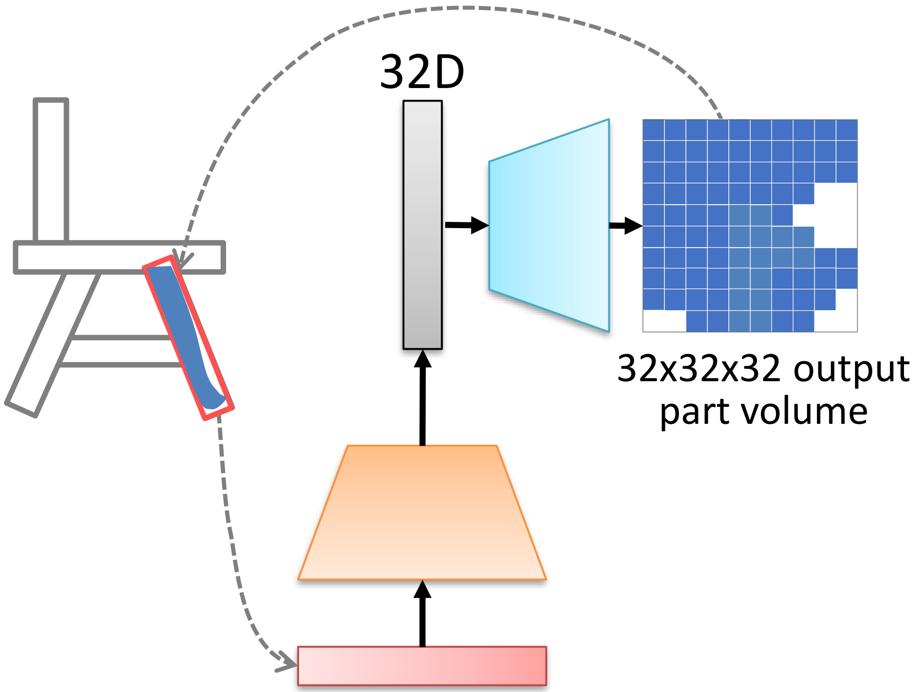
Adversarial training

- Benefit of VAE-GAN

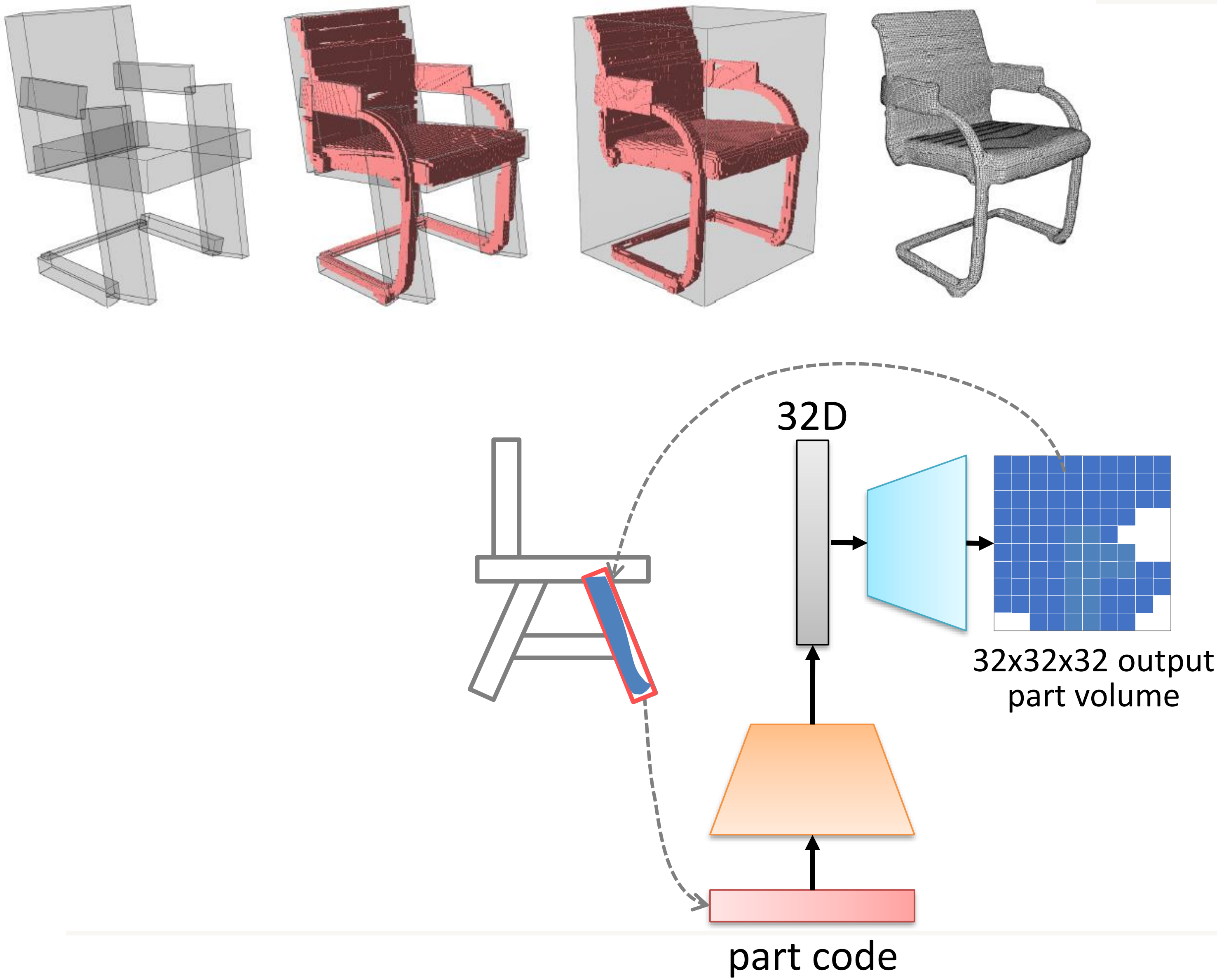


Synthesize geometry within a box

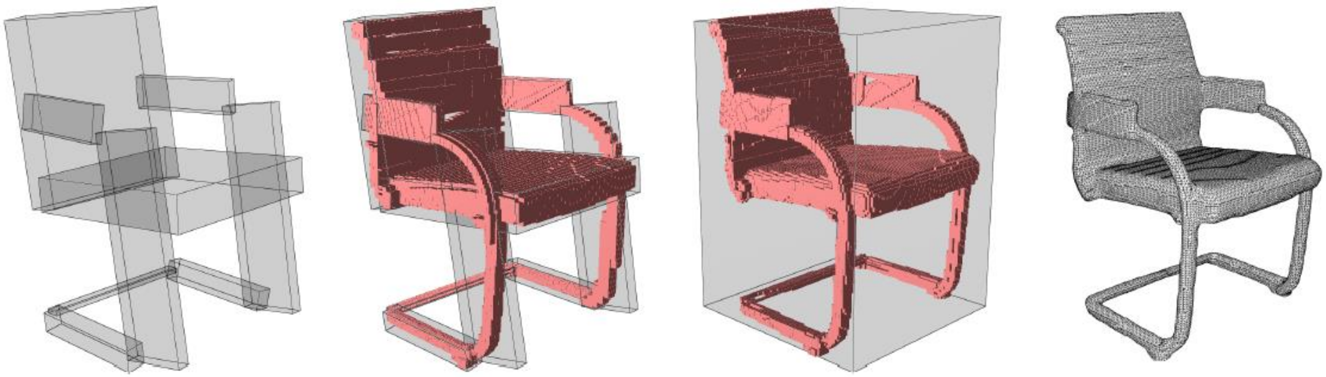


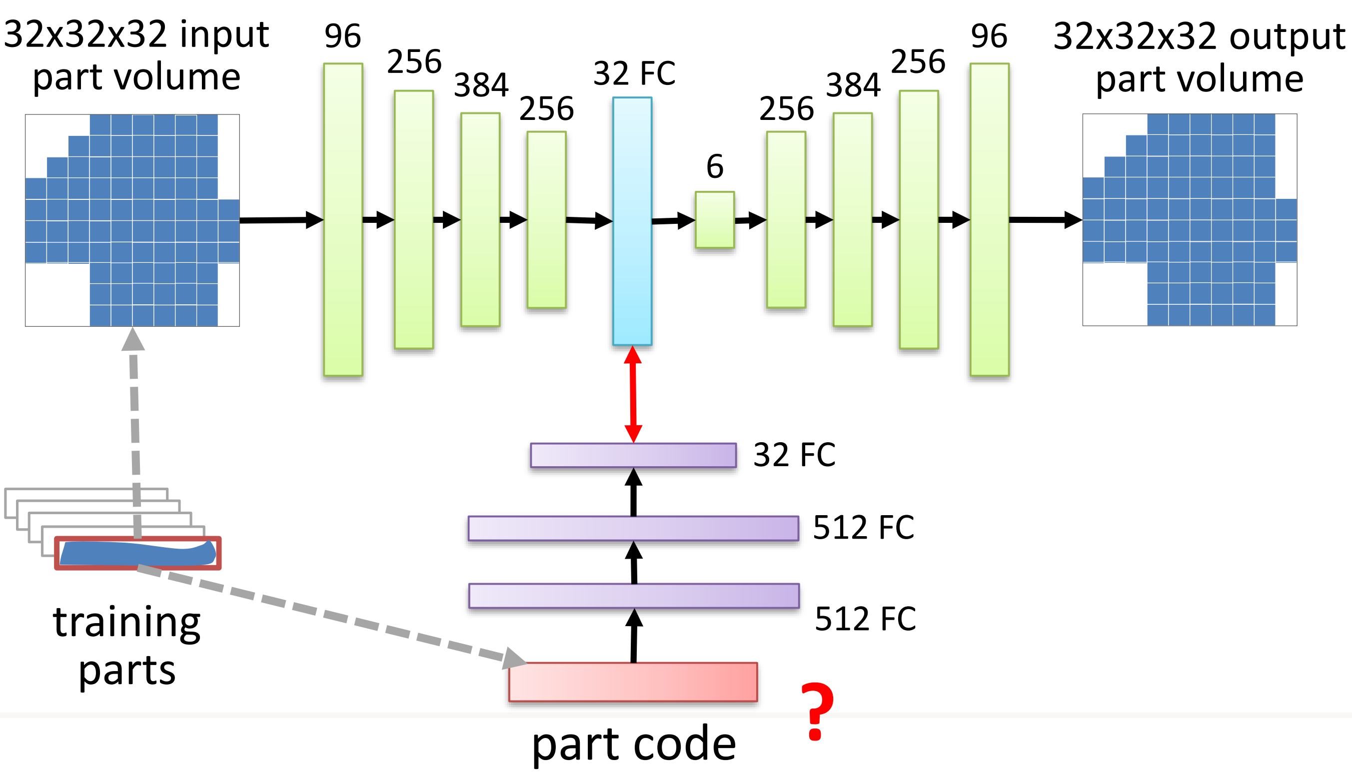


Synthesize geometry within a box

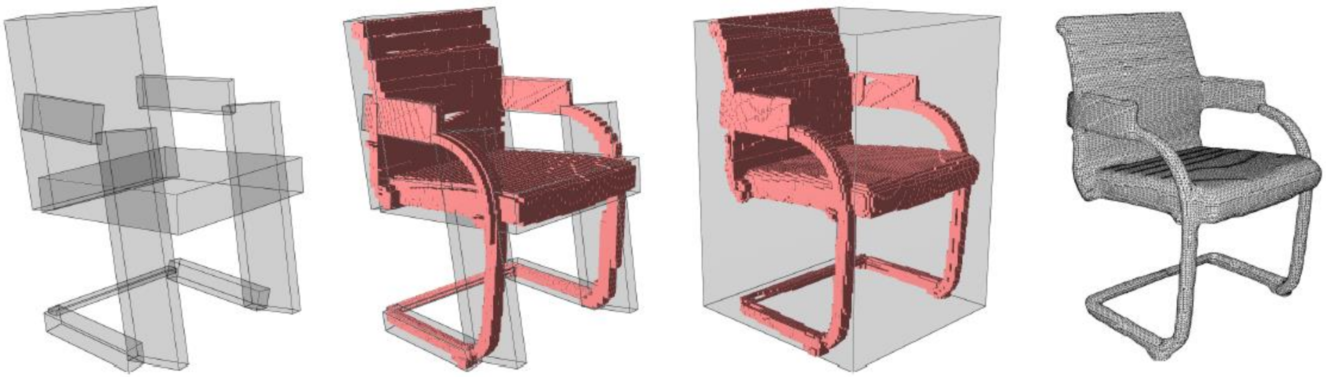


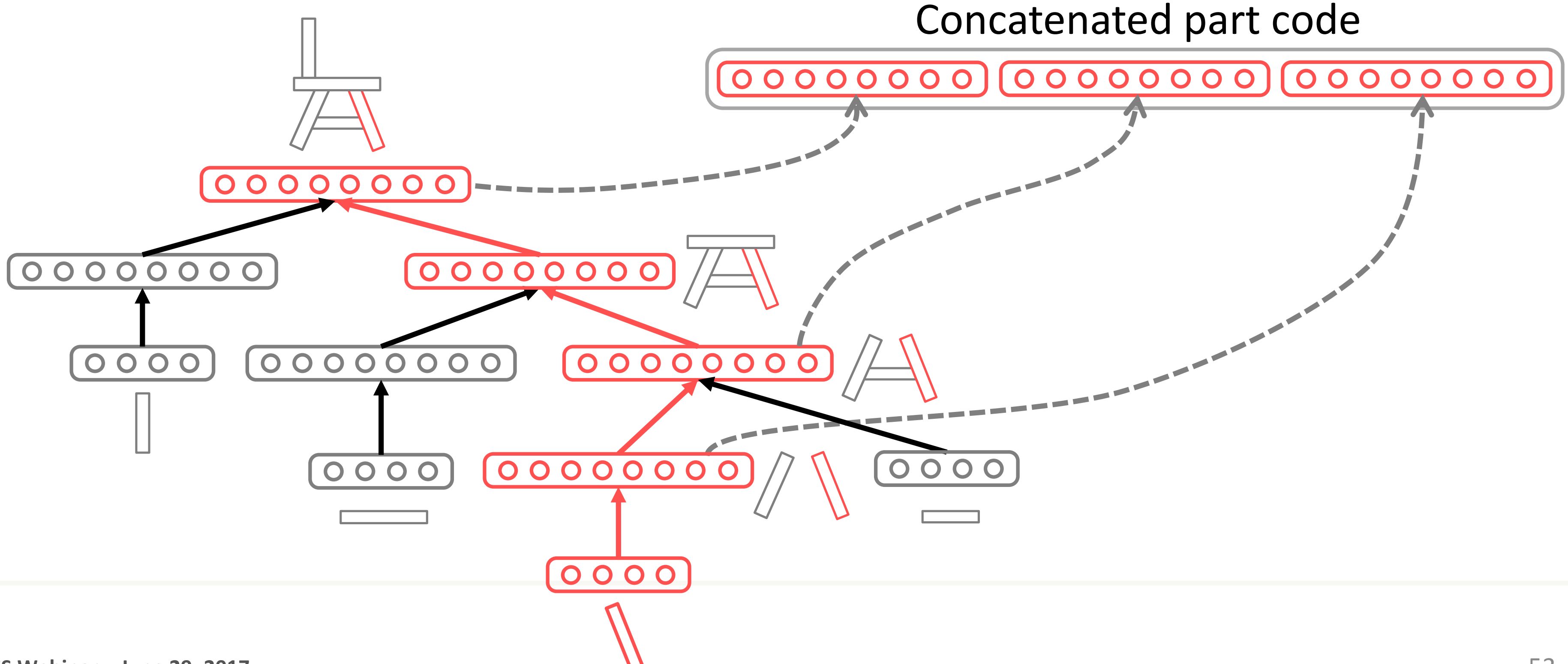
Synthesize geometry within a box



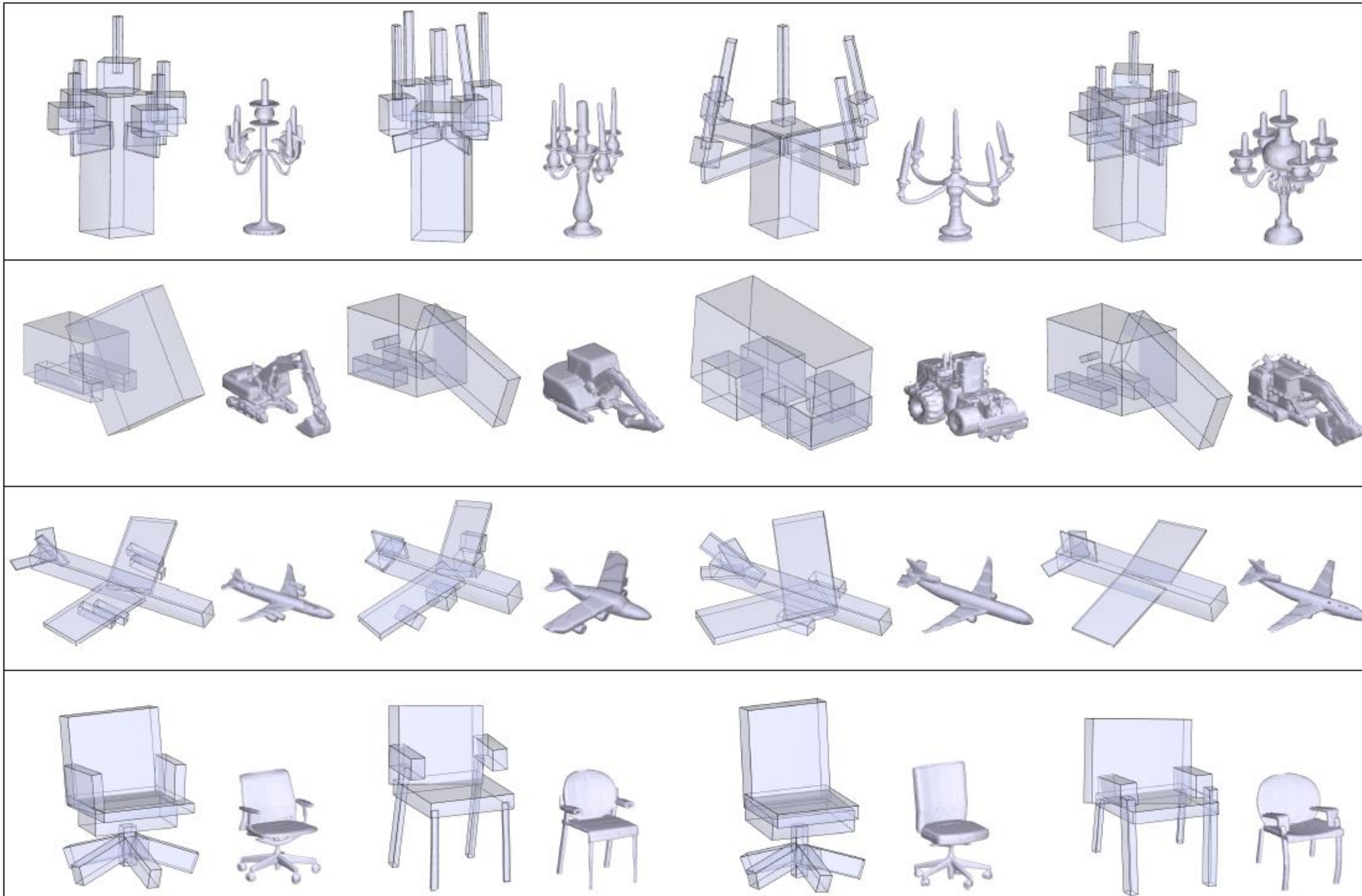


Synthesize geometry within a box

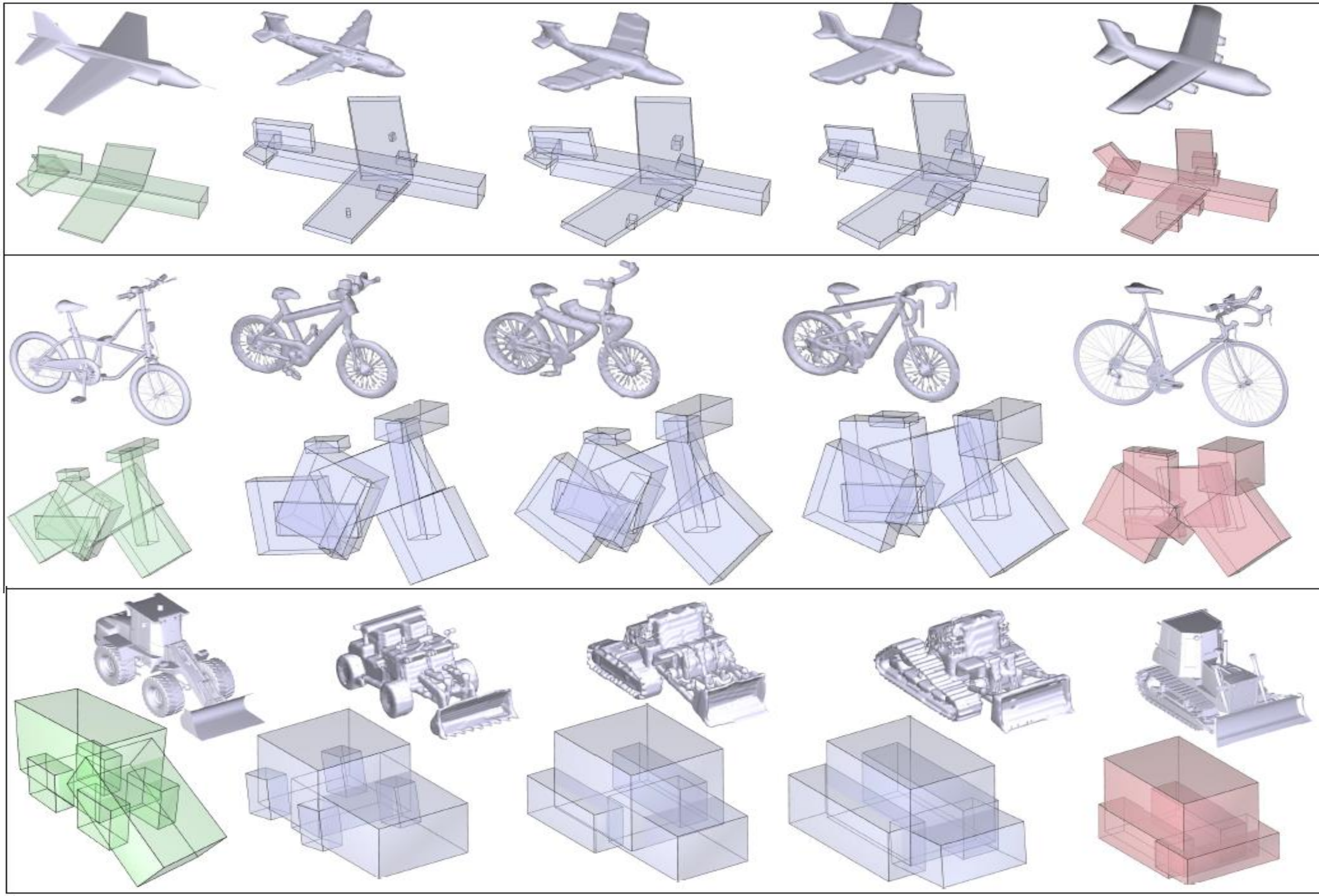




Synthesis results



Interpolation results



Today

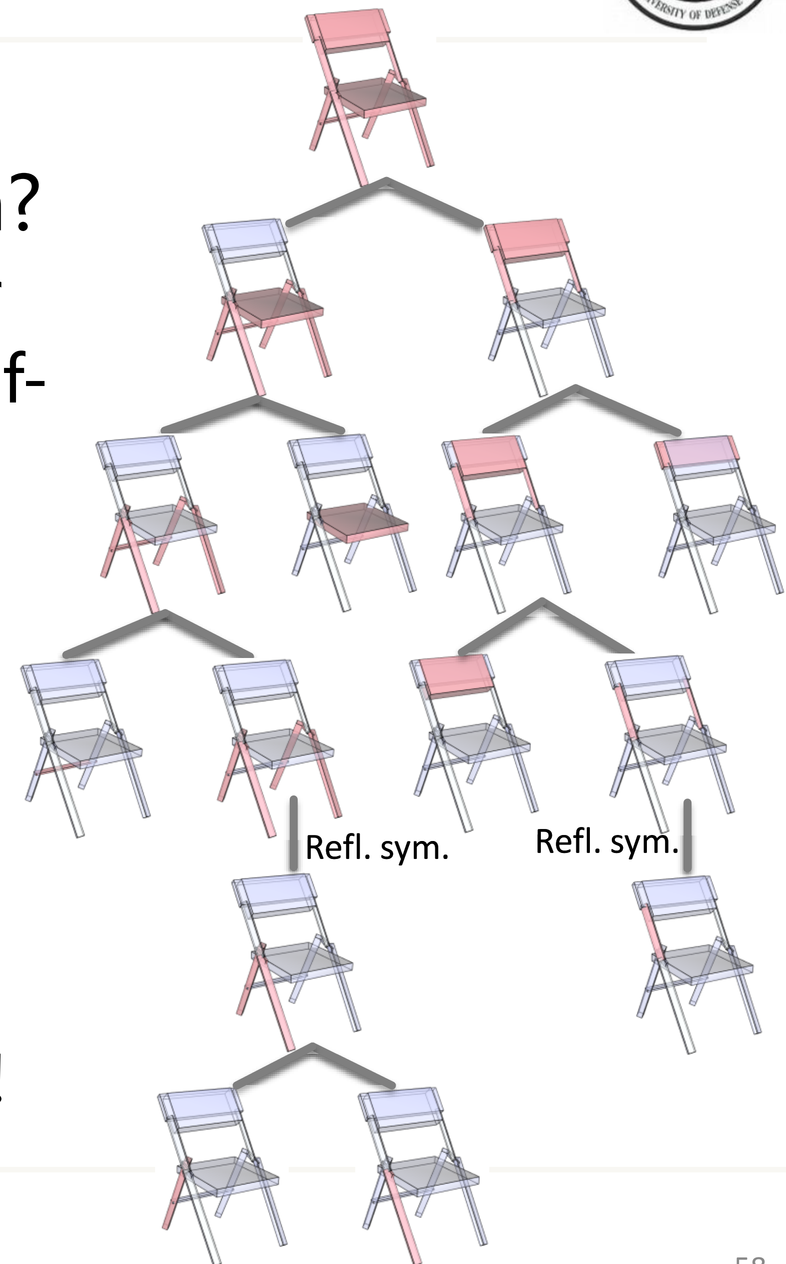
- What is special in generating 3D shapes?
- What is a good generative model for 3D shapes?
- What do we learn from building such a model?

Discussion

- What is a good representation for 3D shapes?
- Depends on task
  - Visual recognition and classification? Multi-view rendering
  - Fusing 2D, 2.5D and 3D data? Volumetric representation
  - Structure-aware tasks? Part representation
- It may not be wise to learn everything from raw data, despite the feature learning power of DNN
  - One never learns to understand natural language from images of characters, but instead works with symbols.
- For 3D shapes, graphics people should think independently

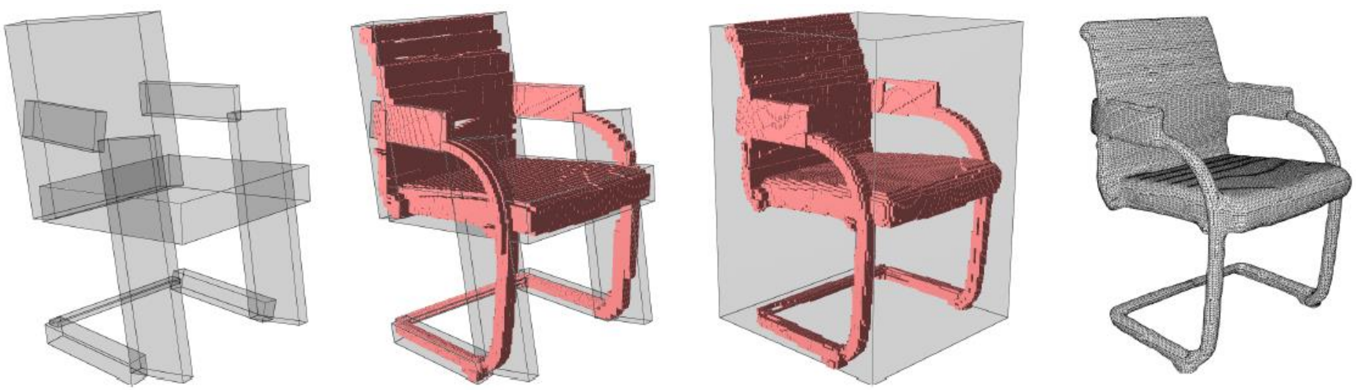
Discussion

- What does our model learn?
  - A hierarchical organization of part structure (minimizing self-reconstruction error)
  - A good way to generate 3D structure
    - Bottom-up generation
    - Creating parts and
    - Hierarchically grouping parts
  - This is exactly how a human modeler creates a 3D model !



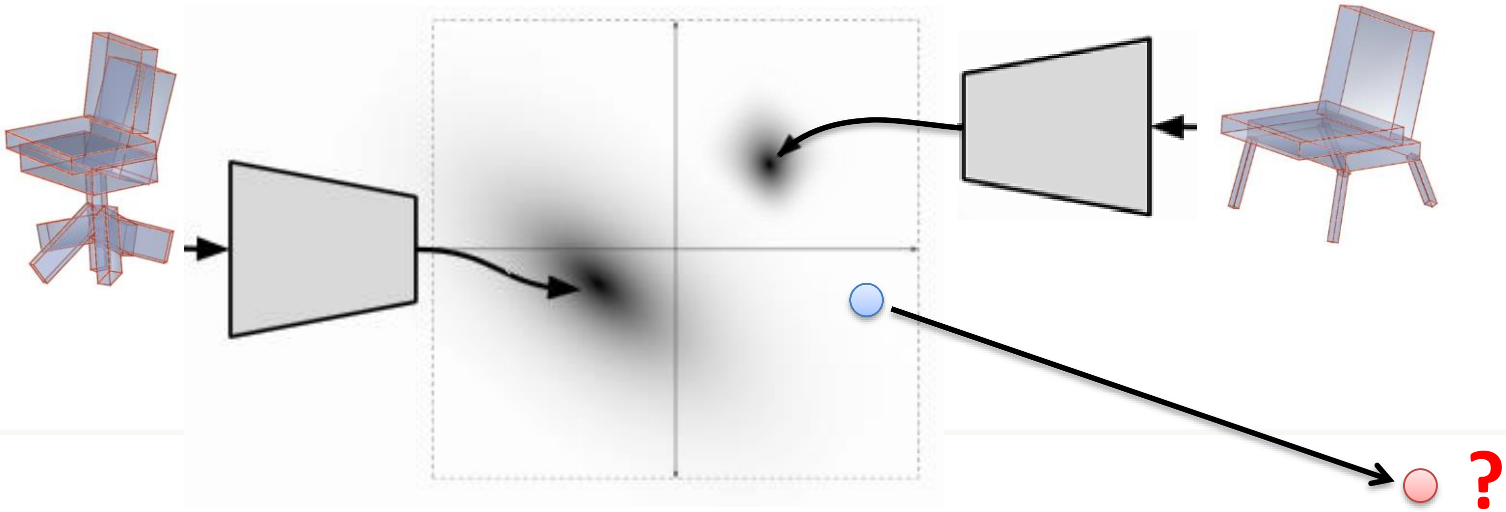
Discussion

- Generally, how to generate things?
- Coarse-to-fine:
  - First generate coarse structure
  - Then generate fine details
  - May employ different representations and models
- A guiding rule for designing a generative model



Discussion

- The trade-off between plausibility and diversity
  - Plausible requires keeping close to input exemplars
    - Is there a definition on plausibility really?
    - Similarity against exemplars?
  - Diverse requires going further away



Discussion

- Is there a low-dim manifold of 3D shape structures?
  - Not every in-between structure is functionally valid
  - However, they may reveal
    - The evolution of design in human brains
    - The exploration of design space in human brains
- We cannot say we model **the manifold** of 3D shape structures.
  - Our ongoing research ...

Acknowledgement

- Reviewers
- Jun Li is a visiting PhD student of University of Bonn, supported by the CSC
- Co-authors
  - Siddhartha Chaudhuri, Ersin Yumer, Hao Zhang and Leonidas Guibas
- Help on data collection
  - Yifei Shi, Min Liu, Chengjie Niu and Yizhi Wang
- Research grants
  - NSFC, NSERC, NSF
  - Google Focused Research Award, Gifts from the Adobe, Qualcomm and Vicarious corporations.

Thank you!

Welcome to try - code & data

www.kevinkaixu.net