Structured Embedding Spaces for 3D Shape Completion and Synthesis

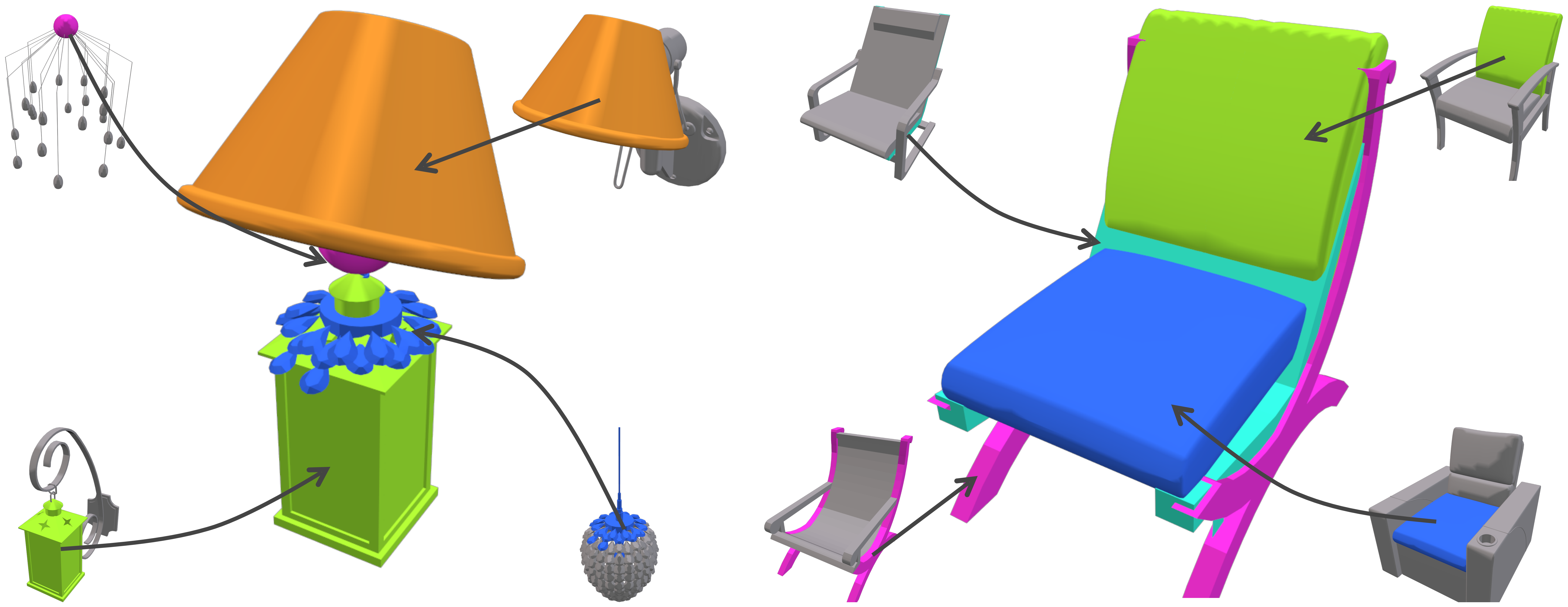
Anastasia Dubrovina

Lyft (previously Stanford)

Joint work with: Minhyuk Sung, Fei Xia, Panos Achlioptas, Leonidas Guibas (Stanford), Vladimir Kim (Adobe)

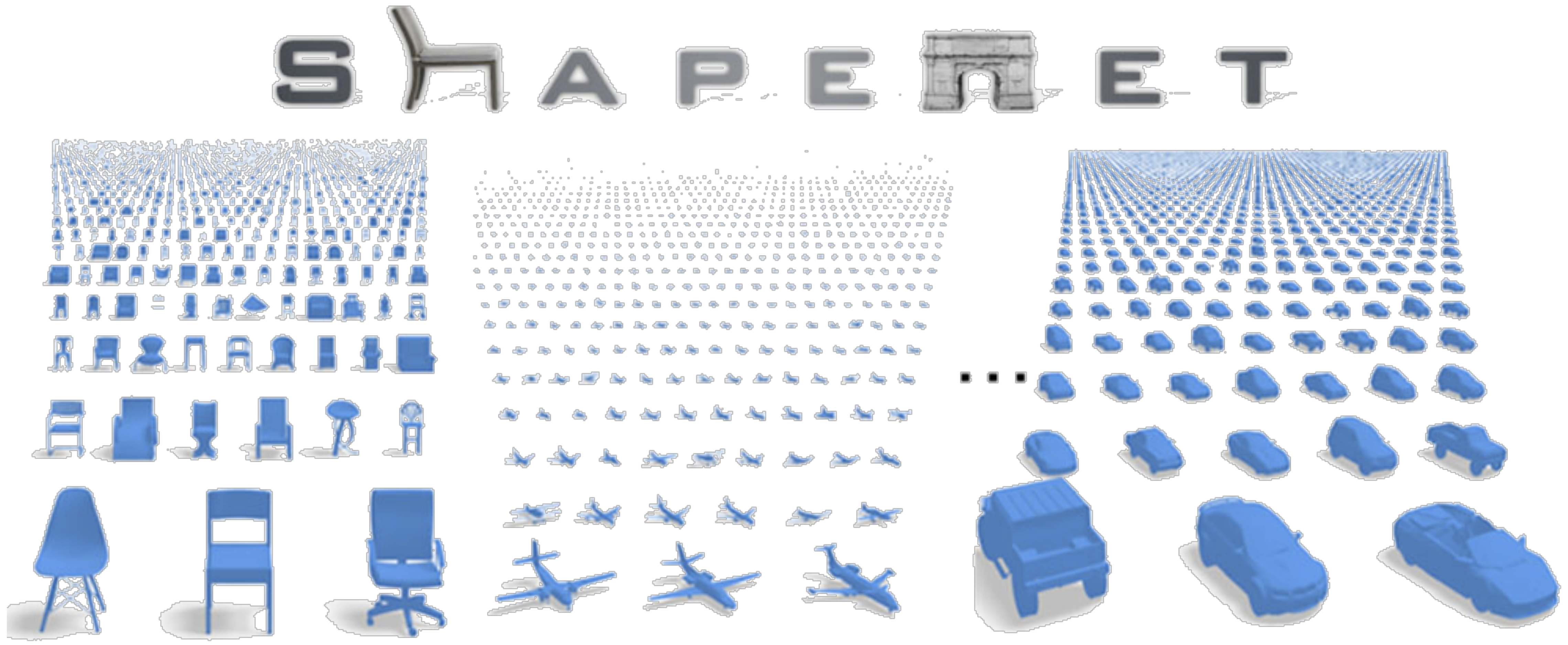
Component-based Shape Modeling

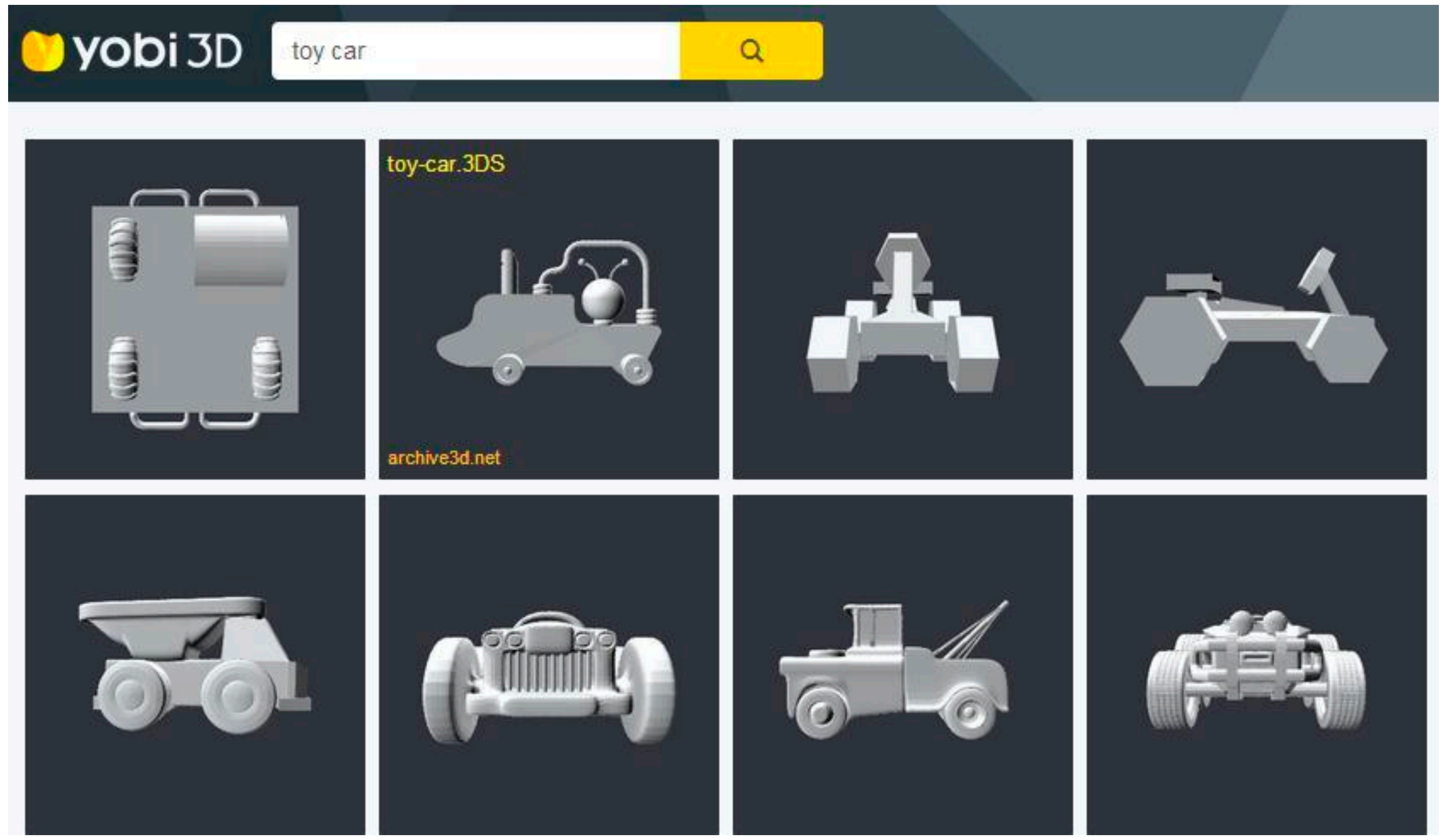
Bring components from different objects to create new object

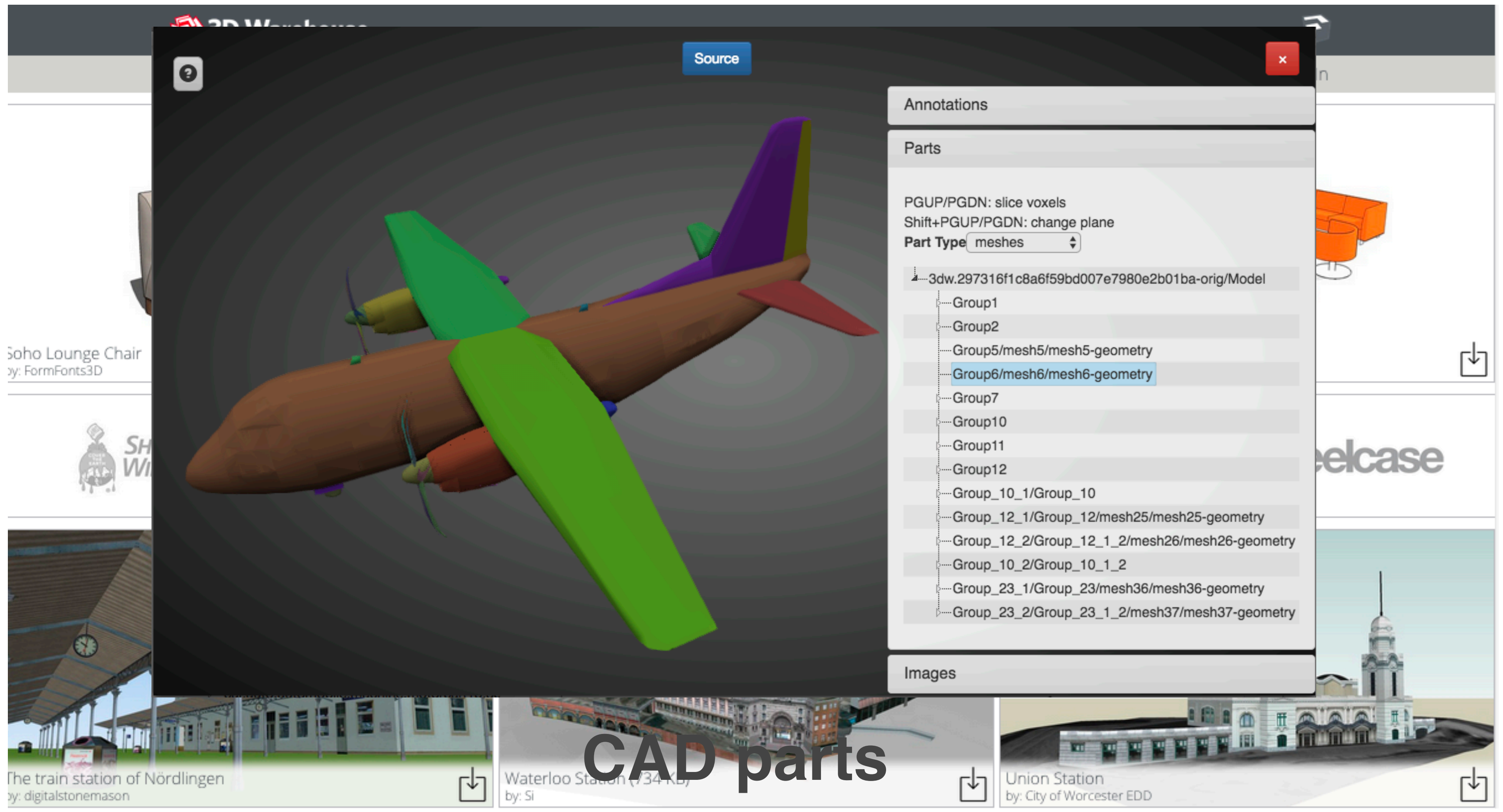


[Sung et al., 2017]

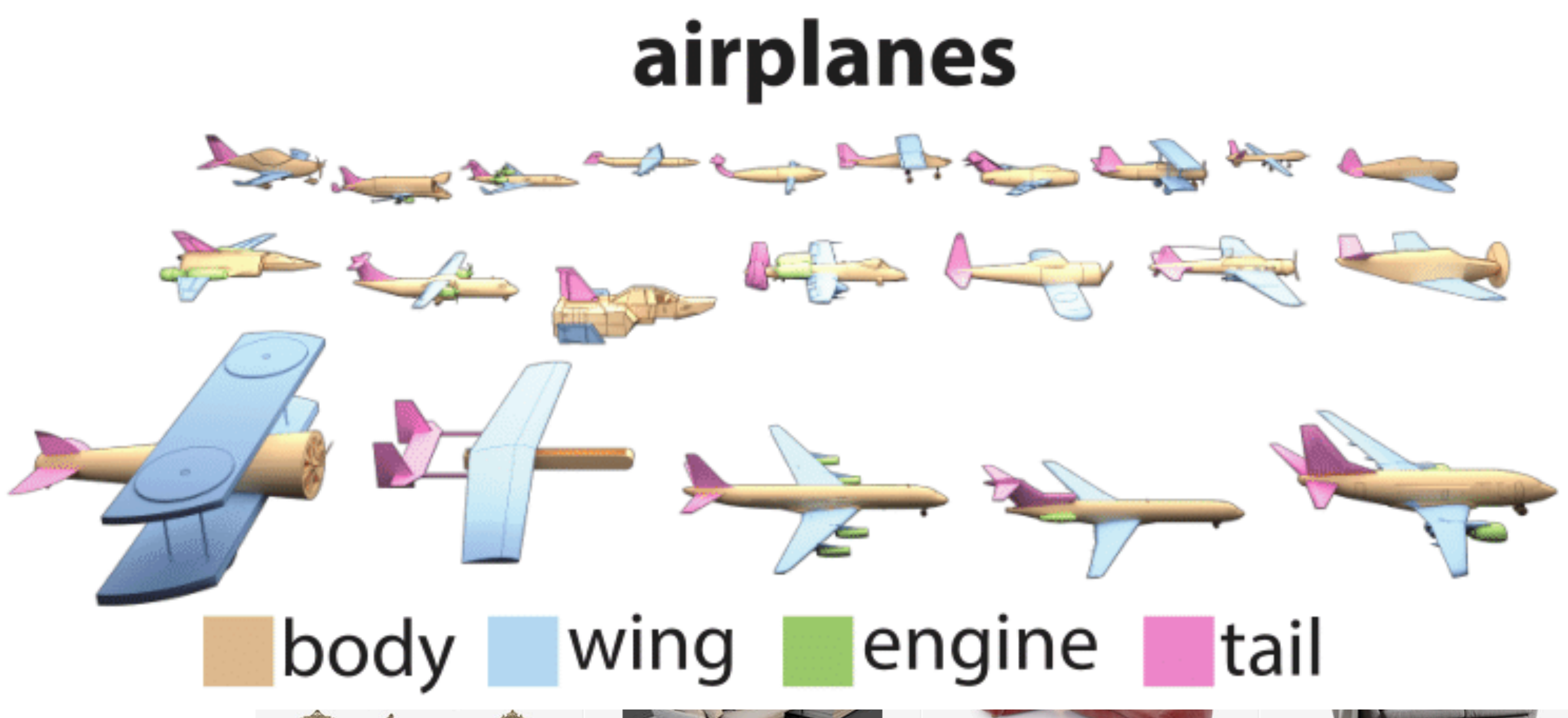
Large-scale 3D Shape Repositories







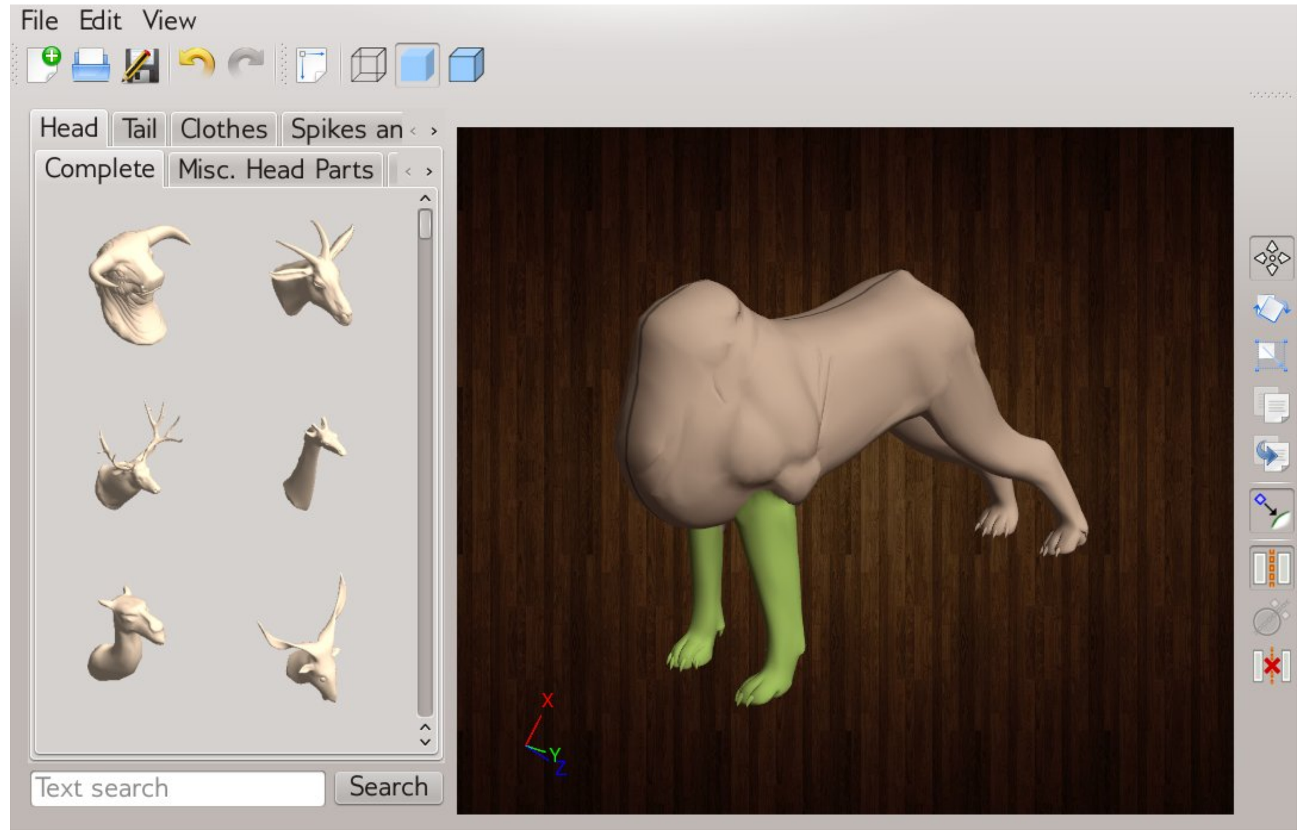
CAD parts



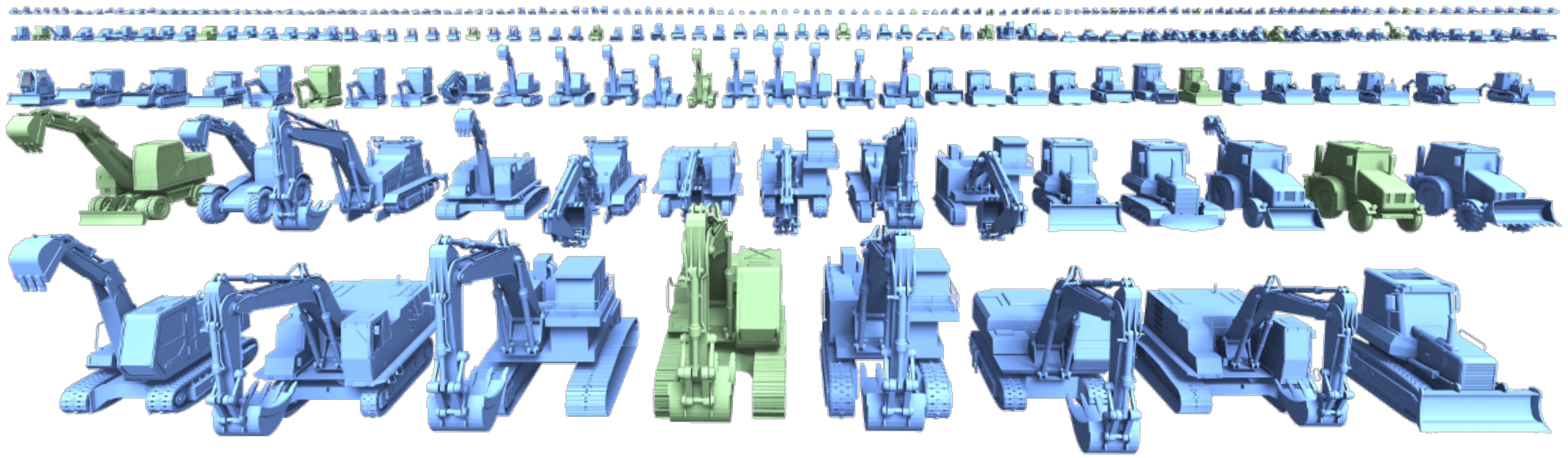
Semantic segmentation [Yi et al., 2016]

Previous work

Modeling by Example
[Funkhouser et al., 2004]



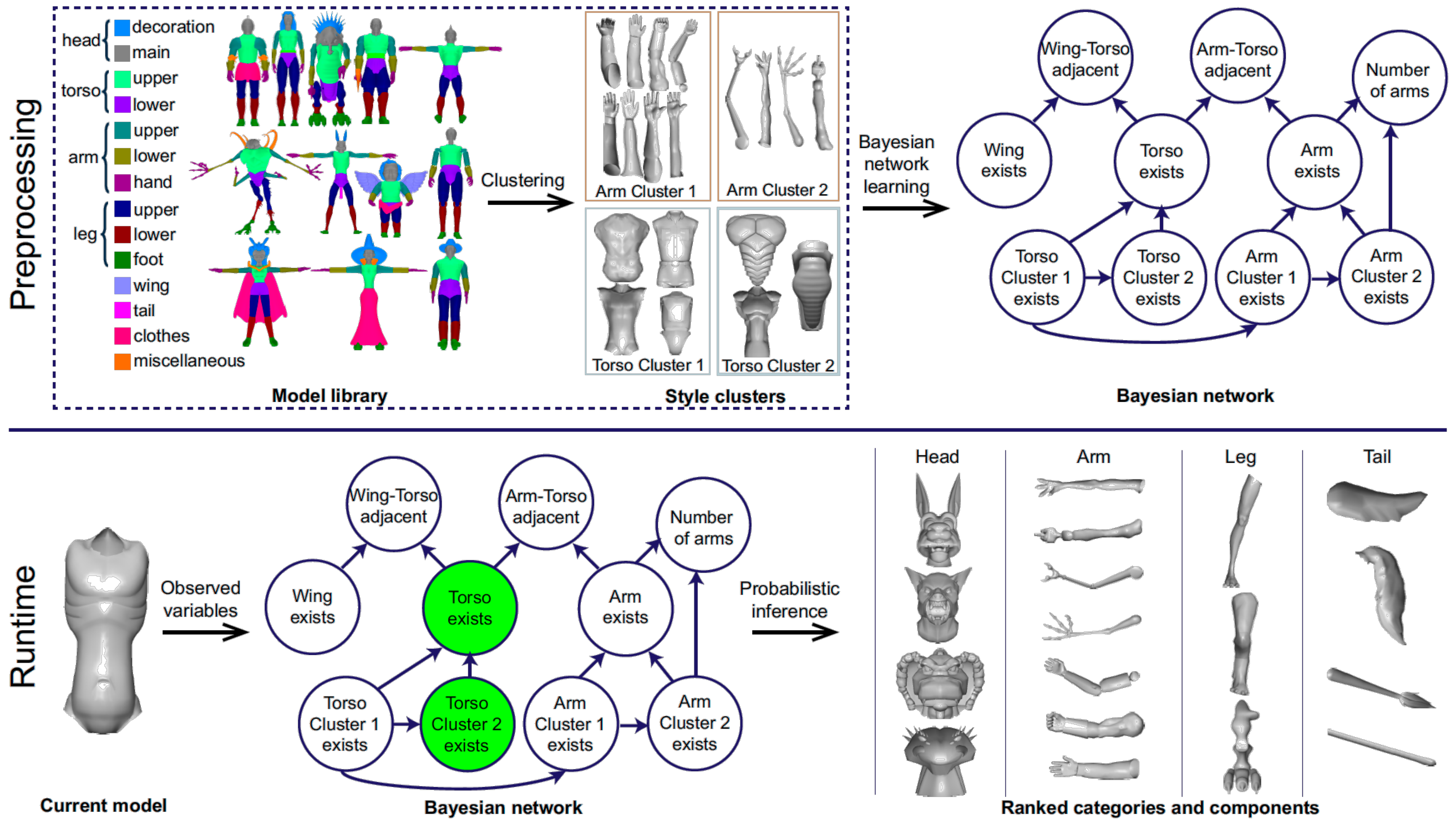
[Chaudhuri et al., 2011]



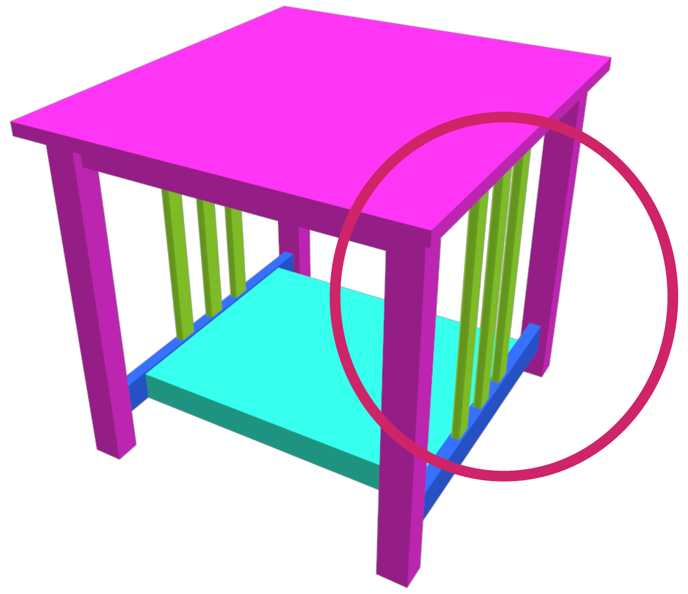
[Kalogerakis et al., 2012]

Challenges

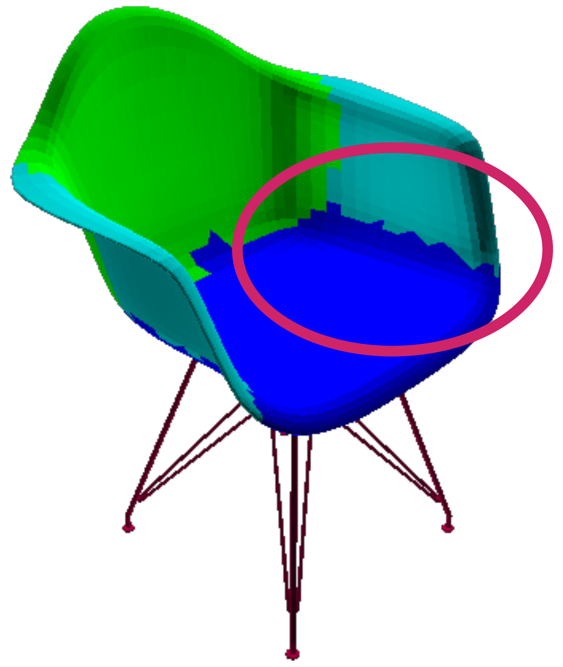
[Funkhouser et al., 2004]



[Chaudhuri et al., 2011]



Name?



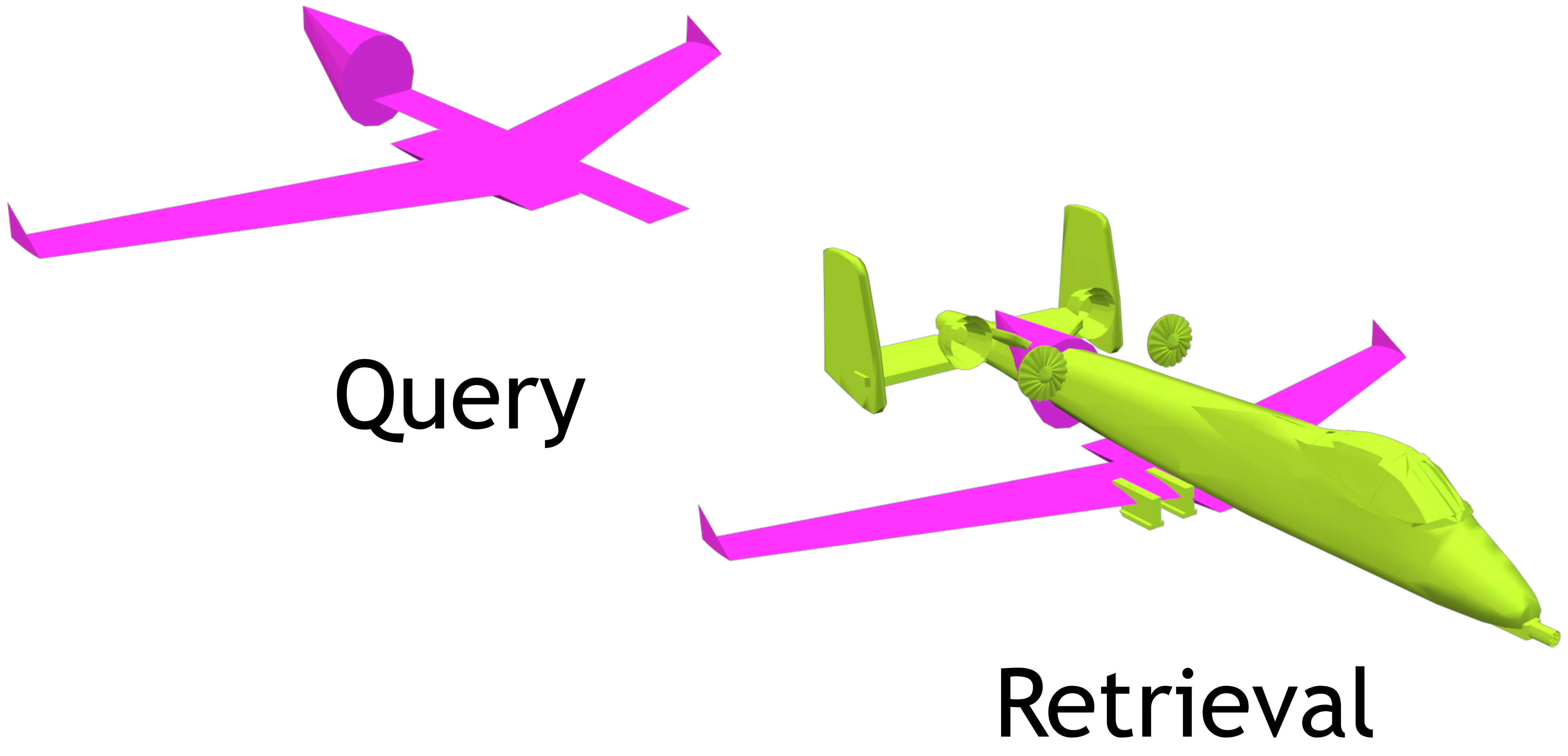
Boundary?

- CAD model segments are inconsistent and unlabeled
- Part annotation is expensive; may be unavailable at test time
- Hard to define proper names and boundaries of parts

This Talk

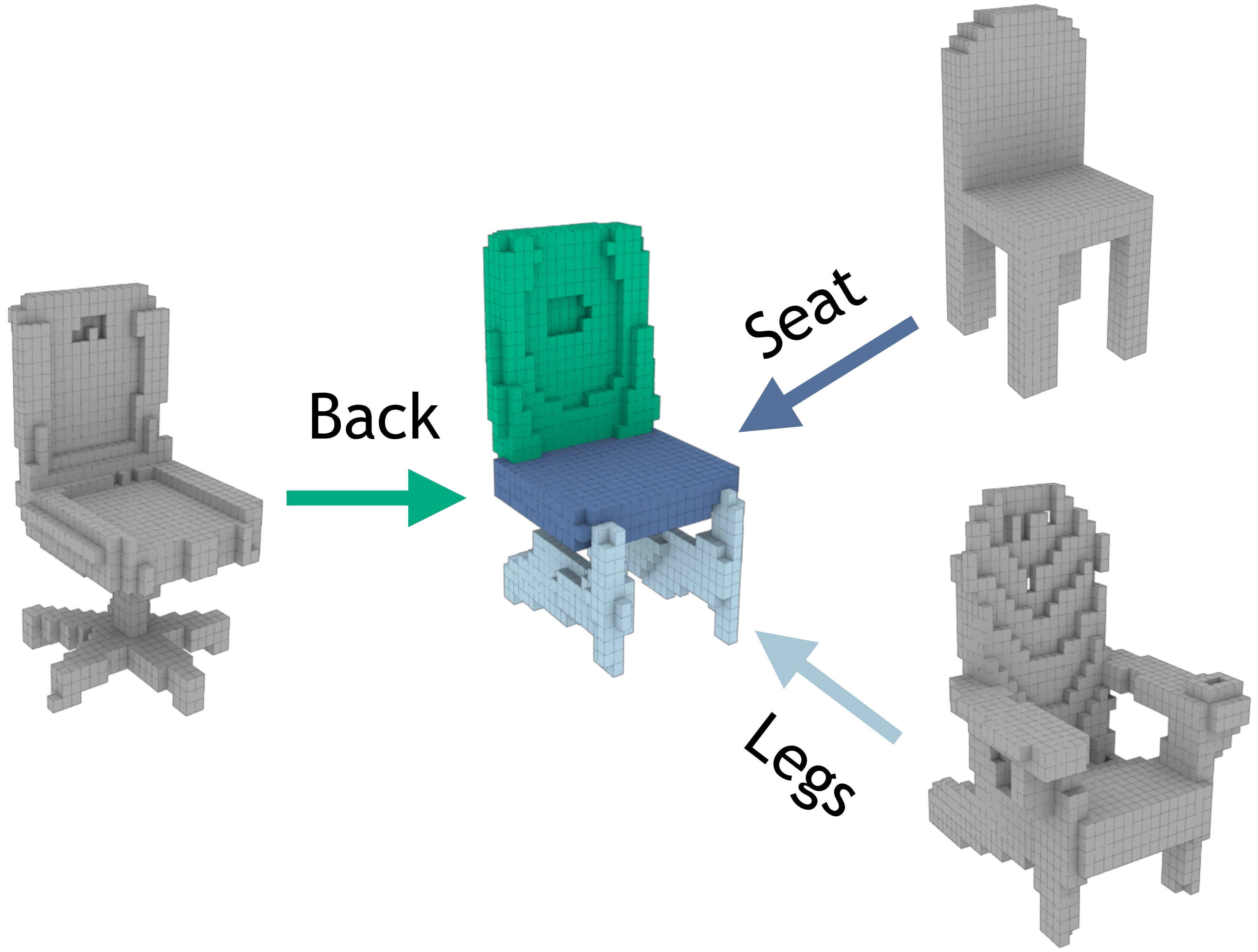
Shape Completion

From dateset with *unlabeled* and *inconsistent* components



Composite modeling

Using *unsegmented* shapes via latent space factorization



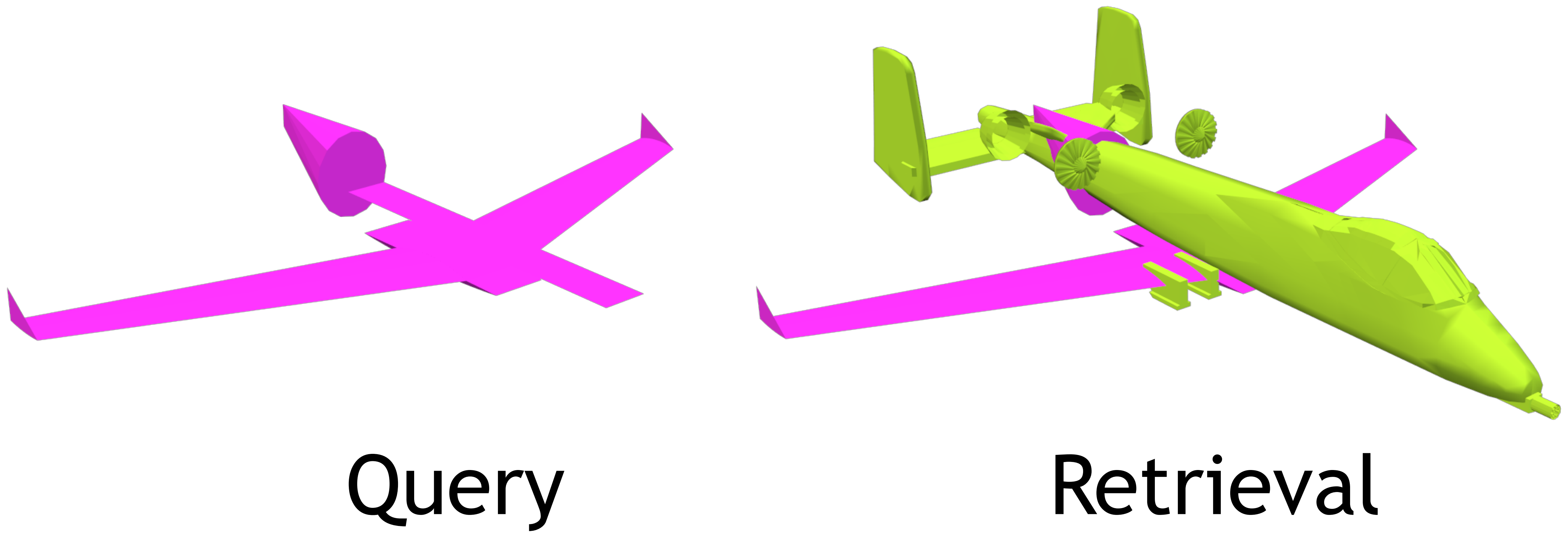
Learning Fuzzy Set Representations of Partial Shapes on Dual Embedding Spaces

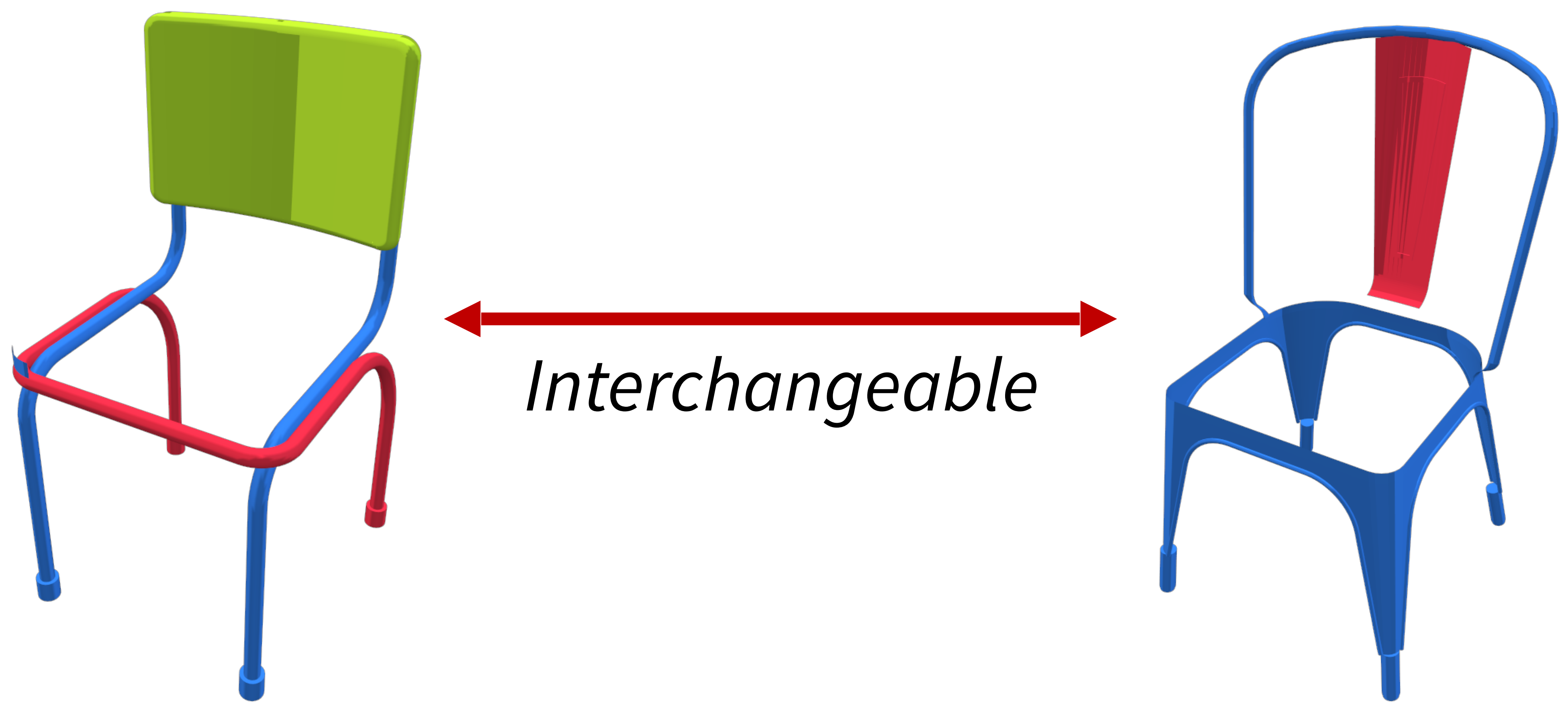
joint work with Minhyuk Sung, Vladimir G. Kim and Leonidas Guibas. SGP 2018.

Problem definition

Learn relations among partial shapes, so that we can

- Complete an object with a single retrieval
- Discover group-to-group relations



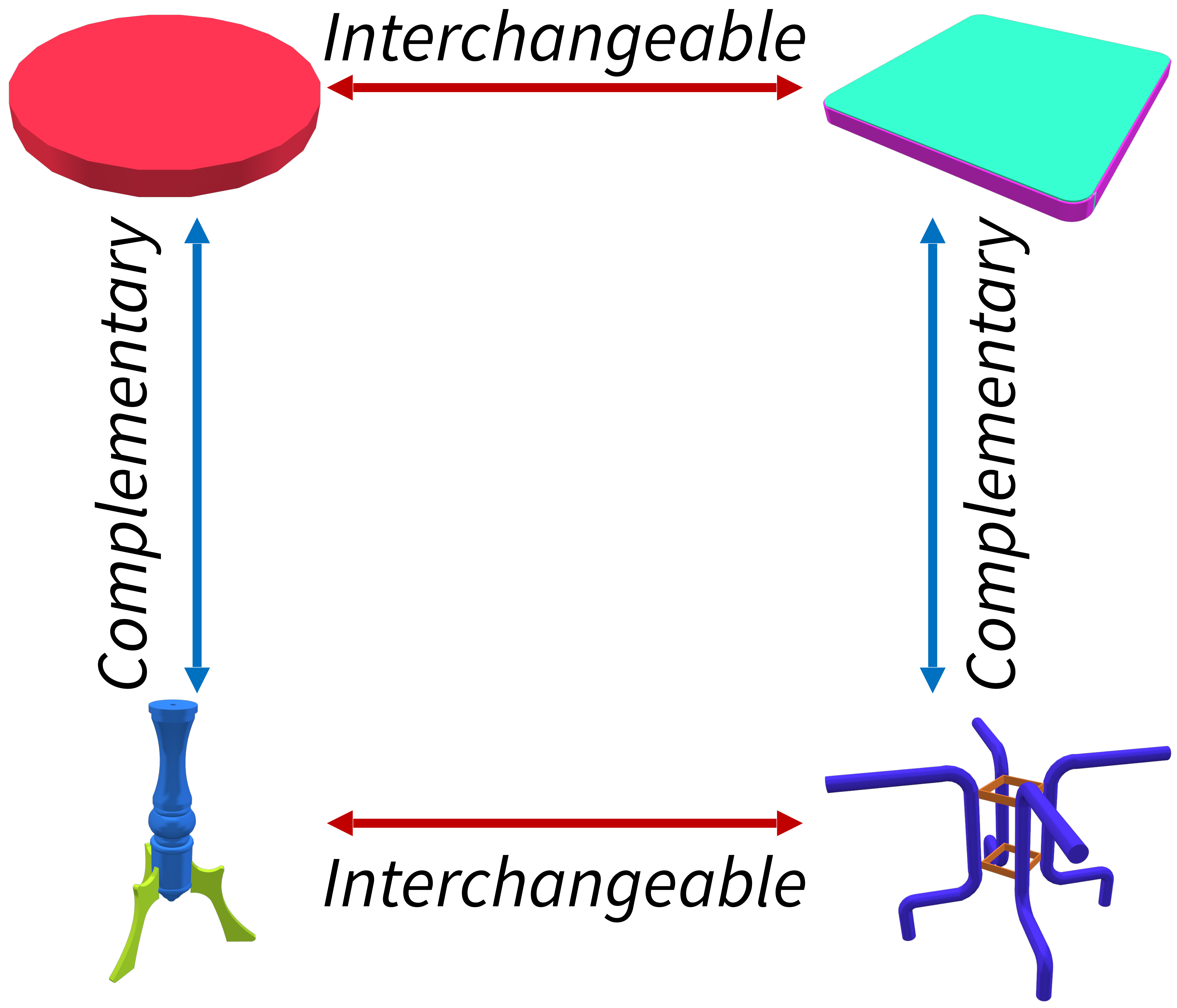


Sung, Dubrovina, et al., SGP 2018

Relations Among Partial Shapes

Learn relations among partial shapes

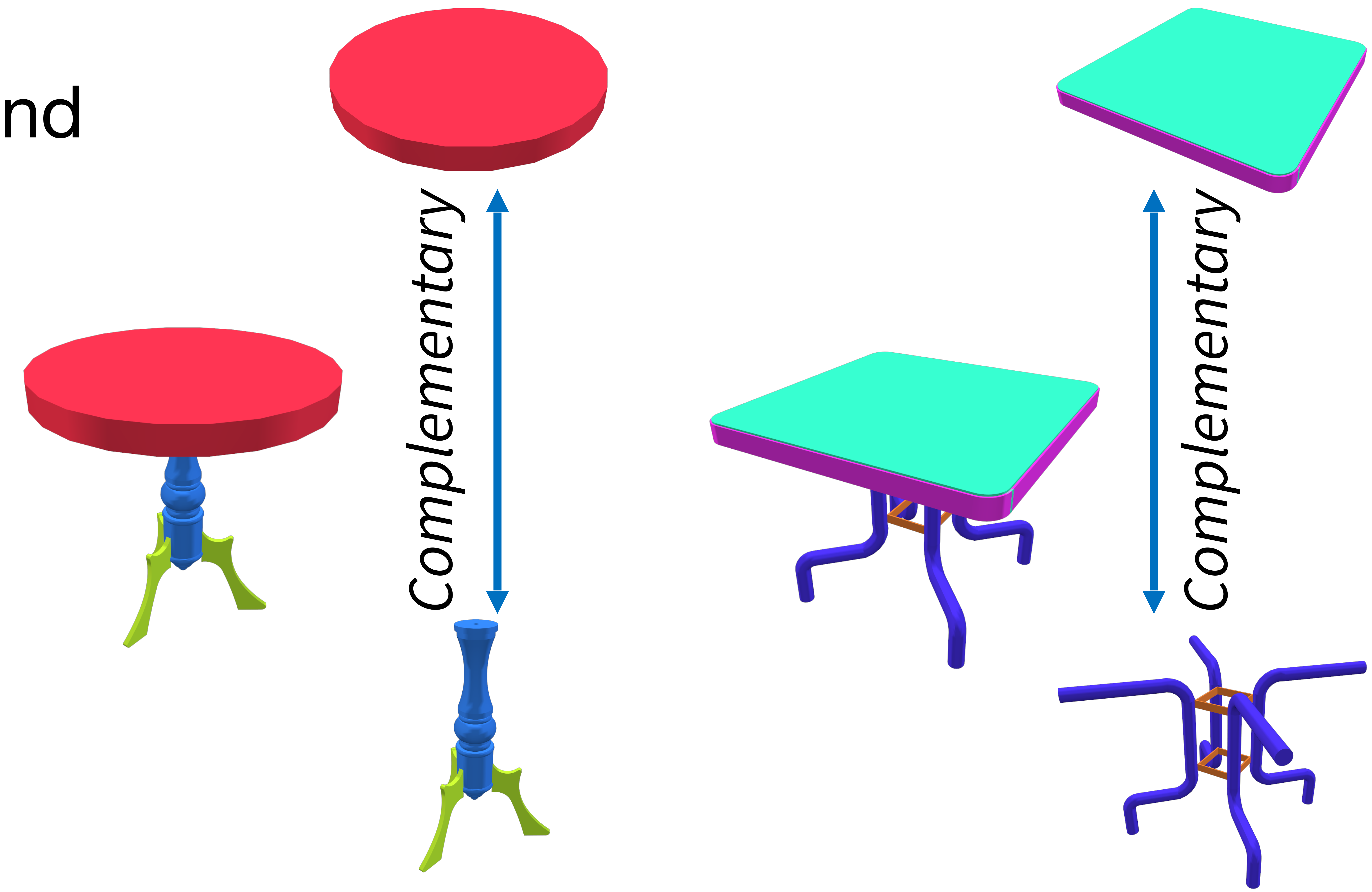
- Complementarity
- Interchangeability



Relations Among Partial Shapes

Complementarity

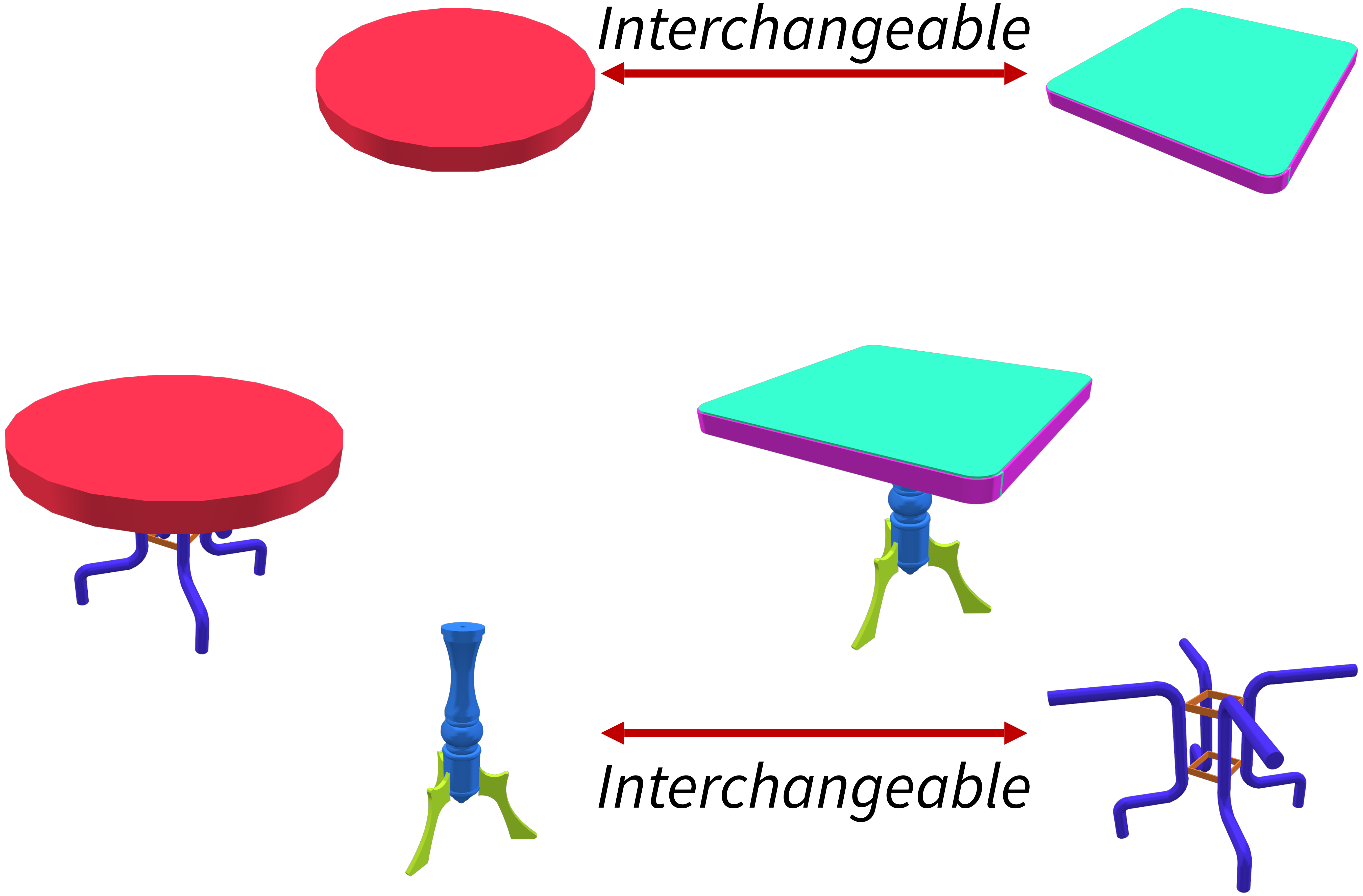
: Two partial shapes can be combined into a complete and plausible object



Relations Among Partial Shapes

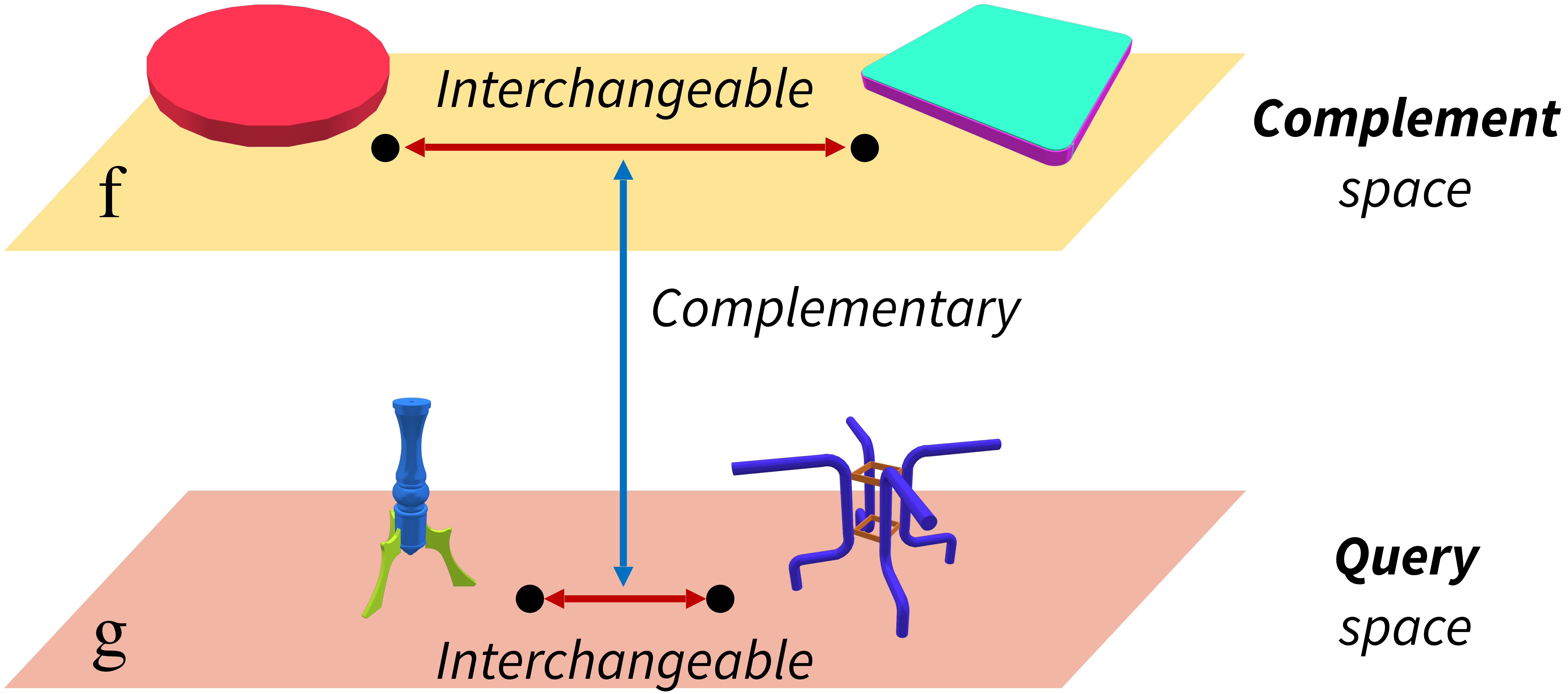
Interchangeability

: Replacing one with the other still produces a plausible object

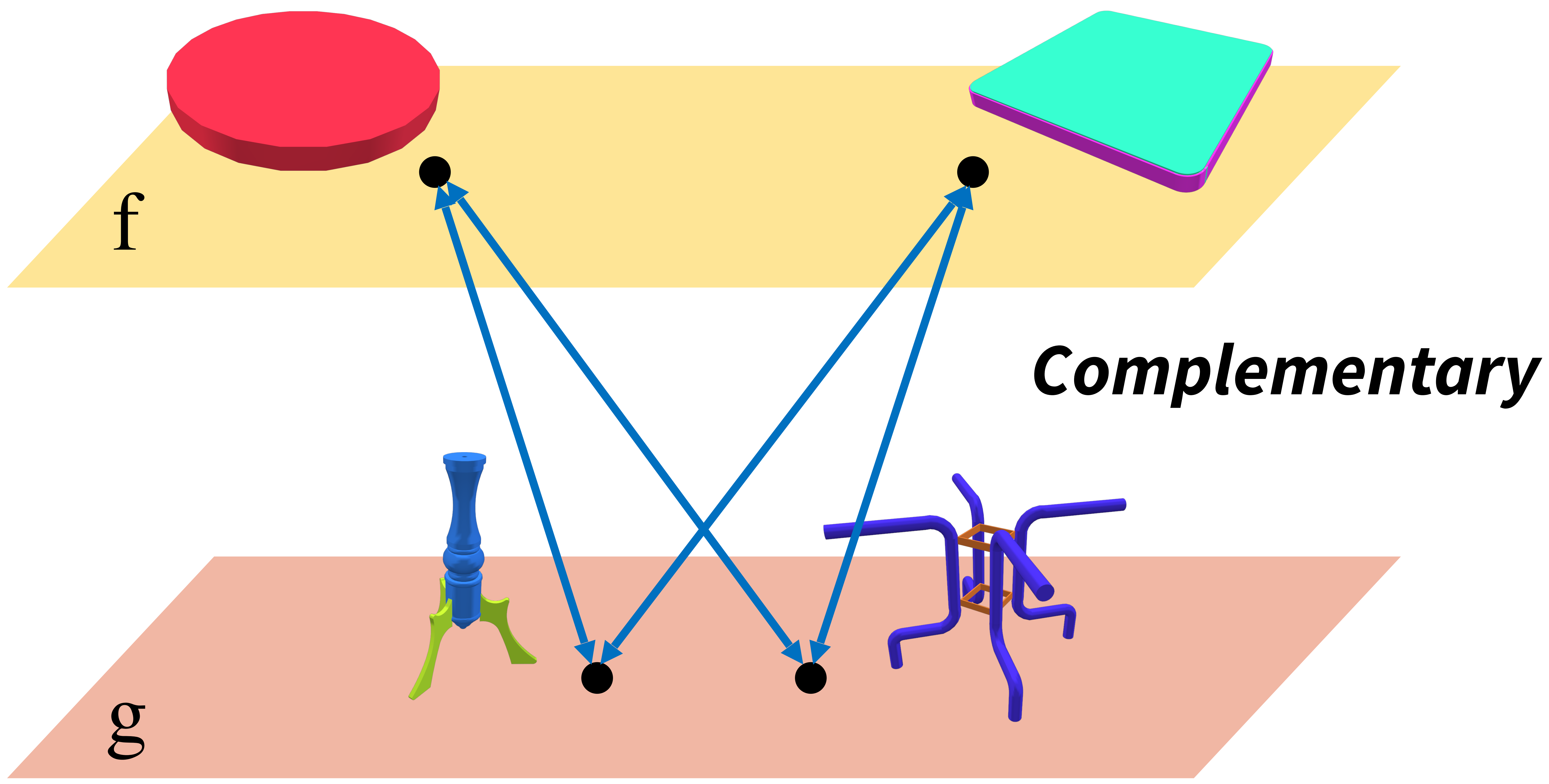


Our Approach

Jointly encode *complementary* and *interchangeable* relations in *dual embedding spaces*

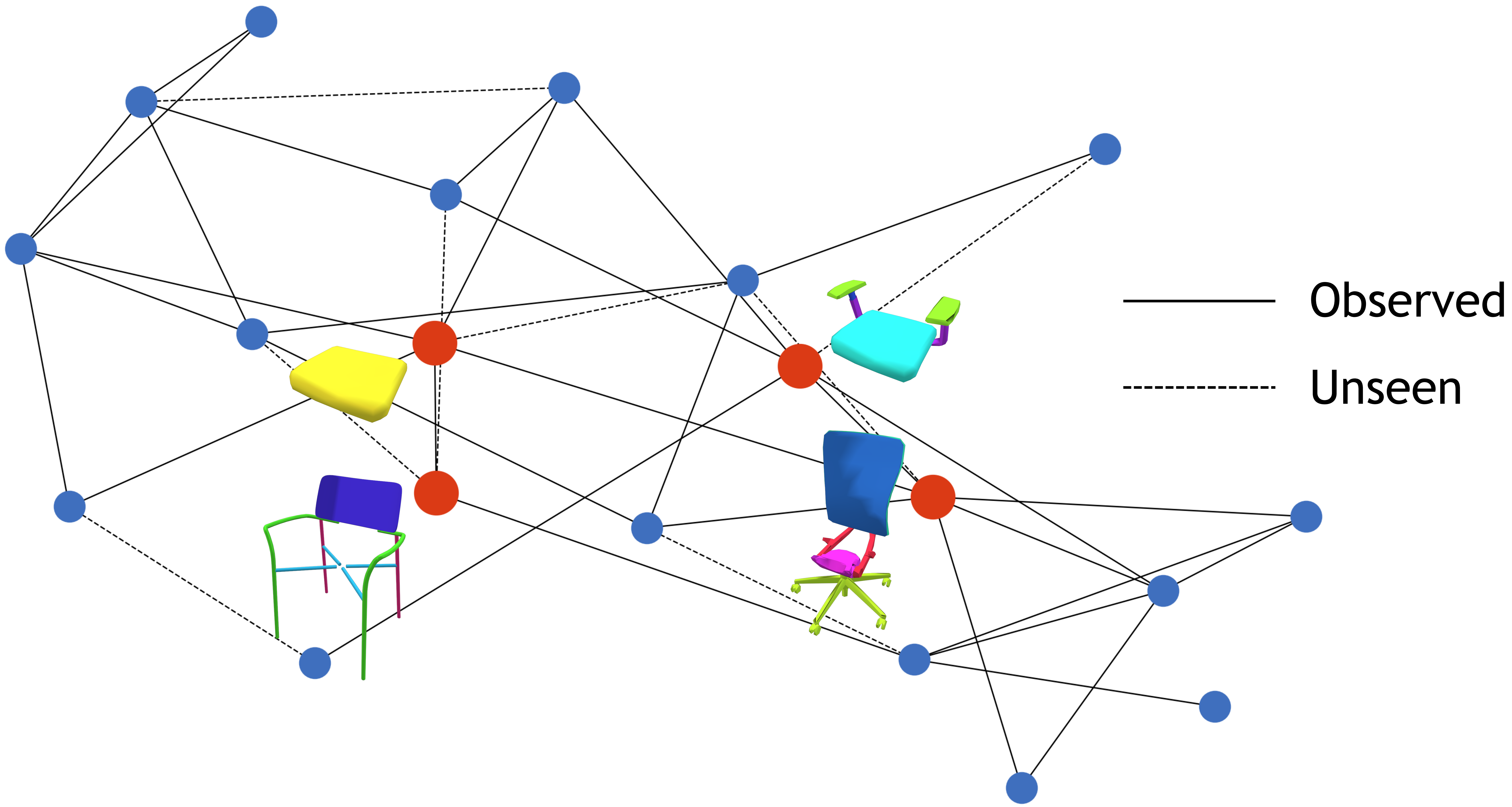


Complementarity



Graph Illustration

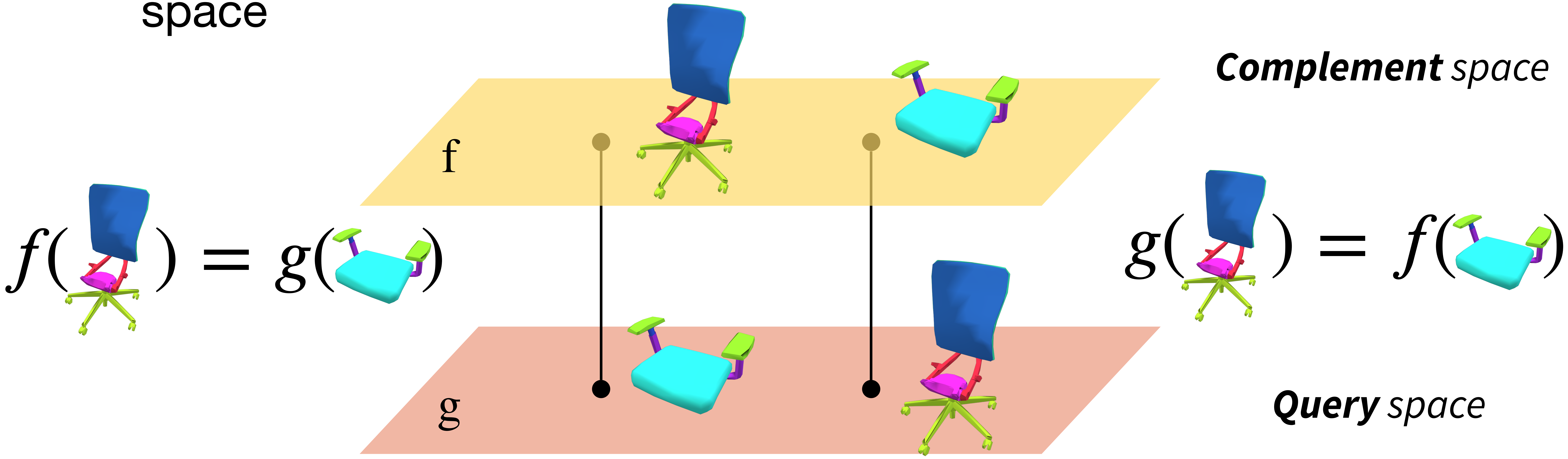
- Consider a *graph* of complementary relations
- Infer *unseen* relations



Dual Embedding Space

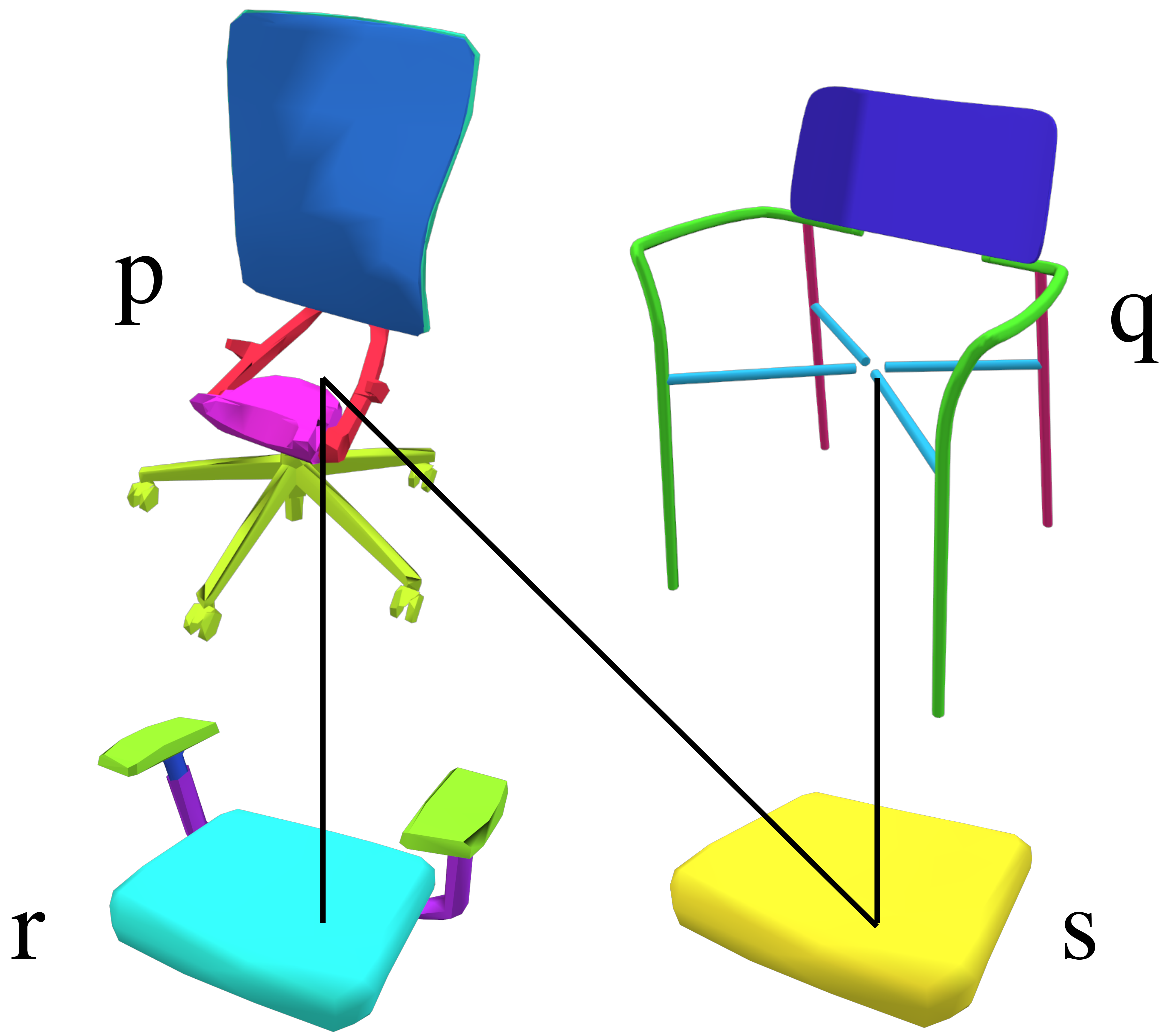
Create *two* embedding spaces

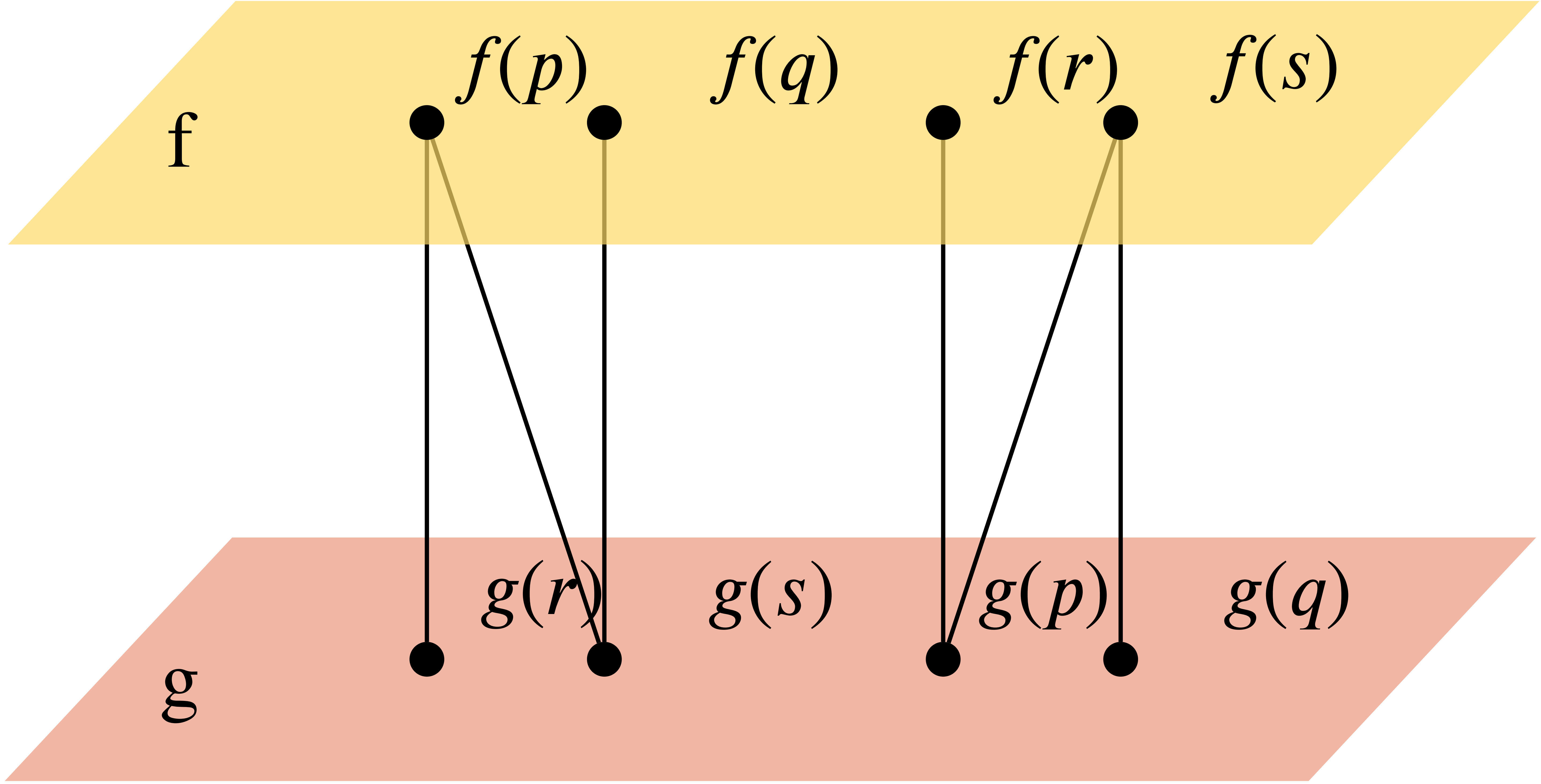
- All partial shapes present in both spaces
- Naïve idea: One space *mirrors* complements from the other space



Dual Embedding Space

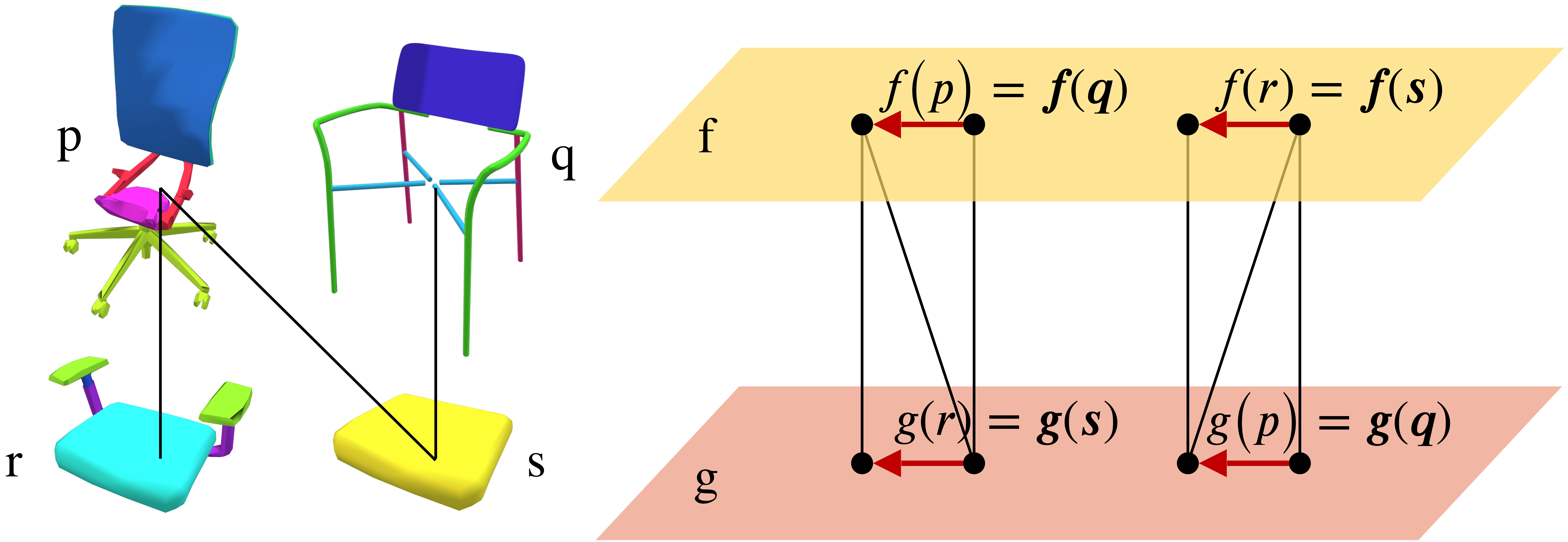
Problem: Embedding collapse





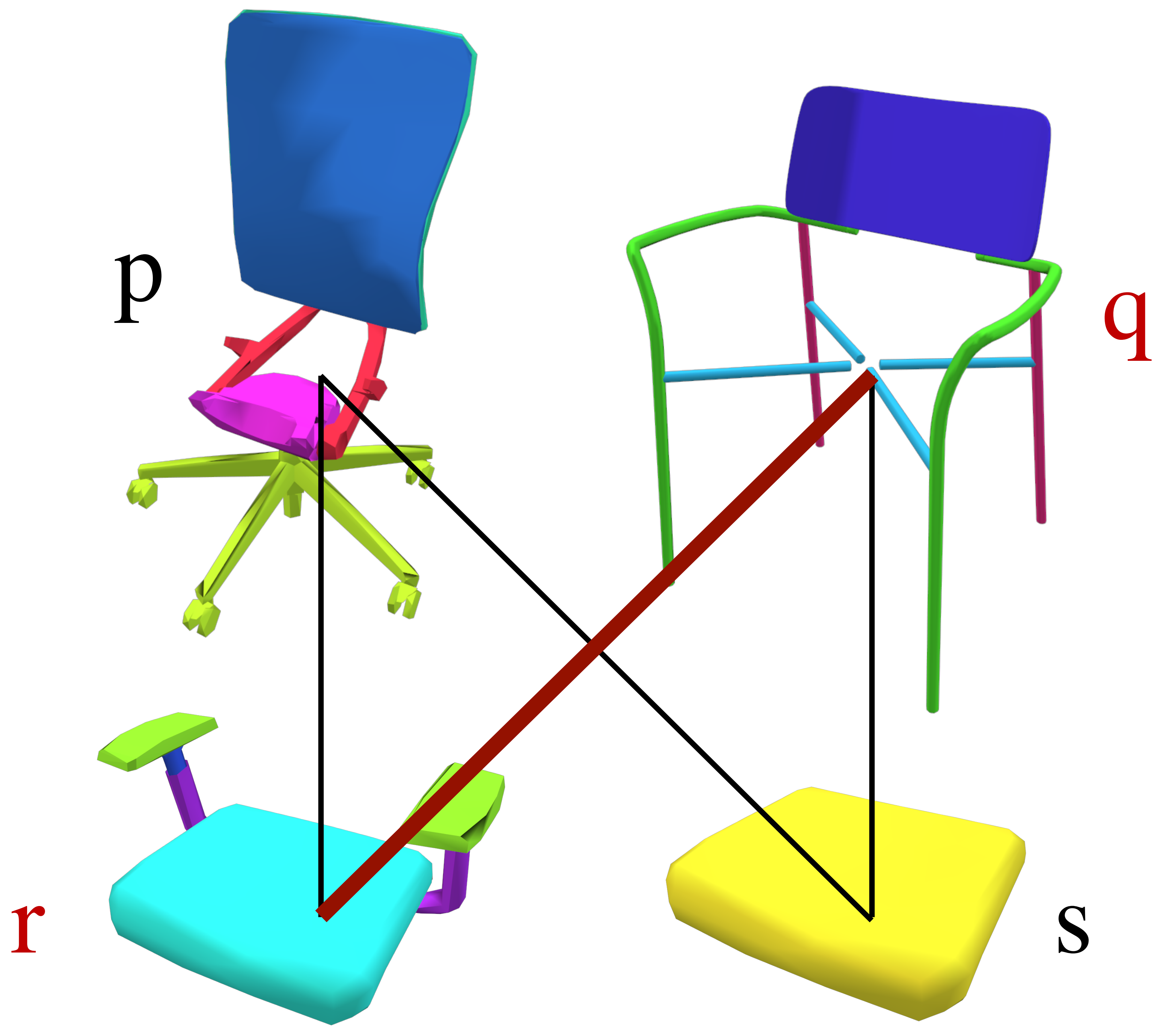
Dual Embedding Space

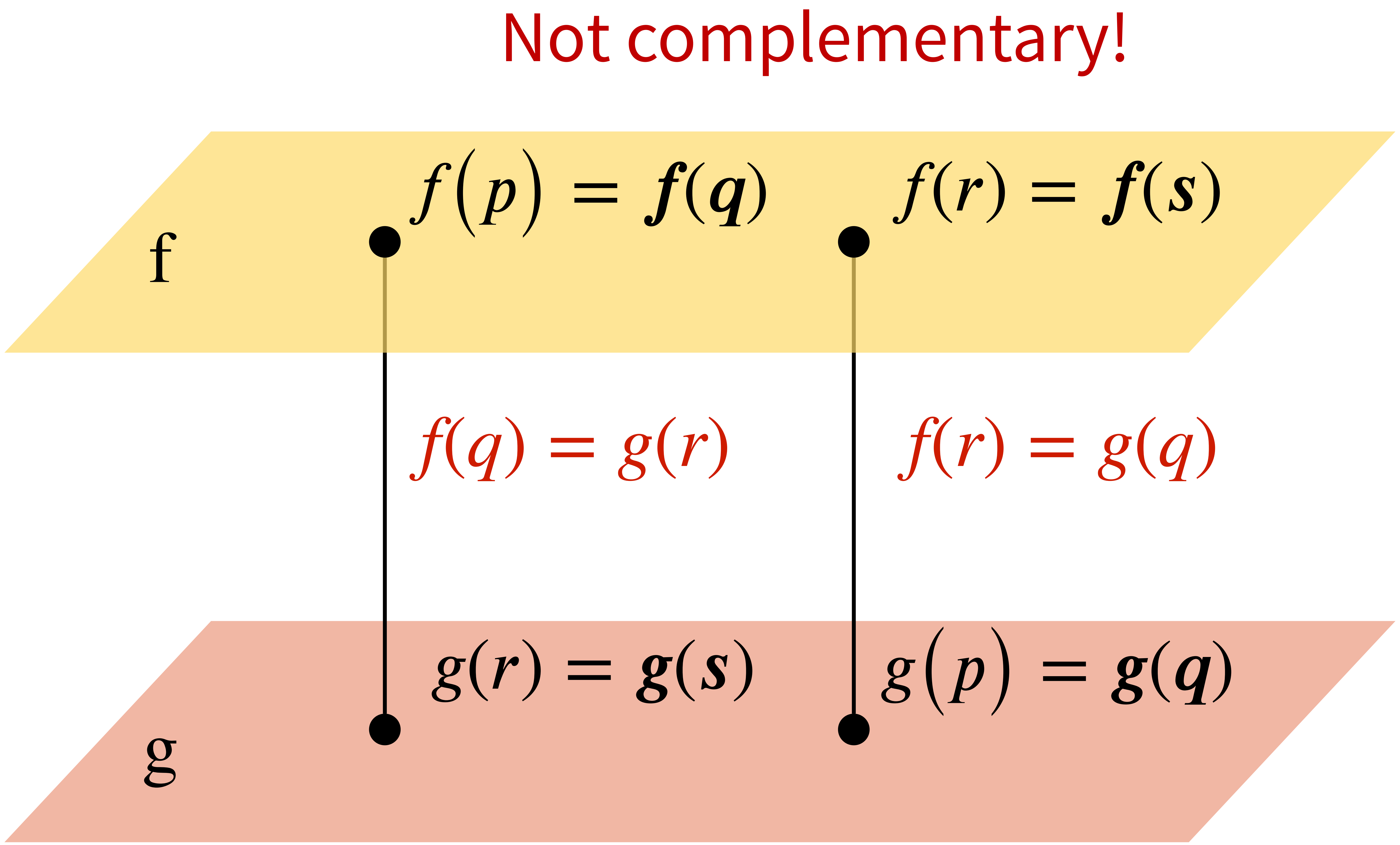
Problem: Embedding collapse



Dual Embedding Space

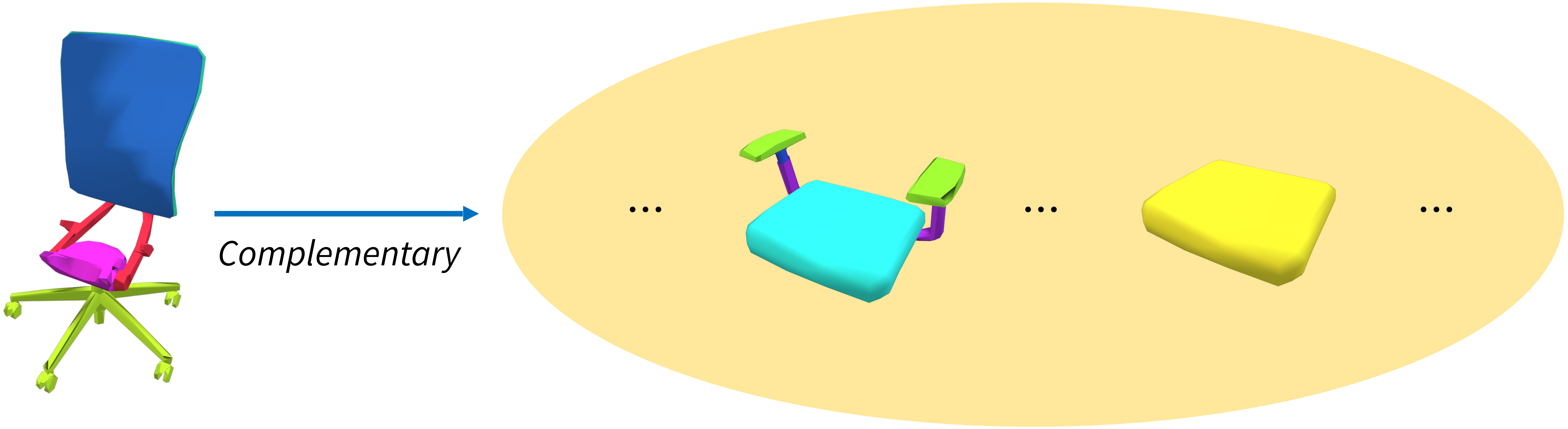
Problem: Embedding collapse





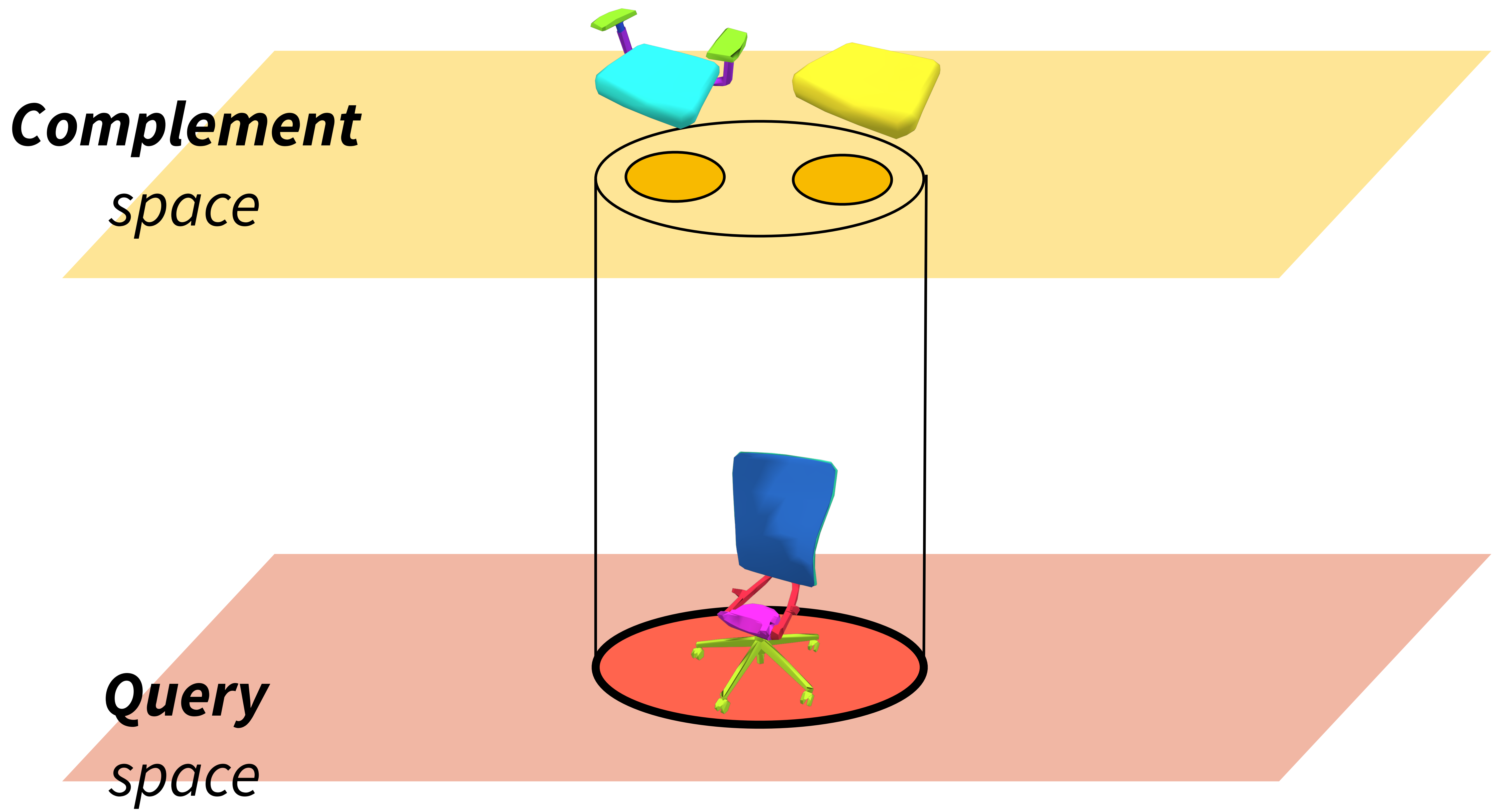
Complementarity as Set Inclusion

- Represent shapes by *sets* in embedding spaces, and
- Encode *1-to-N* mapping as *set inclusion*



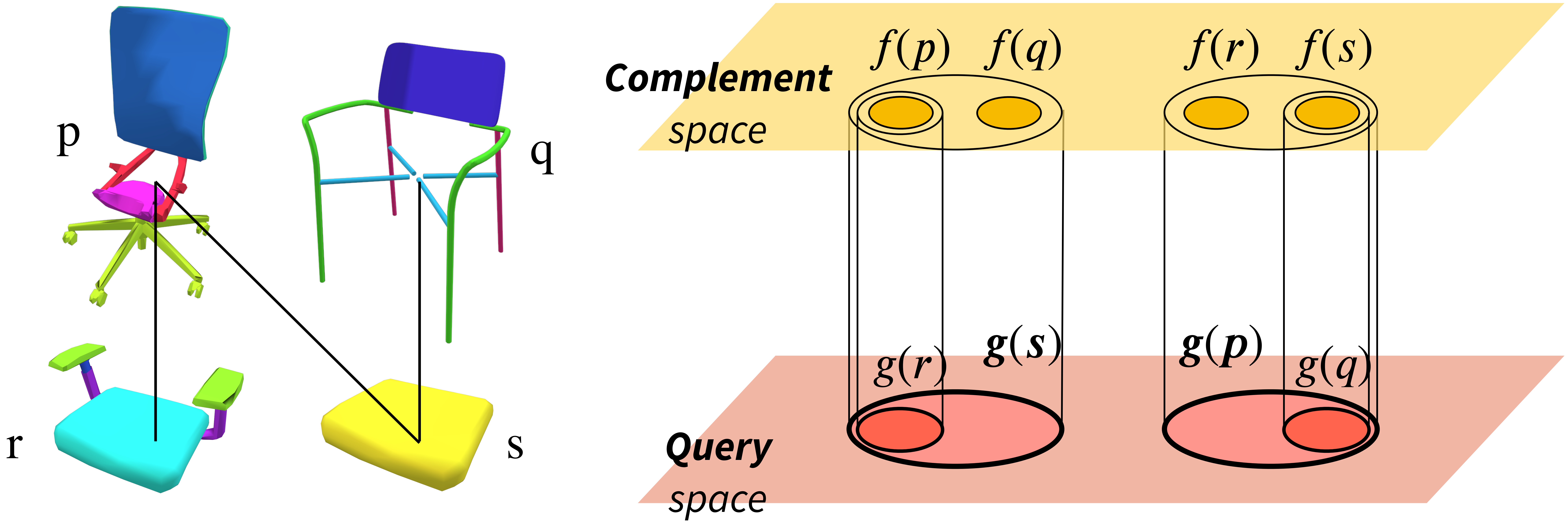
Complementarity as Set Inclusion

Encode *1-to-N* mapping as *set inclusion*



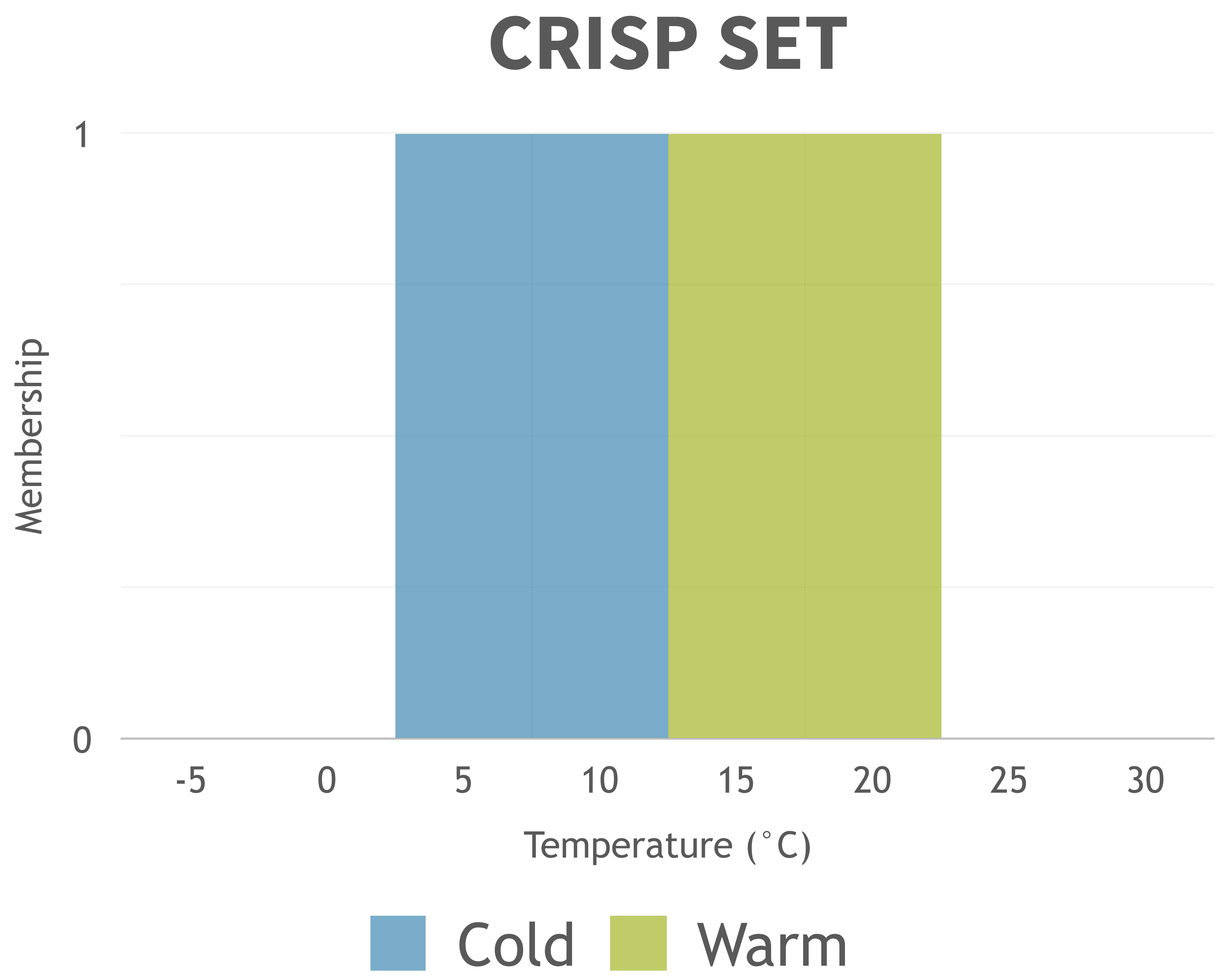
Complementarity as Set Inclusion

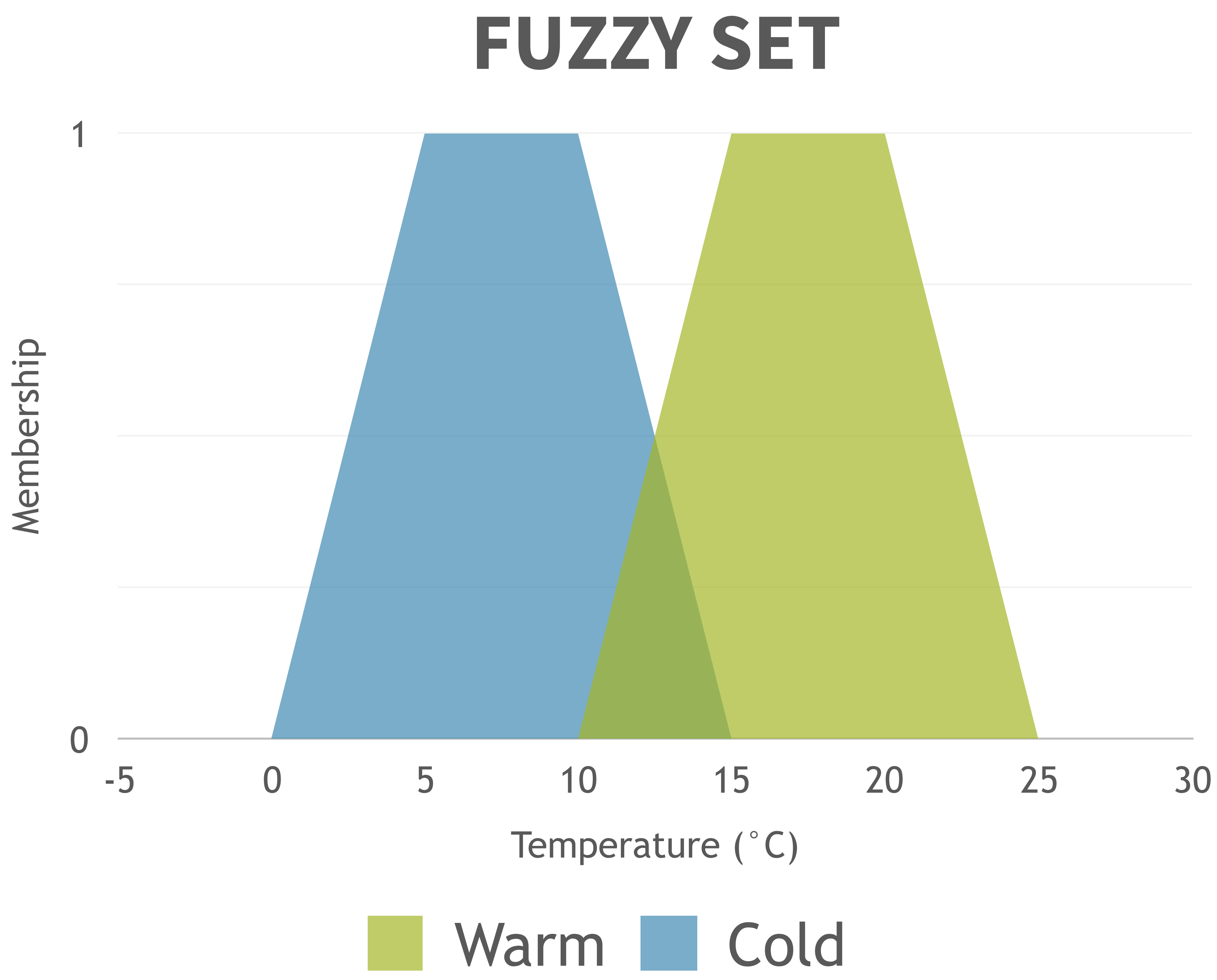
Encode *1-to-N* mapping as *set inclusion*



Fuzzy Sets

Fuzzy set theory [Zadeh, 1965]





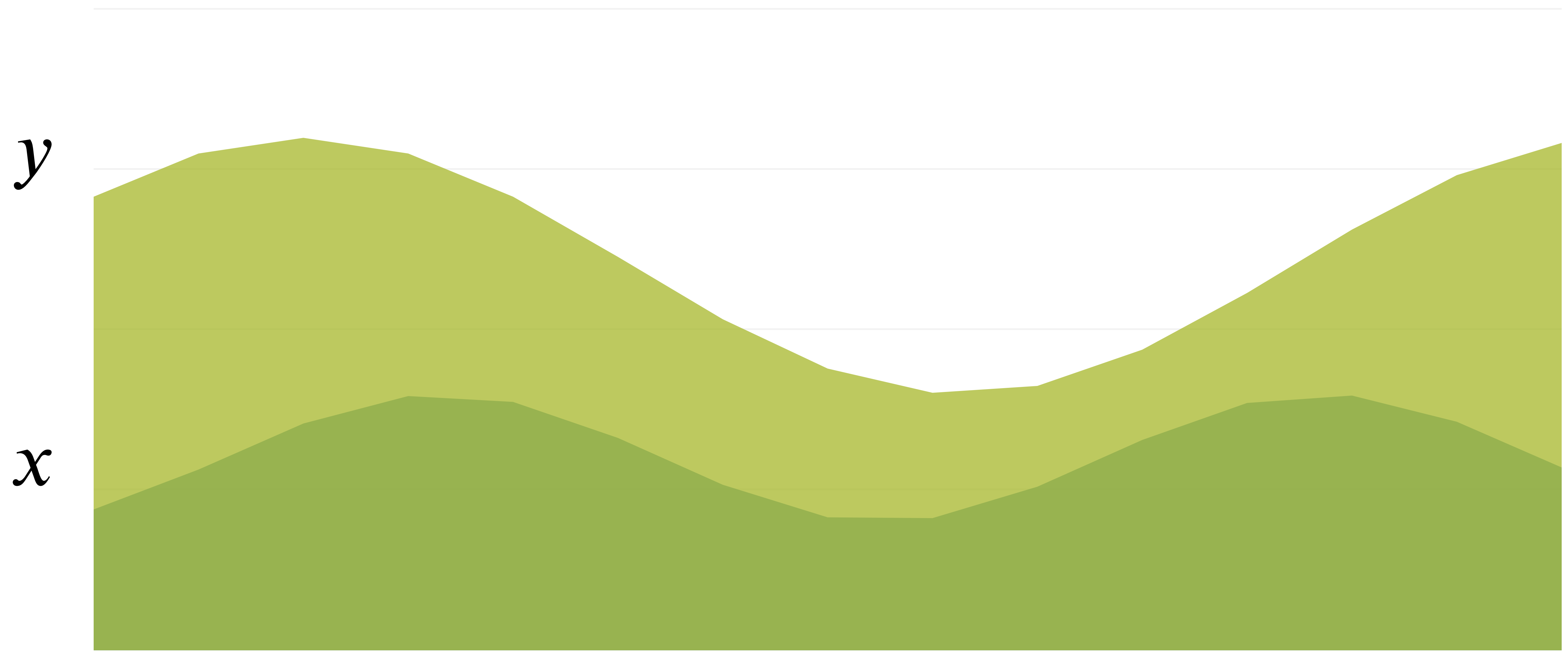
Fuzzy Set Inclusion

$$x \subseteq y \Longleftrightarrow \bigwedge_{i=1}^{D} x_i \leq y_i$$

Inclusion indication operator

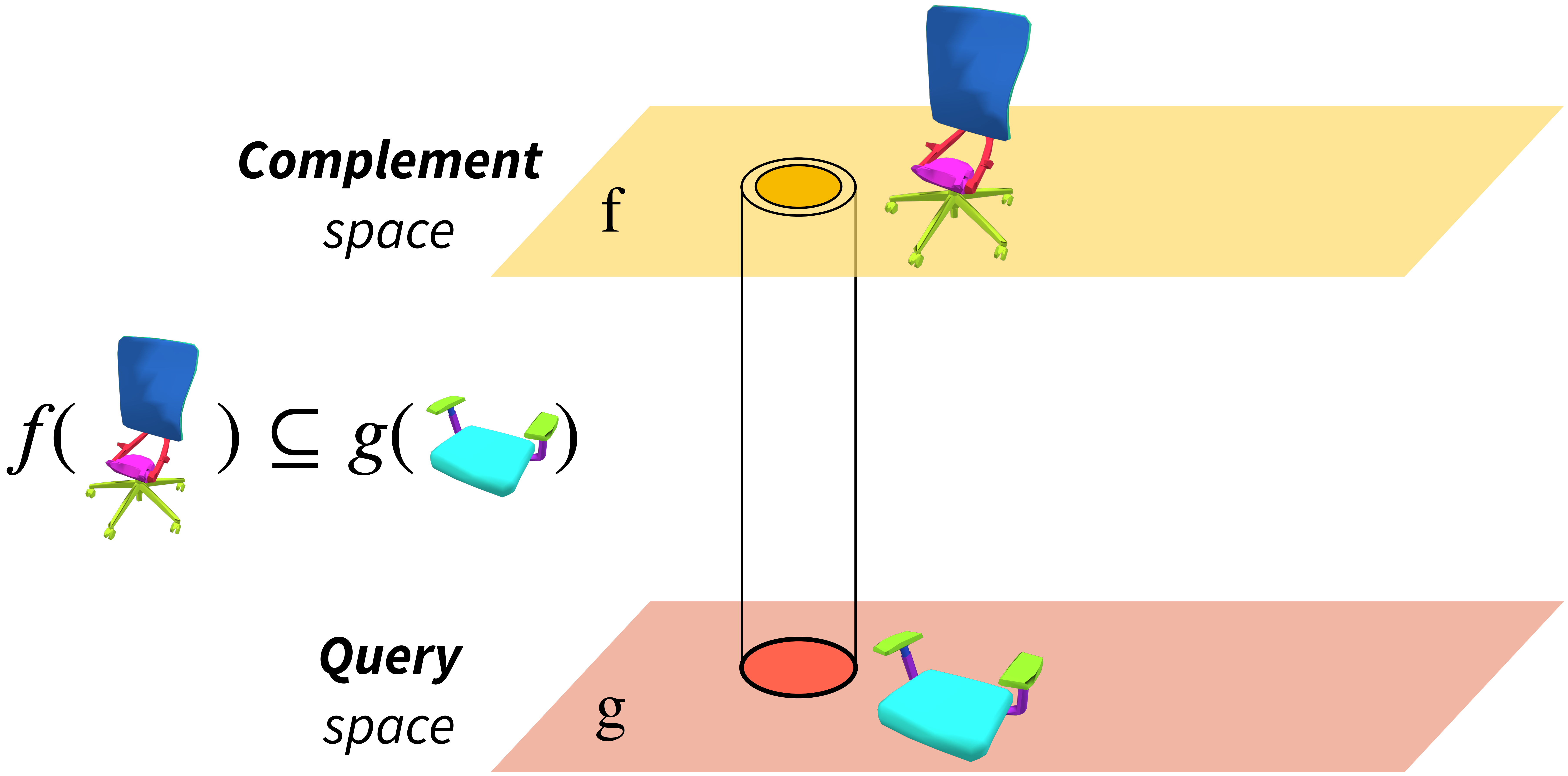
[Vendrov et al., 2016]

$$(x \subseteq y)_i = (x_i \leq y_i)$$



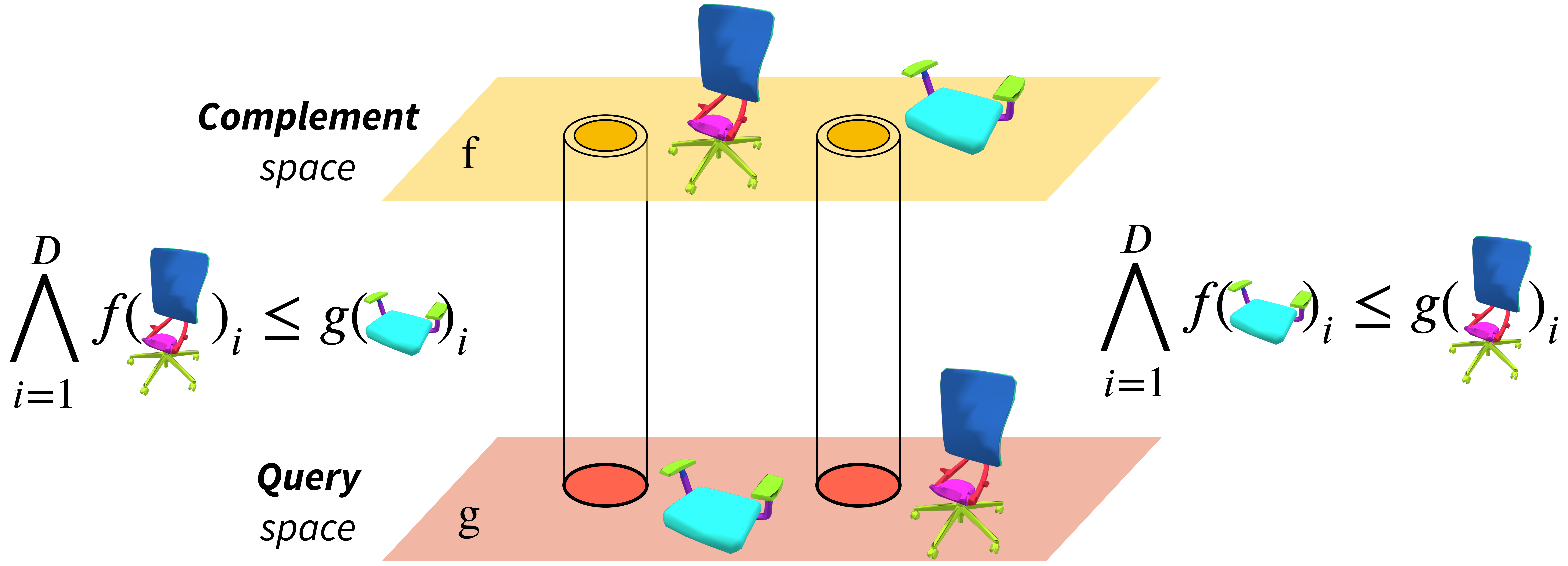
Embedding as Set Inclusion

Inclusion representation in the dual embedding space



Embedding as Set Inclusion

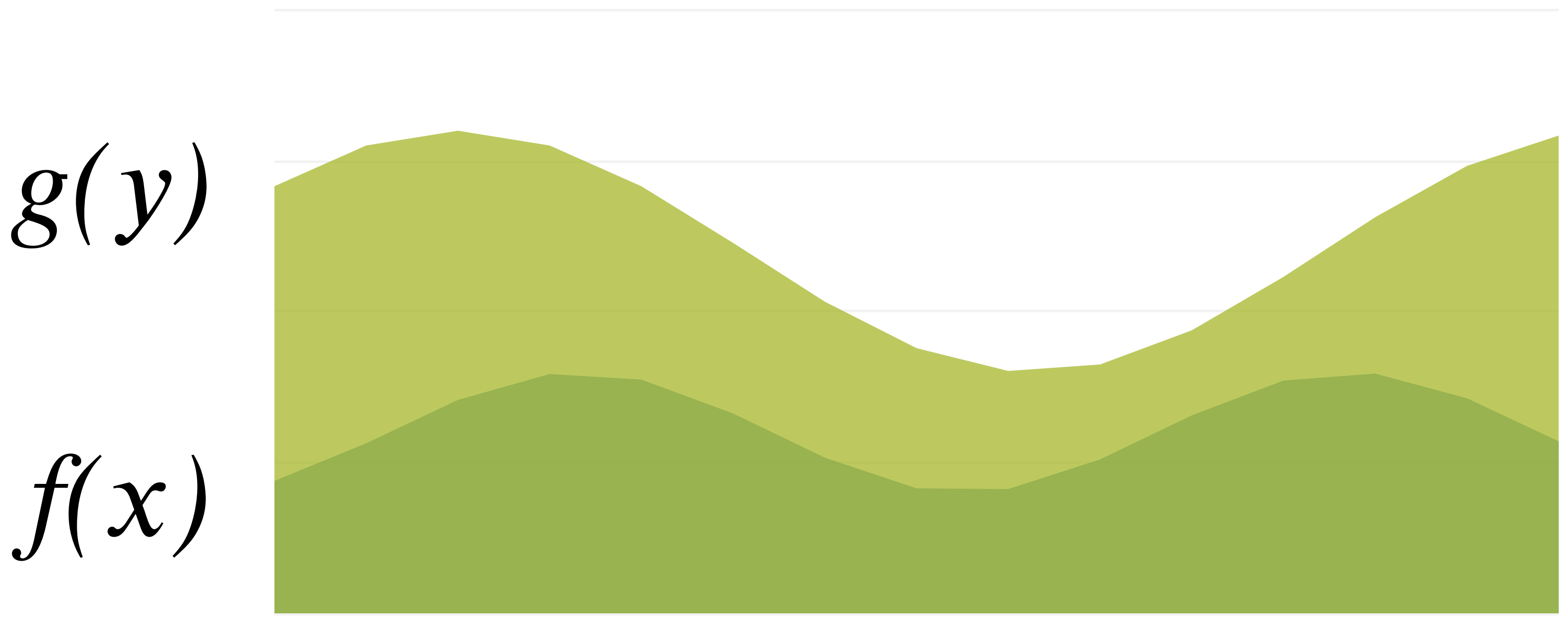
Inclusion representation in the dual embedding space



Complementarity Energy Function

- Inclusion Energy Function

$$E(f(x) \subseteq g(y)) = \sum_i \underbrace{\max\left(0, f(x)_i - g(y)_i\right)^2}_{\textbf{ReLU}}$$



$$(f(x) \subseteq g(y))_i = (f(x)_i \leq g(y)_i)$$

Complementarity Energy Function

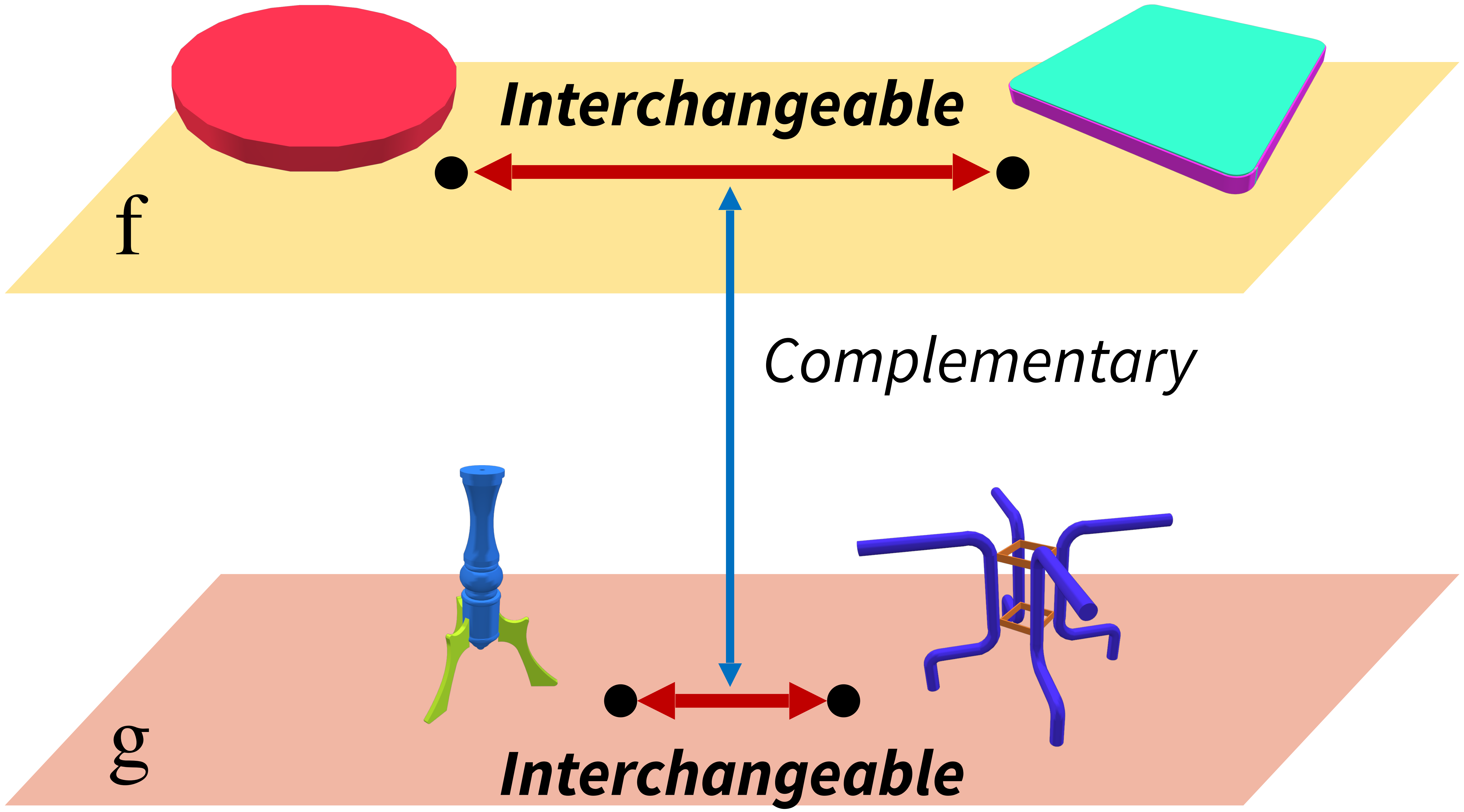
- Inclusion Energy Function

$$E(f(x) \subseteq g(y)) = \sum_i \underbrace{\max\left(0, f(x)_i - g(y)_i\right)^2}_{\textbf{ReLU}}$$

- Complementarity Energy Function

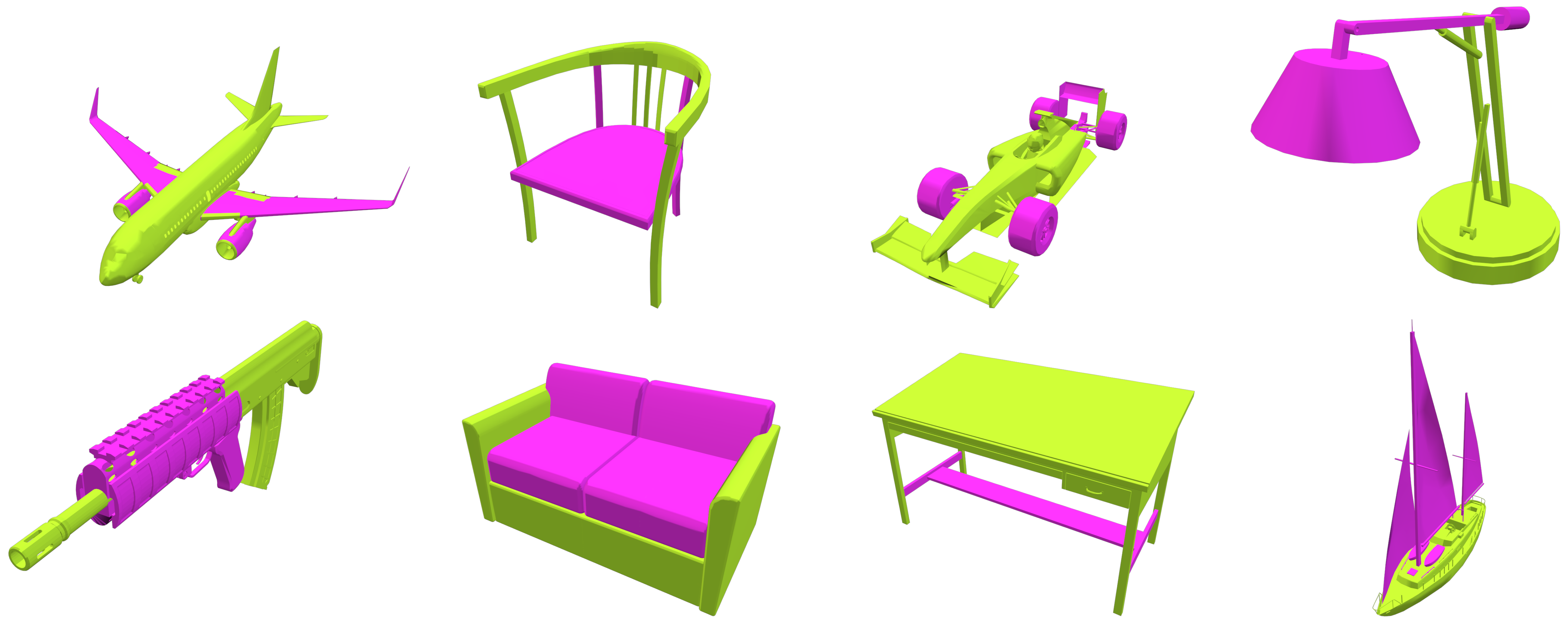
$$E_c(x, y) = E\left(f(x) \subseteq g(y)\right) + E\left(f(y) \subseteq g(x)\right)$$

Interchangeability



Our Training Data

- *Complementary* pairs are created by splitting objects
- No supervision for *interchangeability* is given
    - Learn interchangeability from complementarity



Interchangeability

Similar embedding fuzzy sets (of x and y)

- Have *large intersection* in the *query* space (g)
- Have a *small union* in the *complement* space (f)

$$\max\left(E\Big(f(x) \subseteq g(z)\Big), E\Big(f(y) \subseteq g(z)\Big)\right) \leq E\Big(\underline{\boldsymbol{f}(\boldsymbol{x}) \vee \boldsymbol{f}(\boldsymbol{y})} \subseteq g(z)\Big)$$

$$\max\left(E\Big(f(z) \subseteq g(x)\Big), E\Big(f(z) \subseteq g(y)\Big)\right) \leq E\Big(f(z) \subseteq \underline{\boldsymbol{g}(\boldsymbol{x}) \wedge \boldsymbol{g}(\boldsymbol{y})}\Big)$$

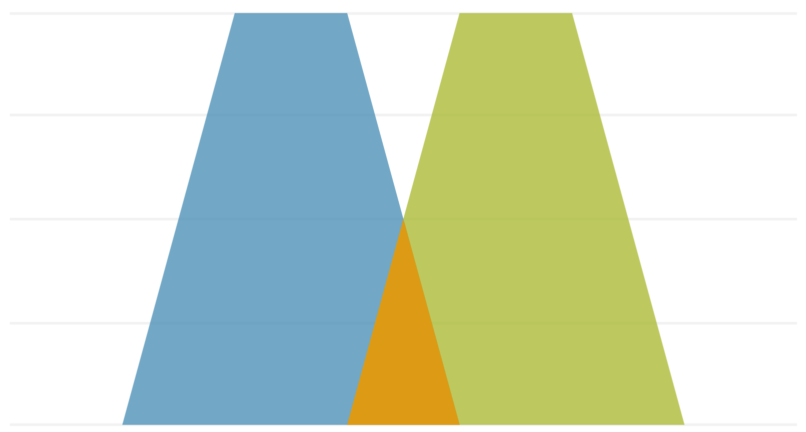
Interchangeability

Similar embedding fuzzy sets (of x and y)

- Have *large intersection* in the *query* space (g)
- Have a *small union* in the *complement* space (f)

$$E_r(x, y) = \underbrace{\|f(x) \vee f(y)\|_2^2}_{\text{small union}} - \underbrace{\|g(x) \wedge g(y)\|_2^2}_{\text{large intersection}}$$

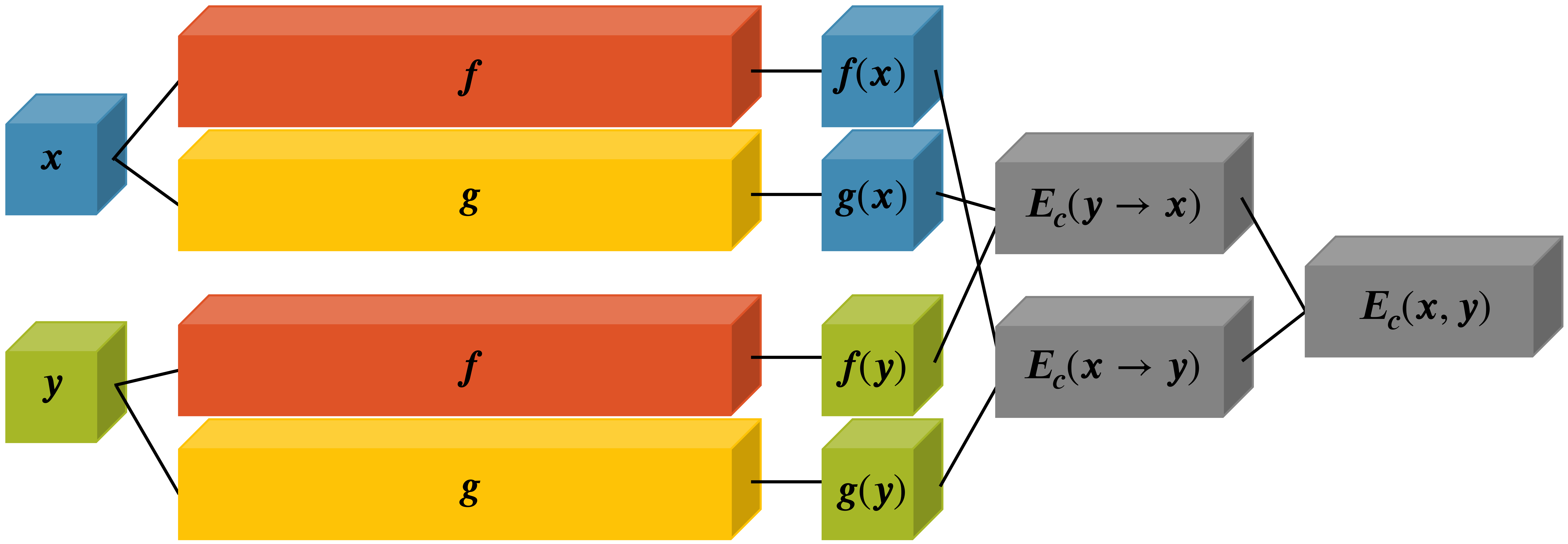
$$(x \wedge y)_i = \min(x_i, y_i)$$



$$(x \vee y)_i = \max(x_i, y_i)$$

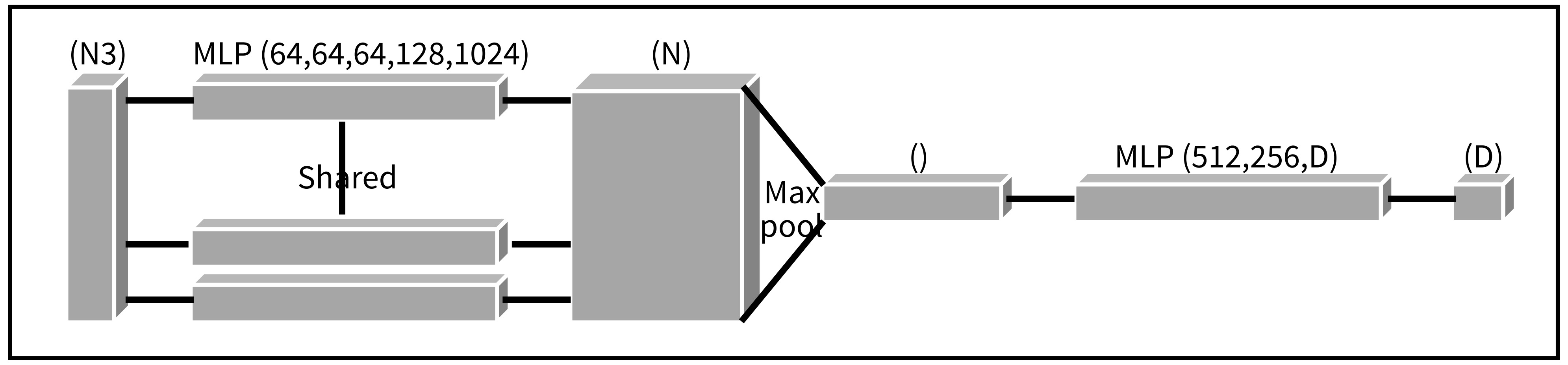
Neural Network Architecture

Dual siamese structure for a pair of complementary shapes



Neural Network Training Details

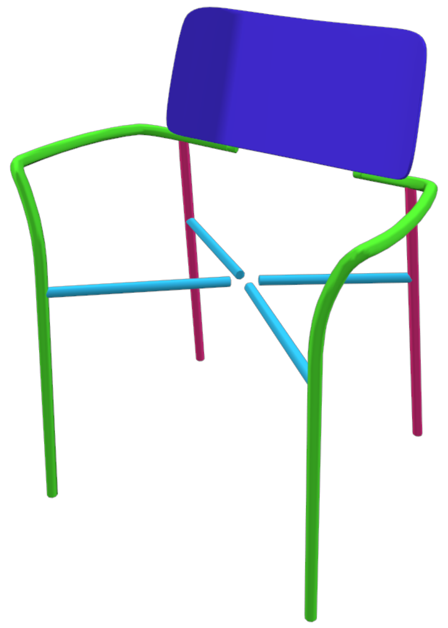
- PointNet [Qi et al., 2017] for f and g
  Convert a mesh to 1K point samples



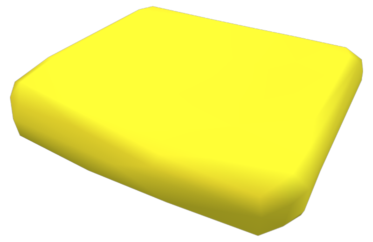
- Ranking Loss [Socher et al., 2014]

  Randomly split shapes into pairs of connected component sets

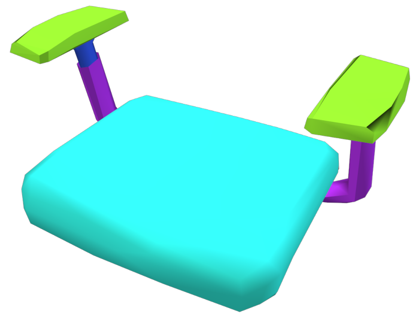
$$E_c(X, Y) + \alpha \leq E_c(X, Z)$$



Query X



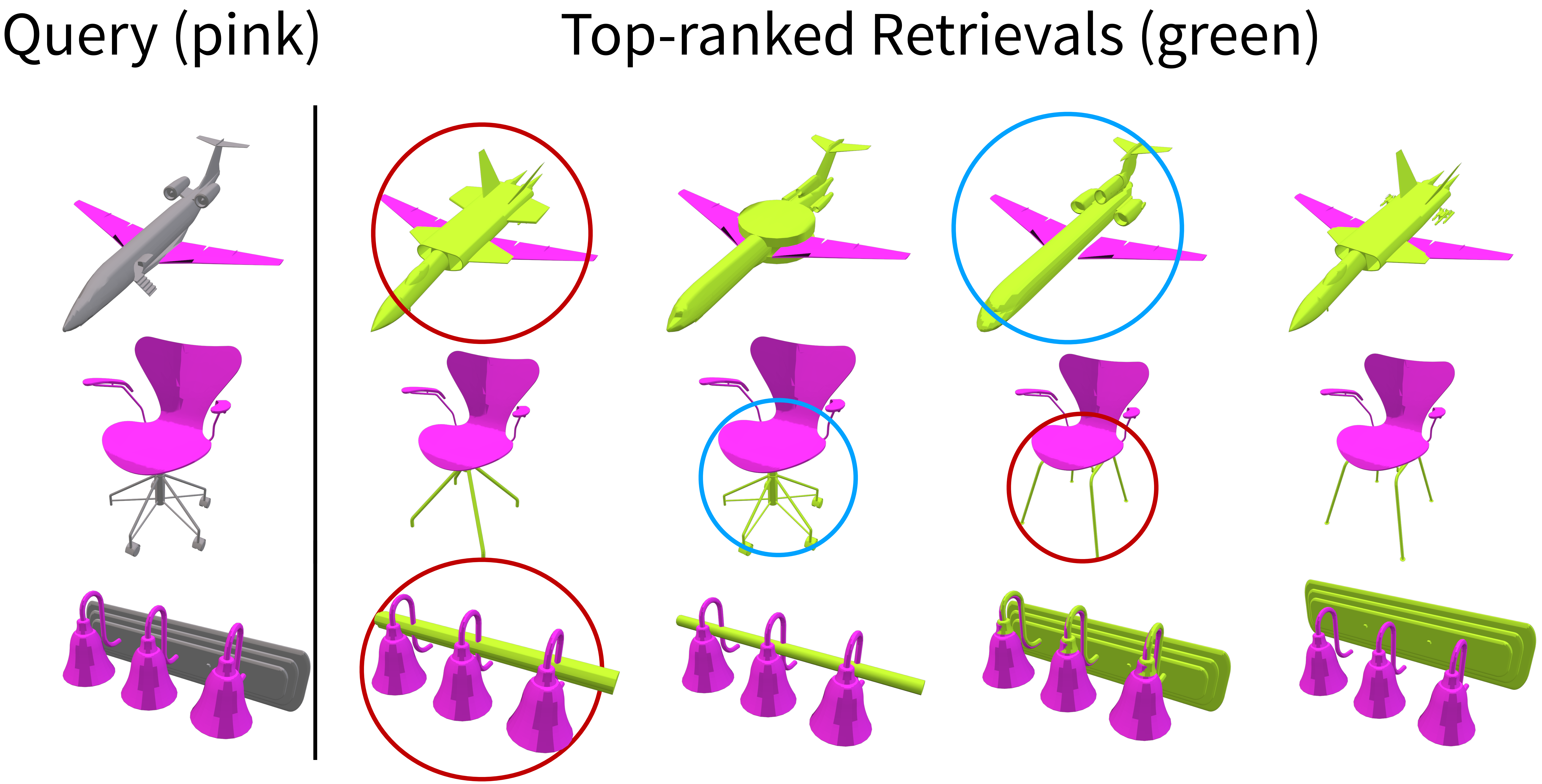
Positive
Complement



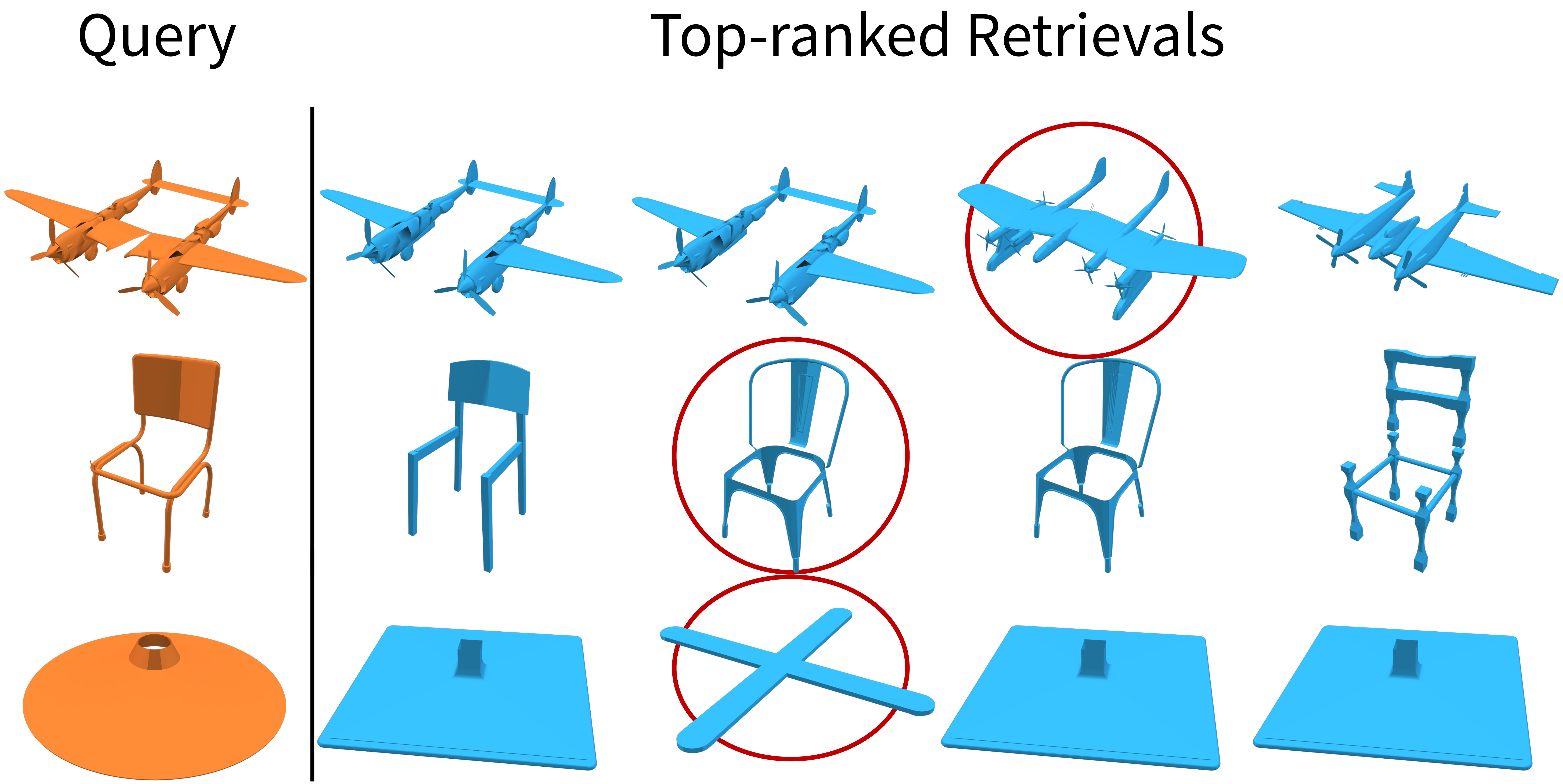
Negative
Complement

Evaluation

Complementary Shape Retrievals



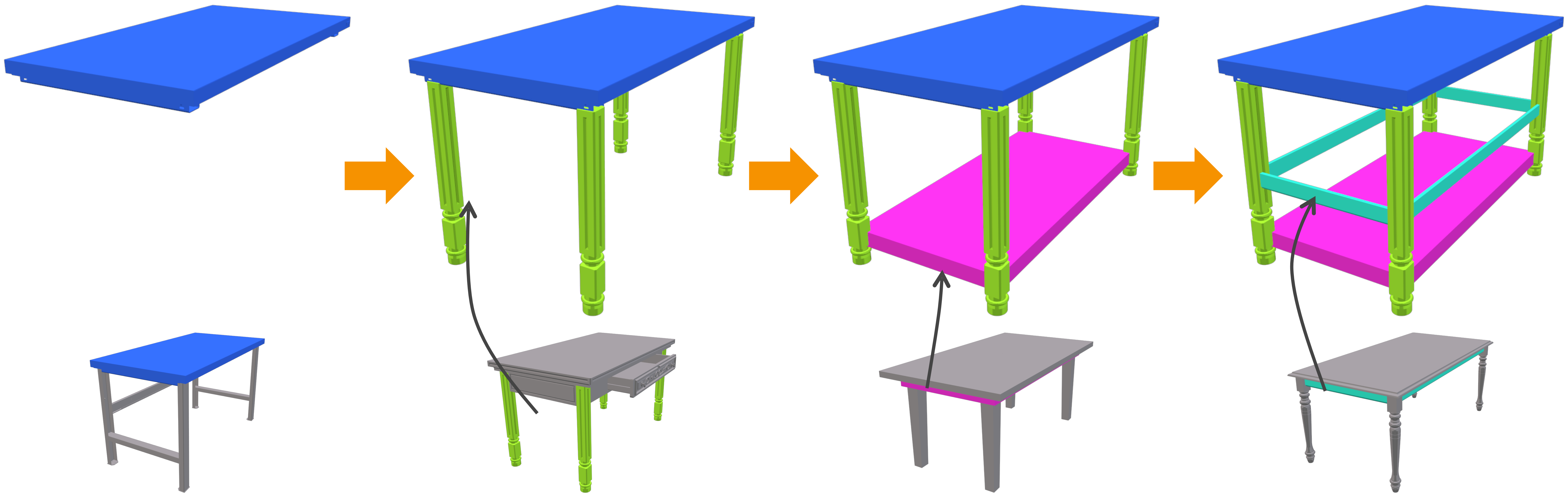
Interchangeable Shape Retrievals



Comparison and Quantitative Evaluation

ComplementMe [Sung et al., 2017]

- Assemble components *iteratively*
- Learn a relation from a partial shape to a single component



Complementarity Evaluation

Evaluation metrics using *ranks* of *ground truth* complements

Category (# Partial Shapes)		Airplane (4140)	Car (5770)	Chair (8374)	Guitar (198)	Lamp (1778)	Rifle (1184)	Sofa (4452)	Table (4594)	Watercraft (1028)	Mean
Recall@1	CM	9.9	2.4	4.9	19.2	1.7	1.9	3.9	2.7	0.7	4.3
	Ours	**17.5**	**5.8**	**8.0**	**23.7**	**5.1**	**7.3**	**6.7**	**4.1**	**3.2**	**7.8**
Recall@10	CM	48.6	15.5	27.2	67.7	11.1	17.1	20.0	15.5	7.3	23.5
	Ours	**61.3**	**30.5**	**35.0**	**72.2**	**19.7**	**23.5**	**30.1**	**19.2**	**14.3**	**32.9**
Median Percentile Rank	CM	99.8	98.8	99.6	97.0	89.6	94.3	98.5	98.3	87.0	97.9
	Ours	**99.9**	**99.5**	**99.7**	**98.5**	**90.4**	**95.8**	**99.2**	**98.5**	**88.7**	**98.4**
Mean Percentile Rank	CM	98.4	96.4	98.3	**94.5**	**81.4**	88.2	94.0	94.9	77.6	94.8
	Ours	**98.5**	**97.2**	**98.5**	93.8	79.9	**89.0**	**94.9**	**95.0**	**78.7**	**95.2**

Recall@N = the percentage of the ground truth complements in the top-N rank retrievals

Percentile Rank = percentage of partial shapes having ranks equal or greater than the rank of the ground truth complements

Limitations

- Can complete only at the level of components
- May not be able to handle small and thin parts

Summary

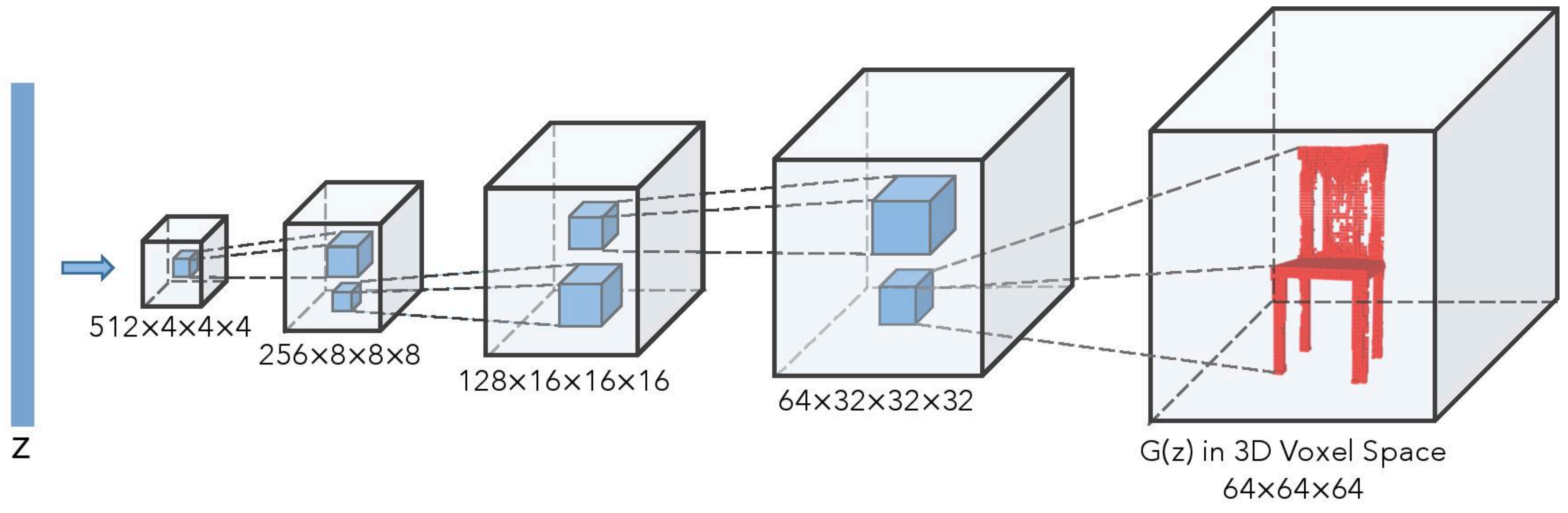
- Proposed to jointly learn *complementarity* and *interchangeability* relations in *dual embedding spaces*
- Encoded complementarity and interchangeability as *inter-* and *intra*-space relations
- Introduced a *fuzzy set representation* for an one-to-*N* mapping

Questions?

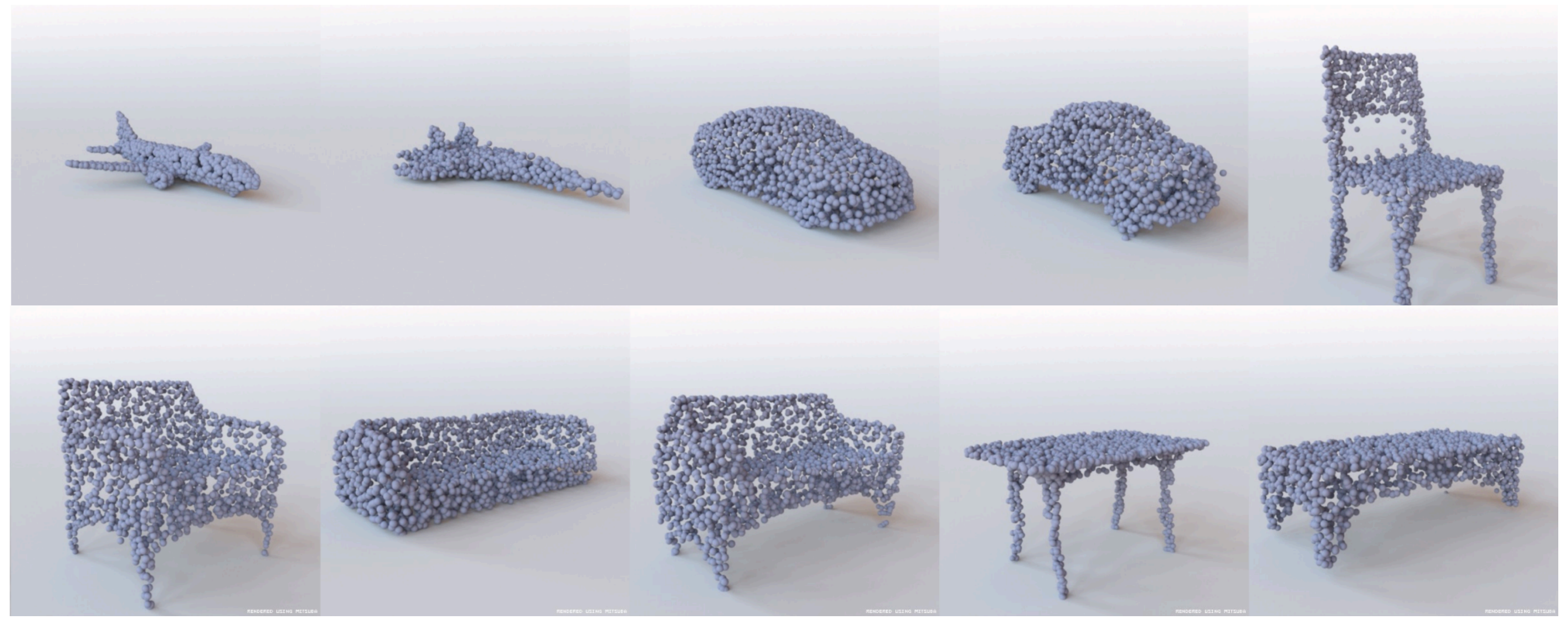
Composite Shape Generation via Latent Space Factorization

joint work with Fei Xia, Panos Achlioptas, Mira Shalah and Leonidas Guibas

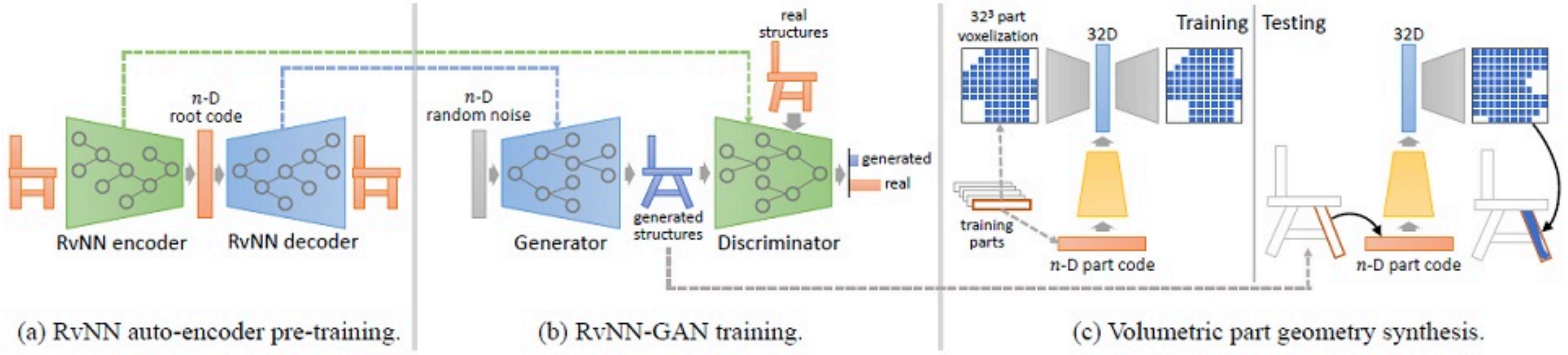
Standard DL Generation Pipeline



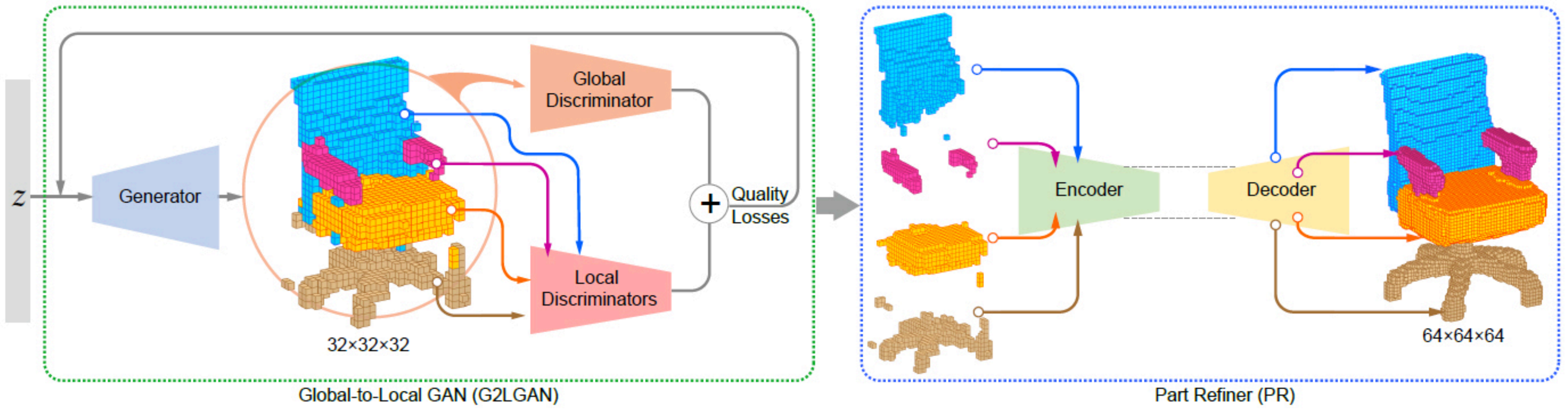
Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling
[Wu et al., NIPS 2016]



Learning Representations and Generative Models for 3D Point Clouds
[Achlioptas et al., ICML 2018]



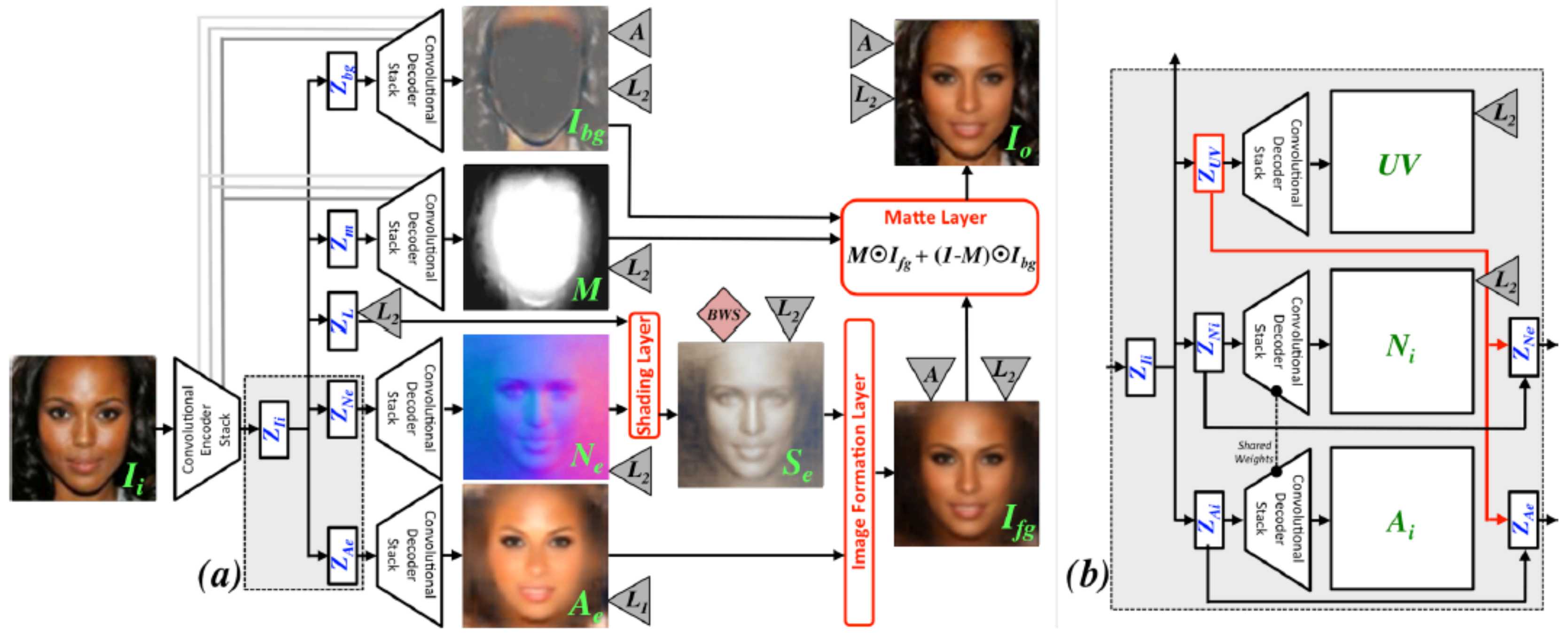
GRASS: Generative Recursive Autoencoders for Shape Structures
[Li et al., SIGGRAPH 2017]



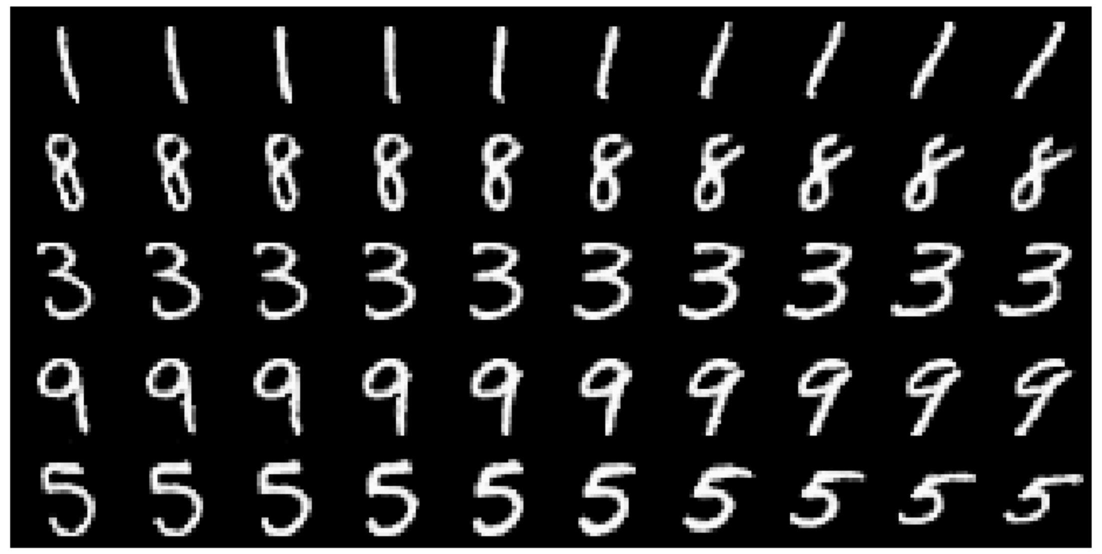
Global-to-Local Generative Model for 3D Shapes
[Wang et al., SIGGRAPH Asia 2018]

Motivation from 2D Image Synthesis

Latent space disentangling in 2D



Neural Face Editing with Intrinsic Image Disentangling
[Shu et al., SVPR 2017]

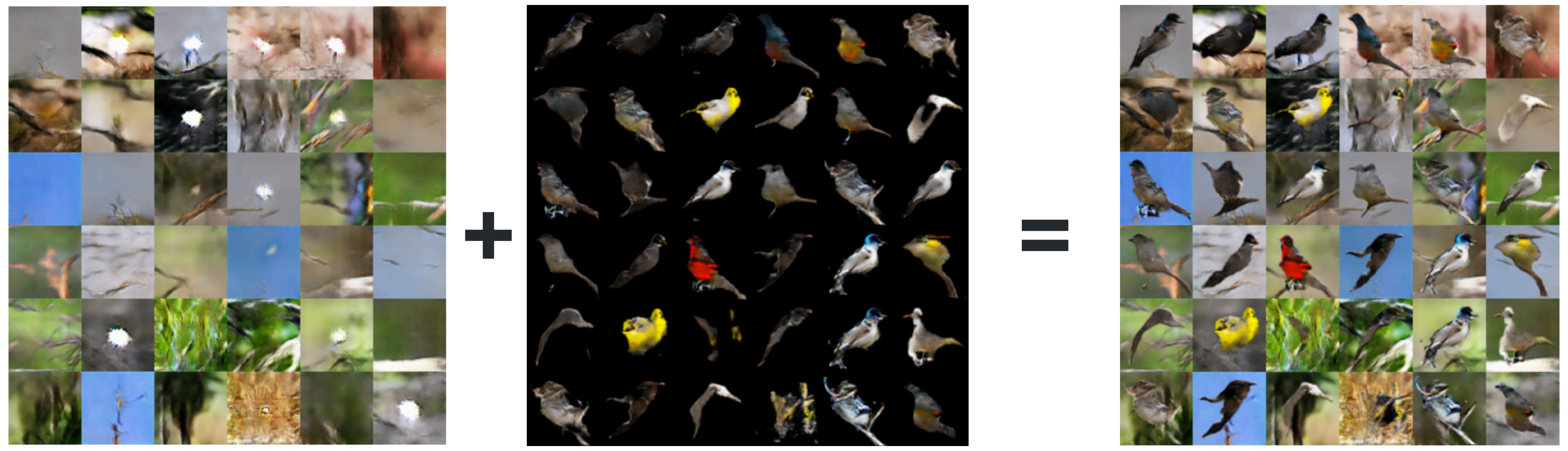


(c) Varying c_2 from -2 to 2 on InfoGAN (Rotation)

InfoGAN [Chen et al., NIPS 2016]

Image composition

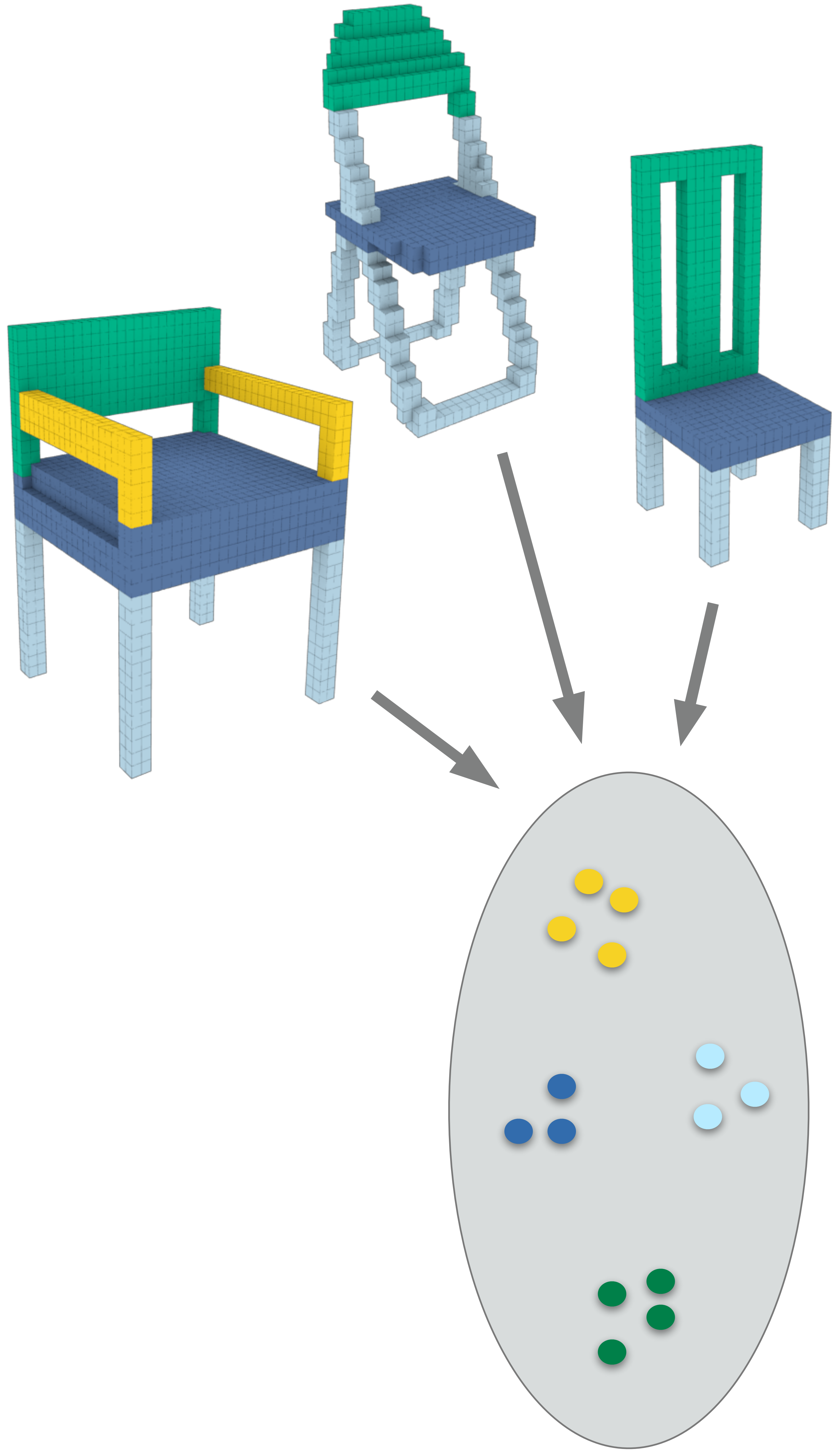
ST-GAN: Spatial Transformer Generative Adversarial Networks for Image Compositing
[Lin et al. CVPR 2018]



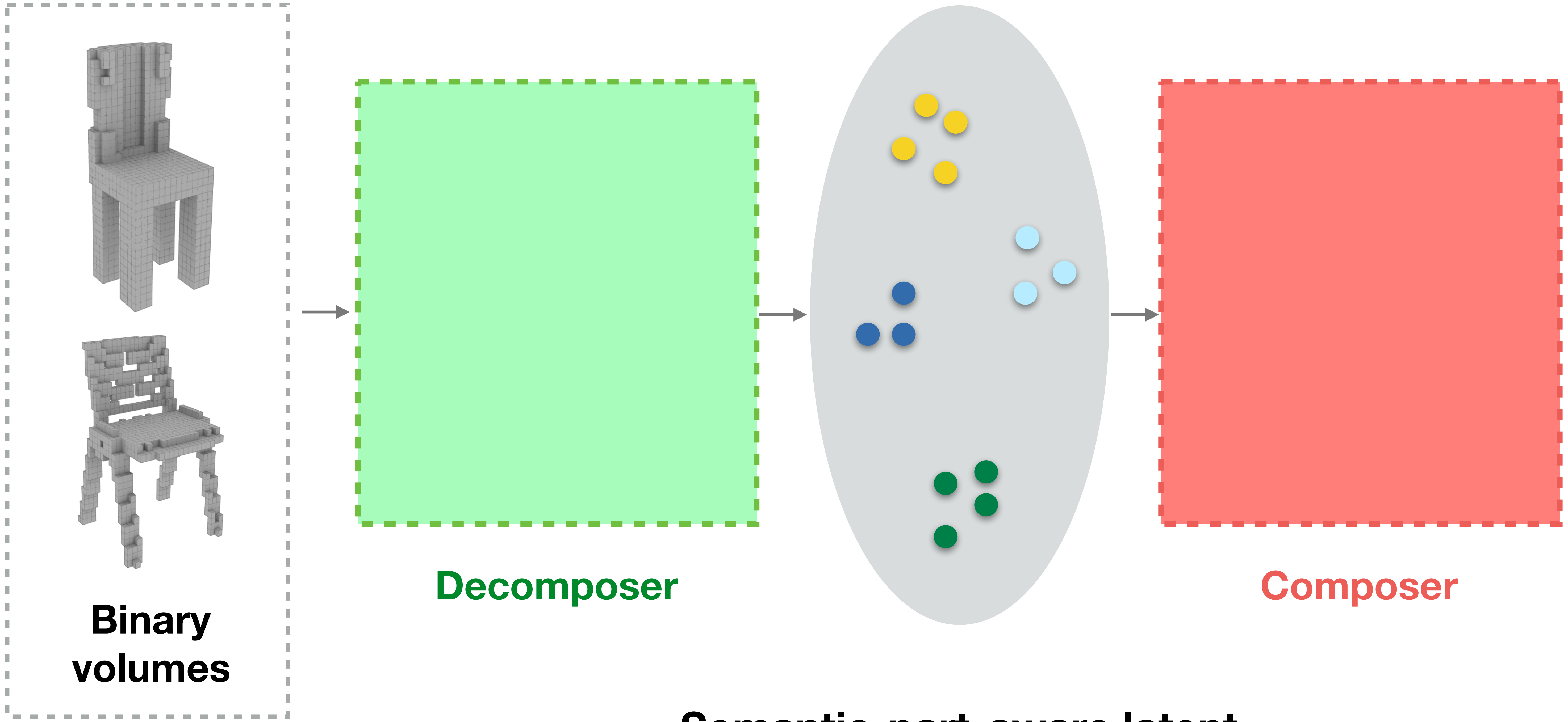
LR-GAN: Layered Recursive GAN for Image Generation
[Yang et al. ICLR 2017]

Our goals

- Construct a *factorized latent space* which encodes both *shape geometry* and its *semantic structure*;
- Operate on *unsegmented data*;
- Perform shape composition and decomposition in latent space.



Decomposer-Composer Network

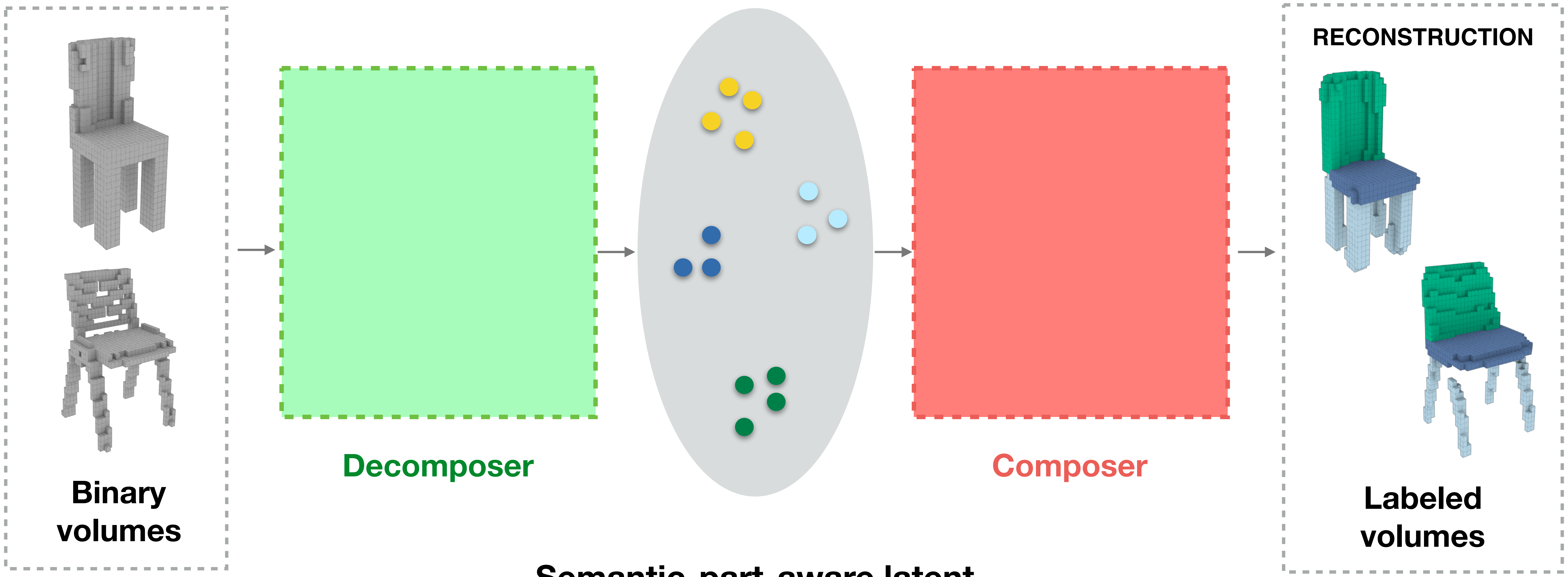


Semantic-part-aware latent space with simple shape composition operator

- armrests
- seats
- backs
- legs

- **Decomposer** maps **unlabeled** shapes into a factorized latent space

Decomposer-Composer Network



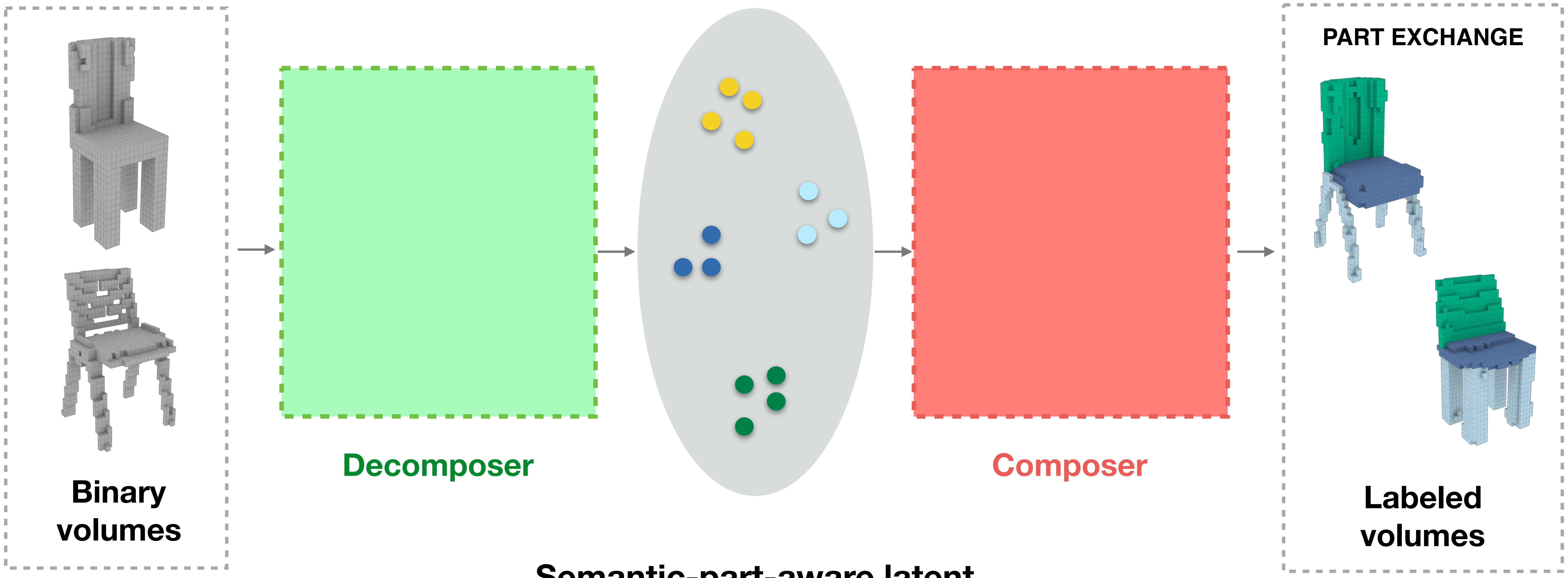
Semantic-part-aware latent space with simple shape composition operator

- **Decomposer** maps **unlabeled** shapes into a factorized latent space

- armrests
- seats
- backs
- legs

- **Composer** reconstructs shapes with semantic part labels from latent part representations

Decomposer-Composer Network



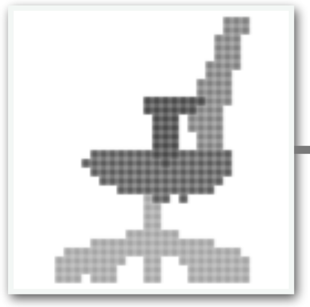
Semantic-part-aware latent space with simple shape composition operator

- **Decomposer** maps **unlabeled** shapes into a factorized latent space

- armrests
- seats
- backs
- legs

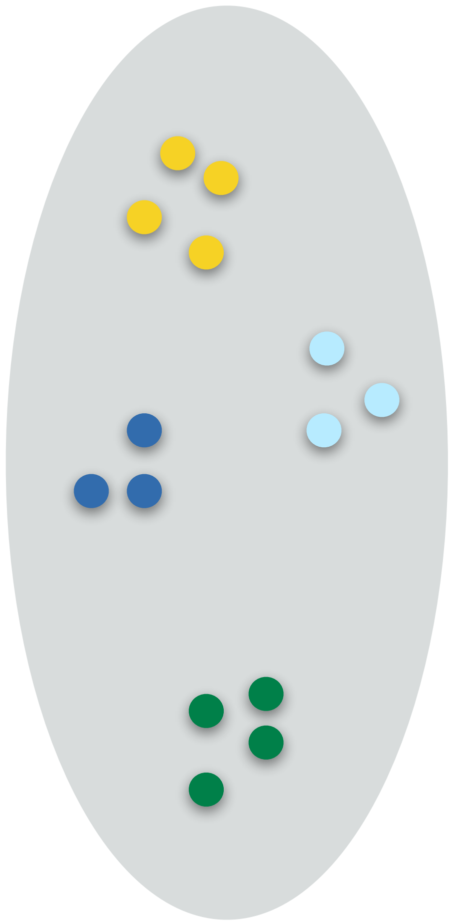
- **Composer** reconstructs shapes with semantic part labels from latent part representations

Simple Decomposer Network



Binary volume

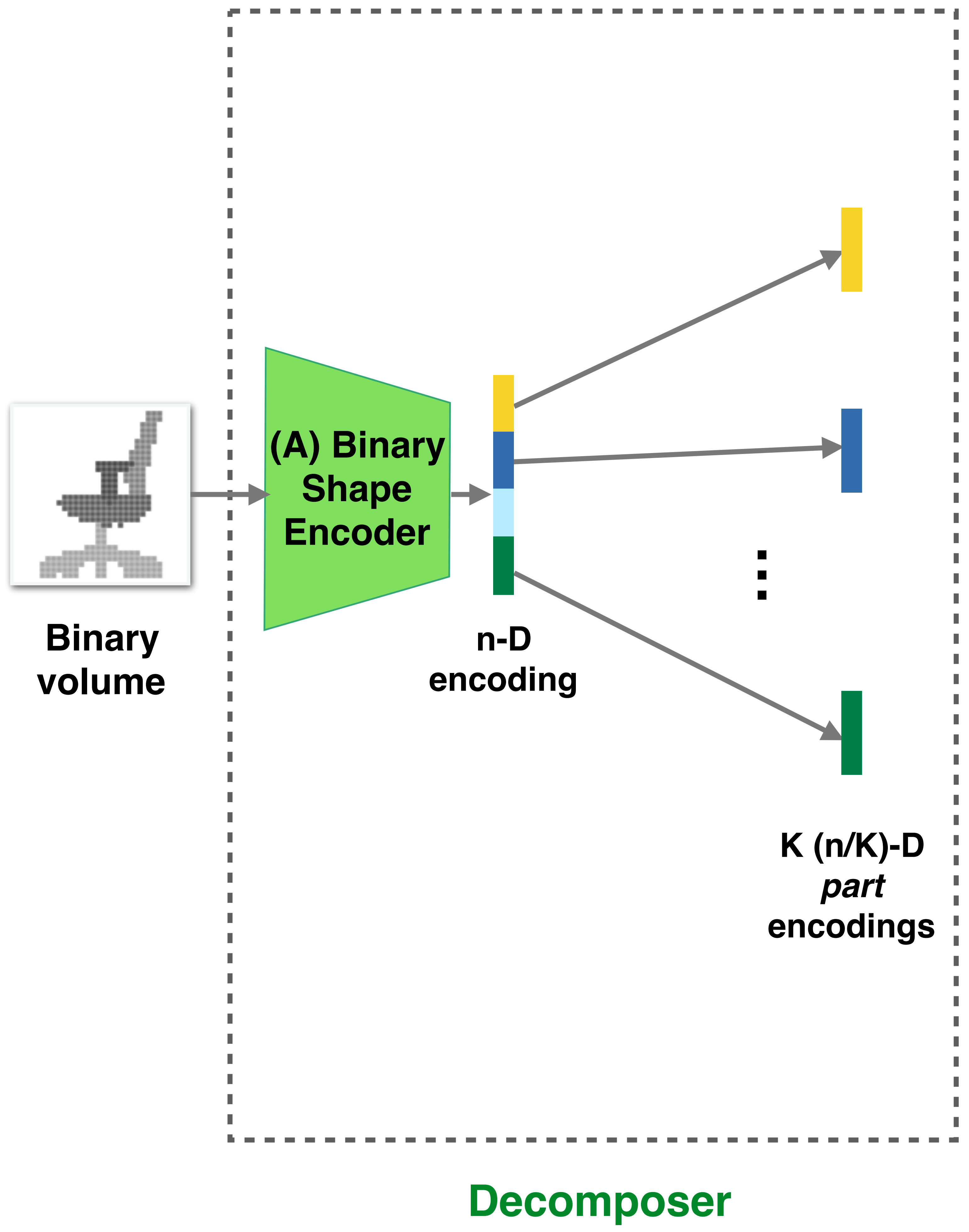
Decomposer

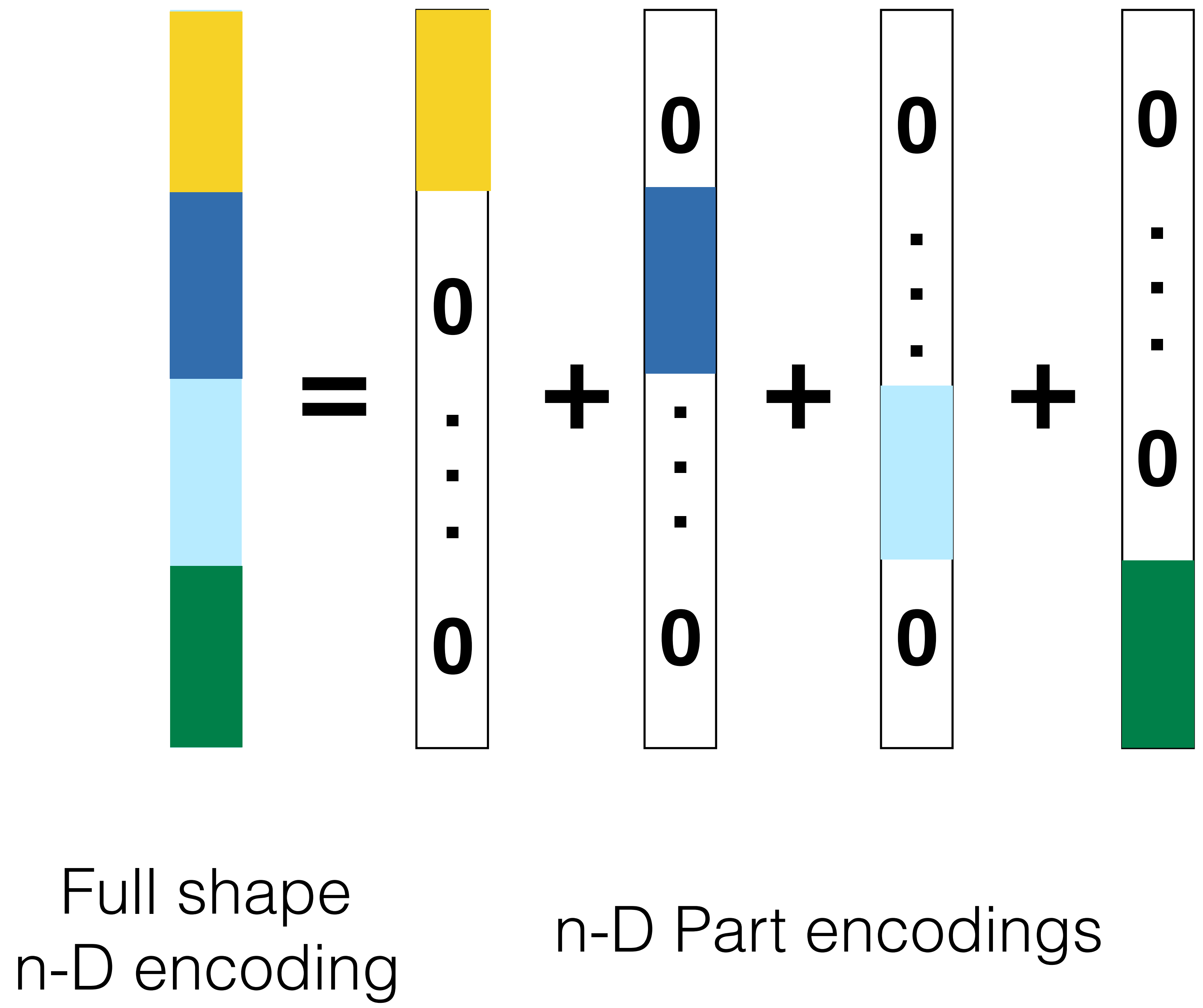


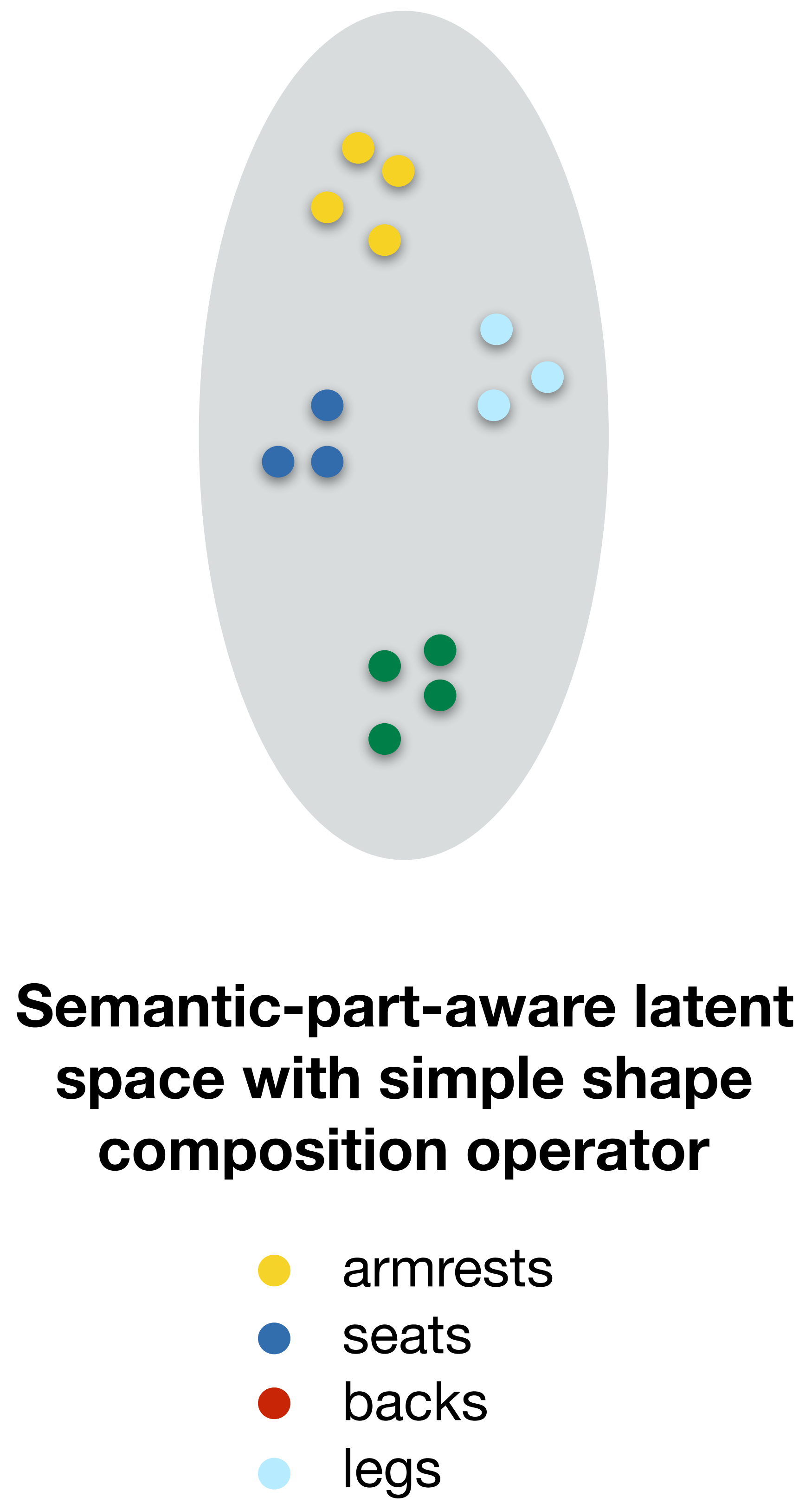
Semantic-part-aware latent space with simple shape composition operator

- armrests
- seats
- backs
- legs

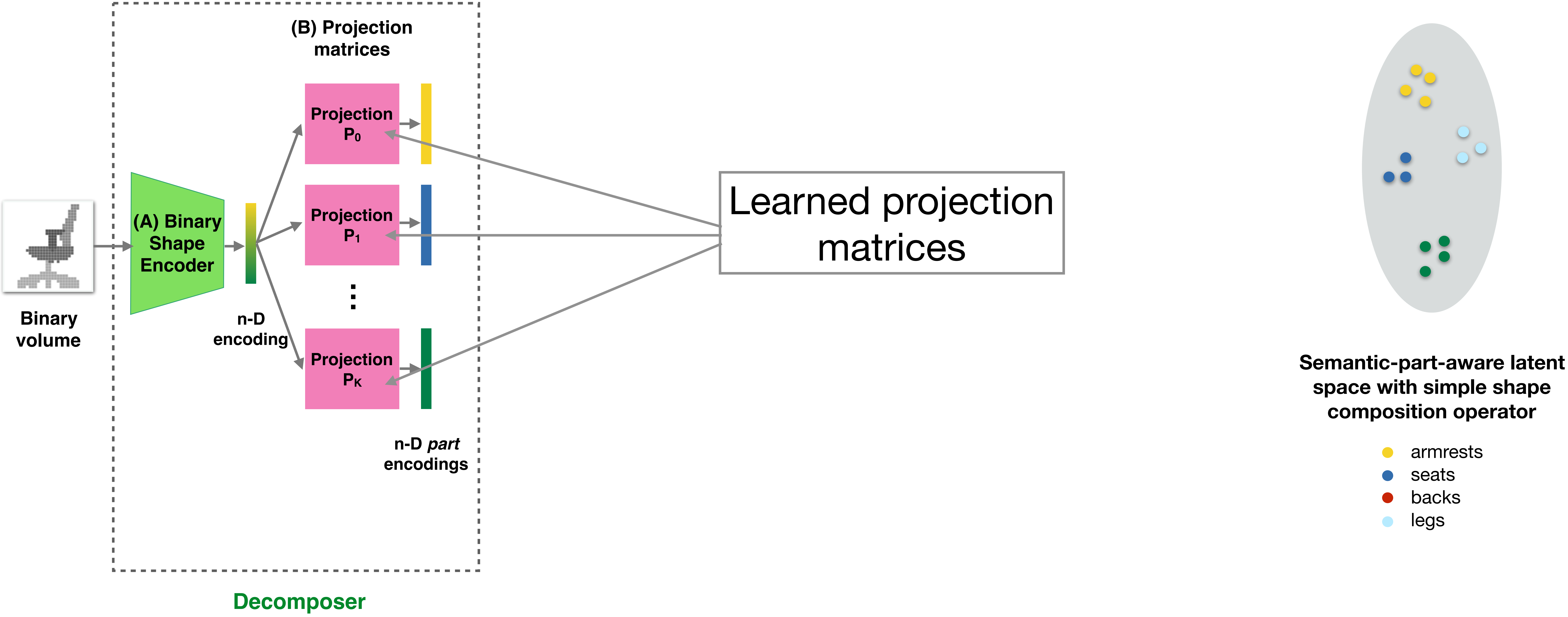
Simple Decomposer Network







Decomposer Network



Partition of the Identity

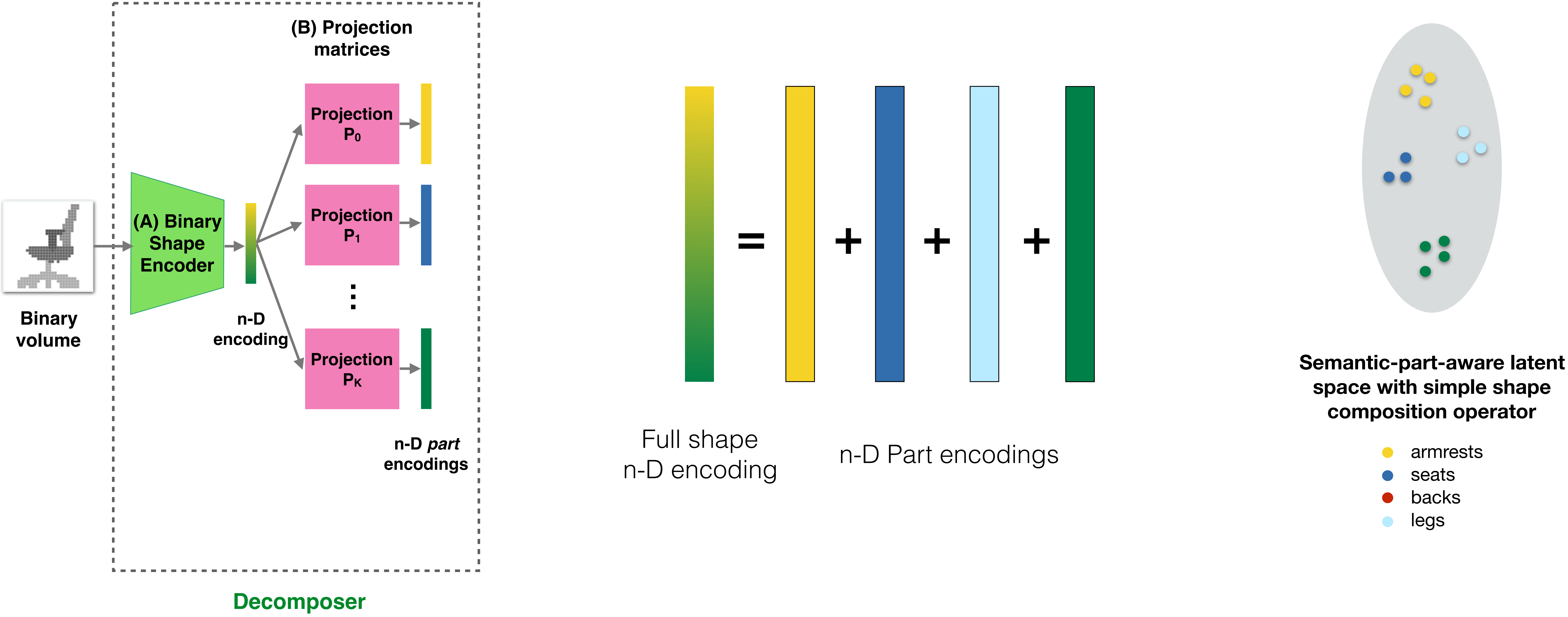
- Embedding decomposition is unique and reversible if $P_1, \ldots, P_k$ form a **partition of the identity**

$$\begin{aligned}
&(1)\ \ P_i^2 = P_i, \forall i, \\
&(2)\ \ P_i P_j = 0 \text{ whenever } i \neq j, \\
&(3)\ \ P_1 + \ldots + P_K = I,
\end{aligned}$$

- Then, $\{P_i\}_{i=1}^{K}$ divide $\mathbb{R}^n$ into K subspaces $\{V_i\}_{i=1}^{K}$ such that

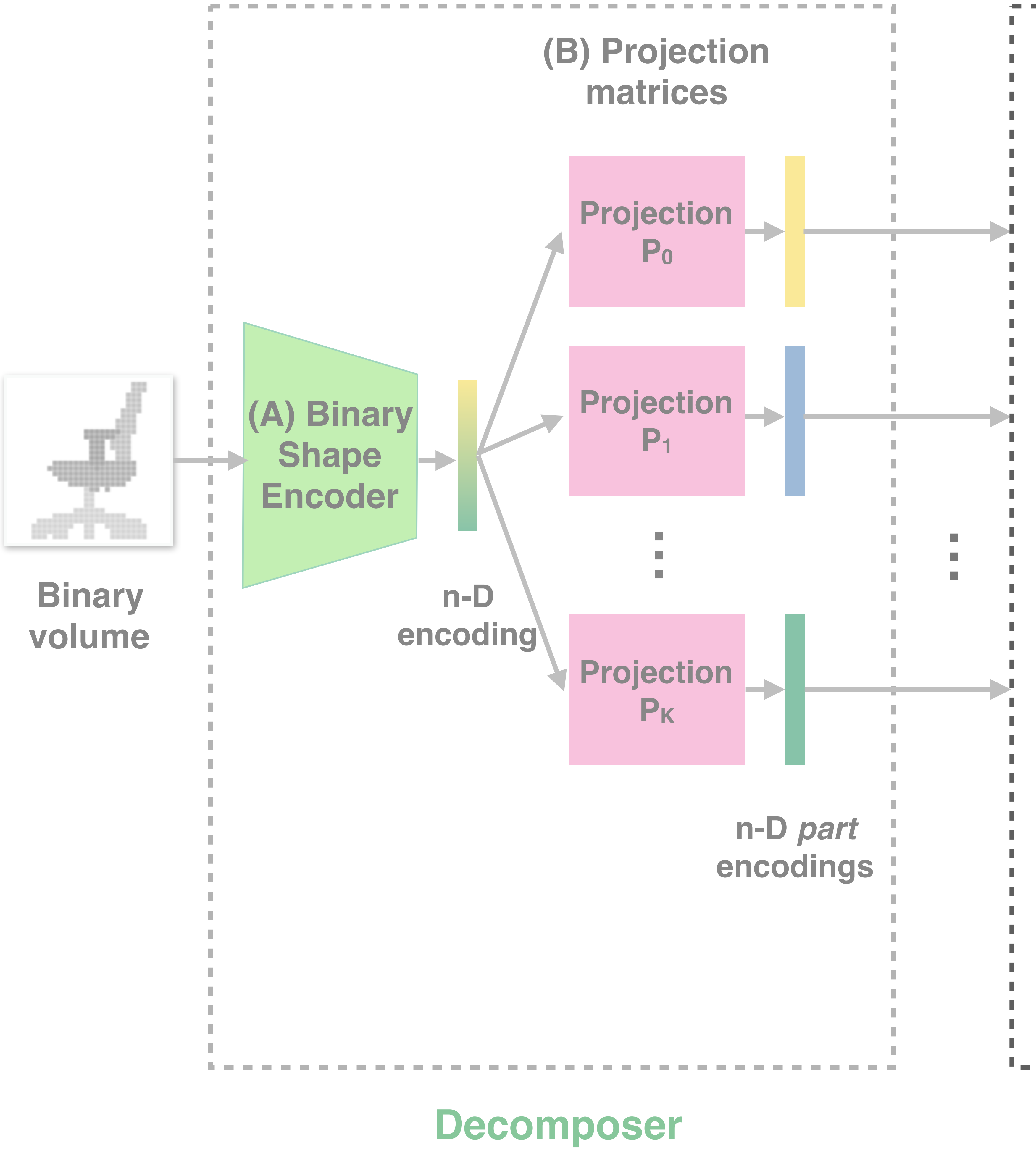
$$\mathbb{R}^n = V_1 \oplus \ldots \oplus V_k$$

Decomposer Network



Simple Composer Network

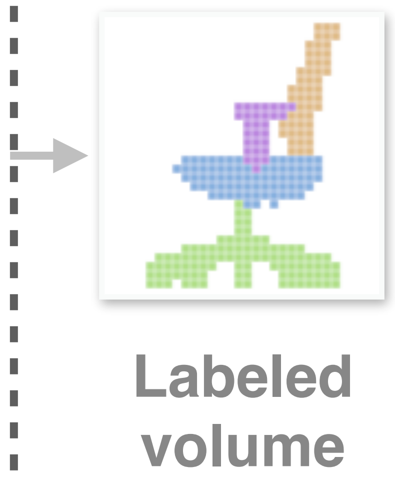
Proposed Composer Network



Proposed approach

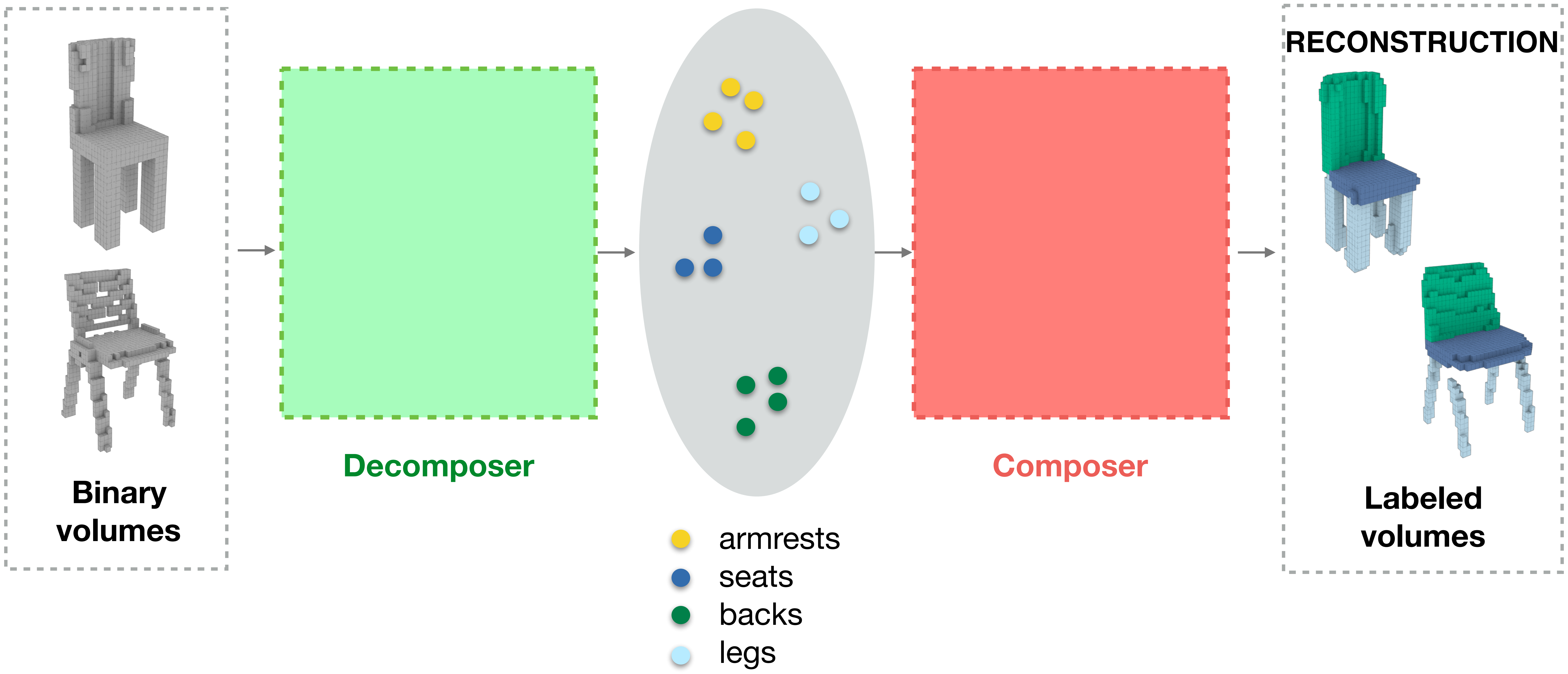
- Reconstruct *scaled* parts *separately*
- Use a 3D Spatial Transformer Net to compose shape from parts
- Assumption: *affine* transformations suffice to compose plausible shape

Composer



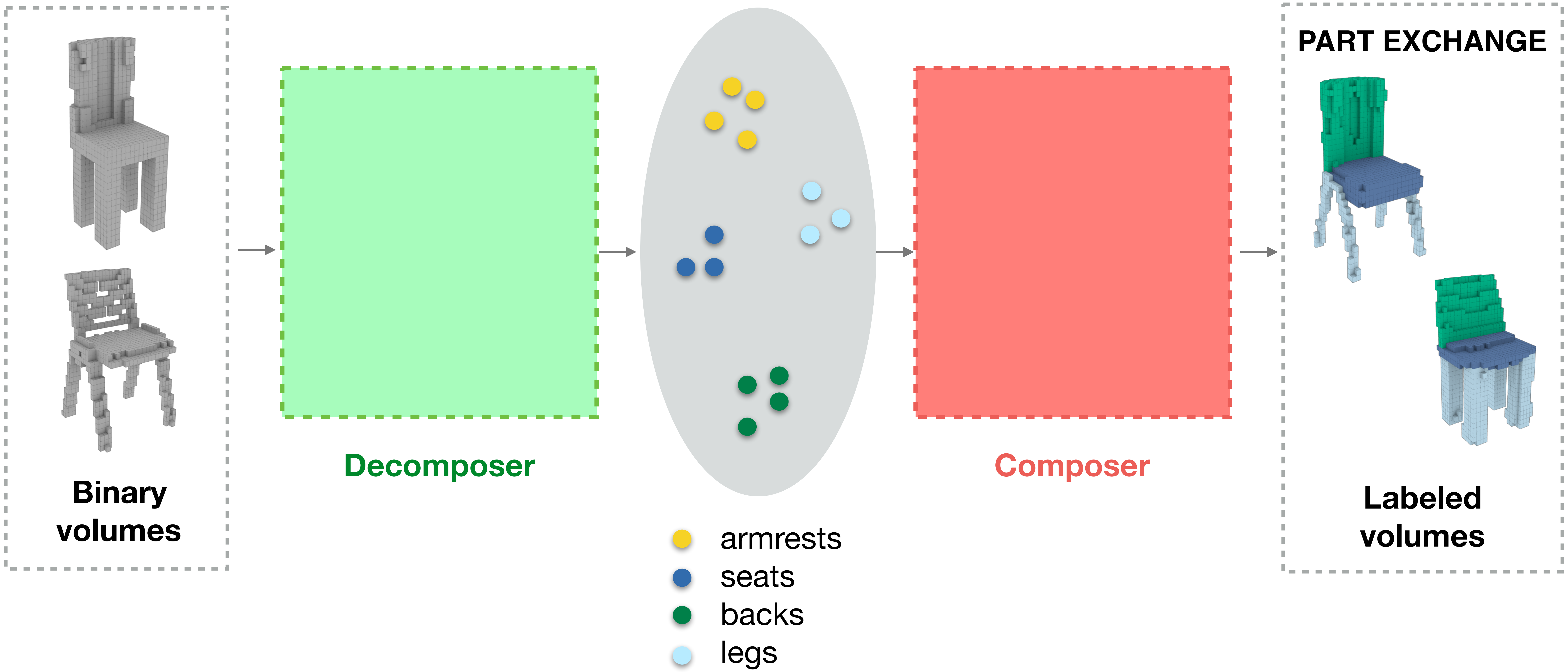
Proposed Composer Network

Shape Decomposition and Composition



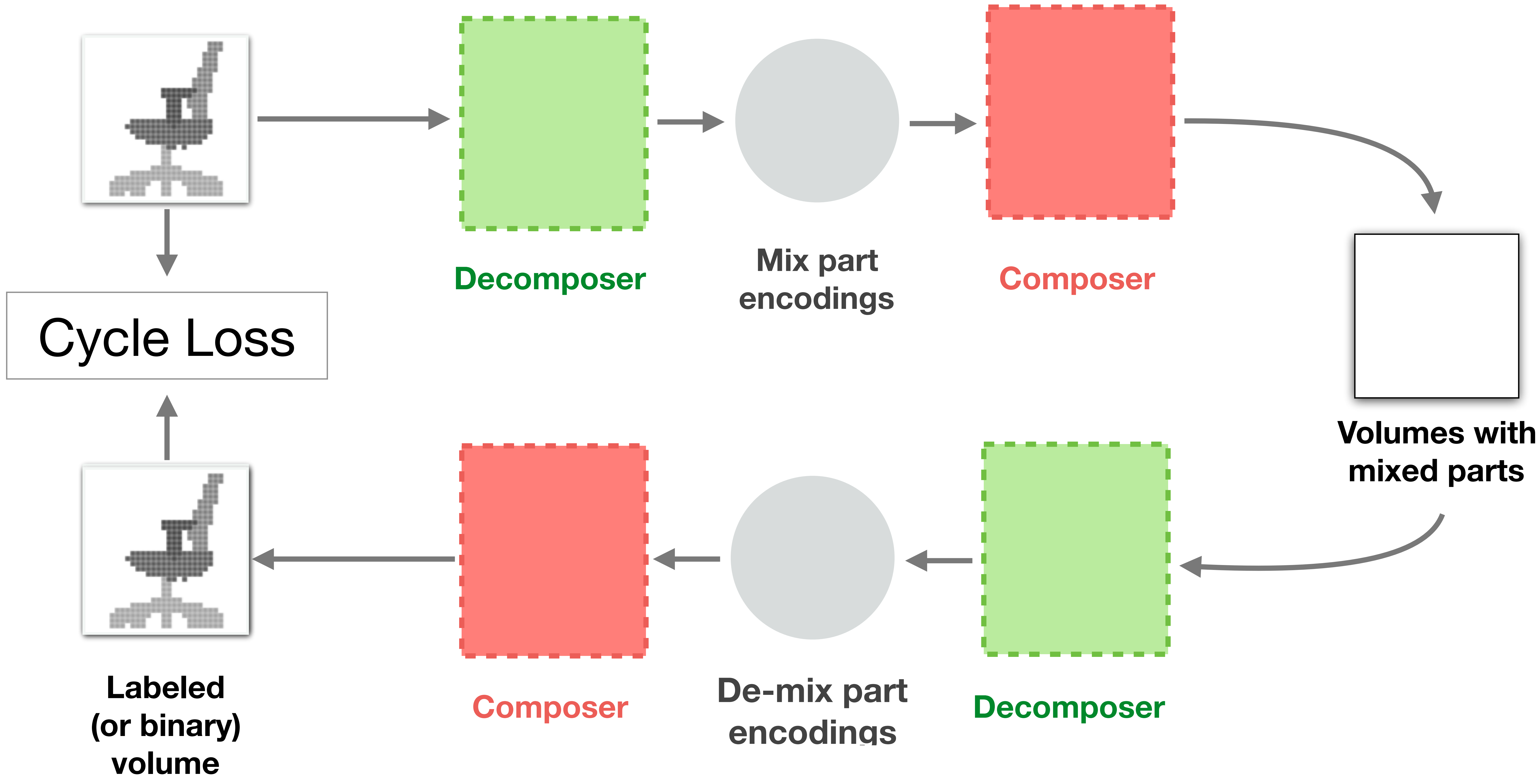
Training shapes provide supervision for shape reconstruction

Shape Decomposition and Composition



No supervision for shape composition from random parts

Cycle Consistency Constraint



- Different approach: GAN-type loss (constraints results to training shapes' manifold)

Loss Function

Per-part cross entropy reconstruction loss

Whole-model cross entropy reconstruction loss

$$L = w_{PI}\underline{L_{PI}} + w_{part}\underline{L_{part}} + w_{trans}\underline{L_{trans}} + w_{whole}\underline{L_{whole}} + w_{cycle}\underline{L_{cycle}}$$

"Partition of the identity"

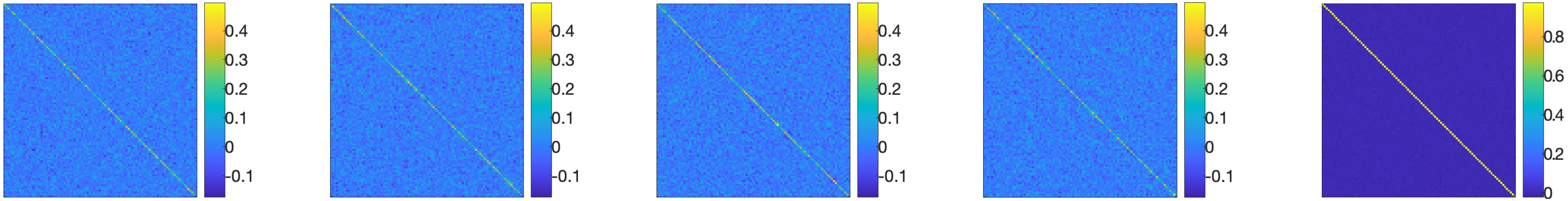
$$L_{PI}(P_1, \ldots, P_k) = \sum_{i=1}^{K} \|P_i^2 - P_i\|_F^2 + \sum_{\substack{i,j=1,\\ i \neq j}}^{K} \|P_iP_j\|_F^2 + \|P_1 + \ldots P_K - I\|_F^2$$

Cycle consistency loss

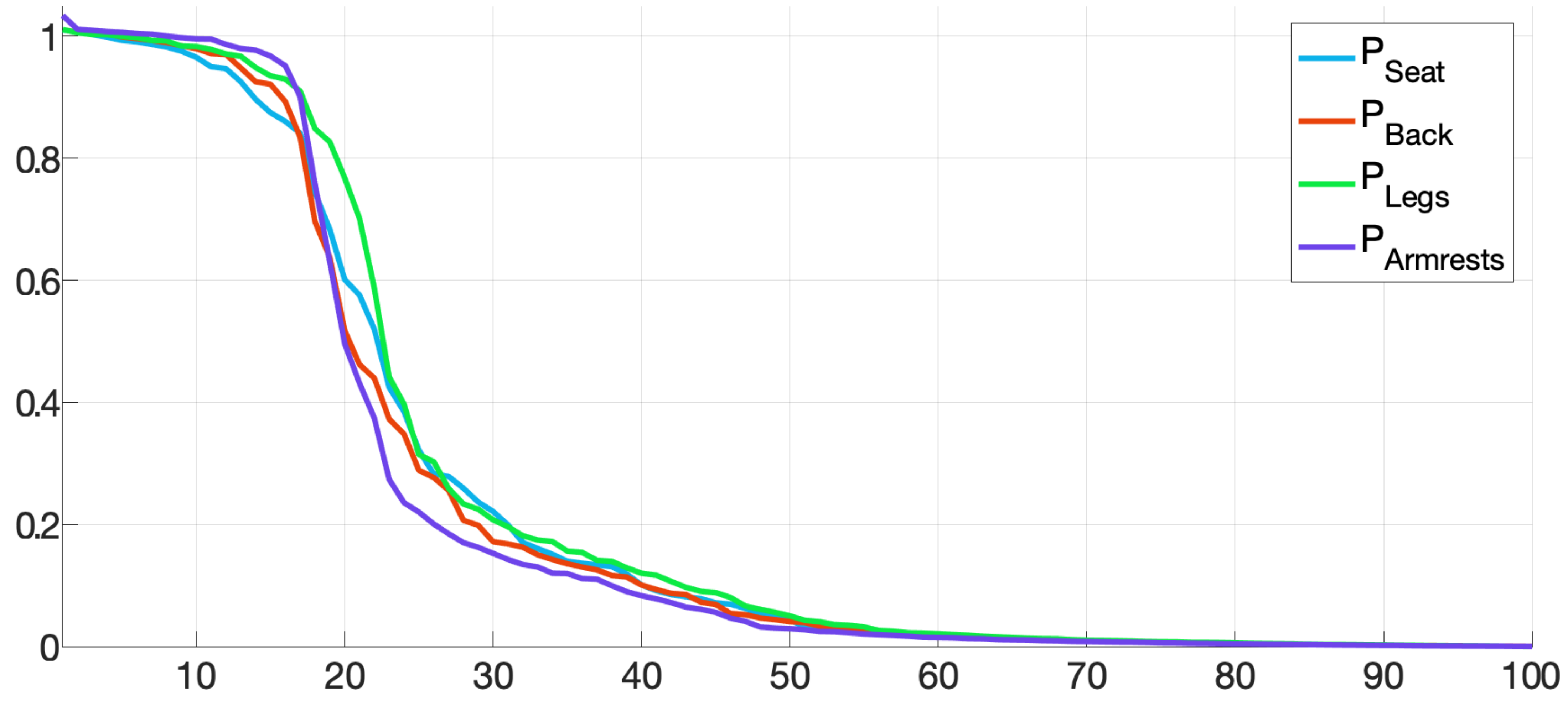
L_2 loss for 12-D affine transformation parameters

Evaluation

Projection matrices

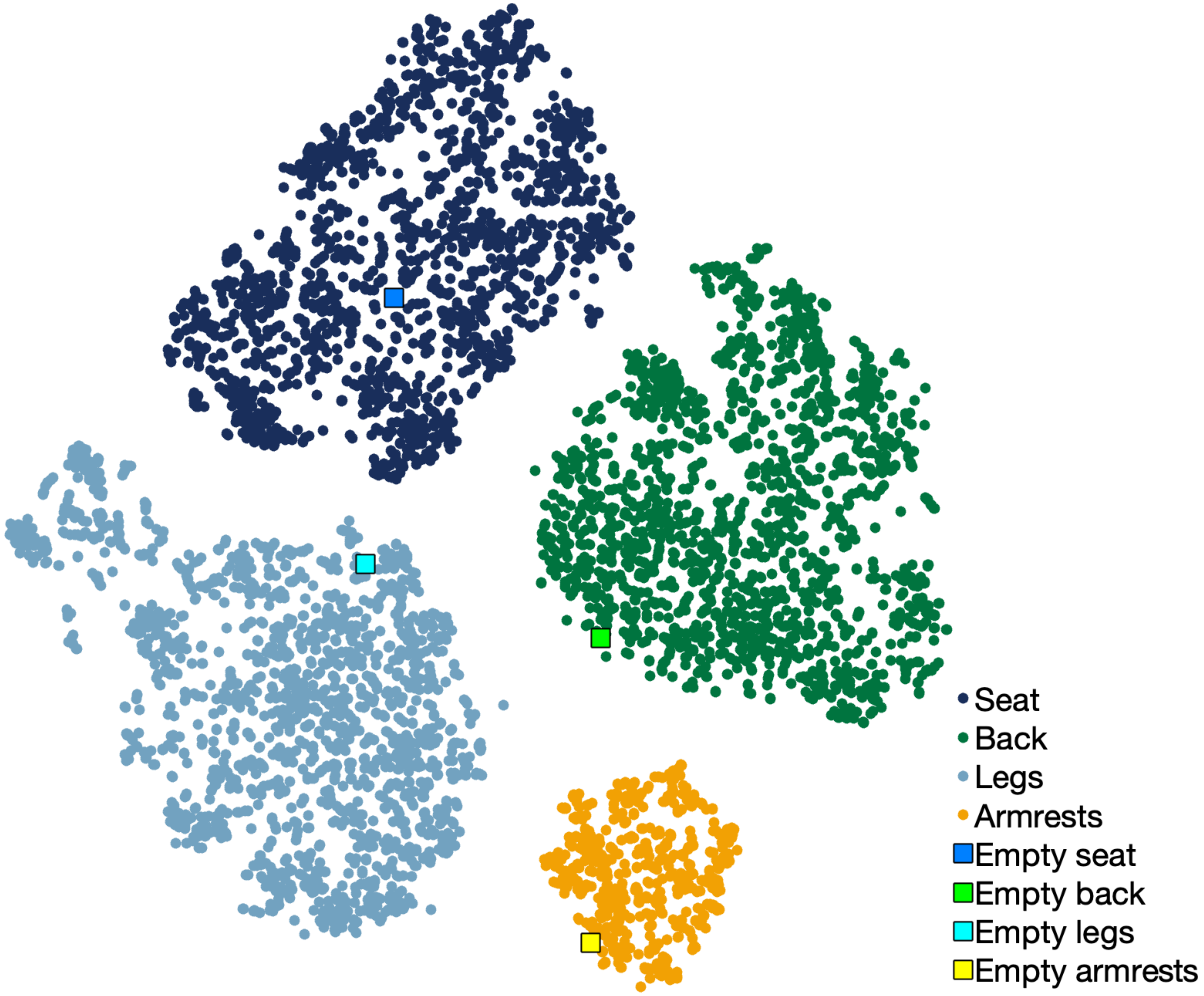


P_{Seat} P_{Back} P_{Legs} $P_{Armrests}$ $\sum P_i$

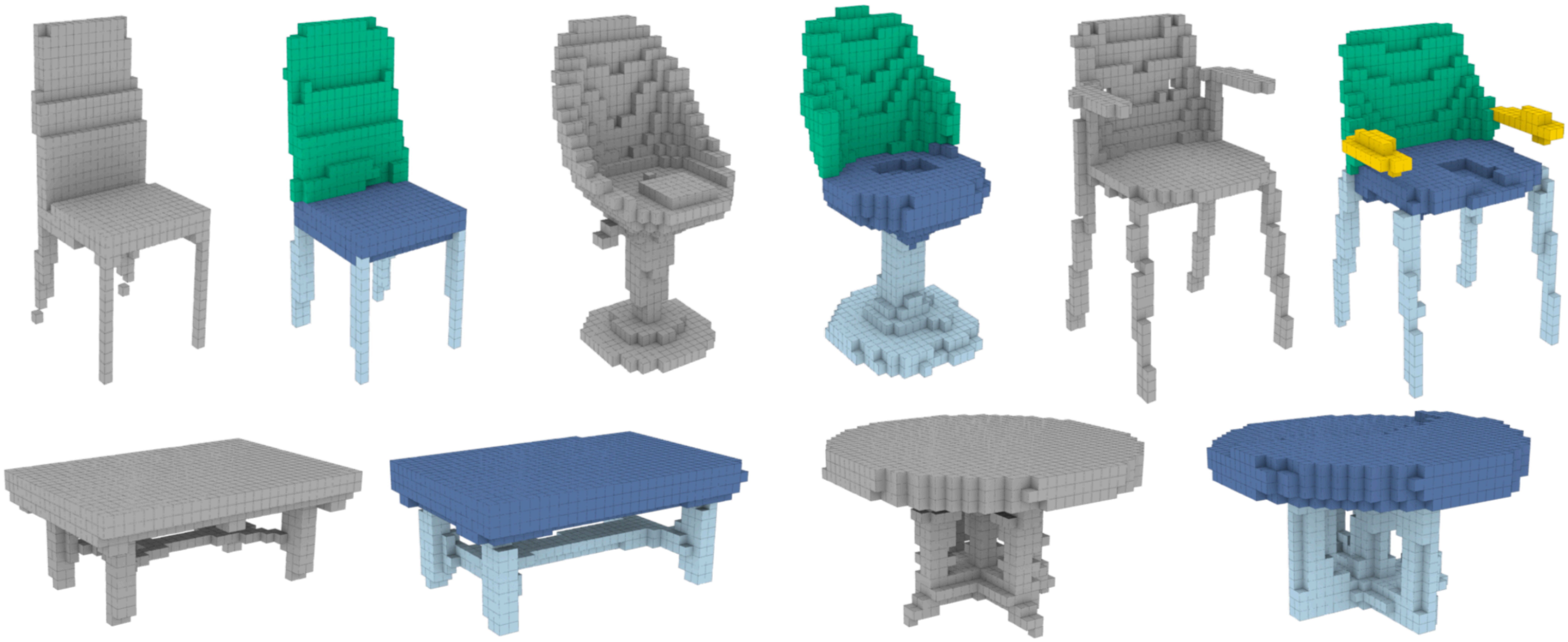


Projection matrix singular values

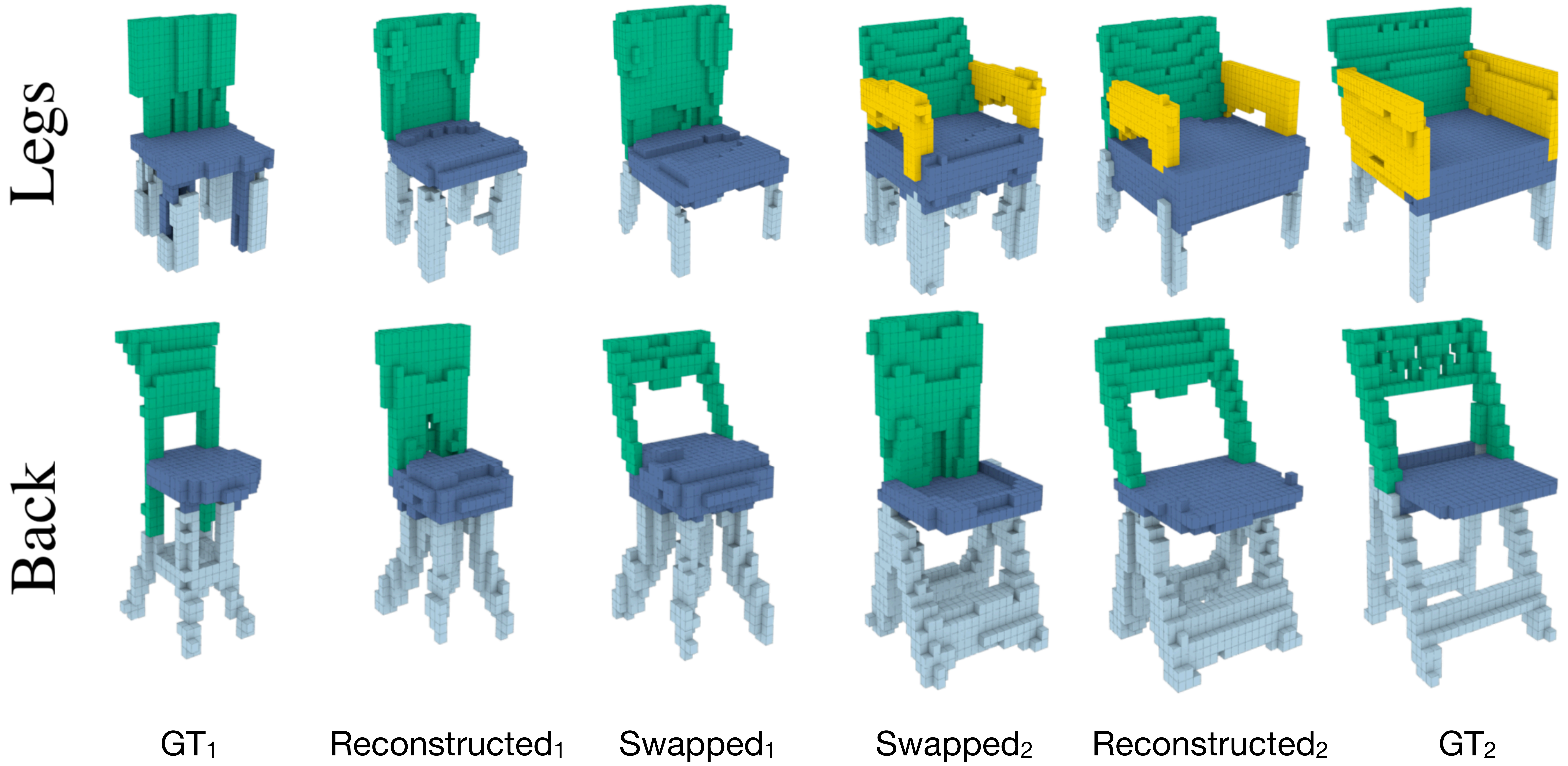
Latent space visualization (t-SNE)



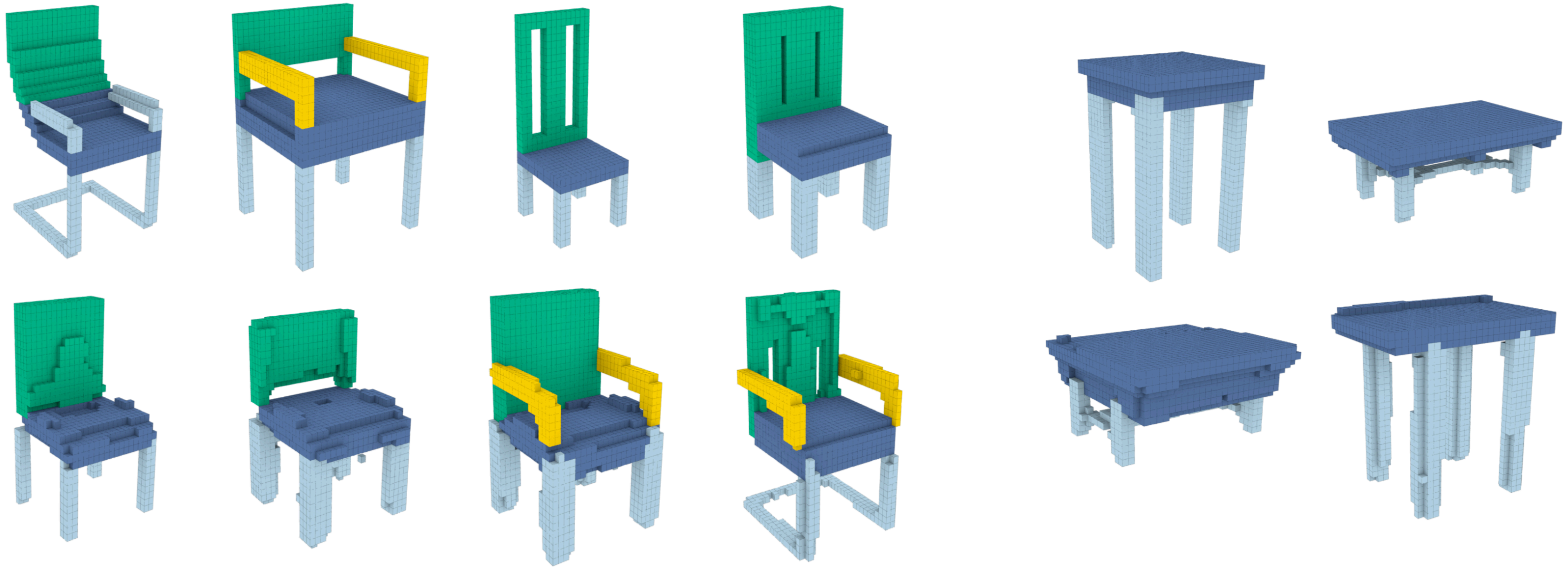
Shape reconstruction



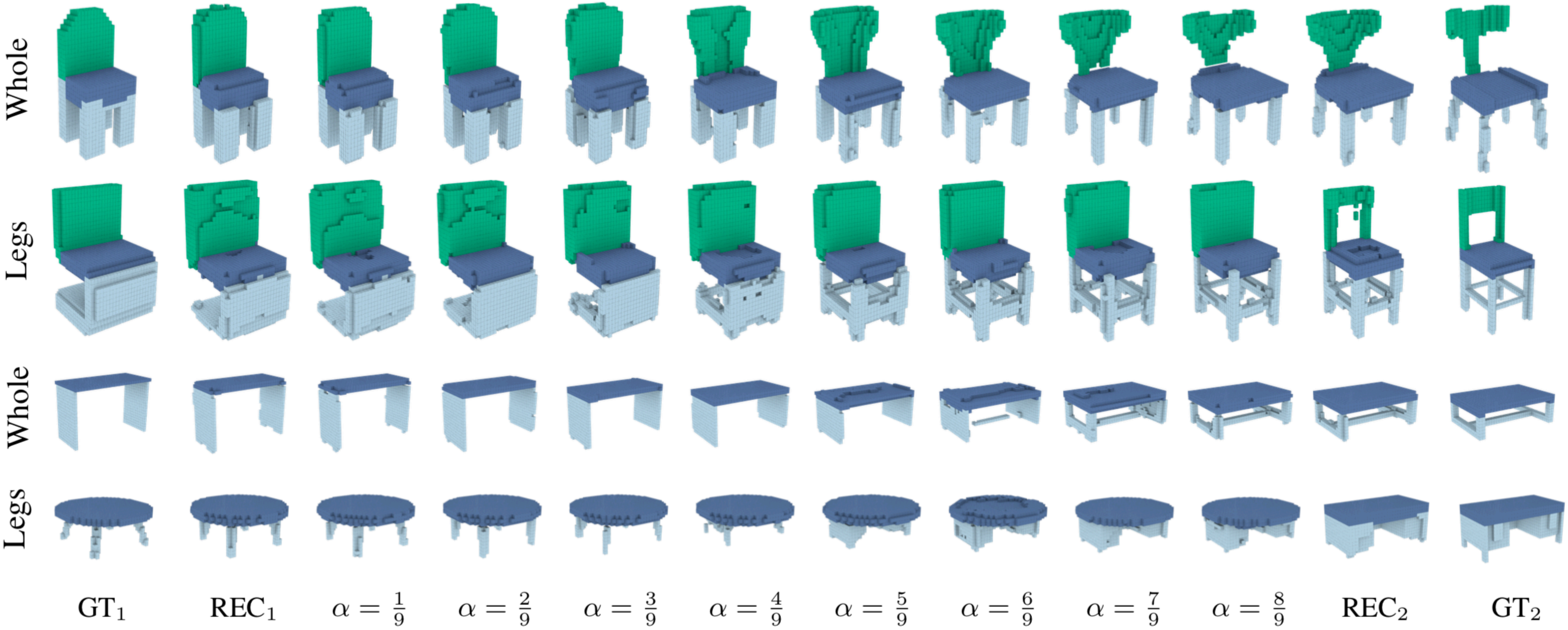
Part exchange



Random part composition



Full and Partial Shape Interpolation



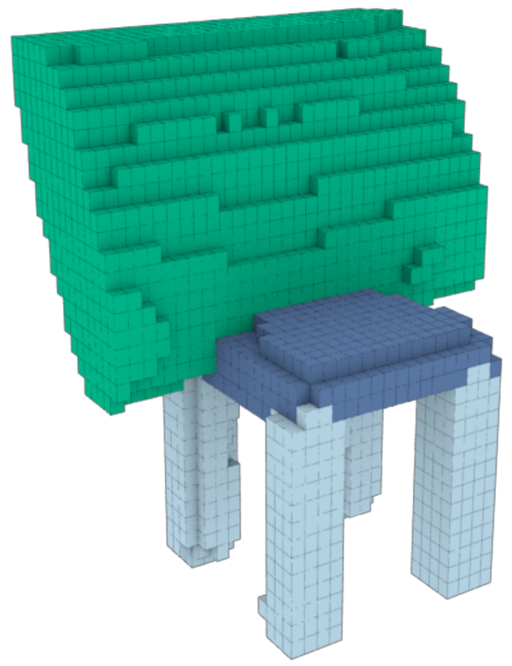
Ablation study

Method \ Metric	mIoU	mIoU (parts)	Connectivity			Classifier accuracy			Symmetry score		
	Rec.	Rec.	Rec.	Swap	Mix	Rec.	Swap	Mix	Rec.	Swap	Mix
Our method	0.64	0.65	**0.82**	**0.71**	**0.65**	**0.95**	0.89	**0.83**	0.95	0.95	0.95
W/o cycle loss	0.63	0.66	0.74	0.62	0.54	0.93	0.84	0.80	0.96	0.96	0.95
Fixed projection	0.63	0.65	0.72	0.61	0.58	0.94	0.86	0.77	0.94	0.95	0.95
Composer w/o STN	**0.75**	**0.8**	0.69	0.48	0.23	**0.95**	**0.9**	0.71	0.95	0.91	0.85
Naive placement	-	-	-	0.68	0.62	0.61	0.47	0.21	-	**0.96**	**0.96**
ComplementMe	-	-	-	**0.71**	0.47	-	0.66	0.43	-	0.66	0.43
Segmentation+STN	-	-	-	0.41	0.64	-	0.64	0.36	-	0.77	0.77

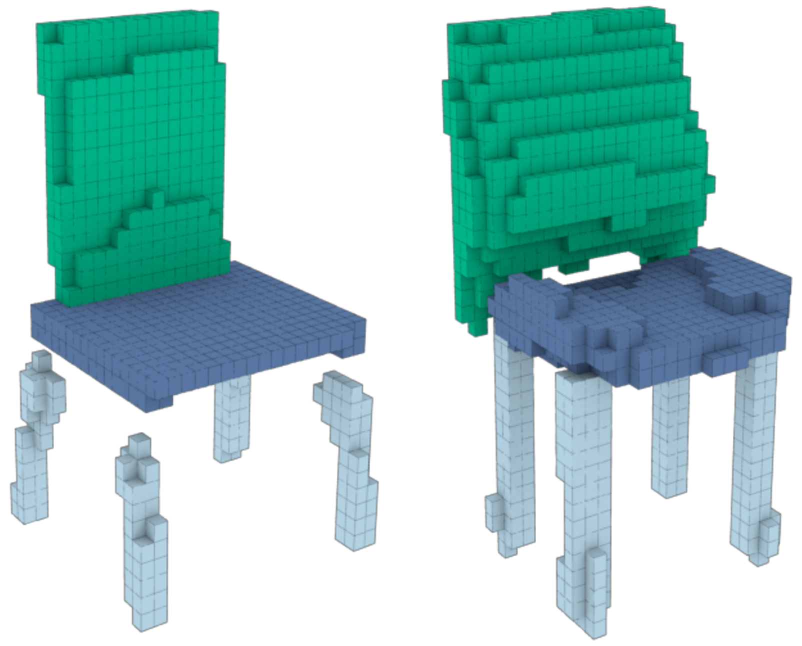
Evaluation metrics: mean Intersection over Union (*mIoU*), per-part mean IoU (*mIoU (parts)*), shape *connectivity measure*, binary *shape classifier accuracy*, and shape *symmetry score*.

Limitations and future work

- Low resolution
  - Use point cloud / memory-efficient voxel representation / mesh
- No measure of part compatibility
  - Model statistical dependence between parts

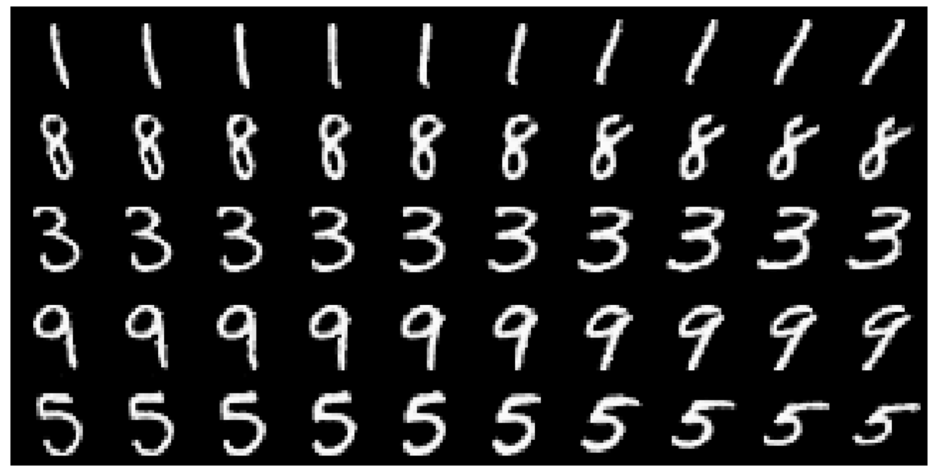


- Separate part reconstruction and placement may fail to produce plausible result
  - Part connectivity constrains / in-network assessment of composition quality / iterative placement



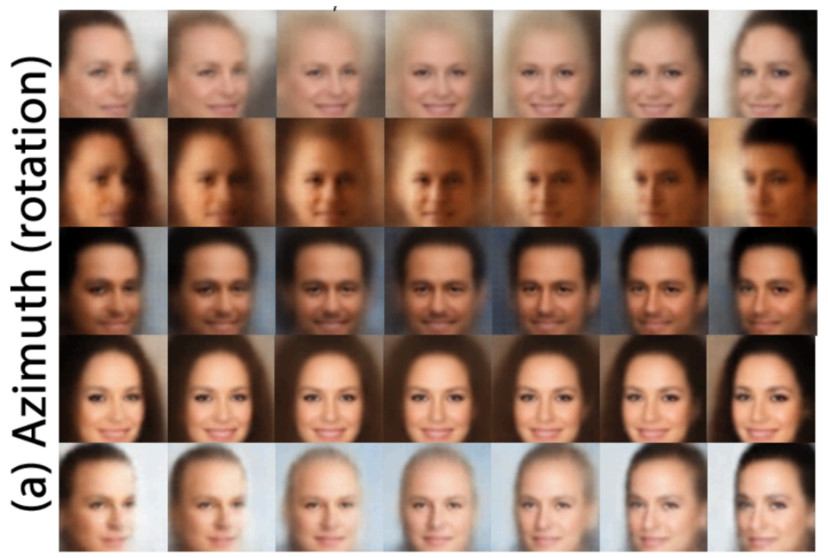
Future Directions

- Relation to other disentangled representations in machine learning
- Unsupervised learning of factors of variation

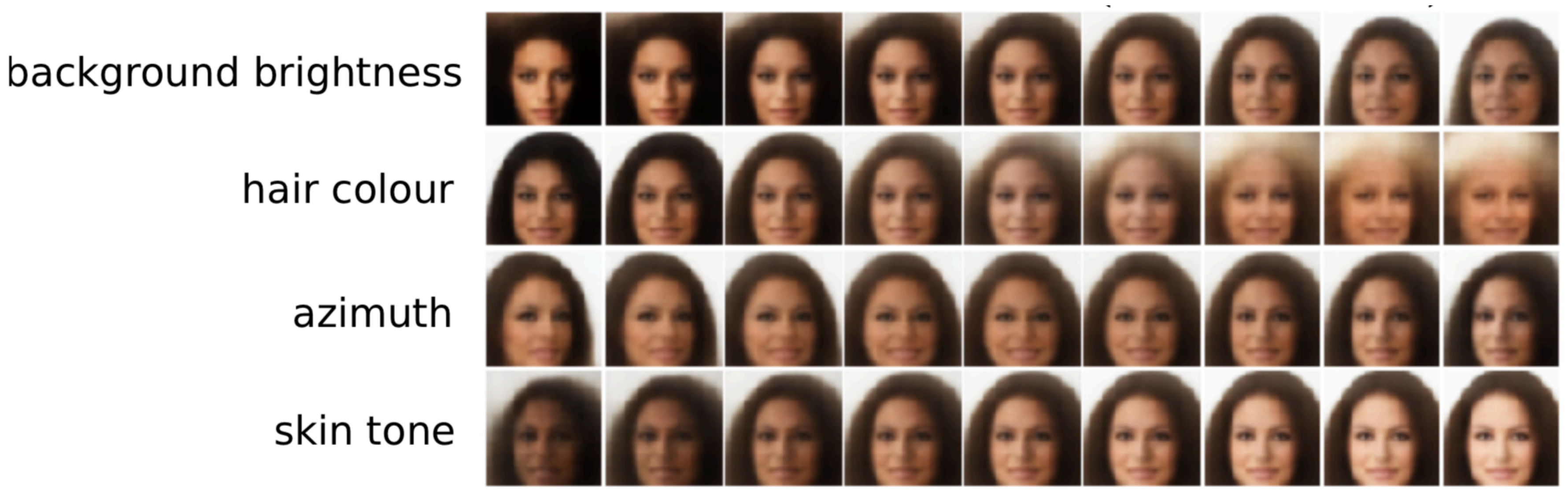


(c) Varying c_2 from -2 to 2 on InfoGAN (Rotation)

InfoGAN [Chen et al., NIPS 2016]



ß-VAE [Higgins et al., ICLR 2017]



Factor-VAE [Kim and Mnih., NIPS 2017 Workshop]

"Having a representation that is well suited to the particular task and data domain can significantly improve the learning success and robustness of the chosen model (Bengio et al., 2013)." [Higgins et al., ß-VAE]

THANK YOU

https://anastasiadk.github.io/