ECCV'20 ONLINE
23-28 AUGUST 2020

Stanford University [1]
Adobe Research [2]

Deformation-Aware 3D Model Embedding and Retrieval


Mikaela Uy [1]


Jingwei Huang [1]


Minhyuk Sung [2]


Tolga Birdal [1]


Leo Guibas [1]

Motivation

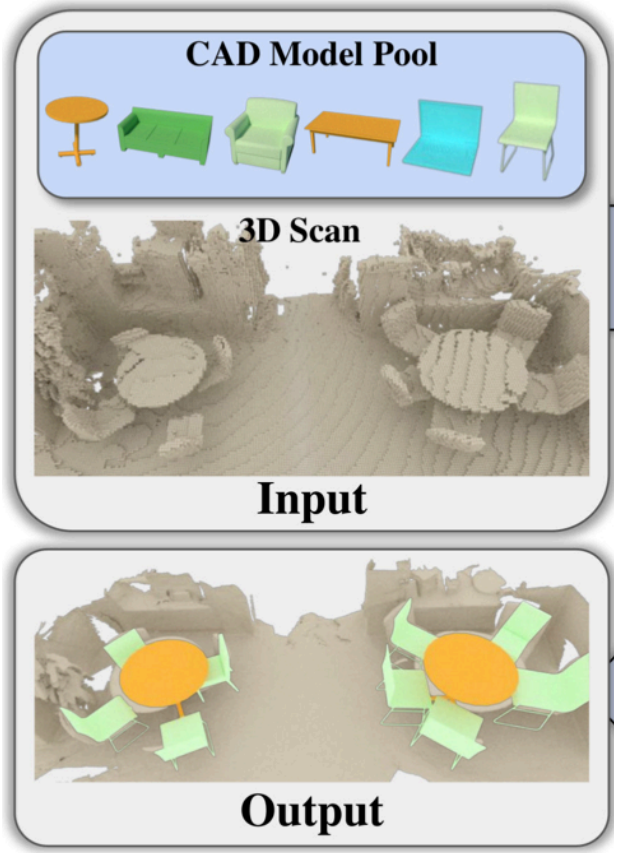


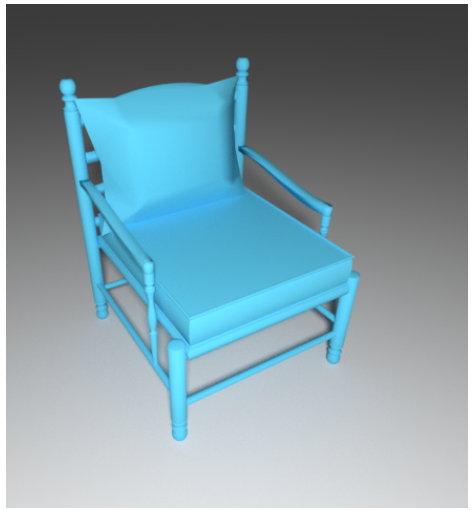
Photo taken from [1]

(a) Real Scan (b) CAD Model (c) Overlay

[1] End-to-End CAD Model Retrieval and 9DoF Alignment in 3D Scans. *Avetisyan et. al., ICCV 2019.*

Goal

Retrieve



Query Model

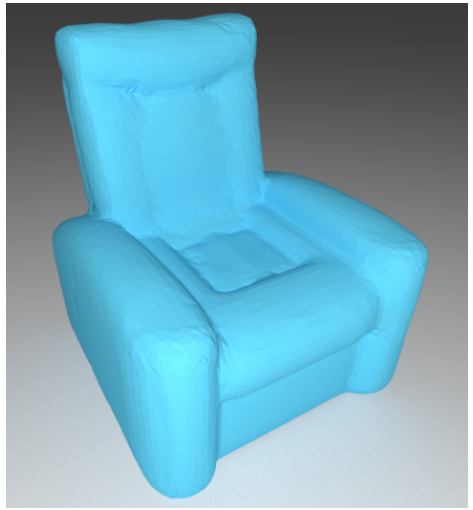
Closest Model

Chamfer Distance:
4.45×10^{-2}

Goal

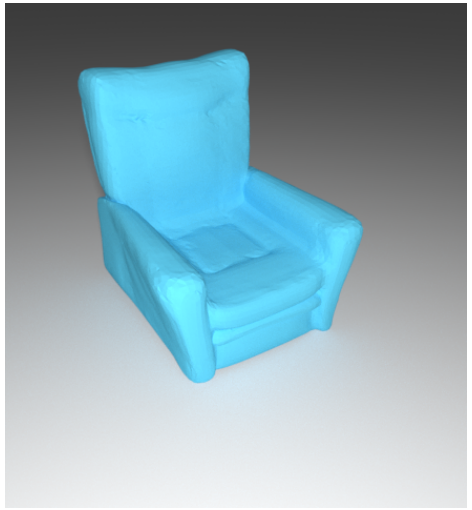
Query Model

Retrieve



Ours Retrieved

Deform



Ours Deformed

Chamfer Distance:
7.09×10^{-2} ↑

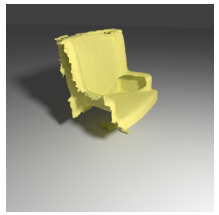
Chamfer Distance:
1.71×10^{-2} ↓

Problem

Input

3D Model

/



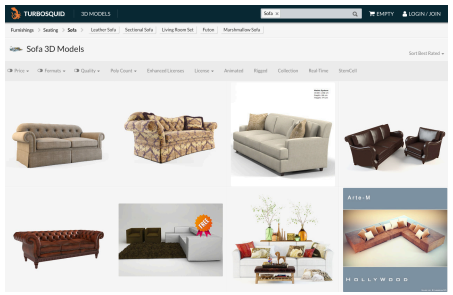
Scan

/

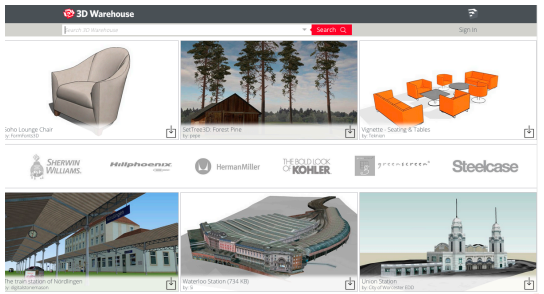
Image

3D Model Database

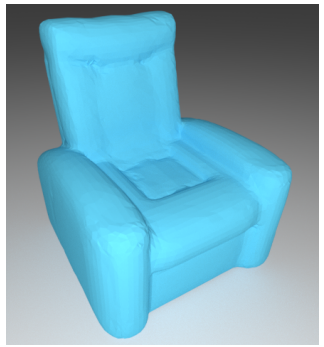
Retrieval



TurboSquid

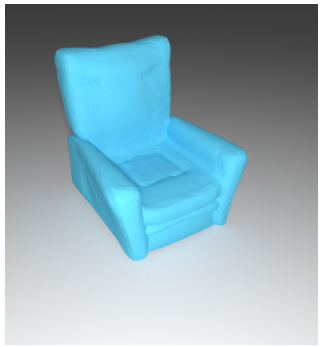


3D Warehouse



Retrieved Model

Deform

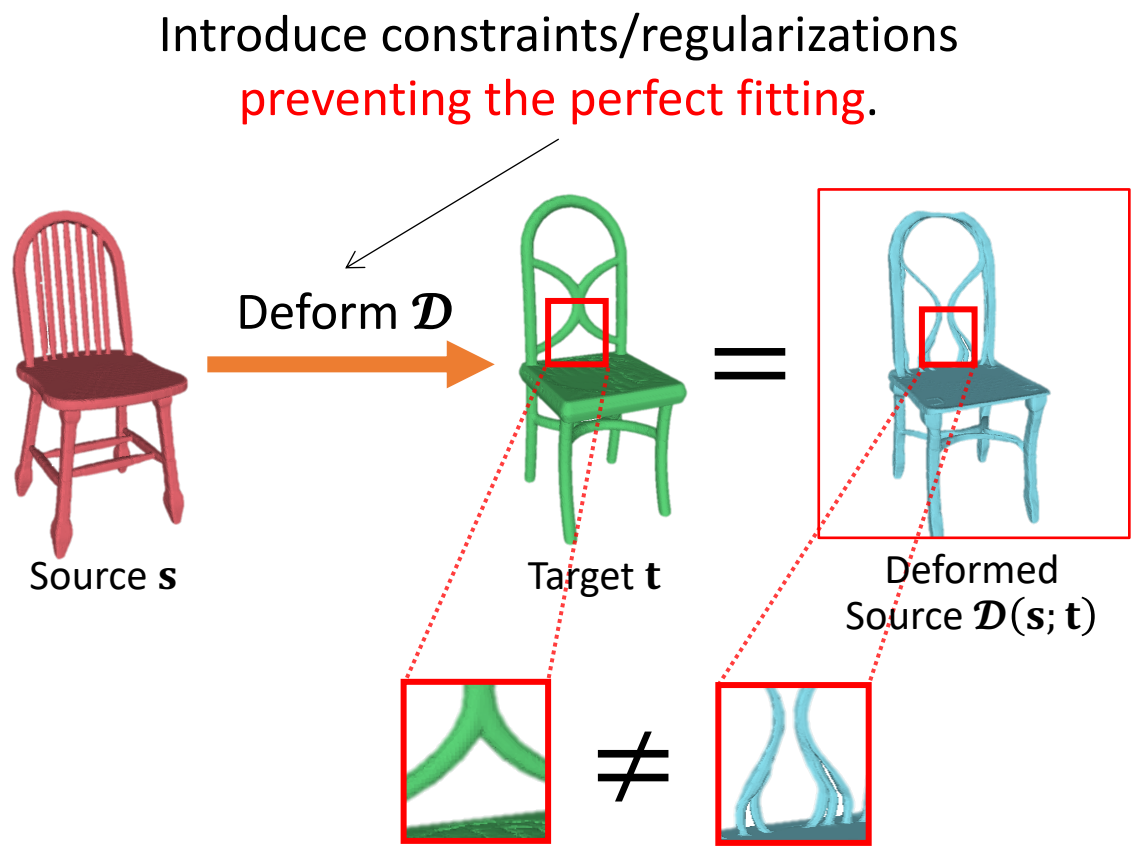


Deformed Model

Fitting Gap

- Deformations introduce **constraints/regularizations** that ensure plausible variations without losing the original CAD model features.

- ***Fitting gap*** $e_{\mathcal{D}}(\mathbf{s}, \mathbf{t}) = \boldsymbol{d}(\boldsymbol{\mathcal{D}}(\mathbf{s}; \mathbf{t}), \mathbf{t})$***:*** Fitting distance (d) ***after*** deforming a database shape ($\mathbf{s}$) to the query ($\mathbf{t}$) using deformation function $\mathcal{D}$.



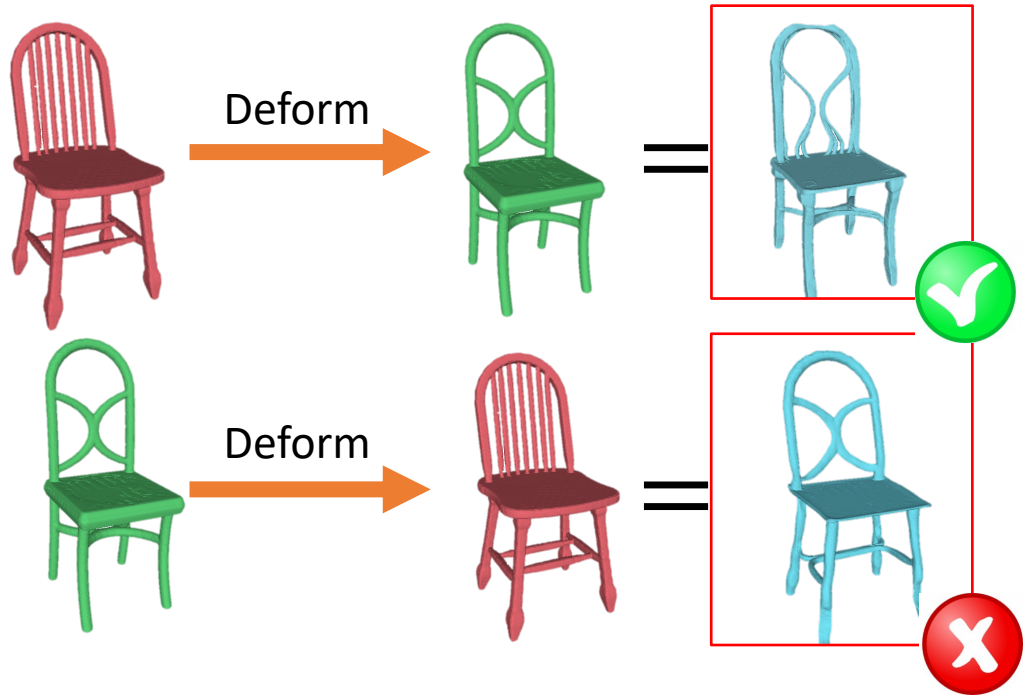
Properties of Fitting Gap

- Fitting gap measures the distance in the real space.

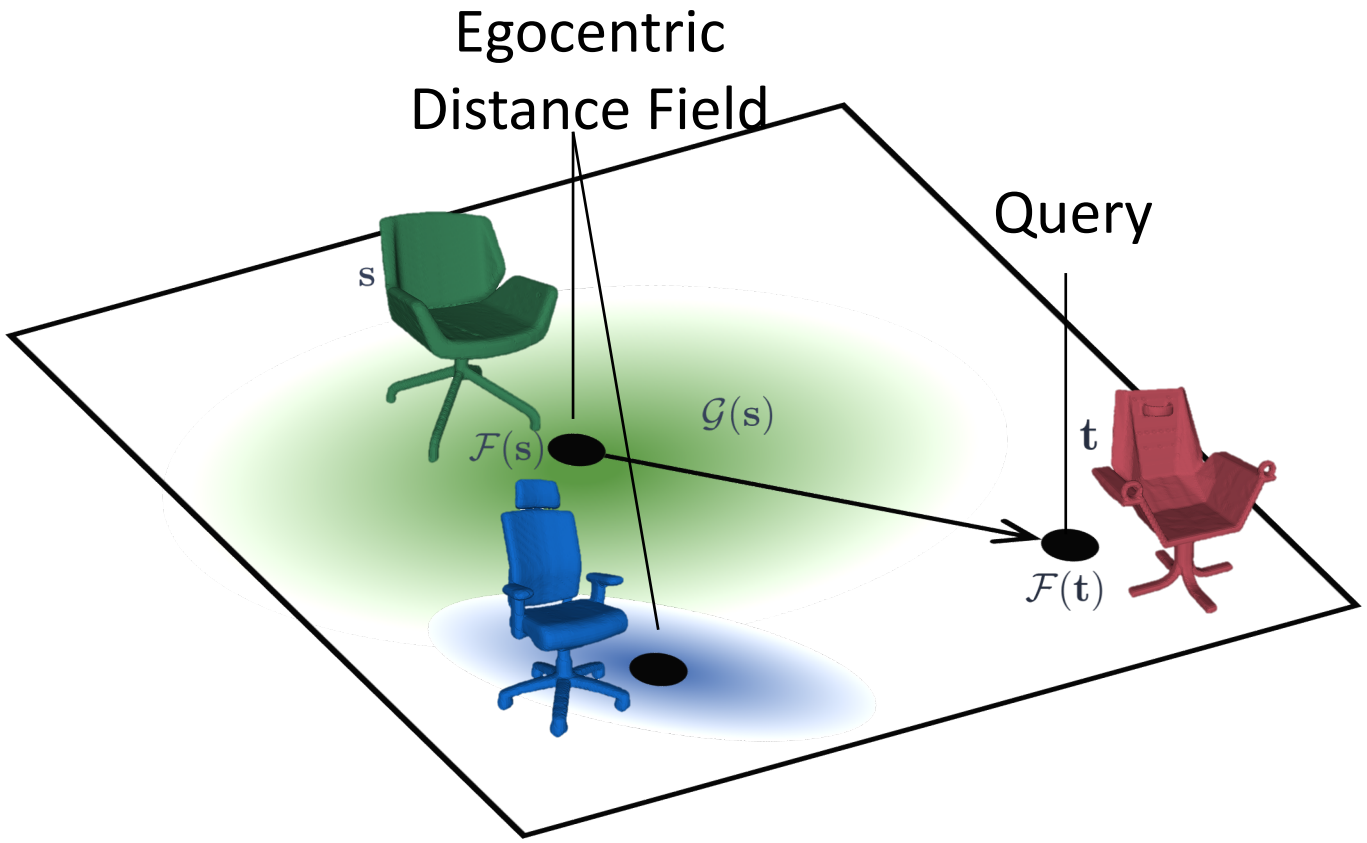
- Properties of fitting gap:

1. (Non-negativity) $e_{\mathcal{D}}(\mathbf{s}, \mathbf{t}) \geq 0$
2. (Identity) $e_{\mathcal{D}}(\mathbf{t}, \mathbf{t}) = 0$
3. *(**Asymmetry**)* $e_{\mathcal{D}}(\mathbf{s}, \mathbf{t}) \neq e_{\mathcal{D}}(\mathbf{t}, \mathbf{s})$

Not a metric!



Egocentric Distance Field $\mathcal{G}(\mathbf{s})$

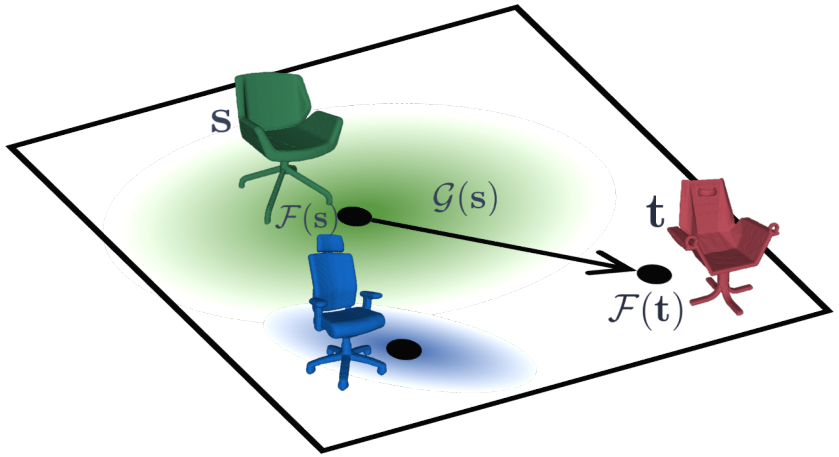


- $\mathcal{G}$ is source-dependent.
- $\mathcal{G}$ is represented with a positive semi-definite matrix.

$$\mathcal{G}(\mathbf{s}) \in \mathcal{S}^k \;\; (\mathcal{G}(\mathbf{s}) \succcurlyeq 0)$$

Distance in Embedding Space δ

$$\delta(\mathbf{s};\mathbf{t}) = \sqrt{(\mathcal{F}(\mathbf{t}) - \mathcal{F}(\mathbf{s}))^T \mathcal{G}(\mathbf{s})(\mathcal{F}(\mathbf{t}) - \mathcal{F}(\mathbf{s}))}$$



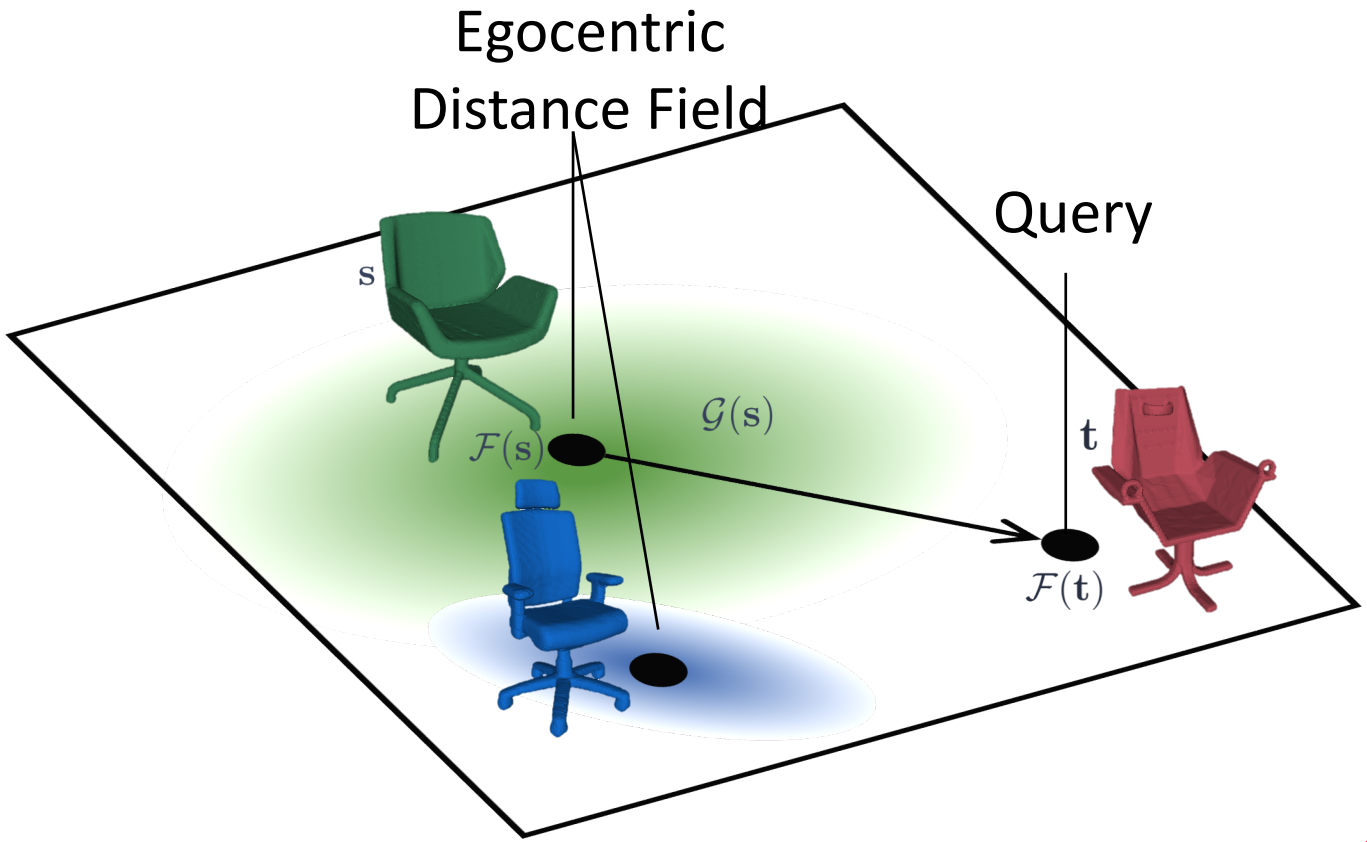
Properties

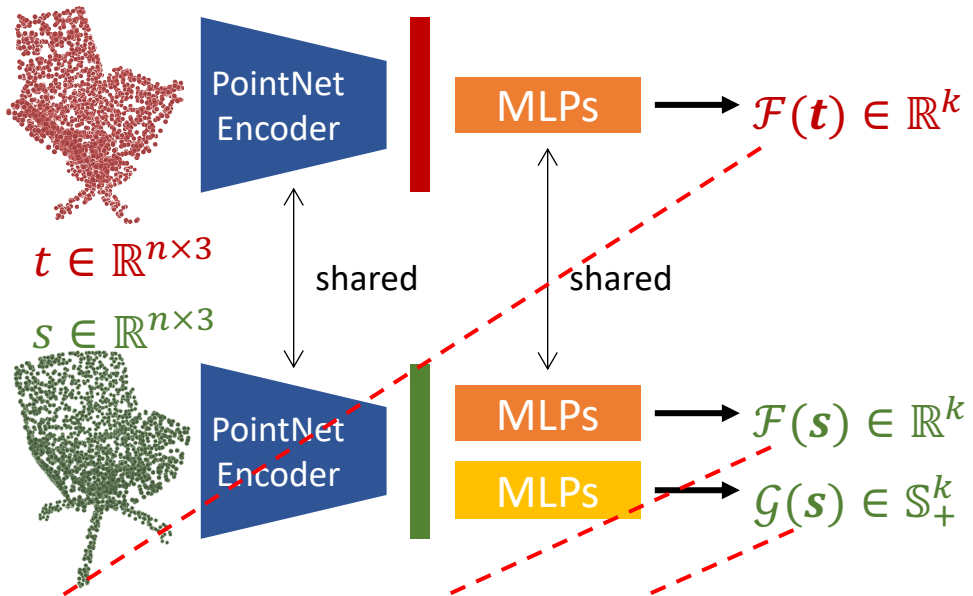
1. (Non-negativity) $\delta(\mathbf{s},\mathbf{t}) \geq 0$
2. (Identity) $\delta(\mathbf{t},\mathbf{t}) = 0$
3. *(**Asymmetry**)*

$$\delta(\mathbf{s},\mathbf{t}) = \sqrt{(\mathcal{F}(\mathbf{t}) - \mathcal{F}(\mathbf{s}))^T \mathcal{G}(\mathbf{s})(\mathcal{F}(\mathbf{t}) - \mathcal{F}(\mathbf{s}))}$$
$$\delta(\mathbf{t},\mathbf{s}) = \sqrt{(\mathcal{F}(\mathbf{s}) - \mathcal{F}(\mathbf{t}))^T \mathcal{G}(\mathbf{t})(\mathcal{F}(\mathbf{s}) - \mathcal{F}(\mathbf{t}))}$$

$$\delta(\mathbf{s},\mathbf{t}) \neq \delta(\mathbf{t},\mathbf{s})$$

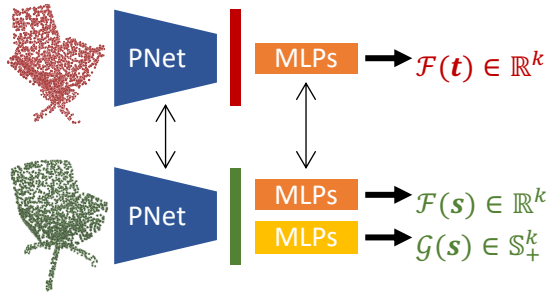
Deformation-Aware Embedding





$$e_{\mathcal{D}}(\mathbf{s},\mathbf{t}) \sim \delta(\mathbf{s};\mathbf{t}) = \sqrt{(\mathcal{F}(\mathbf{t})-\mathcal{F}(\mathbf{s}))^T \mathcal{G}(\mathbf{s})(\mathcal{F}(\mathbf{t})-\mathcal{F}(\mathbf{s}))}$$

Network Training

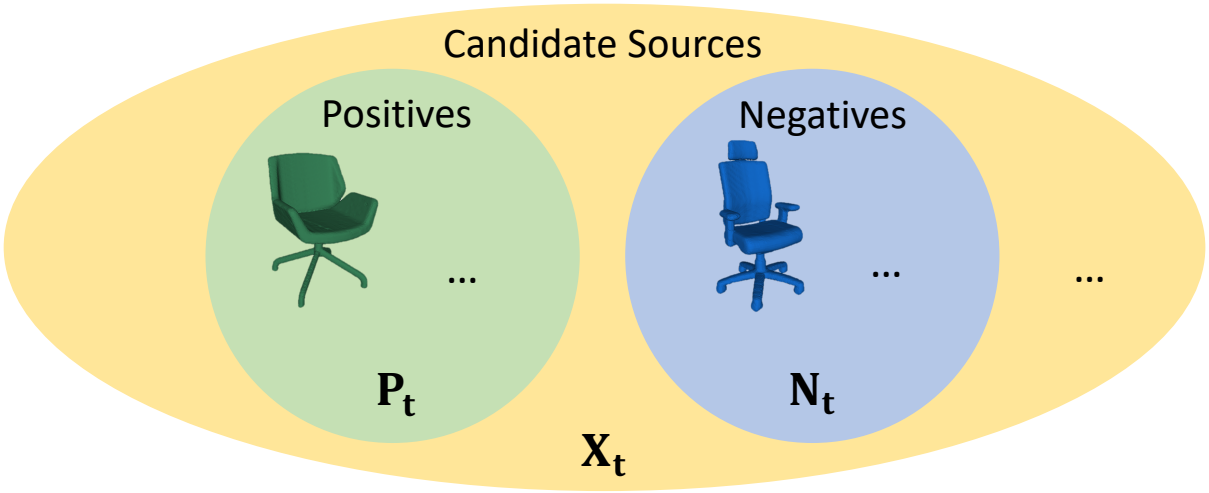


$$\delta(\mathbf{s};\mathbf{t}) = \sqrt{(\mathcal{F}(\mathbf{t}) - \mathcal{F}(\mathbf{s}))^T \mathcal{G}(\mathbf{s})(\mathcal{F}(\mathbf{t}) - \mathcal{F}(\mathbf{s}))}$$

We precompute the fitting gap ($e_{\mathcal{D}}$).

Target

$\mathbf{t}$



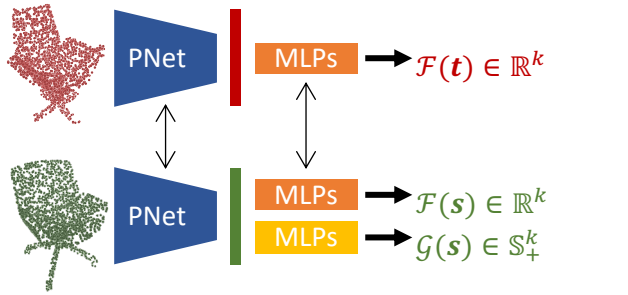
- **Margin-loss-based approach**

$$\mathrm{P_t} = \{\mathbf{s} \in \mathbf{X_t} | e_{\mathcal{D}}(\mathbf{s},\mathbf{t}) \leq \sigma_P\} \qquad \mathrm{N_t} = \{\mathbf{s} \in \mathbf{X_t} | e_{\mathcal{D}}(\mathbf{s},\mathbf{t}) > \sigma_N\}$$

$$\sum_{\mathbf{n} \in \mathrm{N_t}} [\max_{\mathbf{p} \in \mathrm{P_t}} (\delta(\mathbf{p};\mathbf{t}) - \delta(\mathbf{n};\mathbf{t})) + m]_+$$

[2] FaceNet: A Unified Embedding for Face Recognition and Clustering. Schroff et. al., CVPR 2015

Network Training

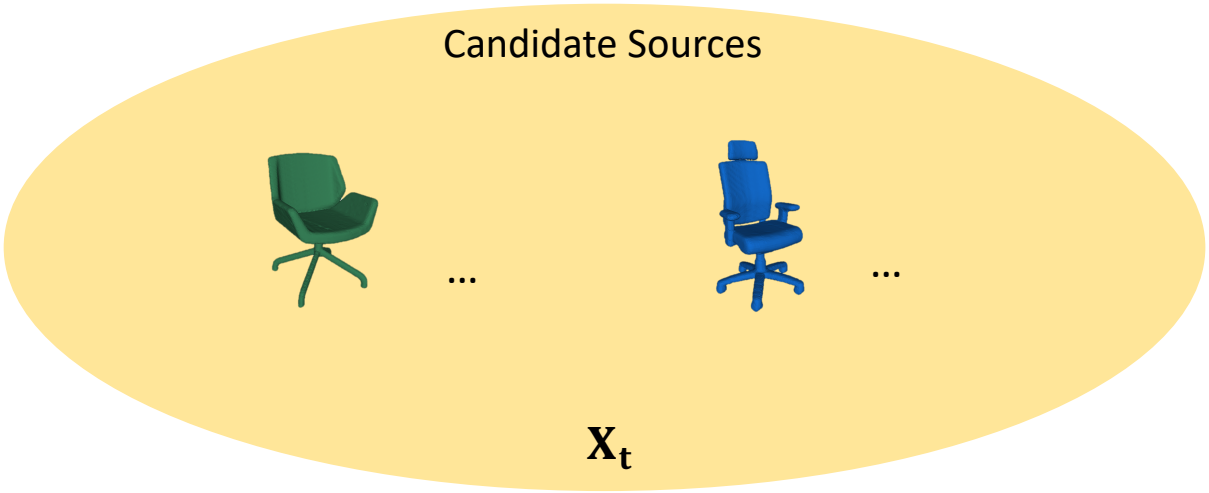


$$\delta(\mathbf{s};\mathbf{t}) = \sqrt{(\mathcal{F}(\mathbf{t}) - \mathcal{F}(\mathbf{s}))^T \mathcal{G}(\mathbf{s})(\mathcal{F}(\mathbf{t}) - \mathcal{F}(\mathbf{s}))}$$

We precompute the fitting gap ($e_{\mathcal{D}}$).

Target

$\mathbf{t}$



- **Regression-based approach**

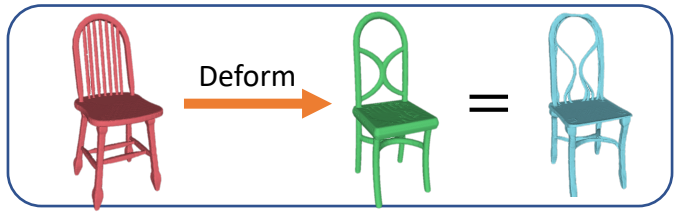
$$p(\mathbf{s};\mathbf{t}) = \frac{\exp(-e_{\mathcal{D}}(\mathbf{s};\mathbf{t})/2\sigma_t^2)}{\sum_{\mathbf{s}' \in \mathbf{X}'_\mathbf{t}} \exp(-e_{\mathcal{D}}(\mathbf{s}';\mathbf{t})/2\sigma_t^2)} \qquad \hat{p}(\mathbf{s};\mathbf{t}) = \frac{\delta^2(\mathbf{s};\mathbf{t})}{\sum_{\mathbf{s}' \in \mathbf{X}'_\mathbf{t}} \delta^2(\mathbf{s};\mathbf{t})}$$

$$\frac{1}{|\mathrm{X}'_\mathbf{t}|} \sum_{\mathbf{s} \in \mathrm{X}'_\mathbf{t}} |\hat{p}(\mathbf{s};\mathbf{t}) - p(\mathbf{s};\mathbf{t})|$$

[3] Stochastic Neighbor Embedding. *Hinton et. al., NeurIPS 2002.*

Summary

1. Fitting gap



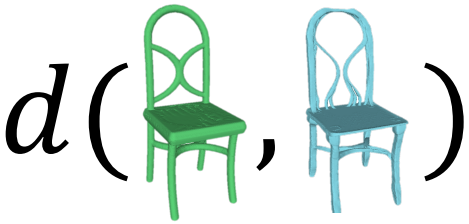
2. Egocentric distance field $\mathcal{G}(\mathbf{s})$

3. Training approaches:
   - Margin-loss-based
   - Regression-based

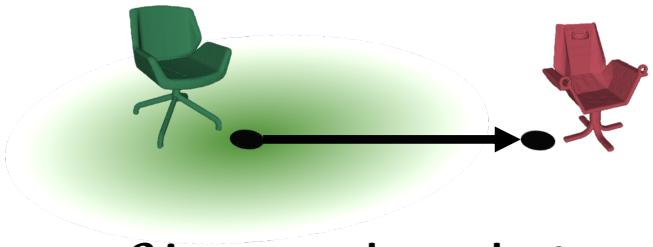
Chamfer Distance
Before Deformation
$\boldsymbol{d}(\mathbf{s}, \mathbf{t})$

$\mathcal{G}$ is fixed to identity.
Symmetric embedding distance

OURS



Chamfer Distance
After Deformation
$e_{\mathcal{D}}(\mathbf{s}, \mathbf{t}) = \boldsymbol{d}(\boldsymbol{\mathcal{D}}(\mathbf{s}; \mathbf{t}), \mathbf{t})$



$\mathcal{G}$ is source-dependent.
Asymmetric embedding distance

Implementation Details

- Training data: ShapeNet (5 categories)
- Backbone architecture: PointNet (sampling points over the meshes)
- Deformation function $\mathcal{D}$: Simplified as-rigid-as-possible (ARAP)

[4] ShapeNet: An Information-Rich 3D Model Repository. Chang *et. al.*, arXiv 2015.
[5] PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *Qi et. al., CVPR 2017.*
[6] As-rigid-as-possible surface modeling. *Sorkine et. al., SGP 2007.*

Quantitative Results

(Mean Chamfer Distance $\times 10^{-2}$ for the best of the top 3 retrieval)
bold = smallest, underline = second smallest

	Before Deformation (B.D.) $\boldsymbol{d(\mathbf{s},\mathbf{t})}$	**After Deformation (A.D.)** $\boldsymbol{e_{\mathcal{D}}(\mathbf{s},\mathbf{t}) = d(\mathcal{D}(\mathbf{s};\mathbf{t}),\mathbf{t})}$

Quantitative Results

Ranked by Chamfer Distance (**Ranked CD**)

- Select the shape with smallest B.D.
- No embedding space.

(Mean Chamfer Distance $\times 10^{-2}$ for the best of the top 3 retrieval)
bold = smallest, underline = second smallest

	Before Deformation (B.D.) $\boldsymbol{d}(\mathbf{s}, \mathbf{t})$	**After Deformation (A.D.)** $e_{\mathcal{D}}(\mathbf{s}, \mathbf{t}) = \boldsymbol{d}(\mathcal{D}(\mathbf{s}; \mathbf{t}), \mathbf{t})$
Ranked CD	**3.025**	**1.104**

Quantitative Results

(Mean Chamfer Distance $\times 10^{-2}$ for the best of the top 3 retrieval)
bold = smallest, underline = second smallest

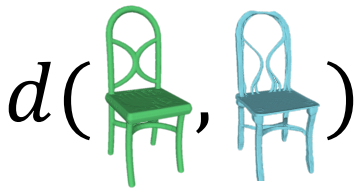
	Before Deformation (B.D.) $\boldsymbol{d(\mathbf{s}, \mathbf{t})}$	**After Deformation (A.D.)** $\boldsymbol{e_{\mathcal{D}}(\mathbf{s}, \mathbf{t}) = d(\mathcal{D}(\mathbf{s}; \mathbf{t}), \mathbf{t})}$
Ranked CD	**3.025**	**1.104**
AE	<u>3.188</u>	<u>1.116</u>

Autoencoder (**AE**)

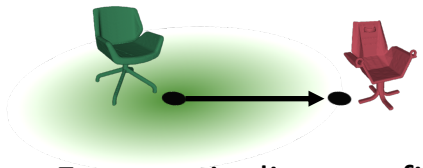
- PointNet autoencoder for reconstruction.
- Use the bottleneck layer as the embedding space.

Quantitative Results

CD-Margin



Fitting gap



Egocentric distance field

Margin-loss is used.

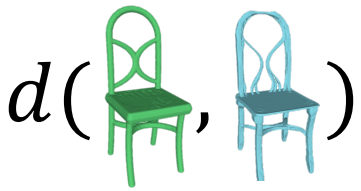
(Mean Chamfer Distance $\times 10^{-2}$ for the best of the top 3 retrieval)
bold = smallest, underline = second smallest

	Before Deformation (B.D.) $\boldsymbol{d(\mathbf{s},\mathbf{t})}$	**After Deformation (A.D.)** $\boldsymbol{e_{\mathcal{D}}(\mathbf{s},\mathbf{t}) = d(\mathcal{D}(\mathbf{s};\mathbf{t}),\mathbf{t})}$
Ranked CD	**3.025**	**1.104**
AE	3.188	1.116
CD-Margin	3.321	1.168
CD-Reg	5.057	2.108

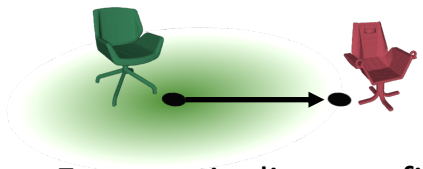
(PointNet encoder is used for the embedding space.)

Quantitative Results

CD-Reg



Fitting gap



Egocentric distance field

Reg-loss is used.

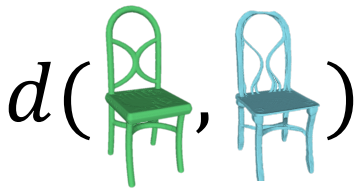
(Mean Chamfer Distance $\times 10^{-2}$ for the best of the top 3 retrieval)
bold = smallest, underline = second smallest

	Before Deformation (B.D.) $\boldsymbol{d(\mathbf{s},\mathbf{t})}$	**After Deformation (A.D.)** $\boldsymbol{e_{\mathcal{D}}(\mathbf{s},\mathbf{t}) = d(\mathcal{D}(\mathbf{s};\mathbf{t}),\mathbf{t})}$
Ranked CD	**3.025**	**1.104**
AE	3.188	1.116
CD-Margin	3.321	1.168
CD-Reg	5.057	2.108

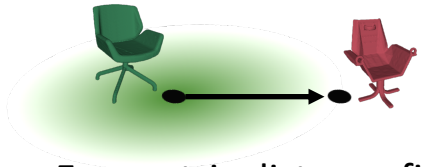
(PointNet encoder is used for the embedding space.)

Quantitative Results

Symm-Margin



Fitting gap



Egocentric distance field

Margin-loss is used.

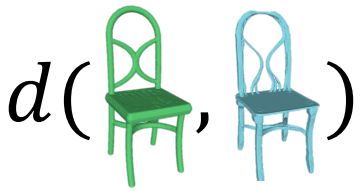
(Mean Chamfer Distance $\times 10^{-2}$ for the best of the top 3 retrieval)
bold = smallest, underline = second smallest

	Before Deformation (B.D.) $\boldsymbol{d(\mathbf{s},\mathbf{t})}$	**After Deformation (A.D.)** $\boldsymbol{e_{\mathcal{D}}(\mathbf{s},\mathbf{t}) = d(\mathcal{D}(\mathbf{s};\mathbf{t}),\mathbf{t})}$
Ranked CD	**3.025**	1.104
AE	3.188	1.116
CD-Margin	3.321	1.168
CD-Reg	5.057	2.108
Symm-Margin	3.537	**1.092**
Symm-Reg	4.649	1.657

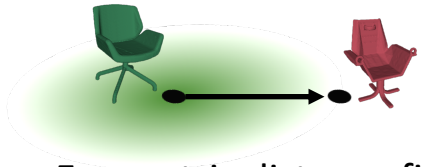
(PointNet encoder is used for the embedding space.)

Quantitative Results

Symm-Reg



Fitting gap



Egocentric distance field

Reg-loss is used.

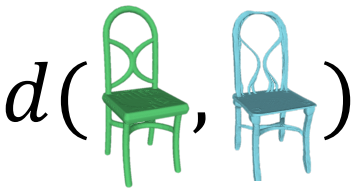
(Mean Chamfer Distance $\times 10^{-2}$ for the best of the top 3 retrieval)
bold = smallest, underline = second smallest

	Before Deformation (B.D.) $\boldsymbol{d(\mathbf{s},\mathbf{t})}$	**After Deformation (A.D.)** $\boldsymbol{e_{\mathcal{D}}(\mathbf{s},\mathbf{t}) = d(\mathcal{D}(\mathbf{s};\mathbf{t}),\mathbf{t})}$
Ranked CD	**3.025**	1.104
AE	3.188	1.116
CD-Margin	3.321	1.168
CD-Reg	5.057	2.108
Symm-Margin	3.537	**1.092**
Symm-Reg	4.649	1.657

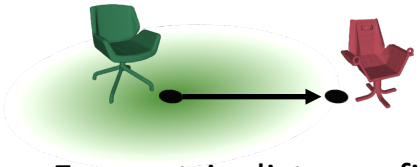
(PointNet encoder is used for the embedding space.)

Quantitative Results

Ours-Margin



Fitting gap



Egocentric distance field

Margin-loss is used.

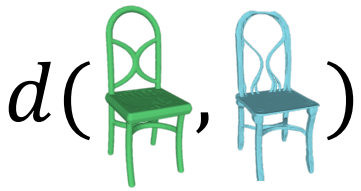
(Mean Chamfer Distance $\times 10^{-2}$ for the best of the top 3 retrieval)
bold = smallest, underline = second smallest

	Before Deformation (B.D.) $d(\mathbf{s}, \mathbf{t})$	After Deformation (A.D.) $e_{\mathcal{D}}(\mathbf{s}, \mathbf{t}) = d(\mathcal{D}(\mathbf{s}; \mathbf{t}), \mathbf{t})$
Ranked CD	**3.025**	1.104
AE	3.188	1.116
CD-Margin	3.321	1.168
CD-Reg	5.057	2.108
Symm-Margin	3.537	1.092
Symm-Reg	4.649	1.657
Ours-Margin	3.587	1.076
Ours-Reg	3.650	**0.984**

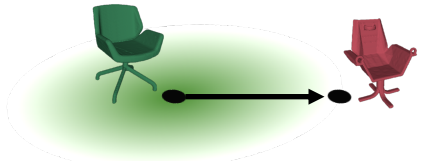
(PointNet encoder is used for the embedding space.)

Quantitative Results

Ours-Reg



Fitting gap



Egocentric distance field

Reg-loss is used.

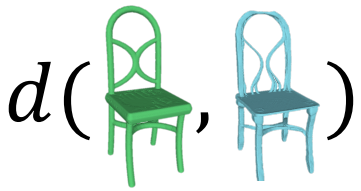
(Mean Chamfer Distance $\times 10^{-2}$ for the best of the top 3 retrieval)
bold = smallest, underline = second smallest

	Before Deformation (B.D.) $\boldsymbol{d(\mathbf{s},\mathbf{t})}$	**After Deformation (A.D.)** $\boldsymbol{e_{\mathcal{D}}(\mathbf{s},\mathbf{t}) = d(\mathcal{D}(\mathbf{s};\mathbf{t}),\mathbf{t})}$
Ranked CD	**3.025**	1.104
AE	3.188	1.116
CD-Margin	3.321	1.168
CD-Reg	5.057	2.108
Symm-Margin	3.537	1.092
Symm-Reg	4.649	1.657
Ours-Margin	3.587	1.076
Ours-Reg	3.650	**0.984**

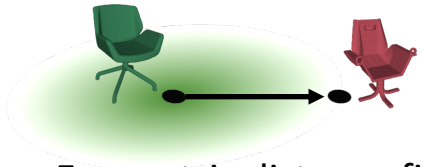
(PointNet encoder is used for the embedding space.)

Quantitative Results

Ours-Reg



Fitting gap



Egocentric distance field

Reg-loss is used.

(Mean Chamfer Distance $\times 10^{-2}$ for the best of the top 3 retrieval)
bold = smallest, underline = second smallest

	Before Deformation (B.D.) $d(\mathbf{s}, \mathbf{t})$	After Deformation (A.D.) $e_{\mathcal{D}}(\mathbf{s}, \mathbf{t}) = d(\mathcal{D}(\mathbf{s}; \mathbf{t}), \mathbf{t})$
Ranked CD	**3.025**	1.104
AE	3.188	1.116
CD-Margin	3.321	1.168
CD-Reg	5.057	2.108
Symm-Margin	3.537	1.092
Symm-Reg	4.649	1.657
Ours-Margin	3.587	1.076
Ours-Reg	3.650	**0.984**

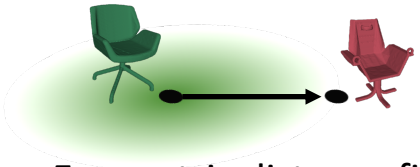
(PointNet encoder is used for the embedding space.)

Quantitative Results

Ours-Reg

$d($  $)$

Fitting gap



Egocentric distance field

Reg-loss is used.

(Mean Chamfer Distance $\times 10^{-2}$ for the best of the top 3 retrieval)
bold = smallest, underline = second smallest

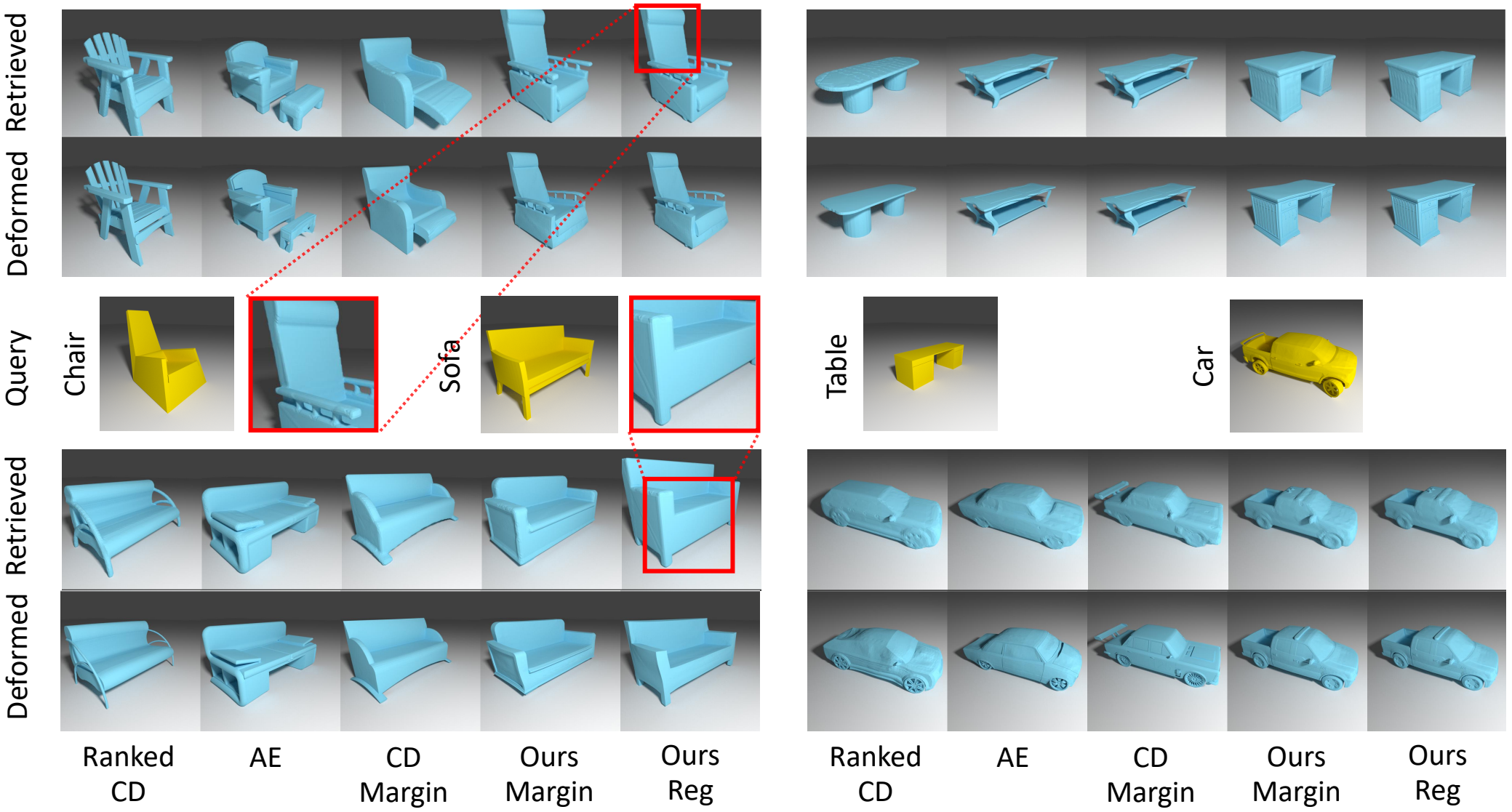
	Before Deformation (B.D.) $\boldsymbol{d(\mathbf{s},\mathbf{t})}$	After Deformation (A.D.) $e_{\mathcal{D}}(\mathbf{s},\mathbf{t}) = \boldsymbol{d}(\mathcal{D}(\mathbf{s};\mathbf{t}),\mathbf{t})$
Ranked CD	**3.025**	1.104
AE	3.188	1.116
CD-Margin	3.321	1.168
CD-Reg	5.057	2.108
Symm-Margin	3.537	1.092
Symm-Reg	4.649	1.657
Ours-Margin	3.587	1.076
Ours-Reg	3.650	**0.984**

(PointNet encoder is used for the embedding space.)

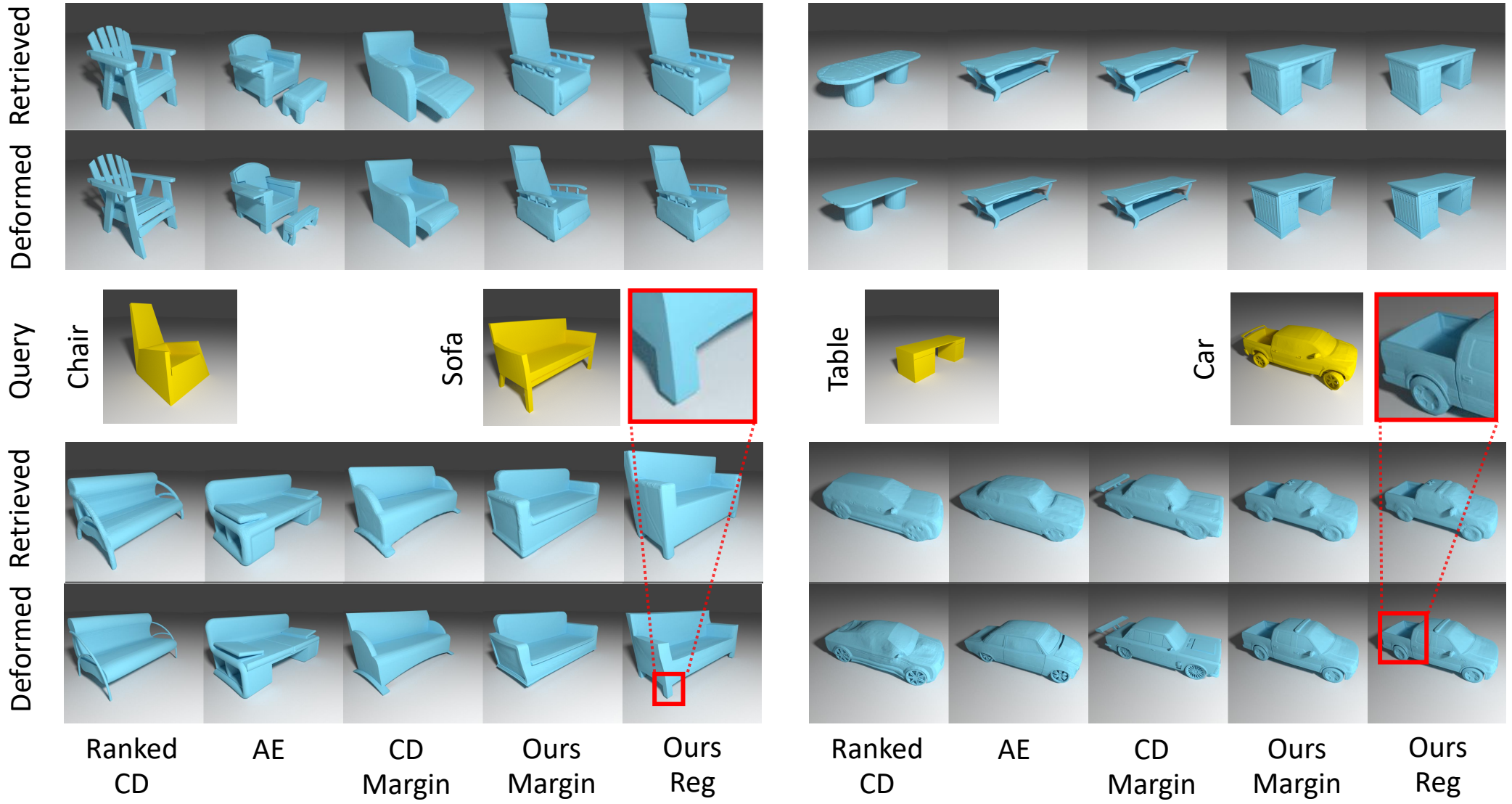
Quantitative Results

	Ranking ↓	**Recall ↑**
Ranked CD	12.32	51.20
AE	12.10	52.15
CD-Margin	14.27	48.06
CD-Reg	39.97	21.02
Symm-Margin	10.61	57.50
Symm-Reg	28.33	38.64
Ours-Margin	9.34	60.94
Ours-Reg	**7.06**	**70.36**

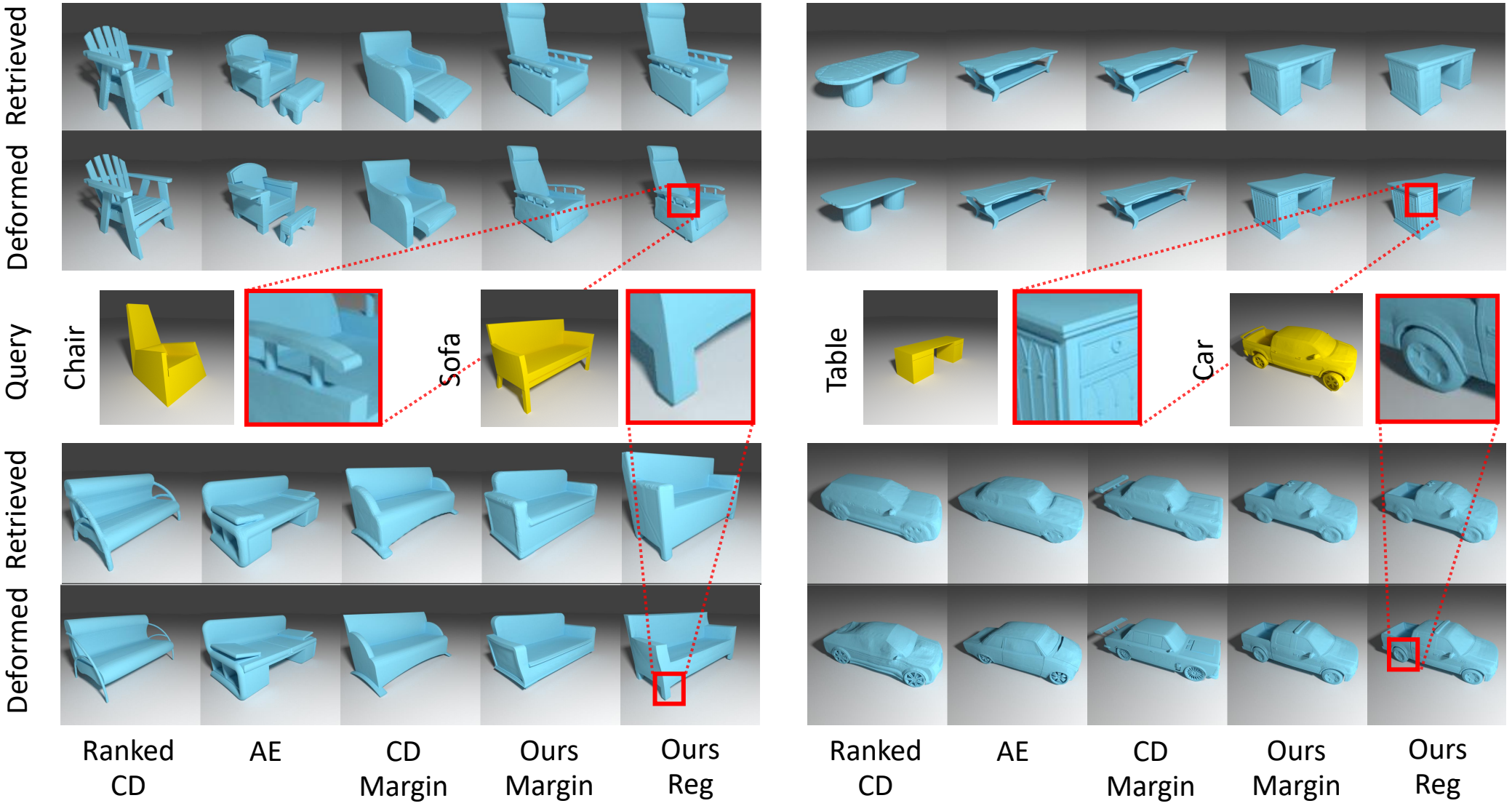
Qualitative Results



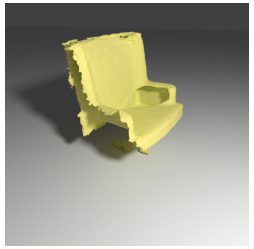
Qualitative Results



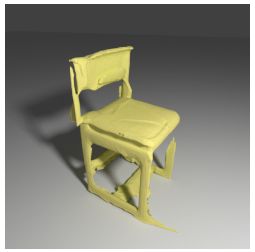
Qualitative Results



Scan2CAD



Input Scan



Input Scan

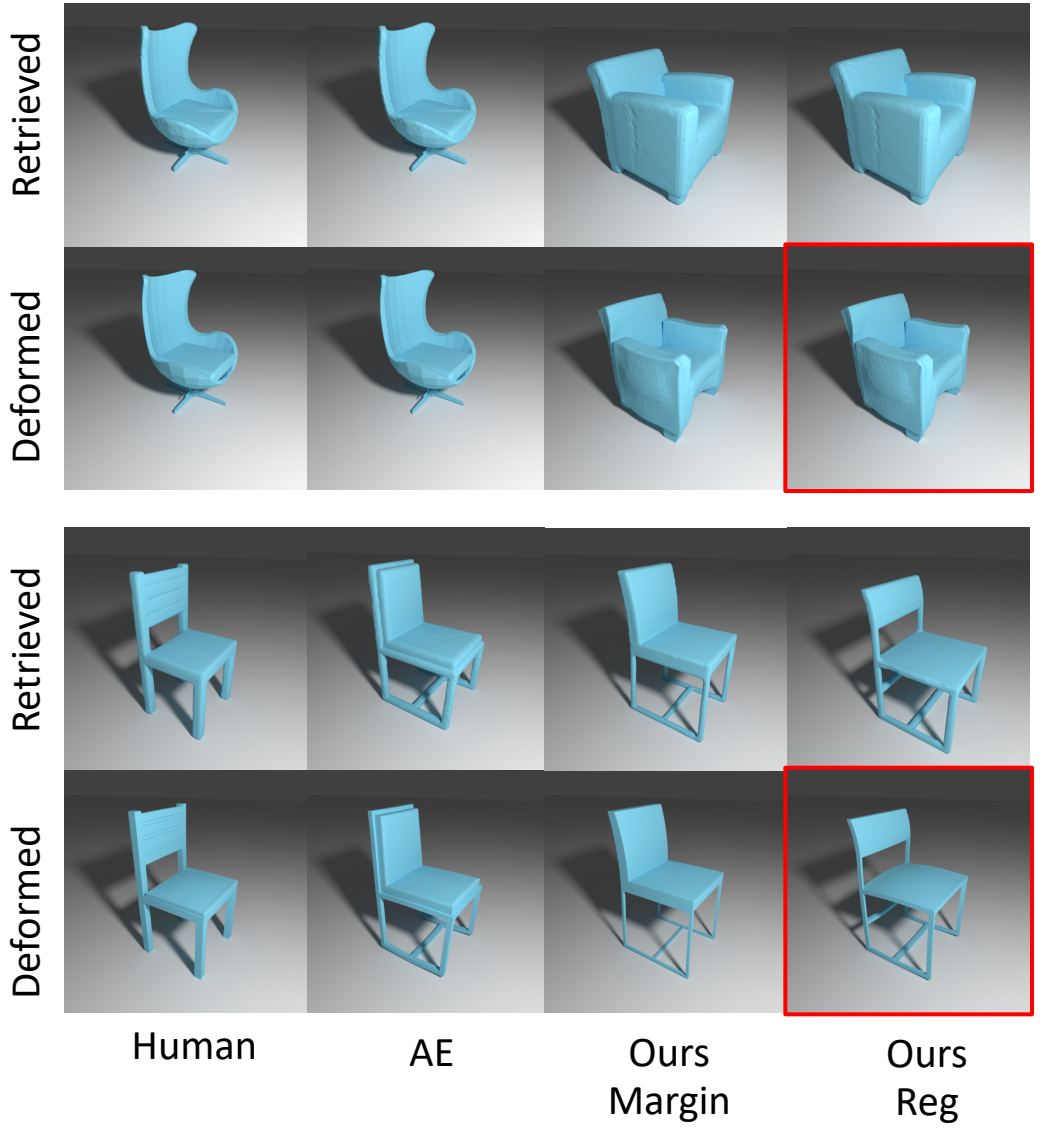
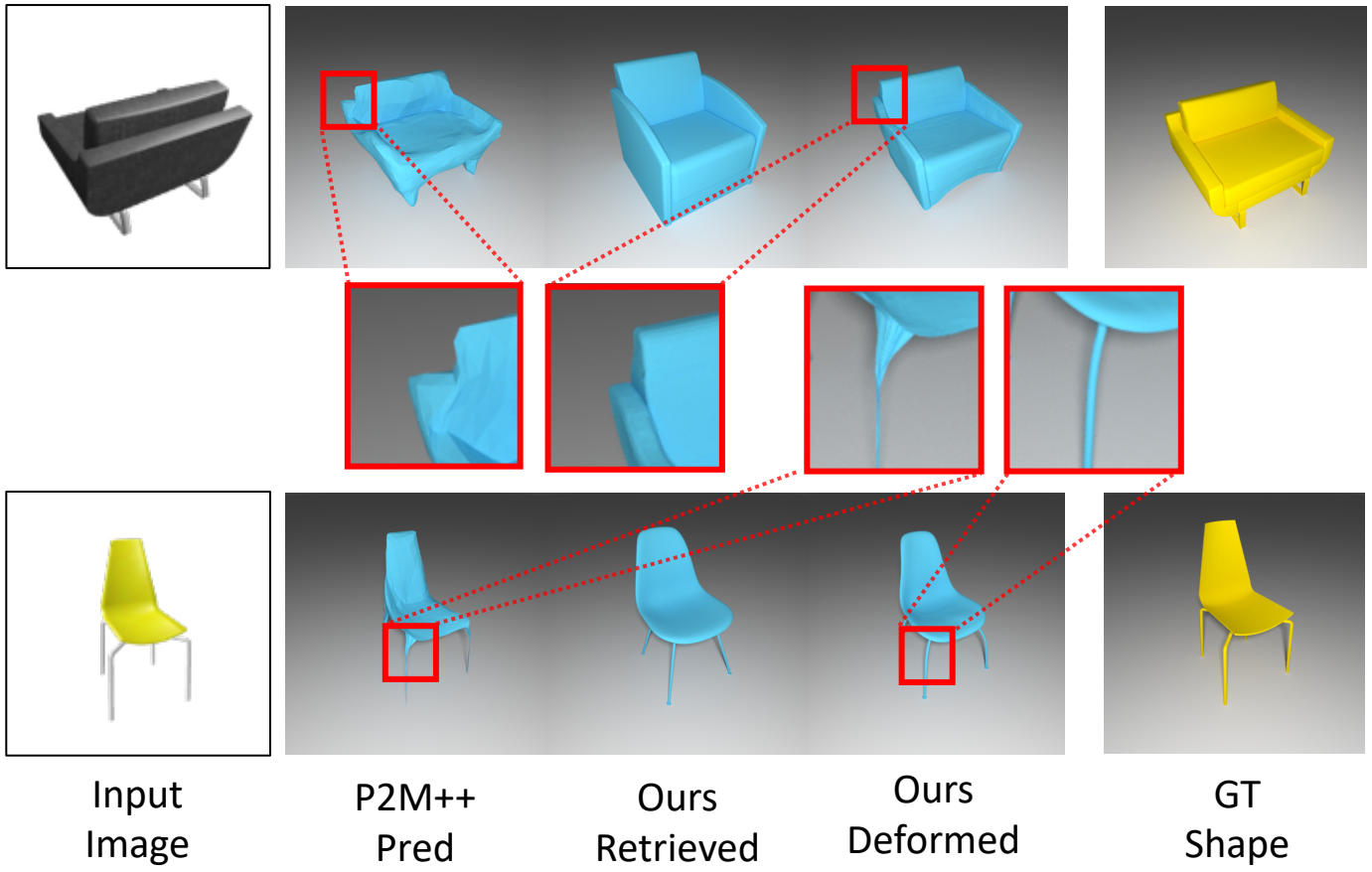


Image2CAD



[7] Pixel2Mesh++: Multi-View 3D Mesh Generation via Deformation. *Wen et. al., ICCV 2019.*

Deformation-Aware 3D Model Embedding and Retrieval

Visit our project page!

https://deformscan2cad.github.io/