Deformation-Aware 3D Model Embedding and Retrieval

Mikaela Uy ¹  Jingwei Huang ¹  Minhyuk Sung ²  Tolga Birdal ¹  Leo Guibas ¹
Motivation

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

Query Model  Retrieve  Closest Model

Chamfer Distance: $4.45 \times 10^{-2}$
Goal

Query Model → Retrieve → Ours Retrieved → Deform → Ours Deformed

Chamfer Distance:
- Query Model to Ours Retrieved: $7.09 \times 10^{-2}$
- Ours Retrieved to Ours Deformed: $1.71 \times 10^{-2}$
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** $e_D(s, t) = d(D(s; t), t)$: Fitting distance ($d$) after deforming a database shape ($s$) to the query ($t$) using deformation function $D$.
Properties of Fitting Gap

• Fitting gap measures the distance in the real space.

• Properties of fitting gap:
  1. (Non-negativity) \( e_D(s, t) \geq 0 \)
  2. (Identity) \( e_D(t, t) = 0 \)
  3. (Asymmetry) \( e_D(s, t) \neq e_D(t, s) \)

Not a metric!
Egocentric Distance Field \( \mathcal{G}(s) \)

- \( \mathcal{G} \) is source-dependent.

- \( \mathcal{G} \) is represented with a positive semi-definite matrix.

\[
\mathcal{G}(s) \in \mathcal{S}^k \quad (\mathcal{G}(s) \succeq 0)
\]
Distance in Embedding Space $\delta$

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

Properties

1. (Non-negativity) $\delta(s, t) \geq 0$
2. (Identity) $\delta(t, t) = 0$
3. (Asymmetry)
   $$\delta(s, t) = \sqrt{(\mathcal{F}(t) - \mathcal{F}(s))^T \mathcal{G}(s) (\mathcal{F}(t) - \mathcal{F}(s))}$$
   $$\delta(t, s) = \sqrt{(\mathcal{F}(s) - \mathcal{F}(t))^T \mathcal{G}(t) (\mathcal{F}(s) - \mathcal{F}(t))}$$
   $$\delta(s, t) \neq \delta(t, s)$$
Deformation-Aware Embedding

\[ e_D(s, t) \sim \delta(s; t) = \sqrt{(F(t) - F(s))^T G(s)(F(t) - F(s))} \]
Network Training

- Margin-loss-based approach

\[ P_t = \{ s \in X_t | e_D(s, t) \leq \sigma_P \} \quad N_t = \{ s \in X_t | e_D(s, t) > \sigma_N \} \]

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

We precompute the fitting gap \( e_D \).

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

- **Regression-based approach**

\[
p(s; t) = \frac{\exp(-e_D(s; t)/2\sigma_t^2)}{\sum_{s' \in X_t} \exp(-e_D(s'; t)/2\sigma_t^2)}
\]

\[
\hat{p}(s; t) = \frac{\delta^2(s; t)}{\sum_{s' \in X_t} \delta^2(s; t)}
\]

\[
\frac{1}{|X_t|} \sum_{s \in X_t} |\hat{p}(s; t) - p(s; t)|
\]

\[
\delta(s; t) = \sqrt{(F(t) - F(s))^T G(s)(F(t) - F(s))}
\]

We precompute the fitting gap ($e_D$).

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

1. Fitting gap

2. Egocentric distance field $G(s)$

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

---

Deformation

Before Deformation

$\mathcal{D}(s)$

After Deformation

$\mathcal{D}(s; t)$

Chamfer Distance

$e_D(s, t) = d(\mathcal{D}(s; t), t)$

**OURS**

$G$ is fixed to identity.

Symmetric embedding distance

$G$ 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.) $d(s, t)$ | After Deformation (A.D.) $e_D(s, t) = d(D(s; t), t)$ |
|-------------------------------------|---------------------------------------------------|
|                                     |                                                   |

Quantitative Results

Ranked by Chamfer Distance (Ranked CD)

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

| Ranked CD | Before Deformation (B.D.) $d(s, t)$ | After Deformation (A.D.) $e_D(s, t) = d(D(s; t), t)$ |
|-----------|------------------------------------|-----------------------------------------------|
| Ranked CD | 3.025                              | 1.104                                         |

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

Autoencoder (AE)
- PointNet autoencoder for reconstruction.
- Use the bottleneck layer as the embedding space.

|                     | Before Deformation (B.D.) | After Deformation (A.D.) |
|---------------------|---------------------------|--------------------------|
|                    | $d(s, t)$                 | $e_D(s, t) = d(D(s; t), t)$ |
| Ranked CD          | **3.025**                 | 1.104                    |
| AE                 | 3.188                     | 1.116                    |

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

(Footnote: Mean Chamfer Distance $\times 10^{-2}$ for the best of the top 3 retrieval)

|                  | Before Deformation (B.D.) | After Deformation (A.D.) |
|------------------|---------------------------|--------------------------|
| $d(s, t)$        | $d(s, t)$                 | $e_D(s, t) = d(D(s; t), t)$ |
| Ranked CD        | 3.025                     | 1.104                     |
| AE               | 3.188                     | 1.116                     |
| CD-Margin        | 3.321                     | 1.168                     |
| CD-Reg           | 5.057                     | 2.108                     |

Bold = smallest, underline = second smallest

(PointNet encoder is used for the embedding space.)
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.) | After Deformation (A.D.) |
|------------------|---------------------------|---------------------------|
| $d(s, t)$        | $e_D(s, t) = d(D(s; t), t)$ |
| 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

(PointNet encoder is used for the embedding space.)

**Symm-Margin**

\[ d(s, t) \]

Fitting gap

Egocentric distance field

Margin-loss is used.

(Mean Chamfer Distance \( \times 10^{-2} \) for the best of the top 3 retrieval)

|                          | Before Deformation (B.D.) | After Deformation (A.D.) |
|--------------------------|---------------------------|--------------------------|
|                          | \( d(s, t) \)             | \( e_D(s, t) = d(D(s; t), t) \) |
| 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                    |

bold = smallest, underline = second smallest
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.) $d(s, t)$ | After Deformation (A.D.) $e_D(s, t) = d(D(s; t), t)$ |
|------------------|--------------------------------------|-----------------------------------------------------|
| 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

(PointNet encoder is used for the embedding space.)

**Ours-Margin**

\[d(s, t)\]  
Fitting gap

Egocentric distance field

Margin-loss is used.

|                        | Before Deformation (B.D.)  | After Deformation (A.D.) |
|------------------------|----------------------------|--------------------------|
|                        | \[d(s, t)\]                | \[e_D(s, t) = d(D(s; t), t)\] |
| 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                    |

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

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

| Metric       | Before Deformation (B.D.) | After Deformation (A.D.) |
|--------------|--------------------------|--------------------------|
| 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.)

**Ours-Reg**

$d(\text{ }, \text{ })$

- **Fitting gap**
- **Egocentric distance field**

Reg-loss is used.
Quantitative Results

(Mean Chamfer Distance $\times 10^{-2}$ for the best of the top 3 retrieval)

|                  | Before Deformation (B.D.) | After Deformation (A.D.) |
|------------------|---------------------------|--------------------------|
|                  | $d(s, t)$                 | $e_D(s, t) = d(D(s; t), t)$ |
| 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

(Mean Chamfer Distance $\times 10^{-2}$ for the best of the top 3 retrieval)

bold = smallest, underline = second smallest

|                      | Before Deformation (B.D.) | After Deformation (A.D.) |
|----------------------|---------------------------|--------------------------|
|                      | $d(s, t)$                 | $e_D(s, t) = d(D(s; t), t)$ |
| 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

Deformed Retrieved

Query
Chair
Sofa
Deformed Retrieved

Ranked CD
AE
CD Margin
Ours Margin
Ours Reg

Ranked CD
AE
CD Margin
Ours Margin
Ours Reg

Table
Car
Qualitative Results

Deformed Retrieved

Query Chair

Deformed Retrieved

Ranked CD AE CD Margin Ours Margin Ours Reg

Ranked CD AE CD Margin Ours Margin Ours Reg
Scan2CAD
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/