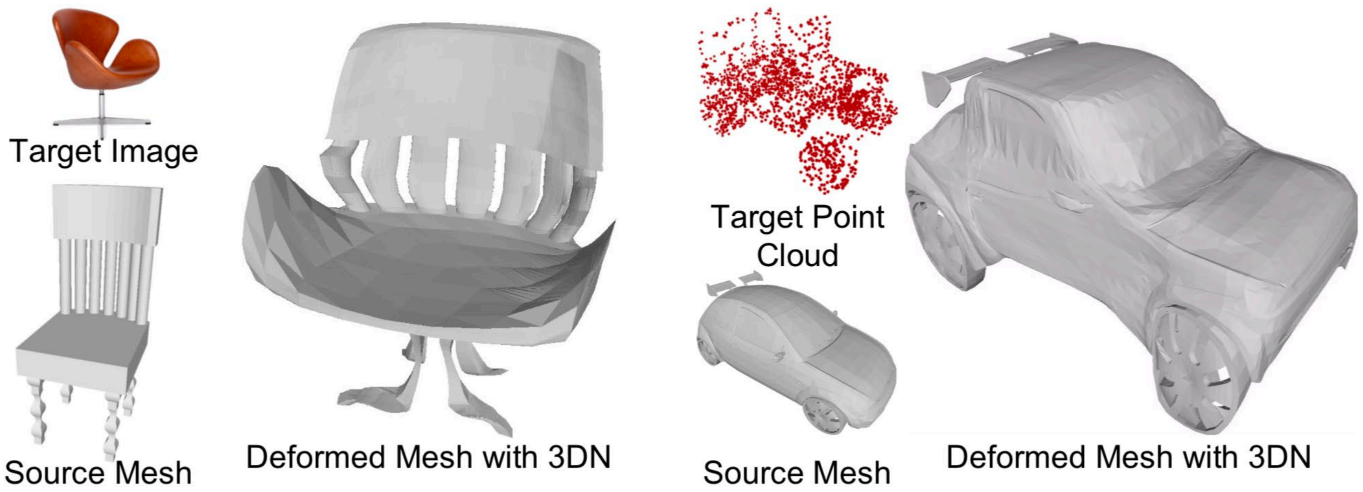
3DN: 3D Deformation Network

Weiyue Wang, Duygu Ceylan, Radomir Mech, Ulrich Neumann

(CVPR 2019)



By: Mikaela Uy
Feb. 28, 2022

Key Idea

- 3D deformation network deforms a **source 3D mesh** based on a **target 2D image, 3D mesh** or a **3D point cloud**
- The network estimates **vertex displacement vectors** (3D offsets) to deform the source while **keeping the vertex connectivity and mesh topology**
- Learns 3D deformation end-to-end and also introduces a differentiable mesh sampling operator

3DN Pipeline

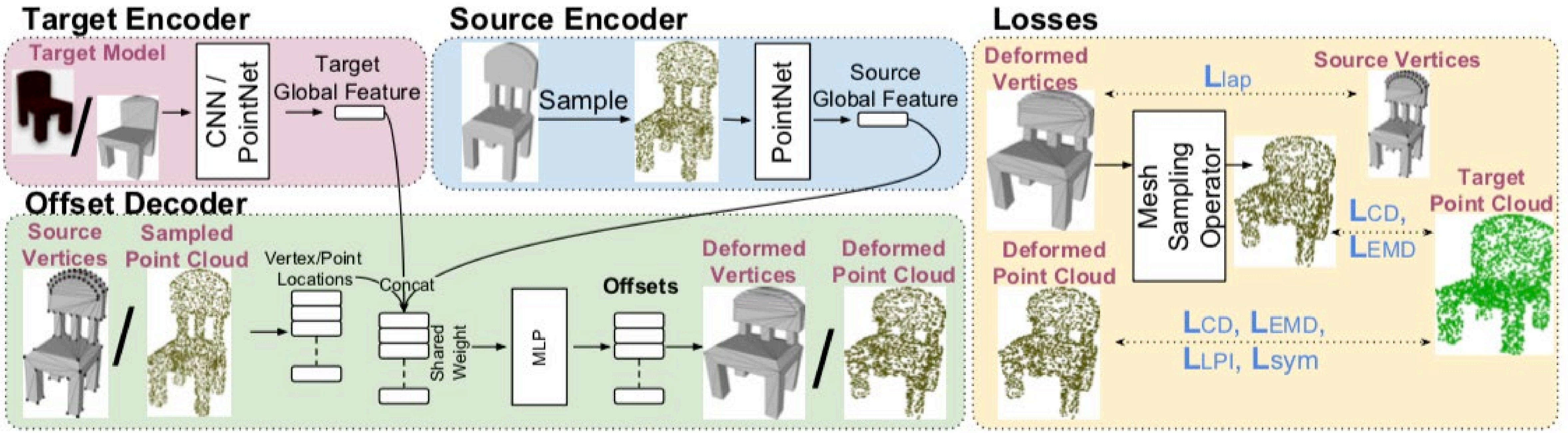
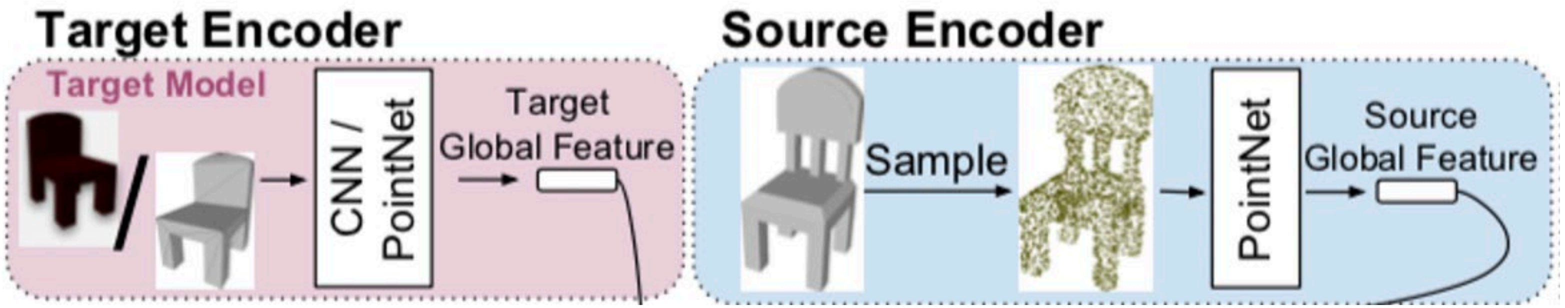


Figure 2: 3DN extracts global features from both the source and target. 'MLP' denotes the '1×1' conv as in PointNet [17]. These features are then input to an offset decoder which predicts per-vertex offsets to deform the source. We utilize loss functions to preserve geometric details in the source ($L_{Lap}, L_{LPI}, L_{Sym}$) and to ensure deformation output is similar to the target (L_{CD}, L_{EMD}).

Encoder

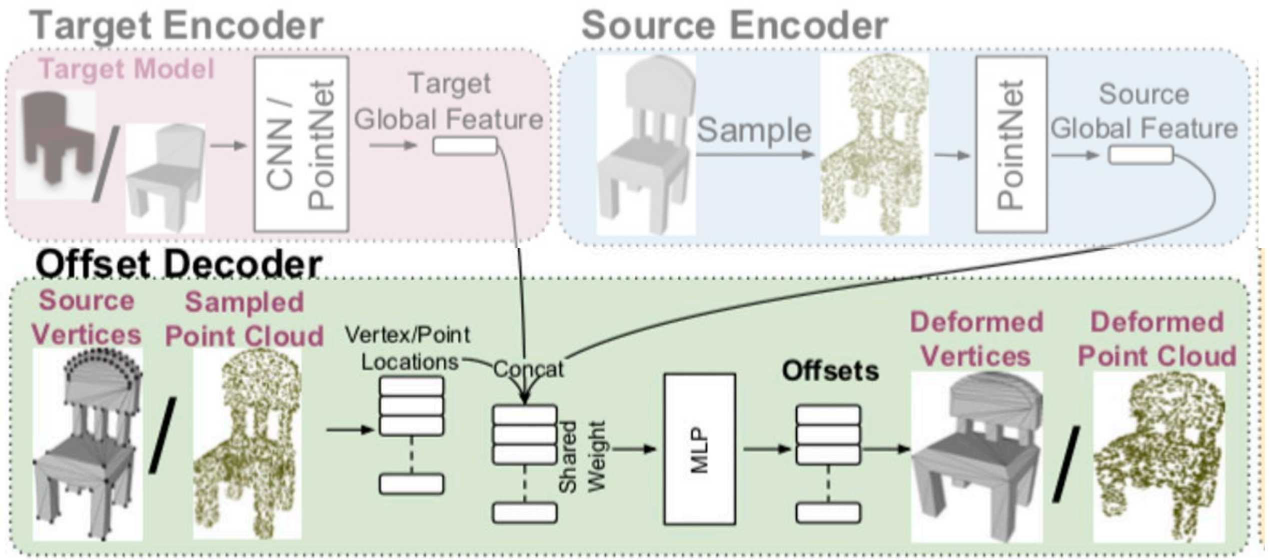


- Source:
  - Mesh: vertex positions, and set of triangles
  - Uniformly sample points and use PointNet encoder

- Target:
  - Image : VGG encoder
  - 3D model: Uniformly sample points and use PointNet encoder

- Extract a global feature for source and target shape

Offset Decoder



- Use PointNet segmentation network, concatenated with original vertex locations
- Final deformed mesh $S' = (V', E)$, where $V' = V + O$ and O is the learned per vertex offset
- Source can be the original mesh or a sampled point cloud

Losses

- Loss should measure the similarity between deformed source S' and target model T
  - **Chamfer** and **Earth Mover** losses
- Even when the target is represented as an image, they train with known target 3D model
- Introduced a **differentiable mesh sampling operator**, to make the losses robust to mesh densities

Differentiable mesh sampling operator

- Point $\boldsymbol{p}$ sampled from face $f = (\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3)$

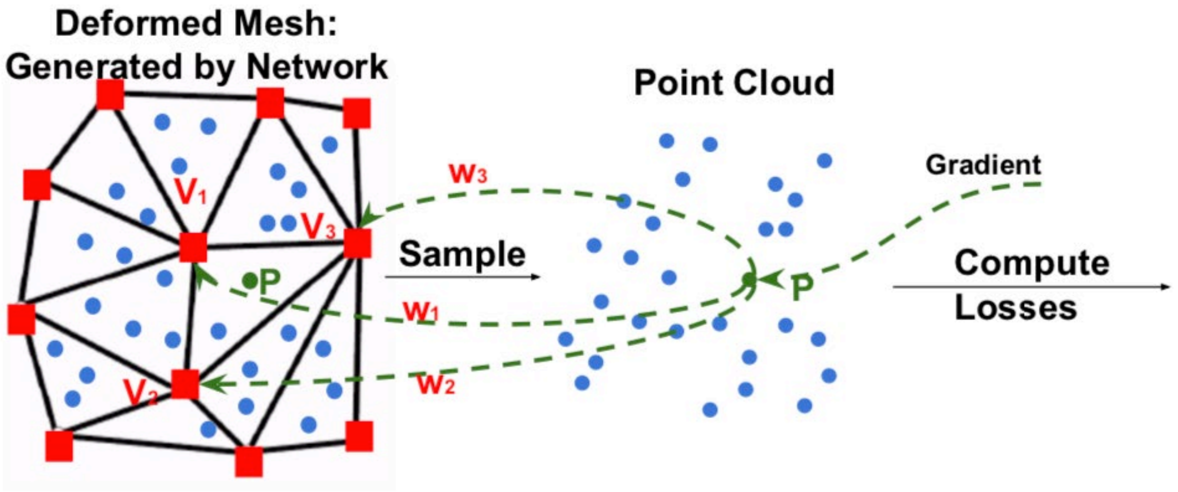
$$\mathbf{p} = w_1\mathbf{v_1} + w_2\mathbf{v_2} + w_3\mathbf{v_3}$$

- Then the offset for $\boldsymbol{p}$ is

$$\mathbf{o_p} = w_1\mathbf{o_{v_1}} + w_2\mathbf{o_{v_2}} + w_3\mathbf{o_{v_3}}$$

- Gradient

$$\boxed{g_{\mathbf{o_{v_i}}} = w_i g_{\mathbf{o_{v_p}}}}$$



Losses

- Shape Loss

$$L_{\text{CD}}^{\texttt{Mesh}}(PC, PC_T) = \sum_{p_1 \in PC} \min_{p_2 \in PC_T} \|p_1 - p_2\|_2^2 + \sum_{p_2 \in PC_T} \min_{p_1 \in PC} \|p_1 - p_2\|_2^2,$$

$$L_{\text{EMD}}^{\texttt{Mesh}}(PC, PC_T) = \min_{\phi: PC \to PC_T} \sum_{p \in PC} \|p - \phi(p)\|_2,$$

- Symmetry Loss

$$L_{\texttt{sym}}(PC, PC_T) = L_{CD}(M(PC), PC_T) + L_{EMD}(M(PC), PC_T).$$

Sample from mirrored deformed output M(PC)

Losses

- Mesh Laplacian

$$L_{\texttt{lap}} = \sum_i ||Lap(S) - Lap(S')||_2$$

- Local permutation invariant loss
  - Preserve the distance between two neighboring points after deformation as well

$$L_{\texttt{LPI}} = -\min\left(F(V+\delta) - F(V), \mathbf{0}\right).$$

- Total loss

$$\boxed{\begin{aligned} L = \omega_{L_1} L_{\texttt{CD}}^{\texttt{Mesh}} + \omega_{L_2} L_{\texttt{EMD}}^{\texttt{Mesh}} + \omega_{L_3} L_{\texttt{CD}}^{\texttt{Points}} + \omega_{L_4} L_{\texttt{EMD}}^{\texttt{Points}} + \\ \omega_{L_5} L_{\texttt{sym}} + \omega_{L_6} L_{\texttt{lap}} + \omega_{L_7} L_{\texttt{LPI}}, \end{aligned}}$$

Experiments

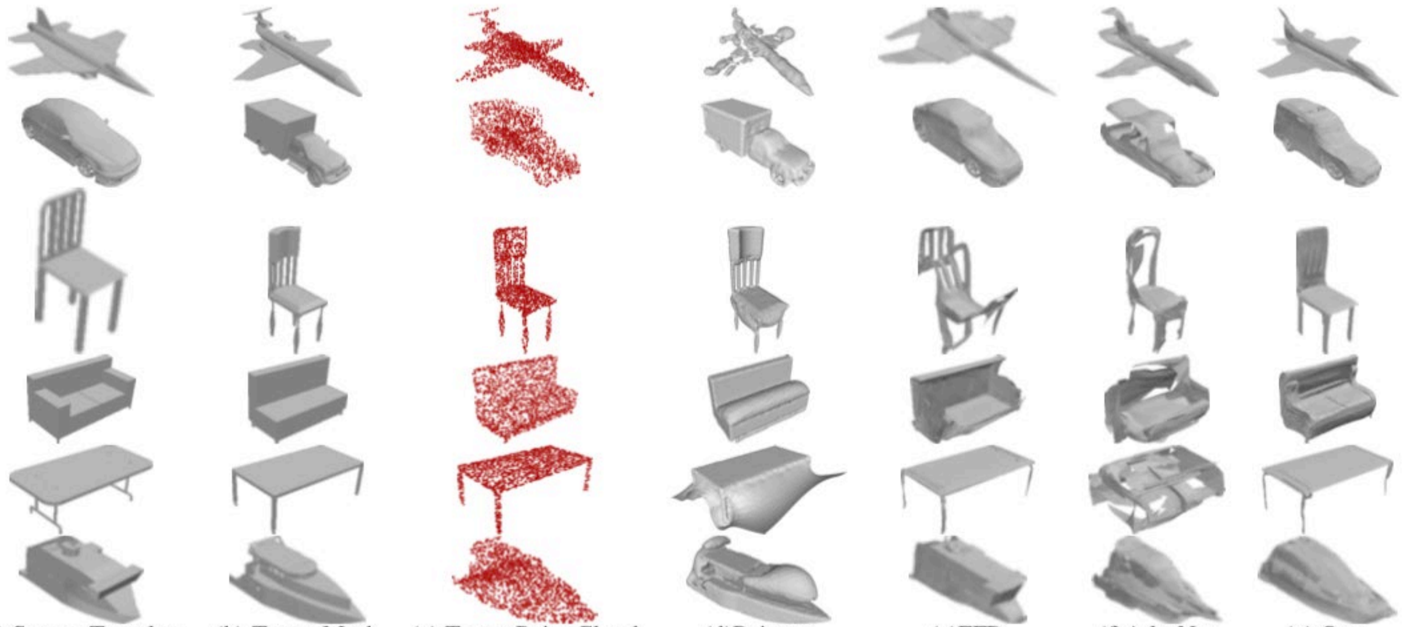
- Source models are taken from a set of template shapes [1]
- To sample source and target pairs:
  - Train a PointNet encoder
  - For each target, choose the nearest neighbor in this embedding from the template set
- For a target image
  - First use PSGN [2] to generate initial point cloud, and retrieve the source model based on this

[1] D. Jack, J. K. Pontes, S. Sridharan, C. Fookes, S. Shirazi, F. Maire, and A. Eriksson. Learning free-form deformations for 3d object reconstruction, ACCV 2018

[2] H. Fan, H. Su, and L. J. Guibas. A point set generation net-work for 3d object reconstruction from a single image , CVPR 2017

Results

- Shape Reconstruction from Point Cloud



(a) Source Template (b) Target Mesh (c) Target Point Cloud (d)Poisson (e)FFD (f)AtlasNet (g) Ours

Figure 5: Given a source (a) and a target (b) model from the ShapeNet dataset, we show the deformed meshes obtained by our method (g). We also show Poisson surface reconstruction (d) from a set of points sampled on the target (c). We also show comparisons to previous methods of Jack et al. (e) and AtlasNet (f).

		plane	bench	box	car	chair	display	lamp	speaker	rifle	sofa	table	phone	boat	Mean
EMD	AtlasNet	3.46	3.18	4.20	2.84	3.47	3.97	3.79	3.83	2.44	3.19	3.76	3.87	2.99	3.46
	FFD	1.88	2.02	2.50	2.11	2.13	2.69	2.42	3.06	1.55	2.44	2.44	1.88	2.00	**2.24**
	Ours	0.79	1.98	3.57	1.24	1.12	3.08	3.44	3.40	1.79	2.06	1.34	3.27	2.27	2.26
CD	AtlasNet	2.16	2.91	6.62	3.97	3.65	3.65	4.48	6.29	0.98	4.34	6.01	2.44	2.73	3.86
	FFD	3.22	4.53	6.94	4.45	4.99	5.98	8.72	11.97	1.97	6.29	6.89	3.61	4.41	5.69
	Ours	0.38	2.40	5.26	0.90	0.82	5.59	8.74	9.27	1.52	2.55	0.97	2.66	2.77	**3.37**
IoU	AtlasNet	56.9	53.3	31.3	44.0	47.9	48.0	41.6	33.2	63.4	44.7	43.8	58.7	50.9	46.7
	FFD	29.0	42.3	28.4	21.1	42.2	27.9	38.9	52.5	31.9	34.7	43.3	22.9	47.7	35.6
	Ours	71.0	40.7	43.6	75.8	66.3	40.4	25.1	49.2	40.0	60.6	57.9	50.1	42.6	**51.1**

Results

- Single View Reconstruction

		plane	bench	box	car	chair	display	lamp	speaker	rifle	sofa	table	phone	boat	Mean
EMD	AtlasNet	3.39	3.22	3.36	3.72	3.86	3.12	5.29	3.75	3.35	3.14	3.98	3.19	4.39	3.67
	Pxel2mesh	2.98	2.58	3.44	3.43	3.52	2.92	5.15	3.56	3.04	2.70	3.52	2.66	3.94	**3.34**
	FFD	2.63	3.96	4.87	2.98	3.38	4.88	7.19	5.04	3.58	3.70	3.56	4.11	3.86	4.13
	Ours	3.30	2.98	3.21	3.28	4.45	3.91	3.99	4.47	2.78	3.31	3.94	2.70	3.92	3.56
CD	AtlasNet	5.98	6.98	13.76	17.04	13.21	7.18	38.21	15.96	4.59	8.29	18.08	6.35	15.85	13.19
	Pixel2mesh	6.10	6.20	12.11	13.45	11.13	6.39	31.41	14.52	4.51	6.54	15.61	6.04	12.66	11.28
	FFD	3.41	13.73	29.23	5.35	7.75	24.03	45.86	27.57	6.45	11.89	13.74	16.93	11.31	16.71
	Ours	6.75	7.96	8.34	7.09	17.53	8.35	12.79	17.28	3.26	8.27	14.05	5.18	10.20	**9.77**
IoU	AtlasNet	39.2	34.2	20.7	22.0	25.7	36.4	21.3	23.2	45.3	27.9	23.3	42.5	28.1	30.0
	Pixel2mesh	51.5	40.7	43.4	50.1	40.2	55.9	29.1	52.3	50.9	60.0	31.2	69.4	40.1	47.3
	FFD	30.3	44.8	30.1	22.1	38.7	31.6	35.0	52.5	29.9	34.7	45.3	22.0	50.8	36.7
	Ours	54.3	39.8	49.4	59.4	34.4	47.2	35.4	45.3	57.6	60.7	31.3	71.4	46.4	**48.7**

Table 2: Quantitative comparison on ShapeNet rendered images. Metrics are CD ($\times 0.001$), EMD ($\times 100$) and IoU (%).

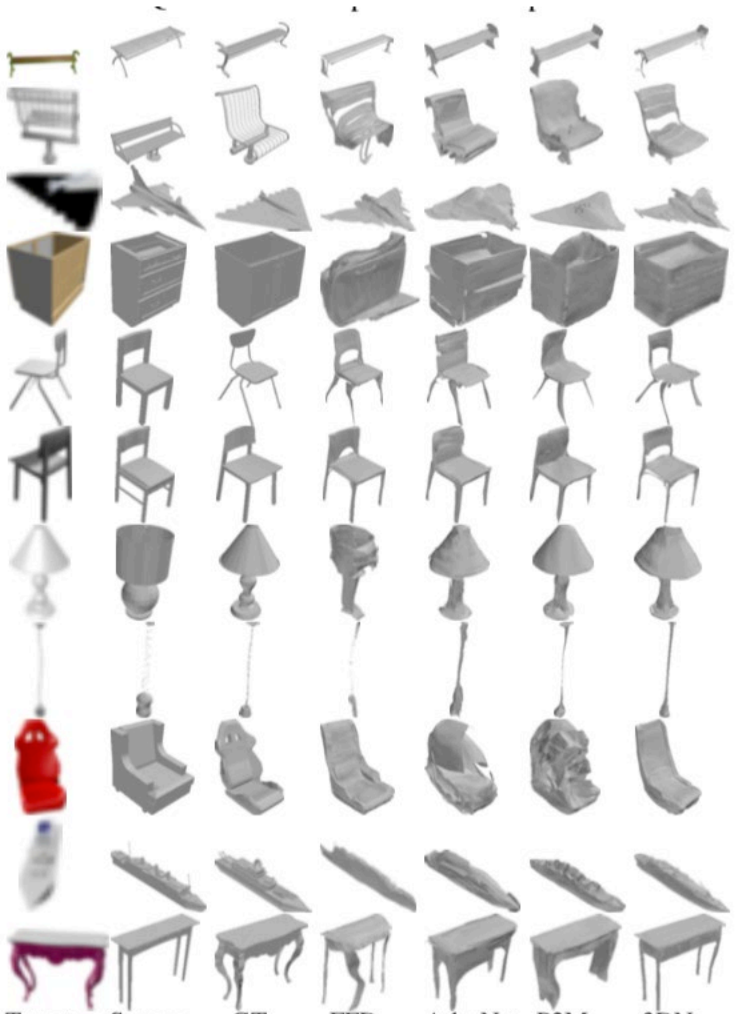


Figure 6: Given a target image and a source, we show deformation results of FFD, AtlasNet, Pixel2Mesh (P2M), and 3DN. We also show the ground truth target model (GT).

Results

Figure 9: Deformation with different source-target pairs. ‘S’ and ‘T’ denote source meshes and target meshes respectively.

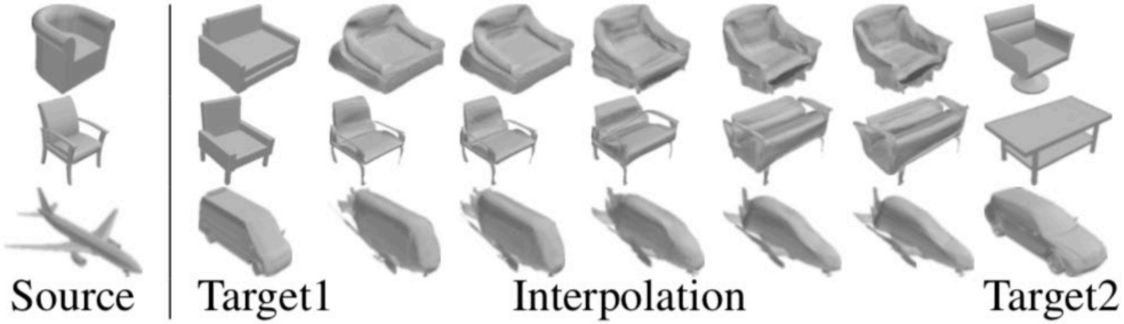


Figure 10: Shape interpolation.

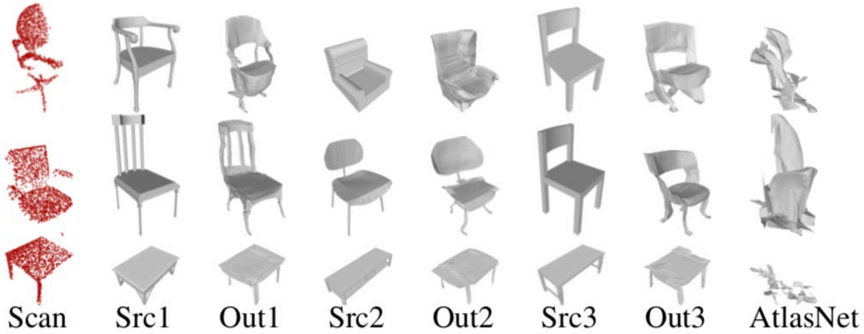


Figure 11: Shape inpainting with real point cloud scan as input. Src means source mesh and ’out’ is the corresponding deformed mesh.

Thank you!