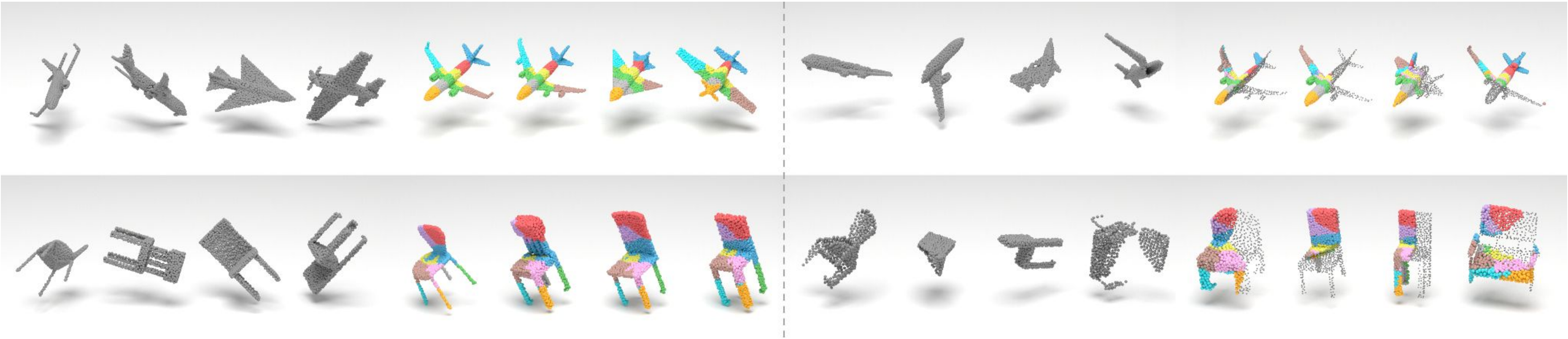
ConDor: Self-Supervised Canonicalization of 3D Pose for Partial Shapes

Rahul Sajnani, Adrien Poulenard, Jivitesh Jain, Radhika Dua, Leonidas J. Guibas, Srinath Sridhar



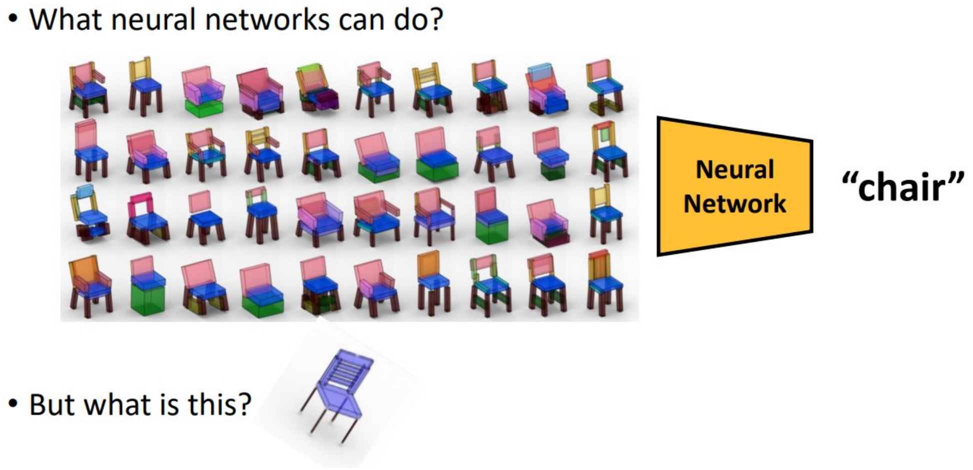
Presentation by Chaitanya Patel

Motivation

- Shapes in current benchmark datasets are in canonical frame.
- But sensors can capture point clouds from any viewpoint.

- Need extensive training with random augmentations to generalize.

- ConDor is a self-supervised method that learns to canonicalize the 3D orientation and position for full and partial 3D point clouds.



TFN: Equivariant Network

- Tensor Field Networks (TFNs) are 3D point cloud architecture that is equivariant to 3D rotation and point permutation, and invariant to translation.
- Given a point cloud X, TFN can compute pointwise or global equivariant features of different types l

Pointwise features

$$f^{\ell}(X)_{\underbrace{i}_{\text{point}},\underbrace{c}_{\text{channel}},:}$$

global features

$$g^{\ell}(X)_{\underbrace{c}_{\text{channel}},:}$$

Definitions

- Instance-level 3D pose canonicalization
  - Find a consistent canonical frame across different poses of the same object instance
- Category-level 3D pose canonicalization
  - A canonical frame that is consistent with respect to the geometry and local shape across different object instances

Idea 1: Rotation Invariant Embedding

- Features F have the same rotation equivariance property as coefficients of spherical functions in the spherical harmonics basis.
- Embed the shape using the spherical harmonics basis and using the global TFN features F as coefficients of this embedding.

$$H_{ij}^{\ell} := \langle F_{:,j}^{\ell}, Y^{\ell}(X_i) \rangle,$$

- H is rotation invariant embedding.

$$X_i^c := \sum_j W_{:,j} H_{ij}^1 = W(F^1)^\top X_i.$$

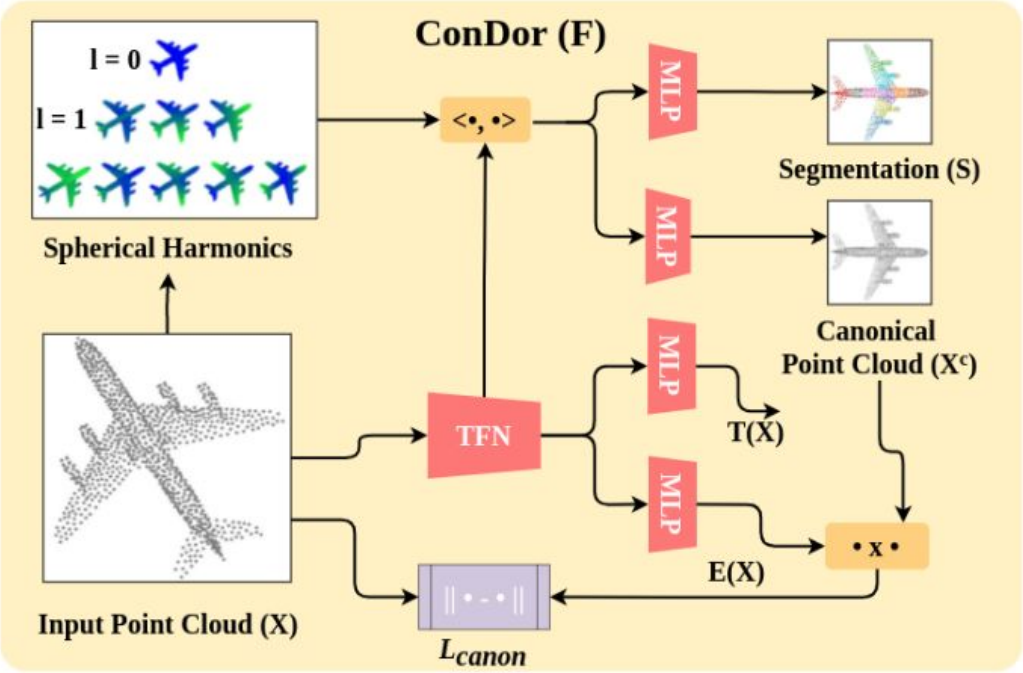
Idea 2: Rotation Equivariant Embedding

- Let TFN output a rotation matrix E(X) for input point cloud X

$$E(R.X) = RE(X)$$

- Supervise such that this rotation matrix encodes
- Using E, transform 3D invariant embedding X_c back to the input equivariant embedding and compare it to the input point cloud X.

Training

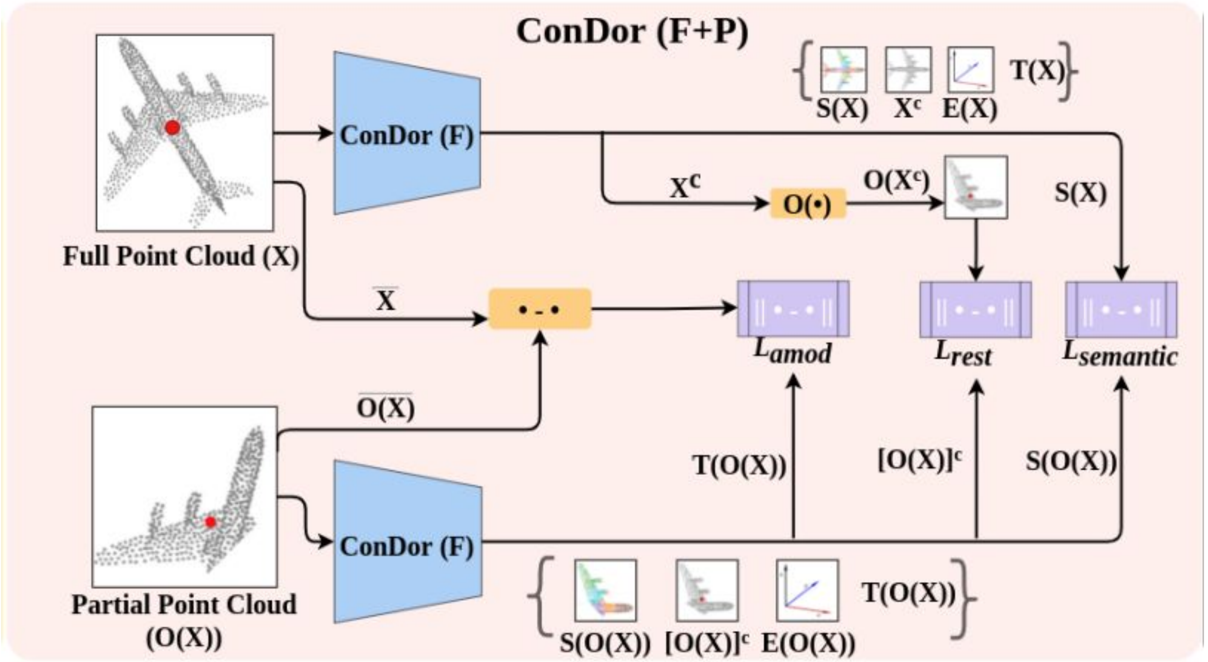


ConDor (F)
l = 0
l = 1
Spherical Harmonics
<•, •>
MLP
MLP
MLP
MLP
Segmentation (S)
Canonical
Point Cloud (X^c)
TFN
T(X)
E(X)
• x •
||• - •||
L_{canon}
Input Point Cloud (X)

Idea 3: Translation Invariance

- Full shapes are centered at their mean.
- For partial shapes,
  - Predict translation to the center of full shape.
  - Network outputs equivariant translation T that follows T (R.X) = R T(X)

Training



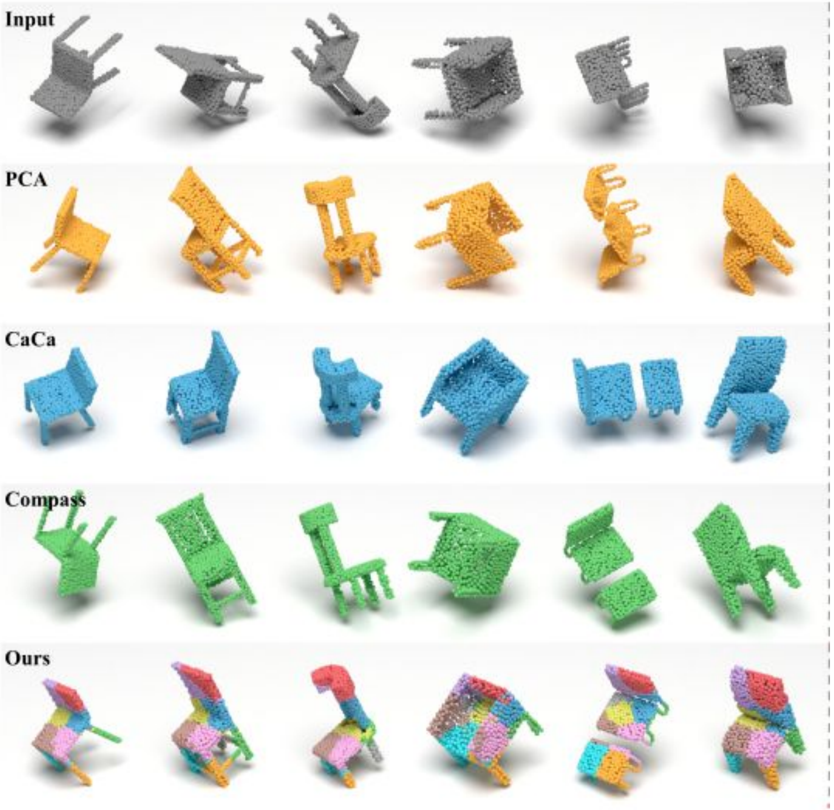
O is slicing operator.

$$\mathcal{L}_{amod} = \|\mathcal{T}(X) - (\overline{X} - \overline{\mathcal{O}[X]})\|_2^2.$$

Other Details

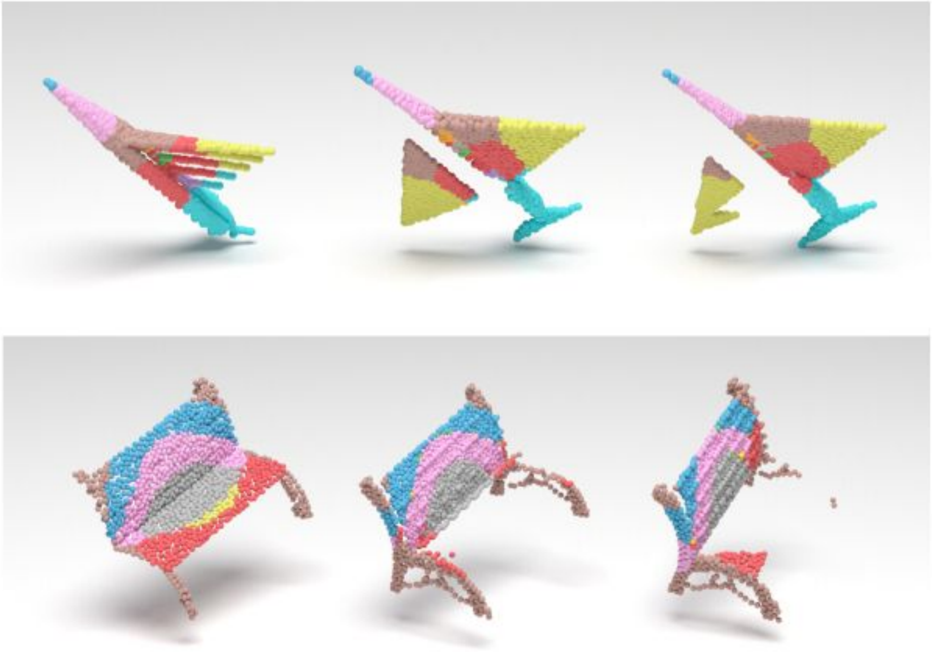
- For semantic segmentation,
  - Predict class probabilities.
  - In the absence of groundtruth, supervise using several part consistency losses
    - Each class should represent roughly the same ‘amount’ of the shape volume
    - Segmentation should match for partial and full shape
- For symmetric objects,
  - Estimate P equivariant rotations and choose the frame that minimizes the L2 norm between corresponding points in the input and the predicted invariant shape

Canonicalization of Full Shapes

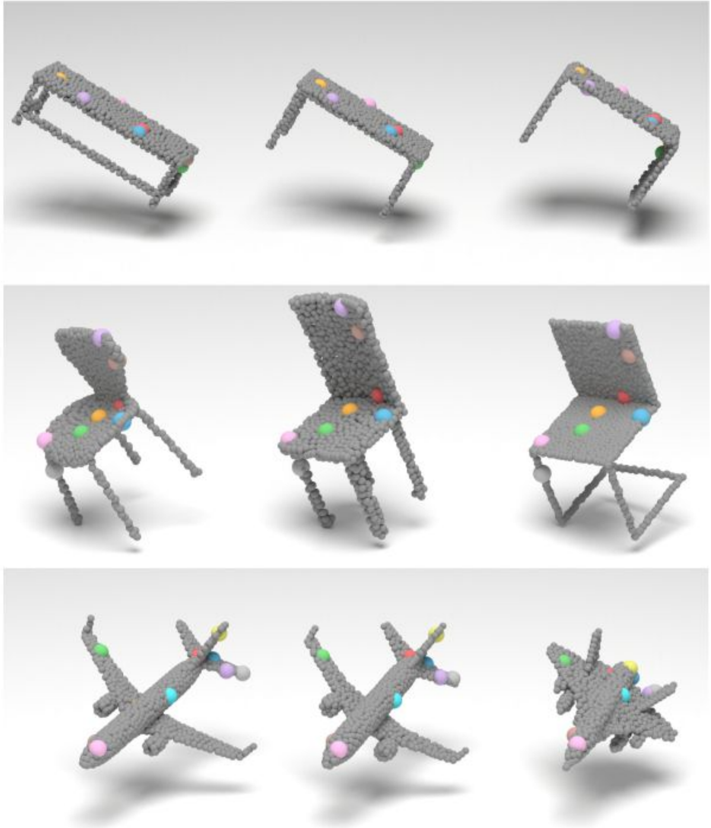


Canonicalization of Partial Shapes

Canonicalization of Partial PC from Depthmaps



Keypoint Transfer



Metrics

- Groundtruth Consistency (GC)
  - Compares the canonicalization when groundtruth canonicalization is available

- When groundtruth is not available…
  - Instance-Level Consistency (IC)
    - Evaluate the quality of canonicalization between different rotated versions of the same instance
  - Category-Level Consistency (CC)
    - Evaluate the quality of canonicalization between different shape instances

Results: Full Shapes

	bench	cabinet	car	cellph.	chair	couch	firearm	lamp	monitor	plane	speaker	table	water.	avg.	multi
Instance-Level Consistency (IC) ↓															
PCA	0.0573	0.0350	0.0477	0.0276	0.0974	0.0628	0.0324	0.0755	0.0480	0.0502	0.0491	0.0727	0.0400	0.0535	0.0535
CaCa [45]	0.0630	0.1567	0.0426	0.0823	0.0253	0.1479	0.0084	**0.0372**	0.0748	**0.0093**	0.1540	0.0787	0.0270	0.0698	0.0395
Compass [42]	0.1030	0.0816	0.0790	0.0664	0.0791	0.0766	0.0748	0.0495	0.0638	0.0610	0.0721	0.0641	0.0430	0.0703	0.0507
Ours (F)	**0.0225**	0.0346	**0.0191**	**0.0234**	**0.0221**	**0.0221**	**0.0081**	0.0454	0.0283	0.0163	0.0787	0.0523	0.0270	**0.0308**	0.0394
Ours (F+P)	0.0696	**0.0288**	0.0230	0.0263	0.0235	0.0222	0.0084	0.0403	**0.0242**	0.0144	**0.0678**	**0.0361**	**0.0236**	0.0314	**0.0329**
Category-Level Consistency (CC) ↓															
Ground truth	0.0980	0.1460	0.0578	0.0733	0.1191	0.0955	0.0536	0.2147	0.1088	0.0673	0.1709	0.1444	0.0915	0.1108	0.1108
PCA	**0.0976**	**0.1055**	0.0654	**0.0600**	0.1389	0.0937	0.0527	0.1802	**0.0970**	0.0731	**0.1397**	0.1479	**0.0816**	0.1026	0.1026
CaCa [45]	0.1134	0.1742	0.0730	0.1033	0.1220	0.1919	**0.0493**	0.1888	0.1186	0.0684	0.1840	0.1660	0.0883	0.1262	0.1132
Compass [42]	0.1654	0.1348	0.1077	0.0931	0.1522	0.1175	0.1258	0.1833	0.1266	0.1019	0.1579	0.1626	0.0942	0.1325	0.1283
Ours (F)	0.1043	0.1067	**0.0575**	0.0612	**0.1135**	**0.0869**	0.0525	**0.1754**	0.0988	**0.0681**	0.1504	0.1475	0.0851	**0.1006**	0.1035
Ours (F+P)	0.1250	0.1065	0.0581	0.0635	0.1145	0.0874	0.0500	0.1844	0.1001	0.0679	0.1477	**0.1432**	0.0912	0.1030	**0.1005**
Ground Truth Consistency (GC) ↓															
PCA	0.0760	0.1047	**0.0208**	**0.0390**	0.1190	0.0799	0.0261	0.1366	0.0862	0.0460	0.1280	0.1267	0.0645	0.0810	**0.0810**
CaCa [45]	0.0761	**0.0688**	0.0529	0.0667	0.0943	0.1812	0.0330	0.1592	0.0897	0.0266	**0.0744**	0.1401	0.0683	0.0870	0.1060
Compass [42]	0.1599	0.1586	0.0892	0.0851	0.1504	0.1160	0.1214	0.1654	0.1231	0.0975	0.1552	0.1554	0.0804	0.1275	0.1247
Ours (F)	**0.0671**	0.1131	0.0257	0.0511	0.0526	0.0585	0.0359	0.1399	0.0674	**0.0255**	0.1505	0.0779	0.0746	0.0723	0.0902
Ours (F+P)	0.1115	0.1134	0.0230	0.0553	**0.0509**	**0.0537**	**0.0223**	**0.1274**	**0.0650**	0.0286	0.1456	**0.0738**	**0.0477**	**0.0706**	0.0843

Results: Partial Shapes

	bench	cabinet	car	cellph.	chair	couch	firearm	lamp	monitor	plane	speaker	table	water.	avg.	multi
Ground Truth Consistency (GC)↓															
PCA	**0.0916**	0.1391	0.0727	0.0879	0.1337	0.0908	0.0371	0.1985	0.0804	0.0915	0.1479	0.1087	0.1021	0.1063	0.1063
Compass*	0.1917	0.1412	0.1020	0.1066	0.1476	0.1115	0.1538	0.1735	0.1194	0.1115	0.1617	0.1709	**0.0737**	0.1358	0.1423
Ours(F+P)	0.1416	**0.1182**	**0.0356**	**0.0685**	**0.0780**	**0.0593**	**0.0300**	**0.1501**	**0.0692**	**0.0360**	**0.1469**	**0.0662**	0.0739	**0.0826**	**0.1016**
Instance-Level Consistency (IC)↓															
PCA	**0.1033**	0.1140	0.1149	0.0828	0.1475	0.1221	0.0517	0.1571	0.0867	0.1000	0.1182	0.1401	0.0756	0.1088	0.1088
Compass*	0.1900	0.0790	0.1183	0.0911	0.1280	0.1053	0.1440	0.1000	0.0836	0.1000	0.1134	0.1080	0.0487	0.1084	0.1247
Ours(F+P)	0.1432	**0.0501**	**0.0349**	**0.0442**	**0.0622**	**0.0478**	**0.0221**	**0.0891**	**0.0442**	**0.0265**	**0.1086**	**0.0739**	**0.0469**	**0.0611**	**0.0792**
Category-Level Consistency (CC)↓															
PCA	**0.1269**	0.1500	0.1253	0.1081	0.1636	0.1367	0.0691	0.2312	0.1178	0.1124	0.1677	0.1769	0.1078	0.1380	0.1380
Compass*	0.2118	0.1300	0.1438	0.1215	0.1612	0.1280	0.1688	**0.1990**	0.1242	0.1255	0.1760	0.1719	**0.0919**	0.1503	0.1647
Ours (F+P)	0.1695	**0.1109**	**0.0632**	**0.0739**	**0.1270**	**0.0935**	**0.0546**	0.2048	**0.1042**	**0.0713**	**0.1666**	**0.1579**	0.0936	**0.1147**	**0.1234**

Thank you.
Questions?