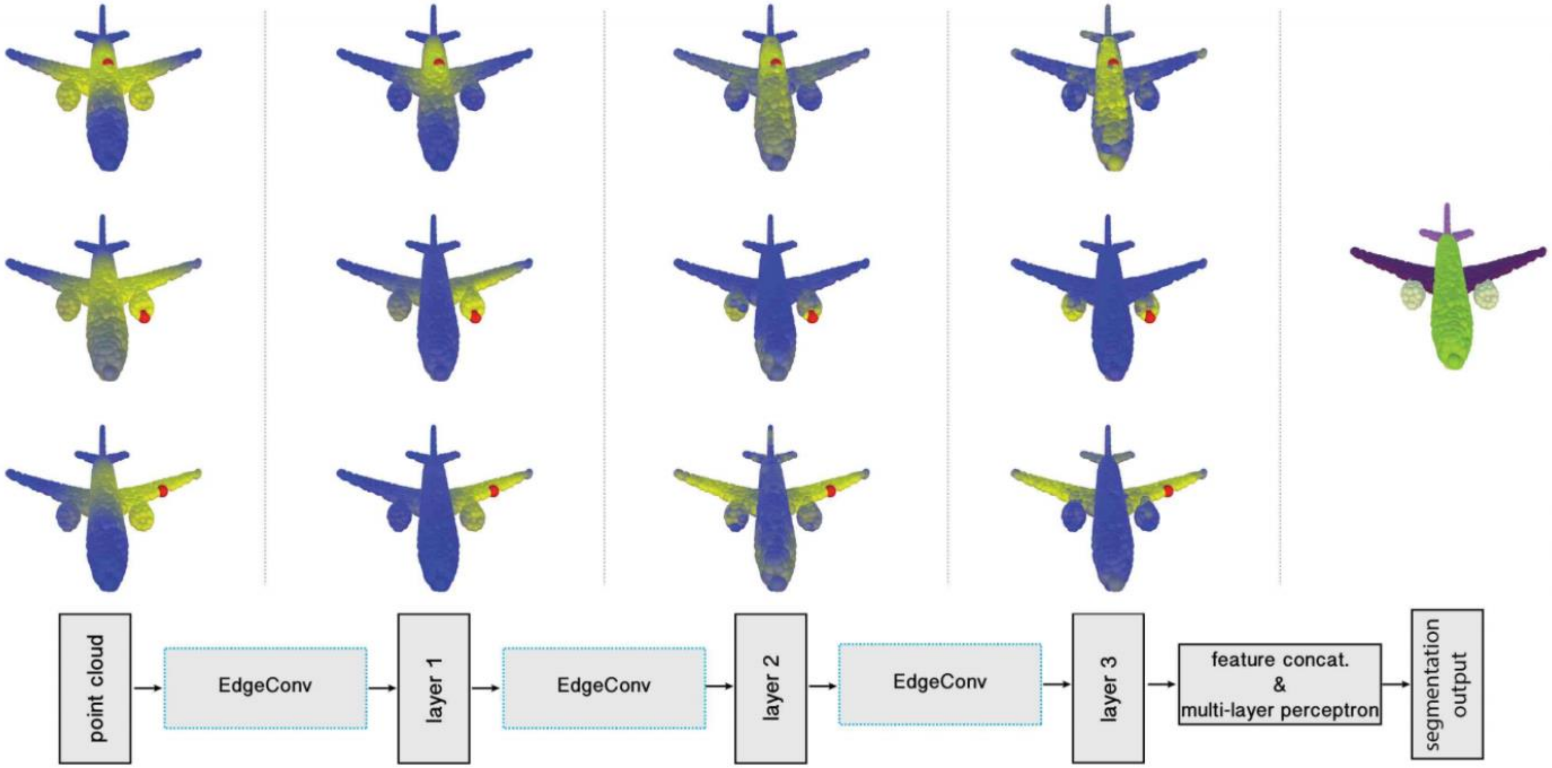
Deep Closest Point: Learning Representation for Point Cloud Registration – ICCV 2019

MINCHANG KIM
제어자동화특강
Seoul National University

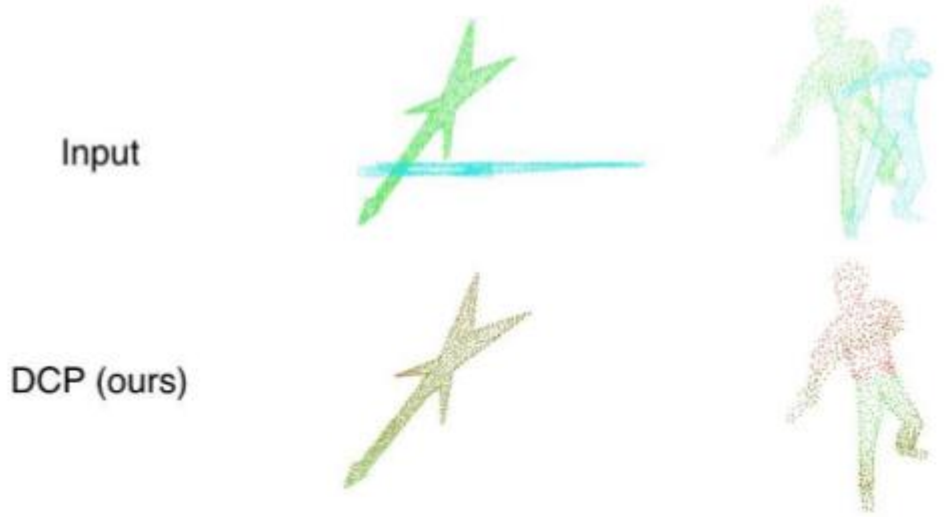
DGCNN (Dynamic graph CNN for learning on point clouds), 2019

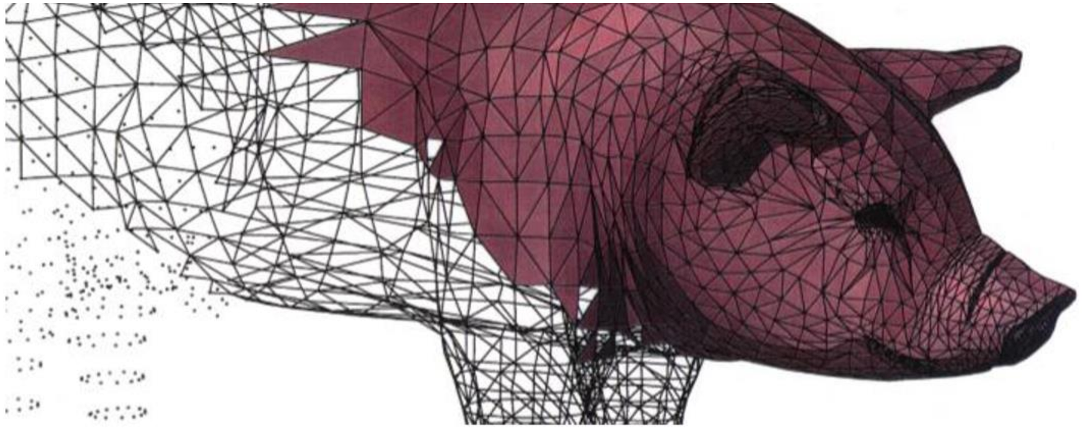
- CNN based high-level task on point cloud
- State-of-the-art on point cloud category classification, semantic segmentation and part segmentation



Registraion?

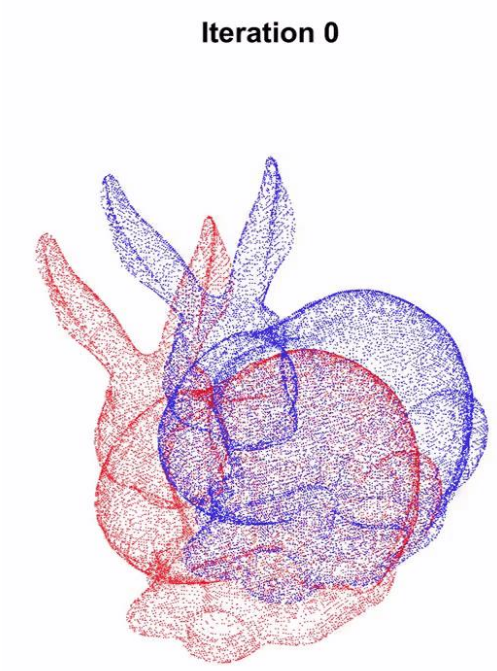
- Process of transforming different sets of data into one coordinate system
- Surface data can be represented by Graph structure
  - Each vertex correspond to each node and edge to edge
- Goal is Finding accurate correspondence between two data





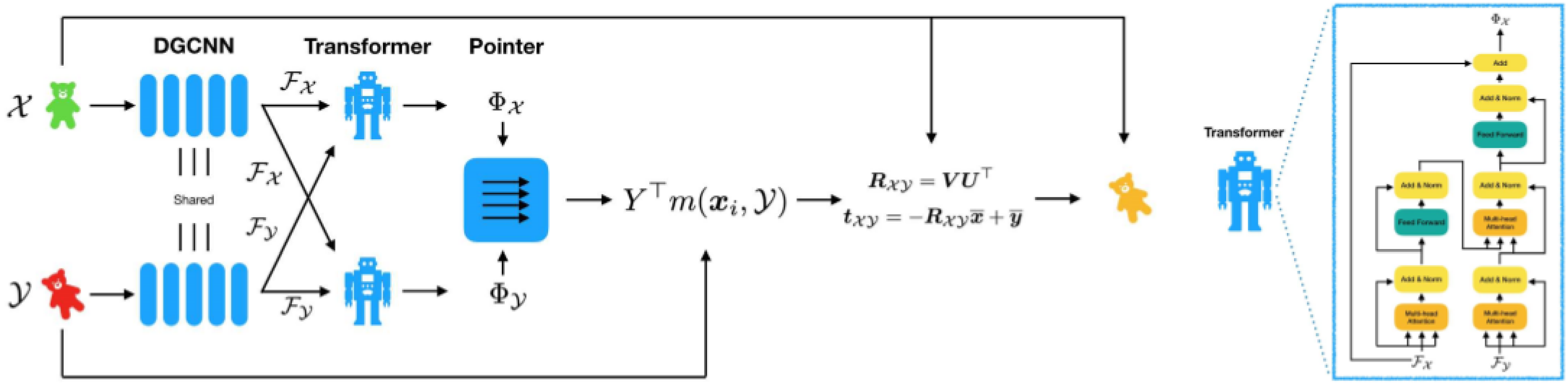
ICP(iterative closest point)

- Set correspondence point as closest point
- Solve problem minimize $$E(\boldsymbol{R}_{\mathcal{XY}}, \boldsymbol{t}_{\mathcal{XY}}) = \frac{1}{N}\sum_{i}^{N} \|\boldsymbol{R}_{\mathcal{XY}}\boldsymbol{x}_i + \boldsymbol{t}_{\mathcal{XY}} - \boldsymbol{y}_i\|^2.$$
- Above problem Can be solved closed form by SVD of cross-covariance matrix
- Easy to fall in local minima if two point cloud far from each other



Deep Closest Point

- Identify sub-network architectures designed to address difficulties in the classical ICP pipeline
- Propose a simple architecture to predict a rigid transformation aligning two point clouds



(a) Network architecture (b) Transformer module

Figure 2. Network architecture for DCP, including the Transformer module for DCP-v2.

Deep Closest Point

- Initial Features
  - Use PointNet and DGCNN
  - PointNet use multilayer perceptron
    - Each vertex are independent

$$x_i^l = h_\theta^l(x_i^{l-1})$$

  - DGCNN use GCN
    - Can more incorporate local neighborhood information than PointNet

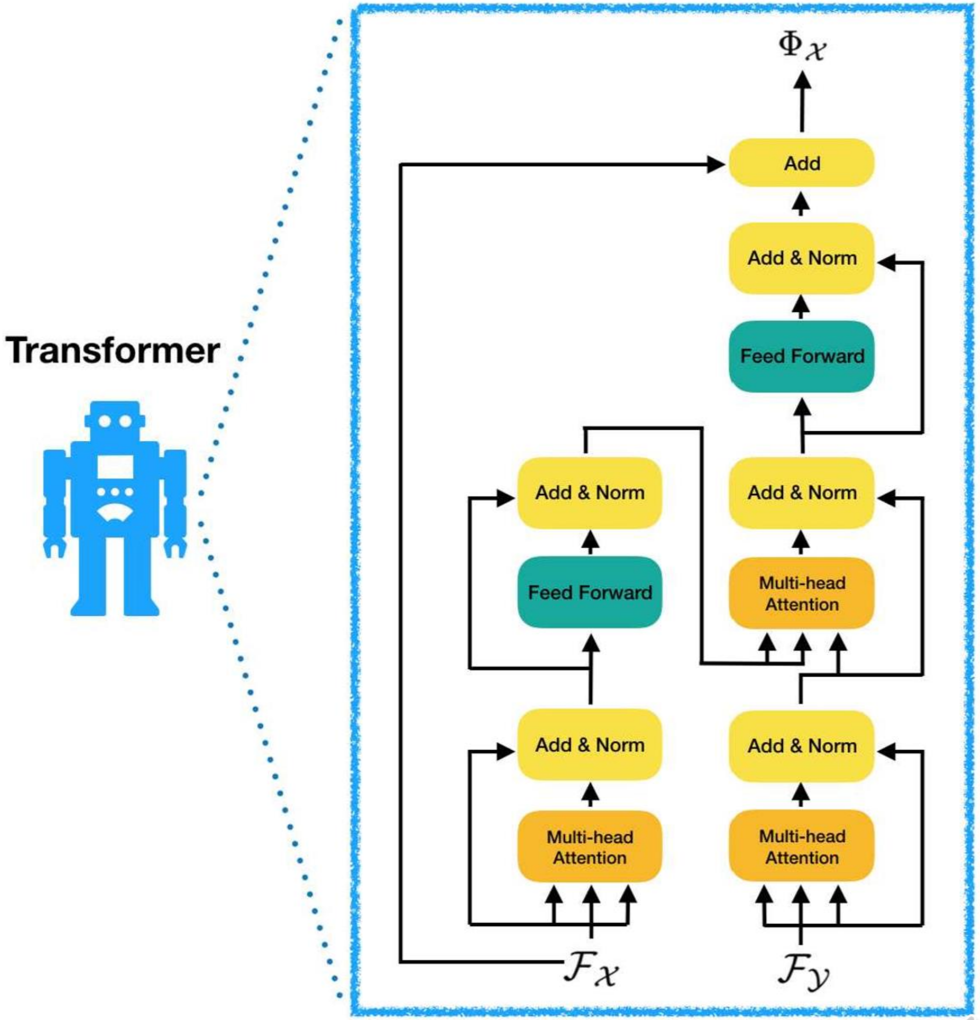
$$x_i^l = f(\{h_\theta^l(x_i^{l-1}, x_j^{l-1})\ \forall j \in \mathcal{N}_i\})$$

Deep Closest Point

- Attention(Transfomer[Ashish 2017])
  - Sequence-to-sequence function
  - Stacked encoder-decoder later
  - Can learn both sequence's information

  - Residual term Depending on the order of inputs
  - Modifies features associated to the points in X in a fashion that is knowledge about structure Y

$$\Phi_{\mathcal{X}} = \mathcal{F}_{\mathcal{X}} + \phi(\mathcal{F}_{\mathcal{X}}, \mathcal{F}_{\mathcal{Y}})$$
$$\Phi_{\mathcal{Y}} = \mathcal{F}_{\mathcal{Y}} + \phi(\mathcal{F}_{\mathcal{Y}}, \mathcal{F}_{\mathcal{X}})$$



Deep Closest Point

- Pointer Generation
  - $m(x_i, Y)$ can be though as a soft pointer from each x_i into the elements of Y
  - Φ_y is embedding of Y

$$m(\boldsymbol{x}_i, \mathcal{Y}) = \mathrm{softmax}(\Phi_{\mathcal{Y}} \Phi_{\boldsymbol{x}_i}^\top)$$

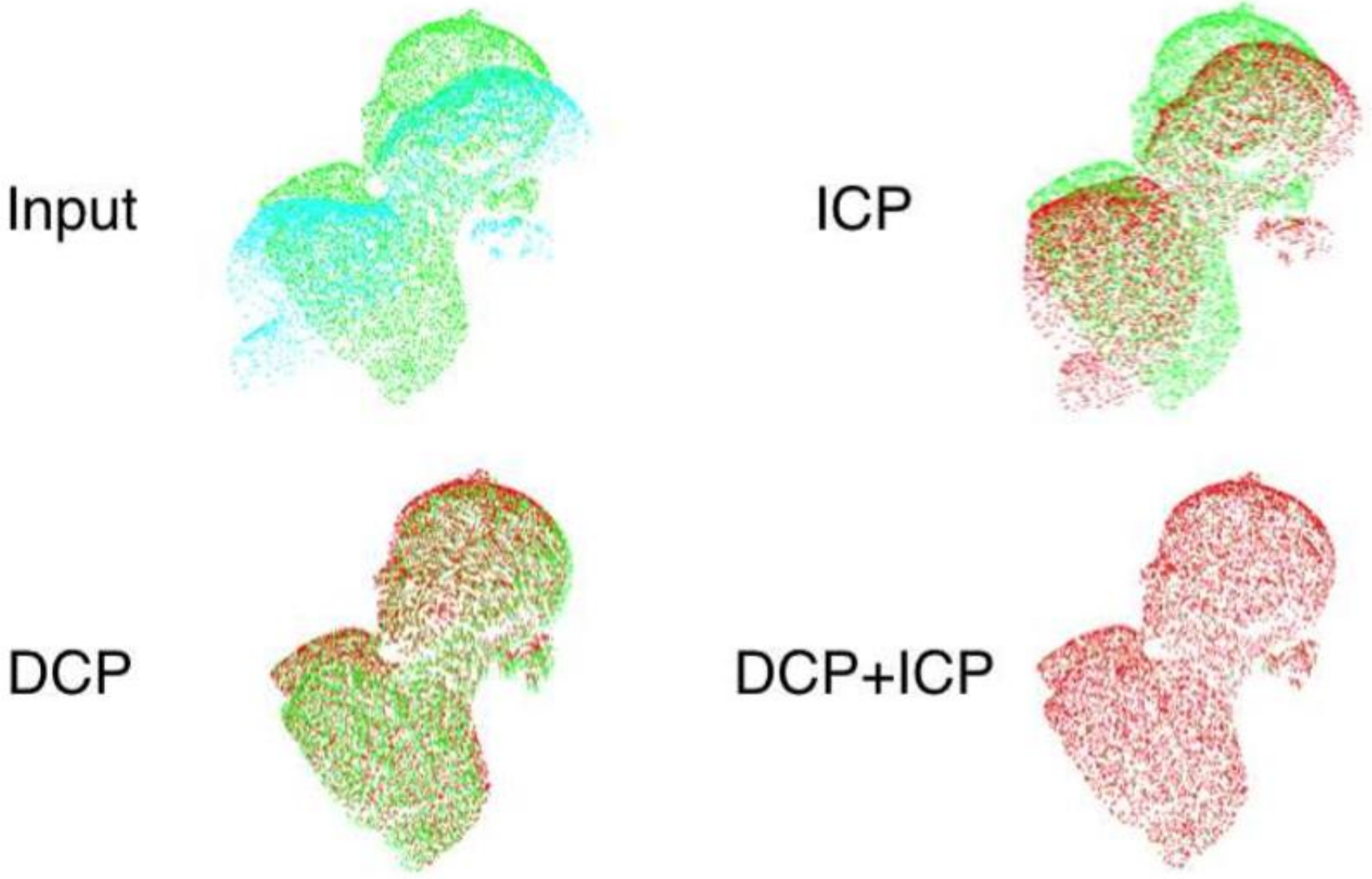
  - Use the soft pointers to generate a matching averaged point in Y for each point in X

$$\hat{\boldsymbol{y}}_i = Y^\top m(\boldsymbol{x}_i, \mathcal{Y}) \in \mathbb{R}^3$$

- Loss

$$\mathrm{Loss} = \|\boldsymbol{R}_{\mathcal{XY}}^\top \boldsymbol{R}_{\mathcal{XY}}^g - I\|^2 + \|\boldsymbol{t}_{\mathcal{XY}} - \boldsymbol{t}_{\mathcal{XY}}^g\|^2 + \lambda\|\theta\|^2$$

Experiment Result



Experiment Result

Model	MSE(R)	RMSE(R)	MAE(R)	MSE(t)	RMSE(t)	MAE(t)
ICP	894.897339	29.914835	23.544817	0.084643	0.290935	0.248755
Go-ICP [55]	140.477325	11.852313	2.588463	0.000659	0.025665	0.007092
FGR [61]	87.661491	9.362772	1.999290	0.000194	0.013939	0.002839
PointNetLK [18]	227.870331	15.095374	4.225304	0.000487	0.022065	0.005404
DCP-v1 (ours)	6.480572	2.545697	1.505548	**0.000003**	**0.001763**	**0.001451**
DCP-v2 (ours)	**1.307329**	**1.143385**	**0.770573**	**0.000003**	0.001786	**0.001195**

Table 1. ModelNet40: Test on unseen point clouds

Model	MSE(R)	RMSE(R)	MAE(R)	MSE(t)	RMSE(t)	MAE(t)
ICP	892.601135	29.876431	23.626110	0.086005	0.293266	0.251916
Go-ICP [55]	192.258636	13.865736	2.914169	0.000491	0.022154	0.006219
FGR [61]	97.002747	9.848997	**1.445460**	0.000182	0.013503	0.002231
PointNetLK [18]	306.323975	17.502113	5.280545	0.000784	0.028007	0.007203
DCP-v1 (ours)	19.201385	4.381938	2.680408	**0.000025**	**0.004950**	**0.003597**
DCP-v2 (ours)	**9.923701**	**3.150191**	2.007210	**0.000025**	0.005039	0.003703

Table 2. ModelNet40: Test on unseen categories

V1 is structure without attention

Experiment Result

Metrics	PN+DCP-v1,	DGCNN+DCP-v1	PN+DCP-v2	DGCNN+DCP-v2
MSE($\boldsymbol{R}$)	17.008427	6.480572	49.863022	1.307329
RMSE($\boldsymbol{R}$)	4.124127	2.545697	7.061375	1.143385
MAE($\boldsymbol{R}$)	2.800184	1.505548	4.485052	0.770573
MSE($\boldsymbol{t}$)	0.000697	0.000003	0.000258	0.000003
RMSE($\boldsymbol{t}$)	0.026409	0.001763	0.016051	0.001786
MAE($\boldsymbol{t}$)	0.01327	0.001451	0.010546	0.001195

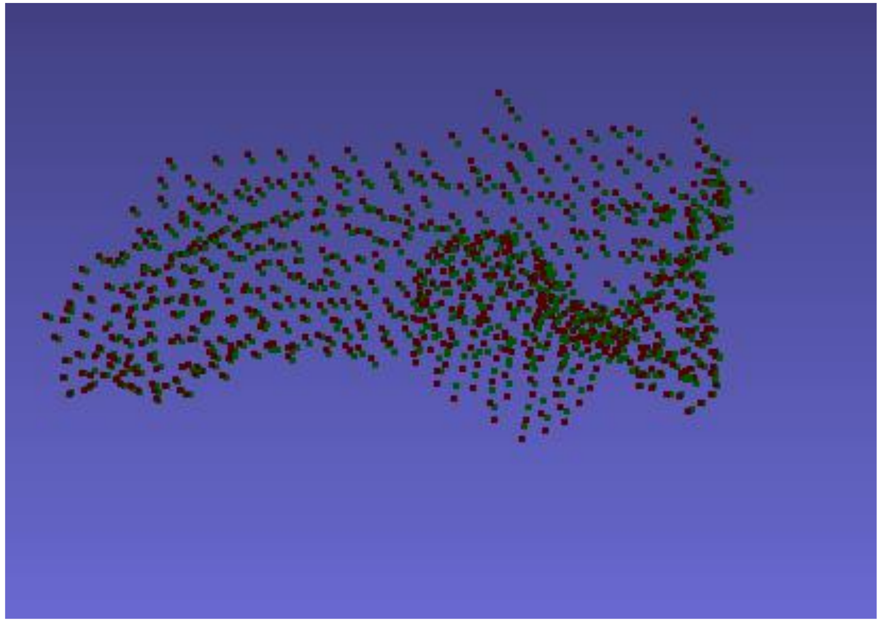
Table 5. Ablation study: PointNet or DGCNN?

Metrics	DCP-v1+MLP	DCP-v1+SVD	DCP-v2+MLP	DCP-v2+SVD
MSE($\boldsymbol{R}$)	21.115917	6.480572	9.923701	1.307329
RMSE($\boldsymbol{R}$)	4.595206	2.545697	3.150191	1.143385
MAE($\boldsymbol{R}$)	3.291298	1.505548	2.007210	0.770573
MSE($\boldsymbol{t}$)	0.000861	0.000003	0.000025	0.000003
RMSE($\boldsymbol{t}$)	0.029343	0.001763	0.005039	0.001786
MAE($\boldsymbol{t}$)	0.022501	0.001451	0.003703	0.001195

Table 6. Ablation study: MLP or SVD?

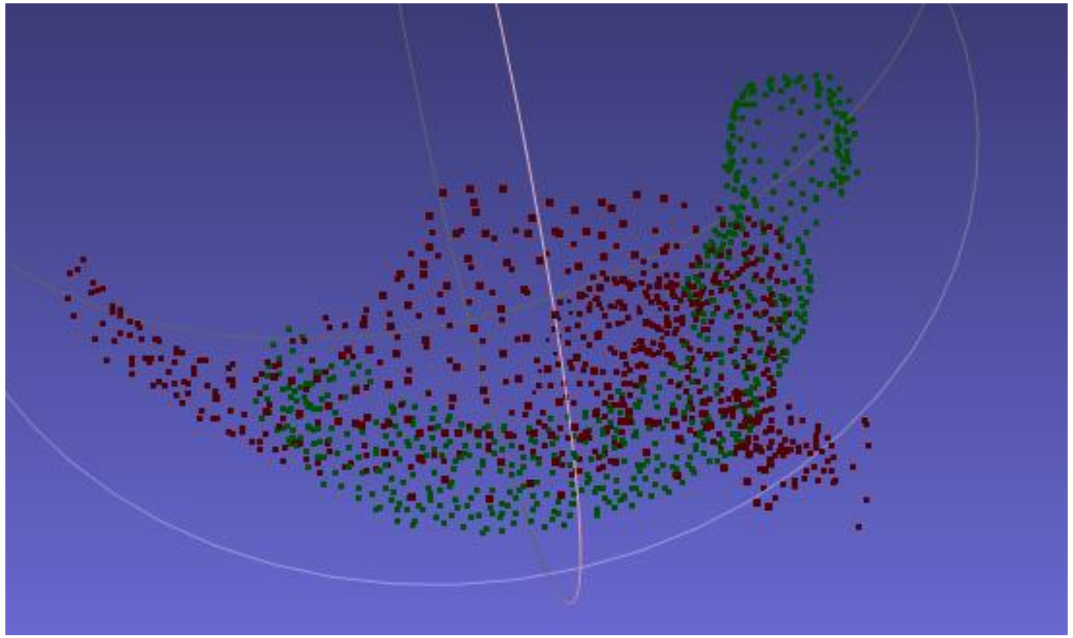
V1 is structure without attention

Experiment Result

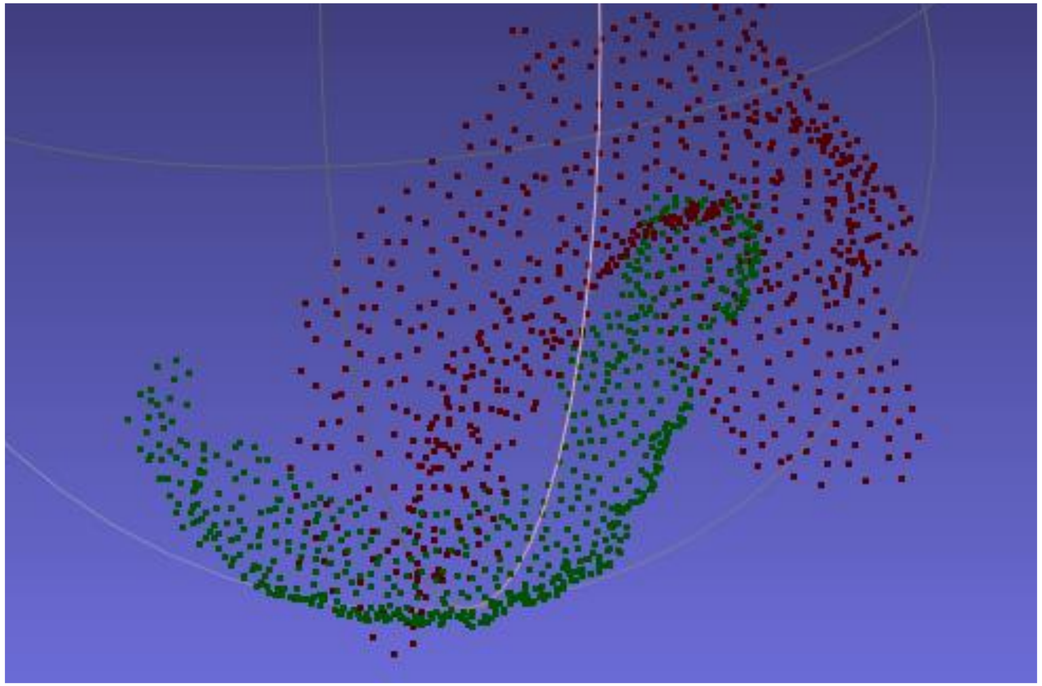


Experiment Result

Modelnet40 trained



Additional trained with our data



V1 is structure ithout attention

Thank you