# Relation-Shape Convolutional Neural Network for Point Cloud Analysis Supplementary Material

## A. Outline

This supplementary material provides further investigations for the proposed RS-CNN. Specifically, three issues on the construction of local neighborhood are discussed in Sec B. More details of the relation learning on 2D views of 3D point cloud are presented in Sec C. All the experiments are conducted on ModelNet40 dataset.

## B. Construction of Local Neighborhood

In the main paper, the local point subset $P_{\text{sub}}$ in Eq. (3) is modeled to be a spherical neighborhood with a sampled point $x_i$ as the centroid, and the surrounding points as its neighbors $\mathcal{N}(x_i)$ (see the left-most part in Fig. 2). Then, the inductive representation $\mathbf{f}_{P_{\text{sub}}}$, which is expected to reason the spatial layout of points in this neighborhood, is obtained by performing the proposed relation-shape convolution to aggregate all the relation between $x_i$ and $\mathcal{N}(x_i)$.

In the above process, there are mainly three issues worth further investigation: (1) How should $\mathcal{N}(x_i)$ be selected? (2) Is it suitable to simply aggregate all the relation between $x_i$ and $\mathcal{N}(x_i)$? (3) Is it reasonable to select the sampled point $x_i$ as the centroid? They are explored as follows.

**(1) Selection of the neighbors $\mathcal{N}(x_i)$.** Two strategies, k-nearest neighbor (k-NN) and random picking in the ball (Random-PIB), are investigated for this issue. Table 1 summarizes the results. Note that the number of neighbors is set to be equal for a fair comparison. As it shows, Random-PIB obtains better classification accuracy. The reason may be k-NN would suffer selection inhomogeneity in some cases, which is adverse to shape-aware learning (the aggregated relation may only focus on dense points and ignore sparse points that are essential for the underlying shape). By contrast, Random-PIB can have a better coverage of points even in the case of inhomogeneous distribution.

**(2) Relation aggregation issue.** To verify this issue, we randomly cut off some relation between $x_i$ and $\mathcal{N}(x_i)$ during training, *i.e.*, randomly setting the learned high-level relation expression $\mathcal{M}(\mathbf{h}_{ij})$ in Eq. (3) to be a zero vector, but using all the relation during testing. This operation is similar to the dropout technique. Table 2 summarizes the results. As can be seen, the best approach is training with all the relation while the second best one is training with relation cut ratio of 0.3. This indicates the dropout-like technique is not suitable for relation learning, probably because RS-CNN can automatically encode the strength of the relation in the learning process.

Table 1. The results (%) of two selection strategies on $\mathcal{N}(x_i)$. Both of them are trained with a single-scale neighborhood. For a fair comparison, the number of neighbors is set to be equal in each layer between the two models.

| method | acc. |
|---|---|
| k-NN | 90.5 |
| Random-PIB | 92.2 |

Table 2. The results (%) of learning with relation in different proportions. "ratio" indicates the cut off relation accounts for the proportion of all the relation between the centroid and the neighbors.

| ratio | 0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
|---|---|---|---|---|---|---|
| acc. | 93.6 | 92.8 | 92.9 | 93.2 | 92.5 | 92.1 |

Table 3. The results (%) of three selection approaches and one fusion strategy of the centroid. The approach of picking in $\mathcal{N}(x_i)$ is performed randomly in each neighborhood. Note that the weight in $\mathcal{M}$ is shared over these approaches in the fusion process.

| centroid | acc. |
|---|---|
| sampled point $x_i$ | 93.6 |
| average of $\mathcal{N}(x_i)$ | 93.6 |
| random picking in $\mathcal{N}(x_i)$ | 92.8 |
| fusion of above | 93.4 |

**(3) Selection of the centroid.** Three types of the centroid: the sampled point $x_i$, the average of $\mathcal{N}(x_i)$ and random picking in $\mathcal{N}(x_i)$, are studied for this issue. Besides, a strategy that fuses all of them is also studied. The results are summarized in Table 3, where the first two strategies obtain the same decent accuracy while random picking performs less well. The reason may be that random picking requires RS-CNN to reason the spatial layout of points from various topological connections, which is quite difficult.

Another promising strategy is fusing a group of relations that are centered on different centroids. This can be achieved by performing element-wise summation of $\mathbf{f}_{P_{\text{sub}}}$ in Eq. (3), with the relation centered on the above three kinds

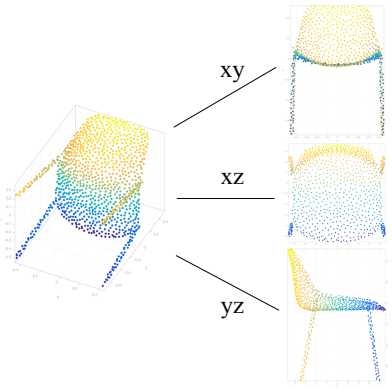


Figure 1. The projection of 3D point cloud onto the 2D plane of XY, XZ and YZ.

Table 4. The results (%) of RS-CNN with the low-level relation $\mathbf{h}$ defined on 2D views (XY-Ed: Euclidean distance in XY plane, $x^{\text{xy}}$: 2D coordinates of $x$ in XY plane, *i.e.*, the value of z is set to be zero). The fusion strategy is achieved by performing element-wise summation of $\mathbf{f}_{P_{\text{sub}}}$ in Eq. (3), with $\mathbf{h}$ defined on three 2D views. Note that the weight in $\mathcal{M}$ is shared over these three views in the fusion process.

| low-level relation $\mathbf{h}$ | channels | acc. |
|---|---|---|
| (XY-Ed, $x_i^{\text{xy}} - x_j^{\text{xy}}$, $x_i^{\text{xy}}$, $x_j^{\text{xy}}$) | 10 | 92.1 |
| (XY-Ed, $x_i^{\text{xz}} - x_j^{\text{xz}}$, $x_i^{\text{xz}}$, $x_j^{\text{xz}}$) | 10 | 92.1 |
| (XY-Ed, $x_i^{\text{yz}} - x_j^{\text{yz}}$, $x_i^{\text{yz}}$, $x_j^{\text{yz}}$) | 10 | 92.2 |
| fusion of above three views | | 92.5 |

of centroids. However, it does not perform better, with an accuracy of 93.4% that is lower than the best single-centroid version of 93.6%.

## C. Low-Level Relation h

More details of the relation learning on 2D views of point cloud (the fourth part in Sec 4.2) are provided in this section. As illustrated in Fig. 1 in this material, the relation among points in the 2D view can also reflect the underlying shape. Therefore, we are interested in how powerfully the proposed RS-CNN to acquire shape awareness from only 2D-view relation of points.

To validate this, the value of one dimension in 3D coordinates is forcibly set to be zero, that is, 3D points are projected onto the 2D plane of XY, XZ and YZ for three 2D views. In addition, a strategy with fusion of these views is also studied. Note that the projection operation is only conducted for the definition of $\mathbf{h}$, the initial input features for $x_j$ in Eq. (3) is still intact 3D coordinates. Table 4 summarizes the results. As can be seen, all single-view relation can achieve an accuracy around 92.2%, which is quite impressive. After fusing them, the result is improved by 0.3%. This shows RS-CNN can also capture the underlying shape well even with relation learning from 2D view (potentially, a group of 2D views) of 3D point cloud, further verifying its effectiveness.