Point Cloud Overlapping Region Co-Segmentation Network

Kexue Fu[*], **Xiaoyuan Luo**[*] and **Manning Wang**[†] MNWANG@FUDAN.EDU.CN
Digital Medical Research Center, School of Basic Medical Science, Fudan University
Shanghai Key Laboratory of Medical Image Computing and Computer Assisted Intervention

Abstract

3D point clouds are being increasingly used in the field of computer vision and many applications involve the processing of partially overlapping point clouds. However, little attention has been paid to the property of partial overlap. In this paper, we propose the concept of co-segmentation of the overlapping region of two 3D point clouds and develop a deep neural network to solve this problem. The proposed network utilizes co-attention mechanism to aggregate information from the paring point clouds so as to find the overlapping region. The co-segmentation of overlapping region can be regarded as a preprocessing step in practical 3D point cloud processing pipelines so that downstream tasks can be better accomplished. We build a dataset of partially overlapping 3D point clouds from ModelNet40 and ShapeNet, which are two widely used 3D point cloud datasets, and the overlapping region can be obtained automatically without manual labelling. We also utilize the real 3D point cloud datasets, 3DMatch and ScanNet, in which the overlapping region can be obtained from the relative pose between point clouds provided in the datasets. We evaluate the performance of the proposed method on co-segmentation of overlapping region on these datasets and its effectiveness in improving one downstream task, 3D point cloud registration, which is very sensitive to partial overlapping

Keywords: Co-segmentation, Point Clouds

1. Introduction

3D point cloud is a common data structure to model 3D objects and is being increasingly used in computer vision, such as for 3D registration Gojcic et al. (2020); Wang and Solomon (2019); Yew and Lee (2020), 3D model reconstruction Chen et al. (2019); Mandikal and Radhakrishnan (2019) and 3D object detection Chen et al. (2020); Qi et al. (2019).Many applications involve processing partially overlapping point clouds, such as point clouds obtained by the lidar of an autonomous driving system or a SLAM system, and partially visible point clouds in 3D model reconstruction. However, existing deep learning methods for point cloud processing do not fully consider the characteristics of partially overlapping point clouds. In this article, we first propose the concept of point cloud co-segmentation, which is to segment the overlapping region of two partially overlapping point clouds, and develop a deep learning method to solve this problem.

The co-segmentation of overlapping region can be regarded as a pre-processing in point cloud processing, which can reduce the difficulty of downstream tasks. For example, it is

[*] These authors have contributed equally to this work.
[†] Corresponding author

easier to register two completely overlapping point clouds than to register two partially overlapping point clouds, and the accuracy and robustness of 3D point cloud registration increase as the overlapping ratio increases. Therefore, if we segment out the overlapping region of the two point clouds to be registered, the registration accuracy and robustness would be further improved. By using co-segmentation of overlapping region, difficult point cloud processing problems would be decomposed into two relatively easy-to-solve sub-problems, so as to be better solved.

The concept of co-segmentation has been proposed in 2D image processing, and there are a lot of research and applications on this topic Banerjee et al. (2019); Zhang et al. (2020); Lu et al. (2019). In 2D image processing, co-segmentation aims to segment common foreground or similar objects. However, because of the difference of data structure between 2D images and 3D point clouds, the methods for 2D image co-segmentation cannot be directly extended to be used on point clouds. At the same time, there are many studies on point cloud segmentation, such as Qi et al. (2017b); Tchapmi et al. (2017); Wang et al. (2018), but they all focus on the segmentation of a single point cloud. There is one latest study solving the segmentation problem of multiple point clouds Zhu et al. (2020), but it deals with part segmentation of similar objects instead of segmenting overlapping region. To the best of our knowledge, there is no research about overlapping region co-segmentation for partially overlapping point clouds.

In this paper, we propose a 3D Point cloud Overlapping region Co-Segmentation Network (POCS-Net), which is composed of a local feature extractor, a co-attention module and a segmentation module. The local feature extractor extracts rotation-invariant local features for each point. The co-attention module takes the local features of each point as input, and integrates the local features of the two point clouds by using co-attention mechanism to obtain the co-attention enhanced features. This module makes the network pay more attention to overlapping region and produces features that better distinguish overlapping and non-overlapping regions. In the segmentation module, we use fully connected network to predict the probability that each point belongs to the overlapping region. POCS-Net is trained to maximizes the similarity of the local features of the overlapping region and encourage producing a larger overlapping region. In addition, we adopt two strategies to prevent the local point features from degeneration. First, we add a regulation term to the loss function to make the features tend to distribute evenly on a sphere. Second, we make POCS-Net share the encoder with a self-supervised task. Finally, we use the ModelNet40 Wu et al. (2015) and ShapeNet Chang et al. (2015) datasets to to construct a new dataset consisting of partially overlapping point clouds to evaluate the accuracy and robustness of the point cloud co-segmentation network, and demonstrate if the co-segmentation of overlapping region is beneficial for one downstream task, 3D point cloud registration. Evaluation is also conducted on real datasets 3DMatch and ScanNet.

Contribution:

- We propose the concept of overlapping region co-segmentation in point cloud processing, and develop an end-to-end network based on co-attention mechanism for this task.
- We propose a loss function based on metrics defined on overlapping region, which eliminates the dependence on manually labeled data. This enables the network pro-

posed in this paper to be trained on a large amount of data. We explore a regulation term in the loss function and simultaneously train POCS-Net with a self-supervised learning network to prevent the features utlized in POSC-Net from degeneration.

- We evaluate the performance of POCS-Net on partially overlapping point clouds constructed from four representative point cloud data sets and demonstrate its effectiveness in improving the downstream task of 3D point cloud registration.

2. Related work

2.1. Image Co-Segmentation

Image co-segmentation is a task of segmenting the common objects from a set of images. The conventional co-segmentation methods are difficult to segment objects with large variability in terms of scale, appearance, pose, viewpoint and background Rother et al. (2006). Thanks to the development of deep learning, co-segmentation methods based on deep neural network have been widely studied and have achieved better results than traditional methods. Li et al. Li et al. (2018)proposed co-segmentation method based on shared weight encoder-decoder network and mutual correlation layer. Banerjee et al. Banerjee et al. (2019) proposed using a metric learning network to find the optimal feature space which minimizes the distance between features extracted among similar objects. Zhang et al. Zhang et al. (2020) used a multi-scale convolution network to deal with objects in different resolution and fuse the features among different images in spatial and semantic level to perform co-segmentation. Co-attention mechanism has been widely applied in visual question answering Lu et al. (2016); Wu et al. (2018b); Xiong et al. (2016) and hashtag recommendation field Zhang et al. (2017); Li et al. (2019). Recently, co-attention has also been introduced into computer vision field. Lu et al. Lu et al. (2019) proposed a collaborative attention Siamese network to segment common objects from a video. Even though many image co-segmentation methods have been developed, these methods cannot be directly applied in 3D point clouds because of the essential difference in data structure between 2D images and 3D point clouds.

2.2. Deep Learning Point Cloud Segmentation

The research of point cloud segmentation based on deep learning can be divided into three categories: multiview-based methods, voxel-based methods, and PointNet-like methods. The multiview-based methods can be seen as an extension of image segmentation Wang et al. (2018); Wu et al. (2018a, 2019). They use multi-view projections to represent 3D point clouds and use 2D CNN to segment the multi-view projections. Finally, the 3D segmentation result is recovered from the 2D CNN segmentation results, such as SnapNet Riegler et al. (2017). Although the multiview approach can solve the problem of irregular point cloud data structure, it causes loss of geometric structure. At the same time, it is difficult to choose suitable views to cover the entire complex scene. Therefore, multiview-based methods are now rarely used for point cloud segmentation. Since point clouds are not in a regular format, many researchers transform such data to regular 3D voxel grids before feeding them to a 3D convolution network, such as SegCloud Tchapmi et al. (2017). The limitations of these methods are also obvious. Voxelization causes the loss of local details and high memory costs. Researches in this category mainly focus on efficiency issues, such

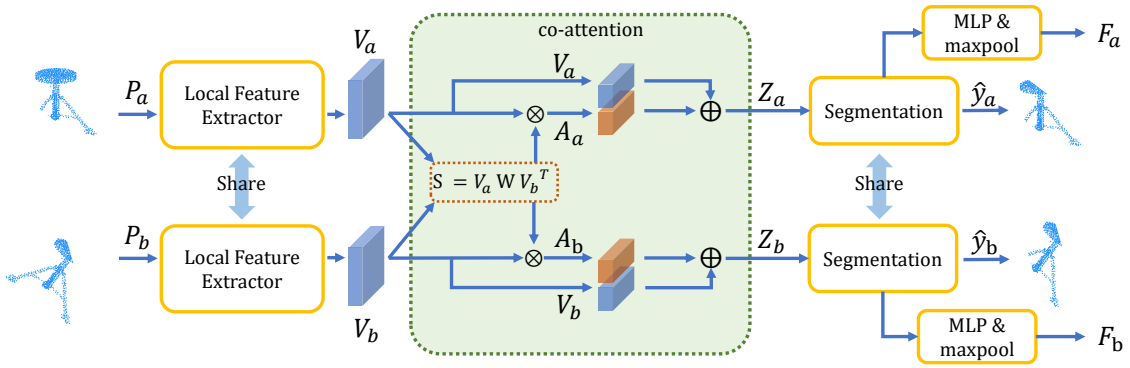

Figure 1: The architecture of the proposed end-to-end network for 3D Point Cloud overlapping region Co-Segmentation, POCS-Net.

as OctNet Riegler et al. (2017), sparse 3D convolution Graham et al. (2018), etc. The PointNet-like methods are the most popular apporach in point cloud segmentation, and these methods originate from the seminal work of PointNet Qi et al. (2017a), which uses MLP to directly process point cloud data. PointNet utilizes the local and global features obtained by MLP to segment the point cloud, but it cannot capture the local geometric structure very well and doesn't have the property of rotation and translation invariance. A series of following researches have been done to achieve better point cloud segmentaiton, such as PointNet++ Qi et al. (2017b), DGCNN Phan et al. (2018), etc. Although there are many researches on 3D point cloud segmentation, they only focus on the segmentation of a single point cloud. A recent study AdaCoSeg Zhu et al. (2020) solves the problem of simultaneous segmentation of multiple point clouds, but it focuses on the point cloud part segmentation of similar objects, and it is not suitable for the overlapping region co-segmentation problem.

3. Methodology

The architecture of the proposed POCS-Net is shown in Fig.1. The network is composed of shared local feature extractor, co-attention module and shared segmentation network. Two partially overlapping point clouds are input to the network at the same time. The local feature extractor extracts features for each point in the two point clouds separately, and then the collaborative attention module enhances the features of each point by fusing both the information from neighboring points in the same point cloud and the information of the points from the other point cloud. Finally, the segmentation network process each point cloud and output the overlapping region. The outputs of local features F_a,F_b are used to calculate the loss.

3.1. Local Feature Extractor

We build two local feature extractor networks with shared parameters to extract local feature of each point in point cloud $P_a \in \mathbb{R}^{\mathrm{M}\times 3}$ and $P_b \in \mathbb{R}^{\mathrm{N}\times 3}$, where M and N represent

the number of points in each point cloud. The network uses the DGCNN structure, and extracts the local features of each point layer by layer from the M/Nx3 input through three EdgeConv layers. In each EdgeConv layer, each point its K-nearest neighbors are grouped to form a directed graph, and then the feature of each edge in the local directed graph is extracted through a fully connected network with shared parameters. Finally, the edge features are convolved with the corresponding vertex features to obtain the local features. In order to enable the extracted features to be invariant to rotation and translation, we replace the original vertex feature of local coordinate with the following PPF features Deng et al. (2018):

$$\mathrm{PPF}\left(\mathbf{x}_c, \mathbf{x}_i\right) = \left(\angle\left(\mathbf{n}_c, \Delta\mathbf{x}_{c,i}\right), \angle\left(\mathbf{n}_i, \Delta\mathbf{x}_{c,i}\right), \angle\left(\mathbf{n}_c, \mathbf{n}_i\right), \left\|\Delta\mathbf{x}_{c,i}\right\|_2\right), \tag{1}$$

$$\Delta\mathbf{x}_{c,i} = \mathbf{x}_i - \mathbf{x}_c \tag{2}$$

Where $\mathbf{n}_c$ and $\mathbf{n}_i$ are the normal of the central point $\mathbf{x}_c$ and point $\mathbf{x}_i$, $\Delta\mathbf{x}_{c,i}$ denotes the neighboring points translated into a local frame by subtracting away the coordinates of the centroid point. After the local features are extracted from multiple cascade EdgeConv layers, the features are embedded into 1024 dimensions through MLP with shared parameters. The output of local feature extractor are V_a and V_b of size M-by-1024 and N-by-1024, respectively.

3.2. Co-Attention Module

Attention mechanism has been used in the co-segmentation of 2D images and fairly good results have been achieved Lu et al. (2019). Studies in both the NLP Vaswani et al. (2017)and the 2D co-segmentation Lu et al. (2019) field indicate that attention is able to establish reasonable correlation between two objects. As shown in Fig.1, we obtain the local feature V_a and V_b through the local feature extractor. However, our task is to segment the overlapping region of the two point clouds. Therefore, we need to learn the correlation between the local features of the two point clouds, which is achieved by using co-attention mechanism. First, we calculate the similarity matrix $S \in \mathbb{R}^{\mathrm{N}\times\mathrm{M}}$ between V_a and V_b:

$$S \;=\; V_b \; W \; {V_a}^T \tag{3}$$

where W is a weight matrix that can be learned. After obtaining the similarity matrix S , we normalize S row-wise and column-wise with a $softmax$ function:

$$S^c \;=\; softmax(S), \quad S^r \;=\; softmax(S^T) \tag{4}$$

where $softmax(\cdot)$ normalizes each column of the input. Therefore, the co-attention feature A_a of P_a can be calculated by the following formula:

$$A_a = V_b S^c = \left[A_a^1, A_a^2, A_a^3, \ldots, A_a^i, \ldots A_a^M\right] \tag{5}$$

$$A_a^i = V_b \otimes S^{c(i)} = \sum_{j=1}^{\mathrm{N}} V_b^j S_{ij}^c \tag{6}$$

$$Z_a = A_a \oplus V_a \tag{7}$$

where A_a^i represents the feature of the i-th point. $\otimes$ denotes the matrix multiplying vector. $S^{c(i)}$ is the i-th column of S^c and reflects the correlation between i-th feature of V_b and V_a. $\bigoplus$ denotes the concatenation operation. Z_a denotes co-attention enhanced feature. Similarly, we compute co-attention enhanced feature for V_b as: $Z_b = A_b \bigoplus V_b$, where $A_b \;= V_a S^r$.

3.3. Segmentation Module

After fusing the information from P_a and P_b through co-attention module and get Z_a, Z_b, two segmentation networks with shared parameters are applied to segment the overlapping regions in P_a and P_b, respectively. We treat the point cloud segmentation as a per-point classification problem and the segmentation network is composed of MLP with shared parameters. The network input is point features of M/N-by-1024 and the output is a per-point classification result of M/N-by-2, which represents the probability that each point belongs to the overlapping and non-overlapping region.

3.4. Loss Function

First, we propose a loss function to minimize the difference between the segmented overlapping regions in two point clouds. The extracted point local features are multiplied with the corresponding probability predicted by the segmentation network to obtain weighted point local features, which are then embedded and pooled to generate the overlapping region features F_a, F_b through a shared MLP and max pooling. The distance between F_a and F_b is used as the first term of our loss function in Eq.8. The second term in Eq.8 is used to encourage the network to output larger overlapping region:

$$Loss = dist\left(F_a, F_b\right) - \alpha \cdot \left(\sum_{i}^{\mathrm{M}} \hat{y}_{ai} + \sum_{j}^{\mathrm{N}} \hat{y}_{bj}\right) \tag{8}$$

where $dist(\cdot, \cdot)$ represent the distance function in feature space and α is the weight of second term, $\hat{y}_{ai}$ and $\hat{y}_{bj}$ represent the probability of belonging to the overlapping region.

However, simply training the network with the loss in Eq.8 may drive the network to degenerate to output the same feature for all points, which can trivially minimize the loss. To avoid the degeneration, we propose the following two strategies:

Feature distribution restriction : Adding a new regulation term to maximize the distance between point features as in Wang and Isola (2020). When keeping all features on a hypersphere, this loss will make them tend to distribute evenly on the sphere instead of clustering at one point

$$L_{reg} = -\left(\sum_{i}^{M}\sum_{j,j\neq i}^{M} V_{ai}^{T} V_{aj} + \sum_{i}^{N}\sum_{j,j\neq i}^{N} V_{bi}^{T} V_{bj}\right) \tag{9}$$

Self−supervised task : Constructing a reconstruction network and making the reconstruction network share the same feature extractor with our co-segmentation network. The reconstruction network is used to reconstruct P_a and P_b from V_a and V_b so as to prevent the feature extractor from degeneration.

4. Experimental Protocol

4.1. Partially Overlapping Point Cloud Dataset

We evaluated the proposed POCS-Net on both synthetic data and real data. For the synthetic data, we will build a co-segmentation dataset named PointCS, which contains

partially overlapping point clouds. PointCS will be built from ModelNet40 and ShapeNet, which are two widely used 3D point cloud data sets. ModelNet40 contains clean 3D models of 40 categories, and ShapeNet contains 3D models of 55 categories collected from online repositories. First, we will sample 1024 points on the 3D models' surface as Chen et al. (2020). Then we will transform the point clouds by sampling three Eular angle rotations in the range $[-180°, 180°]$ and translations in the range $[-0.05, 0.05]$ on each axis to obtain a series of full point cloud. For a full point cloud, a random plane is used to cut it into two parts and the bigger one is retained as a partial point cloud. The cutting plane can be shifted such that a certain percentage of points are retained. Then for each full point cloud, a series of partial point clouds are generated and they are used as partially overlapping points for training and testing. In this way, the overlapping region of each pair of partial point clouds are known without manual labelling, and the overlapping ratio can also be calculated. For the real data, we choose the publicly available benchmark datasets 3DMatch and ScanNet. The overlapping region of a pair of point clouds in 3DMatch and ScanNet are obtained according to the relative pose between point clouds provided in the datasets. There are official splits of training and test sets in all four datasets.

4.2. Performance of POCS-Net on Point Cloud Co-Segmentation

The following three types of experiments will be conducted to evaluate the performance of POCS-Net on point cloud co-segmentation. Firstly, we will evaluate the accuracy of POCS-Net segmentation under different minimal overlapping ratios. We will build four datasets with different minimal overlapping ratios of 20%, 40%, 60% and 80%. A minimal overlapping ratio of 20% means that the overlapping ratio of all the point cloud pairs to be co-segmented belongs to a uniform distribution in the range of [20%, 100%]. Each point will be added a Gaussian with a mean of 0 and a standard deviation of 0.01. e will use 80% training data to train POCS-Net and use the rest 20% training data for validation. The co-segmentation performance will be evaluated according to the IoU between the network segmentation and the ground truth on the test set. Secondly, we will evaluate the accuracy of POCS-Net segmentation under different noise level. We will build four datasets with 40% minimal overlapping ratio but with four different noise variances of 0.005, 0.01, 0.03 and 0.05. The training and evaluating protocol will be same as above. Lastly, we will evaluate the accuracy of POCS-Net segmentation in unseen objects. We will build co-segmentation dataset based on ModelNet40 and select the first 20 categories objects to train POCS-Net and use the last 20 categories to test POCS-Net. The minimal overlapped ratio of the dataset will be set to 40% and the noise will be set to Gaussian with a mean of 0 and a variance of 0.01.

4.3. The Effectiveness of POCS-Net for Point Could Registration

We will evaluate the impact of using POCS-Net as a pre-processing procedure on existing registration methods. We choose two representative traditional registration methods ICP Baker and Matthews (2004) and FRG Zhou et al. (2016), and two deep learning based registration methods DCP Wang and Solomon (2019) and RPM-Net Yew and Lee (2020) as the registration framework. In the two deep learning based methods, DCP doesn't consider the partially overlapping problem, and RPM-Net propose a solution to deal with

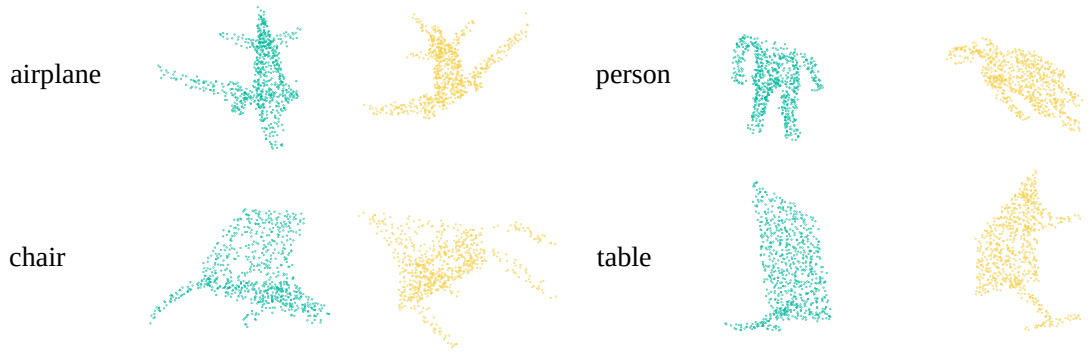

Figure 2: Part of category from the PointCS.

the partially overlapping problem. We tested four different minimal overlapping ratio of the point clouds, 20%, 40%, 60% and 80%. The noise will be set to as Gaussian with a mean of 0 and a variance of 0.01. We will use 80% data to train the network and use the rest of 20% data for testing The overall workflow is to first co-segment the overlapping region from the two point clouds to be registered and then used each algorithm to register the overlapping regions. We will analyze if registering only the overlapping region can improve the accuracy and robustness of 3D point cloud registration.

5. Experimental Results

5.1. Partially Overlapping Point Cloud Dataset

In order to evaluate the proposed POCS-Net, we built a co-segmentation dataset named PointCS, which contains partially overlapping point clouds. PointCS was built from ModelNet40 and ShapeNet, which are two widely used 3D point cloud data sets. ModelNet40 contains clean 3D models of 40 categories, and ShapeNet contains 3D models of 55 categories collected from online repositories. First, we sampled 1024 points on the 3D models' surface as Chen et al. (2020). Then we transformed the point clouds by sampling three Eular angle rotations in the range $[-180°, 180°]$ and translations in the range $[-0.05, 0.05]$ on each axis to obtain a series of full point cloud. For a full point cloud, a random plane is used to cut it into two parts and the bigger one is retained as a partial point cloud. The cutting plane can be shifted such that a certain percentage of points are retained. Then for each full point cloud, a series of partial point clouds are generated and they are used as partially overlapping points for training and testing as show in Fig.2. In this way, the overlapping region of each pair of partial point clouds are known without manual labelling, and the overlapping ratio can also be calculated.

5.2. Co-segmentation Results Under Different Overlapping Ratio

We evaluated the accuracy of POCS-Net segmentation under different minimal overlapping ratios. we built four datasets with different minimal overlapping ratios of 20%, 40%, 60% and 80%. A minimal overlapping ratio of 20% means that the overlapping ratio of all the point cloud pairs to be co-segmented belongs to a uniform distribution in the range of

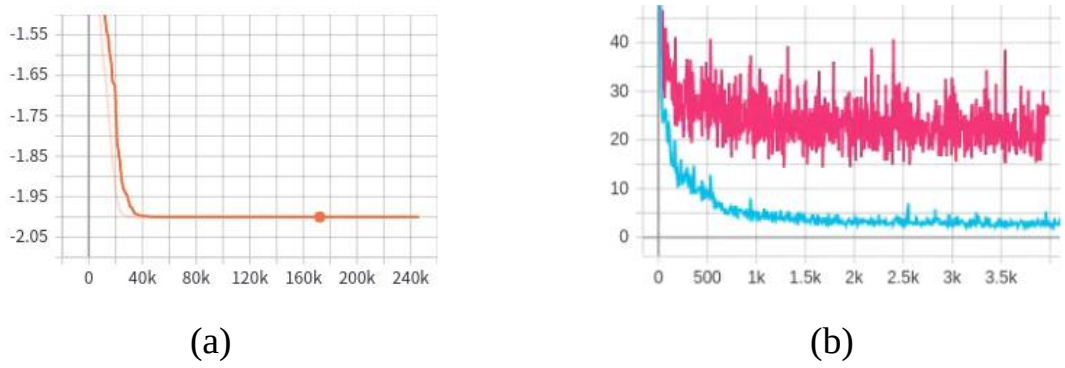

Figure 3: (a) The loss of L_{reg} during training. (b) The training loss of reconstruction task, the pink curve represents the loss of reconstruct a partial point cloud and the blue curve represents the loss of reconstruct a complete point cloud.

Table 1: Co-segmentation accuracy under different overlapping ratio

Overlapping ratio	20%	40%	60%
Accuracy	66.9%	75.2%	76.7%

[20%, 100%]. Each point was added a Gaussian with a mean of 0 and a variance of 0.01. In each dataset, we used 80% data to train the POCS-Net and used the rest of 20% data to evaluate the network segmentation performance according to the IoU between the network segmentation and the ground truth. The co-segmentation accuracy is shown in Table.1.

The results suggest the feature extractor degenerated and the segmentation network only achieved co-segmentation accuracy around 0.7. To avoid degeneration, we first add term L_{reg} as in Wang and Isola (2020) in loss function, but the feature extractor still degenerated and the training loss of the L_{reg} shows in Fig.3(a). We further applied reconstruction task to prevent the degeneration. However, as the input point clouds are random cropped, the reconstruction network was unable to reconstruct the input point clouds correctly and the reconstruction loss during training is shown in Fig.3(b).

5.3. Co-segmentation Results Under Different Noise Level

we further evaluated the accuracy of POCS-Net segmentation under different noise level. We built four datasets with 40% minimal overlapping ratio but with four different noise variances of 0.005, 0.01, 0.03 and 0.05. The training and evaluating protocol were same as above. The co-segmentation accuracy in shown in Table.2. As the degeneration cannot be avoided, POCS-Net also achieved an accuracy around 0.7 under different noise level.

Table 2: Co-segmentation accuracy under different noise level

noise variances	0.005	0.01	0.03	0.05
Accuracy	77.2%	76.3%	75.9%	75.2%

5.4. Co-segmentation Results in Unseen Objects

Lastly, we evaluated the accuracy of POCS-Net segmentation in unseen objects. We built co-segmentation dataset based on ModelNet40 and selected the first 20 categories objects to train POCS-Net and used the last 20 categories to test POCS-Net. The minimal overlapped ratio of the dataset was set to 40% and the noise was set to Gaussian with a mean of 0 and a variance of 0.01.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 81701795.

References

Simon Baker and Iain Matthews. Lucas-kanade 20 years on: A unifying framework. *International journal of computer vision*, 56(3):221–255, 2004.

Sayan Banerjee, Avik Hati, Subhasis Chaudhuri, and Rajbabu Velmurugan. Cosegnet: Image co-segmentation using a conditional siamese convolutional network. In *IJCAI*, pages 673–679, 2019.

Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

Rui Chen, Songfang Han, Jing Xu, and Hao Su. Point-based multi-view stereo network. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1538–1547, 2019.

Yilun Chen, Shu Liu, Xiaoyong Shen, and Jiaya Jia. Dsgn: Deep stereo geometry network for 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12536–12545, 2020.

Haowen Deng, Tolga Birdal, and Slobodan Ilic. Ppfnet: Global context aware local features for robust 3d point matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 195–205, 2018.

Zan Gojcic, Caifa Zhou, Jan D Wegner, Leonidas J Guibas, and Tolga Birdal. Learning multiview 3d point cloud registration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1759–1769, 2020.

Benjamin Graham, Martin Engelcke, and Laurens Van Der Maaten. 3d semantic segmentation with submanifold sparse convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9224–9232, 2018.

Weihao Li, Omid Hosseini Jafari, and Carsten Rother. Deep object co-segmentation. In *Asian Conference on Computer Vision*, pages 638–653. Springer, 2018.

Yang Li, Ting Liu, Jingwen Hu, and Jing Jiang. Topical co-attention networks for hashtag recommendation on microblogs. *Neurocomputing*, 331:356–365, 2019.

Jiasen Lu, Jianwei Yang, Dhruv Batra, and Devi Parikh. Hierarchical question-image co-attention for visual question answering. In *Advances in neural information processing systems*, pages 289–297, 2016.

Xiankai Lu, Wenguan Wang, Chao Ma, Jianbing Shen, Ling Shao, and Fatih Porikli. See more, know more: Unsupervised video object segmentation with co-attention siamese networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3623–3632, 2019.

Priyanka Mandikal and Venkatesh Babu Radhakrishnan. Dense 3d point cloud reconstruction using a deep pyramid network. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1052–1060. IEEE, 2019.

Anh Viet Phan, Minh Le Nguyen, Yen Lam Hoang Nguyen, and Lam Thu Bui. Dgcnn: A convolutional neural network over large-scale labeled graphs. *Neural Networks*, 108: 533–543, 2018.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017a.

Charles R Qi, Or Litany, Kaiming He, and Leonidas J Guibas. Deep hough voting for 3d object detection in point clouds. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9277–9286, 2019.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems*, pages 5099–5108, 2017b.

Gernot Riegler, Ali Osman Ulusoy, and Andreas Geiger. Octnet: Learning deep 3d representations at high resolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3577–3586, 2017.

Carsten Rother, Tom Minka, Andrew Blake, and Vladimir Kolmogorov. Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 993–1000. IEEE, 2006.

Lyne Tchapmi, Christopher Choy, Iro Armeni, JunYoung Gwak, and Silvio Savarese. Segcloud: Semantic segmentation of 3d point clouds. In *2017 international conference on 3D vision (3DV)*, pages 537–547. IEEE, 2017.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

Tongzhou Wang and Phillip Isola. Understanding contrastive representation learning through alignment and uniformity on the hypersphere. *arXiv preprint arXiv:2005.10242*, 2020.

Yuan Wang, Tianyue Shi, Peng Yun, Lei Tai, and Ming Liu. Pointseg: Real-time semantic segmentation based on 3d lidar point cloud. *arXiv preprint arXiv:1807.06288*, 2018.

Yue Wang and Justin M Solomon. Deep closest point: Learning representations for point cloud registration. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3523–3532, 2019.

Bichen Wu, Alvin Wan, Xiangyu Yue, and Kurt Keutzer. Squeezeseg: Convolutional neural nets with recurrent crf for real-time road-object segmentation from 3d lidar point cloud. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1887–1893. IEEE, 2018a.

Bichen Wu, Xuanyu Zhou, Sicheng Zhao, Xiangyu Yue, and Kurt Keutzer. Squeezesegv2: Improved model structure and unsupervised domain adaptation for road-object segmentation from a lidar point cloud. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 4376–4382. IEEE, 2019.

Qi Wu, Peng Wang, Chunhua Shen, Ian Reid, and Anton Van Den Hengel. Are you talking to me? reasoned visual dialog generation through adversarial learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6106–6115, 2018b.

Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920, 2015.

Caiming Xiong, Victor Zhong, and Richard Socher. Dynamic coattention networks for question answering. *arXiv preprint arXiv:1611.01604*, 2016.

Zi Jian Yew and Gim Hee Lee. Rpm-net: Robust point matching using learned features. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11824–11833, 2020.

Kaihua Zhang, Jin Chen, Bo Liu, and Qingshan Liu. Deep object co-segmentation via spatial-semantic network modulation. In *AAAI*, pages 12813–12820, 2020.

Qi Zhang, Jiawen Wang, Haoran Huang, Xuanjing Huang, and Yeyun Gong. Hashtag recommendation for multimodal microblog using co-attention network. In *IJCAI*, pages 3420–3426, 2017.

Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Fast global registration. In *European Conference on Computer Vision*, pages 766–782. Springer, 2016.

Chenyang Zhu, Kai Xu, Siddhartha Chaudhuri, Li Yi, Leonidas J Guibas, and Hao Zhang. Adacoseg: Adaptive shape co-segmentation with group consistency loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8543–8552, 2020.