This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Persistent Homology based Graph Convolution Network for Fine-grained 3D Shape Segmentation

Chi-Chong Wong
University of Macau
amilton.wong@connect.um.edu.mo

Chi-Man Vong
University of Macau
cmvong@um.edu.mo

Abstract

Fine-grained 3D segmentation is an important task in 3D object understanding, especially in applications such as intelligent manufacturing or parts analysis for 3D objects. However, many challenges involved in such problem are yet to be solved, such as i) interpreting the complex structures located in different regions for 3D objects; ii) capturing fine-grained structures with sufficient topology correctness. Current deep learning and graph machine learning methods fail to tackle such challenges and thus provide inferior performance in fine-grained 3D analysis. In this work, methods in topological data analysis are incorporated with geometric deep learning model for the task of fine-grained segmentation for 3D objects. We propose a novel neural network model called Persistent Homology based Graph Convolution Network (PHGCN), which i) integrates persistent homology into graph convolution network to capture multi-scale structural information that can accurately represent complex structures for 3D objects; ii) applies a novel Persistence Diagram Loss ($\mathcal{L}_{PD}$) that provides sufficient topology correctness for segmentation over the fine-grained structures. Extensive experiments on fine-grained 3D segmentation validate the effectiveness of the proposed PHGCN model and show significant improvements over current state-of-the-art methods.

1. Introduction

Fine-grained 3D semantic segmentation is a task to semantically classify the labeling of each 3D point input in detailed levels. It is an essential task for many applications for detailed processing and analysis of 3D shapes, such as intelligent manufacturing, automatic interior design and furniture arrangement, autonomous robotic manipulation, human-machine interaction, 3D clothing analysis.

Segmenting fine-grained 3D objects involves many challenges, due to the specific properties in fine-grained 3D objects, such as i) complex structures located in different regions; ii) shape-dependent topological structures (e.g, handles of objects, doorknobs, device wires). These properties always exhibit in slim parts or multiple small connected components, which are semantically important to downstream tasks, such as robotic manipulation. Sufficiently interpreting such two main structures is essential for accurate 3D fine-grained semantic segmentation task. Failing to tackle such challenges will drastically lower the performance of semantically understanding the 3D fine-grained objects and produce incoherent segmentation output, which is vital for intelligent manufacturing and robotic manipulation.

In recent years, deep neural network based methods [6, 16, 21] and geometric learning methods [15, 27, 28] have become the mainstream methods in 3D point cloud understanding tasks, from general 3D object classification to semantic segmentation on objects and scenes. In retrospect to these methods, it is found that they are not designed specifically for the task of understanding fine-grained 3D objects with complex structures or shape dependent topological structures. Methods in [15, 27, 28] apply graph neural network (GNN) or graph convolutional network (GCN) model to extract features from geometrical structures in 3D point cloud. However, such approach only captures the pairwise relations represented by edges, since the neighboring graphs constructed in GNN/GCN model only represent the pairwise relationships among 3D point clouds. As a result, high dimensional relationships existed in complex structures of fine-grained 3D objects cannot be finely captured. The recent work PartNet model [31] applies cascade binary labeling to represent a top-down recursive parts decomposition for hierarchical segmentation. However, the representation capability of binary labeling is limited by the hierarchy depth and thus suffers from handling 3D objects with multiple complex structures.

In fact, the geometric and topological information existed in complex structure is the essential clue to understand the shapes of fine-grained objects. *Topological Data Analysis* (TDA) [3] is an emerging field which infers relevant topological and geometric features from complex data.

TDA uses a mechanism called complex filtration to construct the multi-scale topological structures for the input point clouds, which extracts the high dimensional relationships existed in complex structures of point cloud, as illustrated in Fig. 1(a). Then, persistent homology, a tool in TDA, is applied on the resulting nested sequence of strictly increasing subcomplexes, which are called filtered complexes, to compute the multi-scale topological features, represented as persistence barcodes and persistence diagram, as shown in Fig. 1(b) and (c). The 0-dim, 1-dim and 2-dim persistent homological features in the resulting persistence diagram correspond to the connected components, cycles and higher dimensional counterparts (e.g, cavities) in point cloud.

In this work, we opt for the TDA tools [10] to propose a novel network model called Persistent Homology based Graph Convolution Network (PHGCN), which incorporates persistent homological features into GCN network to enhance its capability to capture multi-scale topological features of complex structures in fine-grained 3D objects.

To further tackle the fine-grained segmentation problem, we found that the shape-dependent topological structures in fine-grained objects, especially the connected parts (e.g, handles, wires, knobs) are always exhibited as small-sized objects or thinly-connected components, which makes coherent segmentation difficult since cross entropy loss generally used in segmentation task may not sufficiently reflect the topological error even the overall loss value is low. To overcome such issue, we propose a Persistence Diagram Loss ($\mathcal{L}_{PD}$), which works as topological constraint to ensure the segmented output with sufficient topology correctness to obtain coherent segmentation output.

To the best of our knowledge, our work is the first work which introduces persistent homology to tackle fine-grained 3D semantic segmentation problem. The main contributions of our work are summarized as follows:

1. With the persistent homological features, a novel GCN network model is able to capture multi-scale topological features of complex structures in fine-grained 3D objects.

2. A novel persistence diagram loss is applied to reinforce the topological correctness in prediction to provide coherent fine-grained segmentation output.

3. The proposed work demonstrates the feasibility on the extension of generic GNN/GCN structure with computational topological methods.

Extensive experiments on fine-grained semantic segmentation are evaluated on challenging 3D object-parts segmentation benchmarks, which demonstrate that the proposed PHGCN model achieves state-of-the-arts results.

2. Preliminaries in Topological Data Analysis

Topological data analysis (TDA) [3] is an emerging field which goal is to capture the related topological and geometric features from data of complex structure. In this section, a brief overview is provided to highlight the mechanism in TDA. The details of TDA can be found in the seminal papers [11, 33].

2.1. Simplicial Complex

As there is no direct approach in extracting topological information from data points, simplicial complex is constructed as a topological approximation to the underlying shape of the sampled points. Simplicial complex can be regarded as a high dimensional extension of a graph, which contains collection of simplices of different dimension. The geometrical realizations of k-dim simplices are vertices(k=0), edges(k=1), triangles(k=2), tetrahedra(k=3) and higher-order counterparts($k \geqslant 4$), respectively.

2.2. Homology Group

To computationally analyze the topological features of simplicial complex, homology groups are assigned to simplicial complex. The homology group is a kind of mathematical group which describes the topological features of simplicial complex in different dimension. The topological features of k-dim homology group refer to connected components (k=0), cycles (k=1), and cavities (k=2), respectively.

2.3. Persistent Homology and Filtration

Persistence is a kind of measurement on how a shape changes when a given parameter varies. Persistent homology provides a way for keeping track of when topological features appear and vanish during the variation of certain parameter, such as the scale of each data point. In this period, a nested sequence of simplical complexes, also known as filtration, is generated as shown in Fig. 1(a). The filtration captures the evolution progress of simplical complex by increasing the scale parameter, which can be considered as the multi-scale topological spaces underlying the data points. Consequently, the lifetime of each multi-scale topological feature is recorded as persistence barcode, as shown in Fig. 1(b). The persistence barcode can then be converted into the birth time and death time of each topological feature and is represented as persistence diagram, as illustrated in Fig. 1(c). As a result, the multi-scale topological information of the shape underlined in the data points is captured.

3. Related Works

In this section, the two main related techniques: deep learning methods on point clouds and persistent homology methods are discussed.

3.1. Deep Learning on Point Clouds

The success of deep neural network based methods in 2D image semantic segmentation tasks [7,23,32], have promoted its feasibility for 3D point cloud input [6, 16, 17, 21, 30]. However, such methods lack of sufficient capability in capturing the connections among the points. Graph based methods such as DGCNN [28], ResGCN-28 [15] explicitly construct the graph from point cloud by assigning each point as node and building the edges through measuring the relatedness between pair of points. However, such methods only capture the pairwise relationships among points and suffer in capturing high-order relationships of complex structures, which are prevalent in fine-grained objects.

3.2. Persistent Homology in Machine Learning

Persistent homology is an essential method in topological data analysis for extracting topological features from geometric realizations at different spatial resolutions. The extracted topological features give the insights of the underlying shapes of data and work as powerful features deployed in machine learning pipeline [1,2,14]. Therefore, the effectiveness of TDA methods have attracted the wide adoption of computational topological methods for various applications, including action recognition [25], medical imaging [8, 22], shape matching [20] and design of neural network [4, 9]. Some recent works [12, 13] also explore the feasibility in differentiability of persistence homology. Inspired by these promising works, a novel PHGCN is proposed to integrate persistent homology mechanism with graph convolution network for capturing the multi-scale structural information in complex structures of fine-grained objects.

4. Methods

In this section, the details of our proposed method for fine-grained 3D semantic segmentation are presented. The proposed method consists of two core modules: 1) Persistent Homology based Graph Convolution Neural Network (PHGCN) for capturing multi-scale structural information in complex structures via the combination of topological persistence (PH) and graph convolution network (GCN), 2) Persistence Diagram Loss ($\mathcal{L}_{PD}$) applied in the optimization to segment the fine-grained structures with reduced topological error. The entire network architecture is illustrated in Fig. 2. The details of each proposed module are described in the following sections.

4.1. General Graph Convolution Network (GCN)

Following the network design of GACNet [27] and AD-Convnet [29], we construct a general graph convolution network (GCN) layer through aggregation step and update step. In **aggregation step**, each point P_i of point cloud input is represented by its 3D position $r_i \in \mathbb{R}^3$ and D-dimensional feature $f_i \in \mathbb{R}^D$. By thresholding the ball radius, a K-neighbors graph of P_i is built by randomly selecting its K neighboring points namely as P_{ij}= (r_{ij}, f_{ij}), within its spherical neighborhood $\mathcal{N}_i$, where $r_{ij} = r_j - r_i \in \mathbb{R}^3$ is the relative position and $f_{ij} \in \mathbb{R}^{D'}$ is the enhanced feature. D' is the dimension number of transformed features. Through soft attention mechanism, the weighted sum of the features of neighboring points is obtained as the aggregated feature $h'_i \in \mathbb{R}^{D'}$ for center point P_i,

$$h'_i = \sum_{j \in \mathcal{N}_i} a_{ij} f_{ij} \in \mathbb{R}^{D'} \tag{1}$$

where the attentional weight a_{ij} can be learned by back-propagation, as detailed in [29]. In **update step**, the aggregated feature h'_i is fed into a multilayer perceptron (MLP) and then followed by a ReLU [19] activation function to obtain the transformed feature $f_i^{out} = \text{ReLU}(\text{MLP}(h'_i)) \in \mathbb{R}^C$, as the output of GCN layer. The collection of such point features $\{f_i^{out}\}\, (i = 1, ..., M)$ constitute a output feature map $F^{out} \in \mathbb{R}^{M \times C}$, where M is the number of sub-sampled points and can be regarded as the spatial size of output feature map. For simplicity in notations, we denote the output pointwise feature map of the last GCN layer as $F^L \in \mathbb{R}^{M \times C}$. The details of the procedure mentioned above can be referenced from [27] and [29].

4.2. Persistent Homology based Graph Convolution Network (PHGCN)

By applying the computation of sequential GCN layers, the resulting local feature map only captures the features of local neighborhood, which is insufficient for understanding the 3D fine-grained objects. In this work, we adopt tools from topological data analysis to extract the essential information hidden in the complex structures of 3D fine-grained objects.

Persistent homology (PH) is a mathematical tool from topological data analysis which is able to extract provably stable topological features of the shapes underlined in point clouds. To overcome the issue that general GCN model cannot capture the complex structures of 3D fine-grained objects, we extend the GCN model by integrating PH module for extracting the essential topological information in complex structures.

The approach on extracting persistent homology features is described in PH module, as illustrated in Fig. 3(b): the input point cloud of N 3D points can be considered as a finite metric space, denoted as $\mathbb{X}_N$, a filtration construction is applied on $\mathbb{X}_N$ to extract a chain of multi-scale filtered simplical complexes via varying scale parameter, denoted as Filt$(\mathbb{X}_N)$. Then, persistent homology is applied to compute the evolution of topological features and the period between the appearance time and disappearance time of topological features are kept, denoted as *birth time*, b and

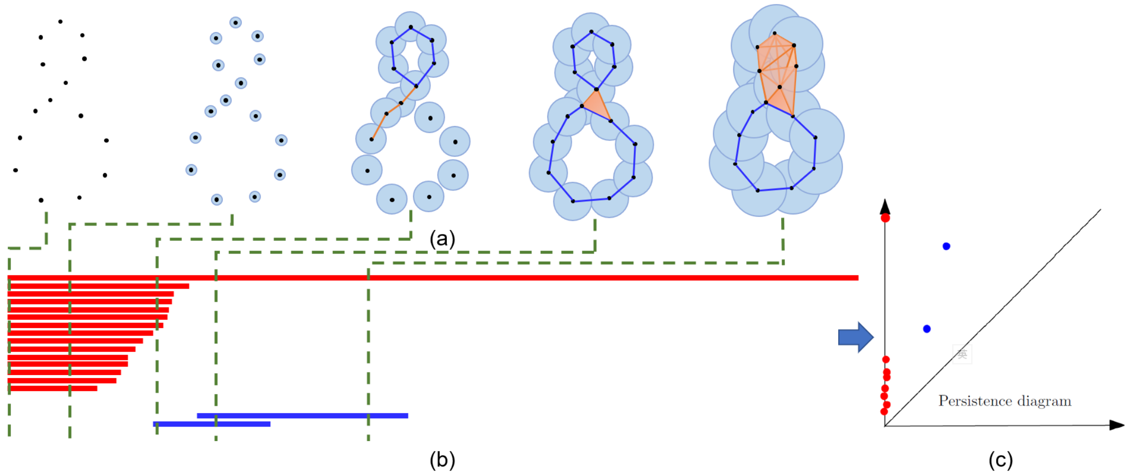

Figure 1. The pipeline of computing persistence diagram. (a). The evolution of filtered subcomplexes (in dark blue and orange color) when varying the scale parameter, which is represented as the radius of the light blue circle around each data point. (b). The corresponding persistence barcode for each filtered subcomplex, which keeps the lifetime for homological class. The topological features of 0-dim and 1-dim homological class are illustrated as red bars and blue bars respectively, for this example of 2D plane case. (c) The final persistence barcode is converted into the persistence diagram.

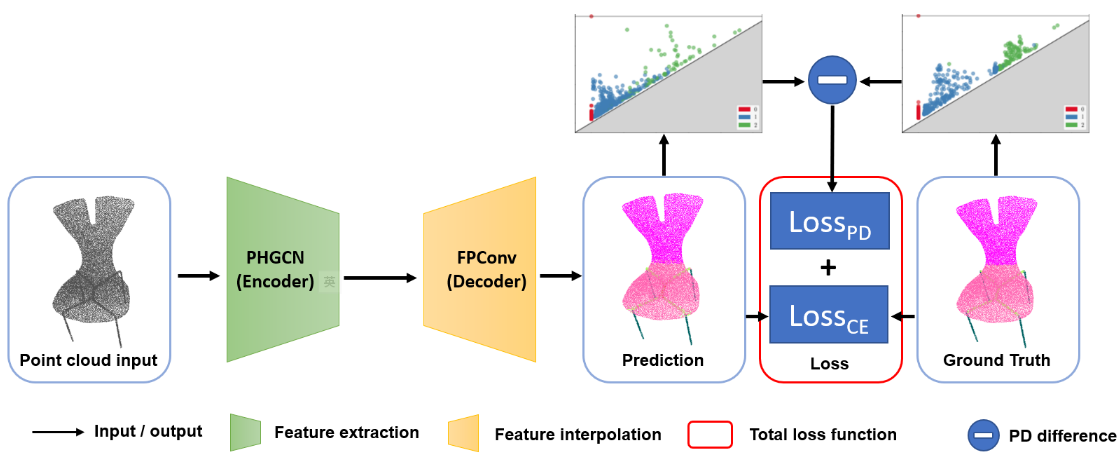

Figure 2. The entire network architecture of PHGCN.

death time, d. Such periods are often depicted by persistence diagrams (PD) which are sets of points in 2D plane, where each point (b, d) represents the k-th persistent homology class which appears at time b and disappears at time d (The dimensions k=0,1,2 refer to connected components, cycles and cavities, respectively). The resulting PD reflects the multi-scale summarized topological information which is essential for interpreting the complex structures. Let $PD = \mathrm{dgm}(\mathrm{Filt}(\mathbb{X}_N))$ denote the resulting persistence diagram obtained from previous filtration construction. We denote PD_k as the persistence diagram for k-dimensional homological features. The counterparts correspond to connected components (k=0), cycles (k=1) and cavities (k=2), respectively. We argue that the essential topological information of complex structure mainly exists in cycles and cavities and hence 1-dimensional and 2-dimensional homological features are used in our method. By applying the TDA pipeline as illustrated in Fig.1, the persistence diagram of 1-dimension and 2-dimension are extracted as PD_1 and PD_2 respectively. As persistence diagram is originally a multi-set but not linear vectors, it is unsuitable to be used as input for general machine learning pipeline. Therefore, PD_1 and PD_2 are further featurized into persistence image [1], denoted as PI_1 and PI_2 , and then flattened as feature vectors. To combine the strength of both 1-dimensional and 2-dimensional homological features, vector concatenation is adopted to obtain the feature vector as a topological descriptor, denoted as $\mathbf{p} \in \mathbb{R}^P$. P is the feature dimension of topological descriptor. For combining with local feature map $F^L \in \mathbb{R}^{M \times C}$, topological descriptor is repeated by M

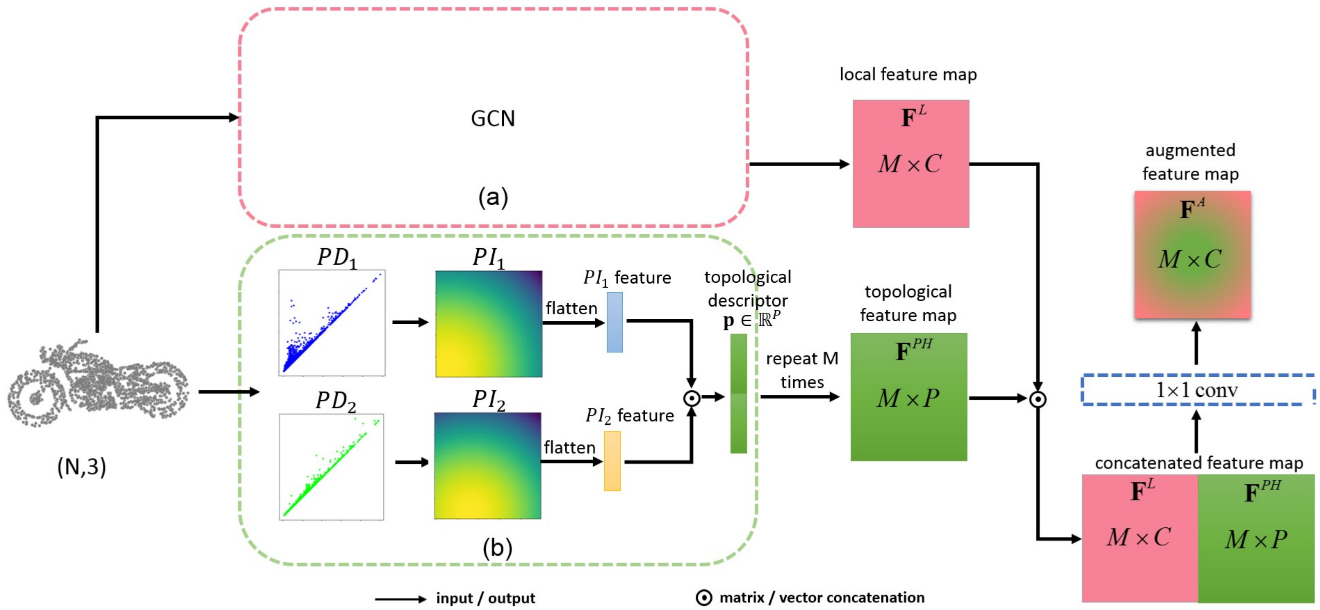

Figure 3. The pipeline of PHGCN module. (a) Data processing flow on GCN module. (b) Data processing flow on PH module.

times to obtain the topological feature map $\mathbf{F}^{PH} \in \mathbb{R}^{M \times P}$. Then topological feature map $\mathbf{F}^{PH}$ is concatenated with local feature map $\mathbf{F}^{L}$. Finally, for learning a better feature map fusion, a learnable 1×1 conv filter is applied on the concatenated feature map to obtain the augmented feature map $\mathbf{F}^{A} \in \mathbb{R}^{M \times C}$. Consequently, the topological information existed in complex structure is adaptively aggregated to augment the pointwise feature representation for better performing the fine-grained semantic segmentation task.

As the feature maps are down-sampled after processed by each GCN layer, the augmented feature map F^A is subsequently interpolated by Feature Propagation Convolution (FPConv) module [21] for recovering the point features as the original scale of input point cloud, which is the final point-wise predictions. The whole proposed network model PHGCN can be trained in an end-to-end manner.

4.3. Persistence Diagram Loss $\mathcal{L}_{PD}$

By insights that it is challenging to obtain coherent segmentation output for fine-grained structures, we further enforce the segmentation output with topological correctness by the utility of persistence diagram. In this work, instead of only using cross-entropy loss which is widely used in semantic segmentation, we integrate a novel topological loss which measures the topological error between the two persistence diagrams of the prediction likelihoods and ground-truth labels. This specific topological loss is called persistence diagram (PD) loss, denoted as $\mathcal{L}_{PD}$. The PD loss $\mathcal{L}_{PD}$ applies the 1-order Wasserstein distance [26] W_1 to find the best match m^* between these two persistence diagrams (persistence diagram for likelihood dgm_L and persistence diagram for ground-truth dgm_{GT}).

$$\begin{aligned} m^* &= \operatorname*{argmin}_m W_1(\mathrm{dgm}_L, \mathrm{dgm}_{GT}) \\ &= \operatorname*{argmin}_m \sum_{(u,v)\in m} \|u - v\|_\infty \end{aligned} \tag{2}$$

where $u \in \mathrm{dgm}_L$, $v \in \mathrm{dgm}_{GT}$ and the Wasserstein distance here uses the L_∞-norm distance between each pair of points $(u, v) \in m$, for each possible match m between diagrams dgm_L and dgm_{GT}. Once the optimal match m^* is found, $\mathcal{L}_{PD}$ is computed as the square distance between each matched pair of points from dgm_L and dgm_{GT}.

$$\mathcal{L}^c_{PD} = \sum_{(u,v)\in m^*} \|u - v\|_2^2 \tag{3}$$

Note that $\mathcal{L}^c_{PD}$ here is computed in terms of each part class of certain category of 3D object. The summation over PD loss of all part classes ($c = 1, 2, ..., N_c$) is performed to obtain the final one,

$$\mathcal{L}_{PD} = \sum_{c=1}^{N_c} \mathcal{L}^c_{PD} \tag{4}$$

then $\mathcal{L}_{PD}$ is integrated with cross entropy loss $\mathcal{L}_{CE}$ to obtain the final objective function $\mathcal{L}$ for optimization.

$$\mathcal{L} = \mathcal{L}_{CE} + \mathcal{L}_{PD} \tag{5}$$

As a result, topological constraints are appended in optimization to enhance the coherency and connectivity in segmentation output, especially for fine-grained objects with shape-dependent topological structures.

4.4. Network Architecture

Taking reference from the network architecture design of Pointnet++ [21], our proposed PHGCN model adopts the encoder decoder style [21,27,29] for part semantic segmentation task. The encoder module contains four graph convolutional network (GCN) layers, which are considered as extractor for local features. Then, the local features extracted are fused with topological features extracted from complex structures, which are captured by the persistent homology (PH) module. For upsampling the features downsampled by the encoder, the decoder module which consists of four FPConv layers is applied for gradually interpolating the features as the original size of the input. For optimization, the predicted likelihood and ground truth for each part class of 3D object are used to compute PD loss $\mathcal{L}_{PD}$, which is then integrated with cross entropy loss $\mathcal{L}_{CE}$.

5. Experiments

5.1. Segmentation on ShapeNet-Part Dataset

The ShapeNet-Part [5] dataset is the first complete benchmark on 3D fine-grained point-wise segmentation. It contains 16,881 CAD shape instances from 16 categories, with part labels annotated on 2,048 sampled points. There are 50 kinds of part labels in total. Each category is annotated with two to five part labels.

For quantitative evaluation, we follow the setting from PointNet [6] to select 14,007 shape instances as the training set, and the remaining 2,874 ones as the validation set for accuracy evaluation. The 3D coordinates of 2,048 sampled points of each instance are used as the input.

Quantitative and Qualitative Results: The quantitative results evaluated for ShapeNet-Part dataset are provided in Table 1. The part-wise Intersection-over-Union (IoU) is used as the metric in our evaluation, and is given for each object category and the mean value (mIoU). It is shown that the proposed PHGCN provides the best result with 89.2% part-wise mean IoU, outperforming all other competitive methods. Specifically, PHGCN achieves significant gains in fine-grained objects with complex structures, such as lamp, guitar, and objects with thin parts, such as ear-phone, mug, table and chair, which are illustrated as the consistent results that the persistent homology mechanism in PHGCN takes effects in interpretating complex structures and shape dependent topological structures of fine-grained objects.

The qualitative segmentation results on validation set of ShapeNet-Part dataset are illustrated in Fig. 4(a), in which the prediction from PHGCN model is greatly consistent with the ground truth part labels annotations, even the ShapeNet-Part dataset contains a lot of shapes in complex structures and thin parts. To give a better intuition on the effectiveness on the extracted topological features, we perform qualitative comparison to two representative geometric deep learning methods (DGCNN, ResGCN-28). In Fig.4(b), for objects (lamp and mug) with fine-grained parts, the segmentations (DGCNN, ResGCN-28) suffer from broken connections and mis-classification on small component, while PHGCN provides coherent segmentation with sufficient topology correctness. For object (motor) with multi-scale and complex structures, the segmentations (DGCNN, Res28GCN) show that handles and parts near the wheel are mis-segmented while the output of PHGCN is accurate enough as the ground-truth.

5.2. Segmentation on PartNet Dataset

The proposed PHGCN model is further evaluated on a larger and more complex benchmark called PartNet [18] dataset. The PartNet [18] dataset contains 26,671 shape instances and divided in 573,585 part instances with fine-grained part annotations. It covers 24 object categories. Among these categories, there are some categories which have complex structures, such as lamp, faucet, chair. Also, some categories such as door, refrigerator, earphone, contain thin and semantically important parts, e.g. the cord of an earphone, the handle of a door or a refrigerator. All of these properties of PartNet dataset bring great challenges for accurate segmentation.

For quantitative evaluation, we follow the setting from PartNet [18] to split the dataset into train set, validation set and test set in the ratios of 70%, 10%, 20%, respectively. Each input instance is the 10,000 points sampled from each CAD model and only 3D coordinates are used as the input. To validate the effectiveness of our proposed method on fine-grained object segmentation, the fine level (level-3) of PartNet (17 categories) is selected for the evaluation.

Quantitative and Qualitative Results: In Table 2, the results of PHGCN are presented with comparisons to several state-of-the-art methods on PartNet dataset. The comparison result shows that the proposed PHGCN model outperforms all prior state-of-the-art methods such as PointNet++ [21], PointCNN [16], ResGCN [15], ADConvnet [29], as reported by the part-wise IoU for each category and the mean IoU across all categories. Particularly, PHGCN provides over 10% relative improvement against graph deep learning methods, such as DGCNN and ResGCN-28 method. It is observed that fine grained object with complex structures such as faucet, lamp and chair can be segmented in improved accuracy. Together with this result, objects with thin parts such as earphone, door and refrigerator, the proposed PHGCN achieves significant accuracy gains compared with other state-of-the-art methods. This is the significance and the most important point for segmenting the fine-grained 3D objects, i.e., thin parts, while the prior state-of-the-art methods may not obtain higher accuracy on thin parts. By applying persistent homology method, the proposed PHGCN captures the structural information of these

Method	mIoU	air plane	bag	cap	car	chair	ear phone	guitar	knife	lamp	laptop	motor	mug	pistol	rocket	skate board	table
PointNet [6]	83.7	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6
PointNet++ [21]	85.1	82.4	79.0	87.7	77.3	90.8	71.8	91.0	85.9	83.7	95.3	71.6	94.1	81.3	58.7	76.4	82.6
SPLATNet [24]	85.4	83.2	84.3	89.1	80.3	90.7	75.5	92.1	87.1	83.9	96.3	75.6	95.8	83.8	64.0	75.5	81.8
PointCNN [16]	86.1	84.1	86.4	86.0	80.8	90.6	79.7	92.3	**88.4**	85.3	96.1	77.2	95.3	84.2	64.2	80.0	83.0
DGCNN [27]	85.0	84.0	83.4	86.7	77.8	90.6	74.7	91.2	87.5	82.8	95.7	66.3	94.9	81.1	63.5	74.5	82.6
ResGCN-28 [15]	85.5	85.3	83.5	86.6	78.1	90.9	75.1	91.5	86.8	83.1	95.9	68.1	95.2	80.9	63.8	74.8	83.1
PartNet [31]	87.4	87.8	**86.7**	89.7	80.5	91.9	75.7	91.8	85.9	83.6	97.0	74.6	97.3	83.6	64.6	78.4	85.8
ADConvnet [29]	87.0	87.4	86.4	89.5	80.1	91.4	78.9	91.5	86.2	82.9	96.9	76.9	97.1	84.0	64.3	79.6	85.4
PHGCN	**89.2**	**88.9**	86.6	**90.2**	**81.5**	**92.8**	**80.2**	**93.1**	88.2	**87.8**	**97.5**	**79.2**	**98.6**	**85.8**	**64.7**	**80.9**	**87.8**

Table 1. Comparison of semantic segmentation on the ShapeNet-Part dataset. Metric is part-wise IoU (%).

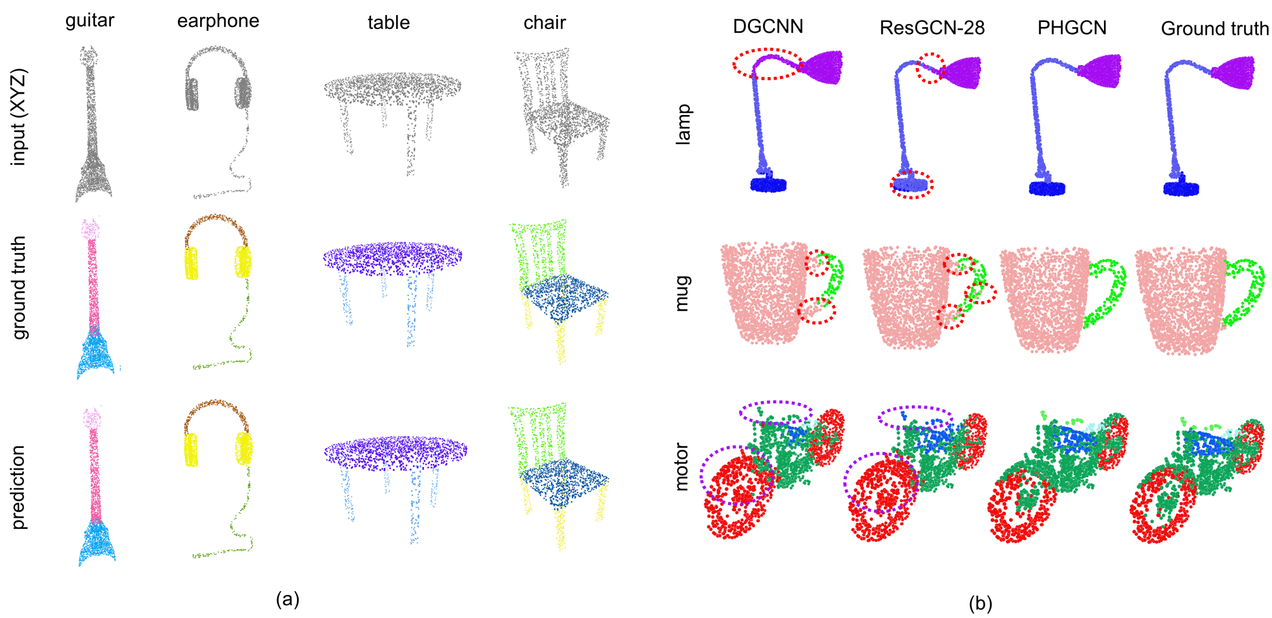

Figure 4. (a) Qualitative results of PHGCN on the ShapeNet-Part [5] validation set. (b) Qualitative comparison to baseline methods (DGCNN, ResGCN-28). Erroneous segmentations are circled in dash lines.

thin parts more effectively and thus gives the reason why PHGCN provides higher accuracy.

5.3. Ablation Analysis

The ablation analyses are conducted on both ShapeNet-Part and PartNet dataset to validate the efficacy of the proposed PHGCN model. The ablation result is presented in Table 3.

(1). Replace PHGCN module with general GCN module. The persistent homology mechanism of PHGCN enables the model to extract the topological information from complex structures in multi-scale manner. For comparison, general GCN layer only captures information in pairwise structures in local neighborhood. As a consequence, the performance is greatly decreased.

(2). Use $\mathcal{L}_{CE}$ only. The $\mathcal{L}_{PD}$ loss function provides sufficient topology correctness for coherent segmentation over the fine-grained structures. By removing $\mathcal{L}_{PD}$ from Eq. (5), the performance is lowered due to the incoherent segmentation output.

The part-wise mIoU scores of all ablated variants are compared in Table 3. It is concluded that: i) The most important ingredient comes from the PHGCN module, since the multi-scale structural information is essential in fine-grained objects. ii) The role of $\mathcal{L}_{PD}$ shows the next important factor in performance, especially for fine-grained objects with thin parts. From this ablation study, it is shown that the proposed module and loss function (which constitute the full PHGCN model) achieve the state-of-the-art accuracy.

5.4. Complexity Analysis on Persistence Diagram

To reduce computational cost while retaining the performance of proposed model, we apply Alpha-complex filtration rather than time-consuming filtrations such as Vietoris-Rips or Cech complex filtration. The running time for com-

Method	mIoU	bed	bottle	chair	clock	dish washer	display	door	ear phone	faucet	knife	lamp	micro-wave	refri-gerator	storage	table	trash can	vase
PointNet [6]	35.6	13.4	29.5	27.8	28.4	48.9	76.5	30.4	33.4	47.6	32.9	18.9	37.2	33.5	38.0	29.0	34.8	44.4
PointNet++ [21]	42.5	30.3	41.4	39.2	41.6	50.1	80.7	32.6	38.4	52.4	34.1	25.3	48.5	36.4	40.5	33.9	46.7	49.8
PointCNN [16]	46.5	41.9	41.8	43.9	36.3	58.7	82.5	37.8	48.9	60.5	34.1	20.1	58.2	42.9	**49.4**	21.3	53.1	58.9
DGCNN [28]	44.6	36.1	48.9	40.7	34.4	55.9	80.4	30.8	45.1	51.5	42.9	22.6	50.9	33.6	46.9	34.2	50.6	53.4
ResGCN-28 [15]	45.1	35.9	**49.3**	41.1	33.8	56.2	81.0	31.1	45.8	52.8	44.5	23.1	51.8	34.9	47.2	33.6	50.8	54.2
ADConvnet [29]	47.0	42.1	41.7	44.2	35.9	58.4	82.9	**38.1**	48.5	58.9	33.3	19.8	58.4	43.4	49.2	34.2	52.9	57.5
PHGCN	**49.8**	**43.0**	48.7	**45.9**	**41.9**	**59.3**	**84.2**	37.5	**50.2**	**62.9**	**46.1**	**26.2**	**59.1**	**44.2**	49.1	**35.2**	**54.3**	**59.5**

Table 2. Comparison of semantic segmentation on fine-grained objects (level=3) in PartNet dataset [18]. Metric is part-wise IoU (%).

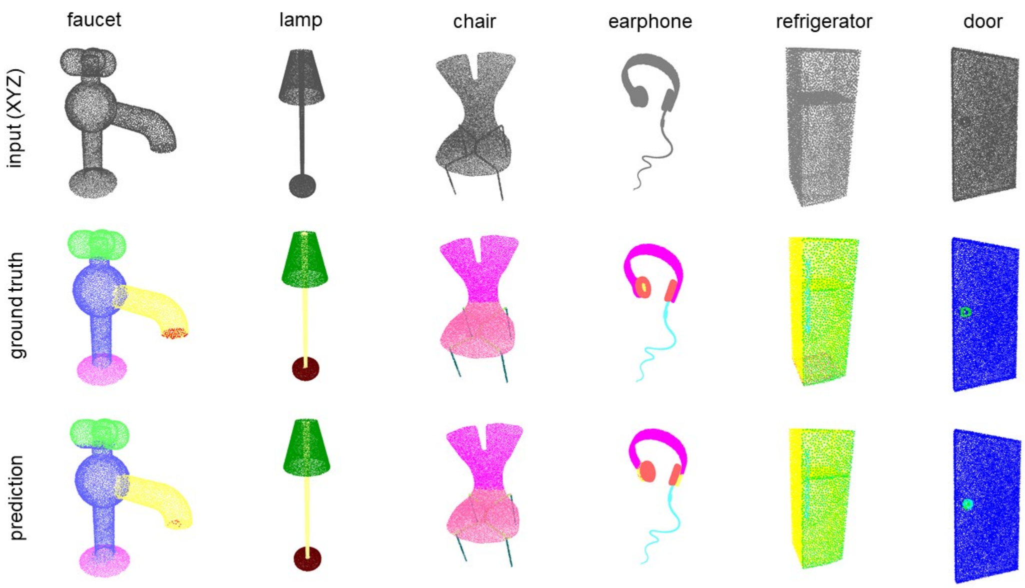

Figure 5. Qualitative results of PHGCN on the PartNet [18] test set.

	mIoU (ShapeNet-Part)	(PartNet)
(1). Use general GCN module	85.2	46.3
(2). Use $\mathcal{L}_{CE}$ only	86.3	47.0
(3). (1)+(2)	83.4	44.2
(4). The full PHGCN model	**89.2**	**49.8**

Table 3. The part-wise mIoU scores of all ablated variants based on the full PHGCN model. Both the ShapeNet-Part and PartNet dataset are selected for evaluation.

puting PD for an instance of ShapeNet-Part is 0.25sec in average (using Intel i7 CPU), which is suitable enough in our scenario.

6. Conclusion

In this work, a novel point cloud based neural network model called PHGCN is proposed which integrates computational topological methods to tackle several challenges in semantic segmentation for fine-grained 3D objects. The proposed PHGCN applies persistent homology mechanism into graph convolution network to handle input with multi-scale complex structures. It also applies $\mathcal{L}_{PD}$ loss function to reinforce the topological correctness in prediction to provide coherent fine-grained segmentation output.

From these two improvements, the segmentation results for fine-grained objects (especially objects with complex structures such as faucet, lamp, chair and objects with thin parts such as earphone, door and refrigerator) have got significant gain in accuracy. The performance of PHGCN is validated in terms of accuracy over two challenging benchmarks. From the experiments, PHGCN outperforms several state-of-the-art point cloud based segmentation methods. The experimental results also validate the contributions of PHGCN: i) persistent homology based GCN is an effective mechanism to capture multi-scale structural information from 3D objects; ii) a more accurate and coherent semantic segmentation with sufficient topology correctness for fine-grained structures; iii) much higher accuracy than state-of-the-art geometric deep learning methods (e.g., more than 10% relative improvement over DGCNN and ResGCN-28 method on PartNet dataset evaluation).

Acknowledgement. This project is funded by the Science and Technology Development Fund, Macau SAR (Files no. 0112/2020/A, and 004/2019/AFJ).

References

[1] Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *Journal of Machine Learning Research*, 18, 2017. 3, 4

[2] Peter Bubenik. Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.*, 16(1):77–102, 2015. 3

[3] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009. 1, 2

[4] Gunnar Carlsson and Rickard Brüel Gabrielsson. Topological approaches to deep learning. In *Topological Data Analysis*, pages 119–146. Springer, 2020. 3

[5] 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. 6, 7

[6] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 77–85, July 2017. 1, 3, 6, 7, 8

[7] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2018. 3

[8] James Clough, Nicholas Byrne, Ilkay Oksuz, Veronika A. Zimmer, Julia A. Schnabel, and Andrew King. A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 3

[9] Ciprian A Corneanu, Sergio Escalera, and Aleix M Martinez. Computing the testing error without a testing set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2677–2685, 2020. 3

[10] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010. 2

[11] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. In *Proceedings 41st annual symposium on foundations of computer science*, pages 454–463. IEEE, 2000. 2

[12] Rickard Brüel Gabrielsson, Bradley J. Nelson, Anjan Dwaraknath, and Primoz Skraba. A topology layer for machine learning. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 1553–1563. PMLR, 26–28 Aug 2020. 3

[13] Christoph Hofer, Florian Graf, Bastian Rieck, Marc Niethammer, and Roland Kwitt. Graph filtration learning. In *International Conference on Machine Learning*, pages 4314–4323. PMLR, 2020. 3

[14] Christoph Hofer, Roland Kwitt, Marc Niethammer, and Andreas Uhl. Deep learning with topological signatures. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. 3

[15] Guohao Li, Matthias Muller, Ali Thabet, and Bernard Ghanem. Deepgcns: Can gcns go as deep as cnns? In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9267–9276, 2019. 1, 3, 6, 7, 8

[16] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. In *Advances in Neural Information Processing Systems*, pages 820–830, 2018. 1, 3, 6, 7, 8

[17] Yongcheng Liu, Bin Fan, Gaofeng Meng, Jiwen Lu, Shiming Xiang, and Chunhong Pan. Densepoint: Learning densely contextual representation for efficient point cloud processing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5239–5248, 2019. 3

[18] Kaichun Mo, Shilin Zhu, Angel X Chang, Li Yi, Subarna Tripathi, Leonidas J Guibas, and Hao Su. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 909–918, 2019. 6, 8

[19] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010. 3

[20] Adrien Poulenard, Primoz Skraba, and Maks Ovsjanikov. Topological function optimization for continuous shape matching. In *Computer Graphics Forum*, pages 13–25. Wiley Online Library, 2018. 3

[21] 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, 2017. 1, 3, 5, 6, 7, 8

[22] Bastian Rieck, Tristan Yates, Christian Bock, Karsten Borgwardt, Guy Wolf, Nick Turk-Browne, and Smit Krishnaswamy. Uncovering the topology of time-varying fMRI data using cubical persistence. In *Advances in neural information processing systems*, pages 5998–6008, 2020. 3

[23] Evan Shelhamer, Jonathan Long, and Trevor Darrell. Fully convolutional networks for semantic segmentation. *IEEE Tansactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651, 2017. 3

[24] Hang Su, Varun Jampani, Deqing Sun, Subhransu Maji, Evangelos Kalogerakis, Ming-Hsuan Yang, and Jan Kautz. Splatnet: Sparse lattice networks for point cloud processing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2530–2539, 2018. 7

[25] Vinay Venkataraman, Karthikeyan Natesan Ramamurthy, and Pavan Turaga. Persistent homology of attractors for action recognition. In *2016 IEEE international conference on image processing (ICIP)*, pages 4150–4154. IEEE, 2016. 3

[26] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008. 5

[27] Lei Wang, Yuchun Huang, Yaolin Hou, Shenman Zhang, and Jie Shan. Graph attention convolution for point cloud semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10296–10305, 2019. 1, 3, 6, 7

[28] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics (TOG)*, 2019. 1, 3, 8

[29] Chi-Chong Wong and Chi-Man Vong. Efficient outdoor 3d point cloud semantic segmentation for critical road objects and distributed contexts. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 499–514. Springer International Publishing, 2020. 3, 6, 7, 8

[30] Wenxuan Wu, Zhongang Qi, and Li Fuxin. Pointconv: Deep convolutional networks on 3d point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9621–9630, 2019. 3

[31] Fenggen Yu, Kun Liu, Yan Zhang, Chenyang Zhu, and Kai Xu. Partnet: A recursive part decomposition network for fine-grained and hierarchical shape segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9491–9500, 2019. 1, 7

[32] Zilong Zhong, Zhong Qiu Lin, Rene Bidart, Xiaodan Hu, Ibrahim Ben Daya, Zhifeng Li, Wei-Shi Zheng, Jonathan Li, and Alexander Wong. Squeeze-and-attention networks for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 3

[33] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005. 2