

ROSS: Robust Learning of One-Shot 3D Shape Segmentation

Shuaihang Yuan
 New York University
 NYU Multimedia and Visual Computing Lab
 sy2366@nyu.edu

Yi Fang*
 New York University
 NYU Multimedia and Visual Computing Lab
 yfang@nyu.edu *

Abstract

3D shape segmentation is a fundamental computer vision task that partitions the object into labeled semantic parts. Recent approaches to 3D shape segmentation learning heavily rely on high-quality labeled training datasets. This limits their use in applications to handle the large scale unannotated datasets. In this paper, we proposed a novel semi-supervised approach, named Robust Learning of One-Shot 3D Shape Segmentation (ROSS), which only requires one single exemplar labeled shape for training. The proposed ROSS can generalize its ability from a one-shot training process to predict the segmentation for previously unseen 3D shape models. The proposed ROSS is composed of three major modules for 3D shape segmentation as follows. The global shape descriptor generator is the first module that utilizes the proposed reference weighted convolution to learn a 3D shape descriptor. The second module is a part-aware shape descriptor constructor that can generate weighted descriptors from a learned 3D shape descriptor according to semantic parts without supervision. The shape morphing with label transferring works as the last module. It morphs the exemplar shape and then transfers labels from the transformed exemplar shape to the target shape. The extensive experimental results on 3D mesh datasets demonstrate the ROSS is robust to noise and incomplete shapes and it can be applied to unannotated datasets. The experiment shows the proposed ROSS can achieve comparable performance with the supervised method.

1. Introduction

Segmentation is the process of partitioning a shape into multiple meaningful parts, making it easier to extend to other applications, such as 3D medical imaging [13], 3D scene understanding [27], 3D object detection [26, 18] and human pose estimation [24]. The desired algorithm should accurately distinguish different parts of the object, even

with structural variations [4, 28] and noise [6]. To address the challenge issued above, traditional methods use a non-learning based feature to represent the shape of objects [23, 2, 8, 7]. However, non-learning based features are often not robust enough to deal with noises and structural variations present in various kinds of 3D models, which limits the generalization capability of a hand-crafted 3D shape segmentation model.

To address the aforementioned limitation, deep learning approaches have been popularly employed to extract learning-based features [31, 10, 32, 21, 29, 17] and the segmentation result outperforms those of traditional methods greatly. The majority of learning-based methods achieve promising performance based on supervised learning from a large volume of a well-annotated dataset. However, it is often costly (sometimes not practical) to have a large volume of well-annotated shape segmentation datasets for the training of deep neural networks. To alleviate the dependency of a well-labeled dataset, we propose a novel semi-supervised learning approach, named Robust Learning of One-shot 3D shape segmentation (ROSS), which requires only one single labeled exemplar shape for training and can generalize its ability to predict 3D segmentation for unseen shape models. To realize the proposed ROSS, three main components, as shown in Figure 1, are included in the development. The first component is a global shape descriptor generator which incorporates neighboring information by our reference weighted convolution and encodes into one informative contextual representation. The second component is a part-aware shape descriptor constructor where attention mechanism is applied to generate the part sensitive signature, namely, part-aware shape descriptor. The third component is shape morphing with label transferring, in which we transform the labeled exemplar shape into one segmented transformed shape with the same shape as input. Then with label transferring from transformed shape to input shape, our model can predict segmentation for unseen shapes. The main contributions of this paper are summarized as follows:

- We propose the reference weighted convolution in the

*Yi Fang is the corresponding author: yfang@nyu.edu

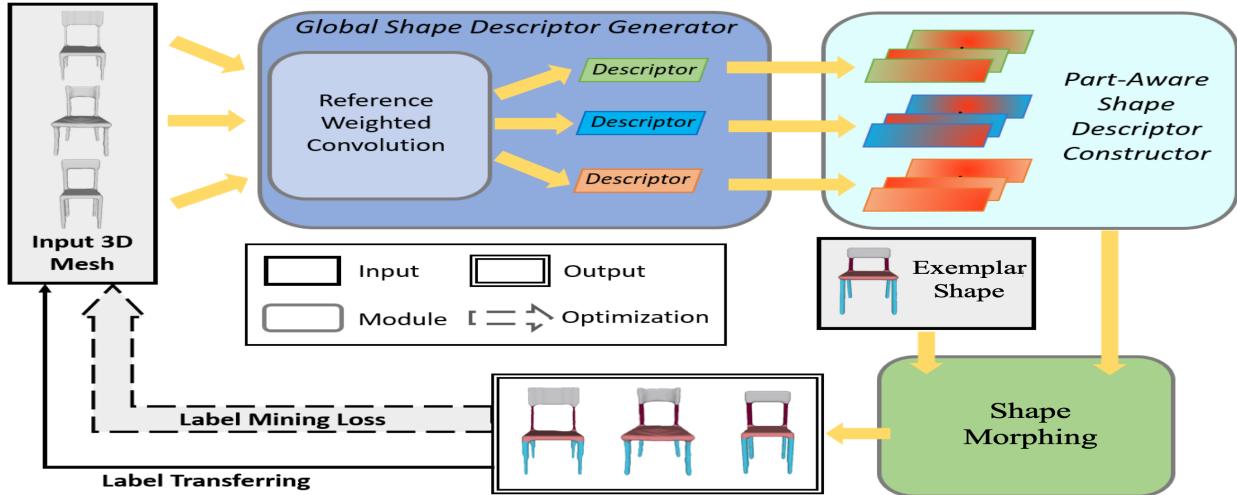


Figure 1: The pipeline of proposed one-shot paradigm which segments 3D mesh by morphing and transferring label from transformed exemplar shape to target shape. The paradigm contains three key components: global shape descriptor generator, part-aware shape descriptor constructor and shape morphing with label transferring.

global shape descriptor generator, which incorporates the surrounding information into the descriptor.

- We introduce the concept of part-aware shape descriptor, which utilizes the attention mechanism to highlight part information.
- We derive the ROSS to deal with the 3D shape segmentation problem by morphing the labeled mesh and segmenting the input mesh by label transferring.
- The proposed ROSS can even outperform some supervised learning methods on multiple 3D datasets.

In the rest of the paper, we first introduce the related work in section 2. Then, we describe the details of the proposed ROSS in section 3. In section 4, we present our experimental results on several datasets. At last, we discuss the limitation of our paradigm and make a brief conclusion in Section 5.

2. Related works

With the rapid development of computer vision, it requires advanced techniques to deal with the increasingly growing of 3D content. Segmentation is one of the long-standing problems in visual recognition. In this part, we will briefly go through some typical non-learning based methods and learning-based methods on 3D mesh segmentation.

2.1. Non-learning based methods on segmentation

As a basic problem in computer vision, a considerably wide range of approaches have been conducted for 3D

shape segmentation. For example, K-Means [23] selects a subset of k seed faces to represent sub-clusters by continually selecting the further face from which has been selected. Fitting Primitives [2] merge the best fitting pairs by approximating every face segment to its adjacent part. Randomized cuts [8] make binary splits by iteratively selecting the most consistent cuts in the randomized set. HMS [7] uses a physics-based approach to characterize vertices on a surface in the heat kernel featured space. Meanwhile, some researchers apply handmade geometric descriptors to mesh segmentation [3, 19]. In the approach proposed by Belongie et al. [3], they attach the descriptor of the whole shape to each point and construct correspondences between two similar shapes. Though it is a novel thought of combining each point with the global shape descriptor, this paper does not provide a reliable way to extract shape descriptor. Later on, data-driven approaches are proposed, which generally extract common global features.

2.2. Learning based methods on segmentation

As Convolutional Neural Network (CNN) has achieved outstanding performance on tasks for 2D images, many recent works apply CNN to 3D model analysis. Since CNN has a stable performance on extracting the geometric feature of 3D shapes, deep learning-based methods presented in [11, 25] outperform many traditional methods. In 2017, Kalogerakis proposed a deep architecture with Fully Convolutional Networks [20] and surface-based Conditional Random Fields (CRF) [16]. However, such a view-based reasoning approach is demanding for the view selections. DCN [32] proposed by Haotian Xu et al., which achieves the

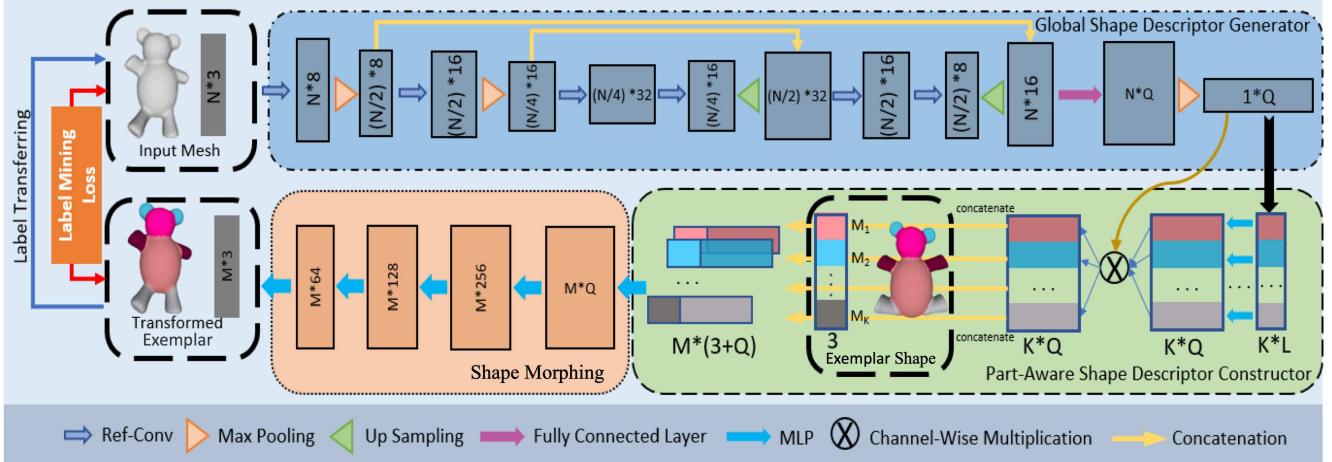


Figure 2: Our one-shot paradigm for 3D mesh object part segmentation task. As the figure shows, the blue box on the top of the figure represents the process of pose feature learning. we adopt several referenced weighted convolution layers, max-pooling layers, and upsampling layers. A global max-pooling layer is added on top of the last linear layer to extract a global shape descriptor where N denotes the total number of points in the mesh and Q represents the length of the global shape descriptor. Then, the part-aware shape descriptor constructor is used to generate a special representation of pose feature for each part by a channel-wise multiplication between global shape descriptor and learned part attention. K denotes different K possible labels. With the informative part-aware shape descriptor, our shape morphing module can generate a precise transformed exemplar shape which consists of M points, and it has nearly uniform pose as input mesh.

state-of-art performance, takes raw geometric features as input then learns a global feature and a local feature with two neural networks and optimizes the segmentation with CRF. Though these learning-based methods generally overcome the shortcomings of non-learning methods and demonstrate persuasive performance on multiple large scale 3D shape datasets, all of them require tons of annotated data in the training stage. The high dependency and high cost of training data annotation drive us to develop the ROSS.

3. Method

In this section, we illustrate our one-shot part segmentation method in detail. The paradigm is presented in Figure 2. We first define our task and approach in section 3.1. Then we propose a novel global shape descriptor learning method in section 3.2. In section 3.3, we demonstrate how we generate a part-aware shape descriptor from the global shape descriptor. Finally, section 3.4 shows the detail of shape morphing and label transferring between the exemplar and input mesh.

3.1. Problem statement

The segmentation task on 3D meshes is to 1) group faces and vertices, which share the same part label k , into a cluster and 2) distinguish which faces and vertices belong to the same part label k . Thus, we propose that the segmentation task can be done by using shape morphing. Since,

first, different parts within exemplar are already been distinctly segmented, which satisfy the requirement of segmentation task. Second, the label of a point in a specific part will not change after transformation, (e.g. a point in the human head may not belong to another class, say chest, after transformation). Third, the shape configuration relationship between different parts may keep the same, (e.g. the head and the hand may not exchange their position). Supposing the transformed mesh is in the same shape as the input mesh, then, the input mesh can be labeled corresponding to the segmentation in transformed mesh naturally. Although our paradigm does not know labels of input 3D mesh, it can still produce accurate segmentation results by using only one labeled exemplar, namely, a one-shot manner. Consequently, our goal is, given input 3D meshes and one exemplar, to return one labeled mesh with the same shape as input and the part labels from the exemplar. Supposing there are n points in a 3D mesh, each point p has the relationship $p \in P$ where P is the point set with $P = \{p_1, p_2, \dots, p_n\}$, $p_i \in \mathbb{R}^D$, and D stands for the dimension number of points set P . Here, we set $D = 3$ to represent our 3D mesh data and $p_i = [x_i, y_i, z_i]$ for the point position in the 3D-coordinate. Compared to other 3D data types, mesh data not only possess substantial point information, but also contain an unique propriety of geometric surface information, formed by connecting adjacent vertices adj where $adj = \{adj_1, \dots, adj_i\}$, $adj_i \in P$. adj_i denotes i^{th} adjacent point of the center point and we have

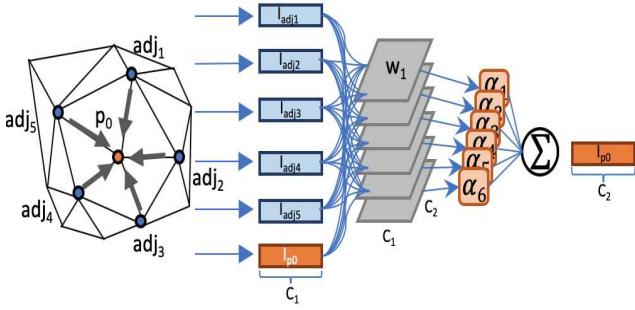


Figure 3: The schema of reference weighted convolution. adj_i represents adjacent points of p_0 and I denotes point signatures. W are weight matrices standing for an adjacent position of point p_0 with $C_1 \times C_2$ parameters. α_i is a learned reference weight by Eq. 2

$adj \subset ADJ$ where ADJ represent all adjacent points for point set P .

3.2. Global shape descriptor generator

To capture the spatial information within the mesh data, much previous work incorporates the calculation of face normal and curvature. However, this approach requires large additional computational cost. Also, since the geometric information in 3D shape meshes and other graph-structured data are not limited in regular grids, simple stacking of CNNs will lose association information within points. Here, we present a novel reference weighted convolution layer that can consider neighboring information and directly taking raw mesh points as input without any data prepossessing nor face normal and curvature calculation. Our mechanism is illustrated in Figure 3.

Reference weighted convolution. For a point p , we approximate the information viewed from the perspective of adjacent points by adopting γ different learnable weights, where γ represents the number of adjacent points. Thus, the point signature I'_p for point p viewed from γ adjacent points is defined as

$$I'_p = \sum_{j=1}^{\gamma} W_j x_p + b. \quad (1)$$

where x_p denotes point signature of p , $W_j x_p$ represents information extracted from j^{th} adjacent point and b represents the bias. However, the information generated from different adjacent points should be considered differently. We adopt a reference attention mechanism that enables our module to pay various attention to the different extracted information. Our approach is presented in Figure 3. We use γ different Multilayer Perceptrons (MLPs) to approximate different weights which are defined in Eq. 2.

$$\alpha_j = MLP_j(e^{|(x_{adj_j} - x_p)|}). \quad (2)$$

Where MLP_j and x_{adj_j} stands for the j^{th} MLP and point signatures for adjacent point j . α_j is the reference attention. After applying the reference weight α_j to I'_p in Eq. 1, the final learned point signature of p is defined as

$$I_p = \sum_{j=1}^{\gamma} \alpha_j W_j x_p + b. \quad (3)$$

Descriptor learning. Inspired by [29], As shown in the blue section of the Figure 2, we use an elaborated structure to compute global features, which contains reference weighted convolution layers, max-pooling layers, upsampling layers, and concatenation. Our network takes a point set of mesh as and the adjacent points as input and then outputs a global shape descriptor Des_{global} of size $1 * Q$. We use the shape descriptor learning function $E(\cdot)$ to simulate the network which is defined below,

$$Des_{global} = E(P, ADJ). \quad (4)$$

As we defined above, P denotes the input mesh point set, and ADJ represents all adjacent points.

3.3. Part-aware shape descriptor constructor

In this part, we present an approach to construct a part-aware shape descriptor by multiplying the learned global shape descriptor with different part attentions which are learned by multiple MLPs.

To generate different part attentions, the learned global shape descriptor is segmented into k parts with each part having the channel size of $L = Q/k$. Segments are represented as $Des_1, Des_2, \dots, Des_k$. As presented in the green box of Figure 2, we use k different MLPs to learn part attentions with hidden layer of size L and Q . The learned k^{th} part attention A_k is defined as

$$A_k = MLP_k(Des_k). \quad (5)$$

Then, k part weighted descriptors are generated by performing a multiplication between the global shape descriptor and different part attentions. The function is

$$D_{PA_k} = Des \odot A_k. \quad (6)$$

Where \odot is a channel-wise multiplication and D_{PA_k} represents k^{th} part weighted descriptors. Finally, the point set P_t of a labeled exemplar is rearrange to k groups according to part labels and each group is represented as $P_t^1, P_t^2, \dots, P_t^k$. We use an concatenation function to fuse exemplar with part weighted descriptor and to construct part-aware shape descriptor D_{PA} , which is

$$D_{PA} = cat(D_{PA_1}, D_{PA_2}, D_{PA_3} \dots D_{PA_k}). \quad (7)$$

where $cat(\cdot)$ is the abbreviation of concatenation.

Table 1: The quantitative results for segmentation accuracy on PSB dataset. “-” represents unreported results. ROSS denotes our proposed method, and ROSS w/ PointNet is model by replacing our global shape descriptor (refer to section 3.2) with PointNet for shape descriptor learning.

Category	Segmentation Accuracy						
	Supervised Method					One-Shot Method	
	Kalogerakis[15]	Wang[30]	Guo[12]	ShapePFCN[14]	DCN[32]	ROSS w/ PointNet	ROSS
Human	0.9320	0.5560	0.9122	0.9380	0.9408	0.9183	0.9274
Cup	0.9960	0.9960	0.9973	0.9370	0.9979	0.9700	0.9900
Glasses	0.9720	-	0.9760	0.9630	0.9869	0.8743	0.9071
Airplane	0.9610	-	0.9667	0.9250	0.9766	0.9472	0.9623
Ant	0.9880	-	0.9880	0.9890	0.9898	0.9477	0.9582
Chair	0.9840	0.9960	0.9867	0.9810	0.9935	0.9326	0.9574
Octopus	0.9840	-	0.9879	0.9810	0.9934	0.9067	0.9166
Table	0.9930	0.9960	0.9955	0.9930	0.9959	0.9389	0.9524
Teddy	0.9810	-	0.9824	0.9650	0.9908	0.9274	0.9372
Hand	0.8870	-	0.8871	0.8870	0.8861	0.7157	0.7589
Plier	0.9620	-	0.9622	0.9570	0.9714	0.9117	0.9417
Fish	0.9560	-	0.9564	0.9590	0.9705	0.8533	0.8933
Bird	0.8790	-	0.8835	0.8630	0.9039	0.8747	0.8943
Armadillo	0.9010	-	0.9227	0.9330	0.9382	0.8012	0.8570
Bust	0.6210	-	0.6984	0.6640	0.7898	0.7187	0.7656
Mech	0.9050	0.9130	0.9560	0.9790	0.9660	0.9287	0.9523
Bearing	0.8660	-	0.9246	0.9120	0.9470	0.8594	0.8835
Vase	0.8580	0.9050	0.8911	0.8570	0.8931	0.8111	0.8211
FourLeg	0.8620	0.5430	0.8702	0.8950	0.8742	0.7995	0.8455
Average	0.9204	0.8436	0.9362	0.9251	0.9476	0.8756	0.9011

3.4. Shape morphing and label transferring

Then we generate a function $\mathcal{M}(\cdot)$ to perform the transformation from exemplar to input source mesh shape. $\mathcal{M}(\cdot)$ is formulated by a multi-layer perceptron network with hidden layer sizes of 1024, 512, 256, 128 and 3. The function $\mathcal{M}(\cdot)$ is defined as

$$P_d = \mathcal{M}(D_{PA}). \quad (8)$$

The function takes learned part-aware shape descriptor from the input source mesh as input and output a transformed exemplar. To make our model an end-to-end learnable model with a one-shot manner, we proposed a one-shot label mining loss for our network. Inspired by [9], our losses adopt three terms including the Chamfer loss $L_c(P_d, P_s)$,

the Edge loss $L_e(P_d, P_s)$ and the Laplacian loss $L_l(P_d, P_s)$. The Chamfer loss makes sure tow shapes of the different poses are as similar as possible while Edge loss and Laplacian loss ensure all clusters are not isolated and reduce the distortion. The final label mining loss is defined as

$$Loss = L_c(P_d, P_s) + L_e(P_d, P_s) + L_l(P_d, P_s). \quad (9)$$

Cluster label assigning. By minimizing three one-shot loss term to optimize the transformation process thus making sure the transformed exemplar P_d is as close to P_s as possible. We use a straight forward method to assign part labels from each point in transformed exemplar to target mesh. To retrieve labels from P_d to P_s , we adopt a label mapping method. That is, for all $p_i \in P_d$, we assign

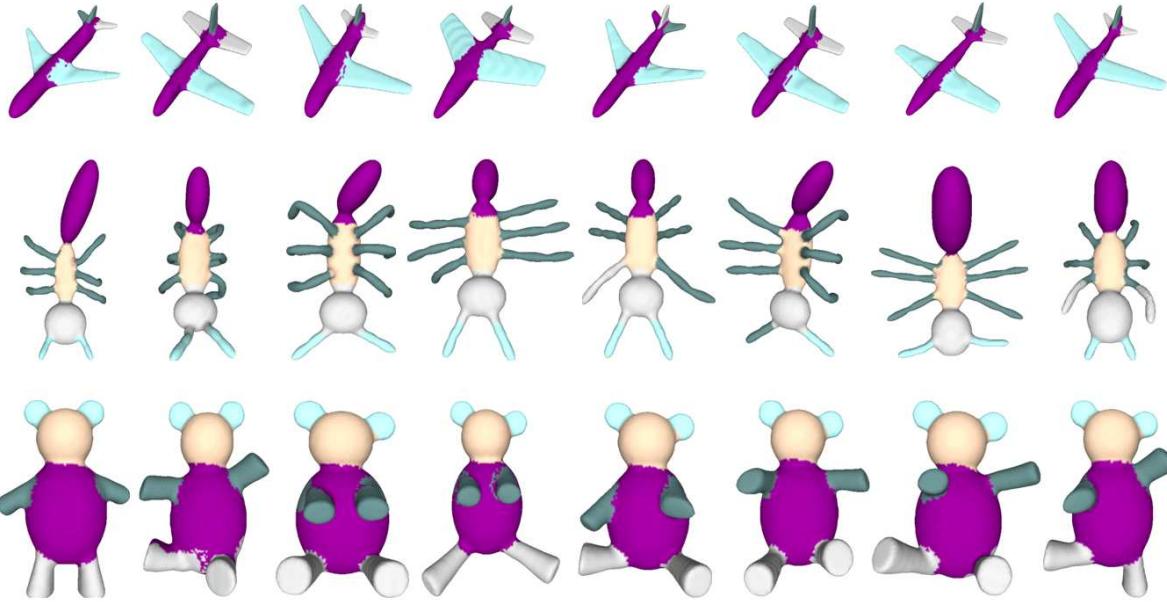


Figure 4: Visualization of part segmentation task results on PSB [5] benchmark dataset for airplane, ant and teddy bear.

the label of p_i to the point which is the nearest one to p_i in target shape P_s .

4. Experiment

4.1. Dataset and experimental setups

In this section, the following benchmark datasets are used to evaluate our proposed model. Princeton Segmentation Benchmark (PSB) [5] is mainly used to compare the ROSS with other different approaches toward 3D mesh segmentation tasks since it is an open dataset and many prior works are evaluated based on this benchmark dataset. PSB dataset carries 19 different object types (including Airplane, Bust, Glasses, Teddy, etc.) and each type has 20 meshes with different shape configurations. Moreover, we use the SCAPE [1] benchmark dataset which contains 70 real human scan meshes with different poses and a skeleton with 12 different kinds of labels for different parts. The SCAPE dataset also provides incomplete human scans that meshes contain holes or missing faces. The usage of the SCAPE dataset shows the proposed ROSS is robust enough to perform part segmentation tasks on incomplete meshes. Note that the SCAPE dataset is originally used for the task of 3D shape completion, which contains 70 unlabeled meshes. To use it as a shape segmentation dataset, we generate part segmentation by transferring the labels from the skeleton to each shape using the ground-truth shape correspondences of the SCAPE dataset.

The setup of our experiment in the PSB dataset is similar to in [5]. For each category in the PSB dataset, we first use the dataset to train our ROSS before the evaluation test. We

sample 2500 points from each category according to their area of adjacent faces to train the model. We select meshes that have the same or similar configuration with our selected exemplar as testing meshes to ensure the consistency between our selected exemplar and testing meshes. Moreover, we evaluate the ROSS on 70 meshes of the SCAPE dataset. We report average accuracy for different categories of the SCAPE testing dataset and the PSB dataset. Following [9], we also adopt a regression step on a part-aware shape descriptor constructor and Shape morphing step to ensure a well-transformed exemplar. The segmentation of our proposed system is at point (vertex) level, and we can transfer it to the facet level. For a given triangular facet, we set the facet label to the majority label of three vertices. If three vertices all have different labels, then we select the first vertex label as the facet label.

The proposed ROSS is implemented using Pytorch 0.4. We optimize our network by Adam optimizer for 4000 iterations which have an initial learning rate of 10^{-3} . Our network is running on a GeForce GTX 1080 Ti GPU.

4.2. Segmentation performance

We carried out a test to verify the performance of ROSS on the shape segmentation task. The proposed ROSS is a semi-supervised approach, which is only trained with one single exemplar shape with ground-truth part segmentation label, and then be used in test shape segmentation. As the proposed ROSS is a one-shot method to address the 3D shape segmentation, we do not have other similar semi-supervised methods as comparison baseline models in this test. Instead, we compared ROSS to supervised shape seg-

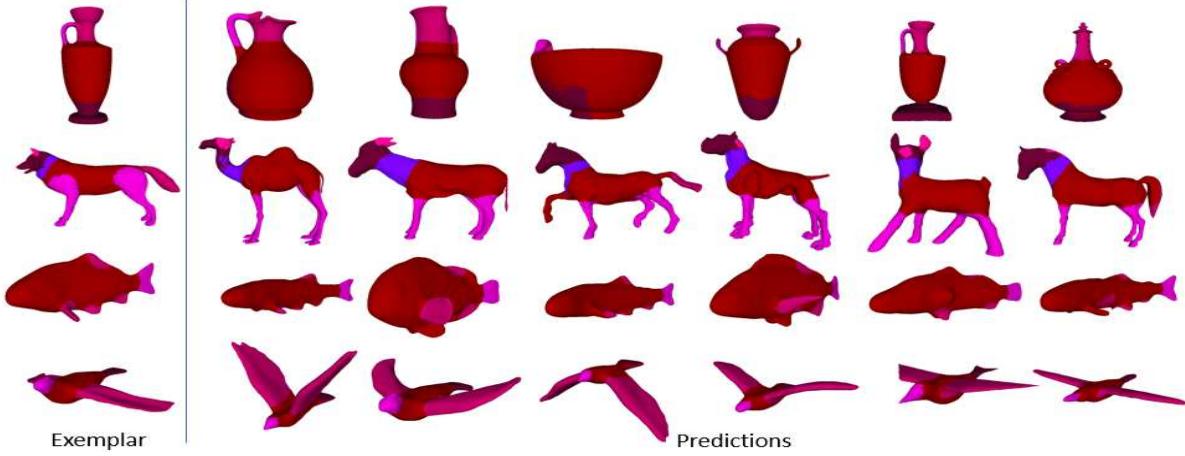


Figure 5: The visualization result of our proposed method. The shape configuration between the selected exemplar and the testing set can be significantly different.

mentation models, which were trained on a large number of annotated shape models, to demonstrate the performance in shape segmentation by ROSS. As what other segmentation tasks reported, the performance is evaluated using facet label accuracy. The formula of accuracy is defined as

$$Acc = \frac{P \cap G}{F}. \quad (10)$$

where P is the prediction made by our network, G is face label ground truth and F is the total number of faces in the mesh. The quantitative results are presented in Table 1 including publicly reported result and our experiment result based on the PSB dataset. As we can see, our proposed method can also achieve outstanding performance. It even outperforms one supervised method [15] by 5.75% and our method is only 4.65% less than the state-of-art supervised model. Moreover, we perform an experiment on the SCAPE dataset for both complete mesh and incomplete mesh. Although the SCAPE dataset contains many different human poses and different human body shapes, the proposed ROSS can still achieve a good quantitative result of 94.02% for the segmentation accuracy on complete human meshes and 92.82% on incomplete human meshes. This experimental test suggests that the ROSS is an effective semi-supervised model in learning of one-shot shape segmentation.

4.3. Visualized examples

In this part of the experiment, we present visualized results of part segmentation task by using our proposed model as well as other published supervised approaches based on the PSB dataset. Figure 4 shows the visualized results of our work. In addition to showing the generality of our model, we also conduct an experiment on the SCAPE dataset beside the PSB dataset. Figure 6 shows our visualized result on the SCAPE dataset. The Figures 4, 6 clearly show our

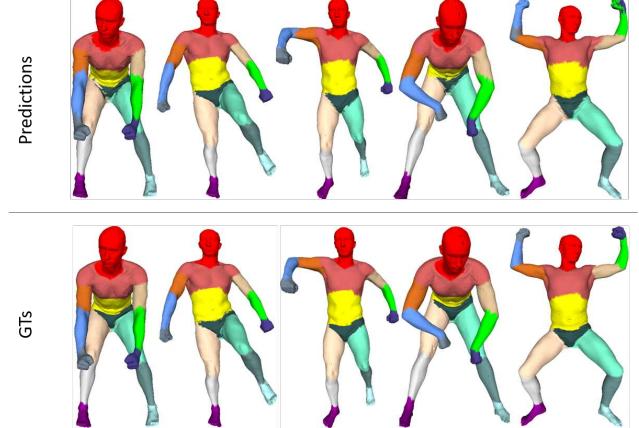


Figure 6: Visualization of part segmentation task result on SCAPE [1] benchmark dataset for complete human meshes. The first row is the prediction result of our model and the second row represents the ground truth.

one-shot method can achieve a robust performance on various types of models from different datasets.

4.4. Robust test

To demonstrate that our model can be applied to various situations (i.e. incomplete shape segmentation task), we provide an experiment that performs the part segmentation task on incomplete meshes with holes randomly distributed on each human scan in the SCAPE dataset. Since the dataset contains no ground truth labels for incomplete human scans, we manually generate ground truth labels for incomplete meshes to calculate the segmentation accuracy. Following the same ground truth generated manner we illustrate in the first subsection, we first align all incomplete meshes and

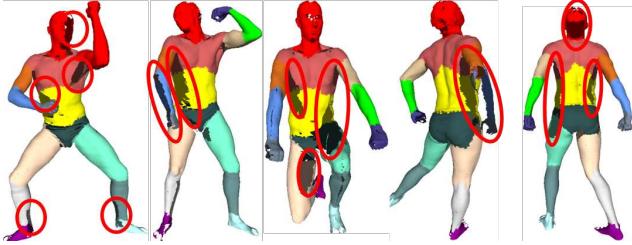


Figure 7: Visualization of part segmentation task results on SCAPE [1] for real human scans with holes. Due to the snapshot condition, dark parts on the human mesh are holes.

complete one to their bounding boxes. Then we generate all labels for complete meshes, finally, we use the nearest neighbor method to map part labels from complete meshes to incomplete meshes. Figure 7 shows the visualized result for the part segmentation task on incomplete meshes based on our generated ground truth. Our proposed ROSS can reach an accuracy of 94.02% for the labeling task in the complete mesh. Although there exist losses of some part information of an incomplete mesh, the proposed ROSS can still capture local parts and global information to deliver a high segmentation performance. It can have the segmentation accuracy of 92.82% for incomplete meshes, which only 1.2% lower than the accuracy on the complete meshes. Apparently, both the quantitative result and visualized results suggest our performance can reach a good level hence our method is a robust approach in various 3D Mesh segmentation situations. Moreover, to test our proposed method that can handle large shape configuration variations, we conduct an experiment on the PSB dataset and figure ?? shows that our proposed method can be applied on various shape configurations of the same categories.

4.5. Ablation study

Table 2: Comparison results between reference weighted convolution-based method and PointNet based method [22] on SCAPE dataset.

Method	Data Type	Average Accuracy
ROSS w/ PointNet	Complete	0.9172
	Incomplete	0.9017
ROSS	Complete	0.9402
	Incomplete	0.9282

In this section, we perform the experiment to evaluate the

effectiveness of our global shape descriptor based on reference weighted convolution (as described in section 3.2) for the shape descriptor learning. Here, we conduct experiments on the SCAPE dataset on both incomplete meshes and complete meshes. We replaced the shape descriptor module in ROSS with a popular feature learning method, PointNet [22], and keep the rest of the ROSS the same to form a new model named ROSS w/ PointNet. We compare the segmentation accuracy between the ROSS w/ PointNet and ROSS in the experiment. Table 2 shows segmentation accuracy on SCAPE dataset with quantitative comparison result. ROSS outperforms the ROSS w/ PointNet by 2.30% and 2.65% respectively on complete and incomplete datasets, which verifies the effectiveness of our global shape descriptor using reference weighted convolution.

4.6. Failing case analysis

According to a careful analysis of visualized results presented in Figure 4, 6 and 7, not only the ROSS but also other prior works including state-of-art method suffer from the low prediction accuracy of labels around edges. This is mainly because we use Euclidean distance to determine the correspondences between the points of an input mesh and transformed exemplar, which is not so stable when points of different parts intersect with each other. Another common failing case comes from the transformation on a pair of shapes with great deformation (e.g. a significant pose change). For instance, as the current setting of ROSS, it is not challenging to align a 3D model with significant different shape configuration as exemplar shape, which leads to the sub-optimal shape segmentation performance as shown in the figure. Our future work will address this through the development of methods for learning shape descriptors, which captures more shape deformation.

5. Discussion and conclusion

In this paper, we proposed a novel semi-supervised approach that requires only one single exemplar labeled shape for training. The proposed approach can generalize its ability from a one-shot training process to predict the segmentation for previously unseen 3D shape models. We introduce a new reference weighted convolution layer which can successfully learn surface information based on adjacent points. The part-aware shape descriptor helps our propose ROSS perform robust and accurate transformation between the exemplar and input 3D mesh. The experiments conducted on PSB [5] dataset indicate that our novel approach can even outperform some supervised methods. Visualized and quantitative results on both complete meshes and incomplete meshes in the unannotated SCAPE [1] dataset show our proposed approach is robust which also proves that our approach can be applied to large scale unlabeled 3D mesh dataset.

References

- [1] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. Scape: Shape completion and animation of people. *ACM Trans. Graph.*, 24(3):408–416, July 2005.
- [2] M. Attene, B. Falcidieno, and M. Spagnuolo. Hierarchical mesh segmentation based on fitting primitives. *The Visual Computer*, 22:181–193, 03 2006.
- [3] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24:509–522, 2001.
- [4] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Efficient computation of isometry-invariant distances between surfaces.
- [5] X. Chen, A. Golovinskiy, and T. Funkhouser. A benchmark for 3D mesh segmentation. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 28(3), Aug. 2009.
- [6] Y. Fang, M. Sun, M. Kim, and K. Ramani. Heat-mapping: A robust approach toward perceptually consistent mesh segmentation. *IEEE Computer Vision and Pattern Recognition*, 01 2011.
- [7] Y. Fang, M. Sun, M. Kim, and K. Ramani. Heat-mapping: A robust approach toward perceptually consistent mesh segmentation. *IEEE Computer Vision and Pattern Recognition*, 01 2011.
- [8] A. Golovinskiy and T. Funkhouser. Randomized cuts for 3D mesh analysis. *ACM Transactions on Graphics (Proc. SIGGRAPH ASIA)*, 27(5), Dec. 2008.
- [9] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry. 3d-coded: 3d correspondences by deep deformation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 230–246, 2018.
- [10] K. Guo, D. Zou, and X. Chen. 3d mesh labeling via deep convolutional neural networks. *ACM Trans. Graph.*, 35(1):3:1–3:12, Dec. 2015.
- [11] K. Guo, D. Zou, and X. Chen. 3d mesh labeling via deep convolutional neural networks. *ACM Transactions on Graphics (TOG)*, 35(1):3, 2015.
- [12] K. Guo, D. Zou, and X. Chen. 3d mesh labeling via deep convolutional neural networks. *ACM Trans. Graph.*, 35(1):3:1–3:12, Dec. 2015.
- [13] T. Heimann. *Statistical Shape Models for 3D Medical Image Segmentation*. VDM Verlag, Saarbrücken, Germany, Germany, 2009.
- [14] E. Kalogerakis, M. Averkiou, S. Maji, and S. Chaudhuri. 3D shape segmentation with projective convolutional networks. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [15] E. Kalogerakis, A. Hertzmann, and K. Singh. Learning 3D Mesh Segmentation and Labeling. *ACM Transactions on Graphics*, 29(3), 2010.
- [16] J. Lafferty, A. McCallum, and F. C. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.
- [17] L. Li, M. Sung, A. Dubrovina, L. Yi, and L. J. Guibas. Supervised fitting of geometric primitives to 3d point clouds. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2652–2660, 2019.
- [18] D. Lin, S. Fidler, and R. Urtasun. Holistic scene understanding for 3d object detection with rgbd cameras. *2013 IEEE International Conference on Computer Vision*, pages 1417–1424, 2013.
- [19] G. Lin, C. Shen, I. D. Reid, and A. van den Hengel. Efficient piecewise training of deep structured models for semantic segmentation. *CoRR*, abs/1504.01013, 2015.
- [20] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
- [21] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 77–85, 2017.
- [22] C. R. Qi, H. Su, K. Mo, and L. 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, 2017.
- [23] S. Shlafman, A. Tal, and S. Katz. Metamorphosis of polyhedral surfaces using decomposition. *Comput. Graph. Forum*, 21(3):219–228, 2002.
- [24] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. *Real-Time Human Pose Recognition in Parts from Single Depth Images*. 2013.
- [25] Z. Shu, C. Qi, S. Xin, C. Hu, L. Wang, Y. Zhang, and L. Liu. Unsupervised 3d shape segmentation and co-segmentation via deep learning. *Computer Aided Geometric Design*, 43:39–52, 2016.
- [26] S. Song and J. Xiao. Deep sliding shapes for amodal 3d object detection in rgbd images. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 808–816, 2016.
- [27] B. soo Kim, P. Kohli, and S. Savarese. 3d scene understanding by voxel-crf.
- [28] H. Tabia, H. Laga, D. Picard, and P.-H. Gosselin. Covariance descriptors for 3d shape matching and retrieval. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR ’14*, pages 4185–4192, 2014.
- [29] N. Verma, E. Boyer, and J. J. Verbeek. Feastnet: Feature-steered graph convolutions for 3d shape analysis. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2598–2606, 2018.
- [30] Y. Wang, M. Gong, T. Wang, D. Cohen-Or, H. Zhang, and B. Chen. Projective analysis for 3d shape segmentation. *ACM Trans. Graph.*, 32(6):192:1–192:12, Nov. 2013.
- [31] Z. Xie, K. Xu, L. Liu, and Y. Xiong. 3d shape segmentation and labeling via extreme learning machine. *Computer Graphics Forum*, 33(5):85–95, 2014.
- [32] H. Xu, M. Dong, and Z. Zhong. Directionally convolutional networks for 3d shape segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2698–2707, 2017.