Xiang Li

Personal Information

Status:	**Post-doctor**
Program:	**Computer Science and Engineering**
School:	**Tandon School of Engineering, New York University**
Website:	**https://xiangli.ac.cn/**
Period:	**From 2019-09 to Present**

Biography

I am a post-doctor at New York University and advised by Professor Yi Fang. During my post doctoral period, I work as a research assistant in NYU Multimedia and Visual Computing (MMVC) Lab. I am broadly interested in 3D Computer Vision and Deep Learning. Now I am a director of project coordination at Inception Institute of Artificial Intelligence, Abu Dhabi.

Research Project: Learning-based Point Correspondence Networks

Description

Point sets correspondence concerns with the establishment of point-wise correspondence for a group of 2D or 3D point sets with similar shape description. Existing methods often iteratively search for the optimal point-wise correspondence assignment for two sets of points, driven by maximizing the similarity between two sets of explicitly designed point features or by determining the parametric transformation for the best alignment between two point sets. In contrast, without depending on the explicit definitions of point features or transformation, we introduce a novel point correspondence neural networks (PC-Net) that is able to learn and predict the point correspondence among the populations of a specific object (e.g. fish, human, chair, etc) in an unsupervised manner. The experimental results demonstrate that PC-Net can establish robust unsupervised point correspondence over a group of deformable object shapes in the presence of geometric noise and missing points.

Method

The proposed PC-Net which is composed of four main components. The first component is "learning global shape descriptor". In this component, the global shape descriptor is learned with a deep neural network to capture global geometric properties. The second component is "forming shape morphing initiator". In this component, a circle or sphere is selected as a template shape that consists of a set of landmark points (i.e. points preserving correspondence between the object and its population). The shape morphing initiator is vector array with each element represented as a vector concatenation of the coordinate of each landmark point with the global shape descriptor. The third component is "Motion-driven Embedding". In this component, landmarks of a template shape morph and conform towards the target shape, guided by the previously characterized shape descriptor. As a result, all landmarks are progressively and coherently drifted from the template shape to corresponding positions on the target shape. In the last component, we map the correspondence of reconstructed landmarks back to the original point sets.

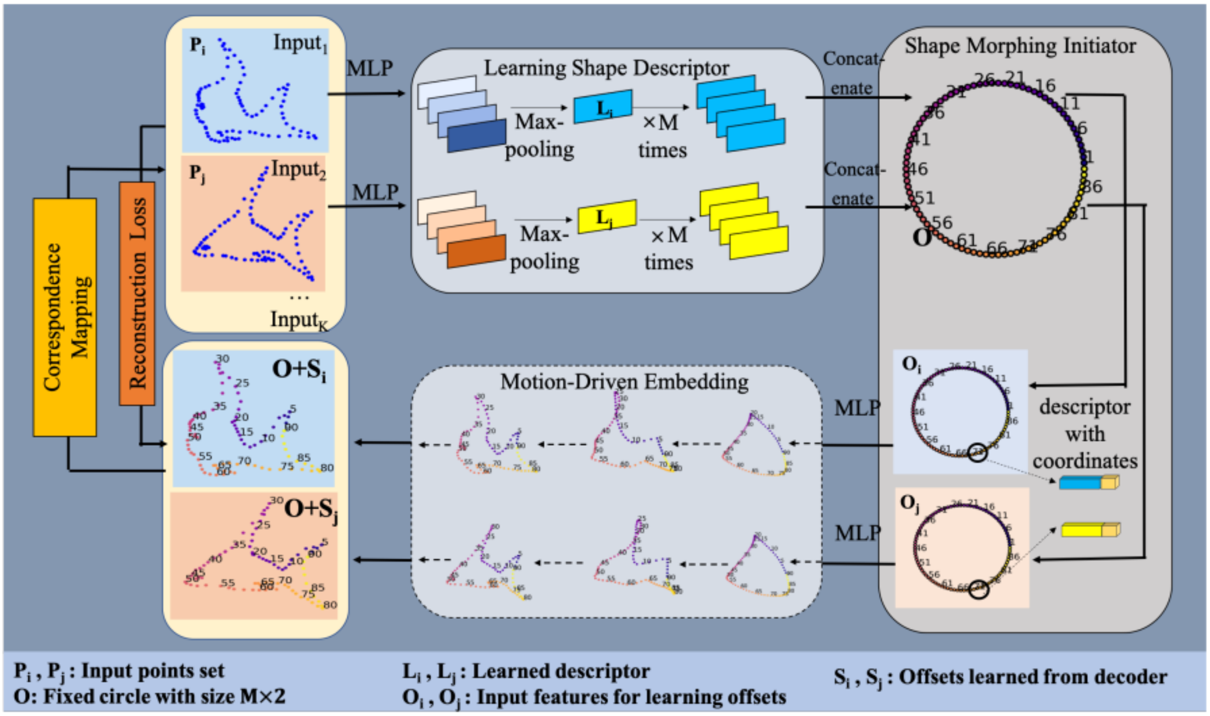


Figure 1: The pipeline of proposed PC-Net model. The pipeline mainly includes four parts.

Results

To evaluate the correspondence performance, we randomly pick out 100 shapes from the training/test dataset and evaluate the pair-wise correspondence among all pairs. For each pair, one is used as a reference shape, and the other one is regarded as the target shape for evaluation. The final correspondence accuracy is calculated as an average over all shape pairs. Experimental results show the robustness of model under deformation, noise and missing points.

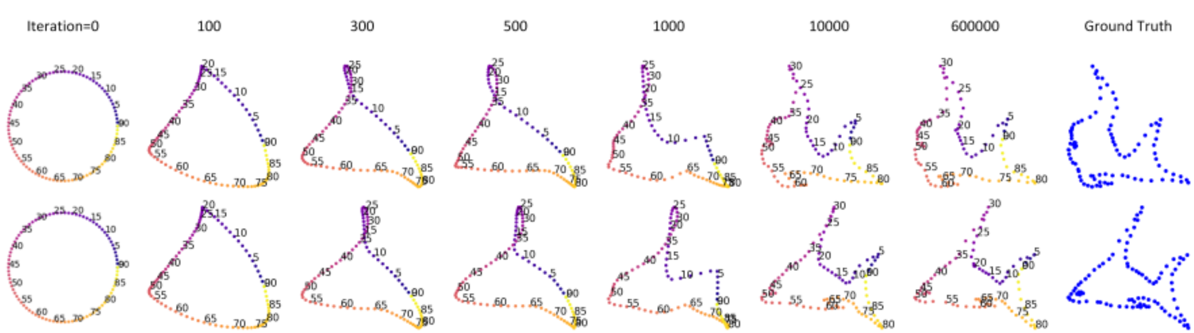


Figure 2: Illustration of our Motion-driven embedding process. Landmark points are numbered with color and the blue fishes in last column are input shapes.

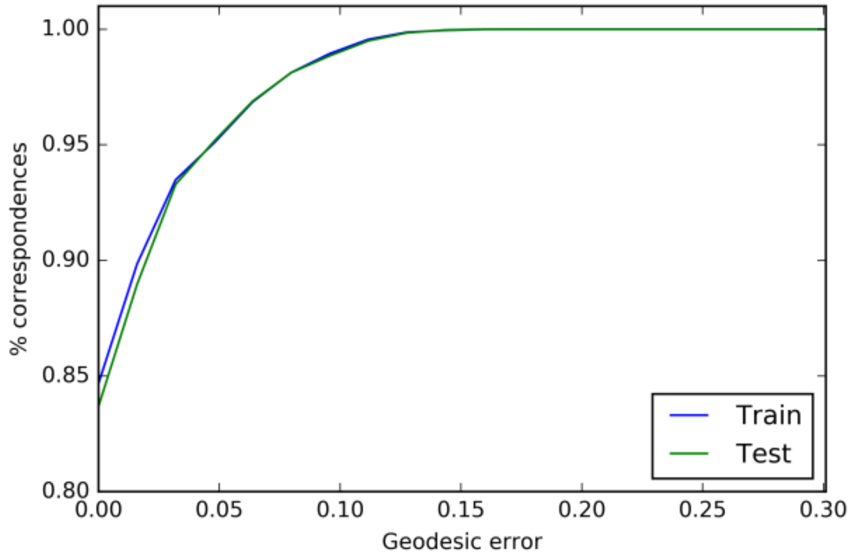
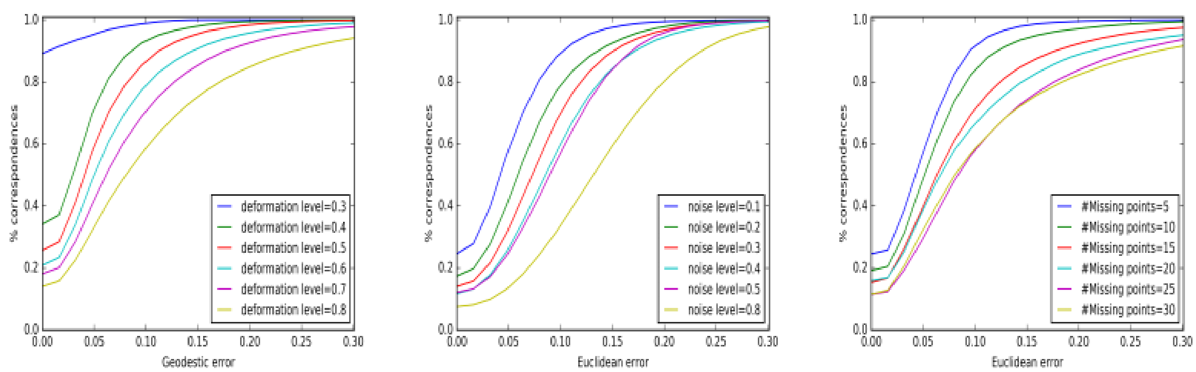


Figure 3: Correspondence performance on test set.



(a) (b) (c)

Figure 4: Robustness test. (a) Correspondence quality at different deformation levels. (b) Correspondences quality at different noise level. (c) Correspondence quality with different number of missing points.

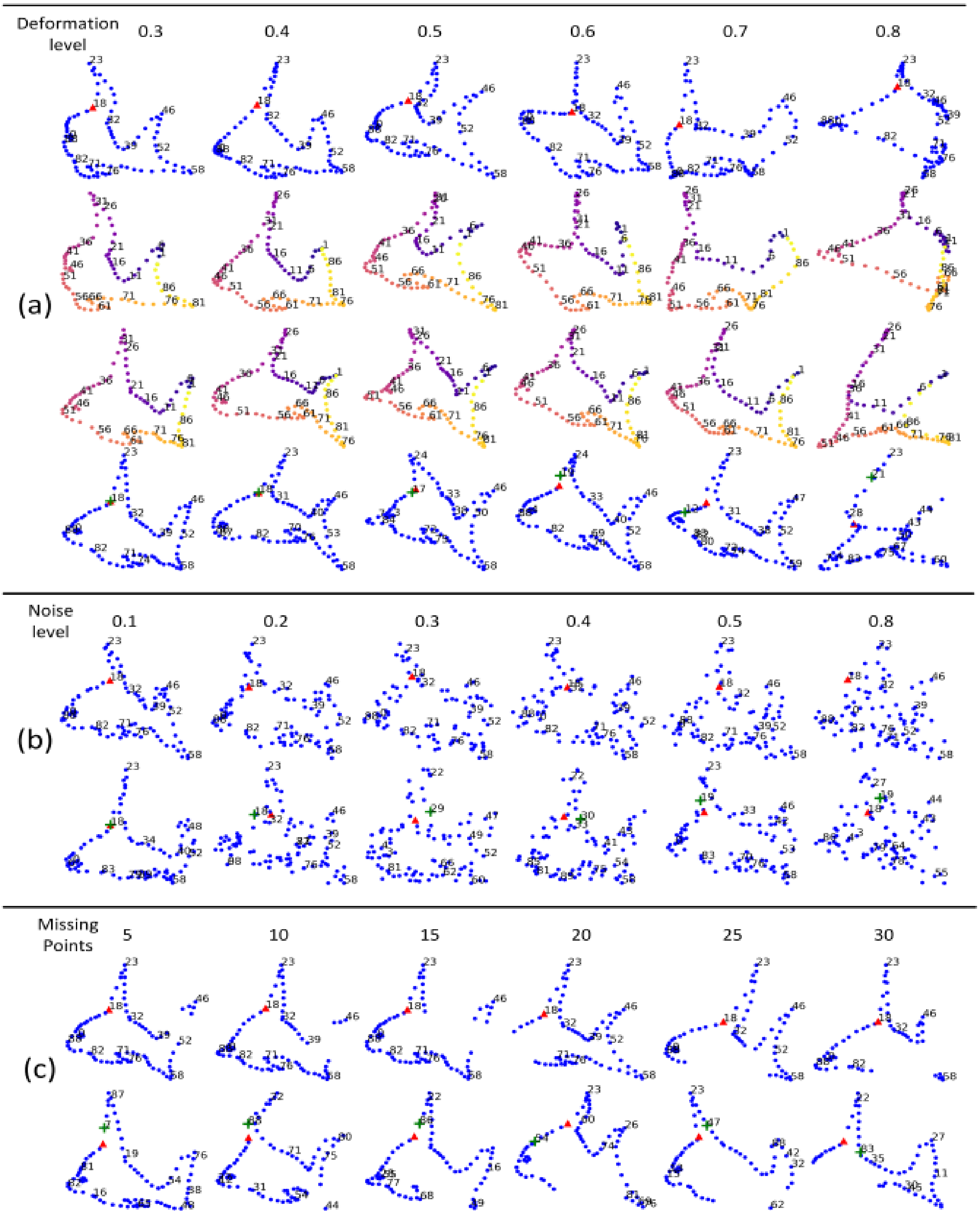


Figure 5: Examples of point correspondence at (a) different deformation level, (b) different noise level, and (c) different number of missing points. The top rows in (a)-(c) show the reference shapes, the middle rows in (a) show the reconstructed shapes, and the bottom rows in (a)-(c) show the predicted shapes with correspondences.

Research Project: Few-shot Learning of Part-specific Probability Space for 3D Shape Segmentation

Description

Most existing supervised methods require a large number of training data with human annotation part labels to guide the training process to ensure the model's generalization abilities on test data. In comparison, we propose a novel 3D shape segmentation method that requires few labeled data for training. Given an input 3D shape, the training of our model starts with identifying a similar 3D shape with part annotations from a minipool of shape templates. With the selected template shape, a novel Coherent Point Transformer is proposed to fully leverage the power of a deep neural network to smoothly morph the template shape towards the input shape. Then, based on the transformed template shapes with part labels, a newly proposed Part-specific Density Estimator is developed to learn a continuous part-specific probability distribution function on the entire 3D space with a batch consistency regularization term. With the learned part-specific probability distribution, our model is able to predict the part labels of a new input 3D shape in an end-toend manner. We demonstrate that our proposed method can achieve remarkable segmentation results on the ShapeNet dataset with few shots.

Method

We propose a novel model, named Weakly Supervised Point Cloud Segmentation Networks (WPS-Net), to realize the 3D point segmentation task assuming the existence of a few labeled training data. Our model aims to directly calibrate a spatially continuous probability function based on a deformed retrieved template to encode part semantic of a 3D point at infinite resolution. The proposed WPS-Net consists of four major components. The first component is "Template Selector." In this component, given an input 3D shape, WPS-Net starts with identifying a similar 3D shape with part annotation from a mini-pool of shape templates. The second component is "Coherent Point Transformer". In this component, a novel Coherent Shape Transformer is proposed to smoothly morph the selected template shape towards the input shape. The third component is "Part-specific Density Estimator". In this component, based on the transformed template shape with part labels, a newly proposed Part-specific Density Estimator is developed to learn a spatially continuous probability function to encode part semantic of a 3D point at infinite resolution with a batch consistency regularization term. The fourth component is "Part-specific Label Predictor." In this component, for a given input shape, the learned Part-specific density estimator is used to assign the part label to each point on the shape. The WPS-Net is able to train and predict the semantic part labels of a 3D shape in an end-to-end manner.

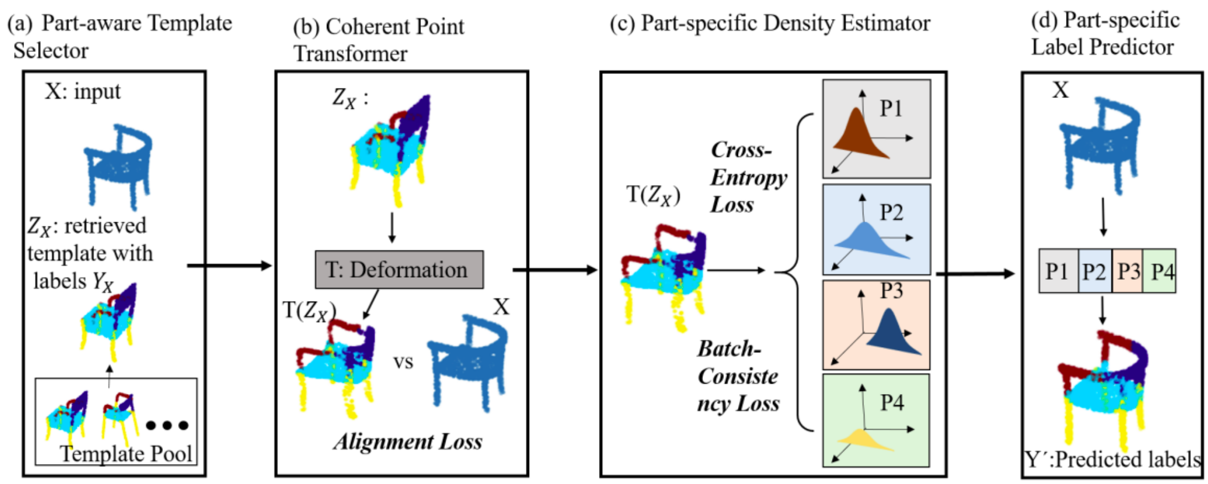


Figure 6: For a given input 3D point cloud, a template shape firstly is retrieved from a template pool by template selector (a). Coherent point transformer (b) morphs the retrieved template towards input shape. In (c), the Part-specific Density Estimator takes the points of deformed templates as input to compute the continuous probability distribution function in 3D space. In (d), for each point of input shape, its label can be predicted by Part-specific label predictor.

Results

We carry out experiments to demonstrate the effectiveness of the modules in our method and evaluate the 3D point cloud segmentation performance of WPS-Net. We report the performance of our method on the test dataset following the official train test split. The mean IoU (Intersection-over-Union) of each category is calculated as an average value over all shapes in that category. The experimental results verifies the effectiveness of each module and achieved a remarkable weakly supervised 3D point cloud segmentation result on the ShapeNet 3D point cloud segmentation dataset.

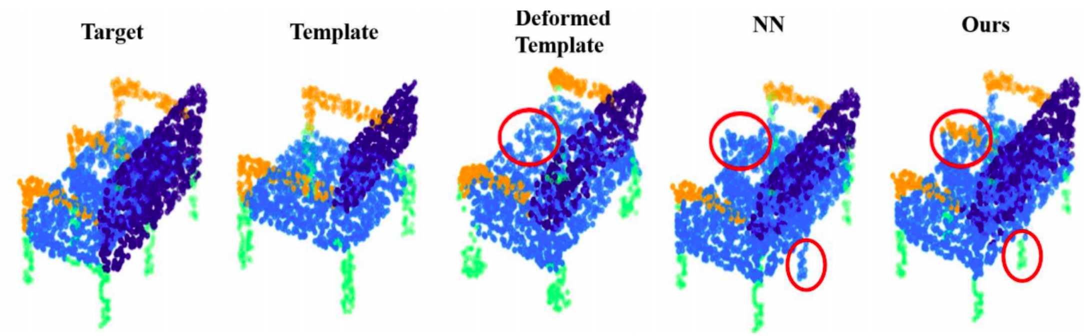


Figure 7: Comparison of the segmentation results based on the deformed template between using Nearest Neighbors method and our part-specific density estimator.

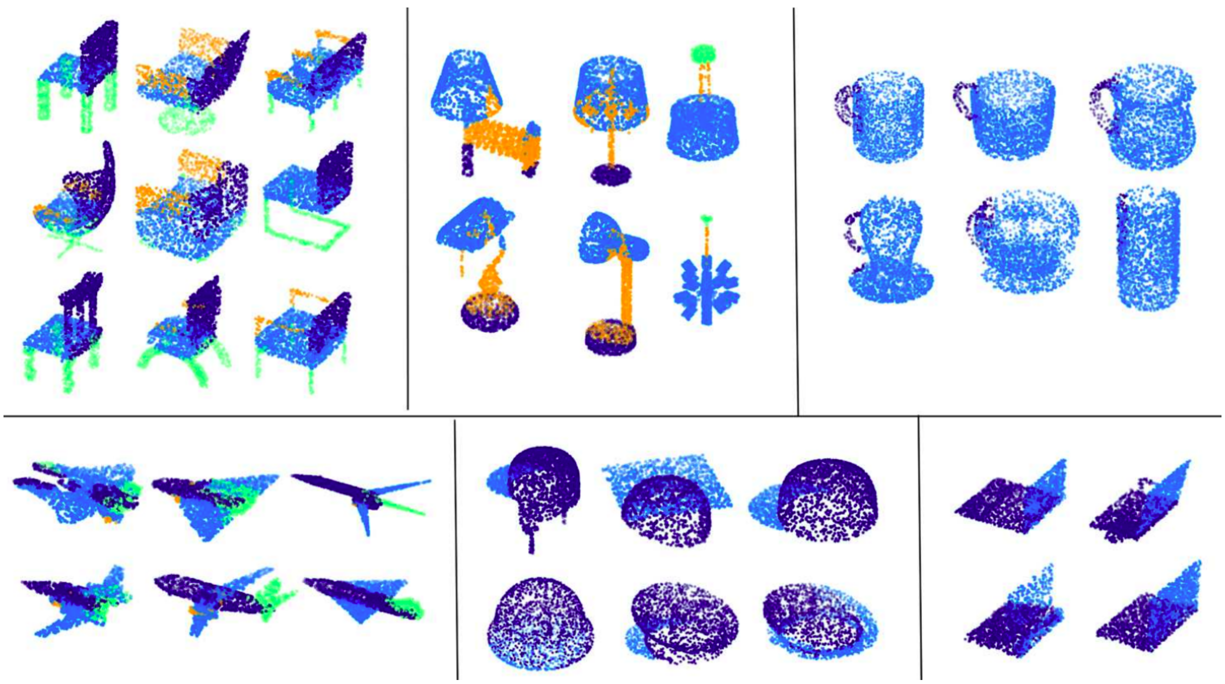


Figure 8: Randomly selected examples of the weakly supervised 3D point cloud segmentation results.

Categories	Ours	[15]	[17]	[35]	[15]
#Samples	10	10	10	10	All
#Parameters	2.6M	3.5M	1.4M	6.9M	3.5M
Airplane	**67.3**	63.3	62.3	65.1	83.4
Bag	**74.4**	64.9	67.4	68.2	78.7
Cap	**86.3**	75.2	80.0	80.7	82.5
Chair	**83.4**	73.8	61.6	66.1	89.6
Lamp	**68.7**	63.8	57.8	60.2	80.8
Laptop	**93.8**	87.3	94.2	93.7	95.3
Mug	**90.9**	80.9	83.1	86.0	93.0
Table	**74.2**	72.2	72.2	72.5	80.6
Mean	**79.8**	72.7	72.3	74.1	85.5

Figure 9: Quantitative result. Comparison with supervised methods on randomly selected small training samples.

Research Project: Topology Constrained Shape Correspondence

Description

Recent studies have developed deep neural networks for robust point descriptor and shape correspondence learning in consideration of local structural information. In this paper, we developed a novel shape correspondence learning network, called TC-NET, which further enhances performance by encouraging the topological consistency between the embedding feature space and the input shape space. Specifically, in this paper, we first calculate the topology-associated edge weights to represent the topological structure of each point. Then, in order to preserve this topological structure in high-dimensional feature space, a structural regularization term is defined to minimize the topology-consistent feature reconstruction loss (Topo-Loss) during the correspondence learning process. Our proposed method achieved state-of-the-art performance on three shape correspondence benchmark datasets.

Method

our proposed TCNET, which is composed of four main components. The first component carries out topological structure mining. In this component, the topological structure is formulated based on the adjacency graph of the input mesh, with the edge weights representing the structural correlation between a node and its neighbors. The second component carries out point feature learning. In this component, a graph based convolutional neural network is adopted to extract high-level point features. The third component carries out topology preserving reconstruction. In this component, the local topological structure in the input shape is transferred into the latent feature space and further used to reconstruct the features of each node. The last component carries out topology-constrained classification. In this component, a topology preserving reconstruction loss function (TopoLoss) is formulated by measuring the similarity between the embedding features of each point and their reconstructions. This regularization term is added to the final loss function to learn more representative point descriptors, which guarantees more robust point correspondence learning.

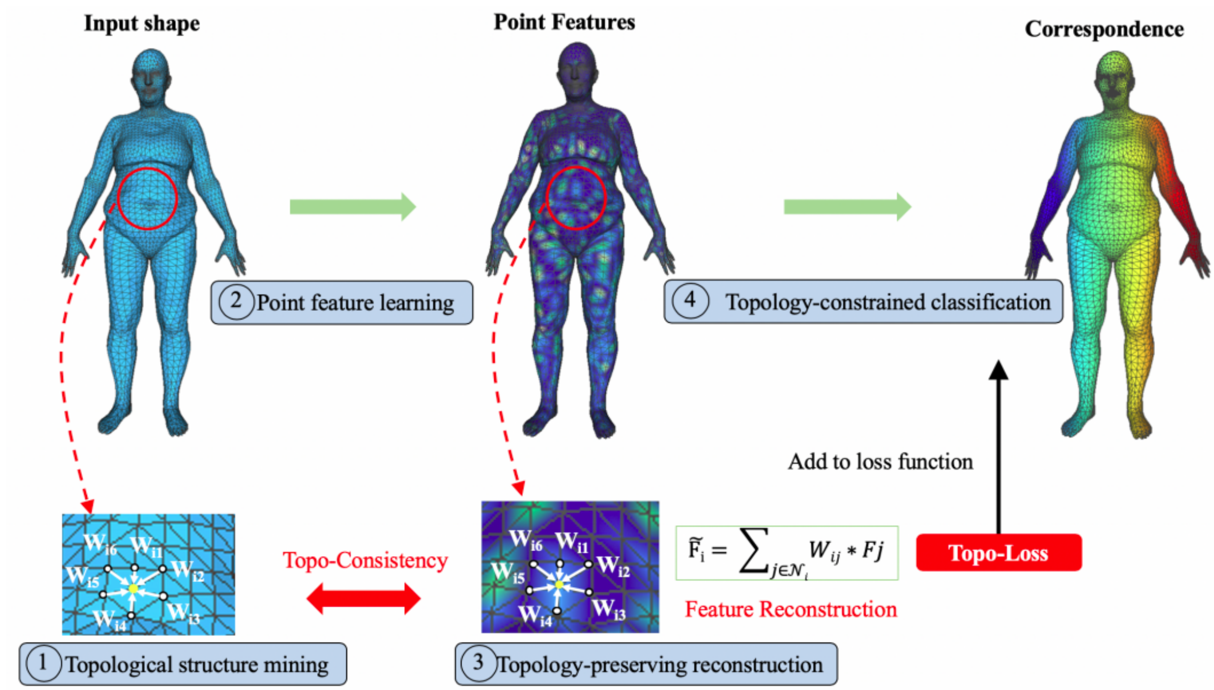


Figure 10: Framework of the topology-constrained shape correspondence learning network. Our method aims to learn topology-preserving point descriptors during the shape correspondence learning process. Given an input shape represented by 3D coordinates and the adjacency graph, our model first extracts the topological structure for each individual point, which is represented by the edge weights connecting the point to its neighbors (step 1). Then, a graph-based convolutional neural network is designed to learn the latent features of each point (step 2). To enforce topological consistency between the input space and the embedding feature space, our model generates reconstructed features for each point by referring to the topological structure in the input space (step 3). We further formulate a Topo-Loss based on the topological similarity between the reconstructed features and the original ones. Finally, the Topo-Loss regularization term is integrated into the vertex classification loss function during shape correspondence learning (step 4). This figure is best viewed in color.

Results

Experiments on FAUST, SCAPE, and TOSCA benchmark datasets for the task of 3D shape correspondence estimation achieve state-of-the-art performance. Experimental results quantitatively and qualitatively demonstrate the effectiveness of the proposed model for shape correspondence learning.

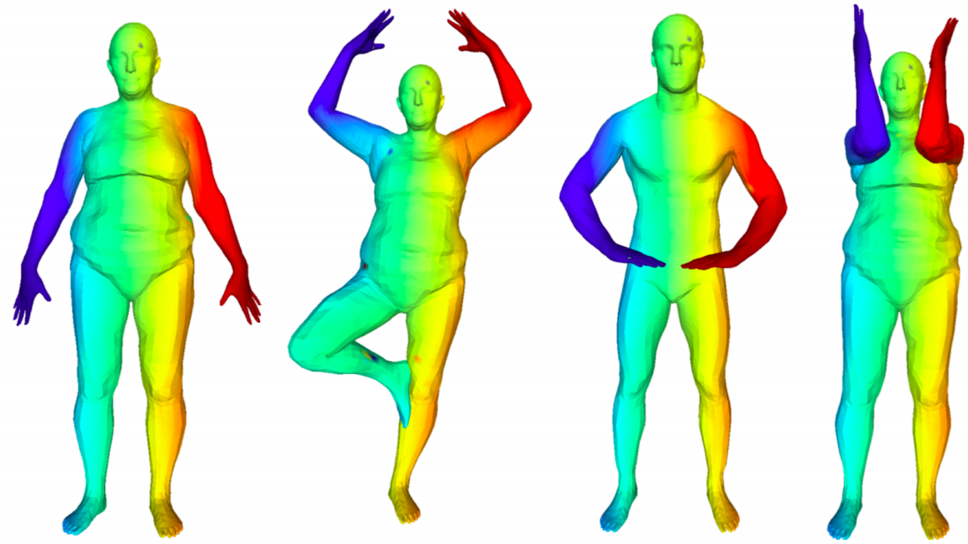


Figure 11: Correspondence results obtained by our model on the FAUST test set. The reference shape is shown on the left, and three deformable shapes with correspondence are shown on the right. Corresponding points are shown in the same color.

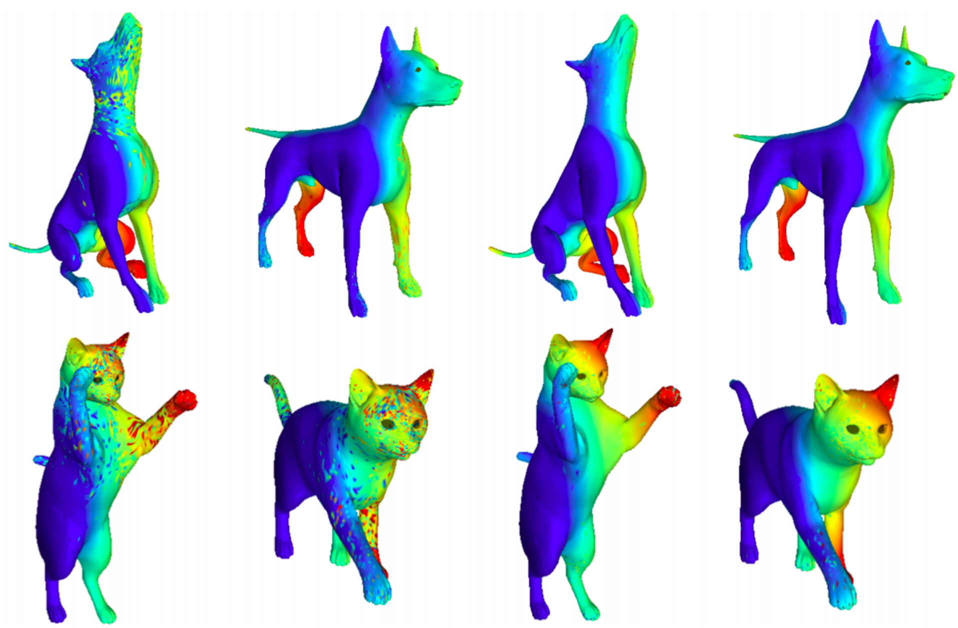


Figure 12: Correspondence results of the non-human shapes from the TOSCA dataset generated by maximum assignment (left columns) and LAP assignment (right columns).

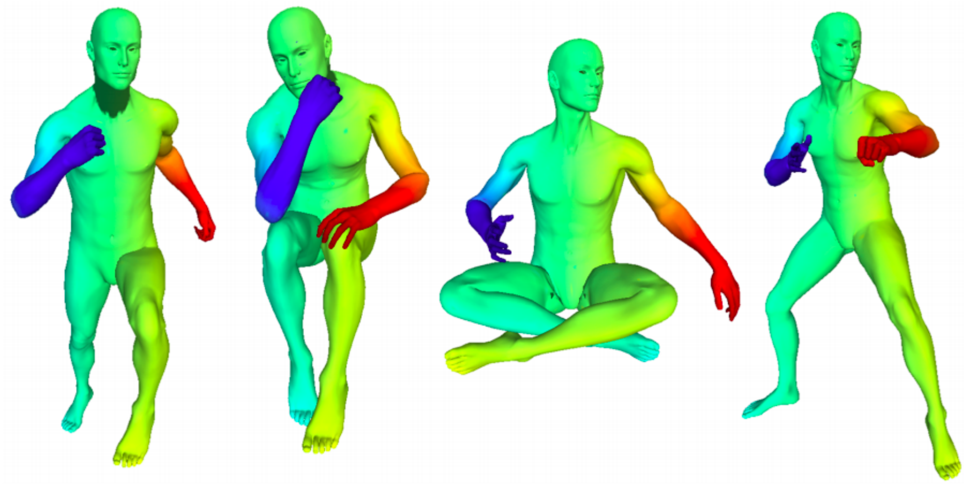


Figure 13: Visualization of correspondence results on the TOSCA human test set. Corresponding points are shown in the same color.