Lingjing Wang

Personal Information

Status: **PH.D. STUDENT**

Program: **Computer Science and Engineering**

School: **Tandon School of Engineering, New York University**

Website: **https://www.math.nyu.edu/~lingwang/**

Period: **From 2014-09 to 2019-05**

Biography

I was a Ph.D. student at New York University and advised by Professor Yi Fang. During my Ph.D. period, I was 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 post-doctoral at NYU MMVC Lab.

Description

We introduce two learning-based registration networks PR- Net and CPD-Net, which can directly predict the desired transformation to align the source and target point sets. In contrast to previous methods, our network jointly learn a registration pattern from a training dataset and is able to instantly predict the desired transformation for an unseen pair from testing dataset without additional iterative optimization process.

Method

We firstly introduce a model-based learning networks for point set registration network PR-Net. We further introduce a model-free learning-based network for deformable point set registration CPD-Net, named coherent point drift networks (CPD-Net).

PR-Net

In PR-Net, we propose a novel technique to learn the global and local shape aware "shape descriptor tensor" directly from the point cloud with irregular and disordered format. A novel shape correlation tensor is proposed to comprehensively evaluate the correlation between two point sets to be registered. We propose a novel statistical alignment loss function that drives our structure to determine the optimal geometric transformation that statistically aligns the source point cloud set and the target one.

CPD-Net

The first component is "Learning Shape Descriptor". In this component, the global shape descriptor is learned with a multilayer perceptron (MLP). The second component is "Coherent PointMorph". In this component, we firstly concatenate the point coordinate of source point, the global shape descriptor of source point set, and global shape descriptor of target point set to form a new

descriptor for each source point. Three successive MLP takes the new descriptor to learn the continuous displacement vector field. The third component is "Point Set Alignment". In this component, a loss function is defined to assess the quality of alignment.
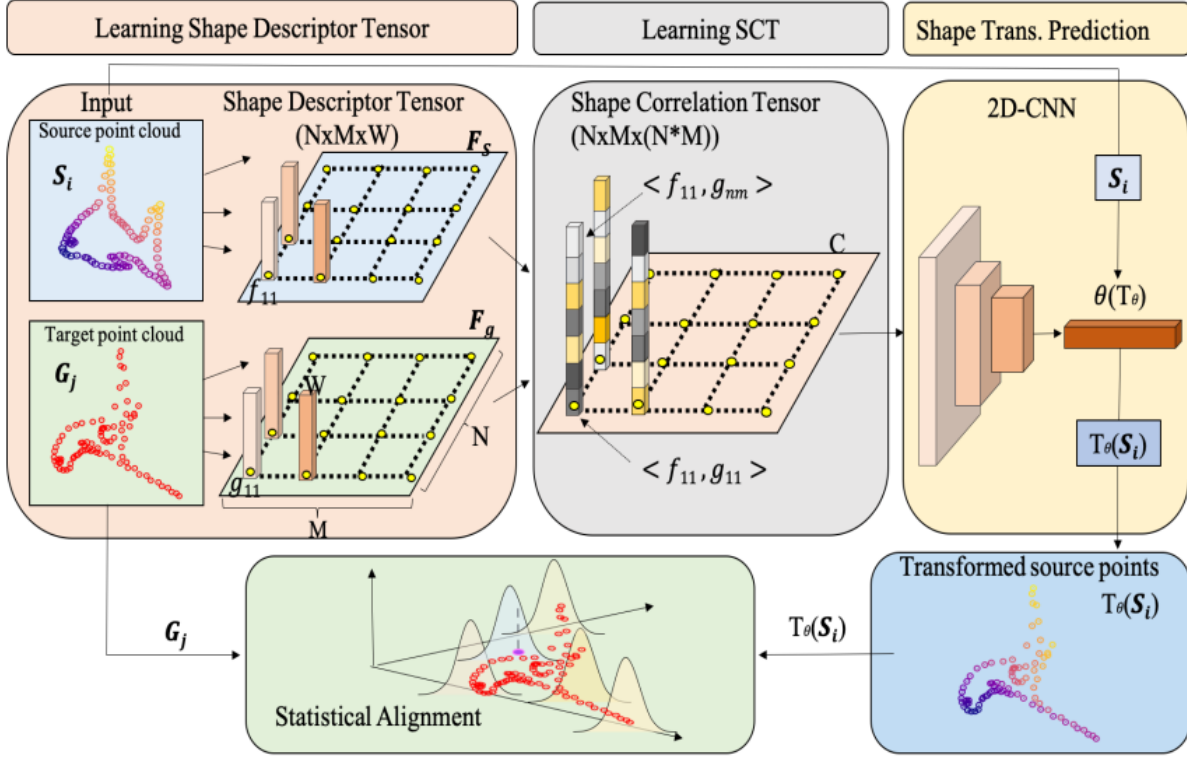

Figure 1: PR-Net pipeline. The proposed PR-Net includes three parts: learning shape descriptor tensor (SCT), learning correlation tensor, and shape transformation prediction.
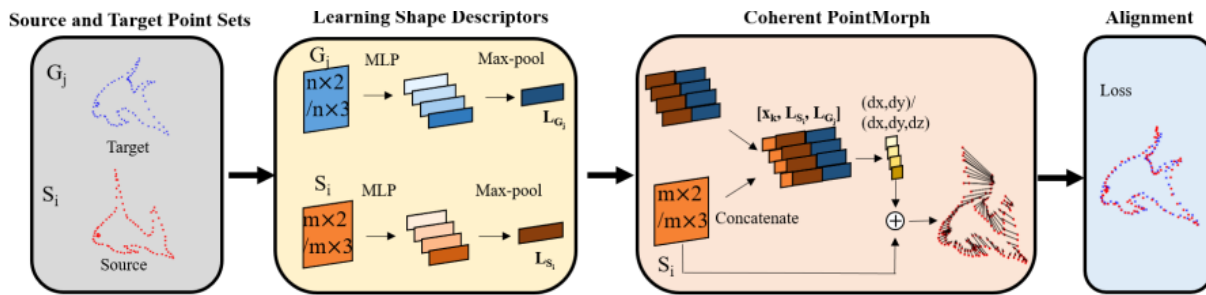

Figure 2: The proposed structure of CPD-Net includes three parts: learning shape descriptor, coherent PointMorph, and the alignment Loss.

Results

PR-Net

We implement a set of experiments to validate the performance of our proposed PR-Net for non-rigid point set registration from different aspects. We compare PR-Net with non-learning based non-rigid point set registration method. We then validate the robustness of PR-Net against the different level of geometric deformation and the robustness of PR-Net against the different types of noise. We further verify that PR-Net can handle registration tasks for various types of dataset.
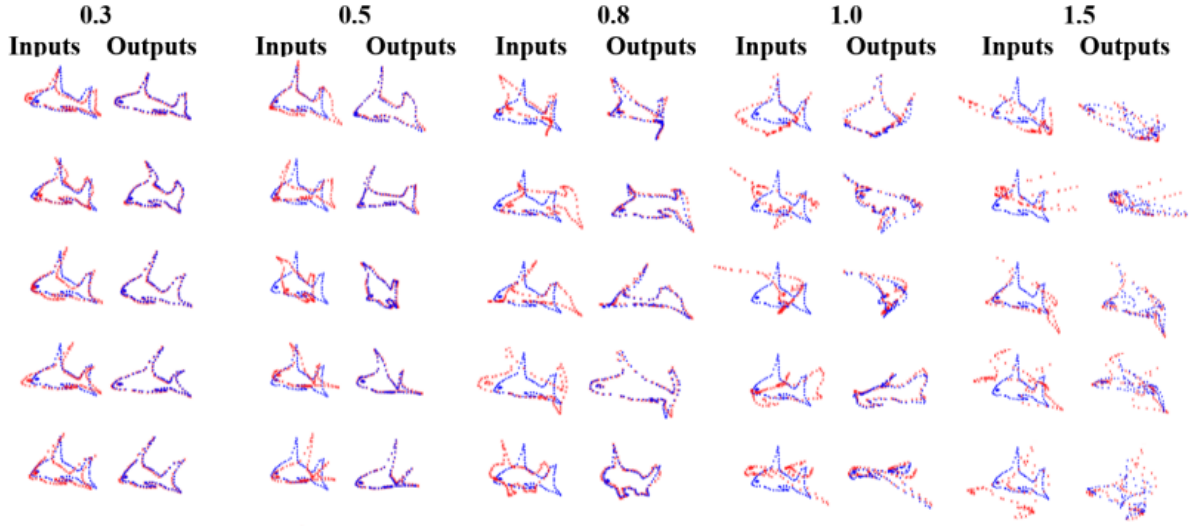

Figure 3: Testing results for 2D fish shape point set registration at different deformation levels. The blue shapes are source point sets and the red shapes are target point sets.
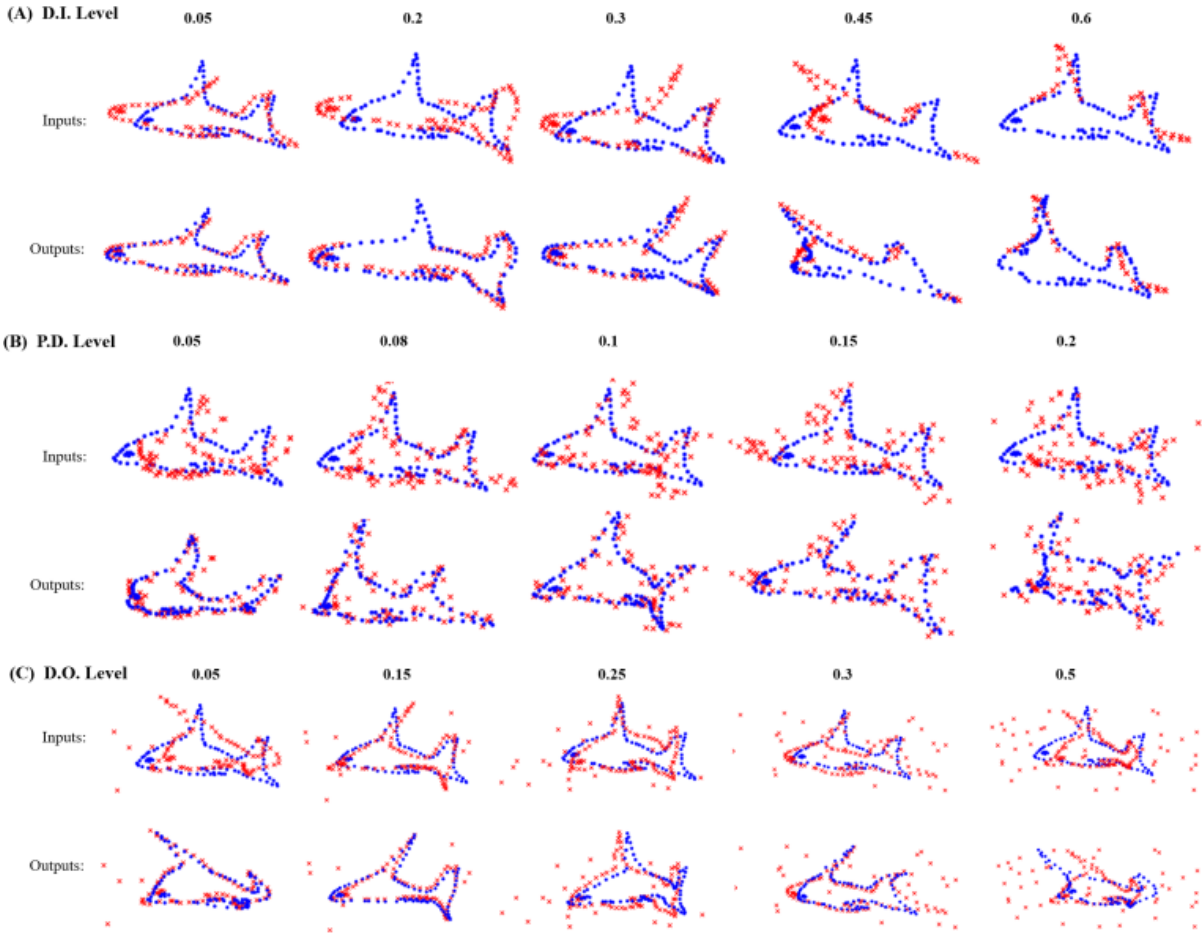
Figure 4: Testing results for 2D fish shape point set registration at deformation level 0.5 in presence of various noise. Blue shapes are source point sets and red ones are target point sets.
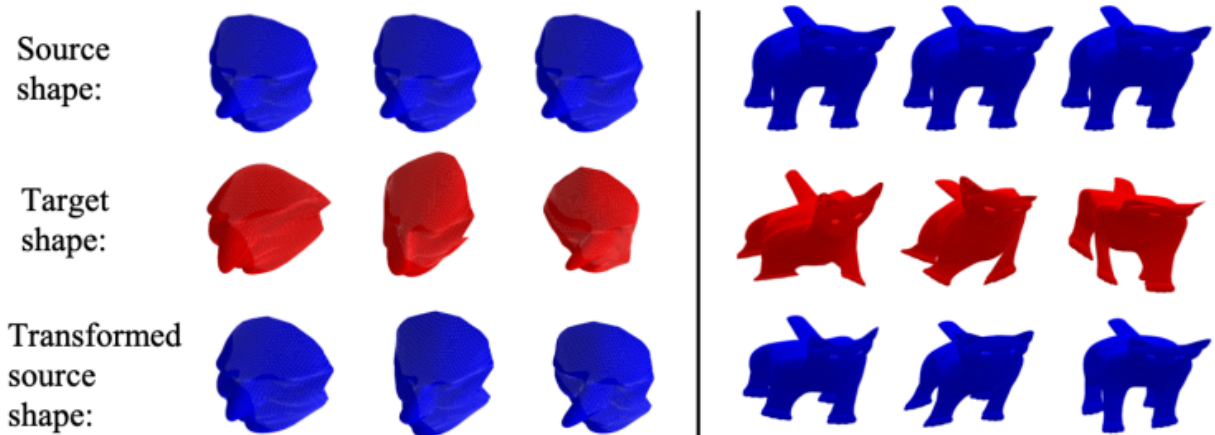

Figure 5: Testing registration performance for 3D face and cat point sets. The blue shapes are source shapes and the red shapes are target ones.

CPD-Net

We report the experimental results to evaluate the performance of our trained CPD-Net on 2D and 3D datasets and shows the resistance of CPD-Net to various types of noise.
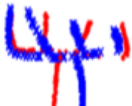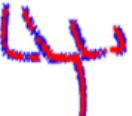

Figure 6: Registration examples for Mushroom, Fork and Face shapes. The blue shape represents target and the red shape represents source point set. The corresponding C.D. score is listed underneath the registered point sets.
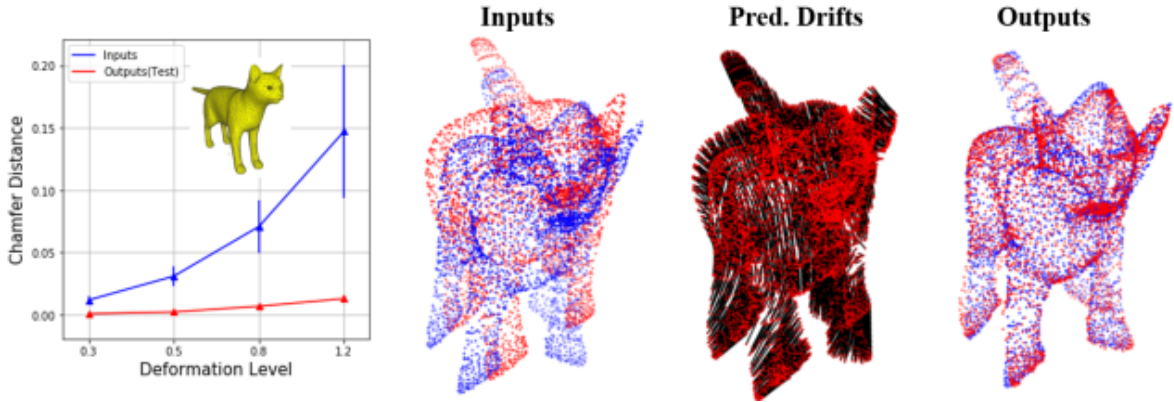
Figure 7: The charts show C.D. between source and target point sets, pre (blue line) and post (red line) registration in left. Selected qualitative registration results are demonstrated in right. The red points represents the source points and the blue ones represent the target points. The black lines represent the predicted drifts for source point set.

Figure 8: The charts of C.D. between transformed source point set and target one in presence of different level of G.D. noise, P.O. noise and D.I.noise are shown in left. The selected qualitative results are demonstrated in right. The red shape represents the source point set and the blue one represents the target point set.

Description

Point sets correspondence concerns with the establishment of point-wise correspondence for a group of 2D or 3D point sets with similar sh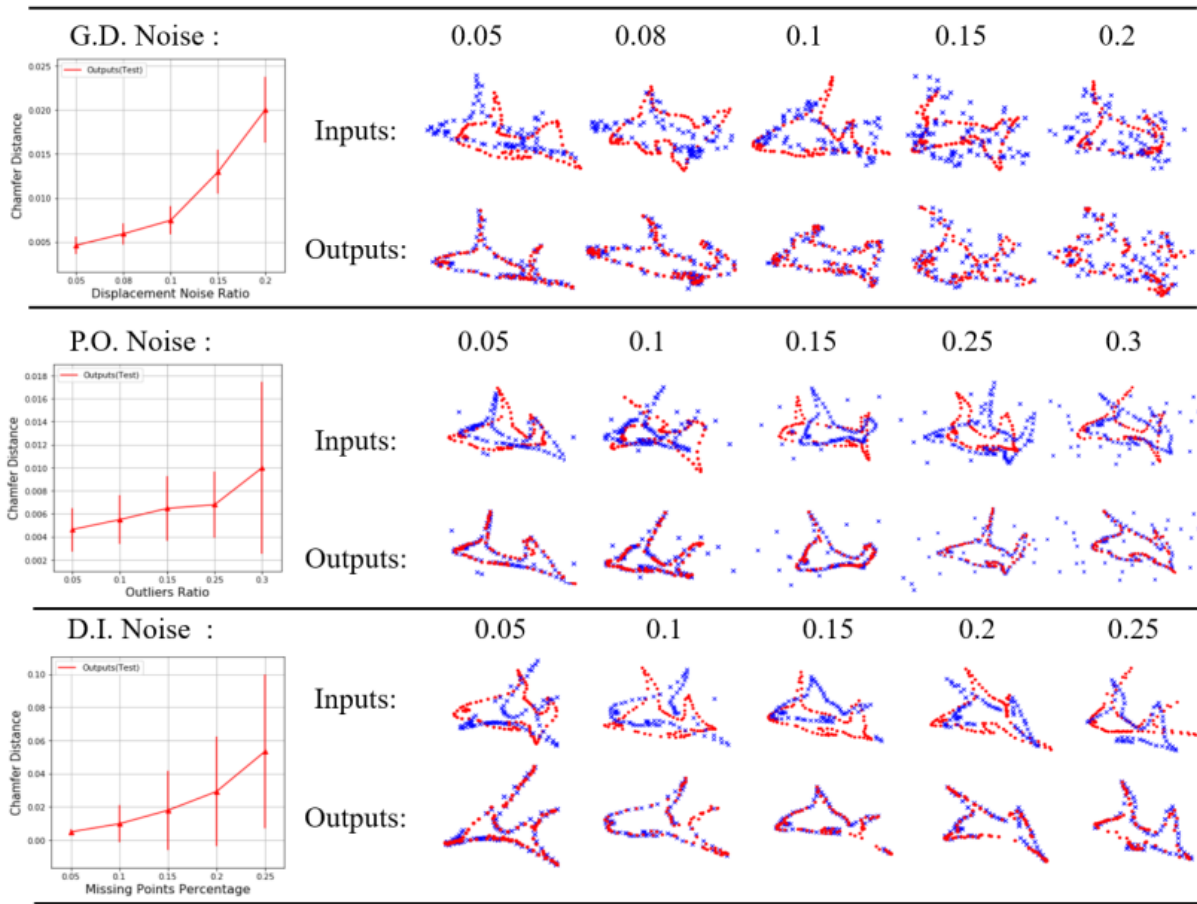ape description. Existing methods often iteratively search for the optimal point-wise correspondence assignment for two sets of points, driven by maximizing the similarity between two sets of explicitly designed point features or by determining the parametric transformation for the best alignment between 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
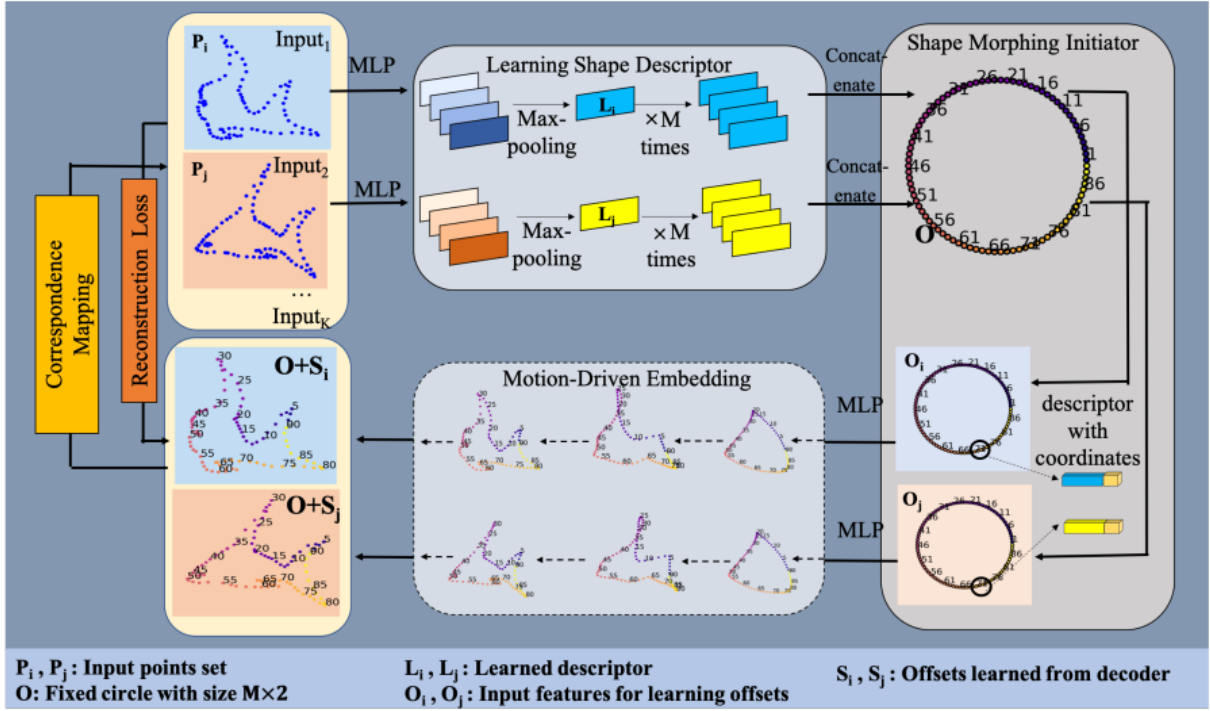
back to the original point sets.

Figure 9: 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.

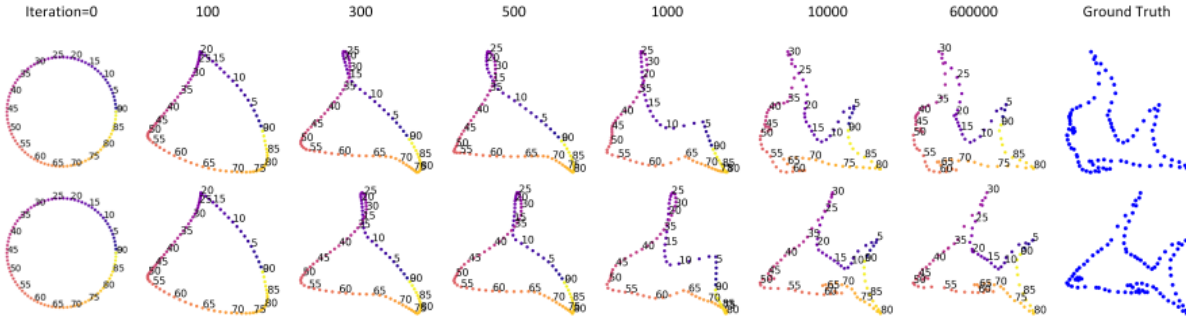Figure 10: Illustration of our Motion-driven embedding process. Landmark points are numbered with color and the blue fishes in last column are input shapes.
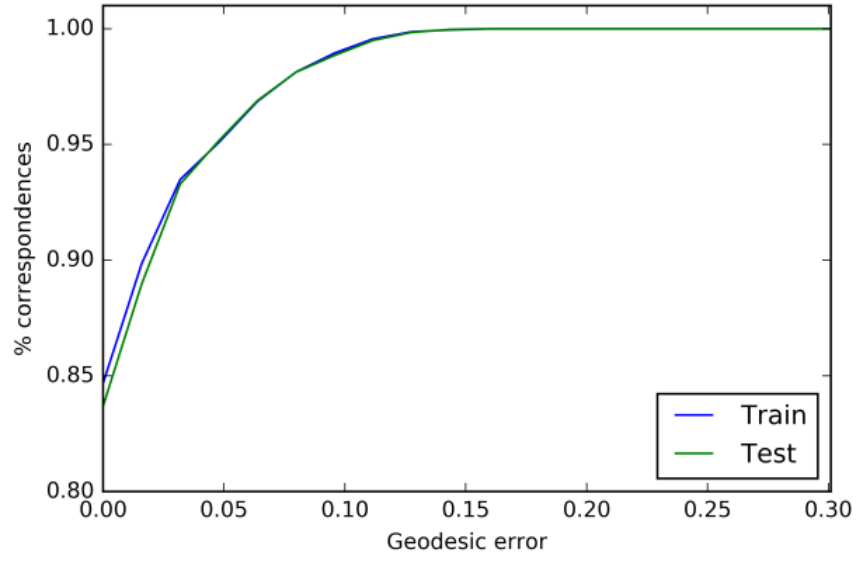
Figure 11: Correspondence performance on test set.
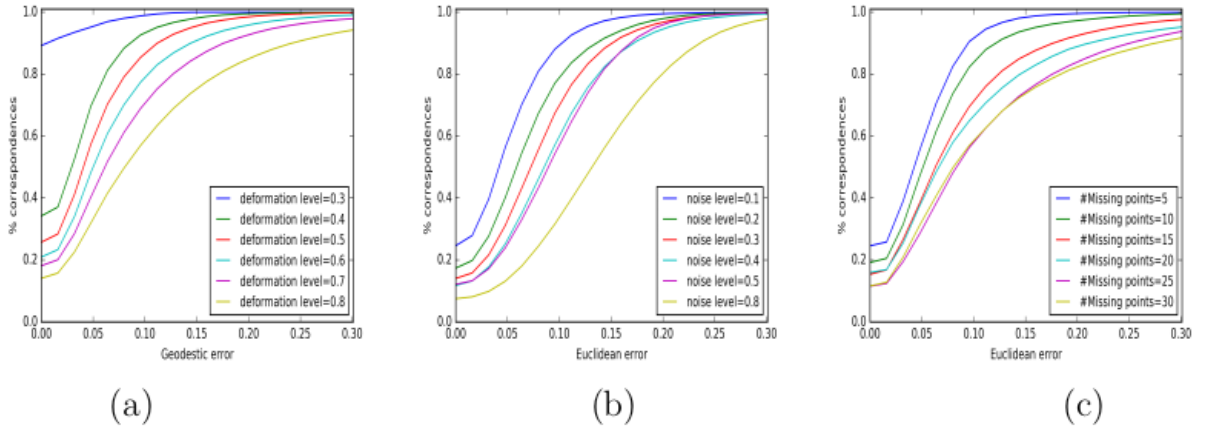
Figure 12: Robustness test. (a) Correspondence quality at different deformation levels. (b) Correspondences quality at different noise level. (c) Correspondence quality with different number of missing points.

Figure 13: 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.

Description

Recent efforts introduce convolutional neural network to learn a geometric model (i.e. an affine or thin- plate spline transformation) for image matching and determin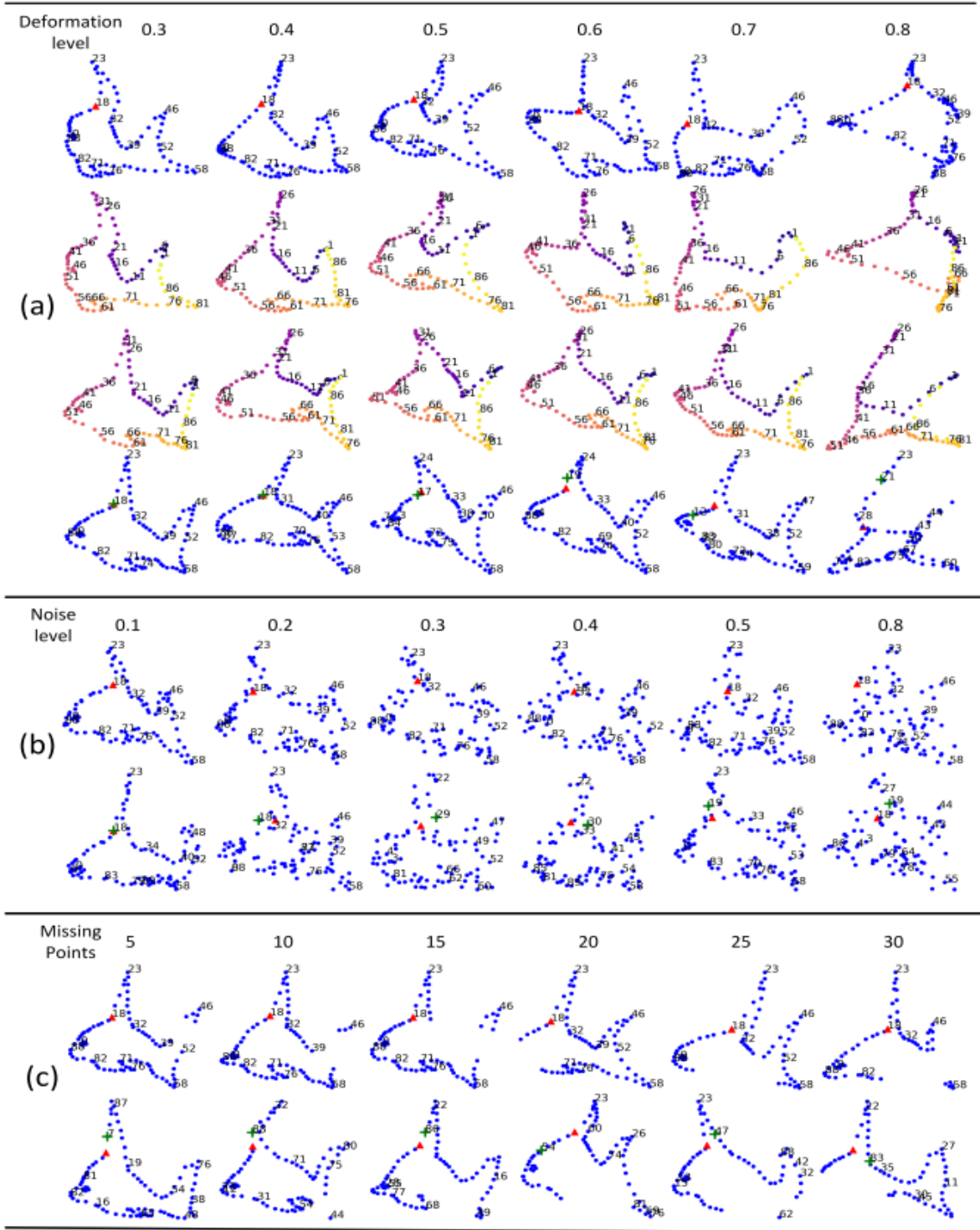e correspondences between two images. The incapability of a geometric model in estimating a high complexity parametric transform limits their use in applications to coarse image alignment/matching. We presents a novel approach to learn a model- free geometric transformation to estimate a continuous smooth displacement field and identify two images with a significant geometric deformation. In contrast to model-based method, our proposed method, named Model-Free Geometric Transformation Networks (MF-GeoNet), can learn displacement vector function to estimate geometric transformation from a training dataset.MF-GeoNet is theoretically proved to learn a continuous displacement vector function to avoid imposing a parametric smoothness constraint by regularizing the displacement field. We conducted experiments over both synthetic and real image dataset. The results demonstrate the superior performance of MF-GeoNet over other state of-the-art methods in identifying the correspondences, even when images are in the presence of significant geometric deformation.

Method

We propose to develop a novel geometric transformation network, named arbitrary continuous geometric transformation networks (Arbicon-Net), to directly predict the dense displacement field that is not formulated by pre-defined hand-crafted geometric models. We design an Arbicon-Net to simultaneously train three major modules, namely front-end geometric feature extractor module, transformation descriptor encoder module and displacement field predictor module, in an end-to-end fashion. The Arbicon-Net firstly extracts dense feature maps from input image pairs and encodes the discriminative local feature correlation into a transformation descriptor. The following predictor module uses the transformation descriptor to decode displacement field for image registration.
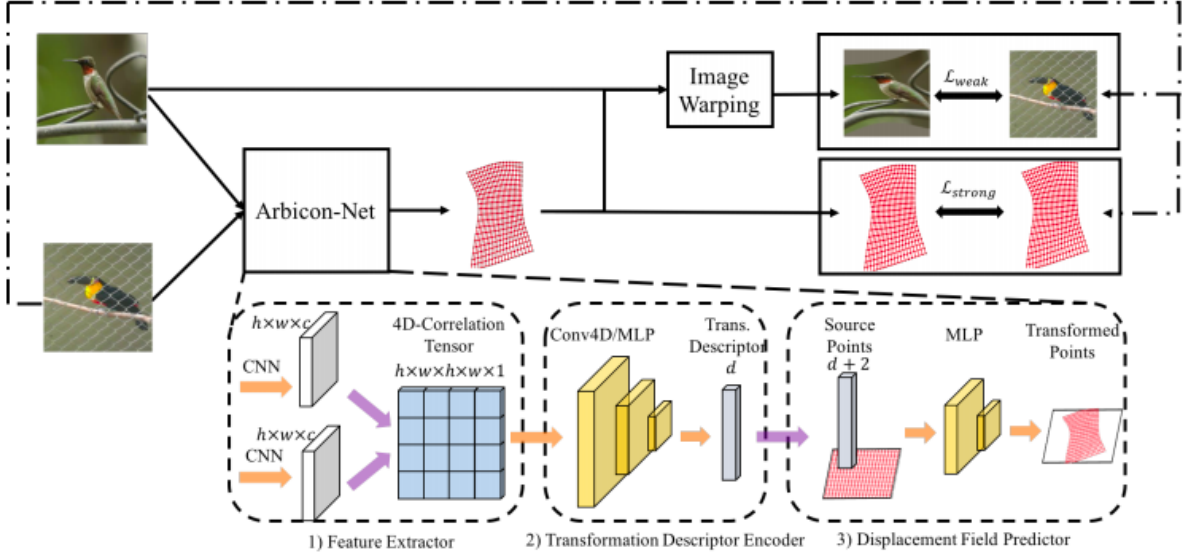
Figure 14: Main Pipeline. Our proposed end-to-end trainable Arbicon-Net has three main components. 1) Geometric Feature Extractor Module; 2) Transformation Descriptor Encoder Module; 3) Displacement Field Predictor Module.

Results

We validate the performance of Arbicon-Net for the estimation of the geometric transformation for real image pairs with weakly supervised training. Experimental results show that our proposed Arbicon-Net achieved superior performance against hand-crafted geometric transformation models with both strong and weak supervision.

Methods	PCK(%)
HOG+PF-LOM [36]	62.5
SCNet-AG+ [28]	72.2
CNNGeo [11]	71.9
A2Net [13]	70.9
WeakAlign [10]	75.8
WeakAlign-4D	76.5
Arbicon-Net	**77.3**

Figure 15: Quantitative results on PFPascal dataset with weakly supervise training.
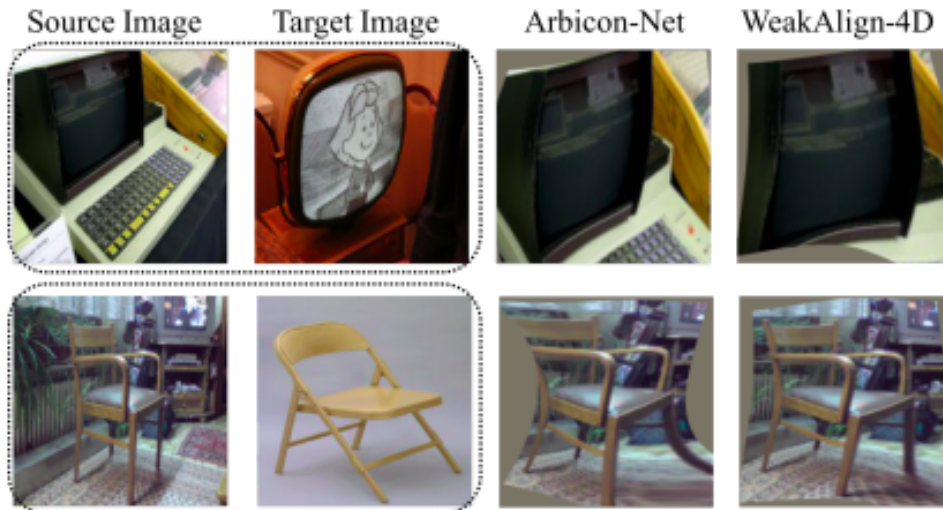
Figure 16: Qualitative results on PF-Pascal dataset with weakly supervise training.

Methods	PCK(%)	E_{smooth}
CNNGeo (aff.)	73.5 $\pm$0.9	0.000
CNNGeo (TPS)	78.2 $\pm$0.7	0.013
CNNGeo (aff.+TPS)	78.5 $\pm$0.9	0.016
Arbicon-Net	**84.3 $\pm$0.6**	0.008

Figure 17: Quantitative comparisons on Proposal Flow dataset (Non-parametric transformation) with strong supervision.

Figure 18: Qualitative comparisons on Proposal Flow [37] dataset with strong supervision.

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.
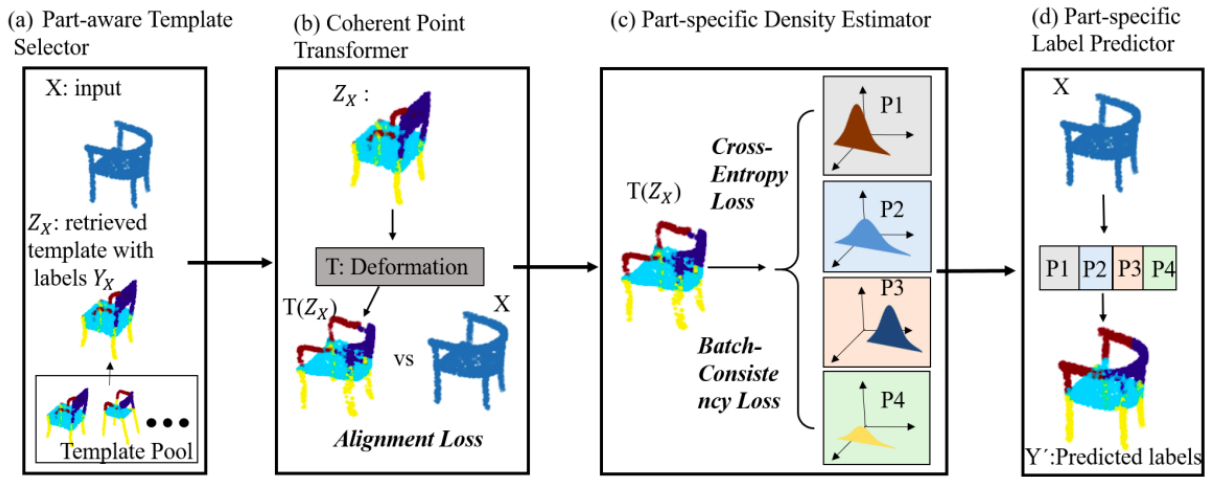

Figure 19: For a given input 3D point cloud, a template shape firstly is retrieved from a template pool by template selector (a). Coherent point transformer (b) morphs the retrieved template towards input shape. In (c), the Part-specific Density Estimator takes the points of deformed templates as input 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.
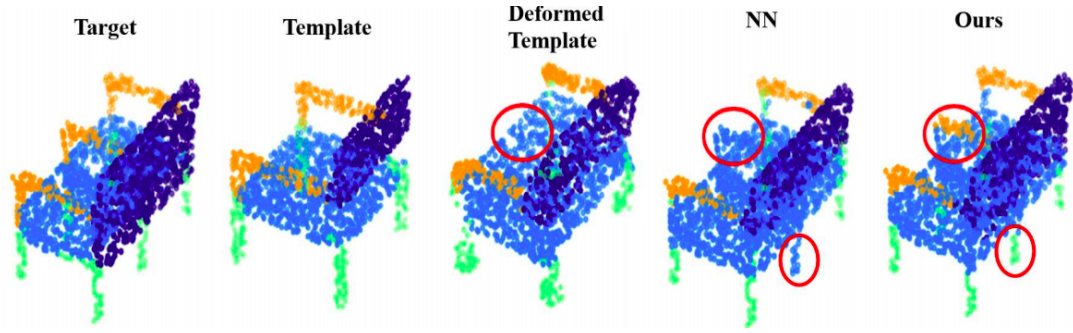
Figure 20: Comparison of the segmentation results based on the deformed template between using Nearest Neighbors method and our part-specific density estimator.
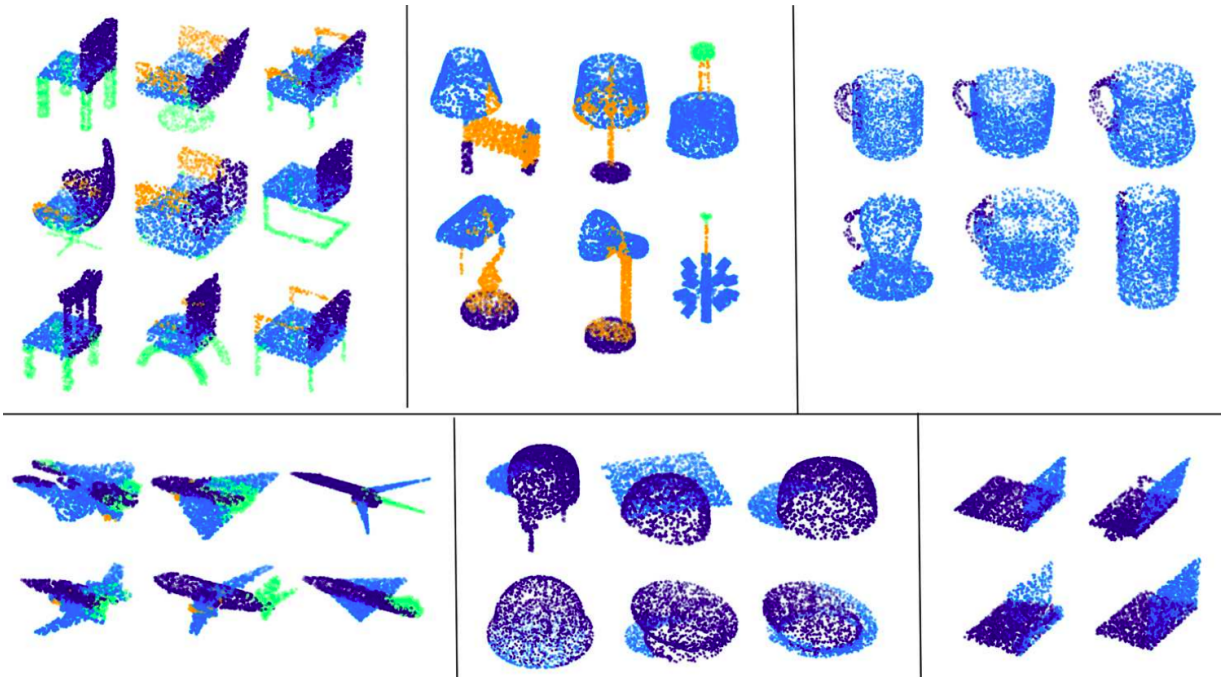

Figure 21: Randomly selected examples of the weakly supervised 3D point cloud segmentation results.

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