Shuaihang Yuan

Personal Information

Status:	PH.D. STUDENT
Program:	Computer Science and Engineering
School:	Tandon School of Engineering, New York University
Website:	https://shalexyuan.github.io/
Period:	From 2020-01 to Present

Biography

I am a Ph.D. student at New York University and advised by Professor Yi Fang. During my Ph.D. 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.

Research Project: 3DMotion-Net: Learning Continuous Flow Function for 3D Motion Prediction

Description

This paper deals with predicting future 3D motions of 3D object scans from the previous two consecutive frames. Previous methods mostly focus on sparse motion prediction in the form of skeletons. While in this paper, we focus on predicting dense 3D motions in the form of 3D point clouds. To approach this problem, we propose a self-supervised approach that leverages the power of the deep neural network to learn a continuous flow function of 3D point clouds that can predict temporally consistent future motions and naturally bring out the correspondences among consecutive point clouds at the same time.

Method

In this paper, to avoid the challenging issue of spatiotemporal feature extraction using an encoder network, we introduce a learnable latent code to represent the temporal-aware shape features which can be optimized during model training. This temporal-aware shape representation is in nature captures both spatial geometry characteristics of 3D shapes and their correlation among consecutive frames. Our method is self-supervised, assuaging the supervision of ground-truth correspondences and predicting future motion and bringing out correspondences of future motion naturally. The pipeline of our proposed method is presented in Fig.1. Three major models are enclosed to learn the desired 3d motion flow function. We first propose a temporal-aware shape descriptor optimizer that can learn the desired spatiotemporal representations implicitly. A temporally consistent motion, Morpher, which takes as input a sequence of two 3D source shapes, is then used to deform the source shapes to the corresponding target shapes based on temporal-aware shape descriptor. This morphing module is built upon a deep neural network. Both temporal-aware shape descriptors and the morphing decoder network are optimized during the network training process using an unsupervised loss function. While in the test stage, the decoder network parameters are fixed, which can be regarded as learned prior information from the training data set, and we only optimize the temporal-aware shape descriptor towards the optimal spatiotemporal representation. The future motion with dense correspondence can be generated using the optimal spatiotemporal representation.

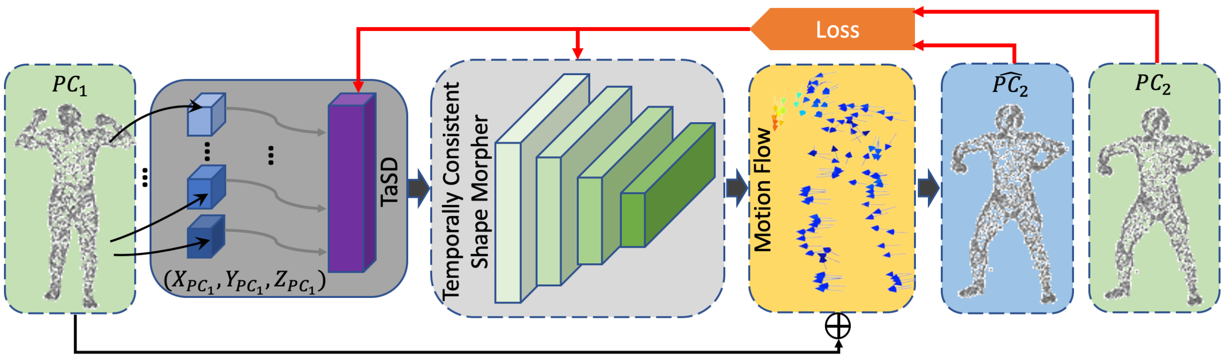

Figure 1: This is the pipeline of the proposed method. We proposed two models to learn a continuous flow function that can predict future motion.

Results

We conduct experiments on several public data sets to show the abilities of our model in predicting 3D motions. Experimental data sets include Dynamic FAUST (D-FAUST), a dynamic human motion data set, which contains 40,000 meshes, a human pose data set SCAPE which consists of a sequence of real human scans. The TOSCA data set contains synthetic nonrigid models of 80 objects with various object category, such as cat, dog, wolf, etc. Note that although the ground truth corresponding points are provided for all these three data sets, our proposed method uses no ground truth information neither in the training nor optimization.

Table 1: Quantitative results on the inconsistent data. Chamfer distance and correspondence ℓ_2-distance are reported. (Lower is better)

3DMotion-Net		
Data Set	Chamfer Distance	Correspond.
SCAPE w/ Nosie Level 0.2	0.029	0.091
SCAPE w/ Nosie Level 0.4	0.037	0.098
SCAPE w/ Holes	0.031	0.092
SCAPE w/ Partial	0.034	0.094

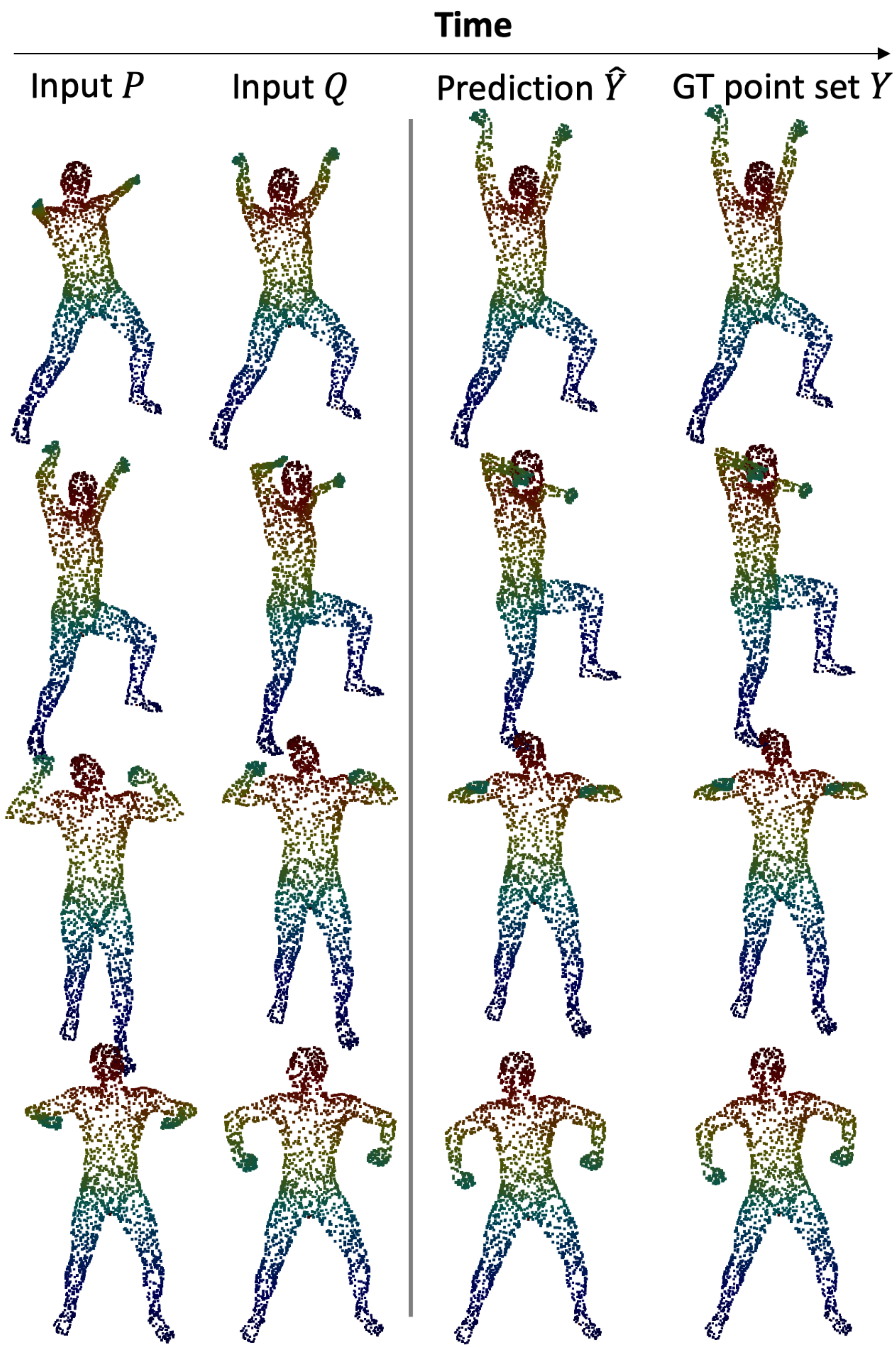

Figure 2: Qualitative results on the temporal consistent SCAPE data set. Corresponding points are painted with same color.

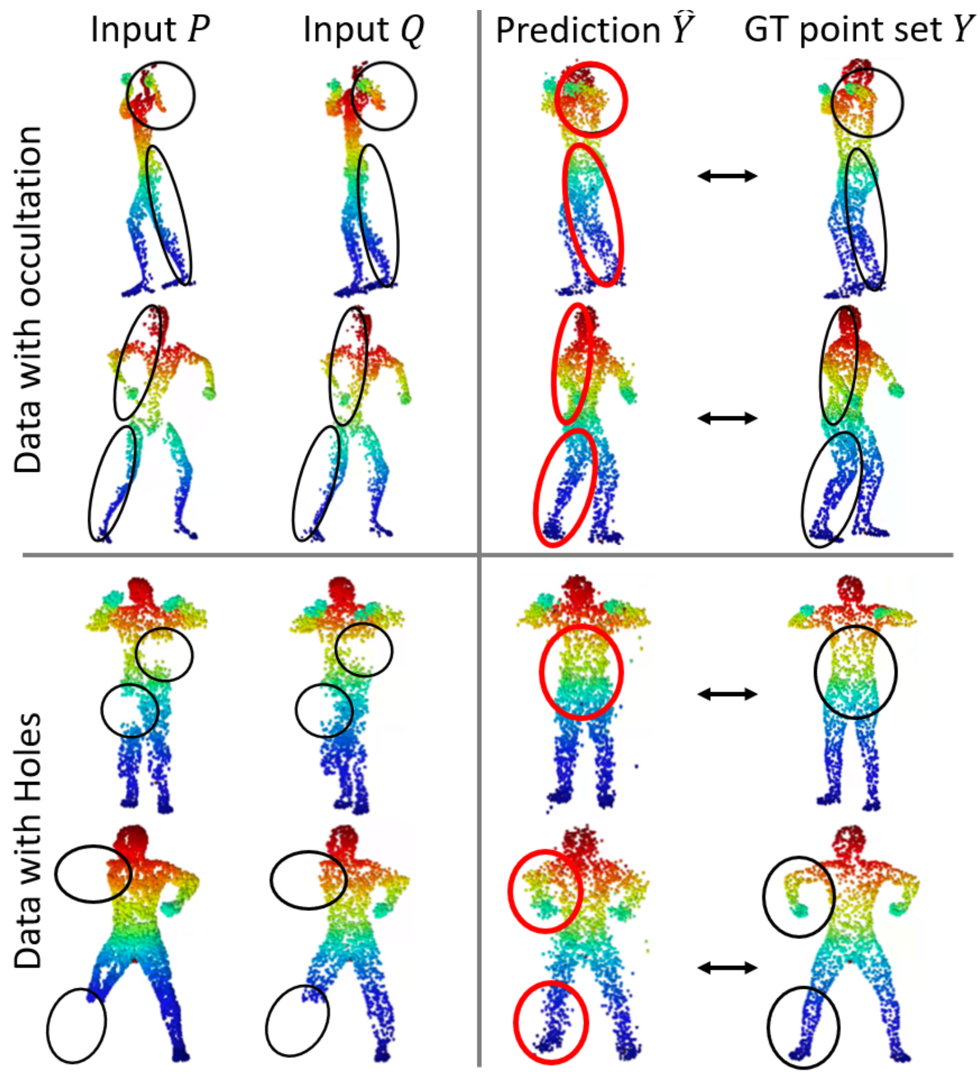

Figure 3: Qualitative results on the inconsistent data with noises and occlusions. The high light part in this figure clearly shows that our proposed method can produce consistent prediction even when the inconsistent inputs are given.

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

Description

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 **R**obust Learning of **O**ne-Shot 3D **S**hape **S**egmentation (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.

Method

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 4, 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.

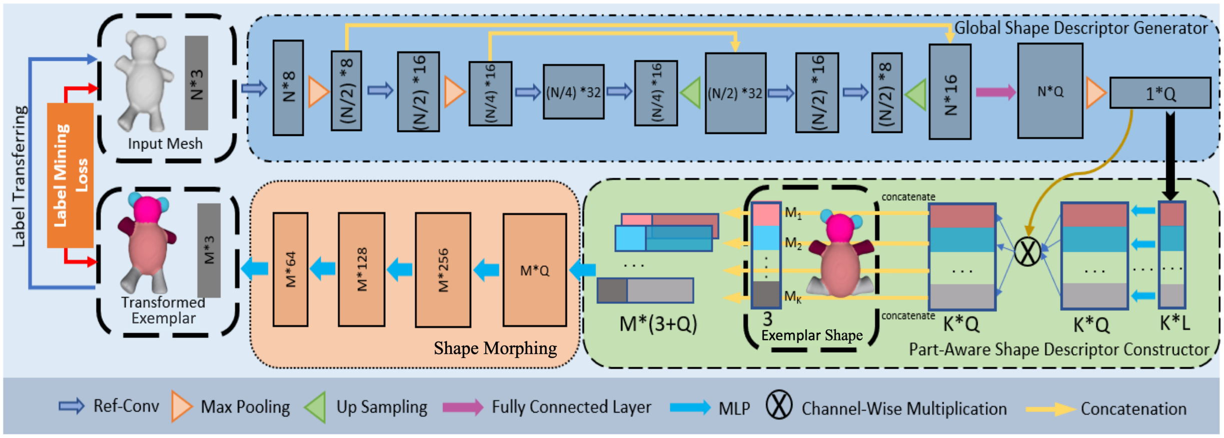

Figure 4: Our one-shot paradigm for 3D mesh object part segmentation task. As the figure shows, the blue box on the top of the figure represents the process of pose feature learning. we adopt several referenced weighted convolution layers, max-pooling layers, and upsampling layers. A global max-pooling 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.

Results

We adopt two dataset to evaluate our proposed method including PSB and SCAPE. 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.

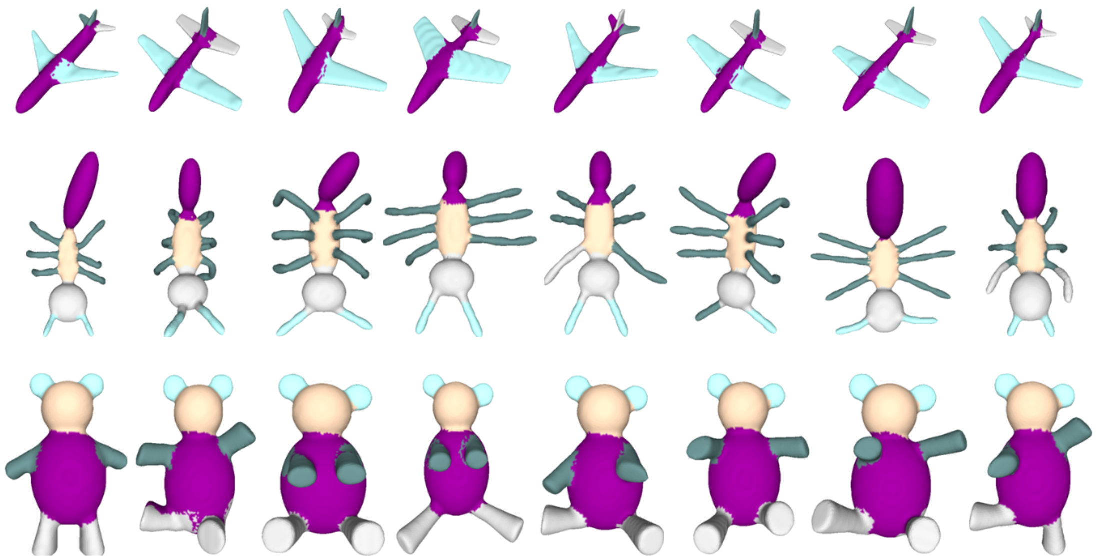

Figure 5: Visualization of part segmentation task results on PSB [**?**] benchmark dataset for airplane, ant and teddy bear.

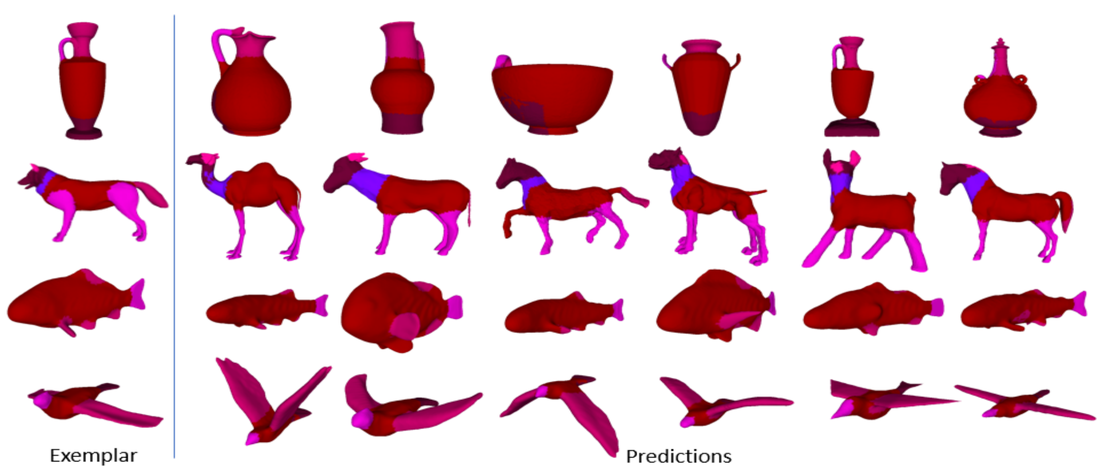

Figure 6: The visualization result of our proposed method. The shape configuration between the selected exemplar and the testing set can be significantly different.