Coherent Point Drift Networks: Unsupervised Learning of Non-Rigid Point Set Registration

Lingjing Wang
lingjing.wang@courant.nyu.edu

Yi Fang
yfang@nyu.edu

Abstract

Given new pairs of source and target point sets, standard point set registration methods often repeatedly conduct the independent iterative search of desired geometric transformation to align the source point set with the target one. This limits their use in applications to handle the real-time point set registration with large volume dataset. This paper presents a novel method, named coherent point drift networks (CPD-Net), for unsupervised learning of geometric transformation towards real-time non-rigid point set registration. In contrast to previous efforts (e.g. coherent point drift), CPD-Net can learn displacement field function to estimate geometric transformation from a training dataset, consequently, to predict the desired geometric transformation for the alignment of previously unseen pairs without any additional iterative optimization process. Furthermore, CPD-Net leverages the power of deep neural network to fit an arbitrary function, that adaptively accommodates different levels of complexity of the desired geometric transformation. Particularly, CPD-Net is proved with a theoretical guarantee to learn a continuous displacement vector function that could further avoid imposing additional parametric smoothness constraint as in previous works. Our experiments verify CPD-Net’s impressive performance for non-rigid point set registration on various 2D/3D datasets, even in presence of significant displacement noise, outliers, and missing points. Our code is available at https://github.com/nyummvc/CPD-Net.

1. Introduction

Point set registration is a fundamental computer vision task, which has wide applications in many fields such as image registration, object correspondence, large-scale 3D reconstruction and so on [2, 3, 5, 11, 15, 16, 19, 23, 26, 30, 31]. Existing approaches often solve the registration problem through an iterative optimization process to search the optimal geometric transformation to minimize a pre-defined alignment loss between transformed source point set and target point set [10, 12–14, 20]. The geometric transformation can be modelled by a specific type of transformation (e.g. thin plate spline) [5], or described as a model-free deformation field [20] where a displacement vector function is defined. The rotation, translation and scaling are often used to model rigid geometric transformation, while the affine and thin plate spline are widely used to model non-rigid geometric transformation. For model-based method, the registration task turns into searching for parameters of the model. For model-free method, the registration task turns into to process to determine the displacement field that moves the source point set towards the target one. Regardless the model-based or model-free point set registration approaches, most existing efforts have to start over a new iterative optimization process to determine the geometric transformation. Consequently, the intensive computation in the iterative routine pose great challenges for existing approaches to handle a large scale dataset in real-time application. To address this challenge, we are motivated to develop a learning based approach to generalize its ability from training process to predict the desired geometric transformation to register previously unseen pair of point sets.

In this paper, we proposed a novel approach, named coherent point drift networks (CPD-Net), for unsupervised learning of displacement vector function to estimate a model-free geometric transformation for point set registration. Figure 1 illustrates the pipeline of the proposed CPD-Net which consists of three major components. 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. The main contributions of our proposed method as follows:

- This paper introduces a novel Coherent Point Drift Networks (CPD-Net) that can be trained in unsupervised manner, consequently it can be generalized to predict geometric transformation for non-rigid point set registration.

- CPD-Net leverages the power of deep neural network to fit an arbitrary function, that is able to accommodate different levels of complexity of the target geometric transformation for best aligning the pair of point sets.

- The CPD-Net theoretically guarantees the continuity of predicted displacement field as geometric transformation, which naturally eliminate the necessity to impose a parametric hand-crafted smoothness constraint.

- The CPD-Net is free of specific geometric model selection for modeling the desired transformation, which avoids the potential mismatch between the transformation described by specific adopted models and the actual transformation required for point set registration.

2. Related Works

2.1. Iterative registration methods.

The development of optimization algorithms to estimate the rigid or non-rigid geometric transformations in an iterative routine attracts attention of numerous researchers in decades. The Iterative Closest Point (ICP) algorithm [5] is one successful solution for rigid registration. It initializes an estimation of a rigid function and then iteratively chooses corresponding points to refine the transformation. However, the ICP algorithm is reported to be vulnerable to the selection of corresponding points for initial transformation estimation, and also incapable of dealing with non-rigid transformation. To accommodate the deformation (e.g. morphing, articulation) between a pair of point sets, many efforts were spent in the development of algorithms for a non-rigid transformation. For non-rigid registration, classical previous methods can usually be divided into parametric and non-parametric by the target transformation. As a robust parametric method, TPS-RSM algorithm was proposed by Chui and Rangarajan [7] to estimate parameters of non-rigid transformation with a penalization on second order derivatives. As a classical non-parametric method, coherence point drift (CPD) was proposed by Myronenko et al. [20], which successfully introduced a process of fitting Gaussian mixture likelihood to align the source point set with the target one. With the penalization term on the velocity field, the algorithm enforces the motion of source point set to be coherent during the registration process. Another non-parametric vector field consensus algorithm was proposed by Ma et al. [14]. This algorithm is emphasized to be robust to outliers. Ma et al. proposed the importance to leverage both local and global structures during the non-rigid point set registration in [15]. The existing algorithms achieved great success for the registration task. Even though all these methods state the registration task as an independent optimization process for each single pair of source and target point sets, they greatly inspire us for designing our learning-based system.

2.2. Learning-based registration methods.

In recent years, learning-based methods achieved great success in many fields of computer vision [1, 17, 18, 21, 25, 29, 32, 33]. Especially recent works started a trend of directly learning geometric features from cloud points (especially 3D points), which motivates us to approach the point set registration problem using deep neural networks by leveraging the possibility of direct learning features from points [4, 17, 21, 24, 29, 32]. The closest research to this work is the unsupervised non-parametric non-rigid registration of 3D volumetric medical image proposed by Balakr-
ishnan et al. [4]. In this research, authors use a voxelMorph CNN architecture to learn the registration field to align two volumetric medical images. Even though, both of our researches engage in learning a non-parametric transformation. In contrast, we discuss a broader task for 2D/3D point set registrations in this paper. In comparison to volumetric data, point clouds have very different characteristics and we target to predict the coherent drifts. Moreover, Our proposed model dose not require additional penalization term on smoothness of predicted transformation. For registration of 2D images, an outstanding registration model was proposed by Rocco et al. [24]. Wang et al. [28] proposed PR-Net for non-rigid point set registration. This works mainly focus on parametric approach for 2D image registration. The parameters of both rigid and non-rigid function (TPS) can be predicted by a CNN-based structure from learning the correlation relationship between a pair of source and target 2D images. Even though these learning-based methods are not exactly designed for solving the point set registration problem, they are closely related to our paper and greatly encourage us to investigate the efficient learning-based structure for model-free point set registration.

3. Approach

We introduce our approach in the following sections. Firstly we define the learning-based registration problem in section 3.1. In section 3.2, our first module is introduced for learning shape descriptor from a 2D/3D source/target point sets. Section 3.3 illustrates coherent PointMorph module for learning the smooth drifts to align the source point set with the target one. In section 3.4, The definition of the loss function is provided. The model configurations and the settings for training are described in section 3.5.

3.1. Problem statement

For iterative methods, the optimization task is defined on a given pair of source and target point sets \( (S_i, G_j) \). Here, unlike interactive methods, for a given training dataset \( D = \{(S_i, G_j) | S_i, G_j \subset \mathbb{R}^N (N = 2 \text{ or } N = 3) \} \), we need to redefine the new optimization task. We assume the existence of a function \( g_\theta(S_i, G_j) = \phi \) using a neural network structure, where \( \phi \) is the coherent drifts for each (sampled) point in source point set. As shown in Figure 1 we demonstrate the overview of our proposed structure. Taking a pair of source and target point sets \( (S_i, G_j) \) as input, our model compute the coherent point drifts \( \phi \) based on a set of parameters \( \theta \), which are weights in the neural network structure. We further transform the input source points by simply moving the drifts and then, measure the similarity between the transformed source and target point sets as the alignment loss to update \( \theta \).

For a given dataset, we use stochastic gradient descent based algorithm to optimize parameters set \( \theta \) for minimizing the expected loss function:

\[
\theta_{\text{optimal}} = \arg\min_{\theta} \mathbb{E}_{(S_i, G_j) \sim D} [\mathcal{L}(S_i, G_j, g_\theta(S_i, G_j))],
\]

where \( \mathcal{L} \) represents a similarity measure. CPD-Net dose not require any ground truth drifts for each point for supervision. For a unseen pair from testing dataset, our model directly predict the desired coherent drifts without an additional iterative process.

3.2. Learning shape descriptor

For a given input point set, we firstly learn a shape descriptor that captures representative and deformation-insensitive geometric features. Let \( (S_i, G_j) \) denotes the input source and target point sets and \( (L_{S_i}, L_{G_j}) \) denotes their shape descriptor, where \( L_{S_i}, L_{G_j} \subseteq \mathbb{R}^{m} \) as shown in Figure 2. To address the problem of irregular format of point set, we introduce the following encoding network, which includes \( t \) successive multi-layer perceptrons (MLP) with ReLu activation function \( \{f_i\}_{i=1,2,...,t} \), such that: \( f_i : \mathbb{R}^{m_i} \rightarrow \mathbb{R}^{m_{i+1}} \), where \( m_i \) is the input layer’s dimension and \( m_{i+1} \) is the output layer’s dimension. The encoder network is defined as: \( \forall (S_i, G_j) \),

\[
L_{S_i} = \text{Maxpool}\{f_t f_{t-1} \ldots f_1(x_i)\}_{x_i \in S_i}
\]

\[
L_{G_j} = \text{Maxpool}\{f_t f_{t-1} \ldots f_1(x_i)\}_{x_i \in G_j}
\]

We use the Maxpool function to extract the order-invariant descriptors from the input point sets. The readers can refer to PointNet [21] for detailed discussion. One can also use other symmetric operators such as summation, average pooling function and so on. This structure can be easily adapted for 3D point set inputs. Other point signature learning architecture such as PointNet++ [22] can be easily implemented here as well.

3.3. Coherent PointMorph architecture

For the next step, we define a PointMorph MLP (multi-layer perceptrons) architecture for learning the coherent
point drifts to move the source point set towards alignment with the target one as shown in Figure 3. This architecture includes successive multi-layer perceptrons (MLP) with ReLu activation function: \( g_i \), \( i = 1, 2, \ldots, n \), such that: \( g_i : \mathbb{R}^{v_i} \to \mathbb{R}^{v_{i+1}} \), where \( v_i \) is the input layer’s dimension and \( v_{i+1} \) is the output layer’s dimension. Therefore, \( \forall (S_i, G_j) \),

\[
dx_i = g_s g_{s-1} \ldots g_1([x_i, L_{S_i}, L_{G_j}])
\]

(4)

\[
S'_i = \phi(S_i) = \{ x_i + dx_i \}_{x_i \in S_i}
\]

(5)

where \( S'_i \) is the transformed source point set and \( dx_i \) represents the predicted drift for each point \( x_i \in S_i \). The notation \([*, *] \) represents concatenation of vectors in same domain. We should notice that due to the high non-linearity of neural networks, there is no difficulty to minimize the similarity loss function between \( S' \) and \( G \). However, the main challenge is to make sure the predicted drifts are coherent [20]. The coherency of drifts is quite important for holding reasonable correspondence and points interpolation for registering large scale point sets as well. Most previous methods deal with this problem by adding a penalization term on smoothness to trade-off target alignment loss for smoothness [4]. Our network naturally predicts smooth drifts because (1) by formula (6), for any neighbor points \( x_k \in S \), with concatenating two identical global descriptors, the inputs \([x_k, L_{S_i}, L_{G_j}] \) should be very close to each other as input for function \( g \). (2) Function \( g \) is a simple linear combination with ReLu activation. This function is continuous and its first derivative should be a constant almost everywhere. Based on the special function \( g \) and our assumption, we have the following theorem.

**Claim:** For a single layer perceptron with ReLu activation function \( g \), \( \forall x_i, x_j \in S, \forall \epsilon > 0, \exists \delta > 0 \), such that \( ||x_i - x_j|| < \epsilon \implies ||dx_i - dx_j|| < \delta \), where \( dx_i \) is defined as \( dx_i = g([x_i, L_{S_i}, L_{G_j}]) \) similarly in equation (6).

**Proof in sketch.** Since \( g \) is a linear function with ReLu activation and we assume its weights \( w \) converges to \( w' \) after training. \( ||x_i - x_j|| = ||[x_i, L_{S_i}, G_j] - [x_j, L_{S_j}, L_{G_j}]|| \) since \( L_{S_j}, L_{G_j} \) are identical for each \( x_i \in S \). Since \( w' \) is constant, \( \exists C > 0 \), s.t. \( ||x_i - x_j|| > C \implies ||w'[x_i, L_{S_i}, G_j] - w'[x_j, L_{S_j}, L_{G_j}]|| \). Since function \( g \) is continuous, if \( w'[x_i, L_{S_i}, L_{G_j}] \geq 0, \exists \delta_i > 0 \) such that \( \forall x_i, \text{ if } ||x_i - x_j|| < \delta_i \text{, } w'[x_i, L_{S_i}, L_{G_j}] \geq 0 \). Similarly if \( w'[x_i, L_{S_i}, L_{G_j}] \leq 0, \exists \delta_i > 0 \) such that \( \forall x_j, \text{ if } ||x_i - x_j|| < \delta_j \text{, } w'[x_j, L_{S_j}, L_{G_j}] \leq 0 \). Therefore, we pick \( \delta = \min(\delta_1, \delta_2, \epsilon/C) \).

\[
||dx_i - dx_j|| = ||g([x_i, L_{S_i}, L_{G_j}]) - g([x_j, L_{S_j}, L_{G_j}])|| = ||\max(w'[x_i, L_{S_i}, L_{G_j}], 0) - \max(w'[x_j, L_{S_j}, L_{G_j}], 0)|| = \max(||w'[x_i, L_{S_i}, L_{G_j}] - w'[x_j, L_{S_j}, L_{G_j}]||, 0) < \max(\frac{\epsilon}{C} ||x_i - x_j||, 0) < \max(0, \epsilon/C) = \epsilon/C < \delta.
\]

4.4. Loss functions

In this part, we define the similarity measure between the transformed source point set \( \phi(S_i) \) and the target point set \( G_j \) as both our loss function and evaluation metric. For two point sets, due to the absence of the corresponding relationship for each point, we cannot adopt the pixel-wise loss in image registration. Fan et al. [7] first proposed Chamfer Distance (C.D.), which is widely used in practice. We define the Chamfer loss on our transformed source point set \( S' \) and target points set \( G \) as:

\[
L(S', G; \theta) = \sum_{x \in S'} \min_{y \in G} ||x - y||^2_2 + \sum_{y \in G} \min_{x \in S'} ||x - y||^2_2
\]

(6)

where \( \theta \) represents all the weights in the our network structure. Chamfer loss (C.D.) is our main choice in this paper. For dataset in presence of outliers and missing points noise, we use the following clipped Chamfer loss:

\[
L(S', G; \theta) = \sum_{x \in S'} \max(\min_{y \in G} ||x - y||^2_2, c) + \sum_{y \in G} \max(\min_{x \in S'} ||x - y||^2_2, c)
\]

(7)

where \( c \) is a hyper-parameter to choose. In our experiment 3, we choose \( c \) equal to 0.1.

3.5. Settings of CPD-Net

We train our network using batch data form training data set \( \{S^i, G^i\} \in D_{i=1,2,\ldots,b} \) b is the batch size and is set to 16. As we explain in section 3.2, for learning the shape descriptor tensor, the input is \( N \times 4/6 \) matrix and we use 5 MLP layers with dimensions (16, 64, 128, 256, 512) and a Maxpool layer to convert it to a 512-dimensional descriptor. For learning the coherent PointMorph discussed in 3.3, we use three layers MLP with dimension (256, 128, 2/3). We use ReLU activation function and implement batch normalization [9] for every layer except the output layer. Learning rate is set as 0.0001 with 0.995 exponential decay with Adam optimizer. The model is trained on single Tesla K80 GPU.
Figure 4. The qualitative registration result for Fish shape at different deformation level. The blue shape is target point set. The red shape is source point set. The black lines are predicted coherent drifts for source point set. Please zoom-in for better visualization.

4. Experiments

In this section, we carry out a set of experiments for non-rigid point set registration and assess the performance of our proposed CPD-Net. In section 4.1, we describe the details of datasets that are used for training and testing of CPD-Net. We report the experimental results to evaluate the performance of our trained CPD-Net on 2D and 3D datasets in section 4.2 and 4.3. Section 4.4 discusses the resistance of CPD-Net to various types of noise. In section 5.5, we compare CPD-Net with non-learning based method.

4.1. Experimental Dataset

A variety of different 2D and 3D shapes (i.e. examples shown in Figure 4, Figure 6, and Figure 7) are used in the experiments to train and test the CPD-Net. In experiments, we prepare the dataset as follows:

- To prepare the deformable shape dataset (as shown in the first column row of Figure 4), we simulate non-rigid geometric transformation on the normalized raw point sets by thin plate spline (TPS) [6] transformation with different deformation levels. The deformation level is defined as the perturbing degree of controlling points in TPS. Specifically, given the deformation level set at \( l \) (e.g. 0.5), a Gaussian random shift with zero-mean and \( 2l \) standard deviation is generated to perturb the controlling points.

- To prepare the Gaussian Displacement (G.D.) noise dataset (as shown in the first row of Figure 8), we simulate the random displacement superimposed on a deformed point set (deformation level at 0.5) by adding an increasing number of Gaussian outliers. The G.D. noise level is defined using the standard deviation of Gaussian.

- To prepare the Point Outlier (P.O.) noise for the shape (as shown in the second row of Figure 8), we simulate the outliers on the deformed point set (deformation level at 0.5) by adding an increasing number of Gaussian outliers. The P.O. noise level is defined as a ratio of Gaussian outliers and the entire target point set.

- To prepare the Data Incompleteness (D.I.) noise (as shown in the second row of Figure 8), we remove an increasing number of neighbouring points from target point set (deformation level at 0.5). The D.I. noise level is defined as the percentage of the points removed from the entire target point set.

4.2. 2D non-rigid point set registration

In this experimental section, we demonstrate the point set registration performance of the CPD-Net on various categories of 2D shapes at different deformation levels.

Experiment Setting: In this experiment, we use four different type of 2D shapes (i.e. fork, face.) to prepare the dataset. For each shape, we first synthesize a set of \( 20k \) deformed shapes at each deformation level. The deformation level ranges from 0.3 to 2.0. To prepare the training dataset, for each type of shape at each deformation level, we split synthesized dataset into two groups. We randomly choose a pair of shapes from group one to form \( 20k \) pairs of training. Similarly, we randomly choose two shapes from the other synthesized dataset to form \( 10k \) testing pairs. Note that there is no intersection between training and testing datasets. To evaluate the registration performance, we use the Chamfer Distance (C.D.) between the transformed source point set and target one as quantitative assessment, and we visualize the pairwise point sets before and after registration for qualitative assessment. We conduct two tests based on the shapes we used in the test. For test one, we use the commonly adopted fish shape, and in test two, we use other 2D shape categories, such mushroom, face, and fork, to assess CPD-Net’s performance, particularly on...
| Def. level | 0.3    | 0.5    | 0.7    | 1.2    | 1.5    | 2.0    |
|------------|--------|--------|--------|--------|--------|--------|
| Fish       | 0.0008±0.0004 | 0.0037±0.0009 | 0.0072±0.0022 | 0.0178±0.0069 | 0.0239±0.0096 | 0.037±0.016 |
| Mushroom   | 0.0006±0.0003 | 0.0031±0.0009 | 0.0051±0.0018 | 0.0103±0.0045 | 0.0129±0.0058 | 0.0196±0.0094 |
| Fork       | 0.0002±0.0001 | 0.0014±0.0011 | 0.0038±0.0019 | 0.0089±0.0048 | 0.0126±0.0074 | 0.0203±0.0127 |
| Face (2D)  | 0.0005±0.0003 | 0.0028±0.0011 | 0.0053±0.0017 | 0.0114±0.0049 | 0.0158±0.0074 | 0.0213±0.01 |

Table 1. Quantitative testing performance for 2D point set registration.

Results of Test 1: After training CPD-Net, we applied the trained model to testing dataset prepared as described above. In Figure 4, shapes from different rows have different deformation levels from 0.1 to 1.5. From the fourth column, we can observe that CPD-Net can predict nearly perfect registration when the deformation level is smaller than 0.9. While we increase the deformation to the level greater than 1.5, the source and target point sets have significant shape structural variation, which dramatically increase the difficulty of point set registration. However, CPD-Net can still reliably transform the source point set to align the main portion of the shape of the target point set. In addition, it is interesting to observe from the second and third columns that source point set moves coherently as a whole towards the target one. This observation verifies that CPD-Net is able to predict a continuous smooth displacement field without necessity to impose additional coherence constraint term. We further provide the mean and standard deviation of C.D. calculated from the 10K testing pairs at each deformation level for quantitative evaluation. In the Figure 5, we plot the mean and standard deviation for the set of C.D. between the source point set and the target one at all deformation levels. We can see from the comparison that the red curve (post registration) is consistently below the blue one (pre registration), which indicates CPD-Net is able to robustly register the source and target point set with a small C.D. Moreover, the red curve stays nearly flat as the deformation increases from 0.1 to 2.0, which indicates that CPD-Net’s robust performance at high deformation level.

The detailed qualitative result is presented in Table 1.

Results of Test 2: In this test, we further use CPD-Net to perform the non-rigid registration on other three types of 2D shapes including Mushroom, Fork, and Face shapes as shown in Figure 6. To visualize the registration result, we compare the pair of testing shapes before and after registration at deformation level 0.5 as shown in Figure 6. All randomly selected samples show nearly perfect registration. Similar to test 1, we present quantitative evaluation using C.D. metric for the non-rigid registration of those three types of shapes in the Table 1. Each row contains the mean and standard deviation of C.D. measurement for all testing pairs of shapes at deformation level from 0.3 to 2.0. Based on the quantitative results shown in the Figure 5 for all the four shapes, CPD-Net demonstrates the remarkable performance of non-rigid registration as evidenced by the fact that the C.D. is dramatically reduced and consistently stays low after alignment at all deformation levels.

4.3. 3D non-rigid point set registration

In this experiment, we take a further step to investigate how well the CPD-Net performs 3D point set registration at different deformation levels since the 3D data have significant contour-based shapes.

The results of this experiment show that CPD-Net can reliably transform the source point set to align the main portion of the shape of the target point set. In addition, it is interesting to observe from the second and third columns that source point set moves coherently as a whole towards the target one. This observation verifies that CPD-Net is able to predict a continuous smooth displacement field without necessity to impose additional coherence constraint term. We further provide the mean and standard deviation of C.D. calculated from the 10K testing pairs at each deformation level for quantitative evaluation. In the Figure 5, we plot the mean and standard deviation for the set of C.D. between the source point set and the target one at all deformation levels. We can see from the comparison that the red curve (post registration) is consistently below the blue one (pre registration), which indicates CPD-Net is able to robustly register the source and target point set with a small C.D. Moreover, the red curve stays nearly flat as the deformation increases from 0.1 to 2.0, which indicates that CPD-Net’s robust performance at high deformation level. The detailed qualitative result is presented in Table 1.

Results of Test 2: In this test, we further use CPD-Net to perform the non-rigid registration on other three types of 2D shapes including Mushroom, Fork, and Face shapes as shown in Figure 6. To visualize the registration result, we compare the pair of testing shapes before and after registration at deformation level 0.5 as shown in Figure 6. All randomly selected samples show nearly perfect registration. Similar to test 1, we present quantitative evaluation using C.D. metric for the non-rigid registration of those three types of shapes in the Table 1. Each row contains the mean and standard deviation of C.D. measurement for all testing pairs of shapes at deformation level from 0.3 to 2.0. Based on the quantitative results shown in the Figure 5 for all the four shapes, CPD-Net demonstrates the remarkable performance of non-rigid registration as evidenced by the fact that the C.D. is dramatically reduced and consistently stays low after alignment at all deformation levels.

4.3. 3D non-rigid point set registration

In this experiment, we take a further step to investigate how well the CPD-Net performs 3D point set registration at different deformation levels since the 3D data have significant contour-based shapes.
been gaining great attention in community with recent advancements in 3D acquisition and computation resources.

Experimental Setting: In this experiment, we use two categories of 3D shapes (i.e., 3D Face and 3D Cat) to prepare the dataset. Similar to 2D data preparation, we synthesize $10k$ training pairs of 3D shapes and $2k$ testing pairs for both 3D Face and Cat shapes at various deformation levels ($0.3$, $0.5$, $0.8$, and $1.2$). We use the same measurement methods for both qualitative and quantitative results (as shown in Figure 7).

Result: In Figure 7 we illustrate the quantitative evaluation curves on the left and visualize one registration cases at deformation level $0.3$ on the right. For both 3D cat and face shapes, CPD-Net demonstrates impressive performance with a quite small pairwise C.D. after registration, and the C.D. measurement remains consistently low while we increase the deformation level. At the level of $1.2$, the mean of C.D. between source and target point sets is nearly $10$ times less after alignment, which indicates that the trained CPD-Net is able to align most portion of the shapes between a source and target point sets. To further visualize the registration performance, on the right side of Figure 7 we show the registration results by plotting the figures for 3D source and target point sets, pre and post registration in first and third columns and the second column depicts the predicted coherent point drift vector. Those plots clearly prove that CPD-Net is able to predict an accurate smooth non-rigid transformation. Due to the limited space, more qualitative results will be showed in supplementary section.

4.4. Resistance to Noise

While using the sensors such as LIDAR sensor and laser scanner, it is unavoidable that the data might be acquired with a variety types of noises. An effective non-rigid registration method should be robust to those noise in addition to the structural variations as discussed in previous section. Therefore, in this section, we focus on testing how well CPD-Net can predict the non-rigid registration from the noisy dataset.

Experimental Setting: In this section, we use fish shape data at deformation level of $0.5$ to prepare the experimental noisy dataset. We simulate three types of noise (i.e., Gaussian Displacement (G.D.) noise, Point Outlier (P.O.) noise and Data Incompleteness (D.I.) noise). For each type, we gradually increase the level of noise added to the deformed target point set of fish dataset as shown in Figure 8. We prepare $10k$ pairs of source and target point sets for each type of noise for testing. As in previous section, the same quantitative and qualitative performance measurement is used in this experiment.

Result: As shown in the Figure 8 we plot the quantitative evaluation curves on the left and visualize five registration
cases at different noise level on the right. For the clean data, the mean of C.D. for the fish shape at deformation level of 0.5 is around 0.08. We need to validate if CPD-Net can significantly reduce the C.D. for pairs of source and target point sets of noisy dataset after registration, and if CPD-Net consistently keep the C.D. comparatively low when the noise level increases.

For the G.D. noise, in the Figure 8, the first row depicts the registration by our CPD-Net for the G.D. noise corrupted data. As we notice the plot, the C.D. after registration remains constantly lower than 0.08, even when the G.D. noise increases to the level of 0.2. CPD-Net can still predict the non-rigid transformation to align the source set (the red points) to the target one (the blue points) with a relatively small C.D. between two point sets, even though the shape was dramatically altered by the Gaussian Displacement noise. For the P.O. noise, as shown in the second row, outlier points is increasingly added to the target point set (blue ones) from left to right in a row. The registration result is impressive that the main body of the source and target shapes can robustly aligned to each other with a small C.D. between them after registration, especially when the P.O. noise level reaches as high as 0.3. For D.I. noise, as shown in the third row, an increasing number of points is removed from the target point set (blue ones) to check if CPD-Net is able to successfully align the source point set to the incomplete portion of the target one. The visualization of pairwise registration result in the third row clearly demonstrates that CPD-Net is able to robustly align the source point set (red) to the incomplete target point set (blue). When D.I. noise level reaches 0.25, the missing part is aligned with a straight line, which is less desired. But the aligned portions from transformed source point set and target one show consistent common geometric structure, which is not affected by the missing portion of the target point set.

4.5. Comparison to CPD

Different from previous efforts, the proposed CPD-Net is a learning-based non-rigid point set registration method, which can learn the registration pattern to directly predict the non-parametric geometric transformation for the point sets alignment. As a learning-based approach to predict the non-rigid registration, it is not applicable to have a direct comparison between CPD-Net and other existing non-rigid iterative registration methods. To compare our method to non-learning based iterative method (i.e. Coherent Point Drift (CPD) [20]), we design the experiment as follows to assess both time and accuracy performance.

Experimental Setting: In this experiment, we use the fish shape at deformation level of 0.5 to prepare the dataset. We synthesize 20k pairwise source and target point sets to form the training set, and another 20k pairs to form the testing set. The CPD-Net is firstly trained with the 20k training dataset. The trained CPD-Net is then applied to directly predict registration for the 20k testing dataset. In contrast, CPD is directly applied to both 20k training and testing dataset.

Result: The C.D. based quantitative comparison is presented in the Table 2. The first and third row list the experimental results for CPD on training and testing dataset respectively. The second row lists the results for the CPD-Net on training dataset, and the fourth row lists the results for the trained CPD-Net on the testing dataset. Based on the comparison between first and third rows, we can see that our model can achieve better performance (i.e. smaller C.D.) than that of CPD within a shorter time (25 minutes versus 22 hours) to align 20k pairs. Unlike the CPD needs to start over a new iterative optimization process to register a new pair of shapes independently, CPD-Net actively learns the registration pattern from training and consequently become capable of handling real-time point set registration or a large volume dataset by direct predicting the geometric transformation. As shown on second and fourth row of the Table 2, we notice that the trained CPD-Net is able to achieve nearly the same training performance, which indicates that CPD-Net has great generalization capability. The trained CPD-Net is able to achieve better performance than CPD on the same dataset with orders of magnitude less time (15 seconds versus 22 hours).

5. Conclusion

In this paper, we report our latest research effort towards non-rigid point set registration. The proposed method, Coherent Point Drift Network (CPD-Net), is presented to the research community as an new alternative solution for non-rigid point registration. In contrast to existing approaches, which often iteratively search for the optimal geometric transformation to register a given pair of point sets, driven by minimizing a pre-defined alignment loss function. The proposed CPD-Net can actively learn the registration pattern from a training dataset as non-parametric geometric transformation in an unsupervised manner. Consequently CPD-Net can predict the desired geometric transformation to align a pair of unseen point sets. To the best of our knowledge, this is the first effort in the community that an algorithm can actually learn to register point sets from training.
in unsupervised fashion. The CPD-Net can be extended to other modality data such as 2D and 3D image and a high-dimensional multimedia data. We will present the extension of the CPD-Net in separate works.

6. Supplementary Materials: Demonstration of Additional Qualitative Results

Due to the space limitation, in the submitted paper we only show one qualitative example for each case in Figure 4 and 7. For better understanding our performance in Figure 4 and 7 of the submitted paper, we randomly select several registration results of testing samples to demonstrate in the supplementary section as we mentioned in the paper. The following Figure 9 is supplementary to Figure 4 of the submitted paper and Figure 10 is supplementary to Figure 7 of the submitted paper.

References

[1] Song and Bai Bai Xiang and Zhou, Gift: A real-time and scalable 3d shape search engine, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 5023–5032.

[2] Xiang and Latecki Bai Longin Jan, Path similarity skeleton graph matching, IEEE transactions on pattern analysis and machine intelligence 30 (2008), no. 7, 1282–1292.

[3] Xiang and Latecki Bai Longin Jan and Liu, Skeleton pruning by contour partitioning with discrete curve evolution, IEEE transactions on pattern analysis and machine intelligence 29 (2007), no. 3.

[4] Guha and Zhao Balakrishnan Amy and Sabuncu, An Unsupervised Learning Model For Deformable Medical Image Registration, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9252–9260.

[5] Paul J and McKay Besl Neil D, Method for registration of 3-D shapes, Sensor Fusion IV: Control Paradigms and Data Structures, 1992, pp. 586–607.

[6] Fred L. Bookstein, Principal warps: Thin-plate splines and the decomposition of deformations, IEEE Transactions on pattern analysis and machine intelligence 11 (1989), no. 6, 567–585.

[7] Haili and Rangarajan Chui Anand, A new algorithm for non-rigid point matching, Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on, 2000, pp. 44–51.

[8] Haoqiang and Su Fan Hao and Guibas, A point set generation network for 3d object reconstruction from a single image, arXiv preprint arXiv:1612.00603 (2016).

[9] Sergey and Szegedy Ioffe Christian, Batch normalization: Accelerating deep network training by reducing internal covariate shift, International Conference on Machine Learning, 2015, pp. 448–456.

[10] Bing and Vemuri Jian Baba C and Pollefeys, Non-rigid point set registration using gaussian mixture models, IEEE transactions on pattern analysis and machine intelligence 33 (2011), no. 8, 1633–1645.

[11] Andreas and Sormann Klaus Mario and Karner, Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure, Pattern Recognition, 2006. 18th International Conference on, 2006, pp. 15–18.

[12] Haibin and Jacobs Ling David W, Deformation invariant image matching, Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, 2005, pp. 1466–1473.

[13] Jiayi and Zhao Ma Ji and Tian, Robust estimation of nonrigid transformation for point set registration, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 2147–2154.

[14] Jiayi and Zhao Ma Ji and Tian, Robust point matching via vector field consensus., IEEE Trans. image processing 23 (2014), no. 4, 1706–1721.

[15] Jiayi and Zhao Ma Ji and Yuille, Non-rigid point set registration by preserving global and local structures, IEEE Transactions on image Processing 25 (2016), no. 1, 53–64.

[16] JB Antoine and Viergever Maintz Max A, A survey of medical image registration, Medical image analysis 2 (1998), no. 1, 1–36.

[17] Jonathan and Boscaini Masci Davide and Bronstein, Geodesic convolutional neural networks on riemannian manifolds, Proceedings of the IEEE international conference on computer vision workshops, 2015, pp. 37–45.

[18] Daniel and Scherer Maturana Sebastian, VoxNet: A 3d convolutional neural network for real-time object recognition, Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on, 2015, pp. 922–928.

[19] Andriy and Song Myronenko Xubo, Image registration by minimization of residual complexity (2009).

[20] Andriy and Song Myronenko Xubo and Carreira-Perpinán, Non-rigid point set registration: Coherent point drift, Advances in Neural Information Processing Systems, 2007, pp. 1009–1016.

[21] Charles R and Su Qi Hao and Mo, Pointnet: Deep learning on point sets for 3d classification and segmentation, Proc. Computer Vision and Pattern Recognition (CVPR), IEEE 1 (2017), no. 2, 4.

[22] Charles Ruizhongtai and Yi Qi Li and Su. Pointnet++: Deep hierarchical feature learning on point sets in a metric space, Advances in Neural Information Processing Systems, 2017, pp. 5099–5108.

[23] Rahul and Frahm Raguram Jan-Michael and Pollefeys, A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus, European Conference on Computer Vision, 2008, pp. 500–513.

[24] Ignacio and Arandjelovic Rocco Relja and Sivic, Convolutional neural network architecture for geometric matching, Proc. CVPR, 2017.

[25] Abbasiak and Grau Sharma Oliver and Fritz, Vcom-dae: Deep volumetric shape learning without object labels, Computer Vision–ECCV 2016 Workshops, 2016, pp. 236–250.

[26] Milan and Hlavac Sonka Vaclav and Boyle, Image processing, analysis, and machine vision, Cengage Learning, 2014.

[27] Michael and Snavely Goesele Noah and Curless, Multi-view stereo for community photo collections, Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, 2007, pp. 1–8.

[28] Lingjing and Chen Wang Jianchun and Li, Non-Rigid Point Set Registration Networks, arXiv preprint arXiv:1904.01428 (2019).

[29] Nutika and Boyer Verma Edmond and Verbeek, FeaStNet: Feature-Steered Graph Convolutions for 3D Shape Analysis, CVPR 2018-IEEE Conference on Computer Vision & Pattern Recognition, 2018.

[30] Yi and Shen Wu Bin and Ling, Online robust image alignment via iterative convex optimization, 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 1808–1814.

[31] Alan L and Grzywacz Yuille Norberto M, A computational theory for the perception of coherent visual motion, Nature 333 (1988), no. 6168, 71.

[32] Andy and Song Zeng Shuran and Nießner, 3dmatch: Learning local geometric descriptors from rgb-d reconstructions, Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, 2017, pp. 199–208.
Figure 9. Supplementary to Figure 4 of the submitted paper. The testing qualitative registration results for Fish shape at different deformation level. The blue shape is target point set. The red shape is source point set. Please zoom-in for better visualization.

[33] Hang and Maji Su Subhransu and Kalogerakis, *Multi-view convolutional neural networks for 3d shape recognition*, Proceedings of the IEEE international conference on computer vision, 2015, pp. 945–953.
Table 10. Supplementary to Figure 7 of the submitted paper. The testing qualitative registration results for 3D shapes at different deformation level. The red points represents the source points and the blue ones represent the target points. Please zoom-in for better visualization.

| Deform. Level | 3D Face  | 3D Cat  |
|---------------|---------|---------|
|               | 0.3     | 0.5     | 0.8     |
| Inputs        | ![Image](image1) | ![Image](image2) | ![Image](image3) |
| Outputs       | ![Image](image4) | ![Image](image5) | ![Image](image6) |
| Inputs        | ![Image](image7) | ![Image](image8) | ![Image](image9) |
| Outputs       | ![Image](image10) | ![Image](image11) | ![Image](image12) |
| Inputs        | ![Image](image13) | ![Image](image14) | ![Image](image15) |
| Outputs       | ![Image](image16) | ![Image](image17) | ![Image](image18) |

Figure 10. Supplementary to Figure 7 of the submitted paper. The testing qualitative registration results for 3D shapes at different deformation level. The red points represents the source points and the blue ones represent the target points. Please zoom-in for better visualization.