Deep-3DAligner: Unsupervised 3D Point Set Registration Network With Optimizable Latent Vector

Lingjing Wang, Xiang Li and Yi Fang

Abstract-Point cloud registration is the process of aligning a pair of point sets via searching for a geometric transformation. Unlike classical optimization-based methods, recent learning-based methods leverage the power of deep learning for registering a pair of point sets. In this paper, we propose to develop a novel model that organically integrates the optimization to learning, aiming to address the technical challenges in 3D registration. More specifically, in addition to the deep transformation decoding network, our framework introduce an optimizable deep Spatial Correlation Representation (SCR) feature. The SCR feature and weights of the transformation decoder network are jointly updated towards the minimization of an unsupervised alignment loss. We further propose an adaptive Chamfer loss for aligning partial shapes. To verify the performance of our proposed method, we conducted extensive experiments on the ModelNet40 dataset. The results demonstrate that our method achieves significantly better performance than the previous state-of-theart approaches in the full/partial point set registration task.

1 Introduction

Point set registration is a challenging but meaningful task, which has wide application in many fields [1], [2], [3], [4], [5], such as mapping, shape recognition, correspondence, large scale scene reconstruction, and so on. Most existing non-learning methods solve the registration problem through an iterative optimization process to search the optimal geometric transformation to minimize a predefined alignment loss between the transformed source point set and target point set [6], [7], [8], [9], [10]. Iterative methods usually treat registration as an independent optimization process for each given pair of source and target point sets, which cannot transfer knowledge from registering one pair to another.

In comparison, as shown in Figure 1, instead of directly optimizing the transformation matrix towards minimization of alignment loss in non-learning based methods, learning-based methods usually leverage modern feature extraction technologies for feature learning and then regress the transformation matrix based on the mutual information and correlation defined on the extracted features of source and target shapes. The most recent model, deep closest point (DCP) [11], leverages DGCNN [12] for feature learning and a pointer network to perform soft matching. Learning-based methods can greatly improve the efficiency by

L.Wang is with MMVC Lab, New York University Abu Dhabi, UAE and Dept. of ECE, e-mail: lingjing.wang@nyu.edu. X.Li is with the MMVC Lab, New York University Abu Dhabi, e-mail: xl1845@nyu.edu. Y.Fang is with MMVC Lab, Dept. of ECE, NYU Abu Dhabi, UAE and Dept. of ECE, NYU Tandon School of Engineering, USA, e-mail: yfang@nyu.edu. Corresponding author. Email: yfang@nyu.edu

direct prediction of the transformation matrix for testing pairs. However, these methods' performance highly depends on the number and quality of the labeled training dataset and the performance may greatly degrade for the testing dataset of unseen categories. In contrast, our proposed network can not only leverage the deep decoding structure to learn the registration pattern from training data but also leverage a directly optimizable SCR feature to further refine the desired transformation in an unsupervised manner, which is different from the DCP that uses the ground-truth transformation parameters (i.e. rotation and translation matrix) for training.

For most cases in practice, input point sets may suffer from various noise such as shape incompleteness [13]. When dealing with partial shapes, classical deep learning-based methods suffer from performance degradation, especially when the overlapping subsets between source and target shapes are small. Wang et al. [14] proposed the first registration method designed for solving the partial point set registration problem. This method still requires the process of detecting key corresponding points from the partial input shapes. When the overlapping area is small, the corresponding points are difficult to be detected, which can lead to inferior registration performance. In contrast, we propose an adaptive Chamfer loss to detect the corresponding subsets between source and target point sets instead of a few key points. Driven by this designed loss, the desired geometric transformation can be gradually optimized based on the detected overlapping subsets between source and target point sets in a coarse-to-fine way.

With the development of the SCR feature, our proposed Deep-3DAligner framework is illustrated in Figure 2, which contains three main components. The first component is an SCR optimizer where the deep SCR feature is optimized from a randomly initialized feature. The second component is a transformation decoder which decodes the SCR feature to regress the transformation parameters for the point sets alignment. The third component is an alignment loss that measures the similarity between the transformed source point set and the target one. In the pipeline, there are two communication routes, indicated by black and red dashed lines. The communication route in black is for the data flow for the Deep-3DAligner paradigm, where the source and target point sets are used as input. The communication route in red is the back-propagation route with which the alignment loss is backpropagated to update the SCR and the transformation decoder. Our contribution is as follows:

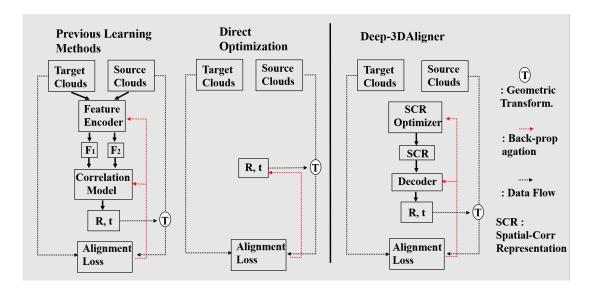


Fig. 1. Comparison of the pipeline between previous learning methods, direct optimization methods and our Deep-3DAligner for point set registration. Our method starts with optimizing a randomly initialized latent spatial correlation representation (SCR) feature, which is then decoded to the desired geometric transformation to align source and target point clouds, integrating the optimization-based SCR optimizer with the learning-based decoder to enhance the model's registration capacity.

- We introduce a novel unsupervised learning approach for the point set registration task.
- We introduce a spatial correlation representation (SCR) feature which can eliminate the design challenges for encoding the spatial correlation between source and target point sets in comparison to learning-based methods.
- We propose an adaptive Chamfer loss to gradually detect the overlapping areas between the transformed source point set and target point set to refine the desired geometric transformation in a coarse-to-fine approach.
- Experimental results demonstrate the effectiveness of the proposed method for point set registration, and even without ground truth transformation for training, our proposed approach achieved superior performance in 3D full/partial point sets registration compared to most recent supervised state-of-the-art approaches.

2 RELATED WORKS

2.1 Iterative registration methods

The development of optimization algorithms to estimate rigid and non-rigid geometric transformations in an iterative routine has attracted extensive research attention in past decades. The iterative closest point (ICP) algorithm [15] 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. Go-ICP [16] was further proposed by Yang et al. to leverage the BnB scheme for searching the entire 3D motion space to solve the local initialization problem brought by ICP. Zhou et al. proposed fast global registration [17] for the registration of partially overlapping 3D surfaces. The TPS-RSM algorithm was proposed by Chui and Rangarajan [18] to estimate parameters of non-rigid transformations with a penalty on second-order derivatives. Coherence point drift (CPD) was further proposed by Myronenko et al. [6] for non-rigid point set registration. Although the independent iterative optimization process limits the efficiency of registering a large number of pairs,

inspiring us to leverage its advantage and equip it with a learning-based system for this task.

2.2 Learning-based registration methods

Recent works have 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 [14], [19], [20]. PointNetLK [21] was proposed by Aoki et al. to leverage the newly proposed PointNet algorithm for directly extracting features from the point cloud with the classical Lucas & Kanade algorithm for the rigid registration of 3D point sets. Liu et al. proposed FlowNet3D [22] to treat 3D point cloud registration as a motion process between points. Wang et al. proposed a deep closest point [11] model, which first leverages the DGCNN structure to exact the features from point sets and then regress the desired transformation based on it. Balakrishnan et al. [14] proposed a voxelMorph CNN architecture to learn the registration field to align two volumetric medical images. In contrast, we first propose a model-free structure to skip the encoding step. Instead, we initialize an SCR feature without pre-defining a model, which is to be optimized with the weights of the network from the alignment loss back-propagation process. PR-Net [13] as the first work proposed a method to detect corresponding points from partial shapes and then solve the desired geometric transformation based on it. In comparison, our method detects the overlapping subsets based on our proposed adaptive Chamfer loss.

3 Approach

We introduce our approach in the following sections. First, we define the learning-based registration problem in section Problem statement. In section Spatial-Correlation Representation, we introduce our spatial-correlation representation. The transformation decoder is illustrated in section Transformation Decoder. In section Loss function, we provide the definition of the loss function. Section Optimization Strategy illustrates the newly defined optimization strategy.

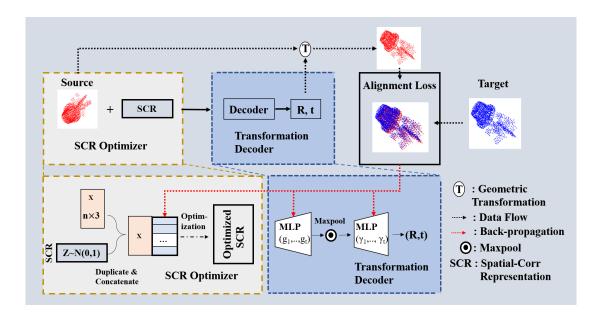


Fig. 2. Our pipeline. For a pair of input source and target point sets, our method starts with the SCR optimization process to generate a spatial-correlation representation feature, and a transformation regression process further decodes the SCR feature to the desired geometric transformation. The alignment loss is back-propagated to update the weight of the transformation decoder and the SCR feature during the training process. For testing, the weights of the transformation decoder remain constantly without updating.

3.1 Problem statement

Given training dataset $\mathbf{D}=\{(S_i,G_i)$, where $S_i\in\mathbf{S},G_i\in\mathbf{G}\}$. \mathbf{S} is the source point set and \mathbf{G} is the target point set, and we have $\mathbf{S},\mathbf{G}\subset\mathbb{R}^N(N=2\text{ or }N=3)$. Assuming the existence of a function $g_{\theta}(S_i,G_i)=\phi$ and g has parameter set θ in a neural network structure. When considering only rigid point set registration, the output ϕ usually contains a homogeneous transformation matrix including a rotation matrix and a translation vector. A model with optimized weights $\theta^{optimal}$ can generate the desired rotation and translation parameters ϕ in g to further align the source and target point sets. The objective loss function $\mathcal L$ is usually a pre-defined similarity metric for evaluation of the alignment quality between transformed source and target point sets. Based on a given dataset $\mathbf D$, a stochastic gradient-descent-based algorithm can be used to update the weight parameters θ and to minimize the pre-defined loss function:

$$\theta^{optimal} = \underset{\theta}{\operatorname{argmin}} [\mathbb{E}_{(S_i, G_i) \sim \mathbf{D}} [\mathcal{L}(S_i, G_i, g_{\theta}(S_i, G_i))]] \quad (1)$$

3.2 Spatial-Correlation Representation

In this paper, we define the spatial correlation representation as the latent feature that characterizes the essence of spatial correlation between a given pair of source and target point sets. As shown in Figure 1, to compute the SCR feature, source and target point sets are usually fed to a feature in previous works for the deep spatial feature extraction, and followed with a pre-defined correlation module. However, the design of an appropriate feature encoder for unstructured point clouds is challenging compared to the standard discrete convolutions assume the availability of a grid structured input (e.g. 2D image). The limitation of the hand-crafted design of modules for the extraction of individual spatial feature and spatial correlation feature motivates us to design a model-free based SCR as described below.

To eliminate the side effects of the hand-craft design in feature encoder and correlation module and to better equip the optimization process within the system, as shown in Figure 2, we define a trainable latent SCR (Spatial-Correlation Representation) feature for each pair of point sets. The design of SCR makes it possible to both leverage the common "knowledge" of registration from the dataset via the shared generator and individual adjustment for each input pair through the optimization of their SCR. After optimization, SCR should contain the spatial correlation information for each input pair. As shown in Figure 2, for a pair of source and target point sets S_i and G_i , the randomly initialized latent vector $z_i \sim \mathcal{N}(0,0.01)$ from Gaussian distribution as an initialized SCR. The initialized SCR is optimized during the training process together with the transformation decoder. The implicit design of SCR allows Deep-3DAligner more flexibility in spatial correlation feature learning that is more adaptive for the alignment of unseen point sets and partial point sets as well.

3.3 Transformation Decoder

Given the above spatial-correlation representation (SCR) feature, we then design a decoding network to regress the desired transformation parameters, as illustrated in Figure 2. More specifically, $\forall x \in S_i$, we stack the coordinates of x with z_i and we note it as $[x,z_i]$. We leverage a multi-layer perceptron (MLP) structure to regress the rotation and translation parameters in the desired transformation. For each layer of the MLP, we define $\{g_i\}_{i=1,2,\dots,s}$, such that $g_i: \mathbb{R}^{v_i} \to \mathbb{R}^{v_{i+1}}$, where v_i is the dimension of input layer and v_{i+1} is the dimension of output layer. For each MLP layer, we use the ReLU activation function. For the output of last layer, we leverage a max pool function to accumulate the point features into a latent vector L_i , calculated as:

$$L_i = Maxpool\{g_s g_{s-1} ... g_1([x_j, z_i])\}_{x_i \in S_i}$$
 (2)

Taking the latent vector L_i as the input, we have a further t successive MLP layers with a ReLU activation function to regress the parameters ϕ_i of the desired transformation. We define

function $\{\gamma_i\}_{i=1,2,...,t}$, such that $\gamma_i: \mathbb{R}^{w_i} \to \mathbb{R}^{w_{i+1}}$, where w_i is the dimension of input layer and w_{i+1} is the dimension of output layer.

$$\phi_i = \gamma_t \gamma_{t-1} \dots \gamma_1(L_i) \tag{3}$$

We further compute the transformed source point set, defined as S_i' ,

$$S_i' = \mathbf{T}_{\phi_i}(S_i) \tag{4}$$

where \mathbf{T}_{ϕ_i} denotes the desired geometric transformation with parameters ϕ_i . Based on the transformed source point set S_i' and the target point set G_i , we further introduce the loss function.

3.4 Loss function

In our unsupervised setting, we do not have the ground truth transformation for supervision and we do not assume a direct correspondence between these two point sets. Therefore, a distance metric between two point sets, instead of the point/pixel-wise loss is desired. In addition, A suitable metric should be differentiable and efficient to compute. In this paper, we adopt the Chamfer distance as our loss function. The Chamfer loss is a simple and effective alignment metric defined on two non-corresponding point sets. We formulate the Chamfer loss between our transformed source point set $S_i' = T_\phi(S_i)$ and target points set G_i as:

$$L_{\text{Chamfer}}(S_i', G_i) = \sum_{x \in S_i'} \min_{y \in G_i} ||x - y||_2^2 + \sum_{y \in G_i} \min_{x \in S_i'} ||x - y||_2^2$$
(5)

For aligning partial-shapes, we further propose an adaptive Chamfer loss. For a time period t during the optimization process, we assume a pre-defined distance threshold σ_t . For the transformed source point set S_i' , we define the overlapping subset of it with the target point set G_i as $S_i'^{(t)} \subset S_i'^{(t-1)} \subset \ldots \subset S_i'^{(0)} = S_i'$, such that $\forall x \in S_i'^{(t-1)}$, if $\min\{||x-y||_2^2\}_{y \in G_i^{(t-1)}} < \sigma_t$, then $x \in S_i'^{(t)}$. Similarly, for the target point set G_i , we define the overlapping subset between it with S_i' as $G_i^{(t)} \subset G_i^{(t-1)} \subset \ldots \subset G_i^{(0)} = G_i$, such that $\forall x \in G_i^{(t-1)}$, if $\min\{||x-y||_2^2\}_{y \in S_i'^{(t-1)}} < \sigma_t$, then $x \in G_i^{(t)}$. Therefore, based on the overlapping subsets $S_i'^{(t)}$ and $G_i^{(t)}$ for time period t, we define the adaptive Chamfer loss for S_i' and G_i as:

$$L_{\text{Adaptive-Chamfer}}^{t}(S_{i}', G_{i}) = \sum_{x \in S_{i}'(t)} \min_{y \in G_{i}^{(t)}} ||x - y||_{2}^{2} + \sum_{y \in G_{i}^{(t)}} \min_{x \in S_{i}'(t)} ||x - y||_{2}^{2}$$
(6)

3.5 Optimization Strategy

In section Spatial-Correlation Representation, we define a set of trainable latent vectors \mathbf{z} , one for each pair of point sets as the SCR feature. During the training process, these latent vectors are optimized along with the weights of network decoder using a stochastic gradient descent-based algorithm. For a given training dataset \mathbf{D} , our training process can be expressed as:

$$\theta^{\mathbf{optimal}}, \mathbf{z^{optimal}} = \operatorname*{argmin}_{\theta, \mathbf{z}} [\mathbb{E}_{(S_i, G_i) \sim \mathbf{D}}[\mathcal{L}(S_i, G_i, g_{\theta}(S_i, z_i))]]$$

where \mathcal{L} represents the pre-defined loss function.

For a given testing dataset W, we fix the network parameters $\tilde{\theta} = \theta^{\text{optimal}}$ and only optimize the SRC features:

$$\mathbf{z}^{\mathbf{optimal}} = \underset{\sigma}{\operatorname{argmin}} [\mathbb{E}_{(S_j, G_j) \sim \mathbf{W}} [\mathcal{L}(S_j, G_j, g_{\tilde{\theta}}(S_j, z_j))]].$$
 (8)

The learned decoder network parameters $\tilde{\theta}$ here provides a prior knowledge for the optimization of SRC. After this optimization process, the desired transformation can be determined by $T_{\phi_i} = T_{g_{\tilde{\theta}}(S_i,z_i^{optimal})}$ and the transformed source shape can be generated by $S_i' = T_{\phi_i}(S_i), \forall S_i \in \mathbf{W}$.

4 EXPERIMENTS

4.1 Dataset

We test the performance of our model for 3D point set registration on the ModelNet40 dataset. This dataset contains 12311 preprocessed CAD models from 40 categories. For each 3D point object, we uniformly sample 1024 points from its surface. Following the settings of previous work, points are centered and re-scaled to fit in the unit sphere. For each source shape S_i we generate the transformed shapes G_i by applying a rigid transformation defined by the rotation matrix which is characterized by 3 rotation angles along the x-y-z-axis, where each value is uniformly sampled from [0,45] unit degree, and the translation which is uniformly sampled from [-0.5,0.5]. At last, we simulate partial point sets by randomly select a point in unit space and keep its 768 nearest neighbors for source and target shapes.

4.2 Settings

We train our network using batch data from the training data set $\{(S_i,G_i)|S_i,G_i\in\mathbf{D}\}_{i=1,2,\ldots,b}$. We set the batch size b to 128. The latent vectors are initialized from a Gaussian distribution $\mathcal{N}(0,0.01)$ with a dimension of 256. For the Deep-3Daligner network, the first part includes 2 MLP layers with dimensions (256,128) and a max pool layer. Then, we use 3 additional MLPs with dimensions of (128, 64, 3) for decoding the rotation matrix and with dimensions of (128, 64, 3) for decoding the translation matrix. We use the leaky-ReLU [23] activation function and implement batch normalization [24] for every layer except the output layer. For adaptive Chamfer, σ_t decreases from 10 to 0.01 in 100 epochs. The learning rate is set as 0.001 with exponential decay of 0.995 at each epoch. We use the mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE) to measure the performance of our model and all comparing methods. Lower values indicate better alignment performance. All angular measurements in our results are in units of degrees. The groundtruth labels are only used for the performance evaluation and are not used during the training/testing process.

4.3 Full 3D point set registration

In this experiment, we follow previous works to test our model for 3D point set registration on unseen point sets in Test 1, and on unseen categories in Test 2.

Experiment Setting: In Test 1, for the 12,311 CAD models from the ModelNet40, following exactly DCP's setting, we split the dataset into 9,843 models for training and 2,468 models for testing. In Test 2, to test the generalizability of our model, we split ModelNet40 evenly by category into training and testing sets

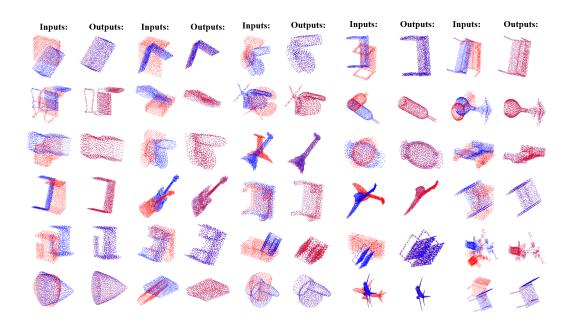


Fig. 3. Randomly selected qualitative results of our model for registration of unseen samples. Left columns: inputs. Right columns: outputs. The red points represent source point sets, and the blue points represent the target point sets.

Model	MSE(R)	RMSE(R)	MAE(R)	MSE(t)	RMSE(t)	MAE(t)
Direct Optimization	406.131713	16.454065	13.932246	0.087263	0.295404	0.253658
ICP [15]	894.897339	29.914835	23.544817	0.084643	0.290935	0.248755
Go-ICP [16]	140.477325	11.852313	2.588463	0.000659	0.025665	0.007092
FGR [17]	87.661491	9.362772	1.999290	0.000194	0.013939	0.002839
PointNetLK [21]	227.870331	15.095374	4.225304	0.000487	0.022065	0.005404
DCPv1+MLP(Supervised) [11]	21.115917	4.595206	3.291298	0.000861	0.029343	0.022501
DCPv2+MLP(Supervised) [11]	9.923701	3.150191	2.007210	0.000025	0.005039	0.003703
DCPv1+SVD(Supervised) [11]	6.480572	2.545697	1.505548	0.000003	0.001763	0.001451
DCPv2+SVD(Supervised) [11]	1.307329	1.143385	0.770573	0.000003	0.001786	0.001195
Ours (MLP-based, Unsupervised)	0.220650	0.350199	0.248512	0.000149	0.008881	0.005021

TABLE 1

ModelNet40: Test on unseen point clouds. Our model is trained in an unsupervised manner without any ground-truth labels. Our model does not require attention mechanism and SVD-based fine-tuning processes.

Model	MSE(R)	RMSE(R)	MAE(R)	MSE(t)	RMSE(t)	MAE(t)
ICP [15]	892.601135	29.876431	23.626110	0.086005	0.293266	0.251916
Go-ICP [16]	192.258636	13.865736	2.914169	0.000491	0.022154	0.006219
FGR [17]	97.002747	9.848997	1.445460	0.000182	0.013503	0.002231
PointNetLK [21]	306.323975	17.502113	5.280545	0.000784	0.028007	0.007203
DCPv1+SVD (Supervised) [11]	19.201385	4.381938	2.680408	0.000025	0.004950	0.003597
DCPv2+SVD (Supervised) [11]	9.923701	3.150191	2.007210	0.000025	0.005039	0.003703
Ours (MLP-based, Unsupervised)	0.280846	0.398275	0.287559	0.000088	0.007547	0.004629

TABLE 2

ModelNet40: Test on unseen categories. Our model is trained in an unsupervised manner without ground-truth labels. Our model does not require SVD-based fine-tuning processes.

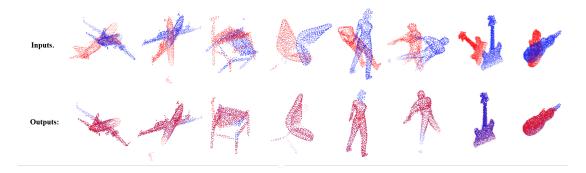


Fig. 4. Qualitative results of partial shapes alignment. From left to right: selected results on the airplane, chair, human, guitar category respectively.

in the same way as DCP. We train our Deep-3DAligner, DCP, and PointNetLK on the divided training dataset and then evaluate the performance on the testing set. ICP, Go-ICP, and FGR are tested directly on the testing dataset. Note that our model is trained without using any ground-truth information and our model does not require the SVD-based fine-tuning processes.

Results of Test 1 (the training/testing split test): We list the quantitative experimental results in Table 1. In this table, we evaluate the performance based on the prediction errors of rotation angles and translation vectors. The first three columns illustrate the comparison results for the rotation angle prediction. For reporting this performance, we ignore the categories of exact symmetric shapes' quantitative results for rotation matrix since in theory there is no unique solution for these cases. For example, for the first cone shown in row 6 of Figure 3, even though the alignment is perfect, we cannot find the unique desired rotation matrix. These categories include bottle, bowl, cone, cup, vase. As shown in Figure 3, we randomly select the qualitative results from the testing dataset. As we can see from the results, our method achieves significantly better performance than the baseline DCPv1+MLP model and also get even better or comparative performance against the DCPv2+SVD version even though we do not require label information for training and we do not require additional SVD layer for fine-tuning. Moreover, considering that DCP assumes the same sampling of points for source and target shapes, we tested that DCP experienced a severe performance degradation in MSE(t) for randomly sampled points of source and target shapes, whereas our model with Chamfer distance is robust to the way of point sampling. For an additional oblation study, we list the performance of direct optimization described in Figure 1. We notice that the performance of direct optimization without a learning-based decoder is not satisfied.

Results of Test 2 (seen/unseen categories test): As shown in Table 2, the quantitative results indicate that our model achieves superior generalization ability on unseen categories as an unsupervised method. For reporting this performance, we ignore the categories of exact symmetric shapes' quantitative results for rotation matrix since in theory there is no unique solution for these cases as explained in the previous part for Test 1. In comparison, all the supervised learning methods experienced a dramatic performance drop compared to the results in Table 1. For example, we can see that PointNetLK and DCPv2+SVD obtain an MSE(R) of 227.87 and 1.31 in "the training/testing split test" as described in the previous experiment (see Table 1). However, the

corresponding values in "seen/unseen categories test" as described in this section increase to 306.32 and 9.92 respectively (see Table 2). The MSE(R) of PointNetLK increased from 227.87 for unseen point clouds to 306.324 for unseen categories. Unsupervised algorithms, such as ICP and FGR, achieve similar accuracy for unseen categories and unseen point clouds. Our method has a small performance drop for the unseen categories compared to the results for unseen point clouds. Particularly, in the prediction of the rotation matrix for unseen categories, our method outperforms state-of-the-art DCPv2-SVD by a large margin (3400% improvement) in MSE(R).

4.4 Partial 3D point set registration

In this experiment, we further verify the performance of our model for registering partial shapes.

Experiment Setting: For this experiment, we use four categories (chair, human, guitar, and airplane) from the ModelNet40 dataset to compare the performance of our model with state-of-the-art methods. In this section, we individually adjust and test our model for each category by controlling the learning rate with the threshold σ_t described in adaptive chamfer loss. For a fair comparison, we follow the settings of PR-Net paper to keep consistency with the previous methods as explained in the setting part of the experiment section. We compare our model with DCP and PR-Net in this section. Furthermore, we show the registration time for registering 100 pairs of 3D shapes between classical iterative method ICP, learning-based methods PR-net and DCP, and our hybrid method. We run DCP, PR-Net, and our model using a single 12-GB Tesla K80 GPU and ICP using CPU.

Results: We list the quantitative experimental results in Table 3. As we can see from the results, our method achieves significantly better performance than PR-Net and DCP models regarding the results of rotation angle and translation prediction for all four categories. Since we individually adjust our model for each category, our model can achieve superior performance in comparison to DCP and PR-Net. Regarding the running time, as shown in Table 4, for aligning 100 shapes from the testing dataset, our model spent 66 seconds. We do sacrifice more time than DCP and PR-Net which only require a single forward step for testing, but our model requires much less time than classical iterative methods such as ICP. More randomly selected qualitative results of our model are demonstrated in Figure 4.

Model	MSE(R)	RMSE(R)	MAE(R)	MSE(t)	RMSE(t)	MAE(t)
DCP [25]	17.078770	4.132647	3.095657	0.001602	0.040024	0.029227
PR-Net [13]	9.120225	3.019970	1.371537	0.000286	0.016917	0.011078
Ours	0.001736	0.041205	0.030796	0.00000003	0.000174	0.000134
DCP [25]	26.890577	5.185613	3.817362	0.001259	0.035477	0.026415
PR-Net [13]	9.430604	3.070928	1.388999	0.000296	0.017214	0.011231
Ours	0.011774	0.107868	0.064698	0.000056	0.007541	0.002581
DCP [25]	26.444530	5.142425	3.424549	0.003018	0.054939	0.039508
PR-Net [13]	15.00834	3.874060	1.41152	0.000296	0.017232	0.011141
Ours	0.021988	0.143298	0.086875	0.000021	0.004583	0.001227
DCP [25]	34.579647	5.880446	4.426307	0.001672	0.040888	0.031120
PR-Net [13]	8.569474	2.927366	1.368731	0.000291	0.017074	0.011149
Ours	0.026013	0.157938	0.075724	0.000029	0.005431	0.001823
			TABLE 3			

Testing performance on partial shapes alignment. From top to bottom: test performance on the chair, airplane, human, guitar category respectively.

Models	ICP	DCP	PR-Net	Ours
Time	571s	4s	5s	66s

TABLE 4

Running time for aligning 100 pairs of 3D point clouds from test dataset.

5 CONCLUSION

This paper introduces a novel approach that integrates a learning-based decoder with one optimizable descriptor to our research community for point set registration. With the newly proposed adaptive chamfer distance, our model can be perfectly applied for aligning partial shapes. We conducted experiments on the ModelNet40 datasets to validate the performance of our method. The results demonstrated that our proposed approach achieved competitive advantages regarding alignment accuracy but sacrifices acceptable computation time in comparison to state-of-the-art approaches.

REFERENCES

- [1] L. Ding and C. Feng, "Deepmapping: Unsupervised map estimation from multiple point clouds," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2019.
- [2] X. Bai, L. J. Latecki, and W.-Y. Liu, "Skeleton pruning by contour partitioning with discrete curve evolution," <u>IEEE transactions on pattern</u> analysis and machine intelligence, vol. 29, no. 3, 2007.
- [3] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," in Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, vol. 3. IEEE, 2006, pp. 15–18.
- [4] J. A. Maintz and M. A. Viergever, "A survey of medical image registration," <u>Medical image analysis</u>, vol. 2, no. 1, pp. 1–36, 1998.
- [5] J. Chen, L. Wang, X. Li, and Y. Fang, "Arbicon-net: Arbitrary continuous geometric transformation networks for image registration," in <u>Advances</u> in Neural Information Processing Systems, 2019, pp. 3410–3420. 1
- [6] A. Myronenko, X. Song, and M. A. Carreira-Perpinán, "Non-rigid point set registration: Coherent point drift," in <u>Advances in Neural Information</u> <u>Processing Systems</u>, 2007, pp. 1009–1016. 1, 2
- [7] J. Ma, J. Zhao, J. Tian, Z. Tu, and A. L. Yuille, "Robust estimation of nonrigid transformation for point set registration," in <u>Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition</u>, 2013, pp. 2147–2154.
- [8] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus." <u>IEEE Trans. image processing</u>, vol. 23, no. 4, pp. 1706–1721, 2014.
- [9] H. Ling and D. W. Jacobs, "Deformation invariant image matching," in Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, vol. 2. IEEE, 2005, pp. 1466–1473.
- [10] L. Wang, J. Chen, X. Li, and Y. Fang, "Non-rigid point set registration networks," arXiv preprint arXiv:1904.01428, 2019.

- [11] Y. Wang and J. M. Solomon, "Deep closest point: Learning representations for point cloud registration," <u>arXiv preprint arXiv:1905.03304</u>, 2019. 1, 2, 5
- [12] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," <u>ACM</u> Transactions on Graphics (TOG), vol. 38, no. 5, pp. 1–12, 2019.
- [13] Y. Wang and J. M. Solomon, "Prnet: Self-supervised learning for partial-to-partial registration," arXiv preprint arXiv:1910.12240, 2019. 1, 2, 7
- [14] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "An unsupervised learning model for deformable medical image registration," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9252–9260. 1, 2
- [15] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in Sensor Fusion IV: Control Paradigms and Data Structures, vol. 1611. International Society for Optics and Photonics, 1992, pp. 586–607. 2, 5
- [16] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-icp: A globally optimal solution to 3d icp point-set registration," <u>IEEE transactions on pattern analysis and machine intelligence</u>, vol. 38, no. 11, pp. 2241–2254, 2015.
- [17] Q.-Y. Zhou, J. Park, and V. Koltun, "Fast global registration," in <u>European</u> Conference on Computer Vision. Springer, 2016, pp. 766–782. 2, 5
- [18] H. Chui and A. Rangarajan, "A new algorithm for non-rigid point matching," in Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on, vol. 2. IEEE, 2000, pp. 44–51.
- [19] A. Zeng, S. Song, M. Nießner, M. Fisher, J. Xiao, and T. Funkhouser, "3dmatch: Learning local geometric descriptors from rgb-d reconstructions," in Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on. IEEE, 2017, pp. 199–208.
- [20] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," Proc. Computer Vision and Pattern Recognition (CVPR), IEEE, vol. 1, no. 2, p. 4, 2017.
- [21] Y. Aoki, H. Goforth, R. A. Srivatsan, and S. Lucey, "Pointnetlk: Robust & efficient point cloud registration using pointnet," in <u>Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition</u>, 2019, pp. 7163–7172. 2, 5
- [22] X. Liu, C. R. Qi, and L. J. Guibas, "Flownet3d: Learning scene flow in 3d point clouds," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 529–537.
- [23] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," <u>arXiv preprint arXiv:1505.00853</u>, 2015. 4
- [24] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in <u>International</u> <u>Conference on Machine Learning</u>, 2015, pp. 448–456. 4
- [25] S. Wang, S. Suo, W.-C. M. A. Pokrovsky, and R. Urtasun, "Deep parametric continuous convolutional neural networks," in <u>Proceedings</u> of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 2589–2597. 7