

GP-Aligner: Unsupervised Groupwise Nonrigid Point Set Registration Based on Optimizable Group Latent Descriptor

Lingjing Wang[✉], Nan Zhou[✉], Hao Huang, Jifei Wang, Xiang Li[✉], Member, IEEE and Yi Fang[✉], Member, IEEE

Abstract—In this article, we propose a novel unsupervised method named GP-Aligner to address the problem of groupwise nonrigid point set registration. Compared to previous nonlearning-based approaches, the proposed method gains competitive advantages by leveraging deep neural networks to effectively and efficiently align a large number of highly deformed 3-D shapes with superior performance. Unlike most learning-based methods that use an explicit feature encoding network to extract per-shape features and their correlations, our model leverages a model-free learnable latent descriptor to characterize shape correlations among groups. More specifically, for a given group, we first define an optimizable group latent descriptor (GLD) to characterize the relationship among a group of point sets. Each GLD is randomly initialized from a Gaussian distribution and then concatenated with the coordinates of each point of the associated point sets in the group. A neural network-based decoder network is further constructed to predict the coherent flow fields to optimally deform the input groups of shapes to the aligned ones. During the optimization process, GP-Aligner jointly updates all GLDs and weight parameters of the decoder network toward the minimization of an unsupervised groupwise alignment loss. After optimization, for each group, our model coherently drives each point set toward a mean position (shape) without specifying one as the target. GP-Aligner does not require large-scale training data for network training, and it can directly align groups of point sets in a one-stage optimization process. GP-Aligner shows both accuracy and computational efficiency improvement in comparison with the state-of-the-art methods for groupwise point set registration. Moreover, GP-Aligner exhibits high efficiency in aligning a large number of groups of real-world 3-D shapes.

Index Terms—Groupwise registration, latent space optimization, nonrigid registration, point set registration.

I. INTRODUCTION

POINT set registration is a fundamental computer vision task, which plays an important role in many applications, such as autonomous driving, pose estimation, robotic manipulation, and scene reconstruction, to name a few [1], [2], [3].

Manuscript received 4 August 2022; revised 23 October 2022; accepted 15 November 2022. Date of publication 20 December 2022; date of current version 29 December 2022. (*Lingjing Wang and Nan Zhou contributed equally to this work.*) (*Corresponding author: Xiang Li.*)

Lingjing Wang and Jifei Wang are with CANAL GLOBAL Pte. Ltd., Singapore 049422.

Nan Zhou is with the Department of Ocean College, Jiangsu University of Science and Technology, Zhenjiang 212100, China.

Hao Huang, Xiang Li, and Yi Fang are with the Department of Electrical and Computer Engineering, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates (e-mail: xiangli92@ieee.org).

Digital Object Identifier 10.1109/TGRS.2022.3230413

Before the era of deep learning, traditional nonlearning-based methods usually search for the geometric transformation by minimizing a predefined alignment loss in an iterative process [4], [5], [6], [7]. These methods can be roughly divided into two branches: feature-based methods and intensity-based methods. Feature-based methods usually leverage handcrafted features to match points in source and target point sets. For example, one of the most popular methods, iterative closest point (ICP) [8], estimates the rigid transformation by finding a set of corresponding points. Other methods formulate a probability density distribution function from one point set and fit the other point set to this distribution to maximize the accumulative density likelihood defined as a similarity metric, e.g., CPD [9]. Intensity-based methods directly optimize the transformation matrix by minimizing a predefined metric between intensity patterns of source and target shapes. Methods falling into this category are widely used in medical images/voxels registration [10]. In recent years, as deep learning-based methods have achieved great success in various visual recognition tasks, researchers are increasingly interested in bringing deep learning-based solutions to the field of point set registration. Thanks to the powerful feature learning/representation abilities, learning-based methods have demonstrated high potential in this field [11], [12], [13], [14].

In this article, we focus on the task of aligning a group of point sets toward one common/mean shape (position). In many real-world applications, such as medical images analysis [15] and 3-D mapping [16], [17], there is an urgent need for developing methods for registering a group of point sets/images. However, direct extension from pair-wise registration to groupwise registration is nontrivial. Recently, Che et al. [15] proposed a learning-based algorithm for the groupwise registration of multispectral fundus images. They enhanced the model proposed in [12] by dynamically computing the templates from a group of transformed input voxel images and comparing each input voxel image in the group with the templates using a predefined loss [12]. Some unsolved problems, such as how to model the relationship among a group of point sets, a proper definition of the groupwise similarity measure, and how to choose a template of the group,¹ are usually not considered in pair-wise registration

¹In practice, the assumption of the existence of an actual template to match against is hard to make.

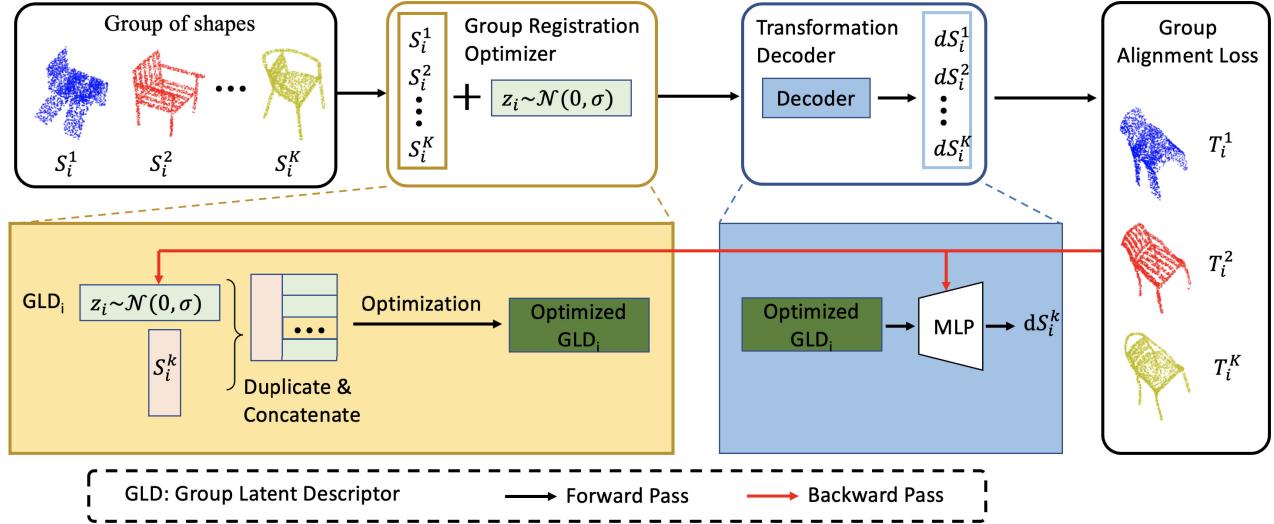


Fig. 1. Pipeline of the proposed method for groupwise point set registration. We show the registration of a single group of shapes $\{S_i^k \in \mathcal{S}_i\}_{k=1}^K$ in this figure for conciseness. Our method contains three main components. The first component is a group registration optimizer where the GLDs are optimized from a set of randomly initialized vectors. The second component is a transformation decoder that decodes the GLDs to the desired transformation for each group. The third component is a groupwise alignment loss that measures the similarity among transformed shapes in each group. The communication route in red represents back-propagation with which the groupwise alignment loss is back-propagated to update the GLDs and the transformation decoder.

task and make the groupwise registration problem even more challenging.

In comparison to the pair-wise point sets registration task, only a few recent researches [18], [19], [20] have studied the problem of groupwise point set registration using classical nonlearning-based methods. In general, these methods fall into two categories. The first category of methods works by selecting a corresponding subset of input point sets as a reference to compute the desired transformations [21], [22]. These methods are less robust and usually have difficulties in dealing with various noises and outliers. The second category of methods [18], [19], [20] directly registers the point sets from the population without detecting correspondence. For these methods, the main challenge is to define an efficient groupwise similarity-based cost function. Chen et al. [19] introduced CDF-HC divergence to quantify the similarity among all cumulative distribution functions (CDFs) estimated from input point sets. Recently, Giraldo et al. [18] provided a closed-form solution to optimize the previously defined cost function proposed in [19] by using Renyi's second-order entropy and improved optimization efficiency. Nevertheless, these methods suffer from unsatisfying performance due to the hand-engineering features involved.

In comparison to classical optimization-based methods, deep learning-based methods have been proved to be more robust for feature learning and enjoy higher efficiency due to powerful GPU parallel computation. In addition to the challenges listed above, learning-based methods usually require a well-labeled dataset for supervised training, but labeled datasets are usually unavailable for groupwise registration tasks. In contrast to the existing methods, we propose a purely unsupervised deep learning-based method for the task of groupwise point set registration. Unlike previous

learning-based methods that primarily fix the network at runtime, we propose to perform runtime optimization following recent works [17], [23]. In this way, our method can not only leverage the power of the neural network structure but also perform runtime optimization for better registration. Moreover, our algorithm does not need a predefined mean shape but can atomically search for the optimal one during the optimization process and aligns all the input groups with high efficiency.

As shown in Fig. 1, our proposed GP-Aligner framework contains three components. In the first component, we introduce an encoder-free structure to learn the geometric feature and groupwise relationship among a group of point sets. To achieve this, a set of optimizable group latent descriptors (GLDs) is randomly initialized from a Gaussian distribution and optimized along with the network parameters toward the minimization of a predefined alignment loss. In the second component, given each input shape and the associated optimized GLD, an MLP-based transformation decoder further decodes the GLD to coherent drift fields as the desired transformations from each of the input shapes among the group to the target position. In the third component, we formulate a groupwise alignment loss to measure the similarity among all the transformed point sets.

Our contributions are listed as follows.

- 1) We introduce a novel unsupervised method for the task of groupwise point set registration. Our method can not only leverage the power of the neural network structure but also perform runtime optimization for better registration.
- 2) We introduce an encoder-free framework to learn GLDs that characterize the geometric features and groupwise relationships among a group of point sets.

- 3) Experimental results demonstrate the effectiveness and high efficiency of the proposed method for groupwise point set registration, especially for the registration of a large number of groups simultaneously.

II. RELATED WORKS

A. Pairwise Registration

Existing point set registration methods mostly focus on the pairwise setting. Given a pair of point sets, a model is designed to find the geometric transformation to align the source point set with the target one. Before the prevalence of deep learning, classical nonlearning-based methods work by iteratively searching for a set of optimal transformation parameters by minimization of a predefined alignment loss. Early efforts mostly focus on rigid point set registration. For example, the ICP algorithm [8] starts with an initial estimation of the optimal rigid transformation and refines the transformation through an iterative process: 1) determines point correspondence based on the estimated transformation and 2) computes a new transformation based on point correspondences. Nevertheless, the performance of the ICP algorithm is sensitive to the initial guess of the desired transformation. Gelfand et al. [24] enhanced the ICP algorithm by developing a robust point set registration method that can be used without any assumptions regarding the initial positions. Go-ICP [25] solved this issue by applying a branch-and-bound (BnB) searching strategy over an entire 3-D motion space.

Another family of methods focuses on registering point sets under nonrigid transformations, such as morphing and articulation. In general, these methods can be divided into parametric and nonparametric categories based on the transformation representation. Chui and Rangarajan [26] introduced a parametric method TPS-RSM for nonrigid point set registration that applies thin-plate spline for the nonrigid mapping. As a pioneering nonparametric method, coherence point drift (CPD) [4] proposed to fix Gaussian mixture models (GMMs) to align source and target point sets. Ma et al. [5] introduced a flexible probabilistic method for nonrigid point registration, which is robust and outlier-adaptive. Li et al. [27] introduced a robust registration method for cross-source (multisource) point clouds with scale changes. To avoid missing optimal hypothesis in the random sample consensus (RANSAC), Quan and Yang [28] introduced a compatibility-guided sampling strategy to eliminate randomness during sampling.

In light of the great success of deep learning-based methods on various vision tasks, recent works have started the trend of directly learning geometric transformations from raw point clouds. The pioneering work, PointNet [29], introduced a simple and effective method for direct feature learning on raw point clouds. In recent years, PointNet and its variants demonstrate remarkable performance for 3-D object classification, object detection, part segmentation, semantic segmentation, and shape correspondences [30], [31], [32], [33], [34], [35], [36]. PointNetLK [37] proposed to use a shared PointNet to extract global shape representation and developed a recurrent deep neural network to predict geometric transformations based on the Lucas & Kanade algorithm.

Wang and Solomon [3] introduced deep closest point (DCP) for rigid point set registration. In the DCP algorithm, a DGCNN network is adopted for per-point feature learning, followed by an attention-based feature matching module to determine point correspondences. PR-Net [38] and JCRNet [39] introduced self-supervised methods for partial point set registration. RPM-Net [40] presented a robust method for rigid point set registration, which is less sensitive to initialization. Sun et al. [41] introduced a graph deep learning framework that fuses the multilayer perceptron (MLP) and the graph convolutional network (GCN) for better keypoint detection and point descriptor learning. Yang et al. [42] introduced a loose-tight geometric voting technique to dynamically score 3-D feature correspondences with better correspondence selection.

For the problem of nonrigid point set registration, Liu et al. [11] introduced a supervised method that can directly estimate the scene flow from a pair of consecutive point clouds in one network forward pass. A flow embedding layer was proposed to learn the feature correlations among two point clouds, which were further used to generate dense flow fields through newly designed set convolutional layers. Wang et al. [43], [44] proposed an unsupervised network for nonrigid point set registration where the Chamfer loss was used as the supervision signal. Unlike previous methods that use an explicit point feature encoding network to learn per-point features, in contrast, we represent the geometric features and relationship of multiple 3-D shapes using an optimizable latent code, which is optimized simultaneously with network parameters during the back-propagation process.

B. Groupwise Registration

Groupwise point set registration aims to register multiple point sets. Although pairwise registration can be applied repeatedly to solve this problem, it achieves inferior performance due to error propagation [45], [46]. To address this issue, researchers have developed numerous methods to register multiple point sets simultaneously without pairwise registration. These methods can be divided into two categories: information theoretic-based methods and probability-based methods. A comprehensive review can be found in [47]. In information theoretic-based methods, some information-theoretic metrics are defined to optimize the distribution of multiple point sets. For example, Wang et al. [20] developed a groupwise point set registration method based on CDF Jensen–Shannon (CDF-JS) divergence. A cost function based on CDF-JS divergence was defined and minimized toward the optimal registration parameters. Chen et al. [19] further enhanced the CDF-JS method by using the Havrda–Charvát (HC) divergence instead of JS for the CDFs to quantify the dissimilarity between estimated CDFs from individual point sets. To further improve the efficiency of the divergence algorithm, Giraldo et al. [18] derived a closed-form formula for the analytic gradient for CDF-HC divergence, which is more efficient compared with the previous algorithms.

In probability-based methods, multiple point sets are assumed to be drawn from predefined probability functions, and these methods treat groupwise point set registration as a

clustering problem. Generally, these methods consist of two separate steps: 1) the construction of a mean shape/point set and 2) the estimation of the transformations from multiple point sets to the predefined mean shape. These two steps are iteratively performed to achieve simultaneous registration of multiple point sets. In [46] and [47], multiple point sets were assumed to be sampled from a GMM, and an EM algorithm was applied to perform clustering and registration. Abtin et al. [48] proposed a groupwise registration technique that leverages soft correspondences between groups of point sets. By learning the shape variations of a group, the learned model of a mean shape can be further aligned toward all shapes among the group after transformation. Chui et al. [22] leveraged a joint clustering and matching algorithm for the computation of a mean shape from multiple unlabelled shapes. Then, all the shapes of the group were forced to move toward the mean shape as an iterative bootstrap process.

Unlike the abovementioned methods that mostly use hand-crafted features for feature matching and require a separate optimization process for each group of point sets, our proposed takes advantage of the powerful feature learning abilities of deep neural networks to automatically extract high-level point/shape features. More importantly, the proposed method can be trained simultaneously for multiple groups of shapes to enable knowledge transfer from registering one group to another.

C. Latent Space Optimization

Although existing deep learning-based methods have achieved ever-increasing performance on shape representation learning using explicit encoder networks [29], [30], [35], [49], [50], it is still a challenging problem for these methods to learn robust shape representation from irregular raw point clouds. An alternative method has been proposed in [51], which introduced a latent code to represent the global feature representation of input data and simultaneously optimize the latent code and decoder network through back-propagation. During inference, the decoder network is fixed, and the latent code is optimized to reconstruct new observations. Similar ideas have been exploited recently for various tasks, such as noise reduction, missing measurement completion, and shape correspondence [52], [53], [54]. For example, Groueix et al. [54] proposed to optimize a latent vector in the inference stage by minimizing an unsupervised loss to enhance the performance of shape correspondence. Park et al. [55] proposed to search for the optimal latent representations for shape reconstruction using a neural implicit representation. In this work, the proposed GLD is inspired by the latent space optimization technique.

III. METHODS

We describe our approach in the following sections. First, we define the groupwise registration problem in Section III-A. In Section III-B, we introduce our newly proposed group registration optimizer. Section III-C illustrates the coherent flow field decoding process. In Section III-D, we provide the definition of our proposed groupwise similarity function. The optimization process is illustrated in Section III-E.

A. Problem Statement

For the groupwise registration task, given a dataset \mathcal{D} of M groups $\{\mathcal{S}_i\}_{i=1,2,\dots,M}$ and for each group, we have K point sets (K can be different for different groups). In this study, we mainly focus on nonrigid deformable point set registration and assume that point sets from each group are prealigned in a canonical frame. Note that, for unaligned point sets, a global registration technique, e.g., ICP [8], can be first applied to compensate for the rigid transformations as a preprocessing step. From the experiments in the following, our model can also be applied to accomplish the rigid registration process. For each group of shapes \mathcal{S}_i , our groupwise registration method aims to find the flow field from each shape from \mathcal{S}_i toward an unknown mean shape $\bar{\mathcal{S}}_i$. A direct solution can be reached by performing pairwise registration between each shape from \mathcal{S}_i and the mean shape $\bar{\mathcal{S}}_i$. However, this requires the predefined mean shape $\bar{\mathcal{S}}_i$. In contrast, our method does not need a predefined mean shape but characterizes its geometric essence by an optimizable GLD, which is atomically searched during the optimization process.

Our method learns the flow fields through an MLP-based parametric function g_θ with parameters θ . $\forall S_i^k \in \mathcal{S}_i$, where $k \in \{1, 2, \dots, K\}$, the desired geometric transformation can be formulated as $T_i^k = g_\theta(S_i^k, z_i)$, where $z_i \sim \mathcal{N}(0, 0.01)$ is an optimizable GLD associated with group \mathcal{S}_i , such that the transformed point sets $\{T_i^k(S_i^k) | S_i^k \in \mathcal{S}_i\}_{k=1,2,\dots,K}$ can reach the minimal groupwise registration loss for all groups $\{\mathcal{S}_i\}_{i=1,\dots,M}$. The geometric transformation in this article is represented by the coordinate drifts from each point in S_i^k to the corresponding location in $T_i^k(S_i^k)$. A gradient descent-based algorithm is used to update the weights parameters θ and latent code $\{z_i\}_{i=1,\dots,M}$ toward the minimization of a predefined loss

$$\begin{aligned} & \tilde{\theta}, \tilde{z}_1, \dots, \tilde{z}_M \\ &= \arg \min_{\theta, z_1, \dots, z_M} \left[\mathbb{E}_{\{\mathcal{S}_i\} \sim \mathcal{D}} \left[\mathbb{E}_{\{S_i^k\} \sim \mathcal{S}_i} \right. \right. \\ & \quad \left. \left. \left[\mathcal{L}(g_\theta(S_1^{(1)}, z_1), \dots, g_\theta(S_i^K, z_i)) \right] \right] \right] \end{aligned} \quad (1)$$

where \mathcal{L} represents a groupwise registration loss function, \tilde{z}_i is the optimized latent descriptor for group \mathcal{S}_i , and $\tilde{\theta}$ is the optimized network parameters trained on all groups. After the optimization process, the transformation $\{\tilde{T}_i^k = g_{\tilde{\theta}}(S_i^k, \tilde{z}_i)\}_{k=1,\dots,K}$ can be regarded as the optimal solution for aligning group \mathcal{S}_i . During the optimization process, we assign different weights to the penalization term on the smoothness and scale of the predicted drifts, and the alignment loss term in function \mathcal{L} .

B. Group Registration Optimizer

For each input group, we define an optimizable GLD to characterize the group geometric feature. This latent descriptor is designed to characterize the relationships among input shapes and guide the transformation prediction process. Unlike volumetric shapes that are regularized in ordered standard voxels, point sets contain sparse geometric coordinates that are not arranged regularly. Even though previous works (e.g., PointNet [29]) have provided techniques to extract shape features from unstructured point sets and recent research

(e.g., [13]) attempts to formulate the correlation between pairwise shapes, it is still challenging to design an effective feature encoder for the registration task with multiple shapes in a group, especially when the target mean shape of the group is not even known. Furthermore, commonly used encoder networks are rather data-/task-specific and inept in dealing with data of large variations.

To eliminate the downside effect of the explicit design of a feature encoding network and correlation module, we introduce an encoder-free structure to learn the geometric features and groupwise relationships among a group of point sets. To achieve this, a set of optimizable GLDs is defined to represent the group features for the groupwise point set registration task. Each GLD is randomly initialized from a zero-mean Gaussian distribution and then concatenated to each point of input shape to guide the transformation prediction process. The GLD is initialized independently for each group in the dataset. Though GLD is not explicitly learned from input shapes, we design it to be optimizable from the groupwise alignment loss together with the decoder network parameters. In this way, our method can not only avoid the challenging problem of explicitly defining the relationships among a group of point sets but also leverage the deep neural network-based structure in the optimization process.

C. Transformation Decoder

We define the geometric transformation T as the coordinate drifts from each point $s \in S_i^k$ to the corresponding location in $\mathcal{T}_k(S_i^k)$

$$T(s) = s + g(s) \quad (2)$$

where $g : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is a “point displacement” function. The groupwise point set registration task requires determining this displacement function such that all point sets in the population can be coherently moved toward a mean point set. To ensure a good groupwise alignment, one would expect that a model can align all point sets in the group to the same mean shape. Besides, according to motion coherent theory (MCT) [56], the function g is required to produce continuous and smooth flow fields. In this work, our deep neural network-based decoder is well designed such that the function g parameterized by our proposed MLP network can not only align the input pair of point sets but also produce continuous and smooth flow fields. Please refer [14] for further explanation.

For registration of a number of groups, given the above-defined GLD for each group, the input of the decoder is formulated by stacking the coordinates of each point s on individual shape S_i^k with the GLD, noted as $[s, z_i], \forall s \in S_i^k$. Then, we define an MLP network with J layers (with the ReLU activation function), $\{g_j\}_{j=1,2,\dots,J}$, to learn coherent drift flows such that $g_j : \mathbb{R}^{v_j} \rightarrow \mathbb{R}^{v_{j+1}}$, where v_j and v_{j+1} are the dimensions of the layer inputs and outputs, respectively, calculated as

$$dS_i^k = g_J g_{J-1}, \dots, g_1([S_i^k, z_i]) \quad (3)$$

$$\mathcal{T}_i^k(S_i^k) = S_i^k + dS_i^k \quad (4)$$

where $\mathcal{T}_i^k(\cdot)$ denotes the flow field as the transformation function for each shape $S_i^k \in \mathcal{S}_i$. After generating all the

transformed point sets from the group, we further define a groupwise alignment loss to optimize the GLDs and the decoder network parameters as described in Section III-D.

D. Loss Function

In our unsupervised setting, we do not have the ground-truth transformation for supervision, and we do not assume correspondences between any pair of point sets. Therefore, a distance metric between two point sets, instead of the pointwise loss, is preferred. In addition, a suitable metric should be differentiable and efficient to compute. In this article, we adopt the Chamfer distance (C.D.) proposed in [57] as our loss function. The Chamfer loss is a simple but effective alignment metric defined on two noncorresponding point sets. We formulate the groupwise Chamfer loss between all transformed point sets as

$$\mathcal{L}_{\text{Group}} = \frac{1}{M} \sum_{i=1}^M \sum_{m,n \in [1,K], m \neq n} \mathcal{L}_{\text{cd}}(\mathcal{T}_i^m(S_i^m), \mathcal{T}_i^n(S_i^n)) \quad (5)$$

where the Chamfer loss between each pair of shapes is formulated as

$$\mathcal{L}_{\text{cd}}(X, Y) = \sum_{x \in X} \min_{y \in Y} \|x - y\|_2^2 + \sum_{y \in Y} \min_{x \in X} \|x - y\|_2^2. \quad (6)$$

We also add a regularization term to regularize the scale of drifts (transformation), and our final regularized groupwise loss is calculated as

$$\mathcal{L}_{\text{all}} = \mathcal{L}_{\text{Group}} + \lambda \sum_{i=1}^M \sum_{k=1}^K \|dS_i^k\|_1 \quad (7)$$

where $\|\cdot\|_1$ denotes the L1 norm and λ is a balance hyperparameter between the alignment loss and the deformation level. This hyperparameter can be fixed or dynamically chosen during the optimization process. By default, λ is set to 0.1 in our experiments.

E. Optimization Strategy

Algorithm 1 gives an illustration of the proposed method for groupwise point set registration. Note that steps 3–8 in Algorithm 1 can be computed using GPUs. In this work, we run experiments on a single Tesla P100 GPU. The learning rate decays from 0.001 to 0.0001 for the decoder network and GLDs during the first 100 steps. The Adam optimizer is utilized for model optimization. We balance the alignment loss with the geometric distance of the drifts. Otherwise, a trivial solution that aligns all shapes toward a single point can be easily reached. Our model requires 100–500 optimization steps until convergence. In this article, we choose latent vector z_i with a dimension of 256 for each group \mathcal{S}_i . For the decoder network described in (3), we use an MLP with dimensions of (258, 128, 64, 2/3) to regress the input $(\{S_i, z_i\}_{i=1,2,\dots,M})$ into a 2-D /3-D drifts flow fields.

IV. EXPERIMENTS

In this experimental section, we demonstrate the performance of our model for groupwise point set registration. In Section IV-A, we demonstrate the performance of our model

Algorithm 1 Groupwise Registration Optimization Process

```

1: Initialize the decoder parameters  $\theta$  and GLDs  $\{z_1, \dots, z_M\}$  and choose hyper-parameters.
2: while not convergence do
3:   for each group  $S_i \in \mathcal{D}$  do
4:     for each shape  $S_i^k \in S_i$  do
5:       Generate the flow field  $dS_i^k$  and transformed point set  $T_i^k(S_i^k, z_i)$  for each input shape by Eq. (3) and Eq. (4) in
         network forward pass.
6:   end for
7: end for
8: Compute the groupwise alignment loss for all transformed shapes  $\{\{T_i^k(S_i^k, z_i)\}_{k=1, \dots, K}\}_{i=1, \dots, M}$  by Eq. (7).
9: Compute the gradients  $d\theta$  and  $dz_i$  w.r.t the loss function  $\mathcal{L}$  and updates  $\theta$  and  $\{z_i\}_{i=1, \dots, M}$ .
10: end while

```

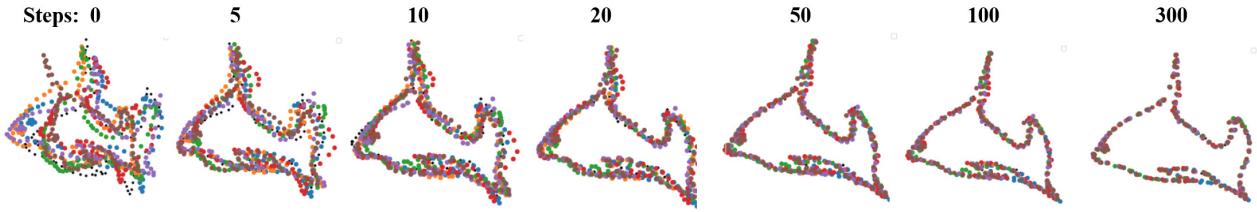


Fig. 2. Groupwise point set alignment process. The inputs include seven nonrigid deformed shapes, and the deformation level of inputs is 0.4.

in aligning a single group of 2-D synthesized point sets and 3-D real-world point sets, and we compare our model with the state-of-the-art methods. In Section IV-B, we test our model on groupwise registration of a large number of groups of 3-D real-world point sets.

A. Registration of Single Group of Shapes

1) Experiments on 2-D Synthesized Dataset: In this section, we demonstrate our model's groupwise point set registration performance on the 2-D synthesized dataset and compare the performance of our model with the previous state-of-the-art methods [18]. We conduct extensive experiments in this section to comprehensively evaluate the performance on both clean and noisy 2-D data.

a) Dataset: For synthesizing 2-D point sets, we simulate nonrigid geometric transformation on the raw point sets through thin plate spline (TPS) [58] transformation with a given deformation level. Given a deformation level l , we randomly perturb the controlling points by sampling noise from a Gaussian distribution $\mathcal{N}(0, 2l)$. When the deformation level is small, previous methods, e.g., [18], can produce good registration performance. In this work, we conduct experiments on more challenging settings with larger deformation levels.

For synthesizing 2-D noise data, three different types of noise are applied to input point sets, including outlier, missing points, and random jittering. To generate point outlier (P.O.) noise data, we add a certain ratio of outlier points with random coordinates to input point sets. The data incompleteness (D.I.) noise data are generated by removing certain parts of the input point points. To generate Gaussian displacement (G.D.) noise data, we perturb the point coordinates using zero-mean Gaussian noise. The generated noise point sets are shown in the first column of Fig. 5.

TABLE I
QUANTITATIVE TESTING PERFORMANCE AND TIME FOR RUNNING
500 STEPS FOR ALIGNING DIFFERENT NUMBERS OF
FISH-SHAPED POINT SETS

Number of Shapes	Groupwise C.D. ($\times 10^{-4}$)	Time
10	0.74	38s
20	0.88	66s
50	1.05	153s
100	0.37	310s

b) Settings: As explained in Section III-E, we choose the latent vector z with a dimension of 256. For the decoder, we use an MLP with dimensions of (258, 128, 64, 2) to regress a 2-D flow field.

First, we examine the alignment process for a group of input point sets by the optimization steps. The inputs include seven nonrigid deformed shapes, and the deformation level of inputs is 0.4. As shown in Fig. 2, our GP-Aligner model can successfully align the main parts of the input shapes after 100 steps. During the experiment, we notice that the convergence speed after 100 steps becomes relatively slower, which can be observed in Fig. 2 as well. All shapes are well-aligned after reaching 300 steps (convergence). We need to point out that the mean shape is different from any input shape and is automatically determined during the optimization process.

Furthermore, we demonstrate the performance of our model for registering a large number of nonrigid deformed shapes simultaneously. For this test, we set the deformation level to 0.2. We randomly generate a set of ten, 20, 50, and 100 shapes as inputs and allow a maximum of 500 steps for each optimization process. As shown in Table I, the quantitative results show that our model achieved comparable performance for aligning

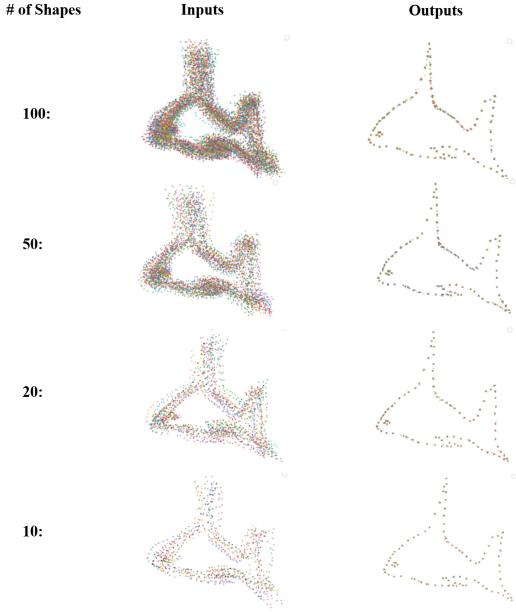


Fig. 3. Qualitative registration results of aligning different numbers of fish-shaped point sets. Inputs include ten to 100 point sets with various nonrigid deformations represented by different colors.

ten to 100 shapes. The registration performance of aligning 100 shapes is the best among these cases, as indicated by the groupwise C.D. Also, the computation time is longer for aligning 100 shapes (310 s) than ten shapes (38 s) as expected. In Fig. 3, we further notice that all the shapes are well-aligned for ten to 100 input shapes without significant differences.

Finally, we compare the performance of our model for groupwise registration with the state-of-the-art models on seven synthesized shapes with different deformation levels. We consecutively increase the deformation level from 0.2 to 0.6. We compare our model with the state-of-the-art models, including Holder [18], IP [18], and Norm-IP [18]. Norm-IP is the previous state-of-the-art model. For quantitative results, we use the groupwise C.D. defined in (5) as the evaluation metric. Furthermore, for comparison of algorithm efficiency between our model with the previous methods, we report the running time for all compared methods.

Fig. 4 shows the registration results of our method and the compared methods at different deformation levels. As shown in this figure, our method successfully aligns all groups of shapes at different deformation levels. When the input point sets are at a lower deformation level, the current state-of-the-art method Norm-IP demonstrates good performance for most shapes except the blue one. When the deformation level increases, the compared methods, including Holder, IP, and Norm-IP, fail to produce consistent registration outputs. In contrast, our model, by leveraging the power of neural networks, shows significantly better performance than previous STOA models.

From the quantitative results shown in Table II, our model demonstrates significantly better performance than all the compared methods under all deformation levels. The alignment C.D. distance of our model is an order of magnitude lower than the current state-of-the-art Norm-IP model. Specifically,

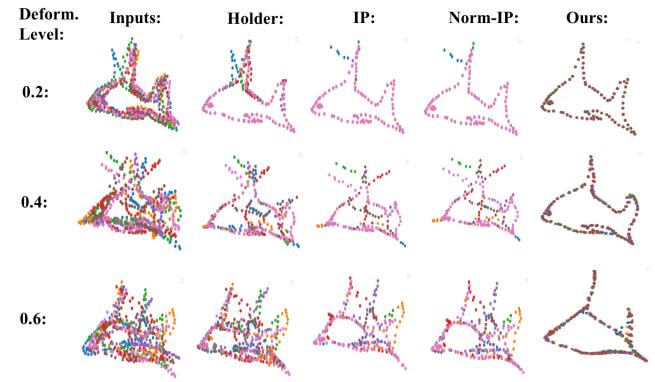


Fig. 4. Registration results for fish shapes at different deformation levels. Inputs include seven fish-shaped point sets with different colors at various deformation levels.

TABLE II
QUANTITATIVE TESTING PERFORMANCE ON FISH-SHAPED POINT SETS AT DIFFERENT DEFORMATION LEVELS. WE REPORT THE PERFORMANCE USING GROUPWISE C.D. ($\times 10^{-4}$)

Def. level	Holder	IP	Norm-IP	Ours
0.2	35.93	24.01	23.67	0.68
0.4	144.03	181.22	176.83	6.42
0.6	474.35	314.09	301.92	10.82

TABLE III
TIME FOR COMPLETING 1000 STEPS FOR GROUPWISE REGISTRATION OF SEVEN FISH-SHAPED POINT SETS AT A DEFORMATION LEVEL OF 0.4

Holder	IP	Norm-IP	Ours
88s	85s	78s	23s

under a deformation of 0.6, the Norm-IP model leads to a large alignment error, with a groupwise C.D. of 301.92×10^{-4} . Our GP-Aligner model achieves a groupwise C.D. of 10.82×10^{-4} , which is $30\times$ lower than the Norm-IP model. Moreover, regarding the algorithm efficiency, as shown in Table III, the state-of-the-art Norm-IP method requires 78 s to register a group of seven fish shapes. In contrast, our model only takes 23 s to accomplish the registration process, which is much faster than Norm-IP. When dealing with real-world 3-D point sets, efficiency improvement becomes more significant, which will be shown in Section IV-A2.

c) Effect of GLD: In this section, to further show the effect of the proposed GLDs, we compare our method with the following two variants. First, instead of using a groupwise latent code, we use a PointNet [29] as an encoder to learn the global feature l_i^k for each shape S_i^k . In the transformation decoder, we formulate the input by concatenating the coordinates of each point $s \in S_i^k$ with its corresponding shape code l_i^k , i.e., $[s, l_i^k]$, and feed the inputs to the transformation decoder to regress flow fields. We mark this method as “Ours w/ PointNet encoder.” For the second method, we use separate optimizable latent code g_i^k for each shape. Similarly, the inputs for the transformation decoder are formulated as $[s, g_i^k]$. We mark this method as “Ours w/ shape-wise GLD.” Experiments are carried out on seven nonrigid deformed shapes with a deformation level of 0.4 and optimized for 500 steps.

Table IV shows the registration performance with and without the proposed GLD. From Table IV, the two models with

TABLE IV

EFFECT OF GLDS. WE SHOW REGISTRATION PERFORMANCE WITH AND WITHOUT THE PROPOSED GLDS ON SEVEN SYNTHESIZED SHAPES AT A DEFORMATION LEVEL OF 0.4

Method	Groupwise C.D. ($\times 10^{-4}$)
Ours w/ PointNet encoder	9.80
Ours w/ shape-wise GLD	8.11
Ours w/ groupwise GLD	6.42

optimizable latent codes achieve significantly better performance than the baseline model using a PointNet encoder. This is probably because a simple PointNet encoder fails to learn a representative global feature of each shape. In contrast, by joint training the optimizable latent codes, our model can effectively learn the global features and correlations among a group of shapes. By comparing the last two rows, it is obvious that our model with groupwise latent codes achieves significantly better performance than the model using independent latent code for each shape. This is because groupwise latent codes can model the correlations among a group of shapes, while the shape-wise latent codes can only learn the shape-specific features.

d) Effect of initialization: In this section, we investigate the impact of different initialization on the registration performance of our model. For this test, the inputs include seven non-rigid deformed fish shapes, and the deformation level of inputs is set to 0.4. We run our experiments 50 times with different initializations and report the quantitative results. For aligning the same group of seven fish shapes, our model achieved an average groupwise C.D. of 6.08×10^{-4} for 50 different initializations, with a small standard deviation of 0.18×10^{-4} . The highest score of these observations is 7.43×10^{-4} , and the lowest score of these observations is 4.73×10^{-4} . Based on this observation, we conclude that different initializations have a small impact on the final registration performance of our model.

e) Effect of input noise: Moreover, we demonstrate the performance of our model on noisy data. We randomly select three point sets at deformation level 0.4, and we separately add P.O., D.I., and G.D. noise at a noise level of 0.4, 0.2, and 0.05, respectively, to the selected shapes to further evaluate the groupwise registration performance on the biased data. For this test, we use the previous state-of-the-art model Norm-IP as the baseline for comparison. In this experiment, we compare the performance of our model and Norm-IP on aligning a group of point sets in the presence of various noise patterns. In Fig. 5, the first row shows input point sets, and the second and third columns show outputs of Norm-IP and our GP-Aligner model, respectively. As shown in Fig. 5, under P.O. noise, we can see that the orange color fish is mistakenly aligned with the other two fishes for Norm-IP, but our model can successfully align them. When the input point sets are with outliers (P.O.), the Norm-IP model aligns the input shapes into two different patterns, while our GP-Aligner model can well align the input shapes into a mean shape. Moreover, when the input point sets are with D.I. or Gaussian noise (G.D.), the Norm-IP model fails to align the top part of input shapes, while our GA-Aligner well registers these point sets with different noise patterns.

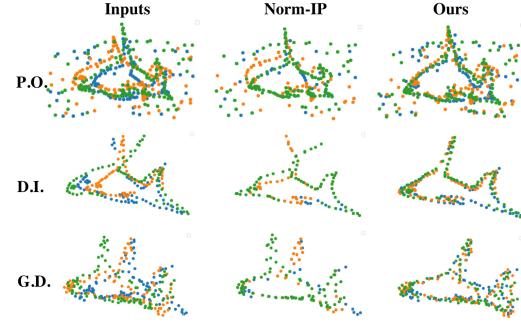


Fig. 5. Qualitative registration results of aligning fish-shaped point sets with different noise patterns. Inputs include three point sets under different nonrigid deformations.

f) Rigid registration: In this section, we show that our model can also be used for registering a group of point sets under rigid transformation. To achieve this goal, in the transformation decoder, instead of using MLP networks to predict 3-D flow fields, in this experiment, a max-pooling layer is applied to generate global features, and then, our model produces 6-D geometric transformation parameters, including three rotation angles and a translation vector. Chamfer loss is also applied for self-supervised training.

We conduct groupwise registration experiments on three fish shapes. For each fish shape, we follow DCP [3] to randomly apply a rigid transformation to generate seven target point sets. The rotations are uniformly sampled from 0° to 45° , and the translations are uniformly sampled in $[-0.5, 0.5]$.

Fig. 6 shows the groupwise registration results under rigid transformations. The first seven columns show independent shapes, and the eighth column shows a merged view. From this figure, our model can successfully align a group of shapes under different poses.

2) Experiments on 3-D Dataset: In this section, we demonstrate the effectiveness of our model for groupwise point set registration on a single group of 3-D real-world shapes.

a) Dataset: We conduct our experiment on the chair category of the ShapeNet [59] dataset. We randomly select three shapes from this category as the inputs, as shown in Fig. 7. Each shape contains 2048 uniformly sampled points.

b) Setting: As explained in Section III-E, we choose the latent vector z with a dimension of 256 for each group. For the decoder network, we use an MLP with dimensions of (259, 128, 64, 3) to generate 3-D flow fields.

First, we compare our model with Norm-IP [18] for groupwise point set registration. Section IV-A1, we use groupwise C.D. defined in (5) as the evaluation metric. The results are listed in Table V. We also report the running time to demonstrate the efficiency of our model. As shown in Table V, our model achieves a groupwise C.D. of 2.20×10^{-4} , which is an order of magnitude lower than the Norm-IP model that achieves a groupwise C.D. of 21.42×10^{-4} . By comparing the last two rows, our model with groupwise latent code achieves better performance than the one using independent latent codes. More importantly, as shown in the second column of Table V, Norm-IP requires 4328 s to run 100 steps, while our model only needs 16 s. This difference in model efficiency can be even larger when increasing the number of shapes or

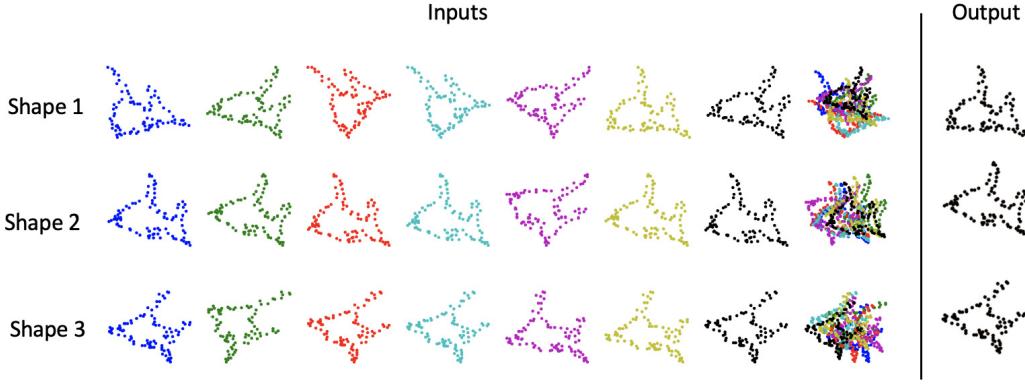


Fig. 6. Groupwise registration results for fish shapes under rigid transformations.



Fig. 7. Qualitative registration results for 3-D real-world point sets. Inputs include three shapes (in different colors) from the chair category of the ShapeNet [59] dataset.

TABLE V
QUANTITATIVE PERFORMANCE AND TIME FOR RUNNING 100 STEPS FOR 3-D POINT SET REGISTRATION. WE RUN NORM-IP FOR 100 STEPS AND GP-ALIGNER FOR 200 STEPS

Methods	Groupwise C.D. ($\times 10^{-4}$)	Time (100 steps)
Norm-IP	21.42	4328s
Ours w/ shape-wise GLD	3.10	16s
Ours w/ groupwise GLD	2.20	16s

the number of sampled points. Qualitative results are shown in Fig. 7. As shown in Fig. 7, our GP-Aligner successfully deforms the three input shapes into a single mean shape (last column). We note that the deformed shapes belong to one style, and the deformation is reasonable.

c) *Effect of different regularizations:* Furthermore, we analyze the impacts of the hyperparameter λ for regularization, as explained in III-D, on the final groupwise registration performance. In this experiment, we set λ to 0.01, 0.1, and 1.2. From Table VI, we note that the groupwise C.D. increases as λ increases from 0.1 to 1.2. To further check the registration performance, we show qualitative results of our model in Fig. 8. As shown in Fig. 8, with a small regularization (e.g., $\lambda = 0.01$), the topological structure of shapes can be dramatically deformed and lose their original semantic meanings. For example, all the legs of the chairs are missing after registration. However, the deformation field is still coherent without local distortion. With a large balance weight, e.g., $\lambda = 1.2$, the deformation freedom is limited, and we can see that the resulting groupwise C.D. is much higher

TABLE VI
QUANTITATIVE PERFORMANCE FOR 3-D POINT SET REGISTRATION WITH DIFFERENT REGULARIZATION WEIGHTS

λ	Groupwise C.D. ($\times 10^{-4}$)	Laplacian loss ($\times 10^{-2}$)
0.01	0.39	8.0
0.1	2.20	6.8
1.2	7.41	4.8

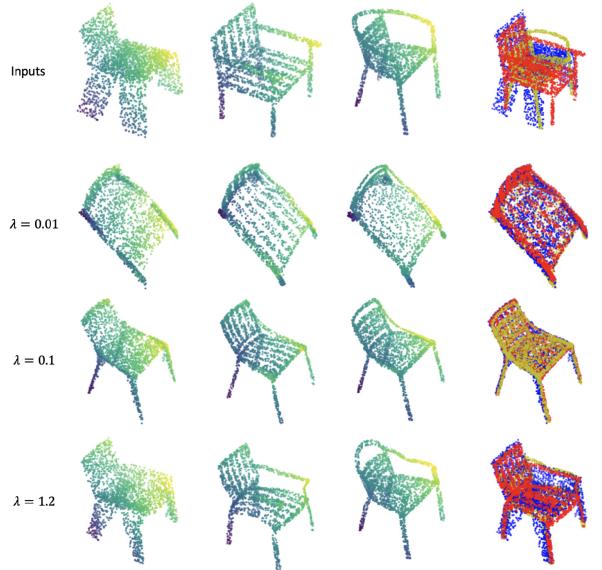


Fig. 8. Qualitative registration results for real-world 3-D point sets with different regularization weights. Inputs include three shapes (in different colors) from the chair category of the ShapeNet [59] dataset. In the first three columns, corresponding points are shown in the same colors. In the last column, different colors show points from different input shapes.

from Table VI in comparison to the other two cases. Therefore, there is always a tradeoff between the alignment loss and the regularization level. A higher regularization level indicates a better topological consistency between input and output shapes but a worse alignment among transformed shapes.

d) *Rigid registration:* We conduct groupwise registration experiments on three chair shapes from the ShapeNet dataset. For each chair, we follow DCP [3] to randomly apply a rigid transformation to generate seven target point sets. The rotations are uniformly sampled from 0° to 45° , and the translations are uniformly sampled in $[-0.5, 0.5]$.

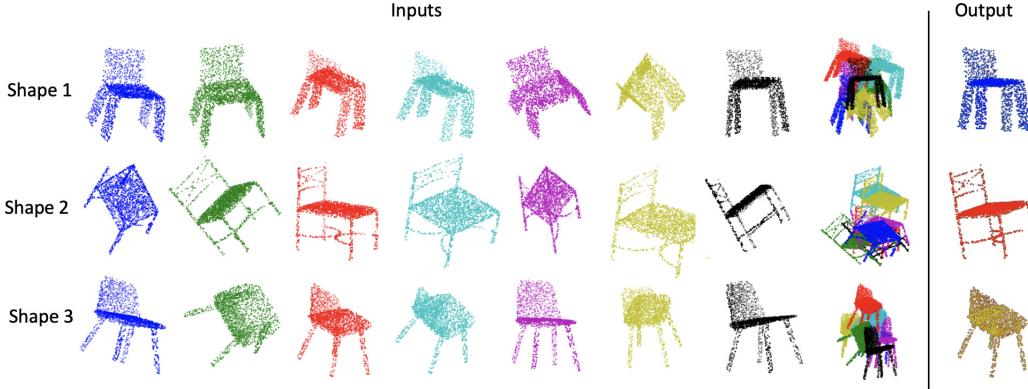


Fig. 9. Groupwise registration results for chair shapes under rigid transformations.

Fig. 9 shows the groupwise registration results under rigid transformations. The first seven columns show independent shapes, and the eighth column shows a merged view. From this figure, our model can successfully align a group of shapes under different poses.

e) Evaluation of shape correspondences: In the first three columns of Fig. 8, we visualize the shape correspondences before and after registration. Corresponding points are shown in the same colors. From this figure, our GP-Aligner model can not only align a group of shapes toward a meaningful mean shape but also maintain the correspondences between input shapes and deformed shapes. This is because our deformation decoder network uses continuous functions for flow field prediction. Moreover, during the optimization process, our model penalizes the flow field using (7) to ensure smoothness. Interested readers may refer to [14] for a detailed discussion on continuity and smoothness.

We also report the distance changes on the kNN graph on the point set during the registration process. We first define a Laplacian coordinate for each point p as $L_p = p - 1/(|\mathcal{N}(p)|) \sum_{q \in \mathcal{N}(p)} q$, where \mathcal{N} denotes the neighborhood of point p . We then calculate the Laplacian loss as $\frac{1}{N} \sum_p \|L_p - L'_p\|_2^2$, where L_p and L'_p denote the Laplacian coordinate before and after registration, $\|\cdot\|_2$ denotes the L2 norm, and N denotes the number of points. In our experiments, we search for the five nearest neighbors to calculate Laplacian loss. Results are listed in Table VI. As shown in the table, the Laplacian loss decreases as expected with the increase in the regularization coefficient. This demonstrates that the proposed regularization term can encourage topology consistency during the registration process. Moreover, the Laplacian losses under different regularizations are at a small magnitude.

f) Discussion: The deformation of a group of 3-D shapes toward a mean shape has a wide range of applications in practice. To align a group of real-world 3-D point sets as shown in this experiment, our method shows significantly superior performance in comparison to the state-of-the-art Norm-IP method. By leveraging the power of GPU computation and an efficient optimization algorithm, our method also shows significantly better performance as indicated by alignment loss and computation efficiency. Moreover, unlike previous

learning-based methods, the proposed network can be regarded as an optimization algorithm, which does not require ground-truth labels of flow fields for model training. The entire alignment process is completed during a one-stage optimization process. These characteristics make our method more applicable to solving real-world industrial problems. The reasonable deformation fields link a group of 3-D shapes together with reasonable correspondences. The groupwise registration result for real-world 3-D shapes can be further used for tasks such as groupwise segmentation and correspondence. Moreover, our method can be easily extended to the registration of a group of volumetric 3-D shapes by replacing the decoder with 3-D deconvolutional layers. Thus, our model can be used for groupwise medical image registration. Future work will extend this method for the registration of a group of volumetric shapes in the domain of medical imaging. In Section IV-B, we extend our model to register a large number of groups at the same time. For aligning multiple groups, our learning-based method with newly designed optimizable groupwise latent vectors demonstrates dramatic improvement in running efficiency.

B. Registration of Multiple Groups

In this section, we demonstrate the performance of our model for the registration of multiple groups of 3-D real-world datasets.

1) Dataset: We conduct experiments on 14 categories of the ShapeNet [59] dataset. For preparing the multiple groups dataset, we randomly select 100 groups of shapes from the dataset where each group includes three randomly selected shapes from the same category. Each shape contains 2048 uniformly sampled points.

2) Setting: As explained in Section III-E, we choose the latent vector z with a dimension of 256 for each group. For the decoder network, we use an MLP with dimensions of (259, 128, 64, 3) to generate 3-D flow fields. We simultaneously optimize all latent vectors and the decoder network on all groups. In this experiment, λ is set to 0.2. Due to a large number of groups, the compared methods (e.g., Holder, IP, and Norm-IP) cannot be applied in this setting. Groupwise C.D. is used to evaluate the performance of our model. We report

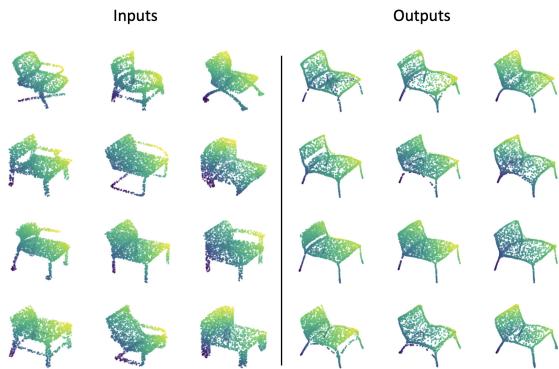


Fig. 10. Qualitative registration results for registration of multiple real-world 3-D point sets from the chair category of the ShapeNet [59] dataset. We randomly select four groups for demonstration. Corresponding points are shown in the same colors.

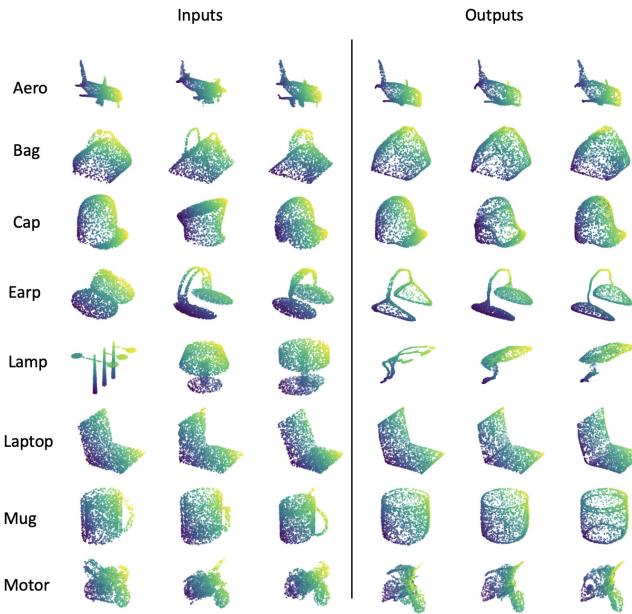


Fig. 11. Qualitative registration results for registration of multiple groups of 3-D real-world point sets in selected categories of the ShapeNet [59] dataset. We randomly select one case from 100 groups for demonstration for each category. Corresponding points are shown in the same colors.

the running time for every 100 groups to demonstrate efficiency.

3) *Results*: Fig. 10 shows the registration results of four groups of input shapes from the chair category. From the results, one can see that our model produces satisfactory alignments for each group of shapes even though the topological structures are quite different for some groups. For example, in the last row of Fig. 10, the legs in this group are well-aligned from three shapes with large structural variations. We show the registration results of the other categories in Fig. 11. As shown in Fig. 11, for most of the other categories, our GP-Aligner can well align all three input shapes to the mean shape.

The quantitative results for the inputs before alignment and outputs after alignment are listed in Table VII. From the table, we can see that the average groupwise C.D. reduces

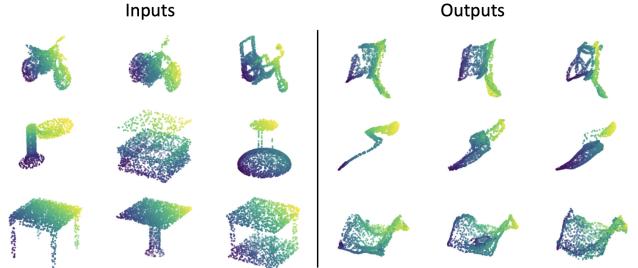


Fig. 12. Selected failed cases for registration of multiple groups of 3-D real-world point sets. Corresponding points are shown in the same colors.

TABLE VII
QUANTITATIVE PERFORMANCE AND TIME FOR ALIGNING 100 GROUPS OF 3-D REAL-WORLD POINT SETS FROM ALL THE CATEGORIES OF THE SHAPENET DATASET. WE REPORT THE GROUPWISE C.D. BEFORE AND AFTER ALIGNMENT WITH ALL VALUES SCALED BY $\times 10^{-4}$

Categories	Groupwise C.D. Before Alignment	Groupwise C.D. After Alignment
aero	77.49	5.59
bag	202.60	6.27
cap	180.80	5.69
car	53.89	4.30
chair	144.84	3.91
earphone	588.86	4.89
guitar	15.47	0.69
knife	33.64	1.01
lamp	753.10	4.97
laptop	56.00	3.34
motor	45.76	3.04
mug	108.14	7.94
pistol	60.37	2.56
rocket	38.09	1.60
skate	28.87	1.96
table	326.58	3.61
Average	169.66	3.83
Running Time	-	27min

from 169.66×10^{-4} to 3.83×10^{-4} after registration, which demonstrates that our GP-Aligner model can successfully align multiple groups of shapes. We notice that, for categories with high similarity in topological structure among input shapes, e.g., guitar, our model shows better registration performance, with a groupwise C.D. less than 1×10^{-4} . For categories with large topological variations, e.g., mug, our model achieves worse registration performance than other categories. Regarding registration efficiency, our model can accomplish the registration tasks of 100 groups of 3-D point sets in 27 min. We do not compare the efficiency with the comparing models due to the quite long computation time required by these methods.

In Fig. 12, we illustrate a few failed cases to show that our model may fail to maintain semantic meaning after deformations, especially when the topological structure dramatically varies among the shapes in the group. In the first row of Fig. 12, the third input motorbike has quite different structures from the other two motorbikes. Therefore, the deformed shape seems to be weird. For the third case, all three input lamps have large structural variations; we cannot expect reasonable alignment in this case.

V. CONCLUSION

In this article, we introduce a novel method, called GP-Aligner, for groupwise nonrigid point set registration. The proposed method is built upon deep neural network architecture and can perform the registration for groups of point sets in a purely unsupervised way. To avoid the explicit design of a shape feature encoding network, our model leverages an encoder-free optimizable latent descriptor to characterize the group relationships. An optimizable GLD is designed to characterize the groupwise relationships among a group of point sets. The GLD is randomly initialized from a Gaussian distribution and optimized along with the network parameters during the optimization process. We conduct extensive experiments for both synthesized shapes and real-world 3-D shapes. Experimental results show that our GP-Aligner model can achieve superior registration performance and running efficiency. We also demonstrate the robustness of our model under different types of input noises. For registering a large number of groups of shapes, our model shows satisfactory registration performance and remarkable running efficiency.

REFERENCES

- [1] J. Ma, J. Zhao, and A. L. Yuille, "Non-rigid point set registration by preserving global and local structures," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 53–64, Jan. 2016.
- [2] J. B. A. Maintz and M. A. Viergever, "A survey of medical image registration," *Med. Image Anal.*, vol. 2, no. 1, pp. 1–36, Mar. 1998.
- [3] Y. Wang and J. Solomon, "Deep closest point: Learning representations for point cloud registration," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3523–3532.
- [4] A. Myronenko, X. Song, and M. A. Carreira-Perpinán, "Non-rigid point set registration: Coherent point drift," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 1009–1016.
- [5] J. Ma, J. Zhao, J. Tian, A. L. Yuille, and Z. Tu, "Robust point matching via vector field consensus," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1706–1721, Apr. 2014.
- [6] H. Ling and D. W. Jacobs, "Deformation invariant image matching," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, vol. 2, Oct. 2005, pp. 1466–1473.
- [7] M. A. Viergever, J. B. A. Maintz, S. Klein, K. Murphy, M. Staring, and J. P. W. Pluim, "A survey of medical image registration—Under review," *Med. Image Anal.*, vol. 33, pp. 140–144, Oct. 2016.
- [8] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes," *Proc. SPIE*, vol. 1611, pp. 586–607, Apr. 1992.
- [9] A. Myronenko and X. Song, "Image registration by minimization of residual complexity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 49–56.
- [10] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1153–1190, Jul. 2013.
- [11] X. Liu, C. R. Qi, and L. J. Guibas, "FlowNet3D: Learning scene flow in 3D point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 529–537.
- [12] G. Balakrishnan, A. Zhao, M. R. Sabuncu, A. V. Dalca, and J. Guttag, "An unsupervised learning model for deformable medical image registration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9252–9260.
- [13] I. Rocco, R. Arandjelovic, and J. Sivic, "Convolutional neural network architecture for geometric matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jul. 2017, pp. 6148–6157.
- [14] J. Chen, L. Wang, X. Li, and Y. Fang, "Arbicon-Net: Arbitrary continuous geometric transformation networks for image registration," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.
- [15] T. Che et al., "Deep group-wise registration for multi-spectral images from fundus images," *IEEE Access*, vol. 7, pp. 27650–27661, 2019.
- [16] L. Wang, Y. Shi, X. Li, and Y. Fang, "Unsupervised learning of global registration of temporal sequence of point clouds," 2020, *arXiv:2006.12378*.
- [17] L. Ding and C. Feng, "DeepMapping: Unsupervised map estimation from multiple point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8650–8659.
- [18] L. G. Sanchez Giraldo, E. Hasanbelliu, M. Rao, and J. C. Principe, "Group-wise point-set registration based on Rényi's second order entropy," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2454–2462.
- [19] T. Chen, B. C. Vemuri, A. Rangarajan, and S. J. Eisenschenk, "Groupwise point-set registration using a novel CDF-based Havrda–Charvát divergence," *Int. J. Comput. Vis.*, vol. 86, no. 1, p. 111, 2010.
- [20] F. Wang, B. C. Vemuri, and A. Rangarajan, "Groupwise point pattern registration using a novel CDF-based Jensen–Shannon divergence" in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2006, pp. 1283–1288.
- [21] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Comput. Vis. Image Understand.*, vol. 89, nos. 2–3, pp. 114–141, Feb. 2003.
- [22] H. Chui, A. Rangarajan, J. Zhang, and C. M. Leonard, "Unsupervised learning of an atlas from unlabeled point-sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 160–172, Feb. 2004.
- [23] X. Li, J. Kaesemola Pontes, and S. Lucey, "Neural scene flow prior," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 7838–7851.
- [24] N. Gelfand, N. J. Mitra, L. J. Guibas, and H. Pottmann, "Robust global registration," in *Proc. Eurographics Symp. Geometry Process.*, vol. 2, no. 3, Vienna, Austria, 2005, p. 5.
- [25] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-ICP: A globally optimal solution to 3D ICP point-set registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 11, pp. 2241–2254, Nov. 2016.
- [26] H. Chui and A. Rangarajan, "A new algorithm for non-rigid point matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2000, pp. 44–51.
- [27] J. Li, Q. Hu, and M. Ai, "Point cloud registration based on one-point RANSAC and scale-annealing biweight estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9716–9729, Nov. 2021.
- [28] S. Quan and J. Yang, "Compatibility-guided sampling consensus for 3-D point cloud registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7380–7392, Oct. 2020.
- [29] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, vol. 1, no. 2, p. 4.
- [30] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "PointNet++: Deep hierarchical feature learning on point sets in a metric space," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5099–5108.
- [31] C. R. Qi, W. Liu, C. Wu, H. Su, and L. J. Guibas, "Frustum PointNets for 3D object detection from RGB-D data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 918–927.
- [32] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. Guibas, "KPConv: Flexible and deformable convolution for point clouds," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 6411–6420.
- [33] C. Wen, L. Yang, X. Li, L. Peng, and T. Chi, "Directionally constrained fully convolutional neural network for airborne LiDAR point cloud classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 162, pp. 50–62, Apr. 2020.
- [34] X. Li, L. Wang, M. Wang, C. Wen, and Y. Fang, "DANCE-NET: Density-aware convolution networks with context encoding for airborne LiDAR point cloud classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 166, pp. 128–139, Aug. 2020.
- [35] X. Li, C. Wen, L. Wang, and Y. Fang, "Topology constrained shape correspondence," *IEEE Trans. Vis. Comput. Graphics*, vol. 27, no. 10, pp. 3926–3937, Oct. 2021.
- [36] O. Halimi, O. Litany, E. R. Rodola, A. M. Bronstein, and R. Kimmel, "Unsupervised learning of dense shape correspondence," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4370–4379.
- [37] Y. Aoki, H. Goforth, R. A. Srivatsan, and S. Lucey, "PointNetLK: Robust & efficient point cloud registration using PointNet," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7163–7172.
- [38] Y. Wang and J. Solomon, "PRNet: Self-supervised learning for partial-to-partial registration," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 8814–8826.
- [39] X. Li, L. Wang, and Y. Fang, "Unsupervised category-specific partial point set registration via joint shape completion and registration," *IEEE Trans. Vis. Comput. Graphics*, early access, Mar. 7, 2022, doi: [10.1109/TVCG.2022.3157061](https://doi.org/10.1109/TVCG.2022.3157061).

- [40] Z. J. Yew and G. H. Lee, "RPM-Net: Robust point matching using learned features," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 11824–11833.
- [41] L. Sun et al., "A weakly supervised graph deep learning framework for point cloud registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–12, 2022.
- [42] J. Yang, J. Chen, S. Quan, W. Wang, and Y. Zhang, "Correspondence selection with loose-tight geometric voting for 3-D point cloud registration," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.
- [43] L. Wang, J. Chen, X. Li, and Y. Fang, "Non-rigid point set registration networks," 2019, *arXiv:1904.01428*.
- [44] L. Wang, X. Li, J. Chen, and Y. Fang, "Coherent point drift networks: Unsupervised learning of non-rigid point set registration," 2019, *arXiv:1906.03039*.
- [45] G. D. Evangelidis, D. Kounades-Bastian, R. Horaud, and E. Z. Psarakis, "A generative model for the joint registration of multiple point sets," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 109–122.
- [46] G. D. Evangelidis and R. Horaud, "Joint alignment of multiple point sets with batch and incremental expectation-maximization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 6, pp. 1397–1410, Jun. 2018.
- [47] H. Zhu et al., "A review of point set registration: From pairwise registration to groupwise registration," *Sensors*, vol. 19, no. 5, p. 1191, Mar. 2019.
- [48] A. Rasoulian, R. Rohling, and P. Abolmaesumi, "Group-wise registration of point sets for statistical shape models," *IEEE Trans. Med. Imag.*, vol. 31, no. 11, pp. 2025–2034, Nov. 2012.
- [49] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen, "PointCNN: Convolution on X-transformed points," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 820–830.
- [50] W. Wu, Z. Qi, and L. Fuxin, "PointConv: Deep convolutional networks on 3D point clouds," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9621–9630.
- [51] S. Tan and M. L. Mayrovouniotis, "Reducing data dimensionality through optimizing neural network inputs," *AIChe J.*, vol. 41, no. 6, pp. 1471–1480, Jun. 1995. [Online]. Available: <https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.690410612>
- [52] A. A. Rusu et al., "Meta-learning with latent embedding optimization," 2018, *arXiv:1807.05960*.
- [53] M. Bouakkaz and M.-F. Harkat, "Combined input training and radial basis function neural networks based nonlinear principal components analysis model applied for process monitoring," in *Proc. IJCCI*, 2012, pp. 483–492.
- [54] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry, "3D-CODED: 3D correspondences by deep deformation," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 230–246.
- [55] J. J. Park, P. Florence, J. Straub, R. Newcombe, and S. Lovegrove, "DeepSDF: Learning continuous signed distance functions for shape representation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 165–174.
- [56] A. L. Yuille and N. M. Grzywacz, "A mathematical analysis of the motion coherence theory," *Int. J. Comput. Vis.*, vol. 3, no. 2, pp. 155–175, 1989.
- [57] H. Fan, H. Su, and L. Guibas, "A point set generation network for 3D object reconstruction from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, no. 4, 2017, p. 6.
- [58] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567–585, Jun. 1989.
- [59] A. X. Chang et al., "ShapeNet: An information-rich 3D model repository," 2015, *arXiv:1512.03012*.



Lingjing Wang received the B.S. degree from Moscow State University, Moscow, Russia, in 2011, and the Ph.D. degree from the Courant Institute of Mathematical Science, New York University, New York, NY, USA, in 2019.

His research interests include deep learning and 3-D visual computing.



Nan Zhou received the B.S. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2014, and the Ph.D. degree in cartography and geographic information system (GIS) from the Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, in 2022.

He is currently an Assistant Professor with the Department of Ocean College, Jiangsu University of Science and Technology, Zhenjiang, China. His research interests include deep learning, computer vision, and remote sensing image recognition.



Hao Huang received the B.S. degree in digital media technology from the Beijing University of Posts and Telecommunications, Beijing, China, in 2012, and the M.S. degree in computer science from the University of Rochester, Rochester, NY, USA, in 2019. He is currently pursuing the Ph.D. degree with New York University, New York, NY, USA.

His research interests include 3-D computer vision and bioinformatics.



Jifei Wang received the B.S. degree from The University of Sydney, Camperdown, NSW, Australia, in 2011, and the Ph.D. degree from the Courant Institute of Mathematical Science, New York University, New York, NY, USA, in 2019.

His research interests include machine learning, computer vision, and financial engineering.



Xiang Li (Member, IEEE) received the B.S. degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2014, and the Ph.D. degree in cartography and geographic information system (GIS) from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2019.

His research interests include computer vision, deep learning, and remote sensing.



Yi Fang (Member, IEEE) received the B.S. and M.S. degrees in biomedical engineering from Xi'an Jiaotong University, Xi'an, China, in 2003 and 2006, respectively, and the Ph.D. degree in mechanical engineering from Purdue University, West Lafayette, IN, USA, in 2011.

He is currently an Associate Professor with the Department of Electrical and Computer Engineering, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates. His research interests include 3-D computer vision and pattern recognition, large-scale visual computing, deep cross-domain and cross-modality multimedia analysis, and computational structural biology.