

Topology Constrained Shape Correspondence

Xiang Li¹, Congcong Wen¹, Lingjing Wang¹, and Yi Fang¹

Abstract—To better address the deformation and structural variation challenges inherently present in 3D shapes, researchers have shifted their focus from designing handcrafted point descriptors to learning point descriptors and their correspondences in a data-driven manner. Recent studies have developed deep neural networks for robust point descriptor and shape correspondence learning in consideration of local structural information. In this article, we developed a novel shape correspondence learning network, called TC-NET, which further enhances performance by encouraging the topological consistency between the embedding feature space and the input shape space. Specifically, in this article, we first calculate the topology-associated edge weights to represent the topological structure of each point. Then, in order to preserve this topological structure in high-dimensional feature space, a structural regularization term is defined to minimize the topology-consistent feature reconstruction loss (Topo-Loss) during the correspondence learning process. Our proposed method achieved state-of-the-art performance on three shape correspondence benchmark datasets. In addition, the proposed topology preservation concept can be easily generalized to other learning-based shape analysis tasks to regularize the topological structure of high-dimensional feature spaces.

Index Terms—Topology preservation, shape correspondence, locally linear embedding, graph convolution

1 INTRODUCTION

THE shape correspondence problem is generally defined as the identification of the homologous points of two or more shapes represented as point sets, i.e., the identification of the homologous points that are attributed with the same or similar structures in terms of their local appearance and geometric context. This problem has been an active research area in the computer vision community for decades [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], and it serves as an essential building block for many applications such as 3D reconstruction, shape registration, time-varying surface reconstruction, shape interpolation and information transfer [12]. Nevertheless, shape correspondence is still challenging due to the low data resolution and high amounts of sensor noise, especially for real-world scanning artifacts that exhibit significant non-rigid deformations and shape variations.

Given mesh data as an input, shape correspondence can be generally established by measuring the similarity of the point descriptors of paired shapes. Early studies have developed various point descriptors based on the extrinsic structures of input shapes that are invariant under rigid transformations. Commonly known methods include the spin image shape context method [13], the spin image method [14], and the integral volume descriptor method [15]. Recent efforts mainly focus on characterizing the intrinsic geometric properties that are invariant under non-rigid transformations. Methods in this family are mostly based on diffusion and

spectral geometry and aim to design isometry-invariant shape descriptors. Notable methods include the heat kernel signature (HKS) method and the wave kernel signature (WKS) method [16], [17].

More recently, with the prevalence of deep learning in various vision tasks, researchers have tried to learn shape correspondences with deep neural networks supervised by human-annotated correspondence labels. Previous works in this direction tried to generalize convolutional operations to irregular graph-structured data by using the local polar pseudo-coordinates around each node [18], [19], [20]. Among these methods, a convolution-like operator is defined to learn intrinsic geometric properties. The encouraging experimental results consistently indicated that effective intrinsic geometric feature learning is crucial to learning shape correspondences.

In this paper, to learn point descriptors with powerful representation abilities and to guarantee accurate shape correspondence learning, we take a further step to investigate the effect of the local topological structure of the input shape during the point feature learning process. The preservation of the intrinsic topological structure between input data and the embedding space has been extensively studied, and its effectiveness in the manifold learning domain has been proven. Those studies naturally motivate us to explore how this concept can contribute to point feature learning by preserving the topological consistency between the input data space and the learned feature spaces. To this end, we develop a novel topology-constrained shape correspondence network (TC-NET) that learns point descriptors while forcing the local topological structure in the embedding feature space to be consistent with the one lying in the raw input shape space.

Fig. 1 illustrates the pipeline of our proposed TC-NET, which is composed of four main components. The first component carries out topological structure mining. In this

• X. Li, L. Wang, and Y. Fang are with NYU Multimedia and Visual Computing Lab, NYU Abu Dhabi, NYU Tandon and Abu Dhabi.
E-mail: {xl1845, lw1474, yfang}@nyu.edu.

• C. Wen is with the NYU Tandon School of Engineering.
E-mail: cw3437@nyu.edu.

Manuscript received 15 June 2019; revised 27 Apr. 2020; accepted 4 May 2020. Date of publication 11 May 2020; date of current version 2 Sept. 2021.

(Corresponding author: Yi Fang.)

Recommended for acceptance by J. Tierny.

Digital Object Identifier no. 10.1109/TVCG.2020.2994013

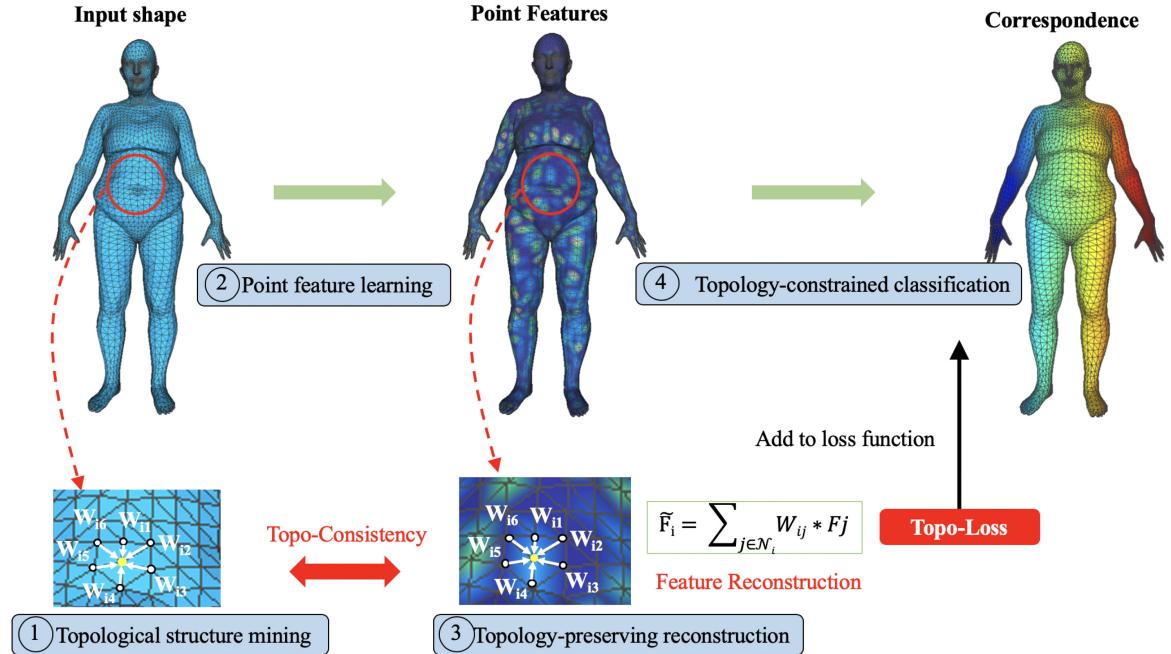


Fig. 1. Framework of the topology-constrained shape correspondence learning network. Our method aims to learn topology-preserving point descriptors during the shape correspondence learning process. Given an input shape represented by 3D coordinates and the adjacency graph, our model first extracts the topological structure for each individual point, which is represented by the edge weights connecting the point to its neighbors (step 1). Then, a graph-based convolutional neural network is designed to learn the latent features of each point (step 2). To enforce topological consistency between the input space and the embedding feature space, our model generates reconstructed features for each point by referring to the topological structure in the input space (step 3). We further formulate a Topo-Loss based on the topological similarity between the reconstructed features and the original ones. Finally, the Topo-Loss regularization term is integrated into the vertex classification loss function during shape correspondence learning (step 4). This figure is best viewed in color.

component, the topological structure is formulated based on the adjacency graph of the input mesh, with the edge weights representing the structural correlation between a node and its neighbors. The second component carries out point feature learning. In this component, a graph-based convolutional neural network is adopted to extract high-level point features. The third component carries out topology preserving reconstruction. In this component, the local topological structure in the input shape is transferred into the latent feature space and further used to reconstruct the features of each node. The last component carries out topology-constrained classification. In this component, a topology preserving reconstruction loss function (Topo-Loss) is formulated by measuring the similarity between the embedding features of each point and their reconstructions. This regularization term is added to the final loss function to learn more representative point descriptors, which guarantees more robust point correspondence learning. Accordingly, our main contributions can be summarized as follows:

- We propose the TC-NET paradigm that explicitly preserves the topology during the point feature learning process and experimentally verifies its effectiveness for 3D shape correspondence.
- By integrating topology-preserving reconstruction loss into the feature learning network as a regularization term, our proposed method can learn topology-constrained point descriptors, which further boosts the shape correspondence learning performance.
- Instead of using maximum indices to get the prediction label for each vertex, we propose to find the

bijective correspondence by solving a linear assignment problem.

- Our method achieves state-of-the-art performance on three shape correspondence benchmark datasets.

2 RELATED WORK

2.1 Handcrafted Point Descriptors

Before deep learning methods were prevalently applied for point descriptor learning, researchers developed various handcrafted point descriptors for characterizing the local structure of each point on a 3D shape. While early methods such as the shape context [21], spin image [14] and multi-scale local signature [22] focused on shape correspondence under rigid transformations, recent efforts are mainly focused on the development of deformation-insensitive point descriptors that can handle the correspondence problem under non-rigid shape deformations [23], [24], [25]. Among these methods, diffusion geometry-based methods obtain point descriptors from the spectral decomposition of the Laplace-Beltrami operator associated with the shape. Rustamov [26] proposed a deformation invariant representation of surfaces in the space of convergent sequences $\ell^2(R)$ by making use of the eigenvalues and eigenfunctions of the Laplace-Beltrami differential operator. Based on the fundamental solution of the heat diffusion equation, Sun *et al.* [16] developed the heat kernel signature (HKS) as a point signature based on the fundamental solutions of the heat equation (heat kernels). Furthermore, based on the heat diffusion distance, Vestner *et al.* [25] presented an efficient deformable

shape correspondence method by solving the matching problem between a set of pairwise and point-wise descriptors. In [17], the wave kernel signature (WKS) was proposed to improve the quality of the point signatures by encoding information from multi-scale Laplace eigenfrequencies. These diffusion-based methods have been reported to show superior performance over other handcrafted point descriptors in numerous shape analysis tasks.

2.2 Correspondence Learning

Learning-based methods have been successful applied to various computer vision tasks. Recent efforts tried to generalize successful deep learning paradigms to non-euclidean data such as manifolds or graphs [27]. Masci *et al.* [19] proposed the first CNN-like neural network on non-euclidean domains (surfaces) based on a local geodesic system of polar coordinates. The framework exhibited promising performance in representative point feature learning for the problems of shape description, retrieval, and correspondence. In [20], the author improved upon the method from [19] by incorporating oriented anisotropic heat kernels. Based on the good properties of [19], [20], Monti *et al.* [28] proposed a generalized framework that allows robust feature learning on non-euclidean domains (graphs and manifolds). More recently, Verma *et al.* [29] proposed the so-called FeaStNet to learn shape correspondences using a novel graph-convolution operator. Some other research [11], [30] tried to generate correspondences by transferring the correspondence labels from template shapes to the reconstructed ones in a deformation process.

Another promising way to learn shape correspondence is the family of methods based on functional maps [31]. Instead of using point-wise descriptors for point-to-point matching, these methods treat the overall shape correspondences as linear compositions between spaces of functions on manifolds and can obtain global yet compact shape correspondences. Litany [32] built a structured prediction model in the space of functional maps along with deep residual networks to provide a compact representation of shape correspondence.

2.3 Topology-Preserving Manifold Learning

Exploiting the intrinsic topological structure of high-dimensional data has been attracting widespread attention in the computer vision community, e.g., [33], [34], [35], [36]. Topology-preserving learning aims to learn a mapping function that transfers a finite set of data into an embedding feature space while preserving the topological structure present in the original data space. Commonly used topology-preserving learning methods include self-organizing mapping (SOM) [37] and locally linear embedding (LLE) [38].

While SOM learns to map the input data into predefined lattices (SOM nodes) in a competitive learning process with a neighborhood function to preserve the local topology on SOM nodes, LLE is capable of preserving the local topology on each point. Furthermore, LLE is inherently invariant to rotation, translation, and scaling [39], and it is widely used in the field of manifold learning. Therefore, it is a promising method for our point feature learning task. Following LLE, several variations have been proposed for improving the quality of feature learning by incorporating multiple weight vectors [40] or Hessian-based quadratic equations [41].

3 OUR METHOD

In this section, we introduce the proposed topology-constrained shape correspondence learning method. In Section 3.1, we illustrate the notations for our shape correspondence learning method. Section 3.2 introduces our formulation of the local topological structure in the input shape space. We present our feature learning network in Section 3.3 and the topological regularization term in Section 3.4. The overall network architecture is presented in Section 3.5, and the final correspondence prediction module is illustrated in Section 3.6.

3.1 Problem Statement

In this paper, we formulate the shape correspondence between the input shape and the given reference shape as a vertex labeling problem, where the classification label of each point on the input shape is defined as the vertex index of the corresponding point on the reference shape. Given an input shape represented by the vertices $P = \{P_1, P_2, \dots, P_N\}$, $P_i \in R^D$ and an adjacency graph $A \in R^{N \times N}$, where N represents the number of points and D stands for the input dimension ($D = 3$ in this paper), our vertex labeling network produces a $N \times N$ matrix, with each row indicating the probability distribution over the corresponding point on the reference shape. A point feature learning network is formulated to extract representative point descriptors $F_i (i = 1, 2, \dots, N)$, $F_i \in R^d$. Traditional shape correspondence learning networks, such as [20], [29], [42], directly predict the vertex labels guided by a classification loss function. In this paper, a topological regularization term is defined on point feature representation F to enforce the topological consistency between the input shape space and the feature space. The topology-preserving module can guarantee more representative point descriptor learning and boost the shape correspondence learning performance.

3.2 Topological Structure Mining

In this paper, we apply local linear embedding (LLE) [38] to explore the topological structure of the input. LLE is one of the most successful and widely used methods for topology-preserving learning in the geometric learning domain. The intrinsic conformal mapping inherently makes LLE invariant to rotation, translation, and scaling; thus, it is a promising technique for the robust point descriptor learning of deformable shapes.

Initially, LLE assumes each data point x_i and its neighboring points \mathcal{N}_i should lie on or in proximity to a locally linear patch of the manifold; thus, they can be reconstructed using a linear combination of all neighbors points. In this paper, the watertight mesh data are naturally subject to this assumption. The reconstruction weights can be computed by solving the following least squares problem:

$$\epsilon(W) = \sum_{i=1}^N \|x_i - \sum_{j \in \mathcal{N}_i} W_{ij} x_j\|^2 \quad (1)$$

$$s.t. \sum_{j \in \mathcal{N}_i} W_{ij} = 1. \quad (2)$$

The weight coefficient W_{ij} indicates the contribution of the j th neighbor point to the i th point reconstruction, and it

can effectively characterize the intrinsic geometric properties of each point and its neighbors. Here, rotation and scaling invariance are intuitively forced by Eq. (1), while the sum-to-one constraint given in Eq. (2) leads to translation invariance.

3.3 Point Feature Learning

To guarantee the representation power of learned point features, a graph-based convolution neural network is commonly adopted to exploit local structural information [19], [20], [28], [29], [43]. Among these methods, the FeaStNet model [29] proposes a graph-convolution operator that uses the learned features of the preceding layer to dynamically determine the association between filter weights and the neighbor nodes. In this paper, we use this graph convolutional layer for point feature learning. Note that other point feature learning networks such as geodesic CNN [19] and anisotropic CNN [20] can also be used.

3.4 Topological Regularization

To enforce the topological consistency between the input shape space and the embedding feature space, we explicitly transfer the topological structure discussed in Section 3.2 to the feature space and further formulate a topological regularization term. Following locally linear feature regularization [44], which regularizes the classification model by penalizing the difference between each node's weight and the reconstructed weight at the same node by referring the neighborhood graph, we intend to preserve the topological structure by simply adding a topology-preserving reconstruction loss function (Topo-Loss) defined for the embedding feature space, formulated as Eq. (3):

$$\mathcal{L}_{Topo} = \frac{1}{N * d} \sum_{i=1}^N |F_i - \sum_{j \in \mathcal{N}_i} W_{ij} F_j|^2, \quad (3)$$

where d denotes the dimension of the embedded feature layer and $F_i \in R^d$ denotes the i th point's feature in this layer. We integrate this regularization loss function into the final classification loss function with a regularization coefficient of λ . The overall object loss function for our shape correspondence learning problem is formulated as

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda \mathcal{L}_{Topo}, \quad (4)$$

where \mathcal{L}_{cls} denotes the cross-entropy loss function defined based on the predicted vertex labels and the ground-truth labels. By adding this topological regularization term, our shape correspondence learning network is forced to minimize the topological reconstruction loss function and thus maintain topological consistency during the training process. Note that our topological regularization term includes no extra parameters or computational costs in the inference stage.

3.5 Network Architecture

Following the network architectures of [19] and [20], we first design a single-scale shape correspondence learning model that consists of a sequence of MLP layers (11 convolutions) and graph convolution layers (GConv; see Section 3.3 for details). For a fair comparison with [29], our single-scale

network is formulated as follows: MLP(8) + GConv(32) + GConv(64) + GConv(128) + MLP(256) + MLP(N), where the numbers indicate the feature dimensions of each layer and N indicates the number of points on the input shape. Our Topo-Loss regularization term is formulated on the last graph convolution layer (i.e., GConv(128)).

In addition, we develop a multi-scale architecture following FeaStNet [29], with pooling and unpooling layers for hierarchy feature learning. The Graclus algorithm [45] is used to define the pooling operation on the graph [46]. Fig. 2 illustrates our multi-scale shape correspondence learning network. First, our network takes as input the mesh data represented by raw 3D XYZ coordinates and the adjacency graph. Then, graph convolution layers are used for point feature learning, followed by pooling layers for neighbor feature aggregation. To recover the down-scaled point features to the original resolution, the fractional stride convolution up-sampling layers are employed, and the outputs of which are concatenated with the features map of the same resolution in the down-sampling path. Note that we enforce local topological preservation on the last up-sampling layer by reconstructing the same topological structure as that of the input shape space on the embedding features. Using successive MLP layers, we obtain a soft matching matrix B , where $B \in R^{N \times N}$, and each row vector indicates the probability distribution over the corresponding point on the reference shape. Cross-entropy loss and topology preserving reconstruction loss (Topo-Loss) terms are combined in the loss function to optimize model parameters.

3.6 Correspondence Prediction

In this paper, two different strategies are used to generate the correspondence prediction. Given the predicted soft matching matrix B , we can take the index of the maximum value as the prediction label for each point i , i.e., $\pi_i = argmax(B_i)$, where $i = 1, 2, \dots, N$, or we can formulate the following bijection assignment problem:

$$P^* = argmax < P, B >, \quad (5)$$

with $P \in \{0, 1\}^{N \times N}$ being a permutation matrix representing the bijection correspondence. The optimal assignment P^* can be estimated by solving a linear assignment problem (LAP) of Eq. (5) following [25].

4 EXPERIMENTS

4.1 Dataset

FAUST. The FAUST dataset [47] was one of the first datasets to introduce high-resolution, real-world scanned meshes of humans with ground-truth correspondences. The experimental dataset consists of 100 watertight meshes with 6,890 vertices each, corresponding to 10 shapes in 10 different poses each. In our experiments, we use the first 80 meshes for network training, and the remaining 20 meshes include 10 poses of two shapes for model evaluation. The ground-truth shape correspondence is known, and the first shape of the first pose is taken as a reference.

TOSCA. The TOSCA dataset [48] consists of high-resolution three-dimensional non-rigid shapes in a variety of poses by deforming template meshes. In the dataset, objects

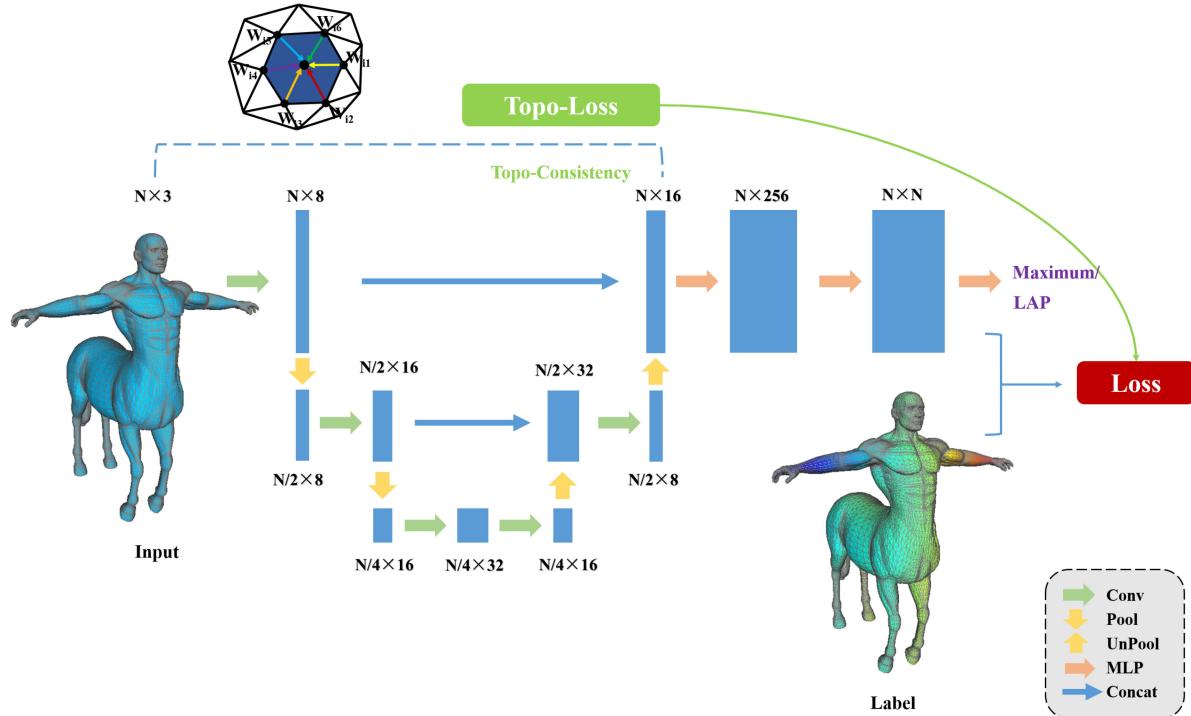


Fig. 2. Overview of the proposed multi-scale point correspondence network. Following [29], our model takes a 3D mesh as an input, represented by the $N \times 3$ matrix of 3D coordinates, along with the adjacency graph. A point feature learning network is formulated to learn the $N \times 16$ embedding point descriptors. Our model enforces topological consistency between the input shape space and the embedding feature space by adding the topology-preserving reconstruction loss function (Topo-Loss) into the final vertex labeling loss function. The topology-constrained point descriptor is further used to generate the $N \times N$ point label outputs. The final correspondence prediction can be generated either by the softmax function or by solving a linear assignment problem. This figure is best viewed in color.

within the same class have the same triangulation and equal numbers of vertices arranged in a compatible way. The Michael dataset, which contains 20 non-rigid human shapes, is selected for our experiments.

To evaluate the performance of our model on non-human shapes, we conduct additional experiments on dog and cat shapes. The dog and cat datasets contain 9 and 11 deformable shapes, respectively, and the points and polygons in those shapes correspond.

SCAPE. The SCAPE dataset [49] contains 71 registered meshes of a particular person in different poses. This dataset contains 25,000 polygon meshes containing morphs created with the correlated correspondence algorithm [49], and the points and polygons in those meshes correspond.

4.2 Experimental Setup

As part of the model input, the initial adjacency graph was generated by selecting 6 nearest neighbor points for each vertex. In the graph convolution layer, we set the number of weight matrix M to 32. Our Topo-Loss regularization term is embedded into the last unpooling layer, and the value of the regularization coefficient will be discussed in Section 5.1. As a default configuration, we set the topological regularization coefficient to 0.001.

Our network is optimized using the Adam optimizer with an initial learning rate of 1e-2, and the learning rate is divided by 10 every 30,000 steps. We set the batch size to 2, momentum to 0.9, and weight decay to 1e-5. For model regularization, we add L2 regularization to the model with a factor of 1e-5. We implement our method with the TensorFlow library on a TESLA K80 GPU.

4.3 Results on FAUST

In Table 1, we compare the correspondence labeling accuracy of our method with those of state-of-the-art methods on the FAUST dataset by adopting the 3D XYZ coordinates or the SHOT descriptor as input. Note that the accuracy reported in Tables 1 and 2 is the correspondence accuracy with zero geodesic error. It can be found that our model achieves 99.7 percent accuracy with raw 3D XYZ coordinates, which is a significantly better result than those of the compared methods. Fig. 3 shows selected examples of the correspondence results between the reference shape and three deformable shapes, represented by the same color of corresponding points.

TABLE 1
Correspondence Accuracies of Our Model and
the State-of-the-Art Approaches
on the FAUST Dataset

Method	Input	Accuracy
PointNet [52]	SHOT	49.7%
ACNN, w/o refinement [20]	SHOT	60.6%
ACNN, w/ refinement [20]	SHOT	62.4%
GCNN, w/o refinement [19]	SHOT	65.4%
GCNN, w/ refinement [19]	SHOT	42.3%
MoNet, w/o refinement [28]	SHOT	73.8%
MoNet, w/ refinement [28]	XYZ	88.2%
FMNet, w/ refinement [32]	SHOT	98.0%
FeaStNet, single-scale, w/o refinement [29]	XYZ	88.1%
FeaStNet, multi-scale, w/o refinement [29]	XYZ	98.6%
Ours, single-scale, w/o refinement	XYZ	94.0%
Ours, multi-scale, w/o refinement	XYZ	99.7%

TABLE 2
Effect of Topological Regularization

Method	λ	Accuracy
FeaStNet [29]	-	88.1%
baseline	-	90.6%
Ours	0.0005	91.0%
Ours	0.001	94.0%
Ours	0.003	92.8%
Ours	0.005	91.7%
Ours, multi-scale	0.001	99.7%

λ indicates the coefficient of the topological regularization term.

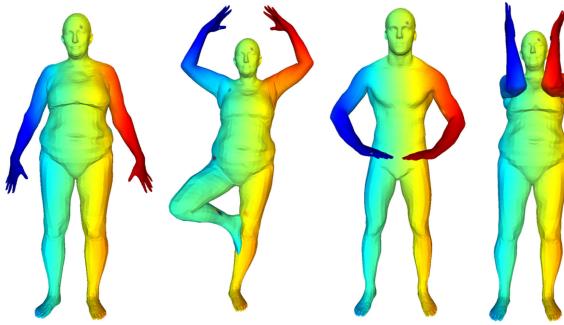


Fig. 3. Correspondence results obtained by our model on the FAUST test set. The reference shape is shown on the left, and three deformable shapes with correspondence are shown on the right. Corresponding points are shown in the same color.

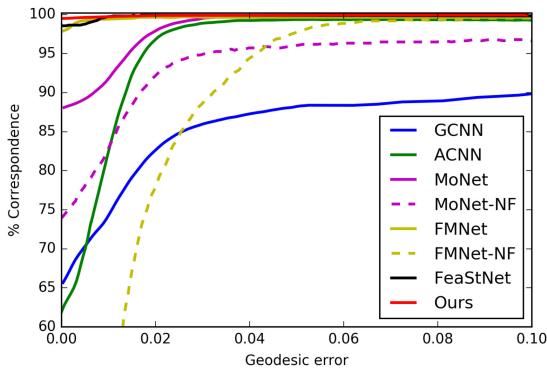


Fig. 4. Comparison with learning-based shape matching approaches on the FAUST human dataset. Our model reaches a correspondence accuracy of 99.7 percent with zero error (top left of the plot).

Furthermore, using the Princeton benchmark protocol [50] evaluation metric, we compare the performance of our model on the FAUST dataset with state-of-the-art learning-based approaches for deformable shape correspondence, including geodesic CNNs (GCNN) [19], anisotropic CNNs (ACNN) [20], MoNet [28], FMNet [32] and FeaStNet [29], as shown in Fig. 4. Our model achieved state-of-the-art correspondence performance on the FAUST dataset. Note that all the comparing methods are post-processed with the refinement technique of [51].

4.4 Results on TOSCA

In this section, we conduct experiments on the TOSCA Michael dataset. The TOSCA Michael dataset contains 20 human shapes, and we randomly select 16 shapes for training and the remaining 4 shapes for model evaluation. The

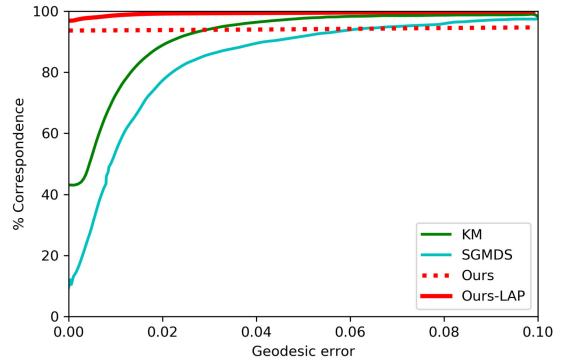


Fig. 5. Comparison with learning-based shape matching approaches on the TOSCA human dataset. Our model reaches a correspondence accuracy of 96.9 percent with zero error (top left of the plot).

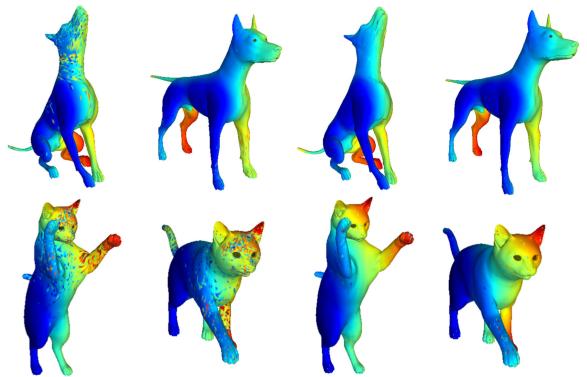


Fig. 6. Correspondence results of the non-human shapes from the TOSCA dataset generated by maximum assignment (left columns) and LAP assignment (right columns).

original human shapes in the TOSCA high-resolution dataset possess over 50K vertices, and we down-sample the data to around 13K vertices due to the GPU memory limitation. Each shape is uniformly sampled and split into 4 shapes, and all sub-sampled shapes are fed into our model for training. During the test stage, we aggregate the predicted labels and map them to the original label space. In Fig. 5, we compare the performance of our model with those of the kernel matching (KM) [25], SGMDS [23] and FMNet [32] models on the TOSCA human dataset. As can be seen in the figure, our model with LAP performs significantly better than the compared methods. More importantly, our model with the maximum assignment and LAP assignment achieves correspondence accuracies with zero error of 93.7 and 96.8 percent, respectively, significantly outperforming all compared methods. The results also prove the superiority of the LAP assignment over the maximum assignment for correspondence prediction.

Additionally, we test the performance of our model on non-human shapes, and the correspondence results are shown in Fig. 6. As can be seen from Fig. 6, our model with both assignment strategies can generate accurate correspondence labels for most of the vertices on test shapes. However, using the maximum assignment can lead to correspondence noise, probably because the maximum assignment considers only the probability distribution at one vertex and not those of its neighbors (i.e., it neglects the useful label correlations between neighboring vertices). By using the LAP assignment,

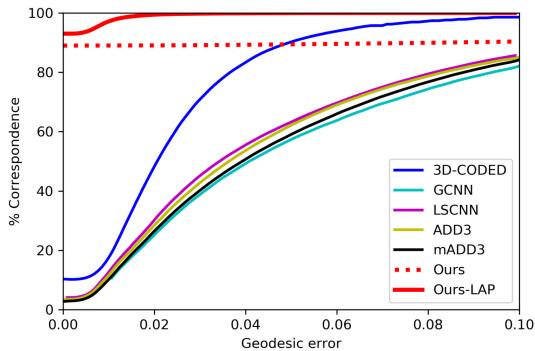


Fig. 7. The correspondence performance of our model and the compared learning-based models on the SCAPE dataset. Our model reaches a correspondence accuracy of 94.4 percent with zero error (top left of the plot).

our model can generate more accurate correspondence predictions.

4.5 Results on SCAPE

On the SCAPE dataset, our model also gets superior performance compared with state-of-the-art models for shape correspondence, including the optimal spectral descriptor (OSD) model [18], geodesic CNNs (GCNN) [19], localized spectral CNNs (LSCNN) [53], two variants of anisotropic diffusion descriptors (ADD3, mADD3) [20], FMNet [32], and 3D-CODED [30], as shown in Fig. 7. More specifically, the proposed TC-NET model achieves a far higher correspondence accuracy at zero error. Note that our model is trained and evaluated using full meshes while other methods use partial meshes. FMNet is trained on the FAUST dataset.

5 DISCUSSION

5.1 Ablation Study

In this section, we conduct further experiments to explore different hyperparameter configuration, and we investigate the effect of local topology preservation. Here, we conduct experiments on the FAUST dataset with the maximum assignment to predict the correspondence label for each vertex.

5.1.1 Effect of Topological Regularization

First, we test the effectiveness of topological regularization for local topology preservation. Table 2 shows the correspondence accuracy under different Topo-Loss regularization

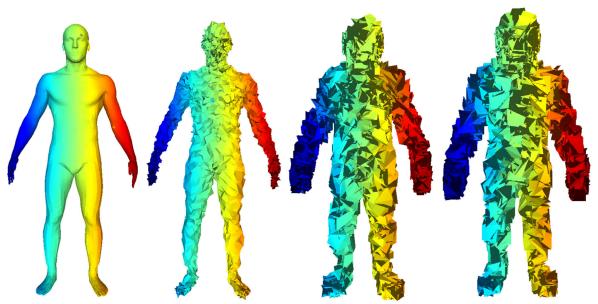


Fig. 9. Visualization of the correspondence results under different levels of Gaussian noise (from left to right: 0, 0.01, 0.1, 0.2).

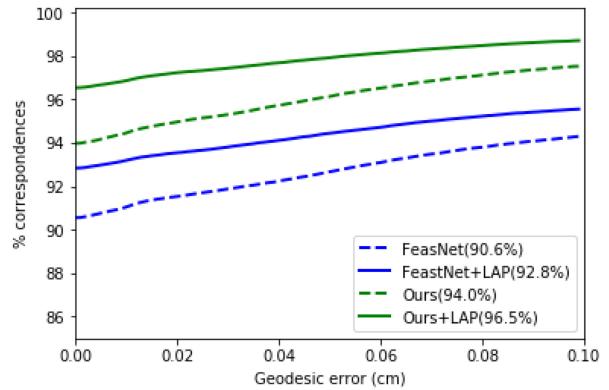


Fig. 10. Correspondence performance under different configurations on the FAUST test set. Numbers indicate the correspondence accuracy with zero error (top left of the plot).

coefficients. As shown in Table 2, with the topological regularization term, our model consistently improves the correspondence performance. Additionally, our model achieves the best performance when the topological regularization coefficient equals 0.001, with an improvement of 3.97 percent over the baseline model without topological regularization. Fig. 10 shows the correspondence performance under four different configurations, including FeaStNet, FeaStNet with LAP, our single-scale model, and our single-scale model with LAP. As shown in this figure, our proposed Topo-Loss regularization can significantly boost the performance by using either softmax or LAP for correspondence prediction.

Fig. 8 shows the geodesic errors of our model with and without topological regularization and the multi-scale network. As

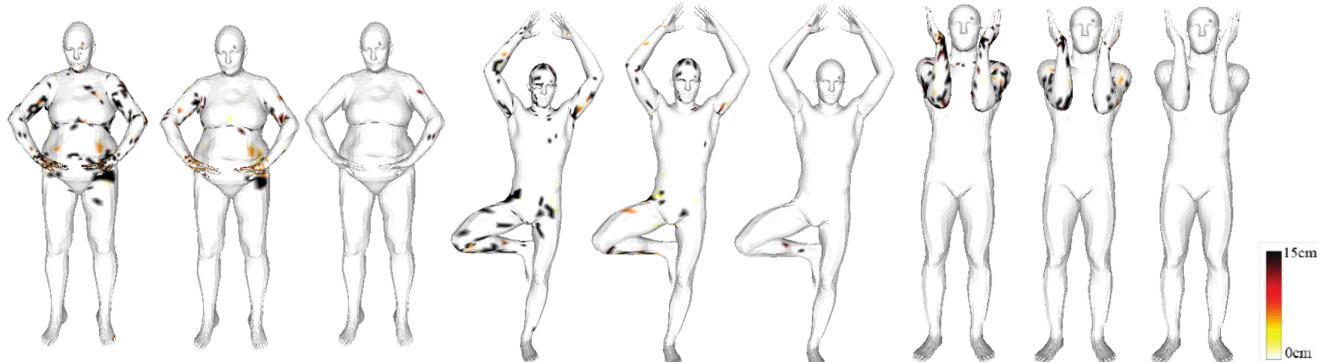


Fig. 8. Visualization of the correspondence errors in terms of the geodesic distance to the ground-truth correspondence on three test shapes, using (from left to right) the single-scale architecture without and with Topo-Loss and the multi-scale architecture with Topo-Loss.

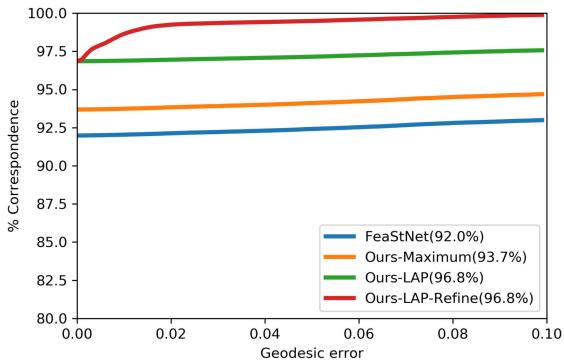


Fig. 11. Correspondence performance under different configurations on the TOSCA human test set. Numbers indicate the correspondence accuracy with zero error (top left of the plot).

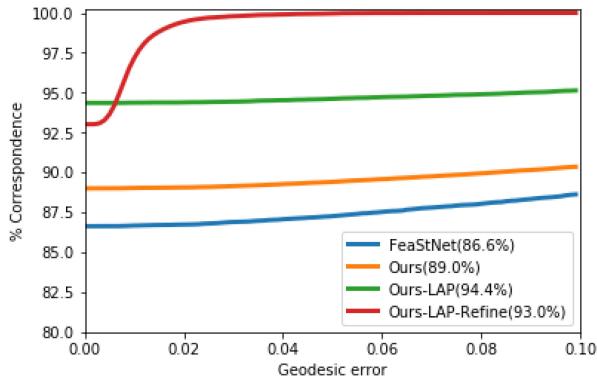


Fig. 12. Correspondence performance under different configurations on the SCAPE test set. Numbers indicate the correspondence accuracy with zero error (top left of the plot).

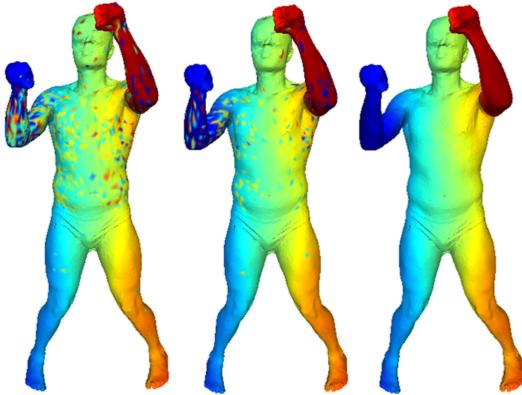


Fig. 13. Illustration of the correspondence results generated by maximum assignment and LAP assignment, and the result after refinement.

can be seen in this figure, by using topological regularization, our model obtains significantly smaller geodesic correspondence errors, which further validates the effectiveness of topological regularization.

5.1.2 Effect of Multiscale Architecture

In Table 2, we show the performance of our model using a single-scale architecture and that using a multiscale architecture. As indicated in this table, our model is further improved by using the multiscale architecture. The geodesic error shown in Fig. 8 also proves the effectiveness of the multiscale architecture.



Fig. 14. Illustration of the feature maps learned by our TC-NET model on different meshes from the FASUT test set. We randomly pick one feature from the last convolutional layer for visualization. Note the consistent pattern of feature maps across poses.

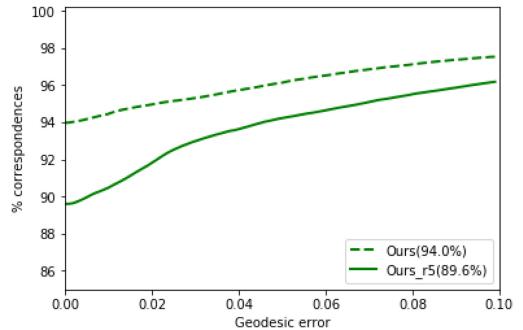


Fig. 15. Correspondence performance under different mesh connectives. The dashed line indicates our model using the complete adjacency graph, and the solid line indicates our model using the perturbed adjacency graph. Numbers indicate the correspondence accuracy with zero error (top left of the plot).

Robustness Test. To evaluate the robustness of the model, we add Gaussian noise to the coordinates of each vertex of the input shapes. We train our model with the noisy data and visualize the results of the multiscale model with TopoLoss on three new shapes in Fig. 9. As shown in Fig. 9, our model obtains quite consistent correspondence results under different noise levels.

5.2 Effect of LAP and Refinement

In Section 3.6, we introduce two strategies for generating the correspondence prediction. The first one directly takes the maximum index of the probability distribution matrix as the prediction label for each vertex. The other one determines the optimal correspondence matches by solving a linear assignment problem (LAP). Figs. 11 and 12 show a comparison between these two strategies on the TOSCA human dataset and the SCAPE dataset. As we can see from the figures, both assignment strategies can generate quite accurate correspondence labels for most of the vertices on test shapes with zero errors, and the correspondence accuracy of LAP assignment suppresses the maximum assignment by a large margin (3.2 and 4.0 percent for the TOSCA human dataset and the SCAPE dataset, respectively). The visualization result in Fig. 13 also proves the superiority of the LAP assignment over the maximum assignment (left two columns).

Furthermore, as shown in Fig. 13, both assignment methods produce prediction labels with local outliers. These outlier vertices have large geodesic errors and lead to the slow increment of the correspondence accuracy in Figs. 11 and 12. We can find an outlier point by simply mapping the center

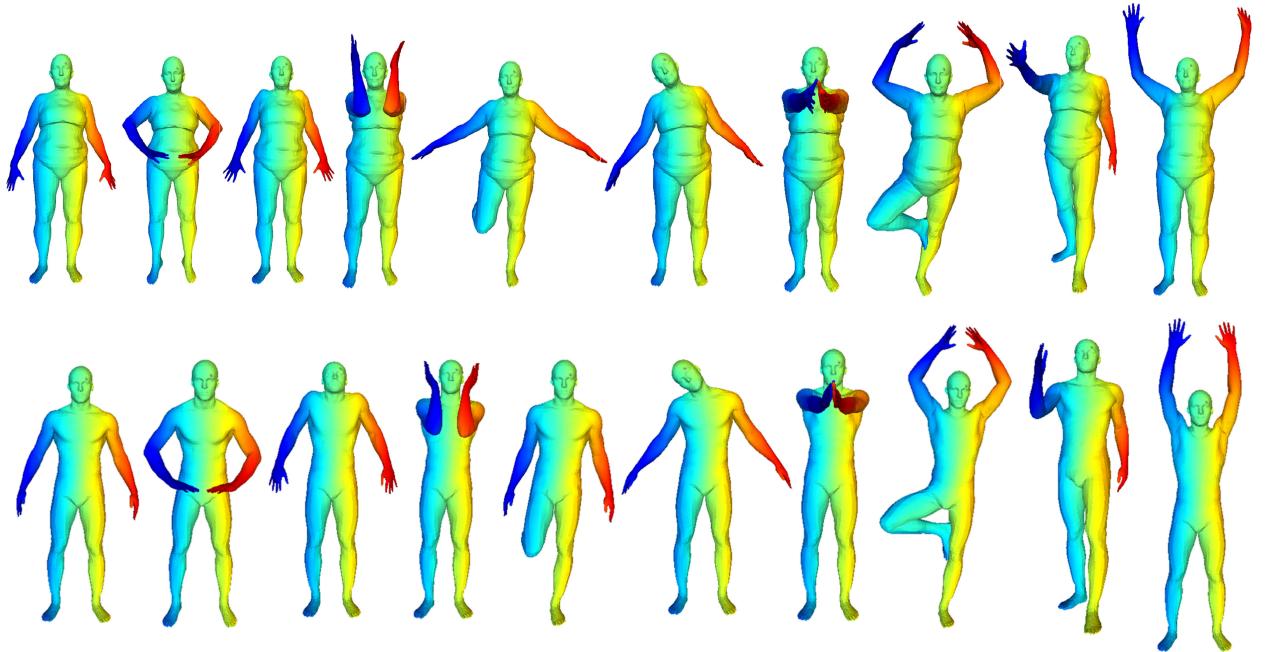


Fig. 16. Visualization of the correspondence results on the FAUST test set. Corresponding points are shown in the same color.

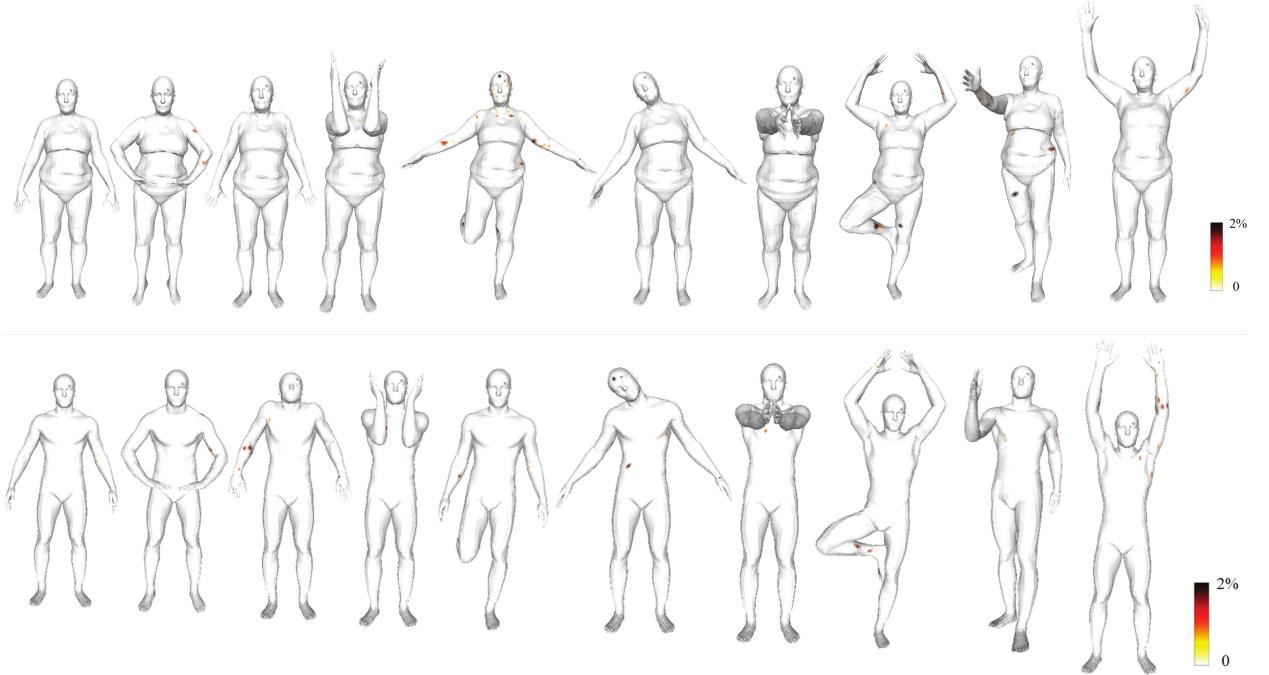


Fig. 17. Visualization of the geodesic error of the TC-Net model on the FAUST test dataset. The error values are saturated at 2 percent of the geodesic diameter, which corresponds to approximately 4 cm. Warm colors represent large errors.

point and its neighbors to the reference shape and calculating the distance of the corresponding points on the reference shape. If a large percentage of its neighbors is far from the center point, then we recognize the center point as an outlier. Specifically, we consider each point x_i on the input shape \mathcal{X} and its corresponding point y_i in the reference shape \mathcal{Y} . We identify point x_i as an outlier point by Eq. (6).

$$O_i = 1 \left\{ \sum_{j \in \mathcal{N}_i} 1 \{ d_{\mathcal{Y}}(y_i, y_{ij}) > \theta * \text{area}(\mathcal{Y})^{1/2} \} \right\} > \beta * |\mathcal{N}_i|. \quad (6)$$

Here, point j denotes the neighbor point of point i , y_{ij} denotes its predicted correspondence point on the reference shape \mathcal{Y} , and $d_{\mathcal{Y}}(\cdot, \cdot)$ denotes the euclidean distance on shape \mathcal{Y} . We use the euclidean distance to ensure computational efficiency. One can also use the geodesic distance. $1\{\cdot\}$ denotes the identity operator, θ denotes the threshold of the euclidean distance for identifying outlier bijections, and β denotes the threshold of the ratio of outlier bijections. In this paper, we set θ to 0.1 and β to 0.5.

After identifying all outlier points, we refine our results by replacing their correspondence predictions with the

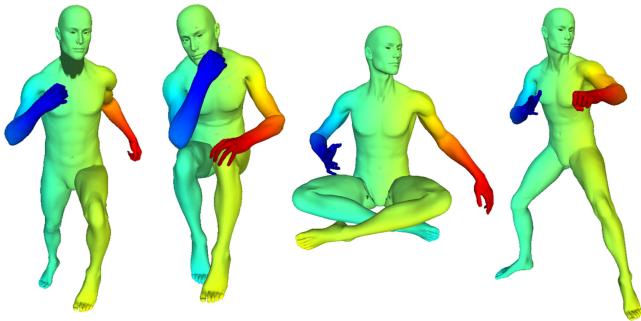


Fig. 18. Visualization of correspondence results on the TOSCA human test set. Corresponding points are shown in the same color.

prediction labels of their nearest points that are not identified as outlier points by Eq. (6). The refinement process can be formulated according to Eq. (7)

$$\pi_i = \arg \min_j d_{\gamma}(y_i, y_j), s.t. O_j = 0, \quad (7)$$

where point j is a point on input shape \mathcal{X} . Fig. 13 (right column) gives an example of our correspondence prediction results after refinement. As we can see from Fig. 13 (right column), our model can produce accurate and smooth correspondence predictions.

Furthermore, using the linear assignment strategy would increase the inference time. Based on our experiments on the FAUST dataset, by using linear instead of maximum assignment, our inference time increases from 6.6 s to 151.5 s for 20 shapes. Experiments were done on a single K80 GPU, and the batch size was set to 10. We solved the linear

assignment problem using MATLAB R2016b, with Intel (R) Xeon(R) CPU E5-2603.

5.3 Resilience to Connectivity Change

Robust determination of correspondence among 3D shapes should be obtained under inconsistent discretization schemes (i.e., different vertices) for the continuous 3D shape surface. While previous experiments are done on meshes with the same connectivity, to demonstrate that our model can learn high-level geometrical features instead of only the adjacency graph, we show some examples of the feature maps learned by our TC-NET model on different meshes from the FAUST test set (please see Fig. 14). We randomly pick one feature from the last convolutional layer and use it to visualize the same person in different poses. The consistent pattern of the feature maps across various poses shows that our TC-Net model successfully learns deformation-invariant geometric features. Moreover, we conduct experiments with a perturbed adjacency graph in which we randomly select 5 neighbors per node to construct the adjacency graph. Both our LLE construction module and the feature learning network take as input the perturbed adjacency graph for model training and evaluation. Fig. 15 shows the comparison between our model using the original adjacency graph and that using the perturbed adjacency graph. Note that we report the performance with maximum assignment for correspondence prediction, and we do not use refinement. As shown in Fig. 15, our model using the perturbed adjacency graph gets a correspondence accuracy of 89.6 percent on the FAUST test set. This demonstrates that even with perturbed adjacency

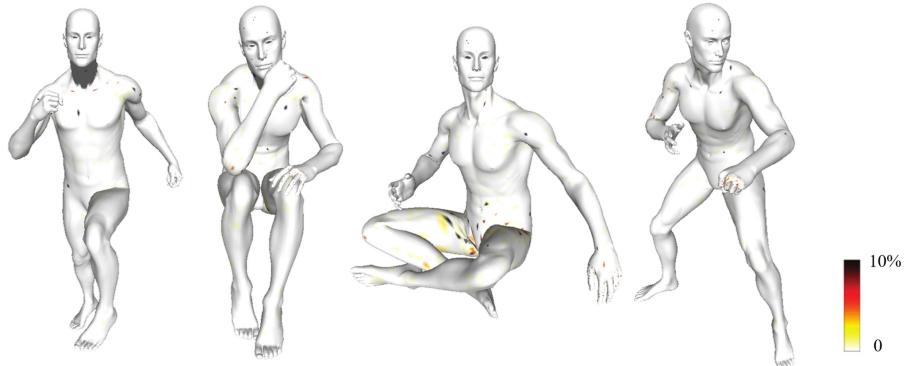


Fig. 19. Visualization of the geodesic error of the TC-Net model on the TOSCA human dataset. The error values are saturated at 10 percent of the geodesic diameter, which corresponds to approximately 15 cm. Warm colors represent large errors.



Fig. 20. Visualization of correspondence results on the SCAPE test set. Corresponding points are shown in the same color.

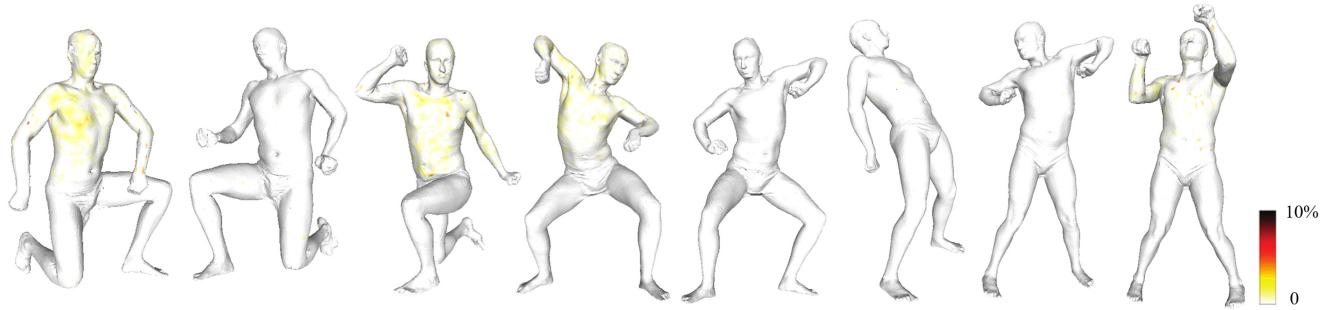


Fig. 21. Visualization of the geodesic error of the TC-Net model on the SCAPE test dataset. The error values are saturated at 10 percent of the geodesic diameter, which corresponds to approximately 15 cm. Warm colors represent large errors.

relations, our model maintains promising performance for correspondence prediction (our model achieves a correspondence accuracy of 94.0 percent when using the complete adjacency graph). This finding further demonstrates that our model is robust to perturbations in mesh connectivity.

5.4 Additional Visualizations

In this section, we visualize the correspondence results on the FAUST, TOSCA human and SCAPE test datasets in Figs. 16, 18 and 20 and the geodesic error maps in Figs. 17, 19 and 21, respectively. The correspondence results and geodesic error maps in these figure consistently demonstrate the effectiveness of our model for shape correspondence prediction.

6 CONCLUSION

In this paper, we propose a novel topology-constrained shape correspondence learning network framework that learns highly geometrically representative features on 3D point clouds by explicitly preserving the local topology. By imposing a topology-preserving reconstruction loss function as a regularization term on the feature layer, our network can enforce topological consistency between the feature space and the original 3D point space. Experiments on FAUST, SCAPE, and TOSCA benchmark datasets quantitatively and qualitatively demonstrate the effectiveness of the proposed model for shape correspondence learning.

ACKNOWLEDGMENTS

This work was supported by the NYU Abu Dhabi Institute (AD131). The authors contributed equally to this work.

REFERENCES

- [1] A. Myronenko, X. Song, and M. A. Carreira-Perpiñán, “Non-rigid point set registration: Coherent point drift,” in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2007, pp. 1009–1016.
- [2] B. Jian and B. C. Vemuri, “Robust point set registration using Gaussian mixture models,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1633–1645, Aug. 2011.
- [3] A. Myronenko and X. Song, “Image registration by minimization of residual complexity,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 49–56.
- [4] 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.
- [5] J. Ma, J. Zhao, J. Tian, Z. Tu, and A. L. Yuille, “Robust estimation of nonrigid transformation for point set registration,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2147–2154.
- [6] 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.
- [7] Y. Wu, B. Shen, and H. Ling, “Online robust image alignment via iterative convex optimization,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 1808–1814.
- [8] H. Ling and D. W. Jacobs, “Deformation invariant image matching,” in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, 2005, pp. 1466–1473.
- [9] X. Bai, L. J. Latecki, and W.-Y. Liu, “Skeleton pruning by contour partitioning with discrete curve evolution,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 449–462, Mar. 2007.
- [10] X. Bai and L. J. Latecki, “Path similarity skeleton graph matching,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1282–1292, Jul. 2008.
- [11] X. Li, L. Wang, and Y. Fang, “PC-Net: Unsupervised point correspondence learning with neural networks,” in *Proc. Int. Conf. 3D Vis.*, 2019, pp. 145–154.
- [12] O. Van Kaick, H. Zhang, G. Hamarneh, and D. Cohen-Or, “A survey on shape correspondence,” *Comput. Graph. Forum*, vol. 30, no. 6, pp. 1681–1707, 2011.
- [13] S. Belongie, J. Malik, and J. Puzicha, “Shape context: A new descriptor for shape matching and object recognition,” in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2001, pp. 831–837.
- [14] A. E. Johnson and M. Hebert, “Using spin images for efficient object recognition in cluttered 3D scenes,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 433–449, May 1999.
- [15] S. Manay, D. Cremers, B.-W. Hong, A. J. Yezzi, and S. Soatto, “Integral invariants for shape matching,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1602–1618, Oct. 2006.
- [16] J. Sun, M. Ovsjanikov, and L. Guibas, “A concise and provably informative multi-scale signature based on heat diffusion,” *Comput. Graph. Forum*, vol. 28, no. 5, pp. 1383–1392, 2009.
- [17] M. Aubry, U. Schlickewei, and D. Cremers, “The wave kernel signature: A quantum mechanical approach to shape analysis,” in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2011, pp. 1626–1633.
- [18] R. Litman and A. M. Bronstein, “Learning spectral descriptors for deformable shape correspondence,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 171–180, Jan. 2014.
- [19] J. Masci, D. Boscaini, M. Bronstein, and P. Vandergheynst, “Geodesic convolutional neural networks on Riemannian manifolds,” in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2015, pp. 37–45.
- [20] D. Boscaini, J. Masci, E. Rodolà, and M. Bronstein, “Learning shape correspondence with anisotropic convolutional neural networks,” in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 3189–3197.
- [21] S. Belongie, J. Malik, and J. Puzicha, “Shape matching and object recognition using shape contexts,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, 2002.
- [22] X. Li and I. Guskov, “Multiscale features for approximate alignment of point-based surfaces,” in *Proc. Symp. Geometry Process.*, 2005, Art. no. 217.
- [23] Y. Aflalo, A. Dubrovina, and R. Kimmel, “Spectral generalized multi-dimensional scaling,” *Int. J. Comput. Vis.*, vol. 118, no. 3, pp. 380–392, 2016.
- [24] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, “Generalized multidimensional scaling: A framework for isometry-invariant partial surface matching,” *Proc. Nat. Academy Sci. United States America*, vol. 103, no. 5, pp. 1168–1172, 2006.
- [25] M. Vestner *et al.*, “Efficient deformable shape correspondence via kernel matching,” in *Proc. Int. Conf. 3D Vis.*, 2017, pp. 517–526.

- [26] R. M. Rustamov, "Laplace-Beltrami eigenfunctions for deformation invariant shape representation," in *Proc. 5th Eurographics Symp. Geometry Process.*, 2007, pp. 225–233.
- [27] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond euclidean data," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 18–42, Jul. 2017.
- [28] F. Monti, D. Boscaini, J. Masci, E. Rodolà, J. Svoboda, and M. M. Bronstein, "Geometric deep learning on graphs and manifolds using mixture model CNNs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5115–5124.
- [29] N. Verma, E. Boyer, and J. Verbeek, "FeaStNet: Feature-steered graph convolutions for 3D shape analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 2598–2606.
- [30] 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.*, 2018, pp. 230–246.
- [31] M. Ovsjanikov, M. Ben-Chen, J. Solomon, A. Butscher, and L. Guibas, "Functional maps: A flexible representation of maps between shapes," *ACM Trans. Graph.*, vol. 31, no. 4, 2012, Art. no. 30.
- [32] O. Litany, T. Remez, E. Rodolà, A. Bronstein, and M. Bronstein, "Deep functional maps: Structured prediction for dense shape correspondence," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 5659–5667.
- [33] E. Corchado, J. M. Corchado, and J. Aiken, "IBR retrieval method based on topology preserving mappings," *J. Exp. Theor. Artif. Intell.*, vol. 16, no. 3, pp. 145–160, 2004.
- [34] A. Angelopoulou, J. G. Rodríguez, and A. Psarrou, "Learning 2D hand shapes using the topology preservation model GNG," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 313–324.
- [35] J. Si, S. Lin, and M.-A. Vuong, "Dynamic topology representing networks," *Neural Netw.*, vol. 13, no. 6, pp. 617–627, 2000.
- [36] C. Szepesvári, L. Balázs, and A. Lőrincz, "Topology learning solved by extended objects: A neural network model," *Neural Comput.*, vol. 6, no. 3, pp. 441–458, 1994.
- [37] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990.
- [38] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [39] L. K. Saul and S. T. Roweis, "An introduction to locally linear embedding," unpublished, 2000. [Online]. Available: <http://www.cs.toronto.edu/~roweis/lle/publications.html>
- [40] Z. Zhang and J. Wang, "MLLE: Modified locally linear embedding using multiple weights," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2007, pp. 1593–1600.
- [41] D. L. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data," *Proc. Nat. Academy Sci. United States America*, vol. 100, no. 10, pp. 5591–5596, 2003.
- [42] G. Te, W. Hu, Z. Guo, and A. Zheng, "RGCNN: Regularized graph CNN for point cloud segmentation," in *Proc. 26th ACM Int. Conf. Multimedia*, 2018, pp. 746–754.
- [43] D. Boscaini, J. Masci, E. Rodolà, M. M. Bronstein, and D. Cremers, "Anisotropic diffusion descriptors," *Comput. Graph. Forum*, vol. 35, no. 2, pp. 431–441, 2016.
- [44] T. Sandler, J. Blitzer, P. P. Talukdar, and L. H. Ungar, "Regularized learning with networks of features," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2009, pp. 1401–1408.
- [45] I. S. Dhillon, Y. Guan, and B. Kulis, "Weighted graph cuts without eigenvectors a multilevel approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 11, pp. 1944–1957, Nov. 2007.
- [46] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 3844–3852.
- [47] F. Bogo, J. Romero, M. Loper, and M. J. Black, "FAUST: Dataset and evaluation for 3D mesh registration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3794–3801.
- [48] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, *Numerical Geometry of Non-Rigid Shapes*. Berlin, Germany: Springer, 2008.
- [49] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis, "SCAPE: Shape completion and animation of people," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 408–416, 2005.
- [50] V. G. Kim, Y. Lipman, and T. Funkhouser, "Blended intrinsic maps," *ACM Trans. Graph.*, vol. 30, no. 4, 2011, Art. no. 79.
- [51] M. Vestner, R. Litman, E. Rodolà, A. Bronstein, and D. Cremers, "Product manifold filter: Non-rigid shape correspondence via kernel density estimation in the product space," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3327–3336.
- [52] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 77–85.
- [53] D. Boscaini, J. Masci, S. Melzi, M. M. Bronstein, U. Castellani, and P. Vandergheynst, "Learning class-specific descriptors for deformable shapes using localized spectral convolutional networks," *Comput. Graph. Forum*, vol. 34, no. 5, pp. 13–23, 2015.



Xiang Li received the BS degree in remote sensing science and technology from Wuhan University, Wuhan, China, in 2014, and the PhD degree in cartography and GIS from the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China, in 2019. He is currently a postdoctoral associate with the Department of Electrical and Computer Engineering, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates. His research interests include deep learning, computer vision, and remote sensing image recognition.



Congcong Wen received the BS degree in geographic information systems from the China University of Petroleum, Shandong, China, in 2016. He is currently working toward the PhD degree with the Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing, China. His research interests include remote sensing image recognition, smart city data mining, deep learning, and computer vision.



Lingjing Wang received the BS degree from Moscow State University, Moscow, Russia, in 2011, and the PhD degree from the Courant Institute of Mathematical Science, New York University, New York, in 2019. He is currently a postdoctoral associate with the Department of Electrical and Computer Engineering, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates. His research interests include deep learning and 3D visual computing.



Yi Fang received the BS and MS degrees in biomedical engineering from Xi'an Jiaotong University, Xi'an, China, in 2003 and 2006, respectively, and the PhD degree in mechanical engineering from Purdue University, West Lafayette, Indiana, in 2011. He is currently an assistant professor with the Department of Electrical and Computer Engineering, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates. His research interests include three-dimensional computer vision and pattern recognition, large-scale visual computing, deep visual computing, deep cross-domain and cross-modality multimedia analysis, and computational structural biology.

▷ For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/csl.