Relation Constrained Capsule Graph Neural Networks for Non-Rigid Shape Correspondence

YUANFENG LIAN, SHOUSHUANG PEI, and MENGQI CHEN, Department of Computer Science and Technology, China University of Petroleum, Beijing, China
JING HUA, Department of Computer Science, Wayne State University, Detroit, MI, USA

Non-rigid 3D shape correspondence aims to establish dense correspondences between two non-rigidly deformed 3D shapes. However, the variability and symmetry of non-rigid shapes usually lead to mismatches due to shape deformation, topological changes, or data with severe noise. To finding an accurate correspondence between 3D dynamic shapes for the local deformation complexity, this article proposes a Relation Constrained Capsule Graph Network (RC-CGNet), which combines global and local features by encouraging the relation constraints between the embedding feature space and the input shape space based on the functional maps framework. Specifically, we design a Diffusion Graph Attention Network (DGANet) to segment the surface into parts with correct edge boundary between two regions. The Minimum Spanning Tree (MST) of geodesic curves among the singularities obtained from the segmented parts is added as relation constraints, which can compute isometric correspondences in both direct and symmetric directions. Besides that, the relation-and-attention constrained neural networks are designed to learn the shape correspondence via attention-aware CapsNet and functional maps under relation constraints. To improve the convergence speed and matching accuracy, we propose an optimized residual network structure based on the Nesterov Accelerated Gradient (NAG) to extract local features, and use graph convolution structure to extract global features. Moreover, a lightweight Gated Attention Module (GAM) is designed to fuse global and local features to obtain a richer feature representation. Since the capsule network has better spatial reasoning ability than the traditional convolutional neural network, our novel network architecture is a dual-route capsule network based on Routing Attention Fusion Block (RAFB), filtering low-discriminative capsules from a holistic view by exploiting geometric hierarchical relationships of semantic parts. Experiments on open datasets show that our method has excellent accuracy and wide adaptability.

CCS Concepts: • **Computing methodologies → Computer graphics; Artificial intelligence**;

Additional Key Words and Phrases: Shape correspondence, attention mechanism, capsule network, graph convolution network

This work was partially supported by the grants: NSFC 61972353, NSF IIS-1816511 and OAC-1910469.
Authors' Contact Information: Yuanfeng Lian (corresponding author), Department of Computer Science and Technology, China University of Petroleum, Beijing, China; e-mail: lianyuanfeng@cup.edu.cn; Shoushuang Pei, Department of Computer Science and Technology, China University of Petroleum, Beijing, China; e-mail: 2020211254@student.cup.edu.cn; Mengqi Chen, Department of Computer Science and Technology, China University of Petroleum, Beijing, China; e-mail: 2021211256@student.cup.edu.cn; Jing Hua, Department of Computer Science, Wayne State University, Detroit, MI, USA; e-mail: jinghua@wayne.edu.

ACM 2157-6912/2024/8-ART0
https://doi.org/10.1145/3688851

ACM Reference format:
Yuanfeng Lian, Shoushuang Pei, Mengqi Chen, and Jing Hua. 2024. Relation Constrained Capsule Graph Neural Networks for Non-Rigid Shape Correspondence. *ACM Trans. Intell. Syst. Technol.* 0, 0, Article 0 (August 2024), 26 pages.
https://doi.org/10.1145/3688851

1 Introduction

The shape correspondence problem is one of the important research topics for scientific visualization, computer graphics, computer vision, shape registration, etc. Considering the deformations of non-rigid shape, it is very difficult to find the correspondence between the shapes before and after deformation due to the low data resolution and high amounts of sensor noise [43], and there are so many variables required to define the dense mapping. Even though some progresses [3, 8, 14, 18, 22, 33, 45, 54, 56, 59] have been made, the task of finding dense shape correspondence is still very challenging.

Given mesh data as an input, the non-rigid shape correspondence problem can be summarized as establishing a point-wise matching by measuring the similarity of the point descriptors of two shapes. Traditional approaches [10, 47, 51, 53] based on distortion minimization between 3D shapes are suitable for enforcing various constraints on mesh sets. However, the features extracted by such methods may produce a dense correspondence with poor subjectivity. Another kind of matching methods attempt to utilize the local search for investigating the invariance of shapes in a parameterized domain or functional space with fewer degrees of freedom by exploring some specified geometric structures, such as conformal geometry [25, 48, 55], isometry [44, 46, 47], spectral quantities [5, 21, 33], etc. Recent efforts mainly focus on functional map frameworks [35, 38–41] that are soft non-rigid transformations. Methods in this family are mostly based on special functions of the respective shapes and formulating function preservation constraints to a linear system of equations. Despite the flexibility of the framework, these methods ignore the relation constraints between models, and the isometric attribute mapping results for deformed shapes may result in severe mismatches.

In this article, we present **Relation Constrained Capsule Graph Network (RC-CGNet)** to jointly train two tasks of networks for shape segmentation and correspondence. Relational constraints can effectively maintain consistent structural relationships between different parts of the shape throughout the matching process. These constraints ensure that the relative positions and connections with other parts of the shape remain unchanged when the shape is stretched or compressed. Moreover, different from existing methods, such as MGCN [57] and CA-CGNet [24], the energy diffusion layers are designed to extract the spatial diffusion features with stronger semantic information from energy convolution to segment the surface into parts with correct edges boundary between two regions. Figure 1 illustrates the flowchart of our proposed RC-CGNet, which is composed of five main stages. The first stage carries out **diffusion graph attention network (DGANet)** to segment the whole surface into a few parts. The second stage selects landmarks of segmented parts on surface. In this stage, **Farthest Point Sampling (FPS)** is adopted to extract high-level point features and take high-density point as segment centers. The third stage computes the **Minimum Spanning Tree (MST)** of geodesic curves among the landmarks. The fourth stage enforces the relation constraints in the functional maps framework, which can compute isometric correspondences in both direct and symmetric directions. In the last stage, the regularization term of relation constrained loss function is formulated by measuring the similarity between the embedding features of

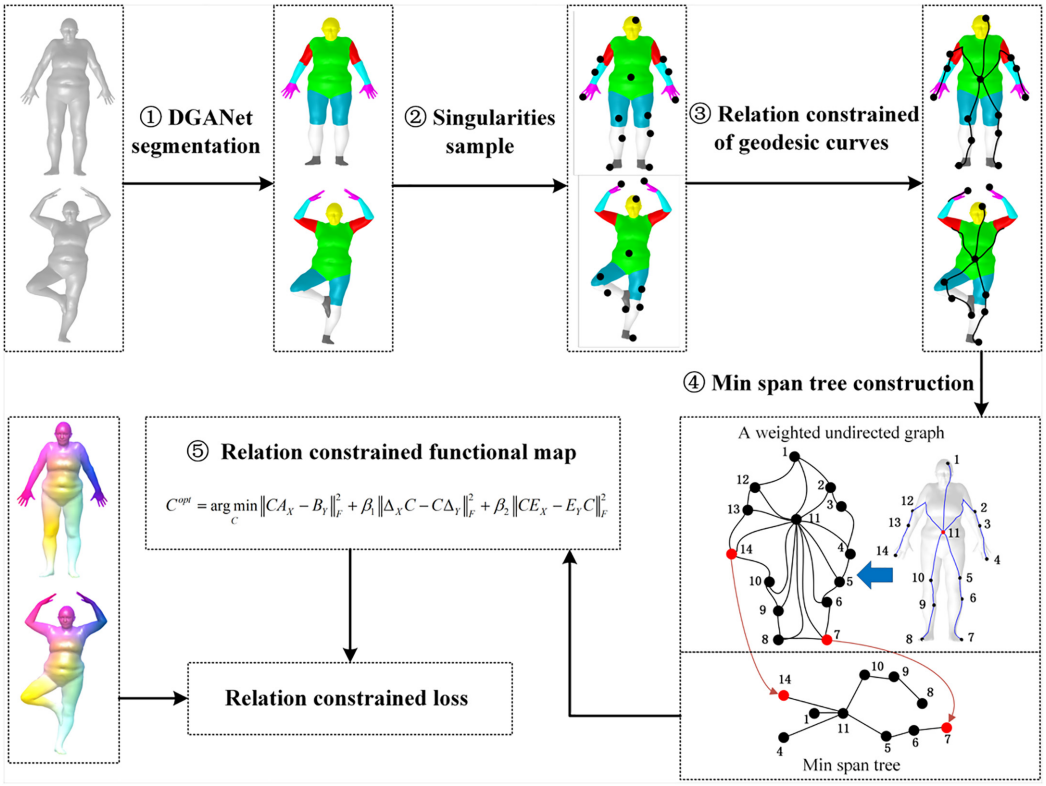

Fig. 1. The framework of our proposed network Relation Constrained Capsule Graph Network (RC-CGNet). Given an input of descriptors and the Laplacian eigenfunctions from a pair of shapes, our model first segments the whole surface into a few parts (step 1). Then, Farthest Point Sampling (FPS) is adopted to extract high-level point features and take high-density point as segment center (step 2). A Minimum Spanning Tree (MST) is constructed based on geodesic curves among the landmarks (step 3). The MST of geodesic curves as relation constraints is applied to the functional maps (step 4). Finally, the relation constrained loss as a regularization term is integrated into the loss function during shape correspondence learning (step 5).

each point in the MST of geodesic curves. Accordingly, our main contributions can be summarized as follows:

- —We present a DGANet composed of multiple **diffusion graph attention blocks (DGAB)**, which operate on the surfaces of scalar values throughout, with each DGAB diffusing the features and fusing vertex energy features through a graph attention mechanism.
- —By adding the MST of geodesic curves as relation constraints in the functional maps framework, the feature learning network is integrated relation constrained loss as a regularization term, which further boosts the shape correspondence learning performance.
- —The RC-CGNet combined global and local features by using attention feature fusion block to enhance the primary capsules is proposed to improve the feature expressiveness by RAFB for higher classification accuracy of capsules.
- —Experiments on accuracy and generalization with different datasets have shown great improvements when integrating the proposed RC-CGNet and the DGANet in the uniquely designed architecture.

2 Related Work

2.1 Shape Descriptors

Non-rigid shape matching includes traditional methods and deep learning methods. Traditional 3D shape matching methods based on statistical features find a point-wise matching between the points on two or more shapes, and many scholars have studied in depth on the basis of feature descriptors. Tombari et al. [52] developed **Signature of Histograms of Orientations (SHOT)** descriptors with rotation and translational invariance by accumulating the normal angles of the key and neighboring points in the neighborhood space. Guo et al. [13] proposed the Rotational Projection Statistics descriptor by rotationally projecting neighboring points onto 2D planes and calculating a set of statistics. Aubry et al. [2] used random forests to classify shapes based on their **Wave Kernel Signature (WKS)** features to solve the shape matching problem and achieve good results. Sun et al. [51] proposed the **Heat Kernel Signature (HKS)** as a point signature based on the fundamental solutions of the heat equation (heat kernels). Wang et al. [57] used graph wavelets to decompose the Dirichlet energy on a surface based on their Wavelet Energy Decomposition Signature. While hand-crafted feature extraction can provide valuable features in certain circumstances, it often performs sub-optimally in learning-based models. In contrast, using deep learning techniques for feature extraction and processing can significantly improve the model's predictive accuracy and generalization ability. More recently, researchers have tried to enhance the representation power of descriptors with deep neural networks supervised by human-annotated correspondence labels [17, 50, 58, 60]. Among these methods, shape descriptors can be learned from a large-scale dataset using deep neural networks to learn intrinsic geometric properties.

2.2 Correspondence with Functional Maps

Instead of using point-wise descriptors for point-to-point matching, functional maps treat the overall shape correspondences as linear compositions between spaces of functions on manifolds [35]. This idea has been extended significantly in [7, 11, 34] recently. Using a spectral basis, the framework of functional map allows to easily incorporate constraints to reach desirable properties, including multi-scale spectral manifold wavelets preservation [15], continuous and orientation-preserving correspondences [39], product preservation [34, 41]. With the advent of deep learning technique, deep functional maps have shown that learned features can improve performance to alleviate this dependence of the initial choice of descriptor functions. Litany [26] designed a structured prediction model called FMNet in the space of functional maps along with deep residual networks to obtain a representation of shape correspondence. Roufosse et al. [43] proposed the SURFMNet model by using an unsupervised loss that enforces the desired structural properties on the resulting map. Halimi et al. [14] introduced the unsupervised correspondence learning approach for deformable 3D shapes by replacing the ground truth requirement with a different geometric criterion based on pair-wise distance. In [11], self-supervised dense mapping between non-isometric shapes was shown to provide a compact representation of shape correspondence. Hu et al. [16] constructed a pointwise mapping by using optimal transport and subsequently convert it into an initial functional mapping. In [28], an efficient optimization algorithm called SmoothFM was proposed to achieve high-quality results in non-equidistant shape correspondence, which used Dirichlet energy to smooth the point-to-point mapping. However, these methods do not fully consider the loss of mesh details, and the corresponding results are usually mismatched when the density of surface mesh vertices changes.

2.3 Capsule Graph Neural Networks

To overcome the disadvantage of CapsNet and a **Graph Neural Network (GNN)** rarely considering local and global relation for feature learning, the **Capsule Graph Neural Networks (CGNNs)** [61] were first introduced to extract the features of the nodes from GNN to generate high-quality graph embeddings. Feng et al. [9] proposed a dual-routing CGNN to explore local-global relations and to preserve detailed properties. In [20], CGNNs were adopted to learn graph capsules for encoding underlying characteristics from the heterogeneous knowledge graph. Some recent work [62, 63] presented hierarchical graph capsule network to jointly learn node embeddings and extract the hierarchical structure of the input graph for preserving detailed graph information. Later in [23], Li et al. introduced a new capsule network with graph routing for learning both relationships, where capsules in each layer were treated as the nodes of a graph. However, the key difference between the methods mentioned above and our proposed approach lies in the fact that our method uses AFM to fuse global features with local features for enhancing the primary capsules. Also, our method aggregates features from dual-routing by RAFB to filter low-discriminative capsules from a holistic view.

3 Our Methods

3.1 Overview

In this work, we develop a novel end-to-end deep neural network named RC-CGNet upon the functional maps framework. The motivation is to leverage the CGNN with attention mechanism for a powerful understanding of the shape correspondence by combining both global and local spectral features to the functional maps with the relation constraints from the learned shape semantic tree representation. The RC-CGNet architecture is specifically designed to incorporate relation constraints, allowing it to effectively learn and represent the complex inter-dependencies present in the shape, leading to improved performance in shape correspondence. The basic pipeline can be described by two major modules, i.e., the geometry-aware semantic feature representation and the relation-and-attention constrained neural networks. The entire network architecture is given in Figure 2.

One of the challenges we address in this article is to find the shape correspondences between Riemannian manifolds of the local deformation complexity with intrinsic symmetry. Although functional maps can solve correspondence between models by constructing linear optimization of functional maps matrices, it cannot describe the geometric relationship between feature points on the source/target model, which leads to local mismatches between the deformed regions, such as hands and legs of the intra-class/inter-class regions in intrinsically symmetric shape model. Moreover, to effectively enforce adding geodesic relation constraints in the functional maps framework, it is necessary to obtain the accurate classification results of the 3D model surfaces. DGANet can obtain accurate segmentation results of 3D models and enhance the expression ability of regularization constraints in the minimization energy function in the functional maps matrix with the support of accurate classification information.

To take advantage of capsnet in shape correspondence for reducing the loss of spatial direction information, we propose an attention-aware capsnet module integrated with **Descriptor Extraction Network (DENet)**, in which the GAM module is designed to learn the fused feature from global feature generated by GCN and local feature generated by NAG-ResNet. Furthermore, we present a DGANet to segment the surface of the 3D mesh models through the diffusion of heat kernel and graph attention mechanism. The point features are refined and the deep feature representation is generated, which effectively improves the discriminative ability and robustness of the algorithm. Since using the functional maps to search the matching points of entire manifold

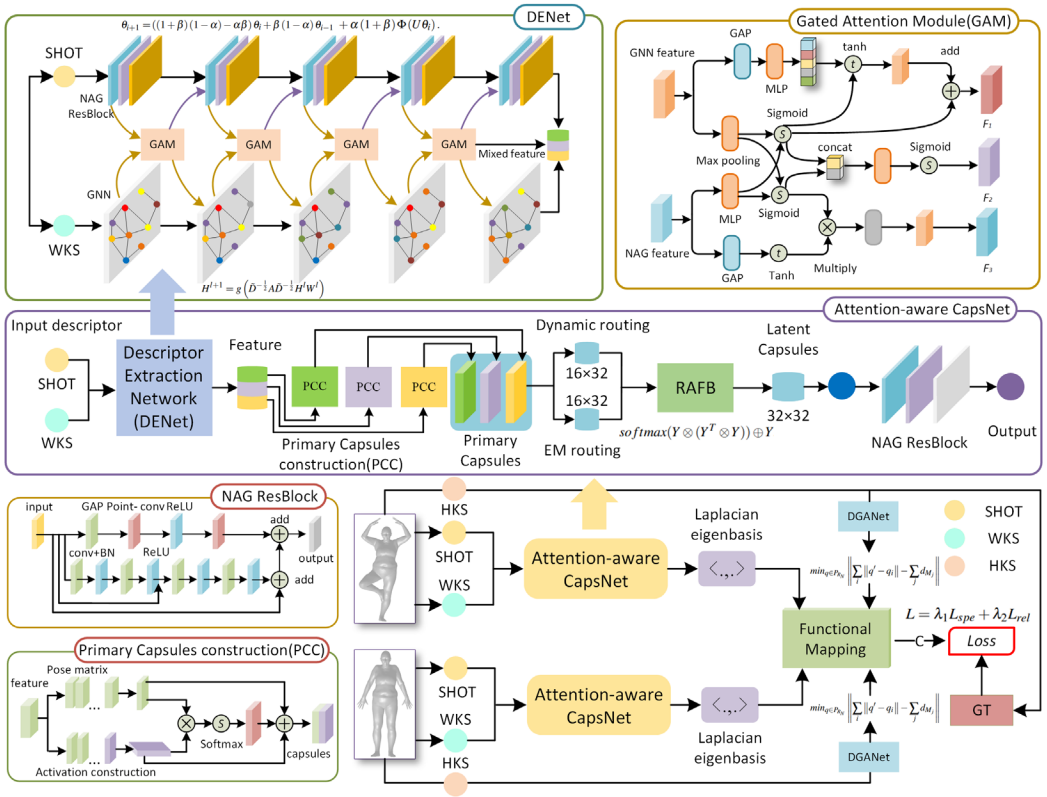

Fig. 2. Detailed structure of our proposed network RC-CGNet. For surface segmentation, a DGANet is applied to segment surface into several parts. Our model takes the Signature of Histograms of Orientations (SHOT) descriptor for NAG-ResNet and WKS descriptor for Graph Neural Network (GNN) with the Laplacian eigenfunctions from a pair of shapes as input. To produce the spectral representations, the Gated Attention Module (GAM) is constructed to extract the global and local attention feature. Then, the dual-routing (a sigmoid dynamic routing and an attention EM routing) is fused via the Routing Attention Fusion Block (RAFB). The relation constrained functional maps are represented by Equation (20) and the relation constrained loss is derived from Equation (22). The two NAG-ResNet respectively share weights for the two sets of data.

may lead to correspondence noise, we propose a shape correspondence method based on relation constraints. The geodesic distance constraint is used as the regularization term of the functional maps model, and the relation constraint loss function is used to train the model, which can enhance the convergence effect of the model and improve the shape matching accuracy. Essentially, our method uniquely combines attention-aware capsnet and DGANet with functional maps aiming at overcoming the aforementioned problems.

3.2 Geometry-Aware Semantic Feature Representation

In this section, we propose a new geometry-aware semantic feature representation for effective and efficient shape correspondence learning, i.e., a shape semantic tree. It includes two components, i.e., shape semantic learning and semantic tree construction.

3.2.1 Shape Semantic Learning. For the purpose of improving the semantic segmentation accuracy of 3D models, based on graph attention mechanism and heat kernel diffusion theory, we

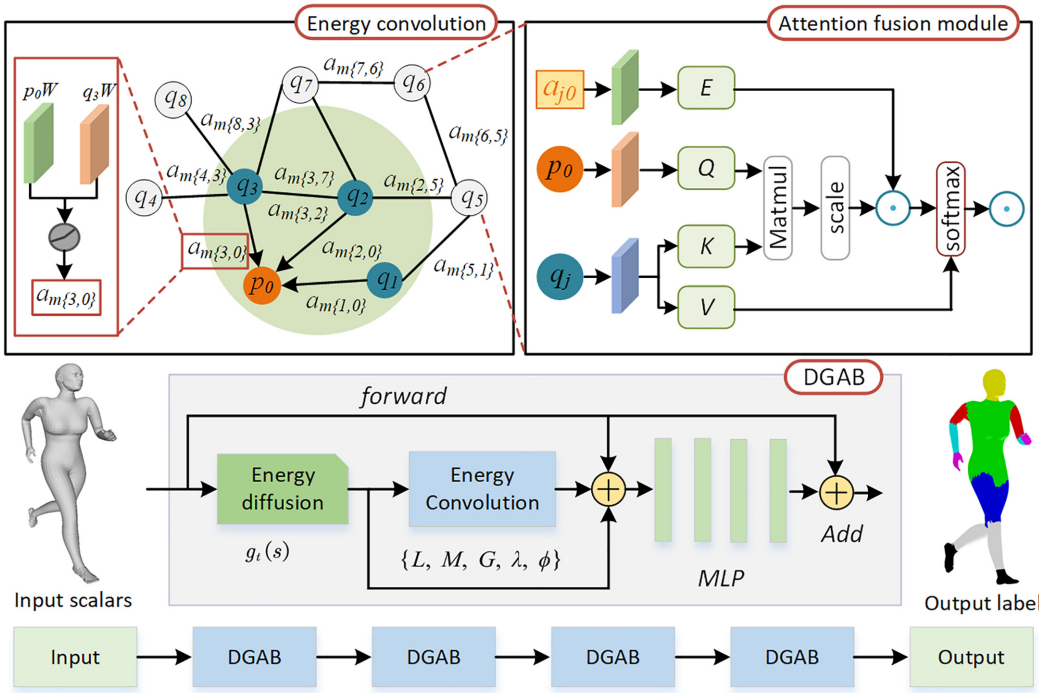

Fig. 3. DGANet is composed of successive identical DGAB. Each block diffuses features over the learned timescale and then aggregates the features by energy convolution. The attention scores are calculated and applied to the features of the nodes and the edges. The self-attention mechanism is shown in detail to aggregated graph attention to generate a normalized attention score a_{i0} for the edge from node q_i to node p.

Table 1. Definition of Main Symbols

Symbol	Meaning	Symbol	Meaning
L	Laplace matrix	M	Mass matrix
G	Spatial gradient matrix	λ	Eigenvalues
q_i	Vertices of the mesh	W	Weight matrix
a_m	Energy convolution score	Ψ	Stacked matrix of eigenvectors
E	Column vector of eigenvalues	TS	Tangent space
X	Tangent vector	e	Basis vector
σ_{leaky}	LeakyReLU function	$\mathcal{P}$	Parallel transport map
$\tilde{D}$	Diagonal node degree matrix	A	Adjacency matrix

develop a DGANet to learn the semantic segmentation of the input mesh. It combines the graph attention network with the fusion features, which take HKS descriptors as inputs. The detailed network architecture of DGANet is given in Figure 3, and Table 1 shows the meaning of the main notations of the model. First, to obtain the spatial diffusion feature, we discretize a surface S by a triangle mesh with vertices V. For every time scale t, the energy diffusion layer is inspired by [49]:

$$g_t(s) = \Psi E \odot (\Psi^T M s), \tag{1}$$

where Ψ denotes the stacked matrix of eigenvectors, E is the column vector made up of $e^{-\lambda_k t}$, and M is the mass matrix. Then, the feature vector of each vertex $p, q \in V$ is then mapped into the tangent space, respectively. Let X_p and X_q be the tangent vectors of $g_t(p)$ and $g_t(q)$ for the two tangent spaces T_pS and T_qS, the angle between two vectors can be expressed as:

$$\theta_{pq} = arccos\left(\frac{X_p \cdot X_q}{\|X_p\| \cdot \|X_q\|}\right). \tag{2}$$

We use the parallel transport map $\mathcal{P}_{pq} : T_qS \rightarrow T_pS$ to translate vector X_q from tangent space T_qS to T_pS, which is defined as:

$$\mathcal{P}_{qp}\left(X_q\right) = e^{i\theta_{pq}} X_p, \tag{3}$$

where $e \in \mathbb{C}$ represents the basis vector. Taking vertex p as the center point, the features in its k-neighborhood are aggregated by the attention mechanism. Similar to [29], the energy convolution score $a_{m(p,q)}$ of the edge (q, p) in a mesh is normalized across all neighboring adjacent vertices $q \in \mathcal{N}_{mp}$ by using softmax function, and the energy convolution function can be defined as:

$$a_{m(p,q)} = softmax(e_{m(p,q)}) = \frac{\exp(e_{m(p,q)})}{\sum_{k \in \mathcal{N}_{mp}} \exp(e_{m(k,q)})}, \tag{4}$$

where $e_{m(p,q)}$ is energy convolution score coefficient, which is calculated by:

$$e_{m(p,q)} = \sigma_{leaky}([p_m W || q_m W]_{q \in \mathcal{N}_{mp}}), \tag{5}$$

where $||$ denotes the concatenation operation, W is the weight for the linear transformation, σ_{leaky} refers to the LeakyReLU function.

In order to utilize the potential of the attention score $a_{m(p,q)}$, we aggregate them alongside the features of the neighboring nodes $\{q_i | q \in \mathcal{N}_p\}$ by using the similar principle of self-attention. The final feature representation $X'_p \in \mathbb{R}^{f'_\eta}$ of the node p can be computed as:

$$X'_p = \sigma\left(\sum\nolimits_{q \in \mathcal{N}_{mp}} softmax\left(\frac{W^a X_p W^\beta \mathcal{P}_{q_i p}}{\sqrt{d_k}} a_{q_i p}\right) W^\varepsilon \mathcal{P}_{q_i p}\right) || X_p, \tag{6}$$

where W^a, W^β and $W^\varepsilon \in \mathbb{R}^{f'_\eta \times f_\eta}$ are learnable parameters, respectively. d_k is the dimension of p, $\sigma(.)$ is an activation function.

In order to enhance feature integration, the filter is considered on the complex plane, we have:

$$\left(X' * b\right)(p) = \int_S e^{i(\theta_{pq}+\phi_q)} b\left(e^{i\theta_{pq}} e^{i(\theta_{pq}-\phi_q)}\right) dq, \tag{7}$$

where b is a filter, $\phi_q = arg(X_q)$ is the angle formed with the real axis.

3.2.2 Semantic Tree Construction. The DGANet is adapted to segment the given 3D non-rigid shape model into different semantic parts. For the convenience of representation, as shown in Figure 4, the output $S = \{s_0, s_1, \cdots, s_T\}$ represents the segmented position set, where T is the number of segmentation parts. In each shape, we select the point with FPS in the parts $R = \{s_H, s_{RH}, s_{LH}, s_{RF}, s_{LF}\}$ of head, right hand, left hand, **right foot (RF)** and left foot as the reference point to get the singularity set $P_R = \{p_H, p_{RH}, p_{LH}, p_{RF}, p_{LF}\}$.

In the remaining segmentation region of the source model M, we select the point with the highest density as the singularity of the interval. For each vertex $p' \in S_M - R_M$, its density can be expressed as:

$$N_{p'} = |\{x \in s_i \mid d\left(p', x\right) \leq \mu\}|, \tag{8}$$

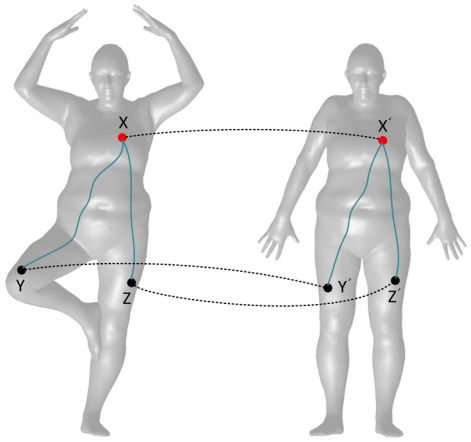

Fig. 4. One example of the semantic tree construction on the source model (left) and the target model (right) in FAUST dataset.

where $d(v, x)$ denotes the geodesic distance from point p' to point x, μ is the threshold. Further, we extract the geodesic distance between the point p' and the reference point $p \in P_{R_M}$ with $d(p', p)$ to construct geodesic distance set $D_M = \{d_{M_1}, d_{M_2}, d_{M_3}, d_{M_4}, d_{M_5}\}$. We can obtain the corresponding vertex $q' \in S_N - P_N$ to the same segmentation part in the target model N by optimizing the geodesic distance constraint:

$$min_{q \in P_{R_N}} \left\| \sum_i \|q' - q_i\| - \sum_j d_{M_j} \right\|. \tag{9}$$

By repeating the above steps, corresponding singularities of the segmentation part of the original model M can be obtained successively in the target model N.

After that a weighted connected indirect graph $G = (V, E)$ is constructed, connecting each singularity in set $P_R \cup P$ with geodesic, and the MST on each model is calculated by solving it with Prim's algorithm. Then, we can represent two shape models by the corresponding computed semantic trees. Figure 4 shows the semantic tree construction process on the source and target models, and Algorithm 1 gives the complete computational steps.

3.3 Relation-and-Attention Constrained Neural Networks

In this section, we introduce the new relation-and-attention constrained neural networks to learn the shape correspondence via attention-aware CapsNet and functional maps with relation constraints from the learned shape semantic tree representations.

3.3.1 Attention-Aware CapsNet. The proposed attention-aware CapsNet has the following four major components, i.e., NAG-ResNet, GNN, **gated attention module (GAM)** and **routing attention fusion block (RAFB)**. The local features of SHOT operator are extracted by NAG-ResNet, and the global features of WKS operator are extracted by GNN. The feature extraction module is composed of five modules, respectively, and the features extracted from each module are processed by GAM module to enhance the integration. Then, a dual-routing is fused via the RAFB. The detailed network architecture is shown in Figure 2.

NAG-ResNet. The SHOT descriptor is one of the inputs in the feature extraction network. The local features of SHOT operator are extracted by NAG-ResNet. Inspired by Li et al. [19], an optimized residual block is designed based on the **Nesterov Accelerated Gradient (NAG)** descent algorithm

Algorithm 1: Algorithm of Semantic Tree Construction

Input: The segmented sets S_M and N_M of shapes M and N
Output: Corresponding semantic trees of shapes M and N

Initialize $R \leftarrow \emptyset, P_R \leftarrow \emptyset, D \leftarrow \emptyset, P \leftarrow \emptyset$
begin
  // *Calculate reference point in each shape*
  for *each shape* **do**
    Construct the set R
    Select the vertex with FPS in each part to form the initial point and join the set P_R
  end
  // *Choose singularities of remaining parts*
  for s_i *in set* $S_M - R_M$ **do**
    Compute the vertex p' with the highest density value in this part using Equation (8)
    Update the point set $P_M \leftarrow P_M \cup \{p'\}$
    // *Calculate geodesic distance from the reference point*
    for $p \in P_{R_M}$ **do**
      Construct geodesic distance set $D_M \leftarrow D_M \cup \{d(p', p)\}$
    end
    Obtain the corresponding point q' on the target model N by using distance constraint Equation (9)
    Update the point set $P_N \leftarrow P_N \cup \{q'\}$
  end
  Obtain a weighted connected indirect graph $G = (V, E)$ by connecting each vertex in set $P_R \cup P$ with a geodesic and solve it by least squares method
  Use Prim's algorithm to obtain the minimum spanning tree
  return *Corresponding semantic tree of shapes M and N*
end

(Figure 2). In the following, we provide proof of the relation between the neural network structure and the NAG optimization algorithm and then describe the derivation process of the new network structure. NAG can be described as:

$$\begin{aligned} y_{k+1} &= x_k + \beta\,(x_k - x_{k-1}), \\ x_{k+1} &= y_{k+1} - \alpha \nabla f\,(y_{k+1}). \end{aligned} \tag{10}$$

Lemma 1. *Suppose that during the propagation of the neural network, the transmission of signal from the first layer to the last layer is expressed as* $\theta_{i+1} = \Phi(U_i\theta_i)$. *Assume that* U *is a symmetric positive definite matrix. Set* $V = \sqrt{U}$, *then, there is a function* $f(\xi)$ *makes the propagation of the neural network equivalent to optimize* $F(\xi) = f(V\xi)$*:*

(1) Define a new function $\Psi(\xi)$, which satisfies $\Psi'(\xi) = \Phi(\xi)$;
(2) Use NAG to optimize $f(\xi)$;
(3) Obtain $\theta_0, \theta_1, ..., \theta_k$ from $\xi_0, \xi_1, ..., \xi_k$ through $\xi = V\theta$

From Lemma 1, the expression can be obtained:

$$\nabla \sum_i \Psi\left(V_j^T \xi\right) = U\Phi\left(U^T\theta\right) = U\Phi(U\theta). \tag{11}$$

Proof. The optimization function $f(\xi)$ is defined as:

$$f(\xi) = \frac{\left\|\xi^2\right\|}{2} - \sum_i \Psi\left(V_j^T \xi\right), \tag{12}$$

where V_j is the jth column of the matrix V. Taking the derivative of both sides of $f(\xi)$, we get:

$$\nabla f(\xi_i) = \xi_i - V\Phi(V\xi_i). \tag{13}$$

The NAG formula can be simplified as an equivalent form of:

$$\xi_{i+1} = \xi_i + \beta(\xi_i - \xi_{i-1}) - \alpha((1+\beta)\nabla f(\xi_i) - \beta\nabla f(\xi_{i-1})). \tag{14}$$

Use the iteration of the NAG algorithm to minimize $f(\xi)$ and recover θ from $\theta = V^{-1}\xi$, we have:

$$\theta_{i+1} = ((1+\beta)(1-\alpha) - \alpha\beta)\theta_i + \beta(1-\alpha)\theta_{i-1} + \alpha(1+\beta)\Phi(U\theta_i). \tag{15}$$

In the derivation, $\Phi(U\theta_i)$ denoted as the ith layer feed-forward network, $\Phi(.)$ is a nonlinear transformation defined by an activation function, U can be regarded as a convolution operation. The whole process of the block is divided into two channels: one is input to the point-wise processing module through global average pooling, and the other is directly input to the point-wise processing module, which includes point-wise convolution and ReLU activation function. The features processed by the point-wise module are calculated and ReLU activated, and then the multiply operation is performed with the initial input features. Finally, the output is cascaded with the extracted features.

GNN. WKS descriptor is the other of the inputs in the feature extraction network. The global features of WKS operator are extracted by GNN. In order to comprehensively leverage the relation among points, the GNN takes the point-level features as nodes. The GNN updates the representation of each node by aggregating information of its neighbors. By definition, a weighted directed GNN is represented as $G = (V, A)$, where $V = (v_1, v_2, \cdots, v_m) \in \mathbb{R}^{M\times d}$ is the set of feature representation of shape, M is the number of points. $A \in \mathbb{R}^{M\times M}$ is the adjacency matrix, which can be learned as follows:

$$A_{ij} = \cot\left(a_{ij}\right) + \cot\left(b_{ij}\right), \tag{16}$$

where A_{ij} represents the correlation between points v_i and v_j, a_{ij}, and b_{ij} are the angles to the edge (i, j), respectively.

In this work, GNN is built upon in an end-to-end manner, which is formulated as:

$$H^{l+1} = g\left(\tilde{D}^{-\frac{1}{2}} A \tilde{D}^{-\frac{1}{2}} H^l W^l\right), \tag{17}$$

where g is the message passing function that includes convolutional layers and BatchNorm layers. $H^l \in \mathbb{R}^{M\times d^l}$ represents features of points on the layer l, and d^l represents the dimension of feature, $W^l \in \mathbb{R}^{d^l \times d^{l+1}}$ is a trainable weight matrix, $\tilde{D}$ is a diagonal node degree matrix. The initialization input of g is H^0 $(l = 0)$, where $H^0 = V$.

DENet. The proposed DENet is dedicated to fuse features from SHOT descriptors and WKS descriptor with the GAM, which consists of two branches to retain the intrinsic feature information while reducing the possibility of noise allocation for shape matching. The inputs of the GAM module are GNN features and NAG features as shown in Figure 2. GAM uses attention mechanisms to weigh the importance of nodes within a neighborhood, which can extract more nuanced features from individual nodes and their relational context.

For GNN features, global average pooling and **Multi-Layer Perceptron (MLP)** are firstly used to process the features to obtain the dynamic combination of input features, which are activated by tanh function and finally cascaded with the features of NAG features activated by sigmoid function

to obtain feature F_1. For the NAG feature, the MLP is used to process the feature, and sigmoid function is used to activate it. Then the global average pooling operation is carried out on the NAG feature. The two groups of features are fused by matrix multiplicative operation, and the MLP is used to process the feature dimension to obtain the output feature F_3. In order to better fuse NAG and GNN features for the feature extraction network, GNN features are firstly maximized, and then NAG features are processed by MLP, and the two features are activated by sigmoid function, and the fused features are obtained after cascading. Finally, the fusion output feature F_2 is obtained through sigmoid activation function again, as shown in Figure 2.

RAFB. The learning features generated by feature extraction module are transformed into N capsules, and each capsule consists of pose matrix and activation value [27]. First, we convert each of the 3-channel feature into a 64-dimension feature map by a convolution layer and reshape the feature map as the pose matrix of the capsules. Second, the pose matrix is transposed and multiplied the activation value, which is computed by a convolution layer for the learning features as input. In the end, we use the softmax function to activate feature map (32*32*4*4), and cascade the 3-channel activated features with the pose matrix and activation value to form the primary capsule. Here, Dynamic routing and EM routing are used to obtain the output of primary capsules, and then the attention mechanism is used to further process advanced capsules to obtain latent capsules with better feature distribution. Finally, the output features of capsule network are obtained through multi-layer NAG ResBlock.

To improve the classification accuracy of capsules,we have proposed a dual-routing program, including sigmoid dynamic routing and EM routing, and cascaded the output feature information to obtain enhanced feature information. The output of self-attention module is denoted as $X_T \in \mathbb{R}^{C\times H\times W}$, and let the input feature be $X \in \mathbb{R}^{C\times H\times W}$. We first feed it into three linear layers separately and reshape the generated features to $K \in \mathbb{R}^{C\times N}$ and $V \in \mathbb{R}^{C\times N}$, where $N = H \times W$. The $Q \in \mathbb{R}^{N\times C}$ is obtained by performing an additional transpose operation. Then we perform a matrix multiplication between Q and K and apply the softmax function to compute as:

$$X_T = softmax(V \otimes (Q^T \otimes K)) \oplus X. \tag{18}$$

The output feature map of channel attention module is denoted as $Y_T \in \mathbb{R}^{C\times H}$ and the input feature $Y \in \mathbb{R}^{C\times H\times W}$ is multiplied by itself twice and then added to obtain the output feature. The expression of channel attention module is:

$$Y_T = softmax(Y \otimes (Y^T \otimes Y)) \oplus Y. \tag{19}$$

The sigmoid routing and EM routing are processed by spatial attention and channel attention, respectively, and enhanced features are obtained by AFEM module.

3.3.2 Functional Maps with Relation Constraints. The functional maps with relation constraints are used to search matching points across the whole manifold, which leads to a high computational complexity for some inaccurate matching. To alleviate this issue, an optimal functional maps $\mathbf{C}_{opt}$ by the least squares method can be expressed as:

$$\mathbf{C}_{opt} = \arg\min_{\mathrm{C}} \|\mathbf{C}\mathbf{A}_X - \mathbf{B}_Y\|_F^2 + \beta_1 \|\Delta_X \mathbf{C} - \mathbf{C}\Delta_Y\|_F^2 + \beta_2 \sum_{i=1}^{N_l-1} \left\|\mathbf{C}\mathbf{E}_{X_i} - \mathbf{E}_{Y_i}\mathbf{C}\right\|_F^2, \tag{20}$$

where $\mathbf{A}_X$ and $\mathbf{B}_Y$ are the point-wise descriptor coefficient matrices. Δ_X and Δ_Y are the eigenvalues of Laplacian-Beltrami operator, N_l is the number of landmarks. $\mathbf{E}_{X_i}$ and $\mathbf{E}_{Y_i}$ are the geodesic distance matrices, and β_1, β_2 are the weights. A detailed explanation of the first and second terms of Equation (20) can be referred to [15]. Theoretically, the geodesic curve pairs between the corresponding point pairs of the isometric mapping are equidistantly mapped, as shown in Lemma 2.

Lemma 2. *If two-dimensional manifolds S_1 and S_2 have an isometric mapping relation $F : S_1 \rightarrow S_2$, the corresponding MST of geodesic lines T_1 on S_1 under F is T_2. Let $\delta(t) : [v_i, v_j] \rightarrow S_1$ is the shortest geodesic curve of the MST on S_1 between vertex v_i and v_j. Let $\epsilon(t) : [v_i', v_j'] \rightarrow S_2$ is the shortest geodesic curve of the MST on S_2 between vertex v_i' and v_j'. If $v_i' = F(v_i)$, $v_j' = F(v_j)$, we get $F(\delta(t)) = \epsilon(t), \forall t \in [0, 1]$.*

Proof. Since F is isometric matching and $\delta(t)$ is the shortest geodesic curve between v_i and v_j, $F(\delta(t))$ is the shortest geodesic curve between v_i' and v_j' on S_2, $F(\delta(t))$ and $\epsilon(t)$ are the same curve for S_2 because the shortest geodesic curve between two locations is unique.

If the parameter t is sampled uniformly, the set of corresponding points on the geodesic curves $\{p_1, p_2, \cdots, p_n\}$ and $\{q_1, q_2, \cdots, q_n\}$ can be obtained, where (p_i, q_i) are the corresponding pairs of equidistant alignment, which indicate their correspondence under equidistant alignment. As a result, more sampled pairs of matching points can be introduced as geometric restrictions to the function mapping.

3.3.3 Relation Constrained Loss Function. To improve the robustness of relation constrained functional maps, we integrate the relation constraints as a regularization term discussed in Section 3.3.1. Given the domain mapping matrix $\mathbf{C} = \{\mathbf{C}_{spe}, \mathbf{C}_{rel}\}$, we adopt a spectral mapping and relational mapping with $\mathbf{C}_{spe}$ and $\mathbf{C}_{rel}$. The ground truth point-wise correspondence is defined as $\mathbf{C}_{gt}$.

A point-to-point map retains the local area if and only if the functional map is orthonormal [35]. Thus, the spectral loss function L_{spe} and relational loss function L_{rel} are defined as:

$$\begin{aligned} L_{spe} &= \left\|\mathbf{C}_{gt} - \mathbf{C}_{spe}\right\|_F^2 + \left\|\mathbf{C}_{spe}^T\mathbf{C}_{spe} - \mathbf{I}\right\|_F^2 \\ L_{rel} &= \left\|\mathbf{C}_{gt} - \mathbf{C}_{rel}\right\|_F^2 + \left\|\mathbf{C}_{rel}^T\mathbf{C}_{rel} - \mathbf{I}\right\|_F^2. \end{aligned} \tag{21}$$

The overall loss function L of shape corresponding is defined as follows:

$$L = \lambda_1 L_{spe} + \lambda_2 L_{rel}, \tag{22}$$

where λ_1 and λ_2 are the hyper-parameters. In our experiments, we set both to be 0.5.

4 Experiments and Discussion

The experimental environment of this algorithm is based on Linux Ubuntu 16.04 operating system, with GeForce GTX 1080Ti GPU processor as hardware support. Extensive experiments are performed on five publicly available non-rigid datasets, FAUST [4], TOSCA [6], SCAPE [1], KIDS [42], and SMAL [66]. All network models are trained only using the FAUST training dataset (the first 80 manifolds). Furthermore, we conduct experiments with raw and resampled data (4,096 points) on the FAUST testing dataset (the last 20 manifolds), and sample all other datasets to 4,096 points for testing the model. The Adam algorithm is used as the optimizer for network training, and the learning rate is changed through a polynomial learning strategy, where the initial learning rate is set as 1e-3, the end learning rate as 1e-5, and the batch size as 1.

4.1 Segmentation

In order to show the accuracy of semantic segmentation of DGANet proposed in this article, we use the composite dataset [31] of human body parts as training sets and test the model on Shrec dataset. Additionally, we cite a variety of reported results from other approaches on this task. According to Table 2, our approach achieves the best performance, with a 0.20% improvement over the original DiffusionNet [49] due to the introduction of graph attention neural network.

Table 2. Mesh Segmentation Accuracy on the Human Body Dataset

Method	Accuracy
GCNN [32]	86.4%
PointNet++ [37]	90.8%
Toric Cover [31]	88.0%
MDGCNN [36]	88.6%
CGConv [64]	89.9%
DiffusionNet-xyz [49]	90.6%
DiffusionNet-hks [49]	91.7%
Ours-hks	**91.9**%

Compared with several existing methods, the proposed method has higher segmentation accuracy.

Fig. 5. Segmentation results of the proposed method on the KIDS and TOSCA dataset. The first row shows the segmentation results on the KIDS dataset, and the second row shows the segmentation results on the TOSCA dataset.

Further, Figure 5 demonstrates the segmentation performance of our model on both the KIDS and TOSCA dataset, which achieves accurate segmentation predictions. Moreover, we visualize the semantic segmentation results, as shown in Figure 6. By magnifying local detail comparison, it is evident that DGANet outperforms DiffusionNet in terms of semantic segmentation accuracy, with more precise and accurate details. The results highlight the effectiveness of our method in achieving high precision in edge segmentation and producing better visual effects on human shapes as input.

Additionally, we test the performance of DGANet on non-human shapes, and the segmentation results are shown in Figure 7. Through observation, our model with energy diffusion and energy convolution can generate accurate segmentation results for most of the vertices on test shapes. We select the models of the hippocampus for the human brain as testing sets to verify the effectiveness of our semantic segmentation method. Figure 8 visualizes the semantic segmentation results of our DAGNet and DiffusionNet on the hippocampus dataset. Our method has better segmentation results and higher accuracy for hippocampus of different shapes.

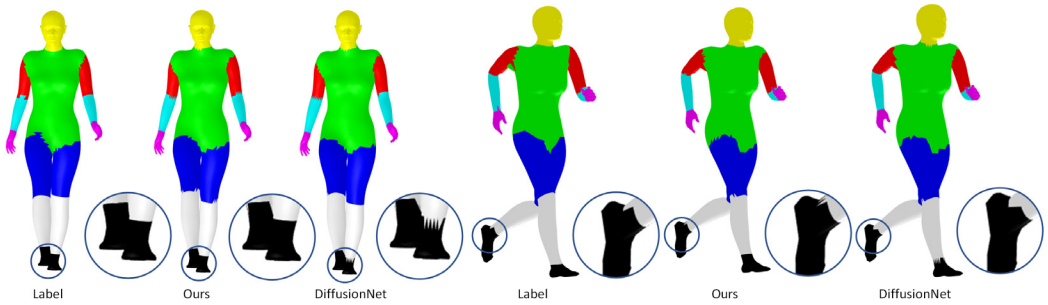

Fig. 6. Segmentation results from the human body dataset. DGANet correctly classified all body parts and gave more accurate boundaries.

Fig. 7. The segmentation results of the proposed method on the animation character models. (a) cat, (b) dog, (c) centaur, and (d) horse.

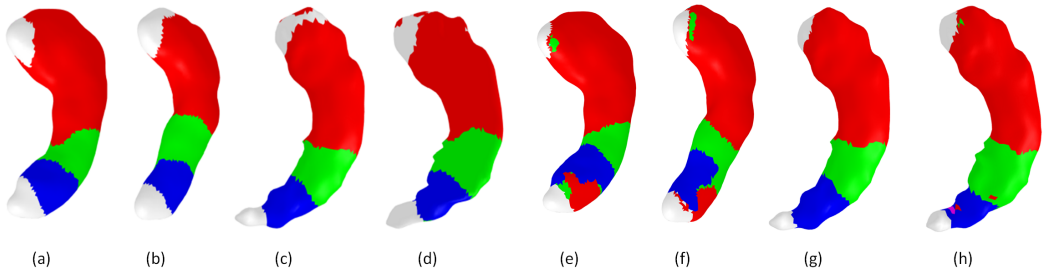

Fig. 8. Segmentation results of the hippocampus dataset. (a)–(d) show the segmentation results of the proposed method, (e)–(h) show the segmentation results of DiffusionNet.

4.2 Shape Matching

After segmentation, we obtain N sampling points by calculating the Gaussian curvature of each point in the region and then traverse the target shape to obtain the optimal corresponding point. Computing the geodesic distance between each point and using the MST algorithm to connect all nodes of a network, the total length of the edge can be minimized. With the process of our approach, we employ intra-subject pairs from FAUST dataset and create 4,096 points on the surface. In Figure 9(a) and (b) show the segmented surfaces by DGANet and the singularities by FPS, respectively. (c) and (d) show the results of the singularities using Vintescu et al.'s and geodesic curves between these points. Then, (e) represents the MST of geodesic curves.

Figures 10 and 11 show the corresponding results of reference shapes and deformable shapes on the original and resampled FAUST dataset. Corresponding points on the two shapes are represented by the same color. We display the matching error of the proposed method on FAUST dataset, as shown in Figure 12. The average geodesic error threshold is set to 0.3, and the error is visualized

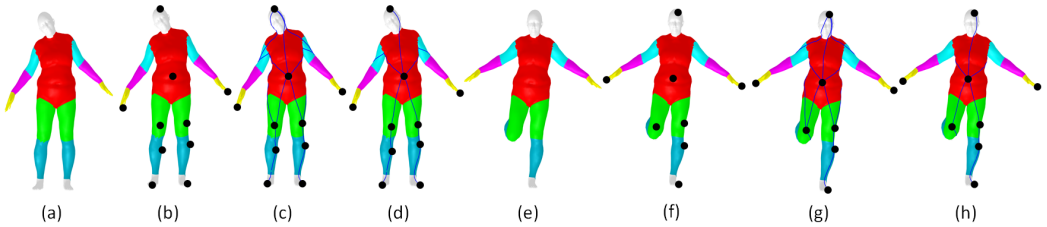

Fig. 9. The process of our method on the FAUST dataset. (a) presents the shape segmented by DGANet, (b) shows the singularity of the region, (c) shows the result of Vintescu et al.'s approach, (d) indicates the connection point of the geodesic curves, and (e) represents the MST of geodesic curves.

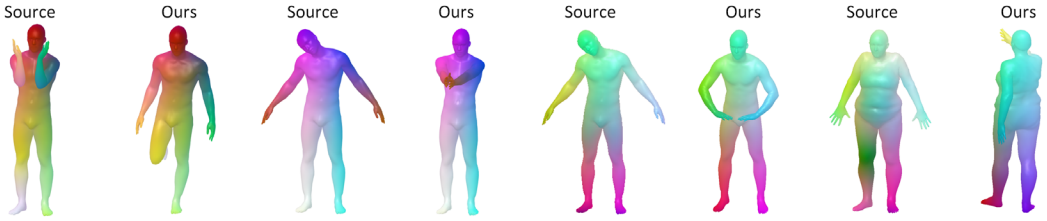

Fig. 10. Visualization of the shape matching result using our method on the original Faust dataset (6,890 points). For each pair of shapes, the left is the reference shape and the right is the target shape with correspondence.

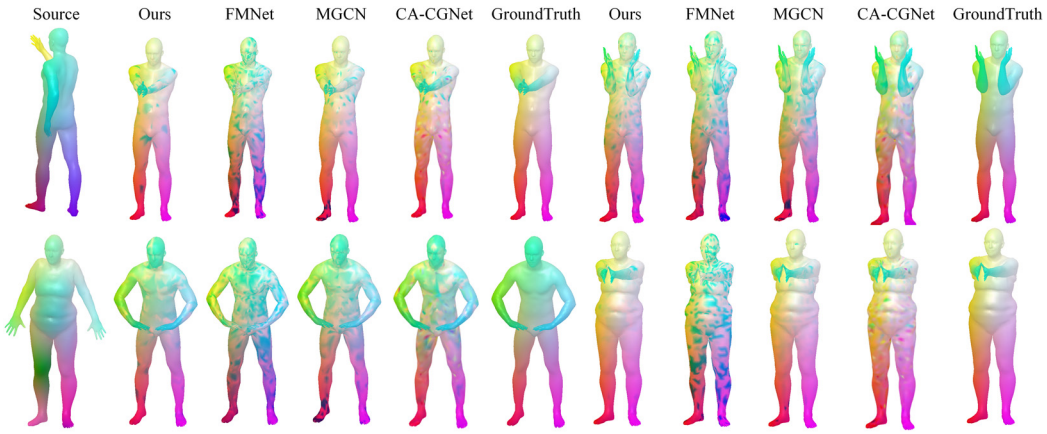

Fig. 11. Shape correspondences learned using different methods on the resampled FAUST dataset. Each column shows the output matching from the source mesh to the target meshes on the top. Corresponding positions share the same color.

and compared with FMNet method. It can be found that the proposed method has lower matching error in FAUST dataset and better shape mapping accuracy than the FMNet method.

Considering the accuracy of quantitatively evaluating our method, Table 3 compares the experimental results of the proposed approach with the other seven methods. The evaluation index is **average error (AE)**. In the intra-FAUST test set, the error of RC-CGNet is reduced by 25.0% compared with FMNet and 24.3% compared with MGCN. In inter-FAUST test set, the error of RC-CGNet was reduced by 4.51% compared with 3D-CODED and 11.52% compared with SP.

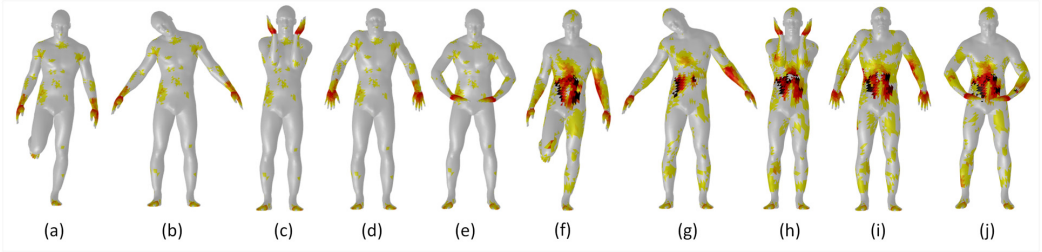

Fig. 12. Geodesic error visualization of the proposed method on the FAUST dataset. (a)–(e) show our results, and (f)–(j) show the results of the FMNet. Hot colors correspond to large errors.

Table 3. Average Error Comparison on FAUST Dataset Using Several Methods

Method	Intra AE ↓	Inter AE ↓	Mean ↓	Standard Deviation ↓
RF [42]	15.04	17.05	16.04	1.42
FMNet [26]	2.44	4.83	3.63	1.69
FARM [30]	2.81	4.12	3.46	0.92
MGCN [57]	2.51	3.65	3.08	0.81
SURFMNet [43]	1.73	3.63	2.68	1.34
3D-CODED [12]	1.98	2.88	2.43	0.64
SP [65]	**1.57**	3.13	2.35	1.10
Ours	1.90	**2.75**	**2.27**	**0.60**

Intra means the same subject with different poses, while Inter means a different subject.

The geodesic error is visualized in Figure 13. It shows the geodesic error index comparison of the test results of RC-CGNet and other corresponding methods on three non-rigid shape TOSCA, SCAPE, and KIDS datasets, where lines (a), (b) and (c) are representative examples tested on SCAPE, TOSCA, and KIDS datasets, respectively. It manifests that the method in this article achieves better generalization performance than the other three methods and proves its effectiveness in these examples.

Figure 14 shows the cumulative geodesic error percentage curve, where (a) and (b) are geodesic errors on 6,890 vertices, respectively, and (c) and (d) are geodesic errors on 4,096 vertices resampled. With the increase of threshold, the matching accuracy gradually improves. Compared with FMNet, Unsupervised FMNet and CA-CGNet, the geodesic error of this model on 6,890 vertices is slightly better than that of other methods, and the geodesic error on resampling dataset is also improved than that of other methods, indicating that this model has stronger robustness on lower resolutions of models.

Here, the model trained on the FAUST training set is used to test the generalization ability of the network on other datasets. In Table 4, the proposed method is quantitatively evaluated by comparing the mean geodesic errors matched on different test set shape pairs. In FAUST(4,096 points) dataset, the error of our RC-CGNet is 79.90% less than FMNet, 51.72% less than RF, 50.33% less than MGCN and 45.59% less than CA-CGNet. In SCAPE dataset, it is 47.41% less than FMNet, 22.66% less than RF, 31.77% less than MGCN and 2.39% less than CA-CGNet. In KIDS dataset, it is 70.45% less than FMNet, 74.08% less than RF, 66.86% less than MGCN and 21.42% less than CA-CGNet. In TOSCA dataset, it is reduced by 45.34% compared with FMNet, 46.60% compared

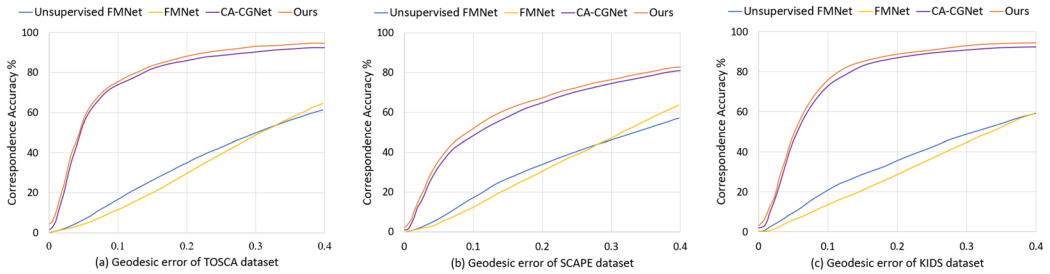

Fig. 13. Quantitative performance of point-wise correspondences of different methods on the TOSCA, SCAPE, and KIDS non-rigid shape dataset. (a) shows geodesic error of the TOSCA dataset, (b) shows geodesic error of the SCAPE dataset, and (c) shows geodesic error of the KIDS dataset.

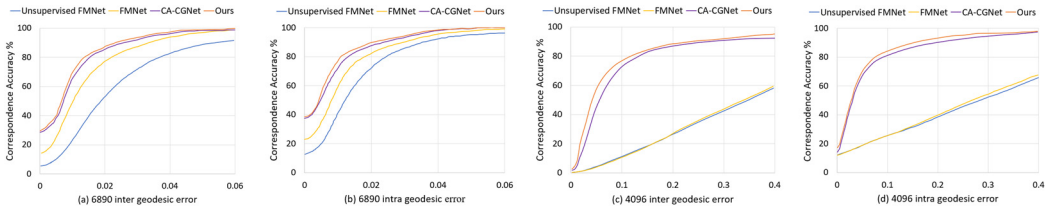

Fig. 14. Quantitative performance of point-to-point correspondences under different configurations on the FAUST dataset. (a) and (b) represent results on the raw dataset, (c) and (d) represent results on the resampled dataset. It shows our method is robust to a low sampling.

Table 4. The Comparison with Some State-of-the-Art Approaches on FAUST, SCAPE, KIDS, and TOSCA Datasets with Eight Object Categories

Method	FAUST	SCAPE	KIDS	TOSCA			Mean	Standard Deviation
				David	Michael	Victoria		
FMNet	0.3413	0.3712	0.3601	0.3100	0.3752	0.3574	0.3525	**0.0240**
RF	0.1421	0.2321	0.4106	0.3156	0.3800	0.3716	0.3087	0.1030
MGCN	0.1381	0.2631	0.3211	0.2787	0.3153	0.3015	0.2696	0.0681
CA-CGNet	0.1261	0.1839	0.1354	0.1937	0.2115	0.1928	0.1739	0.0347
Ours	**0.0686**	**0.1795**	**0.1064**	**0.1878**	**0.2014**	**0.1807**	**0.1541**	0.0535

The datasets are resampled to 4,096 vertices. Metric is the mean geodesic error (cm).

with RF, 36.36% compared with MGCN and 4.70% compared with CA-CGNet. Experimental results show that our RC-CGNet has a good generalization on these datasets and is superior to the methods in the previous chapter, such as FMNet and CA-CGNet.

By comparison with MGCN and FMNet, the proposed method in this article has better matching results on the TOSCA dataset of low-resolution grids, as shown in Figure 15. To better demonstrate the comparison, Figure 16 shows the matching results of the proposed method, CA-CGNet [24], MGCN [57], and FMNet [26] on the resample KIDS and SCAPE dataset with 4,096 vertices, and the matching errors are shown by staining model. It can be seen that on the resample KIDS and SCAPE test set, the performance of our approach is better than the other three methods, indicating that the proposed method has stronger robustness.

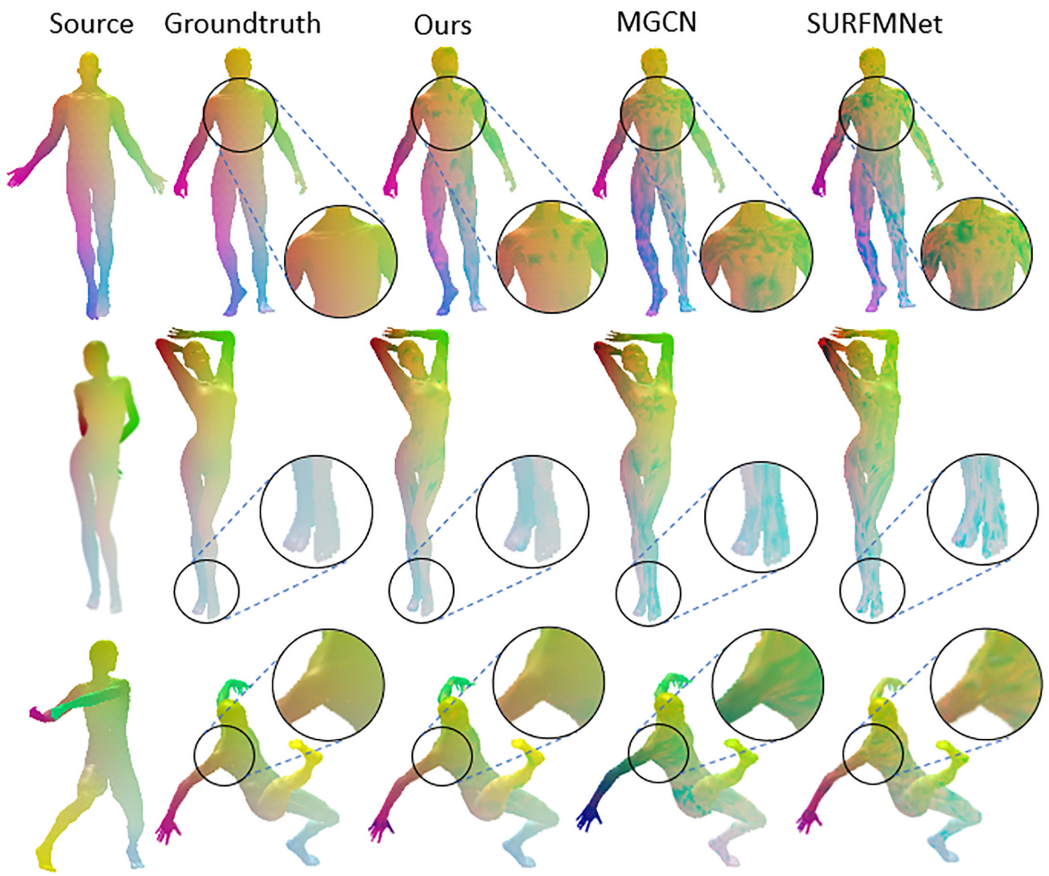

Fig. 15. Visualization of the proposed method correspondence results on the TOSCA datasets. Corresponding positions share the same color.

Furthermore, we select several different kinds of non-human shapes on resample TOSCA dataset to test the performance of our model, the correspondence results are shown in Figure 17. Compared with state-of-the-art learning-based approaches for deformable shape correspondence, the correspondence results from our proposed method are closer to the Groundtruth and have better visual effects for these non-human shapes. On the non-human models, our method also achieves better performance compared to the learning-based approaches. As shown in Figure 18, our proposed method corresponds to minimal geodesic errors and can significantly generate accurate corresponding labels for most vertices on the non-human shapes. Figure 19 visualizes the geodesic error of our method, CA-CGNet and MGCN on resampled TOSCA dataset with a color scale of 0–0.4. The experimental results show that our method achieves high-quality matching on both human models and non-human models with large postural variations, thus demonstrating better generalization of RC-CGNet.

We further test RC-CGNet on the resampled SMAL dataset, which is a non-isometric dataset. In the SMAL dataset, the volume of the object changes before and after deformation, that is, the surface may undergo tensile compression deformation. The corresponding results of different methods on the small dataset are shown in Figure 20. We see that even in this challenging example with strong non-isometric distortions, our approach produces a very close matching effect for incorporating

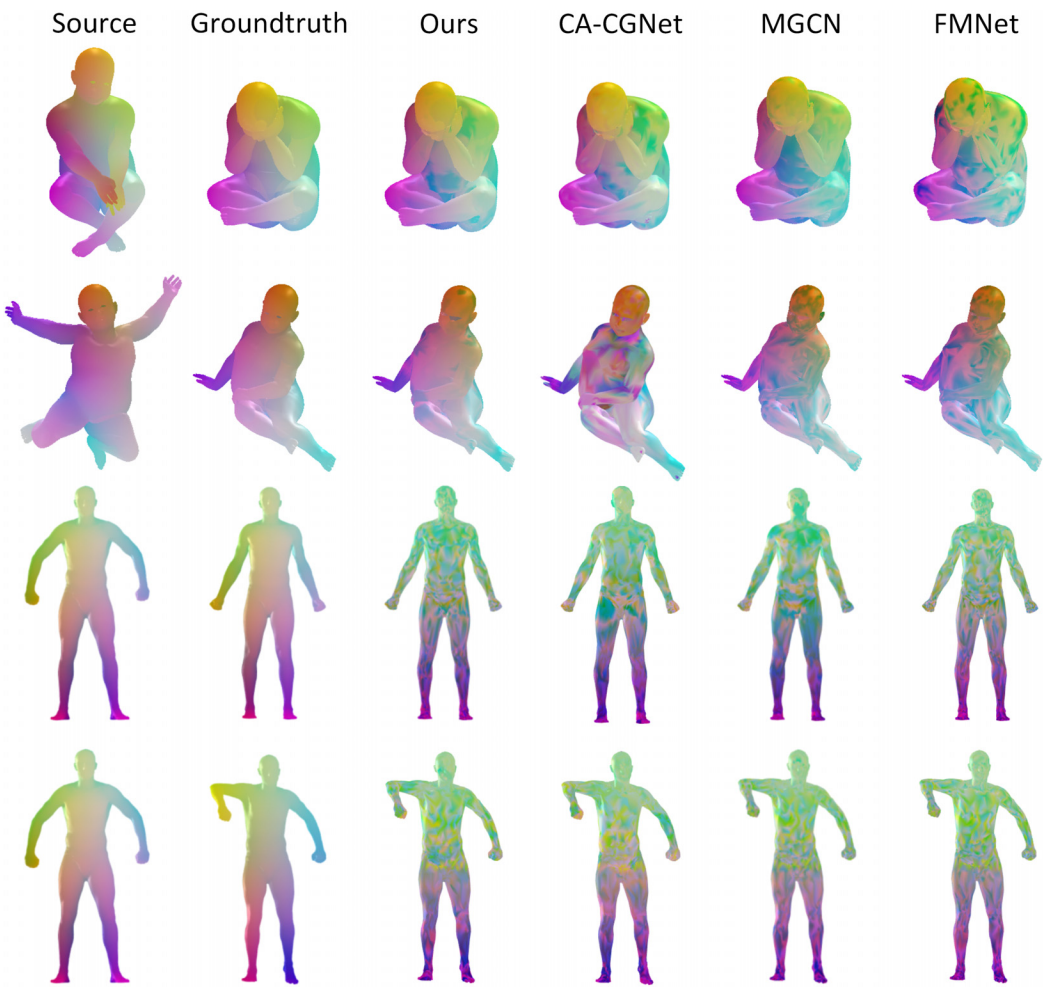

Fig. 16. Visualization of the proposed method correspondence results on the resample KIDS and SCAPE datasets. Each row shows the output matching from the source mesh to the target meshes on the right. Corresponding positions share the same color.

geometric methods and making use of semantic trees instead of triangulating details to generate well-oriented maps.

4.3 Ablation Experiments

4.3.1 Effectiveness RAFB. In Section 3.3.1, the dual-route fusion mechanism is adopted to process the master capsule. In order to verify the effectiveness of the dual-route fusion algorithm, EM routing and sigmoid routing are used to replace dual-route fusion, respectively. Train the network separately and test the evaluation indicators in multiple datasets, as shown in Table 5. Through observation, only EM routing has the highest AE of 0.1874 in all datasets. The AE of sigmoid routing algorithm is 0.1943, slightly lower than that of EM routing algorithm. The fusion of the two routing mechanisms achieves the best performance in most datasets with an AE of 0.1813, which is 3.25% less than that of sigmoid routing only and 6.69% less than that of EM routing.

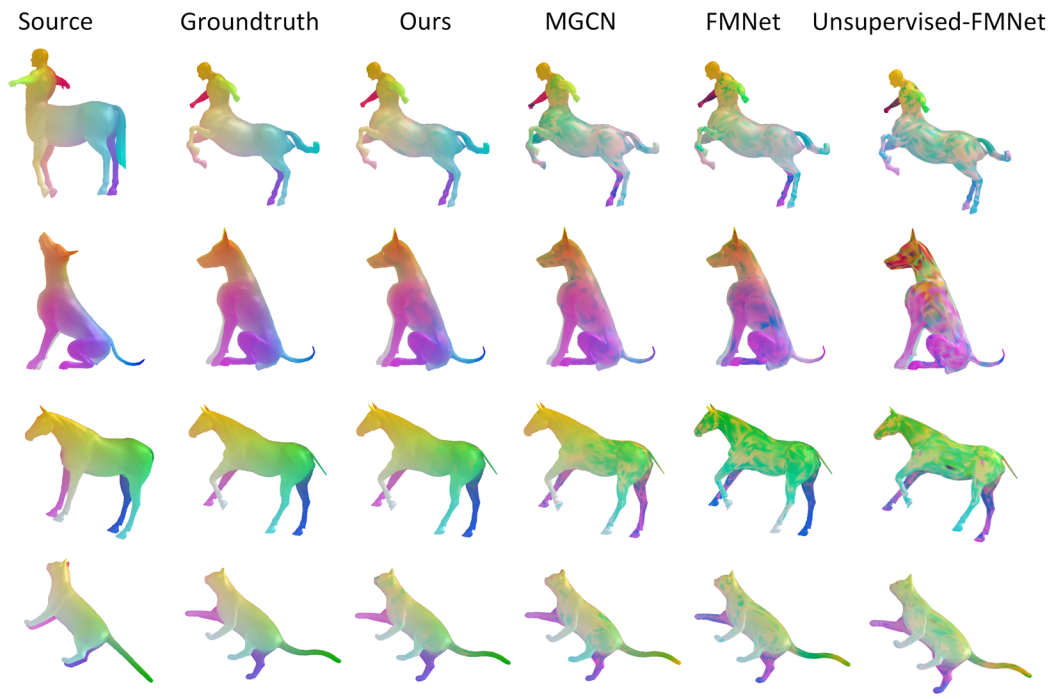

Fig. 17. Visualization of the correspondence results on the non-human shapes computed with our proposed method. Each row shows the output matching from the source mesh to the target meshes on the right.

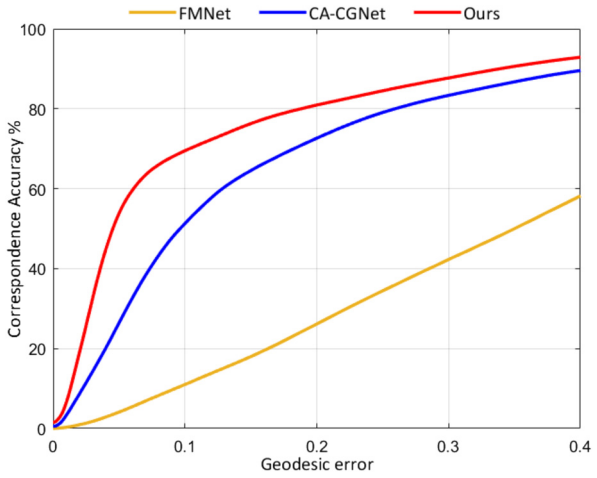

Fig. 18. The correspondence performance of our model and the compared learning-based models on the non-human dataset. Numbers on the left of the plot indicate correspondence accuracy with zero error.

Therefore, the dual route fusion mechanism can effectively improve the classification accuracy of digital capsules.

4.3.2 Effectiveness of the Global and Local Feature Fusion. In Section 3.3.1, the attention feature fusion module is used to enhance the feature extraction capability of the model. In order to verify

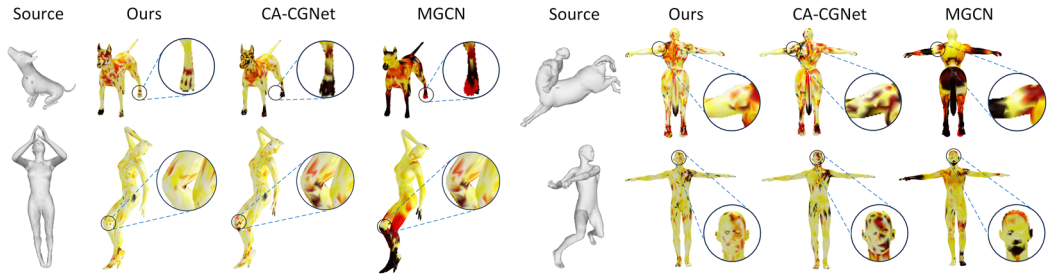

Fig. 19. Visualization of geodesic errors on human models and non-human models among compared methods. Hot color means a large error.

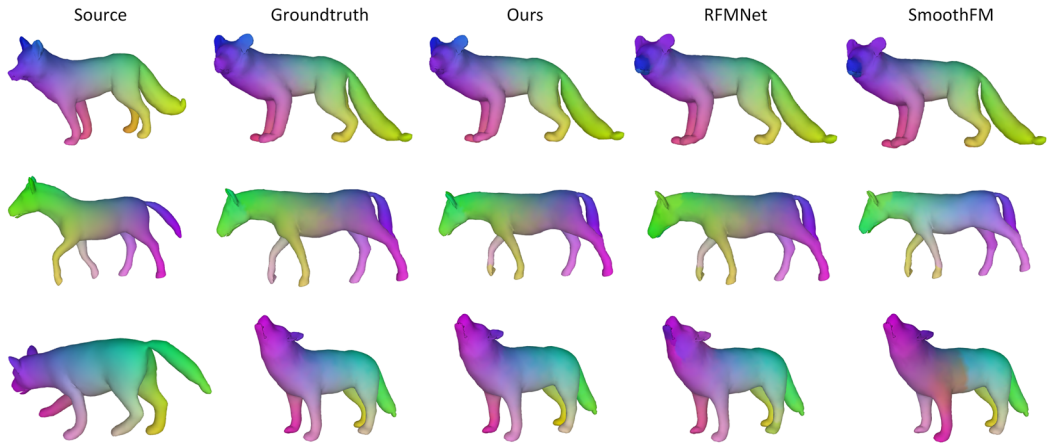

Fig. 20. Qualitative results on the SMAL dataset. The shapes are non-isometric from each other, and same parts correspond to the same colors.

Table 5. Comparison of Performance in Terms of the Mean Geodesic Errors (cm) on Remeshed TOSCA Dataset Using Different Routing Procedures in Capsule Networks

Method	David	Michael	Victoria	Mean	Standard Deviation
DR	**0.1734**	**0.1695**	0.1802	0.1744	**0.0054**
EM	0.2142	0.2016	0.1745	0.1967	0.0203
Ours	0.1807	0.1784	0.1609	**0.1733**	0.0108

the effectiveness of the attention-feature fusion module, we remove the attention-feature fusion module, only uses NAG-ResNet to extract the features of descriptors, and retrains the network on the FAUST dataset, and then tests the model on the TOSCA dataset. The experimental geodesic error results are shown in Table 6. It can be observed that the geodesic error of the model decreases after the attention feature fusion algorithm is adopted, which is 3% higher than the sigmoid routing algorithm. Further, Figure 21 shows the visualization results of geodesic error. The model using the attention feature fusion module achieves a lower geodesic error, thus proving the effectiveness of the attention feature fusion module.

Table 6. Comparison of Performance in Terms of the Mean Geodesic Errors (cm) on Remeshed TOSCA Dataset Using Different Feature Extraction Procedures

Method	David	Michael	Victoria	Mean	Standard Deviation
SURFMNet	0.3659	0.3895	0.3727	0.3760	0.0121
Ours without NAG-ResNet	0.2184	0.2317	0.2702	0.2401	0.0269
Ours	**0.1807**	**0.1784**	**0.1609**	**0.1733**	**0.0108**

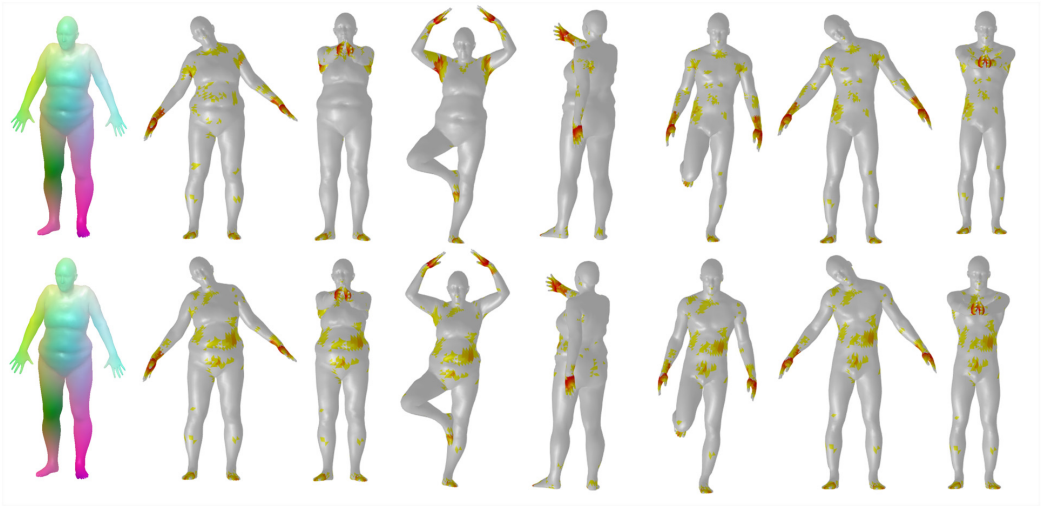

Fig. 21. Visualization of geodesic errors of several methods tested on remeshed shape pairs from FAUST dataset.

5 Conclusion

In this article, we propose a novel relation constrained shape correspondence network that embeds feature space and the input shape space based on the functional maps with MST of geodesic curves as relation constraints. By integrating relation constrained loss as a regularization term, our network can further boost the shape correspondence learning performance. Our approach proved rigorously and demonstrated the effectiveness on several challenging datasets, and can be adapted to different shape categories, especially for regions of hands and feet as shown in Figure 15. Currently, our method cannot be applied to other kinds of models without limbs, in which the number of categories for mesh segmentation is difficult to determine for training the DGANet. Hence, the geodesic relation constraints of singularities among the segmentation parts cannot be embedded in the framework of functional maps. Moreover, we have observed that our model incurs high computational costs and depends heavily on the quality of initial segmentation. In the future study, we plan to extend CGNNs to the partial setting for models without limbs and search for relation-aware routing mechanism in capsule network. Besides, we are exploring unsupervised and semi-supervised learning approaches that can potentially reduce the reliance on pre-segmentation and simplify models, which can enable the network to learn more effective correspondence strategies directly from the data.

References

[1] Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. 2005. Scape: Shape completion and animation of people. In Proceedings of the *ACM SIGGRAPH 2005 Papers*, 408–416.

[2] Mathieu Aubry, Ulrich Schlickewei, and Daniel Cremers. 2011. The wave kernel signature: A quantum mechanical approach to shape analysis. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*. IEEE, 1626–1633.

[3] Florian Bernard, Zeeshan Khan Suri, and Christian Theobalt. 2020. Mina: Convex mixed-integer programming for non-rigid shape alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13826–13835.

[4] Federica Bogo, Javier Romero, Matthew Loper, and Michael J. Black. 2014. FAUST: Dataset and evaluation for 3D mesh registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3794–3801.

[5] Davide Boscaini, Jonathan Masci, Emanuele Rodolà, and Michael Bronstein. 2016. Learning shape correspondence with anisotropic convolutional neural networks. *Advances in Neural Information Processing Systems* 29 (2016).

[6] Alexander M. Bronstein, Michael M. Bronstein, and Ron Kimmel. 2008. *Numerical Geometry of Non-Rigid Shapes*. Springer Science & Business Media.

[7] Etienne Corman, Maks Ovsjanikov, and Antonin Chambolle. 2014. Supervised descriptor learning for non-rigid shape matching. In *Proceedings of the European Conference on Computer Vision*. Springer, 283–298.

[8] Nicolas Donati, Abhishek Sharma, and Maks Ovsjanikov. 2020. Deep geometric functional maps: Robust feature learning for shape correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8592–8601.

[9] Yangbo Feng, Junyu Gao, and Changsheng Xu. 2022. Learning dual-routing capsule graph neural network for few-shot video classification. *IEEE Transactions on Multimedia* (2022).

[10] Vignesh Ganapathi-Subramanian, Boris Thibert, Maks Ovsjanikov, and Leonidas Guibas. 2016. Stable region correspondences between non-isometric shapes. In *Computer Graphics Forum*, Vol. 35. Wiley Online Library, 121–133.

[11] Dvir Ginzburg and Dan Raviv. 2020. Cyclic functional mapping: Self-supervised correspondence between non-isometric deformable shapes. In *Proceedings of the European Conference on Computer Vision*. Springer, 36–52.

[12] Thibault Groueix, Matthew Fisher, Vladimir G. Kim, Bryan C. Russell, and Mathieu Aubry. 2018. 3D-CODED: 3D correspondences by deep deformation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 230–246.

[13] Yulan Guo, Ferdous Sohel, Mohammed Bennamoun, Min Lu, and Jianwei Wan. 2013. Rotational projection statistics for 3D local surface description and object recognition. *International Journal of Computer Vision* 105, 1 (2013), 63–86.

[14] Oshri Halimi, Or Litany, Emanuele Rodola, Alex M. Bronstein, and Ron Kimmel. 2019. Unsupervised learning of dense shape correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4370–4379.

[15] Ling Hu, Qinsong Li, Shengjun Liu, and Xinru Liu. 2021. Efficient deformable shape correspondence via multiscale spectral manifold wavelets preservation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14536–14545.

[16] Ling Hu, Qinsong Li, Shengjun Liu, Dong-Ming Yan, Haojun Xu, and Xinru Liu. 2023. RFMNet: Robust deep functional maps for unsupervised non-rigid shape correspondence. *Graphical Models* 129 (2023), Article 101189.

[17] Haibin Huang, Evangelos Kalogerakis, Siddhartha Chaudhuri, Duygu Ceylan, Vladimir G. Kim, and Ersin Yumer. 2017. Learning local shape descriptors from part correspondences with multiview convolutional networks. *ACM Transactions on Graphics* 37, 1 (2017), 1–14.

[18] Varun Jain, Hao Zhang, and Oliver Van Kaick. 2007. Non-rigid spectral correspondence of triangle meshes. *International Journal of Shape Modeling* 13, 1 (2007), 101–124.

[19] Huan Li, Yibo Yang, Dongmin Chen, and Zhouchen Lin. 2018. Optimization algorithm inspired deep neural network structure design. In *Proceedings of the Asian Conference on Machine Learning*. PMLR, 614–629.

[20] Junyi Li, Siqing Li, Wayne Xin Zhao, Gaole He, Zhicheng Wei, Nicholas Jing Yuan, and Ji-Rong Wen. 2020a. Knowledge-enhanced personalized review generation with capsule graph neural network. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 735–744.

[21] Qinsong Li, Shengjun Liu, Ling Hu, and Xinru Liu. 2020b. Shape correspondence using anisotropic Chebyshev spectral CNNs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14658–14667.

[22] Xiang Li, Congcong Wen, Lingjing Wang, and Yi Fang. 2020c. Topology constrained shape correspondence. *IEEE Transactions on Visualization and Computer Graphics* 27, 10 (2020), 3926–3937.

[23] Yang Li, Wei Zhao, Erik Cambria, Suhang Wang, and Steffen Eger. 2021. Graph routing between capsules. *Neural Networks* 143 (2021), 345–354.

[24] Yuanfeng Lian and Mengqi Chen. 2023. CA-CGNet: Component-aware capsule graph neural network for non-rigid shape correspondence. *Applied Sciences* 13, 5 (2023), 3261.

[25] Yaron Lipman and Thomas Funkhouser. 2009. Möbius voting for surface correspondence. *ACM Transactions on Graphics* 28, 3 (2009), 1–12.
[26] Or Litany, Tal Remez, Emanuele Rodola, Alex Bronstein, and Michael Bronstein. 2017. Deep functional maps: Structured prediction for dense shape correspondence. In *Proceedings of the IEEE International Conference on Computer Vision*, 5659–5667.
[27] Yi Liu, Dingwen Zhang, Qiang Zhang, and Jungong Han. 2021. Part-object relational visual saliency. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).
[28] Robin Magnet, Jing Ren, Olga Sorkine-Hornung, and Maks Ovsjanikov. 2022. Smooth non-rigid shape matching via effective Dirichlet energy optimization. In *Proceedings of the International Conference on 3D Vision (3DV)*. IEEE, 495–504.
[29] Sazan Mahbub and Md Shamsuzzoha Bayzid. 2022. EGRET: Edge aggregated graph attention networks and transfer learning improve protein–protein interaction site prediction. *Briefings in Bioinformatics* 23, 2 (2022), Article bbab578.
[30] Riccardo Marin, Simone Melzi, Emanuele Rodola, and Umberto Castellani. 2020. Farm: Functional automatic registration method for 3d human bodies. In *Computer Graphics Forum*, Vol. 39. Wiley Online Library, 160–173.
[31] Haggai Maron, Meirav Galun, Noam Aigerman, Miri Trope, Nadav Dym, Ersin Yumer, Vladimir G. Kim, and Yaron Lipman. 2017. Convolutional neural networks on surfaces via seamless toric covers. *ACM Transactions on Graphics* 36, 4 (2017), 71–1.
[32] Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. 2015. Geodesic convolutional neural networks on riemannian manifolds. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 37–45.
[33] Simone Melzi, Jing Ren, Emanuele Rodola, Abhishek Sharma, Peter Wonka, and Maks Ovsjanikov. 2019. Zoomout: Spectral upsampling for efficient shape correspondence. arXiv:1904.07865.
[34] Dorian Nogneng, Simone Melzi, Emanuele Rodola, Umberto Castellani, Michael Bronstein, and Maks Ovsjanikov. 2018. Improved functional mappings via product preservation. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 179–190.
[35] Maks Ovsjanikov, Mirela Ben-Chen, Justin Solomon, Adrian Butscher, and Leonidas Guibas. 2012. Functional maps: A flexible representation of maps between shapes. *ACM Transactions on Graphics* 31, 4 (2012), 1–11.
[36] Adrien Poulenard and Maks Ovsjanikov. 2018. Multi-directional geodesic neural networks via equivariant convolution. *ACM Transactions on Graphics* 37, 6 (2018), 1–14.
[37] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. 2017. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 652–660.
[38] Jing Ren, Mikhail Panine, Peter Wonka, and Maks Ovsjanikov. 2019. Structured regularization of functional map computations. In *Computer Graphics Forum*, Vol. 38. Wiley Online Library, 39–53.
[39] Jing Ren, Adrien Poulenard, Peter Wonka, and Maks Ovsjanikov. 2018. Continuous and orientation-preserving correspondences via functional maps. *ACM Transactions on Graphics* 37, 6 (2018), 1–16.
[40] Emanuele Rodolà, Luca Cosmo, Michael M Bronstein, Andrea Torsello, and Daniel Cremers. 2017. Partial functional correspondence. In *Computer Graphics Forum*, Vol. 36. Wiley Online Library, 222–236.
[41] Emanuele Rodolà, Zorah Lähner, Alexander M. Bronstein, Michael M. Bronstein, and Justin Solomon. 2019. Functional maps representation on product manifolds. In *Computer Graphics Forum*, Vol. 38. Wiley Online Library, 678–689.
[42] Emanuele Rodola, Samuel Rota Bulo, Thomas Windheuser, Matthias Vestner, and Daniel Cremers. 2014. Dense non-rigid shape correspondence using random forests. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4177–4184.
[43] Jean-Michel Roufosse, Abhishek Sharma, and Maks Ovsjanikov. 2019. Unsupervised deep learning for structured shape matching. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1617–1627.
[44] Yusuf Sahillioğlu. 2018. A genetic isometric shape correspondence algorithm with adaptive sampling. *ACM Transactions on Graphics* 37, 5 (2018), 1–14.
[45] Yusuf Sahillioğlu. 2020. Recent advances in shape correspondence. *The Visual Computer* 36, 8 (2020), 1705–1721.
[46] Yusuf Sahillioğlu and Yücel Yemez. 2012. Minimum-distortion isometric shape correspondence using EM algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34, 11 (2012), 2203–2215.
[47] Y. Sahillioğlu and Y. Yemez. 2011. Coarse-to-fine combinatorial matching for dense isometric shape correspondence. *Computer Graphics Forum* 30, 5 (2011), 1461–1470. DOI: https://doi.org/10.1111/j.1467-8659.2011.02020.x
[48] Konstantinos Sfikas, Theoharis Theoharis, and Ioannis Pratikakis. 2012. Non-rigid 3D object retrieval using topological information guided by conformal factors. *The Visual Computer* 28, 9 (2012), 943–955.
[49] Nicholas Sharp, Souhaib Attaiki, Keenan Crane, and Maks Ovsjanikov. 2022. Diffusionnet: Discretization agnostic learning on surfaces. *ACM Transactions on Graphics* 41, 3 (2022), 1–16.

[50] Yi Shi, Mengchen Xu, Shuaihang Yuan, and Yi Fang. 2020. Unsupervised deep shape descriptor with point distribution learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9353–9362.
[51] Jian Sun, Maks Ovsjanikov, and Leonidas Guibas. 2009. A concise and provably informative multi-scale signature based on heat diffusion. In *Computer Graphics Forum*, Vol. 28. Wiley Online Library, 1383–1392.
[52] Federico Tombari, Samuele Salti, and Luigi Di Stefano. 2010. Unique signatures of histograms for local surface description. In *Proceedings of the European Conference on Computer Vision*. Springer, 356–369.
[53] Matthias Vestner, Zorah Lähner, Amit Boyarski, Or Litany, Ron Slossberg, Tal Remez, Emanuele Rodola, Alex Bronstein, Michael Bronstein, Ron Kimmel, and Daniel Cremers. 2017. Efficient deformable shape correspondence via kernel matching. In *Proceedings of the International Conference on 3D Vision (3DV)*. IEEE, 517–526.
[54] Kangkan Wang, Guofeng Zhang, Huayu Zheng, and Jian Yang. 2021. Learning dense correspondences for non-rigid point clouds with two-stage regression. *IEEE Transactions on Image Processing* 30 (2021), 8468–8482.
[55] Sen Wang, Yang Wang, Miao Jin, Xianfeng David Gu, and Dimitris Samaras. 2007. Conformal geometry and its applications on 3D shape matching, recognition, and stitching. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 7 (2007), 1209–1220.
[56] Yiqun Wang, Jianwei Guo, Dong-Ming Yan, Kai Wang, and Xiaopeng Zhang. 2019. A robust local spectral descriptor for matching non-rigid shapes with incompatible shape structures. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6231–6240.
[57] Yiqun Wang, Jing Ren, Dong-Ming Yan, Jianwei Guo, Xiaopeng Zhang, and Peter Wonka. 2020. MGCN: Descriptor learning using multiscale gcns. *ACM Transactions on Graphics* 39, 4 (2020), 122–1.
[58] Lingyu Wei, Qixing Huang, Duygu Ceylan, Etienne Vouga, and Hao Li. 2016. Dense human body correspondences using convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1544–1553.
[59] Rui Xiang, Rongjie Lai, and Hongkai Zhao. 2021. A dual iterative refinement method for non-rigid shape matching. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15930–15939.
[60] Jin Xie, Guoxian Dai, Fan Zhu, Edward K. Wong, and Yi Fang. 2016. Deepshape: Deep-learned shape descriptor for 3d shape retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 7 (2016), 1335–1345.
[61] Zhang Xinyi and Lihui Chen. 2018. Capsule graph neural network. In *Proceedings of the International Conference on Learning Representations*.
[62] Jinyu Yang, Peilin Zhao, Yu Rong, Chaochao Yan, Chunyuan Li, Hehuan Ma, and Junzhou Huang. 2020. Hierarchical graph capsule network. arXiv:2012.08734.
[63] Rui Yang, Wenrui Dai, Chenglin Li, Junni Zou, and Hongkai Xiong. 2020. NCGNN: Node-level capsule graph neural network. arXiv:2012.03476.
[64] Zhangsihao Yang, Or Litany, Tolga Birdal, Srinath Sridhar, and Leonidas Guibas. 2021. Continuous geodesic convolutions for learning on 3d shapes. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 134–144.
[65] Silvia Zuffi and Michael J. Black. 2015. The stitched puppet: A graphical model of 3d human shape and pose. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3537–3546.
[66] Silvia Zuffi, Angjoo Kanazawa, David W. Jacobs, and Michael J. Black. 2017. 3D menagerie: Modeling the 3D shape and pose of animals. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 6365–6373.

Received 18 December 2023; revised 22 June 2024; accepted 26 July 2024