IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received October 14, 2018, accepted November 15, 2018, date of publication November 22, 2018,
date of current version December 27, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2882711

Non-Rigid 3D Model Retrieval Based on Quadruplet Convolutional Neural Networks

HUI ZENG, YANRONG LIU, JIWEI LIU, AND DONGMEI FU
Beijing Engineering Research Center of Industrial Spectrum Imaging, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

Corresponding author: Hui Zeng (hzeng@ustb.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grants 61375010, in part by the Natural Science Foundation of Hebei Province under Grant F2018205102, in part by the Fundamental Research Funds for the Central Universities under Grant FRF-BD-17-002A, and in part by the China Scholarship Council under Grant 201706465032.

ABSTRACT Non-rigid 3-D model retrieval is a challenging problem in 3-D shape analysis. Recently, deep learning-based 3-D feature extraction methods have been studied and have achieved better performance than the previous state-of-the-art methods. Inspired by the quadruplet neural networks proposed for learning local image feature descriptors, we propose a novel non-rigid 3-D model retrieval method based on quadruplet convolutional neural networks. For training the proposed networks, the quadruplet samples are first selected using the online sampling method. For each 3-D model, the wave kernel signature descriptor of each vertex is computed, and its corresponding multi-energy shape distribution matrix is constructed as the input of the network. Then, the quadruplet convolutional neural networks are trained using our improved quadruplet loss function, which not only preserves the advantages of existing quadruplet loss functions but also decreases the risk of underfitting. For the query sample, the 3-D shape features are computed using one branch of the trained quadruplet networks. Finally, the retrieval results are obtained by the L2 distance measure. Extensive experimental results have validated the effectiveness of the proposed method.

INDEX TERMS Non-rigid 3D model retrieval, convolutional neural network, quadruplet loss function, wave kernel signature, multi-energy shape distribution.

I. INTRODUCTION

As the fourth generation of multimedia data type following audio, image and video, the 3D model plays an increasingly important role in many fields, such as biometrics, virtual reality, medical diagnosis, self-driving cars, and intelligent robots. Currently, the explosive growth of the number of 3D models has led to an urgent need for efficient 3D model retrieval. Generally, the 3D model includes rigid 3D models and non-rigid 3D models. As the non-rigid 3D models have many kinds of different deformations, how to analyze the non-rigid 3D model effectively is a challenging problem to be solved. Non-rigid 3D model retrieval has become an important research topic in the field of computer vision and multimedia information processing [1].

Recently, extensive efforts have been dedicated to non-rigid 3D model retrieval. Existing non-rigid 3D model retrieval methods can be divided into two categories: multi-view-based methods and local feature-based methods [2]. The multi-view-based methods first describe the 3D model by a collection of 2D projective images and then it compares the similarity of the images for retrieval. Although this kind of method can convert the non-rigid 3D model retrieval problem into multiple image-based retrieval problems, it needs to transform the non-rigid 3D model into the canonical form. The local feature-based method does not need the above transformation step and has better robustness to occlusions and mesh resolution. Therefore, in recently, local feature-based non-rigid 3D model retrieval methods have attracted increasing attention from researchers. The typical algorithm flow first computes the local descriptors, and then encodes the local descriptors to obtain the global shape representation of the non-rigid 3D model. Finally, the similarity of the encoded global shape features is computed for retrieval.

The 3D local descriptor is used for characterizing local surfaces effectively, and it can directly influence the performance of the 3D model retrieval system. Recently, many kinds of 3D local descriptors have been proposed, such as spin images [3], [4], 3D shape context [5], rotational projection statistics [6], local surface patches [7], [8], point feature histograms [9], fast point feature histograms [10],

2169-3536 © 2018 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

and signature histograms of orientations [11], [12]. Although the above descriptors have achieved good retrieval results, most of them are not invariant to non-rigid 3D deformations and are not suitable for non-rigid 3D model retrieval. To solve this problem, Sun *et al.* [13] proposed heat kernel signatures (HKS) to describe non-rigid 3D local surfaces, which are based on diffusion scale-space analysis and defined as an exponentially weighted combination of the Laplace-Beltrami (LB) eigenfunctions. It is invariant to isometric deformations and robust to small non-isometric deformations, but it is sensitive to the scale changes of the 3D model. Bronstein and Kokkinos [14] proposed scale-invariant heat kernel signatures (SI-HKS), which satisfies both the merits of the HKS descriptor and scale invariance. Similar to the HKS descriptor, the SI-HKS descriptor includes low-frequency information describing the global structure of the shape. Aubry *et al.* [15] proposed the wave kernel signature (WKS), which describes the average probability of quantum mechanics at a specific location on the surface. The WKS is invariant to non-rigid transformations and clearly separates the influence of different frequencies, which allows access to both low-frequency information and high-frequency information. As the WKS descriptor is appropriately parameterized by a theoretical stability analysis, it is not only highly informative but also robust to non-isometric perturbations of the shape. Therefore, in this paper, we adopt the WKS descriptor to describe the local surfaces of the non-rigid 3D models.

Generally, the global feature of the non-rigid 3D model can be obtained by encoding the 3D local descriptors. Similar to the field of image retrieval, the earliest widely used encoding method was the bag of features (BoF) model-based methods. Then, researchers proposed several improved models based on the framework of the BoF model, for example, the spatially sensitive bag of words (SS-BoW) [16], bag of feature graphs (BoFG) [17], bag of phrases (BoP) [18], and weighted bag of phrases (W-BoP) [19]. Tabia *et al.* [20] used the covariance matrices of the local descriptors and the generalized BOF paradigm to represent the 3D model. Lavoué [21] proposed a bag of words (BoW)-based 3D shape retrieval algorithm, which uses a uniform sampling of feature points associated with a new local Fourier descriptor. EINaghy proposed a non-rigid 3D model retrieval method based on the bag of compact HKS-based feature descriptors [22], which includes five steps: HKS computation, feature point detection, feature point description, bag of features and the matching phase. Lian *et al.* [22] used the Fisher vector encoding method to describe the global shape for 3D model retrieval. Litman *et al.* [23] used the sparse coding method to learn encoded representation coefficients for retrieval. Agathos *et al.* [24] proposed a graph-based representation for 3D object retrieval, which uses an attributed relational graph to obtain the structural description of the 3D object. Papadakis *et al.* [25] proposed a 3D object retrieval method using an efficient and compact hybrid shape descriptor, which is composed of 2D features based on depth buffers and 3D features based on spherical harmonics. Tabia *et al.* [26] used the vector of locally aggregated tensors (VLAT) technique to aggregate the local descriptors of object depth maps. Then, they reduced their size using principal component analysis (PCA) on VLAT vectors.

Recently, as the deep learning technique has achieved superior performance in the field of image retrieval, researchers have begun to investigate the application of the deep learning technique in 3D model retrieval. Bu *et al.* [27] propose a multilevel 3D shape feature extraction framework using deep learning techniques. The 3D local descriptors are first encoded into a geometric BoW, and then the shape features are learned via deep belief networks for shape classification and retrieval. Dai *et al.* [28] used the locality-constrained linear coding (LLC) algorithm to encode the SI-HKS descriptor of each vertex to form the global shape representation. Then, a discriminative shape descriptor was learned for retrieval via a many-to-one encoder. Xie *et al.* [29] proposed a deep unsupervised shape descriptor using a supervised progressive shape distribution encoder (SPSDE). First, they developed a shape distribution representation based on the HKS descriptor. Then, multiple SPSDEs are stacked to characterize the intrinsic structures of 3D shapes. Finally, all neurons in the middle hidden layers of the network are concatenated to form a shape descriptor for 3D model retrieval. After that, they proposed the discriminative autoencoder-based shape descriptor (DASD) to extract high-level shape features more efficiently [30]. It uses a multiscale shape distribution as input to the autoencoder and imposes the Fisher discrimination criterion on the neurons in the hidden layers.

Motivated by the fact that the deep neural network-based 3D shape feature learning methods have achieved better performance, this paper introduces a novel non-rigid 3D model retrieval method based on the quadruplet convolutional neural networks (QCNNs), which is inspired by the quadruplet networks proposed in [31] for learning local image feature descriptors. The proposed networks include four branches, and each branch is a convolutional neural network (CNN). The differences between our proposed networks and the networks proposed in [31] are the structure of the CNN and the quadruplet loss function. To summarize, the contributions of this paper are listed as follows. (1) We propose the WKS descriptor-based multi-energy shape distribution construction method, which can be used as the input of the CNN. (2) We design the structure of the CNN referring to VGG [32] and ResNet [33], which has better efficiency for 3D shape feature learning. (3) We propose an improved quadruplet loss function, which can minimize the intraclass distances and increase the interclass distances so that they are greater than a given threshold. Compared with existing quadruplet loss functions, our proposed quadruplet loss function can effectively reduce the dependence on training samples and decrease the risk of overfitting and underfitting. Our experimental results show that the proposed non-rigid 3D model retrieval method has better performance than the state-of-the-art methods.

The rest of the paper is organized as follows. In Section II, we briefly introduce the wave kernel signature descriptor and the deep quadruplet network. In Section III, we present the proposed non-rigid 3D model retrieval method, including multi-energy shape distribution and the quadruplet convolutional neural networks. Section IV performs extensive experiments, and Section V concludes this paper.

II. BACKGROUND

A. WAVE KERNEL SIGNATURE

The WKS is a kind of non-rigid 3D local shape descriptor derived from the framework of quantum mechanics, which is represented by the average probabilities of quantum particles of different levels [15]. The evolution of a quantum particle on the surface is defined by its wave function $\psi(x, t)$, which can be computed from the following Schrödinger equation:

$$\frac{\partial \psi}{\partial t}(x, t) = i \Delta \psi(x, t) \tag{1}$$

where Δ is the LB operator of the 3D shape, t is time, and i is the energy-related angular quantum number. Then, the wave function of the particle can be expressed using the solution of the Schrödinger equation:

$$\psi_E(x, t) = \sum_{k=0}^{\infty} e^{iE_k t} \phi_k(x) f_E(E_k) \tag{2}$$

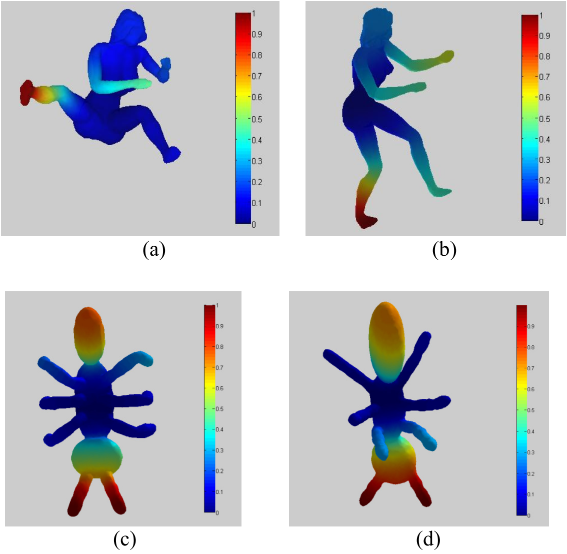
where $\phi_k(x)$ is the eigenvector of the LB operator Δ, and $f_E(E_k)$ is the energy probability density function with expectation energy value E_k. As the probability of the particle at the point x at time t is $|\psi_E(x, t)|^2$, the WKS can be defined as the average probability over time:

$$WKS(E, x) = \lim_{T \to \infty} \frac{1}{T} \int_0^T |\psi_E(x, t)|^2 = \sum_{k=0}^{\infty} \phi_k(x)^2 f_E(E_k)^2 \tag{3}$$

Then, the WKS descriptor at a point x on the surface can be computed based on the logarithmic energy scale $e = \log(E)$ using the following real-valued function:

$$\begin{cases} WKS(x, \cdot) : \mathbb{R} \to \mathbb{R} \\ WKS(x, \cdot) = \left(\sum_k e^{\frac{-(e - \log E_k)^2}{2\sigma^2}} \right)^{-1} \sum_k \phi^2(x) e^{\frac{-(e - \log E_k)^2}{2\sigma^2}} \end{cases} \tag{4}$$

From Equation (4) we can see that the WKS descriptor is a function of the energy levels. Large energies are mostly influenced by the local shape structure, and the small energies are mostly influenced by the global shape structure. Therefore, compared with the more widely used HKS descriptor, the WKS descriptor not only remains robust to non-rigid deformations but also captures more information about the shape differences at finer scales. It can clearly separate the influences of different frequencies and treat all frequencies equally. Furthermore, it does not require shape alignment and builds local coordinates in the application of 3D model retrieval. Therefore, in this paper, we adopt the WKS descriptor to compute the local descriptor of each vertex of the non-rigid 3D model. Figure 1 shows the WKS descriptors of the non-rigid 3D models from two classes: human and ant. The color maps are projected according to the value of one dimension of the WKS descriptors. From Figure 1, we can see that the WKS descriptors can describe different shape structures for different classes, and they have good robustness to non-rigid transformations.



(a) (b)

(c) (d)

FIGURE 1. Examples of the WKS descriptors of four non-rigid 3D models from two classes. (a) human1. (b) human2. (c) ant1. (d) ant2.

B. DEEP QUADRUPLET NETWORK

Although existing deep learning-based 3D model retrieval methods have achieved better performance than previous state-of-the-art methods, both loss function and network structure are still worth exploring for further improvement. Recently, the deep triplet networks and the deep quadruplet networks have been proposed for image feature extraction or ranking tasks, such as learning local feature descriptors, face recognition, and person reidentification [34], [35]. Compared with the one-branch deep network, the multibranch deep network can make full use of the relations between the samples, which are less prone to overfitting and have better training efficiency. The non-rigid 3D model retrieval can be considered a ranking problem; thus, we investigate how to design the multibranch deep network for non-rigid 3D model retrieval.

Balntas *et al.* [36] proposed a conjoined triple deep network for learning local image descriptors. The network has three parallel inputs, where two of the inputs are positive patches, and the third input is a negative patch. The loss function is defined using three distances from the three training patches, which can exploit the relations within the triplets. Kumar *et al.* [37] proposed a triplet convolutional network with a global loss that can minimize the overall classification

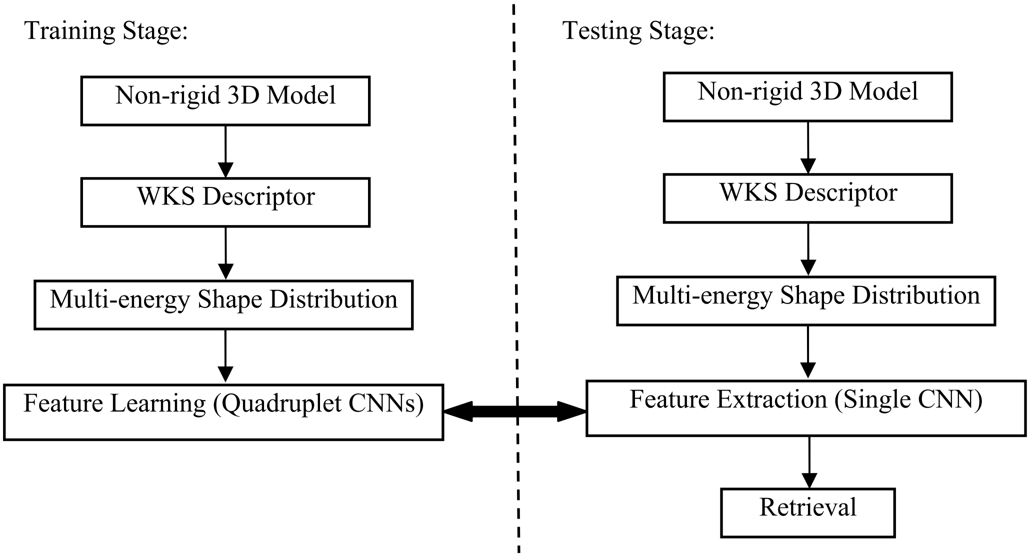

Training Stage:
Non-rigid 3D Model
WKS Descriptor
Multi-energy Shape Distribution
Feature Learning (Quadruplet CNNs)
Testing Stage:
Non-rigid 3D Model
WKS Descriptor
Multi-energy Shape Distribution
Feature Extraction (Single CNN)
Retrieval

FIGURE 2. The flowchart of the proposed non-rigid 3D model retrieval method.

error and improve the generalization capability of the model. Because the triplet network is excessively dependent on the "anchor" samples and has the margin varying problem that increases the risk of overfitting, Zhang *et al.* [31] proposed a deep convolutional neural network with quadruplet ranking loss to learn local feature descriptors. It has four branches with tied weights, which receive four image patches as inputs. The output of each branch is the feature vector of the corresponding image patch. Each branch is designed based on ResNet [33], and it consists of two residual blocks along with other layers. For the sample x, suppose its corresponding network output is $f(x)$. Let (p_1, p_2, n_1, n_2) be a sample quadruplet, and then the quadruplet loss function is defined as follows:

$$L(p_1, p_2, n_1, n_2) = \max(0, I + \|f(p_1) - f(p_2)\|_2 - \|f(n_1) - f(n_2)\|_2) \tag{5}$$

where (p_1, p_2) is the positive sample pair, (n_1, n_2) is the negative sample pair, and $I$is the given interval. Compared with the triplet loss function, this quadruplet loss function can mitigate the margin varying problem to some degree. It can flexibly use any combination of positive and negative pairs, which can improve the capacity to utilize the limited training data.

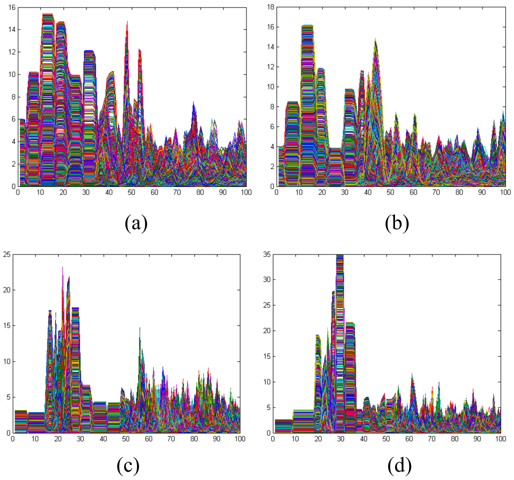
III. THE PROPOSED NON-RIGID 3D MODEL RETRIEVAL METHOD

In this section, we detail the proposed quadruplet CNNs-based non-rigid 3D model retrieval method. As shown in Figure 2, it consists of two stages: the training stage and the testing stage. For each non-rigid 3D model, we first compute the WKS descriptor of each vertex and then construct its corresponding multi-energy shape distribution. In the training stage, we use the quadruplet samples to train the proposed quadruplet networks using our improved quadruplet loss function. In the testing phase, we use only a single branch of the quadruplet networks to compute the 3D shape feature of the non-rigid 3D model. Finally, the retrieval results are obtained according to the similarity of the 3D shape features.

A. MULTI-ENERGY SHAPE DISTRIBUTION

As the numbers of the vertices of different 3D models are different, the feature matrices that are directly connected by the WKS descriptors have different dimensions for different 3D models. Therefore, they cannot be used as the input of the CNN. If we sample the 3D models with the same number of vertices, the directly connected feature matrices can have the same dimensions. However, if the number of sampling vertices is too small, considerable effective discriminative information will be lost. If the number of sampling vertices is too large, the structure of the CNN will be large, and an optimal search will be difficult. Inspired by the multiscale shape distribution [30], we propose a construction method for the multi-energy shape distribution of the WKS descriptor to obtain the global representation of the 3D model. For different 3D models, their corresponding multi-energy shape distribution matrices have the same dimensions. Therefore, the proposed multi-energy shape distribution is used as the input of the CNN.

Similar to the multiscale shape distribution of the HKS descriptor [30], we estimate the probability distribution of the WKS descriptor to form the shape distribution matrix. Suppose the non-rigid 3D model X has N vertices, and the dimension of the WKS descriptor is D. That is, the energy level of the WKS descriptor is D. For a non-rigid 3D model, let S_i be the WKS descriptor of the ith vertex, where $S_i = [S_i^1, S_i^2, \cdots S_i^D]$ and S_i^k is the WKS value at energy level $k(k = 1, \cdots, D)$. First, find the maximum value $S_{\max}$ and the minimum value $S_{\min}$of all the WKS values, and equally divide the interval range $[S_{\min}, S_{\max}]$ into M bins. For energy level k, we count the number of times that the WKS value S_i^k falls into each bin. Then, L1-normalization is performed on the statistical result, and it is taken as the *kth* column of the global shape distribution matrix. Finally, we can obtain an $M \times D$ multi-energy shape distribution matrix to describe the 3D model. Figure 3 illustrates four multi-energy shape distributions of four non-rigid 3D models from two classes. Here, $M = 128$, $D = 100$, the horizontal axis represents the energy levels, the vertical axis represents the number of times that the WKS values fall into each bin. From Figure 3, we can see that the multi-energy shape distributions are similar for the non-rigid 3D models from the same class, and the multi-energy shape distributions have obvious differences for the non-rigid 3D models from different classes. Therefore, the proposed multi-energy shape distribution can effectively describe the discriminative information of the non-rigid 3D models.



(a) (b)

(c) (d)

FIGURE 3. The multi-energy shape distributions of four non-rigid 3D models from two classes. (a) human1. (b) human2. (c) ant1. (d) ant2.

B. QUADRUPLET CONVOLUTIONAL NEURAL NETWORKS

Inspired by the quadruplet convolutional neural network for learning local feature descriptors [31], we propose a quadruplet CNN for non-rigid 3D model retrieval. As shown in Figure 4, the proposed networks have four parallel CNNs, which share the same weights. The reason for using the same weights is to ensure the consistency of the mapping of different CNNs. First, we use the online sampler proposed in [31] to obtain the quadruplet samples. A quadruplet sample set (p_1, p_2, n_1, n_2) includes four samples. Among them, the sample p_1 and p_2 are from the same class and are called a positive sample pair. The sample n_1 and n_2 are from different classes and are called a negative sample pair. Then, we compute the multi-energy shape distributions of the four samples and send them to the four CNNs.

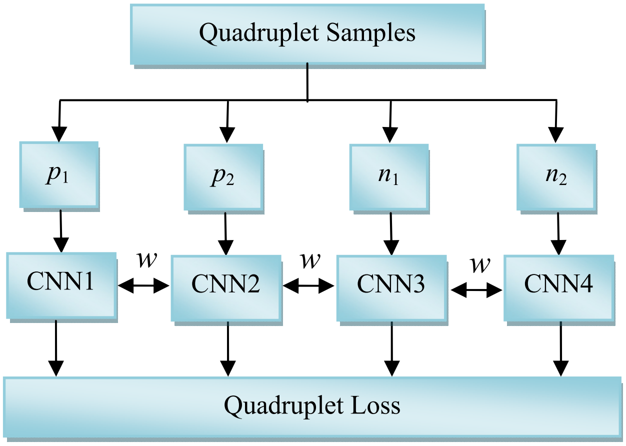


FIGURE 4. The framework of the proposed quadruplet convolutional neural networks.

Each CNN is designed referring to VGG and ResNet [32], [33]. As shown in Figure 5, each CNN consists of three convolutional layers (conv1, conv2, conv3), four pooling layers (pooling1, pooling2, pooling3, pooling4), two residual blocks (block1, block2), and two fully connected layers (fc1, fc2). For each deep network, we use the max-pooling method and ReLU activation function. Furthermore, we repeatedly use the convolution kernel with the same size for the first two continuous convolutional layers, which can reduce network parameters without degrading performance. When the depth of the deep network increases, the gradients of the front layers will become small. To solve this gradient vanishing problem, we adopt the residual blocks to optimize the network. To obtain more inputs for the network, we use the online sampling method to enlarge the training set. Then, the output of the second fully connected layer is the final 3D shape feature for non-rigid 3D model retrieval.

In this paper, the aim of the loss function is to decrease the outputs of the intraclass distances and increase the outputs of the interclass distances. The quadruplet loss function proposed in [31] enforces that the distances between the positive sample pairs are smaller than the distances between the negative sample pairs. It is suitable for data with different distributions and has fewer restrictions, but it easily converges to the local optimum. Usually, the network needs sufficient quadruplet samples to ensure convergence to global optimality. The generalization ability of the network is determined by the maximum distance of the positive sample pairs and the minimum distance of the negative pairs of the training sets. Compared with existing triplet loss functions, although the quadruplet loss function proposed in [31] can decrease the risk of overfitting, it also has the risk of underfitting. To solve this problem, we propose an improved quadruplet loss function by adding a threshold constraint, which is defined as follows:

$$L(p_1, p_2, n_1, n_2) = w \left\| f(p_1) - f(p_2) \right\|_2^2 + \max\left(0, T - \left\| f(n_1) - f(n_2) \right\|_2^2\right) \quad (6)$$

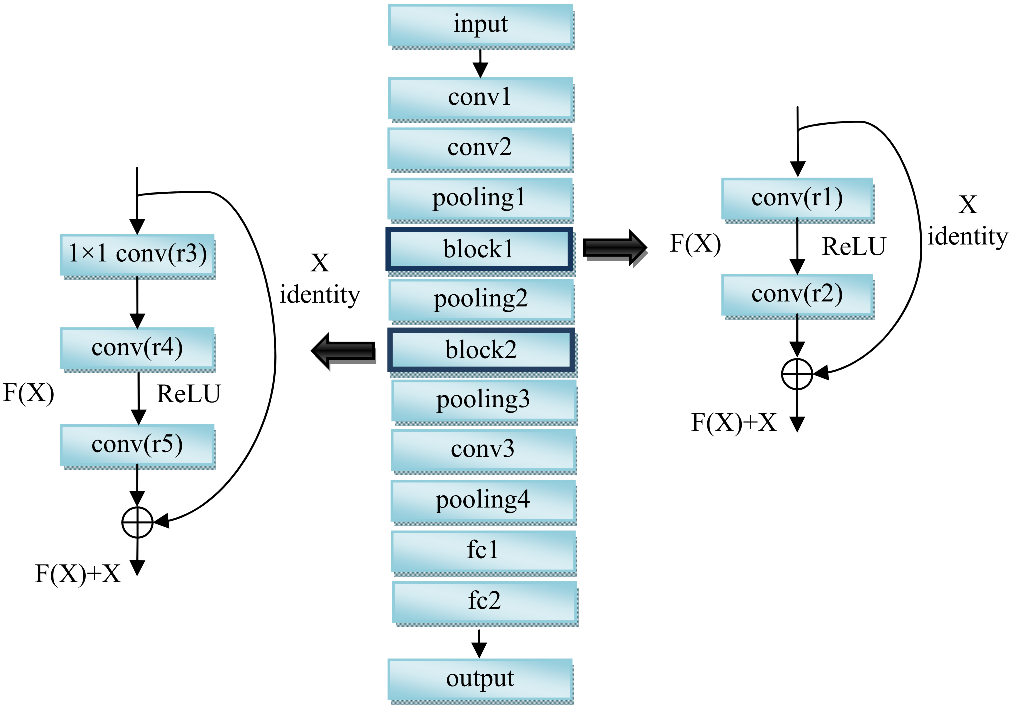


FIGURE 5. The structure of the CNN of the proposed quadruplet convolutional neural networks.

where w is the weight, and $T$is the threshold. From Equation (6), we can see that the proposed loss function makes the distance of the positive sample pair as small as possible, and the distance of the negative sample pair is greater than a given threshold. Thus, our improved loss function not only preserves the advantages of the existing quadruplet loss function but also effectively reduces the dependence on the training samples. The proposed network is trained by the stochastic gradient descent (SGD) method with back propagation, and the gradient of the loss function can be derived as:

$$\frac{\partial L}{\partial f(p_1)} = 2\left[f(p_1) - f(p_2)\right] \tag{7}$$

$$\frac{\partial L}{\partial f(p_1)} = 2w\left[f(p_1) - f(p_2)\right] \tag{8}$$

$$\frac{\partial L}{\partial f(n_1)} = 2\left[f(n_1) - f(n_2)\right] \cdot 1_A\left[T - \|f(n_1) - f(n_2)\|_2^2\right] \tag{9}$$

$$\frac{\partial L}{\partial f(n_2)} = -2\left[f(n_1) - f(n_2)\right] \cdot 1_A\left[T - \|f(n_1) - f(n_2)\|_2^2\right] \tag{10}$$

where $1_A(x) = \begin{cases} 1, x >= 0 \\ 0, x < 0 \end{cases}$.

The proposed non-rigid 3D model retrieval method can be summarized as follows. In the training stage, we first select quadruplet samples and enlarge the training set using the online sampling method. Then, for each training sample, we compute the WKS descriptor of each vertex, and its corresponding multi-energy shape distribution matrix. Finally, we train the proposed quadruplet networks using the improved quadruplet ranking loss function and the SGD method, and the final 3D shape feature of each training sample is obtained. For the testing sample, we first compute the WKS descriptor of each vertex and its corresponding multi-energy shape distribution matrix. Then, the 3D shape feature is computed using one branch of the trained quadruplet network. Finally, we use the L2 distance measure to obtain the retrieval results.

IV. EXPERIMENTAL RESULTS

To evaluate the effectiveness of the proposed non-rigid 3D model retrieval method, we compare it to the state-of-the-art methods on two datasets: McGill 3D shape benchmark [38] and SHREC'11 non-rigid 3D model dataset [39]. In this paper, we denote our proposed non-rigid 3D model retrieval method by "QCNN" and use the following four measures: nearest neighbor (NN), the first tier (FT), the second tier (ST) and the discounted cumulative gain (DCG) to evaluate the retrieval performance.

A. EXPERIMENTAL SETTINGS

In the experiments, the number of eigenvalues of the LB operator k is 300, and 100 values of the energy scale are evaluated. Therefore, the dimension of the WKS descriptor is 100. Let 128 bins be equally divided between $S_{\min}$and $S_{\max}$, and then the multi-energy shape distribution is a 128×100 matrix, which is used as the input to the CNN. Table 1 presents a detailed description of each CNN. All the convolutional layers and the pooling layers adopt the "SAME" convolution

TABLE 1. Description of each conventional neural network.

Layer	Type	Patch size	Stride	Number of feature maps	Output size
x	Input	-	-	-	128×100
conv1	Convolution	3×3	1	16	128×100
conv2	Convolution	3×3	1	32	128×100
max pooling1	max pooling	2×2	2	32	64×50
block1	Convolution	3×3	1	32	64×50
	Convolution	3×3	1	32	64×50
	F(X)+X	-	-	32	64×50
max pooling2	max pooling	2×2	2	32	32×25
	Convolution	1×1	1	64	32×25
	Convolution	3×3	1	64	32×25
	Convolution	3×3	1	64	32×25
	F(X)+X	-	-	64	32×25
max pooling3	max pooling	2×2	2	64	16×13
conv3	Convolution	2×2	1	128	16×13
max pooling4	max pooling	2×2	2	128	8×7
fc1	fully connected	-	-	1	7168
fc2	fully connected	-	-	1	1000

method so that the dimensions are consistent after convolution. The activation function of all the convolutional layers and the fully connected layers is a ReLU function. The convolution kernel weights are initialized using the truncated positive distribution, and the bias term is initialized to 0. We use the batch training method, and the batch size is 5. The SGD method with back-propagation is used to minimize the proposed quadruplet loss. The learning rate is 0.001, and the proposed quadruplet networks converge after 200,000 training iterations according to the judging from the loss curve.

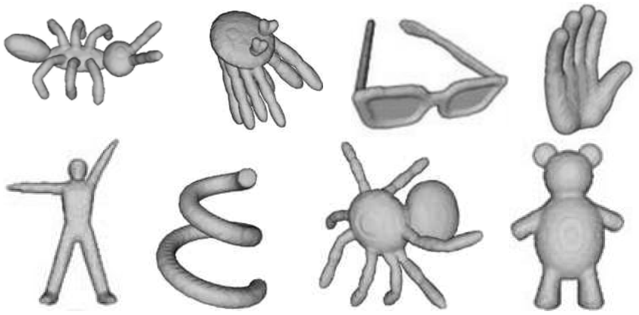


FIGURE 6. Example non-rigid 3D models of the McGill 3D shape benchmark.

B. COMPARISON EVALUATION

1) McGill 3D SHAPE BENCHMARK

The McGill 3D shape benchmark contains 255 non-rigid 3D models from 10 different classes, including: ‘‘ant’’, ‘‘crab’’, ‘‘spectacle’’, ‘‘hand’’, ‘‘human’’, ‘‘octopus’’, ‘‘plier’’, ‘‘snake’’, ‘‘spider’’ and ‘‘teddy-bear’’ [38]. Each class has 20 to 30 3D models. These models include rotational transformation, scale transformation and non-rigid deformation. Some example non-rigid 3D models are shown in Figure 6.

First, we analyze the learned 3D shape features of the proposed ‘‘QCNN’’ method. Figure 7 shows the two classes of the 3D models and their corresponding learned 3D shape features. From Figure 7, we can see that for different 3D models of the same class, their corresponding learned 3D shape features are very similar. For the different 3D models of different classes, their corresponding learned 3D shape features are distinctive. Therefore, we can conclude that our proposed quadruplet convolutional neural networks can extract robust and discriminative shape features. Figure 8 shows some retrieval results that include the given query sample and the top 15 retrieved 3D models; mistakes are highlighted in red. From Figure 7, we can see that there are large non-rigid deformations within the intraclass samples, and some interclass samples look similar. For example, when the query is a ‘‘spider’’, it has an incorrectly retrieved 3D model from the class ‘‘ant’’. They look very similar.

Finally, we compared our proposed ‘‘QCNN’’ method with the covariance descriptor-based method [20], the hybrid 2D/3D method [21], the CBoFHKS-based method [22], the graph-based method [24], the hybrid BOW method [25], the PCA-based VLAT method [26], the DASD-based method [29], the SPSDE-based method [30], the NMLM-based method [40] and the QCNN_origin method. The QCNN_origin method use same network and training method

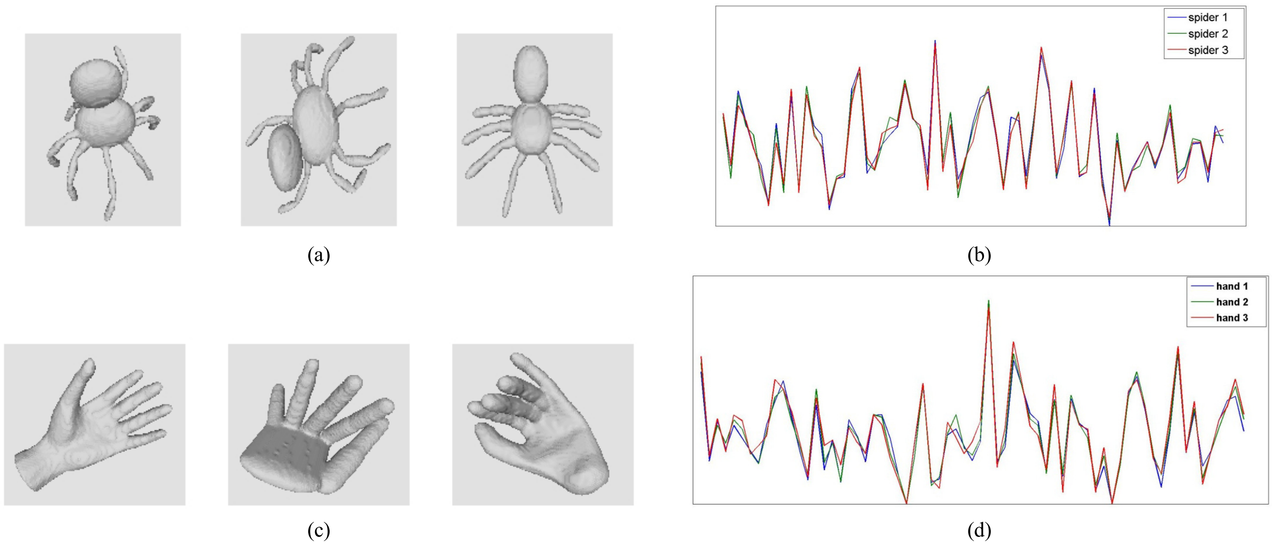


FIGURE 7. The learned 3D shape features for the spider models and the hand models of the McGill 3D shape benchmark. (a) Spider models: spider1, spider2, spider3. (b) The learned 3D shape feature for the spider models. (c) Hand models: hand1, hand2, hand3. (d) The learned 3D shape feature for the hand models.

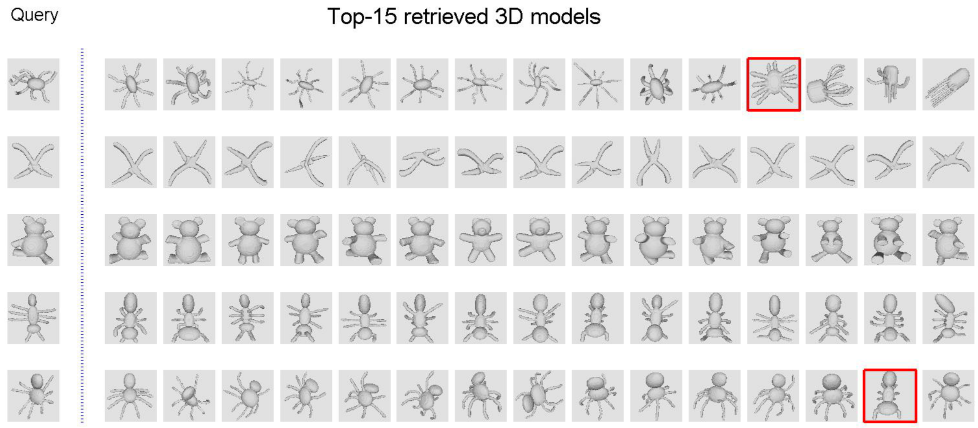


FIGURE 8. Examples showing the given query and the top 15 retrieved 3D models; mistakes are highlighted in red.

as our proposed method, but it use the quadruplet loss function shown in Equation (5). The comparative retrieval results of our method and the other state-of-the-art methods are listed in TABLE 2. From this table, we can see that our proposed method achieved the best performance on the NN, FT, ST and DCG measures. Compared with the QCNN_origin method, our proposed method also performs better. So our improved quadruplet loss function outperforms the original quadruplet loss function. In summary, our proposed method achieved better performance than the other methods.

2) SHREC'11 NON-RIGID 3D MODEL DATASET

The SHREC'11 non-rigid 3D model dataset consists of 600 non-rigid 3D models that are equally classified into 30 categories, among which 26 classes of objects are collected from several freely accessible repositories while the other 4 classes of objects are created using Autodesk 3d Max. Some example non-rigid 3D models are shown in Figure 9.

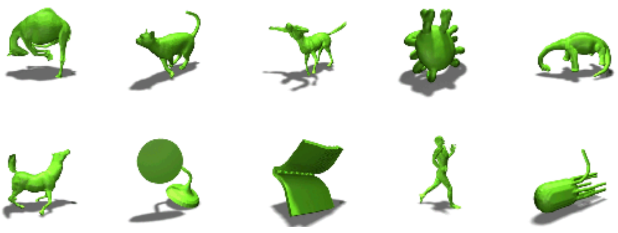


FIGURE 9. Example non-rigid 3D models of the SHREC'11 non-rigid 3D model dataset.

First, six non-rigid 3D models and their corresponding learned 3D shape features are illustrated in Figure 10. Similar to Figure 7, we can conclude that the learned shape features

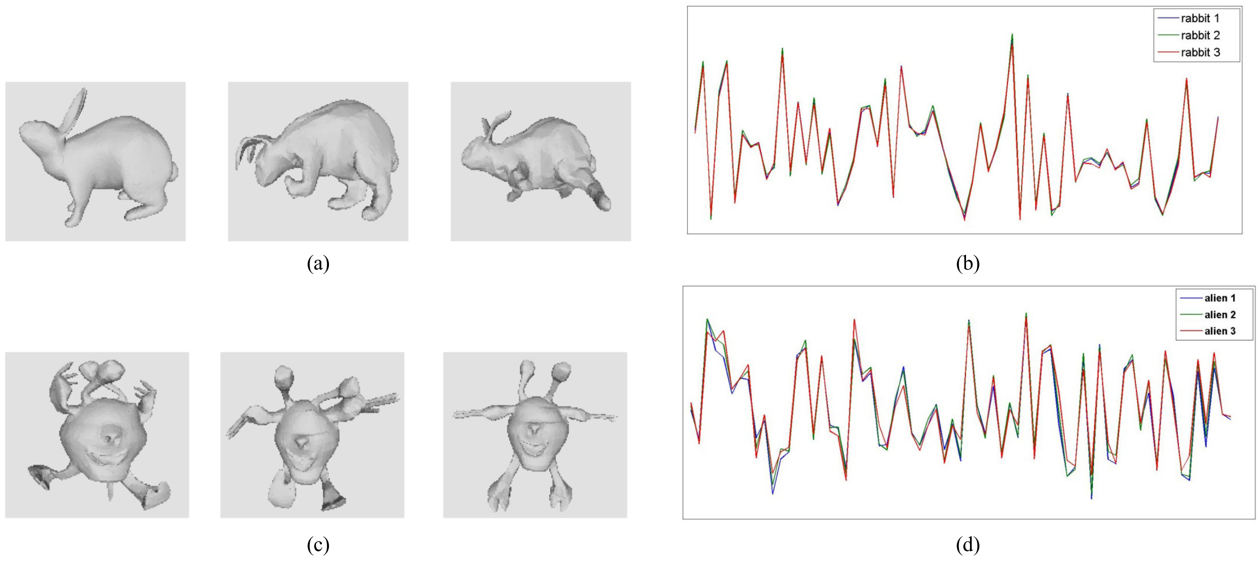


FIGURE 10. The learned 3D shape features for the rabbit models and the alien models of the SHREC'11 non-rigid 3D model dataset. (a) Rabbit models: rabbit1, rabbit2, rabbit3. (b) The learned 3D shape features for the rabbit models. (c) Alien models: alien1, alien2, alien3. (d) The learned 3D shape features for the alien models.

TABLE 2. Retrieval results compared with other methods.

Method	NN	FT	ST	DCG
Covariance descriptor[20]	0.977	0.732	0.818	0.937
Hybrid 2D/3D[21]	0.925	0.557	0.698	0.850
CBoFHKS[22]	0.901	0.778	0.876	0.891
Graph-based method[24]	0.976	0.741	0.911	0.933
Hybrid BOW[25]	0.957	0.635	0.790	0.886
PCA-based VLAT[26]	0.969	0.658	0.781	0.894
DASD[29]	0.988	0.782	0.834	0.955
SPSDE[30]	0.986	0.883	0.911	0.952
MMLM[40]	0.971	0.916	0.991	0.973
QCNN_origin	0.962	0.968	0.985	0.975
QCNN	**0.991**	**0.990**	**1.000**	**0.995**

TABLE 3. Retrieval results compared with other methods.

Method	NN	FT	ST	DCG
FOG+MRR[39]	0.9600	0.8810	0.9461	0.9586
LSF[39]	0.9950	0.7988	0.8631	0.9432
BOGH[39]	0.9933	0.8111	0.8839	0.9493
MeshSIFT[39]	0.9950	0.8844	0.9617	0.9804
OrigM-n12-normA[39]	0.9917	0.9153	0.9569	0.9783
MDS-CM-BOF[39]	0.9950	0.9127	0.9691	0.9822
SD-GDM[39]	**1.0000**	0.9622	0.9840	0.9936
SD-GDM-meshSIFT[39]	**1.0000**	0.9720	0.9901	0.9955
QCNN_origin	0.9905	0.9584	0.9860	0.9797
QCNN	0.9933	**0.9911**	**0.9987**	**0.9961**

of the 3D models belonging to the same class are very similar, and the leaned shape features of the 3D models belonging to different classes are distinctive. Thus, our proposed network can effectively learn the 3D shape features. Figure 11 shows some retrieval results that include the given query and the top 15 retrieved 3D models; mistakes are highlighted in red. We can see that there are two mistakes. When the query samples are "cat" and "women", there are two incorrectly retrieved 3D models respectively from the class "dog" and "man". For the two mistakes, both the query sample and the wrong retrieved 3D model look very similar.

Then, we compared our proposed "QCNN" method with other state-of-the-art methods [39] and the QCNN_origin method. From Table 3 we can see that our proposed method has the best performance with the FT, ST and DCG metrics. The NN metric of our proposed method is slightly lower than that of the "SD-GDM" method and the "SD-GDM-meshSIFT" method. Our proposed method outperforms the QCNN_origin method with the four metrics. Thus, our improved quadruplet loss function is better than the original quadruplet loss function, and our proposed method performs better than the other methods.

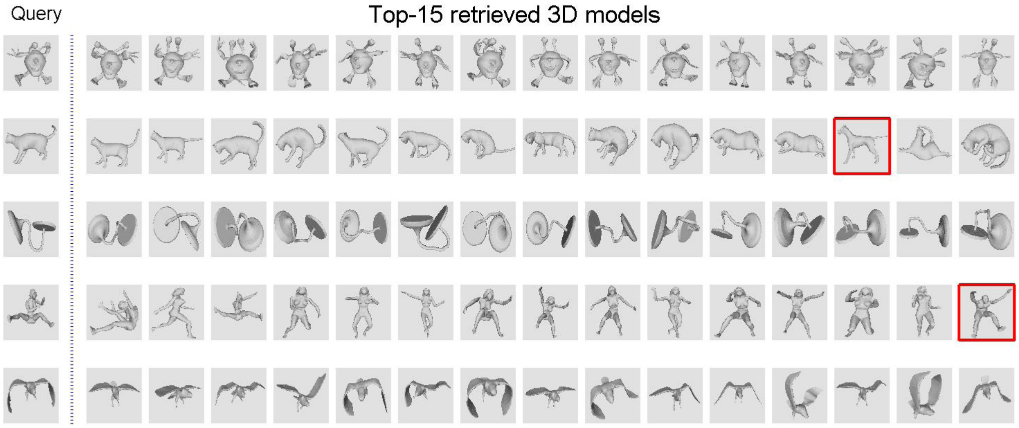


FIGURE 11. Examples showing the given query and the top 15 retrieval 3D models; mistakes are highlighted in red.

V. CONCLUSIONS

In this paper, we present a novel non-rigid 3D model retrieval method based on quadruplet convolutional neural networks. First, the WKS descriptor and the multi-energy shape distribution matrix are computed. Then, four branches of the proposed networks are trained using quadruplet samples. Finally, the final 3D shape feature is obtained using only one branch of the proposed networks. The contributions of the paper are that we design the construction method for the multi-energy shape distribution and the structure of the proposed networks. Furthermore, we propose an improved quadruplet loss function, which can decrease the risk of overfitting and underfitting. Our extensive experimental results show that the proposed method performs better than other state-of-the-art methods.

REFERENCES

[1] K. Lu, N. He, J. Xue, J. Dong, and L. Shao, "Learning view-model joint relevance for 3D object retrieval," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1449–1459, May 2015.

[2] Z. Lian *et al.*, "A comparison of methods for non-rigid 3D shape retrieval," *Pattern Recognit.*, vol. 46, no. 1, pp. 449–461, Jan. 2013.

[3] A. E. Johnson and M. Hebert, "Surface matching for object recognition in complex three-dimensional scenes," *Image Vis. Comput.*, vol. 16, nos. 9–10, pp. 635–651, 1998.

[4] 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.

[5] A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik, "Recognizing objects in range data using regional point descriptors," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 224–237.

[6] Y. Guo, F. Sohel, M. Bennamoun, M. Lu, and J. Wan, "Rotational projection statistics for 3D local surface description and object recognition," *Int. J. Comput. Vis.*, vol. 105, no. 1, pp. 63–86, 2013.

[7] H. Chen and B. Bhanu, "3D free-form object recognition in range images using local surface patches," *Pattern Recognit. Lett.*, vol. 28, no. 10, pp. 1252–1262, Jul. 2007.

[8] H. Chen and B. Bhanu, "Human ear recognition in 3D," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 718–737, Apr. 2007.

[9] R. B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz, "Aligning point cloud views using persistent feature histograms," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, Sep. 2008, pp. 3384–3391.

[10] R. B. Rusu, N. Blodow, and M. Beetz, "Fast point feature histograms (FPFH) for 3D registration," in *Proc. IEEE Int. Conf. Robot. Autom.*, Kobe, Japan, May 2009, pp. 3212–3217.

[11] F. Tombari, S. Salti, and L. Di Stefano, "Unique signatures of histograms for local surface description," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 356–369.

[12] S. Salti, F. Tombari, and L. Di Stefano, "SHOT: Unique signatures of histograms for surface and texture description," *Comput. Vis. Image Understand.*, vol. 125, pp. 251–264, Aug. 2014.

[13] 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.

[14] M. M. Bronstein and I. Kokkinos, "Scale-invariant heat kernel signatures for non-rigid shape recognition," in *Proc. IEEE Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1704–1711.

[15] 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*, Barcelona, Spain, Nov. 2011, pp. 1626–1633.

[16] A. M. Bronstein, M. M. Bronstein, L. J. Guibas, and M. Ovsjanikov, "Shape Google: Geometric words and expressions for invariant shape retrieval," *ACM Trans. Graph.*, vol. 30, no. 1, pp. 623–636, 2011.

[17] T. Hou, X. Hou, M. Zhong, and H. Qin, "Bag-of-feature-graphs: A new paradigm for non-rigid shape retrieval," in *Proc. Int. Conf. Pattern Recognit.*, Nov. 2012, pp. 1513–1516.

[18] Y. J. Li, "Non-rigid 3D model retrieval and tagging based on HKS," M.S. thesis, School Comput. Inf. Technol., Beijing Jiaotong Univ., Beijing, China, 2012.

[19] H. Zeng, H. Wang, S. Li, and W. Zeng, "Non-rigid 3D model retrieval based on weighted bags-of-phrases and LDA," in *Proc. 7th Chin. Conf. Pattern Recognit.*, 2016, pp. 449–460.

[20] H. Tabia, H. Laga, D. Picard, and P.-H. Gosselin, "Covariance descriptors for 3D shape matching and retrieval," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 4185–4192.

[21] G. Lavoué, "Combination of bag-of-words descriptors for robust partial shape retrieval," *Vis. Comput.*, vol. 28, no. 9, pp. 931–942, Sep. 2012.

[22] Z. Lian *et al.*, "SHREC'15 track: Non-rigid 3D shape retrieval," in *Proc. Eurograph. Workshop 3D Object Retr.*, 2015, pp. 107–120.

[23] R. Litman, A. Bronstein, M. Bronstein, and U. Castellani, "Supervised learning of bag-of-features shape descriptors using sparse coding," *Comput. Graph. Forum*, vol. 33, no. 5, pp. 127–136, 2014.

[24] A. Agathos, I. Pratikakis, P. Papadakis, S. Perantonis, P. Azariadis, and N. Sapidis, "Retrieval of 3D articulated objects using a graph-based representation," in *Proc. Eurographics Workshop 3D Object Retr.*, 2009, pp. 29–36.

[25] P. Papadakis, I. Pratikakis, T. Theoharis, G. Passalis, and S. J. Perantonis, "3D object retrieval using an efficient and compact hybrid shape descriptor," in *Proc. Eurograph. Workshop 3D Object Retr.*, 2008, pp. 9–16.

[26] H. Tabia, D. Picard, H. Laga, and P.-H. Gosselin, "Compact vectors of locally aggregated tensors for 3D shape retrieval," in *Proc. Eurograph. Workshop 3D Object Retr.*, 2013, pp. 17–24.

[27] S. Bu, Z. Liu, J. Han, J. Wu, and R. Ji, "Learning high-level feature by deep belief networks for 3-D model retrieval and recognition," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2154–2167, Dec. 2014.

[28] G. Dai, J. Xie, F. Zhu, and Y. Fang, "Learning a discriminative deformation-invariant 3D shape descriptor via many-to-one encoder," *Pattern Recognit. Lett.*, vol. 83, no. 11, pp. 330–338, Nov. 2016.

[29] J. Xie, F. Zhu, G. Dai, L. Shao, and Y. Fang, "Progressive shape-distribution-encoder for learning 3D shape representation," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1231–1242, Mar. 2017.

[30] J. Xie, G. Dai, F. Zhu, E. K. Wong, and Y. Fang, "DeepShape: Deep-learned shape descriptor for 3D shape retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1335–1345, Jul. 2017.

[31] D. Zhang, L. Zhao, D. Xu, and D. Lu, "Learning local feature descriptors with quadruplet ranking loss," in *Proc. CCF Chin. Conf. Comput. Vis.*, 2017, pp. 206–217.

[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[34] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 815–823.

[35] W. Chen, X. Chen, J. Zhang, and K. Huang, "Beyond triplet loss: A deep quadruplet network for person re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 403–412.

[36] V. Balntas, E. Johns, L. Tang, and K. Mikolajczyk. (2016). "PN-Net: Conjoined triple deep network for learning local image descriptors." [Online]. Available: https://arxiv.org/abs/1601.05030

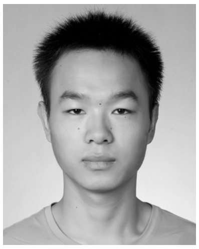
[37] B. G. V. Kumar, G. Carneiro, and I. Reid, "Learning local image descriptors with deep siamese and triplet convolutional networks by minimising global loss functions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 5385–5394.

[38] K. Siddiqi, J. Zhang, D. Macrini, A. Shokoufandeh, S. Bouix, and S. Dickinson, "Retrieving articulated 3-D models using medial surfaces," *Mach. Vis. Appl.*, vol. 19, no. 4, pp. 261–274, 2008.

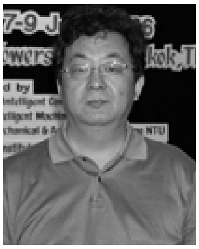
[39] Z. Lian *et al.*, "SHREC'11 Track: Shape retrieval on non-rigid 3D watertight meshes," in *Proc. Eurograph./ACM SIGGRAPH Symp. 3D Object Retr.*, 2011, pp. 1–10.

[40] J. Xie, G. Dai, and Y. Fang, "Deep multimetric learning for shape-based 3D model retrieval," *IEEE Trans. Multimedia*, vol. 19, no. 11, pp. 2463–2474, Nov. 2017.

HUI ZENG received the B.S. and M.S. degrees from Shandong University in 2001 and 2004, respectively, and the Ph.D. degree from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, in 2007. She is currently an Associate Professor with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, China. Her main research interests include computer vision, pattern recognition, and machine learning.



YANRONG LIU received the B.S. degree from Shanxi University in 2015. He is currently pursuing the M.S. degree with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, China. His current research interests include computer vision and pattern recognition.



JIWEI LIU received the B.Sc. degree from the University of Science and Technology of China in 1984 and the M.Sc. degree from the University of Science and Technology Beijing in 1997. He is currently an Associate Professor with the University of Science and Technology Beijing. His main research interests include image processing and pattern recognition.

DONGMEI FU received the M.S. degree from Northwest Polytechnical University in 1984 and the Ph.D. degree in automation science from the University of Science and Technology Beijing (USTB) in 2006. From 2002 to 2012, she took charge of several national projects about corrosion data mining and infrared image processing. She is currently a Professor and a Doctoral Supervisor with the School of Automation and Electrical Engineering, USTB, China. Her main research interests include automation control theory, image processing, and data mining.