Computers & Graphics 81 (2019) 82–91

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Special Section on CAD & Graphics 2019

A dimensional reduction guiding deep learning architecture for 3D shape retrieval

Zihao Wang[a], Hongwei Lin[a,b,*], Xiaofeng Yu[b], Yusuf Fatihu Hamza[a]

[a] School of Mathematics, Zhejiang University, Hangzhou 310027, China
[b] State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou 310058, China

ARTICLE INFO

Article history:
Received 8 March 2019
Revised 29 March 2019
Accepted 7 April 2019
Available online 27 April 2019

Keywords:
Shape retrieval
Shape descriptor
Dimensionality reduction
ResNet

ABSTRACT

The state-of-the-art shape descriptors are usually lengthy for gaining high retrieval precision. With the rapidly growing number of 3-dimensional models, the retrieval speed becomes a prominent problem in shape retrieval. In this paper, by exploiting the capabilities of the dimensionality reduction methods and the deep convolutional residual network (ResNet), we developed a method for extracting short shape descriptors (with just 2 real numbers, named 2-*descriptors*) from lengthy descriptors, while keeping or even improving the retrieval precision of the original lengthy descriptors. Specifically, an attraction and repulsion model is devised to strengthen the direct dimensionality reduction results. In this way, the dimensionality reduction results turn into desirable labels for the ResNet. Moreover, to extract the 2-descriptors using ResNet, we transformed it as a classification problem. For this purpose, the range of each component of the dimensionality reduction results (including two components in total) is uniformly divided into n intervals corresponding to n classes. Experiments on 3D shape retrieval show that our method not only accelerates the retrieval speed greatly but also improves the retrieval precisions of the original shape descriptors.

1. Introduction

3D shape retrieval is a crucial topic in computer vision and pattern recognition, which aims to retrieve the most relevant shapes to the query shape, based on the shape descriptors. In the field of shape retrieval, there are two competing ingredients: retrieval accuracy and retrieval speed. While high retrieval accuracy requires long shape descriptors, quick retrieval speed needs short descriptors. The state-of-the-art studies on shape retrieval mainly focus on improving accuracy, thus producing lots of lengthy shape descriptors, sometimes with high dimensions [1–3]. Although the accuracies of the retrieval algorithms based on these lengthy descriptors are acceptable, the retrieval speeds are usually very slow in huge shape database. More seriously, when the traditional index structures are employed for the fast query, the space and time requirements grow exponentially in the dimension [4]. Although there are indexing methods that avoid the curse of dimensionality [5], they do not support the exact search, and the dimension is also an important factor especially when facing a huge shape database. Therefore, due to the rapidly growing number of 3D shapes, it has become an urgent task to develop high-speed shape retrieval algorithms for the huge shape database, while keeping high retrieval accuracy.

In this paper, we develop an effective method for extracting short shape descriptors from lengthy or high dimensional descriptors, while keeping or even improving the retrieval accuracy. Specifically, given a shape database with classification, the descriptors of the shapes in the database consist of a point set in a high dimensional space. First, as a high dimensional data set, the set of shape descriptors are projected to a 2-dimensional plane using the dimensionality reduction (DR) methods. Moreover, an *attraction and repulsion model* is devised to reinforce the classification results in the plane, by enlarging the inter-class margin and reducing the intra-class variance. Then, taking the original shape descriptors as input, a deep convolutional residual network (ResNet) is constructed and trained. To improve the fitting precision of the ResNet, each component of the projected 2-dimensional shape descriptors is transformed into a vector, and taken as a label. Using the ResNet, a short shape descriptor with just two real numbers, named *2-descriptor*, can be extracted from lengthy or high dimensional shape descriptors. In our implementation, the shapeDNA (one-dimensional lengthy descriptor) and the Fourier shape descriptor (three-dimensional descriptor) are taken as the original shape descriptors, respectively, and two kinds of ResNets are

* Corresponding author at: School of Mathematics, Zhejiang University, Hangzhou 310027, China.
E-mail address: hwlin@zju.edu.cn (H. Lin).

https://doi.org/10.1016/j.cag.2019.04.002
0097-8493/

devised. It is shown by experiments that, the 2-descriptor generated by our method not only accelerated the retrieval speed significantly, but the retrieval precision is also improved as well, compared with the original lengthy or high dimensional descriptors.

The structure of this paper is as follows. In Section 2, some related work is briefly reviewed. In Section 3, the generation method of the 2-descriptor is introduced in detail, including the attraction and repulsion model and the construction and training of the ResNets. Moreover, the implementation details and results are presented in Section 4. Finally, Section 5 concludes the paper.

2. Related work

In this section, we briefly review the related work on 3 dimensional (3D) shape descriptors, dimension reduction (DR) methods, and descriptor learning methods by deep networks, respectively.

2.1. 3D shape descriptors

3D shape descriptors are compact discrete data to describe and represent 3D objects, usually represented as vectors or tensors, which provide an easy way to analyze and compare shapes. Because 3D shape descriptors encode geometrical or topological information of objects and are easy to process, they play a vital role in computer vision and graphics applications, including shape matching, recognition, and retrieval [6–8].

According to different levels in shape representation, 3D shape descriptors can be divided into two categories: global descriptors and local descriptors, which are surveyed in the following, respectively.

The global shape descriptors catch the information of the entire shape, and are generally invariant with respect to isometric transformation. An important topic on global shape descriptors focuses on describing shapes with only one real number, thus generating the scale shape descriptors, including *eccentricity* [9], *rectilinearity* [10], and *convexity* [11,12], etc. Although the study on scale shape descriptors is significant in theory, their retrieval accuracy is restricted. To improve the retrieval accuracy, recent work prefers lengthy shape descriptors, such as, light-field descriptor [13], graph based descriptor (Reeb Graph) [14], spectrum based descriptors (shapeDNA) [15], SD-GDM [1], and medial axis spectrum [16], etc. Although the retrieval accuracy is improved by these lengthy global shape descriptors, the retrieval speed is very slow, especially in the huge shape database.

On the other hand, local shape descriptors capture important geometric information on local regions of 3D shapes, and are frequently employed in local shape analysis such as partial shape retrieval [17,18], and shape correspondence [19,20]. Representative work on local shape descriptors includes spin images [21], intrinsic shape signatures [22], heat kernel signature [6], wave kernel signature [19], etc. Usually, the local shape descriptors have more complicated structure and longer length than those of the global shape descriptors, and strategies like partial aggregation and bag of features [7,23] can be used to combine them together.

2.2. Dimensionality reduction methods

In our method, the lengthy shape descriptors in high dimensional space are projected onto a plane using the DR methods. The basic motivation of DR is to reduce the complexity of data and extract more compact and representative information. With the increasing complexity and high dimension of data, DR has become a basic operation across a variety of fields, e.g., feature extraction, data visualization and high-dimensional nearest neighbor query.

Some classical DR methods regard the data point set to be on a manifold in a high dimensional space and devote to preserve the relative distances in local neighborhoods. The key ingredient in these DR methods is the way of local neighborhood construction and low-dimensional embedding. For example, Isomap [24] extends the classical multidimensional scaling (MDS) [25], and the pairwise geodesic distances between points are strived to be kept. In LLE [26], the linear structure in the local neighborhood of each data point is tried to be preserved. Laplacian Eigenmaps [27] tries to minimize the low-dimensional distance between nearby data points. TSNE [28] tries to keep the distribution of data instead of distance. Moreover, the objective of LTSA [29] is to hold the local coordinate representation.

However, the direct DR results are not discriminative enough for the ResNet learning. Therefore, we develop an attractive and repulsion model to enlarge the inter-class margin and reduce the intra-class variance. In this way, the discrimination of the DR results are improved, and the ResNet learning process is better guided.

2.3. Descriptor learning via deep networks

With the success of deep learning architecture in the area of computer vision, especially the pretty good results of recognition challenge on imageNet [30], there has been a significant trend to learn descriptors according to different tasks or objectives by deep networks. For 3D objects, there are mainly two kinds of methods to extract features through the deep learning architecture, which are elucidated in the following.

The first kind of methods directly process the representation data of 3D objects, which are represented as collections of multi-view images or voxels. In this way, the classical CNNs can be employed. Exploiting the rendered views, Su et al. [31] developed the multi-view CNN to combine input views of an object through a view pooling layer. Moreover, Huang et al. [32] extended the view-based method to per-point representation. The multi-view image based methods are highly informative, but they are not concise, and sensitive to isometric geometric transformations. Also, by the multi-view image based methods, the size of data is significantly increased. On the other hand, Wu et al. [33] proposed the 3D shapeNets, a convolutional deep belief network, which regards a 3D shape as a distributions of binary variables on a 3D voxel grid. These methods can take the latest advancements in image processing. However, 2D views ignore the global 3D geometry information, and the voxel-based representations are generally coarse to ensure a satisfied computational cost.

Different from the first kind of methods, the second kind of approaches handle the features of the 3D objects. From the existing features of 3D objects, more discriminative high-level shape features are extracted as the shape descriptors, using the deep network. In Ref. [34], two low-level local 3D shape descriptors are encoded into geometric bag-of-features, from which the high-level descriptors are learned by a deep belief network. Fang et al. [35] developed the many-to-one encoder to force the input heat shape descriptors of the same class to be mapped to a pre-computed unique target descriptor. Ben-Shabat et al. [36] use the continuous generalization of Fisher vectors to represent 3D point clouds in their convolutional neural network. In [37], Han et al. represent 3D meshes as local function energy distributions, and learn discriminative hierarchical features by a convolutional deep belief network.

3. Method

The goal of our work is to find a high discriminative low-dimensional representation of 3D shapes from existing shape descriptors.

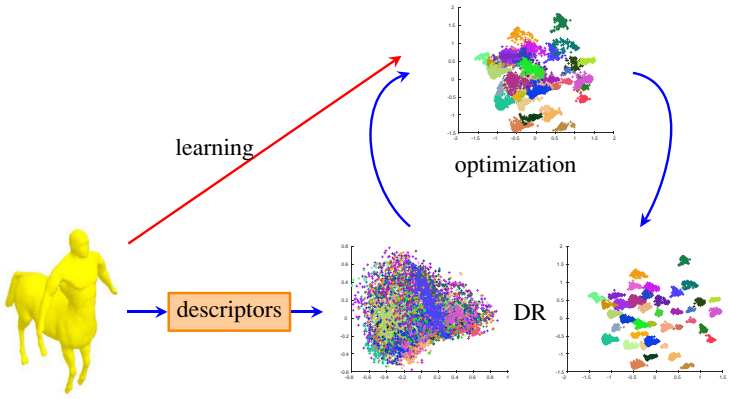

Fig. 1. The pipeline of our approach on shape retrieval.

Mathematically, for each kind of shape descriptor, we aim to find a function f that takes any descriptor X_s extracted from shape s as input and output a feature $Y_s = f(X_s) \in R^d$ of much lower dimensionality as the new representation of shape s. Through this function, descriptors of 3D shapes can be quickly mapped to low-dimensional features.

Our architecture (refer to Fig. 1) encompasses three distinct steps:

- Different shape descriptors are extracted to represent 3D objects. In the final step, these high-dimensional descriptors are taken as the inputs of the ResNet.
- We reduce the dimensionality of the space of each kind of 3D shape descriptors to d (in this paper, we select $d = 2$) such that 3D shapes can be represented in low-dimensional space, and devise an attraction and repulsion model to iteratively optimize the DR result. Then the result of the attraction and repulsion model is transformed as the labels of the ResNet for training. This procedure guarantees high speed retrieval while keeping an acceptable precision.
- We then develop ResNet for each kind of shape descriptor to learn the process of dimensionality reduction.

3.1. Shape descriptors

The choice of shape descriptors is one of the most critical factors that affects the final retrieval performance. Here we select three standard descriptors as inputs to show our architecture, and they are different in data format.

3.1.1. CSD

The CSD [38] is a combination of seven shape attributes. Each of the shapes attributes: eccentricity [9], rectilinearity [10], convexity [11,12], compactness [39] and fractal dimension [40] intuitively describes the attribute of 3D objects in a certain aspect. The descriptor is a seven dimensional vector.

3.1.2. ShapeDNA

The ShapeDNA [15] is the first spectral shape signature for non-rigid shape analysis. It is the normed beginning sequence of the Laplace spectrum. The work has shown that several geometrical and topological properties, like area, volume, Euler characteristic are determined by the spectrum of the Laplacian. Since the spectrum is isometric invariant, it can be useful in non-rigid shape comparison.

Mathematically, if we have the solutions of the Laplacian eigenvalue problem $\Delta u = -\lambda u$, the first N smallest non-zero eigenvalues $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ are taken as the shapeDNA. Here Δ denotes the Laplace–Beltrami operator on the surface. The shapeDNA descriptor is a high-dimensional vector of numbers.

3.1.3. The Fourier descriptor

To extract the Fourier descriptor, 3D geometric shapes are represented as discrete signed distance fields (SDF). The discrete SDF is a scalar field, where each grid point stores the distance to the surface of 3D shape, and the negative or positive value indicates that the grid point is outside or inside the 3D shape respectively. The grid size of SDF is $128 \times 128 \times 128$ with two extra cells of padding in both directions to reduce border artifacts. We then select the top m spectrum coefficients along each dimension to form a Fourier descriptor of grid size $m \times m \times m$, which is a tensor representation.

3.2. Dimensionality reduction to extract low-dimensional feature vectors

There are some considerations that we make to utilize the dimensional reduction method. As the examples we have mentioned in 2.1, most of the existing shape descriptors, either hand-craft or learned shape descriptors, are high-dimensional data. In the application of shape retrieval, we compare shapes by comparing corresponding shape descriptors. However, traditional index structures designed for the fast query are not suitable for high-dimensional data [4]. Approximate nearest neighbor search methods [5] speed up high-dimensional retrieval, but the accuracy is sacrificed and the dimension is also not to be neglected.

Based on the above discussion, we think that reducing the dimension of shape descriptors before retrieval is necessary, especially in the era of rapid data growth. Ideally, the points obtained through DR should ensure that similar objects are close to each other, and shapes belong to different categories are far apart. Therefore, we design an attraction and repulsion model to optimize the results of DR, and propose a learning based schema to obtain well behaved 2D points. The learning process is guided by a valid 2D representation from iterative optimization of the result of manifold learning method, which keeps the relative position of points in local neighborhoods before and after DR. After that, a deep neural network is applied to learn the mapping from shape descriptors to their corresponding 2D points.

3.3. Attraction and repulsion model

Existing manifold learning methods like Isomap [24], LLE [26] generally have good performance on high-dimension

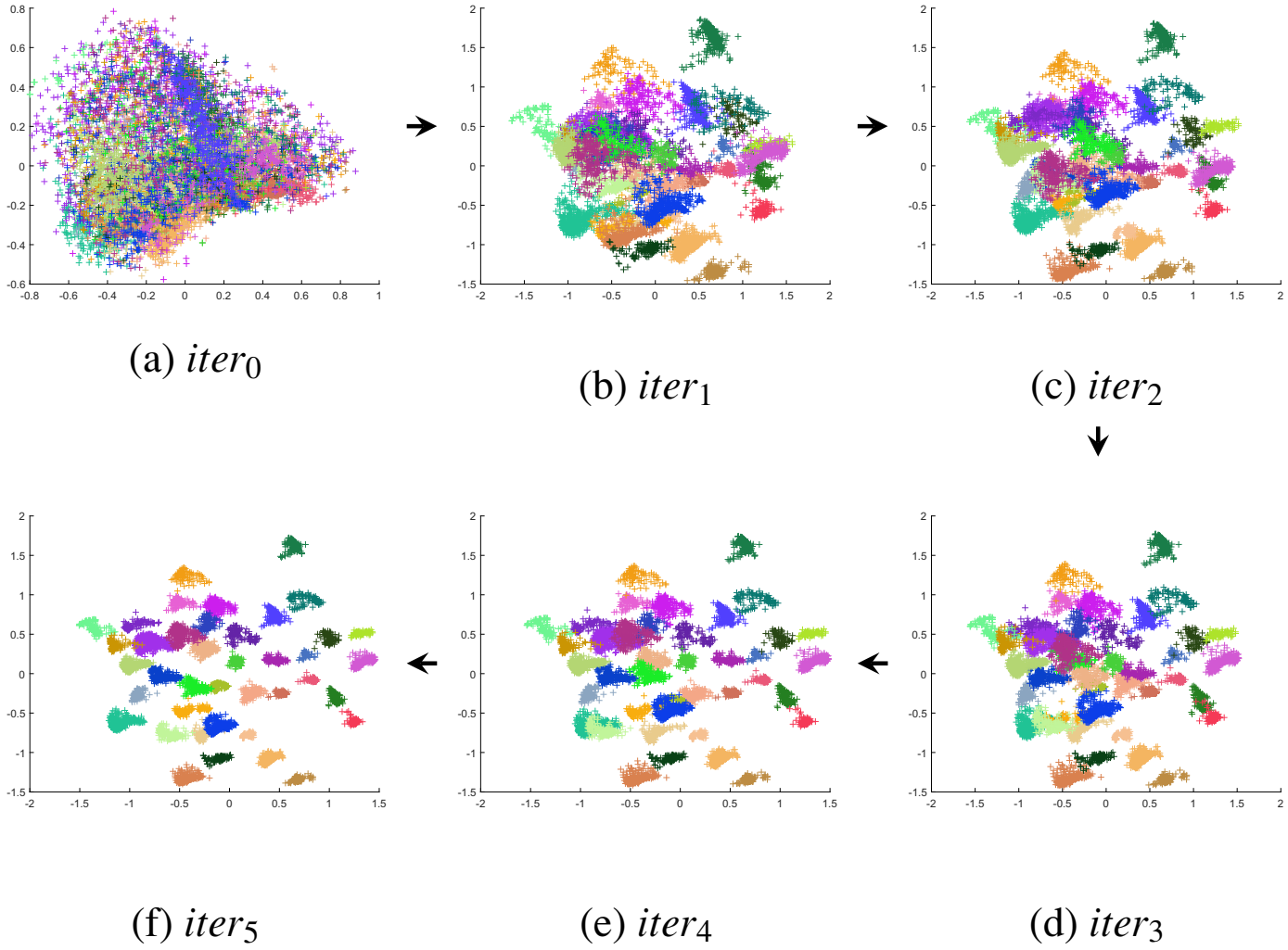

Fig. 2. The result of the attraction and repulsion process. The initial state($iter_0$) is the result of Isomap on the Fourier shape descriptors. Points in different color belong to different classes.

datasets with desirable low-dimensional embedding structures. When facing complex problems, there may exist overlapping instances in the low-dimensional subspace. A bright idea to solve the overlapping is to dispart features belong to different categories and keep relative intra-class distance at the same time, as shown in Fig. 2.

Thus, to train the best feature vectors such that the low dimensional representation of geometrically and semantically similar shapes are close to each other, we propose an attraction and repulsion post-process to iteratively adjust the result of dimensional reduction work in order to guide deep learning in the next section.

In detail, suppose the 3D descriptors for training are classified according to the shape categories, the capacity of the ith class is n_i, p_j^i, $j = 1, 2, \ldots, n_i$ denotes the jth low-dimensional feature vector of 3D objects that belongs to class i, and C_i denotes the ith class center. The adjustment process is an iterative process, and each iteration encompasses two steps:

- The repulsion process. This process aims to improve the discrimination among feature vectors belong to different classes. For each class i, the class center $C_i = \frac{\sum_j p_j^i}{n_i}$ is first calculated as the mean of the features. Then a transfer vector $v_i^P = \sum_j w_i^j (C_j - C_i)$ is calculated, which is regarded as the resultant of repulsive forces that all the other classes apply on class i. Here w_i^j is the weight that determines the magnitude of each force, and we set

$$w_i^j = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-\|C_j - C_i\|^2}{2\sigma^2}} & C_j \text{ is the } k\text{NN of } C_i, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

  which means that the magnitude of the force that acts on class i is negatively correlated with the corresponding distance between classes, and the effects of distant classes (not in the k nearest neighbor of class i) are ignored. Points which belong to class i are then updated by the same transfer vector v_i^P.
- The attraction process. This process aims to reduce the variation of each class. The points p_j^i, $j = 1, \ldots, n_i$ in each class i move toward the class center C_i with a displacement vector $v_i^A = \min\{e^{-len/2}, 1 - e^{-len/2}\}(C_i - p_i^j)$, where len denotes the maximum distance between class centers.

In this paper, the iteration is terminated if the maximum number of iterations is reached, or the coefficient of variation (the ratio of variance to mean of data points) no longer increase. Fig. 2 shows the performance of our adjustment process using the result of Isomap as the initial state. It can be seen that the adjustment of the dimensionality reduction process gradually improves the data overlapping and successfully enlarges the inter-class margin while reducing the intra-class variance.

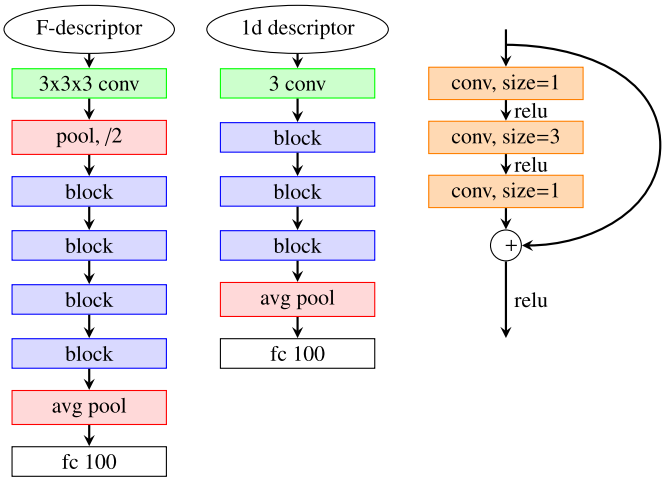

Fig. 3. The structure of the ResNet to process the Fourier descriptor (left), the CSD and shapeDNA (middle). The right subfigure shows the structure of block.

3.4. Dimensionality reduction training

This procedure aims to find the function that maps a high dimensional shape descriptor to a low-dimensional representation while keeping good recognition power. Since the shape descriptors we select in Section 3.1 to represent 3D objects are in different formats, we developed networks in different structures for the training procedure.

We use the architecture of the ResNet [41] to train the dimensionality reduction mapping. Residual learning techniques have exhibit improved performance in computer vision problems [41,42]. The basic idea of the residual network is to insert a shortcut connection that skip one or more layers. The shortcut connection is selected as the identity mapping and it turns the output of stack layers into the input plus a residual item. By optimizing the residual mappings, the network address the possible degradation problem result from superimposed layers.

The input and output of the network are pre-computed high-dimensional descriptors and high-recognition low-dimensional points, respectively, and the training procedure is the least squares fitting problem. Since the ResNet is more appropriate for classification rather than fitting, we transform the regression problem into multiple classification problems. Because the output dimension is $d = 2$, two classification networks are devised to train the mappings for the two dimension, i.e., the x and y coordinates, respectively, and labels are obtained from the output data. For the first dimension, the range of x-coordinates are uniformly divided into M intervals (left-open and right-closed). Then, we record the value of interval length Δ_x and the minimal value of x-coordinates min_x, and train a classification network of M classes according to the x-coordinates. If the x-coordinate lies in the ith interval, we set its class label of x as $label_x = i$ and it is represented as standard softmax variables in the net. The same process applies to y.

For the test data, the 2-descriptor (x', y') is extracted from the inverse of the above process, i.e., $x' = label_{x'} \cdot \Delta_x + \min_x$, $y' = label_{y'} \cdot \Delta_y + \min_y$. In this way, we get better fitting precision than directly training the regression problem.

As indicated in Fig. 3 left, our network for the Fourier descriptor consists of one convolutional layer, one max-pooling layer followed by four blocks, one average pooling layer and one fully connected layer with softmax. The convolution layers mostly have 3×3 filters, and the number of filters follows the rules in [41]. Fig. 3 right shows the building block, which is the stack of three convolution layers with a shortcut connection skipping them. For the jth dimension of the target value, the objective function of the proposed network is the cross entropy function:

$$argmin_{W,b}\left(-\sum_i Y_j^i \cdot log_2 h(s_i, W, b)\right), \tag{2}$$

where W is the weight matrix of the multiple-hidden-layer neural network, b is the bias matrix, s_i represents the ith training sample and Y_j^i in vector format is the jth coordinate of the corresponding target value, $h(s_i, W, b)$ in general is a non-linear mapping from the input s_i to the output.

For the CSD and shapeDNA, the two descriptors are in vector format. The network is designed to be similar to the network designed for the Fourier descriptor. Fig. 3 middle shows their network structure. The only difference is that there is no max-pooling layer after the first convolution layer, and all the operations are one dimensional.

4. Implementation, results and discussions

4.1. Datasets

We carry out several experiments for shape retrieval to assess the performance of our dimensionality reduction method. The Princeton ModelNet dataset is a collection of 3D CAD models for objects which contains 127,915 3D models from 662 categories. In our implementation, modelNet40, a clean and well annotated subset of modelNet, is employed to evaluate the effect of our DR method on the Fourier descriptor and CSD. ModelNet40 is provided on the ModelNet website [43] with 12311 shapes from 40 common categories. In our experiments, we use the same training and test split of ModelNet40 as in [33].

For the shapeDNA descriptor, we conduct related experiments on the Shrec'15 database. Shrec'15 contains 1200 watertight meshes that are derived from 50 original models. Each of the original model and 23 deformed versions of it constitute one category of this database. We randomly choose 1/3 models as the test set, and the remaining are used for training.

4.2. Implementation details

Before experimentation, we need to pre-process the raw data. When computing shapeDNA, the spectrum of non-manifold meshes maybe completely wrong. To solve this, one way is to convert these non-manifold meshes into the SDF zero-value isosurface or to compute shapeDNA on manifold datasets. And that is why we chose Shrec'15 instead of ModelNet to conduct related experiments.

In the initial DR stage, there are also things to be considered. When we apply traditional DR methods on the original shape descriptors, we choose cosine distance for shapeDNA, and Euclidean distance for CSD and Fourier descriptor. Also, several outliers of the initial DR result are deleted first to better guide deep learning. In our experiments, outliers should be removed only when we perform Isomap and MDS on the CSD.

In addition, before the attraction and repulsion process, the range of data should be scaled to a proper size. The weight of each repulsive force and its magnitude is determined by the distance between class centers. As shown in Fig. 4, the magnitude of repulsive force increases at the early stage and then decreases with distance increasing. If the distance between class centers is far enough, the magnitude of the repulsive force will be too small to have enough influence on the position of 2D points. Thus, the initial state of the iterative dimensional reduction needs a small range, which we generally choose to be not wider than $[-2, 2]$.

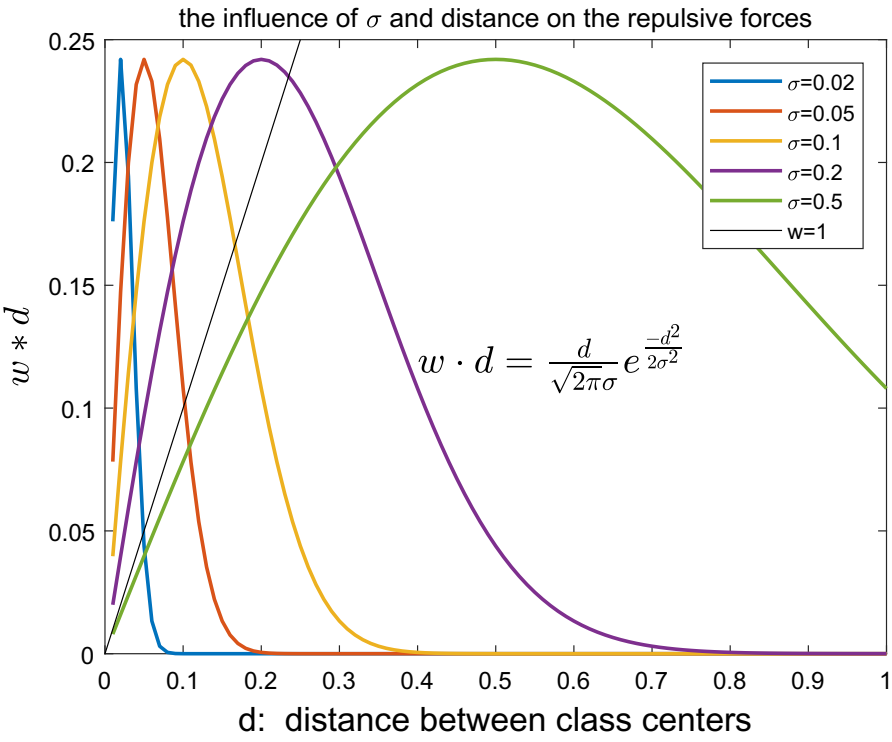

Fig. 4. The influence of σ and center distance on the actual repulsive forces.

Table 1
Parameters of the attraction and repulsion process.

CSD					Fourier				ShapeDNA			
DR	Isomap	MDS	Laplacian	tSNE	Isomap	MDS	Laplacian	tSNE	Isomap	MDS	Laplacian	tSNE
σ	0.2	0.1	0.18	0.038	0.1	0.1	0.1	0.1	0.2	0.15	0.25	0.2
k	15	12	15	10	10	12	12	12	12	18	12	18
scale	1	1	100	1/150	1	1	100	0.01	500	100	80	1/60

Fig. 4 shows the influence of the parameter σ in Eq. (1) on the repulsive forces. It is clear that the curves corresponding to a larger σ will have larger inflection point and wider scope of valid variable d. Since the distance between class centers mainly influences the weight w and magnitude of each repulsive force $w \cdot d$, we used different model parameters for each experiment. We list these parameters in Table 1, including the parameter σ in the weight, the number of nearest neighbor k and the scale of original DR results.

For the most part, small σ is an appropriate choice. On the one hand, to dispart 2D points belongs to different classes, closer classes supposed to have more influence on each other. On the other hand, it is not uncommon that in some experiments, several class centers are very close to each other (with class center distance $d < 0.1$), in this case, it is better to consider smaller σ.

The choice of M, the number of intervals mentioned in Section 3.4, is very crucial, M should be large enough to distinguish data from different classes. Moreover, oversizing M may split several classes into multiple smaller sets. In our experiments, the choice $M = 100$ is suitable in most cases.

In the experiments, we also test the time consumption of our architecture in the training and test phase. Before training, the time consuming of the attraction and repulsion model is determined by the size of datasets and the number of iterations, and we show the time cost of its initialization and iterations in Table 2.

As for the training process, it takes 53 min 28 s for the Fourier descriptor, 16 min 36 s for CSD and 20min10s for the shapeDNA to train 30,000 steps. In the test stage, we input shape descriptors of one shape into the ResNet, and record the time of running the trained model many times. As an average of 10,000 times, it respectively costs 8.5 ms, 8.6 ms and 6.7 ms for the three descriptors CSD, shapeDNA and Fourier descriptor to generate their corresponding 2-descriptors. This shows the high efficiency of our method.

4.3. Retrieval precision and retrieval speed

In the experiment, we first evaluate the performance of the extracted 2-descriptors from 3D shape retrieval experiments. In Fig. 5, the precision-recall curves of descriptors are shown to evaluate their retrieval precision. Here, Isomap is chosen as the initialization of DR to produce the 2-descriptors. Experiments about CSD and Fourier descriptor are performed on modelNet40, and those about shapeDNA are performed on Shrec'15. For convenience, we denote the 2-descriptor generated from descriptor A by DR methods B with the attraction and repulsion model as A-B. From the comparison, 2-descriptors CSD-Isomap and Fourier-Isomap have better performance than the original shape descriptors CSD and Fourier descriptor, and only the 2-descriptors shapeDNA-Isomap performs a little bit worse than the original shapeDNA. In Fig. 6, we compare the original descriptors with their corresponding 2-descriptors generated by Isomap on an example of shape retrieval. The retrieval behavior also supports the above conclusion from Fig. 5.

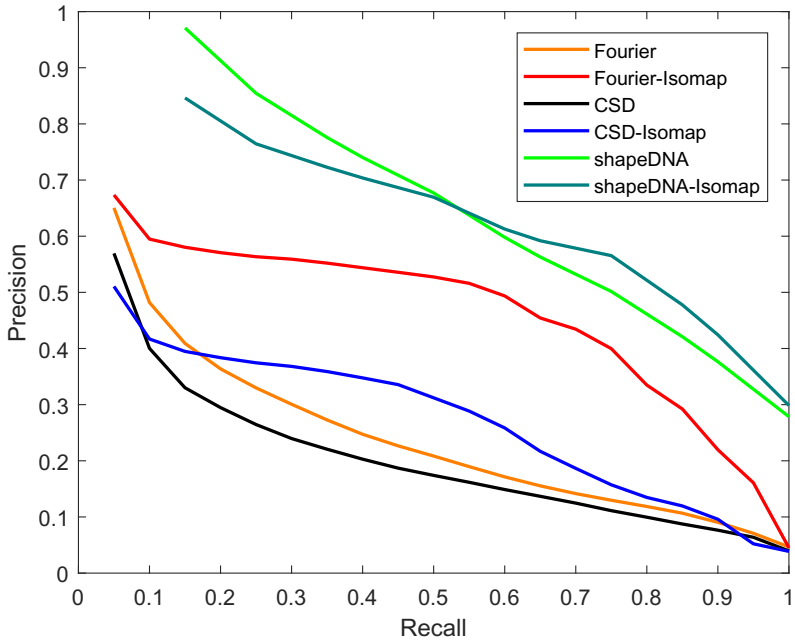

Fig. 5. The PR curve of the original shape descriptors and their corresponding 2-descriptors.

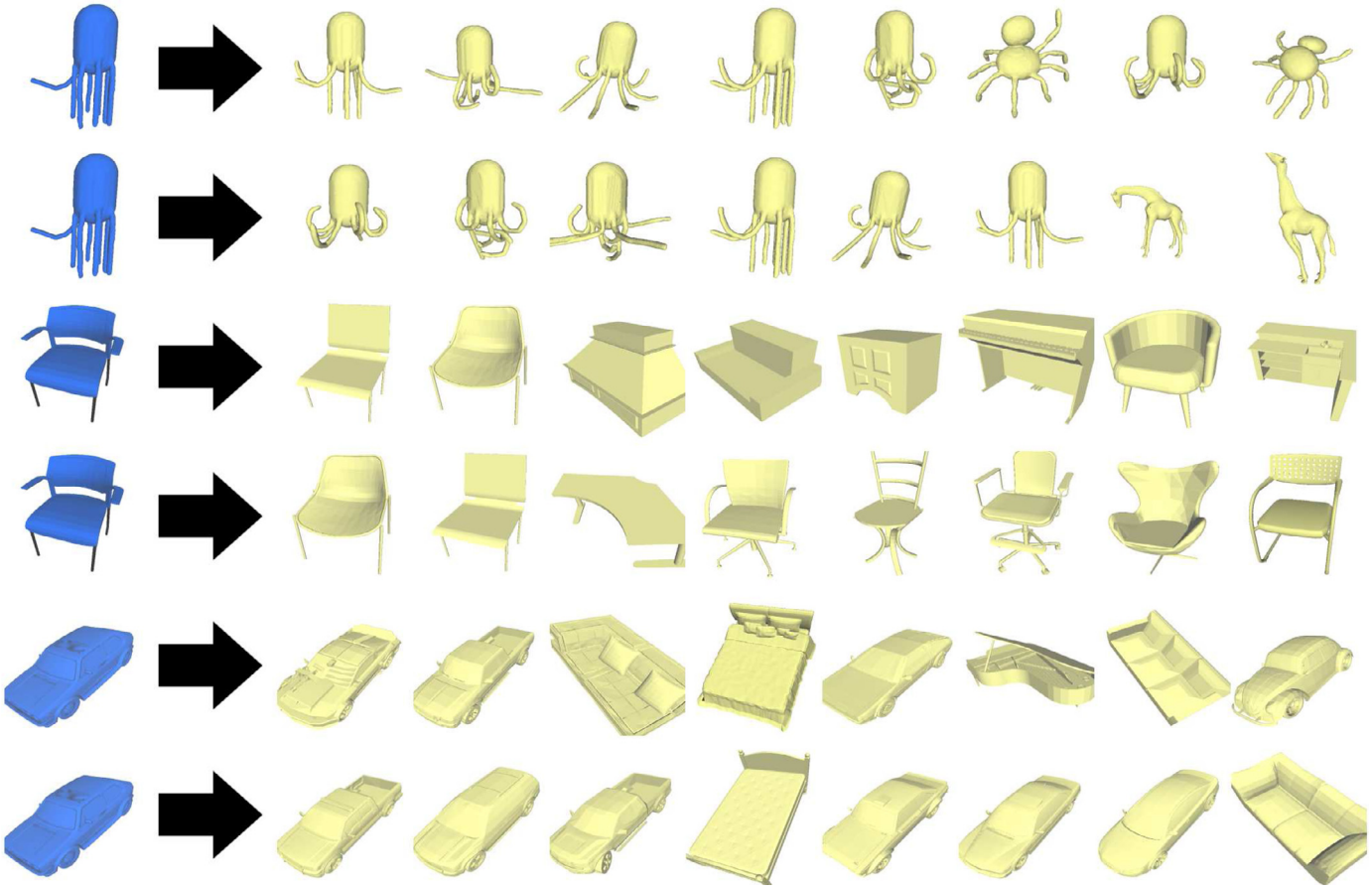

Fig. 6. The retrieval behavior of descriptors and their corresponding 2-descriptor generated by Isomap. The blue ones indicate the query shape, and yellow ones are the corresponding query results. The method from top to bottom in each row is: shapeDNA, shapeDNA-Isomap, Fourier, Fourier-Isomap, CSD, CSD-Isomap (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

Table 2
The time cost of the attraction and repulsion process in the training phase.

Time	Isomap|iterations	MDS|iterations	Laplacian|iterations	tSNE|iterations
Fourier	731.20 s|5.36 s	12.90 s|3.12 s	28.57 s|4.56 s	2273.77 s|3.88 s
CSD	759.46 s|4.63 s	0.36 s|4.62 s	14.58 s|3.82 s	2266.56 s|5.34 s
shapeDNA	0.65 s|0.04 s	0.06 s|0.04 s	0.10 s|0.04 s	12.08 s|0.04 s

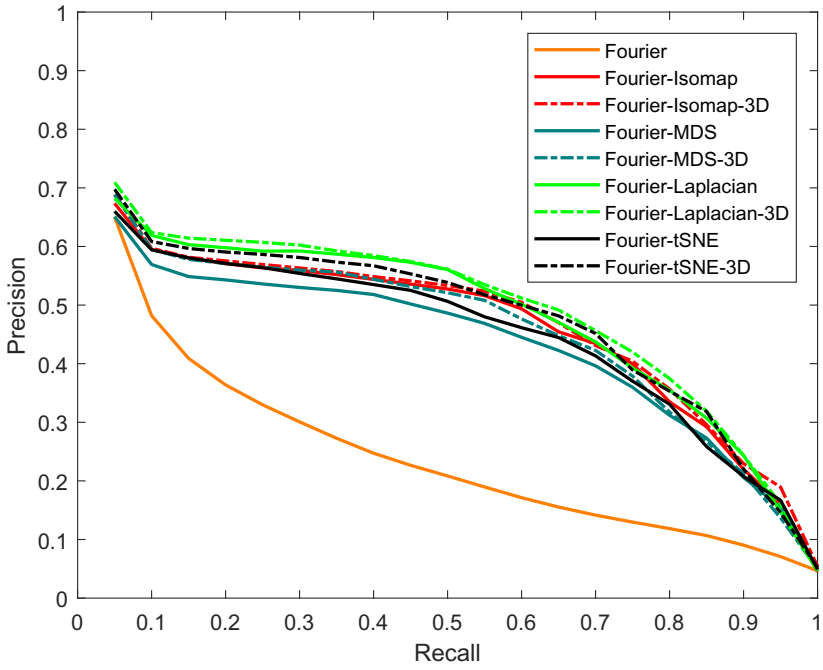

Fig. 7. The PR curve of the Fourier descriptor and its corresponding 2-descriptors and 3-descriptors.

Table 3
Retrieval time (average of 10,000 times 10-NN queries, unit: ms).

	CSD		shapeDNA		Fourier	
	2D retrieval	retrieval	2D retrieval	retrieval	2D retrieval	retrieval
Time	0.005	18.4	0.005	143.5	0.005	3053

It is worth noting that our framework that generates 2-descriptors can be extended to generate higher dimensional short descriptors, for example, 3-descriptor (short descriptor with just 3 real numbers). However, the attraction and repulsion model has successfully separated DR results (see Fig. 2), which makes the trained dimensionality reduction mapping able to keep the metric of descriptor space. As shown in Fig. 7, 2-descriptors and their corresponding 3-descriptors have similar retrieval performance. So, the increase in dimension does not lead to higher discriminatory ability. Here, we use different parameters σ and k for 3-descriptors, the parameter pairs (σ, k) for Fourier-Isomap-3D, Fourier-MDS-3D, Fourier-Laplacian-3D and Fourier-tSNE-3D are respectively (0.3,15), (0.1,12), (0.3,15) and (0.25,10).

We evaluate the retrieval time comparing the 2-descriptors with the three original descriptors in different formats. Results are listed in Table 3. In the experiment, we use the kd-tree as index structure for descriptor query and conduct an approximate query for high dimensional descriptors through an open source library called FLANN (Fast Library for Approximate Nearest Neighbors) [44]. The data in Table 3 are evaluated as the mean of 10,000 times 10-NN queries on the whole dataset ModelNet40. The results show that our 2-descriptor accelerates the retrieval speed greatly relative to high-dimensional shape descriptors.

The above experiments show the advantage of our 2-descriptors, that is to say, our trained DR methods successfully keep the retrieval ability of original shape descriptor while reducing the time consumption at the same time.

4.4. Influence of different DR methods

It is clear that the original descriptor and DR method influence the retrieval precision. Since our architecture mainly focus on the improvement of a fast query of high dimensional descriptors, we will now evaluate the effect of different DR methods on the recognition ability of shape descriptors, and compare the retrieval behavior of shape descriptors and their corresponding 2-descriptors using the four scores, NN, 1-Tier, 2-Tier, DCG.

We list their scores in Tables 4 and 5. Four kinds of DR methods: MDS [25], Isomap [24], Laplacian [27] and tSNE [28] are used to evaluate the effect of different DR methods. It is clear that nearly all 2-descriptors have improvements on the sore FT, ST and DCG of original descriptors. The 2-descriptors extracted from the Fourier descriptor especially show significant improvement, which indicates that the convolution structure is more suitable for our proposed Fourier descriptor than others. For the CSD and shapeDNA, they do not have well improvements like the Fourier

Table 4
Retrieval performance on the ModelNet40 test dataset.

Methods	NN(%)	FT(%)	ST(%)	DCG(%)
CSD	46.8	21.3	31.6	62.0
CSD-Isomap	35.9	30.2	36.7	63.6
CSD-MDS	34.6	29.2	35.7	63.0
CSD-Laplacian	28.3	29.5	35.7	63.3
CSD-tSNE	31.2	29.8	35.6	63.3
CSD-IsoKRR	27.8	22.6	35.4	61.2
CSD-S-LE	29.3	19.5	29.7	58.7
CSD-LDA	23.5	18.6	30.8	58.7
Fourier	61.1	24.1	32.6	64.2
Fourier-Isomap	53.3	46.3	52.7	73.5
Fourier-MDS	50.3	43.7	50.7	72.1
Fourier-Laplacian	51.3	48.1	53.4	74.3
Fourier-tSNE	54.8	45.8	51.3	72.9
Fourier-IsoKRR	27.1	19.3	29.6	58.1
Fourier-S-LE	41.1	25.3	34.7	63.0
Fourier-LDA	23.0	12.2	20.5	55.8

Table 5
Retrieval performance on the shrec15 test dataset.

Methods	NN(%)	FT(%)	ST(%)	DCG(%)
shapeDNA	81.3	48.6	60.6	73.9
shapeDNA-Isomap	66.5	56.5	60.8	71.9
shapeDNA-MDS	63.2	49.2	53.3	68.0
shapeDNA-Laplacian	53.0	47.2	57.1	65.8
shapeDNA-tSNE	66.0	51.2	59.5	70.9
shapeDNA-IsoKRR	55.5	43.5	53.6	65.5
shapeDNA-S-LE	58.5	37.9	50.7	63.0
shapeDNA-LDA	24.0	15.6	23.7	41.6

descriptor. Perhaps because the training is not convergent to a good minimal value in the learning process, but we can see that the 2-descriptors still keep the precision while reducing the dimension.

Besides, from Tables 4, 5 and Fig. 7, another conclusion is that 2-descriptors generated by different DR methods of the same shape descriptor have similar performance on shape retrieval. This means that our dimensionality reduction architecture is less influenced by the choice of DR methods. It can be seen that with more high discriminative hand-craft descriptors proposed in the future, the architecture we introduced is sure to be more meaningful for high-speed shape query.

4.5. The importance of the attraction and repulsion model

In Section 3.3, we introduce the attraction and repulsion model. It is clear that the discriminatory ability of the generated 2-descriptors is determined by hand-crafted shape descriptors and DR methods. Thus, a good DR result is helpful for training a mapping that reduces the dimension of high dimensional descriptors, while keeping the metric of the space of shape descriptors at the same time. Because the attraction and repulsion model improves the DR result, it is beneficial to training a good mapping. For demonstrating the importance of the attraction and repulsion model, we compare the retrieval performance of 2-descriptors generated with and without the attraction and repulsion model and show the results in Table 6, where the 2-descriptor generated from descriptor A by DR method B without the attraction and repulsion model is denoted by A-B-noAR, and the 2-descriptor generated from descriptor A by DR methods B with the attraction and repulsion model as A-B. It is clearly that the retrieval accuracy has significantly improved by the attraction and repulsion model.

Table 6
Comparison of the 2-descriptor generated with and without the attraction and repulsion model.

Methods	NN(%)	FT(%)	ST(%)	DCG(%)
Fourier-Isomap-noAR	23.5	15.3	25.2	55.5
Fourier-Isomap	53.3	46.3	52.7	73.5
Fourier-MDS-noAR	17.1	12.7	21.7	53.3
Fourier-MDS	50.3	43.7	50.7	72.1
Fourier-Laplacian-noAR	21.9	17.9	26.5	56.6
Fourier-Laplacian	51.3	48.1	53.4	74.3
Fourier-tSNE-noAR	44.1	23.8	32.4	62.0
Fourier-tSNE	54.8	45.8	51.3	72.9

4.6. Comparison to other supervised DR methods

The proposed method utilizes unsupervised DR methods as the initial state of our attraction and repulsion model. Although there exist supervised DR methods, our methods outperform them on the retrieval application.

We conduct retrieval experiments for three supervised DR methods IsoKRR [45], S-LE [46] and LDA [47], on the datasets mentioned above. The out-of-sample mappings are applied for the test data after the embedding of the training points are learned. The retrieval performance of these three methods is listed at the tail of Tables 4 and 5, from which we can see a clear advantage of our method over them gain.

The three supervised DR methods have relatively good performance on the training set. However, they behave worse on the test dataset. The reason may be that they only preserve the pairwise distance of the training data, and can not extends well to the whole data manifold. When dealing with out-of-sample sets, they have weaker generalization ability than our method.

5. Conclusion

In this paper, we proposed a new architecture for shape retrieval. Our architecture combines the advantages of dimensionality reduction and deep learning. Taking a lengthy shape descriptor as input, our method can generate a 2-descriptor with only two components effectively and efficiently. By the 2-descriptors, not only is the retrieval speed accelerated significantly, but the retrieval precision is improved as well, compared with the original lengthy shape descriptors. The 2-descriptors developed in this paper will play a more and more important role in large scale shape retrieval in huge shape data set with rapidly growing number of shapes.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant no. 61872316, and the National Key R&D Plan of China under Grant no. 2016YFB1001501.

References

[1] Smeets D, Fabry T, Hermans J, Vandermeulen D, Suetens P. Isometric deformation modelling for object recognition. In: Proceedings of the international conference on computer analysis of images and patterns. Springer; 2009. p. 757–65.
[2] Lian Z, Godil A, Sun X, Zhang H. Non-rigid 3d shape retrieval using multidimensional scaling and bag-of-features. In: Proceedings of the ICIP; 2010. p. 3181–4.
[3] Giachetti A, Lovato C. Radial symmetry detection and shape characterization with the multiscale area projection transform. In: Proceedings of the computer graphics forum, 31. Wiley Online Library; 2012. p. 1669–78.
[4] Marimont R, Shapiro M. Nearest neighbour searches and the curse of dimensionality. IMA J Appl Math 1979;24(1):59–70.
[5] Har-Peled S, Indyk P, Motwani R. Approximate nearest neighbor: towards removing the curse of dimensionality. Theory Comput 2012;8(1):321–50.

[6] Sun J, Ovsjanikov M, Guibas L. A concise and provably informative multi-scale signature based on heat diffusion. In: Proceedings of the computer graphics forum, 28. Wiley Online Library; 2009. p. 1383–92.
[7] Bronstein A, Bronstein M, Guibas L, Ovsjanikov M. Shape Google: geometric words and expressions for invariant shape retrieval. ACM Trans Graph 2011;30(1):1.
[8] Wohlkinger W, Vincze M. Ensemble of shape functions for 3d object classification. In: Proceedings of the 2011 IEEE international conference on robotics and biomimetics (ROBIO). IEEE; 2011. p. 2987–92.
[9] Papadakis P, Pratikakis I, Perantonis S, Theoharis T. Efficient 3d shape matching and retrieval using a concrete radialized spherical projection representation. Pattern Recognit 2007;40(9):2437–52.
[10] Lian Z, Rosin P, Sun X. Rectilinearity of 3d meshes. Int J Comput Vis 2010;89(2–3):130–51.
[11] Zunic J, Rosin P. A new convexity measure for polygons. IEEE Trans Pattern Anal Mach Intell 2004;26(7):923–34.
[12] Lian Z, Godil A, Rosin P, Sun X. A new convexity measurement for 3d meshes. In: Proceedings of the 2012 IEEE conference on computer vision and pattern recognition (CVPR). IEEE; 2012. p. 119–26.
[13] Chen D, Tian X, Shen Y, Ouhyoung M. On visual similarity based 3d model retrieval. In: Proceedings of the computer graphics forum, 22; 2003. p. 223–232.
[14] Biasotti S, Giorgi D, Spagnuolo M, Falcidieno B. Reeb graphs for shape analysis and applications. Theor Comput Sci 2008;392(1–3):5–22.
[15] Reuter M, Wolter F, Peinecke N. Laplace-Beltrami spectra as "Shape-DNA" of surfaces and solids. Comput Aided Des 2006;38(4):342–66.
[16] He S, Choi Y, Guo Y, Guo X, Wang W. A 3d shape descriptor based on spectral analysis of medial axis. Comput Aided Geom Des 2015;39:50–66.
[17] Liu Y, Zha H, Qin H. Shape topics: A compact representation and new algorithms for 3d partial shape retrieval. In: Proceedings of the 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06), 2. IEEE; 2006. p. 2025–32.
[18] Wan L, Zou C, Zhang H. Full and partial shape similarity through sparse descriptor reconstruction. Visual Comput 2017;33(12):1497–509.
[19] Aubry M, Schlickewei U, Cremers D. The wave kernel signature: A quantum mechanical approach to shape analysis. In: Proceedings of the 2011 IEEE International Conference on computer vision workshops (ICCV Workshops). IEEE; 2011. p. 1626–33.
[20] Remil O, Xie Q, Wu Q, Guo Y, Wang J. Intrinsic shape matching via tensor-based optimization. Comput Aided Des 2019;107:64–76.
[21] Johnson A, Hebert M. Using spin images for efficient object recognition in cluttered 3d scenes. IEEE Trans Pattern Anal Mach Intell 1999;21(5):433–49.
[22] Zhong Y. Intrinsic shape signatures: A shape descriptor for 3d object recognition. In: Proceedings of the IEEE international conference on computer vision workshops; 2010. p. 689–96.
[23] Kuang Z, Li Z, Liu Y, Zou C. Shape similarity assessment based on partial feature aggregation and ranking lists. Pattern Recognit Lett 2016;83:368–78.
[24] Tenenbaum J, De Silva V, Langford J. A global geometric framework for nonlinear dimensionality reduction. Science 2000;290(5500):2319–23.
[25] Borg I, Groenen P. Modern multidimensional scaling: theory and applications. J Educ Measur 2003;40(3):277–80.
[26] Roweis S, Saul L. Nonlinear dimensionality reduction by locally linear embedding. Science 2000;290(5500):2323–6.
[27] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Proceedings of the advances in neural information processing systems; 2002. p. 585–91.
[28] van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008;9(Nov):2579–605.
[29] Zhang Z, Zha H. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. SIAM J Sci Comput 2004;26(1):313–38.
[30] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. Int J Comput Vis 2015;115(3):211–52.
[31] Su H, Maji S, Kalogerakis E, Learned-Miller E. Multi-view convolutional neural networks for 3d shape recognition. In: Proceedings of the IEEE international conference on computer vision; 2015. p. 945–53.
[32] Huang H, Kalogerakis E, Chaudhuri S, Ceylan D, Kim V, Yumer E. Learning local shape descriptors from part correspondences with multiview convolutional networks. ACM Trans Graph (TOG) 2018;37(1):6.
[33] Wu Z, Song S, Khosla A, Yu F, Zhang L, Tang X, et al. 3d shapenets: a deep representation for volumetric shapes. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 1912–20.
[34] Bu S, Liu Z, Han J, Wu J, Ji R. Learning high level feature by deep belief networks for 3d model retrieval and recognition. IEEE Trans Multimed 2014;16(8):2154–67.
[35] Fang Y, Xie J, Dai G, Wang M, Zhu F, Xu T, et al. 3d deep shape descriptor. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2015. p. 2319–28.
[36] Ben-Shabat Y., Lindenbaum M., Fischer A. 3d point cloud classification and segmentation using 3d modified fisher vector representation for convolutional neural networks. arXiv:171108241 2017.
[37] Han Z, Liu Z, Han J, Vong C, Bu S, Chen C. Mesh convolutional restricted Boltzmann machines for unsupervised learning of features with structure preservation on 3-d meshes. IEEE Trans Neural Netw Learn Syst 2017;28(10):2268–81.
[38] Wang Z, Lin H, Xu C. Data driven composite shape descriptor design for shape retrieval with a vor-tree. Appl Math A J Chin Univ 2018;33(1):88–106.
[39] Bribiesca E. An easy measure of compactness for 2d and 3d shapes. Pattern Recognit 2008;41(2):543–54.
[40] Jiang J, Zhu W, Shi F, Zhang Y, Lin L, Jiang T. A robust and accurate algorithm for estimating the complexity of the cortical surface. J Neurosci Methods 2008;172(1):122–30.
[41] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–8.
[42] Kim J, Lee J, Lee K. Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 1646–54.
[43] http://modelnet.cs.princeton.edu; 2018. [accessed 12 May 2018].
[44] Muja M., Lowe D. FLANN: Fast library for approximate nearest neighbors. http://www.cs.ubc.ca/research/flann/; 2018. [accessed 12 September 2018].
[45] Orsenigo C, Vercellis C. Kernel ridge regression for out-of-sample mapping in supervised manifold learning. Expert Syst Appl 2012;39(9):7757–62.
[46] Raducanu B, Dornaika F. A supervised non-linear dimensionality reduction approach for manifold learning. Pattern Recognit 2012;45(6):2432–44.
[47] Mika S, Ratsch G, Weston J, Scholkopf B, Mullers K. Fisher discriminant analysis with kernels. In: Proceedings of the 1999 IEEE signal processing society workshop on neural networks for signal processing IX. IEEE; 1999. p. 41–8.