CROSS-MODAL GUIDANCE NETWORK FOR SKETCH-BASED 3D SHAPE RETRIEVAL

*Weidong Dai and Shuang Liang**

School of Software Engineering, Tongji University, China
{18wddai, shuangliang}@tongji.edu.cn.

ABSTRACT

The main challenge of sketch-based 3D shape retrieval is the large cross-modal differences between 2D sketches and 3D shapes. Most recent works employed two heterogeneous networks and a shared loss to directly map the features from different modalities to a common feature space, which failed to reduce the cross-modal differences effectively. In this paper, we propose a novel method that adopts a teacher-student strategy to learn an aligned cross-modal feature space indirectly. Specifically, our method first employs a classification network to learn the discriminative features of 3D shapes. Then, the pre-learned features are considered as a teacher to guide the feature learning of 2D sketches. In order to align the cross-modal features, 2D sketch features are transferred to the pre-learned 3D feature space. Our experiments on two benchmark datasets demonstrate that our method obtains superior retrieval performance than the state-of-the-art approaches.

***Index Terms*—** sketch, 3D shape retrieval, cross-modal differences, guidance network, feature alignment

1. INTRODUCTION

Compared with keywords and 3D shapes, sketches are intuitive and convenient for users to search for 3D shapes. Therefore, sketch-based 3D shape retrieval has been receiving more and more attention in the community of computer graphics and computer vision [1–3].

Despite of its flexibility and convenience, sketch-based 3D shape retrieval is a very challenging task due to the sever cross-modal differences between 2D sketches and 3D shapes. Fig.1 shows the significant differences between two the modalities. In recent years, the research of sketch-based 3D shape retrieval has achieved great progress because of the deep learning techniques [3–9]. Most of these deep learning methods are based on Siamese network architecture [10]. Specifically, they adopt two different Convolution Neural Networks (CNNs) to learn the features of sketches and 3D shapes respectively. Then the features from different modalities are directly mapped into a common embedding subspace under a shared loss function. However, these methods are unable to reduce the cross-modal differences effectively because the cross-modal features can not be aligned compactly in the common embedding subspace.

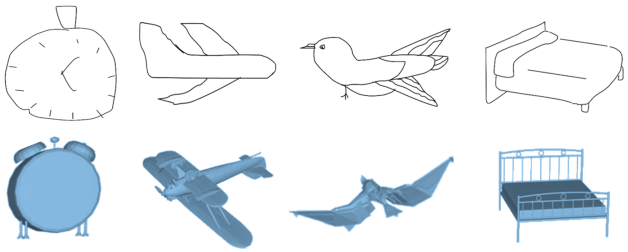

Fig. 1. Some examples of sketches and 3D shapes on two benchmark datasets.

Recently, distillation network and its variants (teacher-student networks) [11–14] are proposed to achieve model compression and knowledge transfer. Distillation network adopts a teacher-student strategy, where a complex and powerful teacher network is employed to teach a simple and weak student network on a given task. That is, the supervised information (e.g. the output of teacher network) is distilled out by a teacher network, which can be used to guide the training of a student network.

In this paper, we proposed a novel framework named Cross-modal Guidance Network (CGN), for sketch-based 3D shape retrieval. In our method, we separate the cross-modal retrieval task into two steps: classification of 3D shapes and feature transfer of 2D sketches. Fig.2 shows the pipeline of our proposed model. We first obtain the features of 3D shapes by a classification network. Then we build a transfer network for feature learning of sketches. The training of the transfer network is conducted under the guidance of the pre-learned features of 3D shapes and the constraint of our formulated guidance loss function. We regard the classification network of 3D shapes as a teacher network and the transfer network of sketches as a student network. As illustrated in Fig.3, our cross-modal guidance network transfer the features of sketches to the pre-learned feature space of 3D shapes, achieving the cross-modal feature alignment. As a result, the cross-modal differences is compensated. To summarize, our main contributions are as follows:

- We separate the task of sketch-based 3D shape retrieval

*Corresponding author.

978-1-7281-1331-9/20/$31.00 ©2020 IEEE

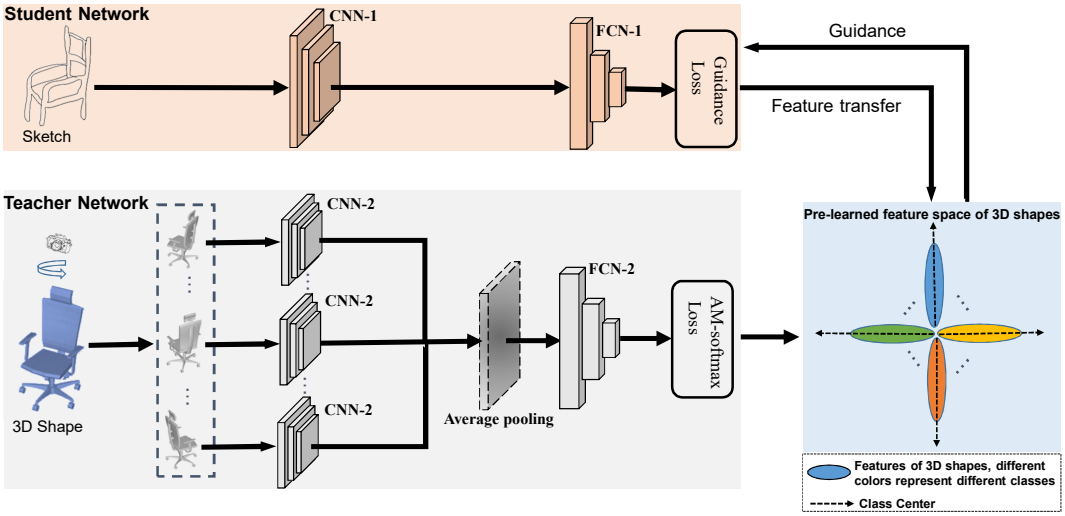

Fig. 2. The framework of our proposed method. Our proposed model consists of teacher/classification network, student/transfer network and the pre-learned feature space of 3D shapes. The training process consists of two steps:1.The pre-learned feature space of 3D shapes is learned by the teacher network. 2. The sketch features are learned by student network under the guidance of pre-learned feature space.

into two steps: classification of 3D shapes and feature transfer of 2D sketches. Then we adopt the distillation network (teacher-student strategy) to effectively combine these two steps. To our best knowledge, this work is the first one that incorporates distillation network into sketch-based 3D shape retrieval.

- We proposed a novel Cross-modal Guidance Network (CGN) for sketch-based 3D shape retrieval. It takes advantage of the pre-learned feature space of 3D shapes to guide the feature learning of sketches, which can effectively reduce the cross-modal differences.
- Our method outperforms the state-of-the-art methods of sketch-based 3D shape retrieval on two large benchmark datasets.

2. RELATED WORK

In this section, we introduce some related works of sketch-based 3D shape retrieval and distillation networks.

2.1. Sketch-based 3D shape retrieval

The early works proposed various methods based on the handdraft features for sketch-based 3D shapes retrieval [**?**, 1, 2, 15, 16]. In recent years, deep learning methods have been explored for sketch-based 3D shape retrieval [3–5, 7–9, 17]. Wang et al. [3] first used Siamese network architecture to address the cross-modal retrieval problem. Zhu et al. [4] proposed Pyramid Cross-Domain Neural Networks (PCDNN) to learn cross-modal features. In [6, 7], the deep correlated metric learning method was proposed. In [5], Wasserstein barycenters of multiple views of 3D shapes are learned to represent 3D shapes. Chen et al. [8] proposed to reduce cross-modal divergence via semantics preserving adversarial learning. He et al. [17] proposed the Triplet-Center loss (TCL) that combined the triplet loss and center loss. Lei et al. [9] proposed a Deep Point-to-Subspace method with an improved center loss function. However, most of them directly mapped the cross-modal features into a common embedding subspace, which failed to reduce the cross-modal differences effectively.

2.2. Distillation networks

Knowledge distillation with neural networks was pioneered by [11, 18], which is a transfer learning method that aims to improve the training of a student network by relying on knowledge learned from a powerful teacher network. Zagoruyko et al. [12] employed the distillation method to deliver the attention information of the teacher network to the student network. An improved knowledge distillation method for metric learning is proposed in [14]. Zhou et al. [13] proposed to use the booster net to supervise the learning of the light network. In this paper, we incorporate the distillation method into sketch-based 3D shape retrieval to reduce the cross-modal differences effectively.

3. PROPOSED METHOD

In this section, we first briefly introduce our motivation. Then we introduce the architecture of our proposed cross-modal guidance network and our formulated guidance loss function.

3.1. Motivation

Sketch-based 3D shape retrieval is a challenging task because it is difficult to learn a common feature space with

small cross-modal differences. However, the methods of image classification are mature due to the development of deep learning. This inspires us to use a classification network to learn a discriminative feature space for a single modality, then transfer the features of another modality to the pre-learned feature space, thus reducing the cross-modal differences between the two modalities. Moreover, as shown in Fig.1, compare with sketches that have limited information and high abstraction, 3D shapes have rich and geometrically realistic information for the classification task. Hench, it is more reasonable to choose the 3D shape modality for the classification network. Ablation study in Sec 4.3 verifies this.

3.2. Network architecture

We propose a cross-modal guidance network (CGN) for sketch-based 3D shape retrieval, which uses the classification results of 3D shapes to guide the feature learning of sketches. Fig.2 shows the architecture of our network. The framework consists of three parts: teacher network, student network and the pre-learned feature space of 3D shapes.

The pipeline of our method contains two steps. In the first step, we train the teacher network to learn the features of 3D shapes. Specifically, following the previous methods [5,8,19], we first adopt the widely used multi-view representation for 3D shapes. That is, a 3D shape is rendered to V views by placing V virtual cameras around the 3D shape evenly (V = 12 in this paper). Subsequently, the multiple views passed through CNN-2 separately, all branches of the CNN-2 share the same parameters. An average pooling layer is used to fuse the features output from CNN-2. We adopt the AM-softmax [20] as the loss function of the classification network. AM-softmax can enlarge the inter-class cosine distance and reduce the intra-class cosine distance. After the training of the teacher network, training data of 3D shapes pass through the teacher network again, and the features are extracted after the FCN-2 to obtain the pre-learned feature space of 3D shapes. Finally, we obtain the class centers by calculating the average value of the features in the same class. In the second step, we train the student network (*i.e.* transfer network) under the guidance of the pre-learned class centers of 3D shapes, which transfers the features of sketches to the pre-learned feature space under the supervision of our formulated guidance loss function. The loss function will be discussed in the next subsection.

In the testing phase, the features of query sketches and galley 3D shapes are extracted after FCN-1 and FCN-2 respectively, then the similarity of the query sketches and galley 3D shapes is calculated and ranked to obtain the rank list. We note that the cosine distance is required for the similarity.

3.3. Guidance loss function

To effectively constrain the feature transfer of sketches, we formulate a guidance loss to optimize the student network.

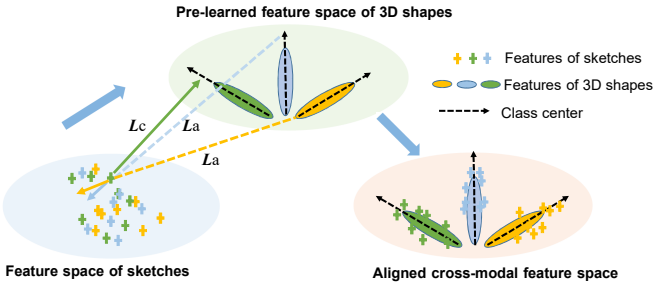

Fig. 3. A toy illustration of the cross-modal feature transfer under the guidance of pre-learned feature space of 3D shapes and the constraints of guidance loss function. The different colors represent different classes.

Our goal is to transfer the features of sketches to the pre-learned feature space of 3D shapes. At the same time, we also need to ensure that the sketch features are aligned with the class centers that has the same semantic information. Based on this goal, the loss function is formulated as:

$$\mathbf{L}_G = \mathbf{L}_c - \lambda \mathbf{L}_a \tag{1}$$

Where the L_G is our guidance loss. The L_c is the cosine distances between the features of sketches and the pre-learned class centers of 3D shapes in the same class, while the L_a represents the sum of cosine distance between the features of sketches and other class centers of 3D shapes. λ is a hyper-parameter to balance the L_c and L_a. In our experiments, the optimal value of λ is 0.1. We note that we calculate the L_c and L_a based on mini-batch sketches.

Specifically, the L_c is formulated as :

$$\mathbf{L_c} = \frac{1}{M}\sum_{i=1}^{M}\frac{\mathbf{f}_i \cdot \mathbf{c}_{y^i}}{\|\mathbf{f}_i\|_2\|\mathbf{c}_{y^i}\|_2} \tag{2}$$

the L_a is defined as:

$$\mathbf{L_a} = \frac{1}{M}\sum_{i=1}^{M}\sum_{j=1,j\neq y^i}^{N}\frac{\mathbf{f}_i \cdot \mathbf{c}_j}{\|\mathbf{f}_i\|_2\|\mathbf{c}_j\|_2} \tag{3}$$

In Eq. 2 and Eq. 3, the M is the size of a mini-batch, N is the number of classes. $\mathbf{f}_i$ denotes the feature vector of sketch, and y^i denotes the label of i-th sample. And $\{\mathbf{c_1}, \mathbf{c_2}, ..., \mathbf{c_N}\}$ are the pre-learned class center vectors of 3D shapes.

As shown in the Fig.3, our guidance loss function aims to cluster the features of sketches toward the class centers of 3D shapes in the same class and meanwhile make the features away from the class centers of different classes. In short, it constrains the learning of sketches to transfer the sketch features to the pre-learned feature space of 3D shapes, which achieves the semantic alignment of cross-modal features. Consequently, the cross-modal differences between 2D sketches and 3D shapes is reduced.

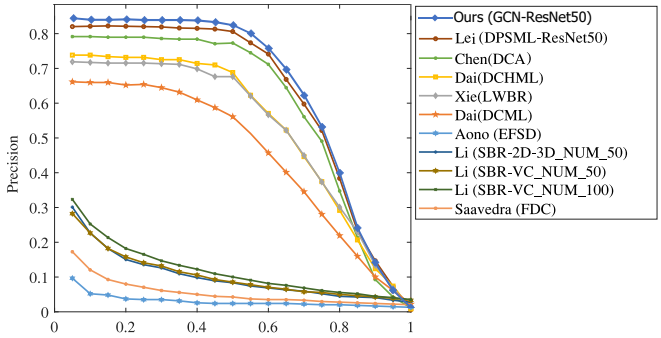

Fig. 4. The precision-recall curves of previous methods and our method on SHREC 2013.

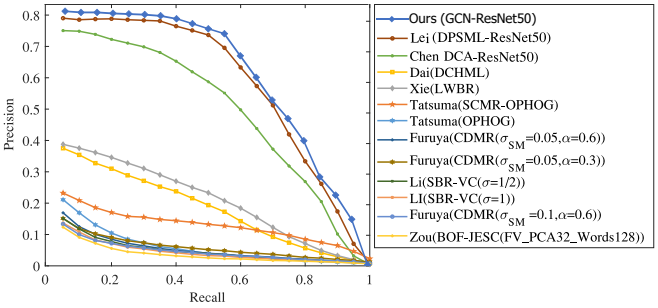

Fig. 5. The precision-recall curves of previous methods and our method on SHREC 2014.

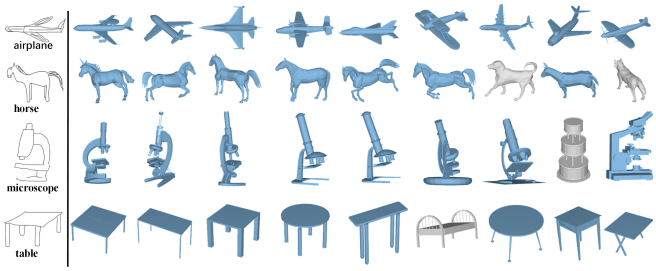

Fig. 6. Retrieval examples of SHREC 2014 dataset. Blue denotes the correct retrievals.

4. EXPERIMENTS

In this section, we introduce two benchmark datasets and evaluate the performance of our proposed approach on them.

4.1. Datasets and experimental settings

We perform our experiments on two wildly uesd benchmark datasets, **SHREC 2013** [1] and **SHREC 2014** [2]. Fig.1 shows some samples of 2D sketches and corresponding 3D shapes from the two datasets.

SHREC 2013. It is a benchmark for sketch-based 3D shape retrieval based on Princeton shape benchmark (PSB) [21]. The dataset contains 7,200 hand-drawn sketches and 1,258 3D shapes, which are divided into 90 classes. For the 3D shapes, the numbers of distinct classes are different, about 14 on average. There are 80 images for each sketch class, where 50 images are for training and 30 are for test.

SHREC 2014. It is a larger dataset than SHREC 2013. The data set contains a total of 13,680 sketches and 8987 3D shapes, which are grouped into 171 classes. There are approximately 53 3D shapes for each class. Similar to SHREC 2013, sketches collection also has 80 images for each class, which are further divided into 50 training images and 30 test images. Due to the more categories and the larger intra-class variations, the SHREC 2014 dataset is quite challenging.

Experimental settings. Our experimental codes are implemented in PyTorch. For fair and comprehensive comparison with the above methods, we employ several backbones for training and evaluation, including AlexNet [22], VGG-16 [23], VGG-19 [23], ResNet-50 [24]. Following the previous methods [3, 8, 9, 17], we use the same backbones for the networks of sketches and 3D shapes. All the backbones are pre-trained on ImageNet [22]. Specifically, for the CNNs (CNN-1 and CNN-2), we adopt the layers of AlexNet before "fc7" layer(inclusive), the layers of VGG1-16 before "fc7" layer (inclusive) and the layers of ResNet-50 before the "pooling5" layer (inclusive). The FCNs (FCN-1 and FCN-2) consist of fully connected layers. (*i.e.*, 4096-1024-256 for Alexnet, VGG-16 and VGG-19, 2048-1024-256 for ResNet-50.) The initial learning rate is set to be 0.001, it decays by 0.9 after every 10 epochs. The maximum epoch number is set to 90. The SGD is employed as an optimizer whose momentum and weight decay are set to 0.9 and 1e-4.

4.2. Experiment results on two benchmark datasets

On the two benchmark datasets, we compare our CGN method with state-of-the-art methods, including the CDMR [15], SBR-VC [1], DCML [6], Siamese [3], LWBR [5], DCHML [7], TCL [17], DCA [8] and the DPSML [9]. We use the standard performance metrics for a comprehensive evaluation, which includes Precision-Recall (PR) curve, Nearest neighbor (NN), First Tier (FT), Second Tier(ST), E-Measure (E), Discounted Cumulated Gain (DCG) and mean Average Precision (mAP) [16].

Retrieval Performance on SHREC 2013. Fig.4 shows the PR curves of the proposed method (ResNet-50 backbone) and compared approaches on the SHREC 2013 dataset. As illustrated, one can see that the performance of the proposed CGN method is significantly superior to those of compared models. We also report NN, FT, ST, E, DCG and mAP of various methods on Table 1. It can be seen that our proposed method outperforms state-of-the-art methods for all the evaluation metrics under the same CNN backbones. Specifically, our method outperforms the current state-of-the-art method by a gain of 3.2% mAP with AlexNet backbone. Even compared with the best ResNet-50 results achieved by DPSML [9], our VGG-19 results exceed them in some eval-

Table 1. The performance (%) on SHREC'13 dataset.

Method	Backbone	NN	FT	ST	EE	DCG	mAP
CMDR [15]	-	27.9	20.3	29.6	16.6	45.8	25.0
SBR-VC [1]	-	16.4	9.7	14.9	8.5	34.8	11.6
Siamese [3]	3-layer CNN	40.5	40.3	54.8	28.7	60.7	46.9
DCML [6]	AlexNet	65.0	63.4	71.9	34.8	76.6	67.4
LWBR [5]	AlexNet	71.2	72.5	78.5	36.9	81.4	75.2
DCHML [7]	AlexNet	73.0	71.5	77.3	36.8	81.6	74.4
TCL [17]	Alexnet	76.3	78.7	84.9	39.2	85.4	80.7
DCA [8]	Resnet50	78.3	79.6	82.9	37.6	85.6	81.3
DPSML [9]	AlexNet	74.1	76.1	82.1	38.5	83.6	78.5
	VGG19	80.1	81.6	85.2	39.8	87.0	83.1
	ResNet50	81.9	83.4	87.5	41.5	89.2	85.7
Ours(CGN)	AlexNet	**77.0**	**80.1**	**84.7**	**39.9**	**86.1**	**81.7**
	VGG19	**81.9**	**83.0**	**87.5**	**41.8**	**88.6**	**84.9**
	ResNet50	**83.2**	**85.3**	**90.2**	**41.9**	**90.1**	**87.0**

Table 2. The performance (%) on SHREC'14 dataset.

Method	Backbone	NN	FT	ST	EE	DCG	mAP
CMDR [15]	-	10.9	5.7	8.9	4.1	32.8	5.4
SBR-VC [1]	-	9.5	5.0	8.1	3.7	31.9	5.0
Siamese [3]	3-layer CNN	23.9	21.2	31.6	14.0	49.6	22.8
DCML [6]	AlexNet	27.2	27.5	34.5	17.1	49.8	28.6
LWBR [5]	AlexNet	40.3	37.8	45.5	23.6	58.1	40.1
DCHML [7]	AlexNet	40.3	32.9	39.4	20.1	54.4	33.6
TCL [17]	Alexnet	58.5	45.5	53.9	27.5	66.6	47.7
DCA [8]	AlexNet	49.8	46.4	51.3	29.4	62.7	50.2
	VGG16	68.2	69.8	72.3	37.5	78.3	71.1
	ResNet50	77.0	78.9	82.3	39.8	85.9	80.3
DPSML [9]	AlexNet	67.7	73.2	79.5	37.9	83.0	75.1
	VGG19	74.8	78.5	83.9	40.6	86.6	80.0
	ResNet50	77.4	79.8	84.9	41.5	87.7	81.3
Ours(CGN)	AlexNet	**73.4**	**74.3**	**80.8**	**39.1**	**84.4**	**77.8**
	VGG16	**76.9**	**78.9**	**83.1**	**41.2**	**86.9**	**80.8**
	VGG19	**77.5**	**79.2**	**84.4**	**41.3**	**87.1**	**81.5**
	ResNet50	**78.9**	**81.1**	**85.0**	**41.8**	**88.1**	**83.0**

uation metrics. For instance, our 41.8% EE is higher than its 41.5% EE. Moreover, our CGN method with ResNet-50 backbone significantly beats the state-of-the-art method in all evaluation metrics.

Retrieval Performance on SHREC 2014. SHREC 2014 dataset is more difficult since it has more classes and larger variations within each class. We also compare our proposed method with the previous methods on this dataset. The PR curves of various methods on SHREC 2014 are plotted in Fig.5, which reveals that our method beats state-of-the-art methods on the SHREC 2014 dataset. Table 2 also lists the comprehensive evaluation of our proposed CGN method and the previous methods. The tables shows that our proposed method also outperforms all the previous methods in all evaluation metrics on SHREC 2014 dataset. For example, Our CGN method with ResNet-50 reaches 82.7% mAP, which is 1.7%, 2.7% higher than DPSML [9], DCA [8]. The performance demonstrates that our method is stable for difficult large-scale datasets.

Fig.6 shows some retrieval examples of SHREC 2014 dataset. The query sketches are listed on the left side, and their retrieved top ten 3D shapes are listed on the right side according to their ranking order. As shown in it, our proposed method achieves promising results for the classes of examples.

4.3. Ablation study

We also conduct more experiments to explore which modality is more suitable to be the teacher. Specifically, we change the student network and teacher network. *i.e.* the network of sketches is the teacher and the network of 3D shapes is the student. And the other settings are the same as before. The Table 3 shows the comparison of the performance on SHREC 2014 dataset. The results demonstrate the clear superiority of our proposed method. The reason is that the hand-drawn sketches are very abstract, the pre-learned features of sketches have large intra-class distance and small inter-class distance, which is unable to effectively guide the feature transfer of 3D shapes to obtain a discriminative cross-modal feature space. Note that, we only report the results using ResNet-50 as the backbone. It is believed that, the other CNN backbones share similar performance trend.

Table 3. Ablation study on SHREC'14 dataset.

Teacher-student	NN	FT	ST	EE	DCG	mAP
3D-sketch	**78.9**	**81.1**	**85.0**	**41.8**	**88.1**	**83.0**
sketch-3D	77.1	79.1	82.7	41.3	86.0	80.9

5. CONCLUSION

In this paper, we proposed a novel cross-modal guidance network for sketch-based 3D shape retrieval, which adopts the idea of distillation networks to separate the cross-modal retrieval into a classification task and a feature transfer task. Our method uses the pre-learned features of 3D shapes to guide the feature learning of sketches. As a result, the features of sketches transfer to the feature space of 3D shapes, resulting in the reduction of cross modal differences. The results of our experiments outperform the state-of-the-art approaches.

Acknowledgments

This work is supported by Shanghai Natural Science Foundation (No. 19ZR1461200) and National Natural Science Foundation of China (No. 61976159). The authors would also like to thank the anonymous reviewers for their valuable comments and suggestions.

6. REFERENCES

[1] Bo Li, Yijuan Lu, Afzal Godil, Tobias Schreck, Masaki Aono, Henry Johan, Jose M Saavedra, and Shoki Tashiro, "Shrec'13 track: large scale sketch-based 3d shape retrieval," in *Eurographics Workshop on 3D Object Retrieval*, 2013, pp. 89–96.

[2] Bo Li, Yijuan Lu, Chunyuan Li, Afzal Godil, Tobias Schreck, Masaki Aono, Martin Burtscher, Hongbo Fu, Takahiko Furuya, Henry Johan, et al., "Shrec'14 track: Extended large scale sketch-based 3d shape retrieval," in *Eurographics workshop on 3D object retrieval*, 2014, pp. 121–130.

[3] Fang Wang, Le Kang, and Yi Li, "Sketch-based 3d shape retrieval using convolutional neural networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1875–1883.

[4] Fan Zhu, Jin Xie, and Yi Fang, "Learning cross-domain neural networks for sketch-based 3d shape retrieval," in *AAAI Conference on Artificial Intelligence*, 2016, pp. 3683–3689.

[5] Jin Xie, Guoxian Dai, Fan Zhu, and Yi Fang, "Learning barycentric representations of 3d shapes for sketch-based 3d shape retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5068–5076.

[6] Guoxian Dai, Jin Xie, Fan Zhu, and Yi Fang, "Deep correlated metric learning for sketch-based 3d shape retrieval," in *AAAI Conference on Artificial Intelligence*, 2017, pp. 4002–4008.

[7] Guoxian Dai, Jin Xie, and Yi Fang, "Deep correlated holistic metric learning for sketch-based 3d shape retrieval," *IEEE Transactions on Image Processing*, vol. 27, pp. 3374–3386, 2018.

[8] Jiaxin Chen and Yi Fang, "Deep cross-modality adaptation via semantics preserving adversarial learning for sketch-based 3d shape retrieval," in *The European Conference on Computer Vision*, 2018, pp. 605–620.

[9] Yinjie Lei, Ziqin Zhou, Pingping Zhang, Yulan Guo, Zijun Ma, and Lingqiao Liu, "Deep point-to-subspace metric learning for sketch-based 3d shape retrieval," *Pattern Recognition*, vol. 96, pp. 106–116, 2019.

[10] Sumit Chopra, Raia Hadsell, Yann LeCun, et al., "Learning a similarity metric discriminatively, with application to face verification," in *IEEE international conference on computer vision*, 2005, pp. 539–546.

[11] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[12] Sergey Zagoruyko and Nikos Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer," *arXiv preprint arXiv:1612.03928*, 2016.

[13] Guorui Zhou, Ying Fan, Runpeng Cui, Weijie Bian, Xiaoqiang Zhu, and Kun Gai, "Rocket launching: A universal and efficient framework for training well-performing light net," in *AAAI Conference on Artificial Intelligence*, 2018, pp. 3786–3790.

[14] Yuntao Chen, Naiyan Wang, and Zhaoxiang Zhang, "Darkrank: Accelerating deep metric learning via cross sample similarities transfer," in *AAAI Conference on Artificial Intelligence*, 2018, pp. 4213–4217.

[15] Takahiko Furuya and Ryutarou Ohbuchi, "Ranking on cross-domain manifold for sketch-based 3d model retrieval," in *2013 International Conference on Cyberworlds*, 2013, pp. 274–281.

[16] Bo Li, Yijuan Lu, Afzal Godil, Tobias Schreck, Benjamin Bustos, Alfredo Ferreira, Takahiko Furuya, Manuel J Fonseca, Henry Johan, Takahiro Matsuda, et al., "A comparison of methods for sketch-based 3d shape retrieval," *Computer Vision and Image Understanding*, vol. 119, pp. 57–80, 2014.

[17] Xinwei He, Yang Zhou, Zhichao Zhou, Song Bai, and Xiang Bai, "Triplet-center loss for multi-view 3d object retrieval," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1945–1954.

[18] Cristian Buciluǎ, Rich Caruana, and Alexandru Niculescu-Mizil, "Model compression," in *ACM SIGKDD international conference on Knowledge discovery and data mining*, 2006, pp. 535–541.

[19] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller, "Multi-view convolutional neural networks for 3d shape recognition," in *IEEE international conference on computer vision*, 2015, pp. 945–953.

[20] Feng Wang, Jian Cheng, Weiyang Liu, and Haijun Liu, "Additive margin softmax for face verification," *IEEE Signal Processing Letters*, vol. 25, pp. 926–930, 2018.

[21] Philip Shilane, Patrick Min, Michael Kazhdan, and Thomas Funkhouser, "The princeton shape benchmark," in *Proceedings Shape Modeling Applications*, 2004, pp. 167–178.

[22] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[23] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[24] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.