

“© 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

MULTI-VIEW PAIRWISE RELATIONSHIP LEARNING FOR SKETCH BASED 3D SHAPE RETRIEVAL

Hanhui Li, Hefeng Wu, Xiangjian He, Shujin Lin, Ruomei Wang, Xiaonan Luo

ABSTRACT

Recent progress in sketch-based 3D shape retrieval creates a novel and user-friendly way to explore massive 3D shapes on the Internet. However, current methods on this topic rely on designing invariant features for both sketches and 3D shapes, or complex matching strategies. Therefore they suffer from problems like arbitrary drawings and inconsistent viewpoints. To tackle this problem, we propose a probabilistic framework based on Multi-View Pairwise Relationship (MVPR) learning. Our framework includes multiple views of 3D shapes as the intermediate layer between sketches and 3D shapes, and transforms the original retrieval problem into the form of inferring pairwise relationship between sketches and views. We accomplish pairwise relationship inference by a novel MVPR net, which can automatically predict and merge the pairwise relationships between a sketch and multiple views, thus freeing us from exhaustively selecting the best view of 3D shapes. We also propose to learn robust features for sketches and views via fine-tuning pre-trained networks. Extensive experiments on a large dataset demonstrate that the proposed method can outperform state-of-the-art methods significantly.

Index Terms— Sketch, 3D Shape Retrieval, Semantic Similarity.

1. INTRODUCTION

With the rapid development in multimedia technologies and computer graphics, numerous 3D shapes have been created in recent decades. Such situation brings about the need of efficiently retrieving 3D shapes, in many real-life applications like 3D printing and computer aided design. Extensive content-based methods [1–3] for 3D shape retrieval have been proposed. These methods focus on measuring the visual similarity between 3D shapes, and need to use 3D shapes as the queries (query-by-shape). Such methods are cumbersome for unskilled users, if they need to modify existing 3D shapes or even create a new one to describe their search intent.

To address the aforementioned issues of conventional query-by-shape methods, sketch-based 3D shape retrieval (query-by-sketch) methods [4–12] have been proposed. Methods of this type use sketches, which are very simple and abstract 2D drawings [13], as the queries. These methods do not require users to be skilled in 3D modeling or painting, and

thus improve the efficiency of query greatly.

However, recent research has pointed out that the performance of query-by-sketch methods are far lower than that of query-by-shape methods [4]. The main reason behind such performance gap is the significant visual differences between sketches and 3D shapes, e.g, the viewpoint of a sketch and is usually different from that of a shape. Moreover, the large intra-class variance among sketches also makes the retrieval task become more difficult. Therefore, even though conventional methods [6–8, 11, 12, 14] have tried to minimize such differences by proposing complicate hand-crafted descriptors or similarity measures, their results are still unsatisfactory. [9] proposes to employ deep Siamese networks to learn features for sketches and 3D shapes, and its has achieved better performance, compared with the conventional methods. But its Siamese networks only use one fixed view of 3D shape, and thus it omits lots of helpful details of 3D shapes.

Our solution is inspired by the recent success of multi-view learning based methods [15, 16] in the query-by-shape task. These methods show that we can effectively describe and distinguish 3D shapes via their multiple views. Therefore, we propose a probabilistic framework based on Multi-View Pairwise Relationship (MVPR) learning to tackle the query-by-sketch problem. Here pairwise relationship refers to whether a given sketch and a shape belong to the same class. In other words, we propose to measure the *semantic* similarity between sketches and 3D shapes, instead of the visual similarity. Measuring semantic similarity allows our method to omit the intra-class variance among sketches and focus on the inter-class variance of sketches and 3D shapes. We propose a novel MVPR net to predict the pairwise relationship score, and then transform the score to generate the probability of retrieving a candidate shape via our probabilistic framework. In our MVPR net, the pairwise relationships among a sketch and multiple views are predicted and merged automatically, so we do not have the trouble of selecting the best view of 3D shapes as in [8, 9]. Besides, we also propose to exploit the advantages of representation learning by fine-tuning pre-trained networks, if the class labels of sketches and 3D shapes are available. In this way, we can learn robust features for both sketches and 3D shapes, and consequently increase the performance of our method.

In summary, the contributions of this paper are threefold: (i) we formulate sketch-based 3D shape retrieval problem in

a probabilistic framework, and solve it via measuring semantic similarity; (ii) within our framework, we propose a novel network to infer the multi-view pairwise relationships among sketches and 3D shapes; (iii) we propose a class-specific method to fine-tune pre-trained networks and learn robust features for both sketches and 3D shapes.

2. RELATED WORK

Current methods on 3D shape retrieval can mainly be classified into query-by-shape methods [3, 16, 17] and query-by-sketch methods [5–12, 14]. The comprehensive study of 3D shape retrieval can be found in [1, 2, 4]. Due to the page limitation, we only focus on query-by-sketch methods here.

Reforming local features is the typical way in conventional query-by-sketch methods: [8] proposes to combine local features generated by Gabor filters with the bag-of-features model, and trains the support vector machines to predict the best view of 3D shapes; [4] introduces the overlapped version of Histograms of Orientation Gradients; [7] proposes the junction-based extended shape context feature.

Introducing new measure to model similarity between sketches and 3D shapes is the mainstream trend on this topic as well. [6] considers shape retrieval as a graph diffusion problem and introduces a cross-modal manifold ranking method to handle it; [5] proposes to co-retrieve a set of 3D shapes of indoor objects; [12] uses collaborative filtering to incorporate user preference in their retrieval system; [10] proposes to embed sketch features and shape features into the same feature space via a neural network.

The idea of exploring semantic similarity is used in [14] and [9] as well. But our method differs from them in several aspects: [14] requires the class labels for sketches and 3D shapes to train its models, while our method only needs to know the pairwise relationships; [9] uses multiple Siamese networks to model the semantic similarity (between sketch and sketch, view and view, sketch and view), while we use a single network (sketch and view) to accomplish our task efficiently. Besides, [9] only works on a single fixed view of 3D shape, while our method can handle multiple views easily.

3. MULTI-VIEW PAIRWISE RELATIONSHIP FOR SKETCH BASED 3D SHAPE RETRIEVAL

In this section, we introduce the Multi-View Pairwise Relationship (MVPR) learning method for the sketch-based 3D shape retrieval problem in detail. We first formulate our retrieval problem in a Bayesian framework in Section 3.1. The core of our framework is a discriminative model that infers the pairwise relationships among sketches and 3D shapes, and our method for optimizing such model is presented in Section 3.2. At last, we introduce a method for learning robust representation for sketch and 3D shape in Section 3.3.

3.1. A probabilistic formulation

In order to tackle the 3D shape retrieval problem, we propose to model the semantic similarity between sketch and 3D shape in the probabilistic way. The advantage of formulating our task in a probabilistic framework is that we can easily include multiple views of 3D shape as an intermediate layer between 3D shape and sketch by the law of total probability. To be specific, let S and M denote the domain (or set) of sketches and 3D shapes, respectively; given a query sketch $s \in S$, assume we project each 3D shape into N views and use a tuple $\mathbf{v} = (v_1, v_2, \dots, v_N)$ to denote the views of any 3D shape, and then the probability of retrieving shape m given sketch s can be defined as follows:

$$p(m|s) = \sum_{\mathbf{v}} p(m|\mathbf{v})p(\mathbf{v}|s) = \sum_{\mathbf{v}} \frac{p(\mathbf{v}|m)p(m)}{p(\mathbf{v})}p(\mathbf{v}|s), \quad (1)$$

where $p(m|\mathbf{v})$ is the posterior of observing views \mathbf{v} given shape m , $p(m)$ is the prior of retrieving shape m , $p(\mathbf{v})$ is the prior of observing views \mathbf{v} , and $p(\mathbf{v}|s)$ is the posterior of observing views \mathbf{v} given sketch s .

The aforementioned formulation provides us with a general way to tackle the query-by-sketch problem: $p(\mathbf{v}|s)$ can be used to measure the visual or semantic similarity between sketch and 3D shape; $p(m)$ can include user preference to encourage the retrieval system to return the frequently queried shapes; $p(m|\mathbf{v})$ and $p(\mathbf{v})$ can serve as the normalization term to alleviate the effect of abnormal views. Such formulation can generalize most conventional methods, and setting $N = 1$ will let our method work on a single view, as in [9].

In our implementation of Eq. (1), we assume that the priors of each shape being retrieved are equal, i.e., $p(m) = \frac{1}{|M|}$, where $|M|$ denotes the cardinality of the set of 3D shapes. As to $p(\mathbf{v}|m)$, we assume the N views of each 3D shape are pre-defined and fixed, so that given a shape m and a set of views \mathbf{v} , $p(\mathbf{v}|m) = 1$ if all N views are generated from m , otherwise $p(\mathbf{v}|m) = 0$. Such setting can reduce the computational cost of our method significantly. We denote the N views of shape m as $\mathbf{v}(m)$ for convenience. The normalization term $p(\mathbf{v})$ can be marginalized over a training sketch set S^t as $p(\mathbf{v}) = \sum_i p(\mathbf{v}|s_i^t)p(s_i^t)$, where $i \in 1, \dots, |S^t|$ and we let $p(s_i^t) = \frac{1}{|S^t|}$. Here we use the superscripts t and o to denote sketches from the training set and the test set, respectively. In this way, $p(m|s^o)$ can be estimated as

$$p(m|s^o) \propto \frac{p(\mathbf{v}|s^o)}{\sum_i p(\mathbf{v}(m)|s_i^t)}. \quad (2)$$

From Eq. (2) we can see that our retrieval framework mainly relies on estimating $p(\mathbf{v}|s)$, the probability of observing views \mathbf{v} given sketch s , no matter s is from the training set or the test set. Therefore, in the subsequent sections, we will focus on inferring $p(\mathbf{v}|s)$.

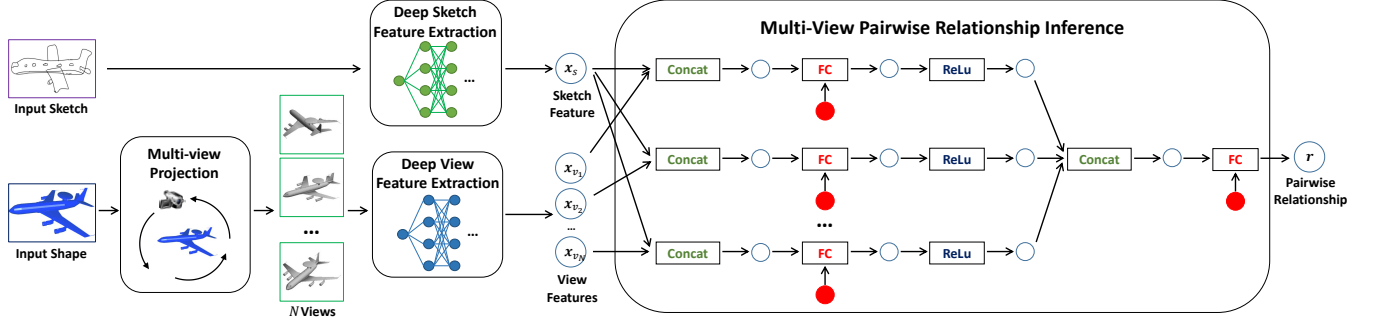


Fig. 1. Demonstration of the proposed framework for sketch-based 3D shape retrieval. Our framework is based on a Multi-View Pairwise Relationship (MVPR) net. The MVPR net is designed to predict the pairwise relationship score between sketch and 3D shape, and its directed acyclic graph representation is demonstrated on the right side of this figure: rectangles denote the three types of computational blocks, including concatenation (Concat), fully connected layer (FC) and rectified linear unit (ReLU). White circles denote inner variables while red circles denote parameters of the MVPR net.

3.2. Multi-View Pairwise Relationship Inference

Directly inferring $p(\mathbf{v}|s)$ is intractable because of the large searching space of \mathbf{v} . Hence, we propose to approximate $p(\mathbf{v}|s)$ via a discriminative model that explores pairwise relationship between sketch and 3D shape. The intuition behind using pairwise relationship to approximate $p(\mathbf{v}|s)$ is straightforward: if a sketch s and a 3D shape m are from the same class, the chance of observing $\mathbf{v}(m)$ given s should be higher. In this way, $p(\mathbf{v}|s)$ is considered as the measure of semantic similarity between sketch and 3D shape.

Assume we have a training set T consisting of triplets $(s_i^t, \mathbf{v}(m_i^t), r_i^t)$, $i \in 1, \dots, |T|$, where r_i^t is the auxiliary variable that $r_i^t = 1$ if the sketch s_i^t and shape m_i^t belong to the same class, otherwise $r_i^t = -1$. We consider the discriminative model as a function $f(\mathbf{v}, s)$ mapping (\mathbf{v}, s) to $(-\infty, +\infty)$. Based on the aforementioned discussions, if f satisfies the constraint that $f(\mathbf{v}(m_i^t), s_i^t) > 0$ when $r_i^t = 1$, while $f(\mathbf{v}(m_i^t), s_i^t) \leq 0$ when $r_i^t = -1$, and then we can approximate $p(\mathbf{v}|s)$ via performing the logistic sigmoid function on f [18]:

$$p(\mathbf{v}|s) \propto \frac{1}{1 + \exp(-f(\mathbf{v}, s))}. \quad (3)$$

With Eq. (3), we can easily see that maximum likelihood estimation with respect to $p(\mathbf{v}|s)$ can be approximated by minimizing the following logistic log loss function:

$$L_p = \sum_i \log(1 + \exp(-r_i^t \cdot f(\mathbf{v}(m_i^t), s_i^t))). \quad (4)$$

To implement $f(\mathbf{v}(m), s)$, we propose a novel Multi-View Pairwise Relationship (MVPR) net, as demonstrated in Figure 1. The MVPR net contains 2 major layers: in the first layer, the sketch feature is concatenated to each view feature to generate the N input features. Our method for learning the feature for sketch and views will be introduced later in

Section 3.3. Each generated feature is filtered by a fully connected layer to be compressed into a 64-D vector. This vector encodes the similarity between the input sketch and one corresponding view of the 3D shape. At the second layer, the N similarity vectors are concatenated together, and filtered by a $(N \times 64)$ -D dimensional fully connected layer to predict the pairwise relationship score. To train the MVPR net, we perform mini-batch based gradient descent to minimize the loss function L_p .

Discussion: the advantages of our MVPR net are dual. On the one hand, our method is trained to automatically combine the multi-view information of 3D shape. Thus it does not need to select the best view of 3D shape, and it can exploit richer information of 3D shape, compared with [9]. On the other hand, the combination of features of sketch and views are on high-level (category information), while in [10], such combination is conducted with low-level image features. Thus our method can better measure the semantic similarity between sketch and 3D shape.

In our experiment, we find that directly using $f(\mathbf{v}, s)$ as the ranking scores will increase the possibility of retrieving outliers. Therefore, we propose to use normalized version of $f(\mathbf{v}, s)$, i.e., $\frac{f(\mathbf{v}, s) - \mu_{\mathbf{v}(m)}}{\sigma_{\mathbf{v}(m)}}$, where μ_m and σ_m denote the mean and the standard deviation of the predicted relationship scores among all training sketches and the 3D shape m , respectively.

3.3. Representation Learning

In our framework, we use the network structure of the first to the “relu7” layers of the VGG-19 network [19] pre-trained on ImageNet [20] to extract features of sketches and views. However, We also propose to fine-tune the pre-trained network, since the training images on the ImageNet [19] are quite different from sketches and views. Our fine-tuning method requires that class labels for sketches and 3D shapes are accessible.

Sketch feature Learning: we attach a fully connected layer to the end of the “relu7” layer of the VGG-19 network to construct a multi-class sketch classifier, and optimize the classifier to fine-tune it. Assume the sketches in the training set can be classified into C classes, and the multi-class sketch classifier is a function that takes a sketch image of 224×224 pixels as its input and outputs a C -D dimensional vector. Each dimension of the output vector represents the pseudo probability of the input sketch belonging to the corresponding class. Given a training example (s_i^t, c_i^t) , where c_i^t denotes the ground-truth class of the i -th training sketch, the multinomial logistic loss of our classifier on it, $L_r(s_i^t, c_i^t)$, can be defined as follows:

$$L_r(s_i^t, c_i^t) = -\log\left(\frac{\exp(h(s_i^t, c_i^t))}{\sum_{k=1}^C \exp(h(s_i^t, k))}\right) \quad (5)$$

where $h(s_i^t, k)$ denotes the pseudo probability of s_i^t belonging to the k -th class. We adopt the mini-batch gradient descent method to minimize L_r . Once the sketch classifier is optimized, we use the 4,096-D outputs of its “relu7” layer as the sketch features.

View feature Learning: As to the view feature, we first set N cameras rotating around the 3D shape at the interval of $360/N$ degrees, and then render each view via the Phong reflection model [21], which is an empirical model estimating the intensity of pixels in projection views of 3D shapes. The image size of each view is 224×224 pixels as well, and we generates $(N \times 4,096)$ -D vector for each 3D shape. To reduce the complexity of the view net, we let the N views share the same network, instead of training a net for each view. The view feature is optimized in the same way as the sketch feature.

Note that our networks for representation learning and that for pairwise relationship learning are trained individually. We do not combine them into a whole network and train it via end-to-end learning, because such network will contain about 280 million parameters and be hard to optimize.

4. EXPERIMENTS

In this section, we validate the effectiveness of the MVPR framework via extensive experiments. Our evaluation mainly focuses on three aspects: (a) the overall performance of the proposed framework; (b) the significance of adopting the multi-view strategy; and (c) the contribution of representation learning to the performance of the whole framework.

Setup: For fair comparison, we evaluate our framework on the LSB benchmark [4] with its evaluation protocol. LSB contains 171 object classes, with a total of 13,680 sketches (50 training instances and 30 test instances for each class) and 8,987 shapes. Thus LSB is large enough to prevent overfitting. We report our experimental results with all evaluation metrics defined on LSB, including Precision and Recall,

Nearest Neighbor (NN), First Tier (FT) and Second Tier (ST), E-Measure (E), Discounted Cumulative Gain (DCG) and Average Precision (AP). The reciprocally weighted version of these metrics are used in our experiments as well. Interested Readers can refer to [4] for more details. Eight state-of-the-art methods are included for comparison, including BF-fGALIF [8], CDMR [6], SBR-VC [14], SCMR-OPHOG [4], BOF-JESC [7], HOAD-CR [12], Siamese CNN [9] and PCDNN [10].

Implementation Details: We use the MatConvNet toolbox [22] to implement our neural networks. To train our neural networks, we employ the mini-batch gradient descent method and the size of a mini-batch is set to 8. Representation learning for sketch and shape is run with all training data for 30 epochs. The learning rate of each epoch is 10^{-4} . The MVPR net is trained for 200 epochs. In each epoch, we randomly sample 10^4 triplets for training (5K positive and 5K negative). The initial learning rate of the MVPR is set to 10^{-2} , and after each epoch, we reduce the learning rate by 10^{-4} to encourage the learning process to converge. The number of views is set to 12 throughout this paper.

The overall performance metrics of the proposed framework are demonstrated in Table 1. From these results, we can see that our MVPR framework successfully obtains significant performance gains with all metrics. Compared with the Siamese CNN, which is similar to our method if we use a single view in our method, our method outperforms the Siamese CNN by 30.7% in NN, 29.4% in FT, 32.6% in ST, 16.1% in E, 22.5% in DCG and 31.5% in AP. Except for the Siamese CNN, another method utilizing neural networks is PCDNN, and our method outperforms it significantly with most reciprocally weighted performance metrics as well. The precision-recall curves in Figure 2 also demonstrate the advantages of the proposed framework. The precision of our method is about 0.629 when the recall is 0.1, which is significantly higher than other methods. In summary, all these experimental results validate that our method can solve the sketch-based 3D shape retrieval problem effectively.

Moreover, since our method stems from the assumption that multi-view mechanism can provide more information to be exploited, therefore, we also conduct an experiment to evaluate the performance of our method with different numbers of views. We use the Siamese CNN as the baseline because it only uses one view. Our experimental results are reported in Table 2. From this table we can see the performance of our method is proportional to the number of views. The AP score of our method with only one view is 0.365, while using three views can increase it to 0.481, which is about the twice the AP score of the Siamese CNN (0.228). Therefore, our opinion that using multiple views can increase the performance of retrieval system is validated.

Besides, we also conduct an experiment to find out the contribution of the representation learning part in our framework. We report the results of our framework without rep-

Table 1. Performance evaluation of the proposed method against state-of-the-art methods on the LSB benchmark [4]. The scores before and after “/” are standard performance metrics and reciprocally weighted performance metrics, respectively. Scores of the latter metrics are measured in 10^{-5} . The proposed MVPR method outperforms the baselines by a large margin.

| Method | NN | FT | ST | E | DCG | AP |
|---|----------------------|----------------------|----------------------|----------------------|-----------------------|----------------------|
| BF-fGALIF [8] | 0.115 / 0.802 | 0.051 / 0.520 | 0.078 / 0.735 | 0.036 / 0.289 | 0.321 / 3.408 | 0.044 / 0.596 |
| CDMR ($\sigma_{SM} = 0.05, \alpha = 0.3$) [6] | 0.109 / 0.789 | 0.057 / 0.526 | 0.089 / 0.773 | 0.041 / 0.330 | 0.328 / 3.430 | 0.054 / 0.626 |
| SBR-VC ($\alpha = 1$) [14] | 0.095 / 0.449 | 0.050 / 0.264 | 0.081 / 0.425 | 0.037 / 0.264 | 0.319 / 3.051 | 0.050 / 0.291 |
| SCMR-OPHOG [4] | 0.160 / 0.993 | 0.115 / 0.743 | 0.170 / 1.035 | 0.079 / 0.541 | 0.376 / 3.676 | 0.131 / 0.886 |
| BOF-JESC (Words800-VQ) [7] | 0.086 / 0.462 | 0.043 / 0.271 | 0.068 / 0.467 | 0.030 / 0.236 | 0.310 / 3.149 | 0.041 / 0.370 |
| HOAD-CR [12] | 0.163 / - | 0.130 / - | 0.170 / - | 0.078 / - | 0.393 / - | 0.140 / - |
| Siamese CNN [9] | 0.239 / - | 0.212 / - | 0.316 / - | 0.140 / - | 0.490 / - | 0.228 / - |
| PCDNN [10] | - / 5.175 | - / 3.285 | - / 4.406 | - / 2.056 | - / 12.39 | - / 3.960 |
| MVPR | 0.546 / 7.839 | 0.506 / 7.306 | 0.642 / 8.725 | 0.301 / 2.617 | 0.715 / 10.882 | 0.543 / 8.248 |

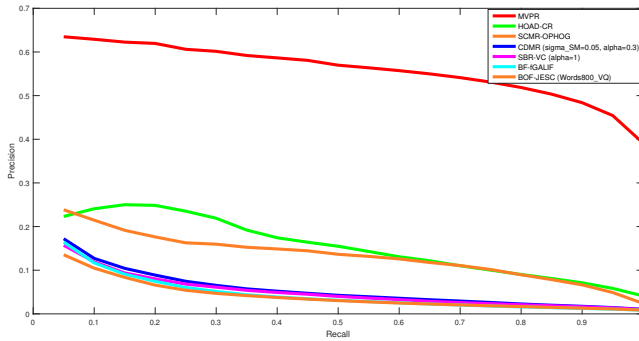


Fig. 2. Precision-recall curves of the proposed framework (MVPR) against state-of-the-art methods on LSB. This figure shows that MVPR can outperform conventional methods significantly.

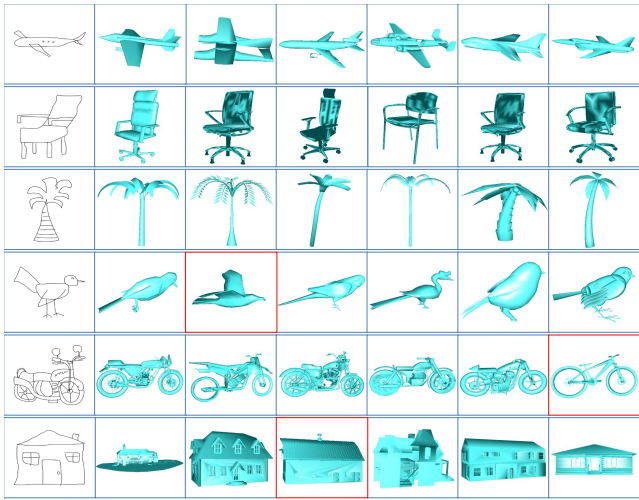


Fig. 3. Qualitative retrieval results of the proposed method. Semantically unrelated results are labeled in red.

resentation learning in Table 2. It can be observed that our method can achieve a good performance even without representation learning. These results also show that the representation learning method can boost the performance of our

framework by averagely 5% with all performance metrics. Therefore, our representation learning method is considerable if class labels are available.

At last, we present some retrieval results of our method in Figure 3. These results show that our method can retrieve satisfactory 3D shapes with respect to sketches of different classes.

5. CONCLUSION

In this paper, we formulate the sketch-based 3D shape retrieval problem within a probabilistic framework, and propose the Multi-View Pairwise Relationship (MVPR) net as a discriminative model to bridge the gap between sketches and 3D shapes. The MVPR net can exploit the rich information of multiple views of 3D shapes, and bring the most significant performance gain to our retrieval system. We also propose to construct multi-class classifiers to fine-tune pre-trained neural networks, so that they can generate robust sketch and 3D shape features to further strengthen our framework. Extensive experiments on the LSB benchmark validate that our method is a practical solution for the sketch-based 3D shape retrieval problem.

6. REFERENCES

- [1] Y. Yang, H. Lin, and Y. Zhang, “Content-based 3-d model retrieval: A survey,” *IEEE Trans. Systems, Man, and Cybernetics, Part C*, vol. 37, no. 6, pp. 1081–1098, 2007.
- [2] Johan W. H. Tangelder and Remco C. Veltkamp, “A survey of content based 3d shape retrieval methods,” *Multimedia Tools Appl.*, vol. 39, no. 3, pp. 441–471, 2008.
- [3] Mohamed Chaouch and Anne Verroust-Blondet, “3d model retrieval based on depth line descriptor,” in *Proceedings of the 2007 IEEE International Conference on Multimedia and Expo*, 2007, pp. 599–602.

Table 2. Performance metrics of the proposed method with different settings. This table includes results of two experiments: (i) the performance of MVPR with different numbers of views, and (ii) the performance of MVPR with / without representation learning. This table demonstrates that the proposed method can obtain distinct performance gains by exploiting multiple views of 3D shapes, and the representation learning method can further boost the performance of our method.

| Method | Number of Views | NN | FT | ST | E | DCG | AP |
|-------------|-----------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Siamese CNN | 1 | 0.239 | 0.212 | 0.316 | 0.140 | 0.490 | 0.228 |
| MVPR | 1 | 0.388 / 0.334 | 0.336 / 0.288 | 0.475 / 0.418 | 0.218 / 0.188 | 0.601 / 0.561 | 0.365 / 0.311 |
| MVPR | 3 | 0.496 / 0.440 | 0.446 / 0.390 | 0.586 / 0.527 | 0.273 / 0.245 | 0.678 / 0.637 | 0.481 / 0.421 |
| MVPR | 4 | 0.515 / 0.458 | 0.470 / 0.402 | 0.608 / 0.541 | 0.281 / 0.250 | 0.693 / 0.647 | 0.504 / 0.435 |
| MVPR | 6 | 0.524 / 0.465 | 0.483 / 0.416 | 0.622 / 0.554 | 0.291 / 0.253 | 0.703 / 0.657 | 0.518 / 0.449 |
| MVPR | 12 | 0.546 / 0.486 | 0.506 / 0.448 | 0.642 / 0.582 | 0.301 / 0.269 | 0.715 / 0.675 | 0.543 / 0.481 |

- [4] et al. Bo Li, “A comparison of 3d shape retrieval methods based on a large-scale benchmark supporting multimodal queries,” *Computer Vision and Image Understanding*, vol. 131, pp. 1–27, 2015.
- [5] Kun Xu, Kang Chen, Hongbo Fu, Wei-Lun Sun, and Shi-Min Hu, “Sketch2scene: sketch-based co-retrieval and co-placement of 3d models,” *ACM Transactions on Graphics*, vol. 32, no. 4, pp. 123:1–123:12, 2013.
- [6] Ryutarou Ohbuchi Takahiko Furuya, “Ranking on cross-domain manifold for sketch-based 3d model retrieval,” in *International Conference on Cyberworlds*, 2013, pp. 274–281.
- [7] Yafei Wen Lei Zhang Changqing Zou, Changhu Wang and Jianzhuang Liu, “Viewpoint-aware representation for sketch-based 3d model retrieval,” *IEEE Signal Processing Letters*, vol. 21, no. 8, pp. 966–970, 2014.
- [8] Tamy Boubekeur Kristian Hildebrand Mathias Eitz, Ronald Richter and Marc Alexa, “Sketch-based shape retrieval,” *ToG*, vol. 31, no. 4, pp. 31:1–31:10, 2012.
- [9] Le Kang Fang Wang and Yi Li, “Sketch-based 3d shape retrieval using convolutional neural networks,” in *CVPR*, 2015, pp. 1875–1883.
- [10] Jin Xie Fan Zhu and Yi Fang, “Learning cross-domain neural networks for sketch-based 3d shape retrieval,” in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 2016, pp. 3683–3689.
- [11] Xavier Muñoz Anna Bosch, Andrew Zisserman, “Representing shape with a spatial pyramid kernel,” in *Proceedings of the 6th ACM International Conference on Image and Video Retrieval*, 2007, pp. 401–408.
- [12] Fei Wang, Shujin Lin, Hefeng Wu, Ruomei Wang, and Xiaonan Luo, “Data-driven method for sketch-based 3d shape retrieval based on user similar draw-style recommendation,” in *SIGGRAPH ASIA 2016 Posters*, 2016, p. 34.
- [13] James Hays Mathias Eitz and Marc Alexa, “How do humans sketch objects?,” *ToG*, vol. 31, no. 4, pp. 44:1–44:10, 2012.
- [14] Bo Li, Yijuan Lu, and Ribel Fares, “Semantic sketch-based 3d model retrieval,” in *2013 IEEE International Conference on Multimedia and Expo Workshops*, 2013, pp. 1–4.
- [15] Evangelos Kalogerakis Hang Su, Subhransu Maji and Erik G. Learned-Miller, “Multi-view convolutional neural networks for 3d shape recognition,” in *ICCV*, 2015, pp. 945–953.
- [16] Haiyun Guo, Jinqiao Wang, Yue Gao, Jianqiang Li, and Hanqing Lu, “Multi-view 3d object retrieval with deep embedding network,” *IEEE Trans. Image Processing*, vol. 25, no. 12, pp. 5526–5537, 2016.
- [17] Pengjie Li, Huadong Ma, and Anlong Ming, “View-based 3d model retrieval with topological structure,” in *Proceedings of the 2011 IEEE International Conference on Multimedia and Expo*, 2011, pp. 1–6.
- [18] Christopher M. Bishop, *Pattern Recognition and Machine Learning*, 2006.
- [19] Karen Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition,” *CoRR*, vol. abs/1409.1556, 2014.
- [20] Richard Socher Jia Deng, Wei Dong, Li-Jia Li, Kai Li, and Fei-Fei Li, “Imagenet: A large-scale hierarchical image database,” in *CVPR*, 2009, pp. 248–255.
- [21] Bui Tuong Phong, “Illumination for computer generated pictures,” *Commun. ACM*, vol. 18, no. 6, pp. 311–317, 1975.
- [22] Andrea Vedaldi and Karel Lenc, “Matconvnet: Convolutional neural networks for MATLAB,” in *Proceedings of the 23rd Annual ACM Conference on Multimedia Conference*, 2015, pp. 689–692.