This CVPR2015 extended abstract is the Open Access version, provided by the Computer Vision Foundation.

Sketch-based 3D Shape Retrieval using Convolutional Neural Networks

Fang Wang[1], Le Kang[2], and Yi Li[1]*
[1]NICTA and ANU; [2]University of Maryland, College Park; *Corresponding author: yi.li@nicta.com.au;
Project website: http://users.cecs.anu.edu.au/~yili/cnnsbsr/index.html

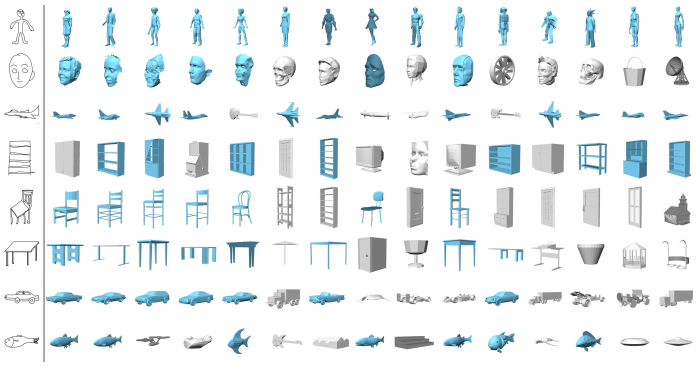

Figure 1: Retrieval examples of the PSB/SBSR dataset [2]. The cyan denotes the correct retrievals.

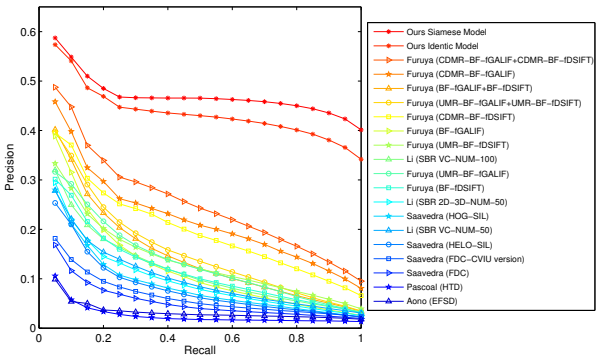

Figure 2: Performance comparison on SHREC'13. Please refer to [3] for the descriptions of the compared methods.

Retrieving 3D models from 2D human sketches (Fig. 1) has received considerable attention in the areas of graphics, image retrieval, and computer vision. However, directly matching 2D sketches to 3D models suffers from significant differences between the 2D and 3D representations.

Almost always in state of the art approaches a large amount of "best views" are computed for 3D models, with the hope that the query sketch matches one of these 2D projections of 3D models using predefined features. We argue that this two stage approach (view selection – matching) is pragmatic but also problematic because the "best views" are subjective and ambiguous, which makes the matching inputs obscure. This imprecise nature of matching further makes it challenging to choose features manually.

Instead of relying on the elusive concept of "best views" and the handcrafted features, we propose to define our views using a minimalism approach and learn features for both sketches and views. Specifically, we drastically reduce the number of views to only two predefined directions for the whole dataset. Once the viewpoints are chosen, we use closed boundaries and Suggestive Contours to render line drawings.

This seeming radical approach triumphs only when the features are learned properly. In principle, this can be regarded as learning representations between sketches and views by specifying similarities, which gives us a semantic level matching. To achieve this goal, we need comprehensive shape representations rather than the combination of a bunch of shallow features that only capture the low level visual information.

Our model is based on the Siamese network [1]. A Siamese network is a particular neural network architecture consisting of two identical subconvolutional networks. The goal of the network is to make the output vectors similar if input pairs are labeled as similar, and dissimilar for the input pairs that are labeled as dissimilar.

Since the sketches and the rendered line drawings have distinctive intrinsic properties, we use two Siamese networks, one for the sketches and one for the rendered line drawings (Fig. 4). The loss function is defined on the within-domain as well as the cross domain similarities.

$$\mathcal{L}(s_1,s_2,v_1,v_2,y) = L(s_1,s_2,y) + L(v_1,v_2,y) + L(s_1,v_1,y), \qquad (1)$$

where $L(\cdot,\cdot,\cdot)$ is defined by $L(s_1,s_2,y) = (1-y)\alpha D_w^2 + y\beta \exp(\gamma D_w)$. This two models strategy gives us more power to capture different properties in different domains (Fig. 3).

Our experiments on three large datasets [2, 3] demonstrated that our method is significantly better than state of the art approaches (Fig. 2), and outperforms them in all conventional metrics.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

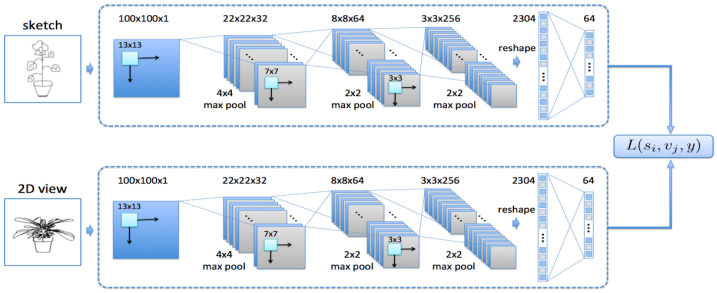

Figure 3: Dimension reduction using Siamese network.

[1] S Chopra, R Hadsell, and Y LeCun. Learning a similarity metric discriminatively, with application to face verification. In *CVPR 2005*, volume 1, pages 539–546.

[2] M Eitz, R Richter, T Boubekeur, K Hildebrand, and M Alexa. Sketch-based shape retrieval. *ACM Trans. Graphics*, 31(4):31:1–31:10, 2012.

[3] B Li, Y Lu, A Godil, T Schreck, B Bustos, A Ferreira, T Furuya, MJ Fonseca, H Johan, T Matsuda, et al. A comparison of methods for sketch-based 3d shape retrieval. *CVIU*, 119:57–80, 2014.

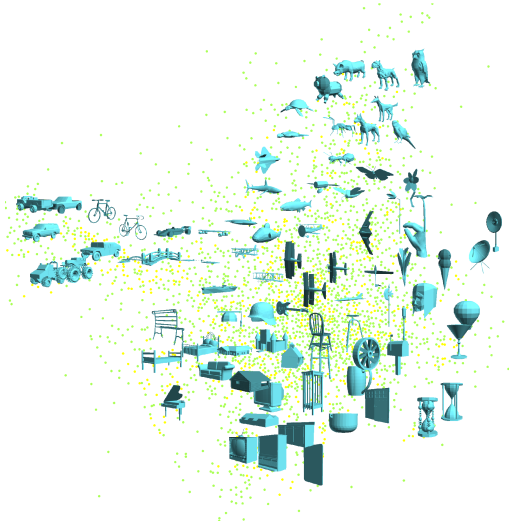

Figure 4: Visualization of feature space on SHREC'13. Sketch and view feature points are shown by green & yellow, respectively.