Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Perspective:

Three-dimensional shape space learning for visual concept construction: challenges and research progress

Xin TONG

Microsoft Research Asia, Beijing 100080, China

E-mail: xtong@microsoft.com

Received July 26, 2022; Revision accepted Aug. 26, 2022; Crosschecked Sept. 6, 2022

https://doi.org/10.1631/FITEE.2200318

Human beings can easily categorize three-dimensional (3D) objects with similar shapes and functions into a set of "visual concepts" and learn "visual knowledge" of the surrounding 3D real world (Pan, 2019). Developing efficient methods to learn the computational representation of the visual concept and the visual knowledge is a critical task in artificial intelligence (Pan, 2021a). A crucial step to this end is to learn the shape space spanned by all 3D objects that belong to one visual concept. In this paper, we present the key technical challenges and recent research progress in 3D shape space learning and discuss the open problems and research opportunities in this area.

1 Introduction

Our real world consists of objects with diverse 3D shapes, structures, and dynamics. Cognitive psychology studies have illustrated that human beings can efficiently recognize 3D shapes and their relationships and represent them as "mental imagery" or "visual knowledge" in their memory (Pan, 2019). With the help of visual knowledge, human beings can directly manipulate these 3D objects and scenes in their mind and easily recognize and reconstruct objects from 2D drawings such as sketches or stylized paintings (Pan, 2019).

ORCID: Xin TONG, https://orcid.org/0000-0001-8788-2453

In past decades, computer graphics and vision researchers have developed numerous mathematical and physical theories and computational methods for modeling 3D shapes and simulating their dynamics in computers (Hughes et al., 2013). Digital representations of 3D objects and the associated 3D modeling and simulation systems result in the prevalence of graphics applications such as computer-aided design (CAD) and computer-aided manufacturing (CAM), movie making, virtual reality, and 3D games. Despite the advances in the past, modeling and reconstructing 3D shapes is still a difficult task, even for skilled artists and engineers.

A plausible reason for the gap between the capability of the human mind and computer graphics techniques is that current 3D modeling methods do not have any prior knowledge or "visual knowledge" of 3D objects, no matter how many 3D shapes have been created. Pan (2021a) pointed out that the core of visual knowledge is a set of "visual concepts," each of which models a set of 3D objects that share similar shapes, appearance, or functions. Typically, a visual concept is composed of a representative prototype (shape and appearance) and a domain for describing object variations in the visual concept (Pan, 2019). Different from 3D representations in traditional graphics that focus on individual 3D objects, a visual concept models the space spanned by all 3D objects in a category.

Therefore, learning the space spanned by 3D objects in one category (i.e., the 3D objects belong to one "visual concept") will be a critical task in computer graphics and computer vision. The learned shape space will bring more "visual knowledge" or "prior" to different downstream graphics and vision tasks, such as 3D shape completion and reconstruction (Jin et al., 2020; Wang PS et al., 2022), 3D shape analysis (Wu et al., 2015), and 3D shape generation and editing (Zheng XY et al., 2022; Yang J et al., 2023). Moreover, 3D shape space learning is a cornerstone for constructing visual knowledge and next-generation artificial intelligence (Pan, 2021a).

In this paper, we focus on techniques developed for learning the shape space of 3D objects and discuss the challenges, research progress, and open problems in this research field. We define the problem of 3D shape space learning, discuss the technical challenges in this task, summarize the research status of 3D shape space learning by reviewing our explorations and other recent research works for tackling each technical challenge, and present the open problems and research opportunities in 3D shape space learning.

2 Challenges of 3D shape space learning

Three-dimensional objects in a category (e.g., human faces, chairs, and airplanes) often share a similar structure or 3D shape and thus form a low-dimensional subspace in the high-dimensional space spanned by all 3D shapes. The goal of 3D shape space learning is to construct a compact and accurate representation of the subspace spanned by all 3D shapes in a category, which can faithfully represent all 3D shapes in the subspace. In addition, the learned 3D shape space should be consistent with human perception so that the nearby 3D shapes in the learned shape space are visually similar to each other. In this way, the learned shape space can be easily controlled by users for 3D shape generation and editing.

Early techniques have been developed for constructing the parametric model of the shape space from manually aligned 3D shapes in specific shape classes, such as human faces (Cao et al., 2014; Egger et al., 2020) and bodies (Loper et al., 2015), as well as animals (Zuffi et al., 2017). The shapes in these classes share the same topology and structure and thus can be easily aligned via human labeling. For buildings or trees, procedural modeling techniques (Měch and Prusinkiewicz, 1996; Müller et al., 2006) have been designed for generating 3D shapes in each specific shape category. All these methods are dedicated to specific object classes and cannot be generalized to other object classes. With the advances in deep learning techniques, deep learning based methods have been developed for learning 3D shape spaces from 3D object collections (Xiao et al., 2020), providing a generic and powerful solution for learning 3D shape spaces of different shape classes. However, learning 3D shape space is still a non-trivial task due to three challenges:

1. No correspondence. For most shape classes, the 3D shapes in the class exhibit large geometric variations and have no correspondence with each other. Due to the large shape variations, manually aligning these shapes is a difficult task. It is unclear how to efficiently learn a compact and meaningful shape space representation from an unaligned 3D shape collection.

2. Structure variation. Three-dimensional objects, especially man-made ones, are always composed of parts with similar layouts. Although human beings can easily recognize the structure (i.e., part layout) of 3D shapes from a small number of exemplars, learning the shape structures is still a challenging task in computer graphics and vision. For classes (e.g., chairs) that include shapes with various structures, learning the shape space that consists of both continuous geometric variations and discrete structure variation is even more challenging.

3. Dimension gap. Because all imaging sensors can capture only one-dimensional (1D) or two-dimensional (2D) signals of 3D scenes, capturing a complete 3D shape is still a labor-intensive task. Moreover, modeling and labeling 3D data via a 2D graphical user interface (GUI) is more difficult than 2D image labeling and authoring. As a result, there are still far fewer 3D models available for deep learning than their 2D observations (i.e., RGB images or depth images). Therefore, we need to develop new methods that learn 3D shape space either from a relatively small 3D dataset or from 2D observations of 3D shapes.

Fig. 1 takes the chair class as an example to illustrate the three technical challenges.

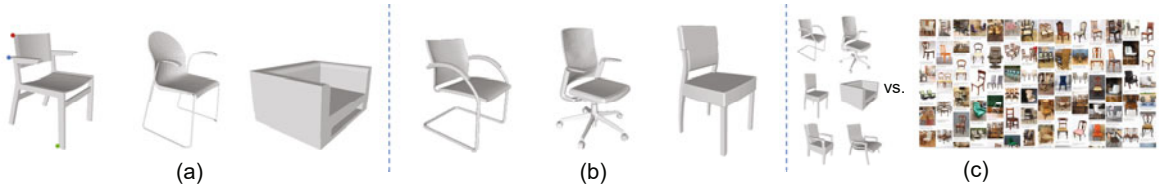

Fig. 1 We take the chair class as an example to illustrate the technical challenges of 3D shape space learning: (a) for 3D chairs with large geometric variations, it is difficult to specify the meaningful point correspondences (e.g., the dots in different colors) between different 3D shapes; (b) the discrete structure variations of 3D chairs make the shape space learning more challenging; (c) there is a dimension gap between the 3D chair models and their 2D observations, with a small number of 3D models (thousands of 3D chairs in the ShapeNet dataset) and a large number of 2D images (tens of thousands of 2D chair images that can be easily found on the Internet). References to color refer to the online version of this figure

3 Research progress of 3D shape space learning

The challenges described above make the 3D shape space learning task different from other representation learning tasks in computer vision. In this section, we review recent research works on 3D shape space learning. Our goal is not to provide a comprehensive survey of 3D shape space learning methods. Instead, we try to classify existing 3D shape space learning methods into different categories according to how they solve each technical challenge and discuss the key idea of the representative methods in each category.

3.1 Learning shape correspondence

To learn 3D shape space from unaligned 3D shapes, existing methods can be grouped into two categories according to how they address the shape correspondence: implicit correspondence based methods and explicit correspondence based methods.

Implicit correspondence based methods encode the shape space via a deep neural network (DNN) based on the implicit shape correspondence defined by their underlying 3D shape representations. Generally, these methods first transform and scale the 3D shapes into a unified bounding box, convert them into a specific 3D representation, and feed the converted 3D shapes into the DNN for learning the latent space of these 3D shapes. In these methods, the underlying 3D representation defines an implicit correspondence between different 3D shapes. For volume-based methods (Wu et al., 2015; Riegler et al., 2017; Wang PS et al., 2017, 2022), which model 3D shapes as the signed distance functions (Park et al., 2019) or occupancy fields (Mescheder et al., 2019), the points of the 3D shapes that share the same spatial coordinate correspond to each other. For multi-view-based methods (Su et al., 2015; Bai et al., 2016; Lun et al., 2017), the points of the 3D shapes that project to the same pixel position correspond to each other. For manifold-based methods (Masci et al., 2015; Sinha et al., 2016) and point-based methods (Qi et al., 2017; Yu LQ et al., 2018), the order of points or surface vertices feeding into the neural network determines the correspondence between different 3D shapes, where the points or vertices that have the same index correspond to each other.

With carefully designed neural networks, these methods can faithfully encode all 3D shapes and offer a good shape space prior to shape reconstruction and completion, as well as shape understanding. However, most of these methods cannot guarantee that each latent code be decoded to a valid 3D shape in the class. Although a set of implicit correspondence based generative methods (Chen and Zhang, 2019; Zheng XY et al., 2022) have been developed for 3D shape generation, the distance in the latent space does not correspond to the perceptual distance of two 3D shapes. As a result, a small change of the latent codes may lead to large 3D shape variations, which hinders its usage in many shape modeling applications. Because these methods do not learn the correct correspondence between 3D shapes, they are difficult to apply for learning the disentangled shapes and appearance space of 3D objects.

Explicit correspondence based methods represent the 3D shapes as deformations of templates

shared by all 3D shapes and learn the shape space and shape correspondence in one task. These methods can be further categorized into surface-based methods and volume-based methods. Surface-based methods (Groueix et al., 2018; Wang NY et al., 2018; Wen et al., 2019) either learn the mapping between 3D shape surfaces and a common 2D atlas or directly learn the correspondence between a pair of shape surfaces (Jiang et al., 2020). However, these methods cannot construct reasonable correspondence between 3D shapes with structure or topology variations, resulting in poor 3D reconstruction quality. This is because the points of a part on one shape surface cannot find their correspondences on the other shape surface without this part.

Volume-based methods (Liu and Liu, 2020; Deng et al., 2021; Zheng ZR et al., 2021) represent the implicit field defined by 3D shapes with the transformed 3D template field and learn the transformations via a DNN. Although in theory the volumetric correspondence between implicit fields of two shapes with different structures can be well defined, designing an algorithm to learn the correspondence from unlabeled 3D shapes is still a non-trivial task. Liu and Liu (2020) proposed a method that exploits the part label of 3D shapes for learning the dense correspondence of 3D shapes. Zheng ZR et al. (2021) introduced a method that uses the deformed implicit field for modeling the dense correspondence of 3D shapes with the same structure.

In Deng et al. (2021), we developed a method for learning the shape space and the dense correspondence between 3D shapes from unlabeled 3D shapes. As shown in Fig. 2, the key idea of our method is to represent the signed distance field of each 3D shape with the deformed template field modified by a correction field. The template field fuses all possible structures of the 3D shapes in the category. The deformation field defines the correspondence between 3D shapes and the template field. The correction field revises the deformed template field to the signed distance field of a 3D shape with a specific structure. Based on this representation, we have learned a template field shared by all 3D shapes and a compact latent space for encoding the deformation fields and the correction fields of all 3D shapes. As shown in Fig. 2, our method successfully learns the shape space and reasonable dense correspondence between 3D shapes with different structures. The learned dense shape correspondence also enables texture transfer between different shapes. As a result, the appearance or texture of all 3D shapes can be separated from the underlying geometry and encoded in the shared template field.

3.2 Learning 3D structures

Learning the common structure of 3D shapes in a class and their variations is a critical task in shape space modeling. From the computational perspective, the structures can be regarded as the abstraction of 3D shapes or a region-level correspondence between 3D shapes. As a result, researchers have developed a set of deep learning based shape abstraction methods and shape segmentation methods for learning the structures of 3D shapes. According to the data used for training, these methods can be classified into supervised methods, semi-supervised methods, and unsupervised methods.

Supervised methods (Niu et al., 2018; Mo et al., 2019; Yu FG et al., 2019) formulate this problem

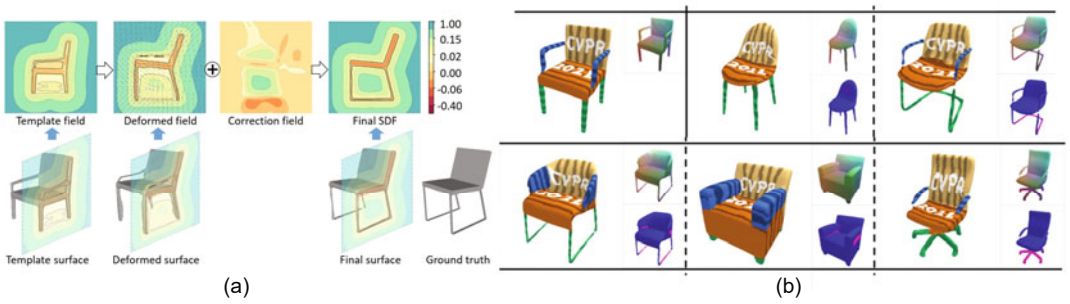

Fig. 2 Learning the space of 3D shapes and their dense correspondence: (a) the deformed implicit field representation in Deng et al. (2021) models the implicit field of a 3D shape by the sum of a deformed template field and a correction field; (b) the dense correspondence of 3D shapes learned by the method in Deng et al. (2021). Reprinted from Deng et al. (2021), Copyright 2021, with permission from IEEE

as 3D semantic segmentation and solve the problem with DNNs trained from the labeled datasets. As the amount of segmented 3D shape data is limited, a set of semi-supervised methods have been developed to learn the networks from a small set of segmented 3D shapes and a large collection of unlabeled 3D shapes. On one hand, these methods can learn and reveal the semantic structure of 3D shapes, especially small parts and parts with large geometric variations. The results are consistent with human perception and understanding with the help of human-labeled ground truth. On the other hand, these methods rely on the labeled dataset for training and are difficult to apply to unseen 3D shape classes.

Unsupervised methods (Tulsiani et al., 2017; Sun et al., 2019; Paschalidou et al., 2021; Yang KZ and Chen, 2021) formulate the structure learning as a shape abstraction problem, where each shape is approximated by a sparse set of 3D primitives (e.g., cuboids) shared by all 3D shapes. In Sun et al. (2019), we proposed an adaptive hierarchical cuboid representation for modeling variant hierarchical structures of 3D shapes in a class. Our key observation is that although 3D objects in a class have various structures, they share a common structure at a higher abstraction level and exhibit various structure details in different detailed abstraction levels. We also developed a neural network for constructing this adaptive hierarchical representation from unlabeled 3D shape data. After training, the learned network can successfully derive the number of cuboids, their adaptive hierarchy, and parameters for each 3D shape in the class and reveal the shape structures. Fig. 3 demonstrates the adaptive hierarchical cuboid representation and the results generated by our method for 3D airplanes. Compared to supervised methods and semi-supervised methods, these unsupervised methods can be used to derive the shape structure of unseen shape collections. Without the supervision of semantic labels, the structures learned by these methods are totally determined by 3D geometry and thus may be inconsistent with human perception.

3.3 Learning 3D shape space from 2D images

With the techniques described above, we can learn the 3D shape space from a large collection of 3D objects modeled by the artist or captured from the real world. Unfortunately, the difficulty of 3D acquisition and 3D modeling limits the quality and scale of the 3D dataset available for deep learning.

To solve this issue, researchers proposed a set of methods for learning 3D shape space from 2D image collections. The key idea of these methods is to train a generative adversarial network (GAN) that can generate 3D representations of the objects whose 2D projections are indistinguishable from the 2D input images. Different from traditional methods that use multiple-view correspondence or shading cues to reconstruct the 3D shape of a 3D object from its images, these methods learn the space of multiple 3D shapes by discriminating the images rendered from the 3D objects sampled in the learned space from the input real images. Early methods (Gadelha et al., 2017; Li et al., 2019) learned a GAN for generating

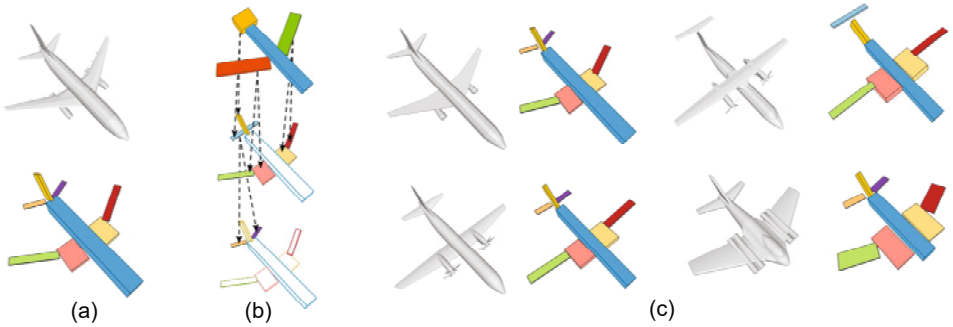

Fig. 3 Unsupervised 3D structure learning: (a) given 3D shapes as input (top), the network in Sun et al. (2019) infers the cuboid abstraction of the input 3D shapes (bottom); (b) the adaptive hierarchical cuboid representation, in which the cuboids at different levels are selected for approximating the input 3D shape; (c) the shape abstraction results generated by the method in Sun et al. (2019), where the different structures of 3D shapes are successfully learned and modeled by the adaptive hierarchical cuboid representation. Reprinted from Sun et al. (2019), Copyright 2019, with permission from the authors, licensed under CC BY

Fig. 4 The 3D shapes and objects generated by the GAN learned from 2D images: (a) the volumetric 3D bird shapes generated by the GAN in Li et al. (2019) (reprinted from Li et al. (2019), Copyright 2019, with permission from IEEE); (b) the images of the 3D neural radiance manifolds of human faces and cat heads rendered from different views, which are generated by the learned GANs in Deng et al. (2022) (reprinted from Deng et al. (2022), Copyright 2022, with permission from the authors, licensed under CC BY)

volumetric 3D shapes from 2D images of a 3D object collection. Recent methods (Chan et al., 2021; Deng et al., 2022) learned a GAN for generating 3D geometry and the appearance of 3D objects encoded by neural radiance fields from a large collection of 2D real images. Fig. 4 illustrates the 3D shapes generated by the GAN learned from 2D real images in Li et al. (2019) and Deng et al. (2022).

4 Looking forward

Three-dimensional shape space learning plays an important role in 3D shape understanding and authoring and is critical for constructing the visual concept of the real-world 3D objects. In this paper, we discuss the main challenges in 3D shape space learning and review recent research works for solving these challenges.

Although researchers have made great progress in developing efficient techniques for 3D shape space learning, there are still many open problems in this area. Below we list a few of them in which we are interested and hope this incomplete list will inspire more discussions and attract more investigations in this area.

1. Three-dimensional learning from 2D images

As described in Section 3.3, GAN-based methods provide a new paradigm for learning the shape and appearance space of 3D objects from 2D image collections. Different from traditional 3D reconstruction techniques that are built on a solid theoretical foundation and have been well studied, there are many unknowns in the theory and practice of this new paradigm. In theory, it is unclear how many images are required for reconstructing a 3D shape space. It is also unclear in what conditions we can reconstruct the 3D shape space from 2D image collections and in what conditions we cannot. In practice, we need to develop efficient representations and robust algorithms for learning the space of detailed 3D shapes from as few 2D images as possible.

2. Disentangled representation of 3D objects

In this paper, we focus mainly on the challenges and techniques for 3D shape space learning. However, real-world 3D objects have multiple attributes, such as 3D shapes, surface appearance, and dynamics. Pan (2019) pointed out that the visual concept of a 3D object class should consist of the prototypes and domains of all the attributes of 3D objects in the class. To this end, we need to develop a disentangled representation of 3D objects for modeling the 3D shape and appearance of 3D objects, as well as their dynamics, in which each attribute can be separately modeled and manipulated. We also need to develop efficient methods for learning the space of 3D objects with disentangled attributes from an unlabeled 3D dataset or image/video collections.

3. Online learning of visual concepts

Although human beings can gradually learn new visual concepts from the surrounding 3D environment and aggregate their visual knowledge, existing DNN models are difficult to update with new inputs after training. It is interesting to design new DNN architectures, develop new training strategies and learning schemes so that we can keep learning and updating the 3D object space model with online input, and finally obtain a model of all 3D objects in the real world.

4. Closing the loop of visual concept learning and application

After 3D shape space learning, the learned 3D shape space can be used for various downstream tasks, such as 3D modeling, shape classification and segmentation, as well as object detection and scene understanding. Furthermore, the learned 3D shape space and the outputs of these downstream tasks can be used for visual concept learning. It is interesting to explore how to leverage the learned visual concept or visual knowledge for 3D shape space learning. Another interesting research topic is how to develop an efficient user interface and interaction methods that can exploit the user's visual knowledge for efficient 3D shape space learning.

5. From visual concept to visual knowledge and beyond

We believe that 3D shape space learning is a small but important step toward constructing the "visual concept" of all 3D objects and scenes in the real world. The techniques and representations for 3D shape space and "visual concept" learning will not only advance the traditional 3D computer graphics with state-of-the-art artificial intelligence techniques, but also make the graphics representation a cornerstone in constructing the visual knowledge for artificial intelligence. Early cognitive studies have shown that "the quantity of visual knowledge present in human memory is larger than that of verbal knowledge, and that "understanding of verbal knowledge further requires assistance from this visual knowledge" (Pan, 2019, 2021b). Inspired by this observation, the constructed visual knowledge will finally be fused with state-of-the-art deep learning representations and techniques developed for vision and natural language processing together and help artificial intelligence evolve into a new era.

Acknowledgements

This paper is based on the author's presentations in the first and second workshops on visual knowledge and visual intelligence. The author would like to thank all workshop attendees for the insightful discussions. The author also thanks Prof. Yunhe PAN and Dr. Heung-Yeung SHUM for their invaluable comments on the topics presented in the paper. Finally, the author thanks all collaborators for our research works presented in Li et al. (2019), Sun et al. (2019), and Deng et al. (2021, 2022).

Compliance with ethics guidelines

Xin TONG declares that he has no conflict of interest.

References

Bai S, Bai X, Zhou ZC, et al., 2016. GIFT: a real-time and scalable 3D shape search engine. IEEE Conf on Computer Vision and Pattern Recognition, p.5023-5032. https://doi.org/10.1109/CVPR.2016.543

Cao C, Weng YL, Zhou S, et al., 2014. FaceWareHouse: a 3D facial expression database for visual computing. *IEEE Trans Visual Comput Graph*, 20(3):413-425. https://doi.org/10.1109/TVCG.2013.249

Chan ER, Monteiro M, Kellnhofer P, et al., 2021. pi-GAN: periodic implicit generative adversarial networks for 3D-aware image synthesis. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.5799-5809. https://doi.org/10.1109/CVPR46437.2021.00574

Chen ZQ, Zhang H, 2019. Learning implicit fields for generative shape modeling. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.5932-5941. https://doi.org/10.1109/CVPR.2019.00609

Deng Y, Yang JL, Tong X, 2021. Deformed implicit field: modeling 3D shapes with learned dense correspondence. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.10286-10296. https://doi.org/10.1109/CVPR46437.2021.01015

Deng Y, Yang J, Xiang J, et al., 2022. GRAM: generative radiance manifolds for 3D-aware image generation. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.10673-10683.

Egger B, Smith WA, Tewari A, 2020. 3D morphable face models past, present, and future. *ACM Trans Graph*, 39(5):157. https://doi.org/10.1145/3395208

Gadelha M, Maji S, Wang R, 2017. 3D shape induction from 2D views of multiple objects. Int Conf on 3D Vision, p.402-411. https://doi.org/10.1109/3DV.2017.00053

Groueix T, Fisher M, Kim VG, et al., 2018. A Papier-Mache approach to learning 3D surface generation. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.216-224. https://doi.org/10.1109/CVPR.2018.00030

Hughes JF, van Dam A, McGuire M, et al., 2013. Computer Graphics: Principles and Practice (3rd Ed.). Addison-Wesley, Upper Saddle River, USA.

Jiang C, Huang J, Tagliasacchi A, et al., 2020. ShapeFlow: learnable deformation flows among 3D shapes. Advances in Neural Information Processing Systems 33, p.9745-9757.

Jin YW, Jiang DQ, Cai M, 2020. 3D reconstruction using deep learning: a survey. *Commun Inform Syst*, 20(4): 389-413. https://doi.org/10.4310/CIS.2020.v20.n4.a1

Li X, Dong Y, Peers P, et al., 2019. Synthesizing 3D shapes from silhouette image collections using multi-projection generative adversarial networks. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.5530-5539. https://doi.org/10.1109/CVPR.2019.00568

Liu F, Liu XM, 2020. Learning implicit functions for topology-varying dense 3D shape correspondence. Proc 34th Int Conf on Neural Information Processing Systems, p.4823-4834.

Loper M, Mahmood N, Romero J, et al., 2015. SMPL: a skinned multi-person linear model. *ACM Trans Graph*, 34(6):248. https://doi.org/10.1145/2816795.2818013

Lun ZL, Gadelha M, Kalogerakis E, et al., 2017. 3D shape reconstruction from sketches via multi-view convolutional networks. Proc Int Conf on 3D Vision, p.67-77. http://arxiv.org/abs/1707.06375

Masci J, Boscaini D, Bronstein MM, et al., 2015. Geodesic convolutional neural networks on Riemannian manifolds. Proc IEEE Int Conf on Computer Vision Workshop, p.832-840. https://doi.org/10.1109/ICCVW.2015.112

Měch R, Prusinkiewicz P, 1996. Visual models of plants interacting with their environment. Proc 23rd Annual Conf on Computer Graphics and Interactive Techniques, p.397-410. https://doi.org/10.1145/237170.237279

Mescheder L, Oechsle M, Niemeyer M, et al., 2019. Occupancy networks: learning 3D reconstruction in function space. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.4455-4465. https://doi.org/10.1109/CVPR.2019.00459

Mo KC, Zhu SL, Chang AX, et al., 2019. PartNet: a large-scale benchmark for fine-grained and hierarchical part-level 3D object understanding. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.909-918. https://doi.org/10.1109/CVPR.2019.00100

Müller P, Wonka P, Haegler S, et al., 2006. Procedural modeling of buildings. ACM SIGGRAPH Papers, p.614-623. https://doi.org/10.1145/1141911.1141931

Niu CJ, Li J, Xu K, 2018. Im2Struct: recovering 3D shape structure from a single RGB image. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.4521-4529. https://doi.org/10.1109/CVPR.2018.00475

Pan YH, 2019. On visual knowledge. *Front Inform Technol Electron Eng*, 20(8):1021-1025. https://doi.org/10.1631/FITEE.1910001

Pan YH, 2021a. Miniaturized five fundamental issues about visual knowledge. *Front Inform Technol Electron Eng*, 22(5):615-618. https://doi.org/10.1631/FITEE.2040000

Pan YH, 2021b. On visual understanding. *Front Inform Technol Electron Eng*, early access. https://doi.org/10.1631/FITEE.2130000

Park JJ, Florence P, Straub J, et al., 2019. DeepSDF: learning continuous signed distance functions for shape representation. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.165-174. https://doi.org/10.1109/CVPR.2019.00025

Paschalidou D, Katharopoulos A, Geiger A, et al., 2021. Neural parts: learning expressive 3D shape abstractions with invertible neural networks. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.3204-3215. https://doi.org/10.1109/CVPR46437.2021.00322

Qi CR, Su H, Mo KC, et al., 2017. PointNet: deep learning on point sets for 3D classification and segmentation. IEEE Conf on Computer Vision and Pattern Recognition, p.77-85. https://doi.org/10.1109/CVPR.2017.16

Riegler G, Ulusoy AO, Geiger A, 2017. OctNet: learning deep 3D representations at high resolutions. IEEE Conf on Computer Vision and Pattern Recognition, p.6620-6629. https://doi.org/10.1109/CVPR.2017.701

Sinha A, Bai J, Ramani K, 2016. Deep learning 3D shape surfaces using geometry images. Proc 14th European Conf on Computer Vision, p.223-240. https://doi.org/10.1007/978-3-319-46466-4_14

Su H, Maji S, Kalogerakis E, et al., 2015. Multi-view convolutional neural networks for 3D shape recognition. IEEE Int Conf on Computer Vision, p.945-953. https://doi.org/10.1109/ICCV.2015.114

Sun CY, Zou QF, Tong X, et al., 2019. Learning adaptive hierarchical cuboid abstractions of 3D shape collections. *ACM Trans Graph*, 38(6):241. https://doi.org/10.1145/3355089.3356529

Tulsiani S, Su H, Guibas LJ, et al., 2017. Learning shape abstractions by assembling volumetric primitives. IEEE Conf on Computer Vision and Pattern Recognition, p.1466-1474. https://doi.org/10.1109/CVPR.2017.160

Wang NY, Zhang YD, Li ZW, et al., 2018. Pixel2Mesh: generating 3D mesh models from single RGB images. Proc 15th European Conf on Computer Vision, p.55-71. https://doi.org/10.1007/978-3-030-01252-6_4

Wang PS, Liu Y, Guo YX, et al., 2017. O-CNN: octree-based convolutional neural networks for 3D shape analysis. *ACM Trans Graph*, 36(4):72. https://doi.org/10.1145/3072959.3073608

Wang PS, Liu Y, Tong X, 2022. Dual octree graph networks for learning adaptive volumetric shape representations. *ACM Trans Graph*, 41(4):103. https://doi.org/10.1145/3528223.3530087

Wen C, Zhang YD, Li ZW, et al., 2019. Pixel2Mesh++: multi-view 3D mesh generation via deformation. IEEE/CVF Int Conf on Computer Vision, p.1042-1051. https://doi.org/10.1109/ICCV.2019.00113

Wu ZR, Song SR, Khosla A, et al., 2015. 3D ShapeNets: a deep representation for volumetric shapes. Proc IEEE Conf on Computer Vision and Pattern Recognition, p.1912-1920. https://doi.org/10.1109/CVPR.2015.7298801

Xiao YP, Lai YK, Zhang FL, et al., 2020. A survey on deep geometry learning: from a representation perspective. *Comput Visual Med*, 6(2):113-133. https://doi.org/10.1007/s41095-020-0174-8

Yang J, Mo KC, Lai YK, et al., 2023. DSG-Net: learning disentangled structure and geometry for 3D shape generation. *ACM Trans Graph*, 42(1):1. https://doi.org/10.1145/3526212

Yang KZ, Chen XJ, 2021. Unsupervised learning for cuboid shape abstraction via joint segmentation from point clouds. *ACM Trans Graph*, 40(4):152. https://doi.org/10.1145/3450626.3459873

Yu FG, Liu K, Zhang Y, et al., 2019. PartNet: a recursive part decomposition network for fine-grained and hierarchical shape segmentation. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.9483-9492. https://doi.org/10.1109/CVPR.2019.00972

Yu LQ, Li XZ, Fu CW, et al., 2018. PU-Net: point cloud upsampling network. IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.2790-2799. https://doi.org/10.1109/CVPR.2018.00295

Zheng XY, Liu Y, Wang PS, et al., 2022. SDF-StyleGAN: implicit SDF-based StyleGAN for 3D shape generation. https://arxiv.org/abs/2206.12055

Zheng ZR, Yu T, Dai QH, et al., 2021. Deep implicit templates for 3D shape representation. Proc IEEE/CVF Conf on Computer Vision and Pattern Recognition, p.1429-1439. https://doi.org/10.1109/CVPR46437.2021.00148

Zuffi S, Kanazawa A, Jacobs DW, et al., 2017. 3D Menagerie: modeling the 3D shape and pose of animals. IEEE Conf on Computer Vision and Pattern Recognition, p.5524-5532. https://doi.org/10.1109/CVPR.2017.586