Caixia LIU, Dehui KONG, Shaofan WANG, Zhiyong WANG, Jinghua LI, Baocai YIN, 2021. Deep 3D reconstruction: methods, data, and challenges. *Frontiers of Information Technology & Electronic Engineering*, 22(5):652-672.
https://doi.org/10.1631/FITEE.2000068

Deep 3D reconstruction: methods, data, and challenges

Key words: Deep learning models; Three-dimensional reconstruction; Recurrent neural network; Deep autoencoder; Generative adversarial network; Convolutional neural network

Corresponding author: Shaofan WANG
E-mail: wangshaofan@bjut.edu.cn
ORCID: https://orcid.org/0000-0002-3045-624X

Motivation

1. Three-dimensional (3D) reconstruction of shapes is an important research topic in the fields of computer vision, computer graphics, pattern recognition, and virtual reality.

2. Existing 3D reconstruction methods suffer from two bottlenecks: (1) they involve multiple manually designed states which can lead to cumulative errors, but can hardly learn semantic features of 3D shapes automatically; (2) they depend heavily on the content and quality of images, as well as precisely calibrated cameras. As a result, it is difficult to improve the reconstruction accuracy of those methods.

3. Three-dimensional reconstruction methods based on deep learning overcome both of these bottlenecks by automatically learning semantic features of 3D shapes from low-quality images using deep networks. However, while these methods have various architectures, in-depth analysis and comparisons of them are unavailable so far.

Summary

We present a comprehensive survey of 3D reconstruction methods based on deep learning.

- Based on different deep learning model architectures, we divide 3D reconstruction methods based on deep learning into four types, recurrent neural network, deep autoencoder, generative adversarial network, and convolutional neural network based methods, and analyze the corresponding methodologies carefully.
- We investigate four representative databases that are commonly used by the above methods in detail.
- We give a comprehensive comparison of 3D reconstruction methods based on deep learning, which consists of the results of different methods with respect to the same database, the results of each method with respect to different databases, and the robustness of each method with respect to the number of views.
- We discuss future development of 3D reconstruction methods based on deep learning.

Methods

Methodologies of deep 3D reconstruction methods:

1. RNN-based 3D reconstruction methods
   - Direct decoding based reconstruction
   - Indirect decoding based reconstruction
2. DAE-based 3D reconstruction methods
   - Database matching based reconstruction
   - Deconvolution decoding based reconstruction

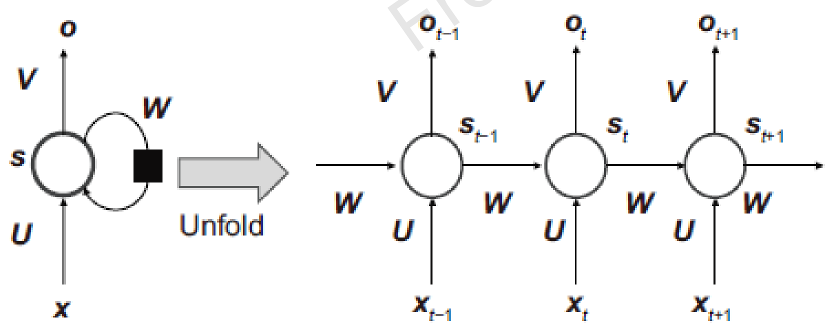

Fig. 4 Structure of a recurrent neural network

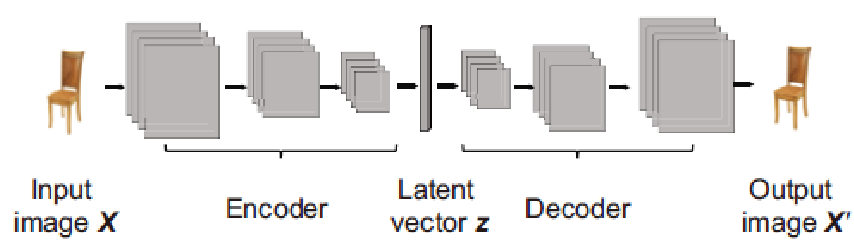

Fig. 5 Structure of deep autoencoder

Methods

3. GAN-based 3D reconstruction methods

- Single GAN based reconstruction
- Fusion GAN based reconstruction

4. CNN-based 3D reconstruction methods

- Skeleton-based reconstruction
- Voxel-based reconstruction
- Point cloud based reconstruction
- Mesh-based reconstruction

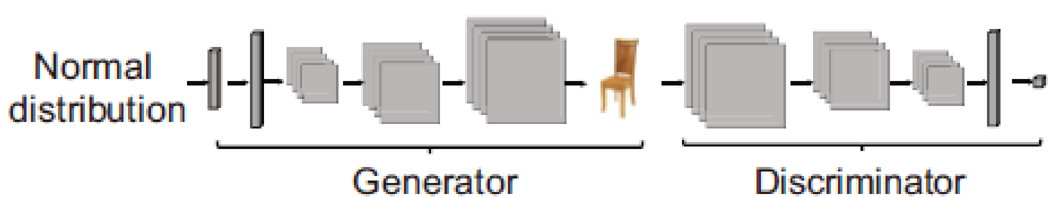

Fig. 6 Structure of a generative adversarial network

Data

Commonly used 3D geometric databases

- ShapeNet database, ModelNet database
- PASCAL3D+ database, IKEA database

Table 1 Public 3D geometric databases

Database	Data type	Number of categories	3D shape		Image		
			Number	Type	Number	Type	Camera
ShapeNet (Chang et al., 2015)	Synthetic	55/270§	51 300/12 000§	Mesh	–	Rendered	✓
ModelNet (Wu ZR et al., 2015)	Synthetic	662	127 915	Mesh	–	Rendered	
PASCAL3D+ (Xiang et al., 2014)	Real-world, indoor, outdoor	12	36 000	Mesh	30 899	Real-world	
IKEA (Lim et al., 2014)	Real-world, indoor	7	219	Mesh	759	Real-world	
Pix3D (Sun XY et al., 2018)	Real-world, indoor	9	418	Mesh	16 913	Real-world	
ObjectNet3D (Xiang et al., 2016)	Real-world, indoor	100	44 147	Mesh	90 127	Real-world	✓
NYUdv2 (Silberman et al., 2012)	Real-world, indoor	894	35 064	Mesh	1449	Real-world	✓

§: Red for ShapeNetCore and blue for ShapeNetSem. ShapeNet: https://www.shapenet.org; ModelNet: http://modelnet.cs.princeton.edu; PASCAL3D+: http://cvgl.stanford.edu/projects/objectnet3d; IKEA: http://ikea.csail.mit.edu; Pix3D: http://pix3d.csail.mit.edu; ObjectNet3D: http://cvgl.stanford.edu/projects/objectnet3d; NYUdv2: https://cs.nyu.edu/~silberman/datasets. "–" denotes that there are no images in the database, but images can be obtained by rendering 3D shapes from the database. References to color refer to the online version of this table

Data

Table 2 Deep learning models used on 3D geometric databases

Method	Model	Literature	ShapeNet	ModelNet	PASCAL3D+	IKEA	ObjectNet3D	NYUdv2	Pix3D	Note
RNN	3D-R2N2	Choy et al. (2016)	✓	–	✓	–	–	–	–	Supervised
RNN	3D-PRNN	Zou et al. (2017)	–	✓	–	–	–	✓	–	Supervised
DAE	TL	Girdhar et al. (2016)	✓	–	–	✓	–	–	–	Supervised
DAE	Image2Mesh	Pontes et al. (2018)	✓	–	✓	–	–	–	–	Supervised
DAE	3D-PCG	Lun et al. (2017)	✓	–	–	–	–	–	–	Supervised
DAE	SketchModeling	Nash and Williams (2017)	✓	–	–	–	–	–	–	Supervised
DAE	PointOutNet	Fan et al. (2017)	✓	–	–	–	–	–	–	Supervised
DAE	Pix2Vox	Xie et al. (2019)	✓	–	✓	–	–	–	–	Supervised
DAE	Base-AttSets	Yang et al. (2020)	✓	–	–	–	–	–	–	Supervised
DAE	HRShapeCompletion	Han XG et al. (2017)	✓	–	–	–	–	–	–	Supervised
DAE	Object-completion	Varley et al. (2017)	✓	–	–	–	–	–	–	Supervised
DAE	MarrNet	Wu JJ et al. (2017)	–	–	–	–	–	–	✓	Supervised
DAE	HSP	Häne et al. (2017)	✓	–	–	–	–	–	–	Supervised
GAN	3D-GAN	Wu JJ et al. (2016a)	–	✓	–	–	–	–	–	Unsupervised
GAN	3D-IWGAN	Smith and Meger (2017)	–	✓	–	–	–	–	–	Unsupervised
DAE-GAN	3D-RecGAN++	Yang et al. (2019)	✓	–	–	–	–	–	–	Supervised
DAE-GAN	3D-RecGAN	Yang et al. (2018)	–	✓	–	–	–	–	–	Supervised
DAE-GAN	3D-VAE-GAN	Wu JJ et al. (2016a)	–	–	–	✓	–	–	–	Supervised
DAE-GAN	3D-VAE-IWGAN	Smith and Meger (2017)	–	✓	–	✓	–	–	–	Supervised
DAE-GAN	U3DRec	Wang LJ and Fang (2017)	–	–	✓	–	–	–	–	Unsupervised
RNN-GAN	McRecon	Gwak et al. (2017)	✓	–	–	–	✓	–	–	Weakly supervised
CNN	3D-INN	Wu JJ et al. (2016b)	–	–	✓	✓	–	–	–	Supervised
CNN	Pixel2Mesh	Wang NY et al. (2018)	✓	–	–	–	–	–	–	Supervised

Data

Performance comparison of deep 3D reconstruction methods

- Comparison of reconstruction based on different deep models
- Performance comparison of reconstruction on different databases

Table 3 Quantitative results of different methods on ShapeNet

Method	Model	Literature	IoU					3D	Resolution	Note
			Bench	Car	Chair	Table	Mean	representation		
RNN	3D-R2N2	Choy et al. (2016)	0.421	0.798	0.466	0.513	0.550	Voxel grid	32^3	Supervised
	HRShapeCompletion	Han XG et al. (2017)	0.611	–	0.524	0.615	0.583	Voxel grid	32^3	Supervised
DAE	HSP	Häne et al. (2017)	–	0.698	0.361	–	0.530	Voxel grid	256^3	Supervised
	Image2Mesh	Pontes et al. (2018)	–	0.664	0.403	–	0.534	Mesh	–	Supervised
	PointOutNet	Fan et al. (2017)	0.550	0.831	0.544	0.606	0.633	Point cloud	32^3	Supervised
	Pix2Vox	Xie et al. (2019)	0.613	0.806	0.599	0.613	0.658	Voxel grid	32^3	Supervised
	Base-AttSets	Yang et al. (2020)	0.569	0.848	0.571	0.597	0.646	Voxel grid	32^3	Supervised
	Object-completion	Varley et al. (2017)	0.653	–	0.619	0.678	0.650	Voxel grid	32^3	Supervised
	3D-RecAE	Yang et al. (2018)	0.800	–	0.790	0.808	0.799	Voxel grid	32^3	Supervised
	3D-RecAE	Yang et al. (2018)	0.733	–	0.736	0.759	0.742	Voxel grid	64^3	Supervised
DAE-GAN	3D-RecGAN++	Yang et al. (2019)	0.806	–	0.793	0.821	0.807	Voxel grid	32^3	Supervised
DAE-GAN	3D-RecGAN++	Yang et al. (2019)	0.745	–	0.741	0.772	0.753	Voxel grid	64^3	Supervised
RNN-GAN	McRecon	Gwak et al. (2017)	0.295	0.562	0.350	0.353	0.390	Voxel grid	32^3	Weakly supervised

Data

- Comparison of reconstruction based on different deep models
- Performance comparison of reconstruction on different databases

Table 4 Quantitative results of different methods on ModelNet

Method	Model	Literature	IoU						3D representation	Resolution	Note
			Chair	Table	Night stand	Stool	Toilet	Mean			
RNN	3D-PRNN	Zou et al. (2017)	0.245	0.188	0.204	–	–	0.212	Point cloud	30^3	Supervised
DAE	Object-completion	Varley et al. (2017)	0.564	–	–	0.273	0.503	0.447	Voxel grid	40^3	Supervised
DAE	3D-RecAE	Yang et al. (2018)	0.633	–	–	0.488	0.520	0.547	Voxel grid	64^3	Supervised
DAE-GAN	3D-RecGAN	Yang et al. (2018)	0.661	–	–	0.501	0.569	0.577	Voxel grid	64^3	Supervised

Table 5 Quantitative results of different methods on PASCAL3D+

Method	Model	Literature	IoU							3D representation	Resolution	Note
			Aeroplane	Bus	Car	Chair	Sofa	TV	Mean			
RNN	3D-R2N2	Choy et al. (2016)	0.544	0.816	0.699	0.280	0.332	0.574	0.541	Voxel grid	32^3	Supervised
DAE	Image2Mesh	Pontes et al. (2018)	0.366	0.280	0.371	0.236	0.207	–	0.292	Mesh	–	Supervised
DAE	Pix2Vox	Xie et al. (2019)	0.690	0.760	0.657	0.593	0.634	0.694	0.671	Voxel grid	32^3	Supervised
DAE-GAN	U3DRec	Wang LJ and Fang (2017)	–	–	0.634	0.241	0.450	0.247	0.393	Voxel grid	32^3	Unsupervised

Data

- Comparison of reconstruction based on different deep models
- Performance comparison of reconstruction on different databases

Table 6 Quantitative results of different methods on IKEA

Method	Model	Literature	AP							3D representation	Resolution	Note
			Bed	Bookcase	Chair	Desk	Sofa	Table	Mean			
DAE	TL	Girdhar et al. (2016)	56.3	30.2	32.9	25.8	71.7	23.3	40.03	Voxel grid	20^3	Supervised
DAE-GAN	3D-VAE-GAN	Wu JJ et al. (2016a)	63.2	46.3	47.2	40.7	78.8	42.3	53.08	Voxel grid	20^3	Supervised
DAE-GAN	3D-VAE-IWGAN	Smith and Meger (2017)	77.7	51.8	56.2	49.8	82.0	52.6	61.68	Voxel grid	32^3	Supervised
RNN-GAN	McRecon	Gwak et al. (2017)	–	–	32.0	28.6	55.7	29.0	36.33	Voxel grid	32^3	Weakly supervised

Table 7 Quantitative results of different 3D representations for CNN-based reconstruction methods on ShapeNet

Method	Literature	CD					EMD					3D representation
		Bench	Car	Chair	Table	Mean	Bench	Car	Chair	Table	Mean	
3D-R2N2	Choy et al. (2016)	1.891	0.845	1.432	1.116	1.321	1.136	1.670	1.466	1.641	1.478	Voxel grid
Pixel2Mesh	Wang NY et al. (2018)	0.624	0.268	0.610	0.498	0.500	0.965	1.297	1.399	1.480	1.285	Mesh
PointOutNet	Fan et al. (2017)	0.629	0.333	0.645	0.517	0.531	1.113	1.747	1.946	2.121	1.732	Point cloud

Data

- Impact of the number of views on 3D reconstruction methods

Table 8 Quantitative results with respect to the number of views on ShapeNet

Method	Model	Literature	Number of views	IoU Chair	IoU Car	IoU Table	IoU Bench	IoU Mean	3D representation	Resolution	Note
RNN	3D-R2N2	Choy et al. (2016)	1	0.466	0.798	0.513	0.421	0.550	Voxel grid	32^3	Supervised
			2	0.515	0.821	0.550	0.484	0.593			
			3	0.533	0.829	0.564	0.502	0.607			
			4	0.541	0.833	0.573	0.516	0.616			
			5	0.550	0.836	0.580	0.527	0.623			
RNN-GAN	McRecon	Gwak et al. (2017)	1	0.350	0.562	0.353	0.295	0.390	Voxel grid	32^3	Weakly supervised
			5	0.437	0.614	0.420	0.401	0.468			

Challenges

Future development of 3D reconstruction:

- Supervised learning versus unsupervised learning models
- Matrix neural networks versus nonmatrix neural networks
- Low-resolution outputs versus high-resolution outputs

Conclusions

1. More knowledge should be mined from existing data by deep networks to improve 3D reconstruction accuracy.

2. Generative models are more conducive to reconstruction than non-generative models because of inherent architectures.

Caixia LIU is a PhD candidate at Beijing University of Technology, Beijing, China. She received her BS and MS degrees from Beijing Union University, China. Her research interests include computer graphics, image processing, and pattern recognition.

Prof. Dehui KONG received her MS and PhD degrees from Beihang University, in 1992 and 1996, respectively. She is a professor at Beijing Institute of Artificial Intelligence, Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Faculty of Information Technology, Beijing University of Technology, Beijing, China. Her research interests include virtual reality, computer graphics, and pattern recognition.

Prof. Shaofan WANG received his BS and PhD degrees in computational mathematics from Dalian University of Technology, Dalian, China, in 2003 and 2010, respectively. He is an associate professor at Beijing Institute of Artificial Intelligence, Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Faculty of Information Technology, Beijing University of Technology, Beijing, China. His research interests include pattern recognition and machine learning.

Prof. Zhiyong WANG received his BS and MS degrees in electronic engineering from South China University of Technology, Guangzhou, China, and his PhD degree from The Hong Kong Polytechnic University, Hong Kong, China. He is currently an associate professor and an associate director of the Multimedia Laboratory with the School of Computer Science, The University of Sydney, Australia. His research interests focus on multimedia computing, including multimedia information processing, retrieval and management, Internet-based multimedia data mining, human-centerd multimedia computing, and pattern recognition.

Prof. Jinghua LI received her MS and PhD degrees from Beijing University of Technology, Beijing, China, in 2005 and 2014, respectively. She is an associate professor at Beijing Institute of Artificial Intelligence, Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Faculty of Information Technology, Beijing University of Technology, Beijing, China. Her research interests include image processing and pattern recognition.

Prof. Baocai YIN received his BS, MS, and PhD degrees in computational mathematics from Dalian University of Technology, Dalian, China, in 1985, 1988, and 1993, respectively. He is currently a professor and dean of Beijing Institute of Artificial Intelligence, Beijing Key Laboratory of Multimedia and Intelligent Software Technology, Faulty of Information Technology, Beijing University of Technology, Beijing, China. His research interests include multimedia, image processing, computer vision, and pattern recognition.