International Journal of Advanced Smart Convergence Vol.12 No.4 142-146 (2023)
http://dx.doi.org/10.7236/IJASC.2023.12.4.142

IJASC 23-4-17

3D Object Generation and Renderer System based on VAE ResNet-GAN

Min-Su Yu [1], Tae-Won Jung [2], GyoungHyun Kim [3], Soonchul Kwon [4], Kye-Dong Jung [5]

[1] Master Student, Department of Smart Convergence, Kwangwoon University, Korea
[2] , Department of Immersive Content Convergence, Kwangwoon University, Korea
[3] Master Student, Department of Interdisciplinary Information System, Graduate School of Smart Convergence, Kwangwoon University, Korea
[4]Associate professor, Department of Interdisciplinary Information System, Graduate School of Smart Convergence, Kwangwoon University, Korea
[5]Professor, Ingenium College of Liberal Arts, Kwangwoon University, Korea
{ srejis[1], onom[2], hyunkim[3], ksc0226[1], gdchung [3]}@kw.ac.kr

Abstract

We present a method for generating 3D structures and rendering objects by combining VAE (Variational Autoencoder) and GAN (Generative Adversarial Network). This approach focuses on generating and rendering 3D models with improved quality using residual learning as the learning method for the encoder. We deep stack the encoder layers to accurately reflect the features of the image and apply residual blocks to solve the problems of deep layers to improve the encoder performance. This solves the problems of gradient vanishing and exploding, which are problems when constructing a deep neural network, and creates a 3D model of improved quality. To accurately extract image features, we construct deep layers of the encoder model and apply the residual function to learning to model with more detailed information. The generated model has more detailed voxels for more accurate representation, is rendered by adding materials and lighting, and is finally converted into a mesh model. 3D models have excellent visual quality and accuracy, making them useful in various fields such as virtual reality, game development, and metaverse.

Keywords: *variational autoencoder; generative adversarial network; residual learning; generation; reconstruction; voxel.*

1. Introduction

Rapid advances in 3D acquisition and modeling technologies allow 3D data to be efficiently captured from the real world or generated with easy-to-use modeling software. Additionally, recent advances in Internet tools, especially online repositories, allow 3D shapes to be shared between users [2]. 3D generative models utilizing

Manuscript Received: october. 14, 2023 / Revised: october. 20, 2023 / Accepted: october. 24, 2023
Corresponding Author: gdchung@kw.ac.kr
Tel: +82-02-940-5288   Fax: +82-02-940-5443
Professor, Ingenium College of Liberal Arts, Kwangwoon University, Korea

GANs can capture complex details, high-level structures, and realistic deformations of 3D objects or scenes. They have the potential to create diverse and visually appealing 3D content, opening application possibilities in entertainment, gaming, metaverse, virtual and augmented reality, computer graphics and other fields.

The proposed method is a system that combines improved object generation and a renderer using VAE ResNet-GAN and focuses on the problem of generating and rendering 3D objects with improved quality. We utilize recent advances in Convolutional Neural Network and Generative Adversarial Network to propose residual learning of the encoder, which reconstructs the layers by applying residual learning to training to generate 3D objects in VAE-GAN learning. It is also tightly coupled with the rendering system to achieve improved image quality.

2. Proposed Method

The system consists of VAE ResNet-GAN, Render, Mesh Transformation modules for object generation and, enhanced rendering for object generation. Our approach combines the latent space, a generator or decoder, which is the core component of VAE and GAN, which constructs the 3D shape, and integrates residual learning and shortcut connections into the encoder to generate 3D models with good quality and wide variation. Encoder training has great potential to reflect image features into voxels by constructing deep layers. Applying residual training instead of deep layers prevents overfitting and gradient vanishing and exploding.

Figure 1 shows the VAE ResNet-GAN based 3D object generation system. The residual VAE and GAN generate stable voxels, and 30 images are extracted by rendering the voxels with a virtual camera. Voxels are rendered into multiple views and re-represented as a mesh. As visualized in Figure 1, the overall architecture consists of two stages: (a) GAN coupled with residualVAE, (b) renderer and mesh converter for object rendering.

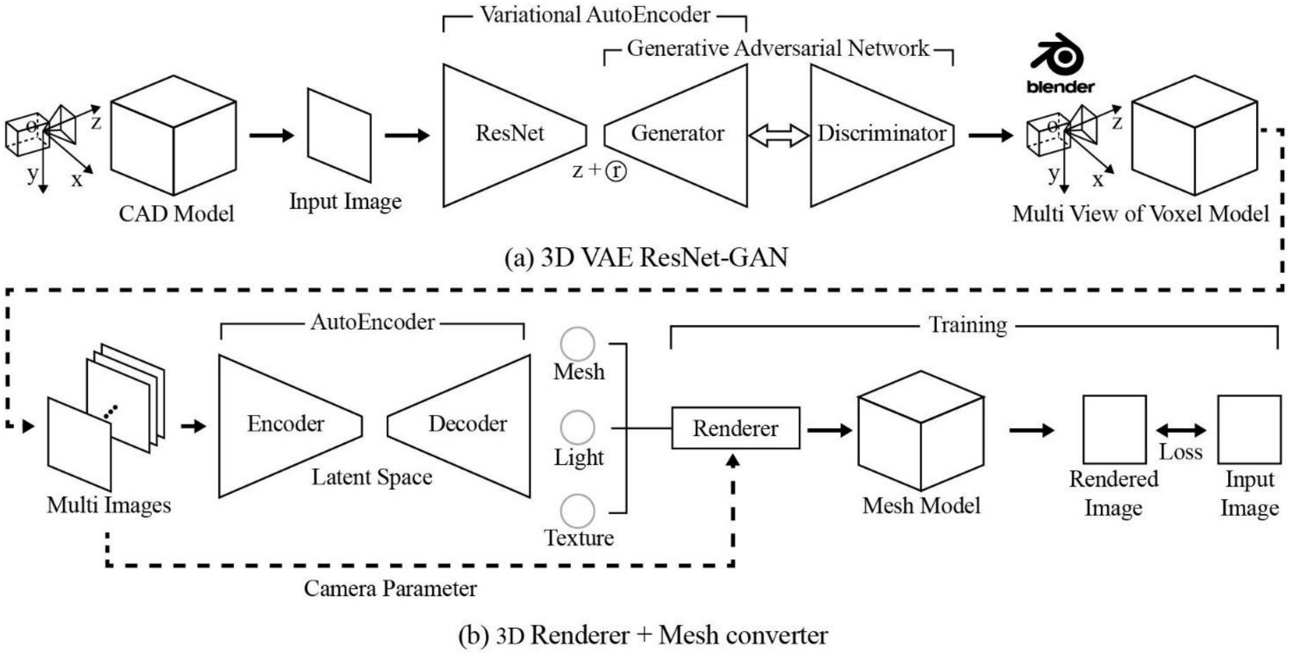

Figure 1. System architecture: (a) VAE Cascade-GAN, and (b) 3D Renderer + Mesh converter.

2.1 3D Object Generation using ResNet-GAN

In VAE, the structures of the encoder and decoder are fixed and learning is performed. Our method is a technology that configures layers by dividing the network into several stages in the encoder. This helps GANs learn more effectively to generate more complex 3D data. ResNet basically connects multiple residual blocks to form one large network. Each block learns by simply combining the result of performing a convolution operation with the result generated from the block in the previous step. Therefore, weight loss/explosion does not occur even in deep layers and the goal is to extract features from large images. In the last layer, the average and standard deviation of the image are extracted from the layer results using the identity and tanh functions, respectively. A 3D model is created through GAN by calculating the latent space with the extracted mean and standard deviation, and as learning progresses, the performance of the generator is supervised to create the model.Improving the performance of the encoder to extract features from the image helps the generator better learn the full distribution of the data, and also improves the discriminator's evaluation of the generated data.

In Figure 2, the encoder network is composed of blocks A, B, C, D, and E. In each block, two layers are paired and used as a residual block. If the stride of the residual block is 1, the output of the previous block is combined as is through skip connection. If it is not 1, skip connection is used using padding(?) to match the shape. For example, in Block B, padding is applied only to the first block for /2, and the remaining residual blocks of Block B are combined with the previous output as is.

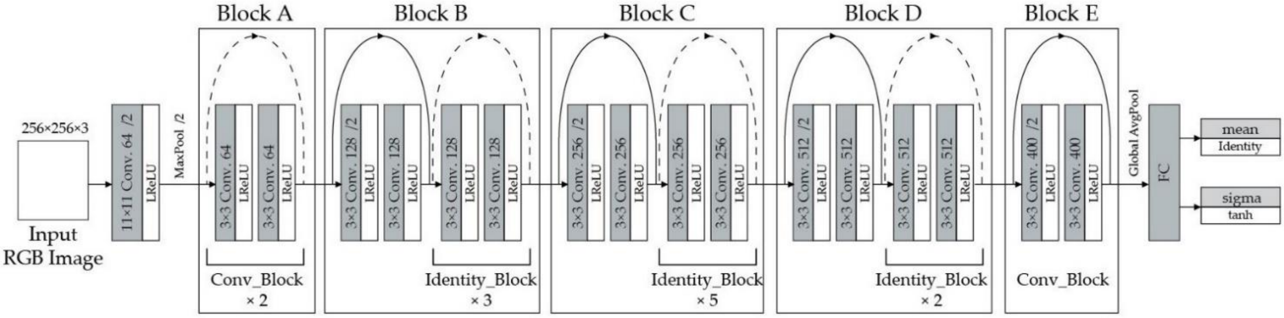

Figure 2. ResNet-34 Architecture of VAE module.

2.2 Rendering and Mesh Transformation

Our method enables more accurate and flexible 3D object prediction. Objects created from a progressively growing GAN can acquire multi-view images. In the image projection stage, an interpolation-based Differentiable Renderer is introduced to enable smoother projection. It can perform sophisticated projections through differentiable operations at the pixel level and achieve better visual quality. In the rendering stage, new methods are introduced to more effectively model light reflection and shadows. This predicts the object's surface properties in detail and produces better simulation results. Our rendering system consists of a render and mesh conversion module, an encoder module, a decoder module to generate 3D data, a loss calculation module for learning, and a prediction output module.

3. Experimental

We experimented with two networks, connecting a 3D Renderer + Mesh converter to each. The two networks under investigation were the proposed VAE ResNet-GAN and the existing 3D-VAE-IWGAN. Both the proposed VAE ResNet-GAN and 3D-VAE-IWGAN used the same hyperparameters. The learning rate was set to 0.0001, the batch size was 16, and the number of epochs was fixed at 4000.

The experiments were conducted on the Ubuntu 18.04 LTS operating system, utilizing the Ryzen 9 3900X CPU and Nvidia Titan XP 12GB GPU. The programs were developed using TensorFlow and PyTorch for deep learning tasks with GPU acceleration.

Accuracy experiments on AP (Average Precision), RMSE(Root Mean Square Error) and CD(Chamfer Distance) were evaluated on the ShapeNets Dataset and the proposed VAE ResNet-GAN improved voxel accuracy. Table 2 compares our results on the ShapeNets Dataset. VAE ResNet-GAN improved voxel generation.

Table 1. Comparison of evaluation of voxel generation results

Method	3D-VAE-IWGAN			VAE ResNet-GAN		
	Chair	Desk	Table	Chair	Desk	Table
AP	0.5648	0.4321	0.1957	0.5810	0.5245	0.4238
RMSE	0.2114	0.3338	0.3021	0.1949	0.2976	0.2431
CD	0.031	0.1339	0.4765	0.0251	0.0845	0.1323

Diversity and quality experiments on FID were evaluated on the ShapeNets Dataset and the proposed 3D object generation using Rendering module improved object generation diversity and quality. Table 2 compares our results on the ShapeNets Dataset. Rendering module improved object generation quality and diversity

Table 2. Comparison of object generation result evaluation by stage

Method	IWGAN-Mesh			ResNet-Mesh		
	Chair	Desk	Table	Chair	Desk	Table
object creation module	256.77	276.62	301.6	251.26	210.41	224.38
rendering module	210.43	181.7	230.62	198.76	166.09	175.15

4. Conclusion

The method we propose generates voxel objects using VAE ResNet-GAN, which applies residual learning and shortcut connections between VAE GAN-based networks and re-expresses the objects in mesh form through a rendering module to improve diversity and quality. To generate objects, VAE ResNet-GAN deeply configures the encoder layers and applies residual learning to prevent problems in deep layers. The rendering module uses DIB-R to re-express it in a more accurate and flexible mesh form, solving artifacts that are a problem with GAN.

In the system, VAE ResNet-GAN and rendering modules use pre-trained weights. The prediction value of VAE ResNet-GAN is output as a value between 0 and 1 and is a confidence value that indicates whether the voxel is filled in the corresponding cell location. To render, the confidence value is transformed by a threshold value and converted into 0 and 1 values. Because the voxels generated in this way are generated by GAN, an effect occurs. To solve this, a rendering module is used.

The rendering module uses the original mesh for training to learn the ability to suppress artifacts, predicts the mesh using the rendered voxel image as input, and re-expresses the voxel as a mesh.

Acknowledgement

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2023-RS-2023-00258639) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation)

References

[1] Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114
DOI: https://doi.org/10.48550/arXiv.1312.6114

[2] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
DOI: https://doi.org/10.1109/CVPR.2016.90

[3] Chen, W., Ling, H., Gao, J., Smith, E., Lehtinen, J., Jacobson, A., & Fidler, S. (2019). Learning to predict 3d objects with an interpolation-based differentiable renderer. Advances in neural information processing systems, 32.
DOI: https://doi.org/10.48550/arXiv.1908.01210

[4] Han, X. F., Laga, H., & Bennamoun, M. (2019). Image-based 3D object reconstruction: State-of-the-art and trends in the deep learning era. IEEE transactions on pattern analysis and machine intelligence, 43(5), 1578-1604.
DOI: https://doi.org/10.1109/TPAMI.2019.2954885

[5] Smith, E. J., & Meger, D. (2017, October). Improved adversarial systems for 3d object generation and reconstruction. In Conference on Robot Learning (pp. 87-96). PMLR.
DOI: https://doi.org/10.48550/arXiv.1707.09557

[6] Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2017). Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196.
DOI: https://doi.org/10.48550/arXiv.1710.10196

[7] Arjovsky, M., Chintala, S., & Bottou, L. (2017, July). Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). PMLR.
DOI: https://doi.org/10.48550/arXiv.1701.07875

[8] Awiszus, M., Schubert, F., & Rosenhahn, B. (2021, August). World-gan: a generative model for minecraft worlds. In 2021 IEEE Conference on Games (CoG) (pp. 1-8). IEEE.
DOI: https://doi.org/10.1109/CoG52621.2021.9619133

[9] Li, Z.; Hoiem, D. Learning without forgetting. IEEE transactions on pattern analysis and machine intelligence 2017, 40(12), 2935-2947.
DOI: https://doi.org/10.1109/TPAMI.2017.2773081

[10] Gadelha, M.; Maji, S.; Wang, R. 3D shape induction from 2D views of multiple objects. In: Proceedings of the International Conference on 3D Vision, Qingdao, China, 10-12 October 2017. pp. 402‒411.
DOI: https://doi.org/10.1109/3DV.2017.00053