# 3D Generative Adversarial Modeling

Hongzhi Liu

July 12, 2018

## Abstract

*We believe a good generative model should be able to synthesize 3D objects that are both highly varied and realistic then can makes a 3D generative model of object shapes appealing. Specifically, for 3D objects to have variations, a generative model should be able to go beyond memorizing and recombining parts or pieces from a pre-defined repository to produce novel shapes. Today, I read a paper written by Jiajun Wu who is from MIT. The team propose a novel framework, namely 3D Generative Adversarial Network (3D-GAN), which generates 3D objects from a probabilistic space by leveraging recent advances in volumetric convolutional networks and generative adversarial nets. Experiments demonstrate that their method generates high-quality 3D objects and the unsupervisedly learned features achieve impressive performance on 3D object recognition, comparable with those of supervised learning methods.*

## 1. Overview of the 3D-GAN

In the past decades, researchers have made impressive progress on 3D object modeling and synthesis [6], mostly based on meshes or skeletons. Many of these traditional methods synthesize new objects by borrowing parts from objects in existing CAD model libraries. Therefore, the synthesized objects look realistic, but not conceptually novel.

Recently, with the advances in deep representation learning and the introduction of large 3D CAD datasets like ShapeNet [1], there have been some inspiring attempts in learning deep object representations based on voxelized objects [2].

In this paper, Jiajun Wu and his team demonstrate that modeling volumetric objects in a general-adversarial manner could be a promising solution to generate objects that are both novel and realistic [7]. Different from traditional heuristic criteria, generative-adversarial modeling introduces an adversarial discriminator to classify whether an object is synthesized or real which could be a particularly favorable framework for 3D object modeling.

## 2. Models of 3D-GAN

Wu's team discuss how they build framework, 3D Generative Adversarial Network (3D-GAN), by leveraging previous advances on volumetric convolutional networks and generative adversarial nets. Then show how to train a variational autoencoder simultaneously so that our framework can capture a mapping from a 2D image to a 3D object.

### 2.1. 3D Generative Adversarial Network

As proposed in [3], the Generative Adversarial Network (GAN) consists of a generator and a discriminator, where the discriminator tries to classify real objects and objects synthesized by the generator, and the generator attempts to confuse the discriminator. In Wu' 3D Generative Adversarial Network (3D-GAN), the generator G maps a 200-

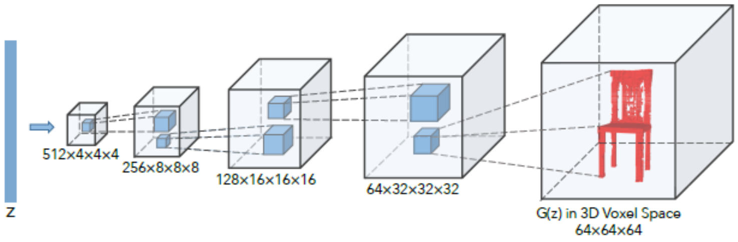

Figure 1. The generator in 3D-GAN. The discriminator mostly mirrors the generator.

Table 1. Number of dataset images for each tag.

| Method | Bed | Bookcase | Chair | Desk | Sofa | Table | Mean |
|---|---|---|---|---|---|---|---|
| AlexNet-fc8 [2] | 29.5 | 17.3 | 20.4 | 19.7 | 38.8 | 16.0 | 23.6 |
| AlexNet-conv4 [2] | 38.2 | 26.6 | 31.4 | 26.6 | 69.3 | 19.1 | 35.2 |
| 3D-VAE-GAN | 49.1 | 31.9 | 42.6 | 34.8 | 79.8 | 33.1 | 45.2 |

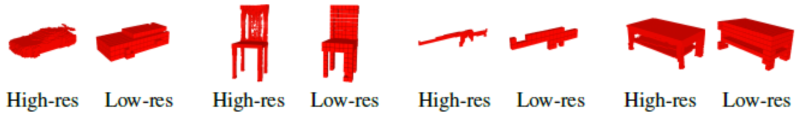

Figure 2. We present each object at high resolution (64 × 64 × 64) on the left and at low resolution (down-sampled to 16 × 16 × 16) on the right. While humans can perceive object structure at a relatively low resolution, fine details and variations only appear in high-res objects.

dimensional latent vector z, randomly sampled from a probabilistic latent space, to a 64 × 64 × 64 cube, representing an object G(z) in 3D voxel space. The discriminator $D$ outputs a confidence value $D(x)$ of whether a 3D object input $x$ is real or synthetic. They use binary cross entropy as the classification loss, and present our overall adversarial loss function as Eq. 1:

$$L_{3D-GAN}(D,G) = \log D(x) + \log(1 - D(G(z))). \quad (1)$$

where $x$ is a real object in a $64 \times 64 \times 64$ space, and $z$ is a randomly sampled noise vector from a distribution $p(z)$. In this work, each dimension of $z$ is an i.i.d. uniform distribution over [0,1].

As shown in Fig. 1, the generator consists of five volumetric fully convolutional layers of kernel sizes $4 \times 4 \times 4$ and strides 2, with batch normalization and ReLU layers added in between and a Sigmoid layer at the end. The discriminator basically mirrors the generator, except that it uses Leaky ReLU instead of ReLU layers. There are no pooling or linear layers in their network. More details can be found in the supplementary material.

## 2.2. 3D-VAE-GAN

Wu and his team introduce 3D-VAE-GAN as an extension to 3D-GAN. They add an additional image encoder $E$, which takes a 2D image $x$ as input and outputs the latent representation vector $z$. This is inspired by VAE-GAN proposed by [4], which combines VAE and GAN by sharing the decoder of VAE with the generator of GAN.

The 3D-VAE-GAN therefore consists of three components: an image encoder E, a decoder (the generator G in 3D-GAN), and a discriminator D. Formally, these loss functions write as Eq. 2:

$$\begin{aligned} L = L_{3D-GAN} + \alpha_1 L_{KL} + \alpha_2 L_{recon}, \\ L_{KL} = D_{KL}[q(z|y)||p(z)], \\ L_{recon} = ||G(E(y)) - x||_2. \end{aligned} \quad (2)$$

where $\alpha_1$ and $\alpha_2$ are weights of the KL divergence loss and the reconstruction loss. And $x$ is a 3D shape from the training set, $y$ is its corresponding 2D image, and $q(z|y)$ is the variational distribution of the latent representation $z$.

## 3. Evaluation of the Model

Wu and his team evaluate their framework from various aspects. THey first show qualitative results of generated 3D objects. And then evaluate the unsupervisedly learned representation from the discriminator by using them as features for 3D object classification. They show both qualitative and quantitative results on the popular benchmark ModelNet [7]. Further, they evaluate the 3D-VAE-GAN on 3D object reconstruction from a single image, and show both qualitative and quantitative results on the IKEA dataset [5].

Compared with previous works, our 3D-GAN can synthesize high-resolution 3D objects with detailed geometries. Fig. 2 shows both high-res voxels and down-sampled low-res voxels for comparison. Note that it is relatively easy to synthesize a low-res object, but is much harder to obtain a high-res one due to the rapid growth of 3D space.

They show the results in Tab. 1, with performance of a single 3D-VAE-GAN on all six categories, as well as the results of six 3D-VAE-GANs separately. In all categories, our model consistently outperforms previous state-of-the-art in voxel-level prediction and other baseline methods.

## References

[1] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, and H. Su. ShapeNet: An information-rich 3D model repository. *Computer Science*, 2015. 1

[2] R. Girdhar, D. F. Fouhey, M. Rodriguez, and A. Gupta. Learning a predictable and generative vector representation for objects. In *ECCV*, 2016. 1, 2

[3] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *NIPS*, 2014. 1

[4] A. B. L. Larsen, H. Larochelle, and O. Winther. Autoencoding beyond pixels using a learned similarity metric. In *ICML*, 2016. 2

[5] J. J. Lim, H. Pirsiavash, and A. Torralba. Parsing IKEA objects: Fine pose estimation. 2

[6] J. W. Tangelder and R. C. Veltkamp. A survey of content based 3D shape retrieval methods. *MT&A*, 2008. 1

[7] J. Wu, C. Zhang, T. Xue, W. T. Freeman, and J. B. Tenenbaum. Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling. In *NIPS*, 2016. 1, 2