3D Bounding Box Generative Adversarial Nets

Ping Kuang[1] and Haoshuang Wang[1]
[1]*University Of Electronic Science And Technology Of China,Chengdu,China*

Keywords: Generative Adversarial Networks, Deep Learning, 3D-GAN.

Abstract: Recently, Generative Adversarial Networks (GANs) gradually applied to the generation of 3D objects and has achieved remarkable success, but at the same time, it also faces some problems, such as the training instability, low-quality samples and mode collapse. We propose a novel framework, namely 3D Bounding Box Generative Adversarial Network (3D-BBGAN), which can reduce the probability space of generation by adding conditional information. According this way, we can get 3D objects with more detailed geometries.

1 INTRODUCTION

The Generative Adversarial Network (GAN) (Goodfellow, 2014) have achieved a great success in generation of pictures. The some flaws in GAN's initial stage are generally remedied by the variations of GAN. To solve the out-sync problem of generator and discriminator network, (Arjovsky, 2017) proposes Wasserstein GAN (WGAN) of clipping the weights of the critic to restrict within a fixed interval [-c, c]. Furthermore, (Gulrajani, 2017) provide the WGAN with gradient penalty (WGAN-GP) to avoid the possible pathological behavior of WGAN. Otherwise, the initial GAN have no any prior information, which leads that we couldn't constrain the generated results. The Conditional Generative Adversarial Net (CGAN) (Mirza, 2014) creatively adds the prior information into network by adding conditions to both the generator and discriminator network.

Recently, the appearance of 3D-GAN (Wu, 2016) indicates that GAN begins to apply to the generation of 3D objects. Compared with the 2D field, the feature information of high-dimensional object is more complex. It is easy to fall into the curse of dimensionality. And it is a bit challenging for the network structure and hardware performance to learn more accurately the characteristic information of high-dimensional objects. At present, the accuracy of improving the 3D model is mainly to improve the complexity of network architecture and reduce the distribution space of 3D objects. (Smith, 2017) (3D-IWGAN) reduces the data space of the original 3D-GAN from 64*64*64 to 32*32*32, at the same time introduce into the WGAN-GP to improve the stability of training.

Base on the methods known as 3D-IWGAN and CGAN, here we propose a new idea to capture these wider and more complicated distributions, which attempt to introduce the prior information into the network architecture. Unlike the CGAN directly feeding same extra information into the both the discriminator and generator as additional input layer, while feeding the generator network the bounding box information of 3D samples which can be represented as double three-dimensional coordinates, we are also feeding the discriminator the mask information corresponding to bounding box. We call our resulting shape model the 3D Bounding Box Generative Adversarial Network (3DBBGAN).

We can demonstrate that our 3D-BBGAN have following advantages:

- Compared with 3D-IWGAN, we can get a more stable and faster procedure of convergence testing on ModelNet10 Dataset (Wu, 2015)
- By adjusting the input coordinate information, we can guide the size of generated object to some extent.

Kuang, P. and Wang, H.
3D Bounding Box Generative Adversarial Nets.
DOI: 10.5220/0008096801170121
In *Proceedings of the International Conference on Advances in Computer Technology, Information Science and Communications (CTISC 2019)*, pages 117-121
ISBN: 978-989-758-357-5

2 RELATED WORKS

2.1 Modelling and Synthesizing 3D Shapes

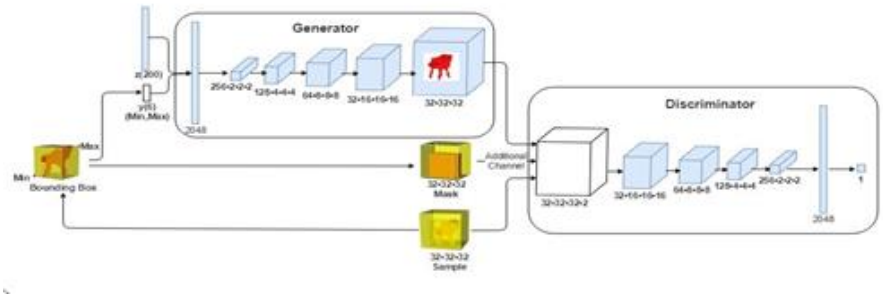

Figure 1: The architecture of 3D-BBGAN.

The generation and synthesizing of 3D object models is a hot spot in computer graphics and vision community. In the past few decades, the method of 3D object modeling and synthesis mainly realized by the combination of primitives or simpler geometric shapes, which exist in CAD model libraries (Wu,2015),(Chang,2015), there is a lot of the relevant literature (Tangelder,2004), (Chaudhuri,2011),(Carison,1982) or through the way of point cloud reconstruction(Alexa,2003) The synthesized objects using these methods look realistic, but it is time-consuming and high-cost. As 3D-GAN (Wu, 2016) is applied to the reconstruction and generation of 3D objects, some researchers have realized the potential of GAN in 3D filed. 3D-IWGAN combines the 3D-GAN and WGAN-GP to improve the stability of training and the effect of synthesized object, (Wang,2017) named as 3d-ED-GAN combines a 3D Encoder-Decoder GAN and a Long-term Recurrent Convolutional Network (Donahue,2015)(LRCN) to construction from broken models result in complete and high-resolution 3D objects, and (Yang,2017) proposes 3D-RecGAN approach to realize the construction from a single 2.5D depth view in a complete 3D objects. Get inspiration from CGAN, our 3D-BBGAN introduces the prior knowledge into the 3D-GAN. We will testify that our model could improve the efficiency of training.

3 PROPOSED MODELS

In this section we introduce our 3D-BBGAN architecture for 3D objection generation. We first get an introduction toward CGAN and WGAN-GP. Then we display our model and explain how we achieve control over the generated objects.

3.1 Conditional Adversarial Nets(CGAN)

Based the (Goodfellow, 2014) CGAN creatively attempt to condition both the generator and discriminator through same extra information y. The information may be meaningless to humans, but the generator and discriminator will learn automatically the meaning of y. To generate verisimilar results, the generator will build a mapping function from real samples, at the same time the discriminator try to distinguish the fake results, as if they are playing the two-player min-max game with value function V (G, D):

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}[\log D(x|y)] + E_{z\sim P_z(z)}[\log(1 - D(G(z|y)))] \quad (1)$$

The generator model G wants to captures the data distribution, and the discriminative model D tries to estimates the probability that a sample came from the training data. Where z represents a prior noise distribution.

3.2 WGAN with Gradient Penalty (WGAN-GP)

(Arjovsky, 2017) (WGAN) argues that cross entropy is not appropriate to measure the distance of distributions with disjoint parts. Instead, the wassertein distance is proposed to measure the distance between the generated data and real data distribution, theoretically solving the problem of the training. In order to satisfy the lipschitz continuity that wassertein distance needs, WGAN clips the weights of the critic to lie within a fixed interval

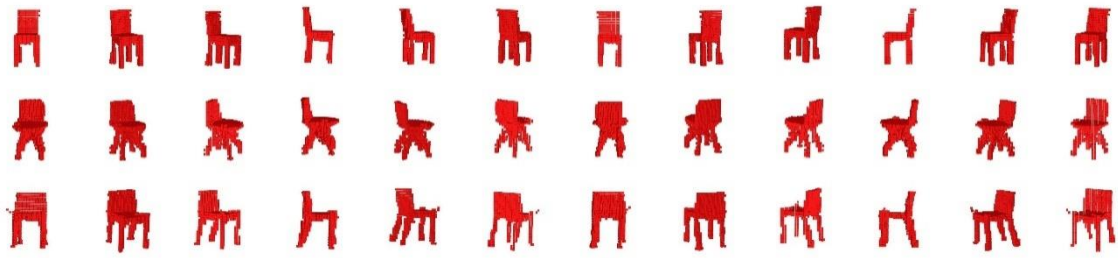

Figure 2: Objects generated by the 3D-BBGAN system trained on the ModelNet10 chair class in 12 orientations. Each row of images come from a same generated object.

[*c,c*]. Determining the size of c is a tricky business. If too large, it may cause gradient explosion. On the contrary, it may cause gradient to disappear.(Gulrajani,2017) proposes a more appropriate method to enforcing the Lipschitz constraint. It directly constrains the gradient norm of the critic's output with respect to its input to realize a soft version of constraint with a penalty on the gradient norm for random samples $\hat{x} \sim P_{\hat{x}}$, where randomly sampled from generator Pg and real data distribution Pr. This results in the following loss function:

$$L = {}_{\tilde{x}\sim P_g}^{E}[D(\tilde{x})] - {}_{x\sim P_r}^{E}[D(x)] + \lambda_{\hat{x}\sim P_{\hat{x}}}^{E}[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2] \quad (2)$$

3.3 Our 3D Bounding Box Generative Adversarial Net (3D-BBGAN)

Inspired by (Mirza, 2014), our 3D-BBGAN architecture introduces the prior information into original network architecture of 3D-IWGAN (Smith, 2017). As shown in 1, every real input is a 32×32×32 voxel space, but the sample could be encircled at a more small area called Bounding Box. The Bounding Box could be represented by double coordinates (Min, Max) in a 3D space, which is inputted into the generator *G* as additional information y. Meanwhile, the every Bounding Box corresponds to a single Mask, inside of the Bounding Box are *1.0* and outside are *0.0*, which is fed into the discriminator *D* as additional input layer. As for network structure, the generator *G* maps a 206-dimensional latent vector (*z, y*), randomly sampled from a probability latent space, to a 32×32×32 cube, representing an object in 3D voxel space. The discriminator *D* is fed the samples and the corresponding mask information, outputing a confidence value of whether an object model is real or generated.

Following (Gulrajani, 2017), the loss function is as follows:

$$L = {}_{G(z,y)\sim P_g}^{E}[D(G(z,y)|Mask(y))] - {}_{x\sim P_r}^{E}[D(x|Mask(y))] + \lambda_{\hat{x}\sim P_{\hat{x}}}^{E}[(\|\nabla_{\hat{x}} D(\hat{x}|Mask(y))\|_2 - 1)^2] \quad (3)$$

where the Mask maps y. The $P_{\hat{x}}$ is the mixed distribution of mixing randomly interpolation of generator distribution P_g and data distribution P_r.

4 EXPERIMENT

In this section, we evaluate our framework on ModelNet10 Dataset (Wu, 2015). We first list 3D-GAN, 3D-IWGAN and our generated results separately. Then we introduce our training details. Finally we show the frame how to control the size of generate object.

4.1 3D Object Generation

For this experiment, we first train our 3D-BBGAN for the highest complexity of chair and table class of ModelNET10 dataset. Then on the all classes of ModlNET10. We trained all experiments on 1080Ti GPU. For generation, we get generated object from the last tanh activation function of generator network,

and render the voxel of which the value exceed the threshold 0.3.

Figure2 are generated by chair class of ModelNet10, which shows that our 3DBBGAN system has ability to generalize the object model of high quality from the complex sample space. Figure 3 shows separately the generated samples of 3DGAN,3D-IWGAN and our 3D-BBGAN. These samples are uniformly sampled and selected intact results. The source code we tested was from the github provided by (Gulrajani, 2017). Figure 4 are generated by training on the entire class of ModelNet10 dataset, which has bathtub, chair, bed, desk, dresser, monitor, table, toilet, nightstand and sofa class. Because our mask information successfully limited

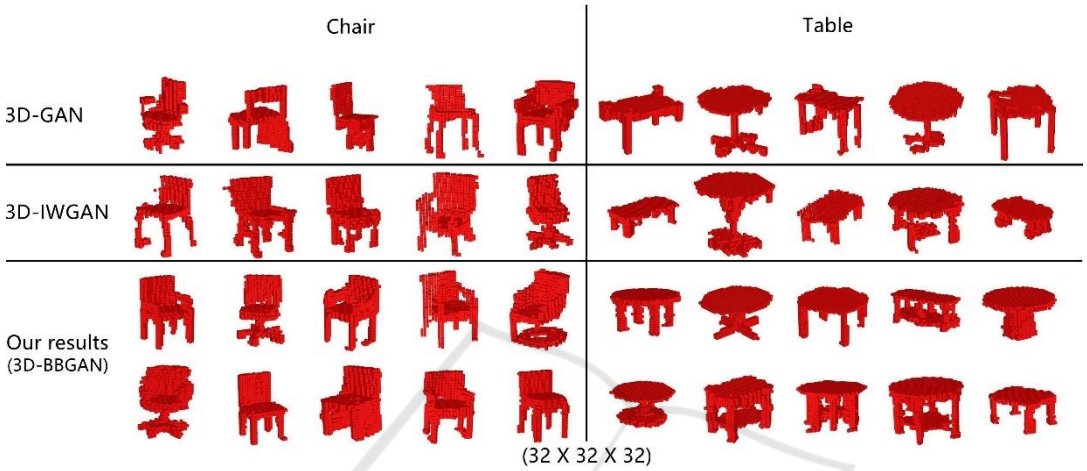

Figure 3: Objects generated by the 3D-GAN ,3D-IWGAN and our 3D-BBGAN models, trained separately on the single chair or table class of ModelNet10 dataset. For comparison, the results are uniformly sampled to represent the generalization ability of the model as much as possible.

the space of data distribution, the model convergence speed was obviously improved.

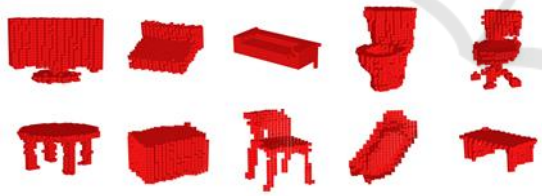

Figure 4: Objects generated by our 3D-BBGAN model, trained on the all ModelNet10 dataset. Each shape above is derived from a different class as order: monitor, bed, sofa, toilet, chair, table, dresser, nightstand, bathtub, desk. The orientations have been altered for optimal viewing.

During the experiment, we found that the 3D-GAN have come across the problem of mode collapse. It's generated types are more less than others and always incomplete.

While comparing the synthesized objects of 3D-IWGAN and 3D-BBGAN generated by same latent vector z, our results can get more detailed geometries and more complete styles. But we have to admitted that there is no effective quantifiable indicators to measure the effect of synthesized objects.

4.2 Mask Information

Here we will show the effect of our model controlling the size of 3D object. We provide the generator with the length information of three sides of the bounding box to guiding the generation of 3D object. The length information can be converted into two coordinates. As Figure 5 shows, the 3D object will be generated in the corresponding bounding box. In fact, the specific generation effect of 3D object is not always good. According our research, the generation effect under bounding box.

Training samples should contain as many 3D object models of various sizes as possible.

- While the mask information provided exist in the training samples, the generated effect tends to be better.
- Under extreme conditions like very short side length, the mask information will lose its guiding effect.

In a word, our 3D-BBGAN model has ability to guide the size of the generate object by limiting the

generated area of the object through two three-dimensional coordinates information.

5 CONCLUSIONS

In this work, we proposed 3D-BBGAN for 3D object generation. We demonstrated that our models are able to generate novel 3D objects with more detailed geometries. Through the addition of the conditional information to the generator and the discriminator in the training process, we have realized the effective limitation of the probability space of the generated object. We effectively limit the probability space of generation object to shorten the training time and improve the generation effect of 3D objects. And by adjusting the information we add to the generator, we can direct the size of the generated object. Next, we will try to add the appropriate conditional information to guide the type of the generated object.

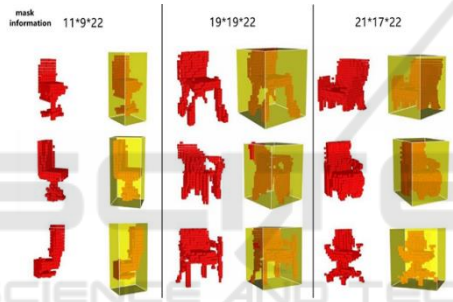

Figure 5: Objects generated by the 3D-BBGAN system trained on the ModelNet10 chair class. Here we list the generate results of the three bounding box information. The graph on the left of each column represents the generated 3D Object, the right of each column shows the render result with the correspond bounding box information.

ACKNOWLEDGEMENTS

This work is supported by Sichuan Science and Technology Program (2015GZ0358, 2016GFW0077, 2016GFW0116, 2018GZ0889) and Chengdu Science and Technology Program (2018-YF05-01138-GX).

REFERENCES

Alexa, M., Behr, J., Cohen-Or, D., Fleishman, S., Levin, D., Silva, C.T.: Computing andrendering point set surfaces. IEEE Transactions on visualization and computer graphics 9(1), 3–15 (2003).

Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein gan. arXiv preprint arXiv:1701.07875 (2017).

Carlson, W.E.: An algorithm and data structure for 3d object synthesis using surface patchintersections. ACM SIGGRAPH Computer Graphics 16(3), 255–263 (1982).

Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S.,Savva, M., Song, S., Su, H., et al.: Shapenet: An information-rich 3d model repository. arXiv preprint arXiv:1512.03012 (2015).

Chaudhuri, S., Kalogerakis, E., Guibas, L., Koltun, V.: Probabilistic reasoning for assemblybased 3d modeling. In: ACM Transactions on Graphics (TOG). vol. 30, p. 35. ACM (2011).

Donahue, J., Anne Hendricks, L., Guadarrama, S., Rohrbach, M., Venugopalan, S., Saenko,K., Darrell, T.: Long-term recurrent convolutional networks for visual recognition and description. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2625–2634 (2015).

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville,A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training ofwasserstein gans. In: Advances in Neural Information Processing Systems. pp. 5769–5779 (2017).

Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014).

Smith, E., Meger, D.: Improved adversarial systems for 3d object generation and reconstruction. arXiv preprint arXiv:1707.09557 (2017).

Tangelder, J.W., Veltkamp, R.C.: A survey of content based 3d shape retrieval methods. In:Shape Modeling Applications, 2004. Proceedings. pp. 145–156. IEEE (2004).

Wang, W., Huang, Q., You, S., Yang, C., Neumann, U.: Shape inpainting using 3d generativeadversarial network and recurrent convolutional networks. arXiv preprint arXiv:1711.06375 (2017).

Wu, J., Zhang, C., Xue, T., Freeman, B., Tenenbaum, J.: Learning a probabilistic latent spaceof object shapes via 3d generative-adversarial modeling. In: Advances in Neural Information Processing Systems. pp. 82–90 (2016).

Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3d shapenets: A deeprepresentation for volumetric shapes. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1912–1920 (2015).

Yang, B., Wen, H., Wang, S., Clark, R., Markham, A., Trigoni, N.: 3d object reconstructionfrom a single depth view with adversarial learning. arXiv preprint arXiv:1708.07969 (2017).