universitätbonn

Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling

Presenter:

Nagesh TR

374559

Mentor:

Martin Garbade

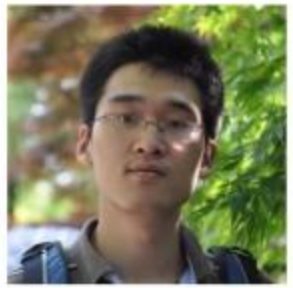
universitätbonn

Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling

NIPS 2016


Jiajun Wu*


Chengkai Zhang*


Tianfan Xue


Bill Freeman


Josh Tenenbaum

MIT CSAIL Google Research

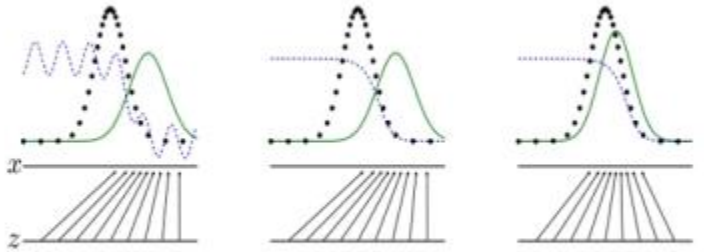
* indicates equal contribution

universitätbonn

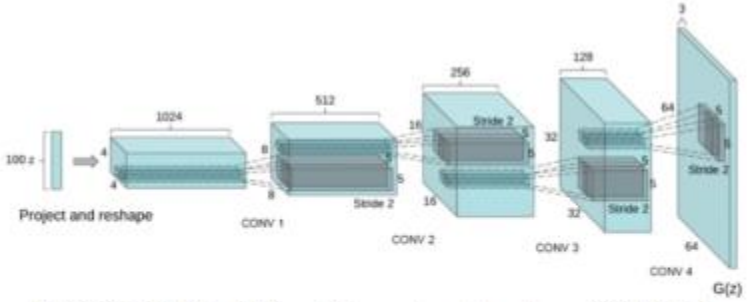
INTRODUCTION

- Paper addresses the problem of 3D object generation.
- A novel framework is proposed, 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.

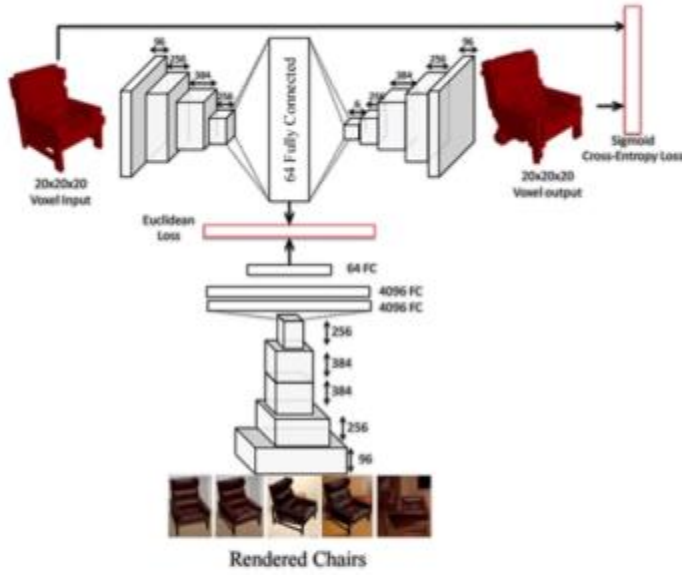
Related Work



GAN [Goodfellow et al., 2014]

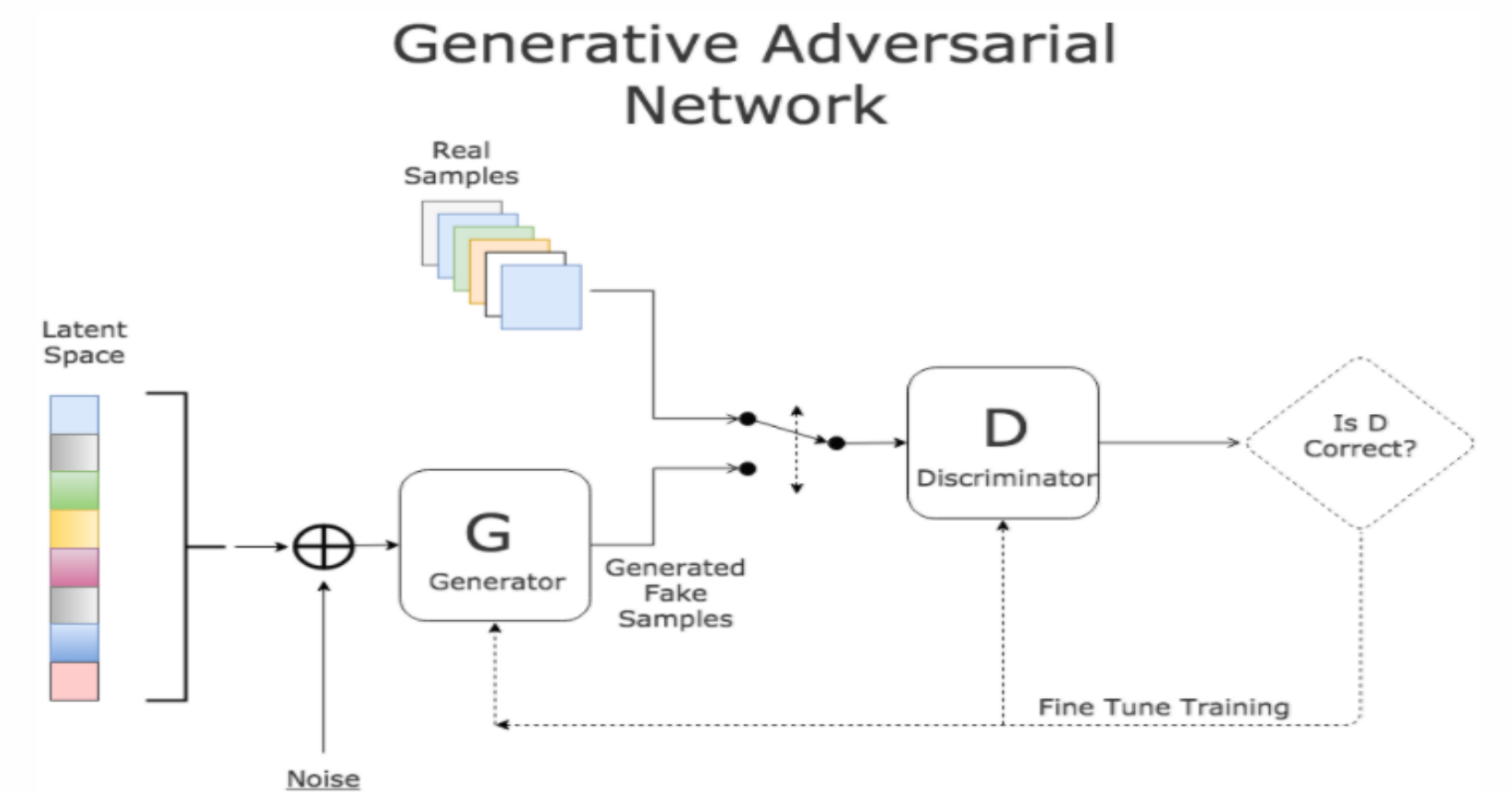


DCGAN [Radford et al., 2016]



T-L Network
[Girdhar et al., 2016]

Related Work 1 - GAN



Architecture of a generative adversarial network. (Image source: https://www.kdnuggets.com/2017/01/generative-adversarial-networks-hot-topic-machine-learning.html

universitätbonn

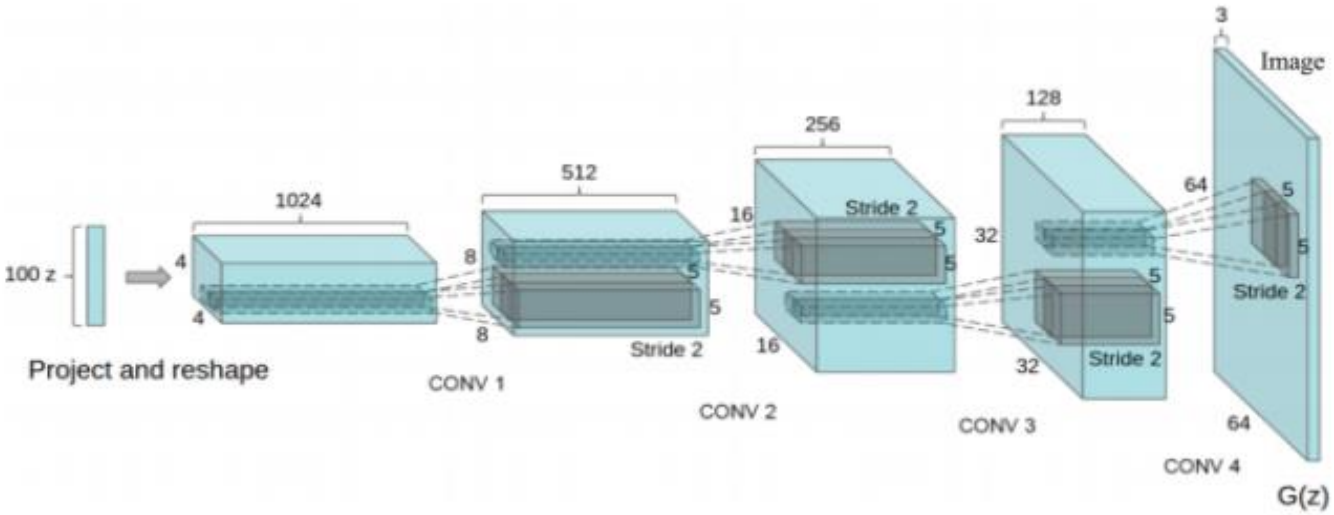
Related Work 1 - GAN

- Generator: Generate fake samples, tries to fool the Discriminator
- Discriminator: Tries to distinguish between real and fake samples
- Train them against each other
- Repeat the process

Related Work 2 - DCGAN

Idea : Deep convolution GAN works in opposite direction of CNN .

- CNN transforms an image to a class label (list of probabilities)
- DCGAN generates an image from random parameters
- Discriminator mirrors the generator



Related Work 3 – T-L Network

Paper : Learning a Predictable and Generative Vector Representation for Objects

Question : What is a good vector representation of an object

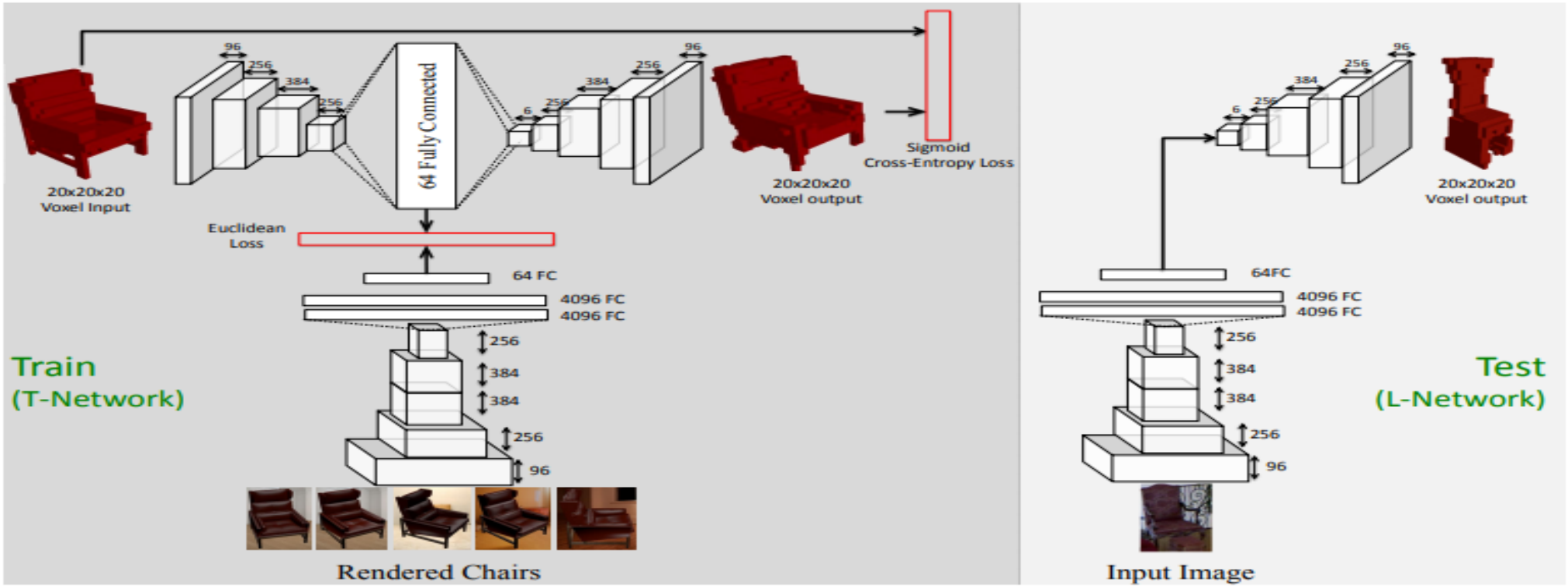
Answer: It must be generative in 3D: We should be able to reconstruct objects in 3D from it.
It must be predictable from 2D: We should be able to easily infer this representation from images.

Approach : T-L Network

An Auto encoder that ensures the representation is generative
CNN that ensures the representation is predictable

The T and L refer to the architecture in the training and testing phase

Related Work 3 – T-L Network



universitätbonn

NOVELTY -1

Modeling and synthesizing 3D shapes

- Carlson, 1982, Tangelder and Veltkamp, 2008, Van Kaick et al., 2011, Blanz and Vetter, 1999, Kalogerakis et al., 2012, Chaudhuri et al., 2011, Xue et al., 2012, Kar et al., 2015, Bansal et al., 2016, Wu et al., 2016 have made inspiring attempts to design or learn 3D object representations
- Many of these shape synthesis algorithms are nonparametric and they synthesize new objects by retrieving and combining shapes and parts from a database
- Huang et al. [2015] explored generating 3D shapes with pre-trained templates and producing both object structure and surface geometry
- 3D GAN framework synthesizes objects without explicitly borrow parts from a repository, and requires no supervision during training

universitätbonn

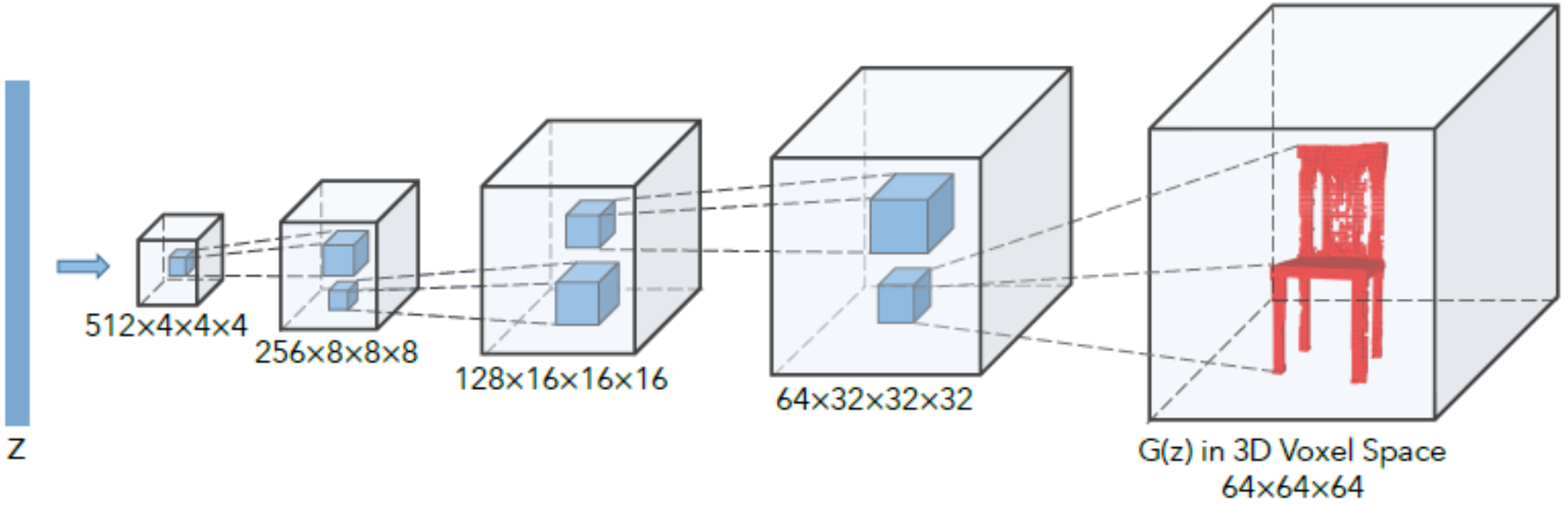
NOVELTY -2

Deep learning for 3D data

- Girdhar et al. [2016], Sharma et al. [2016] explored autoencoder-based networks for learning voxel-based object representations
- Many of these networks can be used for 3D shape classification , 3D shape retrieval, and single image 3D reconstruction mostly with full supervision.
- 3D GAN requires no supervision for training, is able to generate objects from a probabilistic space, and comes with a rich discriminative 3D shape representation.

universitätbonn

3D GAN - Network Structure



In our 3D Generative Adversarial Network (3D-GAN), the generator G maps a 200-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

universitätbonn

3D GAN – Explanation 1

- GAN - Generator + Discriminator
- Generator : G : z -> G(z)
  where z : latent vector (200 dimensions)
  G(z) : 3D object in 3D voxel space (64*64*64 cube)
- Discriminator D : Output a confidence value of whether an input value is real or synthetic
- Overall adversarial loss: $L3D\text{-}GAN = \log D(x) + \log(1 - D(G(z)))$

3D GAN – Explanation 2

Network Structure :

- 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.
- The discriminator basically mirrors the generator.

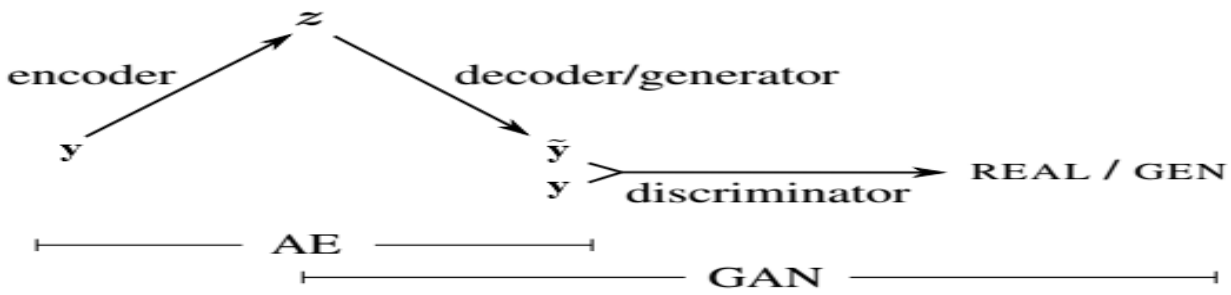
Training :

- A straightforward training procedure is to update both the generator and the discriminator in every batch. However, the discriminator usually learns much faster than the generator
- Adaptive training strategy: For each batch, the discriminator only gets updated if its accuracy in the last batch is not higher than 80%

3D VAE GAN

• 3D GAN addressed how to generate 3D objects by sampling a latent vector z and mapping it to the object space.

• In practice, it would also be helpful to infer these latent vectors from observations. For example, if there exists a mapping from a 2D image to the latent representation, we can then recover the 3D object corresponding to that 2D image



3D VAE -GAN – Loss Function

• Loss function consists of three parts: an object reconstruction loss Lrecon, a cross entropy loss L3D-GAN for 3D-GAN, and a KL divergence loss LKL to restrict the distribution of the output of the encoder.

• Formally, these loss functions write as

$$L = L3D\text{-}GAN + \alpha 1 LKL + \alpha 2 Lrecon$$

where $\alpha 1$ and $\alpha 2$ are weights of the KL divergence loss and the reconstruction loss.

$$L3D\text{-}GAN = \log D(x) + \log(1 - D(G(z)))$$
$$LKL = DKL(q(z|y) \,\|\, p(z))$$
$$Lrecon = \|G(E(y)) - x\|2$$

where 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. The KL-divergence pushes this variational distribution towards to the prior distribution p(z), so that the generator can sample the latent representation z from the same distribution p(z)

universitätbonn

Model Evaluation

Paper shows qualitative results of generated 3D objects.

Secondly ,evaluates the unsupervisedly learned representation from the discriminator by using them as features for 3D object classification
DataSet : ModelNet

Further evaluate 3D-VAE-GAN on 3D object reconstruction from a single image
DataSet : IKEA

Model Evaluation-Dataset

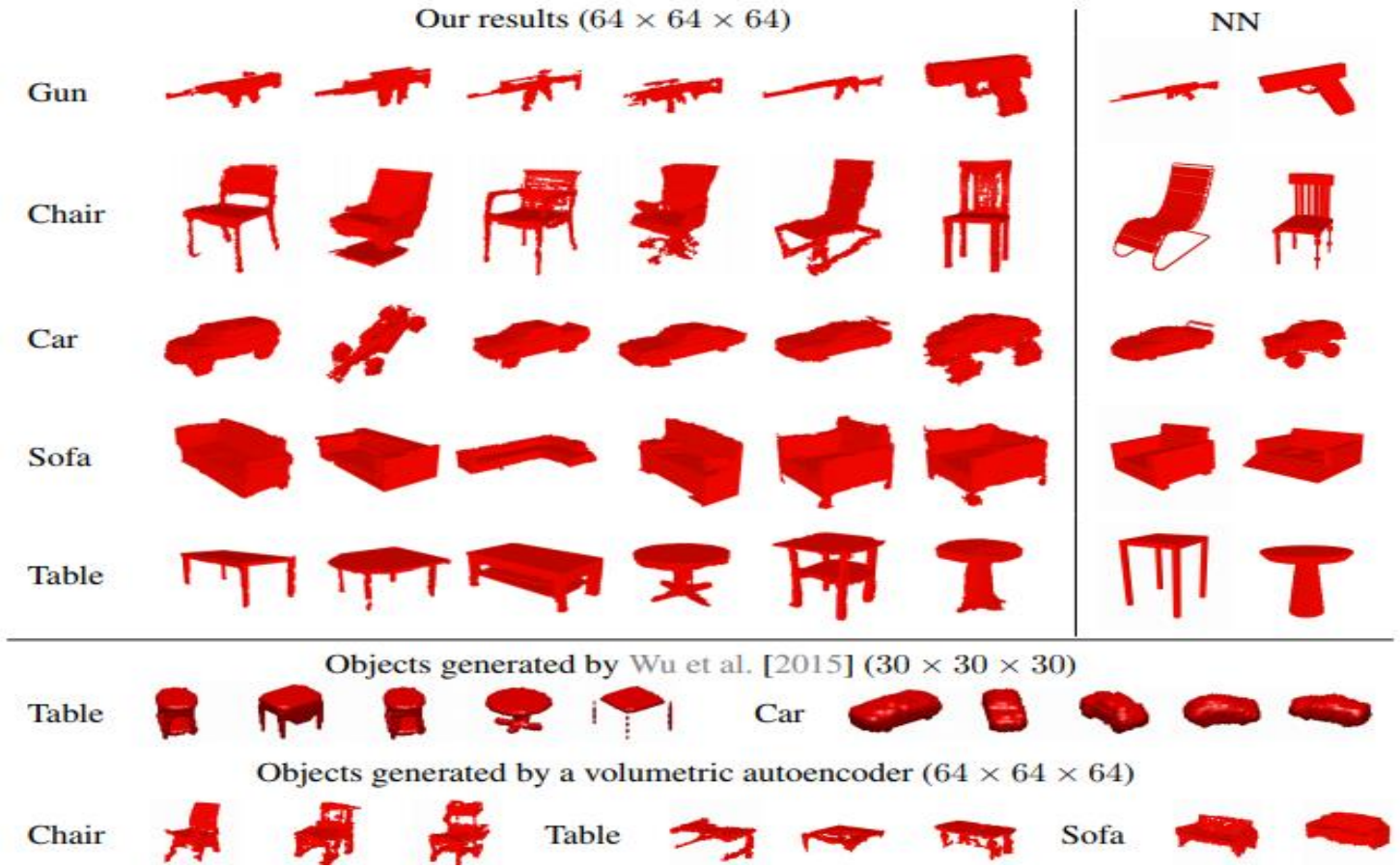
•Princeton Modelnet dataset which features various mesh models of different objects (e.g airplane, bookshelf).

• It comes in a smaller 10 classes dataset (~2.5k examples) and a 40 class set (~10k examples) already split into train/test. Categories of 40 classes are airplane, bathtub, bed, bench, bookshelf, bottle, bowl, car, chair, cone, cup, curtain, desk, door, dresser, flower pot, glass box, guitar, keyboard, lamp, laptop, mantel, monitor, night stand, person, piano, plant, radio, range hood, sink, sofa, stairs, stool, table, tent, toilet, tv stand, vase, wardrobe, xbox.

• IKEA dataset contains about 759 images and 219 3D models. All 759 images are annotated using available models (about 90 different models)

Model Evaluation -1

3D Object Generation :



universitätbonn

Model Evaluation -2

3D Object Classification:

A typical way of evaluating representations learned without supervision is to use them as features for classification

Dataset : Train a single 3D-GAN on the seven major object categories (chairs, sofas, tables, boats, airplanes, rifles, and cars) of ShapeNet and use ModelNet for testing

Supervision	Pretraining	Method	Classification (Accuracy)	
			ModelNet40	ModelNet10
Category labels	ImageNet	MVCNN [Su et al., 2015a]	90.1%	-
		MVCNN-MultiRes [Qi et al., 2016]	**91.4%**	-
	None	3D ShapeNets [Wu et al., 2015]	77.3%	83.5%
		DeepPano [Shi et al., 2015]	77.6%	85.5%
		VoxNet [Maturana and Scherer, 2015]	83.0%	92.0%
		ORION [Sedaghat et al., 2016]	-	**93.8%**
Unsupervised	-	SPH [Kazhdan et al., 2003]	68.2%	79.8%
		LFD [Chen et al., 2003]	75.5%	79.9%
		T-L Network [Girdhar et al., 2016]	74.4%	-
		VConv-DAE [Sharma et al., 2016]	75.5%	80.5%
		3D-GAN (ours)	**83.3%**	**91.0%**

Table 1: Classification results on the ModelNet dataset. Our 3D-GAN outperforms other unsupervised learning methods by a large margin, and is comparable to some recent supervised learning frameworks.

Model Evaluation -3

Single Image 3D Reconstruction:

3D-VAE-GAN can perform well on single image 3D reconstruction

Dataset : IKEA
Result : Performance of 3D-VAE-GAN on 6 different categories .

Method	Bed	Bookcase	Chair	Desk	Sofa	Table	Mean
AlexNet-fc8 [Girdhar et al., 2016]	29.5	17.3	20.4	19.7	38.8	16.0	23.6
AlexNet-conv4 [Girdhar et al., 2016]	38.2	26.6	31.4	26.6	69.3	19.1	35.2
T-L Network [Girdhar et al., 2016]	56.3	30.2	32.9	25.8	71.7	23.3	40.0
3D-VAE-GAN (jointly trained)	49.1	31.9	42.6	34.8	**79.8**	33.1	45.2
3D-VAE-GAN (separately trained)	**63.2**	**46.3**	**47.2**	**40.7**	78.8	**42.3**	**53.1**

Table 2: Average precision for voxel prediction on the IKEA dataset.†

Analysis – Generative Representations 1

We look deep into the representations learned by both the generator and the discriminator of 3D-GAN

Visualizing the object vector :

To visualize the semantic meaning of each dimension, we gradually increase its value, and observe how it affects the generated 3D object

Each Column corresponds to one dimension of the object vector, where the red region marks the voxels affected by changing values of that dimension.

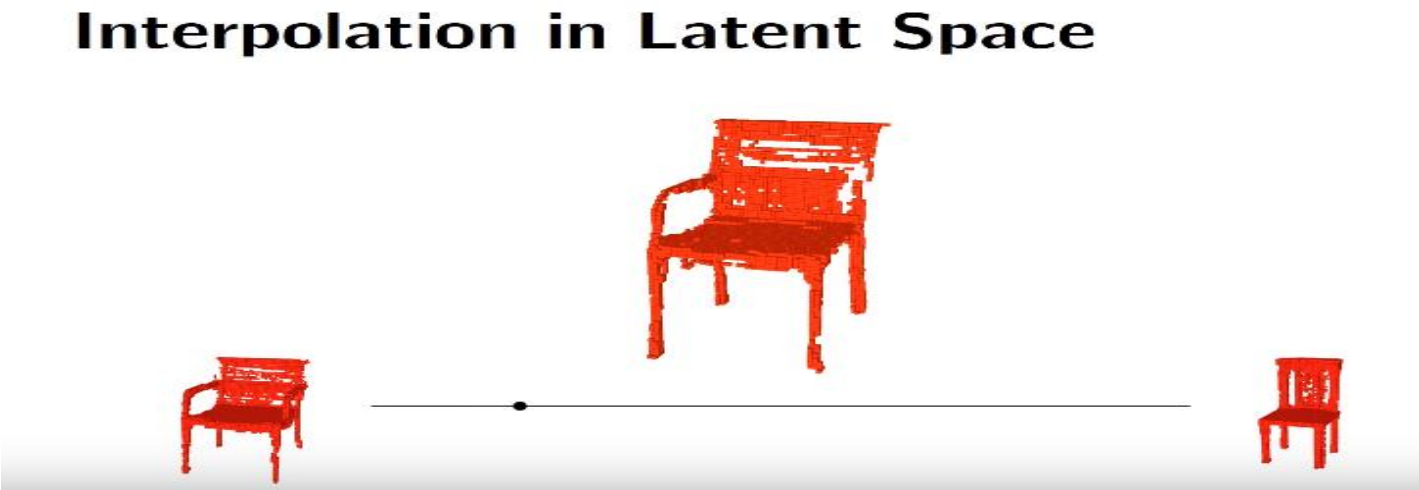
We observe that some dimensions in the object vector carries semantic knowledge of the object, e.g., the thickness or width of surfaces

Figure 5: The effects of individual dimensions of the object vector

Analysis – Generative Representations 2

We look deep into the representations learned by both the generator and the discriminator of 3D-GAN

Interpolation : Interpolations both within and across object categories

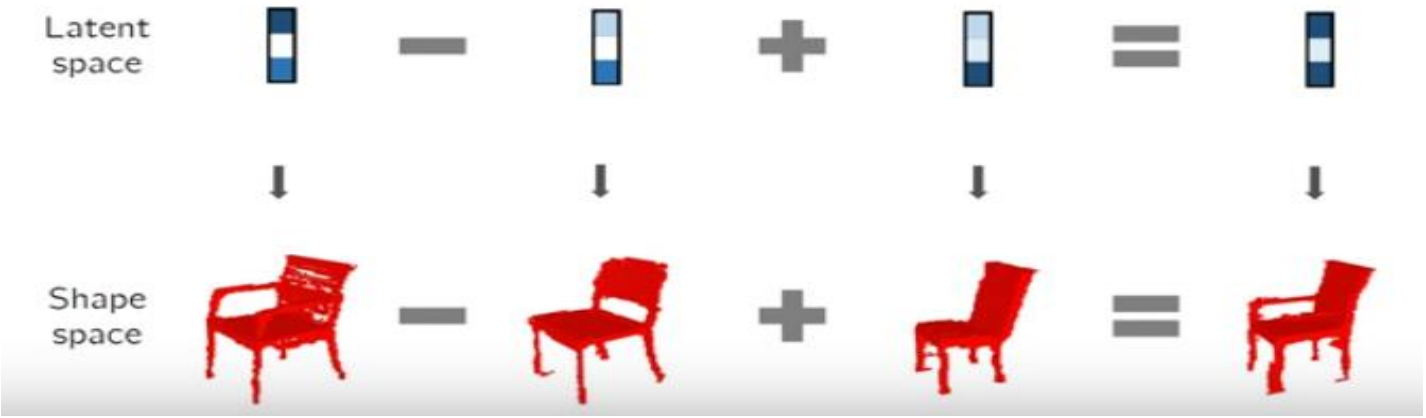


Analysis – Generative Representations 3

We look deep into the representations learned by both the generator and the discriminator of 3D-GAN

Arithmetic : Another way of exploring the learned representations is to show arithmetic in the latent space .

Ex : Shape arithmetic of chairs



Analysis – Discriminative Representations 1

Here we are interested in showing what input objects, and which part of them produce the highest intensity values for each neuron.

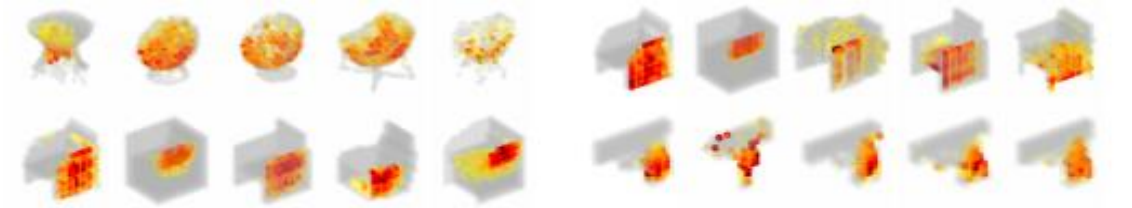


Figure 9: Objects and parts that activate specific neurons in the discriminator. For each neuron, we show five objects that activate it most strongly, with colors representing gradients of activations with respect to input voxels.

There are two main observations:

• First, for a single neuron, the objects producing strongest activations have very similar shapes, showing the neuron is selective in terms of the overall object shape.

• Second, the parts that activate the neuron, shown in red, are consistent across these objects

universitätbonn

Conclusion

• Paper proposed 3D-GAN for 3D object generation, as well as 3D-VAE-GAN for learning an image to 3D model mapping.

• Paper showed that the discriminator in GAN, learned without supervision, can be used as an informative feature representation for 3D objects, achieving impressive performance on shape classification

• Paper also explored the latent space of object vectors, and presented results on object interpolation, shape arithmetic, and neuron visualization

universitätbonn

Thank You ..