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



Jiajun Wu*



Chengkai Zhang*

Tianfan Xue



Bill Freeman



Josh Tenenbaum

NIPS 2016



(* indicates equal contribution)

3D Shape Synthesis

Templated-based model

- Synthesizing realistic shapes
- Requiring a large shape repository
- Recombining parts and pieces



3D Shape Synthesis

Voxel-based deep generative model

- Synthesizing new shapes
- Hard to scale up to high resolution
- Resulting in not-as-realistic shapes



Image credit: 3D ShapeNet [Wu et al., CVPR 2015]

3D Shape Synthesis

Realistic

New



Realistic + New

Adversarial Learning



Generative adversarial networks [Goodfellow et al., NIPS 2014]



DCGAN [Radford et al., ICLR 2016]



Our Synthesized 3D Shapes



3D Generative Adversarial Network



3D Generative Adversarial Network



Generator Structure





Results from 3D ShapeNet



Randomly Sampled Shapes

Tables





Results from 3D ShapeNet



Interpolation in Latent Space





Interpolation in Latent Space

















Unsupervised 3D Shape Descriptors



3D Shape Classification



3D Shape Classification Results

Supervision	Pretraining	Mathad	Classification (Accuracy)		
Supervision			ModelNet40	ModelNet10	
Category labels	ImageNet	MVCNN [Su et al., 2015]	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%	

3D Shape Classification Results

Supervision	Pretraining	Mathad	Classification (Accuracy)		
Supervision		Method	ModelNet40	ModelNet10	
Category labels	ImageNet	MVCNN [Su et al., 2015]	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%	
		SPH [Kazhdan et al., 2003]	68.2%	79.8%	
		LFD [Chen et al., 2003]	75.5%	79.9%	
Unsupervised	-	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%	

3D Shape Classification Results

Supervision	Pretraining	Mathad	Classification (Accuracy)			
Supervision		Method	ModelNet40	ModelNet10		
Category labels	ImageNet	MVCNN [Su et al., 2015]	90.1%	_		
		MVCNN-MultiRes [Qi et al., 2016]	91.4%	_		
		3D ShapeNets [Wu et al., 2015]	77.3%	83.5%		
	None	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%		
		SPH [Kazhdan et al., 2003]	68.2%	79.8%		
		LFD [Chen et al., 2003]	75.5%	79.9%		
Unsupervised	-	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%		

Limited Training Samples



Comparable with best unsupervisedly learned features with about 25 training samples/class

Comparable with best voxel-based supervised descriptors with the entire training set

Discriminator Activations



Units respond to certain object shapes and their parts.

Extension: Single Image 3D Reconstruction



Model: 3D-VAE-GAN



A variational image encoder maps an image to a latent vector for 3D object reconstruction. VAE-GAN [Larson et al., ICML 2016], TL-Network [Girdhar et al., ECCV 2016]

Model: 3D-VAE-GAN



We combine the encoder with 3D-GAN for reconstruction and generation.

Single Image 3D Reconstruction

















Input image Reconstructed 3D shape

Input image Reconstructed 3D shape

Single Image 3D Reconstruction

	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
Our 3D-VAE-GAN (jointly trained)	49.1	31.9	42.6	34.8	79.8	33.1	45.2
Our 3D-VAE-GAN (separately trained)	63.2	46.3	47.2	40.7	78.8	42.3	53.1

Average precision on IKEA dataset [Lim et al., ICCV 2013]

Contributions of 3D-GAN

- Synthesizing new and realistic 3D shapes via adversarial learning
 - Exploring the latent shape space
- Extracting powerful shape descriptors for classification
- Extending 3D-GAN for single image 3D reconstruction

