DeepHolo: Recognizing 3D Objects using a Binary-weighted Computer-Generated Hologram

Naoya Muramatsu
University of Tsukuba
Pixie Dust Technologies, Inc.
sh.mn.nat@gmail.com

Chun Wei Ooi
University of Tsukuba
made3102@gmail.com

Yuta Itoh
University of Tsukuba
s1711428@s.tsukuba.ac.jp

Yoichi Ochiai
University of Tsukuba
Pixie Dust Technologies, Inc.
wizard@slis.tsukuba.ac.jp

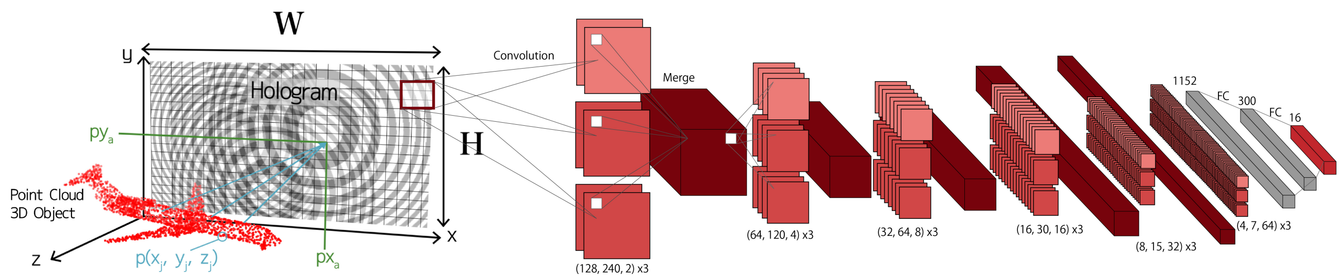

Figure 1: Framework of DeepHolo network.

ABSTRACT

Three-dimensions (3D) models contain a wealth of information about every object in our universe. However, it is difficult to semantically recognize the media forms, even when they featured in simplest form of objects. We propose the DeepHolo network using binary-weighted computer-generated holograms (CGHs) reconstructed from point cloud models. This neural network facilitates manipulation of 3D point cloud form, and allows it to be processed as Two-dimensions (2D) data. We construct the network using hologram data, which is simpler and contains smaller volume of information than their point cloud data(PCD) counterpart, leading to a smaller number of parameters required. The Deep Neural Network (DNN) is trained to recognize holograms of 3D objects, so that it resembles a point attributed to a 3D model of a similar object to the one depicted in the hologram. DeepHolo network allows high precision object recognition, as well as processing 3D data using much lesser computer resources. We evaluate our method on a recognition task space efficiency, and outperforming state-of-the-art methods.

SA '17 Posters, November 27-30, 2017, Bangkok, Thailand

ACM ISBN 978-1-4503-5405-9/17/11.
https://doi.org/10.1145/3145690.3145725

CCS CONCEPTS

• **Computing methodologies → Classification and regression trees**;

KEYWORDS

object recognition, convolutional neural network, computer-generated hologram

ACM Reference Format:
Naoya Muramatsu, Chun Wei Ooi, Yuta Itoh, and Yoichi Ochiai. 2017. DeepHolo: Recognizing 3D Objects using a Binary-weighted Computer-Generated Hologram. In *Proceedings of SA '17 Posters.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3145690.3145725

1 INTRODUCTION

The understanding of 3D environment is an important element of the latest computer vision research ranging from automotive to autonomous robotic application scenarios. In order to handle 3D information, however, large amount of memory is required.

In this paper, we propose a simple and effective neural network using computer-generated holograms for 3D object recognition and reconstruction. In conventional studies, various deep learning approaches to 3D recognition and construction have been attempted. VoxNet [Maturana and Scherer 2015] uses high number of parameters. Also, these methods require much complex networks.

We use computer generated holograms, they are the result of simulation of light interference between a coherent light source and point clouds of a 3D object. The resulted holograms are a 2D interference pattern with depth information which can be reconstructed

Table 1: Classification results on ModelNet40.

	input	#params	#views	accuracy avg. class	accuracy overall
3DShapeNets [Wu et al. 2015]	volume	-	1	77.3	84.7
VoxNet [Maturana and Scherer 2015]	volume	-	12	83.0	85.9
Subvolume [Qi et al. 2016b]	volume	16.6M	20	86.0	**89.2**
LFD [Wu et al. 2015]	image	-	10	75.5	-
MVCNN [Su et al. 2015]	image	16.6M	80	**90.1**	-
PointNet [Qi et al. 2016a]	point	3.5M	1	86.2	**89.2**
Ours Network	image	**0.67M**	1	80.2	83.3

into a 3D hologram when illuminated with a suitable light source. Our neural network takes these interference patterns as input, and learn the feature of input objects in 2D hologram patterns.

Instead of learning 3D objects as they are, we thought that it as possible to handle information densely by dropping the data into a 2D image. Also, in this method, DeepHolo can be more space efficient because we are able to reduce the parameters expressing the features.

2 APPROACH

The main idea of DeepHolo is to build up an end-to-end fashion deep neural network. As illustrated in Figure 1, our DeepHolo network mainly consists of two parts: CGH and 2D CNN. We first convert a 3D PCD into a hologram with a 128×240 resolution. The 2D hologram data is then fed into the 2D CNN. For 2D CNN, we use inception modules [Szegedy et al. 2015] to classify the 2D hologram image of the original 3D PCD. Key to DeepHolo is the implementation of the CGH program. The CGH program can generate binary hologram correctly, and can be trained with standard back-propagation, allowing for end-to-end fashion.

2.1 Computer Generated Hologram

We used a simple gradation representation method using a binary-weighted computer-generated hologram (CGH) [Fujiwara et al. 2017] to convert 3D PCD into binary hologram. This method uses multiple bit planes including binary weighted CGH with various pulse widths. The object point of the 3D object is assigned to multiple bit planes according to the gray level. Bit planes are sequentially displayed in a time division multiplexing manner. Therefore, the method realizes a gradation representation of the reconstructed 3D object.

2.2 Compression of binary hologram

The size of our CGH is full HD(1024×1920) that is large file capacity for DNNs. We need to compress the size of binary hologram and replaced 8×8 as one pixel of 64 bit expression. To solve this problem, we compressed the size of binary hologram by dividing every hologram into 128×240 parts of 8×8 pixels and regarding each of them as one pixel of 64bit expression.

2.3 2D CNN Architecture

Figure 1 shows the framework of 2D CNN. Our primary goal is to recognize a 3D point cloud data. We tackle the complication by proposing a CGH recognizing network based on some inception modules [Szegedy et al. 2015] that consist of three convolutional layers. The modules can be expected to have fewer spatially expanded clusters that can be covered by convolution over larger patches and that the number of patches will decrease over larger areas. DeepHolo has a simple structure of six inception modules followed by average pooling layers. At the end of the network, there are three fully connected layers and a sigmoid layer placed consecutively.

3 RESULTS

To evaluate the performance of view selection, we compare some DNNs against two alternative methods. The recognition network is trained on ModelNet40 [Wu et al. 2015] shape classification benchmark. Table 1 shows some of the recognition results and summarizes space (number of parameters in the network) complexity of our classification network. We also compare DeepHolo to a representative set of volumetric, multiview and point cloud based architectures in previous works. DeepHolo is much more space efficient than the others in terms of #params in the network.

REFERENCES

Masato Fujiwara, Naoki Takada, Hiromitsu Araki, Chun Wei Ooi, Shohei Ikawa, Yuki Maeda, Hiroaki Niwase, Takashi Kakue, Tomoyoshi Shimobaba, and Tomoyoshi Ito. 2017. Gradation representation method using a binary-weighted computer-generated hologram based on pulse-width modulation. *Chinese Optics Letters* 15, 6 (2017), 060901.

Daniel Maturana and Sebastian Scherer. 2015. Voxnet: A 3d convolutional neural network for real-time object recognition. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on.* IEEE, 922–928.

Charles Ruizhongtai Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas. 2016a. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. *CoRR* abs/1612.00593 (2016). http://arxiv.org/abs/1612.00593

Charles R Qi, Hao Su, Matthias Nießner, Angela Dai, Mengyuan Yan, and Leonidas J Guibas. 2016b. Volumetric and multi-view cnns for object classification on 3d data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 5648–5656.

Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. 2015. Multi-View Convolutional Neural Networks for 3D Shape Recognition. In *The IEEE International Conference on Computer Vision (ICCV).*

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 1–9.

Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 2015. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* 1912–1920.