Low cost GAN depth estimation for 3D object modeling

Arif Suganda

September 21, 2025

Abstract

Depth estimation, one of the leading research in computer vision, from autonomous driving to 3D modeling, play key role to artifical perception of the machine. This paper present reconstruction of 3D object with the optimal cost for accuracy, using simple Generative Adversarial Network (GAN) on NYU Depth v2 and KITTI dataset. Our approach in optimization achieved xx% accuracy with only xx epochs training resulted in xx times yield. We concluded this approach cost very inexpensive compared to other method.

1 Introduction

Depth estimation (DE) research in robotic [1], a-utonomous driving, and 3d modeling has varied in result with overall accuracy consistently high. There are many approach explored by the current researches from optimizing hardware [1], improving dataset [3], and/or using new set of architecture [3], [4]. The common approach on the current research is using deep learning (DL) as primary architecture in research.

DL is still the leading architecture model for monocular depth estimation (MDE) [5]. There are many variety used on MDE such as convolutional neural networks (CNNs), recurrent neural net-works (RNNs), variational auto-encoders (VAEs) and generative adversarial networks (GANs) [6]. Eigen et al. [2] are the pioneers for using CNN for MDE, by using local and global information from single image. GAN has become a hot research direction in recent years [6], commonly use generator to estimate global and local 3D structure, and discriminator to predict depth of monocular image. GAN also varied in framework, such as stacked GAN, conditional GAN [4] and Cycle GAN [6].

These varied method and framework still face a critical problem, high economic cost to create data-sets with ground truth [6], and high cost of tra-ining. Our approach aims to create 3D object modeling and reduce the cost of training, optimizing cost to accuracy and error function using GAN architecture framework, leveraging on GAN ability for capturing global structure with fewer training data than a standard en-coder-decoder CNN [4].

2 Related Work

3 Proposed Method

4 Experiment

5 Discussion

6 Conclusion

7 Acknowledgments

References

[1] Julie Chang and Gordon Wetzstein. Deep optics for monocular depth estimation and 3d object detection. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 10192–10201, 2019.

[2] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network, 2014.

[3] Vitor Guizilini, Rareș Ambruș, Sudeep Pillai, Allan Raventos, and Adrien Gaidon. 3d packing for self-supervised monocular depth estimation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2482–2491, 2020.

[4] Andi Hendra and Yasushi Kanazawa. Tpgan: Simple adversarial network with additional player for dense depth image estimation. *IEEE Access*, 11:44176–44191, 2023.

[5] Armin Masoumian, Hatem A. Rashwan, Julián Cristiano, M. Salman Asif, and Domenec Puig. Monocular depth estimation using deep learning: A review. *Sensors*, 22(14), 2022.

[6] ChaoQiang Zhao, QiYu Sun, ChongZhen Zhang, Yang Tang, and Feng Qian. Monocular depth estimation based on deep learning: An overview. *Science China Technological Sciences*, 63(9):1612–1627, June 2020.