

Analyzing Single Image 3D Model Reconstruction Methods: MarrNet and 3DRL

Introduction

A single image 3D reconstruction framework generates a 3D model given a 2D image. Two frameworks are proposed in literature: MarrNet and Modeling 3D Shapes by Reinforcement Learning (3DRL). Both methods break the problem into sequential steps, generating an intermediate representation of the model and then creating a 3D model. To understand the importance and advantages of intermediate models I design experiments that compare both approaches. My original objective was to do a direct comparison of MarrNet and 3DRL's results on a chair dataset. However, due to complications with the provided code, I was unable to produce results.

Comparing MarrNet and 3DRL

	MarrNet	3DRL
Input	2D RGB image	2D depth image
Intermediate steps	2.5D estimator and 3D estimator	Prim-Agent and Mesh-Agent
Learning approaches	Supervised learning is used to train estimation components	Combination of imitation and reinforcement learning to train agents
Metric	Voxel IOU and human studies	Voxel IOU and Chamfer Distance (CD)
Compared	Direct predictions	Direct predictions

Table 1. Comparison of MarrNet and 3DRL

Suggested Experiments

MarrNet

- Does MarrNet's performance improve given the ground truth depth, surface normals, and silhouette maps as inputs?
- Estimate 3D object shape using only depth map
- Estimate 3D object shape using only surface normals
- Would MarrNet improve using an imitation and reinforcement learning approach?

3DRL

- Would 3DRL's performance improve if the input considered depth, surface normals and silhouette maps?
- How would Mesh-Agent perform if the Agent was trained on ground truth dataset?

MarrNet

- MarrNet consists of two main estimators. A 2.5D sketch estimator and a 3D shape estimator [1].
- 2.5D estimator: estimates depth, surface normal, and silhouette maps
- 3D estimator: creates a 128x128x128 voxel-based reconstruction of the input image.

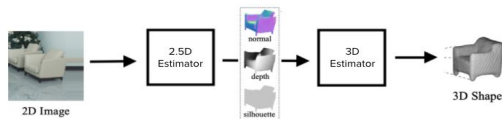


Figure 1. An overview of MarrNet's framework

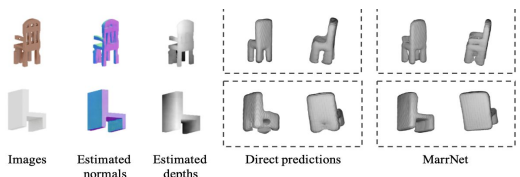
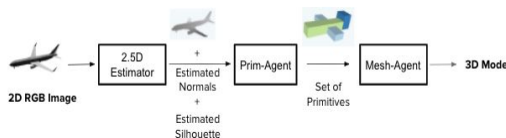


Figure 2. MarrNet vs. Direct Prediction

Proposed Combined Framework

After comparing MarrNet and the direct prediction baseline, I suspect the surface normal map is in fact important. To investigate this, I propose a framework that combines MarrNet and 3DRL.



3D Modeling using Reinforcement Learning

- 3DRL consists of Prim-agent and Mesh-agent. [2]
- Prim-Agent: approximates 3D shape with primitives
- Mesh Agent: given a set of primitives, Mesh-Agent edits the vertices using edge loops to produce higher mesh qualities

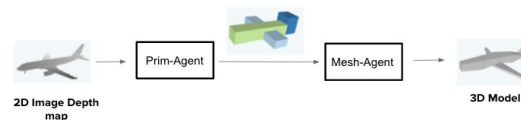


Figure 3. An overview of the framework



Figure 4. A step-by-step procedure of 3D modeling

Conclusion

- To further understand the importance of intermediate steps we design experiments that evaluate all parts of MarrNet and 3DRL's frameworks
- Neural networks are difficult to implement due to massive amounts of data needed, large computational power required, and outdated neural network programs

References

- [1] WU, J., WANG, Y., XUE, T., SUN, X., FREEMAN, W. T., AND TENENBAUM, J. B. Marrnet: 3d shape reconstruction via 2.5d sketches, 2017.
- [2] LIN C., FAN, T., WANG, W., AND NIESSNER, M. Modeling 3d shapes by reinforcement learning, 2020