Otto-von-Guericke University Magdeburg

Faculty of Computer Science

Master's Thesis

Synth2Real : 3D-Furniture Reconstruction in Ersatz Environment (S2R:3D-FREE)

Author:

Kartik Prabhu

October 15, 2021

Advisors:

Prof. X

National Aeronautics and Space Administration

Dr. No

Department of Technical and Business Information Systems

Otto-von-Guericke University Magdeburg

Prabhu, Kartik:
Synth2Real : 3D-Furniture Reconstruction in Ersatz Environment (S2R:3D-FREE)
Master's Thesis, Otto-von-Guericke University Magdeburg, 2021.

Abstract

Besides the title, the abstract is the most important part of your thesis, as most readers will only read title and abstract. Your goal is to advertise the rest of the thesis for potential readers. For that, you briefly explain what you are focusing on in the thesis. With a misleading abstract, you will miss interested readers and maybe even attract readers with wrong assumptions about your work which will stop reading soon. The abstract should describe the general area as well as the most interesting insights of the work. It is crucial to find the right level of abstraction and length. An abstract typically consists of one paragraph that is significantly shorter than the introduction.

Abstracts typically follow the same structure. You start by describing the research area as well as the general and the specific problem you are focusing on. Then, you outline how you approach the problem in terms of concepts and evaluations. Finally, you close with the most interesting insights that you gained and why they are relevant for the research area.

Acknowledgments

If you want to, you can thank your advisors and/or anyone else who supported you during the thesis in some way.

Contents

List of Figures **ix**

List of Tables **xi**

List of Code Listings **xiii**

List of Acronyms **xv**

List of Symbols **xvii**

1 Introduction **1**

1.1 Domain Adaptation 1
1.2 Volumetric representations of 3D shapes 1
1.3 GameEngines 1
1.4 Background and motivation 1
1.5 Goal of this Thesis 1
1.6 Structure of this Thesis 2

2 Related Work **3**

2.1 State of the art for 3d-reconstruction 3
2.2 Synthetic dataset generation 3
2.3 GAN based style transfer 3
2.4 Domain adaptatin 3

3 Concept **5**

3.1 Structure of this Thesis 5
3.2 Pix3D: A large-scale benchmark 5
3.2.1 Disadvantages of Pix3D 5
3.2.2 Why Pix3D? 7
3.3 Role of SceneNet 7
3.4 S2R:3D-FREE, a Pix3D based sythetic dataset 7

4 Implementation **9**

4.1 3D-Scene framework 9
4.2 3d-Reconstruction framework 9
4.3 Implement a domain adaptation technique 9

5 Evaluation **11**

5.1 Domain gaps - Qualitative 11

5.2 Domain gaps - Quantitative 11
5.3 Performance 11
5.4 Domain shift technique 11
5.5 Ablation study on chairs 11

6 Conclusion **13**
6.1 Summary 13
6.2 Limitations of game engines 13
6.3 Future Work 13

Appendix **15**

B Tutorial **17**
B.1 This is a section 17
B.1.1 This is a subsection 18

List of Figures

3.1 Distribution of pix3D **?** models 6

A.1 Caption in the list of figures 15

B.1 Drawing with tables and TikZ 17

List of Tables

B.1 Short caption for list of tables 18

List of Code Listings

List of Acronyms

FOL First-Order Logic
SLR Systematic Literature Review
SPL Software Product Line

List of Symbols

General

$\alpha[x\backslash y]$	Substitution of x in α with y
$\mathcal{P}(A)$	Power set of a set A
$f\|_{A'}$	Restriction of a function $f\colon A \to B$ to a smaller domain $A' \subseteq A$

First-Order and Dynamic Logic

$A \land B$	Formalization of the sentence "P and Q"

1. Introduction

As Deep Learning field goes deeper and wider, the need for abundant data has become the basic requirement. But in practice, collecting data manually is time-consuming, expensive and needs lot of effort. Synthetic data can be defined as data created artificially and not from real world events. These data can be created in abundance and hence are frontrunners for Deep Learning training. Synthetic data has another major benefit; it enables data privacy, which further enables companies in building tools without any exposure of personal data. Synthetic approaches are also means of getting data which might be difficult to collect and annotate in real world.

1.1 Domain Adaptation

Why is it needed? How can it be achieved? Disadvantages of not achieving it?

1.2 Volumetric representations of 3D shapes

1.3 GameEngines

Advantageous of using game engine like unity Advantages, uses, shaders, rendering

1.4 Background and motivation

Why synthetic data is necessary

1.5 Goal of this Thesis

The following research questions will be answered:

1. Can domain gap be reduced between synthetic and real images using game engines?
2. Can Ersatz environment from a game engine like Unity replace real data as training data for 3d reconstruction?

1.6 Structure of this Thesis

The remainder of this thesis is structured as follows:

- Chapter 2, we visit some of the related work on synthetic data generation and state of the art 3d reconstruction networks.
- Chapter 3 deals with concepts and design choices. x
- In Chapter 4, we discuss in detail the 3DScene tool developed using Unity game engine for creating ersatz environment.
- Chapter 5 is dedicated to reviewing and discussing evaluation results obtained by comparison of synthetic and real dataset.
- Finally in Chapter 6, we conclude this thesis with results of the study, highlight few limitations and discuss future improvements.

2. Related Work

2.1 State of the art for 3d-reconstruction

Voxel based, mesh based Pix2vox,OctNet,Mesh rcnn,Pixel2Mesh,Occupancy networks, etc

2.2 Synthetic dataset generation

Synthetic datasets available Explain briefly Scenenet,Openrooms,Hyperism,3dfront,NVIDIA Deep learning Dataset Synthesizer (NDDS),Habitat: A Platform for Embodied AI Research,BlenderProc: Reducing the Reality Gap with Photorealistic Rendering,SynthDet: An end-to-end object detection pipeline using synthetic data None of them seem to provide 3d reconstruction except for 3dfront for some extent. A table to explain diffs between various methods(only if 3D-FREE has more advantage)

2.3 GAN based style transfer

2.4 Domain adaptatin

Discuss other methods used to mitigate domain shift Example: distance learning, subspace matching

3. Concept

In this section we discuss about dataset and design choice,and the rationale behind reducing the domain gap.

3.1 Structure of this Thesis

SceneNet,room models used

3.2 Pix3D: A large-scale benchmark

In **?**, a large-scale benchmark for 2D-3D alignment was introduced. The raw images from web search engines were collected and labelled keypoints were used to align the 2D images with the corresponding 3D shapes. The 3D models are extension from IKEA dataset **?**, which is a collection of high-quality IKEA furniture. The dataset also provides with masks and keypoints for the object under observation. For adding more images to the IKEA dataset, the authors of **?** conducted manual web search on Google, Bing, and Baidu, using Ikea model name as keywords. to get around 104,220 images which were further filtered by removing irrelevant images with the help of Amazon Mechanical Turk (AMT) workers. After this manual experiment the total images in Pix3D, only 14,600 images were selected for the 219 Ikea models. For our experiment on synthetic to real dataset, we chose to select only furniture classes from Pix3D, leaving out ”misc and tools classes, which were significantly less to begin with.

3.2.1 Disadvantages of Pix3D

The distribution of models in pix3d is as shown in 3.1. As we can see, the dataset distribution is uneven across classes and more than 50% of classes have less than 1000 images.

Though Pix3D set a benchmark for 2D-3D alignment, here are few disadvantages of using this real dataset.

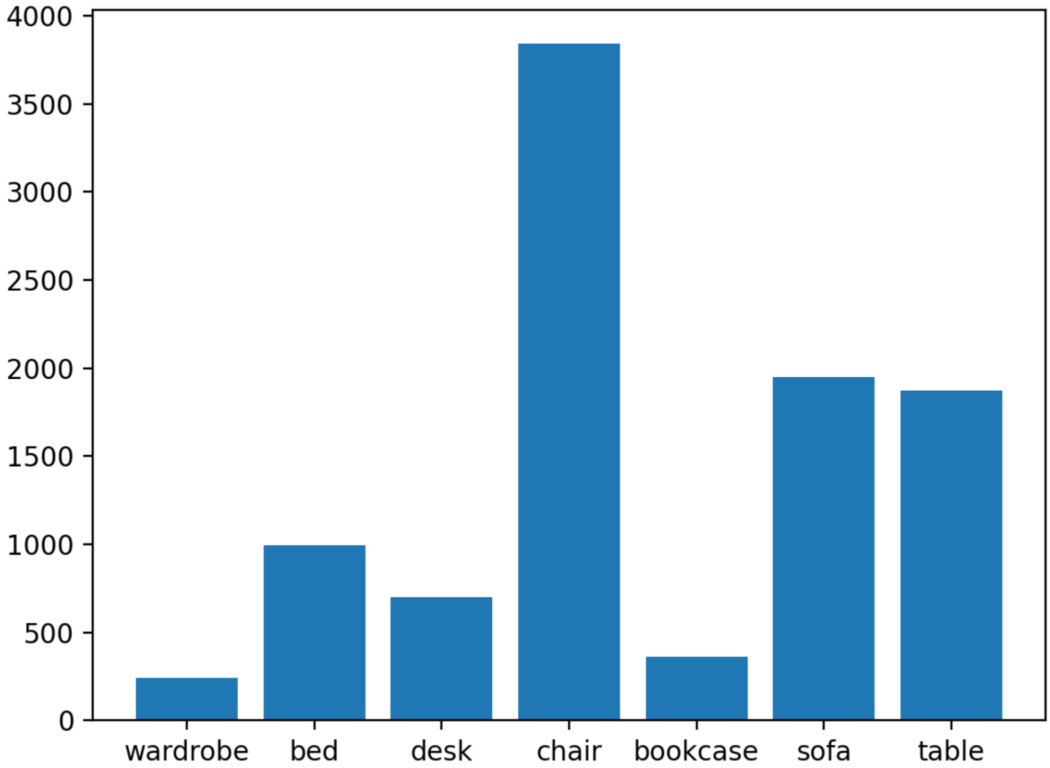


Figure 3.1: Distribution of pix3D **?** models

1. For Deep Learning approches, we need large-scale data and 14,600 might not be sufficient.
2. The orientation of object is not randomised.
3. The dataset does not provide 2.5D information (i.e. depth and normal) which can be crucial for 3D learning.

3.2.2 Why Pix3D?

Pix3D is a perfect example of having limited real world data. And since 3D models are available for each furniture,a synthetic data can be generated in abundance from those models.

3.3 Role of SceneNet

SceneNet **?** is a large collection photorealistic indoor scene trajectories. The dataset provides with images and videos of indoor scene which can be used for tasks like SLAM, semantic and instance segmentation, object detection and further enhanced for other vision problems like optical flow depth and pose estimation **?**). They use ShapeNet **?** models to fill-in 57 indoor scenes which gives the scene unlimeted configurations. Unfortunately there is no mapping from scene to 3D model for tasks like3D reconstructions. In our approach, we utilize the scene provided by SceneNet as layout for our indoor scenes. This include initial shapenet models with furniture placement. We then replace the class under observation with a corresponding model from Pix3D.

3.4 S2R:3D-FREE, a Pix3D based sythetic dataset

. We randomise the indoor scene from SceneNet **?** along with textures provided by them.

4. Implementation

4.1 3D-Scene framework

Replacing models, lighting(indoor and outdoor), randomisation of texture, camera distance, camera angles, outdoor skybox, UI controls Manual inputs, automated Unity - ML-ImageSynthesis UML diagram is needed? (If it helps)

4.2 3d-Reconstruction framework

Development environment Training setup(hardware, voxel size. . .) Mixed precision using apex If mixed precision helped performance, memory, speed, evaluation metric

4.3 Implement a domain adaptation technique

To serve as a proof if more domain gap reduction is needed

5. Evaluation

5.1 Domain gaps - Qualitative

Visualize domain gap between S2R:3D-FREE and real image(pix3d) Visualize domain gap between other synthetic dataset like 3dfront, scenenet tsne

5.2 Domain gaps - Quantitative

Maximum Mean Discrepancy Kullback-Leibler divergence

5.3 Performance

Performance of model trained on synthetic dataset by testing model with real dataset(pix3d) Qualitative: Voxel output comparisons Quantitative: IOU, Dice

5.4 Domain shift technique

Compare the differences in domain shift of models learnt on S2R:3D-FREE and a traditional domain shift learning.

5.5 Ablation study on chairs

Check if randomising textures of chair and background helps improve performance. Example: Create dataset with constant room texture vs randomising texture, Check if lighting helps (indoor, outdoor). Example: Create a dataset with constant lighting(only outdoor lighting) vs indoor lighting. Iou per category: as in https://arxiv.org/pdf/1905.03678.pdf

6. Conclusion

6.1 Summary

6.2 Limitations of game engines

6.3 Future Work

Appendix

This table is automatically generated from a CSV file. Of course, you can also create tables from scratch.

	List_ Insert	List_ Search	Ordered_ Insert	Ordered_ Search	Ordered_ Min	Ordered_ Sort	Set_ Insert
coarse, Ex., Strict	2400	1418	10608	10824	2871	10169	23366
coarse, Ex., Def.	2400	1418	9652	20172	2871	14043	35884
coarse, Ex., None	2400	1418	14347	20160	2871	12687	54907
coarse, L.S., Strict	2400	1418	10608	10824	2871	10169	21047
coarse, L.S., Default	2400	1418	9430	20172	2871	13835	24688
coarse, L.S., None	2400	1418	13802	20160	2871	12687	28501
coarse, Complete	2400	1418	4504	10149	2871	10047	16114
coarse, Product	9600	2836	9010	20298	5742	20094	16114
fine, Ex., Strict	2211	1321	11210	16984	2501	6500	12901
fine, Ex., Def.	2211	1321	10299	16971	2501	6420	14695
fine, Ex., None	2211	1321	10881	16836	2501	7115	23005
fine, L.S., Strict	2211	1321	11210	16984	2501	6500	12301
fine, L.S., Default	2211	1321	10251	16971	2501	6420	13173
fine, L.S., None	2211	1321	11116	16836	2501	7115	20053
fine, Complete	2211	1321	2018	16531	2501	6613	9336
fine, Product	8844	2642	4036	33062	5002	13226	9336

Figure A.1: Caption in the text.

B. Tutorial

Acronyms like software product lines (SPLs) and first-order logic (FOL) are automatically added to the list of acronyms above. To reference a chapter, subsection, figure, etc., put a label after the command (see above) and then use Chapter 1.

You can mark stuff as TODO, which is useful for writing drafts.[1]

Or put notes to your advisors in the margin.

Theorem B.1: My Cool Theorem

You can also add definitions and theorems in mathematical theses.

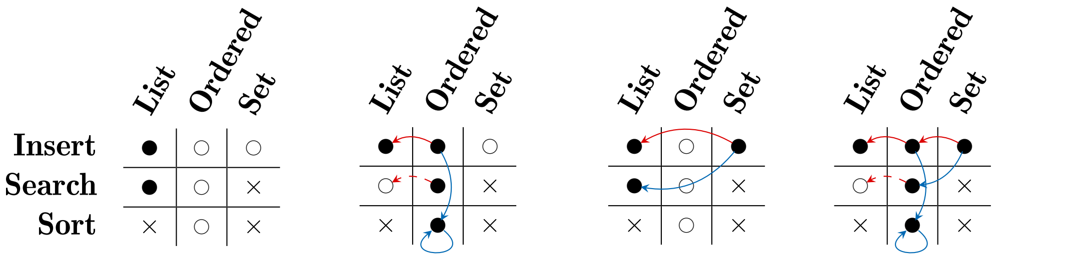
Proof. You can also add proofs, though most theses do not require any. □

In Figure B.1, we show an example of how to typeset a complex figure only with LaTeX by using TikZ.

B.1 This is a section

For citing, put entries in the BibTeX file (one example is there) and use the following to reference the authors directly: **?** conducted a systematic literature review (SLR)

[1]Footnotes are also possible, but are rarely used in computer science.



(a) List **(b)** List, Ordered **(c)** List, Set **(d)** List, Ordered, Set

Figure B.1: Drawing with tables and TikZ.

Table B.1: Long caption for a table.

Column1	Column2	Column3	Column4
Left	Right	Centered	Left, fixed width
Multi column			
Multi row	X		
	Y		

(this is an acronym defined in the main file), or use the following for references in parenthesis [**?**]. **? ?**

B.1.1 This is a subsection

Numerated list

1. One
2. Two
3. Three

Bullet list

- One
- Two
- Three

I herewith assure that I wrote the present thesis independently, that the thesis has not been partially or fully submitted as graded academic work and that I have used no other means than the ones indicated. I have indicated all parts of the work in which sources are used according to their wording or to their meaning.

Magdeburg, 15th October 2021