3D ImageNet: A data collection and labeling tool for Depth and RGB Images

Gurjeet Singh
Department of EECS
Oregon State University
Corvallis, OR 97331
singhg@oregonstate.edu

Sifan Zhou
School of Automation
Southeast University
Nanjing 210096, China
zhousifan@seu.edu

James Qian
State Key Laboratory of ASIC and system
Fudan University
Shanghai, China
jqian20@fudan.edu.cn

Patrick Chiang
PhotonIC Technologies
Shanghai, China
pchiang@photonic-tech.com

Abstract

3D-sensing is increasingly being used everywhere, including tablets, smartphones, robots, and autonomous vehicles. One major limitation to the usage and application of 3D-depth data is that very few databases have clean and accessible data, preventing researchers from building new applications and algorithms. This paper proposes 3D-ImageNet – an analogue to the original ImageNet [2], which spurred the 2D image-processing AI explosion. 3D-ImageNet's goal is to do the same for 3D-depth-sensing as was experienced for 2D images. This paper describes an open-source system that multiple cellphone users can use to collect and label a large amount of 3D data.

1 Introduction

One of the significant limitations to 3D-sensing, from the datacentric-AI approach, is different datasets collected from different 3D-sensors, but no unified methodology for 3D-data collection, cleaning, and labeling. Just like the time period before ImageNet [2] was announced, the amount of 3D-depth AI algorithms and application scenarios is very limited, sparse, singular, and disparate.

Conventional research on 3D images is based on small datasets collected from different sensor types for different scenes. To make such data usable, we have to either use transfer learning, which means transfer from weights of existing 2D images, or limit the 3D for essential feature extraction work in conjunction with RGB images. The ability to create and label large datasets can usher in new algorithms for 3D data and remove the problem of non-availability of ground truth for depth data.

This paper proposed 3D-Imagenet, whose goal is to provide an open-source framework for researchers world-wide to unify their 3D-data collection, cleaning, and labeling, so that the state of 3D-sensing AI-algorithms can experience the growth that was previously seen for 2D-images through the original ImageNet.

2 Prior Art

This section discusses the previous work done in this area. We will look into two aspects of the existing open-source systems and existing 3D datasets available

Preprint. Under review.

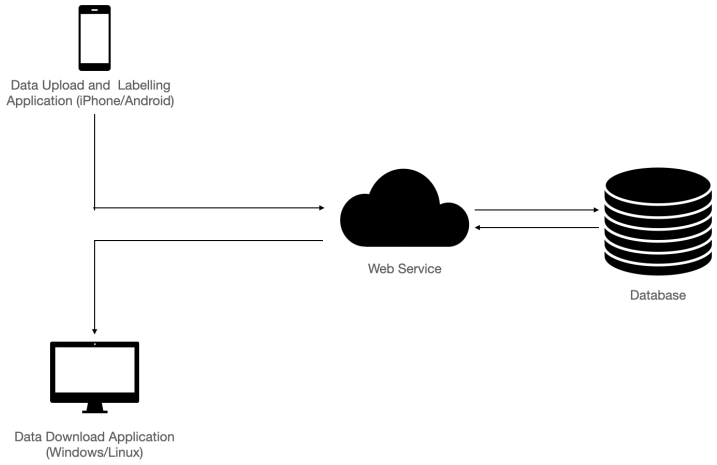


Figure 1: Basic blocks of 3DImageNet

2.1 ImageNet

ImageNet is one of the most popular open-source datasets used for various deep learning [52]. As the name suggests, it is a dataset with millions of images for training and benchmarking. ImageNet is a hierarchical dataset based on synonyms sets or synsets available in WordNet [3]. Each synset has about 1000 images. When ImageNet was published, it also started a Large Scale Visual Recognition Challenge (LSVRC). ImageNet created a hierarchy by downloading images from internet search engines and carefully labeling them using the Amazon Mechanical Turk (mTurk) service [1]. ImageNet first started with the Image Recognition challenge and later expanded to various other research fields such as Object Recognition, semantic segmentation, etc. Several deep learning applications use the ImageNet database as the training database and later use different datasets to achieve their goal.

3 System Architecture

Figure 3 shows the basic blocks of the overall system. Currently, the system is already available for Android cellphones that already have a ToF (Time-of-Flight) sensor, and is currently being ported to iOS. It is divided into several sub-blocks which are explained below.

3.1 Data Upload Application

The CameraManager interface of Android is used to access depth data. CameraManager allows depth information to be obtained from the sensor using the DEPTH16 format, but its resolution depends on the actual sensor. Therefore, the resolution will be assigned dynamically according to the device to accommodate most of the resolutions available in the market.

3.2 Web Service

Web Service is the interface through which various applications access the database, such as Microsoft SQL Server. For a mobile application, it is difficult to access the database directly. So, Web Service acts as an intermediary between the Database and applications. The Applications send their data in the form of web service calls. Each call to the Database causes the data to be uploaded to the Database. In the Desktop application case, it brings the stored data from the database and updates the database with labels.

3.3 Database

Our design uses Microsoft SQL Server. Unlike ImageNet, which uses a hierarchical structure, we use a Relational Database Management System (RDBMS). The reason for this choice is the data is not being stored according to the hierarchical design of wordnet. Instead, the dataset is being updated dynamically according to the requirements. Microsoft SQL server is capable of handling multiple requests. So, it is capable of handling a large number of requests at the same time. So, if many users are uploading data to the system, it can handle all such large data loads.

4 Discussion and Future Work

3DImageNet is a long-term project goal that will have multiple stages. Our goal is to provide a system where any user can upload and label the objects and get the benefit of the work of other contributors. We are planning to be the most significant source of diverse depth maps. The goal is to develop the system on a diverse range of devices with depth sensors so that users can easily don't have to write any program and upload it to the database. We are starting the first version of the system with three subgroups and will continue to add subgroups according to the requirements of the community.

References

[1] Amazon Mechanical Turk,https://www.mturk.com(accessed: 21st March 2021).

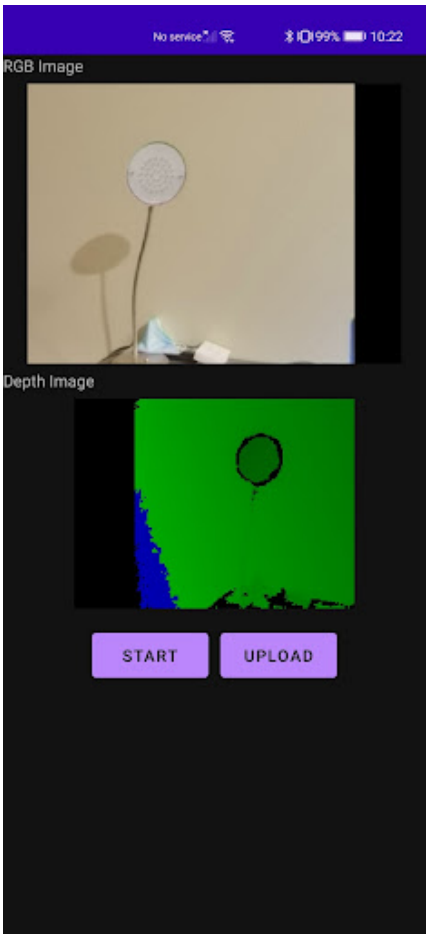
[2] Deng, Jia, et al. "Imagenet: A large-scale hierarchical image database." 2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009.

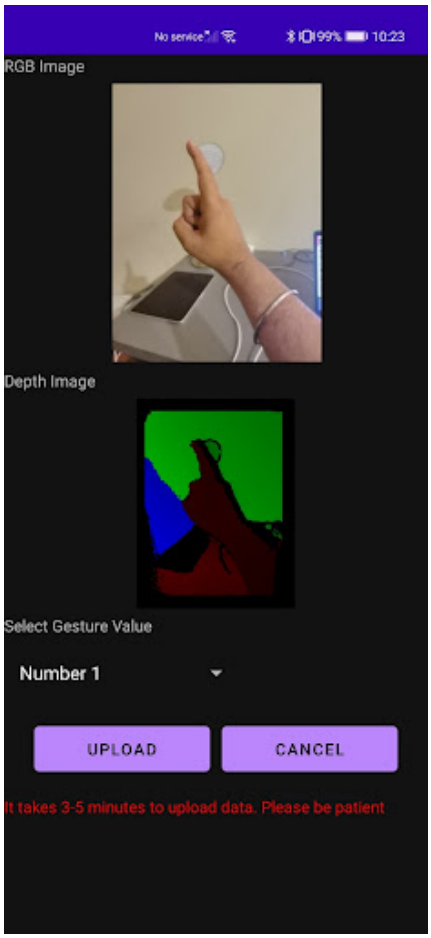
[3] Miller, George A. "WordNet: a lexical database for English." Communications of the ACM 38.11 (1995): 39-41.

[4] "Sketchfab", https://sketchfab.com, accessed 26th March 2021.

A Appendix

A.1 Mobile Application Screenshots





A.2 Database Download Application