Project 1 Detecting 3D Objects based on ARKit

Zhiao Zhou U65882678

Problem Statement:

The topic is about Detecting 3D Objects in AR. In computer vision, detecting 3D objects refers to the technology of recognizing and determining 3D objects captured by the camera. The basis of detecting 3D objects is image recognition, but it is much more complicated than 2D image recognition, because its visual feature points are distributed in 3-dimensional space, the shapes and textures of objects seen from different angles are different, it requires much more data and higher computing resources than 2D image recognition. In ARFoundation, detecting 3D objects searches for corresponding 3D real objects in the real environment by pre-recording the spatial feature information of 3D objects. Similar to image recognition, 3D object recognition in ARFoundation also requires a reference object library, and each object in this reference object library is the spatial feature information of a 3D object. To obtain the spatial feature information of the reference object, you can scan the real 3D object to collect its feature information, and generate the .arobject reference object spatial feature information file. The .arobject file only includes the reference object spatial feature information, and this file cannot be used to restore the reference object structure. The spatial feature information of reference objects plays a key role in quickly and accurately recognizing 3D objects. This topic is important because we always need 3D information. Such as when we trying to develop a new 3D application or when we need to identify the distance with a 3D obstacle. So, in my opinion that researching and developing better 3D object detection approach will help us process the information in the 3D environment more accurately as we do with 2D images now and can be applied in more situations where 3D object detection can be used.

Applications:

From my own perspective, detecting 3D objects can help us obtain more accurate positioning information and it can be applied in multiple situations, such as 3D games and simulation modeling of furniture in AR environments. This technology will help people as the carrier of the combination between virtual and reality and provide convenience for people in life. For instance, virtual try on shoes. Amazon, Gucci and lots of brand, they provide an application that their customer could see how the shoes look like on their feet by their device. Recently, Nike developed a new application which could match the best shoe size for user. They do this by acquiring user foot data through 3D object detection. In addition, the most promising application direction of 3D objects detection and tracking is autonomous cars. The autonomous driving

problem is one of the most actual problems in both research and industrial fields. So, in autonomous cars, 3D objects detection can improve driving safety by tracking and identifying static and dynamic obstacles. The societal significance of all these researches is to better provide convenience and security for all people's lives. Above all, I am very interested in the direction of this project, I would love to continue study and research.

The area that I am interested in this topic:

Since we know that we can use Lidar to get point cloud 3D data, I would like to do more research on how to catch the 3D object we want more effectively. As we already know that our 2D image detection is much faster and accurately by new approach are developed. I think this will be reimplemented once in 3D object detection. I would like to do some study and research in this area.

Literature review:

After reading several papers, we know that there are two main methods could help us to accomplish 3D objects detection. One is detected 2D image to reconstruction 3D model, the other is by using lidar to get 3D point cloud data, then by train an object detection system to identify the object we want.
In paper 'Integrating YOLO Object Detection with Augmented Reality for iOS Apps' [4]. The author introduces an MLModel built using TinyYOLO that could detect the identity and location of the pre-trained objects in an image. TinyYOLO is a much faster but less accurate detection system than YOLO, which means we can detect every frame of the video. If the MLModel detected an object, it will record the 2D coordinates of the object in the image, then using the ARKit HitTest API to convert 2D coordinates to 3D coordinates. After we get the 3D coordinates, we could overlay the 3D graphics by using ARKit which means we accomplished the reconstruction. As we said, this method depends on the trained model, so its accuracy decided by how much training this model did. Similar ideas have also appeared in 'Augmented Reality and Machine Learning Incorporation Using YOLOv3 and ARKit' [5], what they did is combines machine learning to detect and track augmented reality maker targets in an applications using deep neural networks. They implement the auto-generated dataset tool to help the system do the training, so the author said that they required less time and labor. Above all, using YOLO to combine machine learning and deep neural network is a very efficient way to restore reality and this method is a good direction could continue to be studied.
Except this, another method had already been proposed for several year. Which is using lidar to collect 3D point cloud data, then by train an object detection system to capture and identify the target object. Author of 'Object Recognition in 3D Point Clouds Using Web Data and Domain Adaptation' [6] showed how to use objects from Google's 3D Warehouse to train an object

detection system. This system can detect in the collected 3D point cloud data. By using the lidar on our device, we could directly get the depth data which means the data we collect is not a 2D data anymore. So, we do not need to train a model to compute the 3D data. So, to some extent this method is more accurate. After we know both methods, bringing 3D objects detection into our lives is the most important thing. There is an upcoming article [7] explains to us the possibility of further application of 3D object detection on smartphones. As a low-cost and highly efficient technique for 3D obstacle detection and object recognition in indoor environments, they gave us a good guide from an application point of view. It is like being a bridge between technology and people's live.

After reading all the articles, I am more interested in 3D objects detection from point cloud data and develop meaningful applications. I would like to try to explore more possibilities of 3D objects detection and develop useful applications.

1. Lee, Wonwoo, Nohyoung Park, and Woontack Woo. "Depth-assisted real-time 3D object detection for augmented reality." *ICAT*. Vol. 11. No. 2. 2011.

2. Rahman, Mohammad Muntasir, et al. "Notice of violation of IEEE publication principles: Recent advances in 3D object detection in the era of deep neural networks: A survey." *IEEE Transactions on image processing* 29 (2019): 2947-2962.

3. A. Agafonov and A. Yumaganov, "3D Objects Detection in an Autonomous Car Driving Problem," 2020 International Conference on Information Technology and Nanotechnology (ITNT), 2020, pp. 1-5, doi: 10.1109/ITNT49337.2020.9253253.

4. S. Mahurkar, "Integrating YOLO Object Detection with Augmented Reality for iOS Apps," 2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2018, pp. 585-589, doi: 10.1109/UEMCON.2018.8796579.

5. Le, H., Nguyen, M., Yan, W.Q. and Nguyen, H., 2021. Augmented reality and machine learning incorporation using YOLOv3 and ARKit. Applied Sciences, 11(13), p.6006.

6. https://journals.sagepub.com/doi/10.1177/0278364910369190

7. Chen, J. and Zhu, Z., 2022, January. Real-Time 3D Object Detection and Recognition using a Smartphone [Real-Time 3D Object Detection and Recognition using a Smartphone]. In Proceedings of the 2nd International Conference on Image Processing and Vision Engineering-IMPROVE.

Open source research:

1.https://github.com/hanleyweng/CoreML-in-ARKit
2.https://github.com/KleinYuan/tf-3d-object-detection
3.https://github.com/yudiz-solutions/Detect-3D-Object-ARKit-2.0
4. https://github.com/xingyizhou/CenterNet
5. https://github.com/open-mmlab/OpenPCDet

Duplicate the results:

App: IKEA Place

By using this application, I further felt the practicality of this project. I think this app is very useful not only because we can use this app to see what the furniture really looks like in our home, but also because this application can consider both the high degree reduction of furniture and the front and back relationship of objects. That is why I think this is an amazing application. However, in my opinion, this application still has parts that can be improved, such as zooming like a camera, which can present more content to customers when the screen size of a mobile phone is limited, thereby replacing customers with moving themselves to control the distance.