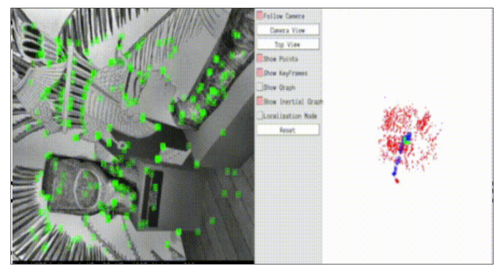
Using Synthetic Data & Simulation to Enhance Object Detection Algorithms in Indoor Environments

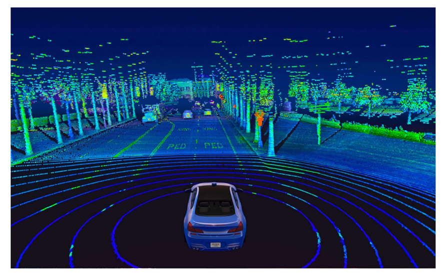
Yuhe (Alice) Hu, Ying Chen, Maria Gorlatova
Intelligent Interactive Internet of Things (I^3T) Lab

Background

Motivation

- Object detection in computer vision is an indispensable component for robotic navigation, autonomous driving, and augmented reality.





- Acquiring extensive and diverse training data is costly, time-consuming, and biased towards historical scenarios [1].

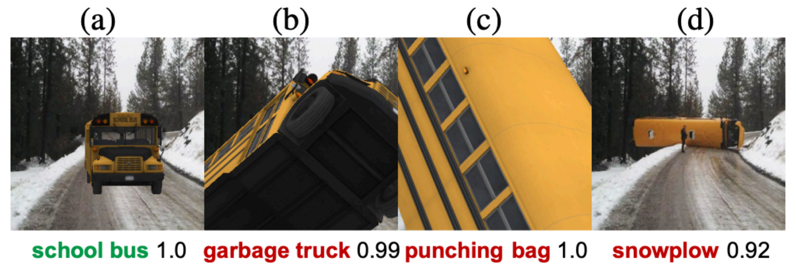
Synthetic Dataset Generation

- Synthetic data generated from simulated environments can significantly improve the accuracy of computer vision models [1].
- This study uses the Unity game engine and Meta Oculus 3 to create high-fidelity 3D reconstructions of real-world indoor environments

Assorted snapshots of high-fidelity 3D reconstructions in Simulated Indoor Environment [2].

Related Works

- High misclassification of object poses in DNNs underlines need for robust training with synthetic datasets to improve accuracy [3].



The Google Inception-v3 classifier [3] correctly labels the canonical poses of objects (a), but fails to recognize out-of-distribution images of objects in unusual poses (b–d) [3]

- Monocular SLAM systems tested on simulated forest terrain miss key environmental complexities due to lighting variability and motion [4].

Photorealistic simulated forest (Left) and region of real forest (Right)

- SceneNet RGB-D's vast synthetic indoor datasets facilitate object detection models' training [2].

Synthetic Dataset Generation

1. Indoor Environments in Unity Game Engine

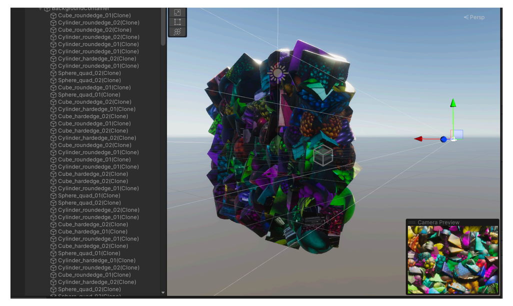
Developed 3 indoor scenes to reflect real-world scenarios

Virtual fire station scene

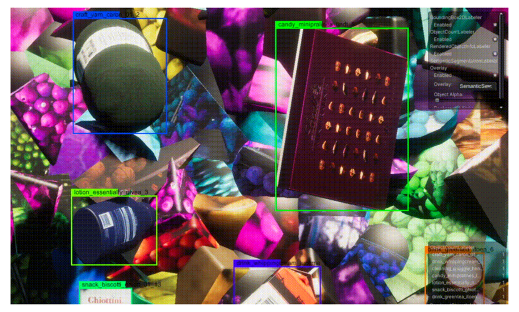
Virtual living room environment

2. Camera Movement, Ground Truth, Randomizations

Built C# script to control camera movement and angles
Implemented object detection, ground-truth annotation



Background Randomizer



Foreground Randomizer with Annotations

3. Collecting Synthetic Dataset

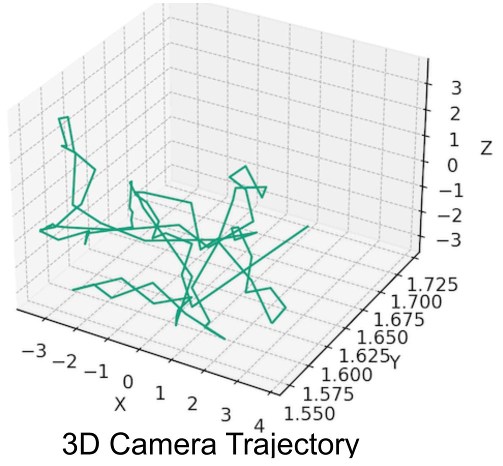
C# script captures 10-min footage per scene with annotations (object label IDs, pixel visibility, camera position and rotation)

Inspection of Generated Synthetic Data

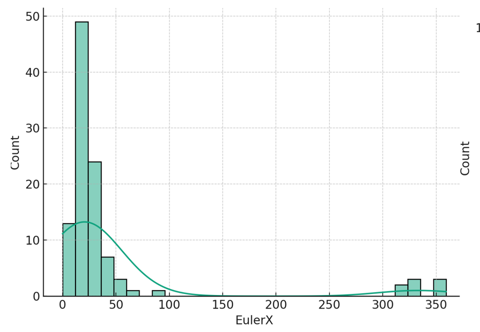
Synthetic Dataset Analysis

Conclusions from Dataset Recording:

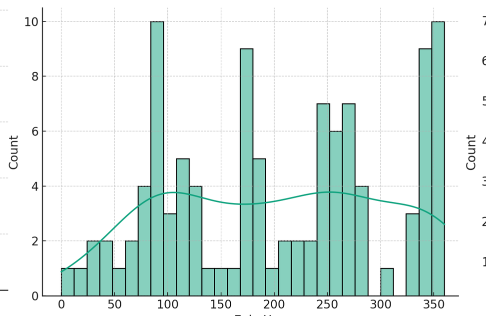
- Central interest point:
  - X: -0.39, Y: 1.55, Z: 0.05
- Preferred camera angles:
  - X: 15°, Y: 90°, Z: 0°
- Synthetic data reveals user preference, interaction patterns; useful for improving VR engagement.



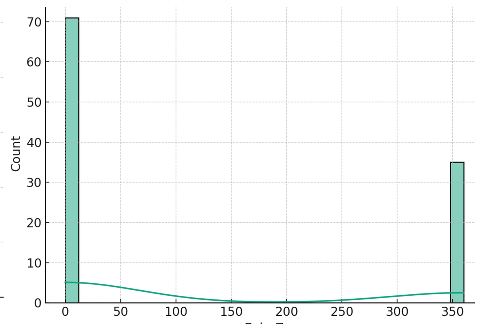
3D Camera Trajectory



Distribution of X-Rotation Angle



Distribution of Y-Rotation Angle



Distribution of Z-Rotation Angle

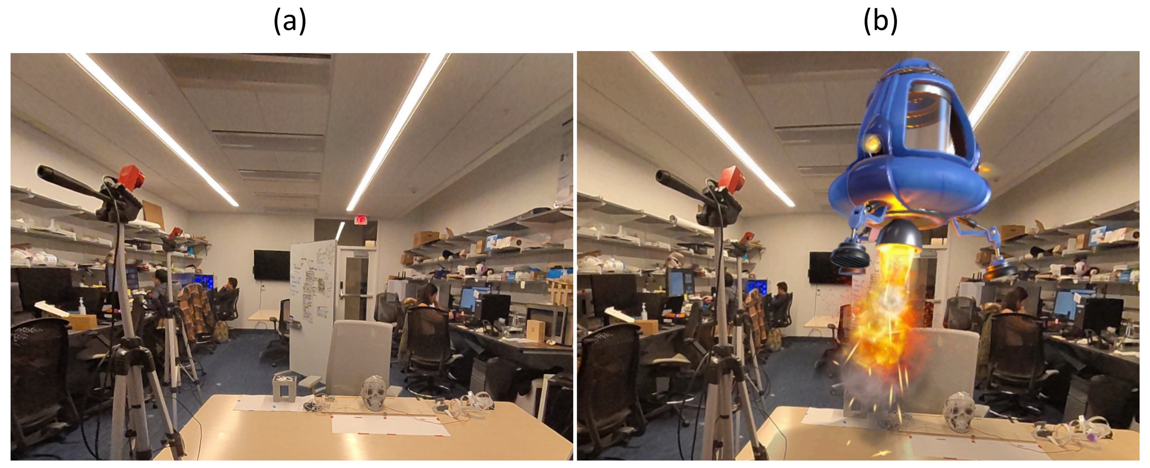
Meta Oculus Quest 3 Integration

Meta Oculus Quest 3 in Synthetic Data Research

- Serves as a <u>high-definition platform</u> and <u>fluid and responsive interface</u> to compile synthetic datasets and simulations
- Adds <u>realism</u> and <u>responsiveness</u> to virtual and augmented reality experiences with 2064 x 2208 resolution (+29.5% from Quest 2)

Meta Oculus 3 Headset, launched 10/10/2023

Transition from VR fire-station scene to real-world setting



Oculus Screenshots: (a) Real-world lab environment (b) AR overlay in action

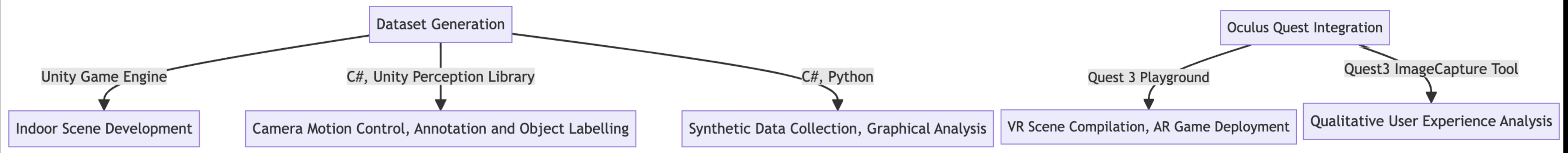
Virtual Reality (VR) Experimentation

- Compiled 3 Unity indoor scenes to Quest3
- Provide enhanced of user engagement metrics due to <u>expansive, cinematic</u> experience

Augmented Reality (AR) Experimentation

- Deployed 3 AR games to Quest3
- Provide insights into realistic <u>user-environment dynamics</u> and <u>spatial awareness</u>

Conclusions



- Synthetic datasets show potential for <u>enhancing object detection</u> in <u>virtual indoor</u> settings
- Variability in object placement, lighting, textures can prove beneficial for algorithmic accuracy
- Unity enabled <u>streamlined annotation and randomization</u>; VR offered an <u>immersive user experience</u>, and AR provided a <u>realistic setting</u> for application testing and simulations

Example Applications:

- Enhance smart home automation and robot navigation using synthetic data, simulations.
- Developing enhanced navigation systems

Future works

- Broaden VR/AR simulations, synthetic data to outdoor environments
- Assess synthetic dataset accuracy

References

[1] Y. Chen, H. Inaltekin, and M. Gorlatova, "AdaptSLAM: Edge-assisted adaptive SLAM with resource constraints via uncertainty minimization," in Proc. IEEE INFOCOM, 2023.
[2] McCormac, John & Handa, Ankur & Leutenegger, Stefan & Davison, Andrew. (2016). SceneNet RGB-D: 5M Photorealistic Images of Synthetic Indoor Trajectories with Ground Truth.
[3] Alcorn, Michael & Li, Qi & Gong, Zhitao & Wang, Chengfei & Mai, Long & Ku, Wei-Shinn & Nguyen, Anh. (2019). Strike (With) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects. 4840-4849. 10.1109/CVPR.2019.00498.
[4] Garforth, James & Webb, Barbara. (2019). Visual Appearance Analysis of Forest Scenes for Monocular SLAM. 1794-1800. 10.1109/ICRA.2019.8793771.

Duke UNIVERSITY