Capstone Project Poster: Detection and Converting Person(Object) from 2D rgb-image(Reality) to 3D mesh model(Virtuality)

Nishank Singhal, Mentor: Dr. Christelle Scharff

Pace University

Abstract

The objective of this capstone project is to use a Deep Learning to convert real-world environments into virtual worlds through object detection and mesh techniques onto Jackal Robot.

Figure 1. Jackal robot

Deep learning algorithms such as Yolo, PIFuHD, and posenet-model are being implemented to accurately recognize and classify objects in the environment. The system is rendering a person from an image in a virtual space based on data gathered by the robot's sensors. The project aims to evaluate the accuracy and efficiency of the system and its potential applications in various fields. Although the project had to be ended prematurely due to a disaster, the system successfully detects and converts persons and chairs to virtual representations. Future work would involve implementing style-gan to handle texture in 3D models. The project seeks to advance the fields of robotics and virtual reality by providing the Jackal Robot with the capability to convert reality into virtuality.

Research Question

- **Object Detection** What is the level of accuracy achieved by the Yolov5 model for object detection on a Jackal Robot in real-time?
- **Mesh Model** How accurately can we generate a 3D model of a person in real-time using PIFuHD and posenet-model on the Jackal Robot while maintaining good quality in the rendered virtual world?
- **Connectivity** How fast and accurate results are reflecting.(Tells run-time)

Related Work

- **Object Detection using Deep Learning: A Survey** by Shengcai Liao et al.
- **Robot perception of the environment: A survey**
  by Xinlei Pan, Yurong You, Ziyan Wang, and Cewu Lu
- **Virtual to Real Reinforcement Learning for Autonomous Driving**
  by Xinlei Pan, Yurong You, Ziyan Wang, and Cewu Lu .
- **CAT-Det: Contrastively Augmented Transformer for Multi-modal 3D Object Detection** in 2021 by
  authors Jianfeng Wang, Yukuan Yang, Zhiliang Tian, Junchi Yan, Zeming Li, Shuai Yi, and Li Erran Li.

Dataset

Dataset was directly obtain from Jackal robot. data consist of a robotics lab pictures captured from every angle by moving the robot across the room. By doing so I was success in obtaining **1640 RGB Imags, 391 RGB Point-cloud data, 10,206 Velodyne Point Cloud, 1628 Depth Image**.

Methodology

The methodology for generating a high-fidelity 3D model of a person from a single RGB image involves several steps. First, the image is processed using an object annotation algorithm to identify objects, followed by the application of the YOLO model to detect the presence of a person. Next, the image is passed through a PoseNet model to estimate the person's pose, generating a 2D skeleton of the person in the image. Finally, the isolated image of the person is passed through a PIFuHD model, which generates a 3D model of the person by reconstructing the 3D surface of the person based on the 2D image input.

Overall, this methodology combines several deep learning techniques, including object detection, pose estimation, and 3D modeling, to generate a high-fidelity 3D model of a person that can be used in various applications, such as virtual reality, gaming, and animation. By using a single RGB image, this methodology provides a practical solution for creating detailed 3D models of individuals, which has numerous potential applications in various fields.

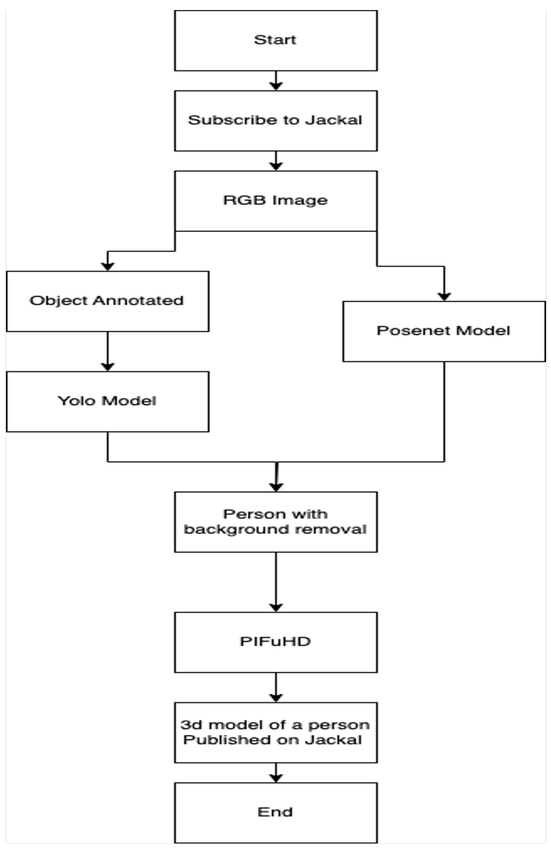

Figure 2. Methodology steps

Result

The accuracy of the YOLOv5 object detection model, PoseNet pose estimation model, and 3D PIFuHD model for generating 3D models of people can vary depending on several factors such as the training data, model architecture, and hyperparameters used. YOLOv5 is a state-of-the-art object detection model that has shown high accuracy on various datasets. PoseNet is also a widely used model for pose estimation that has shown high accuracy on various datasets. The 3D PIFuHD model is a state-of-the-art model for generating high-fidelity 3D models of people from 2D images. The accuracy of these models can be high when trained on suitable datasets and with appropriate hyperparameters but can also vary depending on the quality of the input image, the complexity of the pose, and other factors.

Figure 3. object and pose-net

Figure 4. mesh 3d model

Conclusions Future Work

The described methodology involves using deep learning models such as YOLOv5 for object detection, PoseNet for pose estimation, and PIFuHD for 3D surface reconstruction to generate a high-fidelity 3D model of a person from an RGB image. Although these models have shown high accuracy in various applications, further research is needed to improve their accuracy and robustness to challenges such as occlusion and complex poses. Future work in this area could focus on improving the texture recovery process, accurately positioning the 3D models in virtual environments, and creating virtual fashion try-on platforms.

Figure 5. future work

References

[1] Couturier, R., Noura, H. N., Salman, O., Sider, A. (2021). A Deep Learning Object Detection Method for an Efficient Clusters Initialization. arXiv preprint arXiv:2104.13634v3.

[2] Toshev, A., Szegedy, C. (2014). DeepPose: Human Pose Estimation via Deep Neural Networks. arXiv preprint arXiv:1312.4659v3.