Object Segmentation with Point Cloud for Autonomous Service Robot Manipulators

Cauan Vinicius Espinha de Sousa[1] and Plinio Thomaz Aquino-Junior[2]

***Abstract*— Robotics has established itself as a versatile and promising technology, finding applications in various fields, from industrial production to medical assistance in hospitals and homes. One of the most common tasks robots perform is manipulating objects, which can be carried out with precision and autonomy in unknown environments through object segmentation techniques. Object segmentation by point cloud is a widely used technique to reconstruct the geometry of objects in three dimensions (3D). This technique employs a set of points in a three-dimensional space to represent the surface of objects, allowing the autonomous robot to identify and locate objects in an unknown environment. In the proposed scientific initiation project, the point cloud approach will be explored for object segmentation and its application in manipulation by autonomous robots. The proposal is to develop a method that allows the robot to recognize different objects in an unknown environment and, based on the segmentation, determine the best way to manipulate them, considering their geometric characteristics. To this end, experimental analyses and computational simulations will be conducted to evaluate the efficiency and precision of the proposed method. The expected results include a robust and efficient point cloud object segmentation technique capable of recognizing and manipulating different types of objects in an unknown environment, thereby contributing to the advancement of autonomous service robotics.**

I. Introduction

Object segmentation with point clouds for autonomous service robot manipulators is a significant challenge in robotics. Object segmentation is identifying regions of interest in an image corresponding to real-world objects to enable a robot to manipulate them. Point clouds are a type of sensor used to capture three-dimensional information about the environment. They are often used in robotics because they allow the robot to obtain more accurate information about the scene than a traditional camera. Such detailed object information enables the robotic platform to use these data for decision-making regarding base positioning, object approach, and gripper opening, among other essential actions in task execution.

Object segmentation with point clouds is a process that involves extracting relevant information from a point cloud data set. This information can include the shape, position, and orientation of objects. Segmentation can be performed using various techniques, such as analysis of colors, textures, and shapes.

[1]Cauan Vinicius Espinha de Sousa, Centro Universitario FEI, Department of Computer Science, S. Bernardo do Campo, SP, Brazil.

[2]Plinio Thomaz Aquino-Junior, Centro Universitario FEI, Department of Computer Science, S. Bernardo do Campo, SP, Brazil. plinio.aquino@fei.edu.br

There are several approaches to object segmentation with point clouds. A common approach is region-based segmentation, which divides the scene into regions corresponding to different objects. Another approach is edge-based segmentation, which detects object edges in the scene and uses these edges to segment the objects.

Object segmentation with point clouds is essential for many robotic applications, such as object manipulation in factories and automated warehouses, cleaning of hazardous environments, and applications in healthcare assistance and home services. Once objects are segmented and identified, the robot can use this information to safely and efficiently perform specific tasks, such as picking up, moving, or stacking objects.

II. Application Context

The ability of robots to interact and manipulate objects in unstructured environments marks a milestone for robotic autonomy and its practical applicability, especially in the services domain. This investigation explores advanced object recognition and manipulation techniques through point cloud analysis, representing a significant step in this direction. This study promises to contribute to the scientific community regarding how robots perceive and interact with their surroundings, empowering them to perform manipulation tasks with precision and adaptability.

The essence of this study lies in its innovative approach to object segmentation, a process that enables robots to identify and locate three-dimensional objects in unknown spaces. Such capability is fundamental for developing robotic systems capable of operating autonomously in various environments, from automated factories to residences. By enhancing the precision and efficiency with which robots identify and manipulate objects, this research advances the field of service robotics. It sets new standards for robotic interaction in complex scenarios.

The relevance of this work to the development of robotic platforms is broad and multidimensional. By providing a methodology for object segmentation in unknown environments, the research directly addresses one of the biggest challenges faced by service robotics: adaptation to dynamic and unpredictable environments. This adaptability is crucial for expanding robot usage in daily tasks and opening new possibilities for applications in home assistance, industrial maintenance, and even healthcare.

The applicability and relevance of this research are exemplified through its integration with the HERA robot [1], developed by the RoboFEI@Home team in the context of

the RoboCup@Home competition [2]. HERA represents an advanced robotic platform designed to meet the challenges of service robotics in domestic and public environments. The implementation of object segmentation with point clouds on this platform demonstrates the technical feasibility of the research. It highlights the potential of such technologies to improve service robots' functionality and efficacy significantly. With HERA, the RoboFEI@Home team not only tests and validates the proposed methods in real scenarios but also provides valuable feedback for the refinement and evolution of these techniques as a continuation of various research efforts that motivate the challenges mapped in this project [3], [4], [5].

III. Fundamental Concepts

This section presents the theoretical concepts that underpin the work, considering topics such as point clouds, object segmentation, neural networks, advanced simulation, robotics, and synthetic datasets.

A. Point Cloud

Point clouds are a three-dimensional representation of objects and scenes formed by many points obtained by sensors such as 3D cameras and laser scanners. They are widely used in computer vision, robotics, augmented reality, and gaming applications. In computer vision, point clouds are used in the 3D reconstruction of objects from 2D images, while in robotics, they are employed for navigation in unknown environments. Heyden [6] provides an in-depth look at 3D reconstruction techniques, image registration, and object detection using point clouds.

B. Object Segmentation

Object segmentation is an essential process in computer vision and robotics involving separating the objects of interest in an image or scene for individual analysis. The image quality and the scene's complexity affect the efficiency of segmentation, and various image processing and computer vision techniques are used to identify the characteristics of objects. In Gonzalez [7], one can find object segmentation techniques and examples of applications in various areas.

C. Convolutional Neural Network - CNN

Convolutional Neural Networks (CNNs) represent an advanced category of deep neural networks, particularly effective in processing images and point clouds. A central feature of CNNs lies in their convolutional layers, which are crucial in extracting significant spatial features from these data. These layers use filters, or kernels, that move over the image or point cloud, analyzing small portions at a time. By applying these convolution operations, CNNs can identify patterns such as edges, textures, and other relevant geometric elements.

This feature extraction process occurs hierarchically: the initial convolutional layers may capture basic features, while subsequent layers integrate this information to identify increasingly complex elements. This approach enables CNNs to construct a detailed understanding of the spatial structure of objects contained in images or point clouds, facilitating tasks such as object recognition, segmentation, and classification.

The importance and efficacy of CNNs in this context are extensively documented, with GoogleFellow cited in [8] being an essential reference in studying deep learning. This work offers a comprehensive view of the techniques used in the field, with particular emphasis on CNNs. It highlights their fundamental role in the revolution of image processing and spatial data analysis through machine learning.

D. Advanced Simulation with NVIDIA Omniverse

NVIDIA Omniverse represents a significant evolution in robot simulation, offering high visual and physical realism. With advanced features like real-time ray tracing, this platform is ideal for simulations requiring precision and detail. Moreover, Omniverse facilitates collaboration and integration with other design, simulation platforms, and robotics frameworks like ROS, becoming a valuable tool for complex robotics and automation projects. The ability to simulate detailed and realistic interactions with the environment opens new possibilities for testing and developing advanced robotic systems.

E. Robotics and Automation

To advance in the field of research in object segmentation using point clouds in service robot manipulators, it is essential to have a solid foundation in various areas of robotics and automation. These areas include, but are not limited to, robot kinematics and control, which provide the fundamentals for understanding how robots move and how to control these movements precisely. Moreover, a deep understanding of perception systems and the integration of sensors and actuators is crucial, allowing robots to interact with their environment efficiently.

A thorough knowledge of motion control systems is also vital, as the precision in object manipulation depends on the robot's ability to perform complex movements and accurately adjust its position. Research in object segmentation with point clouds, specifically, requires a robust understanding of these concepts and skills in computer vision and image processing. This knowledge enables researchers to develop algorithms to interpret three-dimensional data and identify objects within a space. In Corke [9], the essential principles of robotics and computer vision are addressed, considering everything from kinematic analysis to motion planning and real-time control, providing a comprehensive overview of the techniques and theories necessary for developing advanced robotic systems.

F. Synthetic Dataset

Synthetic datasets, artificially generated through algorithms or simulations, aim to replicate characteristics of accurate data for use in training, testing, and validating machine learning models, as well as offering ethical alternatives for handling sensitive data and mitigating privacy

issues. These sets are created using statistical models and simulations to advanced artificial intelligence techniques, such as Generative Adversarial Networks (GANs), allowing for broad application in areas ranging from software development to scientific research. Although they offer solutions for the scarcity of accurate data and ethical issues, they face challenges related to realism, representativeness, and the inadvertent introduction of bias, requiring critical use to ensure their effectiveness and accuracy.

IV. Related Works

Tang [10] proposes a new unsupervised learning method for point cloud segmentation. The method consists of learning the boundaries between different classes of points using contrastive learning. The algorithm trains a model to predict whether two points belong to the same class based on their spatial proximity and point features. Experiments show that the method is effective and outperforms other point cloud segmentation methods.

In [11], a new model based on the *Transformer* for processing 3D point clouds is presented. The model is designed to extract contextual information at different scales by dividing the point cloud into strata and applying a stratified *Transformer* layer to each stratum. The model also uses multi-head attention mechanisms to deal with point density variability. Experiments show that the model is effective in the classification and segmentation tasks of 3D point clouds.

Cheng's work [12] introduces a new method for semantic segmentation of 3D point clouds using semi-supervised learning. The method proposes a new parameter transfer technique that allows the model to leverage information from a small amount of labeled data to improve performance on unlabeled data. The method also uses a point transformation network to increase the model's robustness to rotation and translation variations. Experiments show that the method is effective in 3D point cloud segmentation tasks with few labeled data.

In Qin [13], a new method for training a reinforcement learning agent for dexterous object manipulation in a simulated environment is verified. The method uses point cloud information to represent the environment configuration, the object position, and the agent's action. The algorithm uses policy optimization techniques to learn a general policy that can be transferred from the simulated environment to the real world. Experiments show that the method effectively performs object manipulation tasks in simulated and real-world environments.

In [14], a new neural network architecture for 3D point cloud tasks, such as classification and semantic segmentation, is proposed. The architecture, called SPH3D-GCN, uses a combination of convolutional filters and spherical point transformations to represent the point cloud geometry locally. Additionally, the method proposes a new batch normalization scheme to improve training stability. Experiments show that the SPH3D-GCN architecture outperforms other state-of-the-art architectures on different datasets and 3D point cloud tasks.

Meanwhile, Wang [15] introduces a new convolutional neural network (CNN) architecture for learning in three-dimensional point clouds. The proposed approach can effectively learn the features of dynamic point clouds. The network is based on a dynamic graph model updated at each layer to capture variable neighborhood information. Experimental results demonstrated that the proposed approach can outperform other point cloud learning techniques regarding accuracy and computational efficiency.

The research [16] highlighted in NVIDIA's "Eureka" project reveals an innovative approach to robot training, utilizing AI-generated reward programs that outperform those written by humans in over 80 percent of tasks. This method, benefiting from trial-and-error learning, results in an average performance improvement of over 50 percent for robots. Utilizing GPU-accelerated simulation in Isaac Gym and visual integration with NVIDIA Omniverse, Eureka efficiently evaluates large sets of reward candidates, continuously improving AI's generation of reward functions. This work presents significant advances in robotic autonomy. It opens new possibilities for creating physically realistic animations, demonstrating the potential of AI-accelerated simulation technologies and the Omniverse in transforming robotic development and digital content production.

This preliminary study verifies that various works have been conducted emphasizing point cloud segmentation, serving as a study object for advancing this proposal. Additional related works will be explored during the execution of this project.

V. Material Resources

For the development of this research project, a specific set of components was necessary to enable the segmentation of objects by point cloud and its application in manipulation by autonomous robots. Firstly, using a computer with an adequate graphics card for processing the required computer vision is essential. The Jetson Orin is a graphics card widely used in robotics projects, as it offers high performance in embedded processing and reduced energy consumption.

Another fundamental component is the use of a depth camera, such as RealSense, capable of capturing three-dimensional images and generating the necessary point cloud for the segmentation and detection of objects. It is essential to highlight that a standard RGB camera cannot provide such information, making RealSense a vital component of the project's effectiveness.

Additionally, it is essential to have a manipulator capable of interacting with the segmented and detected objects by the system. This manipulator must be designed to execute the necessary movements to manipulate the objects precisely and safely, ensuring the task's efficacy. The manipulator to be used in this project was developed by the RoboFEI@Home Team, utilizing the HERA robotic platform [1], [2].

Finally, combining these components with the proposed knowledge and techniques enables advanced research in point cloud processing and computer vision, with practical applications in various scenarios, such as industrial process

automation and mobile robotics. It is important to emphasize that using these components will provide greater efficiency and safety in the tasks performed by the robot, bringing benefits to the industry and other sectors that utilize robotic technologies.

VI. Methodology

The proposed methodology begins with creating a synthetic dataset using the Omniverse Replicator. This advanced tool is designed to generate three-dimensional environments that reflect, with high fidelity, the graphical and physical representations of the real world. Through the Replicator, it will be possible to accurately simulate essential details of the environment, such as textures, lighting, and the physics of objects.

To enrich the diversity and quality of the generated data, specific code will be developed to randomize various aspects of the environment. This randomization includes the dynamic alteration of the objects in the scene, lighting conditions, and even the position and angle of the capture cameras. Such an approach ensures a wide variety of data, which is essential for the robust training of artificial intelligence models.

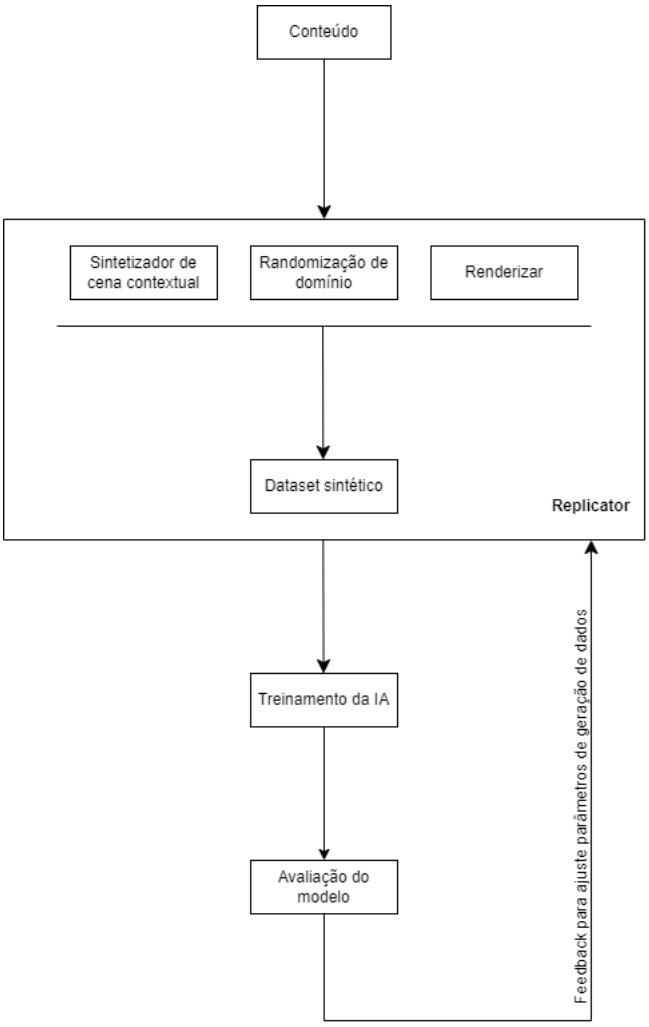

Fig. 1. Methodology Using Replicator

After generating and properly annotating the data, the next step involves training a convolutional neural network (CNN). This network will be meticulously trained with the produced synthetic dataset, enabling it to detect and segment objects accurately in real-time.

With the segmentation completed, the process moves to the motion planning stage, where the motion plan is calculated to determine the best trajectory for the robot to manipulate the object safely and efficiently. This planning considers the manipulator's physical characteristics, previously described, to determine whether the object can or cannot be manipulated by the hardware in question.

Experimental analyses and computational simulations are conducted to evaluate the proposed method's efficiency and precision, considering different types of objects, environments, and lighting conditions. This process is crucial to assess the robustness and generalization of the proposed algorithm.

Furthermore, great importance is placed on producing quality research documentation, with version-controlled reports and code shared and documented on Github, thus facilitating the monitoring and replicability of the research.

Finally, all research findings are shared and tested on the HERA robotic platform of the RoboFEI@Home Team. This final step aims to validate the effectiveness of the proposed methodology in practical scenarios, thereby contributing to the advancement of knowledge in the field of point cloud processing and, enabling object manipulation in various scenarios, optimizing processes, and increasing the efficiency of autonomous robotic systems.

VII. Preliminary Results

To effectively implement the "Replicator" framework, sophisticated code incorporating a wide range of 3D models representing different types of bushes and vegetation was developed. This set is enriched with diverse textures meticulously chosen to exemplify potential applications. The analysis focuses on the human figure strategically positioned within the scene.

In the initial scenario, a static light source was chosen. However, an opportunity for enhancement was recognized by introducing a dynamic lighting system. This evolution would allow for a more realistic simulation, adapting to the variable movement of objects within the scene, similar to the flexibility already implemented in their positioning.

Example of defining random and static objects:

```
def worker():
    worker = rep.create.from_usd(WORKER,
        semantics=[('class', 'worker')])

    with worker:
        rep.modify.pose(
            position=rep.distribution.
                uniform((-500, 0, -500),
                (500, 0, 500)),
            rotation=rep.distribution.
                uniform((-90,-45, 0),
                (-90, 45, 0)),
        )
```

```
        return worker

    rep.randomizer.register(env_props)
    rep.randomizer.register(worker)

    env = rep.create.from_usd(ENVS)
    surface = rep.create.from_usd(SURFACE)

    camera = rep.create.camera(
        focus_distance=800,
        f_stop=0.5
    )
    render_product = rep.create.
        render_product(camera, resolution
        =(899, 899))
```

Listing 1. Exemplo de definição de elementos estáticos e randômicos

One of the most significant contributions of this method is its scalability in data production. Remarkably, it eliminates the need for manual annotations, whether through detailed segmentations or bounding box delimitations. An exeperient labaler take four hours to label a thousand images, otherwise with framework's engine it would take twenty minutes, because the Replicator autonomously generates these metadata, almost one image labeled per second, as the graphic 5 exemples. Freeing users from the burdensome manual labor that, until recently, was an indispensable step, regardless of the data collection methodology. This innovation democratizes access to a vast volume of data and eliminates the requirement for laborious manual processes previously mandatory in data preparation for analysis.

Fig. 2. The output of segmentation from step 1

The synthetic dataset generated so far has proved to be a valuable tool for the preliminary training of convolutional neural networks. However, a detailed analysis revealed some limitations, particularly in the variety and complexity of environmental conditions and the realistic representation of objects. For example, the variation in lighting and textures still needs to fully capture the range of scenarios encountered in natural environments.

To address these limitations, efforts have been concentrated on enhancing the randomization code. This code is critical in generating a wider diversity of scenarios and is crucial for effectively training artificial intelligence models. We are introducing more advanced algorithms to simulate a broader range of variations.

Additionally, efforts are concentrated on refining the textures and physical properties of objects in the dataset. This includes adjustments in reflectance, roughness, and other material properties to increase the visual authenticity of objects. These improvements are iteratively tested to assess the impact on the performance of neural network models, focusing on increasing the accuracy and generalization capacity of detection and segmentation algorithms.

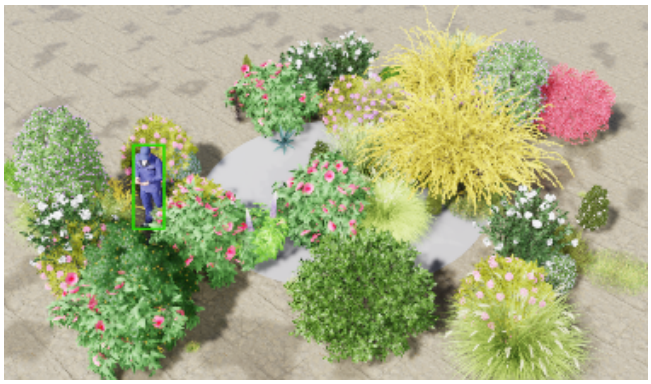

Fig. 3. The output of 2D marking from Step 1

The following steps include a closer integration of sensory feedback into the simulation process, aiming to generate data that looks realistic and behaves consistently with the laws of physics. This is crucial for applications in robotics, where interaction with natural objects is unpredictable and complex.

The preliminary results indicate significant progress in creating a robust and realistic synthetic dataset for simulating robots in 3D environments. The continuous enhancement of the randomization code and the refined representation of environmental elements are essential steps toward achieving a neural network training model that can be effectively applied to real-world scenarios. This iterative process of evaluation and continuous improvement is fundamental to the success of the research.

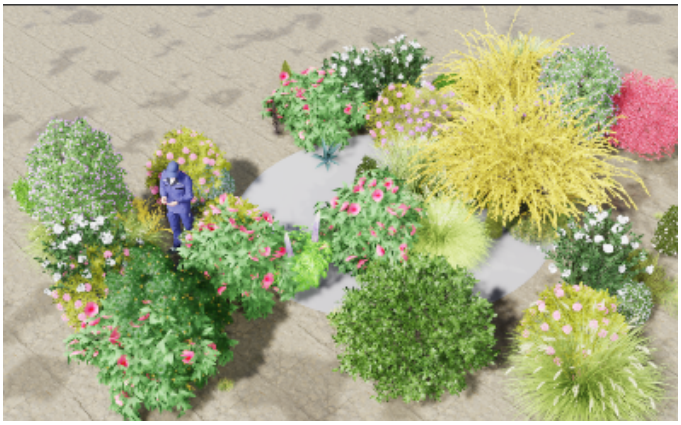

Fig. 4. The output of the camera from Step 1

VIII. Preliminary Conclusions

The preliminary conclusions obtained at this project stage offer a bifurcated view, demonstrating both the proposed

methodology's promising potential and imminent limitations. The success of the plane/object segmentation approach in simulated environments highlights the theoretical solidity of the method, outlining a promising path for the autonomous identification and manipulation of objects by robots. However, the decreased performance observed in natural environments underscores a critical area that requires methodological refinement.

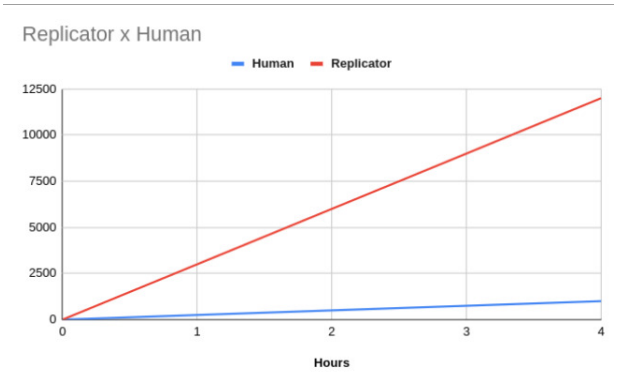

Fig. 5. Comparative between images labeled by human and replicator per hour

The primary difficulty in dealing with oscillatory and high-magnitude noise illustrates the need to develop robust noise filtering strategies and, possibly, enhanced segmentation algorithms that can be resilient to the uncertainties and complexities inherent in natural environments. The disparity between simulated and actual results also suggests the need for a more thorough evaluation of the assumptions and parameters employed in the current methodology.

Furthermore, the preliminary conclusions point to the importance of exploring machine learning algorithms and other advanced techniques that may offer better robustness against adversities encountered in real-world scenarios. Integrating additional sensory modalities and using a more diverse set of sensory data could be valuable guidelines for enhancing the robustness and accuracy of segmentation.

IX. Future Works

Building on this research, it's important to overcome the current methodology's shortcomings, particularly the challenge of noise in natural settings. Future efforts will focus on developing a new detection technique using edge identification enhanced by machine learning. This approach aims to provide more accurate and noise-resilient segmentation by leveraging machine learning's ability to identify patterns and adapt to data variations, enhancing object detection in complex environments.

Integrating edge detection with machine learning addresses previous challenges, aiming to create a segmentation system that excels in simulations and real-world conditions where uncontrolled variables like noise are prevalent. This more sophisticated system is expected to significantly enhance segmentation precision, boosting the robot's capability for effective autonomous object manipulation.

Additionally, comprehensive experiments and evaluations in real settings will be conducted to validate the new method's effectiveness and robustness. These tests confirm system performance across various scenarios before practical deployment.

References

[1] P. T. A. Junior, B. d. F. V. Perez, R. Meneghetti, F. d. A. M. Pimentel, G. N. Marostica, J. G. R. Amorim, L. C. Neves, L. I. Gazignato, M. Y. Gonbata, R. Souza, *et al.*, "Hera: Home environment robot assistant," in *II BRAHUR and III Brazilian workshop on service robotics*, 2019.

[2] G. Nicolau Marostica, N. A. Grotti Meireles Aguiar, F. d. A. Moura Pimentel, and P. T. Aquino-Junior, "Robofei@home: Winning team of the robocup@home open platform league 2022," in *RoboCup 2022: Robot World Cup XXV*, A. Eguchi, N. Lau, M. Paetzel-Prüsmann, and T. Wanichanon, Eds. Cham: Springer International Publishing, 2023, pp. 325–336.

[3] F. Pimentel and P. Aquino, "Performance evaluation of ros local trajectory planning algorithms to social navigation," in *2019 Latin American Robotics Symposium (LARS), 2019 Brazilian Symposium on Robotics (SBR) and 2019 Workshop on Robotics in Education (WRE)*, 2019, pp. 156–161.

[4] F. d. A. M. Pimentel and P. T. Aquino-Jr, "Evaluation of ros navigation stack for social navigation in simulated environments," *Journal of Intelligent & Robotic Systems*, vol. 102, no. 4, p. 87, Jul 2021.

[5] F. Pimentel and P. Aquino-Jr, "Study on the comfort of people in spatial interactions with a social robot," in *Anais Estendidos do XIV Simpósio Brasileiro de Robótica e XIX Simpósio Latino-Americano de Robótica*. Porto Alegre, RS, Brasil: SBC, 2022, pp. 13–24.

[6] A. Heyden, F. Kahl, K. Åström, G. Sparr, and G. Sanniti di Baja, *3D Computer Vision: Efficient Methods and Applications*. Springer Science & Business Media, 2009.

[7] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Pearson Education, 2008.

[8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[9] P. Corke, *Robotics, Vision and Control: Fundamental Algorithms in MATLAB*. Springer Science & Business Media, 2011.

[10] L. Tang, Y. Zhan, Z. Chen, B. Yu, and D. Tao, "Contrastive boundary learning for point cloud segmentation," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8479–8489, 2022.

[11] X. Lai, J. Liu, L. Jiang, L. Wang, H. Zhao, S. Liu, X. Qi, and J. Jia, "Stratified transformer for 3d point cloud segmentation," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8490–8499, 2022.

[12] M. Cheng, L. Hui, J. Xie, and J. Yang, "Sspc-net: Semi-supervised semantic 3d point cloud segmentation network," *ArXiv*, vol. abs/2104.07861, 2021.

[13] Y. Qin, B. Huang, Z.-H. Yin, H. Su, and X. Wang, "Dexpoint: Generalizable point cloud reinforcement learning for sim-to-real dexterous manipulation," in *Conference on Robot Learning*, 2022.

[14] X. Ma, C. Qin, H. You, H. Ran, and Y. R. Fu, "Rethinking network design and local geometry in point cloud: A simple residual mlp framework," *ArXiv*, vol. abs/2202.07123, 2022.

[15] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph cnn for learning on point clouds," *ACM Transactions on Graphics (TOG)*, vol. 38, pp. 1 – 12, 2018.

[16] L. J. Fan *et al.*, "Eureka: A research breakthrough in robot learning with generative ai and omniverse," *NVIDIA Blog*, 2023. [Online]. Available: https://blogs.nvidia.com/blog/2023/10/03/eureka-robot-learning/