Scope
Volume 13 Number 02 June 2023

A Review of Object Grasp Prediction Techniquesand Grasp Net

Vickram karthick R[1], Dr. Sunil Sikka[1], & Nisu Sethi[1]

Amity University Haryana

Abstract. This paper discusses about significance of object grasping research and its potential to bridge the gap between perception and ac- tion, enabling machines to perceive and understand their surroundings and perform physical interactions with objects. By advancing the state- of-the-art in object grasping, researchers aim to overcome challenges such as occlusions, cluttered environments, object variability, and uncertain sensory data. Effective object grasping algorithms and systems have the potential to revolutionize automation, robotics, and human-robot inter- action, leading to improved productivity, safety, and efficiency in varioussectors.

Keywords: Object grasping, Grasp poses, Clutter

1 Introduction

Object grasping is of significant importance in computer vision and robotics, en- abling machines to interact with and manipulate objects in the physical world. Through the development of accurate grasp prediction algorithms, researchers aim to enhance automation, robotic manipulation, and human-robot collabo- ration, leading to advancements in various industries and domains. GraspNet serves as a valuable resource in this pursuit, providing a standardized bench- mark to evaluate and improve object grasping capabilities. Object grasping plays a crucial role in the fields of computer vision and robotics, addressing the fun- damental challenge of enabling machines to interact with and manipulate the physical world. Grasping involves the ability to perceive objects, plan appro- priate hand movements, and execute precise control to achieve a stable and effective grip. In computer vision, object grasping aims to develop algorithms and techniques that enable machines to understand and interpret the 3D ge- ometry, shape, and appearance of objects in a scene. By accurately estimating the pose and location of objects, computer vision systems can facilitate object recognition, scene understanding, and robotic manipulation tasks. Grasping is a fundamental step toward achieving more complex tasks such as object manipulation, assembly, and pick-and-place operations.[1] In the field of robotics, object grasping is essential for robots to interact with the physical world, perform tasks autonomously, and assist humans in various domains. Robots equipped with grasping capabilities can handle objects, manipulate tools, and perform intricate operations in industrial automation, household chores, healthcare, agriculture,

and other areas. Grasping enables robots to exhibit dexterity, adaptability, and efficiency, revolutionizing industries and improving quality of life.[2] GraspNet provides researchers with a standardized evaluation platform, enabling the com- parison of different approaches and fostering collaboration and innovation in the field. The availability of large-scale, diverse, and annotated datasets like Grasp-Net facilitates the training and evaluation of grasp prediction models, promoting advancements in perception, learning-based techniques, and real-world robotic applications.[3]

2 Working Principles of Grasp Net

Grasp Net offers a rich and comprehensive resource for studying object grasping in diverse scenarios. The inclusion of object models, scenes, and detailed grasp annotations enables researchers to explore different grasp planning algorithms, evaluate grasp stability, and develop learning-based approaches. The dataset's structure promotes standardized evaluation and fosters innovation in the field of object grasping research[4] The data-set structure and contents of GraspNetare essential aspects that contribute to its comprehensiveness and effectivenessas a part of object grasping research. GraspNet provides researcher's with a diverse collection of 3D object models, scenes, and associated grasp annotations. This section provides a detailed overview of the dataset's structure and its key components.

Object Models

GraspNet includes a wide range of object models representing various categories, such as household items, tools, and industrial objects. These object models are represented in 3D format, typically in mesh or point cloud representations. Each object model is accompanied by metadata, including category labels, object di- mensions, and other relevant information which are useful for getting accurateresults.[5][6]

Scenes

GraspNet comprises a large number of scenes that simulate real-world environ- ments for object grasping. These scenes consist of the object models placed in different configurations, orientations, and cluttered scenarios and were captured with variations in lighting conditions, background settings, and object place- ments to enhance the dataset's diversity and realism.[4]

Grasp Annotations

One of the primary components of GraspNet is the grasp annotations that pro- vide ground truth information about object grasping. Grasp annotations include the position and orientation of the gripper relative to the object, known as the grasp pose. Annotations are present for both stable and force-closed grasps, allowing researchers to study different types of grasping strategies. The anno- tations may also include additional information such as contact points, grasp quality metrics, and stability evaluations.

Grasp Stability Evaluation

GraspNet incorporates techniques to evaluate the stability of grasps within the dataset. Stability evaluations assess the likelihood of a grasp being successful interms of maintaining a stable grip on the object. Various factors such as contact forces, friction coefficients, and object geometry may be considered in evaluating grasp stability. Dataset can be altered for including custom values pertaining tothe tools being used.

Grasp Quality Metrics

GraspNet defines evaluation metrics to quantify the quality and effectiveness of grasps. These metrics provide a standardized way to measure the success of different grasping algorithms and techniques. Grasp quality metrics can include measures of stability, force closure, contact quality, or a combination of these factors.

3 Grasp prediction algorithms

Grasp prediction algorithms play a crucial role in the field of robotic grasping, enabling robots to plan and execute effective grasps on objects in various sce- narios. These algorithms aim to determine the optimal grasp configuration that ensures stable and successful grasping. This section explores some of the promi- nent grasp prediction algorithms that have been developed and utilized in the research community. These algorithms utilize different approaches, including an- alytical methods, learning-based techniques, and optimization strategies.Grasp prediction algorithms continue to evolve, driven by advancements in machine learning, computer vision, and robotics. Further future research aims to address challenges such as generalization to unseen objects and environments, incorpo- rating tactile feedback, and improving real-time grasp planning and execution. By developing more accurate and robust grasp prediction algorithms the capa- bilities of robotic systems in tasks requiring object manipulation and interaction with the physical world can be enhanced. [7][8]

Analytical Grasp Prediction

Analytical methods formulate grasp prediction as an optimization problem, aim- ing to find the optimal grasp parameters that maximize stability and force clo- sure. One such popular approach is the Grasp Quality Measures (GQM), which defines analytical metrics to evaluate grasp quality based on contact forces, fric- tion, and stability criteria. Another commonly used analytical method is the Grasp Wrench Space (GWS) analysis, which characterizes the set of wrenches that a grasp can resist without slipping or losing stability. Analytical methods provide insights into grasp stability and can be computationally efficient, but they often rely on simplified assumptions and may struggle with complex objectgeometries and uncertain environments.[9][10]

Learning-Based Grasp Prediction

Learning-based approaches leverage machine learning techniques to predict grasps based on training data. Convolutional Neural Networks (CNNs) have been widely used for learning-based grasp prediction, where the network learns to map visual input (e.g., RGB or depth images) to grasp parameters. Point cloud-based approaches use PointNet or PointNet++ architectures to process 3D point cloud data and predict grasp configurations. Reinforcement Learning (RL) methods have also been employed for grasp prediction, where an agent learns to interactwith the environment and optimize grasping actions through trial and error. Learning-based approaches have shown promising results, especially in complex and unstructured environments, but they typically require large amounts of la- beled training data and may suffer from generalization issues.[11][12]

Hybrid Approaches

Hybrid approaches combine multiple techniques, such as analytical methods, learning-based models, and optimization strategies, to benefit from their respec- tive strengths. For instance, a hybrid approach may use learning-based modelsto provide initial grasp candidates and then apply optimization techniques to re- fine and optimize the grasps. These approaches aim to leverage the advantagesof different methods to improve grasp prediction accuracy and robustness.[13]

4 Grasp quality assessment

Grasp quality assessment is a critical aspect of robotic grasping that aims to evaluate the effectiveness and reliability of a grasp configuration. Assessing grasp quality allows robotic systems to select and execute grasps that are stable, force-closed, and capable of performing the desired manipulation tasks. This sectionexplores various approaches and metrics used for grasp quality assessment. As- sessing the quality of a grasp is a challenging undertaking that entails taking into account various elements such as stability, force closure, robustness, and the specific requirements of the task at hand. The selection of evaluation metrics and approaches relies on the particular application and the desired characteristics of the grasp. Precise evaluation of grasp quality enables robotic systems to choosedependable grasps, enhance their manipulation capabilities, and improve overall performance across a range of tasks.

Grasp Stability Metrics

Grasp stability metrics are which evaluate the ability of a grasp to resist ex- ternal perturbations and maintain a stable grip on the object. One commonly used metric is the center of mass (CoM) wrench, which measures the maximum wrench that can be applied to the object before the grasp loses stability. The robustness index quantifies the magnitude of disturbances that a grasp can tol-erate while remaining stable. Other stability metrics include the wrench space volume, which measures the space of external wrenches that the grasp can resistwithout slipping or losing contact.

Force-Closure Metrics

Force-closure metrics assess the ability of a grasp to achieve force closure, where the contact forces applied by the gripper result in a net closing force on the ob-ject.The wrench-based force-closure measure quantifies the magnitude of forces that can be applied without inducing object motion. Other force-closure metricsinclude the grasp wrench space volume, which evaluates the space of contact forces that can maintain force closure.

Grasp Robustness

Grasp robustness metrics evaluate the ability of a grasp to tolerate uncertainties and variations in object pose, shape, and other factors. The robustness metric quantifies how sensitive a grasp is to changes in the object's position and orien-tation.Monte Carlo simulations and perturbation-based analysis are often usedto assess grasp robustness by sampling various object poses and measuring thegrasp's success rate.

Grasp Quality Measures

Grasp quality measures aim to provide a single scalar value that represents the overall quality of a grasp configuration. This quality measures combine multiple factors, such as stability, force closure, and robustness, into a unified metric. Examples of grasp quality measures include the Grasp Quality Measures (GQM)framework, which combines stability, force closure, and other criteria into a single value. Other measures, such as the Power Grasp Quality Measure (PGQM), incorporate physical properties and geometric features of the object to assess grasp quality.

Learning-Based Grasp Quality Assessment

Learning-based approaches utilize machine learning techniques to predict grasp quality based on training data. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have been employed to learn grasp quality directly from visual input or grasp configurations. These approaches leverage large datasets with labeled grasp quality information to train models that can predict grasp quality for new grasp configurations.

5 Task-oriented grasp evaluation

Task-oriented grasp evaluation is a crucial aspect of robotic grasping that fo- cuses on assessing the suitability of a grasp configuration for specific manipu- lation tasks. It goes beyond traditional grasp quality assessment by considering the grasp's effectiveness in achieving task-related objectives. Learning-based ap- proaches have revolutionized the field of robotic grasping by leveraging the power of machine learning algorithms to improve grasp planning and execution. These approaches utilize large datasets of labeled grasp examples to train models thatcan predict and generate effective grasps for a wide range of objects and sce- narios. This section explores various learning-based approaches used in robotic grasping.

Convolutional Neural Networks

Convolutional Neural Networks have been widely used in learning-based grasp planning and prediction.CNNs are capable of processing visual input, such as RGB or depth images, and learning complex patterns and features relevant to grasping. These networks are trained on large datasets of labeled grasp examples, where the input images are associated with corresponding grasp configurations. The trained CNN models can then be used to predict grasp poses or generate grasp proposals for new objects based on their visual representations.

Point Cloud-Based Approaches

Point cloud data, obtained from depth sensors or 3D scanners, provides rich ge-ometric information about objects, which is crucial for grasp planning. PointNet and PointNet++ architectures have been widely adopted for processing point cloud data in learning-based grasp prediction tasks. These architectures can capture local and global geometric features of objects, allowing the model to learn grasp-relevant patterns from the point cloud data. By training on labeledpoint cloud data with associated grasp configurations, these models can predictgrasp poses or generate grasp proposals directly from point cloud inputs.

Reinforcement Learning

Reinforcement Learning(RL) techniques have been applied to grasp planning, where an agent learns to interact with the environment and optimize grasping actions through trial and error. RL methods typically involve an agent, a re- ward function, and a policy network. The agent explores different grasp actions,receives feedback in the form of rewards based on grasp success or failure, and adjusts its policy network to maximize long-term cumulative rewards. RL-basedgrasp planning allows the agent to learn complex grasping strategies and adaptto varying object shapes and environments.[14]

Generative Adversarial Networks

Generative Adversarial Networks (GANs) have been utilized for generating grasp proposals or augmenting existing grasp datasets. GANs consist of a generator network and a discriminator network, which compete against each other in a training process. The generator network generates synthetic grasp configura- tions, while the discriminator network distinguishes between real and synthetic grasps.By training the GAN on real grasp data, the generator network can learn to generate realistic and diverse grasp proposals, augmenting the training dataset for other learning-based approaches.

Transfer Learning and Domain Adaptation

Transfer learning techniques have been applied to leverage pre-trained models on large-scale datasets for grasp planning tasks with limited labeled data. By fine-tuning pre-trained models on a smaller labeled dataset specific to the target task, the models can effectively generalize to new objects and scenarios. Domain adaptation techniques aim to adapt grasp models trained on one domain (e.g., synthetic data) to perform well in a different domain (e.g., real-world data) by minimizing the domain gap.

Task Specification

Task-oriented grasp evaluation begins with specifying the manipulation task or objective that the robot needs to accomplish.The task can vary depending on the application, such as picking and placing objects, manipulating tools, or performing assembly tasks. The specification of the task provides guidance forevaluating grasp configurations based on their ability to achieve the desired taskoutcomes.

Task Relevance Metrics

Task relevance metrics quantify the extent to which a grasp configuration aligns with the requirements of the manipulation task.These metrics consider factors such as object pose, object properties, contact forces, joint torques, and task- specific constraints. For example, a metric may evaluate how well the grasp enables a robot to manipulate an object into a desired configuration or complete a specific action sequence.Task relevance metrics are typically application-specific and designed to capture the essential aspects of the task at hand.

6 Active vision-based object localization

Active vision-based object localization is a process in computer vision and robotics that involves actively selecting and controlling the viewpoint of a camera or sen- sor system to improve the accuracy and efficiency of object localization. Unlike passive vision-based methods that rely on fixed or pre-determined viewpoints, active vision systems dynamically adjust their viewpoint to gather more informa- tive data and reduce uncertainty in object localization.Active vision-based object localization techniques have shown promising results in various applications, in- cluding robot manipulation, augmented reality, or autonomous navigation. By actively controlling the viewpoint and gathering informative data, these systems improve the efficiency and accuracy of object localization, particularly in challenging scenarios with occlusions, cluttered environments, or ambiguous object appearances. As research in active vision continues to advance, incorporating more sophisticated algorithms and sensor technologies, the capabilities of active vision-based object localization are expected to further enhance, enabling robots and autonomous systems to interact with the environment more effectively.This section will will explore the concept of active vision-based object localization and discuss various techniques used in this area.[15][16]

Viewpoint Selection

Viewpoint selection is a crucial aspect of active vision-based object localiza- tion, where the system decides on the optimal camera viewpoint for capturing informative images. This selection is typically guided by specific criteria, such as reducing uncertainty, maximizing information gain, or minimizing the ex- pected localization error. Different strategies can be employed for viewpoint se- lection, including heuristic-based methods, exploration-exploitation algorithms,or Bayesian optimization techniques.

Uncertainty Reduction

Active vision systems aim to reduce uncertainty in object localization by select- ing viewpoints that provide the most informative data. Uncertainty can arise due to factors such as occlusions, viewpoint limitations, or ambiguous object appear- ances. Methods such as uncertainty sampling, Bayesian inference, or entropy- based measures can be used to quantify and prioritize uncertain regions for exploration

Feature Selection

Active vision systems can actively select and extract relevant features from the acquired data to improve object localization. These features can include edges,corners, keypoints, or texture descriptors that are informative for object recog- nition and localization. Feature selection can be performed based on factors such as saliency, distinctiveness, or discriminative power.

Sensor Planning

Active vision-based object localization can involve planning the motion of the camera or sensor system to acquire multiple views of the object. Sensor plan- ning algorithms optimize the trajectory of the camera or sensor to minimize the number of viewpoints required for accurate localization. Techniques such as view planning, sensor placement optimization, or motion planning algorithms can be employed for efficient sensor motion.

Feedback and Iteration

Active vision-based object localization often involves an iterative process wherethe system updates its belief about the object's location based on acquired dataand continuously refines the estimation. Feedback mechanisms, such as online learning or adaptive sampling, can be incorporated to adjust the viewpoint se- lection strategy based on previous observations and the current estimation.

Integration with Object Recognition

Active vision-based object localization is closely related to object recognition, as accurate localization requires the detection and identification of the object. The integration of object recognition techniques, such as deep learning-based ob- ject detectors or feature-based recognition algorithms, enables the active vision system to accurately localize the object in the acquired images.

7 Limitations

- Lack of Contextual Information: Isolated object grasping ignores the contex- tual information surrounding the object. In real-world scenarios, objects are often found in cluttered environments or in the presence of other objects. Ignoring this contextual information can limit the robot's ability to plan andexecute successful grasps.
- Limited Adaptability: Isolated object grasping techniques typically rely on predefined grasping strategies or models trained on specific object categories. This limits the adaptability of the system to handle novel objects or unfore-seen scenarios. The lack of generalization can hinder the robot's ability to grasp a wide range of objects effectively
- Occlusion and Partial Visibility: Isolated object grasping assumes that the entire object is visible and not occluded. However, in practical scenarios, objects may be partially occluded or have obstructed views. Dealing with occlusion and partial visibility is a challenging problem that requires addi-tional perception and planning capabilities.[17]
- Grasp Failure in Complex Shapes: Isolated object grasping techniques may struggle to grasp objects with complex shapes or irregular geometries. Ob-jects with concave surfaces, thin structures, or asymmetrical shapes can pose challenges for grasp planning and execution. Specialized grasp strategies or adaptive grasping mechanisms may be required to handle such objects effec- tively[18]

- One of the limitation in the field of computer vision and robotic graspingis the limited diversity of objects and scenes available in existing datasets and benchmarks. While datasets like GraspNet and Contact-GraspNet have contributed significantly to advancing the research in object grasping, they still have limitations in terms of the diversity of objects and scenes they cover. This limitation can impact the applicability and robustness of grasping algorithms in real-world scenarios.

  Limited diversity of objects refers to the lack of representation of various object categories, shapes, sizes, textures, and material properties within thedataset. Existing datasets often focus on common household objects or spe- cific industrial items, which may not fully represent the vast range of objects encountered in real-world settings. This can result in grasping algorithms that are biased towards the objects present in the dataset, leading to poor performance on unseen or novel objects.
- Lack of tactile feedback and force-based grasping:

  One significant limitation in current grasping systems is the lack of tactile feedback and reliance on vision-based approaches. While vision-based per- ception plays a crucial role in object recognition and pose estimation, it provides limited information about the physical properties of objects, such as their texture, hardness, or slipperiness. Tactile feedback and force-based grasping, on the other hand, involve using sensors and force/torque mea- surements to perceive and control the interaction between the gripper andthe object.[19]

  Tactile feedback enables the system to detect and respond to subtle changes in object properties during the grasping process, such as detecting object slip, adjusting grip force, or ensuring a secure and stable grasp. Force-based grasping allows the system to actively control the applied forces and adapt the grasp according to the object's physical characteristics. Incorporating tactile feedback and force-based grasping can significantly improve the ro- bustness, stability, and dexterity of grasping systems, especially when dealing with objects with complex shapes, fragile materials, or uncertain properties.

8 Experiments and Results

This section discusses about the experiments which are performed in order to minimize the limitation which occurs due to Occlusion and Partial Visibility and cluttering. A cluttered scene is captured using Zed camera to provide stereo and Depth image. This image contains various indoor objects which are of interest to us and to which can perform Grasp prediction techniques.

To select the object of interest within the scene which has multiple objects i,e. from the cluttered scene Segmentation was performed to single out the object of interest.[20]

Various Segmentation Models were studied and tabulated IN Table 1 to com- pared and select appropriate Model which would help for segmenting out the

Table 1. Comparision of Segmentation models

ibute	**NN**	**t**	**Net**	**Lab**
t Resolution	ium			
el Capacity			ium	
ning Speed		ium		
ence Time				
ndary Clarity			ium	
il Preservation				
ct Detection				
antic Segmentation				
nce Segmentation				
rmance	erate		erate	

Object of interest[8][21][22] Deeplabv3 ADE20k dataset was used to segment the images from clutters and single out the object of interest. DeepLabv3 is a deep learning architecture for semantic image segmentation developed by Google. It achieves state-of-the-art performance on various segmentation tasks, including the ADE20K dataset.[23]

The ADE20K dataset is a large-scale scene parsing dataset that contains 150 classes for semantic segmentation.[24]. The combination of DeepLabv3 with the ADE20K dataset allows for accurate pixel-level segmentation of objects and scenes in images. The model can learn to differentiate between different classes and generate high-quality segmentation masks for each object present in the im- age. By this method the object of interest was succesfully singled out and pixel values of the objects were also obtained.[21] A mask for the object of interest was created and the pixel value of other objects were set to 0. Same was done for the depth image which was obtained from the Zed camera along with the stereo image. Few examples are shown in Fig.1.

After successfully Segmenting the Color and Depth images to find area of interest the next step is to obtain the Grasp poses using GraspNet. GraspNet can be modified to use custom data set for which information of the camera intrinsic properties have to be changed. The camera intrinsic values of Zed cam were updated and used for predicting grasp poses.

GraspNet takes as input a depth image of the scene,which is obtained from Zed camera. The depth image represents the 3D geometry of the objects in the scene.It utilizes convolutional neural network (CNN) as the backbone architec- ture for feature extraction from the input depth image. The features extracted from the backbone network are fed into a series of fully connected layers to pre- dict the grasp pose parameters. These parameters typically include the grasp position (3D coordinates), grasp orientation (e.g., Euler angles or quaternions),and other relevant information such as the grasp width.[25]

In addition to the grasp pose parameters, GraspNet also predicts the quality or success probability of each grasp pose, which output a probability score indi- cating the likelihood of a successful grasp. It is trained using labeled data that

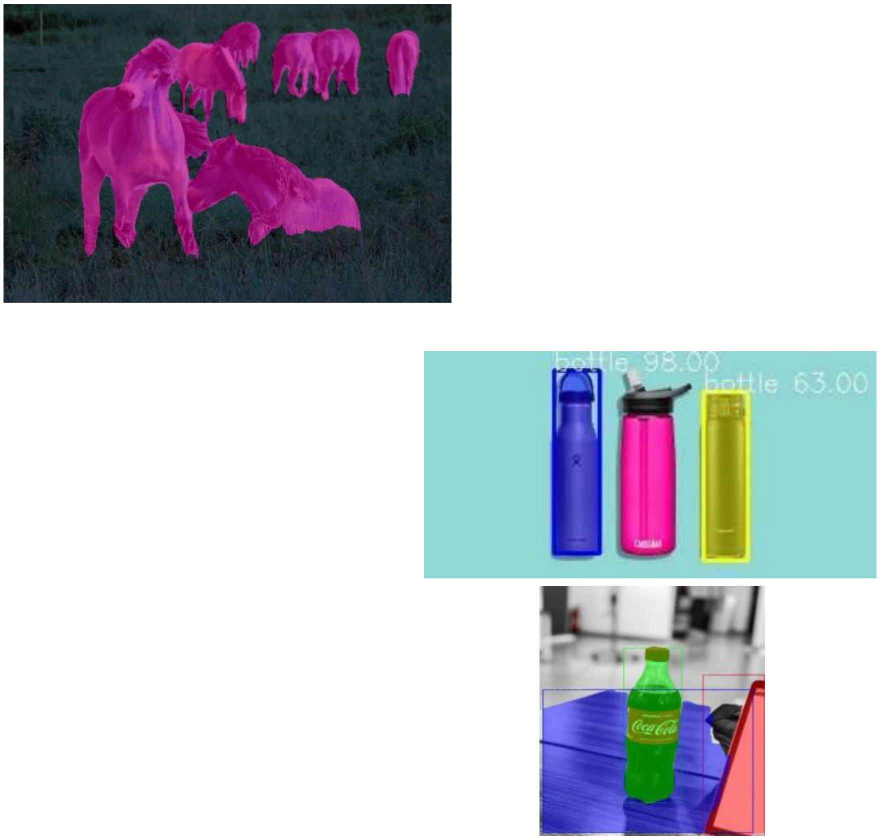

Fig. 1. Segmentation Using Deeplabv3ade20k

includes depth images of scenes along with ground truth grasp poses and their associated quality labels (e.g., success or failure). The network is trained to min- imize a loss function that measures the discrepancy between the predicted grasp poses and the ground truth poses, as well as the predicted grasp quality and theground truth labels. The grasp poses can be visualised using open3d after the successful prediction of grasp poses.

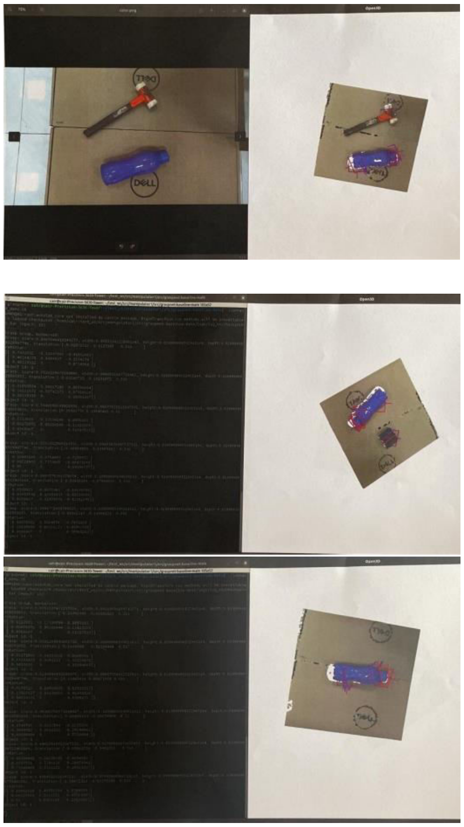

Fig. 2. Grasp Prediction using GraspNet

9 Conclusion & Future Scope

- Experiments were performed to compare the cluttered scenes with and with- out the use of segmentation to improve accuracy in such situations
- Particular Object of interest was able to be identified using the segmentation method which resulted in getting grasps for the said object

- Further study can be done to accurately find out the difference between clutter scene with different segmentation methods
- Further limitations can be rectified to improve the accuracy of grasps whichare obtained through graspnet

References

1. Li Wanyan and Ruan Guanqiang. Scene prediction and manipulator grasp pose estimation based on yolo-graspnet. In *2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT)*, pages 496–501. IEEE, 2023.
2. Samarth Brahmbhatt, Ankur Handa, James Hays, and Dieter Fox. Contactgrasp: Functional multi-finger grasp synthesis from contact. In *2019 IEEE/RSJ Inter- national Conference on Intelligent Robots and Systems (IROS)*, pages 2386–2393.IEEE, 2019.
3. Umar Asif, Jianbin Tang, and Stefan Harrer. Ensemblenet: Improving grasp de- tection using an ensemble of convolutional neural networks. In *BMVC*, page 10, 2018.
4. Yiye Chen, Yunzhi Lin, and Patricio Vela. Keypoint-graspnet: Keypoint- based 6-dof grasp generation from the monocular rgb-d input. *arXiv preprint arXiv:2209.08752*, 2022.
5. Hao-Shu Fang, Chenxi Wang, Minghao Gou, and Cewu Lu. Graspnet-1billion: A large-scale benchmark for general object grasping. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11444–11453, 2020.
6. Umar Asif, Jianbin Tang, and Stefan Harrer. Graspnet: An efficient convolutional neural network for real-time grasp detection for low-powered devices. In *IJCAI*, volume 7, pages 4875–4882, 2018.
7. Lerrel Pinto and Abhinav Gupta. Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours. In *2016 IEEE international conference on robotics and automation (ICRA)*, pages 3406–3413. IEEE, 2016.
8. Dengfeng Chai, Shawn Newsam, Hankui K Zhang, Yifan Qiu, and Jingfeng Huang. Cloud and cloud shadow detection in landsat imagery based on deep convolutional neural networks. *Remote sensing of environment*, 225:307–316, 2019.
9. Robert Krug, Achim J Lilienthal, Danica Kragic, and Yasemin Bekiroglu. Ana- lytic grasp success prediction with tactile feedback. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 165–171. IEEE, 2016.
10. Máximo A Roa and Raúl Suárez. Grasp quality measures: review and performance. *Autonomous robots*, 38:65–88, 2015.
11. Zhen Xie, Xinquan Liang, and Canale Roberto. Learning-based robotic grasping: A review. *Frontiers in Robotics and AI*, 10:1038658, 2023.
12. Dmitry Kalashnikov, Alex Irpan, Peter Pastor, Julian Ibarz, Alexander Herzog, Eric Jang, Deirdre Quillen, Ethan Holly, Mrinal Kalakrishnan, Vincent Vanhoucke, et al. Scalable deep reinforcement learning for vision-based robotic manipulation. In *Conference on Robot Learning*, pages 651–673. PMLR, 2018.
13. Pablo Bermejo, Jose A Gámez, and Jose M Puerta. A grasp algorithm for fast hy- brid (filter-wrapper) feature subset selection in high-dimensional datasets. *PatternRecognition Letters*, 32(5):701–711, 2011.
14. Shirin Joshi, Sulabh Kumra, and Ferat Sahin. Robotic grasping using deep rein- forcement learning. In *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, pages 1461–1466. IEEE, 2020.

15. H de Ruiter, M Mackay, and B Benhabib. Autonomous three-dimensional track- ing for reconfigurable active-vision-based object recognition. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*,224(3):343–360, 2010.
16. Gian Luca Mariottini and Stergios I Roumeliotis. Active vision-based robot local- ization and navigation in a visual memory. In *2011 IEEE International Conferenceon Robotics and Automation*, pages 6192–6198. IEEE, 2011.
17. Sung-Kyun Kim and Maxim Likhachev. Planning for grasp selection of partially occluded objects. In *2016 IEEE International Conference on Robotics and Au- tomation (ICRA)*, pages 3971–3978. IEEE, 2016.
18. Bo Lu, Bin Li, Wei Chen, Yueming Jin, Zixu Zhao, Qi Dou, Pheng-Ann Heng, and Yunhui Liu. Toward image-guided automated suture grasping under complex en- vironments: A learning-enabled and optimization-based holistic framework. *IEEE Transactions on Automation Science and Engineering*, 19(4):3794–3808, 2021.
19. Pascal Weiner, Felix Hundhausen, Raphael Grimm, and Tamim Asfour. Detecting grasp phases and adaption of object-hand interaction forces of a soft humanoid hand based on tactile feedback. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3979–3986. IEEE, 2021.
20. Bolei Zhou, Hang Zhao, Xavier Puig, Tete Xiao, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Semantic understanding of scenes through the ade20k dataset. *International Journal of Computer Vision*, 127:302–321, 2019.
21. Abolfazl Abdollahi, Biswajeet Pradhan, and Abdullah M Alamri. An ensem- ble architecture of deep convolutional segnet and unet networks for building se- mantic segmentation from high-resolution aerial images. *Geocarto International*, 37(12):3355–3370, 2022.
22. Mat Nizam Mahmud, Muhammad Hiszarul Azim, Mohd Hisham, Muham- mad Khusairi Osman, Ahmad Puad Ismail, Fadzil Ahmad, Khairul Azman Ahmad, Anas Ibrahim, and Azmir Hasnur Rabiani. Altitude analysis of road segmentation from uav images with deeplab v3+. In *2022 IEEE 12th International Conference on Control System, Computing and Engineering (ICCSCE)*, pages 219–223. IEEE, 2022.
23. Saziye Ozge Atik and Cengizhan Ipbuker. Instance segmentation of crowd detection in the camera images. In *Proceeding of Asian Conference on Remote Sensing 2020(ACRS 2020)*, 2020.
24. Salih Can Yurtkulu, Yusuf Hüseyin Şahin, and Gozde Unal. Semantic segmen- tation with extended deeplabv3 architecture. In *2019 27th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4. IEEE, 2019.
25. Edward Johns, Stefan Leutenegger, and Andrew J Davison. Deep learning a grasp function for grasping under gripper pose uncertainty. In *2016 IEEE/RSJ Inter- national Conference on Intelligent Robots and Systems (IROS)*, pages 4461–4468.IEEE, 2016.