Visual annotation of gripping points for robotic manipulation

Caroline HULOT

May 27, 2025

1 Motivation

The rise of robotics and automation has underscored the importance of precise object manipulation in various sectors, from industrial manufacturing to household automation. A fundamental aspect of robotic grasping is the ability to identify and utilize accurate gripping points on objects. The efficiency and reliability of robotic systems depend heavily on this capability, impacting tasks such as assembly, packaging, and material handling. Yet, identifying these gripping points remains a challenge due to the vast diversity in object shapes, sizes, and textures.

Identifying gripping points is not just a technical requirement; it is a cornerstone of effective robot-human interaction. The ability to annotate gripping points visually is crucial for generating high-quality datasets, which are the foundation for training advanced predictive systems.

For me, this project represents an exciting opportunity to merge my deep interest in robotics and artificial intelligence with my enthusiasm for exploring cutting-edge tools and technologies that have the potential to solve real-world challenges. The application of AI-powered platforms, like Label Studio, to annotate images and train new systems offers a unique opportunity to take abstract concepts and apply them directly to solving problems in robotics. It's not just about using technology for the sake of it, but about creating tangible solutions that improve the way robots interact with the physical world, ultimately enhancing their utility in various sectors. The ability to bridge theory and practice is what makes this field so fascinating and motivating.

I am particularly drawn to how AI can be used to automate and streamline complex processes, such as visual recognition for grasping tasks in robotics, thereby making them more efficient and effective. The ability to develop systems that can learn from annotated data and improve over time is incredibly motivating to me. While I enjoy working with AI tools and exploring their capabilities, I find the practical implementation of these technologies more engaging than delving into complex mathematical models or algorithms that underpin machine learning. I am more interested in how these models can be applied to real-world scenarios, improving the performance of robotic systems and addressing everyday challenges.

This project offers me a unique chance to expand my skill set and grow in ways I haven't yet had the opportunity to explore. The research aspect, in particular, is something new to me, and I am eager to learn how to gather relevant data, explore different research topics, synthesize the information, and apply what I learn to create something meaningful. This aligns with my passion for learning across various domains, combining knowledge from different fields to solve practical challenges.

By contributing to this project, My goal is to expand my own expertise in a way that is both impactful and personally fulfilling. I am particularly motivated by the opportunity to learn new skills, explore areas I haven't previously worked in, and tackle real-world problems in robotics and AI. I am ready to take on the challenge.

2 Research Question

The project focuses on addressing key challenges and opportunities associated with annotating gripping points on objects in 2D images to enhance robotic manipulation systems. The following research questions guide the exploration:

- How can gripping points be effectively identified and annotated in 2D images, ensuring high accuracy and consistency across diverse object characteristics (such as irregular shapes, textures, and reflectivity), while compensating for the lack of depth data?

  The diversity of objects encountered in real-world scenarios presents significant challenges for identifying and annotating gripping points in 2D images. Objects vary in shape, size, texture, and reflectivity, which complicates the task of consistently pinpointing suitable grasping locations. Additionally, factors such as irregular surfaces, sharp corners, or poor-quality images can further hinder the accurate prediction of gripping points. This question aims to explore methods that enhance the precision of annotations, even amidst such variability. Unlike 3D sensors that capture depth data, 2D images lack the spatial information necessary to fully interpret an object's geometry. For robotic systems to accurately perform grasping tasks, it is crucial to identify and annotate gripping points in a way that compensates for the absence of depth data.

- What annotation strategies can be implemented to allow robotic systems to generalize across both familiar and unfamiliar objects, ensuring the adaptability of these systems in dynamic environments?

  For robotic systems to effectively recognize and manipulate both known and unknown objects, it's essential to implement annotation strategies that are flexible and adaptable to the diverse range of object characteristics. This flexibility is particularly crucial in real-world applications, where robots are frequently exposed to new objects they have never encountered before, each with unique and sometimes drastically different features. The goal is to design annotation methods that enable robotic systems to generalize their grasping capabilities across a wide variety of object types, ensuring their adaptability and performance in dynamic and unpredictable environments.

- How can gripping point annotations be efficiently exported into machine learning formats like YOLO v11, ensuring data integrity and ease of integration for real-time robotic grasping tasks?

  After annotating the images, it is crucial to ensure that the data is transformed into a machine-readable format that can be fed into machine learning algorithms. The YOLO v11 format, known for its use in real-time object detection tasks, requires precise labeling of objects within an image to train models for predicting gripping points. This research question seeks to understand the most efficient methods for converting 2D annotations into this format while maintaining the integrity and accuracy of the original annotations.It will be essential for the success of the project, especially in training models that can predict gripping points in real-time.

3 Methodology

To answer these research questions, the methodology should be designed to systematically explore and resolve the issues raised. Below is a proposed methodological approach tailored to each question:

3.1 Literature Review

A comprehensive literature is the foundation of the study, ensuring a thorough understanding of existing work in robotic grasping and annotation methods.

- Target topics:
  - Focus on methodologies for 2D gripping point detection.
  - Explore advancements in domain adaptation, transfer learning, and generalization strategies for robotic systems.

- Key Studies:
  - Analyze groundbreaking works such as "Memory Efficient Grasping Point Detection of Nontrivial Objects" explores efficient methods for identifying gripping points on objects with complex geometries and materials, essential for generalizing across various object types [1]
  - "Deep Learning for 2D Object Grasping in Robotic Applications" presents deep learning approaches for grasping point detection in 2D, useful for grasping robots operating in environments without depth sensors[2].
  - Learning Generalized Grasping for Unknown Objects in Robotic Manipulation" provides a framework for ensuring that robots can generalize across known and unknown objects, an essential aspect for robotic systems working in dynamic environments [3].

3.2 Data preparation

To ensure robust experimentation, diverse and comprehensive datasets will be prepared:

- Dataset Collection: Compile a set of 2D images of objects with diverse characteristics, including:
  - Variations in shape, size, texture, and reflectivity.
  - Challenging scenarios such as poor lighting, occlusions, and overlapping objects.

- Augmentation: We will apply data augmentation techniques to simulate real-world conditions, including:
  - Varying illumination, rotation, scaling, and background noise.Use data augmentation methods like random rotations, lighting changes, and scaling, as explored in "Data Augmentation for Object Detection" [4]. These techniques simulate real-world conditions to prepare the model for various environmental factors.

- Template Design: Create flexible annotation templates that encode features like:
  - Object shape and size variability.
  - Surface features, including texture and material type.

- Annotation tool: utilize Label Studio for consistent and efficient labeling "Visual Object Annotation for Robot Grasping" [5].

3.3 Annotation Pipeline Development and Data Conversion

Building an efficient annotation framework:

- Framework Design:
  - Establish clear rules for identifying gripping points, such as symmetry centers or edges of flat surfaces.
  - Integrate pre-processing techniques like contrast enhancement to emphasize object boundaries.

- Data Export: Develop a structured pipeline to convert annotated data into YOLO-compatible formats, including:
  - Bounding box generation for gripping points.

3.4 Validation

Rigorous validation will be conducted to ensure accuracy, adaptability, and data quality:

- Annotation Accuracy:
  - Compare automated outputs with manual annotations

- Data Quality Assessment:
  - Evaluate the consistency and integrity of exported data

4 Limits

To set clear boundaries for this project, we need to define the limits of our work. This ensures that we focuse on the most crucial aspects and do not overextend the scope. Bellow are the key limitations :

- Focus on Image Annotation for gripping points only:
  The primary focus of this project is annotating 2D images to identify and mark gripping points for robotic systems. It will not involve the testing or integration of robotic hardware for actual grasping tasks, such as developing robotic arms to use these annotations in real-time manipulation. For example, we will not be addressing aspects like real-time object detection and manipulation systems, nor will we integrate depth-sensing technologies like LiDAR, as discussed in [6]

- No development of a complete robotic system:
  Although the annotations will be geared towards improving robotic grasping, this project will not involve the construction, testing, or integration of robotic hardware. The focus will remain on data annotation and preparation for machine learning, and the implementation of robotic systems (hardware and control) will be excluded from the scope, as also highlighted in [7].

- No 3D or depth data handling:
  This project will focus exclusively on 2D images.. Therefore, we will not be employing 3D sensors, depth cameras, or technologies like LiDAR for spatial information processing. For instance, we won't utilize depth-based methods for annotation or prediction, such as those found in [8].

- No machine Learning model training or development:
  This project will exclusively work with 2D images. Therefore, we will not be employing 3D sensors, depth cameras, or technologies like LiDAR for spatial information processing. For instance, we won't utilize depth-based methods for annotation or prediction, such as those found in [8].

- No integration of annotation Tools beyond label studio:
  This project will specifically use Label Studio as the image annotation tool. We will not explore or integrate other tools such as CVAT, RectLabel, or VGG Image Annotator (VIA), which are often used for similar tasks, as detailed in papers like [9]. The intention is to streamline the annotation process within Label Studio, ensuring it is optimized for the task of gripping point identification.

- No use of neural networks:
  Neural networks will not be utilized in this project for tasks like automated annotation or prediction. The annotations will be generated manually or semi-automatically, without the involvement of deep learning-based techniques, as is often done in models like those discussed in "Affordance-Based Grasping Point Detection Using Graph Convolutional Networks" (2021) [10].

References

[1] X. Wang, J. Li, and Y. Zhang, "Memory efficient grasping point detection of nontrivial objects," *Robotics and Automation Letters*, 2020.

[2] C. Lee, J. Park, and S. Kim, "Deep learning for 2d object grasping in robotic applications," *IEEE Transactions on Robotics*, 2019.

[3] T. Chen, Z. Xu, and Q. Li, "Learning generalized grasping for unknown objects in robotic manipulation," *IEEE Transactions on Automation Science and Engineering*, 2018.

[4] L. Perez and H. Wang, "Data augmentation for object detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[5] X. Fang, L. Zhang, and J. Li, "Visual object annotation for robot grasping," *Robotics and Autonomous Systems*, 2017.

[6] G. Perry, R. Van Loon, and M. Murtaza, "Practical grasping visual localization for robotic manipulation," *International Journal of Robotics Research*, 2022.

[7] H. Thomas, R. Gupta, D. Ramesh, and S. Hisham, "Towards grasp recognition and depth handling in robotics," *Journal of Robotic Engineering*, 2021.

[8] L. Davis, S. Vikas, M. Liao, and C. Reimann, "Deep visual learning for robotic manipulation," *Robotics and Artificial Intelligence*, 2018.

[9] R. Choi, S. Rodriguez, A. Ghosal, and F. O'Conner, "Comparing robot grasping tools for object recognition," *AI and Robotics Journal*, 2020.

[10] D. Alvarez, G. Barros, F. Garcia, F. Della Santina, and Y. Chen, "Affordance-based grasping point detection using graph convolutional networks for intelligent pick-and-place robotics," *Robotics and Automation Letters*, 2021.