Emre Girgin
emre.girgin.ms@gmail.com
+90 (554) 864 64 99
İstanbul, Türkiye

Automated Battery Disassembly

Case Study

February 2024

OVERVIEW

The objective of the case study is to develop an automated battery disassembly system that combines computer vision subtasks with a robotic manipulator system. The system will disassemble each battery module by removing protective plates, unscrewing, and unplugging connectors. To achieve this, a reliable object detector is necessary to differentiate the components that hold the pack together, and a task planner is needed to generate ordered disassembly instructions. The system will then estimate the component location and communicate it to the robotic subsystem through a suitable interface, such as ROS. Given the significant risk of electrical shock and fire associated with disassembling a high-voltage battery pack, it is essential to implement additional precautions enabled by the vision system.

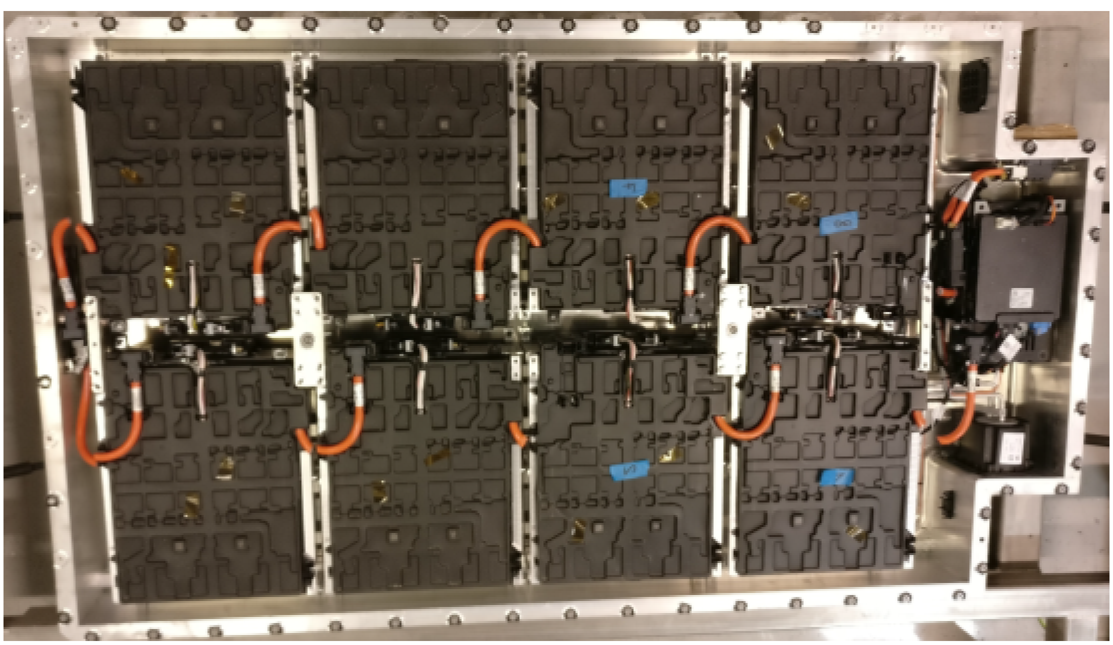


Figure 1. An example interior of BMW i3 battery pack. [3]

GOALS

1. Localization of battery pack
2. Removal of casing
    a. Detection of screws and grabbing point of the upper case
3. Task planning
    a. In which order to dismantle components
    b. 2D detection and 3D localization of components: screw heads, screw holes, fasteners, bolts, nuts, metal and plastic plates, wires, battery modules, cable connectors, BMS unit, thermo sensors, busbar, cable guides, gas vent, etc.
4. Visual Precautions
    a. Detection of potential sparks and overheats
5. Send the location of detected components to the robot manipulator over ROS.

REQUIREMENT SPECIFICATIONS

Functional Requirements

1. Software Requirements
    - 1.1. Vision System
        - 1.1.1. The vision system shall visually cover the entire battery pack.
        - 1.1.2. The vision system shall detect the support point of the cover case to be removed by a robot manipulator.
        - 1.1.3. The vision system shall detect and localize screw heads, screw holes, fasteners, bolts, nuts, metal and plastic plates, wires, battery modules, cable connectors, BMS unit, thermo sensors, busbar, cable guides, and gas vent.
        - 1.1.4. The vision system shall estimate the 3D orientation of screw heads, screw holes, fasteners, bolts, nuts, metal and plastic plates, wires, battery modules, cable connectors, BMS unit, thermo sensors, busbar, cable guides, and gas vent.
        - 1.1.5. The vision system shall estimate the 3D location of screw heads, screw holes, fasteners, bolts, nuts, metal and plastic plates, wires, battery modules, cable connectors, BMS unit, thermo sensors, busbar, cable guides, and gas vent.
        - 1.1.6. The vision system shall plan an ordered disassembly instruction set to dismantle components without damage.
        - 1.1.7. The vision system should estimate the time to be spent on disassembly.

1.2. Robotic Manipulator Subsystem
  1.2.1. The robot manipulator shall solve the inverse kinematic to plan a trajectory to the desired 3D position and orientation.
  1.2.2. The robot manipulator controller shall change the end-effector depending on the component to be dismantled.
  1.2.3. The robot manipulator should drop the unmounted component to the predetermined location safely.

2. Hardware Requirements
  2.1. Vision System
    2.1.1. The system shall have two stereo cameras tilted towards the center of the conveyor band within a range of at least 12 meters.
    2.1.2. The system shall have a wrist camera attached to the manipulator.
    2.1.3. The system shall survey the disassembly region with a thermal camera that can cover all the battery modules in the battery pack.
  2.2. Robotic Manipulator System
    2.2.1. The system shall have a robotic manipulator with at least ±0,5 mm precision.
    2.2.2. The robotic manipulator shall lift at least 50 kg of payload.
    2.2.3. The robotic manipulator shall have a gripper, screwdriver, and vacuum end-effectors.

Non-functional Requirements

1. Adaptability
  1.1. The vision system shall cover a 2 by 3-meter area to handle various sizes and shapes of battery packs.
2. Safety
  2.1. The end effector of the robotic manipulator shall be coated with an insulator material.
  2.2. Each end effector shall be equipped with a thermal sensor to detect possible sparks and fires.
  2.3. The thermal camera on the vision system shall emit a hazard signal in case of overheating detected.
3. Efficiency
  3.1. The detection, pose estimation, and position estimation algorithms must run in real-time.
  3.2. The visual task planning system shall plan an ordered disassembly instruction set within 10 seconds.
4. Precision
  4.1. The object detection system shall have at least 90% mAP for each class.

4.2. The pose estimation system shall have a rotation error of less than 5% for each dimension.

4.3. The position estimation system shall have a displacement error of less than 0.5mm.

4.4. The visual task planning system shall have a delay error of less than 10 seconds.

Vision Capabilities

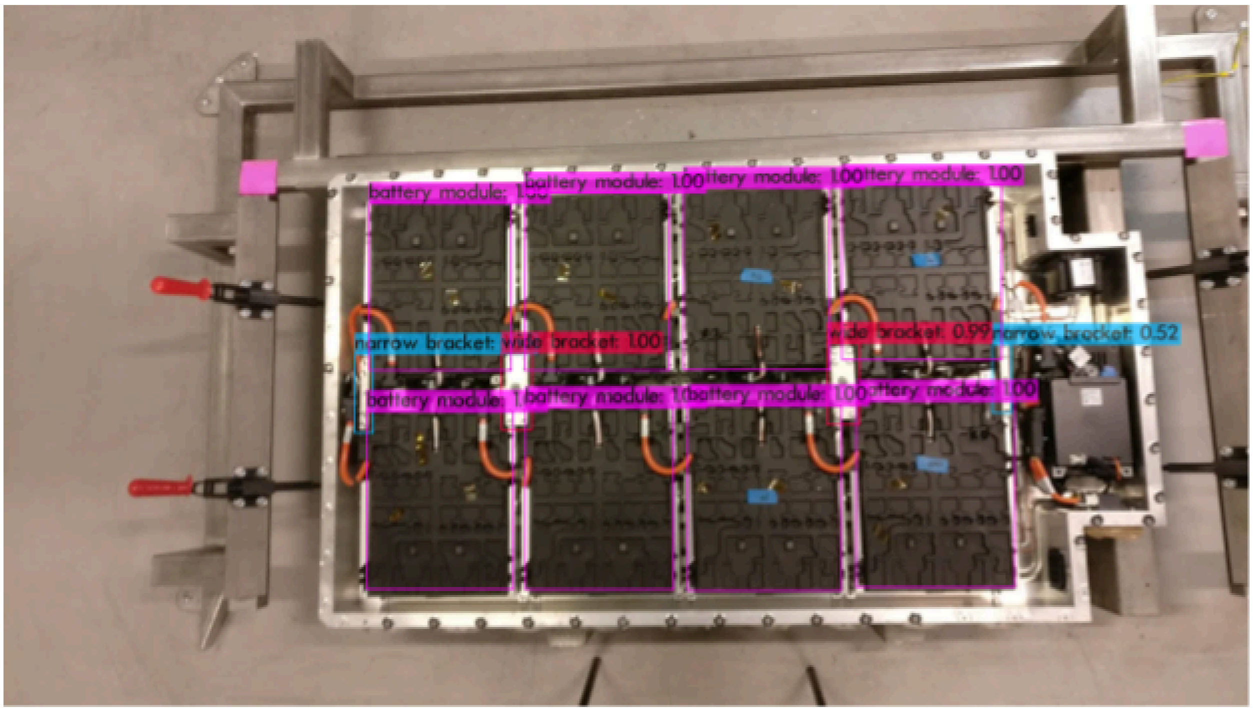


Figure 2. An example output of object detection on battery pack. [3]

1. Object Detection

To disassemble the battery pack, it is necessary to have an object detection mechanism that can detect its various components, including screws, fasteners, bolts, nuts, plates, wires, battery modules, cable connectors, and cable guides. The literature suggests using deep learning-based solutions to spatially localize and classify these components in 2D images. A real-time object detection algorithm would be beneficial for this task. A CNN-based YOLO algorithm might be a popular option since they have been extensively studied during the last decade. However, vision transformers may not be a suitable solution for this use case since they run slower, except in cases where severe occlusions disturb the spatial continuity of components. It is important to note that a labeled dataset is required for deep learning. The diversity of data is crucial for achieving generalization. While illumination and reflectance of surfaces are important factors, the size and color of components should reflect the distribution found in the commercial product.

Data augmentation can be used to achieve this, such as randomly adjusting brightness, contrast, saturation, and hue of an image or rotating and scaling it. On the other hand, traditional methods such as template or feature matching combined with ORB or HOG transformers may have low precision and slow inference. However, they can still be useful for saliency guidance, which can lead to faster inference and coarse auto-labeling. (See Figure 4)

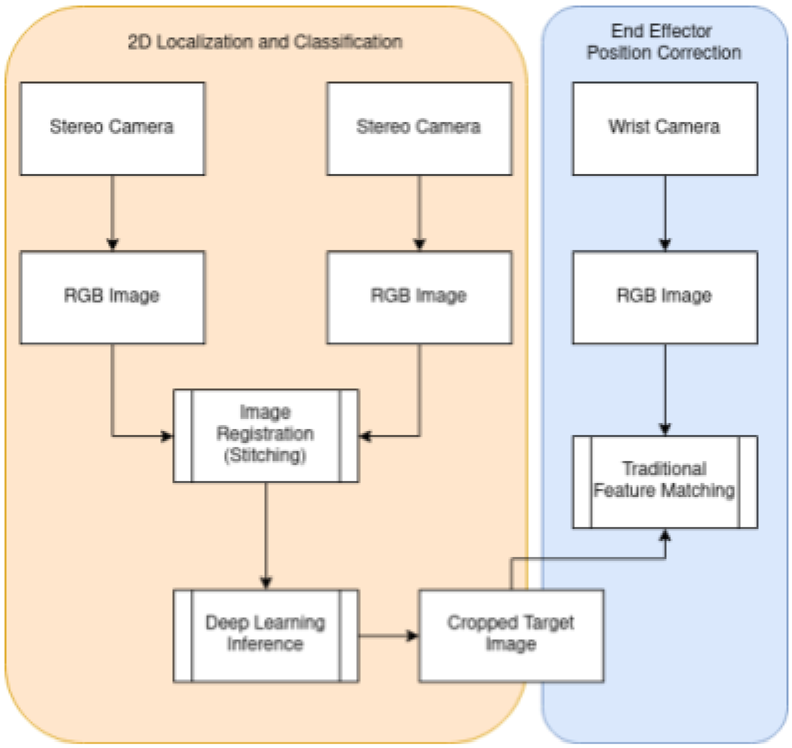


Figure 3. Object detection workflow.

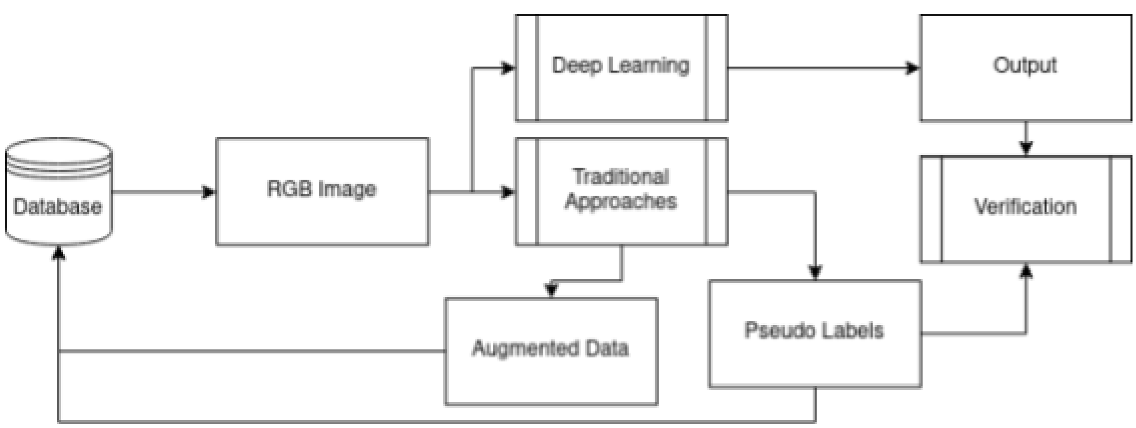


Figure 4. Cooperation with traditional methods.

2. Pose Estimation

Determining the orientation of components is crucial for successfully removing modules from the pack. For example, each module is connected to another in a series using a wire (orange in Figures 1 and 2) with a connector at the end. To release the connector clips, the robotic manipulator must press the connector from the correct angle. When dealing with planar surfaces that have obvious patterns, such as checkerboards, estimating the object's pose with a single image is easier. However, stronger techniques are required because the components in the battery pack may not have that structure. Popular approaches aim to predict the pose of objects visible from multiple angles and placed on a surface, such as a table. However, when an object is inside a case, only the upper face is visible, making it difficult to apply these methods in this domain. To determine the pose, it may be helpful to consider the relative position of the components. Therefore, pose estimation accuracy is closely linked to the accuracy of the object detector. End-to-end deep learning pipelines are commonly used for predicting 6-DoF object pose. However, most existing studies do not address objects that are packed inside a box-like casing.

Figure 5. Pose Estimation Workflow.

3. 3D Localization and Point Cloud Registration

To disassemble using a robotic arm, the 3D position of the component must be provided. To obtain depth information in stationary recording, stereo cameras can produce a point cloud or depth map, unlike monocular cameras. At least two stereo cameras are required to cover two sides of the battery pack and overcome limitations of illumination and reflectance from non-Lambertian surfaces. To register the point clouds from two cameras, an affine transformation can be constructed using the extrinsic parameters of both cameras. If the cameras can move within small displacements, the Iterative Closest Point (ICT) algorithm can be used.

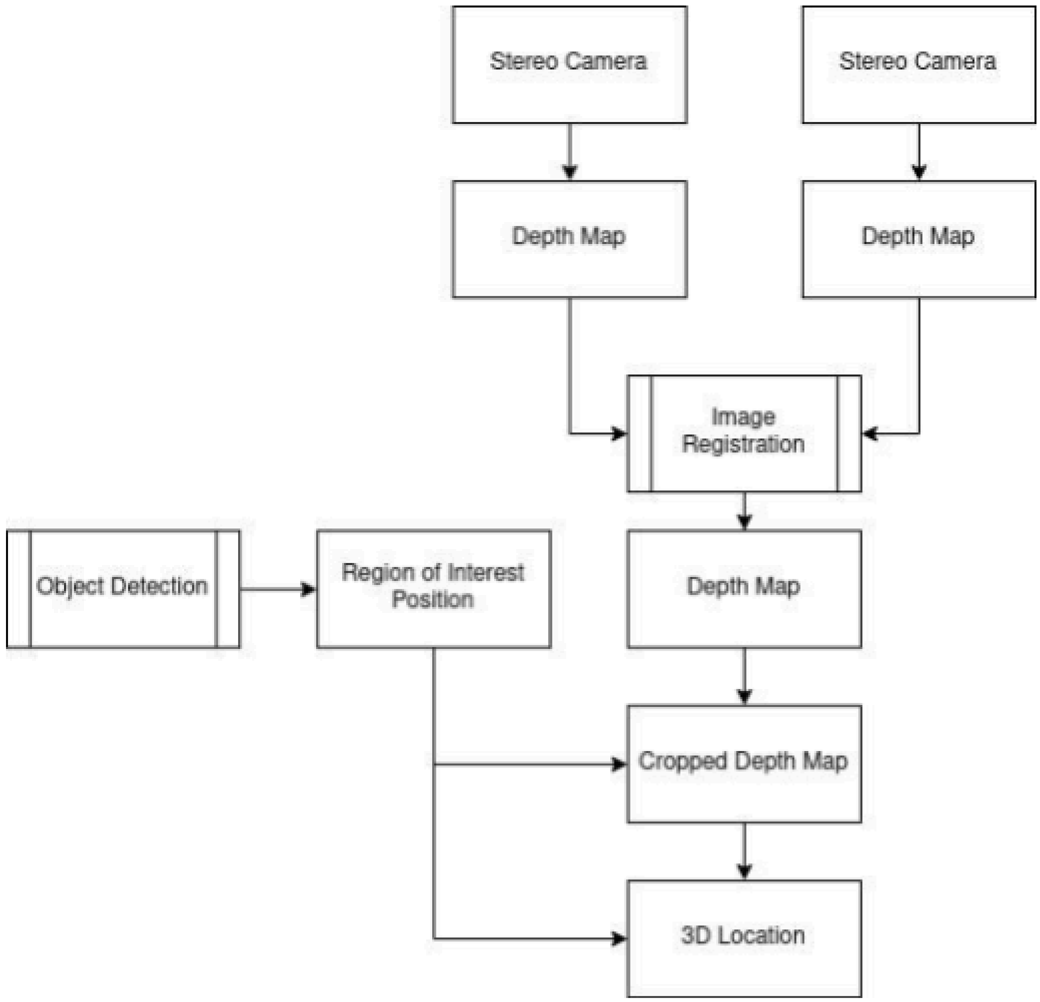


Figure 6. Position estimation workflow.

4. Visual Task Planner

The process of disassembly involves reversing the steps taken during the assembly line. For example, to remove the battery modules, you must first remove the brackets, plates, and other screws that secure the modules to the casing. Therefore, a task planner is necessary to determine the proper sequence for dismantling. While this can be a difficult task, the order of the components' depth can be utilized. The next component to be removed can be determined by combining the results of object detection on RGB images with registered depth maps from stereo cameras. Also, the disassembly instructions can be planned in the PDDL domain.

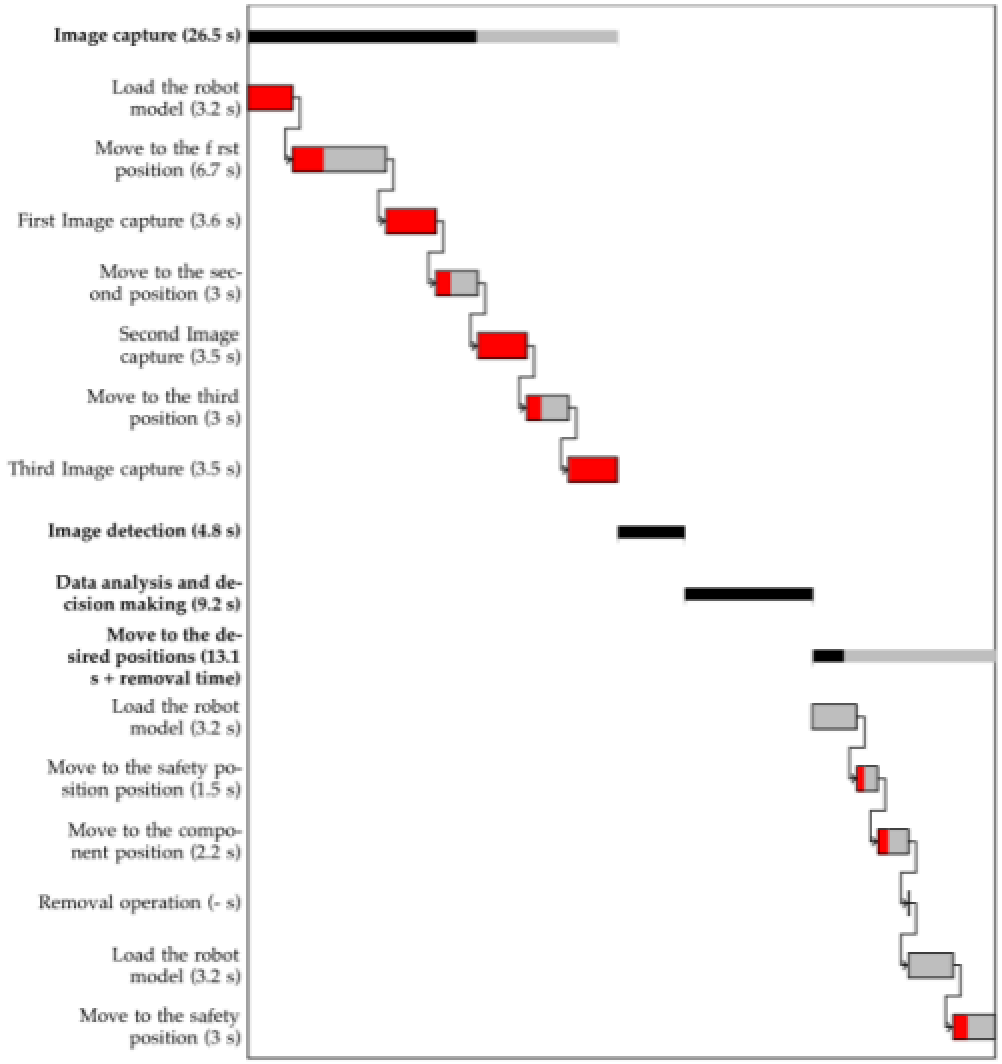


Figure 7. An example planned task from [1].

5. Visual Safety Precautions

The high-voltage battery pack poses additional safety risks. During the disassembly process, the operation zone must be monitored by a thermal camera in case of a short circuit. By combining the object detection output and thermal image, the temperature of the battery modules and cells can be monitored.

Deployment

Due to the need for multiple RGB images, thermal images, and point clouds, transferring this data to a cloud processor with low latency may not be feasible. Therefore, edge solutions running on a single-board computer, such as Jetson Nano, would be more suitable. For object detection tasks that require deep learning solutions, real-time capabilities are often necessary. In this domain, edge hardware is typically sufficient. However, it is important to note that these devices must be connected to the internet for updates and diagnostic purposes. Quantization and optimization tools are useful for deploying deep learning algorithms on edge devices. TensorRT is a popular tool used to accelerate and optimize deep learning models for Nvidia products.

Sensor Integration

The required sensors are the following:

- Stereo cameras (x2): Stereo cameras provide RGB images, depth maps, and corresponding point clouds in real time. They are used for object detection and pose estimation algorithms, as well as 3D position estimation and visual task planning. At least two cameras should observe battery packs from a tilted view to capture an additional dimension. (See Figure 3) Registering the point clouds helps to reduce errors in the 3D location estimate.
- Thermal camera: Thermal data is important to detect unintended sparks automatically and the temperature of the battery modules in the case of a short circuit.
- Thermal sensor: As a redundancy, the temperature of the medium should be monitored by a thermal sensor.

Figure 8. Illustration of proposed system hardware.

Continuous Improvement

Generative AI solutions have become increasingly popular. Two common use cases are dataset augmentation and synthetic data generation. When generating battery pack data, it is best to create components separately and combine them in a procedural manner. This approach allows for automatic generation of data annotations and avoids unrealistic samples. In addition, reconstructing the battery packs in 3D provides a new perspective. Novel approaches such as NeRFs and Gaussian splats generate view-dependent illumination outputs, which can be used to enhance the dataset for those dimensions.

Additional Suggestions

Object Affordances

Affordance learning is a popular field of research in robot learning.[6-12] It involves categorizing objects that are interacted with in a similar way. For example, screws, bolts, and nuts have similar interactions. Predicting object affordance from images is a useful tool for estimating disassembly instructions for out-of-distribution components in the battery pack.

Graph Neural Networks

Graphs are useful for showing the relationship between components in a battery pack. Recent advancements in graph neural networks (GNNs) and their variations can improve the task planning step when representing components in a graph. However, preprocessing requires converting object locations and poses to a graph, which can be challenging.

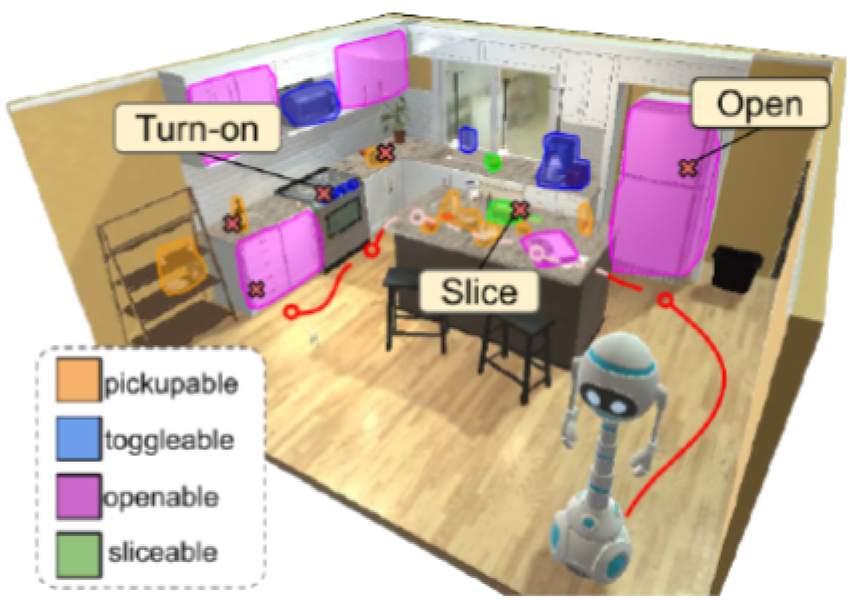


Figure 9. An example of affordances.

Imitation Learning / Learning from Demonstration

LfD is a subfield of robot learning where a human expert demonstrates how to execute a command, and the robot learns and generalizes this behavior under appropriate conditions. This approach can be used to trigger a robotic manipulator by directly processing the given image or disassembly instructions in an end-to-end manner.

Example Implementations

Morphological Segmentation

In this setting, morphological operations such as erosion and dilation are tuned to the specific input. After successfully locating the battery modules and corresponding wires, the pose is aligned from the center of the modules to the wires.

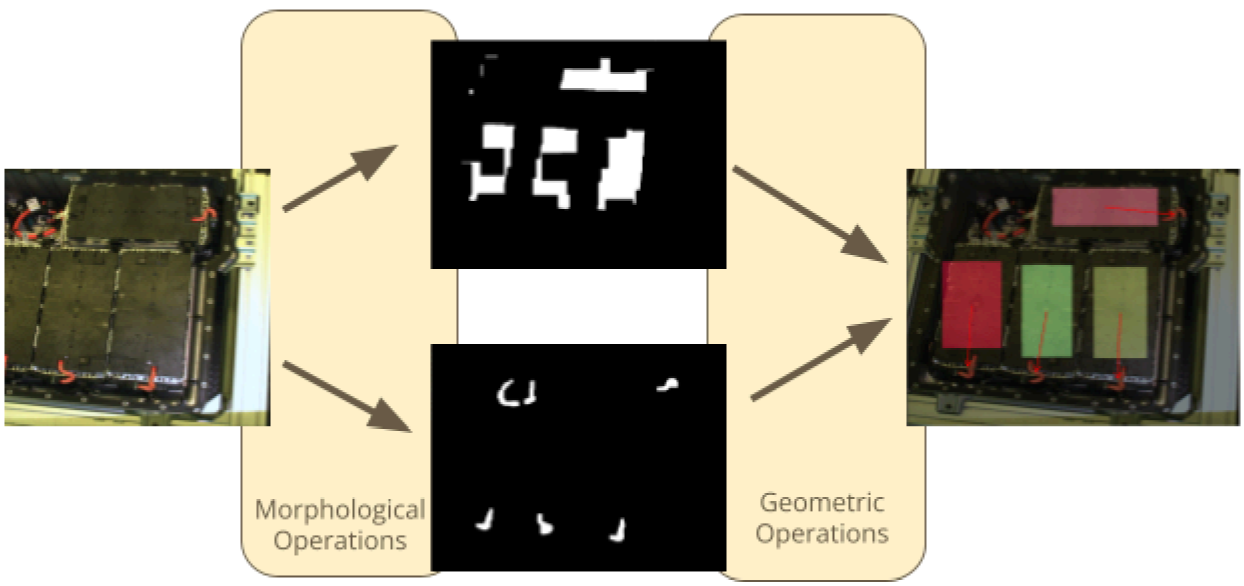
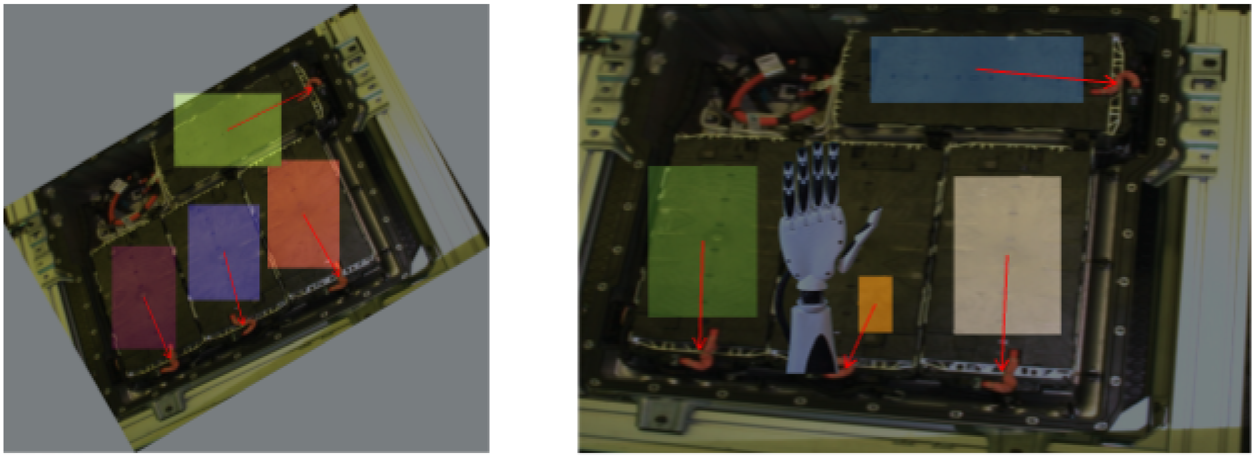


Figure 10. An overview of morphological segmentation and pose estimation implementation.

The system's output under rotation and occlusion is reported in Figure 5. While the system is able to find modules and poses under 30 degrees of rotation, the object detection precision degrades under severe occlusion.

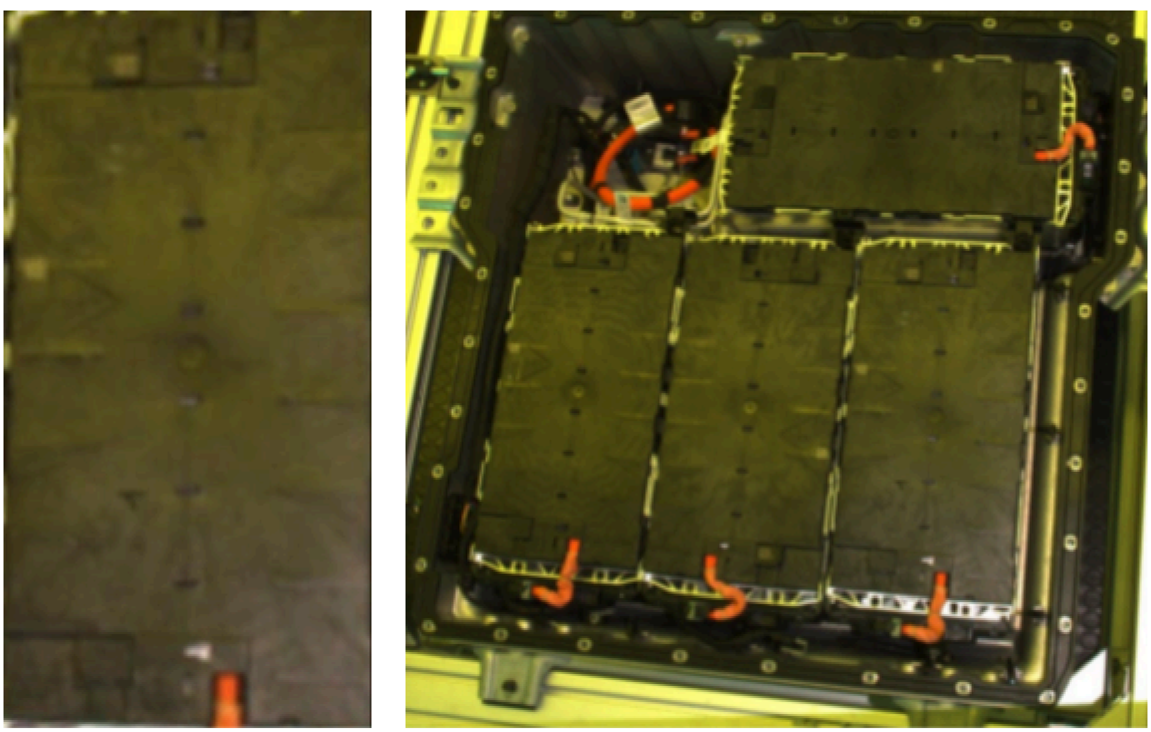


Rotation　　Occlusion

Figure 11. Morphological operation robustness under rotation and occlusion.

Descriptor Matching by Sliding Window

In this implementation, an example target image of a battery module's features are matched with features extracted by the sub-sections of the battery pack image by sliding window. As feature extractor we both utilized ORB and SIFT detectors.



a) Target battery module.　　b) Traversed battery pack.

Figure 12. Images used for object detection by feature matching.

The pipeline first corrects the white balance on the image and traverses the battery pack image by sliding window approach. After calculating the similarity between the target and the window by the number of good matches, a similarity score is obtained. Note that the 2 sliding window anchors are perpendicular to each other.

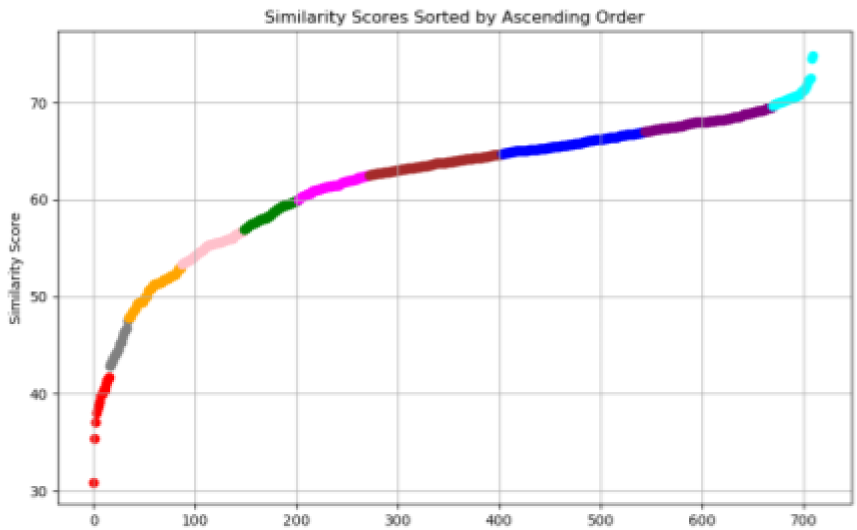


Figure 13. K-means clustered sorted similarity scores.

After the windows are sorted according to similarity scores, a K-means algorithm divides the predictions on 10 clusters. The cluster representing the most similar windows are fed to a Non-Max Suppression algorithm to eliminate multiple detections.

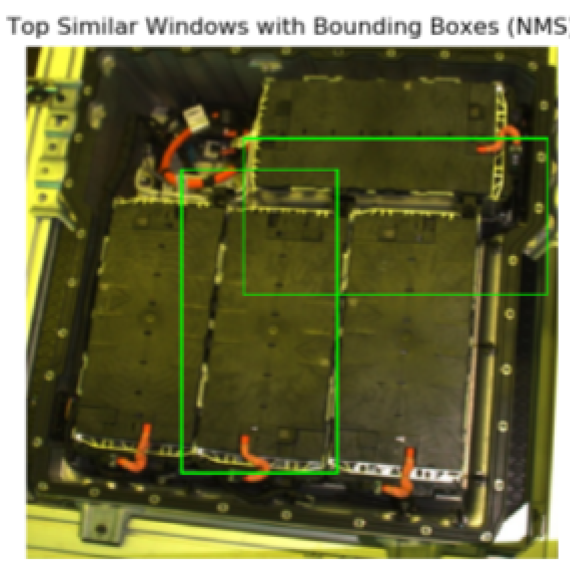


Figure 14. Output of object detection by feature matching.

References

1) Choux, Martin, Eduard Marti Bigorra, and Ilya Tyapin. "Task planner for robotic disassembly of electric vehicle battery pack." Metals 11.3 (2021): 387.
2) Kay, Ian, et al. "Robotic disassembly of electric vehicles' battery modules for recycling." Energies 15.13 (2022): 4856.
3) Rehnholm, Jonas. "Battery pack part detection and disassembly verification using computer vision." (2021).
4) Tan, Wei Jie, et al. "A hybrid disassembly framework for disassembly of electric vehicle batteries." International Journal of Energy Research 45.5 (2021): 8073-8082.
5) Zorn, Merle, et al. "An approach for automated disassembly of lithium-ion battery packs and high-quality recycling using computer vision, labeling, and material characterization." Recycling 7.4 (2022): 48.
6) P. Mandikal and K. Grauman, "Learning dexterous grasping with object-centric visual affordances," in 2021 IEEE international conference on robotics and automation (ICRA). IEEE, 2021, pp. 6169–6176.
7) A. Zeng, S. Song, K.-T. Yu, E. Donlon, F. R. Hogan, M. Bauza, D. Ma, O. Taylor, M. Liu, E. Romo et al., "Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching," The International Journal of Robotics Research, vol. 41, no. 7, pp. 690–705, 2022.
8) G. Schiavi, P. Wulkop, G. Rizzi, L. Ott, R. Siegwart, and J. J. Chung, "Learning agent-aware affordances for closed-loop interaction with articulated objects," in 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023, pp. 5916–5922
9) Y. Geng, B. An, H. Geng, Y. Chen, Y. Yang, and H. Dong, "Rlafford: End-to-end affordance learning for robotic manipulation," in 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023, pp. 5880–5886.
10) S. Cheng, K. Mo, and L. Shao, "Learning to regrasp by learning to place," in Conference on Robot Learning, 8-11 November 2021, London, UK, ser. Proceedings of Machine Learning Research, A. Faust, D. Hsu, and G. Neumann, Eds., vol. 164. PMLR, 2021, pp. 277–286. [Online]. Available: https://proceedings.mlr.press/v164/cheng22a.htm
11) Y.-C. Lin, P. Florence, A. Zeng, J. T. Barron, Y. Du, W.-C. Ma, A. Simeonov, A. R. Garcia, and P. Isola, "Mira: Mental imagery for robotic affordances," in Conference on Robot Learning. PMLR, 2023, pp. 1916–1927
12) J. Borja-Diaz, O. Mees, G. Kalweit, L. Hermann, J. Boedecker, and W. Burgard, "Affordance learning from play for sample-efficient policy learning," in 2022 International Conference on Robotics and Automation (ICRA). IEEE, 2022, pp. 6372–6378