Overcoming Data Scarcity in EV Battery Disassembly with Augmented Deep Learning and Structural Reasoning

Marina Baucks
wbk Institute of Production Science
Karlsruhe Institute of Technology
Karlsruhe, Germany
https://orcid.org/0009-0004-3321-9460

Haoran Sun
wbk Institute of Production Science
Karlsruhe Institute of Technology
Karlsruhe, Germany
uuycw@student.kit.edu

Florian Koessler
wbk Institute of Production Science
Karlsruhe Institute of Technology
Karlsruhe, Germany
https://orcid.org/0000-0002-1267-9423

Juergen Fleischer
wbk Institute of Production Science
Karlsruhe Institute of Technology
Karlsruhe, Germany
https://orcid.org/0000-0003-0961-7675

Abstract*—The steadily increasing demand for electric vehicles (EVs) has led to a heightened need for efficient, automated battery disassembly systems for sustainable recycling. However, the variability of battery designs and the insufficient availability of training data pose significant challenges for machine learning in the context of disassembly automation. In this study, we propose an image-based approach that integrates deep learning and structured reasoning to enable the autonomous generation of disassembly sequences for EV battery packs. A YOLOv8-based pipeline for object detection and instance segmentation is trained using two different approaches to data augmentation: conventional image transformations are compared with synthetic image generation using Segment Anything Models (SAM). Object-specific augmentation using SAM leads to higher precision in object recognition than general, conventional augmentation techniques. Structural relationships between components are then derived using both bounding box heuristics and pixel-level segmentation masks. This enables reliable extraction of spatial and connection data. The information extracted from the image data can then be further processed to derive well-founded and adaptable disassembly sequences. This process is possible without the use of CAD models or extensive prior knowledge. The experimental results showed improved recognition accuracy by using SAM-based data augmentation compared to data augmentation with conventional image transformation (mean Average Precision at an Intersection-over- Union threshold of 0.5 (mAP@0.5) increase: 2.0 %) and compared to not using augmentation techniques (mAP@0.5 increase: 3.3 %) on a small-sized dataset containing 190 images (original size, + 50 images for SAM-based augmentation; + 160 images for conventional augmentation).

***Index Terms*—electric vehicles, batteries, disassembly, computer vision**

I. Introduction

In the context of electric mobility, the globally increasing number of registered electric vehicles means that an equally rising number will reach the end of their service life in the coming years [1]. These vehicles' batteries contain valuable raw materials, including lithium, nickel, manganese, and cobalt, that are only available in limited quantities. Therefore, maximizing the recovery of these materials through recycling is essential so they can be reused in new batteries [2], [3]. To enable this recovery, end-of-life batteries must first be disassembled into their individual components at least to the module level which allows each material stream (e.g., aluminum, steel, copper, plastic, and active battery materials) to be separated. The more effectively this separation is performed, the purer the recovered materials will be [4].

Currently, however, the disassembly of batteries remains predominantly a manual process [5], [6]. The main obstacle to automation lies in the diversity of battery pack designs and the unpredictable conditions of batteries and their components at the end of their life. As a result, rigid, standardized processes like those used in assembly are unsuitable. Instead, successful automation demands a process that is both flexible and adaptive, capable of accommodating different product variants and varying component conditions [2], [7].

Bridging this gap between adaptability and automation requires advanced sensing and decision-making capabilities, making computer vision an essential enabling technology for flexible, automated disassembly systems. It allows image data from the product currently being disassembled to be processed and evaluated, thereby facilitating the analysis of its components and the identification of any potential issues. This enables different product variants, component configurations and conditions to be recognized, and a disassembly sequence to be created that is adapted to them [8], [9]. Modern computer vision approaches are often based on machine learning methods. Models such as "You Only Look Once" (YOLO) [10] are pre-trained neural networks that can recognize and segment objects with pixel-precise location [8], [11]. In order to adapt these models to specific objects, they must be

retrained using data sets containing those objects. However, to achieve a high degree of prediction accuracy, these datasets must contain a sufficient number of instances. For objects of high complexity, such as a wide variety of variants, different arrangements, occlusions and variations in the scene, many images are required to represent the complexity accordingly. Unfortunately, such large amounts of data are unavailable in the field of battery disassembly [12]. To utilize the advantages of image-processing machine learning models in this domain, it is necessary to investigate how accurate object detection can be achieved with small image datasets.

One option is to enlarge the data set through data augmentation, a process in which new, modified data is generated from the existing data [13]. This paper investigates the influence of two different augmentation methods on the accuracy of object detection, which are described in more detail in Section IV. Images from a dataset are modified using the two augmentation methods and then added back to the original dataset, creating two new datasets. A YOLOv8 model is then trained for object detection using the original dataset and the two augmented datasets for the classes 'module' and 'busbar', and the accuracy of object detection is compared. A segmentation model is then trained on the dataset with which the object recognition model achieves the highest accuracy in order to obtain the pixel-precise position of the components. Finally, we show how the information obtained from object recognition and segmentation can be used to extract the topological properties of a battery pack. Based on rules, both the arrangement of the components (next to each other or on top of each other) and whether the components are connected to each other are analyzed. This information is required, for example, to derive the disassembly sequence for automated disassembly.

II. State of the Art

Previous work has already investigated various computer vision approaches for automating battery disassembly. Choux et al. [8] use a YOLOv3 model to detect individual components and joints such as modules, the battery management system and screws of a A3 Sportback e-tron hybrid battery pack. The 2D pixels containing the objects are then converted into 3D points to localize the components in three-dimensional space. They then plan the disassembly process 'from top to bottom' so that the components at the top are removed next. Further work conducted in the context of this research suggests that the dataset used in this research contained 89 images of the same battery pack from different angles [14].

Zorn et al. [11] investigated an approach for a computer vision pipeline containing instance segmentation and point cloud registration for various battery pack components. They compared two instance segmentation network architectures and concluded that a Mask R-CNN with a Swin transformer backbone is best suited for component segmentation. Multiple datasets containing 30-200 images of eight component classes were created and synthetically enlarged using conventional data augmentation techniques such as cropping and rotation. In addition, the authors developed an approach for processing three-dimensional image data, aligning a detected scene point cloud of a component with its corresponding model point cloud to accurately estimate the component's pose, e.g. for calculating robotic grasping movements during disassembly.

Gerlitz et al. [15] also deal with the processing of 3D point clouds. However, this work is less concerned with ML-based methods. The authors developed a computer vision pipeline that can be used to compare 3D point clouds recorded from the actual disassembly situation with the CAD model of the battery module to be disassembled in order to determine the actual position of the components and joints predefined in the CAD model.

Zheng et al. [16] use computer vision-based approaches to detect components in the battery pack during disassembly and guide the worker through the process with instructions during manual disassembly. Facing the limited amount of domain-specific data, they use the Segment Anything Model to segment all present components in a first step, benefiting from the visual general knowledge contained in the pre-trained model. The segmented components are then classified in a second step using a CNN containing domain-specific information about the appearance of battery pack components.

Overall, the reviewed studies demonstrate that computer vision and deep learning methods, particularly object detection, instance segmentation, and 3D point cloud processing, are effective tools for identifying and localizing components in battery packs, which in turn enables the automation or guided execution of disassembly processes. However, a recurring limitation across these approaches is their dependence on extensive and high-quality datasets, as well as on detailed CAD models of the battery systems. For instance, Choux et al. rely on a manually collected dataset with limited diversity, while Zorn et al. address data scarcity through data augmentation techniques but still depend on controlled image sets of known components. Gerlitz et al. bypass learning-based methods entirely, instead relying on CAD-to-point cloud alignment, which again presupposes the availability of accurate digital models. Although Zheng et al. introduce a promising approach using a pre-trained foundation model to reduce the need for domain-specific training data, their method still requires a subsequent classification step based on labeled battery-specific images. In summary, while the state of the art illustrates the feasibility and potential of automated or assisted battery disassembly through computer vision, it also reveals a significant gap: current methods are constrained by their reliance on either richly annotated datasets or comprehensive CAD models which are resources that are often unavailable in practice, particularly for diverse or less-documented battery systems. This work addresses this gap by exploring a new approach to mitigate data scarcity through domain-specific creation of synthetic data, aiming to enable robust component detection even in data-limited scenarios. Based on this, further steps are investigated to determine the topology of the battery packs to be disassembled in order to enable the automated derivation of a disassembly sequence.

III. Prerequisites

A. Dataset

In this study, a dataset composed of 190 freely available online images is utilized. The images originate from a variety of sources, including real-world disassembly scenarios, informational graphics, and CAD renderings, and cover 15 distinct types of battery packs. All images depict battery packs in an open state (i.e., with the housing cover removed) or display detailed views of internal sections. The battery packs represented in the dataset follow a pack-to-module structural design, though the presence, configuration, and visibility of individual components vary significantly across the dataset. Additionally, the image resolution is generally low, which limits the visibility of fine structures and small-scale components.

Given these constraints, this work focuses on the detection and localization of two particularly critical component types: high-voltage (HV) connections (i.e., HV cables or busbars) and battery modules. HV connections link the individual modules in a series configuration and are typically the first components to be removed upon opening the pack, in order to safely lower the system voltage and eliminate the immediate risk posed by residual high voltage. Battery modules, on the other hand, represent the primary target of most disassembly operations, as they contain the battery cells and, consequently, the highest concentration of valuable raw materials. Disassembly operations generally proceed at least to the module level, as this enables the efficient separation of high-value cell materials from other components such as aluminum housings, copper wiring, and plastic electronics. The two classes were annotated manually using the CVAT data annotation platform.

B. Model Selection and Configuration

YOLOv8 was selected as the base vision model since it provides an optimal balance between high detection accuracy and computational efficiency [10], making it well-suited for battery pack component recognition in complex industrial environments.

The experimental setup for the model training is described in Table I. In order to ensure compatibility with other research projects, we used older versions of Python and YOLO instead of the latest versions (which are Python 3.13 and YOLOv12 to this date).

TABLE I
Hardware and Software Specifications

Component	Specification
Operating System	Windows 10
GPU Model	NVIDIA RTX 3050 Ti (VRAM: 4 GB)
CUDA Version	11.8
Programming Language	Python
Python Version	3.8
Deep Learning Framework	PyTorch 2.0.1 + Ultralytics YOLOv8

In addition to the system configuration, training parameters were selected with consideration of available resources and typical practices for object detection tasks. The dataset was split into 160 images for model training and 30 images for validation. The model was trained over 100 epochs, which appeared sufficient to allow meaningful learning without leading to overfitting. A batch size of 4 was chosen, taking into account the resolution of the training images (640×640) and the limitations of GPU memory.

The input image size was set at 640 pixels, aiming to balance accuracy and processing time. For optimization, the Stochastic Gradient Descent (SGD) algorithm was employed. The initial learning rate was set to 0.01, following commonly used values for YOLO models.

IV. Data Augmentation and Object Detection

Given the relatively limited availability of labeled battery pack images in publicly accessible datasets as described in Section I, this work investigated and compared two different data augmentation techniques to increase the diversity of training data and improve the generalization capability of the object detection model. These techniques include a conventional augmentation pipeline based on image transformation operations and a synthetic data generation approach utilizing the Segment Anything Model (SAM).

A. Conventional Augmentation Techniques

To simulate the variability encountered in real-world battery disassembly environments, a set of five augmentation transformations was applied to the training data. These include geometric modifications such as cropping, translation, small-angle rotation (limited to $\pm 10^{\circ}$), and horizontal flipping, along with pixel-level brightness variation. Brightness adjustments were constrained to a $\pm 20\%$ range and while not exceeding the minimum and maximum RGB values (0; 255) to mimic different lighting conditions. The workflow is described in Figure 1.

These transformations were chosen to reflect typical disturbances found in industrial settings, such as changes in camera angle, equipment layout, partial occlusion, and variable lighting. Cropping helped simulate cases where only part of a battery was visible due to spatial constraints, while translation and rotation allowed the model to adapt to positional shifts. Brightness variation, on the other hand, addressed challenges related to inconsistent illumination.

Each original image from the training dataset was augmented using one or more of these techniques. The number of techniques used (from at least one to all five) and the degree of variation of the parameters were chosen randomly for each image. The newly generated training sample was then added to the training dataset. This process doubled the size from 160 to 320 images and increased the variability of the training dataset while maintaining the semantic integrity of the visual content.

B. SAM-Based Synthetic Augmentation

In addition to conventional augmentations, this study implemented a synthetic augmentation strategy inspired by the Context-Aware Copy-Paste (CACP) method proposed by Guo

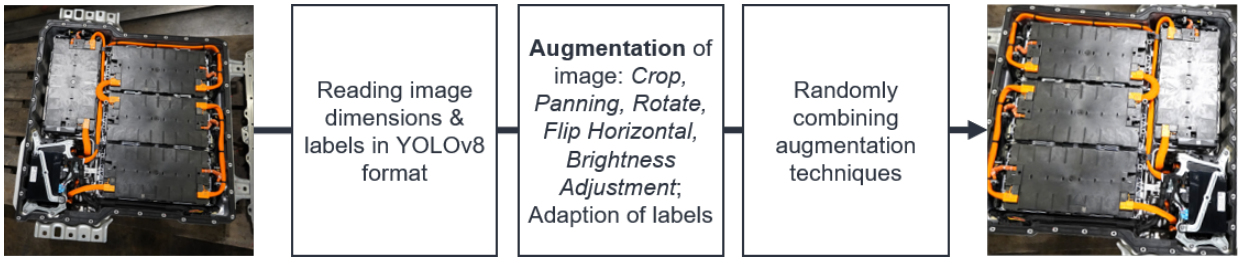

Fig. 1. Workflow illustrating conventional data augmentation methods

et al. [17] and the Simple Copy-Paste approach from Ghiasi et al. [18]. The method involves leveraging the SAM model to extract and reposition key components from existing images into new backgrounds.

A pre-trained YOLOv8 model was utilized as the starting point for the procedure. The model was configured to detect components and provide bounding boxes, which were then passed as prompts to the SAM model for segmentation. In cases where segmentation required refinement, manual inputs such as boundary clicks or region guidance were used to ensure high-quality masks.

After segmentation, the components were overlaid onto a variety of background images manually, selected to represent diverse visual settings. The placement of components was randomized in terms of location, scale, and orientation to mimic realistic spatial arrangements. 50 new images were created by the SAM-based method, using components and backgrounds from the original training data. These were added to the original training dataset, enlarging it to 210 images in total. The workflow of the SAM-based data synthesis augmentation method is illustrated in Figure 2.

This synthetic augmentation approach addresses data scarcity fundamentally differently from conventional augmentation techniques by generating novel training samples rather than merely transforming existing ones. While conventional methods such as rotation and scaling create variations constrained by the original dataset's limitations in scene diversity and contextual variations, the implemented strategy overcomes these constraints by extracting components from limited original data and systematically recombining them with diverse background environments, increasing the variety of visual contexts in which target objects appear. The SAM-based segmentation preserves authentic visual characteristics of target components while expanding contextual diversity, preventing common artifacts associated with simpler copy-paste approaches. This combinatorial expansion is particularly valuable in specialized domains where acquiring comprehensive real-world data is cost-prohibitive, allowing the model to develop robustness to domain-specific variations without requiring extensive additional data collection efforts.

C. Experimental Design for Augmentation Comparison

To assess the impact of different augmentation strategies, three experimental datasets were prepared: (1) the original dataset without augmentation, (2) the dataset enhanced using conventional image transformation methods, and (3) the dataset augmented through SAM-based synthesis.

Each dataset was used to train a YOLOv8 small (YOLOv8s) model under identical conditions, including the same hyperparameters and computing environment, as described in Section III-B. The evaluation metric used to compare performance was mean Average Precision (mAP) at an Intersection-over-Union (IoU) threshold of 0.5. The detection performance was analyzed for two specific object classes: battery modules and busbars.

Figure 3 shows the classification loss and box loss for training and validation over 100 epochs for two training runs. The model trained on the dataset created by SAM-based augmentation was selected as an example. The classification loss describes how well the model assigns the detected objects to the correct classes, while the box loss indicates how accurately the predicted bounding boxes are placed compared to the ground truth. A low loss generally indicates a more accurate prediction by the model. The classification loss curves show convergence, with training and validation curves closely aligned throughout the epochs. This indicates that the training process does not show signs of substantial overfitting. The box loss curves show a steady decrease, with the validation curves exhibiting additional fluctuations. The behavior of the curves indicates that the bounding box regression has not yet fully converged and may have required additional training epochs. Overall, the curves suggest that the classification of the detected objects can potentially be performed more reliably than the positioning of the bounding boxes.

D. Performance Comparison of Augmentation Methods

Table II presents the average precision (AP) values for both object categories and the overall mAP scores across all three datasets. The results indicate that the SAM-based augmentation strategy led to the highest detection performance, especially for the more difficult-to-detect busbar category.

The mAP score achieved with the SAM-based dataset was 0.744, reflecting an improvement of approximately 4.6% over the baseline and outperforming the conventionally augmented dataset by 2.0%. Notably, the busbar category showed an AP increase from 0.527 to 0.574, underscoring the value of synthetic augmentation in enhancing model sensitivity to subtle or low-contrast features. In summary, the SAM-based data synthesis approach demonstrated consistent advantages

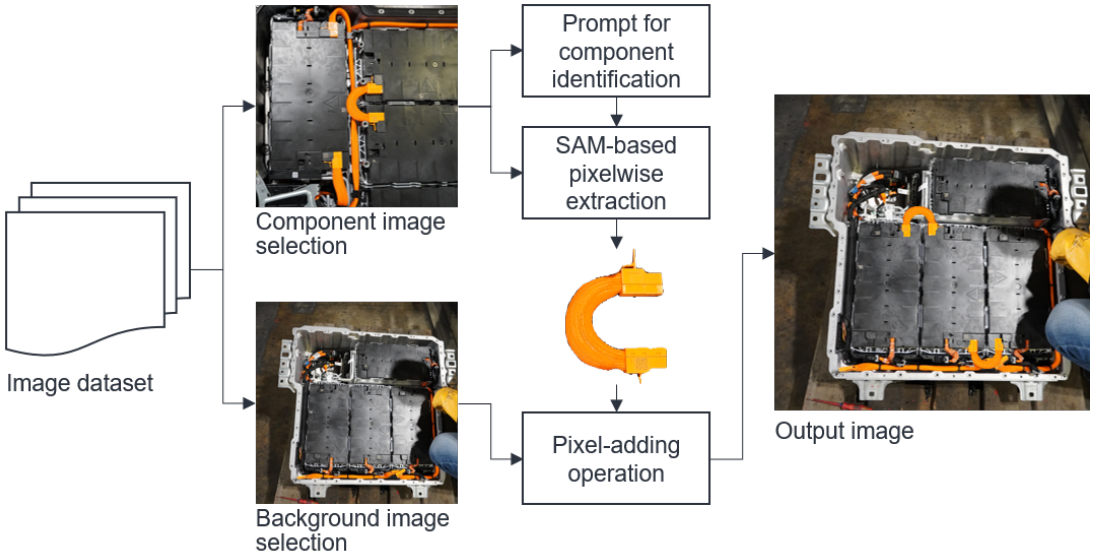

Fig. 2. Workflow of SAM-based data synthesis augmentation method

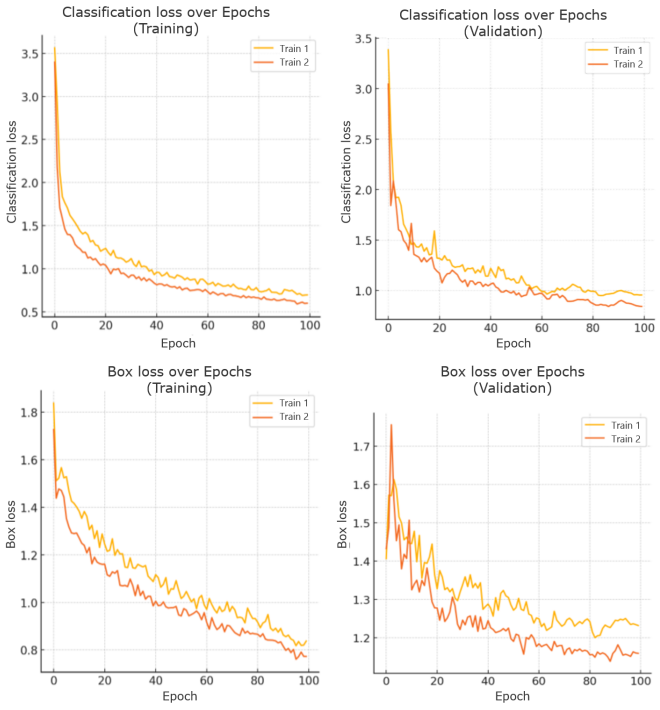

Fig. 3. Training and validation loss curves for classification loss (top row) and box loss (bottom row) over 100 epochs for two training runs (Train 1 and Train 2) of a YOLOv8s model with the dataset augmented through SAM-based synthesis

across all evaluation metrics, offering an effective strategy to augment limited datasets in industrial object detection scenarios. This method was particularly beneficial in improving model robustness and accuracy for complex component identification tasks.

TABLE II
COMPARISON OF DETECTION PERFORMANCE UNDER DIFFERENT DATA AUGMENTATION STRATEGIES

Augmentation Method	**AP** (Battery Module)	**AP** (Busbar)	**mAP@0.5**
Original Dataset (Baseline)	0.895	0.527	0.711
Conventional Augmentation	0.884	0.563	0.724
SAM-Based Synthesis	0.914	0.574	0.744

V. INSTANCE SEGMENTATION

To localize the components not only by bounding box coordinates but on a pixel-level, the object detection experiments were followed by an instance segmentation task which was carried out using the YOLOv8s-seg model. This model retains the foundational architecture of the YOLOv8s object detection model, including the backbone and the feature fusion neck, while incorporating an additional branch dedicated to producing segmentation masks. Due to the structural and functional similarities between the object detection and instance segmentation models, the validated training configuration from the detection experiments were transferred to this segmentation task. Moreover, the dataset enlarged by SAM-based synthesis was used, as it had already achieved better results in object detection. However, in this experiment, the class ”HV connections” was divided into the classes ”HV cables” and ”busbars” because the model missed the ability to generalize across the two types of HV connections, leading to the combined class showing low prediction accuracy.

To assess the model’s performance, Precision-Recall curves were analyzed for both bounding box predictions and segmentation masks. After completing 100 training epochs, the YOLOv8s-seg model achieved a strong performance overall:

- **Bounding Box Evaluation:** The model reached an overall mAP@0.5 of 0.847. Specifically, the battery module achieved an mAP of 0.930, the high-voltage cable scored 0.995, and the busbar obtained a relatively lower value of 0.615.
- **Segmentation Mask Evaluation:** The model achieved an overall mAP@0.5 of 0.822. In this evaluation, the battery module scored 0.924, the high-voltage cable maintained a high score of 0.995, and the busbar dropped to 0.548.

The results indicate that the model is particularly effective at identifying and delineating large, well-defined components such as battery modules and high-voltage cables. These components consistently exhibited high precision across both evaluation metrics, demonstrating the model's capacity for robust feature learning and boundary recognition.

In contrast, the busbar presented significant challenges. Its relatively low mAP scores in both evaluations suggest that this component's characteristics—small size, elongated form, and frequent occlusion by neighboring elements—limit the model's ability to capture its features accurately. The lower performance in segmentation mask evaluation (0.548) compared to bounding box evaluation (0.615) further highlights the increased difficulty in achieving precise pixel-level delineation for such fine structures.

Several contributing factors may explain this discrepancy. First, the busbar's minimal size and occlusion make it difficult for the model to acquire sufficient feature information during training. Second, segmentation tasks generally require high-resolution inputs to accurately represent fine component edges. Since a portion of the training data was derived from publicly available sources with limited image clarity, this likely hindered the model's ability to generalize fine-grained patterns associated with smaller components.

Despite these limitations, the YOLOv8s-seg model showed a solid overall performance. Its ability to generate both bounding boxes and detailed segmentation masks provides a strong foundation for extracting structural features essential for spatial and relational analysis within battery packs. Given these strengths, the instance segmentation model was selected as the final visual module for structural feature extraction, supporting a comprehensive and data-driven representation of battery pack assemblies.

VI. Structural Feature Extraction

In order to capture and model the internal structure of battery packs, this study investigated two complementary approaches to structural feature extraction. The first method, based on object detection outputs, infers relative spatial positions from bounding box relationships. The second, more detailed method leverages instance segmentation outputs to identify pixel-level contact and physical connections between components. Together, these techniques allow for a multi-layered understanding of the battery pack's structural configuration.

A. Bounding Box-Based Spatial Relationship Inference

Building upon the detection results produced by the YOLOv8s object detection model trained using the SAM-augmented dataset, this approach uses bounding box outputs to infer spatial layouts between different components.

Each detection includes a component label and a bounding box, which inherently provides a spatial reference. By analyzing the positional relationships and overlaps between these boxes, structural layouts such as "stacked" or "side-by-side" arrangements can be inferred. For example, in the case of battery modules and busbars, two common layouts were identified across the dataset: one in which busbars sit atop battery modules, and another in which they appear adjacent.

To quantify this spatial relationship, the overlap ratio between bounding boxes is computed as the area of intersection divided by the area of the smaller bounding box. The following rule set is applied:

- **Overlap Ratio** $r > 0.5$**:** Components are classified as having a "stacked" (above) relationship.
- **Overlap Ratio** $0 < r \leq 0.5$**:** Components are considered to be "side-by-side" (beside).

This method is illustrated schematically in Figure 4. However, the technique's accuracy is highly sensitive to the image's perspective. When images are captured from a top-down view, spatial inference is generally reliable. But with angled or distorted viewpoints, bounding boxes may misleadingly overlap, producing incorrect spatial classifications.

Furthermore, since bounding boxes provide only coarse localization, they lack the granularity needed to assess direct physical contact or actual connectivity between components. These limitations motivated the use of a finer segmentation-based method.

B. Pixel-Level Connection Analysis via Instance Segmentation

To address the limitations of bounding box-based spatial reasoning, this study employed a more refined instance segmentation approach using the YOLOv8s-seg model. This model generates pixel-level masks for each detected object, enabling a more nuanced understanding of component boundaries and contact points.

Connection relationships were identified based on two primary criteria:

- **Direct Mask Overlap:** If the segmentation masks of two components intersect at the pixel level, a connection is inferred.
- **Minimum Edge Distance:** If no overlap exists but the distance between the mask boundaries is below a threshold (15 pixels), a connection is still considered valid.

This dual-criteria approach ensures robustness against minor segmentation inaccuracies and occlusions. Spatial relationships ("stacked" or "side-by-side") continue to be inferred using bounding box overlap ratios, as in the previous method.

As shown in Figure 5, segmentation masks enable clear identification of components, their counts, spatial layouts, and physical connection patterns. In particular, the flexible and

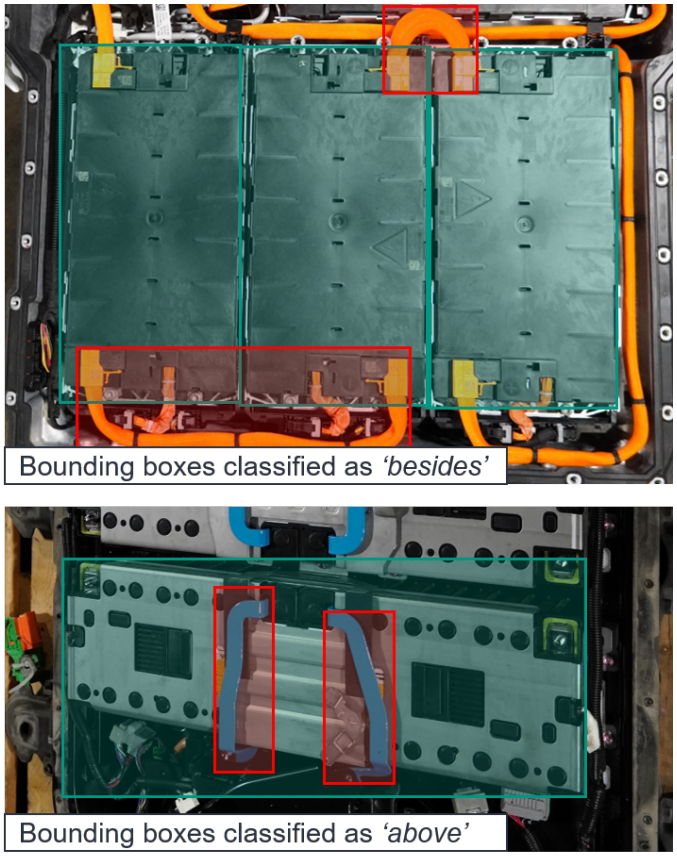

Fig. 4. Schematic illustration of spatial analysis of busbars (red) and modules (green). Top: Side-by-side layout of a BMW 330e battery pack; Bottom: Stacked layout of a Mitsubishi Outlander PHEV battery pack

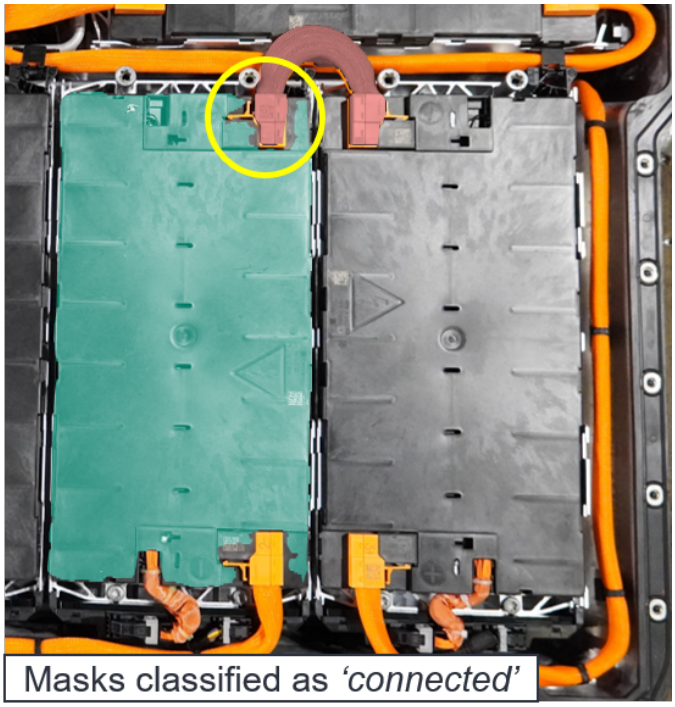

Fig. 5. Schematic illustration of the instance segmentation and connection analysis in a BMW 330e battery pack image

interconnected nature of high-voltage cables like they can be found in the BMW 330e battery image was accurately captured, which was not possible with bounding boxes alone.

The additional resolution provided by segmentation masks allowed the system to determine contact more precisely, improving the reliability of structural modeling. This proves especially valuable in scenarios involving compact, complex layouts or fine-grained component interactions.

In conclusion, the instance segmentation-based method provided more accurate and comprehensive structural feature extraction compared to the bounding box approach. By combining pixel-level contours with spatial heuristics, it enhances the system's ability to model real-world battery pack structures, laying a solid foundation for further disassembly reasoning and planning.

VII. Discussion

A. Assessment

While the proposed visual perception system achieved promising results, several important limitations must be acknowledged. The system demonstrated effectiveness through the YOLOv8-based detection and segmentation models, with the final YOLOv8s model achieving an overall mAP@0.5 of 0.744, particularly benefiting from SAM-based data augmentation for smaller or complex components. The segmentation masks proved more robust than bounding box overlap methods in identifying physical adjacency and contact relationships, especially under non-ideal viewing conditions.

However, a core challenge lies in the system's reliance on two-dimensional image data for structural feature extraction, which inherently lacks the ability to directly perceive depth information. Spatial reasoning based on surface-level cues such as object projection and relative positioning cannot accurately represent real-world physical distances or vertical arrangements, especially under conditions of occlusion or perspective distortion. While top-down imagery produced more reliable spatial classification results, angled views or limited resolution frequently led to misclassification, emphasizing the need for 3D data sources such as point clouds or stereo vision for precise modeling of adjacency and component stacking.

Additionally, the models performed well for large and distinct components like battery modules and high-voltage cables, but struggled with smaller or partially occluded elements such as small-sized busbars due to the variability and limited detail of publicly sourced training images. The current reasoning logic, while functional in constrained scenarios, relies on relatively simple geometric heuristics and lacks the flexibility and contextual depth of more advanced reasoning frameworks such as graph-based or deep learning relational models. Future iterations would benefit from enhanced reasoning capabilities and 3D depth-enabled methods to overcome these limitations.

In addition, the approach presented in this paper was only

tested on two different components that are mainly found in battery packs with a pack-to-module design. To enable a complete analysis of a battery pack, the object detection and segmentation models would have to be trained on additional classes, which would increase the effort required to create the underlying data set. If additional pack designs such as cell-to-pack were to be considered, it would potentially be advantageous to create a separate data set for each design and train a separate model due to the differences in the visual appearance of the pack designs.

Notwithstanding, the integrated system effectively demonstrated its ability to detect, segment, and interpret the spatial configuration of battery components. This framework constitutes the visual basis for a more extensive, information-driven disassembly framework, with applications in structural modeling, knowledge representation, and task planning.

B. Outlook

Looking ahead, several directions for refinement and expansion can be derived. First, incorporating high-resolution cameras and depth-aware sensors such stereo cameras could help resolve ambiguities in detecting small-sized components, vertical positioning and occlusion.

In terms of reasoning capabilities, transitioning from heuristic-based to learned relational models such as graph neural networks could provide the contextual depth necessary for complex structural analysis. This would enable flexible sequencing of disassembly operations even for unknown battery packs.

Furthermore, future experiments should extend beyond static image testing and incorporate dynamic, real-world disassembly environments. This would allow for a more rigorous evaluation of the system's robustness, adaptability, and real-time performance under practical constraints.

Finally, the proposed computer vision-based system could be integrated into automated disassembly lines, enabling real-time decision-making and dynamic adaptation to variable battery layouts. In order to accomplish this, the real-world coordinates of the components and joints, as well as feasible disassembly steps, must be extracted from the image data. Then, they can be transferred to a system for robotic path and task planning and ultimately contribute to the flexible, autonomous execution of disassembly operations.

References

[1] IEA, "Global EV Outlook 2024," IEA, Paris, Tech. Rep., Apr. 2024. [Online]. Available: https://www.iea.org/reports/global-ev-outlook-2024/trends-in-electric-cars

[2] K. Meng, G. Xu, X. Peng, K. Youcef-Toumi, and J. Li, "Intelligent disassembly of electric-vehicle batteries: a forward-looking overview," *Resources, Conservation and Recycling*, vol. 182, p. 106207, Jul. 2022. doi: 10.1016/j.resconrec.2022.106207 .

[3] A. Beaudet, F. Larouche, K. Amouzegar, P. Bouchard, and K. Zaghib, "Key Challenges and Opportunities for Recycling Electric Vehicle Battery Materials," *Sustainability*, vol. 12, no. 14, p. 5837, Jan. 2020. doi: 10.3390/su12145837 .

[4] S. Wu, N. Kaden, and K. Dröder, "A Systematic Review on Lithium-Ion Battery Disassembly Processes for Efficient Recycling," *Batteries*, vol. 9, no. 6, p. 297, Jun. 2023. doi: 10.3390/batteries9060297 .

[5] M. Beghi, F. Braghin, and L. Roveda, "Enhancing Disassembly Practices for Electric Vehicle Battery Packs: A Narrative Comprehensive Review," *Designs*, vol. 7, no. 5, p. 109, Oct. 2023. doi: 10.3390/designs7050109 ., publisher: Multidisciplinary Digital Publishing Institute.

[6] A. Al Assadi, T. Götz, A. Gebhardt, O. Mannuß, B. Meese, J. Wanner, S. Singha, L. Halt, P. Birke, and A. Sauer, "Automated Disassembly of Battery Systems to Battery Modules," *Procedia CIRP*, vol. 122, pp. 25–30, Jan. 2024. doi: 10.1016/j.procir.2024.01.005 .

[7] E. Gerlitz, M. Greifenstein, J. Hofmann, and J. Fleischer, "Analysis of the Variety of Lithium-Ion Battery Modules and the Challenges for an Agile Automated Disassembly System," *Procedia CIRP*, vol. 96, pp. 175–180, Jan. 2021. doi: 10.1016/j.procir.2021.01.071 .

[8] M. Choux, E. Marti Bigorra, and I. Tyapin, "Task Planner for Robotic Disassembly of Electric Vehicle Battery Pack," *Metals*, vol. 11, no. 3, p. 387, Feb. 2021. doi: 10.3390/met11030387 .

[9] Y. Lu, M. Maftouni, T. Yang, P. Zheng, D. Young, Z. J. Kong, and Z. Li, "A novel disassembly process of end-of-life lithium-ion batteries enhanced by online sensing and machine learning techniques," *Journal of Intelligent Manufacturing*, vol. 34, no. 5, pp. 2463–2475, Jun. 2023. doi: 10.1007/s10845-022-01936-x .

[10] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[11] M. Zorn, C. Ionescu, D. Klohs, K. Zähl, N. Kisseler, A. Daldrup, S. Hams, Y. Zheng, C. Offermanns, S. Flamme, C. Henke, A. Kampker, and B. Friedrich, "An Approach for Automated Disassembly of Lithium-Ion Battery Packs and High-Quality Recycling Using Computer Vision, Labeling, and Material Characterization," *Recycling*, vol. 7, no. 4, p. 48, Aug. 2022. doi: 10.3390/recycling7040048 .

[12] D. Klohs, C. Offermanns, H. Heimes, and A. Kampker, "Automated Battery Disassembly—Examination of the Product- and Process-Related Challenges for Automotive Traction Batteries," *Recycling*, vol. 8, no. 6, p. 89, Dec. 2023. doi: 10.3390/recycling8060089 .

[13] K. Man and J. Chahl, "A Review of Synthetic Image Data and Its Use in Computer Vision," *Journal of Imaging*, vol. 8, no. 11, p. 310, Nov. 2022. doi: 10.3390/jimaging8110310 .

[14] E. M. Bigorra, "Design and Implementation of a Computer Vision System for Robotic Disassembly of Electric Vehicle Battery Pack," Master's thesis, University of Agder, 2020, accepted: 2020-10-02T09:31:01Z Publication Title: 150. [Online]. Available: https://uia.brage.unit.no/uia-xmlui/handle/11250/2680868

[15] E. Gerlitz, L.-E. Enslin, and J. Fleischer, "Computer vision application for industrial Li-ion battery module disassembly," *Production Engineering*, vol. 18, no. 3, pp. 393–401, Jun. 2024. doi: 10.1007/s11740-023-01231-5 .

[16] H. Zheng, S. Liu, H. Zhang, J. Yu, and J. Bao, "Visual-triggered contextual guidance for lithium battery disassembly: a multi-modal event knowledge graph approach," *Journal of Engineering Design*, pp. 1–26, Jan. 2024. doi: 10.1080/09544828.2024.2301876 .

[17] Q. Guo, S. Wang, C.-P. Chang, and J. Rambach, "CACP: Context-Aware Copy-Paste to Enrich Image Content for Data Augmentation," in *Proceedings of the 1st Workshop on Exploring the Next Generation of Data. CVPR Workshop on Exploring the Next Generation of Data (NeXD-25)*. Nashville, TN, United States: IEEE, 2025.

[18] G. Ghiasi, Y. Cui, A. Srinivas, R. Qian, T.-Y. Lin, E. D. Cubuk, Q. V. Le, and B. Zoph, "Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation," 2020. doi: 10.48550/ARXIV.2012.07177 .