scientific data



OPEN

DATA DESCRIPTOR

# RecyBat24: a dataset for detecting lithium-ion batteries in electronic waste disposal

Ximena Carolina Acaro Chacón[1,2]✉, Fabrizio Lo Scudo[1,2]✉, Gregorio Cappuccino[1] & Carmine Dodaro[1]

**In recent years, deep learning techniques have been extensively used for the identification and classification of lithium-ion batteries. However, these models typically require a costly and labor-intensive labeling process, often influenced by commercial or proprietary concerns. In this study, we introduce RecyBat24, a publicly accessible image dataset for the detection and classification of three battery types: Pouch, Prismatic, and Cylindrical. Our dataset is designed to support both academic research and industrial applications, closely replicating real-world scenarios during the acquisition process and employing data augmentation techniques to simulate various external conditions. Additionally, we demonstrate how the RecyBat24's detection-oriented annotations can be used to create a second version of RecyBat24for instance-segmentation tasks. Finally, we demonstrate that recent lightweight machine learning models achieve high accuracy, highlighting their potential for classification and segmentation applications where computational resources are constrained.**

## Background & Summary

The global production capacity of Lithium-Ion batteries (LIBs) is anticipated to exceed 1.3 TWh by 2030[1]. However, the forecast demand for these batteries is expected to exceed the supply of rare metal resources[2]. Furthermore, the average lifespan of LIBs currently ranges from 5 to 8 years, resulting in a substantial increase in the number of decommissioned batteries in various nations. Ineffective management of this surplus could lead to significant resource wastage and environmental harm. Therefore, the development of advanced battery recycling technologies is crucial[3]. The imperative for recycling is further underscored by the necessity of reclaiming valuable materials.

The retrieval of LIBs from electronic waste is extremely important, particularly concerning the extraction and reuse of valuable materials. LIBs contain considerable amounts of rare elements that can be recovered using specialized recycling methods[4]. In particular, LIBs contain critical materials such as lithium, cobalt, nickel, and manganese, all of which are essential for manufacturing new batteries and other electronic components[5,6].

Given the growing demand for these materials in the production of electric vehicles, renewable energy storage systems, and portable electronics, recycling batteries provides a sustainable solution to resource scarcity. By extracting these materials from spent batteries, the need to extract new raw materials is significantly reduced, which is not only environmentally damaging but also geopolitical complex due to the concentration of these resources in a few countries. However, for the extraction of these materials, the correct identification of LIBs is important.

In the preliminary stage of the recycling process, human-assisted identification of battery types is typically used. However, this method may prove economically impractical as the recycling industry grows[1]. Thus, it is crucial to develop an automated and cost-effective system capable of sorting battery types based on readily accessible field data[7]. With the extensive adoption of Artificial Intelligence (AI) tools, there is an increasing tendency to apply machine learning, particularly deep learning (DL) techniques, to address this challenge. Furthermore, machine learning has been successfully implemented in various battery-related applications, such as predicting the remaining useful life[8–11] and evaluating the state of health[12,13].

[1]University of Calabria, Arcavacata, Italy. [2]These authors contributed equally: Ximena Carolina Acaro Chacón, Fabrizio Lo Scudo. ✉e-mail: carolina.acaro@dimes.unical.it; fabrizio.loscudo@unical.it

Within battery recycling research field, the usage of DL techniques is a somewhat under-researched area, with existing studies primarily falling into two main categories: methods that leverage features derived from LIBs and techniques based on computer vision (CV).

In the first category, we mention a recent study by Tao *et al.*[3] which proposes a federated machine learning approach to mitigate privacy concerns among recycling partners of different organizations. The authors suggest a sorting model that relies solely on a single cycle of end-of-life charging and discharging data rather than historical data, while still maintaining the data privacy budgets of various battery recycling partners. Previously, Garg *et al.*[14] outlined an approach for assessing the suitability of retired lithium batteries for secondary use that involves a thorough evaluation of several parameters such as capacity, resistance, voltage, and temperature. The topic of organizing LIBs according to specific attributes and performance is also discussed in the works[15–17], which employs a multistage deep sorting strategy derived from the clustering of static and dynamic features.

In the second category, which aligns more closely with the objective of this work, Liu *et al.*[18] introduce a disassembly platform enhanced by online sensing and DL technologies. Although they use CV techniques, their study is limited by the size of the presented dataset. Furthermore, the authors have not disclosed any code or data related to their study. BatSort, a transfer learning-based solution for automatic battery type classification, was recently presented in[19]. The researchers claimed that they have created a unique dataset containing more than 500 images covering 9 different battery types as a case study. However, a thorough examination of the released dataset shows that it mainly comprises a single category of battery, specifically cylindrical, with images sourced from the Internet depicting brand-new batteries. We notice that the features of these images do not closely match the typical conditions found in real-world battery recycling environments. Finally, Ueda *et al.*[20] introduced an integrated sorting system capable of identifying batteries within electronic waste through the application of X-ray imaging and DL methodologies. A key shortcoming of this method pertains to X-ray scanners, which encounter difficulties when dealing with highly dense materials and small batteries.

In general, most of the studies in this discussion emphasize the prevalent challenge of limited data availability. We hypothesize that this scarcity of image data is the primary reason for the slow adoption of DL solutions in LIB recycling. In fact, the success and effectiveness of statistical models are significantly dependent on the quality of the datasets used[21,22]. This is especially valid in the context of CV settings[23]. However, the creation of high-quality datasets often requires manual labeling, a process that is not only labor intensive and costly, but also prone to errors[24], and significant privacy issues arising from commercial or proprietary interests.

Although a dataset's quality is frequently linked to the quality of its annotations, aspects such as its size and variety also greatly influence it. The initial step in constructing a dataset involves a data preparation stage, where data are discovery and cleaned[25]. This procedure can be quite costly, depending on the context of its application. Subsequently, after gathering the samples, this data often requires annotation[26]. An affordable and scalable method for gathering large quantities of annotated data is crowd-sourcing[27]. This approach is effective in straightforward scenarios where annotation tasks do not necessitate specialized knowledge or involve data privacy concerns. In specialized fields like medicine, annotators must possess a specific level of expertise. Consequently, this often leads to high levels of disagreement, which results in labels that are more subjective and inconsistent[28].

In addition to concerns associated with the labeling process, the dataset's size also influences the model's performance[29]. Although, overall, the key aspect lies in how well a dataset reflects the original distribution, as opposed to its magnitude. The main issue here is that each model needs a different amount of data in order to maximize its performance. In the domain of DL, we can certainly leverage transfer learning[30] by employing models that have been trained on extensive datasets. This would diminish the requirement for an extensive dataset because the *low-level* features, since we are working with natural images, are presumed to have been already captured by the initial layers of the models.

Finally, sample diversity is crucial for enhancing a model's robustness and generalizability. Recently, Yang *et al.*[31] highlighted the importance of incorporating varied datasets during training to enhance the robustness and adaptability of building extraction models in various settings. Within the scope of this study, the research presented in[32,33], although primarily focused on novel techniques for predicting battery lifespan using tabular data, offers nevertheless valuable insights on the quality of various datasets on batteries.

Thus, data scarcity poses a major challenge for training DL models because a large amount of data is necessary to achieve industry-standard performance.

This study aims to tackle the challenges posed by limited data availability by presenting a publicly accessible dataset, named RecyBat24, specifically designed to train DL models in the field of lithium-ion battery recycling. This dataset features a comprehensive and diverse compilation of samples, accompanied by a high-quality labeling process for object detection tasks. Additionally, we illustrate the application of this detection-centered annotation for crafting a subsequent version of RecyBat24tailored to instance segmentation task.

However, despite the fact that we have devoted considerable effort, RecyBat24inherently has limitations regarding its coverage of the real data distribution. The primary issue lies in the variation of LIB designs among manufacturers, as each individual design incorporates unique configurations, materials, and assembly methods. This requires DL models to adjust to a diverse range of structures for accurate identification. A second challenge arises as a result of the aging of LIBs. For example, physical deformations may affect both the appearance and structural stability of battery components. Therefore, the model must be able to adapt to a range of conditions, encompassing different types of damage.

Although we acknowledge that these factors are less significant compared to their impact on LIB volumetric imaging tasks, as discussed in[34], they nevertheless poses a significant barrier to industrial implementation. Still, as far as we are aware, RecyBat24is the first publicly accessible dataset comprising natural images of disposal LIB. Consequently, it serves as a concrete asset for applied research, which will facilitate future industrial applications.

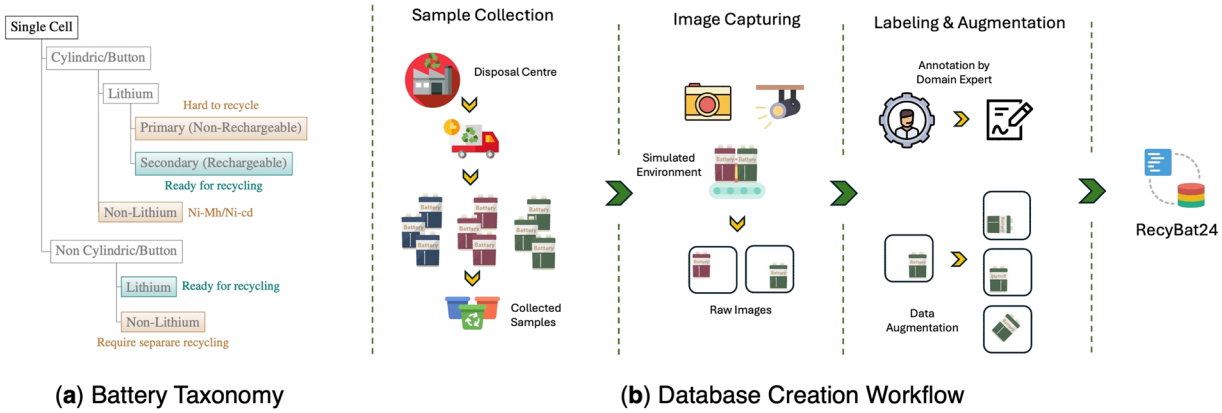


**Fig. 1** (**a**) The Battery Taxonomy created to facilitate the battery type selection to be used in RecyBat24. With *Non Cylindric* we refer to the cell types Prismatic and Pouch. **(b)** Visual representation of the workflow used to build RecyBat24.

In addition, issues related to aging and occlusions can be addressed with recent advances in the field of DL. Data augmentation techniques that introduce artificial occlusions can enhance the robustness of deep learning models by enabling them to generalize effectively to scenarios where critical features of battery components are partially obscured. Similarly, the production of synthetic data using modern AI models that mirror the physical attributes of old or deteriorated batteries aids in the creation of more robust models. They should generalize better across batteries at various stages of their life-cycle, thereby enhancing their utility in practical recycling and diagnostic applications.

## Methods

RecyBat24is created as a dataset intended to aid in identifying and categorizing LIBs. In building this dataset, we examine various battery types, including those without Lithium, and categorize them into a taxonomy, depicted in Fig. 1a. This taxonomy was designed to facilitate the selection of discarded batteries. Our main priority in this study is rechargeable LIBs, as they contain important materials such as lithium, cobalt, and nickel. Consequently, we exclude non-Lithium batteries and *Primary* or non-rechargeable Cylindric batteries because they involve non-intercalating lithium metal and reactive electrolytes. Thus, primary batteries contain lithium in metallic forms; this makes them more challenging to recover. Secondary batteries, instead, are equipped with cathode materials such as NMC (Nickel-Manganese-Cobalt), LFP (Lithium Iron Phosphate), and LCO (Lithium Cobalt Oxide), which can be recycled through pyrometallurgical, hydrometallurgical, and direct recycling techniques. From an economic perspective, lithium-ion battery recycling is viable because of the higher concentration of reusable materials compared to primary batteries[35].

Once the target cells are selected, we devise the workflow to build a dataset that can enhance both research and industrial efforts to make the recycling process more efficient. The workflow involves three primary stages: Sample Collection, Image Capturing, and finally Labeling and Augmentation. We provide a more detailed explanation of each stage in the following sections. A visual representation is shown in Fig. 1b.

**Samples Collection.** In electronic waste disposal centers, individual LIBs are commonly found due to several factors related to the lifecycle and disposal practices of electronic devices. Many electronic devices, such as laptops, smartphones, power tools, and tablets, rely on LIBs for power. Over time, these batteries degrade and lose their capacity to hold a charge, prompting users to replace them. When devices are discarded, the batteries often remain inside, but in some cases, batteries are removed before disposal either by users or during preliminary recycling processes. The growing awareness and regulations around e-waste recycling have led to more organized collection efforts.

Nowadays, many e-waste recycling programs and drop-off centers specifically ask consumers to separate batteries from electronic devices to avoid potential hazards. LIBs, if damaged or improperly handled, can pose significant safety risks, including fires and chemical leaks. By separating these batteries from the devices, recycling centers can manage these risks more effectively.

In general, the presence of individual LIBs in electronic waste disposal centers is a reflection of both consumer behavior and industry practices. The separation and collection of these batteries are driven by safety considerations, regulatory requirements, and the economic value of recycling the materials they contain.

The initial phase of building RecyBat24involved the gathering of various samples of LIBs. Our goal was to ensure that RecyBat24encompasses a wide range of battery conditions, types, and degradation levels commonly found. However, this task proved to be quite intricate due to safety and regulatory compliance. For example, as the distance from the disposal center increases, expenses related to the delivery and handling of used LIBs tend to rise significantly. This limiting factor forced us to narrow down the list of available centers and select the one that could provide us the largest variety of batteries according to the information obtained from the European and Italian Consortium. As a result, automotive cells are absent from our dataset because they are only available at a few centers that have formed exclusive agreements with manufacturers.


**Fig. 2** Raw samples of batteries present in the dataset. All capture settings, including a dark background and fixed camera position, are designed to mimic real-world operational conditions.

From the selected disposal center, we collected a total of 110 batteries, comprising 59 Pouch-type, 22 prismatic, and 29 cylindrical batteries. The spectrum of collected samples includes those in excellent working order with minimal visible defects, to those that exhibit notable signs of deterioration. Unfortunately, we could not gather batteries of the button variety.

Pouch-type batteries, known for their flexible and lightweight design, were the most numerous in the collection. These are commonly found in consumer electronics such as smartphones and tablets. Prismatic batteries, typically sourced from laptops and some electric vehicles, are characterized by their rectangular shape and robust construction. We claim that these two types are well-represented in RecyBat24, and the primary differences with batteries outside the dataset can be typically ascribed to the external color of the casing, as their size and shape conform to standardized specifications.

The other battery class that was collected is the Cylindrical 18650 lithium-ion. These batteries are rechargeable cells with a standardized size of 18 mm in diameter and 65 mm in length. They are widely used in portable electronics, electric vehicles, and energy storage systems, because they offer high energy density, long cycle life, and stable performance.

As previously mentioned, while the collection of batteries from the disposal center is not comprehensive, it should nonetheless provide a representative sample of various battery types and their realistic conditions. Moreover RecyBat24, as the first publicly accessible LIBs dataset for training and evaluating DL models, is ideally suited to gauge insights about model performance in a broad range of real-world scenarios. In Fig. 2, we illustrate the three types of batteries collected from the recycling center.

**Image Capturing.** In order to enhance the practical utility of the dataset, we constructed an experimental setup that closely mimics the actual conditions found in electronic waste recycling settings. The arrangement involved placing the batteries in ways that simulate typical e-waste situations, while also including a variety of lighting scenarios. These steps were implemented to guarantee the dataset's resilience and its wide-ranging applicability across different recycling scenarios.

For image capture, we used a standard full HD 1080p camera, which included an adjustable ring light and a CMOS photo sensor (The camera's features listed a 3x maximum focal length, an aperture setting of f/2.1, and a screen measuring 2 inches). Different scenarios are simulated by employing three distinct lighting conditions: low, medium, and high. During training, these variations in illumination offer valuable signals to the models, enhancing their ability to distinguish effectively under the varied lighting conditions usually found in industrial settings.

We recreated industrial environments by placing batteries at random on a conveyor belt simulated by a dark background. Every photograph was captured in a square format with a resolution of $640 \times 640$ pixels, keeping a steady distance of 40 cm from the battery to the camera to guarantee consistent image quality and perspective. Improving image uniformity increases the applicability of RecyBat24to DL applications while effectively replicating the conveyor belt environments where batteries are usually classified and handled.

To create the training and evaluation sets, we manually built two distinct subsets from all captured images using the following protocol. For a battery with two separate sides, one side was assigned to the training dataset, while the other was designated for the evaluation dataset. For batteries lacking distinct sides, we chose half of the samples from every battery category at random for the training dataset, while the rest were used to construct the evaluation dataset.

The filenames were structured as *id_n_F/R_LightLevel_ClassType. jpg*, where *id* signifies the sample ID, *n* identifies the specific image number of that sample {1, 2, 3, 4, 5}, *F/R* indicates whether the front or rear battery is shown, *LightLevel* describes the lighting condition as low, medium, or high (range in {1, 2, 3}), and $ClassType \in \{cyl, po, pri\}$ specifies the battery classification, being cylindrical, Pouch or prismatic.

| Class label | Cylindric | Pouch | Prismatic |
|---|---|---|---|
| Originals Images | 435 | 1740 | 660 |
| Augmented Images | 5220 | 20880 | 10560 |
| Image Size | 640 × 640 | 640 × 640 | 640 × 640 |

**Table 1.** The table provides a breakdown of the dataset by individual images. An augmentation process is applied to the original images to boost the number of samples in the dataset.

**Labeling and Augmentation.** The images collected belong to three types of batteries: cylindrical, Pouch, and prismatic. In total, 2835 images were collected, comprising 660 prismatic, 1740 Pouch, and 435 cylindrical batteries. No distinction was made with respect to battery size or state (such as new, used, or damaged).

The annotation process was performed by an expert who carefully specified the battery type and location for each image. In particular, the annotation operation consists of assigning a specific label to a spatial region within the image. The annotation of images was performed effectively using publicly accessible annotation tools such as LabelMe (https://github.com/labelmeai/labelme) and Roboflow (https://roboflow.com). These tools were used to generate bounding boxes and class labels for different types of batteries.

The COCO format has been used for the release of the final annotations of RecyBat24. Annotations in the COCO format are stored within a *.json* file. Each annotation contains the category label, bounding box coordinates, and image size specifications (height and width).

Data augmentation techniques[36] were also employed to mitigate the challenges associated with acquiring a large and diverse dataset. By generating variations of existing images, these techniques enable the model to be exposed to a broader range of examples, thereby reducing overfitting while simultaneously enhancing robustness and overall performance.

We selected augmentation techniques to simulate potential real-world differences in battery appearance, including aspects such as dirtiness and occlusions. In addition to implementing fundamental rotation techniques for the development of rotation-invariant features, we chose to incorporate brightness modification and exposure alteration to address various lighting conditions. These transformations ensured that the essential characteristics of the primary objects were subtly modified, thereby preserving the critical informative signal required for training while simultaneously increasing the dataset.

To mimic potential battery overlap or occlusion scenarios, we explored traditional methods, such as CutOut[37] or Random Erasing[38], which involve randomly obscuring square regions of the input image throughout the training process. However, it became apparent to us that CutOut and Random Erasing can be excessively aggressive, erasing essential features. Generally speaking, any technique that directly eliminates or obscures portions of an object needs careful calibration before application. Therefore, we chose to simulate a type of soft occlusion using *mixup*[39]. This approach can help to manage occlusion indirectly by presenting the model with blended samples that integrate mixed pixel values from various classes. This blending result can result in situations where essential aspects of one object are somewhat hidden by another. Additionally, the objects in the combined images might resemble a state of dirtiness, which should facilitate the model to generalize in real-word application. It is important to note that the dataset itself does not contain pre-generated mixup-ed images. Instead, these blended samples are dynamically constructed at the batch level during training, ensuring that the augmentation is applied adaptively throughout the learning process.

Table 1 shows the overall count of images and associated labels for each type of battery.

Subsequently, the annotated images were cross-verified by an independent expert to identify and correct discrepancies. The used quality control protocol required all batteries to be enclosed within their designated boxes and the segmentation to be precise, with no visible artifacts or inaccuracies. Regarding the latter criterion, due to the absence of a reference standard, it is not possible to report or discuss any quantitative evaluation metrics. This evaluation led to the exclusion of seven images due to minor annotation errors or insufficient image quality. Automated validation scripts were also employed to detect common annotation errors such as missing labels or incorrect bounding box formats. The expert further assessed the consistency of the automatically generated segmentation masks, which we will discuss later. Finally, an additional inspection was performed by the same expert on randomly selected samples from the augmented dataset to evaluate the overall quality of the annotation, thus guaranteeing that the dataset meets the criteria necessary for the development of DL applications.

## Data Records

The RecyBat24dataset is publicly accessible via Zenodo[40]. This dataset comprises exclusively non-augmented images, with a total of 1421 samples in the *train* directory and 1407 in the *val* directory. An extended version, referred to as RecyBat24-aug, includes 18370 training samples and 18206 validation samples. Users can download the dataset from the repository and conduct experiments locally on their own environments. Table 2 provides information on the structure of the datasets.

The repository contains the files as given in Table 3:

## Technical Validation

**Classification Benchmark.** To evaluate the efficacy of various object detection models in the dataset RecyBat24, a standardized evaluation protocol has been formulated, which covers two recent object detection methodologies. This framework is designed to deliver a consistent metric for the evaluation and comparison of the detection capabilities of advanced models. For an extensive review of recent advances in object detection, the reader is referred to the complete survey[41].

| | Train | | | | Val | | | |
|---|---|---|---|---|---|---|---|---|
| | *Tot.* | *Cylindric* | *Pouch* | *Prismatic* | *Tot.* | *Cylindric* | *Pouch* | *Prismatic* |
| RecyBat24 | **1421** | 225 | 866 | 330 | 1407 | 207 | 870 | 330 |
| RecyBat24-aug | **18370** | 2700 | 10390 | 5280 | 18206 | 2484 | 10442 | 5280 |

**Table 2.** Details of the structure of the datasets.

| File Name | Description |
|---|---|
| *classification.zip* | This file holds the configuration files necessary for conducting classification experiments. |
| *recybat24.tar.gz* | This file contains the compressed version of the RecyBat24dataset. It features two primary directories, *train* and *val*, each including two annotation files: *annotations.json* provides classification annotations, while *instance_segmentation_annotations.json* includes annotations for the instance segmentation task. |
| *recybat24_aug.tar.gz* | This file contains the compressed version of the RecyBat24-aug dataset. Similar structure as above. |
| *Readme.md* | This file contains the details information to perform the experiments. |
| *segmentation.zip* | This file includes a script to convert the provided dataset from COCO format to YOLO segmentation format required from Ultralytics library. The other file conducts the segmentation test utilizing Fast SAM. |

**Table 3.** Overall structure of the RecyBat24repository.

Since 2012, the advent of DL has opened a new phase for computer vision and specifically convolutional neural networks (CNN), allowing the acquisition of complex and flexible representations of visual features of images[42]. This development led to the evolution of object detection techniques into two major categories: two-stage detectors and one-stage detectors.

The two-stage detectors handle the detection task by moving from coarser analysis to more detailed analysis. The introduction of Region-based CNN (RCNN) by Girshick *et al.*[43] marks a pivotal development in this approach. The fundamental concept of RCNN entails first producing a collection of object proposals (candidate regions) using selective search[44]. These proposals are subsequently resized to a standard image size and evaluated by a CNN model trained on ImageNet[42] to derive pertinent features. Following this, a linear SVM classifier analyzes each region's features to identify and classify the objects present. Fast R-CNN[45] enhanced this approach by allowing the concurrent training of a detector and a bounding box regressor within a unified network structure, even though its speed was still hindered by the proposal detection stage. To overcome this limitation, Faster R-CNN[46] incorporated a region proposal network (RPN) to more effectively create region proposals. The evolution from RCNN to Faster R-CNN involved merging different parts of the object detection system, such as proposal generation, feature extraction, and bounding-box regression, into a single end-to-end learning architecture. Although this coarse-to-fine processing technique yields high precision-with the coarse stage enhancing recall and the fine stage refining localization and separation-these methods are rarely used in real-world applications. Their complexity and slow speed prevent their applicability in environments with limited resources. Conversely, single-stage detectors simplify the procedure by identifying all objects in a single inference step.

One-stage object detection models represent a class of detection frameworks that perform the detection process in a single stage, bypassing the region proposal phase characteristic of two-stage models. Instead, they execute detection directly across a dense grid of potential locations. These models generally provide faster inference times; however, this speed often comes at the expense of detection accuracy.

YOLO[47], an acronym for *You Only Look Once*, is one of the most widely adopted single-stage methods in the field. YOLO is distinguished by its remarkable speed, achieved using a single neural network to process the entire input image. The network divides the image into a grid and concurrently predicts the bounding boxes and associated probabilities for each grid cell. Although YOLO significantly accelerates detection, it generally exhibits reduced localization accuracy relative to two-stage detectors, especially for smaller objects. This limitation has been addressed in subsequent iterations of the YOLO framework[48,49]. Currently, an enhanced version has been released as the next iteration of YOLO, named YOLOX[50].

Another notable single-stage method is RTMDet[51], which represents an advanced generation of real-time object detection models. RTMDet is designed to surpass the YOLO series in both efficiency and versatility, encompassing tasks such as instance segmentation and rotated object detection. The architecture of RTMDet features large-kernel depth-wise convolutions, which enhance its ability to capture global context while maintaining a balance between the model's depth and width. This design choice ensures rapid inference speeds without requiring re-parameterization of the model. Furthermore, RTMDet employs soft labels for dynamic label assignment, which improves accuracy by facilitating high-quality matching and reducing label noise. This results in superior performance metrics, including 52.8% AP on COCO at speeds exceeding 300 FPS, thus outpacing existing detectors in terms of both speed and precision. Additionally, RTMDet can be readily adapted for tasks such as instance segmentation and rotated object detection with minimal modifications.

Object detection has recently incorporated Transformer[52], a relatively modern architecture widely used in various contexts. By integrating this novel architecture, it becomes possible to train the detector without relying on anchor boxes or anchor points. In 2020, Carion *et al.* redefined the object detection task as a set prediction challenge and introduced an end-to-end detection network named DETR[53]. When presented with an image, the model is tasked with forecasting an unordered set (or list) of all the objects present, each denoted by its class, alongside a precise bounding box encompassing each object. The authors combine a convolutional neural

| Model Type | Input Shape | FLOPs | Parameters |
|---|---|---|---|
| *Yolo X* | 416 × 416 | 3.2G | 5.0M |
| *RTM Det* | 640 × 640 | 8.0G | 4.8M |
| *RT-DERT-l* | 640 × 640 | 103.4G | 32.0M |

**Table 4.** Details of the model versions used in the experiments. The FLOPs count is estimated by the library and highlights the difference in the computational costs between the models.

network (CNN) for extracting local information from the image with a Transformer encoder-decoder architecture to analyze the image as a whole and subsequently make the prediction. Recently, Zhao *et al*. introduced the Real-Time DEtection TRansformer (RT-DETR)[54], a solution designed to overcome the challenges of high computational expenses that restrict the practicality of earlier Transformer-based approaches. In their study, the researchers enhanced the DETR framework by crafting an efficient hybrid encoder to substitute the standard Transformer encoder. This innovation noticeably accelerates inference speed by separating intra-scale interactions from cross-scale feature fusion across various scales. Additionally, RT-DETR allows for adaptable speed adjustments to fit different real-time applications without the need for retraining.

For the purposes of this study, RT-DETR's performance was compared with that of YOLOX and RTMDet, as all models are readily applicable to real-world scenarios. We conducted our experiments only on the augmented version of RecyBat24.

*Experimental settings.* In our experiments, we use the implementations of the selected architectures provided by MMDetection[55] and Ultralytics[56]. MMDetection, an open-source library developed by the Multimedia Laboratory (MMLab) at the Chinese University of Hong Kong,supports a variety of computer vision tasks, and is used since it offers an optimized implementation of YOLOX and RTMDet. This library is designed with user accessibility and adaptability in mind, supporting a range of detection frameworks that include single-stage and two-stage detectors. Its modular architecture enables users to construct custom models by integrating various components, such as the backbone, neck, and heads (In contemporary neural architecture designs for CV, the network is often conceptually divided into three parts: the backbone, neck, and head. The neck functions as an intermediate processing stage, in which features extracted by the backbone are further refined before final processing by the head). The library offers a wide array of cutting-edge pre-trained models along with extensive documentation, establishing itself as a valuable resource for both academic research and real-world implementations in the field of CV. At the time of writing this document, RT-DERT was not yet implemented in MMDetection. Consequently, we used the code provided by Ultralytics for our experiments, as both Ultralytics and MMDetection share similar objectives in the field of applied CV.

Finally, given our focus on lightweight architectures, we opt for the *tiny* version for YOLOX and RTMDet models, and the smallest available version for RT-DERT. The main characteristics are reported in Table 4.

The experiments were conducted on a computational node equipped with a Tesla V100 GPU featuring 16 GB of VRAM. The training was carried out using the MMDetection[55] and Ultralytics frameworks[56]. All training hyperparameters were maintained as specified in the default configuration files provided by the library (the original configuration files are: *rtmdet_tiny_8xb32-300e_coco.py* and *yolox_tiny_8xb8-300e_coco.py*). The only modification made was the adaptation of the dataset reference, with all other parameters remaining unchanged.

*Results.* In evaluating the performance of object detection models, we compare YOLOX, RTMDet, and RT-DERT using precision, mean Average Precision at IoU 0.5 ($mAP_{50}$), and mean Average Precision across multiple IoU thresholds ($mAP_{50:95}$). While the $mAP_{50}$ determines the average precision over all categories at an Intersection over Union (IoU) threshold set at 0.5, the $mAP_{50:95}$ serves as a stricter measure, evaluating average precision across a variety of IoU thresholds, from 0.5 to 0.95, in steps of 0.05. A greater score in the latter indicates that the model shows improved localization abilities across different levels of overlap between its predictions and the actual ground truth.

The findings, presented in Tables 5, 6, reveal that RT-DERT demonstrates superior precision compared to RTMDet. This suggests that RT-DERT produces fewer false positive detections, likely due to its more sophisticated architecture. RTMDet, however, surpasses RT-DERT in both $mAP_{50}$ and $mAP_{50:95}$, demonstrating superior overall detection capability, particularly in terms of recall and localization accuracy across different IoU thresholds.

With superior *mAP* scores, RTMDet stands out as the top-performing model in this assessment. Nonetheless, RT-DERT could be more appropriate in situations where minimizing false positives is crucial. Finally, YOLOX exhibits the lowest scores across all three metrics, suggesting that it is less efficient in both precision and overall detection accuracy.

YOLOX may struggle with finer distinctions or variations among battery types, which RTMDet and RT-DERT handle more effectively. The performance gap is particularly pronounced in the detection of cylindrical batteries, where YOLOX shows significantly lower accuracy, indicating potential limitations in the identification of this specific shape.

From this analysis, we can conclude that RT-DERT demonstrates a more cautious approach in its predictions, resulting in a reduced number of false positives and thereby enhancing precision. However, its lower performance on the *mAP* scores suggests a tendency to overlook more objects compared to RTMDet. The latter provides a more balanced precision and recall, delivering improved overall detection performance, as demonstrated

| | Yolo X | | | RTM Det | | | RT-DERT | | |
|---|---|---|---|---|---|---|---|---|---|
| | P | $mAP_{50}$ | $mAP_{50:95}$ | P | $mAP_{50}$ | $mAP_{50:95}$ | P | $mAP_{50}$ | $mAP_{50:95}$ |
| Cylindric | 0.795 | 0.964 | 0.795 | 0.853 | **0.988** | **0.853** | **0.975** | 0.969 | 0.821 |
| Pouch | 0.875 | 0.972 | 0.875 | 0.920 | **0.994** | **0.92** | **0.942** | 0.988 | 0.918 |
| Prismatic | 0.878 | 0.955 | 0.878 | 0.925 | **0.983** | **0.925** | **0.976** | 0.93 | 0.878 |

**Table 5.** The models' object detection performance was evaluated on the test set, with the results indicating comparable performance for both models. The best results are highlighted in bold.

| | Yolo X | | | RTM Det | | | RT-DERT | | |
|---|---|---|---|---|---|---|---|---|---|
| | P | $mAP_{50}$ | $mAP_{50:95}$ | P | $mAP_{50}$ | $mAP_{50:95}$ | P | $mAP_{50}$ | $mAP_{50:95}$ |
| Tot. | 0.849 | 0.9640 | 0.850 | 0.899 | **0.989** | **0.900** | **0.964** | 0.941 | 0.842 |

**Table 6.** Summarized outcomes over the classes for the two models presented as mean precision and recall.

by its increased mAP scores in all classes. This enhances RTMDet's suitability for practical applications that demand dependable and precise object detection across diverse battery types, particularly in real-world scenarios where both accuracy and high detection rates are crucial.

To validate this final assertion, Fig. 3 presents the confusion matrices for all models. These matrices illustrate the models' performance by displaying the true labels along the *y*-axis and the predicted labels along the *x*-axis, with the diagonal entries representing the correctly classified cases.

RTMDet successfully recognized 99.2% of Cylindric samples, with a misclassification rate of 0.6% as Pouch and 0.2% as Prismatic. The Pouch batteries were flawlessly categorized, whereas the Prismatic batteries were accurately identified 93.4% of the time, with a 6.6% misclassification rate as Pouch. RT-DERT, conversely, excels with this final class, attaining an accuracy of 96%, with 6.6% inaccurately classified as Pouch. However, it encounters difficulties with Cylindric samples, achieving just 94.1% accuracy in this category.

YOLOX achieved 90.1% accuracy for Cylindric samples, with 7.8% incorrectly identified as Pouch and 2.1% as Prismatic. All Pouch samples were accurately classified. Given that this is the dominant class, it could suggest that the model may be exhibiting overconfidence because of the imbalance of the dataset. It is recommended that future research explore this and implement suitable countermeasures. Finally, the Prismatic samples achieved a classification accuracy of 83%, with a misclassification rate of 17%, primarily as Pouch. This outcome may be attributed to the more straightforward architecture design of YOLOX, which, when compared to others, presents a challenge in capturing the fine details necessary for distinguishing between Prismatic and Pouch.

In summary, RTMDet outperforms both YOLO X and RT-DERT by demonstrating superior results. It reduces errors in misclassification and improves the accuracy of Cylindric, establishing itself as a more trustworthy model for this particular classification task.

**Automatic Segmentation.** In recent years, substantial foundational pre-trained models[57] have gained recognition as instruments for a variety of downstream applications, such as automated data annotation tools. In the domain of computer vision, a newly introduced model known as the segment anything model (SAM) has greatly impacted the field[58]. The primary innovation of this model lies in allowing users to specify the image segments through prompts, thus allowing diverse segmentation tasks without the need for additional training.

SAM uses a pre-trained ViT-H[59] as an image encoder that processes each image individually to generate an image embedding. It also includes a prompt encoder that encodes input prompts such as text or bounding boxes. Subsequently, a compact transformer-based mask decoder predicts object masks using the embeddings generated from images and prompts. In a very short period of time, SAM has become a fundamental component for various advanced applications, including image segmentation, image captioning, and image editing. This is due to its foundation on a Transformer model trained with the SA-1B dataset[58], enabling it to manage diverse scenes and objects.

The results presented by Kirillov *et al.*[58] indicate that SAM is capable of producing high-quality masks which are only slightly less accurate than manually annotated ground truth. An example task of these robust quantitative and qualitative results is the zero-shot Instance Segmentation, where the bounding box (generated automatically) can be used as a prompt. The mask with the highest IoU with the bounding box is then selected as the predicted mask.

We used the ground-truth bounding box from the RecyBat24dataset to prompt SAM, resulting in a new dataset suitable for segmentation tasks. Examples of the outcomes produced by SAM can be observed in Fig. 4.

*Experiments with Fast Automatic Segmentation.* Implementing SAM for instance segmentation in real-time industrial settings, where computational resources may be constrained, poses significant challenges. While the visual Transformer is key to SAM's effectiveness, it also brings considerable limitations for broad use because of its high computational requirements. Zhao *et al.*[60] propose FastSAM, a real-time solution that uses a segmentation-focused variant of YoloV8 (YOLOv8-seg), where a specific branch of the object detector is allocated for instance segmentation.

The authors' innovative approach divides the segmentation task into two procedural phases: all-instance segmentation and prompt-driven selection. In the initial phase, a Convolutional Neural Network (CNN)-based detector is used to generate segmentation masks for all instances within the image. Subsequently, in the following

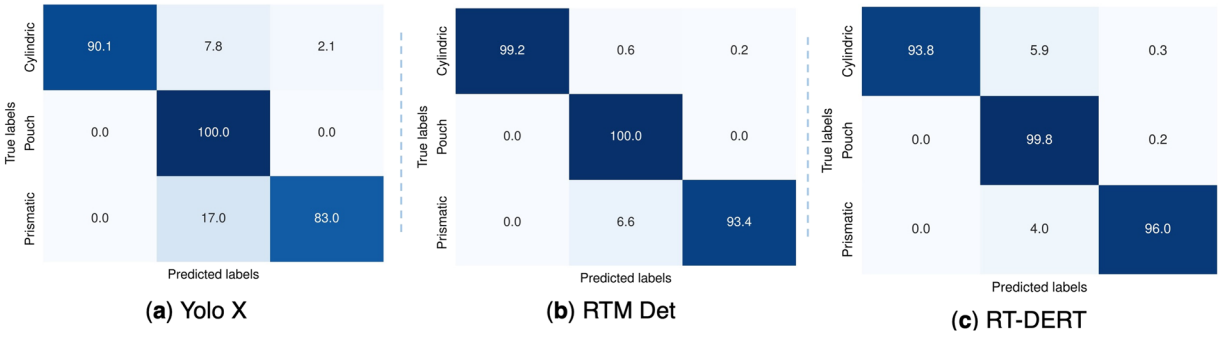


**Fig. 3** Normalized confusion matrices obtained from the test data are presented for Yolo X (**a**), RTM Det (**b**), and RT-DERT (**c**). RTM Det outperforms both YOLO X and RT-DERT, especially in accurately identifying Cylindric samples. Conversely, YOLO X tends to incorrectly classify Cylindric and Prismatic cells as Pouch cells. While RT-DERT shows greater accuracy with Prismatic cells, it is more likely to misclassify Cylindric samples.

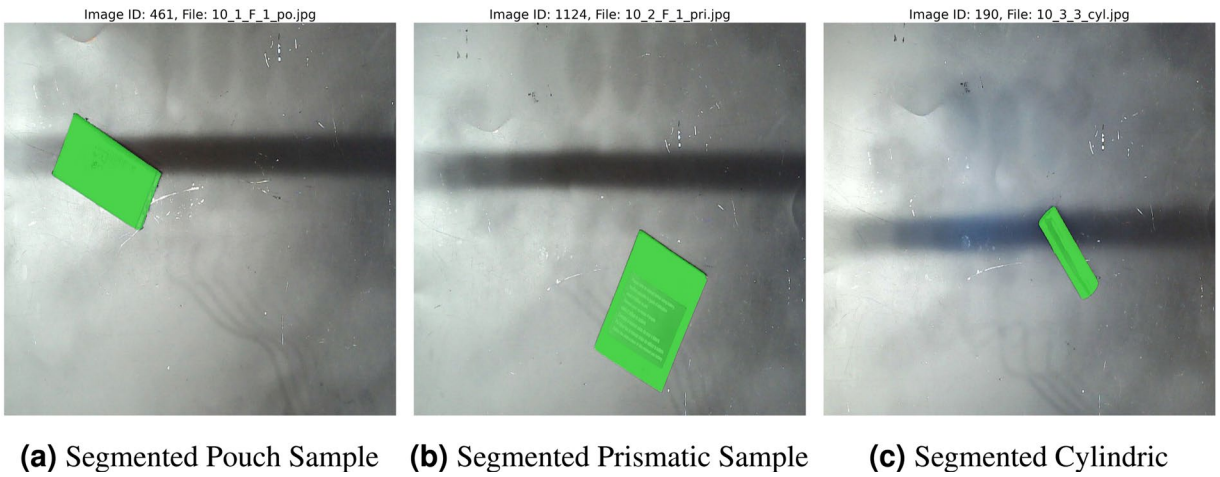


**Fig. 4** Examples of segmented battery using SAM and the bounding boxes as prompt.

| | Instances | Fast SAM | | | |
|---|---|---|---|---|---|
| | | Precision | Recall | $mAP_{50}$ | $mAP_{50:95}$ |
| All types battery | 36574 | 0.809 | 0.801 | 0.802 | 0.76 |

**Table 7.** The performance of FastSAM in instance segmentation was assessed across the whole dataset. The model does not classify instances, but treats every instance as if it belongs to a single category. The *mAP* values are calculated based on the reference masks generated by SAM.

phase, the model identifies and outputs the region-of-interest prompted by the user. Using CNNs' computational efficiency, FastSAM offers performance similar to SAM, but with significantly lower computational and resource requirements, improving real-time execution. For this reason, we chose this model as a possible candidate to solve the segmentation task and present the results obtained by FastSAM, using the public implementation provided by *ultralytics*[60] on our automatically annotated dataset.

The FastSAM findings, detailed in Table 7, indicate acceptable performance in instance segmentation tasks, with a significant equilibrium between precision and recall. In particular, FastSAM achieved a precision and recall of approximately 0.8, showcasing its effectiveness in accurately identifying instances. Furthermore, the model recorded an $mAP_{50}$ of 0.8, highlighting its strong capability in object localization. However, performance decreased slightly to an $mAP_{50:95}$ of 0.76 when assessed against various IoU thresholds, which illustrates the challenge of consistently achieving high accuracy under stricter conditions.

Overall, the metrics suggest that FastSAM, as a smaller version of SAM, efficiently performs instance segmentation with a robust level of accuracy, underlining its potential applicability in segmentation scenarios where computational resources are limited.

## Code availability

The data repository folder contains the Python files used for conducting experiments on the data. The authors declare that no custom code or software was used in the generation or processing of the data presented in this study.

Received: 30 September 2024; Accepted: 15 May 2025;
Published online: 22 May 2025

## References


1. Zheng, M. *et al*. Intelligence-assisted predesign for the sustainable recycling of lithium-ion batteries and beyond. *Energy & Environmental Science* **14**, 5801–5815 (2021).
2. Tao, Y., Rahn, C. D., Archer, L. A. & You, F. Second life and recycling: Energy and environmental sustainability perspectives for high-performance lithium-ion batteries. *Science advances* **7**, eabi7633 (2021).
3. Tao, S. *et al*. Collaborative and privacy-preserving retired battery sorting for profitable direct recycling via federated machine learning. *Nature Communications* **14**, 8032 (2023).
4. Harper, G. *et al*. Recycling lithium-ion batteries from electric vehicles. *nature* **575**, 75–86 (2019).
5. Alipanah, M., Reed, D., Thompson, V., Fujita, Y. & Jin, H. Sustainable bioleaching of lithium-ion batteries for critical materials recovery. *Journal of Cleaner Production* **382**, 135274 https://www.sciencedirect.com/science/article/pii/S095965262204848X (2023).
6. Raj, T. *et al*. Recycling of cathode material from spent lithium-ion batteries: Challenges and future perspectives. *Journal of Hazardous Materials* **429**, 128312 https://www.sciencedirect.com/science/article/pii/S0304389422001005 (2022).
7. Baum, Z. J., Bird, R. E., Yu, X. & Ma, J. Lithium-ion battery recycling-overview of techniques and trends (2022).
8. Severson, K. A. *et al*. Data-driven prediction of battery cycle life before capacity degradation. *Nature Energy* **4**, 383–391 (2019).
9. Hu, X., Xu, L., Lin, X. & Pecht, M. Battery lifetime prognostics. *Joule* **4**, 310–346 (2020).
10. Tao, S. *et al*. Battery cross-operation-condition lifetime prediction via interpretable feature engineering assisted adaptive machine learning. *ACS Energy Letters* **8**, 3269–3279 (2023).
11. Fu, S. *et al*. Data-driven capacity estimation for lithium-ion batteries with feature matching based transfer learning method. *Applied Energy* **353**, 121991 (2024).
12. Ng, M.-F., Zhao, J., Yan, Q., Conduit, G. J. & Seh, Z. W. Predicting the state of charge and health of batteries using data-driven machine learning. *Nature Machine Intelligence* **2**, 161–170 (2020).
13. Jones, P. K., Stimming, U. & Lee, A. A. Impedance-based forecasting of lithium-ion battery performance amid uneven usage. *Nature Communications* **13**, 4806 (2022).
14. Garg, A., Yun, L., Gao, L. & Putungan, D. B. Development of recycling strategy for large stacked systems: Experimental and machine learning approach to form reuse battery packs for secondary applications. *Journal of cleaner production* **275**, 124152 (2020).
15. Liu, T., Chen, X., Peng, Q., Peng, J. & Meng, J. An enhanced sorting method for retired battery with feature selection and multiple clustering. *Journal of Energy Storage* **87**, 111422 https://www.sciencedirect.com/science/article/pii/S2352152X24010077 (2024).
16. Pan, R. *et al*. Multi-stage deep sorting strategy for retired batteries based on the clustering of static and dynamic features. *Journal of Energy Storage* **99**, 113387 https://www.sciencedirect.com/science/article/pii/S2352152X24029736 (2024).
17. Wu, Q., Ye, W., Liang, R., Pan, R. & Ma, T. Two-step sorting and regrouping method of retired lithium-ion battery cells based on branches' topological structures. *Energy & Fuels* **38**, 21122–21133 https://doi.org/10.1021/acs.energyfuels.4c03320 (2024).
18. Lu, Y. *et al*. A novel disassembly process of end-of-life lithium-ion batteries enhanced by online sensing and machine learning techniques. *Journal of intelligent manufacturing* **34**, 2463–2475 (2023).
19. Zhao, Y. *et al*. Batsort: Enhanced battery classification with transfer learning for battery sorting and recycling. In *2024 IEEE Annual Congress on Artificial Intelligence of Things (AIoT)*, 201–206 https://doi.org/10.1109/AIoT63253.2024.00047 (2024).
20. Ueda, T., Koyanaka, S. & Oki, T. In-line sorting system with battery detection capabilities in e-waste using combination of x-ray transmission scanning and deep learning. *Resources, Conservation and Recycling* **201**, 107345 https://www.sciencedirect.com/science/article/pii/S0921344923004792 (2024).
21. Chai, C., Wang, J., Luo, Y., Niu, Z. & Li, G. Data management for machine learning: A survey. *IEEE Trans. Knowl. Data Eng.* **35**, 4646–4667, https://doi.org/10.1109/TKDE.2022.3148237 (2023).
22. Alzubaidi, L. *et al*. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *Journal of Big Data* **10**, 46 (2023).
23. Emam, Z. *et al*. On the state of data in computer vision: Human annotations remain indispensable for developing deep learning models. *CoRR***abs/2108.00114** https://arxiv.org/abs/2108.00114. (2021).
24. Dufek, E. J., Tanim, T. R., Chen, B.-R. & Kim, S. Battery calendar aging and machine learning. *Joule* **6**, 1363–1367 (2022).
25. Ilyas, I. F. & Chu, X.*Data Cleaning*, vol. 28 of *ACM Books* https://doi.org/10.1145/3310205 (ACM, 2019).
26. Platanios, E. A., Al-Shedivat, M., Xing, E. P. & Mitchell, T. M. Learning from imperfect annotations. *CoRR***abs/2004.03473** https://arxiv.org/abs/2004.03473 (2020).
27. Simula, H. The rise and fall of crowdsourcing? In *46th Hawaii International Conference on System Sciences, HICSS 2013, Wailea, HI, USA, January 7-10, 2013*, 2783–2791 https://doi.org/10.1109/HICSS.2013.537 (IEEE Computer Society, 2013).
28. Elmore, J. G. *et al*. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *Jama* **313**, 1122–1132 (2015).
29. Althnian, A. *et al*. Impact of dataset size on classification performance: An empirical evaluation in the medical domain. *Applied Sciences* **11** https://www.mdpi.com/2076-3417/11/2/796 (2021).
30. Zhuang, F. *et al*. A comprehensive survey on transfer learning. *Proc. IEEE* **109**, 43–76, https://doi.org/10.1109/JPROC.2020.3004555 (2021).
31. Yang, S., Song, C., Bao, L. & Zhao, G. An exploration of the impact of training datasets on deep learning-based building extraction. In *6th International Conference on Internet of Things, Automation and Artificial Intelligence, IoTAAI 2024, Guangzhou, China, July 26-28, 2024*, 541–545 (IEEE, 2024). https://doi.org/10.1109/IoTAAI62601.2024.10692949.
32. Zhang, H. *et al*. Battery lifetime prediction across diverse ageing conditions with inter-cell deep learning. *Nature Machine Intelligence* 1–8 (2025).
33. Zhao, J., Wang, Z., Wu, Y. & Burke, A. F. Predictive pretrained transformer (ppt) for real-time battery health diagnostics. *Applied Energy* **377**, 124746 (2025).
34. Müller, S. *et al*. Deep learning-based segmentation of lithium-ion battery microstructures enhanced by artificially generated electrodes. *Nature Communications* **12**, 6205, https://doi.org/10.1038/s41467-021-26480-9 (2021).
35. Yan, B., Ma, E. & Wang, J. Research on the high-efficiency crushing, sorting and recycling process of column-shaped waste lithium batteries. *Science of The Total Environment* **864**, 161081 https://www.sciencedirect.com/science/article/pii/S0048969722081840 (2023).
36. Xu, M., Yoon, S., Fuentes, A. & Park, D. S. A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognit.* **137**, 109347, https://doi.org/10.1016/j.patcog.2023.109347 (2023).
37. Devries, T. & Taylor, G. W. Improved regularization of convolutional neural networks with cutout. *CoRR***abs/1708.04552** http://arxiv.org/abs/1708.04552 (2017).
38. Zhong, Z., Zheng, L., Kang, G., Li, S. & Yang, Y. Random erasing data augmentation. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*, 13001–13008 https://doi.org/10.1609/aaai.v34i07.7000 (AAAI Press, 2020).
39. Zhang, H., Cissé, M., Dauphin, Y. N. & Lopez-Paz, D. mixup: Beyond empirical risk minimization. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings* (OpenReview.net, 2018). https://openreview.net/forum?id=r1Ddp1-Rb.

40. Acaro Chacón, X. C., Lo Scudo, F., Cappuccino, G. & Dodaro, C. Recybat24: Battery dataset for recycling https://zenodo.org/records/13126689 (2024).
41. Zou, Z., Chen, K., Shi, Z., Guo, Y. & Ye, J. Object detection in 20 years: A survey. *Proceedings of the IEEE* **111**, 257–276 (2023).
42. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* **25** (2012).
43. Girshick, R., Donahue, J., Darrell, T. & Malik, J. Region-based convolutional networks for accurate object detection and segmentation. *IEEE transactions on pattern analysis and machine intelligence* **38**, 142–158 (2015).
44. Uijlings, J. R., Van De Sande, K. E., Gevers, T. & Smeulders, A. W. Selective search for object recognition. *International journal of computer vision* **104**, 154–171 (2013).
45. Girshick, R. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 1440–1448 (2015).
46. Ren, S., He, K., Girshick, R. & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* **28** (2015).
47. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788 (2016).
48. Redmon, J. & Farhadi, A. Yolov3: An incremental improvement. CoRR abs/1804.02767 (2018).
49. Bochkovskiy, A., Wang, C. & Liao, H. M. Yolov4: Optimal speed and accuracy of object detection. CoRR abs/2004.10934 (2020).
50. Ge, Z., Liu, S., Wang, F., Li, Z. & Sun, J. YOLOX: exceeding YOLO series in 2021. CoRR abs/2107.08430 (2021).
51. Lyu, C. et al. Rtmdet: An empirical study of designing real-time object detectors. CoRR abs/2212.07784, 10.48550/ARXIV.2212.07784 (2022).
52. Vaswani, A. *et al*. Attention is all you need. *Advances in neural information processing systems* **30** (2017).
53. Carion, N. *et al*. End-to-end object detection with transformers. In *European conference on computer vision*, 213–229 (Springer, 2020).
54. Zhao, Y. *et al*. Detrs beat yolos on real-time object detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024*, 16965–16974 (IEEE, 2024). https://doi.org/10.1109/CVPR52733.2024.01605.
55. Chen, K. *et al*. Mmdetection: Open mmlab detection toolbox and benchmark. *CoRR* **abs/1906.07155**, http://arxiv.org/abs/1906.07155 (2019).
56. Baldovino, R. G. *et al*. Comprehensive analysis on ultralytics-supported YOLO models for detection and recognition of large office objects for indoor navigation. In Toro, C., Ríos, S. A., Howlett, R. J. & Jain, L. C. (eds.) *Knowledge-Based and Intelligent Information & Engineering Systems: Proceedings of the 28th International Conference KES-2024, Seville, Spain, 11-13 September 2023*, vol. 246 of *Procedia Computer Science*, 3851–3858 (Elsevier, 2024). https://doi.org/10.1016/j.procs.2024.09.158.
57. Zhou, C. et al. A comprehensive survey on pretrained foundation models: A history from BERT to chatgpt. CoRR abs/2302.09419, 10.48550/ARXIV.2302.09419 (2023).
58. Kirillov, A. *et al*. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4015–4026 (2023).
59. Dosovitskiy, A. *et al*. An image is worth 16 × 16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations* (2020).
60. Zhao, X. *et al*. Fast segment anything. CoRR **abs/2306.12156**, https://doi.org/10.48550/ARXIV.2306.12156 (2023).


## Acknowledgements

This work was funded by the Next Generation EU - Italian NRRP, Mission 4, Component 2, Investment 1.5, call for the creation and strengthening of "Innovation Ecosystems", building "Territorial R& D Leaders" (Directorial Decree n. 2021/3277) - project Tech4You - Technologies for climate change adaptation and quality of life improvement, n. ECS0000009. This work reflects only the authors' views and opinions, neither the Ministry for University and Research nor the European Commission can be considered responsible for them.

## Author contributions

X.C.A.C. contributed by creating the dataset. F.L.S. was responsible for designing and conducting the experiments. G.C. and C.D. provided supervision, overseeing the project's progress and ensuring its successful completion. All authors were involved in writing and revising the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to X.C.A.C. or F.L.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.