Waste Management 138 (2022) 274–284

Contents lists available at ScienceDirect

Waste Management

journal homepage: www.elsevier.com/locate/wasman

Deep learning-based waste detection in natural and urban environments

Sylwia Majchrowska [a], Agnieszka Mikołajczyk [b,*], Maria Ferlin [b], Zuzanna Klawikowska [b], Marta A. Plantykow [c], Arkadiusz Kwasigroch [b], Karol Majek [d]

[a] Wrocław University of Science and Technology, wybrzeże Stanisława Wyspiańskiego 27, 50-370 Wrocław, Poland
[b] Gdańsk University of Technology, Gabriela Narutowicza 11/12, 80-233 Gdańsk, Poland
[c] Intel Technology Poland Sp. z o.o., Juliusza Słowackiego 173, 80-298 Gdańsk, Poland
[d] CuFiX, Legionowa 16, 05-270 Marki, Poland

ARTICLE INFO

Keywords:
Object detection
Semi-supervised learning
Waste classification benchmarks
Waste detection benchmarks
Waste localization
Waste recognition

ABSTRACT

Waste pollution is one of the most significant environmental issues in the modern world. The importance of recycling is well known, both for economic and ecological reasons, and the industry demands high efficiency. Current studies towards automatic waste detection are hardly comparable due to the lack of benchmarks and widely accepted standards regarding the used metrics and data. Those problems are addressed in this article by providing a critical analysis of over ten existing waste datasets and a brief but constructive review of the existing Deep Learning-based waste detection approaches. This article collects and summarizes previous studies and provides the results of authors' experiments on the presented datasets, all intended to create a first replicable baseline for litter detection. Moreover, new benchmark datasets *detect-waste* and *classify-waste* are proposed that are merged collections from the above-mentioned open-source datasets with unified annotations covering all possible waste categories: *bio, glass, metal and plastic, non-recyclable, other, paper*, and *unknown*. Finally, a two-stage detector for litter localization and classification is presented. EfficientDet-D2 is used to localize litter, and EfficientNet-B2 to classify the detected waste into seven categories. The classifier is trained in a semi-supervised fashion making the use of unlabeled images. The proposed approach achieves up to 70% of average precision in waste detection and around 75% of classification accuracy on the test dataset. The code and annotations used in the studies are publicly available online[1].

1. Introduction

Environmental pollution related to solid waste mismanagement is a global problem. Massive production of disposable goods in the last years has resulted in significant increase in the produced garbage, which according to the European household waste collection (Eurostat) report 5.2 tonnes of waste were generated per EU inhabitant in 2018 (Eurostat, 2018). Additionally, the World Bank (WB) report (Kaza et al. 2018), is expected to reach above 3 billion tons of waste per year by 2050. WB states that only 13.5% of global waste is recycled, while about 33% of garbage is thrown away openly without any preliminary classification (Kaza et al. 2018). This leads to the fact that different types of litter are freely scattered in a wide variety of environments. The biggest concern is plastic waste, as it is the most widespread and the long-term environmental harm (Li et al. 2016). To prevent further pollution of the environment and, as a consequence, protect human and wild organisms' life, immediate steps are needed to facilitate wise collection and segregation of rubbish.

One of the ways to support waste sorting is machine learning (ML). In recent years -ML-based systems that can support or fully cover sorting processes were implemented, accelerating this procedure as a result. The most commonly used solution includes smart self-sorting waste bins, which are capable of classifying one object located on a clear background at a time (White et al. 2020; Sheng et al. 2020). When a single compartment is used, the camera is usually located at the top of the upper container. The deep learning (DL) model assigns proper class based on the image and the garbage is moved to the appropriate bottom container (White et al. 2020). Another way is to mount a camera or a sensor above few separate bins, and direct the consumer handling waste to the correct one (Tavakoli et al. 2018). In this approach, the rubbish must be well exposed to the imaging device. On the other hand, artificial intelligence (AI) can even help with identification of illegal dumping

* Corresponding author.
E-mail addresses: sylwia.majchrowska@pwr.edu.pl (S. Majchrowska), agnieszka.mikolajczyk@pg.edu.pl (A. Mikołajczyk).
[1] https://github.com/wimlds-trojmiasto/detect-waste.

https://doi.org/10.1016/j.wasman.2021.12.001
Received 5 June 2021; Received in revised form 23 November 2021; Accepted 2 December 2021

sites by tracking the undesirable behavior of residents (Lu 2019) or utilizing satellite data to locate agricultural plastic waste (Lanorte et al. 2017). Recently, deep learning started to be used to detect and identify waste in natural environment with drones or cameras mounted on vehicles (Kraft et al. 2021).

The reported successes show that deep learning applications can speed up litter sorting and detecting. However, a question might arise: how can those approaches be compared? Garbage identification is ambiguous even for humans; as it is difficult to state whether an object is a garbage without any context. On top of that, trash can be atypical or deformed, or spotted under various uncontrolled natural conditions. This diversity of objects requires well-annotated data for comparison which distribution should match the target use case. Numerous waste datasets were introduced, yet each covers various types of waste, with different annotations and waste categories. Moreover, most of them usually do not provide a large enough number of annotated images, or they represent only a single type of waste (e.g. bottles or cigarette butts).

The inconsistency in used data and metrics hinders fair and valid comparison of methodologies; the reported research uses, as a rule, different datasets and incoherent metrics and data splits. To authors' knowledge, today, there are no representative and public benchmarks in waste detection. The current lack of benchmarks slows down the optimization cycle of deep learning models and lowers the chances of constant improvement and knowledge sharing in academics and industry. For beginners in the field, benchmarks summarize the best practices, showing the right direction to follow. For domain experts, it is a good baseline provider, giving a chance for relevant comparison.

The article provides the first comprehensive review of existing waste datasets. Moreover, two benchmark datasets: detect-waste and classify-waste are introduced, which utilize the advantages of the existing open-source datasets to the fullest. The publicly available datasets of waste observed in different environments are unified, filtered, and merged. Inspired by waste segregation principles in Gdansk (Poland), the authors propose seven well-defined categories for sorting litter: *bio, glass, metal and plastic, non-recyclable, other, paper*, and *unknown*. The baselines for all reviewed datasets are provided, including the introduced classify-waste and detect-waste benchmarks.

Additionally, a holistic approach is proposed to localizing and classifying waste in images in realistic scenarios that can be used as a baseline for future studies. A two-stage DL-based framework has been implemented for waste detection that consists of two separate neural networks: detector and classifier. The proposed framework is freely available and can be used for different purposes, such as monitoring changes in distribution of waste in nature. To the authors' knowledge, the experiments presented in this article are the first that allow for such universal litter detection and classification. The main contributions of this article are: proposition of relevant benchmarks for litter detection, comprehensive review of the existing datasets, and presentation of baseline results with the two-stage framework for all datasets.

An overview of the actual work for deep learning classification and detection, with an exhausting review of the existing waste datasets, is given in Section 2. Section 3 illustrates the applied framework to detect and classify waste from images. More specifically, the used data is described in detail, and the characteristics of the proposed waste detection system are identified. Section 4 provides training details of the chosen neural networks and reports the obtained results. Finally, conclusions are drawn, and future work is outlined in Section 5.

2. Related works

ML and DL techniques empower many aspects of modern society, like recommendation systems, text-to-speech devices, or objects identification in images (LeCun, Bengio, and Hinton 2015; Konovalenko et al. 2020). The waste management problem has attracted a lot of interest, where the main goal was to create an ML-based image recognition system to sort litter (White et al. 2020; Sheng et al. 2020; Glouche and Couderc 2013). The majority of the proposed approaches are based on deep learning algorithms utilized in the computer vision field. This section describes selected techniques used in classification and object detection challenges and presents several publicly available waste datasets.

2.1. Classification

Convolutional neural networks (CNNs) have had a massive impact on large-scale image classification tasks and made it possible to achieve significantly higher accuracy than solutions based on classical image processing. There has been a constant improvement in the quality of image recognition structures, and many new architectures were presented, among which, few were identified as suitable for the waste classification problem.

In 2016 the ResNet family (He et al. 2016) was proposed that is based on residual connections that introduce no extra parameters or computational effort. Instead, this connectivity pattern facilitates the gradient flow, which enables effective training of network consisting of as many as 200 layers.

Then, the DenseNet structure (Huang et al. 2017) has facilitated training and accuracy by using the feature maps of all preceding layers as inputs. Its main advantages are: minimizing the vanishing gradient problem, improving feature propagation, encouraging feature reuse, and reducing the number of parameters.

The next architecture is EfficientNet (Tan and Le 2019) that consists of modules constructed by the neural architecture search process that optimizes accuracy and FLOPS. The main building block is mobile inverted bottleneck MBCon (Sandler et al. 2018). The mobile-size baseline model called EfficientNet-B0 was built by stacking those modules. Scaling of a convolutional neural network is most often performed in one of the following dimensions: width, depth, or resolution. The compound scaling strategy is used to produce larger and more complex and accurate models EfficientNet-B1-7. This scaling strategy significantly improves the efficiency and accuracy of the model.

EfficientNet-B2 consists of one input 3x3 convolution layer with input size of 224x224. Then, there are seven stages with mobile inverted bottlenecks MBCon with squeeze-and-excitation optimization that have jointly 16 layers. It ends with the 1x1 convolution layer with pooling and fully-connected layer. It is all scaled up with the compound coefficient $\phi = 2$. The EfficientNet-B2 architecture was used in the present research without model modifications, therefore this article provides a brief description of this architecture, however it is recommended to read (Tan and Le 2019) for more information.

EfficientNetv2 (Tan and Le, 2021) improves EfficientNet by eliminating EfficientNet bottlenecks and thus providing faster training and better parameter efficiency. Neural architecture search is used to optimize accuracy, training speed, and parameter size. Unlike standard EfficientNet, EfficientNetv2 uses non-uniform scaling of depth, resolution, and width. Moreover, to limit computational cost, the increase in resolution is limited.

2.2. Object detection

Object detection is a well-studied task in the computer vision field. This task is defined as localization of an Axis-Aligned Bounding Box (AABB) and classification - assignment of a single or multi-label (Zou et al. 2019). In many previous works, object detection was approached using two techniques, namely one-stage and two-stage detection. Two-stage detectors find class-agnostic object proposals first, and then classify them into the class-specific detections as the second stage. At the same time one-stage architectures provide both locations and classes for each object in a single step.

Two-stage detectors were the first object detection methods. They used the sliding window approach in the image pyramid to generate object proposals in multiple scales. Then, in the second stage, a classifier

such as a cascade classifier (Viola and Jones 2001) was used. Over the years, several algorithms have been presented such as Histograms of Oriented Gradient (HOG) and Support Vector Machine (SVM) (Dalal and Triggs 2005), *recognition using regions* paradigm (Gu et al. 2009) and Selective Search algorithm (Uijlings et al. 2013), based on which new architectures were created.

Selective Search was used in R-CNN (Girshick et al. 2014) to extract region proposals, on which CNN features were computed to classify regions using per class SVMs. Fast R-CNN (Girshick 2015) solved R-CNN's main problems by computing convolutional features for the whole image in the first step and then integrating classification into the network architecture. In the Faster R-CNN (Ren et al. 2015) architecture the Region Proposal Network (RPN) State-of-the-art concept was used instead of the Selective Search algorithm. In the following years, many researchers used Faster R-CNN while replacing its backbone (feature extraction CNN) with newer architectures. Like its predecessors, Faster R-CNN is a two-stage object detection method.

Single-stage detection was popularized in Deep Learning mainly by two detector architectures: Single Shot MultiBox Detector (SSD) (Liu et al. 2016) and You Only Look Once (YOLO v1, 9000, v3, v4, scaled V4, and later PP-YOLOv2, YOLOR: (Redmon et al. 2016; Redmon and Farhadi 2017; Redmon and Farhadi 2018; Bochkovskiy, Wang, and Liao 2020; Wang, Bochkovskiy, and Liao 2020; Huang, Wang et al. 2021; Wang, Yeh, and Liao 2021)). In this method, many false detections need to be removed considering objectness or classification score and overlaps between detections.

Detection Transformer (DETR) introduced Transformers known from Natural Language Processing tasks (Vaswani et al. 2017) into object detection while preserving image processing with CNNs. DETR predicts a sparse, fixed number of objects in training matched with ground truth labels using bipartite matching. DETR's main limitations were slow convergence and limited feature spatial resolution. In the following years, these limitations were mitigated in Deformable DETR (Zhu et al. 2020) thanks to the introduced deformable multi-head attention module.

EfficientDet (Tan, Pang, and Le 2020) is a single-stage object detector. It consists of an EfficientNet backbone, a weighted bi-directional feature pyramid network (BiFPN), and class and box prediction networks. Its main goal was to improve the efficiency of object detection models. The improvement consists in using a compound scaling method from EfficientNet (Tan and Le 2019) to jointly scale up the resolution/depth/width for the backbone, the feature pyramid networks, and the box/class prediction networks. EfficientDet is proposed in eight variants referenced as D0-D7 with backbone networks EfficientNet B0-B7. EfficientDet-D2 is based on the EfficientNet-B2, then, there are 5 BiFPN layers with 112 channels for feature selection and 3 box/class prediction layers. This approach reduces the number of parameters and latency in the object detection network, and simultaneously increases the Average Precision metrics.

The EfficientDet-D2 architecture has been used in this research without model modifications. A brief description of this architecture is provided; however it is recommended to read (Tan, Pang, and Le 2020) for more information.

Instance Segmentation is another task for automated image processing. The goal of this task is to provide a segmentation mask for each object instance. Here, Mask R-CNN (He et al. 2017), being an extension of Faster R-CNN, was introduced. In the network, the instance mask branch is added parallel to the classification and bounding box regression branch.

2.3. Deep learning for waste recognition

In recent years multiple attempts have been made to detect, classify and segment waste using deep learning. However, the number of images or categories of the garbage used differed significantly between the approaches making them incomparable. Image-based litter classification of common waste categories has been attempted using a few pre-trained convolutional neural networks – AlexNet (Yang and Thung 2016; Chu et al. 2018), MobileNet, InceptionResNetV2, DenseNet, Xception (Bircanoğlu and Arıca 2018) – achieving average accuracy in ranges 22% (Yang and Thung 2016) and 98.2% (Chu et al. 2018) for pictures of waste on a plain background. Recently an accuracy of 99.95% was obtained on the diverse waste dataset reconstructed with the AutoEncoder network (Toğaçar, Ergen, and Cömert 2020). However, in that study, waste types were split into only two classes (organic and recyclable). Moreover, many used images had misleading labels or contain both types of waste, making one-class classification studies ineffectual.

Experiments have also been conducted on litter detection and segmentation in the streets and homes, using different DL-based architectures. In those experiments, use was made of typical neural networks, including Faster R-CNN (Awe, Mengistu, and Sreedhar 2017; Fulton et al. 2019; Hong, Fulton, and Sattar 2020), SSD (Fulton et al. 2019), different types of YOLO (even Tiny-YOLO) (Liu et al. 2018; Fulton et al. 2019; Carolis, Ladogana, and Macchiarulo 2020; Kraft et al. 2021), EfficientDet (Kraft et al. 2021), Cascade R-CNN and ATSS (Liang and Gu 2021) for detection, Mask R-CNN (Proença and Simões 2020; Hong, Fulton, and Sattar 2020) for instance segmentation, and DeepLab (Wang et al., 2020a; 2020b) for semantic segmentation. The calculated mean average precision (mAP) score varied between different datasets and architectures from 15.9% for TACO (Proença and Simões 2020) with Mask R-CNN, up to 81% for Trash-ICRA19 (Fulton et al. 2019) with Faster R-CNN. The above-mentioned research focused on detecting specific types of garbage (mainly one category – litter) in a single environment. The quality of detection in various environments was not tested. Table comparing results of described above experiments is presented in Supplementary Table 1.

For the purpose of this research, over ten different datasets have been found in the literature which represent three main scenarios: outdoor (natural/urban environment) (Lynch 2018; "Waste Pictures" 2019; Foundation 2016; Proença and Simões 2020; Kraft et al. 2021), indoor (Yang and Thung 2016; Serezhkin 2020; Wang et al., 2020a; 2020b), and underwater (Fulton et al. 2019; Hong, Fulton, and Sattar 2020). A brief description of their main features is provided below.

Open Litter Map. Open Litter Map (Lynch 2018) is a free, open, and crowd-sourced dataset with over 100k images taken by phone cameras. All images are provided with information such as type of presented litter, coordinates, timestamp, or phone model. The images come from all over the world, and were taken by different people. Therefore, they differ significantly from one another.

Waste Pictures. The Waste Pictures dataset ("Waste Pictures" 2019) contains almost 24k waste images scraped from Google search, divided into 34 classes. The type of images is very diverse, including even x-rays and drawings of garbage. The sizes also differ significantly. Most of the photos are below the size of 2000x2000px. Due to the origin of images, they should be manually reviewed for use in a classification task.

TrashNet. The TrashNet dataset (Yang and Thung 2016) contains over 2100 labeled images. Each image belongs to one of the six classes: glass, paper, cardboard, plastic, metal, and trash. The pictures were taken by mobile phone camera using sunlight and/or room lighting. The photographed objects were placed on a white background or fulfilled the whole view (cardboard). All images have the size of 512x384px.

Extended TACO. Trash Annotations in Context (TACO) (Proença and Simões 2020) is a crowd-sourced dataset of waste in the wild with high-resolution mobile phone images. The TACO dataset contains 1500 annotated images with almost 5000 objects. All trash has been assigned to one of 60 classes that belong to 28 super (top) categories, including *Unlabeled litter* for hard to recognize or heavily obscured objects. The annotations are provided in the well-known COCO format (Lin et al.

Table 1
Statistics of selected public waste datasets used for object detection and classification.

Type	Dataset	# classes	# images	# instances	Annotation type	Environment
Classification	Open Litter Map	>100	>100 k	>100 k	multilabels	outdoor
	TrashNet	5	2194	2194	labels	indoor
	Waste Pictures	34	23,633	23,633	labels	outdoor
	Places	205	2.5 M	2.5 M	labels	background
Detection	Drinking waste	4	4810	5058	bounding box	indoor
	Extended TACO	7	4562	14,286	bounding box	outdoor
	MJU-Waste	1	2475	2532	instance masks	indoor
	TrashCan	8	7212	6214	instance masks	underwater
	Trash-ICRA	7	7668	6706	bounding box	underwater
	UAVVaste	1	772	3718	instance masks	outdoor
	Wade-AI	1	1396	2247	instance masks	outdoor

2014) on the instance segmentation level with an extra background description - Trash, Vegetation, Sand, Water, Indoor, Pavement. Additionally, TACO offers around three thousand unannotated images, which of over 3000 were annotated on the detection level[2] achieving over 14 000 instances in total. A great advantage is that TACO is characterized by various litter types and high diversity of backgrounds, from tropical beaches to London streets. However, due to the crowd-sourcing nature of the dataset, labels may contain some user-induced errors and bias, i.e., not all objects in TACO can be categorized strictly as litter as their category is often based on context.

Wade-AI. The Wade-AI dataset (Foundation 2016) contains images of waste in the wild, provided by Google Street View. It consists of nearly 1400 images with 2200 manually labeled instance masks annotations in COCO format with only one class, called rubbish. The environment and size of the images vary due to the source of the images. Most images are less than 1000x1000.

UAVVaste. Another publicly available dataset, which also provides instance segmentation masks in COCO format, is the UAVVaste (Kraft et al. 2021) dataset. It contains 772 hand-labeled aerial images of waste with over 3700 objects of one class - "rubbish". The data was collected in the cities and nature, e.g., streets, parks, and lawns using Unmanned Aerial Vehicles (UAV). The annotated litter is usually relatively small (the median of object shape is 76x68px, while the median of image shape is 3840x2160px).

Trash-ICRA and TrashCan. Trash-ICRA19 (Fulton et al. 2019) and TrashCan 1.0 (Hong, Fulton, and Sattar 2020) are both the datasets containing underwater images. They are video frames of trash, undersea flora and fauna, from the perspective of remotely operated underwater vehicles (ROVs). Both datasets are sourced from the JAMSTEC E-library of Deep-sea Images (J-EDI) dataset (J. A. 2012) curated by the Japan Agency of Marine-Earth Science and Technology (JAMSTEC). The images were recorded in real-world environments, providing a variety of objects. The clarity of the water and the quality of the light varies significantly between images creating a diverse dataset. The image sizes in these datasets are 480x270px and 480x360px. The provided annotations are in COCO format. The TrashCan dataset is annotated on the instance segmentation level (7212 images and 6214 annotations) with 16 classes for Material Version or 22 for Instance Version. On the other hand, the Trash-ICRA19 dataset is annotated on the detection level (7668 images and 6706 annotations). It contains seven categories based on the material of the objects.

Drinking Waste. The Drinking Waste dataset (Serezhkin 2020) contains over 4800 images of drinking waste belonging to 4 classes: Aluminium Cans, Glass bottles, PET bottles, and HDPE. The provided bounding box annotations are in YOLO format. The dataset was created with a 12 MP phone camera. The images look similar – there is usually one object in the center on the indoor, plain background. Most of the images have the size of 512x683px.

MJU-Waste. The MJU-Waste dataset (Wang et al., 2020a; 2020b) is comprised of 2475 indoor trash images manually annotated in the form of an instance mask in COCO format. It allows two-class semantic segmentation of waste and background. For each color image, the co-registered depth image captured using an RGBD camera is provided. The objects are hand-held and situated mostly in the center of the image. In most cases, there is only one object per image. The only image size is 640x480px.

Places. The Places dataset (Zhou et al. 2018) is a repository of 10 million scene photographs, labeled with 434 scene semantic categories, comprising a large and diverse list of types of locations encountered in the world. The images were downloaded by online image search engines using Google Images, Bing Images, and Flickr. The minimal size of images is 200x200px. Although this is not a trash dataset, it can be used to identify natural and urbanized places without trash.

Additionally, the crucial statistics of images, classes and instances for each dataset used in the research are presented in Table 1, considering its purpose, type of annotations and environment.

3. Methodology

This section proposes new waste detection and classification benchmarks and presents their statistics. The details behind the proposed two-stage waste detection framework are outlined. Finally, the methodology behind the training procedure of the proposed model is described.

3.1. Proposed waste benchmarks

For the past few years, considerable attempts have been made toward the development of various waste datasets, yet each is presented with different annotations and ambiguous waste categories. Moreover, most of them usually do not provide a large enough number of annotated images, or they represent only a single type of waste (e.g. bottles or cigarette butts). Furthermore, there is no unified and commonly-followed experimental protocol. This inconsistency in used data and metrics hinders fair and valid comparison of methodologies; reported research uses different datasets, and incoherent metrics and data splits.

Thus, there is an urgent need for a complete and comprehensive suite of real-world benchmarks that combine diverse datasets of different sizes from diverse environments. Evaluation procedure and consistent data splits are essential for measuring progress in a reproducible way. Last but not least, benchmarks need to provide different types of tasks, such as localization, detection, and classification.

Considering those limitations a new benchmark datasets detect-waste for litter detection and classify-waste for litter classification were proposed. It addresses the above-mentioned limitations not only by simply raising the number of data instances but additionally by evening the distribution of waste locations and, most importantly, waste types. Proposed combination of different datasets is open and accessible,

[2] Only the primary part (about 1.5k images) of the dataset is annotated with instance masks, the authors provided bounding box annotation for the rest.

clearly addresses the task, names suitable prediction error evaluation methods, comes with a baseline and sample solution, and is well documented.

Data preparation. Firstly, all subset data annotations were unified to one standard style. Segmentation and instance segmentation masks, bounding boxes, were converted to standardized COCO bounding box annotations. The annotations have been cleaned and fixed; bounding boxes with coordinates outside the scope of the image were normalized. Next, the dataset was examined in terms of labeling consistency. It was decided that each object should be annotated with a separate bounding box, including closely neighboring trash. Hence, inconsistent samples were identified and corrected, or removed (e.g., piles of garbage with a single bounding box, graffiti annotated as waste).

Moreover, according to household waste segregation principles in Gdansk, the annotation guidelines were prepared, making it an excellent, unambiguous tool to compare new solutions of waste classification in real-life applications. All waste labels of all subsets were modified to match the categories proposed in the article. Each dataset split was done carefully, following the distribution of each subset. Then, resulting training, validation and test subsets were concatenated into detect-waste and classify-waste datasets. The distribution of classes in the dataset is connected to the actual distribution of the waste types produced by humans. Although it is imbalanced, there are still many representatives of each class. Finally, appropriate metrics were selected and proposed making it measurable and allowing valid comparison of methodologies. Proposed datasets come with a baseline described in the next section.

Detect-waste. The proposed detect-waste benchmark is a merged collection of Extended TACO (dataset extended by the authors) and publicly available datasets with detection annotations: Wade-AI, UAVVaste, TrashCan, Trash-ICRA, Drinking-Waste, and MJU-Waste. It was ensured that detect-waste contains over 28 000 images and over 40 000 objects with a unified bounding box annotation and a single label *litter*. The main advantage of the created detect-waste is its diversity provided by a combination of existing datasets. The images come from three main environmental categories: indoor, outdoor (urban and natural), and underwater. Moreover, the images were taken in various lighting conditions, using different instruments, and with a wide range of object sizes lowering the chances of bias in data. The collected data is represented by various waste types gathered from around the world, which ensures that models trained on this dataset will have satisfying generalization ability. For the complete comparison, the authors suggest testing models on proposed datasets and every subset separately. The detect-waste subsets are presented in Fig. 1

Classify-waste. The proposed classify-waste benchmark is a merged collection of publicly available datasets with eight classification labels. The first six are based on the recycling rules in Gdańsk, Poland. The categories with corresponding examples are as follows:

- *bio*: food waste such as fruit, vegetables, herbs, used paper towels and tissues,
- *glass*: glass objects such as glass bottles, jars, cosmetics packaging,
- *metals and plastic*: scrap metal and non-ferrous metal, beverage cans, plastic beverage bottles, plastic shards, plastic food packaging, or plastic straws,
- *non-recyclable*: residual rubbish such as disposable diapers, pieces of string, polystyrene packaging, polystyrene elements, blankets, clothing, or used paper cups,
- *other*: construction and demolition, large-size waste (e.g. tires), used electronics and household appliances, batteries, paint and varnish cans, or expired medicines,
- *paper*: paper, cardboard packaging, receipts, newspapers, catalogues, and books,
- *unknown* waste: (highly decomposed and hard-to-recognize litter),
- and extra class *background* label without any litter: a sidewalk, a forest path, a lawn.

The proposed annotation type unification is based on household waste categories and can be applied in real-life solutions, making it a substantial advantage over the existing datasets. Contrary to the existing datasets, all waste types are covered in the dataset, with particular attention to recyclable waste. They are similar to other proposed household waste categories (Slagstad and Brattebø 2013; Sahimaa et al. 2015, Zorpas et al. 2015). The classify-waste dataset contains over 21000 waste instances coming from Extended TACO, drinking-waste, waste-pictures, Google search, TrashNet, and Places. Most of the trash comes from *metal and plastic* or an unknown category which is closely related to the distribution of the waste types produced by humans. Nevertheless, it provides a wide range of trash and it will ensure the generalization ability of a model trained on this dataset. Some samples were rejected manually due to their poor quality. Adding Places dataset was considered as a virtue, because it ensures a good distinction between trash and its surroundings. The details behind the dataset categories and removed samples are provided in the open-source repository. As in the case of detect-waste, it is suggested to test models on proposed

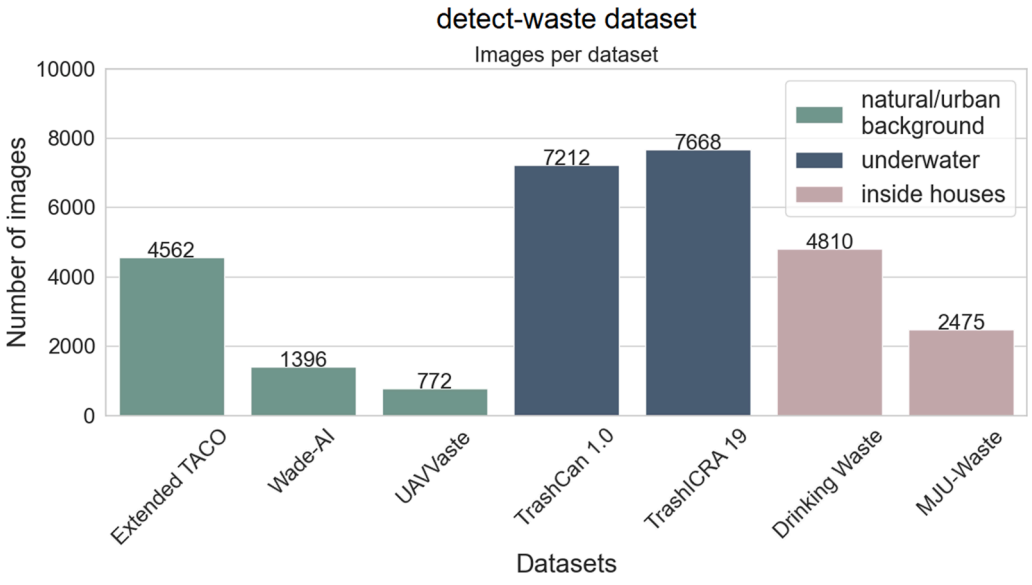

Fig. 1. Numbers of images in individual datasets included in the detect-waste dataset used for detection tasks.

datasets and every subset separately. The dataset is presented in Table 2.

3.2. Proposed detection framework

Despite many advantages, object detectors still require huge amounts of data with proper annotations. Such data is often hard to collect, especially when most available datasets do not meet the requirements of having both bounding box annotations and rich-enough classification labels. As presented in the literature review, some datasets provide only classification labels without bounding boxes. Hence, a decision was made to train localization and classification models separately, to fully utilize the available data.

In this study the problem of incomplete data annotations is addressed by dividing the detection process into two separate stages: litter localization and litter classification. At first, the object detector searches for possible regions with litter. Then, the proposed regions are cropped and forwarded to the classifier. The classifier is used to determine the recyclable litter type. Additionally, except for seven main litter types defined in this research, an additional class *background* is proposed to eliminate false positive predictions. Finally, the user is presented with the image showing all detected litter and its type. The region classified as *background* is excluded from the visualization. The developed pipeline is presented in Fig. 2.

The training procedure is as follows. In the first stage, the neural network is trained to detect regions with *litter*. Afterward, the trained object detector is used to find waste in the images with only classification labels. The provided region proposals from the images without detection annotations are cropped and saved as the classification dataset. The instances with already provided bounding boxes are cut out using ground-truth annotations to provide more stable training. Next, in the second stage, the classification model is trained to assign a proper class to the detected litter.

In contrary to the object detector, the classifier is trained in a semi-supervised fashion. Here, advantage is taken of thousands of unannotated data by applying the pseudo-labeling technique (Lee 2013). The main concept of pseudo-labeling is to use a model to label data without annotations to take advantage of it as a ground truth. Hence, firstly, the unlabeled data is pseudo-labeled by the model trained on target data. Then, pseudo-labels are merged with target data and used as ground-truth in further training. Both steps can be repeated after every batch or epoch. This procedure provides a considerably larger, yet pseudo-annotated, dataset. Moreover, it exposes the model to more diverse data, which could not be used in fully-supervised training.

Three popular object detection networks have been analyzed: EfficientDet (Wightman 2020), DETR (Carion et al. 2020), and Mask R-CNN (Ren et al. 2015). The first two architectures allow for object detection, whereas the third one also implements segmentation. All models were trained for waste detection. The average precision of litter detection obtained using the detect-waste dataset with one class ranged from 28.0% for Mask R-CNN with ResNet-50 backbone to 65.5% for EfficientDet-D2. Due to a significant gap observed between EfficientDets and other tested architectures, the remaining experiments were limited to this network family. It is noteworthy that some more complex networks from this family (i.e., EfficientDet-D3) which were also tested exhibited a similar performance. Since the EfficientDet-D2 network reached the best evaluation results, and taking into consideration that this network has fewer parameters and requires less computing power, a decision was made to proceed with it exclusively. Table comparing results from previous studies is presented in Supplementary Table 2.

For classification, EfficientNet-B2 was used as it achieved the best results compared to ResNet-50 and EfficientNet-B4. Results from our previous studies are presented in an open-source repository (Mikołajczyk et al., 2021).

4. Experiments and results

This section presents the conducted experiments and shows the results of the proposed methodology. Various architectures and numerous hyperparameters have been compared to ensure the efficiency of the solution.

4.1. Data

The detect-waste dataset was used as a primary litter detection benchmark. Additionally, the object detector was trained and tested on all detect-waste subsets: Extended TACO, UAVVaste, TrashCan, Trash-ICRA, MJU-Waste, drinking-waste, and wade-ai.

For the classifier, the classify-waste dataset was used. Moreover, around 55k unlabeled images from Open Litter Map (Lynch 2018) were used to train the classifier in a semi-supervised fashion. It is worth noticing that the data instances were not passed straightforwardly to the classifier. Instead, each image was passed through the object detector to find regions with litter. Then, the predicted bounding boxes were cut from the input image and used to feed the classifier.

80% of the original images were randomly selected as the training set, 10% as validation, and 10% as the test set. To avoid data leakage, each subset of detect-waste and classify-waste was split separately, as described in the Methodology section. The same split was preserved for the common subsets between detect-waste and classify-waste. Additionally, training and test data distributions were ensured by preserving the percentage of samples for each class.

4.2. Performance metrics

The Mean Average Precision (mAP) was used for detection evaluation. Mean Average Precision metric shows how well the predicted box fits the Ground Truth (localization), and whether the class label is correctly predicted (classification). For each bounding box, the intersection over union (IoU) between the predicted bounding box and the Ground Truth was calculated; if the Area of Overlap to the Area of Union was higher than the threshold, the prediction was estimated as correct. The mAP score was averaged for all categories.

The mAP with IoU threshold 0.50 (mAP50) was used as a basic evaluation metric in the detection task. Additionally, the mAP with 0.75 IoU threshold level (mAP75) and the mAP integrated over IoUs were also used in the range from 0.5 to 0.95 with step 0.05 (AP). For better evaluation of how different-sized objects are detected, AP was used for small (AP_S), medium (AP_M), and large (AP_L) objects.

Table 2
Number of images per class in classify-waste and its subsets. Only the proposed subsets (Extended TACO, additional images from Google search) contain images from more than two waste categories.

Dataset	Background	bio	glass	Metals and plastic	Paper	Non-recyclable	Other	Unknown
Extended TACO	–	69	592	6057	601	2802	154	3258
Drinking waste	–	–	1162	3604	–	–	–	–
waste-pictures & Google search	–	92	49	–	203	–	366	–
TrashNet	–	–	501	–	801	–	–	–
Places	1017	–	–	–	–	–	–	–
Classify-waste	1017	161	2304	9661	1605	2802	520	3258

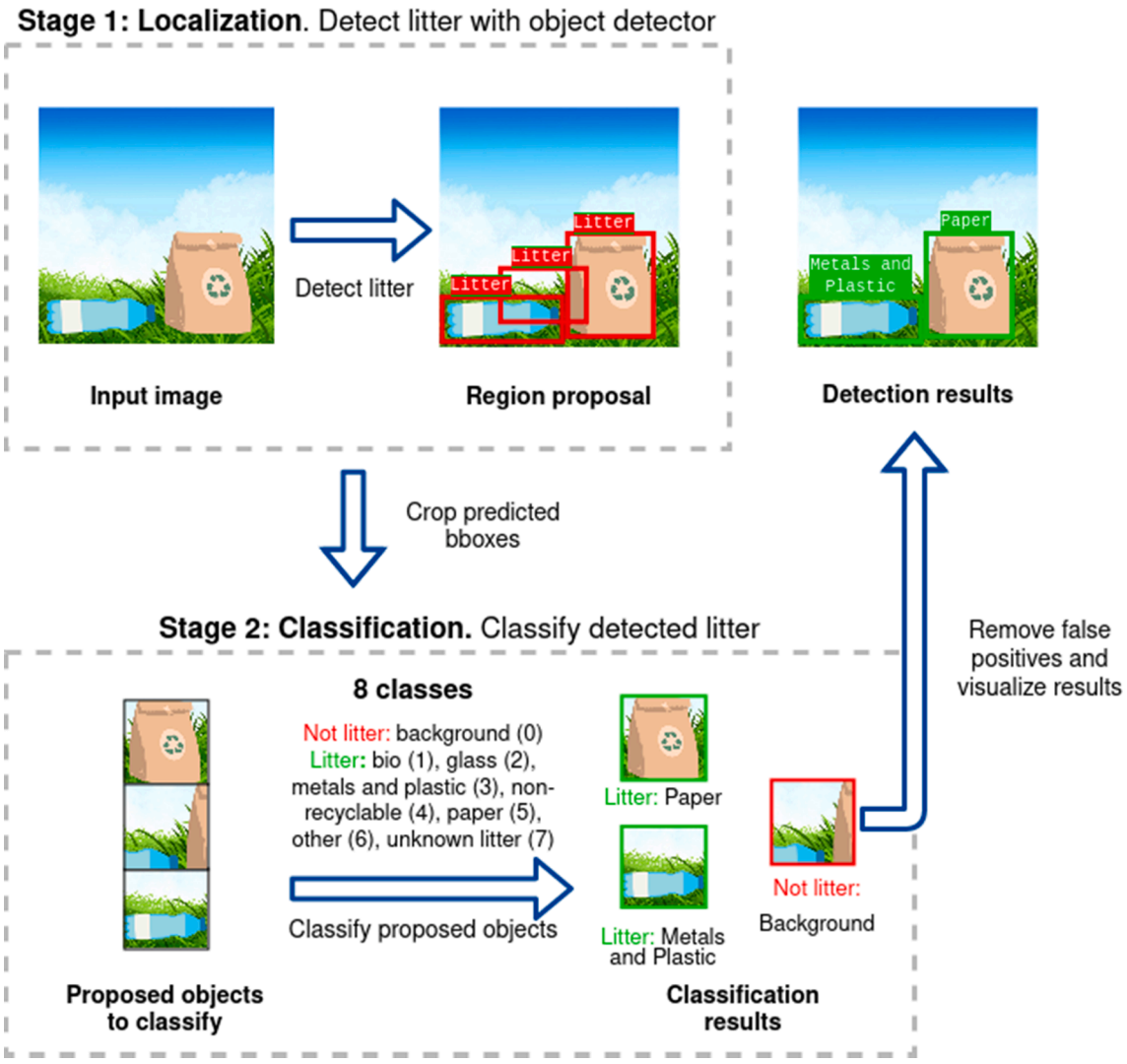

Fig. 2. The pipeline of two-stage framework to detect and sort litter.

For classification, the F_1 score was mostly used, as it balances Precision and Recall metrics $\left(F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}\right)$. Precision shows how accurate predictions are, while Recall shows how many samples from this class were correctly predicted. The F_1 score is a widely used measure for uneven class distribution, which is the case in most waste datasets.

Table 3
Results obtained on different datasets using EfficientDet-D2. The achieved results can be compared with those previously reported in literature mAP50 equalled to 15.9% for TACO with Mask R-CNN (Proença and Simões 2020), or 55.4% for TrashCan with Mask R-CNN (Hong, Fulton, and Sattar 2020).

Train subset	Test subset	#	mAP	mAP50	mAP75	AP_S	AP_M	AP_L
Drink waste	Drink waste	4	85.6	99.4	98.3	–	79.4	86.3
Extended TACO	Extended TACO	7	11.9	16.2	13.0	6.4	9.4	15.0
Extended TACO	Extended TACO	1	39.7	55.7	42.7	18.9	37.3	50.2
MJU-Waste	MJU-Waste	1	82.2	97.9	95.4	71.3	81.8	87.6
TrashCan	TrashCan	8	66.0	91.3	74.9	52.9	70.6	69.3
Trash-ICRA	Trash-ICRA	7	3.9	7.3	3.9	–	7.4	5.9
UAVVaste	UAVVaste	1	40.5	74.1	38.6	9.2	42.8	74.2
Wade-AI	Wade-AI	1	46.8	71.5	52.9	14.8	41.1	55.9
detect-waste	Drink waste	1	85.2	99.1	98.4	–	78.1	85.8
detect-waste	Extended TACO	1	45.4	62.4	49.9	20.8	41.5	58.1
detect-waste	MJU-Waste	1	78.2	97.3	89.0	61.3	78.3	85.5
detect-waste	TrashCan	1	64.5	94.8	72.2	57.0	66.7	70.5
detect-waste	Trash-ICRA	1	32.3	58.2	31.8	–	37.5	39.8
detect-waste	UAVVaste	1	25.9	56.1	20.6	3.7	26.3	55.9
detect-waste	Wade-AI	1	18.1	35.3	16.2	4.7	11.4	25.7
detect-waste	indoor subset	1	84.6	99.0	97.8	60.0	78.0	85.8
detect-waste	outdoor subset	1	37.5	55.4	39.8	15.6	31.6	52.3
detect-waste	underwater subset	1	42.1	68.0	44.9	5.9	50.9	47.3
detect-waste	detect-waste	1	45.8	65.5	50.2	5.9	49.1	59.7

4.3. Waste localization in various environments

In the first stage, the proposed framework localizes litter in the image without recognizing its class. The results of a comprehensive study using EfficientDet-D2 and selected datasets altogether and separately are summarized in Table 3.

The table is divided into two distinctive groups depending on the applied train and test procedure. In the first group, the model was trained and tested on each subset separately, while in the second group, the model was trained on the detect-waste dataset and evaluated on individual waste subsets. The presented results were obtained on the indoor, outdoor, and underwater subsets coming from different environments.

The mAP50 calculated per dataset was the highest for images presenting indoor scenarios, namely the datasets Drink-waste (cans, plastic and glass bottles) and MJU-Waste (one hand-held waste object per image). Moreover, EfficientDet-D2 also reached a very high score, mAP50 above 90%, for TrashCan with selected 8 underwater waste categories. The worst result, with mAP50 below 10%, was achieved for the blurred underwater images from the Trash-ICRA dataset, which could be related to poor quality of the images. That is the reason why this dataset was excluded from the classify-waste dataset. On the other hand, the detection performance for images taken in natural or urban background ranged from 16.2% for Extended TACO (trash in various environments) to 74.1% for UAVVaste (small objects constituting over 80% of the dataset shown from a bird's eye view), which proved that precise detection of garbage in different environments is possible. The corresponding sample predictions are shown in Fig. 3.

The achieved results have confirmed the importance of data quality in the learning process of DL-based systems. Apart from different quality of images and different environments, the number and kind of waste classes varied depending on the analyzed dataset. Moreover, the annotated trash objects differed in shape and size. In the case of images presenting waste in an indoor scenario, using the detect-waste dataset for training slightly worsened the results. On the contrary, the same procedure for a more diverse background improved the detection accuracy by almost 50 pp in the case of Trash-ICRA dataset. High results illustrating the average detection precision for three different environments suggest that one-class detection outperforms multi-class detection.

Similarly, an experiment on seven waste classes was also performed to detect trash in natural and urban environments. However, this significantly reduced mAP in the analyzed evaluation metrics, reaching almost a 4 times smaller value than the result achieved while training on the detect-waste dataset. Regarding the size of the detected objects, for tiny objects the multi-label detector trained on the Extended TACO dataset reached better results (6.4%) than that trained on the detect-waste dataset with one class (5.9%). This may be because in Extended TACO, approximately 45% of instances are small (area $< 32^2$), while for detect-waste it is only about 25% of the whole dataset. It is worth emphasizing that a single stage, 7-class detector demonstrated a significantly lower mAP score than a one-class model. It achieved a similar score of AP_{50} equal to 43.3% only for one category – *metals and plastic* – leaving the rest with scores between 0.1% for *bio* and 8.9% for *unknown* class. For that reason, it was decided to perform classification in a separate stage.

4.4. Classification results using the CNN

At the second stage of the proposed approach, semi-supervised, multi-class classification for seven waste categories was performed. As for the supervised part, the classification networks were trained on the classify-waste benchmark. For the unsupervised part (pseudolabeling), unlabeled litter from the openlittermap dataset was used. To imitate the two-stage system, the trash was cut-out with previously trained waste localizer, described in the previous subsection. The boundaries of the cropped litter were established using bounding boxes for annotated images and objects detected with the trained detection model - EfficientDet-D2 (Tan, Pang, and Le 2020). Combined waste instances were applied as input images to solve the classification problem.

The experiments have shown that updating pseudo-labels every batch can slightly increase the accuracy. When analyzing the confusion matrices of each training, it was noticed that applying a weighted sampler provides more balanced results for each class. As a result, the accuracy of 73% has been achieved (and 86.7% on the training set), while almost 25% of the used dataset were test images.

Although the confusion matrix clearly shows that most of the predictions are accurate (see Fig. 4, and Table 4), it also indicates a significant data imbalance. The *metals and plastic* class was predicted with the highest precision of 87%, which is connected to the large representatives of this class. Still, it also resulted in a relatively low recall, which means that many objects were misclassified as *metals and plastic*.

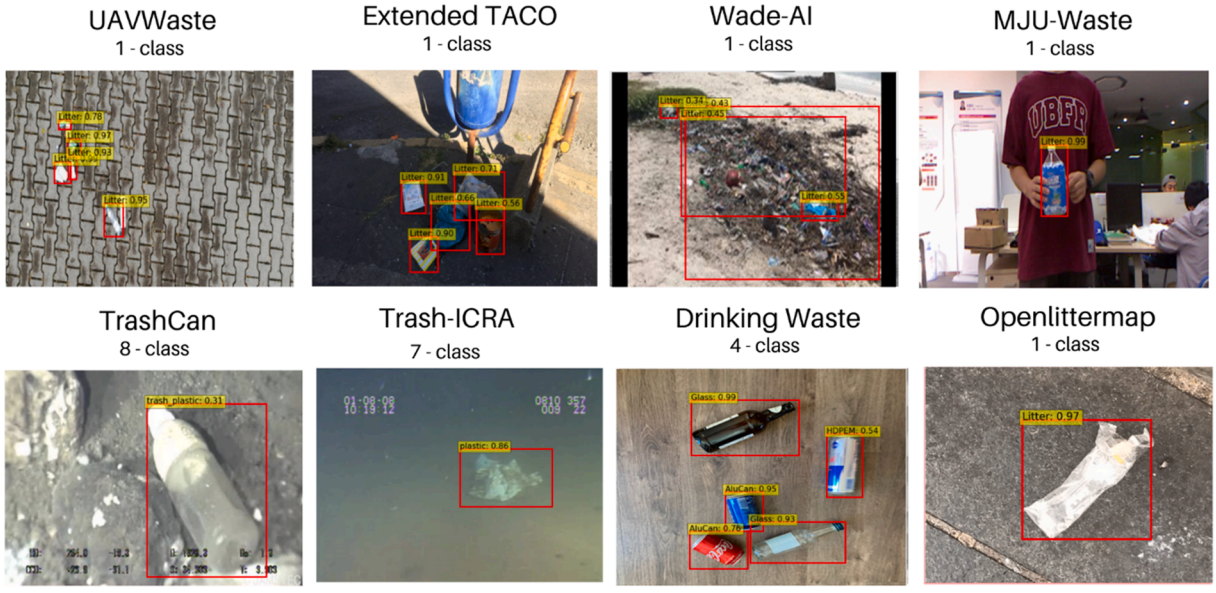

Fig. 3. Example EfficientDet-D2 predictions for diverse waste datasets. The images were taken in different locations such as a beach, pavement, indoor, and underwater. The detected objects vary in size and number – the images show from one to five small, medium, or large objects.

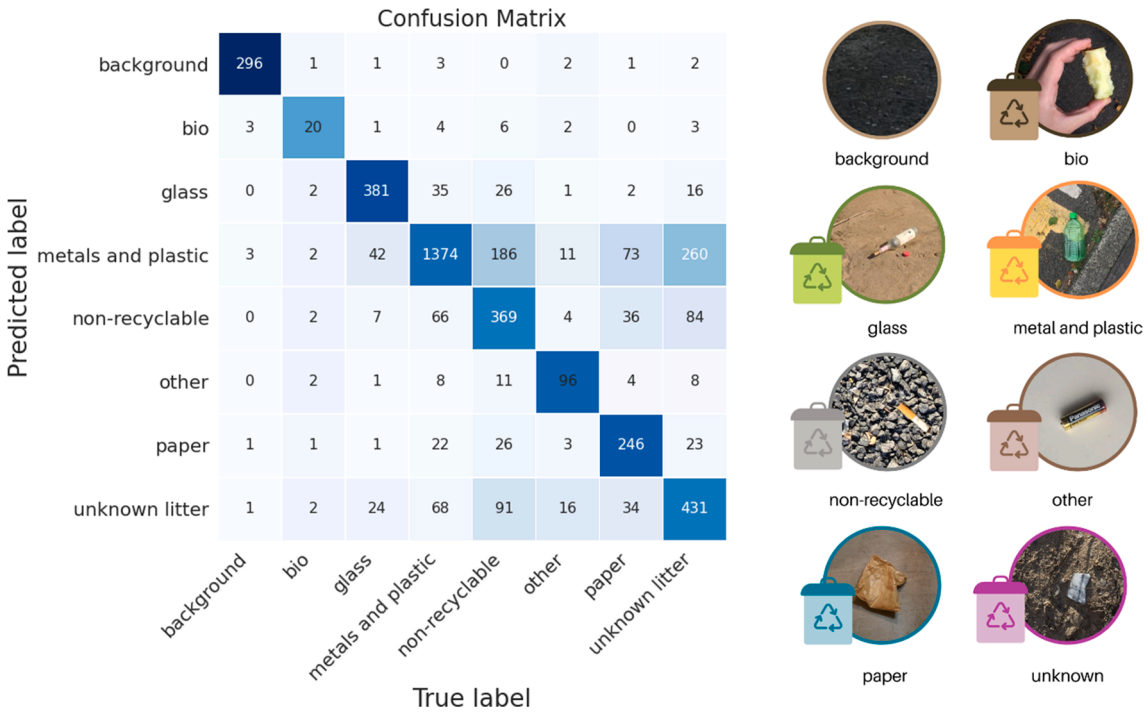

Fig. 4. Evaluation of the classification accuracy of a form of confusion matrix for EfficientNet-B2, weighted sampler, and pseudo-labeling per batch.

Table 4
Summary of precision, recall, and F_1-score per category on classify-waste.

Class name	Precision	Recall	F_1-score
background	0.97	0.97	0.97
bio	0.62	0.51	0.56
glass	0.83	0.82	0.83
metals and plastic	0.87	0.70	0.78
non-recyclable	0.52	0.65	0.58
other	0.71	0.74	0.72
paper	0.62	0.76	0.68
unknown litter	0.52	0.65	0.58

There was a noticeable problem with identifying the *unknown* and *non-recyclable* classes, in which precision was equal to 52%. The other classes were recognized with higher precision but still not fully correct due to data imbalance.

The biggest confusion was observed between the categories *metals and plastic* and *unknown*. Probably this was because of partly degraded or destroyed trash. All the classes were rarely mistaken for *glass*, evidenced by the high recall value – 82%. The F_1-score metric for this class achieved the utmost result – 83%, balancing between false positives and negatives. It is worth noticing that eliminating the background from the rest of the waste was extremely successful - at the level of 97% for precision and 97% for recall. Adding a separate class for background improved the performance and, as assumed, reduced the number of false positives.

Training details: Both EfficientDet-D2 (Tan, Pang, and Le 2020) and EfficientNet-B2 (Tan and Le 2019) were implemented in PyTorch. EfficientDet-D2 was trained with Adam optimizer with the decay rate set to 0.95. The EfficientDet-D2 training stage started with the learning rate of 1e-3 and lambda scheduler, and continued for 20 epochs with a batch size of 16. During all experiments with EfficientNet-B2, the learning rate was set to 1e-4 and the batch size to 16, and each network was trained for 20 epochs. The input images were normalized and resized to 260x260. Additionally, the data augmentation techniques such as crop, flip, rotate, brightness/contrast, and cutout, were randomly applied.

Inference time: The inference time was studied as the function of the localized object count, assuming that all detected objects can be classified in a single batch. The experiments were conducted using the EfficientNetB2 model with input resolution 260, single precision computations FP32 and FP16 on Nvidia Tesla T4 GPU and batch sizes varying from 1 to 120. 100 inference times were averaged to fit the linear function 4.429 ms • batch +44.2 ms with MSE = 17.42 (FP16: 0.071 ms • batch + 260.7 ms with MSE = 2.37). The average time needed to classify 120 objects was 576 ms (FP16: 269.2 ms) which made it possible to classify up to 208 (FP16: 445) detected objects per second. The proposed two-stage method can be used at 30fps with the limit of 7 (FP16: 15) objects per image. The performance can be further improved by using INT8 quantization.

5. Conclusions and future work

There is a visibly increasing demand for artificial intelligence in numerous human activities. Following that, one of the first benchmarks to classify litter into seven household waste categories, observed in the wild environment, with accuracy up to 75% are presented. That addressed the need for objective comparison between different approaches, consequently accelerating the optimization cycle of deep learning models, increasing the chances of constant improvement and knowledge sharing in academics and industry. Also, a DL-based framework that can localize trash in the image and then identify its class using two separate neural networks is proposed. Achieved results for trash localization on mixed datasets outperform previous studies reported for used datasets separately. The two framework's modules were tested individually by conducting exhausting studies on publicly available waste data in diverse environments: inside houses, in natural or urban environment, and underwater. A wide range of baseline results obtained by the authors regarding these environments and various object sizes will help other researchers in future experiments in the field.

The presented framework shows the great potential of the DL-based methodology for waste management in households by determining the right waste category using a mobile application. In the future, with the assistance of DL models, it would be possible to mount robotic arms in waste management plants to automatically distinguish between

different classes of objects and sort garbage without human intervention. This could lead to automation of waste recycling in both waste sorting plants and households using DL algorithms. However, it is important to mention that industrial waste sorting is often more complicated than household one (Wilts et al. 2021), therefore regarding the application, it is required to define more specific classes like different types of plastic (Bobulski and Kubanek 2019), metal alloys (Díaz-Romero et al. 2021), batteries (Sterkens et al. 2021) or combustible waste to produce energy from waste (Brunner and Rechberger, 2015). Proposed work can be a baseline in waste detection and classification for further research. Additionally, high precision of litter localization in a large variety of environments shows the possibility of using neural networks for waste monitoring in cities or detecting illegal dumps in nature, for example, with drones. This will allow for environment monitoring by automated measurement of the environmental pollution level, even underwater. Special robots equipped with waste detection and classification modules could be sent to the most polluted places (also inaccessible for humans) to clean them. All this will lead to minimizing the cost of maintaining the cleanliness of our surroundings.

As Artificial Intelligence is required to be more accurate than a human, the main future direction for the proposed system will be to improve its performance. Selected detectors work well when localizing medium and large objects, but recognition of small litter is still challenging. For that reason, exploring different state-of-the-art approaches, such as Deformable DETR (Zhu et al. 2020), seems to be a good idea. On the other hand, a more balanced dataset and the use of the latest EfficientNetv2 (Tan and Le, 2021), could also boost the classification accuracy. Moreover, further reductions in predictive time are worth exploring since they continue to be a challenge.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

SM, AM, MF and ZK contributed to the final version of the open-source code. SM, AM, MF, and ZK analyzed and interpreted the waste datasets regarding the waste sorting rules in Gdańsk (Poland). AM and SM planned the experiments and adjusted the models. SM prepared the results of the experiments. AM led the team during the project. SM took the lead in the preparation of the manuscript. AM prepared the illustrations. AK and KM helped supervise the project.

SM wrote the Introduction. SM and AM conducted the research on available datasets. AK conducted and wrote the review on the deep learning classification. KM conducted and wrote the review on deep learning object detection. ZK analyzed the data and wrote the datasets sections. MP and MF described the proposed approach and experiments. SM concluded the project and proposed future directions.

All authors provided critical feedback and helped shape the research, analysis, and manuscript.

The 5-month (October 2020 – February 2021) project "Detect Waste in Pomerania" was organized and led by Agnieszka Mikołajczyk, Magdalena Kortas, and Ewa Marczewska from Women in Machine Learning & Data Science Trojmiasto. A team of carefully selected members, nine female data scientists, analysts, and machine learning engineers supported by five industry mentors, studied and worked together on developing a model for trash detection. The authors acknowledge other team members: Anna Brodecka, Magdalena Kortas, Katarzyna Łagocka, Ewa Marczewska, Pedro F. Proença, Adam Kaczmarek and Iwona Sobieraj who contributed to the project.

The authors acknowledge the infrastructure and support of Digital Innovation Hub dih4.ai in Gdańsk. The authors wish to express their thanks to Voicelab.ai for the financial support and Epinote for data annotation.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wasman.2021.12.001.

References

Awe, Oluwasanya, Robel Mengistu, and Vikram Sreedhar. 2017. Smart Trash Net: Waste Localization and Classification. CS229 Project Report. Stanford University: Stanford.

Bircanoğlu, Cenk, Nafiz Arıca, 2018. A Comparison of Activation Functions in Artificial Neural Networks. In: 2018 26th Signal Processing and Communications Applications Conference (Siu), 1–4. doi:10.1109/SIU.2018.8404724.

Bobulski, J., Kubanek, M., 2019. Waste Classification System Using Image Processing and Convolutional Neural Networks. In: Rojas I., Joya G., Catala A. (eds) Advances in Computational Intelligence. IWANN 2019. Lecture Notes in Computer Science, vol 11507. Springer, Cham. https://doi.org/10.1007/978-3-030-20518-8_30.

Bochkovskiy, Alexey, Chien-Yao Wang, Hong-Yuan Mark Liao, 2020. Yolov4: Optimal Speed and Accuracy of Object Detection. arXiv Preprint arXiv:2004.10934.

Brunner, P.H., Rechberger, H., 2015. Waste to energy – key element for sustainable waste management. Waste Management 37, 3–12.

Carion, Nicolas, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko, 2020a. End-to-End Object Detection with Transformers. In: European Conference on Computer Vision, 213–29. Springer.

Carolis, B.D.F. Ladogana, N. Macchiarulo, YOLO Trashnet: Garbage Detection in Video Streams, In: 2020 Ieee Conference on Evolving and Adaptive Intelligent Systems (Eais), 1–7. doi: 10.1109/EAIS48028.2020.9122693.

Chu, Y., Huang, C., Xie, X., Tan, B., Kamal, S., Xiong, X., 2018. Multilayer hybrid deep-learning method for waste classification and recycling. Computat. Intell. Neurosci. 2018, 1–9. https://doi.org/10.1155/2018/5060857.

Dalal, Navneet, Triggs, Bill, 2005. Histograms of Oriented Gradients for Human Detection. In: 2005 Ieee Computer Society Conference on Computer Vision and Pattern Recognition (Cvpr'05), 1:886–93. Ieee.

Dillam Díaz-Romero, Wouter Sterkens, Simon Van den Eynde, Toon Goedemé, Wim Dewulf, Jef Peeters, Deep learning computer vision for the separation of Cast- and Wrought-Aluminum scrap, Resources, Conservation and Recycling, Volume 172, 2021, 105685, ISSN 0921-3449, https://doi.org/10.1016/j.resconrec.2021.105685.

European household waste collection (Eurostat). 2018. Waste statistics [Data file]. Retrieved from https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Waste_statistics.

Foundation, Let's Do It. 2016. "Wade-Ai." *GitHub Repository*. https://github.com/letsdoitworld/wade-ai; GitHub.

Fulton, M., Hong, J., Islam, M.J., Sattar, J., 2019. Robotic Detection of Marine Litter Using Deep Visual Detection Models. In: 2019 International Conference on Robotics and Automation (Icra), 5752–8. doi:10.1109/ICRA.2019.8793975.

Girshick, Ross, 2015. Fast R-Cnn. In: Proceedings of the IEEE International Conference on Computer Vision, 1440–8.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587.

Glouche, Yann, Couderc, P., 2013. A smart waste management with self-describing objects. In: The Second International Conference on Smart Systems, Devices and Technologies (Smart'13). Rome, Italy.

Gu, Chunhui, Lim, Joseph J., Arbeláez, Pablo, Malik, Jitendra. 2009. Recognition using regions. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 1030–7. IEEE.

He Kaiming, Gkioxari Georgia, Dollár Piotr, Girshick Ross. 2017. Mask R-Cnn. In: Proceedings of the IEEE International Conference on Computer Vision, 2961–9.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (Cvpr), pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

Hong, Jungseok, Fulton Michael, Sattar Junaed, 2020. "TrashCan: A Semantically-Segmented Dataset Towards Visual Detection of Marine Debris.".

Huang, Gao, Liu, Zhuang, Van Der Maaten, Laurens, Weinberger, Kilian Q., 2017. Densely connected convolutional networks. In: 2017 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr), 2261–9. doi:10.1109/CVPR.2017.243.

Huang, Xin, Wang, Xinxin, Lv, Wenyu, Bai, Xiaying, Long Xiang, Deng, Kaipeng, Dang, Qingqing et al., 2021. PP-YOLOv2: A Practical Object Detector." arXiv preprint arXiv:2104.10419 (2021).

J. A. for Marine-Earth Science, Technology, Deep-sea debris database, 2012.

Kaza, Silpa, Lisa C. Yao, Perinaz Bhada-Tata, and Frank Van Woerden. 2018. *What a Waste 2.0*. World Bank Publications 30317. The World Bank. https://ideas.repec.org/b/wbk/wbpubs/30317.html.

Konovalenko, I., Maruschak, P., Brezinová, J., Viňáš, J., Brezina, J., 2020. Steel surface defect classification using deep residual neural network. Metals 10 (6), 846. https://doi.org/10.3390/met10060846.

Kraft, M., Piechocki, M., Ptak, B., Walas, K., 2021. Autonomous, onboard vision-based trash and litter detection in low altitude aerial images collected by an unmanned aerial vehicle. Remote Sens. 13 (5) https://doi.org/10.3390/rs13050965.

Lanorte, A., De Santis, F., Nolè, G., Blanco, I., Loisi, R.V., Schettini, E., Vox, G., 2017. Agricultural plastic waste spatial estimation by Landsat 8 satellite images. Comput. Electron. Agric. 141, 35–45. https://doi.org/10.1016/j.compag.2017.07.003.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.
Lee, Dong-Hyun. 2013. Pseudo-Label : The simple and efficient semi-supervised learning method for deep neural networks. ICML 2013 Workshop: Challenges in Representation Learning (WREPL), July.
Li, W.C., Tse, H.F., Fok, L., 2016. Plastic waste in the marine environment: a review of sources, occurrence and effects. Sci. Total Environ. 566-567, 333–349.
Liang, S., Yu, G.u., 2021. A deep convolutional neural network to simultaneously localize and recognize waste types in images. Waste Manage. 126, 247–257. https://doi.org/10.1016/j.wasman.2021.03.017.
Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft Coco: common objects in context. In: Computer Vision – Eccv 2014, edited by David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars, 740–55. Cham: Springer International Publishing.
Liu, Wei, Anguelov, Dragomir, Erhan, Dumitru, Szegedy, Christian, Reed, Scott, Fu, Cheng-Yang, Berg, Alexander C, 2016. Ssd: Single Shot Multibox Detector. In: European Conference on Computer Vision, 21–37. Springer.
Liu, Y., Ge, Z., Lv, G., Wang, S., 2018. Research on automatic garbage detection system based on deep learning and narrowband internet of things. J. Phys.: Conf. Ser. 1069, 012032. https://doi.org/10.1088/1742-6596/1069/1/012032.
Lu, W., 2019. Big data analytics to identify illegal construction waste dumping: a Hong Kong study. Resour. Conservat. Recycl. 141, 264–272. https://doi.org/10.1016/j.resconrec.2018.10.039.
Lynch, S., 2018. OpenLitterMap.com – Open Data on Plastic Pollution with Blockchain Rewards (Littercoin). Open Geospatial Data, Softw. Stand. 3 (6) https://doi.org/10.1186/s40965-018-0050-y.
Mikołajczyk, Agnieszka, Majchrowska, Sylwia, Ferlin, Maria, Marczewska, Ewa, Klawikowska, Zuzanna, Plantykow, Marta, Kortas, Magdalena, Brodecka, Anna, & Łagocka, Katarzyna. 2021. "wimlds-trojmiasto/detect-waste: Detect waste in Pomerania (v1.1-alpha)". Zenodo. https://doi.org/10.5281/zenodo.5375461.
Proença, Pedro F, Simões, Pedro, 2020. TACO: Trash Annotations in Context for Litter Detection. arXiv Preprint arXiv:2003.06975.
Redmon, Joseph, Farhadi, Ali, 2017. YOLO9000: Better, Faster, Stronger. In: Proceedings of the Ieee Conference on Computer Vision and Pattern Recognition, 7263–71.
Redmon, Joseph, Divvala, Santosh, Girshick, Ross, Farhadi, Ali, 2016. You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 779–88.
Ren, Shaoqing, He, Kaiming, Girshick, Ross, Sun, Jian, 2015. Faster R-Cnn: towards real-time object detection with region proposal networks. arXiv Preprint arXiv:1506.01497.
Sahimaa, O., Hupponen, M., Horttanainen, M., Sorvari, J., 2015. Method for residual household waste composition studies. Waste Management 46, 3–14.
Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., 2018. Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520.
Serezhkin, Arkadiy. 2020. "Drinking Waste Classification." https://www.kaggle.com/arkadiyhacks/drinking-waste-classification; Kaggle.
Sheng, T.J., Islam, M.S., Misran, N., Baharuddin, M.H., Arshad, H., Islam, M.R., Chowdhury, M.E.H., Rmili, H., Islam, M.T., 2020. An internet of things based smart waste management system using lora and tensorflow deep learning model. IEEE Access 8, 148793–148811. https://doi.org/10.1109/ACCESS.2020.3016255.
Slagstad, H., Brattebø, H., 2013. Influence of assumptions about household waste composition in waste management LCAs. Waste Manage. 33 (1), 212–219.
Sterkens, W., Diaz-Romero, D., Goedemé, T., Dewulf, W., Peeters, J.R., 2021. Detection and recognition of batteries on X-ray images of waste electrical and electronic equipment using deep learning. Resour. Conserv. Recycl. 168, 105246. https://doi.org/10.1016/j.resconrec.2020.105246.
Tan, M., Pang, R., Le, Q.V., 2020. EfficientDet: Scalable and Efficient Object Detection. In: 2020 IEEE/Cvf Conference on Computer Vision and Pattern Recognition (Cvpr), pp. 10778–10787. https://doi.org/10.1109/CVPR42600.2020.01079.
Tan, Mingxing, Le, Quoc, 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: Proceedings of the 36th International Conference on Machine Learning, edited by Kamalika Chaudhuri and Ruslan Salakhutdinov, 97:6105–14. Proceedings of Machine Learning Research. PMLR.
Tan, Mingxing, Le, Quoc V., 2021. EfficientNetV2: Smaller Models and Faster Training.
Tavakoli, Arash, Ashrafi, Amir, Heydarian, Arsalan, Behl, Madhur, 2018. The internet of wasted things (Iowt). In: IOT '18: Proceedings of the 8th International Conference on the Internet of Things. IOT '18. New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3277593.3277627.
Toğaçar, M., Ergen, B., Cömert, Z., 2020. Waste classification using autoencoder network with integrated feature selection method in convolutional neural network models. Measurement 153, 107459. https://doi.org/10.1016/j.measurement.2019.107459.
Uijlings, J.R.R., van de Sande, K.E.A., Gevers, T., Smeulders, A.W.M., 2013. Selective search for object recognition. Int. J. Comput. Vis. 104 (2), 154–171.
Vaswani, Ashish, Shazeer, Noam, Parmar, Niki, Uszkoreit, Jakob, Jones, Llion, Gomez, Aidan N., Kaiser, Lukasz, Polosukhin, Illia, 2017. Attention Is All You Need. arXiv Preprint arXiv:1706.03762.
Viola, Paul, Jones, Michael, 2001. Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 Ieee Computer Society Conference on Computer Vision and Pattern Recognition. Cvpr 2001, 1:I–I. IEEE.
Wang, Chien-Yao, Bochkovskiy, Alexey, Liao, Hong-Yuan Mark, 2020. Scaled-Yolov4: Scaling Cross Stage Partial Network. arXiv Preprint arXiv:2011.08036.
Wang, Chien-Yao, Yeh, I-Hau, Mark Liao, Hong-Yuan. You only learn one representation: unified network for multiple tasks. arXiv preprint arXiv:2105.04206 (2021).
Wang, T., Cai, Y., Liang, L., Ye, D., 2020b. A multi-level approach to waste object segmentation. Sensors 20 (14). https://doi.org/10.3390/s20143816.
"Waste Pictures." 2019. https://www.kaggle.com/wangziang/waste-pictures; Kaggle.
White, Gary, Cabrera, Christian, Palade, Andrei, Li, Fan, Clarke, Siobhan, 2020. WasteNet: Waste Classification at the Edge for Smart Bins.
Wightman, Ross, 2020. EfficientDet (Pytorch). GitHub Repository. https://github.com/rwightman/efficientdet-pytorch; GitHub.
Wilts, H., Garcia, B.R., Garlito, R.G., Gómez, L.S., Prieto, E.G., 2021. Artificial intelligence in the sorting of municipal waste as an enabler of the circular economy. Resources 10 (4), 28. https://doi.org/10.3390/resources10040028.
Yang, M., Thung, G., 2016. TrashNet. GitHub Repository. https://github.com/garythung/trashnet; GitHub.
Zorpas, A.A., Lasaridi, K., Voukkali, I., Loizia, P., Chroni, C., 2015. Household waste compositional analysis variation from insular communities in the framework of waste prevention strategy plans. Waste Manage. 38, 3–11.
Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A., 2018. Places: a 10 million image database for scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. 40 (6), 1452–1464. https://doi.org/10.1109/TPAMI.2017.2723009.
Zhu, Xizhou, Su, Weijie, Lu, Lewei, Li, Bin, Wang, Xiaogang, Dai, Jifeng. 2020. Deformable Detr: Deformable Transformers for End-to-End Object Detection. arXiv Preprint arXiv:2010.04159.
Zou, Zhengxia, Shi, Zhenwei, Guo, Yuhong, Ye, Jieping, 2019. Object detection in 20 years: a survey. arXiv Preprint arXiv:1905.05055.