Deep learning for detecting macroplastic litter in water bodies
A review

Jia, Tianlong; Kapelan, Zoran; de Vries, Rinze; Vriend, Paul; Peereboom, Eric Copius; Okkerman, Imke; Taormina, Riccardo

DOI
10.1016/j.watres.2023.119632

Publication date
2023

Document Version
Final published version

Published in
Water Research

Water Research 231 (2023) 119632

Contents lists available at ScienceDirect

Water Research

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

Deep learning for detecting macroplastic litter in water bodies: A review

Tianlong Jia [a,*], Zoran Kapelan [a], Rinze de Vries [b], Paul Vriend [c], Eric Copius Peereboom [c], Imke Okkerman [c], Riccardo Taormina [a]

[a] Delft University of Technology, Faculty of Civil Engineering and Geosciences, Department of Water Management, Stevinweg 1, 2628 CN Delft, The Netherlands
[b] Noria Sustainable Innovators, Schieweg 13, 2627 AN Delft, The Netherlands
[c] Rijkswaterstaat, Ministry of Infrastructure and Water Management, Griffioenlaan 2, 3526 LA Utrecht, The Netherlands

ARTICLE INFO

Keywords:
Artificial intelligence
Computer vision
Environmental monitoring
Neural networks
Macroplastics
Pollution

ABSTRACT

Plastic pollution in water bodies is an unresolved environmental issue that damages all aquatic environments, and causes economic and health problems. Accurate detection of macroplastic litter (plastic items >5 mm) in water is essential to estimate the quantities, compositions and sources, identify emerging trends, and design preventive measures or mitigation strategies. In recent years, researchers have demonstrated the potential of computer vision (CV) techniques based on deep learning (DL) for automated detection of macroplastic litter in water bodies. However, a systematic review to describe the state-of-the-art of the field is lacking. Here we provide such a review, and we highlight current knowledge gaps and suggest promising future research directions. The review compares 34 papers with respect to their application and modeling related criteria. The results show that the researchers have employed a variety of DL architectures implementing different CV techniques to detect macroplastic litter in various aquatic environments. However, key knowledge gaps must be addressed to overcome the lack of: (i) DL-based macroplastic litter detection models with sufficient generalization capability, (ii) DL-based quantification of macroplastic (mass) fluxes and hotspots and (iii) scalable macroplastic litter monitoring strategies based on robust DL-based quantification. We advocate for the exploration of data-centric artificial intelligence approaches and semi-supervised learning to develop models with improved generalization capabilities. These models can boost the development of new methods for the quantification of macroplastic (mass) fluxes and hotspots, and allow for structural monitoring strategies that leverage robust DL-based quantification. While the identified gaps concern all bodies of water, we recommend increased efforts with respect to riverine ecosystems, considering their major role in transport and storage of litter.

1. Introduction

Plastic pollution in water bodies is a growing concern with the potential to cause environmental and economic damage, and possible effects on human health (González-Fernández et al., 2021; Lebreton et al., 2018; van Emmerik and Schwarz, 2020). Estimated global emissions of plastic waste to aquatic ecosystems range from 19 to 23 million metric tons in 2016 (Bellou et al., 2021; Borrelle et al., 2020; Jambeck et al., 2015). Rivers are the main source of marine plastic pollution, with estimated riverine plastic emissions of 0.8 to 2.7 million metric tons into oceans each year (Lebreton et al., 2017; Meijer et al., 2021). Furthermore, marine plastic litter may wash up on beaches and shores, and substantial amounts of discarded plastic litter has also been detected in lakes (Imhof et al., 2018; van Emmerik and Schwarz, 2020).

Detecting and quantifying macroplastic litter (plastic items >5 mm) is necessary to assess water pollution levels and develop monitoring strategies for designing effective preventive measures, conducting targeted cleaning campaigns, and devising mitigation interventions (van Emmerik et al., 2022b). Common detection methods include in situ methodologies such as sampling and visual observations (Grøsvik et al.,

Abbreviations: AE, Artificial environment; AI, Artificial intelligence; AUV, Autonomous underwater vehicle; CNN, Convolutional neural network; COCO, Common Objects in Context; CV, Computer vision; DA, Data augmentation; DL, Deep learning; DSGC, Device setup generalization capability; EGC, Environmental generalization capability; GGC, Geographical generalization capability; IC, Image classification; IoU, Intersection over union; IS, Image segmentation; mAP, mean average precision; ML, Machine learning; MLOps, Machine learning operations; MLP, Multilayer Perceptron; NGC, Non-aquatic generalization capability; OA, Overall accuracy; OD, Object detection; OSPAR, Oslo and Paris Conventions; ROV, Remotely operated vehicle; TL, Transfer learning; UAV, Unmanned aerial vehicle.

* Corresponding author.
E-mail address: T.Jia@tudelft.nl (T. Jia).

https://doi.org/10.1016/j.watres.2023.119632
Received 31 May 2022; Received in revised form 3 November 2022; Accepted 15 January 2023
Available online 16 January 2023

2018; Hardesty et al., 2017; van Emmerik and Schwarz, 2020). More recently, methods based on Computer Vision (CV) have been proposed to replace these accurate, but time-consuming and labor-intensive in situ methodologies with automatic detection from images (Kylili et al., 2020, 2019; van Lieshout et al., 2020). In particular, researchers focused on Deep Learning (DL) methods, a subset of flexible machine learning (ML) techniques reaching state-of-the-art performances in many fields of science and technology, including water resources management and engineering (Sit et al., 2020). DL belongs to the family of representation learning techniques that replace manual feature engineering via automatic discovery of the representations needed for feature detection from raw data (LeCun et al., 2015). This allows DL models to reach state-of-the-art performances in all the generic computer vision (CV) tasks related to aquatic macroplastic litter detection, such as image classification (IC), object detection (OD) and image segmentation (IS) (Chai et al., 2021). While preliminary results are promising, this specific field is still in its infancy (van Emmerik and Schwarz, 2020). The researchers must increase their efforts to devise DL-based applications that can help tackling plastic pollution in water bodies at scale. To help build the road ahead, this paper presents a critical review outlining the state-of-the-art of DL-based detection of macroplastic litter in water bodies. The main purposes of this review are (1) to improve researchers' understanding of DL-based macroplastic litter detection by conducting a thorough assessment of the current state of the field; (2) to identify key knowledge gaps and (3) to provide future research directions concerning promising methodologies and novel applications to address these gaps.

The paper is structured as follows. The methodology used to select and analyze the reviewed papers is described in Section 2. Section 3 thoroughly reviews the selected papers with critical discussion points. Section 4 identifies a series of knowledge gaps linked to suggested future research directions. Finally, Section 5 provides conclusions.

2. Methodology

2.1. Search methodology

In this review, we analyzed 34 peer-reviewed journal papers and conference proceedings retrieved from the "Scopus" and "Web of Science" databases. We employed the following steps to identify these papers. Firstly, we searched for papers published until the end of 2021 by employing three sets of keywords: (1) Deep learning-related keywords included "deep learning", "neural network", "artificial intelligence" and "machine learning"; (2) Macroplastic litter-related keywords included "plastic", "trash", "litter", "debris" and "garbage"; (3) Water bodies keywords included "marine", "sea", "ocean", "beach", "shore", "river", "channel", "canal", "waterway" and "lake". The literature search identified papers containing combinations of these terms in their titles, keywords, and abstract. After reviewing the abstracts of all papers matching the inclusion criteria, we selected 33 papers, with the first publication dating back to 2016 (Valdenegro-Toro, 2016). Finally, we conducted a snowball search by checking the citations of these publications. The procedure yielded a total of 34 papers, which are listed in Table 1 along with the most important details. These papers are ordered by the type of water bodies and then by publication year within type in Table 1. If the study features different CV tasks, the review considers each of these tasks separately. When multiple architectures are tested, we report only the architecture achieving the highest performances, which is listed in the "Model architecture" column of Table 1.

2.2. Review methodology

Fig. 1 shows the most relevant factors used to classify and analyze the selected 34 papers. The factors are: (i) the water body(ies) polluted by macroplastic litter (reviewed and discussed in Section 3.1); (ii) dataset on macroplastic litter in water including the dataset source(s), dataset label(s), the dataset size and dataset split (Section 3.2); (iii) details on the CV task(s) performed to detect macroplastic litter, including the type of CV task(s) and model architecture(s) for each CV task (Section 3.3); (iv) whether the authors resorted to data augmentation (DA) and transfer learning (TL) techniques to improve model performances (Section 3.4); (v) the generalization capability of macroplastic litter detection models (Section 3.5); and (vi) details on the metric(s) used for performance evaluation (Section 3.6).

3. Review and discussion

3.1. Water bodies polluted by macroplastic litter

Fig. 2 shows the distribution of water bodies and dataset sources in the reviewed literature. Some papers are counted multiple times since they either consider multiple water bodies (Jakovljevic et al., 2020; Kylili et al., 2020; Panwar et al., 2020; Watanabe et al., 2019; Wolf et al., 2020) or employ multiple dataset sources (Hegde et al., 2021; Watanabe et al., 2019; Wu et al., 2020). Most studies dealt with macroplastic pollution in real settings, with the exception of two studies that considered a controlled artificial environment (AE). Valdenegro-Toro (2016a) and Wu et al. (2020) collected data in a water tank for model training and test, which is time-saving and cost-effective. However, they did not investigate the generalization capability of DL models from studies in AE to field case studies. Field applications are different from studies in AE, because it is generally difficult to replicate the wide variety of litter and environmental conditions (e.g., natural lighting) witnessed in real settings (Valdenegro-Toro, 2016). We encourage further studies to test the performance of DL models trained with AE data when applied to real scenarios to assess the overall suitability of this approach and to allow future benchmarking of different methods.

We categorized the bodies of water examined in the reviewed papers into (i) beaches and shores, (ii) marine underwater, (iii) marine surface, (iv) rivers (including natural rivers, waterways, urban channels and their banks), and (v) lakes. Most of the studies (18 out of 34 papers) focused on macroplastic pollution in marine environments, including marine underwater (11 papers), marine surface (7 papers), and beaches and shores (12 papers). Fewer studies were concerned with river pollution (6 papers), and only one study dealt with macroplastic litter in lakes. This is somewhat expected since, contrary to the consolidated body of knowledge for macroplastic pollution in marine environments, the first scientific studies on the quantification of riverine macroplastic litter date only to the early 2010s (Blettler et al., 2018; van Emmerik and Schwarz, 2020); the first studies concerning lakes are even more recent (Imhof et al., 2018).

3.2. Dataset on macroplastic litter

3.2.1. Employed dataset sources

Researchers collected input data using imaging devices such as standard digital cameras, underwater cameras, cameras mounted on unmanned aerial vehicles (UAVs) or manned aircrafts, cameras mounted on phones, as well as satellite cameras and sonar technologies. As shown in Fig. 2, camera images (12 out of 34 papers), underwater camera images (11 papers), and airborne imagery (8 papers) are the three most popular dataset sources, while 4 studies have used phone images and only 1 study resorted to sonar images or satellite imagery.

Due to affordable costs and user-friendliness, digital cameras are popular data gathering devices regardless of the studied body of water (Mustafah et al., 2012). Fixed cameras are installed on bridges to monitor floating macroplastic litter on the river surface (van Lieshout et al., 2020). One disadvantage is the limited coverage due to their fixed positions and limited viewpoints. Cameras attached to a vessel can survey broader areas (de Vries et al., 2021; Kylili et al., 2019), although these surveying activities are time-consuming and labor-intensive compared to fixed installations. Five studies used camera images which were partially or completely retrieved from structured databases

Table 1
Details of reviewed papers.

Reference	Water body	Dataset				Computer vision task		TL	DA	GC	Performance evaluation[1]	
		Source	Size (# images)	Split (%)	No. classes	Type	Model architecture				Metric	Performance
Wu et al. (2020)	AE	phone, camera und.[2]	1400	80/0/20	3	OD	YOLO v4	✓	✓		mAP	mAP=82.7%
Valdenegro-Toro (2016a)	AE	sonar	22,446	70/15/15	2, 6	IC	CNN		✓		OA, recall, confusion matrix	OA=97.1% (6 object classes)
						OD	CNN with sliding windows				recall	recall=80.8% (binary OD)
Xue et al. (2021b)	marine und.	camera und.[2]	10,000	85/0/15	7	OD	YOLO v3 with ResNet50 backbone[3]	✓	✓		mAP, AP, F1-score, kappa, confusion matrix	mAP_{50}=53.8%
Bajaj et al. (2021)	marine und.	camera und.[2]	2900	85/0/15	3	OD	InceptionResNetV2	✓				
Tian et al. (2021)	marine und.	camera und.	6600	94/6/0	3	OD	Improved YOLO v4[3]				mAP, AP	
Hegde et al. (2021)	marine und.	camera und.[2], camera[2]	10,000	80/20/0	4	OD	SSD MobileNet V2	✓	✓		precision, recall, F1-score	
Marin et al. (2021)	marine und.	camera und.[2]	2395	80/0/20	6	IC	InceptionResNetV2[3]	✓	✓		OA, F1-score, kappa, confusion matrix, macro precision, macro recall, macro F1-score, weighted precision, weighted F1-score	OA=91.4%
Politikos et al. (2021)	marine und.	camera und.	635	80/15/5	11	OD	R-CNN with the MobileNetV1 backbone CNN	✓	✓		mAP, AP	mAP_{50}=62%
Xue et al. (2021a)	marine und.	camera und.[2]	13,914	70/15/15	7	IC	Shuffle-Xception[3]		✓		OA, average accuracy, precision, recall, F1-score, kappa, confusion matrix	
Deng et al. (2021)	marine und.	camera und.[2]	7212		22	OD	Improved Mask R-CNN[3]		✓		mAP	mAP_{50}=65%
						IS						mAP_{50}=60.2%
Musić et al. (2020)	marine und.	camera und.[2]	~2600	60/20/20	5	IC	VGG16[3]	✓	✓		OA	OA=85%
						OD	YOLO v3					
Panwar et al. (2020)	marine und., shores	camera[2]	369	80/0/20	4	OD	RetinaNet with ResNet-101-FPN backbone	✓		N	mAP, AP	mAP_{88}=81.48%
Fulton et al. (2019)	marine und.	camera und.[2]	6540	87/0/13	3	OD	YOLO v2[3]	✓			mAP, AP, Average IoU	mAP=47.9%
Mifdal et al. (2021)	marine sur.	satellite[2]			2	IS	U-Net		✓		pixel accuracy, F1-score, kappa	pixel accuracy=84.28%
Garcia-Garin et al. (2021)	marine sur.	airborne	796	90/0/10	2	IC	CNN		✓		OA, precision, recall, F1-score	OA=81%
de Vries et al. (2021)	marine sur.	camera	100,000		2	OD	YOLO v5[3]	✓				
Kylili et al. (2020)	marine sur., shores	camera[2]	1600	79/20/1	8	IC	VGG16	✓	✓		OA	OA=90%
Battula et al. (2020)	marine sur.	camera[2]	2467		2	OD	Resnet-50			N		
Watanabe et al. (2019)	marine sur., shores	camera, phone	189	80/0/20	4	OD	YOLO v3				mAP	mAP_{50}=77.2%
Kylili et al. (2019)	marine sur.	camera	750	79/20/1	3	IC	VGG16	✓	✓		OA	OA=86%
Kylili et al. (2021)	shores	camera[2]	2000	67/16/17	7	OD	YOLO v5	✓	✓			

(continued on next page)

Table 1 (*continued*)

Reference	Water body	Dataset				Computer vision task		TL	DA	GC	Performance evaluation[(1)]	
		Source	Size (# images)	Split (%)	No. classes	Type	Model architecture				Metric	Performance
						IS	YOLACT++					
Song et al. (2021)	shores	phone	846	70/17/13	7	OD	YOLO v5			D	AP, mAP	
Martin et al. (2021)	shores	airborne	750[(4)]		2, 14	OD	Faster R-CNN			G	precision, recall, F1-score	F1-score=44.2% (binary OD)
Papakonstantinou et al. (2021)	shores	airborne	22,760[(4)]	54/13/33	2	IC	VGG19[(3)]	✓	✓	G	OA, precision, recall, F1-score	OA=77.6%
Wolf et al. (2020)	shores, rivers	airborne	12,918[(4)]	80/0/20	6, 18	IC	CNN	✓	✓		OA, precision, recall, F1-score, confusion matrix	OA=83% (6 class objects), OA=71% (18 class objects)
Gonçalves et al. (2020)	shores	airborne			2	IC	DenseNet			G	precision, recall, F1-score	
Kako et al. (2020)	shores	airborne		64/36/0	2	IS	MLP			G	pixel accuracy	
Fallati et al. (2019)	shores	airborne			2	OD	CNN			G, E	precision, recall, F1-score	recall=67%
Thiagarajan and Satheesh Kumar (2019)	shores	camera	135		2	IC	CNN		✓		OA, precision, recall	
						OD	CNN with sliding windows					
Putra and Prabowo (2021)	rivers	phone		90/10/0	2	OD	YOLO v3 with darknet-53 backbone	✓			AP, mAP	
Lin et al. (2021)	rivers	camera	2400	91/0/9	8	OD	FMA-YOLO v5s[(3)]		✓		AP, mAP	mAP=79.41%
Tharani et al. (2021)	rivers	camera	13,500	93/7[(5)]	3	OD	M2Det(VGG)[(3)]	✓			AP, mAP	mAP=45.8%
						IS	Improved U-Net[(3)]					
van Lieshout et al. (2020)	rivers	camera	1272	85/0/15	2	OD	Faster R-CNN with Inception V2	✓	✓	G, D	recall	recall=68.7%
Jakovljevic et al. (2020)	rivers, lakes	airborne	2608	80/20/0		IS	ResUNet50[(3)]	✓	✓		precision, recall, F1-score	

Acronyms: Transfer learning (TL), Data augmentation (DA), Generalization capability (GC), Artificial Environment (AE), Image classification (IC), Object detection (OD), Image segmentation (IS), Geographical generalization capability (G), Environmental generalization capability (E), Device setup generalization capability (D), Non-aquatic generalization capability (N), overall accuracy (OA), average precision (AP), mean average precision (mAP), Intersection Over Union (IoU).

[(1)] The "Metric" column is not populated when the study does not report common CV metrics. The "Performance" column is populated with the test value of the most representative metric.

[(2)] Part of or all the images in studies are retrieved from public databases or internet.

[(3)] The study features multiple deep learning architectures for computer vision tasks.

[(4)] The authors in studies cut the raw images into image tiles. In the "Dataset size" column, we report the total number of image tiles in datasets.

[(5)] Tharani et al. (2021) split their dataset into 93% for training and validation and 7% for testing.

(e.g., ImageNet dataset (Deng et al., 2009)) or via internet (e.g., Google). Two of these studies (Kylili et al., 2021, 2020) extracted images from the ImageNet dataset. Panwar et al. (2020) retrieved images from the TACO dataset (Proença and Simões, 2020), and directly utilized the images with annotations from the dataset. Battula et al. (2020) extracted images from a Kaggle dataset[1] and labeled images with bounding boxes to train and test OD model. Hegde et al. (2021) retrieved unlabelled images from Google and manually created the annotations to develop and test their models.

Researchers mainly used underwater cameras to collect images under the water surface, e.g., cameras attached to remotely operated vehicles (ROVs) (Wu et al., 2020) or vessels (Politikos et al., 2021). Nine studies used underwater camera images which were partially or completely retrieved from public databases or internet. This drastically reduces the cost of sampling activities in marine underwater environments where sampling requires laborious diving operations, expensive ROVs and/or autonomous underwater vehicles (AUVs) (Valdenegro-Toro, 2016). Five of these studies (Bajaj et al., 2021; Fulton et al., 2019; Marin et al., 2021; Xue et al., 2021b, 2021a) extracted data from the deep-sea debris database[2] provided by Japan Agency for Marine-earth Science and Technology. Four studies (Hegde et al., 2021; Marin et al., 2021; Musić et al., 2020; Wu et al., 2020) retrieved images from internet. While authors of these studies had to manually produce the annotations, one study (Deng et al., 2021) directly utilized the images with annotations from the TrashCan dataset (Hong et al., 2020).

Researchers collected airborne imagery using cameras mounted on UAVs (7 papers) or placed under manned aircrafts (1 paper). UAVs grant versatility since operators can easily customize the flight route and flight height to obtain images at different locations and with different ground sample distances (Fallati et al., 2019). Furthermore, UAVs allows surveying otherwise hard-to-reach locations (Zhang et al., 2017) and eliminate the limitations of fixing sensors on bridges or other infrastructure. However, no-fly zones restrict flying UAVs (e.g., nearby airports). Researchers may also need special training and licenses to

[1] https://www.kaggle.com/asdasdasasdas/garbage-classification

[2] Japan Agency for Marine Earth Science and Technology, "Deep-sea Debris Database", available at http://www.godac.jamstec.go.jp/catalog/dsdebris/e/index.html

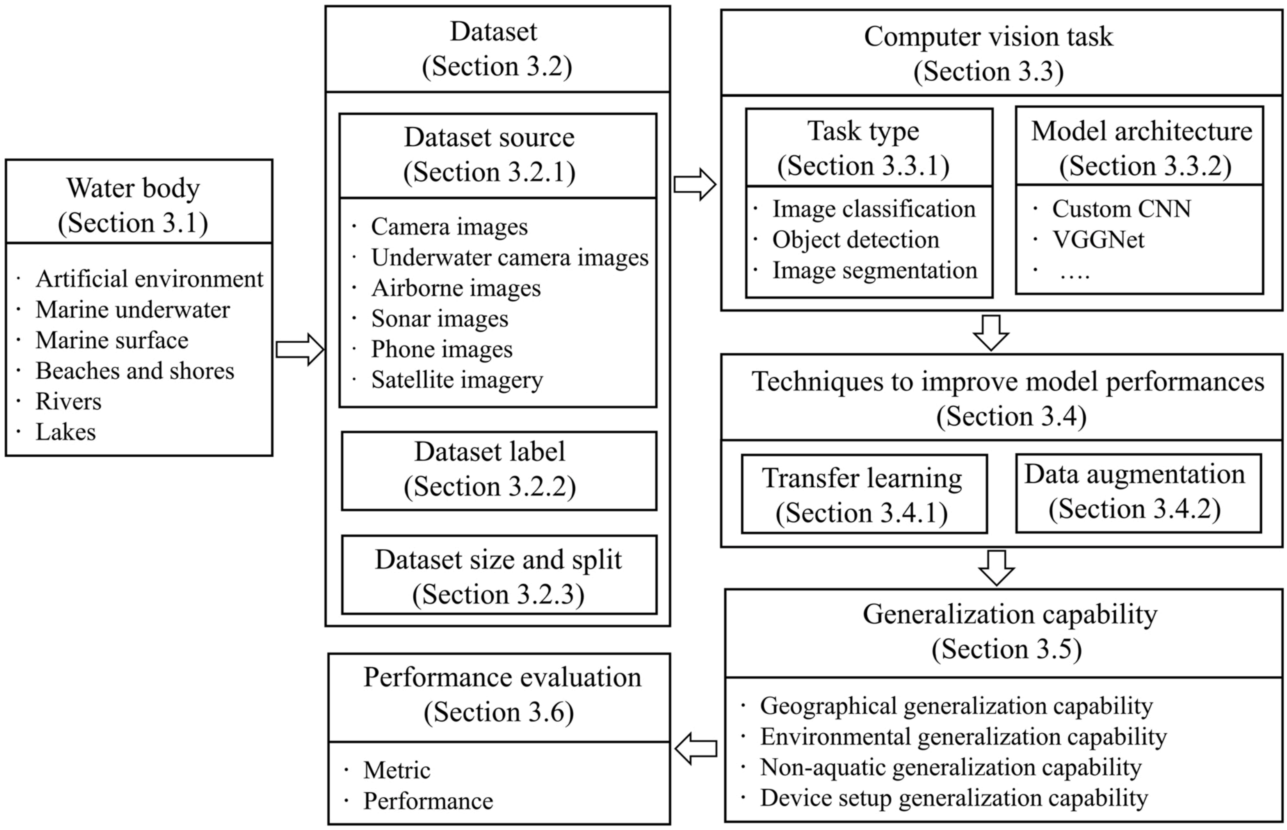

Fig. 1. Factors reviewed in the reviewed literature.

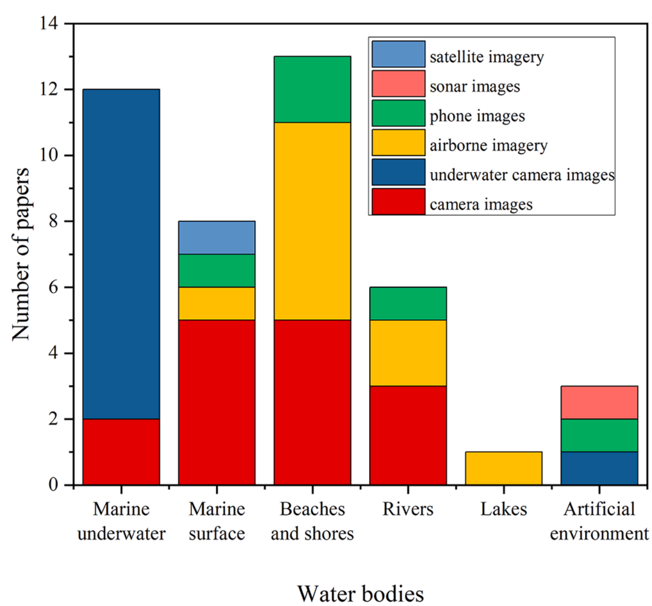

Fig. 2. Distribution of water bodies and dataset sources. Some papers are counted multiple times since they either consider multiple water bodies or multiple dataset sources.

operate UAVs, thus increasing the operational costs and, in some cases/countries, making the use of UAVs difficult, if not impossible. As shown in Fig. 2, UAVs are particularly suitable for field sampling activities along beaches, since these present fewer flight restrictions and obstacles that can potentially interfere with the UAVs flight (e.g., buildings).

Only four studies used mobile phones to collect images. Phones are easily available for citizens, thus could substantially contribute to citizen science initiatives for data collection. Modern smart phones with high-resolution cameras can obtain high-quality images and thus meet the needs of accurate sampling.

While sonar devices are preferred instruments for target and object recognition in underwater environments, e.g., fish classification and fishery assessment (McCann et al., 2018), only one study in the reviewed literature applied sonar devices to collect underwater images (Valdenegro-Toro, 2016). Although the relatively higher sampling costs hinder the development of autonomous detection and classification directly using sonar images (Qin et al., 2021), sonar devices are promising for underwater plastic monitoring as suggested by a recent study (Broere et al., 2021). Sonar sampling can cover a larger area underwater where ROVs or divers cannot safely dive (Neupane and Seok, 2020) because sound waves travel further in water. For these reasons, we encourage further studies to assess their suitability for detecting macroplastic litter under water surface, especially in real-world settings.

One study (Mifdal et al., 2021) retrieved Sentinel 2 imagery on floating marine macroplastic litter from the Google Earth Engine dataset catalog. These data are globally available and free of charge, but do not contain specific annotations for macroplastic litter. After collecting and labeling the satellite imagery, the authors trained DL models to detect floating objects on the sea surface. Compared with other dataset sources, satellite imagery can provide broader geographical coverage that is significant for hotspots monitoring and global environmental monitoring. On the other hand, satellite imagery is not appropriate to detect small and isolated macroplastic litter floating on the vast sea surface, and cannot be used for observing underwater litter (Watanabe et al., 2019).

3.2.2. *Dataset labels*

Authors do not usually categorize macroplastic litter and other types of litter in their datasets using labels that reflect international guidelines and standards. If we consider the categories defined by the Oslo and Paris Conventions (OSPAR) (Wenneker and Oosterbaan, 2010), we identify 12 categories of macroplastic litter (i.e., bags, bottles, nets, caps/lids, industrial packaging/plastic sheeting, cups, buckets, cutlery/trays/straws, shoes/sandals, containers, rope, and floats/buoys), and 5 categories of other litter (i.e., glass, paper/cardboard, rubber, metal, and cloth) across all surveyed datasets. When OSPAR defines multiple sub-categories of one plastic product and the specific sub-category is unclear in reviewed papers, we only report its general category. For example, while "bags" is categorized into 6 sub-categories in OSPAR (e.g., small plastic bags and fertilizer/animal feed bags), we only report "bags" in this paper.

Several studies (13 out of 34 papers) detected plastic in a binary fashion. Among these, only three studies (Garcia-Garin et al., 2021; Kako et al., 2020; van Lieshout et al., 2020) specifically detected the presence of macroplastic litter in images. The remaining studies detected macroplastic litter by including it in a generic "litter" or "trash" or "debris" category. A larger group of studies detected more than 2 classes (22 papers). Among these, one study (Kylili et al., 2019) detected different types of plastic products. Nine studies provided a refined categorization for other types of litter. For example, Panwar et al. (2020) categorized the objects into glass, metal, paper, and plastic. One study (Tharani et al., 2021) detected macroplastic litter by considering three different sizes, included in three generic categories (i.e., small trash, medium trash and large trash). The remaining studies (11 papers) detected different plastic products as well as other object categories. For instance, Watanabe et al. (2019) classified the objects as plastic bottles, plastic bags, drift wood, and other debris. Gathering a balanced dataset with accurate labels becomes challenging as the number of classes increases. We identify 9 studies (Marin et al., 2021; Martin et al., 2021; Musić et al., 2020; Politikos et al., 2021; Tharani et al., 2021; Thiagarajan and Satheesh Kumar, 2019; Tian et al., 2021; Wolf et al., 2020; Xue et al., 2021b) working with unbalanced datasets, featuring classes with very scarce data (e.g., shoes, plastic cups, string and cord). Depending on the sensor used and its resolution, small objects (e.g., straws, toothpicks, and cotton buds) may be far less visible than others (Tharani et al., 2021). To improve detection of rare items or small items, we need to collect more data at higher resolutions (Wolf et al., 2020).

3.2.3. Dataset size and split

The "Dataset size" column of Table 1 reports the size of dataset used in the reviewed papers, not including the data generated via DA. For one study (Wu et al., 2020), only the dataset size including the data generated with DA could be reported. Fig. 3 shows the distribution of dataset sizes (the number of images) per dataset source, as reported in 29 of the reviewed papers (see Table 1). The dataset size of phone images is small because two other datasets containing phone images and another kind of dataset source (Watanabe et al., 2019; Wu et al., 2020) are featured in the "multiple" category. The size of another dataset containing phone images (Putra and Prabowo, 2021) is unclear, thus it was not reported in Fig. 3. One dataset containing 100,000 images (de Vries et al., 2021) is much larger than all the others. These time-lapse images were collected at intervals between 2 s and 10 s during The Ocean Cleanup's North Pacific Mission 3 research expedition. The average dataset length is of around 9000 images. According to Arya et al. (2020), IC generally requires more than 5000 labeled images for each class to train a model with acceptable performances. For binary detection problems, this entails that at least 10,000 images are needed to develop a sufficiently robust detection model. In the reviewed literature, 11 studies conducted IC tasks (see Table 1). Apart from two studies (Gonçalves et al., 2020; Thiagarajan and Satheesh Kumar, 2019), all other 9 studies reported both the specific dataset size and the number of classes in datasets. According to the suggestions of Arya et al. (2020), with the exception of (Papakonstantinou et al., 2021; Valdenegro-Toro, 2016), these studies did not collect sufficient raw data for model training and validation

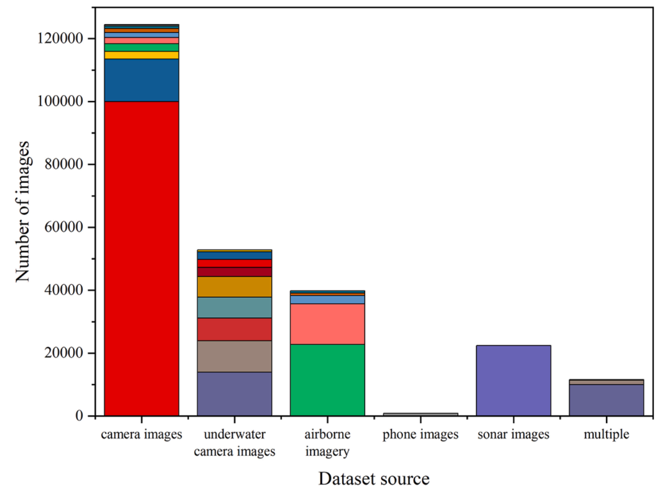

Fig. 3. Distribution of dataset size (the number of images) per dataset source identified in 29 reviewed papers. Each different block identifies a different dataset. The label "multiple" identifies datasets obtained from multiple data-set sources.

considering the number of classes. Therefore, all studies lacking sufficient data adopted TL and/or DA to improve the performances (see Section 3.4). Similar considerations may be drawn also for studies presenting OD and IS applications.

Multiple researchers (e.g., Martin et al. (2021) and van Lieshout et al. (2020)) stressed the importance of a large-scale dataset for DL-based detection of macroplastic litter. Furthermore, three studies (Kylili et al., 2019; Musić et al., 2020; van Lieshout et al., 2020) showed that the increase of training dataset size leads to superior detection performance. For example, van Lieshout et al. (2020) developed a DL model to detect floating macroplastic litter in rivers across Jakarta, Indonesia using a binary classification approach. The precision (i.e., the proportion of objects correctly identified as macroplastic litter with respect to total detections) raised from 49.4% to 59.4% when increasing the number of labels in the training dataset from about 2000 to 10,000. This study also showed that increasing dataset size further (from 10,000 to 24,000) resulted in smaller improvements (from 59.4%). Indeed, after training with sufficient data to learn basic representations, performance for CV tasks tend to grow logarithmically with dataset size (Sun et al., 2017). Therefore, we suggest gathering and labeling training data with respect to the level of performance required to address the specific challenge.

The "Dataset split" column of Table 1 reports the train/validation/test splits of the dataset used. Among 26 papers reporting dataset split, 10 reported the use of both a validation and a test dataset. The validation dataset is commonly used to select the best model by monitoring overfitting during training; on the other hand, the test dataset is employed to assess the generalization capability of the model for "unseen" data. It is not clear whether the remaining studies used part of the training data for validation and model selection, or if they used the test dataset for that purpose. Similarly, some papers report the use of a validation dataset, but not that of a test dataset. In general, we recommend to split the dataset into training (∼80%), validation (∼10%) and test (∼10%) datasets to facilitate robust model selection and unbiased estimation of the generalization error on unseen data.

3.3. Computer vision tasks for macroplastic litter detection

3.3.1. CV task types

General CV tasks are image classification (IC), object detection (OD) and image segmentation (IS) (Chai et al., 2021). Fig. 4 shows an example of DL model architecture, the typical labeling procedure used, and the output of different CV tasks. All reviewed studies resorted to supervised

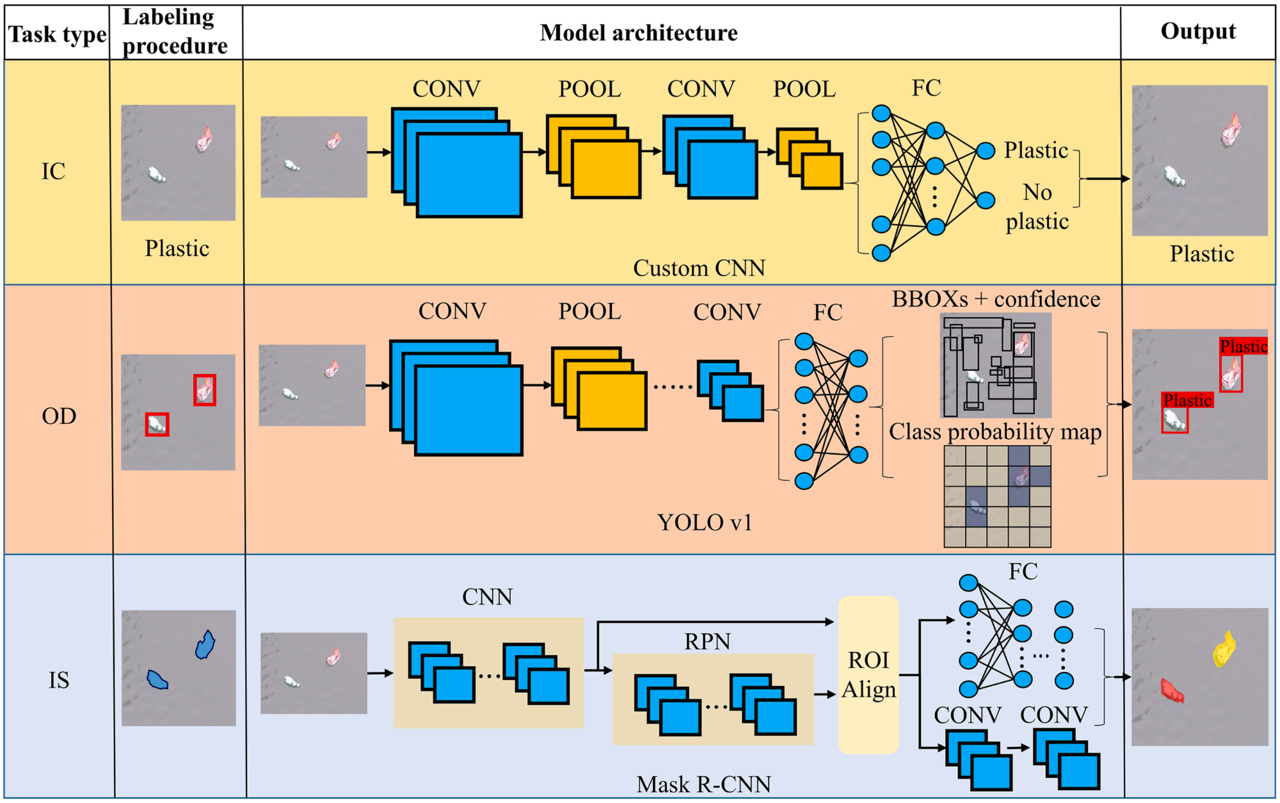

Fig. 4. Labeling procedure, selected typical model architecture and output of different computer vision tasks. The "IC" row shows an example of binary classification, while the "IS" row shows an example of instance segmentation. Acronyms used: Convolutional layer (CONV), Pooling layer (POOL), Fully connected layer (FC), Bounding boxes (BBOXs), Convolutional neural network (CNN), Region Proposal Network (RPN), Image classification (IC), Object detection (OD), Image segmentation (IS).

learning for developing the detection models. In supervised learning, the model is trained to perform its task from examples of paired input/output data, where the output data is carefully labeled, or annotated, by humans.

IC is the process of classifying the entire image into one category (single-label classification) or multiple categories (multi-label classification) (Wei et al., 2016). The labeling procedure of IC includes annotating a given image with one class label or multiple class labels (see Fig. 4, top panel). On the other hand, OD algorithms automatically identify the class and location of different objects in images. The labeling task of OD requires the annotation of objects with class labels and bounding boxes (see Fig. 4, middle panel). Consequently, the output of OD models are bounding boxes and class labels for each detected instance. IS divides an image into multiple segments with similar characteristics, enabling a pixel-by-pixel identification of objects of interest. The labeling task requires assigning corresponding labels and pixel-wise masks to target objects (see Fig. 4, bottom panel) (Chai et al., 2021). We identify two types of IS among reviewed papers: semantic segmentation (Jakovljevic et al., 2020; Kako et al., 2020; Mifdal et al., 2021; Tharani et al., 2021) and instance segmentation (Deng et al., 2021; Kylili et al., 2021). Semantic segmentation assigns category labels to each pixel in images, while instance segmentation assigns category labels and instance identities to each object pixel (Chai et al., 2021). Thus, semantic segmentation is more suitable to quantify the area occupied by macroplastic litter, while instance segmentation is more appropriate to discriminate different macroplastic items.

Table 1 shows that researchers prefer OD methods to detect macroplastic litter in aquatic environments (23 out of 34 papers). OD can concurrently identify the type and location of objects in images, thus estimating the number of macroplastic items in an image (van Lieshout et al., 2020). IC (11 papers) is also popular since it is simpler to implement, especially by deploying one of the many successful architectures already available from the CV literature. Only 6 studies resorted to IS, arguably because of the substantial amount of time required to properly label the datasets (Jabari et al., 2021). Referring to IC models, all reviewed papers employed single-label algorithms, i.e., binary classifiers and multi-class classifiers. These methods can only process images containing one type of object at a time. On the other hand, multi-label classifiers can identify multiple categories of objects in one image (e.g., macroplastic litter, metal, and rubber) (Chai et al., 2021). Although these classifiers can better capture the diversity of litter in natural environments, no reviewed paper resorted to multi-label IC.

Although most studies (24 out of 34 papers) conducted CV tasks only for detection purposes, 10 studies also attempted the quantification of macroplastic litter. Of these, 9 papers quantified the number of macroplastic items via OD (de Vries et al., 2021; Martin et al., 2021; Song et al., 2021; van Lieshout et al., 2020), IC (Garcia-Garin et al., 2021; Gonçalves et al., 2020; Papakonstantinou et al., 2021; Wolf et al., 2020) or IS (Kylili et al., 2021). For example, Gonçalves et al. (2020) cut one original image into small portions, and performed IC to classify each portion into "litter" or "no litter". The number of macroplastic items in one image was then calculated by the sum of the number of "litter" portions. However, if there are portions containing more than one item, performing IC tasks will lead to the deviation between the predicted results and the ground truth. Some studies also post-processed the model results to compute spatial litter concentrations (in items/m^2 or items/km^2, 5 papers), fluxes (in items/min/m, 1 paper), mass concentrations (in g/m^2, 1 paper) or mass (in kg, 1 paper). For instance, Martin et al. (2021) computed the concentrations of plastic bottles in beaches by averaging the number of correctly detected bottles over the tested area, and computed its mass concentrations by multiplying concentrations with respect to the median weights of bottles retrieved from Martin et al. (2019). Kylili et al. (2021) computed the total mass of litter in beaches by tallying the known mass of all macroplastic items predicted by the DL model. van Lieshout et al. (2020) computed the macroplastic fluxes floating in rivers by dividing the number of items detected per time unit by the river width. Kako et al. (2020) computed macroplastic volumes via IS. They first detected the edges of macroplastic litter on images, which were then superimposed on a digital surface model containing location and altitude data over a beach. This allowed the litter volume to be computed from the heights and base area surrounded by the edges.

3.3.2. Model architectures for each CV task

Most reviewed publications (33 out of 34 papers) used Convolutional Neural Networks (CNNs) based architectures, such as YOLO networks (Xue et al., 2021b) and VGG networks (Kylili et al., 2020). Only one study (Kako et al., 2020) used a more conventional, three-layered Multilayer Perceptron (MLP) neural network. Compared with MLP, CNN can take advantage of the spatial patterns implicit in raw images. Besides, the properties of CNN (i.e., local connections and weight sharing) enable it to learn representations with fewer trainable parameters than MLP (Ming and Xiaolin 2015).These characteristics have allowed CNN to outperform MLP (Zhang et al., 2018) and to be more widely used for CV. However, none of the reviewed papers featured current state-of-the-art architectures such as Vision Transformers, e.g., Swin Transformer (Liu et al., 2022a) and ConvNeXts (Zhuang Liu et al., 2022b).

In IC tasks, four studies employed the VGGNet architecture, probably because this architecture was proposed in 2014, and has been applied since then successfully in many fields (Ajit et al., 2020). Custom CNN architectures (4 papers) are also popular, mainly to develop parsimonious models with limited parameters that better match data availability and largely reduce computational efforts. For example, one study (Valdenegro-Toro, 2016) employed a custom 4-layered CNN with 930,000 parameters, much less than the 143.47 million parameters of a deeper VGG model with 19 layers (Marin et al., 2021).

CNN-based OD algorithms are divided into two-stage algorithms and one-stage algorithms. In two-stage algorithms, the first stage generates a set of bounding box proposals that are classified and detected in the second stage (Chai et al., 2021). On the other hand, one-stage algorithms perform classification and bounding box prediction concurrently in a single forward pass of the network. YOLO networks (11 papers) are the most frequently used OD architectures among the reviewed papers. YOLO networks are popular one-stage architectures thanks to their fast processing speed, which can reach the standards required for real-time video processing (Redmon et al., 2016). Although YOLO networks are faster than other architectures, its accuracy may be lower than that of some two-stage OD algorithms such as Faster R-CNN, which has been used in two reviewed papers.

U-Net (3 papers) is the most frequently employed IS architecture (Huang et al., 2020). One study (Jakovljevic et al., 2020) employed ResUNet50, which is based on a hybrid between the popular ResNet (He et al., 2016) and U-Net architectures. Building blocks of ResNet pre-trained on the ImageNet dataset are added to the U-Net.

Two studies deployed DL models, Resnet-50 neural network (Battula et al., 2020) and SSD MobileNet V2 (Hegde et al., 2021) to perform OD in edge computing devices, e.g., processing boards connected to fixed cameras or installed in ROVs. For example, Hegde et al. (2021) stored a trained detection model in a Raspberry Pi board. The device used the model to detect macroplastic litter from the surrounding environment as sampled by the attached underwater camera.

Model complexity plays an important role for DL models that will eventually run in real-time or on edge computing devices. Researchers should thus further investigate the suitability of small architectures with good classification performances, such as MobileNetV2 (~2.4 million parameters) (Dong et al., 2020), and SqueezeNetV1 (~1.2 million parameters) (Gholami et al., 2018). These "light" architectures can be easily transferred to edge devices and play a significant role in tackling macroplastic pollution in water bodies. Pruning algorithms can successfully reduce model complexity and enable edge computing. For instance, Tian et al. (2021) proposed a pruned YOLO v4 capable of accurate OD for underwater camera images with only 7% of the original parameters.

3.4. Techniques to improve DL model performances

3.4.1. Transfer learning

Transfer learning (TL) involves the transfer of prior knowledge from a related task to a new task (Pan and Yang, 2010). When applied to DL models, TL entails reusing parts of a model pre-trained on very large datasets (e.g., ImageNet dataset) using computing clusters. This operation improves learning of the new task by: (1) providing a better starting point for training and preventing the model from falling into local minima (Fulton et al., 2019); (2) limiting the number of parameters to be optimized to a subset of the layers of the network; and (3) reducing data-labeling efforts by reducing the amount of training data needed to reach satisfactory performances on the new task.

Most of the reviewed papers (19 out of 34 papers) adopted TL, regardless of the CV task performed. For example, Musić et al. (2020) performed IC task to detect five categories of litter, i.e., plastic, glass, metal, paper and cardboard. They pre-trained the VGG16 on ImageNet dataset and fine-tuned the final layers of the VGG16 on a new dataset containing images from these five categories of litter. Although the objects in ImageNet are quite different from the detected litter, the pre-trained model improves detection because it recognizes generic features (e.g., edges, and basic shapes) in its early layers.

Researchers usually used models pre-trained on the ImageNet dataset or the Common Objects in Context (COCO) dataset (Lin et al., 2014). ImageNet is preferred for IC (5 papers), but also applied for OD (2 papers) and IS (2 papers). The COCO dataset is a common choice for OD (6 papers). The CIFAR-10 dataset (Recht et al., 2018) and the PASCAL VOC dataset (Everingham et al., 2010) have also been used.

While several authors resorted to TL to develop their models, with the exception of (Marin et al., 2021), no studies have thoroughly assessed its benefits with respect to training from scratch or fine-tuning the entire architecture (not just the classifier). Such investigations can be justified by the reported good performances of small architectures such as MobileNetV2 and SqueezeNetV1. Furthermore, the representation learned on large open source datasets may not always reflect typical features of images with macroplastic litter (e.g., variety of litter, presence of water in the background).

3.4.2. Data augmentation

Data augmentation (DA) reduces model overfitting by increasing the amount of available training data via augmentation or transformation of the images in the original training dataset (Shorten and Khoshgoftaar, 2019). This technique can also improve the performances of models when dealing with imbalanced datasets by creating more samples of underrepresented classes.

DA usually involves automatic procedures performing geometrical transformations as well as color space transformations on available images. The DA methods used in reviewed papers include flipping (11 papers), rotation (10 papers), zooming in/out (4 papers), shifting (4 papers), noise addition (3 papers), cropping (2 papers), shearing (2 papers), copy-paste augmentation (2 papers), changes in brightness (1 paper), and mosaic data augmentation (1 paper). Fig. 5 shows an example of several DA techniques. Copy-paste augmentation is an advanced DA technique, whose purpose is to copy objects from a source image and paste them to a target image (Ghiasi et al., 2021). For example, Lin et al. (2021) employed such technique to superimpose labeled target objects cropped from real-world images against realistic backgrounds. Mosaic data augmentation combines 4 cropped images to create a synthetic image, that is often used for data augmentation in OD

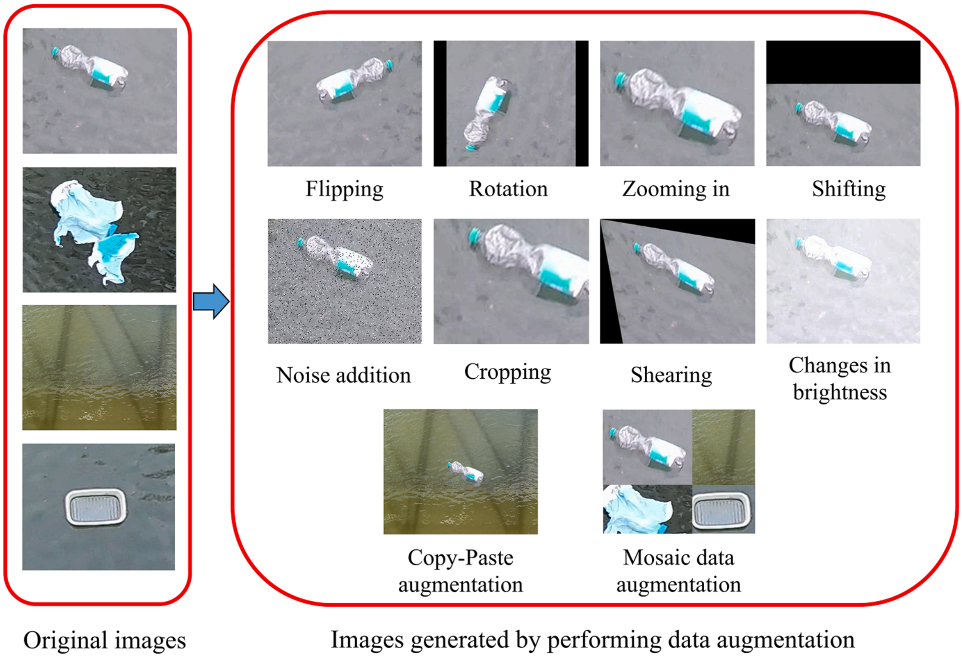

Fig. 5. Examples of data augmentation techniques used in reviewed papers to improve model performances. Left: original images; Right: images generated by performing geometrical transformations (e.g., flipping, rotation, zooming in, shifting, cropping, and shearing), and other basic (e.g., changes in brightness, and noise addition), or advanced transformations (e.g., copy-paste augmentation, and mosaic data augmentation).

tasks (Lin et al., 2021).

Flipping is the most popular choice as it preserves the original features of macroplastic litter in the images and maintains fidelity with respect to the original label. On the other hand, the addition of noise or changes in brightness may alter the original images too much, thus degrading model performances. Rotation, zooming in, shifting, cropping, shearing, and mosaic data augmentation may instead lead to the omission of some of the originals objects of interest in the new images, forcing relabeling and partially nullifying the benefits of DA (Shorten and Khoshgoftaar, 2019).

While most studies (20 out of 34 papers) applied DA (see Table 1), only three studies have thoroughly evaluated the benefits of DA with respect to training the same architecture on the original dataset. van Lieshout et al. (2020) showed that model precision marginally raised from 59.4% to 63.4% when using flipping data augmentation methods. Lin et al. (2021) also showed the model performances increased slightly when employing mosaic data augmentation. Musić et al. (2020) used copy-paste augmentation by superimposing computer-generated macroplastic litter on realistic backgrounds. However, adding these images to the training dataset resulted in poorer prediction performances on the real-world dataset. Thus, researchers should discuss the benefits of different DA method for macroplastic litter detection models in more depth.

3.5. Generalization capability

DL models for CV exploit spatial inductive biases and shared weights to recognize features and objects regardless of their position in the image (Battaglia et al., 2018). While this favors generalization to unseen data, good detection performances at a single location or for similar environmental conditions do not guarantee that the model can be successfully applied or "transferred" to other situations and case studies. Achieving satisfactory out-of-domain generalization capability is a prerequisite for deploying large scale monitoring strategies based on DL, especially with respect to transferability across different bodies of water, locations, and device setups.

We identify four different forms of out-of-domain generalization capability in the reviewed papers: (1) geographical generalization capability, (2) environmental generalization capability, (3) non-aquatic generalization capability, and (4) device setup generalization capability. Geographical generalization capability represents the generalization capability of the model at different locations under roughly the same environmental conditions (such as weather, presence of waves, wind conditions, and terrain shading). Environmental generalization capability refers model testing in different environmental conditions. Non-aquatic generalization capability involves models trained with data from non-aquatic environments and tested on aquatic environments (or vice versa). Lastly, device setup generalization capability represents the generalization capability for different device setups, such as the flight altitude of UAVs, or the setting angle between a fixed camera and the water surface.

Despite the importance of generalization, only few studies (9 out of 34 papers) directly addressed these aspects, with two studies (Fallati et al., 2019; van Lieshout et al., 2020) considering two different forms of generalization capability (see Table 1). The majority of these 9 studies are with respect to geographical generalization capability (6 papers). Five papers (Fallati et al., 2019; Kako et al., 2020; Martin et al., 2021; Papakonstantinou et al., 2021; van Lieshout et al., 2020) studied geographical generalization capability by training and testing on different case studies, respectively. For example, Papakonstantinou et al. (2021) trained DL models on UAV images captured from certain beaches, and tested it on UAV images collected from different beaches. Compared with geographical generalization capability, there are less studies concerning non-aquatic generalization capability (2 papers), device setup generalization capability (2 papers), and environmental generalization capability (1 paper). For instance, Panwar et al. (2020) trained a model on images of macroplastic litter gathered across streets and forests, and tested it on images with macroplastic litter under the sea surface and on beaches. Song et al. (2021) assessed device setup generalization capability by using a phone mounted on a tripod to collect training and test data from different heights at one beach. Fallati et al. (2019) evaluated environmental generalization capability by collecting training and test data at different time of the day. The authors also used a UAV to collect training and test data at different beaches to assess the geographical generalization capability of the model.

Among the 9 papers addressing generalization capability, 4 papers (Battula et al., 2020; Kako et al., 2020; Panwar et al., 2020; Song et al., 2021) did not discuss the performances of DL models trained and tested in different conditions. Only 1 paper (Papakonstantinou et al., 2021) reported promising geographical generalization capability, with a precision metric of 83%. The models in the remaining studies did not show satisfactory generalization performances when tested for different geographical, environmental, or device setup conditions with reported precision between 20% and 63.8%. For example, van Lieshout et al. (2020) showed that the performances of a trained model working reasonably well for one location deteriorated quickly for an unseen location, with a decrease in precision from 68.7% to 54%. These new images featured substantially more organic material (e.g., leaves and branches) than those used for training. The presence of organic material, unaccounted for during training, thus hindered robust detection of floating macroplastic litter. The authors also showed that the generalization performances increased when including images from different locations in the training dataset. In general, we believe the community should increase efforts to develop DL models with robust generalization that can operate well across different conditions.

3.6. Performance evaluation

The "Metric" column of Table 1 reports the performance metrics used by the authors when these reflect common options used for CV (Padilla et al., 2020; Wambugu et al., 2021) and are unambiguous.

For IC tasks, the majority of studies used the overall accuracy (OA) metric (9 out of 11 papers) to evaluate performances over all classes. Precision (6 papers), recall (7 papers), and F1-score (6 papers) were the most popular choices to evaluate performances for each class. These metrics should be preferred for imbalanced datasets since OA misrepresents the minority classes. For example, Wolf et al. (2020) worked on an imbalanced dataset including 18 categories of objects. Although good average performance were reported for all classes (OA=71%), minority classes such as carton (25 images in total) were poorly detected (F1-score=0.46).

For binary OD, common metrics include recall (4 out of 8 papers), precision (2 papers), and F1-score (2 papers). For multi-class OD, the majority of studies employed average precision (AP, 8 out of 17 papers) and mean average precision (mAP, 11 papers) to assess performances for each class object and over all classes, respectively. The value of these metrics depends largely on the selected threshold for determining the Intersection Over Union (IoU), a number that quantifies the degree of overlap between the predicted and ground-truth bounding boxes. With some exceptions (Deng et al., 2021; Panwar et al., 2020; Politikos et al., 2021; Putra and Prabowo, 2021; Song et al., 2021; Watanabe et al., 2019; Xue et al., 2021b), these important thresholds are rarely reported in reviewed papers. Based on common benchmarks (e.g., COCO and PASCAL VOC), we recommend using a threshold IoU=0.5 when estimating fluxes (e.g., number of items across the river width per unit of time), while higher thresholds (e.g., up to 0.95) should be used to quantify mass concentrations (e.g., hotspot areas).

For binary semantic segmentation tasks, two studies (Kako et al., 2020; Mifdal et al., 2021) used pixel accuracy metrics to assess performances on detecting macroplastic litter. For multi-class semantic segmentation tasks, one paper (Jakovljevic et al., 2020) used precision, recall, and F1-score metrics to evaluate performances for each class. No papers reported results in terms of IoU or mean IoU, which are the preferred metrics for semantic segmentation as they account for unbalanced datasets. For multi-class instance segmentation, one paper (Deng et al., 2021) employed mAP to evaluate performances over all classes.

The "Performance" column of Table 1 reports the test value of the most representative metric across all classes. However, since the proposed methodologies have been tested on different macroplastic datasets in disparate experimental settings, a direct comparison is unfeasible. More interestingly, some papers report encouraging evidence on the effectiveness of DL methods with respect to accurate, but time-consuming, sampling methods. For instance, de Vries et al. (2021) found a satisfactory correlation (R^2=0.7) between DL-detected spatial concentrations of macroplastics on the sea surface and manta-trawling ground truth observations. Song et al. (2021) reported a small error (<5%) between the number of litter items on a beach yielded by actual counting and those detected by Yolo v5. Kako et al. (2020) reported similar figures (<5%) for the volumetric difference of beached plastic debris between surveys and MLP-based IS. These results suggest that using DL for automatic detection and quantification of macroplastic litter is a valid alternative to traditional sampling methodologies.

4. Knowledge gaps and future directions

Our review shows that the majority of reviewed papers focus on detecting macroplastic litter in marine environments, while less attention is devoted to detecting freshwater macroplastic litter. Recent research indicated that most plastic debris leaking into the environment does not reach the oceans, but instead accumulates in river systems (Tramoy et al., 2020; van Emmerik et al., 2022a; Weideman et al., 2020), resulting in damaged ecosystems (Blettler et al., 2018). Monitoring the source, transport, and sink points of riverine macroplastic litter is thus essential to quantify global macroplastic pollution transport and effectively reduce pollution (van Emmerik and Schwarz, 2020). Therefore, we advocate for greater efforts on applying DL to tackle riverine macroplastic pollution problems in the future.

Based on the findings reported in Section 3, we identify three major knowledge gaps regardless of the body of water:

(1) There is a lack of DL-based detection models with robust generalization performances. This includes models that can detect macroplastic litter for a certain water body for different geographical/environmental/device setup conditions as well as models that can generalize across different case studies (e.g., waterway networks within a country).
(2) The current literature mainly focuses on quantifying the number of macroplastic items. There is a lack of DL-based methods for the quantification of macroplastic (mass) fluxes and hotspots. Accurate quantification is essential to estimate pollution impacts, devise targeted cleaning interventions, and evaluate the success of mitigation efforts.
(3) Stemming from the first two gaps, we emphasize the lack of research to inform the design of structural monitoring strategies exploiting robust DL-based quantification.

Based on the above knowledge gaps, we propose three future research directions to bridge the corresponding gaps. These include (i) the development of a general DL model to detect macroplastic litter; (ii) future DL-based applications for the quantification of macroplastic (mass) fluxes and hotspots; and (iii) the development of DL-based monitoring strategies.

While these knowledge gaps and research directions concern all bodies of water, we underline their importance with respect to riverine ecosystems. This consideration reflects their major role in transport and storage of macroplastic litter (van Emmerik et al., 2022a), as well as feasibility with respect to extensive monitoring of marine ecosystems.

4.1. Development of a general macroplastic litter detection model

Section 3.5 shows that, despite some promising initial efforts, research on the generalization capability of DL models for detecting macroplastic litter is insufficient. Most papers focus on detecting macroplastic litter on a specific case study, but lack an in-depth analysis on generalization. Besides, review results show that the models proposed so far do not retain satisfactory performances under different geographical, environmental, or device setup conditions. Researchers should thus increase efforts to bridge this gap, possibly by exploiting new trends in ML such as *data-centric AI* and *semi-supervised machine learning*, as suggested in the following paragraphs.

4.1.1. Data-centric artificial intelligence

Regardless of the CV task, our review shows that several known model architectures perform reasonably well for detecting macroplastic litter in individual case studies. Further improvements can be obtained by resorting to the latest state-of-the-art models, such as Vision Transformers (Paul and Chen, 2022) or ConvNeXts (Liu et al., 2022b). However, we suggest that achieving higher generalization performances may require shifting the focus from model architectures to data. This is the core idea behind the emerging field of data-centric artificial intelligence (AI), which aims to improve model performances by training on cleaner and more informative datasets (Motamedi et al., 2021). Several studies have shown the benefits of employing these approaches for a wide variety of CV-related industrial applications (Im et al., 2021; Tang et al., 2021;Zhou et al., 2020b). These approaches usually entail improving the quality of existing data by resorting to pre-processing techniques, systematic labeling, and expert knowledge. For instance, sun glints on the surface of rivers can lead to the misclassification of floating objects (Jakovljevic et al., 2020). Some pre-processing techniques, such as Contrast Limited Adaptive Histogram Equalization, can dilute the effects of these unwanted reflections and boost model performances, as shown already for applications in defect detection and eye tracking (Im et al., 2021; Singvi et al., 2012).

Enforcing consistency in the labeling procedure can results in similar improvements (Jain et al., 2021). Manual labeling may introduce significant human error and bias in data, which in turn may severely undermine model performances. Clear guidelines (with illustrative examples) and cross-checking between multiple labelers strengthen the consistency of labeled data towards achieving reliable performances (Lavitas et al., 2021). Similar to what discussed in Section 3.2.2, labels for DL-based litter detection should reflect OSPAR categories (Wenneker and Oosterbaan, 2010) or plastic categories used in CrowdWater (van Emmerik et al., 2020) to facilitate their usage and allow for joining multiple sources to train models with better generalization performances.

The data-centric approach also favors the collection of additional data from in-situ experiments whenever possible. In particular, we suggest gathering more training images at various sampling locations under different environmental conditions (van Lieshout et al., 2020), and extending the collection to different devices and instrumental settings for extensive monitoring applications (e.g., river networks, nation-wide initiatives). We also recommend collecting more data about rare and small items (e.g., straws, cups, shoes, strings, and cords). The choice of data gathering devices can refer to their characteristics reported in Section 3.2.1, the monitoring aims and available resources. The researchers should prevent privacy violation when using phones, cameras, or UAVs to sample in a public location (e.g., urban waterways). Researchers can resort some privacy-preserving approaches in human and human activity recognition, e.g., image style transformation, and differential privacy (Jung, 2020). Alternatively, researchers should resort to DA techniques to increase the number of the images collected in the field. Review results show that most authors employ traditional augmentation techniques (e.g., flipping transformation), with few studies assessing the benefits of advanced augmentation techniques such as copy-paste augmentation described in Section 3.4.2. Contrary to some traditional techniques, this augmentation procedure does not change the original features of target macroplastic litter or omits objects in the newly generated images. While Musić et al. (2020) did not report improved performances, studies form other fields suggest substantial benefits for different CV tasks (Dwibedi et al., 2017; Ghiasi et al., 2021; Xu et al., 2021). By following the data-centric approach, researchers can develop a large open dataset with carefully labeled images that can be used for model development, pre-training and fair benchmarking.

4.1.2. Semi-supervised learning methods

All reviewed studies employ supervised learning methods to detect macroplastic litter. To achieve good generalization performances, supervised learning requires large amount of labeled data obtained with substantial efforts, professional knowledge and skills. Recently, the AI researchers are increasing efforts to develop alternative methods based on semi-supervised learning that greatly reduce the need of costly supervision (Misra and van der Maaten, 2020). This approach entails a preliminary step based on self-supervised learning, e.g., the process of learning meaningful representations from images via pretext tasks that do not require annotations. After pre-training, the model is fine-tuned for prediction on a specific downstream task, which requires a limited amount of labelled data.

Fig. 6 shows an example of semi-supervised learning for macroplastic litter detection adapted from Noroozi and Favaro (2016). Firstly, we need to carefully label a few images and keep the remaining unlabeled. These images are chosen to maximize the informative content of the labeled dataset (e.g., images with good representations of different macroplastic litter, different sampling locations and different environmental conditions). Next, a CNN is trained on a large number of unlabeled images by solving the jigsaw puzzle problem as a pretext task. In the jigsaw task, we randomly crop a window (red dashed box in Fig. 6) from an image, divide it into a 3×3 grid and randomly select a tile (yellow dashed box) in each cell. These tiles numbered from 1 to 9 are reordered by a randomly selected permutation (e.g., 9, 8, 7, 6, 1, 2, 3, 4, 5) from a predefined permutation set. The training dataset for the pretext task is generated by retaining a subset of all potential permutations (e.g., 100) for each image. The pretext task entails reconstructing the original image from its permutations. After training, the representations of unlabeled data learned by the CNN are transferred to the downstream task (i.e., macroplastic litter detection) by fine-tuning the CNN via supervised learning on a limited number of carefully labeled images. Semi-supervised learning can be a better alternative to supervised learning methods because of the lower cost of annotating data and the competitive model performances.

4.2. Quantification of riverine macroplastic (mass) fluxes and hotspots

Our review shows that the current literature mainly focuses on detecting macroplastic litter in water, and only few studies link DL-based detection to the quantification of macroplastic litter, mainly with respect to the number of macroplastic items. Only one study (van Lieshout et al., 2020) quantifies the floating macroplastic fluxes and no studies quantifies the extension of floating macroplastic hotspots, although stakeholders require this information to design cleaning campaigns, and mitigate the impact of pollution on the environment and human health (Tasseron et al., 2020; van Emmerik et al., 2018). Considering the significance of riverine macroplastic pollution, future studies should focus on the development of new methods for quantifying macroplastic (mass) fluxes and hotspots in riverine environments.

(1) Macroplastic (mass) fluxes

Macroplastic fluxes and mass fluxes can be expressed as the number and the mass of macroplastic items across the river width per unit of time, respectively (van Emmerik et al., 2018). To quantify them, we recommend gathering images by installing fixed cameras on hydraulic infrastructures (e.g., bridges) at various locations along the river (or river network), and train the DL model on these images. Performing OD tasks can precisely quantify the number of items by detecting each item in images with a bounding box (see Fig. 4). The latest versions of YOLO networks (e.g., YOLO v7 (Wang et al., 2022)) are the most promising architectures to detect litter fast and accurately. Then, a limited number

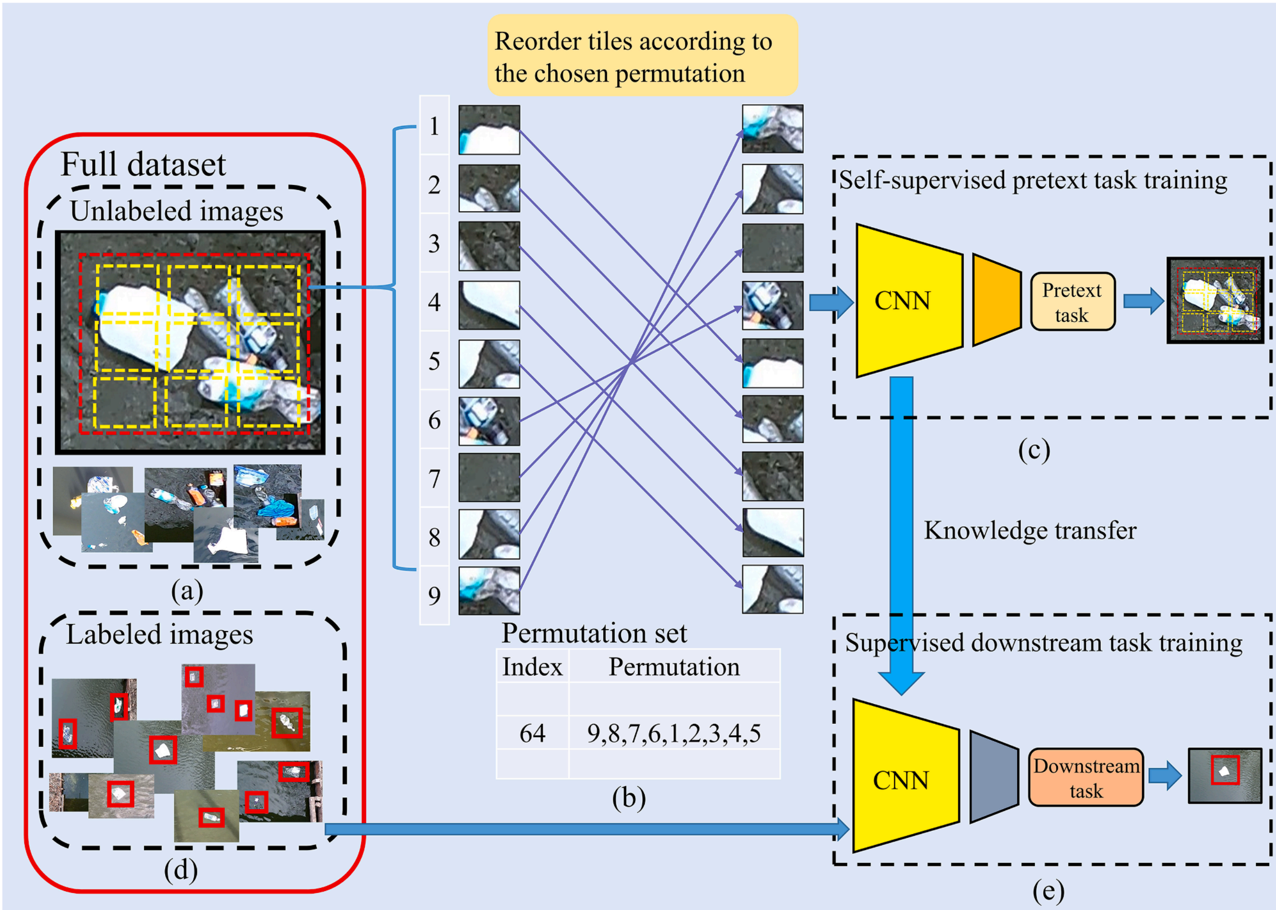

Fig. 6. The schematic illustration of a semi-supervised learning method adapted from Noroozi and Favaro (2016). (a) Unlabeled images are subdivided in 9 tiles (yellow dashed box), which are extracted and shuffled randomly to create different permutations (b); the model (e.g., CNN) is pre-trained by solving the jigsaw puzzle problem of reordering the tiles (pretext task) (c); a limited amount of labeled images (d) is used to fine-tune the DL model to learn a specific plastic detection downstream task (e).

of experiments must be conducted by (i) sampling litter using nets, (ii) counting the number of samples, and (iii) weighing them. The sampled average densities can be computed by dividing the weight of litter by the number of items in experiments. Finally, we can predict the number of riverine litter on tested area, and then post-process the results to compute the fluxes by dividing by the recording time (van Lieshout et al., 2020). Mass fluxes can be obtained by multiplying fluxes with respect to sampled average densities (van Emmerik et al., 2018).

(2) Mass of macroplastic litter in hotspots

Macroplastic hotspots are locations where a large amount of macroplastic litter accumulates on the water surface due to favorable morphological and environmental conditions (Moy et al., 2018). We suggest collecting hotspot images at various locations along rivers using UAVs, that can provide an overview of pollution with low human labor costs and high-resolution images (Vriend et al., 2020), and train the DL model on these images. Studies from other fields suggest DL-based semantic segmentation methods can precisely quantify the area of target objects (Kang et al., 2020; Zhou et al., 2020a). Thus, we believe that semantic segmentation can provide reliable estimations of the area occupied by hotspots. Next, limited experiments are needed to (i) collect hotspot images and measure the true area of them, (ii) sample litter in these hotspots (e.g., using nets), and (iii) weigh litter. The spatial average densities of hotspots can be computed by dividing the weight of the litter by the true area occupied by them. After accounting for image resolution (e.g., pixel to area ratio), the mass of macroplastic litter in hotspots can be obtained by multiplying the pixels identified as hotspots by the IS algorithm by the spatial average densities.

4.3. DL-based monitoring of riverine macroplastic litter

Monitoring the source, transport, distribution, sink points and trends of riverine macroplastic litter is essential for decision-makers to devise mitigation strategies and conduct targeted cleaning campaigns (van Emmerik et al., 2022b; Vriend et al., 2020). Designing effective monitoring strategies helps implement these activities more efficiently. Nonetheless, the review shows that no papers in the literature considered the design and implementation of structural monitoring strategies exploiting DL-based quantification.

To devise automated strategies for long-term monitoring of riverine macroplastic litter, we propose integrating robust DL-based macroplastic litter quantification in the "Roadmap" proposed by van Emmerik et al. (2022b). The roadmap consists of three steps (or levels): (i) method development, (ii) baseline assessment, and (iii) long-term monitoring. The first step focuses on assessing available monitoring techniques and methodologies in order to lay the foundations for developing a suitable strategy. The second step aims at establishing baseline measurements to get a first estimate on the magnitude of the problem, as well as providing insights for improving the monitoring protocol in the long run. The latter step is concerned with the exploitation of structural monitoring for higher-level tasks, such as inferring trends, estimating the effects of policy changes on the level of pollution, or mapping transport pathways.

While these steps are sequential, the roadmap is a cyclical approach for continuous improvement of structural monitoring via incorporation of new insights, monitoring goals, priorities, and data. This iterative design can be successfully mapped to the continuous integration/ continuous deployment process implemented for ML-based solution, which is known as machine learning operations (MLOps) (Ruf et al., 2021). MLOps enables long-term utilization and refinement of ML-based solutions by automating all key phases such as data management, model deployment, and model validation. The integration of MLOps into the roadmap of van Emmerik et al. (2022b) could results in the following steps leading to DL-based structural monitoring of macroplastic litter. The first step could include the selection of DL tasks (e.g., IC, OD, and IS), DL architectures (e.g., YOLO) and monitoring devices (e.g., cameras and drones) for the specific problem. In the second step, the MLOps infrastructure could be initially deployed on selected pilot projects to train baseline DL models and validate their performance. This will require systematic data gathering and ground truth measurements (e.g., visual inspection of recordings, comparison against visual counting, and sampling litter distributions from clean-up initiatives). After establishing a satisfactory baseline, long-term monitoring on the selected pilot study(ies) can start. Concurrently, the infrastructure can be strategically extended, employing the baseline DL models for monitoring at new locations. Following the data-centric AI approach (see Section 4.1.1), we can add novel, accurately labeled images at these new (or existing) locations to improve the generalization capability of the baseline models. This could also include adding litter categories of interests underrepresented in the training dataset or performing tailored ground truth validation for more accurate quantification. As witnessed in other fields of application (Ruf et al., 2021), we believe several iterations of the proposed MLOps approach may lead to robust and automated structural monitoring of macroplastic litter.

5. Conclusions

This paper reviewed the current research concerning deep learning (DL)-based detection of macroplastic litter in water bodies, and proposed key knowledge gaps and future directions. The following knowledge gaps were identified based on the critical review and discussion of 34 reviewed papers:

(1) The lack of DL models with satisfactory generalization capability, that are able to detect macroplastic litter in a given water body in a robust manner under different geographical, environmental and device setup conditions.
(2) In terms of applications, there is a lack of DL-based methods for the quantification of macroplastic (mass) fluxes and hotspots.
(3) No reviewed papers perform the design of structural monitoring strategies exploiting robust DL-based quantification.

To address above gaps, the following research directions are suggested:

(1) Future research is required to develop a robust DL model that has better generalization capabilities, i.e. that is able to detect macroplastic litter in a more reliable and consistent manner irrespective of geographical, environmental and other conditions. This can be done by exploiting new methods in machine learning, such as data-centric artificial intelligence (AI) methods and semi-supervised machine learning.
(2) More efforts should go into developing better methods for quantifying macroplastic (mass) fluxes and hotspots. We identify potential methods of quantifying the macroplastic (mass) fluxes by performing object detection tasks, and quantifying the mass of macroplastic litter in hotspots by conducting semantic segmentation tasks.
(3) The community should focus on developing and validating automated DL-based quantification for structural monitoring of macroplastic litter. We propose to approach this important task by integrating MLOps technologies in the general framework for long-term monitoring proposed by van Emmerik et al. (2022b).

While the identified gaps and suggested research directions concern all bodies of water, we highlight their importance with respect to riverine ecosystems, which are currently understudied despite their major role in transport and storage of litter. This review has the following limitations: (1) comparing the performances across different studies was difficult since models are evaluated on different dataset, and results are reported with different metrics; (2) the code and data used in most studies are not open to the public, therefore we could not perform basic checks on their quality; and (3) this work considered only peer-reviewed journal papers and conference proceedings; as such, we did not take into account project reports or other type of publications (e.g., project blogs, conference abstracts) which may present interesting results concerning the application in the field of reviewed (or other) DL models.

We expect DL to be a promising method to advance automatic detection of macroplastic litter in all bodies of water. This is an emerging research field that requires more efforts, including multidisciplinary research, and data and technology sharing around the world.

CRediT authorship contribution statement

Tianlong Jia: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Formal analysis, Visualization, Project administration, Funding acquisition. **Zoran Kapelan:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition. **Rinze de Vries:** Conceptualization, Methodology, Writing – review & editing, Supervision. **Paul Vriend:** Writing – review & editing, Supervision. **Eric Copius Peereboom:** Writing – review & editing, Supervision. **Imke Okkerman:** Writing – review & editing, Supervision. **Riccardo Taormina:** Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Formal analysis, Project administration, Funding acquisition.

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.

Data availability

No data was used for the research described in the article.

Acknowledgements

The work is supported by China Scholarship Council (No. 202006160032) and the Directorate-General for Public Works and Water Management of The Netherlands (Rijkswaterstaat). Special thanks are given to the anonymous reviewers and editors for their constructive comments.

References

Ajit, A., Acharya, K., Samanta, A., 2020. A Review of Convolutional Neural Networks. In: 2020 International Conference on Emerging Trends in Information Technology and Engineering (Ic-ETITE). Presented at the 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE), pp. 1–5. https://doi.org/10.1109/ic-ETITE47903.2020.049.

Arya, D., Maeda, H., Ghosh, S.K., Toshniwal, D., Mraz, A., Kashiyama, T., Sekimoto, Y., 2020. Transfer learning-based road damage detection for multiple countries. arXiv:2008.13101 [cs].

Bajaj, R., Garg, S., Kulkarni, N., Raut, R., 2021. Sea Debris detection using deep learning : diving deep into the sea. In: 2021 IEEE 4th International Conference on Computing,

Power and Communication Technologies (GUCON). Presented at the 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), pp. 1–6. https://doi.org/10.1109/GUCON50781.2021.9573722.
Battaglia, P.W., Hamrick, J.B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., Gulcehre, C., Song, F., Ballard, A., Gilmer, J., Dahl, G., Vaswani, A., Allen, K., Nash, C., Langston, V., Dyer, C., Heess, N., Wierstra, D., Kohli, P., Botvinick, M., Vinyals, O., Li, Y., Pascanu, R., 2018. Relational inductive biases, deep learning, and graph networks. (No. arXiv:1806.01261).
Battula, S., Kumar, M., Panda, S.K., Rao, U.M., Laveti, G., Mouli, B., 2020. Online Ocean Monitoring using Edge IoT. In: Global Oceans 2020: Singapore – U.S. Gulf Coast. Presented at the Global Oceans 2020. Gulf Coast, Singapore – U.S, pp. 1–7. https://doi.org/10.1109/IEEECONF38699.2020.9389430.
Bellou, N., Gambardella, C., Karantzalos, K., Monteiro, J.G., Canning-Clode, J., Kemna, S., Arrieta-Giron, C.A., Lemmen, C., 2021. Global assessment of innovative solutions to tackle marine litter. Nat. Sustain. 4, 516–524. https://doi.org/10.1038/s41893-021-00726-2.
Blettler, M.C.M., Abrial, E., Khan, F.R., Sivri, N., Espinola, L.A., 2018. Freshwater plastic pollution: recognizing research biases and identifying knowledge gaps. Water Res. 143, 416–424. https://doi.org/10.1016/j.watres.2018.06.015.
Borrelle, S.B., Ringma, J., Law, K.L., Monnahan, C.C., Lebreton, L., McGivern, A., Murphy, E., Jambeck, J., Leonard, G.H., Hilleary, M.A., Eriksen, M., Possingham, H. P., De Frond, H., Gerber, L.R., Polidoro, B., Tahir, A., Bernard, M., Mallos, N., Barnes, M., Rochman, C.M., 2020. Predicted growth in plastic waste exceeds efforts to mitigate plastic pollution. Science 369, 1515–1518. https://doi.org/10.1126/science.aba3656.
Broere, S., van Emmerik, T., González-Fernández, D., Luxemburg, W., de Schipper, M., Cózar, A., van de Giesen, N., 2021. Towards Underwater Macroplastic Monitoring Using Echo Sounding. Front. Earth Sci. 9.
Chai, J., Zeng, H., Li, A., Ngai, E.W.T., 2021. Deep learning in computer vision: a critical review of emerging techniques and application scenarios. Mach. Learn. Appl. 6, 100134 https://doi.org/10.1016/j.mlwa.2021.100134.
de Vries, R., Egger, M., Mani, T., Lebreton, L., 2021. Quantifying floating plastic debris at sea using vessel-based optical data and artificial intelligence. Remote Sens (Basel) 13, 3401. https://doi.org/10.3390/rs13173401.
Deng, H., Ergu, D., Liu, F., Ma, B., Cai, Y., 2021. An embeddable algorithm for automatic garbage detection based on complex marine environment. Sensors 21, 6391. https://doi.org/10.3390/s21196391.
Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. ImageNet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition. Presented at the 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. https://doi.org/10.1109/CVPR.2009.5206848.
Dong, K., Zhou, C., Ruan, Y., Li, Y., 2020. MobileNetV2 Model for Image Classification. In: 2020 2nd International Conference on Information Technology and Computer Application (ITCA). Presented at the 2020 2nd International Conference on Information Technology and Computer Application (ITCA), pp. 476–480. https://doi.org/10.1109/ITCA52113.2020.00106.
Dwibedi, D., Misra, I., Hebert, M., 2017. Cut, paste and learn: surprisingly easy synthesis for instance detection. In: 2017 IEEE International Conference on Computer Vision (ICCV). Presented at the 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, Venice, pp. 1310–1319. https://doi.org/10.1109/ICCV.2017.146.
Everingham, M., Gool, L.V., Williams, C.K.I., Winn, J., Zisserman, A., 2010. The pascal visual object classes (VOC) challenge. Int. J. Comput. Vis. 88, 303–338. https://doi.org/10.1007/s11263-009-0275-4.
Fallati, L., Polidori, A., Salvatore, C., Saponari, L., Savini, A., Galli, P., 2019. Anthropogenic marine debris assessment with unmanned aerial vehicle imagery and deep learning: a case study along the beaches of the Republic of Maldives. Sci. Total Environ. 693, 133581 https://doi.org/10.1016/j.scitotenv.2019.133581.
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). Presented at the 2019 International Conference on Robotics and Automation (ICRA), pp. 5752–5758. https://doi.org/10.1109/ICRA.2019.8793975.
Garcia-Garin, O., Monleón-Getino, T., López-Brosa, P., Borrell, A., Aguilar, A., Borja-Robalino, R., Cardona, L., Vighi, M., 2021. Automatic detection and quantification of floating marine macro-litter in aerial images: introducing a novel deep learning approach connected to a web application in R. Environmen. Pollut. 273, 116490 https://doi.org/10.1016/j.envpol.2021.116490.
Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T.-Y., Cubuk, E.D., Le, Q.V., Zoph, B., 2021. Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Nashville, TN, USA, pp. 2917–2927. https://doi.org/10.1109/CVPR46437.2021.00294.
Gholami, A., Kwon, K., Wu, B., Tai, Z., Yue, X., Jin, P., Zhao, S., Keutzer, K., 2018. SqueezeNext: hardware-Aware Neural Network Design. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Presented at the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, Salt Lake City, UT, pp. 1719–171909. https://doi.org/10.1109/CVPRW.2018.00215.
Gonçalves, G., Andriolo, U., Pinto, L., Duarte, D., 2020. Mapping marine litter with unmanned aerial systems: a showcase comparison among manual image screening and machine learning techniques. Mar. Pollut. Bull. 155, 111158 https://doi.org/10.1016/j.marpolbul.2020.111158.
González-Fernández, D., Cózar, A., Hanke, G., Viejo, J., Morales-Caselles, C., Bakiu, R., Barceló, D., Bessa, F., Bruge, A., Cabrera, M., Castro-Jiménez, J., Constant, M., Crosti, R., Galletti, Y., Kideys, A.E., Machitadze, N., Pereira de Brito, J., Pogojeva, M., Ratola, N., Rigueira, J., Rojo-Nieto, E., Savenko, O., Schöneich-Argent, R.I., Siedlewicz, G., Suaria, G., Tourgeli, M., 2021. Floating macrolitter leaked from Europe into the ocean. Nat. Sustain. 4, 474–483. https://doi.org/10.1038/s41893-021-00722-6.
Grøsvik, B.E., Prokhorova, T., Eriksen, E., Krivosheya, P., Horneland, P.A., Prozorkevich, D., 2018. Assessment of marine litter in the barents sea, a part of the joint Norwegian–Russian ecosystem survey. Front. Mar. Sci. 5 https://doi.org/10.3389/fmars.2018.00072.
Hardesty, B.D., Lawson, T.J., Velde, T.van der, Lansdell, M., Wilcox, C., 2017. Estimating quantities and sources of marine debris at a continental scale. Front Ecol. Environ. 15, 18–25. https://doi.org/10.1002/fee.1447.
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). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.
Hegde, R., Patel, S., Naik, R.G., Nayak, S.N., Shivaprakasha, K.S., Bhandarkar, R., 2021. Underwater Marine Life and Plastic Waste Detection Using Deep Learning and Raspberry Pi. In: Kalya, S., Kulkarni, M., Shivaprakasha, K.S. (Eds.), Advances in VLSI, Signal Processing, Power Electronics, IoT, Communication and Embedded Systems, Lecture Notes in Electrical Engineering. Springer, Singapore, pp. 263–272. https://doi.org/10.1007/978-981-16-0443-0_22.
Hong, J., Fulton, M.S., Sattar, J., 2020. TrashCan 1.0 An Instance-Segmentation Labeled Dataset of Trash Observations. doi:10.13020/g1gx-y834.
Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y.-W., Wu, J., 2020. UNet 3+: a Full-Scale Connected UNet for Medical Image Segmentation. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Presented at the ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1055–1059. https://doi.org/10.1109/ICASSP40776.2020.9053405.
Im, D., Lee, S., Lee, H., Yoon, B., So, F., Jeong, J., 2021. A Data-Centric Approach to Design and Analysis of a Surface-Inspection System Based on Deep Learning in the Plastic Injection Molding Industry. Processes 9, 1895. https://doi.org/10.3390/pr9111895.
Imhof, H.K., Wiesheu, A.C., Anger, P.M., Niessner, R., Ivleva, N.P., Laforsch, C., 2018. Variation in plastic abundance at different lake beach zones - a case study. Sci. Total Environ. 613–614, 530–537. https://doi.org/10.1016/j.scitotenv.2017.08.300.
Jabari, O., Ayalew, Y., Motshegwa, T., 2021. Semi-Automated X-ray Transmission Image Annotation Using Data-efficient Convolutional Neural Networks and Cooperative Machine Learning. In: 2021 The 5th International Conference on Video and Image Processing. Presented at the ICVIP 2021: 2021 The 5th International Conference on Video and Image Processing. ACM, Hayward CA USA, pp. 205–214. https://doi.org/10.1145/3511176.3511207.
Jain, S., Smit, A., Ng, A.Y., Rajpurkar, P., 2021. Effect of radiology report labeler quality on deep learning models for chest X-ray interpretation. arXiv:2104.00793 [cs, eess].
Jakovljevic, G., Govedarica, M., Alvarez-Taboada, F., 2020. A deep learning model for automatic plastic mapping using unmanned aerial vehicle (UAV) data. Remote Sens (Basel) 12, 1515. https://doi.org/10.3390/rs12091515.
Jambeck, J.R., Geyer, R., Wilcox, C., Siegler, T.R., Perryman, M., Andrady, A., Narayan, R., Law, K.L., 2015. Plastic waste inputs from land into the ocean. Science 347, 768–771. https://doi.org/10.1126/science.1260352.
Jung, I.Y., 2020. A review of privacy-preserving human and human activity recognition. Int. J. Smart Sens. Intell. Syst. 13, 1–13. https://doi.org/10.21307/ijssis-2020-008.
Kako, S., Morita, S., Taneda, T., 2020. Estimation of plastic marine debris volumes on beaches using unmanned aerial vehicles and image processing based on deep learning. Mar. Pollut. Bull. 155, 111127 https://doi.org/10.1016/j.marpolbul.2020.111127.
Kang, D., Benipal, S.S., Gopal, D.L., Cha, Y.-J., 2020. Hybrid pixel-level concrete crack segmentation and quantification across complex backgrounds using deep learning. Automat. Construc. 118, 103291 https://doi.org/10.1016/j.autcon.2020.103291.
Kylili, K., Artusi, A., Hadjistassou, C., 2021. A new paradigm for estimating the prevalence of plastic litter in the marine environment. Mar. Pollut. Bull. 173, 113127 https://doi.org/10.1016/j.marpolbul.2021.113127.
Kylili, K., Hadjistassou, C., Artusi, A., 2020. An intelligent way for discerning plastics at the shorelines and the seas. Environ. Sci. Pollut. Res. 27, 42631–42643. https://doi.org/10.1007/s11356-020-10105-7.
Kylili, K., Kyriakides, I., Artusi, A., Hadjistassou, C., 2019. Identifying floating plastic marine debris using a deep learning approach. Environ. Sci. Pollut. Res. 26, 17091–17099. https://doi.org/10.1007/s11356-019-05148-4.
Lavitas, L., Redfield, O., Lee, A., Fletcher, D., Eck, M., Janardhanan, S., 2021. Annotation quality framework - accuracy, credibility, and consistency.
Lebreton, L., Slat, B., Ferrari, F., Sainte-Rose, B., Aitken, J., Marthouse, R., Hajbane, S., Cunsolo, S., Schwarz, A., Levivier, A., Noble, K., Debeljak, P., Maral, H., Schoeneich-Argent, R., Brambini, R., Reisser, J., 2018. Evidence that the Great Pacific Garbage Patch is rapidly accumulating plastic. Sci. Rep. 8, 4666. https://doi.org/10.1038/s41598-018-22939-w.
Lebreton, L.C.M., van der Zwet, J., Damsteeg, J.-W., Slat, B., Andrady, A., Reisser, J., 2017. River plastic emissions to the world's oceans. Nat. Commun. 8, 15611. https://doi.org/10.1038/ncomms15611.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.
Lin, F., Hou, T., Jin, Q., You, A., 2021. Improved YOLO based detection algorithm for floating debris in waterway. Entropy 23, 1111. https://doi.org/10.3390/e23091111.
Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft COCO: common Objects in Context. In: Fleet, D.,

Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), Computer Vision – ECCV 2014, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 740–755. https://doi.org/10.1007/978-3-319-10602-1_48.

Liu, Ze, Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., Ning, J., Cao, Y., Zhang, Z., Dong, L., Wei, F., Guo, B., 2022a. Swin transformer V2: scaling up capacity and resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, p. 11.

Liu, Zhuang, Mao, H., Wu, C.-.Y., Feichtenhofer, C., Darrell, T., Xie, S., 2022b. A ConvNet for the 2020s. In: Presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11976–11986.

Marin, I., Mladenović, S., Gotovac, S., Zaharija, G., 2021. Deep-feature-based approach to marine debris classification. Appl. Sci. 11, 5644. https://doi.org/10.3390/app11125644.

Martin, C., Almahasheer, H., Duarte, C.M., 2019. Mangrove forests as traps for marine litter. Environmen. Pollut. 247, 499–508. https://doi.org/10.1016/j.envpol.2019.01.067.

Martin, C., Zhang, Q., Zhai, D., Zhang, X., Duarte, C.M., 2021. Enabling a large-scale assessment of litter along Saudi Arabian red sea shores by combining drones and machine learning. Environmen. Pollut. 277, 116730 https://doi.org/10.1016/j.envpol.2021.116730.

McCann, E., Li, L., Pangle, K., Johnson, N., Eickholt, J., 2018. An underwater observation dataset for fish classification and fishery assessment. Sci. Data 5, 180190. https://doi.org/10.1038/sdata.2018.190.

Meijer, L.J.J., van Emmerik, T., Ent, R.van der, Schmidt, C., Lebreton, L., 2021. More than 1000 rivers account for 80% of global riverine plastic emissions into the ocean. Sci Adv 7, eaaz5803. https://doi.org/10.1126/sciadv.aaz5803.

Mifdal, J., Longépé, N., Rußwurm, M., 2021. Towards detecting floating objects on a global scale with learned spatial features using sentinel 2. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci. 285–293. https://doi.org/10.5194/isprs-annals-V-3-2021-285-2021. V-3–2021.

Ming, L., Xiaolin, H., 2015. Recurrent convolutional neural network for object recognition. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Boston, MA, USA, pp. 3367–3375. https://doi.org/10.1109/CVPR.2015.7298958.

Misra, I., van der Maaten, L., 2020. Self-supervised learning of pretext-invariant representations. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Seattle, WA, USA, pp. 6706–6716. https://doi.org/10.1109/CVPR42600.2020.00674.

Motamedi, M., Sakharnykh, N., Kaldewey, T., 2021. A Data-Centric Approach for Training Deep Neural Networks with Less Data. arXiv:2110.03613 [cs].

Moy, K., Neilson, B., Chung, A., Meadows, A., Castrence, M., Ambagis, S., Davidson, K., 2018. Mapping coastal marine debris using aerial imagery and spatial analysis. Marine Pollution Bulletin, SI: Japanese Tsunami Debris 132, 52–59. https://doi.org/10.1016/j.marpolbul.2017.11.045.

Musić, J., Kružić, S., Stančić, I., Alexandrou, F., 2020. Detecting underwater sea litter using deep neural networks: an initial study. In: 2020 5th International Conference on Smart and Sustainable Technologies (SpliTech). Presented at the 2020 5th International Conference on Smart and Sustainable Technologies (SpliTech), pp. 1–6. https://doi.org/10.23919/SpliTech49282.2020.9243709.

Mustafah, Y.M., Azman, A.W., Ani, M.H., 2012. Object Distance and Size Measurement Using Stereo Vision System. AMR 622–623, 1373–1377. https://doi.org/10.4028/www.scientific.net/AMR.622-623.1373.

Neupane, D., Seok, J., 2020. A Review on Deep Learning-Based Approaches for Automatic Sonar Target Recognition. Electronics (Basel) 9, 1972. https://doi.org/10.3390/electronics9111972.

Noroozi, M., Favaro, P., 2016. Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (Eds.), Computer Vision – ECCV 2016. Springer International Publishing, Cham, pp. 69–84. https://doi.org/10.1007/978-3-319-46466-4_5.

Padilla, R., Netto, S.L., da Silva, E.A.B., 2020. A Survey on Performance Metrics for Object-Detection Algorithms. In: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP). Presented at the 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), pp. 237–242. https://doi.org/10.1109/IWSSIP48289.2020.9145130.

Pan, S.J., Yang, Q., 2010. A Survey on Transfer Learning. IEEE Trans Knowl Data Eng 22, 1345–1359. https://doi.org/10.1109/TKDE.2009.191.

Panwar, H., Gupta, P.K., Siddiqui, M.K., Morales-Menendez, R., Bhardwaj, P., Sharma, S., Sarker, I.H., 2020. AquaVision: automating the detection of waste in water bodies using deep transfer learning. Case Studies in Chemical and Environmental Engineering 2, 100026. doi:10.1016/j.cscee.2020.100026.

Papakonstantinou, A., Batsaris, M., Spondylidis, S., Topouzelis, K., 2021. A Citizen Science Unmanned Aerial System Data Acquisition Protocol and Deep Learning Techniques for the Automatic Detection and Mapping of Marine Litter Concentrations in the Coastal Zone. Drones 5, 6. https://doi.org/10.3390/drones5010006.

Paul, S., Chen, P.-.Y., 2022. Vision Transformers Are Robust Learners. Proceedings of the AAAI Conference on Artificial Intell. 36, 2071–2081. https://doi.org/10.1609/aaai.v36i2.20103.

Politikos, D.V., Fakiris, E., Davvetas, A., Klampanos, I.A., Papatheodorou, G., 2021. Automatic detection of seafloor marine litter using towed camera images and deep learning. Mar. Pollut. Bull. 164, 111974 https://doi.org/10.1016/j.marpolbul.2021.111974.

Proença, P.F., Simões, P., 2020. TACO: trash annotations in context for litter detection. arXiv:2003.06975 [cs].

Putra, F.F., Prabowo, Y.D., 2021. Low resource deep learning to detect waste intensity in the river flow. Bulletin EEI 10, 2724–2732. https://doi.org/10.11591/eei.v10i5.3062.

Qin, X., Luo, X., Wu, Z., Shang, J., 2021. Optimizing the sediment classification of small side-scan sonar images based on deep learning. IEEE Access 9, 29416–29428. https://doi.org/10.1109/ACCESS.2021.3052206.

Recht, B., Roelofs, R., Schmidt, L., Shankar, V., 2018. Do CIFAR-10 Classifiers Generalize to CIFAR-10? arXiv:1806.00451 [cs, stat].

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Presented at the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Las Vegas, NV, USA, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91.

Ruf, P., Madan, M., Reich, C., Ould-Abdeslam, D., 2021. Demystifying MLOps and Presenting a Recipe for the Selection of Open-Source Tools. Appl. Sci. 11, 8861. https://doi.org/10.3390/app11198861.

Shorten, C., Khoshgoftaar, T.M., 2019. A survey on image data augmentation for deep learning. J Big Data 6, 60. https://doi.org/10.1186/s40537-019-0197-0.

Singvi, M., Dasgupta, A., Routray, A., 2012. A real time algorithm for detection of spectacles leading to eye detection. In: 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI). Presented at the 2012 4th International Conference on Intelligent Human Computer Interaction (IHCI), pp. 1–6. https://doi.org/10.1109/IHCI.2012.6481837.

Sit, M., Demiray, B.Z., Xiang, Z., Ewing, G.J., Sermet, Y., Demir, I., 2020. A comprehensive review of deep learning applications in hydrology and water resources. Water Sci. Technol. 82, 2635–2670. https://doi.org/10.2166/wst.2020.369.

Song, K., Jung, J.-.Y., Lee, S.H., Park, S., 2021. A comparative study of deep learning-based network model and conventional method to assess beach debris standing-stock. Mar. Pollut. Bull. 168, 112466 https://doi.org/10.1016/j.marpolbul.2021.112466.

Sun, C., Shrivastava, A., Singh, S., Gupta, A., 2017. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. arXiv:1707.02968 [cs].

Tang, Z., Tian, E., Wang, Y., Wang, L., Yang, T., 2021. Nondestructive defect detection in castings by using spatial attention bilinear convolutional neural network. IEEE Transac. Indust. Inform. 17, 82–89. https://doi.org/10.1109/TII.2020.2985159.

Tasseron, P., Zinsmeister, H., Rambonnet, L., Hiemstra, A.-.F., Siepman, D., van Emmerik, T., 2020. Plastic hotspot mapping in urban water systems. Geosciences (Basel) 10, 342. https://doi.org/10.3390/geosciences10090342.

Tharani, M., Amin, A.W., Rasool, F., Maaz, M., Taj, M., Muhammad, A., 2021. Trash Detection on Water Channels. In: Mantoro, T., Lee, M., Ayu, M.A., Wong, K.W., Hidayanto, A.N. (Eds.), Neural Information Processing, Lecture Notes in Computer Science. Springer International Publishing, Cham, pp. 379–389. https://doi.org/10.1007/978-3-030-92185-9_31.

Thiagarajan, S., Satheesh Kumar, G., 2019. Machine learning for beach litter detection. In: Tanveer, M., Pachori, R.B. (Eds.), Machine Intelligence and Signal Analysis, Advances in Intelligent Systems and Computing. Springer, Singapore, pp. 259–266. https://doi.org/10.1007/978-981-13-0923-6_22.

Tian, M., Li, X., Kong, S., wu, L., Yu, J., 2021. Pruning-Based YOLOv4 algorithm for underwater gabage detection. In: 2021 40th Chinese Control Conference (CCC). Presented at the 2021 40th Chinese Control Conference (CCC), pp. 4008–4013. https://doi.org/10.23919/CCC52363.2021.9550592.

Tramoy, R., Gasperi, J., Colasse, L., Silvestre, M., Dubois, P., Noûs, C., Tassin, B., 2020. Transfer dynamics of macroplastics in estuaries – New insights from the Seine estuary: part 2. Short-term dynamics based on GPS-trackers. Mar. Pollut. Bull. 160, 111566 https://doi.org/10.1016/j.marpolbul.2020.111566.

Valdenegro-Toro, M., 2016. Submerged marine debris detection with autonomous underwater vehicles. In: 2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA). Presented at the 2016 International Conference on Robotics and Automation for Humanitarian Applications (RAHA), pp. 1–7. https://doi.org/10.1109/RAHA.2016.7931907.

van Emmerik, T., Kieu-Le, T.-.C., Loozen, M., van Oeveren, K., Strady, E., Bui, X.-.T., Egger, M., Gasperi, J., Lebreton, L., Nguyen, P.-.D., Schwarz, A., Slat, B., Tassin, B., 2018. A Methodology to Characterize Riverine Macroplastic Emission Into the Ocean. Front. Mar. Sci. 5, 372. https://doi.org/10.3389/fmars.2018.00372.

van Emmerik, T., Mellink, Y., Hauk, R., Waldschlaeger, K., Schreyers, L., 2022a. Rivers as plastic reservoirs. Front. Water 3, 786936. https://doi.org/10.3389/frwa.2021.786936.

van Emmerik, T., Schwarz, A., 2020. Plastic debris in rivers. WIREs Water 7, e1398. https://doi.org/10.1002/wat2.1398.

van Emmerik, T., Seibert, J., Strobl, B., Etter, S., den Oudendammer, T., Rutten, M., bin Ab Razak, M.S., van Meerveld, I., 2020. Crowd-based observations of riverine macroplastic pollution. Front. Earth Sci. 8.

van Emmerik, T., Vriend, P., Copius Peereboom, E., 2022b. Roadmap for long-term macroplastic monitoring in rivers. Front. Environ. Sci. 9, 802245 https://doi.org/10.3389/fenvs.2021.802245.

van Lieshout, C., van Oeveren, K., van Emmerik, T., Postma, E., 2020. Automated River Plastic Monitoring Using Deep Learning and Cameras. Earth Space Sci. 7, e2019EA000960 https://doi.org/10.1029/2019EA000960.

Vriend, P., Roebroek, C.T.J., van Emmerik, T., 2020. Same but Different: a Framework to Design and Compare Riverbank Plastic Monitoring Strategies. Front. Water 2, 563791. https://doi.org/10.3389/frwa.2020.563791.

Wambugu, N., Chen, Y., Xiao, Z., Tan, K., Wei, M., Liu, X., Li, J., 2021. Hyperspectral image classification on insufficient-sample and feature learning using deep neural networks: a review. Int. J. Appl. Earth Observation Geoinform. 105, 102603 https://doi.org/10.1016/j.jag.2021.102603.

Wang, C.-.Y., Bochkovskiy, A., Liao, H.-Y.M., 2022. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors (No. arXiv:2207.02696).

Watanabe, J., Shao, Y., Miura, N., 2019. Underwater and airborne monitoring of marine ecosystems and debris. J. Appl. Rem. Sens. 13, 1. https://doi.org/10.1117/1.JRS.13.044509.

Wei, Y., Xia, W., Huang, J., Ni, B., Dong, J., Zhao, Y., Yan, S., 2016. CNN: single-label to multi-label. IEEE Trans. Pattern Anal. Mach. Intell. 38, 1901–1907. https://doi.org/10.1109/TPAMI.2015.2491929.

Weideman, E.A., Perold, V., Ryan, P.G., 2020. Limited long-distance transport of plastic pollution by the Orange-Vaal River system, South Africa. Sci. Total Environ. 727, 138653 https://doi.org/10.1016/j.scitotenv.2020.138653.

Wenneker, B., Oosterbaan, L., 2010. Guideline For Monitoring Marine Litter On the Beaches in the OSPAR Maritime Area, Edition 1.0. (Report). OSPAR Commission. https://doi.org/10.25607/OBP-968.

Wolf, M., Berg, K.van den, Garaba, S.P., Gnann, N., Sattler, K., Stahl, F., Zielinski, O., 2020. Machine learning for aquatic plastic litter detection, classification and quantification (APLASTIC-Q). Environ. Res. Lett. 15, 114042 https://doi.org/10.1088/1748-9326/abbd01.

Wu, Y.-.C., Shih, P.-.Y., Chen, L.-.P., Wang, C.-.C., Samani, H., 2020. Towards Underwater Sustainability using ROV Equipped with Deep Learning System. In: 2020 International Automatic Control Conference (CACS). Presented at the 2020 International Automatic Control Conference (CACS), pp. 1–5. https://doi.org/10.1109/CACS50047.2020.9289788.

Xu, Z., Meng, A., Shi, Z., Yang, W., Chen, Z., Huang, L., 2021. Continuous Copy-Paste for One-Stage Multi-Object Tracking and Segmentation. In: Presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 15323–15332.

Xue, B., Huang, B., Chen, G., Li, H., Wei, W., 2021a. Deep-sea debris identification using deep convolutional neural networks. IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing 14, 8909–8921. https://doi.org/10.1109/JSTARS.2021.3107853.

Xue, B., Huang, B., Wei, W., Chen, G., Li, H., Zhao, N., Zhang, H., 2021b. An efficient deep-sea debris detection method using deep neural networks. IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing 14, 12348–12360. https://doi.org/10.1109/JSTARS.2021.3130238.

Zhang, C., Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J., Atkinson, P.M., 2018. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. ISPRS Journal of Photogrammetry and Remote Sensing. Geospatial. Comp. Vision 140, 133–144. https://doi.org/10.1016/j.isprsjprs.2017.07.014.

Zhang, G., Shang, B., Chen, Y., Moyes, H., 2017. SmartCaveDrone: 3D cave mapping using UAVs as robotic co-archaeologists. In: 2017 International Conference on Unmanned Aircraft Systems (ICUAS). Presented at the 2017 International Conference on Unmanned Aircraft Systems (ICUAS), pp. 1052–1057. https://doi.org/10.1109/ICUAS.2017.7991499.

Zhou, L., Li, Z., Zhou, J., Li, H., Chen, Y., Huang, Y., Xie, D., Zhao, L., Fan, M., Hashmi, S., Abdelkareem, F., Eiada, R., Xiao, X., Li, L., Qiu, Z., Gao, X., 2020a. A Rapid, Accurate and Machine-Agnostic Segmentation and Quantification Method for CT-Based COVID-19 Diagnosis. IEEE Trans Med Imaging 39, 2638–2652. https://doi.org/10.1109/TMI.2020.3001810.

Zhou, X., Wang, Y., Zhu, Q., Mao, J., Xiao, C., Lu, X., Zhang, H., 2020b. A Surface Defect Detection Framework for Glass Bottle Bottom Using Visual Attention Model and Wavelet Transform. IEEE Transac. Indust. Inform. 16, 2189–2201. https://doi.org/10.1109/TII.2019.2935153.