

LETTER • OPEN ACCESS

## Machine learning for aquatic plastic litter detection, classification and quantification (APLASTIC-Q)

To cite this article: Mattis Wolf *et al* 2020 *Environ. Res. Lett.* **15** 114042

View the [article online](#) for updates and enhancements.

# Environmental Research Letters



## LETTER

### OPEN ACCESS

RECEIVED  
5 May 2020

REVISED  
28 August 2020

ACCEPTED FOR PUBLICATION  
30 September 2020

PUBLISHED  
16 November 2020

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



# Machine learning for aquatic plastic litter detection, classification and quantification (APLASTIC-Q)

Mattis Wolf<sup>1,2</sup> , Katelijn van den Berg<sup>3</sup>, Shungudzemwoyo P Garaba<sup>1,2</sup> , Nina Gnann<sup>1</sup> , Klaus Sattler<sup>3</sup>, Frederic Stahl<sup>1,4</sup>  and Oliver Zielinski<sup>1,2</sup> 

<sup>1</sup> Marine Perception Research Group, German Research Centre for Artificial Intelligence (DFKI), Marie-Curie-Str. 1, 26129, Oldenburg, Germany

<sup>2</sup> Marine Sensor Systems Group, Institute for Chemistry and Biology of the Marine Environment, Carl von Ossietzky University of Oldenburg, Schleusenstraße 1, Wilhelmshaven 26382, Germany

<sup>3</sup> Environment, Natural Resources and Blue Economy, The World Bank, 1818 H Street, NW, Washington, DC 20433, United States of America

<sup>4</sup> Department of Computer Science, University of Reading, Reading, United Kingdom

E-mail: [Mattis.Wolf@dfki.de](mailto:Mattis.Wolf@dfki.de)

**Keywords:** convolutional neural networks, plastic litter, remote sensing, detection, machine learning, river and beach ecosystems, Cambodia

Supplementary material for this article is available [online](#)

## Abstract

Large quantities of mismanaged plastic waste are polluting and threatening the health of the blue planet. As such, vast amounts of this plastic waste found in the oceans originates from land. It finds its way to the open ocean through rivers, waterways and estuarine systems. Here we present a novel machine learning algorithm based on convolutional neural networks (CNNs) that is capable of detecting and quantifying floating and washed ashore plastic litter. The aquatic plastic litter detection, classification and quantification system (APLASTIC-Q) was developed and trained using very high geo-spatial resolution imagery ( $\sim 5 \text{ pixels cm}^{-1} = 0.002 \text{ m pixel}^{-1}$ ) captured from aerial surveys in Cambodia. APLASTIC-Q was made up of two machine learning components (i) plastic litter detector (PLD-CNN) and (ii) plastic litter quantifier (PLQ-CNN). PLD-CNN managed to categorize targets as water, sand, vegetation and plastic litter with an 83% accuracy. It also provided a qualitative count of litter as low or high based on a thresholding approach. PLQ-CNN further distinguished and enumerated the litter items in each of the classes defined as water bottles, Styrofoam, canisters, cartons, bowls, shoes, polystyrene packaging, cups, textile, carry bags small or large. The types and amounts of plastic litter provide benchmark information that is urgently needed for decision-making by policymakers, citizens and other public and private stakeholders. Quasi-quantification was based on automated counts of items present in the imagery with caveats of underlying object in case of aggregated litter. Our scientific evidence-based machine learning algorithm has the prospects of complementing net trawl surveys, field campaigns and clean-up activities for improved quantification of plastic litter. APLASTIC-Q is a smart algorithm that is easy to adapt for fast and automated detection as well as quantification of floating or washed ashore plastic litter from aerial, high-altitude pseudo satellites and space missions.

## 1. Introduction

Plastic pollution is a ‘wicked environmental problem’ with annual estimates indicating global rivers discharging several million metric tonnes of plastic waste into the oceans (Balint *et al* 2011, Jambeck *et al* 2015, Lebreton *et al* 2017). These studies have also reported that river systems of Asian nations including

Cambodia, transport substantial amounts of plastics into the open ocean (Sethy *et al* 2014, Lebreton *et al* 2017, van Emmerik *et al* 2019). Clearly, these plastic polluted waterways do not only pose localized health and environmental problems but a global threat to the blue economy (Todd *et al* 2010, Blettler *et al* 2018).

During the rainy season, flooding and episodes of extreme weather events are possible. As a

result of these events, it is suspected that plastic leakage is enhanced into rivers, and subsequently shallow sea and the ocean. After these periods of flooding or even high tides, plastic litter is washed ashore or trapped by vegetation, whereas the remainder is transported offshore. Combating plastics should be made in the rivers or even before they are reaching the aquatic environment (Hohn *et al* 2020). Innovative and affordable monitoring strategies need to be put in place for improved waste and plastic management. These strategies should hinge on scientific evidence-based research highlighting sources and abundances of litter in various towns, states and countries. Furthermore, plastic litter descriptors like polymer types constitute actionable information to develop targeted policies and legislation for priority plastic items and investments into improved plastic waste collection and recycling. This is in line with large scale political initiatives like the EU Marine Strategy Framework Directive's descriptor 10 (Galgani *et al* 2013), the Single-use Plastics Directive 2019 (EUPPD 2019), UN Sustainable Development Goal 14 target 14.1 to 'prevent and significantly reduce marine pollutions of all kinds' (Recuero Virto 2018), even the UN 'Decade of Ocean Science for Sustainable Development (2021–2030)' aiming to 'support efforts to reverse the cycle of decline in ocean health and create improved conditions for sustainable development of the ocean' (Ryabinin *et al* 2019, IOC-UNESCO 2019).

Society has shown an increasing interest in advanced and automated monitoring strategies relevant to the plastic litter pollution (Garaba *et al* 2018, Smail *et al* 2019, Maximenko *et al* 2019). Already, remote sensing technologies combined with machine learning algorithms have been at the forefront of advancing scientific knowledge about plastic litter distributions as well as bridging the gap from net trawl surveys to numerical simulations. Unmanned aerial systems (UAS) and satellite imagery captured at high to very high geo-spatial resolution on the shoreline and along beaches have been shown to be useful in monitoring washed ashore plastic litter. Using these captured images, state-of-the-art machine learning algorithms have been evaluated for potential applications in automated quantification and identification of plastics litter (Acuña-Ruz *et al* 2018, Martin *et al* 2018, Bak *et al* 2019, Kylili *et al* 2019, Gonçalves *et al* 2020). A Random Forest classifier trained with histograms of gradients was implemented to recognize isolated dry plastic objects on a beach in the Northern parts of the Red Sea as well as Cabedelo beach, Portugal (Martin *et al* 2018, Gonçalves *et al* 2020). Washed ashore plastics were detected, counted and grouped in the Chiloé islands using support vector machine (SVM) algorithms (Acuña-Ruz *et al* 2018). Beached and floating plastic litter were also detected based on Visual Geometry Group-16 and SegNet architectures (Bak *et al* 2019, Kylili *et al* 2019). Other approaches for monitoring

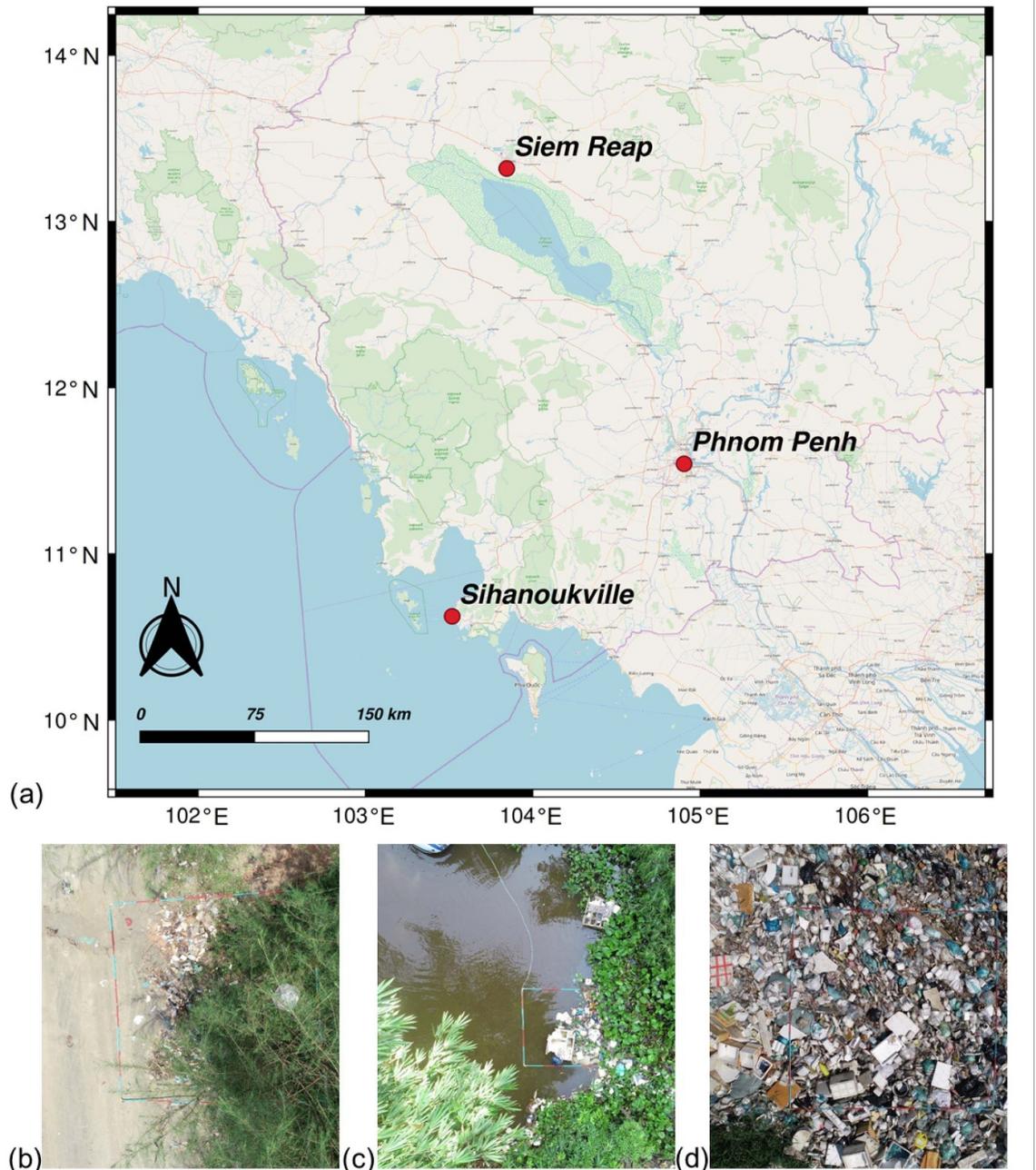
plastic litter have employed visual inspection by trained observers and statistical analysis techniques for imagery from aerial surveys (Kataoka *et al* 2012, 2018, Lebreton *et al* 2018, Garaba *et al* 2018, Bao *et al* 2018).

The scope of our research was to advance the development of machine learning algorithms for related studies about plastic distributions by presenting APLASTIC-Q. It was aimed at providing estimations of litter in a survey region as counts of litter items and predict plastic types from the imagery. We also investigated plastics using convolutional neural networks (CNNs) in various surroundings such as rivers with few plastics, river carpets and aggregated litter on beaches. We examine different machine learning techniques, which have previously been used to detect pollution on beaches and conclude that CNNs produce the most promising results, especially for the classification of plastic types. Thus, the here developed APLASTIC-Q algorithm is based on CNN technology. We also compare our classification metrics with current automated marine litter detection systems for drone imagery. As a result, APLASTIC-Q outperforms these systems with respect to various classification performance metrics. However, there are some challenges in comparing the classification results with related work in the literature. Hence, we raise the need for a framework enabling fair comparison. Our novel APLASTIC-Q algorithm is easy to adapt to potential applications for processing imagery from smartphones, handheld cameras, fixed observatories, and manned aerial and space platforms. Objectives of image processing will be automated in terms of counting and classifying macroplastic (diameter > 25 mm) items in litter. Our study thus aims to further advance the application of CNNs in monitoring plastic litter in the aquatic environments using very high geo-spatial true colour aerial images.

## 2. Methods and materials

### 2.1. Aerial survey

Aerial surveys were completed using a DJI 4 Phantom Pro photography UAS with a 20 MP RGB (red, green and blue color scheme) imaging sensor over Phnom Penh, Sihanoukville and Siem Reap in Cambodia in October 2019 (figure 1). Plastic litter was observed floating, trapped in vegetation, washed ashore on beaches and accumulated forming plastic river carpets (figures 1(b)–(d)). Images were captured at a pixel resolution of  $4864 \times 3648$  with ISO values between 100 and 400 pixels, shutter speed and aperture were set to automatic. The nadir viewing angle of the imaging sensor was  $0^\circ$  at a flight altitude of 6 m with a vertical GPS hover accuracy of 0.5 m. Flight altitude was chosen after analyses of imagery from pre-flight tests ranging between 3 and 60 m, it provided sufficient wide area coverage at sufficient resolution of objects (length > 2.5 cm). A Topcon



**Figure 1.** (a) Map of Cambodia showing areas survey by a DJI 4 Phantom Pro with a 20 MP RGB camera in October 2019, (b) beach litter in Sihanoukville, (c) floating and trapped litter in waterways of Siem Reap (d) plastic river carpets in Phnom Penh. A red or blue marker grid indicates a length of 0.5 m.

GR-5 global navigation satellite receiver system was used to mark the ground control points.

The points were used to optimize geo-referencing and mosaicking of imagery. Post-processing of the collected aerial imagery was performed using Pix4Dmapper version 4.5.3. It involved automated point cloud densification, 3D mesh generation, digital surface modelling, orthomosaic and digital terrain modelling. A visual meter scale was provided in some images to complement estimates of sizes in captured scenes. No atmospheric correction was applied to the images.

## 2.2. Detection and quantification of plastic litter algorithms

True colour RGB images collected during the aerial survey were partitioned into tiles of  $100 \times 100 \times 3$  pixels and  $50 \times 50 \times 3$  pixels. Tile size selection was based on an assessment from a prior study (Martin *et al* 2018). The plastic litter detector (PLD-CNN) algorithm is used to analyse the  $100 \times 100 \times 3$  pixel tiles. It was trained to distinguish the various targets in the tile part-wise as (i) water, (ii) vegetation, (iii) litter-low, (iv) litter-high, (v) sand or (vi) other (figure 2). Tiles with fewer than three

objects were defined as litter-low and those with at least three items were labelled litter-high. We defined these thresholds to optimize counts of litter items by the algorithm after considering that litter objects in images varied from almost none in rivers with natural surroundings to thousands of litter objects in plastic river carpets.

As for the plastic litter quantifier (PLQ-CNN) algorithm,  $50 \times 50 \times 3$  pixel tiles were selected as a divider of the  $100 \times 100 \times 3$  pixel tiles used by PLD-CNN. Applying a divider was aimed at optimizing the algorithm and thus reducing information loss. PLQ-CNN only evaluated tiles highlighted to have any amount of litter. Within these tiles PLQ-CNN was further trained to distinguish and enumerate individual litter items. A total of 18 classes were output by PLQ-CNN, these expand on the six PLD-CNN categories and include cans, cartons, plastic bags, bottles, cups, canister, polystyrene packaging, shoes, Styrofoam, strings and textiles (figure 2). Plastic bags were divided into large bags or small bags that included sweet wrappers, noodle packages, crisps bags. Identification of items was consistent with updated international marine litter classification protocol of the United States National Oceanic and Atmospheric Administration agency (GESAMP 2019). These items or plastic objects were validated by visual inspection during clean-up activities conducted following the aerial survey.

### 2.3. Neural network architecture and training

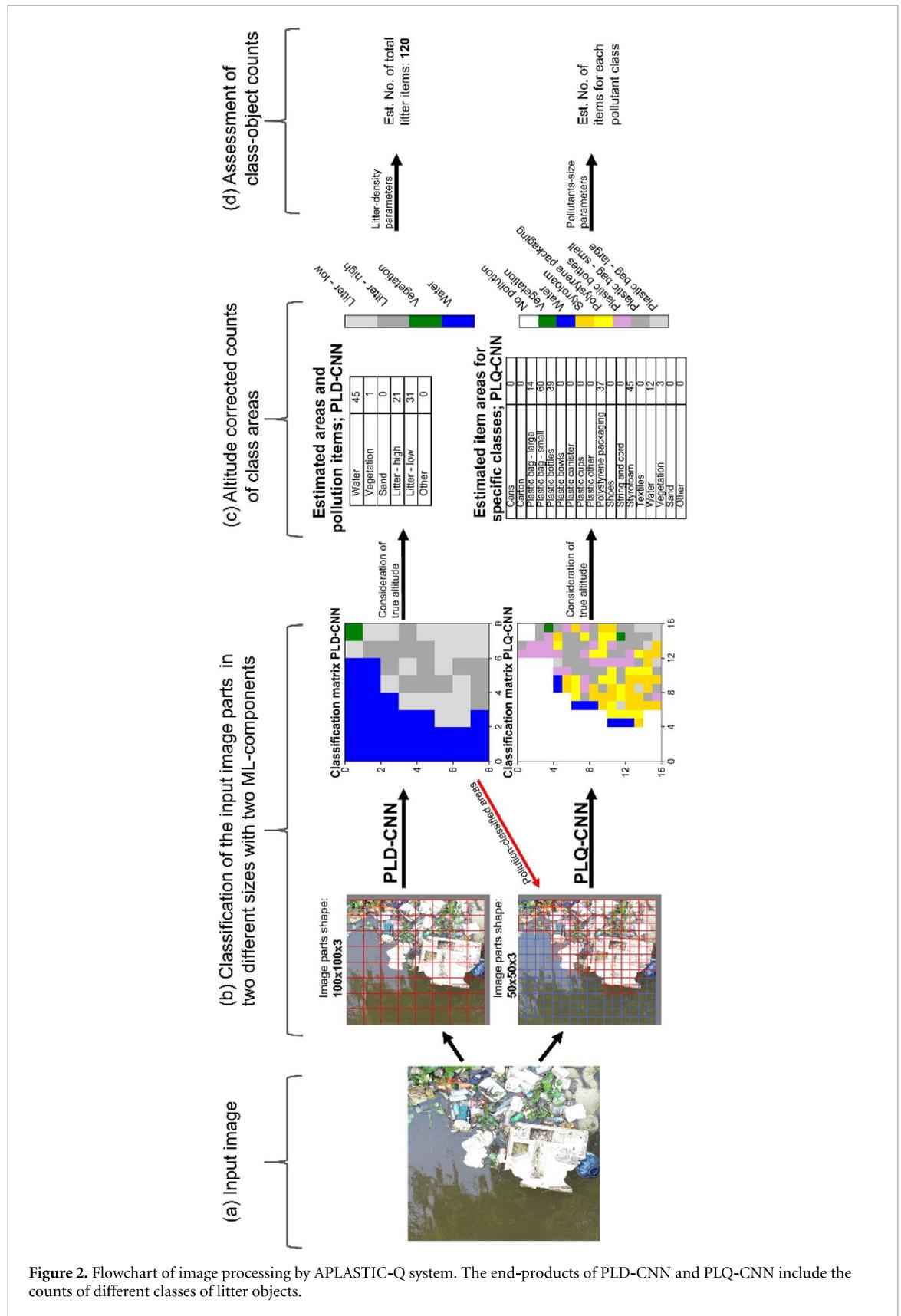
Modelling was executed on an Intel® quad-core i5 8250 U processor utilising two threads per core, a clock rate of 1.6 GHz and 8 GB random access memory. PLD-CNN and PLQ-CNN have a similar architecture except that the training tiles' sizes were different with  $100 \times 100 \times 3$  pixel and  $50 \times 50 \times 3$  pixel respectively (figure 3). It consists of four 2D CNN layers. The first two 2D convolutional layers consisted of  $32 \ 3 \times 3$  kernels and the last two 2D convolutional layers consisted of  $64 \ 3 \times 3$  kernels. 2D convolutional layers are known to work well on image data by adequately preserving some of the pixel's locality (Krizhevsky *et al* 2012, Lecun *et al* 2015). After every 2D convolutional layer there was a  $2 \times 2$  max pooling layer preceding a dropout neural network layer (dropout rate = 25%). Implementing dropouts in neural networks is known to mitigate the challenges related to overfitting and non-optimal co-adaptation. Therefore dropouts minimize generalization uncertainties (Krizhevsky *et al* 2012, Srivastava *et al* 2014). Fully connected dense neural network layers complete the architecture, the first layer comprises 512 units followed by a dropout neural network layer (dropout rate = 50%). The second dense neural network layer had units matching the number of classes, 6 units for PLD-CNN and PLQ-CNN had 18 units.

For training purposes, we randomly sampled 80% of the tiles without replacement for each class and

then tested the algorithms using the remaining 20%. Rectified linear unit (ReLU) activation functions were applied to the 2D convolutional and the dense neural network layers, but after the last dense neural network layer a *softmax* activation function was used. We utilized ReLU activation functions because they have shown to be robust in shortening the training time span compared with alternative activation function with *tanh* units (Krizhevsky *et al* 2012). Our algorithms were further established after training using CIFAR10 dataset on a Keras framework utilizing a TensorFlow backend (Krizhevsky 2009, Chollet *et al* 2015, Abadi *et al* 2016).

Training of PLD-CNN and PLQ-CNN was performed with a batch size of 32 and the categorical cross entropy was optimized using the Adam optimizer (learning rate = 0.001, beta1 = 0.9 and beta2 = 0.999), consistent with a prior study (Kingma and Ba 2015). We also applied data augmentation to randomly flip the training tiles horizontally and vertically. The probability of a vertical, horizontal flip or both was 75%. Data augmentation is a widely used technique that further decreases the overfitting phenomenon on image processing tasks (Chollet *et al* 2015, Perez and Wang 2017, Shorten and Khoshgoftaar 2019). Steps per epoch were a result of dividing the number of training tiles with the batch size parameter for each CNN. Appropriate epoch parameters were determined after an assessment training of both algorithms with 100 epochs. We obtained the least validation losses at 20 epochs for PLD-CNN and at 40 epochs for PLQ-CNN. These epoch parameters were selected for the final training due to low model biases and overfitting effects. Training of 20 PLD-CNN epochs using a sample set of 5515 tiles took 70 min translating to about 210 s per epoch. Training of PLQ-CNN used 4828 sample tiles that required 45 s per epoch or 30 min for 40 epochs. PLD-CNN dataset was established from eight of the 16 RGB true colour images. However, seven images were used to create the dataset for PLQ-CNN. Selection of the tiles for training involved extensive visual inspection that was intended to best detect and differentiate observed litter items (figures 1(b)–(d)). The remaining tiles were used for qualitative analyses.

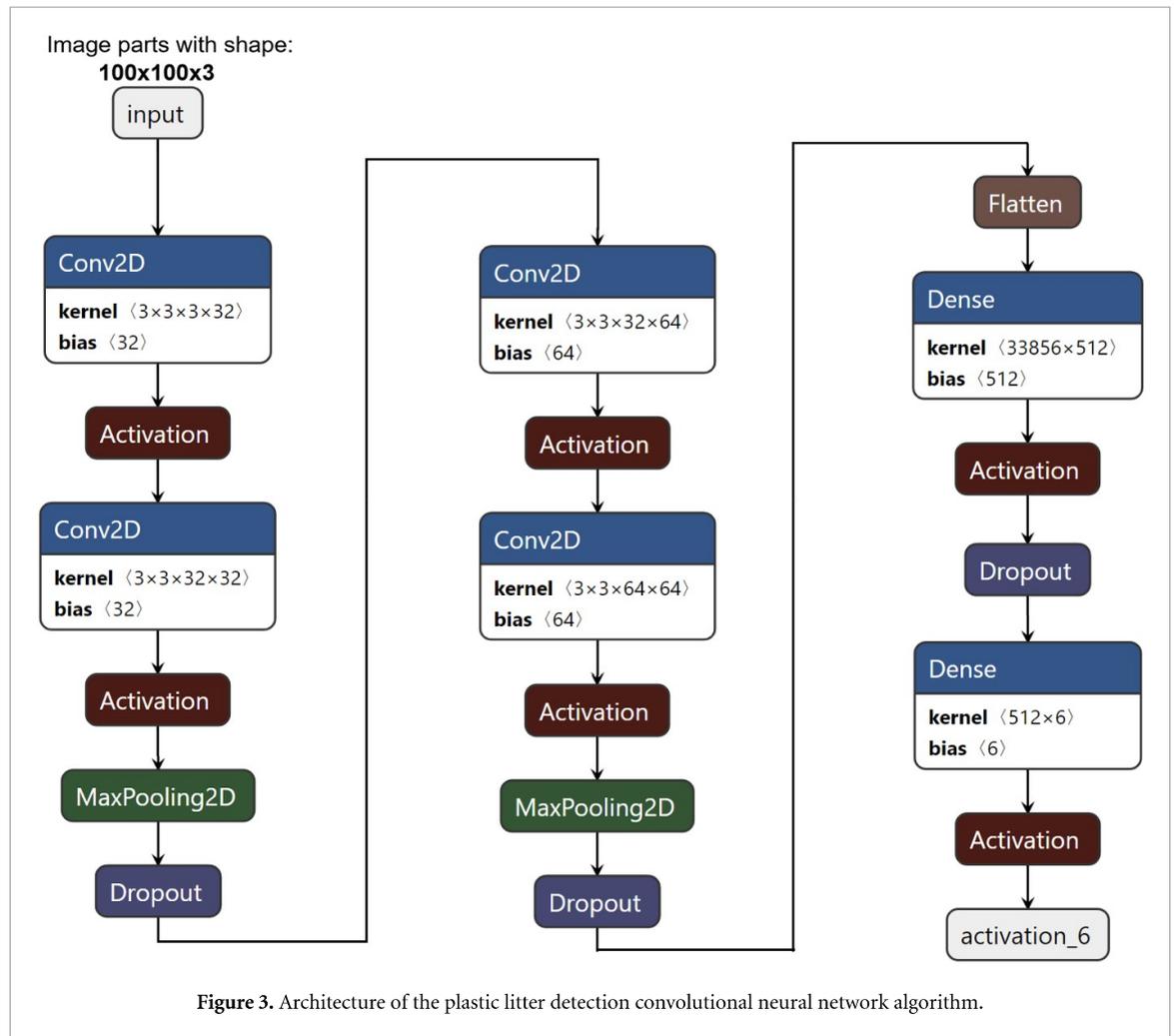
In addition to the training of the PLQ-CNN mentioned before, three further PLQ-CNNs have been trained with cost sensitive learning to mitigate the effects of class imbalances of the PLQ dataset. During the training cycles of these PLQ-CNNs, identical architecture and training parameters of PLQ-CNN were used. However, class weights were enabled during training to increase the training loss for samples of underrepresented classes. Initially, we experimented with a balanced weight factor which was calculated for each class as the fraction of number of samples for the class with the highest number of tiles: Number of samples for each class. However, these class weights resulted in harming the training process by



**Figure 2.** Flowchart of image processing by APLASTIC-Q system. The end-products of PLD-CNN and PLQ-CNN include the counts of different classes of litter objects.

getting stuck in a local minimum in the first epoch and not training any further. Hence, for training of the further PLQ-CNNs, these balanced class weights have been potentiated with the values 0.2, 0.4 and 0.6, respectively to consider both, class imbalances and to mitigate harmful effects for training.

The machine learning components of APLASTIC-Q (figure 2) can be exchanged with CNNs of different architectures or training parameters or with other machine learning methods. Besides the described CNNs, we investigated the overall accuracy on the PLD and PLQ datasets for three types of SVM



kernels: radial basis function, polynomial and linear. Moreover, Random Forest classifiers with 100 estimators were examined. Our PLD-CNN and PLQ-CNN have been compared with the aforementioned machine learning methods, because they are the best performing classification methods used in a marine debris detection study on beaches (Acuña-Ruz *et al* 2018).

#### 2.4. Final augmentations

The accuracy in counts from PLD-CNN and PLQ-CNN was enhanced by considering the true flight altitude at which imagery was collected. An altitude correction factor ( $F_{ac}$ ) was derived as a ratio of the true geo-spatial resolution ( $True_{gsr}$ ): estimated geo-spatial resolution ( $Estimated_{gsr}$ ) of  $0.002 \text{ m pixel}^{-1}$  (equation 1).

$$F_{ac} = \frac{True_{gsr}^2}{Estimated_{gsr}^2} \quad (1)$$

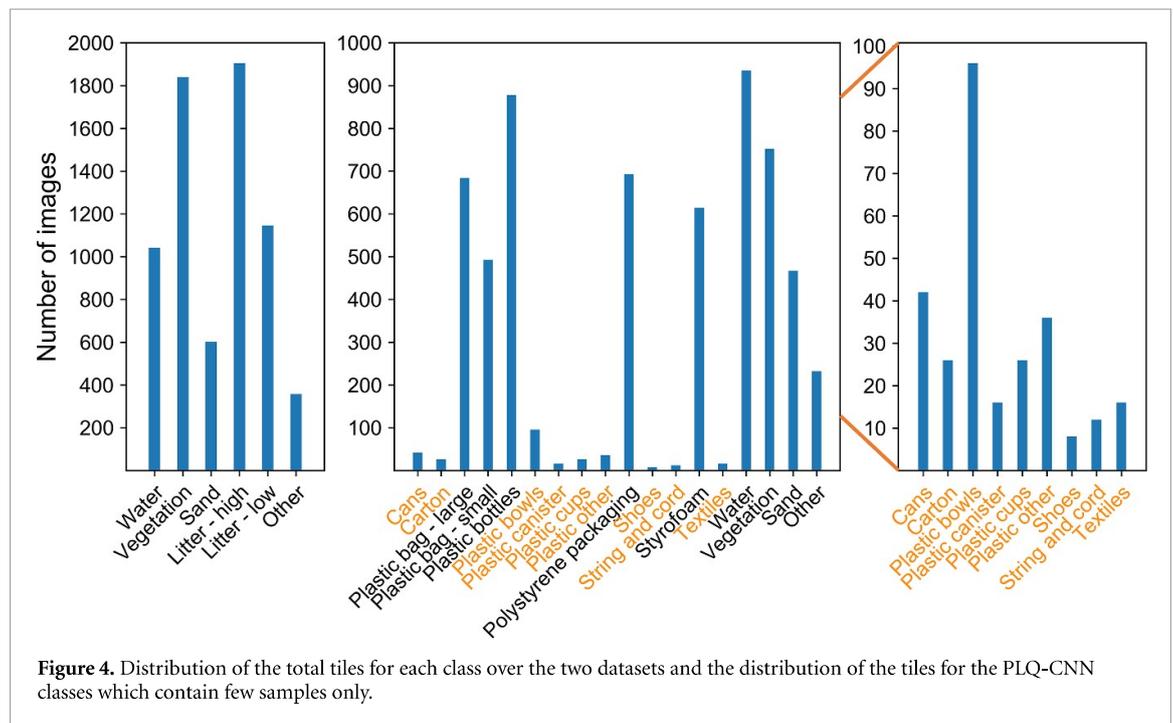
A corrected number of litter objects was then derived by multiplying PLD-CNN counts as follows:  $1.5 \times$  litter-low and  $3.5 \times$  litter-high. Following our threshold levels, litter-low had a limit of 2 items, thus the average of 1 or 2 objects was 1.5 items per tile.

Litter-high was set for object counts of more than 3, and we found out that each tile had a maximum of 4 visible items thus the average was 3.5 items per tile. The PLQ-CNN algorithm involves an additional correction for the average size of objects in the identified pollutant classes. A tile of size  $50 \times 50 \times 3$  pixels covered an area  $\sim 0.01 \text{ m}^2$ . Thus, if a tile section consists of a plastic bottle (size  $\sim 0.02 \text{ m}$ ) extending into a second tile, and this bottle is also covered by some other objects, it would be assumed to be half a bottle after correction in one tile portion. These adjusted counts of the 14 pollutant classes are partly translated into the United States National Oceanic and Atmospheric Administration agency classification system (GESAMP 2019).

### 3. Results

#### 3.1. Litter class distribution in datasets

PLD-CNN dataset was composed of 6892 tiles, a total of 1905 tiles contained high amounts of plastic litter were grouped as litter-high. Other targets that included building structure were found in a total of 357 tiles. The dataset of the PLQ-CNN had 6026 tiles. Plastic bottles were found in most tiles (878), followed by plastic bags, Styrofoam and polystyrene packaging



**Figure 4.** Distribution of the total tiles for each class over the two datasets and the distribution of the tiles for the PLQ-CNN classes which contain few samples only.

found in at least 400 tiles for each item class (figure 4). Vegetation, sand and water were found in several tiles consistent with the fact that the aerial images were collected over waterways and beaches generating more than 400 tiles for each class. During the creation of both datasets, tiles have been selected to represent the individual classes in varied shapes, sizes, and colours. For example, the water class contains tiles which cover different types of turbidity or contain sun glitter and the Styrofoam class contains items with different degrees of weathering and various shapes. However, for underrepresented classes in the PLQ dataset, this could not always be ensured. We believe more data would be needed to improve identification and quantification of the items with low counts such as shoes, cans, plastic canisters, textiles, strings, cords and cartons.

### 3.2. Performance assessment of PLD-CNN and PLQ-CNN

Statistics derived from PLD-CNN suggested promising applications for general classification of floating or washed ashore litter with a precision greater than 0.67 (table 1). Water, vegetation, sand and litter-high were easy to identify due to their inherent abundance, shape, form and colour which attributed to the good statistical results obtained (precision, recall,  $F1$ -score > 0.81).

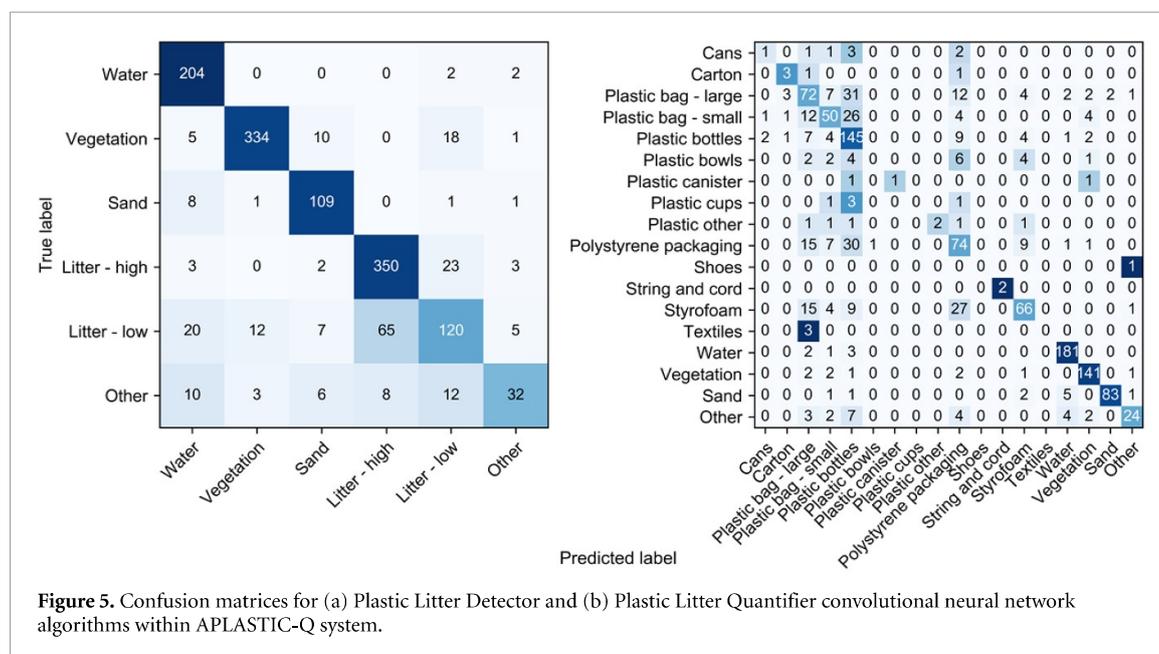
We therefore assume that with moderate to high amounts of plastic litter PLD-CNN will perform with good accuracy and high recall. Although the precision was lower for other classes and few litter objects, our algorithm was still capable of identifying the plastics with reasonable recall. Despite challenges in differentiating between litter-high and litter-low, PLD-CNN managed to correctly identify plastic litter (figure 5).

**Table 1.** Precision, recall and  $F1$ -scores for each class of the PLD-CNN.

Class	Precision	Recall	$F1$ -score	Number of tiles
Water	0.82	0.98	0.89	208
Vegetation	0.95	0.91	0.93	368
Sand	0.81	0.91	0.86	120
Litter-high	0.83	0.92	0.87	381
Litter-low	0.68	0.52	0.59	229
Other	0.73	0.45	0.56	71

For each class, the PLD-CNN algorithm performed with good accuracy, good precision and an  $F1$ -score of at least 0.55 in discriminating the targets into water, vegetation, sand, plastic litter, and others. For both ‘pollution’ classes merged the precision was high with 0.92. It was a similar scenario for the recall, it was generally higher than 0.9 for all classes except for ‘other’ (0.45) and litter-low (0.52). The overall recall of both plastic classes is 0.77, indicating that a large proportion of plastic in the images was detected. It indicated also that tiles polluted with plastic were found at a precision of 0.92 and a recall of 0.77. The low recall and precision values of class ‘other’ were presumed negligible in cases or scenarios of low counts within this category.

The PLQ-CNNs with the class weight exponents 0, 0.2, 0.4 and 0.6 achieved an overall accuracy of 71%, 71%, 66% and 60% respectively on the test dataset, the results are shown in the supplementary material (available online at [stacks.iop.org/ERL/15/114042/mmedia](https://stacks.iop.org/ERL/15/114042/mmedia)). The decrease of the overall accuracy with the increase of the class weight exponent was expected. This is because during training, the focus on classes with many tiles shifts to classes with fewer tiles. Classes



**Figure 5.** Confusion matrices for (a) Plastic Litter Detector and (b) Plastic Litter Quantifier convolutional neural network algorithms within APLASTIC-Q system.

with few tiles did not achieve a better *F1*-score using PLQ-CNN trained with an increased weight exponent. The only exception is for weight exponent 0.2, here some plastic litter types were classified with slightly higher accuracy, showing the highest *F1*-scores for plastic bags, plastic bottles and plastic bowls. A potential reason for PLQ-CNNs not performing better is the already high interdependency of classes with high weights due to few training tiles being available. Based on the overall accuracy and *F1*-score metrics for classes, the PLQ-CNN with class weight exponent 0 was selected. It achieved the highest overall accuracy and outperformed the other PLQ-CNNs on the *F1*-score for specific classes nine times.

The selected PLQ-CNN managed to identify and count the litter objects with a reasonable performance. It had a bias towards plastic bottles with moderate precision = 0.55 and high recall = 0.83, possibly due to overfitting as plastic bottles were common in over 878 tiles of the images analysed. Shoes, textiles, plastic bowls and plastic cups were not detected likely due to the limited quantities in the collected tiles. Additionally, many pollutant classes were not always found by the algorithm as they shared some resemblances in shape and form compared with other pollution classes, such classes were often mixed up. In some cases, plastic cups were classified as plastic bottles whilst plastic bowls and Styrofoam were identified as polystyrene packing. Generally, tiles containing pollution types were rarely mixed up with the four pollution free classes (see figure 5).

**3.3. Comparison with other machine learning techniques**

Both PLD-CNN and PLQ-CNN performed significantly better than other machine learning techniques which have been investigated in an anthropogenic

**Table 2.** Overall accuracy for alternative machine learning techniques, including the PLD-CNN and PLQ-CNN.

Dataset	PLD	PLQ
PLD-CNN/PLQ-CNN	<b>83%</b>	<b>71%</b>
SVM radial	78%	50%
SVM polynomial	63%	46%
SVM linear	47%	40%
Random Forest	71%	53%

marine debris study via satellite imagery (Acuña-Ruz et al 2018). The PLD-CNN achieved a 5% higher overall accuracy than the next best performing machine learning technique being SVM with radial basis kernel (table 2). PLQ-CNN outperformed the second best machine learning algorithm with 18%. A similar trend was observed for the performance of SVMs as in marine debris literature (Acuña-Ruz et al 2018): Radial basis function kernel achieved the best results, linear kernel the poorest. However, the Random Forest Classifier outperformed SVM with polynomial kernel and linear kernel for the PLD dataset and all SVMs in the PLQ dataset. The Random Forest Classifier was 26 times faster compared with PLD-CNN and and 14 times faster compared with PLQ-CNN. This indicates that a future investigation into ensemble models may be promising, due to being computationally more efficient than CNN approaches.

**3.4. Comparison with other automated marine litter detection systems**

The reported classification metrics are in line with recent work on automated marine litter detection on beaches of the Red Sea, the Maldives and Portugal (Martin et al 2018, Fallati et al 2019, Gonçalves et al 2020). To put our work in context, we compared the results of our PLD-CNN with aforementioned work.

**Table 3.** Comparison among different automated litter detection algorithms; an extended table of Gonçalves *et al* (2020). The values for Marine Litter, True Positive, False Negative and False Positive are the added values of our PLD-CNN results for the classes litter-high and litter-low. The metrics for precision, recall and *F1*-score for our work are the weighted averages for the two PLD classes. Reprinted from Gonçalves *et al* (2020), © 2019 Elsevier B.V. All rights reserved, with permission from Elsevier.

Reference	Method	Marine Litter	True Positive	False Negative	False Positive	Precision	Recall	<i>F1</i> -score
Martin <i>et al</i> (2018)	Random Forest (HoG)	415	164	251	1941	0.08	0.40	0.13
Fallati <i>et al</i> (2019)	CNN (pixelwise classification)	132	57.9	74	48.9	0.54	0.44	0.49
Gonçalves <i>et al</i> (2020)	Random Forest (four colour spaces)	159	118	41	44	0.73	0.74	0.75
Our work	CNNs (tile wise classification)	610	470	140	129	0.77	0.77	0.77

**Table 4.** Comparison of flight altitudes, camera resolution, drone models used and investigated scenes among these studies.

Reference	Flight altitude	Camera resolution	Drone model	Investigated scenes
Martin <i>et al</i> (2018)	10 m	12 MP	DJI Phantom 4	Beach, Red Sea
Fallati <i>et al</i> (2019)	10 m	12 MP	DJI Phantom 4	Beaches, Maldives
Gonçalves <i>et al</i> (2020)	20 m	20 MP	DJI Phantom 4 Pro	Beach, Portugal
Our work	6 m	20 MP	DJI Phantom 4 Pro	Beaches and riverine systems, Cambodia

We added the number of samples for litter-high and litter-low which resulted in the Marine Litter samples (table 3). Our results for True Positive, False Negative and False Positive were the added counts of these metrics for Litter-high and Litter-low. This means if a sample of Litter-high was misclassified as Litter-low, or vice versa, this would be counted as False Positive. Our results for precision, recall and *F1*-score are computed through the weighted averages of these metrics for litter-high and litter-low presented in table 1. The comparison is displayed in table 3 which is an extended version of the table created by Gonçalves *et al* (2020). The weighted average metrics for precision, recall and *F1*-score of PLD-CNN outperform the current best performing automated plastic detection systems by 0.04 (precision), 0.03 (recall) and 0.02 (*F1*-score). Moreover, compared with other work in the literature, the number of samples available to the here presented study is larger.

Nevertheless, we have to point out that such comparisons need to be viewed with caution because the settings of the mentioned studies are different, meaning that flight altitudes vary from 6 to 20 m and the camera resolutions vary from 12 to 20 MP (table 4). But most importantly, the investigated scenes in this study comprise beaches, rivers and even plastic carpets and are therefore quite different from the studied scenes in the literature. On the one hand, the fact that the scenes in our work are very diverse makes the detection of plastics harder. This is because the classification algorithms need to learn feature representations for multiple environments. On the other hand, the geo-spatial resolution used in this study was the highest compared with studies in the literature, which should generally help to ease the plastic detection. Because of the many factors which need to be

considered when comparing different automated detection systems, including performance metrics, speed of algorithm, utilized imagery and geo-spatial resolution or difficulties through environment, we raise the need for a framework which enables a fair comparison between works in this area.

### 3.5. Prospects of machine learning in monitoring plastic litter

Our novel APLASTIC-Q system consisting of PLD-CNN and PLQ-CNN components is capable to identify and quantify litter objects with high precision. In polluted tiles PLQ-CNN was able to detect five major pollutant classes (i) plastic bag—large, (ii) plastic bag—small, (iii) plastic bottles, (iv) polystyrene packaging and (v) Styrofoam. These top five litter items were very common in the imagery collected during the survey (figure 4). The results of our APLASTIC-Q system complemented by coordinated scientific field survey and clean-up efforts could produce benchmark information crucial for policymakers and stakeholders to pinpoint problematic plastic types polluting the natural environment. Concurrent field surveys have been conducted to this drone survey, however these were not always aligned with the drone surveys. Therefore, results of the field survey, which investigated 1 m<sup>2</sup> in regular intervals, could not be used as ground truth since the examined sections of the images could not be clearly assigned. We plan overcome these issues in future survey projects. Furthermore, the counts can be used for operational monitoring, including installed cameras, and creating baseline parameters for measuring the efficacy of waste and plastic management policies.

We encountered shadows in captured imagery that introduced biases in our PLQ-CNN algorithm. It

was biased towards plastic bag—large, plastic bag—small, plastic other or plastic bottle. However, in the bright sections of the shadowed image, the PLQ-CNN was more inclined to classify litter as polystyrene packaging or Styrofoam. These two classes consisted of white coloured items that could have played a role in biasing the algorithm. As for PLD-CNN, in tiles identified as litter-low we observed more items than prescribed in the thresholding steps. Fortunately, an additional run of the tiles marked as litter-low in PLQ-CNN proved to resolve this problem.

Training neural networks with imagery contaminated with shadows or surface reflected glitter has been considered problematic (Martin *et al* 2018, Fallati *et al* 2019). We therefore recommend imagery collection during optimal meteorological weather conditions and nadir viewing angles to avoid glitter and shadows. Inevitable shadow could be mitigated in imagery by applying filters such as gamma correction or statistical algorithms (Silva *et al* 2018, Xue *et al* 2019). Additional metadata (flight height, humidity, temperature, aerosol content, wind speed, geolocation) would also be useful to improve atmospheric correction of images and to accurately determine geo-spatial resolution. Future surveys are expected to provide more data to better train our APLASTIC-Q system thus improving its accuracy and prospective application in different regions of the world.

Although, RGB true colour images have shown benefits in automated monitoring of plastic litter, we propose to further explore the value in multispectral imaging technologies. Multispectral sensors with wavebands in the near infrared spectrum could further improve the identification of plastics in the natural environment. Already, studies using hyperspectral sensor technologies have confirmed diagnostic wavebands of plastics in the near to short-wave infrared spectrum (Kühn *et al* 2004, Garaba *et al* 2018, Garaba and Dierssen 2020). We assume that with these additional vectors of information from multispectral imagery, APLASTIC-Q can better differentiate litter in comparison with the three waveband RGB imagery. Furthermore, ocean colour remote sensing approaches might benefit especially in discriminating bright targets such as whitecaps, sea foam and breaking waves from plastics (Martínez-Vicente *et al* 2019, Dierssen and Garaba 2020). Therefore, future research is expected to explore various CNN architectures and involve hyper-parameter tuning. We suggest future works to explore benefits of tailored loss functions that could penalize more if pollution classes get mixed up with non-pollution classes and less if non-pollution classes get mixed up. Moreover, we echo the need for standardized data acquisition guidelines for UAS surveys to mitigate the challenges highlighted in our study.

## 4. Conclusions

We presented a novel machine learning system that performed reasonably well in identifying and quantifying floating and washed ashore plastic litter in terms of covered areas and counts of litter items. Our automated detection and quantification algorithms contribute towards monitoring strategies that complement counts from net trawl, field surveys, clean-up efforts and numerical distribution solutions. Combining the end-products of APLASTIC-Q, we believe we add value to scientific evidence-based knowledge that is important in the decision-making and legislation by policymakers, stakeholders including citizens in improving the blue economy as well as health of the blue planet. Litter estimates from our algorithms echo the need for upscaling monitoring efforts in developing nations. The vertical distribution of the plastic litter especially in the aquatic environment cannot be easily resolved from our algorithm. However, extrapolating the horizontal distributions of litter in different locations can be applied if basin types and prior *in situ* surveys are conducted, an important source of information urgently required in regions with plastic river carpets. There is a need to expand the capabilities of our algorithms to satellite imagery from very high geo-spatial resolution capable missions such as PlanetScope, Skysat, Pleiades and WorldView missions.

## Acknowledgments

We are thankful for the support from Key Consultants Cambodia (KCC) for the aerial survey, GoGreen Cambodia and Grace Smith for the field surveys and detailed true colour images documenting plastic litter. Funding was from the World Bank Group, Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Projektnummer 417276871. Ministry of Science and Culture of Lower Saxony and the VolkswagenStiftung through ‘Niedersachsen Vorab’ (ZN3480), German Ministry of Economic Cooperation and Development (BMZ) acting through the German International Cooperation (GIZ).

## Author contributions

KvdB, SPG, KS, OZ conceived and developed the study. SPG wrote the final manuscript based on the original draft by MW, FS. MW developed APLASTIC-Q under supervision from OZ, FS. MW, FS, NG, OZ validated APLASTIC-Q. MW, NG, SPG prepared the figures. All authors reviewed and approved the manuscript text.

## Disclaimer

The information in this paper reflects the views of the authors and does not reflect the official views of the World Bank, its Executive Directors, or the countries they represent.

## Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

## ORCID iDs

Mattis Wolf  <https://orcid.org/0000-0003-3863-8467>

Shungudzemwoyo P Garaba  <https://orcid.org/0000-0002-9656-3881>

Nina Gnann  <https://orcid.org/0000-0001-8117-7673>

Frederic Stahl  <https://orcid.org/0000-0002-4860-0203>

Oliver Zielinski  <https://orcid.org/0000-0002-6018-5030>

## References

- Abadi M et al 2016 TensorFlow: large-scale machine learning on heterogeneous distributed systems arXiv:1603.04467v2 pp 1–19
- Acuña-Ruz T, Uribe D, Taylor R, Amézquita L, Guzmán M C, Merrill J, Martínez P, Voisin L and Mattar B C 2018 Anthropogenic marine debris over beaches: spectral characterization for remote sensing applications *Remote Sens. Environ.* **217** 309–22
- Bak S H, Hwang D H, Kim H M and Yoon H J 2019 Detection and monitoring of beach litter using UAV image and deep neural network ISPRS - *Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-3/W8* pp 55–58
- Balint P J, Stewart R E, Desai A and Walters L C 2011 *Wicked Environmental Problems: Managing Uncertainty and Conflict* (Washington, DC: Island Press) p 272
- Bao Z, Sha J, Li X, Hanchiso T and Shifaw E 2018 Monitoring of beach litter by automatic interpretation of unmanned aerial vehicle images using the segmentation threshold method *Mar. Pollut. Bull.* **137** 388–98
- Blettler M C M, Abrial E, Khan F R, Sivri N and Espinola L A 2018 Freshwater plastic pollution: recognizing research biases and identifying knowledge gaps *Water Res.* **143** 416–24
- Chollet F et al 2015 Keras: the python deep learning library ([https://keras.io/examples/cifar10\\_cnn/](https://keras.io/examples/cifar10_cnn/)) (Accessed 24 March 2020)
- Dierssen H M and Garaba S P 2020 Bright Oceans: spectral differentiation of whitecaps, sea ice, plastics, and other flotsam *Recent Advances in the Study of Oceanic Whitecaps: Twixt Wind and Waves* ed P Vlahos and E C Monahan (Berlin: Springer)
- EUPPD 2019 Directive 2019/904 of the European Parliament and of the council on the reduction of the impact of certain plastic products on the environment *Off. J. Eur. Union* **L155** 1–19
- Fallati L, Polidori A, Salvatore C, Saponari L, Savini A and 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
- Galgani F, Hanke G, Werner S and De Vrees L 2013 Marine litter within the European Marine Strategy Framework Directive *ICES J. Mar. Sci.: Journal Du Conseil* **70** 1055–64
- Garaba S P, Aitken J, Slat B, Dierssen H M, Lebreton L, Zielinski O and Reisser J 2018 Sensing ocean plastics with an airborne hyperspectral shortwave infrared imager *Environ. Sci. Technol.* **52** 11699–707
- Garaba S P and Dierssen H M 2020 Hyperspectral ultraviolet to shortwave infrared characteristics of marine-harvested, washed-ashore and virgin plastics *Earth Syst. Sci. Data* **12** 77–86
- GESAMP 2019 Guidelines on the monitoring and assessment of plastic litter and microplastics in the ocean. (IMO/FAO/UNESCO-IOC/UNIDO/WMO/IAEA/UN/UNEP/UNDP/ISA Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection) *Report and Studies GESAMP No. 99* ed P J Kershaw, A Turra and F Galgani (London: International Maritime Organization)
- Gonçalves G, Andriolo U, Pinto L and Bessa F 2020 Mapping marine litter using UAS on a beach-dune system: a multidisciplinary approach *Sci. Total Environ.* **706** 135742
- Hohn S, Acevedos-Trejos E, Abrams J F, de Moura J F, Spranz R and Merico A 2020 The long-term legacy of plastic mass production *Sci. Total Environ.* **746** 141115
- IOC-UNESCO 2019 The science we need for the ocean we want: the United Nations Decade of Ocean Science for Sustainable Development (2021–2030) *IOC Brochure (IOC/BRO/2018/7 Rev)* Paris, France
- Jambeck J R, Geyer R, Wilcox C, Siegler T R, Perryman M, Andrady A, Narayan R and Law K L 2015 Plastic waste inputs from land into the ocean *Science* **347** 768–71
- Kataoka T, Hinata H and Kako S I 2012 A new technique for detecting colored macro plastic debris on beaches using webcam images and CIELUV *Mar. Pollut. Bull.* **64** 1829–36
- Kataoka T, Murray C C and Isobe A 2018 Quantification of marine macro-debris abundance around Vancouver Island, Canada, based on archived aerial photographs processed by projective transformation *Mar. Pollut. Bull.* **132** 44–51
- Kingma D P and Ba J 2015 Adam: a method for stochastic optimization *3rd Int. Conf. on Learning Representations (San Diego, CA, USA, May 7–9)* ed Y Bengio and Y Lecun pp 1–15
- Krizhevsky A 2009 *Learning multiple layers of features from tiny images* Master of Science Thesis, University of Toronto
- Krizhevsky A, Sutskever I and Hinton G E 2012 Imagenet classification with deep convolutional neural networks *Advances in Neural Information Processing Systems, 26th Annual Conf. on Neural Information Processing* ed F Pereira, C J C Burges, L Bottou and K Q Weinberger
- Kühn F, Oppermann K and Hötig B 2004 Hydrocarbon Index – an algorithm for hyperspectral detection of hydrocarbons *Int. J. Remote Sens.* **25** 2467–73
- Kylili K, Kyriakides I, Artusi A and Hadjistassou C 2019 Identifying floating plastic marine debris using a deep learning approach *Environ. Sci. Pollut. Res.* **26** 17091–9
- Lebreton L C M, van der Zwet J, Damsteeg J-W, Slat B, Andrady A and Reisser J 2017 River plastic emissions to the world's oceans *Nat. Commun.* **8** 15611
- Lebreton L et al 2018 Evidence that the Great Pacific Garbage Patch is rapidly accumulating plastic *Sci. Rep.* **8** 4666
- Lecun Y, Bengio Y and Hinton G 2015 Deep learning *Nature* **521** *Nature* 521 (7553), S. 436–444. *Processing Systems (Lake Tahoe, Nevada, USA, December 3–6)* (Curran Associates, Inc.) pp 436–44
- Martin C, Parkes S, Zhang Q, Zhang X, McCabe M F and Duarte C M 2018 Use of unmanned aerial vehicles for efficient beach litter monitoring *Mar. Pollut. Bull.* **131** 662–73
- Martínez-Vicente V et al 2019 Measuring marine plastic debris from space: initial assessment of observation requirements *Remote Sens.* **11** 2443

- Maximenko N *et al* 2019 Toward the Integrated Marine Debris Observing System *Front. Mar. Sci.* **6** 447
- Perez L and Wang J 2017 The effectiveness of data augmentation in image classification using deep learning arXiv: [1712.04621](https://arxiv.org/abs/1712.04621)
- Recuero Virto L 2018 A preliminary assessment of the indicators for Sustainable Development Goal (SDG) 14 “Conserve and sustainably use the oceans, seas and marine resources for sustainable development” *Mar. Policy* **98** 47–57
- Ryabinin V *et al* 2019 The UN decade of ocean science for sustainable development *Front. Mar. Sci.* **6** 470
- Sethy S, Sothun C and Wildblood R 2014 Municipal solid waste management in Cambodia *Municipal Solid Waste Management in Asia and the Pacific Islands: Challenges and Strategic Solutions* ed A Pariatamby and M Tanaka (Berlin: Springer)
- Shorten C and Khoshgoftaar T M 2019 A survey on Image Data Augmentation for Deep Learning *J. Big Data* **6** 60
- Silva G F, Carneiro G B, Doth R, Amaral L A and de Azevedo D F G D 2018 Near real-time shadow detection and removal in aerial motion imagery application *ISPRS J. Photogramm. Remote Sens.* **140** 104–21
- Smail E A *et al* 2019 An introduction to the ‘Oceans and Society: blue Planet’ initiative *J. Oper. Oceanogr.* **12** S1–11
- Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R 2014 Dropout: A simple way to prevent neural networks from overfitting *J. Mach. Learn. Res.* **15** 1929–58
- Todd P A, Ong X and Chou L M 2010 Impacts of pollution on marine life in Southeast Asia *Biodivers. Conserv.* **19** 1063–82
- van Emmerik T, Strady E, Kieu-Le T-C, Nguyen L and Gratiot N 2019 Seasonality of riverine macroplastic transport *Sci. Rep.* **9** 13549
- Xue L, Yang S, Li Y and Ma J 2019 An automatic shadow detection method for high-resolution remote sensing imagery based on polynomial fitting *Int. J. Remote Sens.* **40** 2986–3007

