Developing a Segmentation Model for Microscopic Images of Microplastics Isolated from Clams

Ji Yeon Baek[1[0000−0003−3486−6972]], Maria Krishna de Guzman[2,3[0000−0002−5754−2041]], Ho-min Park[1,6[0000−0001−9937−8617]], Sanghyeon Park[1[0000−0002−9854−0085]] [⋆⋆], Boyeon Shin[2], Tanja Cirkovic Velickovic[2,3,4,5], Arnout Van Messem[1,7[0000−0001−8545−7437]], and Wesley De Neve[1,6[0000−0002−8190−3839]]

[1] Center for Biotech Data Science, Ghent University Global Campus, Incheon, South Korea
[2] Center for Food Chemistry and Technology, Ghent University Global Campus, Incheon, South Korea
[3] Department of Food Technology, Safety and Health, Ghent University, Ghent, Belgium
[4] Department of Chemistry, University of Belgrade, Serbia
[5] Serbian Academy of Sciences and Arts, Serbia
[6] IDLab, Department of Electronics and Information Systems, Ghent University, Ghent, Belgium
[7] Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium

{jiyeon.baek, mariakrishna.deguzman, homin.park, sanghyeon.park, boyeon.shin, tanja.velickovic, arnout.vanmessem, wesley.deneve}@ghent.ac.kr

Abstract. Microplastics (MP) have become a major concern, given the threat, they pose to marine-derived food and human health. One way to investigate this threat is to quantify MP found in marine organisms, for instance making use of image analysis to identify ingested MP in fluorescent microscopic images. In this study, we propose a deep learning-based segmentation model to generate binarized images (masks) that make it possible to clearly separate MP from other background elements in the aforementioned type of images. Specifically, we created three variants of the U-Net model with a ResNet-101 encoder, training these variants with 99 high-resolution fluorescent images containing MP, each having a mask that was generated by experts using manual color threshold adjustments in ImageJ. To that end, we leveraged a sliding window and random selection to extract patches from the high-resolution images, making it possible to adhere to input constraints and to increase the number of labeled examples. When measuring effectiveness in terms of accuracy, recall, and F_2-score, all segmentation models exhibited low scores. However, compared to two ImageJ baseline methods, the effectiveness of our segmentation models was better in terms of precision,

[⋆⋆] The first four authors contributed equally

$F_{0.5}$-score, F_1-score, and mIoU: U-Net (1) obtained the highest mIoU of 0.559, U-Net (2) achieved the highest F_1-score of 0.682, and U-Net (3) had the highest precision and $F_{0.5}$-score of 0.594 and 0.626, respectively, with our segmentation models, in general, detecting less false positives in the predicted masks. In addition, U-Net (1), which used binary cross-entropy loss and stochastic gradient descent, and U-Net (2), which used dice loss and Adam, were most effective in discriminating MP from other background elements. Overall, our experimental results suggest that U-Net (1) and U-Net (2) allow for more effective MP identification and measurement than the macros currently available in ImageJ.

Keywords: Deep learning · Environmental monitoring · Image segmentation · Microplastics.

1 Introduction

The production of plastics has increased rapidly since the 1940s, mainly due to the attractive properties of plastic goods (durable, lightweight, corrosion resistant) and inexpensive methods of manufacturing. At present, however, plastics have become a major environmental concern [4]. Specifically, in the marine environment, microplastics or MP (< 5 mm) are currently the most dominant form of aquatic plastic litter [5], originating from synthetic polymers that are primarily manufactured in small sizes (primary source) or from the degradation of large plastic fragments (secondary source) [25].

Due to their small size, MP can be ingested by marine organisms during feeding, which raises significant concerns regarding the safety of seafood [27]. As such, MP consumption by marine biota is typically investigated by extracting and isolating MP through filtration. Using a microscope, particles are then manually sorted and counted, often requiring the involvement of more than one researcher to avoid bias in the measurements. Since this method is labor intensive and time consuming, the MP-VAT (microplastics visual analysis tool) macro in ImageJ [24] was developed in 2019. With MP-VAT, the quantity, size, and shape of MP can be automatically and rapidly measured [18, 19].

Despite the increased throughput enabled by MP-VAT, this macro is prone to errors caused by background fluorescence and fluorescence halos from very bright particles [19]. In this context, additional corrections have to be made to the image beforehand (i.e., adjustment of the color threshold). This introduces an extra manual step in the process that is subjective and also time consuming. Hence, there is a need for a solution that performs better in terms of MP recognition and measurement.

In this study, we propose a deep learning-based approach that embeds the calibration criteria used by researchers, facilitating the automatic measurement of MP in terms of quantity, size, and shape. Specifically, we created an image segmentation model using deep neural networks, obtaining training data from high-resolution microscopic images of fluorescently-dyed MP isolated from clams.

2 Background

2.1 Measurement of microplastics in seafood using MP-VAT

Plastic litter has been transferred, directly or indirectly, to the marine environment due to poor waste disposal and management. Because of this, plastic debris is now ubiquitous in all areas of the ocean [1]. As such, MP could be ingested intentionally by marine biota when they are mistakenly seen as plankton or other prey, or accidentally by filter feeding [27]. Bivalves (mussels, clams, and oysters) are commonly used subjects for MP ingestion research since they are non-selective feeders, abundant worldwide, and easy to handle [27]. Aside from these observations, bivalves are usually eaten whole, making them suitable for the assessment of health risks associated with human consumption of food contaminated with MP [9].

Through visual sorting and counting using a microscope, MP measurement was initially performed manually. By making use of fluorescent staining and subsequent image processing in ImageJ, the throughput of MP measurement could be greatly improved, hereby also reducing bias [11]. Nonetheless, until 2019, image analysis still had to be mainly done manually [8]. With the introduction of the MP-VAT macro in ImageJ, the size, shape, and quantity of MP could also be measured automatically [19]. However, a major drawback is the presence of bright image areas, significantly hampering the effectiveness of MP-VAT. This issue is aggravated when images are taken under a microscope, which is the case in most laboratories. Indeed, as a result of high magnification, fluorescence halos and bright areas in the background become more prominent, leading to MP-VAT producing compromised results. Because of this, additional image editing is needed and optimal photography settings have to be determined.

To improve MP-VAT, MP-VAT 2.0 [18] was developed. This second macro eliminates white reflections by subtracting the green from the red channel, which improves the detection of red fluorescent particles. In addition, the color threshold method changed from maximum entropy to Renyi entropy [23]. However, as in the previous case, photographic conditions greatly affect the results, making it necessary to exactly replicate the conditions used by the developers to maximize the effectiveness of both MP-VAT and MP-VAT 2.0. Since there is limited flexibility in this aspect, there is room for an improved automated MP measurement approach based on images.

2.2 Image segmentation and deep learning

Image segmentation refers to the process of dividing an image into several meaningful sets of pixels. For example, in everyday life, segmentation is used to divide the pixels in a personal photograph into human and background pixels, and in the medical field, segmentation can be used to make a distinction between cancerous and non-cancerous pixels in magnetic resonance imaging (MRI) slice. In general, this is done by finding a threshold that can distinguish between the region of interest and the background. However, for images containing noise and occlusion, various conditions may have to be considered simultaneously. To solve this problem, many machine learning models have been developed [17]. Given the recent success of deep neural networks, several deep learning-based models have been proposed as well [3, 16, 20].

3 Dataset Acquisition

3.1 Wet-lab phase

Sampling Manila clam samples (*Ruditapes philippinarum*) were bought at the Incheon Complex Fish Market (Incheon, Korea) in May 2019. Immediately after the purchase, the samples were kept on ice. After transport to the laboratory, the clams were wrapped in aluminum foil and stored in a $-20\,^{\circ}\mathrm{C}$ freezer.

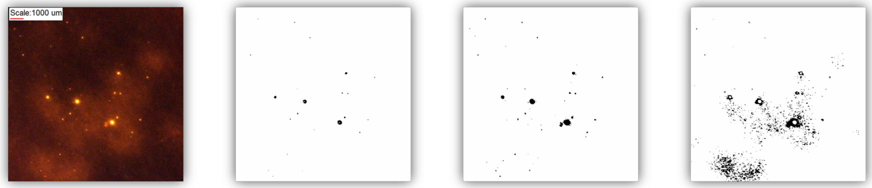
Extraction (digestion) and purification All steps were performed inside a clean bench to avoid contamination. Frozen whole clam tissue was separated from the shell and organic matter was dissolved by incubation in 250 mL 10 % KOH at $60\,^{\circ}\mathrm{C}$, applying stirring for 24 hours [6]. Through this digestion step, the microplastics that were inside the organism were obtained. Once digestion was completed, samples were vacuum filtered and MP were collected on GF/A (glass microfiber) filters. To separate MP from marine contaminants (sand and silt), filter papers were resuspended in 1.37 g/ml Zinc Chloride with sonication. After three repetitions, the solutions were centrifuged and the supernatant was filtered to recover MP [14].

Staining Nile Red (1 mg/mL in acetone) was added to the purified sample solution at a final concentration of 10 μg/ml. The dye was incubated with the sample at $50\,^{\circ}\mathrm{C}$ for 30 minutes with constant mixing. Solutions were vacuum filtered using a GF/A filter and excess dye was washed with absolute ethanol until minimal background fluorescence was achieved [14].

3.2 Dry-lab phase

Capturing and stitching Stained microplastics on filter paper were viewed under a stereomicroscope (Olympus SZX10) equipped with a fluorescence filter unit (RFP filter Ex 545-580 nm, Em 610 nm). Because the filter paper was too large to be captured as a single image, photos were taken by sections (Figure 1(a)). These sections were combined using Microsoft Image Composite Editor to create a complete image of the sample.

Binarization Each composite image was loaded in ImageJ. First, the scale was set (Analyze > Set Scale) and a representative fluorescent particle was selected to determine the optimal RGB threshold (Image > Adjust > Color threshold) for the automated selection of microplastics. When necessary, RGB values were adjusted manually to ensure the selection of all MP. Afterward, a “mask” image was generated (Edit > Selection > Create mask).



(a) Original image. (b) Ground truth mask. (c) MP-VAT. (d) MP-VAT 2.0.

Fig. 1. (a) A section of the filter paper image with stained MP. In this section, only the very bright particles are MP and the rest is background. (b) The corresponding mask image generated by color threshold adjustments and binarization in ImageJ. In this image, the background noise is removed and only MP are present. (c) The output of MP-VAT [19]. (d) The output of MP-VAT 2.0 [18].

Counting and measuring The MP-ACT macro [19] was used to automatically measure the size, shape, and quantity of MP in each mask (Plugins > Macros > MP-ACT). MP-ACT is an ImageJ macro that functions similar to MP-VAT, but that allows for manual color thresholding by the user, as opposed to the maximum entropy thresholding technique of MP-VAT.

4 Methods

As shown in Figure 1(c) and Figure 1(d), a major drawback of using the MP-VAT macro is the overestimation of MP. To prevent this from happening, researchers resort to manual color thresholding, as described in Section 3.2, which is time consuming, subjective, and prone to error. To solve this problem, we propose a deep learning-based approach that can generate a binarized image (mask) in a few seconds. This approach automatically learns the expert criteria for identifying MP using the 99 original images and their corresponding masks.

4.1 Problem definition

Given an image dataset $\mathcal{D} = (x_i, y_i)_{i=1}^n$, where $x_i \in \mathbb{R}^{l \times l}$ is a patch extracted from an original image and where $y_i \in \mathbb{R}^{l \times l}$ is its corresponding mask, let $\mathcal{M}$ be a segmentation model that takes x_i as input and that predicts $\hat{y}_i \in \mathbb{R}^{l \times l}$ as an approximation of the true mask y_i:

$$\mathcal{M}(x_i; \theta) = \hat{y}_i \approx y_i \,, \tag{1}$$

where θ are the model parameters. The difference between the predicted mask $\hat{y}_i$ and the ground truth mask y_i is quantified by a loss function L. Given a segmentation model $\mathcal{M}$ and a loss function L, we want to find the values for θ that minimize the total loss based on the dataset $\mathcal{D}$. Ideally, when a new patch x_j comes in, the trained model $\mathcal{M}$ makes a prediction $\hat{y}_j$ that is close to y_j.

4.2 Dataset characteristics

The dataset is composed of 99 original microscopic images, varying in size from $1,280\times960$ to $7,140\times5,424$ pixels, annotated with corresponding labels (masks). Considering the total number of pixels over all masks, 99.985 % of these pixels are background pixels, whereas 0.015 % of these pixels are MP pixels, pointing to a substantial imbalance in the type of pixels available. Hence, the original microscopic images and their corresponding masks were divided into five datasets in such a way that each dataset comes with a similar total number of MP pixels. As a result, 19 images were placed in Dataset 1, which was used as the test set, and 20 images were placed in each of the Datasets 2, 3, 4, and 5, and where the latter were used as training and validation sets.

Since each input image should come with a fixed size of $l \times l$ (see Section 4.1), which is a common requirement for machine learning methods, the 99 high-resolution images were cropped into patches with a fixed size of 256×256 pixels, coming with corresponding masks of the same size. Note that, before calculating any performance metrics, the original images (each with a predicted mask) need to be reconstructed from the patches in order to be able to correctly determine the MP quantity, size, and shape per filter paper image.

Patches are extracted by leveraging a sliding window method, using a step size of 30 pixels for both the x and y coordinates. A patch is saved if at least one of its pixels represents MP (as indicated in the corresponding mask). Denoting the number of patches with MP as N, then $\lceil N \times 0.05 \rceil$ additional patches without MP are saved, selecting the coordinates for cropping in a random way. A total of 157,474 patches were generated after cropping, with each dataset receiving on average $31,494.8$ ($\pm5,303.6$) patches, of which, on average, $30,003.8$ ($\pm5,051.8$) contain MP and $1,491.0$ (±251.8) do not contain MP. The number of patches per dataset ranged from $21,645$ to $36,508$.

Given that the number of patches differed from dataset to dataset, the different datasets were adjusted in order to each have a number of patches equal to 20,000. This adjustment was carried out by random deletion of patches, hereby keeping the number of patches without MP at approximately 5 % of the number of patches with MP. Thus, each dataset now consists of 20,000 patches, where, on average, $19,084.2$ (±13.3) patches contain MP, and 915.8 (±13.3) patches do not contain MP. Overall, 95.4 % of the 100,000 patches contain MP, while the remaining 4.6 % does not contain MP.

4.3 Model training

We used U-Net [20], a popular convolutional neural network, as our segmentation model $\mathcal{M}$. In particular, U-Net is an encoder-decoder model that was initially developed for medical image segmentation. For our experiments, we chose a U-Net model with a ResNet-101 [10] encoder, pre-trained on ImageNet [22]. A high-level Application Programming Interface (API) [28] has been utilized to build our segmentation model.

Optimization methods Gradient descent was used to update the parameters θ of our segmentation model in the negative direction of the gradient of the objective function $\nabla_\theta J(\theta)$, with the goal of minimizing the objective function $J(\theta)$, where $J(\theta)$ corresponds to the average loss. Two different optimization methods were used: (1) Stochastic Gradient Descent (SGD) and (2) Adaptive Moment Estimation (Adam). For both methods, the model parameters are optimized by using the objective function obtained over mini batches consisting of 10 patches.

Stochastic Gradient Descent The mathematical form of SGD is, for instance, described in the technical documentation of PyTorch[8]. Two hyperparameters of SGD need to be determined, namely the momentum λ and the learning rate α. λ allows accelerating the optimization performed by SGD with less fluctuation, whereas α determines the step size made when updating the parameters in each iteration [21]. These hyperparameters were set to 0.9 and 0.003, respectively.

Adaptive Moment Estimation When making use of Adam, the learning rate is computed individually for each parameter. Unlike SGD, which uses gradients directly, Adam utilizes exponentially moving averages of the gradient and the squared gradient [13]. The four hyperparameters of Adam, which are the exponential decay rates β_1 and β_2, the learning rate α, and the constant for numerical stability ϵ, were set to their default values: $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\alpha = 0.001$, and $\epsilon = 10^{-8}$.

Loss functions Our experiments used the three loss functions discussed below.

Binary cross-entropy (BCE) with logits loss BCE with logits loss is a modification of the regular BCE loss function [12], combining the sigmoid activation function and the BCE loss function into a single layer. The use of BCE with logits loss is numerically more stable than the use of the sigmoid activation function and the BCE loss function independently. To mitigate severe class imbalance, the hyperparameter p can be used to weigh the positive samples. In our experiments, p was set to 9.

Dice loss The dice coefficient, which is widely used for evaluating the effectiveness of segmentation models [26], is a measure for the overlap between a predicted and a ground truth segmentation map. This measure can be used as a loss function which is insensitive to data imbalance [15]. The hyperparameter ϵ, which avoids numerical issues when both the prediction and the ground truth are 0, was set to 1.

BCE with dice loss As discussed in [7], BCE and dice loss can be combined. The dice loss focuses on the similarity between the prediction and the ground truth at the image level, whereas the BCE loss focuses on the pixel-wise differences between the prediction and the ground truth.

[8] `https://pytorch.org/docs/stable/optim.html`

Model definition We created three different segmentation models, with all three models using a ResNet-101 encoder pre-trained on ImageNet: U-Net (1) uses BCE with logits loss and SGD; U-Net (2) uses dice loss and Adam; and U-Net (3) uses BCE with dice loss and Adam. The effectiveness of the three aforementioned segmentation models is compared to the effectiveness of the ImageJ macros MP-VAT and MP-VAT 2.0, with the latter two macros acting as our baselines.

Cross-validation Among the five available datasets, Dataset 1 was set aside as our test set. Four-fold cross-validation was performed using Datasets 2, 3, 4, and 5, meaning that in each round, one of the four available datasets was used as a validation set and the other three formed the training set. For each of the three segmentation models, four parameter sets were thus obtained. For each segmentation model, the parameter set that gave the lowest validation loss was selected as the optimal parameter set, leading to optimal models for U-Net (1), U-Net (2), and U-Net (3). Using different performance metrics, the effectiveness of these optimal models was then evaluated and compared to the effectiveness of the baselines MP-VAT and MP-VAT 2.0.

4.4 Performance metrics

To evaluate the effectiveness of our segmentation models, we used seven metrics: balanced accuracy, precision, recall, intersection over union (IoU), and three variations of the F_β-score ($F_{0.5}$-, F_1-, and F_2-score).

$$\text{Precision} = \frac{TP}{TP+FP} \qquad \text{Balanced Accuracy} = \frac{1}{2} \times \left(\frac{TP}{TP+FN} + \frac{TN}{FP+TN}\right)$$

$$\text{Recall} = \frac{TP}{TP+FN} \qquad \text{IoU} = \frac{|A \cap B|}{|A \cup B|} \qquad F_\beta\text{-score} = \frac{(1+\beta^2) \times (\text{precision} \times \text{recall})}{\beta^2 \times \text{precision} + \text{recall}}$$

Fig. 2. The performance metrics used. TP denotes the number of true positives, TN the number of true negatives, FP the number of false positives, and FN the number of false negatives. For IoU, A denotes the number of MP pixels in the ground truth mask, whereas B denotes the number of MP pixels in the predicted mask. In case of the F_β-score, β values of 0.5, 1, and 2 were used to calculate the $F_{0.5}$-, F_1-, and F_2-score, respectively.

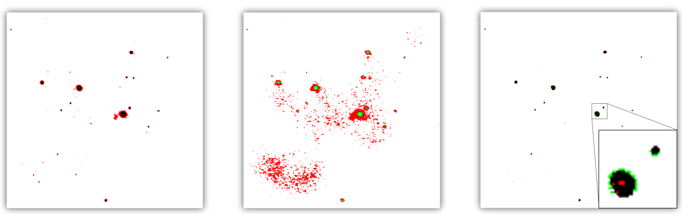
Note that all metrics, with the exception of IoU, are calculated using the average TP, TN, FP, and FN values, as obtained from four-fold cross validation. In this context, the obtained balanced accuracy value is referred to as the mean balanced accuracy (mAcc). On a similar note, the average IoU value, as obtained for the 19 images in the test set, is denoted as mean IoU (mIoU).

5 Results and Discussion

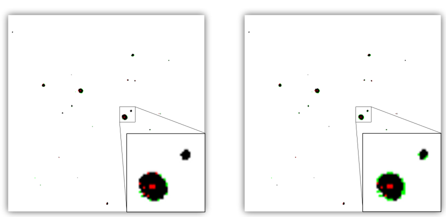
Our experimental results are summarized in Table 1. Overall, the effectiveness of MP-VAT 2.0 was found to be low for all performance metrics used. As can be seen in Figure 1(d), holes are prominently present in the center of several MP elements. Indeed, MP-VAT 2.0 seems to have removed the highest values in the image when subtracting both the green and blue channels. As a result, the center of several MP elements is missing in a mask, making it difficult for these MP elements to be properly identified.

Table 1. Performance evaluation.

Model	mAcc	Precision	Recall	mIoU	$F_{0.5}$-score	F_1-score	F_2-score
MP-VAT	0.986	0.431	0.971	0.392	0.485	0.597	0.777
MP-VAT 2.0	0.719	0.002	0.455	0.139	0.003	0.005	0.012
U-Net (1)	0.934	0.505	0.868	0.559	0.551	0.638	0.759
U-Net (2)	0.906	0.587	0.812	0.557	0.622	0.682	0.755
U-Net (3)	0.899	0.594	0.799	0.519	0.626	0.681	0.747



(a) MP-VAT. (b) MP-VAT 2.0. (c) U-Net (1): BCE with logits loss and SGD.



(d) U-Net (2): Dice loss and Adam. (e) U-Net (3): BCE with dice loss and Adam.

Fig. 3. Visual comparison of the segmentation quality obtained for each segmentation model: black denotes MP that has been correctly predicted (TP), red denotes predicted MP that is not MP (FP), and green denotes MP that has not been predicted (FN).

MP-VAT, on the other hand, comes with higher effectiveness, achieving an mAcc of 0.986, a recall of 0.971, and an F_2-score of 0.777. As shown in Figure 1(c), highly fluorescent MP are intact and visible in the mask. However, similar to MP-VAT 2.0, some areas of the background are also mistaken to be MP. Because of this, MP-VAT scores are low in terms of mIoU, $F_{0.5}$-score, F_1-score, and precision. In comparison, all U-Net-based models score higher with respect to these four metrics. U-Net (1) and U-Net (2) have the best scores for mIoU and F_1-score, respectively. U-Net (3) obtains the best scores for $F_{0.5}$-score and precision. The increase in precision using the U-Net-based models greatly reduces the number of FPs, thereby improving the discrimination between MP and background. This fact is also observed through the $F_{0.5}$-score, since all U-Net-based models obtain higher values than MP-VAT. Within the U-Net-based models, both U-Net (2) and U-Net (3) show better handling of FPs than U-Net (1). As illustrated in Figure 3(c), Figure 3(d), and Figure 3(e), the predicted masks are closer to the ground truth mask. Moreover, the impact of background fluorescence and fluorescence halos was considerably reduced, as compared to Figure 3(a) and Figure 3(b).

Based on the values obtained for the different metrics, it is not straightforward to select which U-Net-based model is most effective. However, U-Net (3) seems to be the least effective among the different models used, coming with the lowest mAcc, recall, F_2-score, and mIoU. In spite of these low scores, the generated masks, with an example shown in Figure 3(e), still demonstrate considerable improvement in discriminating MP from other elements in a fluorescent image.

In summary, our results show that the U-Net-based models are more effective at predicting MP, compared to MP-VAT and MP-VAT 2.0. Specifically, despite a lower balanced accuracy and recall, U-Net (1), U-Net (2), and U-Net (3) introduce significant gains in terms of mIoU, $F_{0.5}$-score, F_1-score, and precision.

6 Conclusions and Future Work

Effective and efficient MP identification and measurement in biological samples remain a challenge. Currently available methods that rely on image processing, such as MP-VAT and MP-VAT 2.0, come with limited flexibility and are prone to errors. In this study, we present a deep learning-based segmentation approach for MP measurement, serving as an alternative to MP-VAT and MP-VAT 2.0. Three variations of U-Net, all using a pre-trained ResNet-101 encoder, were tested in terms of generating masks that depict true MP.

Compared to the MP-VAT macros, all U-Net-based models exhibited better mIoU, $F_{0.5}$-score, F_1-score, and precision values, with the masks generated by these U-Net-based models containing less false positives. In this context, the effectiveness of U-Net (1) and U-Net (2) was found to be slightly better than the effectiveness of U-Net (3). Furthermore, since the U-Net-based models have a response time of less than two seconds after input of the original image, their usage is expected to significantly reduce the time and effort required to produce a mask, which takes 10 to 30 minutes per image in the original process using MP-VAT, including false MP correction.

In this initial research effort, the models tested were all based on U-Net. Future approaches will explore the use of other models, for instance paying attention to local thresholding methods [29]. In terms of deep learning, other state-of-the-art models such as DeepLabV3 [2] will be considered. Moreover, the U-Net-based models presented in this study will be further improved by modifying the model training strategy and enriching the training dataset used.

Acknowledgements The research and development activities described in this paper were funded by Ghent University Global Campus (GUGC) and by the Special Research Fund (BOF) of Ghent University (grant no. 01N01718).

References

1. Anbumani, S., Kakkar, P.: Ecotoxicological effects of microplastics on biota: a review. Environmental Science and Pollution Research **25**(15), 14373–14396 (2018)
2. Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv preprint arXiv:1706.05587 (2017)
3. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV) (2018)
4. Cole, M., Lindeque, P., Halsband, C., Galloway, T.S.: Microplastics as contaminants in the marine environment: A review. Marine Pollution Bulletin **62**(12), 2588–2597 (2011)
5. Cressey, D.: The plastic ocean. Nature **536**(7616), 263–265 (2016)
6. Dehaut, A., Cassone, A.L., Frère, L., Hermabessiere, L., Himber, C., Rinnert, E., Rivière, G., Lambert, C., Soudant, P., Huvet, A., et al.: Microplastics in seafood: Benchmark protocol for their extraction and characterization. Environmental Pollution **215**, 223–233 (2016)
7. Deng, R., Shen, C., Liu, S., Wang, H., Liu, X.: Learning to Predict Crisp Boundaries. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 562–578 (2018)
8. Erni-Cassola, G., Gibson, M.I., Thompson, R.C., Christie-Oleza, J.A.: Lost, but Found with Nile Red: A Novel Method for Detecting and Quantifying Small Microplastics (1 mm to 20 µm) in Environmental Samples. Environmental Science & Technology **51**(23), 13641–13648 (2017)
9. Galloway, T.S.: Micro-and Nano-plastics and Human Health. In: Marine Anthropogenic Litter, pp. 343–366. Springer, Cham (2015)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)
11. Isobe, A., Kubo, K., Tamura, Y., Nakashima, E., Fujii, N., et al.: Selective transport of microplastics and mesoplastics by drifting in coastal waters. Marine Pollution Bulletin **89**(1-2), 324–330 (2014)
12. Jadon, S.: A survey of loss functions for semantic segmentation. arXiv preprint arXiv:2006.14822 (2020)

13. Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980 (2014)
14. Maes, T., Jessop, R., Wellner, N., Haupt, K., Mayes, A.G.: A rapid-screening approach to detect and quantify microplastics based on fluorescent tagging with Nile Red. Scientific Reports **7**(1), 1–10 (2017)
15. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV. pp. 565–571. IEEE (2016)
16. Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D.: Image Segmentation Using Deep Learning: A Survey. arXiv preprint arXiv:2001.05566 (2020)
17. Murphy, K.P.: Machine learning: a probabilistic perspective. MIT press (2012)
18. Prata, J.C., Alves, J.R., da Costa, J.P., Duarte, A.C., Rocha-Santos, T.: Major factors influencing the quantification of Nile Red stained microplastics and improved automatic quantification (MP-VAT 2.0). Science of The Total Environment **719**, 137498 (2020)
19. Prata, J.C., Reis, V., Matos, J.T., da Costa, J.P., Duarte, A.C., Rocha-Santos, T.: A new approach for routine quantification of microplastics using Nile Red and automated software (MP-VAT). Science of The Total Environment **690**, 1277–1283 (2019)
20. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer International Publishing (2015)
21. Ruder, S.: An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747 (2016)
22. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV) **115**(3), 211–252 (2015). https://doi.org/10.1007/s11263-015-0816-y
23. Rényi, A.: On measures of entropy and information. In: Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics. pp. 547–561. The Regents of the University of California, University of California Press (1961)
24. Schneider, C.A., Rasband, W.S., Eliceiri, K.W.: NIH Image to ImageJ: 25 years of image analysis. Nature Methods **9**(7), 671–675 (2012)
25. Silva, A.B., Bastos, A.S., Justino, C.I., da Costa, J.P., Duarte, A.C., Rocha-Santos, T.A.: Microplastics in the environment: Challenges in analytical chemistry - A review. Analytica Chimica Acta **1017**, 1–19 (2018)
26. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Cardoso, M.J.: Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, pp. 240–248. Springer (2017)
27. Wesch, C., Bredimus, K., Paulus, M., Klein, R.: Towards the suitable monitoring of ingestion of microplastics by marine biota: A review. Environmental Pollution **218**, 1200–1208 (2016)
28. Yakubovskiy, P.: Segmentation Models. `https://github.com/qubvel/segmentation_models` (2019)
29. Yan, F., Zhang, H., Kube, C.R.: A multistage adaptive thresholding method. Pattern Recognition Letters **26**(8), 1183 – 1191 (2005). https://doi.org/https://doi.org/10.1016/j.patrec.2004.11.003, `http://www.sciencedirect.com/science/article/pii/S0167865504003290`