

# Plastic Waste on Water Surfaces Detection Using Convolutional Neural Networks

Yurii Kryvenchuk, Andrii Marusyk

Lviv Polytechnic National University, 12 Stepan Bandera Street, Lviv, 79013, Ukraine

## Abstract

This paper delves into the use of state-of-the-art convolutional neural network (CNN) architectures for automatically detecting plastic waste on water surfaces. The study extensively examines the efficiency of various CNN models for object detection, specifically targeting the identification of plastic waste. Prior to model training, extensive preprocessing of the dataset was conducted, which comprises imagery representing four distinct categories of plastic litter, namely 'plastic bags,' 'plastic bottles,' 'other plastic waste,' and non-plastic waste. Multiple configurations of YOLO (You Only Look Once) architecture models were trained either from inception or fine-tuned with diverse hyperparameters and varying numbers of epochs. The training process leveraged PyTorch framework and CUDA technology to enhance computational efficiency. Model assessment was conducted utilizing established CNN performance metrics, including precision, mean Average Precision (mAP), recall, and F1 score. The outcomes reveal superior performance of select models or models exhibiting promising results, substantiated by the evaluation metrics employed. Additionally, the study furnishes insights into the strengths and limitations of the trained models, accompanied by recommendations for refinement and avenues for future research.

## Keywords

Computer Vision, Object Detection, Convolutional Neural Networks, YOLO, Automatic Systems, Artificial Intelligence.

## 1. Introduction

In the contemporary era, the escalating issue of plastic pollution in water bodies demands immediate attention. This concern is underscored by the increasing prevalence of contaminated water bodies [1], with approximately 80% of oceanic plastic originating from terrestrial sources [2]. Upon entering water bodies, plastic undergoes gradual degradation, yielding microplastics that permeate consumable water sources. This contamination poses significant health risks by perturbing endocrine systems and precipitating adverse health outcomes [2]. To address this pressing environmental challenge, there is a growing reliance on advanced technologies, particularly in the pursuit of automated solutions.

This study acknowledges the substantial advancements in utilizing convolutional neural networks (CNNs) to address computer vision tasks, particularly in object detection. From the inception of rudimentary convolutional networks such as the seminal 'LeNet' (1998), to contemporary models like YOLO, the evolution in this field is remarkable and pivotal.

The aim of the study is creating a fast and precise model that can be used for creating complex autonomous plastic waste detection systems.

The goals of this work are to analyze state-of-the-art solutions which might be used or were used for solving object detection problem in similar domains, preprocess and prepare dataset, develop and evaluate models based on CNNs, make conclusions with experiment results and propose ideas for future studies and improvements.

---

COLINS-2024: 8th International Conference on Computational Linguistics and Intelligent Systems, April 12–13, 2024, Lviv, Ukraine

✉ yurii.p.kryvenchuk@lpnu.ua (Yu. Kryvenchuk); Andrii.Marusyk.KNM.2020@lpnu.ua (A. Marusyk)

ORCID 0000-0002-2504-5833 (Yu. Kryvenchuk); 0009-0004-5459-9896 (A. Marusyk)



© 2024 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

## 2. Related Works

Considering the pressing environmental issue of plastic pollution in water and recent advancements in machine learning, particularly in computer vision, over the past 2-3 years, research has focused on creating systems for automated plastic waste recognition. These endeavors encompass classical machine learning approaches as well as cutting-edge solutions that employ deep learning methodologies and usage of CNNs.

In this scientific study [2], the authors, Maharjan, N., Miyazaki, H., Pati, B.M., Dailey, M.N., Shrestha, S., Nakamura propose the utilization of deep convolutional neural networks for the automated recognition of plastic waste in rivers. A significant portion of their investigation involved constructing a dataset sourced from drone-collected data in Thai rivers. Following data preparation, various convolutional neural network (CNN) models were compared. The research findings indicate that the YOLOv5 (You Only Look Once) model with pre-trained weights exhibited superior performance, achieving a mAP50 (Mean Average Precision) of 0.71.

In another study [2], Gilroy Aldric Sio, Dunhill Guantero, Jocelyn Villaverde (2022), the authors utilized the YOLOv5 model, dedicating most of their efforts to establishing a data collection system employing a Raspberry Pi microcontroller equipped with a 5MP camera. Subsequently, they conducted data collection along rivers in the Philippines and created corresponding maps. Upon model training, their results yielded comparable accuracy metrics to those discussed in the aforementioned study; their model achieved an mAP50 value of 0.68.

In a different paper [3], the authors advocate for employing classical machine learning classifiers, specifically Random Forest and Support Vector Machine, in conjunction with feature engineering to pinpoint large clusters of plastic in satellite images of the ocean. Although this methodology yielded a commendable 80% accuracy for their dataset, it is limited in its capacity to identify individual pieces of debris, instead highlighting sizable clusters of white pixels within the images. Furthermore, the authors themselves acknowledge the potential for their algorithm to misinterpret plastic debris as white stones on water surfaces, deeming this approach unsuitable for addressing the research objectives outlined in this study.

In yet another study [4], Colin Lieshout, Kees Oeveren, Tim Emmerik, Eric Postma (2020), advocate for the utilization of convolutional neural networks to identify plastic debris in water bodies. They curated a dataset comprising 1,200 images to support their research endeavors. To enhance model performance, they implemented data augmentation techniques, resulting in a maximum accuracy of 68% (compared to 59% accuracy without data augmentation).

This study [5] offers a comprehensive analysis, focusing solely on the examination of existing literature. The authors reviewed over 30 articles pertaining to plastic waste detection utilizing convolutional neural networks. Through comparative analysis, three models emerged as noteworthy: InceptionResNetV2, VGG16 (Visual Geometry Group), and YOLOv5, with YOLOv5 demonstrating the highest accuracy rates. Most of reviewed studies center around the application of convolutional neural networks (CNNs) for plastic waste detection, highlighting the prominence of CNN models as state-of-the-art solutions in object detection.

While some studies explore classical machine learning methods, it is evident that these approaches are outdated and inadequate for the development of a complex system like automated plastic waste detection.

The related works analysis reveals several key findings. It suggests that optimal development of such a system involves utilizing a convolutional neural network based on the YOLOv8 architecture, given the superior efficiency demonstrated by previous versions of YOLO-type models compared to other CNNs examined in the studies. Effective training of the model requires a dataset of at least two thousand images to achieve accuracy rates of at least 65%. The primary metric for evaluating object recognition model accuracy is mAP, complemented by F1 score, Recall, and Confidence metrics. Additionally, data augmentation emerges as a crucial technique for enhancing model accuracy, particularly in scenarios with limited data availability.

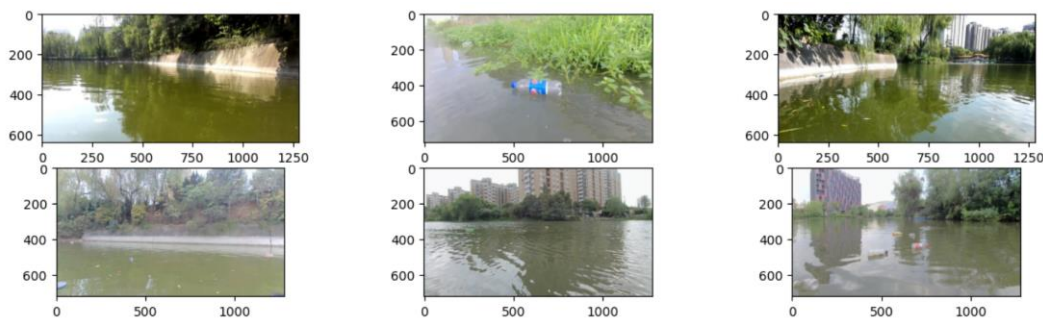
Considering insights gleaned from prior researches in automated plastic waste detection systems, our approach will involve developing a system based on modern convolutional neural

networks trained on an open dataset. This strategy aims to achieve at least the same or improved recognition accuracy results compared to reviewed researches.

### 3. Methods and Materials

#### 3.1. Dataset description

For this study, we selected the dataset “Kili Technologies: plastic\_in\_river” [7], accessible for download on the 'Hugging Face' platform. This dataset stands as the largest publicly available resource for identifying plastic waste, comprising 4259 images, each annotated with markings denoting plastic waste objects. The dataset was partitioned into three subsets: a training set containing 3407 images, a test set with 427 images, and a validation set comprising 425 images. Examples of images from training subset are shown in Figure 1.



**Figure 1:** Example images from the test dataset "Kili Technologies: plastic\_in\_river" [7]

The dataset comprises four distinct classes, denoted by numbers from 0 to 3, representing the following object types in sequential order: “plastic\_bag”, “plastic\_bottle”, “other\_plastic\_waste”, and “not\_plastic\_waste”. The images in the dataset possess high resolution, with widths exceeding 1000 pixels and heights surpassing 800 pixels. This dimensional aspect of the images enables experimentation with hyperparameters such as 'image size' during model training.

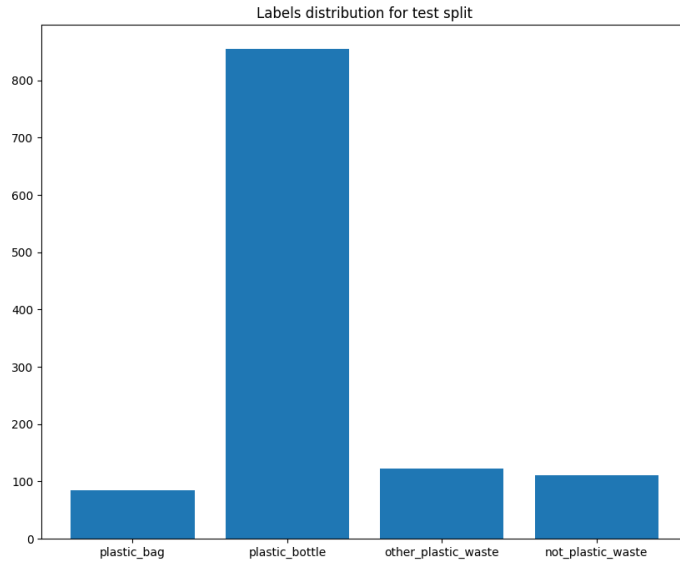
In Table 1. the number of images and text files with annotations for each data set is given.

**Table 1**

**Number of annotations by class**

Class	Train	Validation	Test
plastic_bag	1250	167	85
plastic_bottle	6276	785	854
other_plastic_waste	3345	296	122
not_plastic_waste	1414	212	111

A drawback of this dataset is the uneven distribution of objects among individual classes in the training dataset. As depicted in Figures 2, the number of 'plastic\_bottle' objects significantly outweighs those in the other classes. This imbalance has the potential to degrade the overall accuracy of the model.



**Figure 2:** Distribution of the number of objects of each of the classes in the training data set of the "Kili Technologies: plastic\_in\_river" dataset.

Among the advantages of this dataset, it is worth noting the variety of images presented: different lighting, different water bodies, and a large number of viewing angles that differ from each other.

From the analysis of this dataset, we conclude that despite the existing shortcomings, it still represents a valid dataset for training a convolutional neural network with the aim of achieving an accuracy of more than 65%.

### 3.2. Efficiency Metrics

The utilization of metrics to assess model performance is a critical facet of research in machine learning. Accurate evaluation metrics are essential for gauging the effectiveness of developed models in real-world scenarios. Therefore, this subsection offers an overview of key metrics employed in automatic object recognition tasks, which will inform the experimental phase of this study.

**Precision**, a fundamental metric, signifies the percentage of correctly classified positive cases among all cases identified as positive by the model. The formula for calculating precision is provided below:

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

where TP (true positive): the number of correctly classified positive cases; FP (false positive): the number of incorrectly classified negative cases.

Accuracy serves as a valuable indicator for minimizing false predictions by a model. However, relying solely on accuracy doesn't offer a complete assessment of model performance, as it overlooks errors of the second kind, specifically False Negatives.

**Recall**, another evaluation metric, gauges the percentage of correctly classified positive cases out of all true-positive cases within the dataset. By incorporating recall, the model can effectively mitigate errors of the second kind, notably False Negatives. The recall is computed using the following formula:

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

where TP (True positive) is the number of correctly classified positive cases. FN (False negative) is the number of incorrectly classified positive cases.

The **F1 metric** acts as a harmonic mean between precision and recall, offering a balanced assessment that safeguards against overfitting to a single type of problem during model training. This metric is particularly advantageous in scenarios like ours, where the dataset comprises unbalanced classes. Below is the formula for calculating the F1 metric.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

**mAP**, an abbreviation for mean Average Precision, stands as a pivotal metric in the evaluation of object detection within the realms of computer vision and machine learning. Revered as a standard benchmark in object recognition tasks, mAP signifies the average value of Average Precision calculated across all classes. Average Precision, in essence, encapsulates the area under the Precision-Recall (PR) curve for each individual class. Importantly, mAP encapsulates the nuanced fluctuations in accuracy with variations in the detection threshold.

### 3.3. Main Methods and Techniques

For this research, we opted for models rooted in the YOLO architecture: YOLO (You Only Look Once) stands as an immensely popular and potent family of algorithms for object recognition. YOLO models, especially versions 5, 7, and 8, represent state-of-the-art (SOTA) comprehensive solutions to computer vision challenges, particularly excelling in real-time object recognition scenarios. A key advantage of the YOLO model over other leading solutions in the convolutional neural network domain lies in its efficiency and speed, achieved by conducting object recognition through the CNN network in a single pass [6].

After conducting an in-depth analysis and gaining profound insights into the functionality of the YOLOv8 model, alongside considerations of the literature reviewed in the initial phase of this study, the decision was made to employ this model. Renowned for its distinctive and efficient architecture, coupled with commendable accuracy metrics and user-friendly software interface, the YOLOv8 model offers an optimal platform for conducting the requisite experiments aimed at addressing the objectives outlined in this research endeavor.

Moreover, a crucial technique employed in this study to enhance accuracy involves data augmentation. This approach aims to augment the training dataset, thereby introducing greater diversity of images. Transformations such as adjusting brightness levels, rotating images by small angles (up to 15 degrees), and modifying scale will be utilized to broaden the spectrum of training instances.

## 4. Experiment

Throughout the experimentation phase, the efficacy of networks founded on YOLOv8n and YOLOv8m architectures was trained and assessed. The investigation encompassed models trained from scratch, those refined through fine-tuning techniques, and variants with pre-trained weights.

A meticulous selection of hyperparameters was vital to attaining optimal outcomes. The chosen hyperparameters included:

- The number of epochs, with models trained over durations ranging from 20 epochs to more than 500 iterations.
- Learning rate, spanning values from 0.01 to 0.0001.
- Momentum, within the range of 0.9 to 0.988.
- Image size, which underwent compression to dimensions of 640 x 640 pixels, 704 x 704 pixels, 800 x 800 pixels, and 1008 x 1008 pixels.

The 'Adam' optimizer was selected based on a comprehensive analysis of literature, highlighting its widespread adoption and effectiveness in conjunction with YOLO models. Training sessions were executed on an NVIDIA GeForce RTX 2070 graphics processor utilizing CUDA technology, with an image batch size set to 16. This batch size optimization was

necessitated by the high-resolution nature of the original dataset images, ensuring efficient processing on the aforementioned graphics processor.

## 5. Results

Let's review the results from the convolutional neural network experiments mentioned earlier.

### 5.1. Experiment Results Using YOLOv8n

The first network was trained from scratch using the YOLOv8n architecture. The training process lasted for 100 epochs on low-resolution images with dimensions of 640 x 640 pixels. The complete list of hyperparameters is provided in Table 2 below.

**Table 2**  
**Hyperparameters of Model 1**

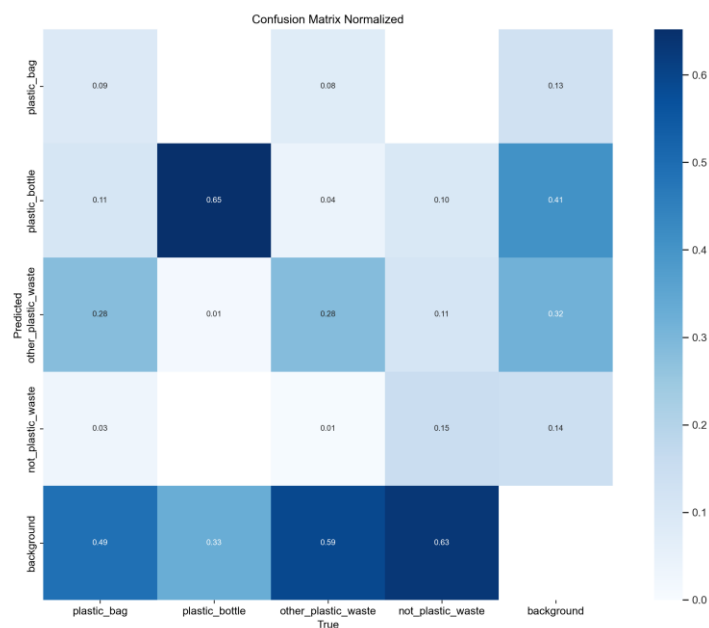
Epochs	Learning rate	Momentum	Image size	YOLO version	Is pre-trained
100	0.001	0.9	640	YOLOv8n	No

The training duration for this model was approximately 4 hours. The collected metrics for evaluating the performance of this model will be presented in Table 3.

**Table 3**  
**Efficiency Metrics of Model 1**

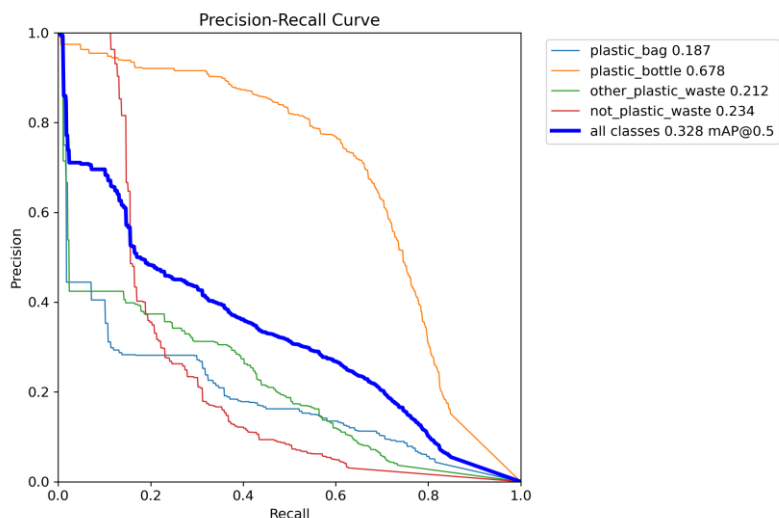
Class	Precision	Recall	mAP
all	0.388	0.395	0.327
plastic_bag	0.281	0.291	0.184
plastic_bottle	0.686	0.674	0.682
other_plastic_waste	0.25	0.409	0.208
not_plastic_waste	0.334	0.208	0.236

Table 3 demonstrates that this model achieves an mAP score of 0.327. Below, graphs are provided for visualizing the collected metrics.



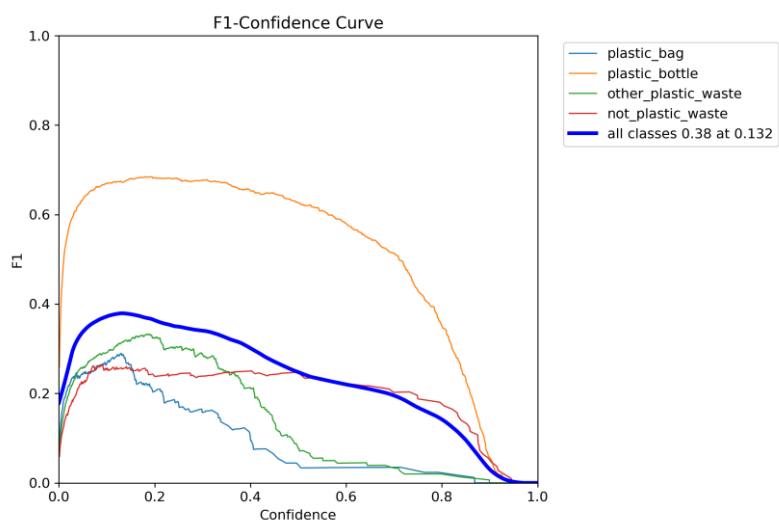
**Figure 3:** Confusion Matrix Not pre-trained YOLOv8n.

From Figure 3, it's evident that this model accurately identifies the 'plastic\_bottle' class in 65% of cases, whereas for the remaining classes, it achieves correct predictions only 20% of the time. This discrepancy is likely attributable to the dataset's imbalance.



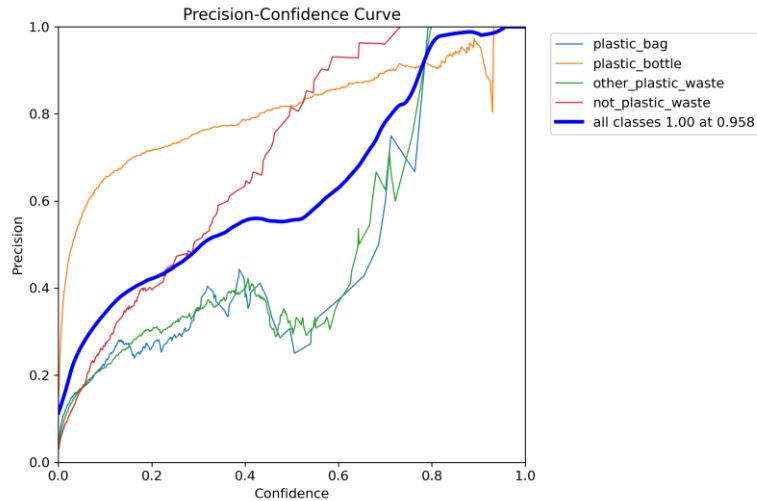
**Figure 4:** "PR-curve" Not pre-trained YOLOv8n.

Figure 4 illustrates the trade-off between precision and recall metrics. The x-axis represents 'recall', while the y-axis represents 'precision'. The graph displays various threshold values of the classifier's decision boundary. Numerical information is obtained using the mAP metric, corresponding to the area under the curve formed by the graph and coordinate axes. Unfortunately, the model exhibits low mAP scores, as listed in Table 4 under the 'mAP' column.



**Figure 5:** "F1-curve" Not pre-trained YOLOv8n.

In Figure 5, it's evident from the "F1" curve that this model achieves its highest F1 metric value at a confidence level of 0.38 for all classes.



**Figure 6:** "P-curve" Not pre-trained YOLOv8n.

The graph depicted in Figure 6 illustrates how accuracy varies with the "confidence" metric. It's apparent from this graph that the model attains its highest accuracy results at the highest confidence level, reaching 1.00 at 0.958.

## 5.2. Experiment Results Using Pre-trained YOLOv8n

The preceding model discussed in Section 5.1 yields suboptimal accuracy indicators. To enhance the achieved results, we employ the 'fine-tuning' technique. Specifically, we utilize a pretrained YOLOv8n model on the extensive COCO dataset and fine-tune it on our dataset, anticipating improved accuracy indicators. Additionally, we experiment with increasing the image resolution to 800 x 800 pixels.

This network undergoes training in two iterations: the initial iteration spans 200 epochs. Subsequently, recognizing the potential for further accuracy improvement through prolonged training, we conduct a second iteration wherein the model undergoes retraining for 300 epochs, with hyperparameters detailed in Table 4.

**Table 4**  
**Hyperparameters of Model 2**

Epochs	Learning rate	Momentum	Image size	YOLO version	Is pre-trained
500	0.0001	0.98	800	YOLOv8n	Yes

The training of this model required approximately 12 hours. Below, we present the results obtained by evaluation metrics for the trained model.

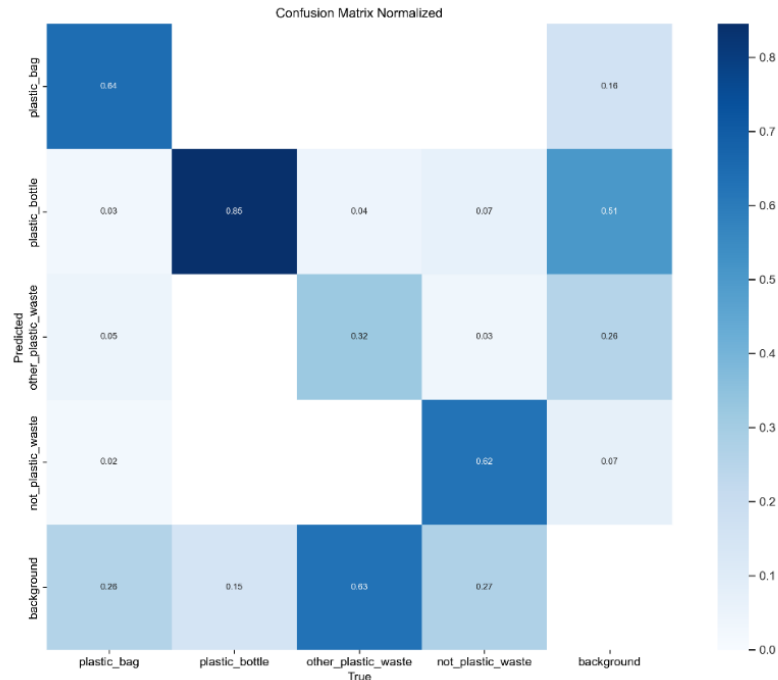
**Table 5**  
**Efficiency Metrics of Model 2**

Class	Precision	Recall	mAP
all	0.799	0.576	0.686
plastic_bag	0.859	0.604	0.753
plastic_bottle	0.848	0.797	0.877
other_plastic_waste	0.674	0.277	0.395
not_plastic_waste	0.816	0.625	0.719

Table 5 demonstrates that our model achieves accuracy indicators (mAP - 0.686; Precision - 0.799 for all classes) closely aligned with the best-performing models (mAP: 0.68-0.71) identified

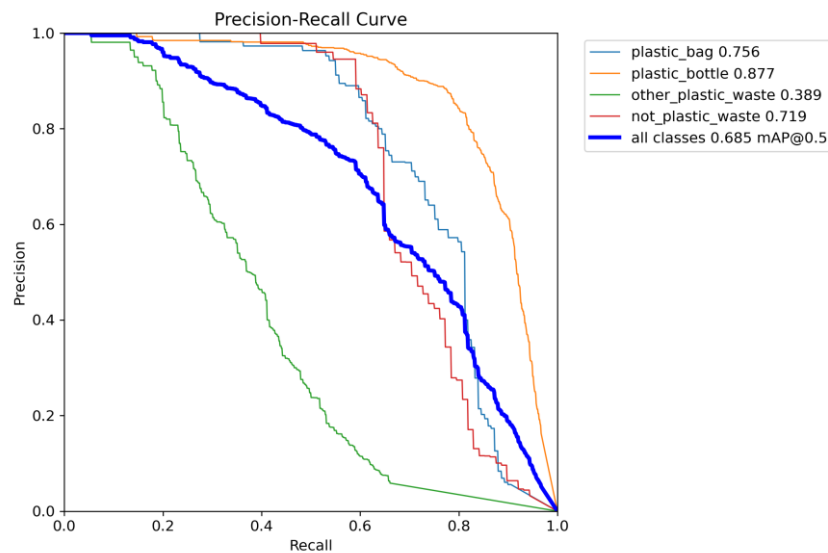


in the studies outlined in the introductory chapter. As anticipated, the model exhibits the highest indicators for the "plastic\_bottle" class - mAP 0.75, given its prominence as the largest class in terms of annotations within the dataset. However, it's apparent that a bottleneck for our model lies in the training data of the "other\_plastic\_waste" class, where the model demonstrates a lower mAP of 0.395. This discrepancy is attributed to the limited number of annotations available for this class



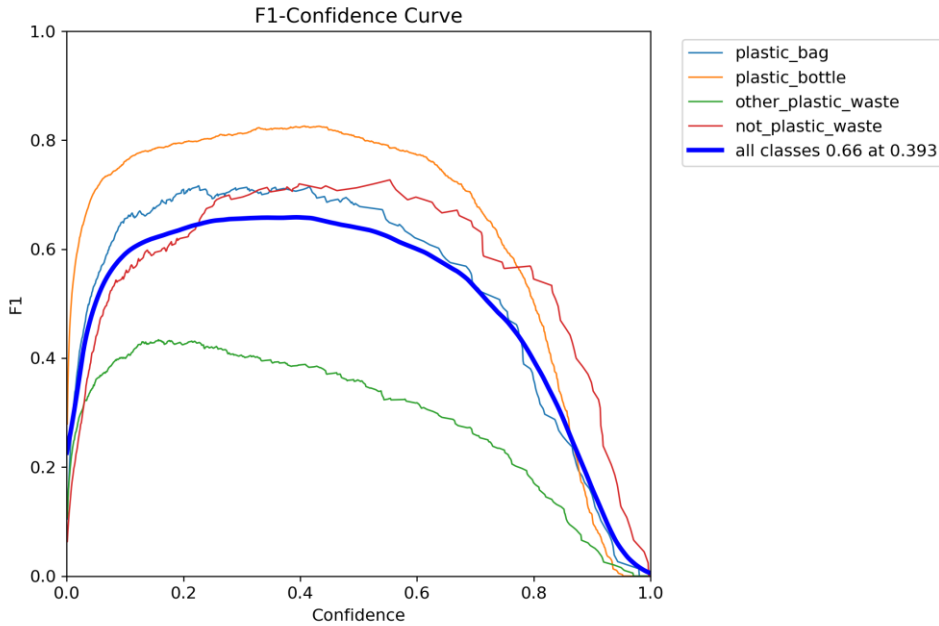
**Figure 7:** Confusion Matrix Pre-trained YOLOv8n.

Based on the confusion matrix depicted in Figure 7, our model generally achieves satisfactory results, correctly classifying the 'plastic\_bottle' class in 85% of cases, 'plastic\_bag' in 64% of cases, and 'not\_plastic\_waste' in 62% of cases



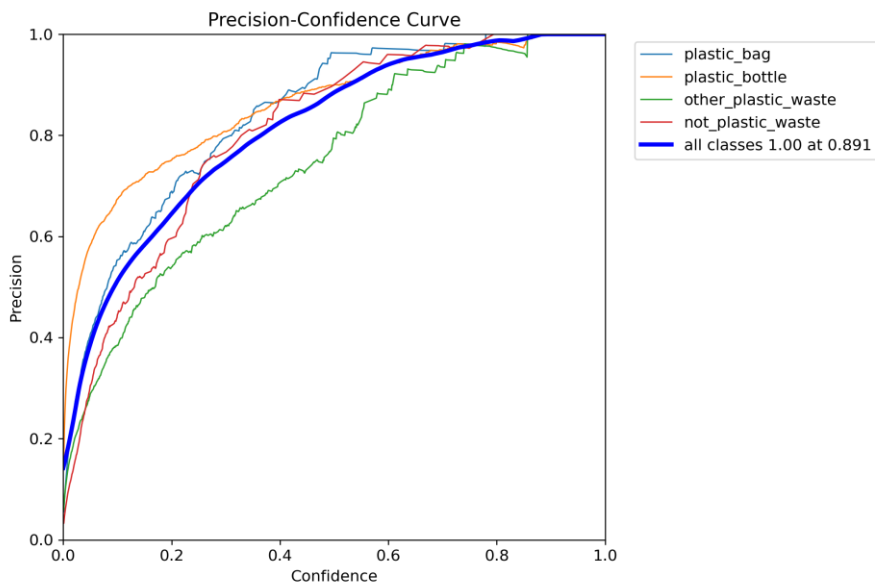
**Figure 8:** "PR-curve" Pre-trained YOLOv8n.

Figure 8 illustrates a clear increase in the mAP values of this model compared to the network from the previous section. The curve corresponding to the 'plastic\_bottle' class forms the largest area, indicating that the model achieves the highest accuracy for this class.



**Figure 9:** "F1-curve" Pre-trained YOLOv8n.

Figure 9 demonstrates that this model achieves its highest F1 metric value of 0.66 at a confidence level of 0.393 for all classes.



**Figure 10:** "P-curve" Pre-trained YOLOv8n.

In Figure 10, we observe that the developed CNN achieves an accuracy of 1.0 with a confidence value of 0.891.

### 5.3. Experiment Results Using Pre-trained YOLOv8m

The final developed model of note is the fine-tuned YOLOv8m CNN network, featuring high image resolution of 1008 x 1008 pixels. YOLOv8m boasts more convolutional layers and parameters compared to YOLOv8n (25,858,636 parameters versus 3,006,428), necessitating substantially more computing resources. Below is a table presenting the comprehensive list of hyperparameters for this model.

**Table 6**  
**Hyperparameters of Model 3**

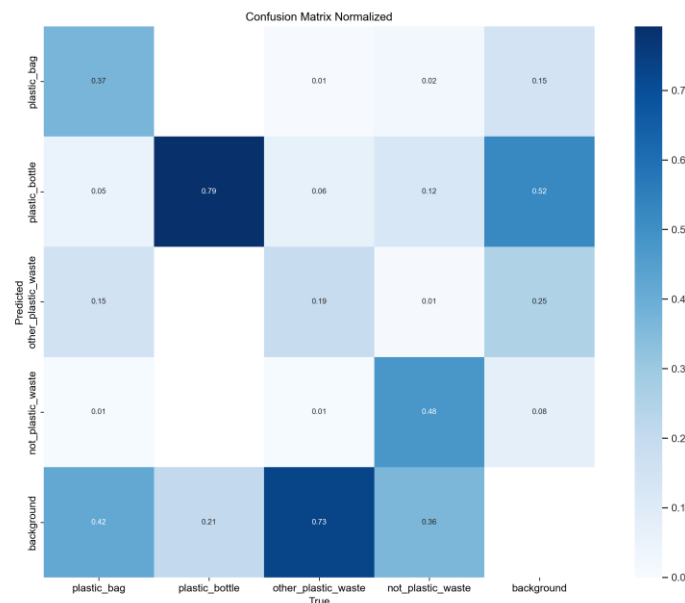
Epochs	Learning rate	Momentum	Image size	YOLO version	Is pre-trained
10	0.001	0.9	1008	YOLOv8m	Yes

Table 6 indicates that the training of this model was limited to only 10 epochs. This constraint arises from the significant computational demand of the model, with each epoch requiring approximately two hours to complete. Moreover, conducting extensive, prolonged training requires more computing power than was available during this study. Prolonged training could risk overheating the hardware complex and potentially lead to failure. Despite the limited number of epochs, the model exhibited promising potential for achieving high results.

**Table 7**  
**Efficiency Metrics of Model 3**

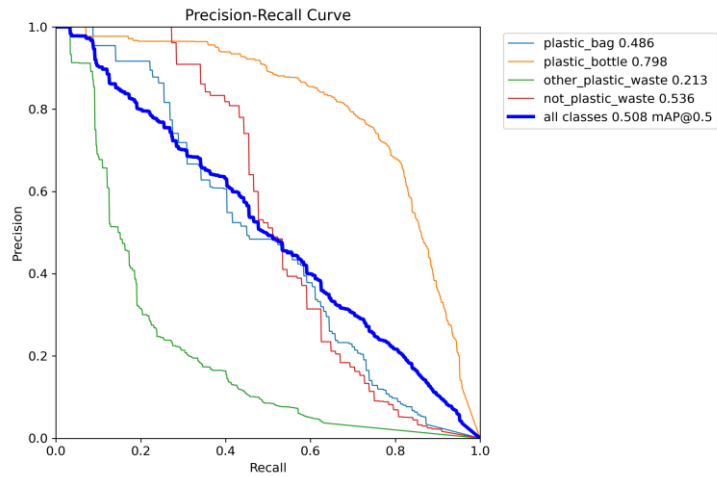
Class	Precision	Recall	mAP
All	0.63	0.437	0.51
plastic_bag	0.623	0.362	0.488
plastic_bottle	0.753	0.75	0.798
other_plastic_waste	0.412	0.181	0.212
not_plastic_waste	0.732	0.455	0.541

The model evaluation results presented in Table 7 showcase that despite the limited number of training epochs, this network attains noteworthy accuracy indicators: a 0.51 mAP is a commendable outcome considering the brevity of training. This success can be attributed to the extensive parameter count and training on high-resolution images.



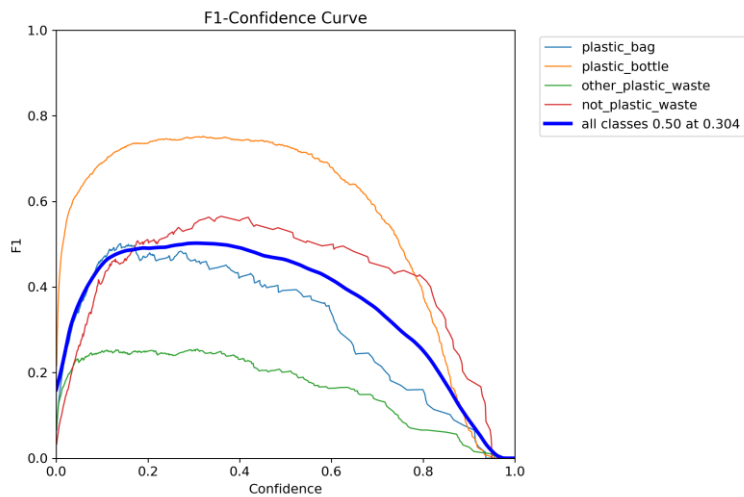
**Figure 11:** Confusion Matrix Pre-trained YOLOv8m.

Figure 11 illustrates that this model accurately identifies the 'plastic\_bottle' class in 79% of cases and correctly recognizes 'not\_plastic\_waste' in 48% of cases. However, it fails to identify 'other\_plastic\_waste' in 73% of cases, possibly attributed to the limited number of training epochs.

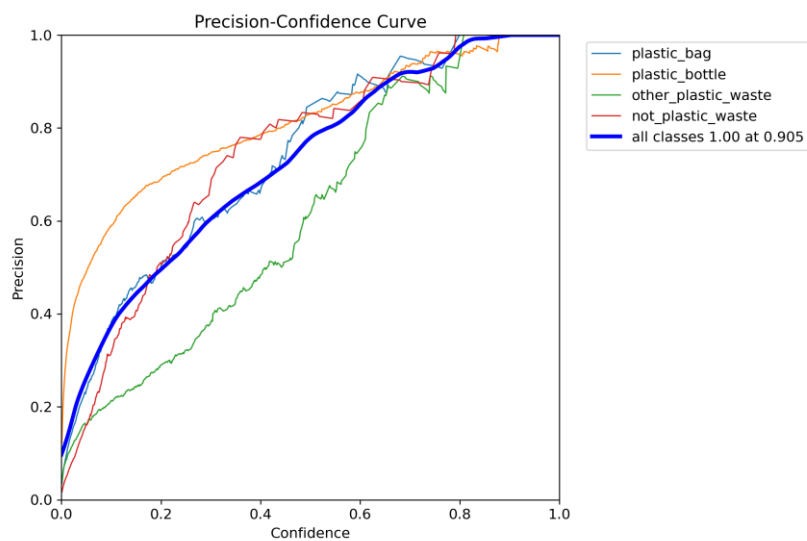


**Figure 12:** "PR-curve" Pre-trained YOLOv8m.

We observe that the 'all classes' line splits this Cartesian plane into two, resulting in an mAP indicator value of 0.5 for all classes.



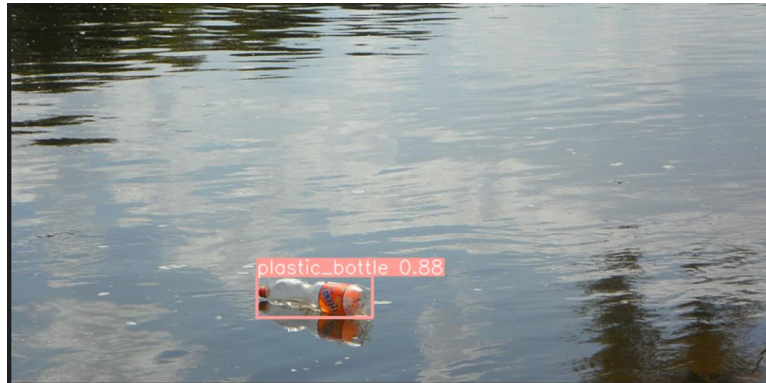
**Figure 13:** "F1-curve" Pre-trained YOLOv8m.



**Figure 14:** "P-curve" Pre-trained YOLOv8m.

It is evident from Figure 13 that the F1 value reaches 0.5 at a confidence level of 0.304, and from Figure 14, the model achieves Precision of 1.0 with a confidence level of 0.905.

#### 5.4. Experiment Results of models on unseen images



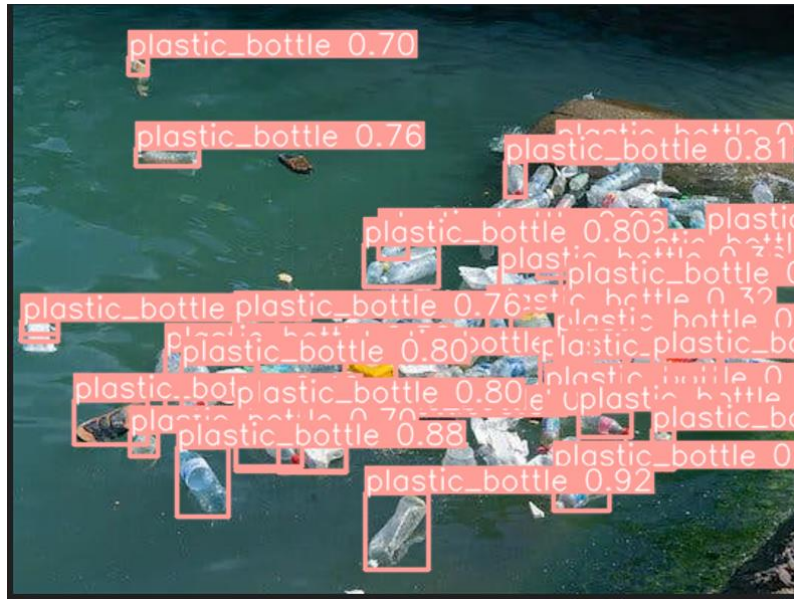
**Figure 15:** Example of detection of one object by the first model.



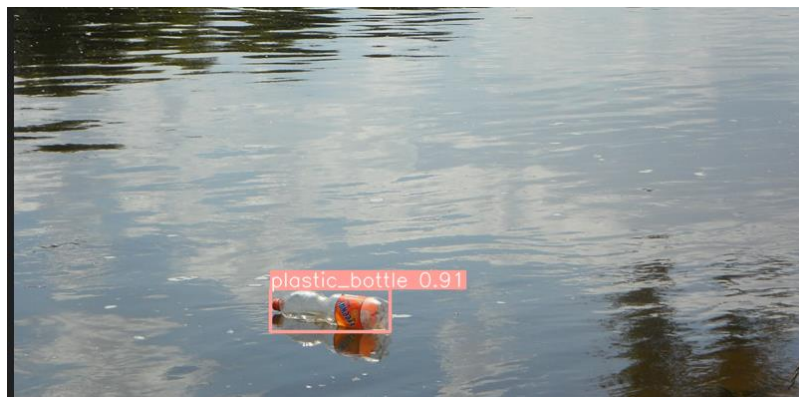
**Figure 16:** Multiple object detection with the first model model.



**Figure 17:** An example of detection of one object by second model.



**Figure 18:** An example of the work of the second model on a large accumulation of plastic waste.



**Figure 19:** Detection of one object by a third model.



**Figure 20:** Detection of garbage accumulation by the third model.

## 6. Discussions

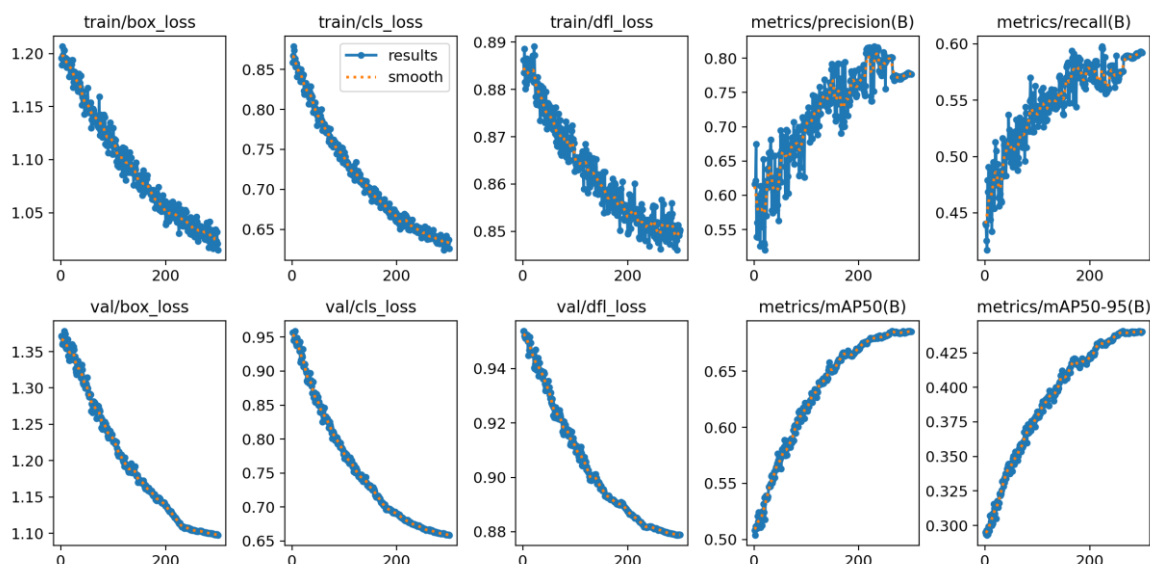
After analyzing the results obtained at the stage of conducting experiments, we conclude that the best model was trained according to the "fine-tuning" principle based on the pretrained YOLOv8n model on the "COCO" dataset with a "learning speed" equal to 0.0001, with a resolution of input images of 800 x 800 pixels and "Adam" optimizer. The general architecture of such a model contains 3006428 parameters. The model itself occupies 5.98 MB. The model was trained for 500 epochs, which took about 12 hours. As a result, it was possible to achieve the results shown in Table 8 and Figure 21.

**Table 8**  
**Efficiency Metrics of The Best Model.**

Class	Precision	Recall	mAP
All	0.799	0.576	0.686
plastic_bag	0.859	0.604	0.753
plastic_bottle	0.848	0.797	0.877
other_plastic_waste	0.674	0.277	0.395
not_plastic_waste	0.816	0.625	0.719

Despite the class annotation imbalance in the original dataset and limited computational resources, a model achieving an 80% accuracy rate and an mAP50 score of 0.686 was developed, as evident from Table 8. While a slightly superior accuracy was achieved in one of the previously analyzed studies [10], it was accomplished through a larger dataset and better hardware infrastructure.

It was determined that the optimal image size, given the available computational resources, is 800 x 800 pixels. Additionally, a significant observation is that models trained on networks with pretrained weights demonstrate notably improved performance. Thus, we infer that even with inferior resources, leveraging a more sophisticated model architecture in this study enabled attainment of results akin to those in advanced domains. This underscores the potential for further advancement in this research field; with enhanced resource allocation, notably superior outcomes are foreseeable compared to those observed in the scrutinized studies.



**Figure 21:** Overall Results of The Best Model.

The following outlines key avenues for potential further exploration and enhancement of results:

- **Expansion of the Training Dataset:** The efficacy of any deep learning model is intricately tied to the size of the training dataset. A fundamental principle dictates that larger datasets correspond to improved accuracy. Therefore, augmenting our dataset by at least 5000 images holds the potential to surpass the 90% accuracy threshold.
- **Class Annotation Balancing:** A primary limitation of the dataset employed in this study is its inadequate class balance. By incorporating images with annotations for items such as plastic bags and other variants of plastic waste, significant enhancements in network accuracy can be achieved, potentially yielding an mAP50 approximation nearing 0.8.
- **Utilization of High-Resolution Images:** Experiment results from the preceding section underscore the promise of training on images exceeding 1000 pixels in resolution. This approach exhibits considerable potential for realizing notable accuracy levels even with a limited number of training epochs.
- **Integration of Models with Enhanced Depth:** The second model, as depicted in the outcomes discussed in the preceding section, exhibits promising accuracy potential but necessitates substantial computational resources due to its augmented layering.
- **Leveraging More Robust Computing Infrastructures:** Engaging in training endeavors involving high-quality images and models with augmented parameters necessitates access to potent computing resources to effectively handle the computational demands.

## 7. Conclusions

Automatic recognition of plastic waste stands as an exceedingly critical problem garnering the attention of numerous researchers. With the rapid advancement of convolutional neural networks (CNNs) in addressing computer vision challenges, their adoption has become a standard practice for implementing object recognition systems. Each passing year witnesses the emergence of newer and more precise models, accompanied by increasingly user-friendly software interfaces, thereby facilitating researchers in various domains to apply them to address pertinent issues.

Following a comprehensive review of literature, it became evident that the YOLO family of models represents an advanced approach for tackling the challenge of automatic recognition of plastic waste. In this study, automatic recognition of plastic waste in water bodies was accomplished using the YOLOv8 model. The model underwent training on the publicly available Kili Technologies dataset, named "plastic\_in\_river" [8], comprising 4259 high-resolution images.

Despite some imbalance inherent in the dataset, rigorous iterations encompassing training, evaluation, and hyperparameter tuning led to the attainment of an accuracy rate approaching 80% after training the model for 500 epochs. This achievement is comparable to the best results reported by other researchers [3, 4, 5, 6], whose studies were considered during the course of this research.

Furthermore, experiments yielded the development of a model demonstrating significant potential for enhancing accuracy outcomes through the utilization of a deeper CNN network and training on high-resolution images. However, complete training of this model necessitates computing resources exceeding those available during this study.

Moreover, it was observed that leveraging pretrained models substantially enhances recognition accuracy post fine-tuning. Subsequent testing of the trained network on real data, distinct from the training set, is anticipated to yield satisfactory results, affirming the success of this research endeavor.

While this work does not entirely address the ongoing need for research on automatic recognition of plastic waste, it serves as a validation of the viability of such a system. Additionally, it delineates potential avenues for future research aimed at enhancing the obtained results.



## References

- [1] J. Solawetz, What is YOLOv8? The Ultimate Guide, 2023. URL: <https://blog.roboflow.com/whats-new-in-yolov8/>
- [2] P. Kershaw, Marine plastic debris and microplastics - global lessons and research to inspire action and guide policy change, (2016): 45-96. doi:10.13140/RG.2.2.30493.51687.
- [3] N. Maharjan, H. Miyazaki, (Eds.), Detection of River Plastic Using UAV Sensor Data and Deep Learning, *Remote Sens* 14, (2022). doi: 10.3390/rs14133049.
- [4] G. Aldric Sio, D. Guantero, J. Villaverde, Plastic Waste Detection on Rivers Using YOLOv5 Algorithm, (ICCCNT), Kharagpur, India (2022). doi: 10.1109/ICCCNT54827.2022.9984439.
- [5] A. Cortesi, M. Masiero, G. Tucci, Random Forest -Based River Plastic Detection With a Handheld Multispectral Camera, *The International Archives of the Photogrammetry* (2021): 101-107. doi: 10.5194/isprs-archives-XLIII-B1-2021-9-2021.
- [6] C. Lieshout, K. Oeveren, Automated River plastic monitoring using deep learning and cameras. *Earth and Space Science*, 7 (2020). doi: 10.1029/2019EA000960.
- [7] J. Tianlong, Z. Kapelan, (Eds.), Deep learning for detecting macroplastic litter in water bodies: A review, *Water Research* 231, (2023). doi: 10.1016/j.watres.2023.119632.
- [8] J. Redmon, S. Divvala, (Eds.), You Only Look Once: Unified, Real-Time Object Detection, 2016. URL: <https://arxiv.org/abs/1506.02640>.
- [9] Hugging Face, Kili Technologies: «plastic\_in\_river» dataset, 2022. URL: [https://huggingface.co/datasets/kili-technology/plastic\\_in\\_river](https://huggingface.co/datasets/kili-technology/plastic_in_river).
- [10] L. Lebreton, V. Zwet (Eds.), Reisser J. River plastic emissions to the world's oceans. *Nat Commun* 8 (2017). doi: 10.1038/ncomms15611.
- [11] G. Jiuxiang, W. Zhenhua, (Eds.), Recent Advances in Convolutional Neural Networks. *Pattern Recognition* (2018): 354-377. URL: <https://arxiv.org/pdf/1512.07108.pdf%C3%A3%E2%82%AC%E2%80%9A>