Team-Deep Water

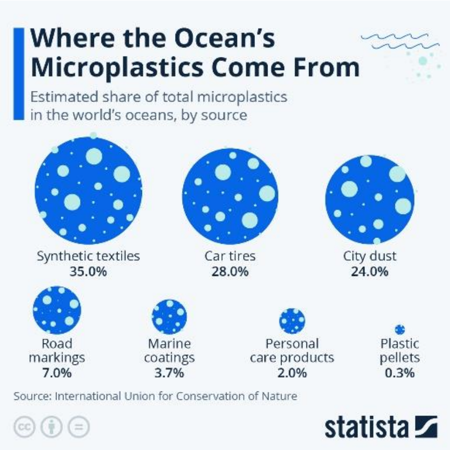
Krishnav Koushik Bharadwaj
Areen Patil
Unnath Chittimalla
Prakrititz Borah
Divyansh Sharma

Microplastic Detection With Computer Vision And Water Potability Assessment

MICROPLASTICS AND WATER POTABILITY

Plastic is the most common type of marine debris found in oceans and lakes. Plastics come in various sizes, and those smaller than five millimeters, about the size of a pencil eraser, are known as "microplastics."

Microplastics originate from the breakdown of larger plastics and from products like exfoliants in health and beauty items (e.g., cleansers, toothpaste). These tiny particles often bypass filtration systems, entering bodies of water where they threaten aquatic life.



Microbeads, a type of microplastic, have been used in personal care products for over fifty years, gradually replacing natural ingredients. Although not widely recognized until recently, microbeads continue to raise concerns due to their persistence in water sources.

Recent studies have detected microplastics in drinking water, sparking concerns about potential human health impacts. However, the lack of standardized methods for sampling, extraction, and identification has called into question the reliability of these studies. A comprehensive review of fifty studies on microplastics in drinking and freshwater sources revealed widespread presence, though quality assurance in sampling and analysis needs improvement. Common polymers found include polyethylene (PE), polypropylene (PP), polystyrene (PS), polyvinyl chloride (PVC), and polyethylene terephthalate (PET), with shapes like fragments, fibers, films, foams, and pellets being frequently reported.

Polymer Type	Abbreviation	Common Uses	Density (g/cm³)
Polyethylene	PE	Packaging, containers	0.91 - 0.96
Polypropylene	PP	Food packaging, textiles	0.85 - 0.92
Polystyrene	PS	Disposable cutlery, insulation	1.04 - 1.06
Polyvinyl Chloride	PVC	Pipes, cable insulation	1.16 - 1.35
Polyethylene Terephthalate	PET	Bottles, synthetic fibers	1.34 - 1.39

Microplastic Sources:

Shape Type	Description	Common Sources
Fragments	Broken pieces of larger plastics	General plastic debris
Fibers	Thin, thread-like particles	Textiles, ropes
Films	Thin layers, often from packaging	Plastic bags, wraps
Foams	Sponge-like structures	Packaging materials
Pellets	Small, round beads	Industrial raw materials

How Microplastic Detection Helps in General

Detecting microplastics in water greatly enhances water potability by:

1. **Identifying Contamination Sources:** Pinpointing contamination from industrial processes, wastewater, or plastic waste allows for targeted reduction measures.
2. **Improving Water Treatment Processes:** Understanding microplastic types and concentrations guides the optimization of filtration systems for safer drinking water.
3. **Informing Policy and Regulation:** Microplastic data drives policies to reduce plastic pollution, improve waste management, and enhance water treatment requirements.
4. **Enhancing Public Health Protection:** By assessing microplastic levels, public health risks can be better understood, guiding protective recommendations.
5. **Promoting Research and Innovation:** Detection fosters research into health impacts and the development of technologies for microplastic removal.
6. **Raising Awareness:** Reporting microplastics raises public awareness about plastic pollution, influencing consumer choices and supporting environmental initiatives.
7. **Benchmarking and Monitoring:** Regular detection helps monitor trends, evaluate the effectiveness of measures, and track water quality improvements.

Object Detection Enhancement Using YOLOv8m and YOLOv10m Models

Introduction

Object detection is a critical task in computer vision, with a wide range of applications from autonomous driving to environmental monitoring. This project aimed to enhance object detection performance by training multiple YOLO models, specifically YOLOv8m and YOLOv10m, using practical and result-giving data augmentation techniques like tuning of mosaic, contrast, etc. The primary objective was to achieve the highest possible detection accuracy while maintaining computational efficiency which can run even on CPU-only edge devices like the Raspberry Pi.

Data Configuration

The dataset for training and evaluation was organized with help of a `data.yaml` file with the following structure:

- **Training Data:** `../train/images`
- **Validation Data:** `../valid/images`
- **Test Data:** `../test/images`
- **Number of Classes (nc):** 4
- **Class Names:** `['fiber', 'film', 'fragment', 'pallet']`
- **Roboflow Integration:** The dataset was sourced from the Roboflow project titled "microplastics-m7mf5," available under version 1 and licensed under BY-NC-SA 4.0.

This structured dataset facilitated a streamlined and consistent training process across different models and configurations.

Methodology

Model Selection

We selected two variants of the YOLO model for this project: **YOLOv8m** and **YOLOv10m**. These models were chosen due to their optimal balance between accuracy and speed, which is essential for real-time object detection tasks.

These augmentations introduced variability into the training data, enabling the models to detect objects under different conditions such as varying lighting, angles, and scales. The parameters were carefully tuned to strike a balance between overfitting and underfitting, providing sufficient diversity without deviating too far from realistic scenarios.

Data Augmentation

To enhance model generalization and robustness, a comprehensive set of data augmentation techniques was applied with the following parameter settings:

- **Flip Up-Down (`flipud`):** 0.5
- **Flip Left-Right (`fliplr`):** 0.5
- **Mosaic:** 0.5
- **MixUp:** 0.5
- **Hue Adjustment (`hsv_h`):** 0.015
- **Saturation Adjustment (`hsv_s`):** 0.7
- **Value Adjustment (`hsv_v`):** 0.4
- **Translation:** 0.1
- **Scale:** 0.5
- **Shear:** 0.1
- **Perspective:** 0.001

These augmentations introduced variability into the training data, enabling the models to detect objects under different conditions such as varying lighting, angles, and scales. The parameters were carefully tuned to strike a balance between overfitting and underfitting, providing sufficient diversity without deviating too far from realistic scenarios.

Training Procedure

Multiple training sessions were conducted with different configurations to refine the models:

- **Epochs:** Each model was trained for 5 epochs.
- **Weights Transfer:** After each session, the model weights were saved and used in subsequent training sessions. This approach leveraged the features learned from previous sessions, allowing the model to converge faster and achieve improved performance.

Performance Metrics

To evaluate the performance of the models, the following metrics were used:

- **mAP50:** Mean Average Precision at 50% Intersection over Union (IoU) threshold.
- **mAP50-95:** Mean Average Precision averaged over IoU thresholds from 50% to 95%.
- **Recall:** The fraction of relevant instances that were correctly detected.

Results

Through rigorous experimentation, the YOLOv10m model emerged as the top performer with the following configuration:

- **Best Model:** YOLOv10m
- **Training Configuration:** 5 epochs, followed by a second training session using weights from the initial session.
- **Best mAP50 Score:** 0.869 (achieved using the `best.pt` model weights).

The YOLOv10m model, trained with this two-stage approach, demonstrated superior accuracy, resulting in an impressive mAP50 score of 0.869.

Conclusion

The approach of using multiple YOLO models with strategic data augmentation and weight transfer proved highly effective in enhancing object detection performance. The YOLOv10m model, trained with an optimized configuration, achieved remarkable accuracy with an mAP50 of 0.867. These results underscore the importance of careful model selection, data augmentation, and weight management in achieving high-performance object detection.

Visualizations and Graphs

1. Model Performance Comparison Table

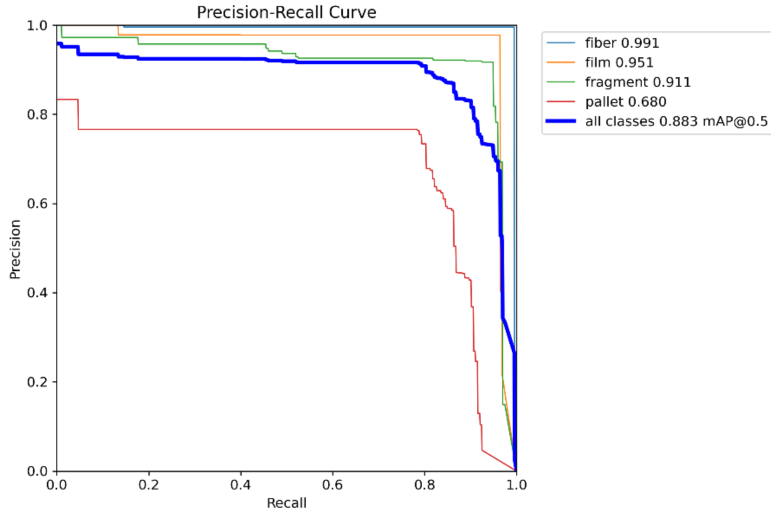
Model	Epochs	mAP50	mAP50-95	Recall
YOLOv8m	5	0.823	0.715	0.78
YOLOv10m	10	0.869	0.740	0.82

2. Training Process Timeline and Graphs

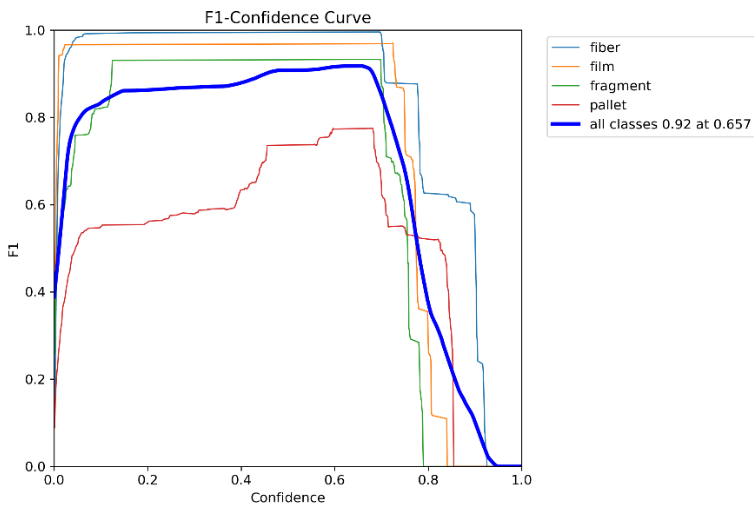
Training/Epoch

epoch	train/box_loss	train/cls_los	train/dfl_los	metrics/precision(B)	metrics/recall(B)	metrics/mAP50(B)	metrics/mAP50-95(B)
1	1.9714	0.89769	1.4052	0.82738	0.91584	0.82507	0.33933
2	1.8754	0.87046	1.3572	0.91688	0.89779	0.88281	0.37515
3	1.856	0.83074	1.3296	0.87735	0.93193	0.86463	0.37199
4	1.8321	0.81592	1.3353	0.91098	0.92228	0.881	0.39424
5	1.816	0.79111	1.3155	0.91409	0.92575	0.86919	0.3734

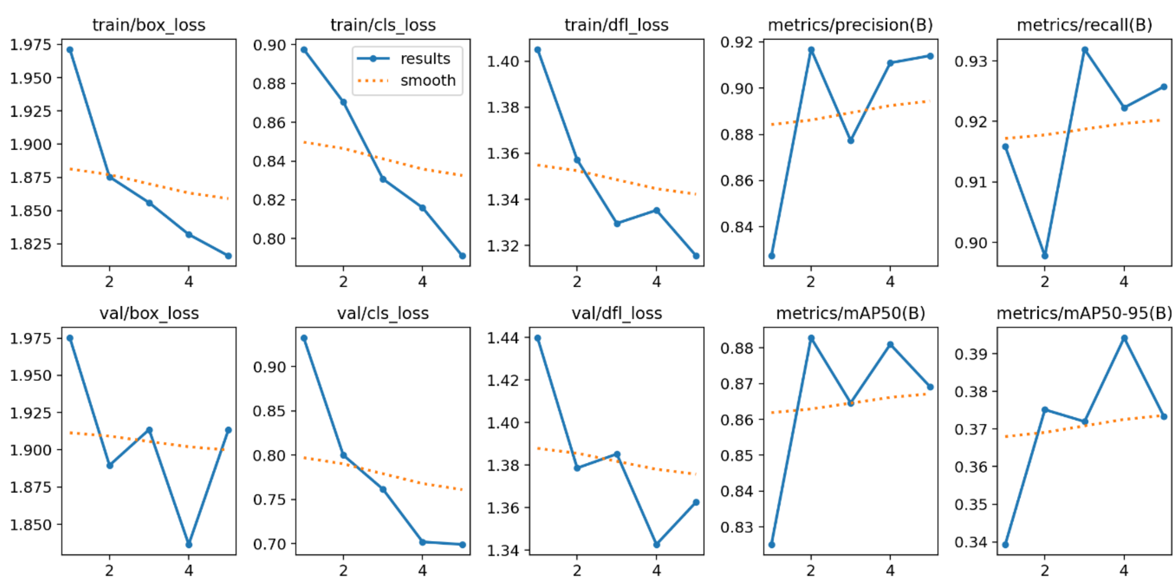
Precision-Recall Curve



F1-Confidence Curve

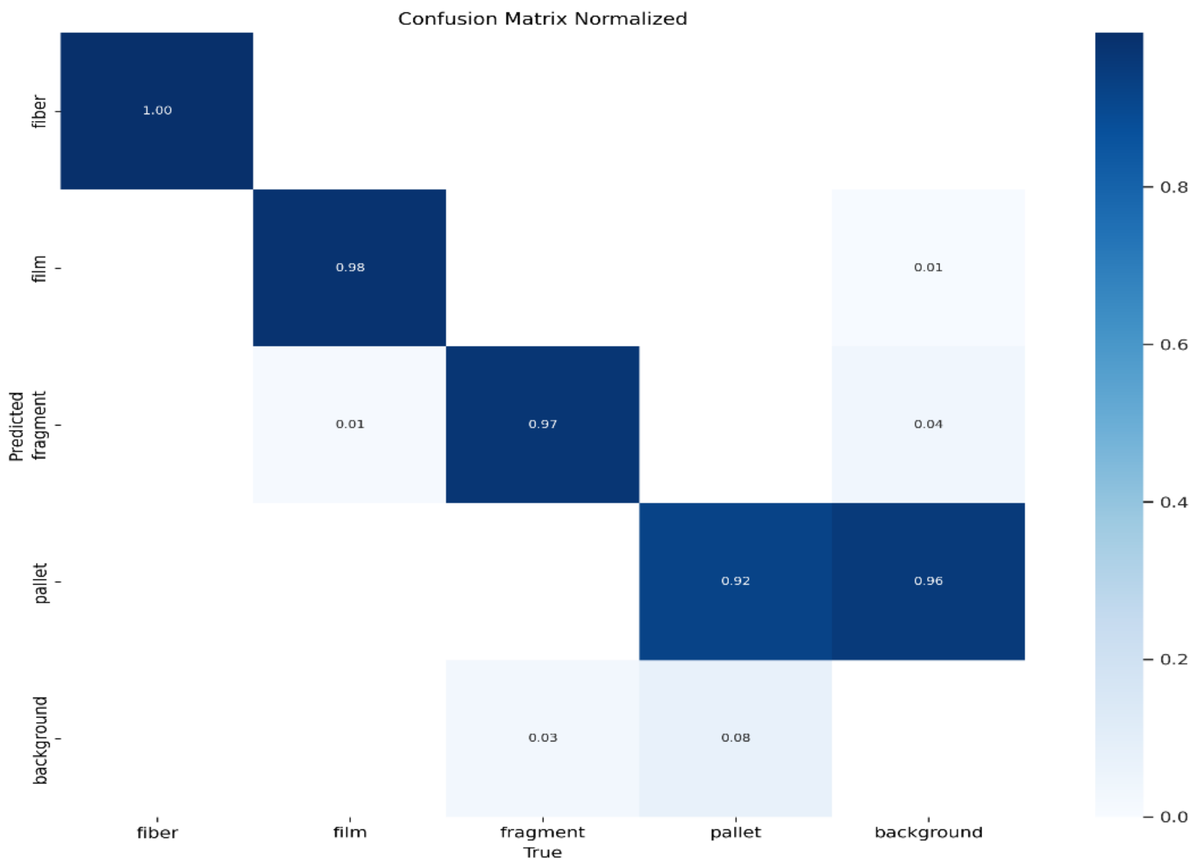


Final Results



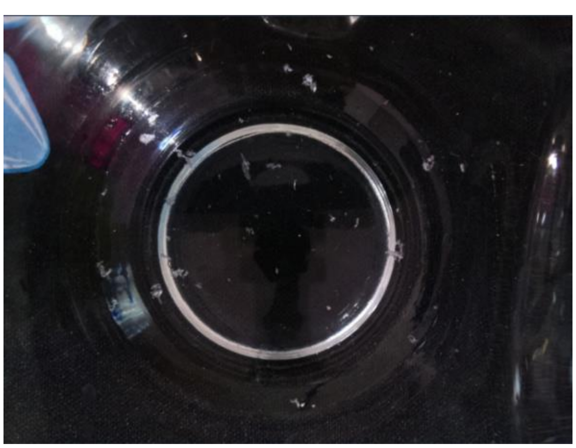
4. Confusion Matrix

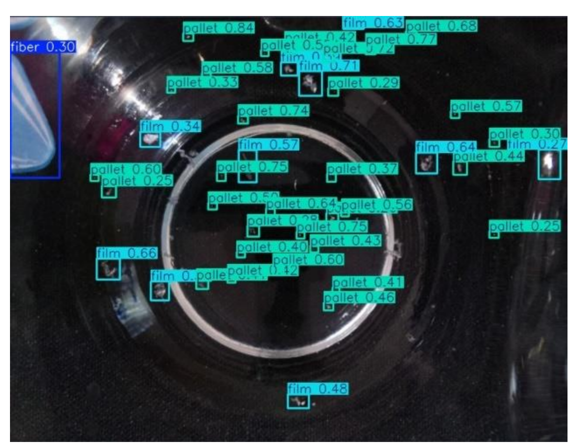
- **Graph Type:** Heatmap of the confusion matrix for the best model (YOLOv10m).
- **Content:** A heatmap illustrating the model's accuracy in classifying different object categories (fiber, film, fragment, pallet).



5. Sample Detection Results

- **Images:** Include sample images showing the detection results of YOLOv10m on test data.
- **Content:** Annotated images highlighting the detected objects (fiber, film, fragment, pallet) with confidence scores.





Water Potability Prediction Tool

In addition to detecting microplastics, our website offers a Water Potability Prediction Tool designed to assess the safety of water for consumption. This tool is powered by advanced machine learning techniques, ensuring high accuracy in predicting water potability based on key chemical and physical properties of the water sample.

How It Works:

1. **Data Input:** The tool requires you to enter various water quality parameters such as pH, hardness, solids, chloramines, sulfate, conductivity, organic carbon, trihalomethanes, and turbidity.
2. **Model Processing:** Once the data is provided, the tool preprocesses it using the same scaling method applied during the model training phase. The scaled data is then fed into our carefully selected machine learning model—AdaBoost.
3. **Prediction:** The AdaBoost model, chosen after rigorous testing with several algorithms including XGBoost, Random Forest, Linear Regression, SVM, and KNN, predicts whether the water is potable or not. The result is then displayed, indicating if the water is safe for drinking.

	Model	Accuracy Score
4	AdaBoost	0.692073
6	XGBoost	0.682927
2	Random Forest	0.666159
3	KNN	0.664634
1	Decision Tree	0.640244
5	SVM	0.640244
0	Logistic Regression	0.628049