Machine Learning for Water Pollutant Detection and Classification

Carter Costic[1*] **Kevin Le**[1*] **William Zhang**[1*]
[1]Princeton University
{cartercostic, kevin_le, william.zhang}@princeton.edu

Abstract

Accurate detection and classification of water pollutants in rivers, lakes, and coastal waters are critical for environmental monitoring and remediation. This research develops a robust computer vision system leveraging machine learning to identify and localize various pollutants like plastic, oil, and sewage from images. The system employs a deep learning model designed for object detection (YOLOv8) and a multi-class classification architecture (ResNet) to achieve high precision, recall, and F1 scores. The system exhibited robust generalization across varying conditions, including water turbidity and occlusions. The research contributes to environmental sustainability by enabling efficient monitoring, pollution source identification, and prioritization of cleanup operations, facilitating timely intervention and ecosystem preservation.

Key Words: Deep Learning, Object Detection, YOLO, ResNet, Waste Classification, etc.

1 Introduction

Water pollution, particularly from plastic waste, poses a severe threat to the health of aquatic ecosystems and human communities relying on these water bodies for everyday use [10]. Plastic pollution in oceans and rivers has reached alarming levels, with an estimated 8 million metric tons of plastic entering the oceans annually [8]. This plastic waste not only endangers marine life but also has detrimental effects on human health, as microplastics can enter the food chain and accumulate in the bodies of organisms [18].

In addition to plastic pollution, oil spills are another significant environmental concern for water bodies worldwide. Oil spills can have devastating consequences on marine and freshwater ecosystems, causing long-lasting damage to flora and fauna [1]. The toxic components in crude oil can disrupt the food chain, contaminate shorelines, and pose risks to human health through the consumption of contaminated seafood [10].

Effective monitoring and timely identification of pollutants are crucial for implementing remediation strategies and mitigating environmental damage. Traditional manual monitoring methods are often labor-intensive, time-consuming, and subject to human error [16]. Machine learning techniques have been explored to mitigate some of these disadvantages by automating the detection and classification processes [7].

This research developed an automated computer vision system that leveraged machine learning techniques such as object detection and multi-class classification to accurately detect and classify various types of water pollutants, including plastic waste and oil spills, from images captured in rivers, lakes, and coastal waters. The system aimed to contribute to global efforts to address plastic pollution and oil spill incidents, which have severe consequences for marine ecosystems, biodiversity, and human health [14, 1].

We found that automated detection and classification of water pollutants significantly improved the efficiency and accuracy of environmental monitoring efforts. By leveraging the power of machine learning algorithms, particularly deep learning techniques, it became possible to process large volumes of image data in a scalable and cost-effective manner. This approach helped identify pollution sources more rapidly, enabling timely intervention and remediation efforts, ultimately contributing to the preservation of aquatic ecosystems and the protection of human health.

The increasing availability of high-resolution satellite imagery and drone footage has further highlighted the potential of computer vision techniques for environmental monitoring applications. However, developing robust and accurate models for detecting and classifying water pollutants posed several challenges, including varying environmental conditions, occlusions, and class imbalances. This research aimed to address these challenges by leveraging various state-of-the-art machine learning techniques and exploring strategies to improve model performance and generalization.

2 Related Work

Previous research in this domain explored the application of object detection and image classification techniques for environmental monitoring purposes. Existing studies demonstrated the potential of deep learning models, such as convolutional neural networks (CNNs), for classifying specific types of water pollutants [12, 3]. CNNs showed remarkable performance in classification tasks due to their ability to automatically learn hierarchical representations from image data, capturing relevant features at different scales.

However, many of these approaches focused on limited types of pollutants or were evaluated on constrained datasets, limiting their generalizability to real-world scenarios [3]. Additionally, few studies tackled the challenge of detecting and classifying multiple types of water pollutants simultaneously, which was essential for comprehensive environmental monitoring [7].

Some researchers explored the use of transfer learning techniques, where pre-trained models on large-scale datasets were fine-tuned for the specific task of water pollutant detection and classification [17]. These approaches leveraged the feature representations learned from diverse image data, allowing for faster convergence and improved performance on the target task.

Other studies investigated the use of ensemble methods, combining predictions from multiple models to enhance overall accuracy and robustness [5]. These techniques have shown promise in addressing class imbalances and improving generalization performance.

Despite these advances, several challenges remained, including handling varying lighting conditions, water turbidity, surface reflections, and occlusions [**?**], which have impacted model performance. Additionally, the availability of large-scale, diverse, and accurately labeled datasets for training and evaluation posed a significant bottleneck in developing robust and generalizable models.

Researchers have also explored the use of attention mechanisms [11] and multi-task learning approaches [2] to improve the performance of water pollutant detection and classification systems. Attention mechanisms have been shown to enhance the model's ability to focus on relevant regions within the input images, potentially improving the classification accuracy for specific pollutant types or under challenging environmental conditions. Multi-task learning, where the model is trained simultaneously on related tasks, such as object detection and classification, has demonstrated improved generalization and robustness.

3 Methodology

The methodology involved the development of a two-stage system: a multi-class classification component for categorizing the detected objects into predefined classes (e.g., plastic and oil) and an object detection component for localizing pollutants within the image frames.

3.1 Multi-class Classification

For the classification stage, we took inspiration from deep neural network architectures like ResNet [6] to create our model. Models like ResNet demonstrated excellent performance in various image

classification tasks, leveraging their depth and efficient design to capture complex patterns and representations for accurate categorization [6].

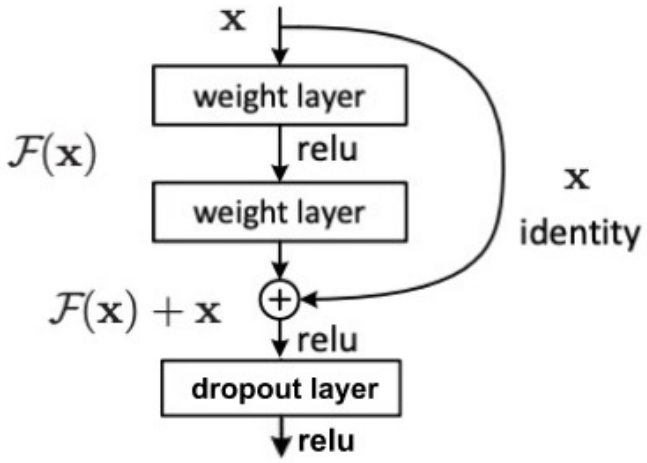

Figure 1: Resnet Architecture with its layers visualized with a dropout modifcation post RELU from paper [6]

Following the developments of models like VGGNet and ResNet, our model leveraged stacks of 3x3 conv2d layers and, like ResNet, captured residual connections by chaining Residual Blocks. While ResNet utilized residual connections to optimize much deeper CNNs, our model follows a similar construction to ResNet18 [6], allowing us to still benefit from the residual connections, effective receptive field of the 3x3 conv2d layers, and chained non-linearities while also keeping training time low.

Unlike ResNet, we incorporated Dropout layers after the first ReLU (at the end of the first Residual Block) and after the second batch normalization (BN), but before the residual addition as seen in Figure 1. Doing so was a stronger regularization technique that we required in our circumstances, unlike ResNet, which was trained on ImageNet. The heterogeneity and imbalance in our dataset created challenges that required stronger regularization in order to ensure more generalizable data. Additionally, adding the Dropout layers was easy to add to PyTorch model.

Another option for training was transfer learning, using existing weights and implementations for a model like ResNet18 [17]. We elected not to do so due to our unique dataset, worried that learned relationships from ImageNet would not transfer effectively to the images of pollutants in water.

For the second model trained, we reduced the size of the train, test, and validation datasets in order to create a more balanced model that gives consideration to all class labels. This significantly reduced the size, as the bottleneck being the number of clean water images limited the inclusion of other images, and breadth of the dataset, but enabled the creation of a more balanced final model.

We then scraped Google Images for images of bodies of water with other pollutants, like oil spills. Further research can expand these categories to include more pollutants, like sewage and chemicals, using similar methods. These scraped images were then integrated into the existing dataset.

Prior to training being completed on M3 Macbook Pro, we leveraged an Nvidia A100 with 80GB of VRAM on Modal Labs. While computation itself was extremely fast, network latency resulted in lengthy iteration times and high training costs; we found training locally on a CPU to be more efficient in this case, although having access to more local (for higher network bandwidth) GPU compute would expand the possibilities for training the classification model, such as a deeper model like ResNet152 [6].

Images were preprocessed by first being resized to 224x224 and then normalized to the mean and standard deviation of the entire dataset to ensure RGB channel variations, such as spikes, would not

skew training results. For the training dataset, images were also randomly horizontally flipped and had a color jitter applied before being normalized. The process of random data augmentation is a stronger form of regularization that allows the model to generalize better on the test set, and was also inspired by the implementations outlined in ResNet Section 3.4.

The same model trained without data augmentation and dropout saw a significant gap between train and test accuracy and loss, which was the impetus for the inclusion of stronger regularization techniques.

We divided our data into three sets: train, validation, and test. The validation set was used to measure the efficacy of the hyperparams and to enable tuning within and between runs. The main limitation of each subset was the amount of clean water images and the homogeneity of scraped oil images.

3.2 Object Detection

For the object detection stage, state-of-the-art models such as YOLOv8 [13] were explored. These models leveraged deep convolutional neural networks to efficiently detect object bounding boxes within images or video frames. By leveraging the hierarchical feature extraction capabilities of CNNs, these models effectively localized objects of interest, even in complex environments with varying backgrounds and occlusions. Below is the model architecture of YOLOv8m, the model we trained to detect waste [9].

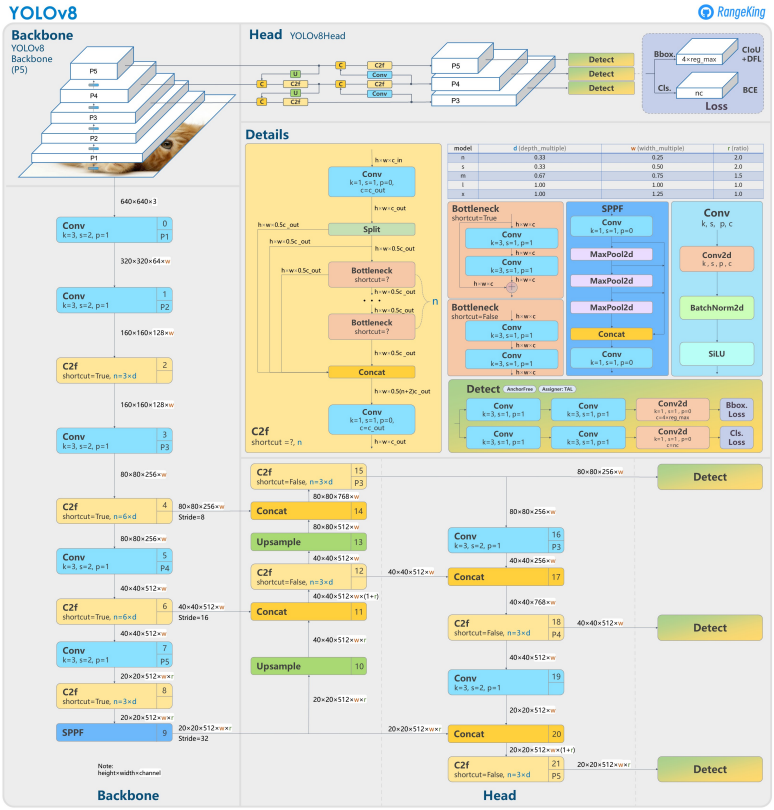

Figure 2: Yolov8 Architecture with its layers visualized on this diagram from Ultralytics [9]

At its core, the convolution layers serve as the building blocks for the CNN, extracting high-level features through convolution operations. The convolution layers are complemented by BatchNorm2D layers, which normalize activations across batches, enhancing the learning process. Additionally, this model architecture utilizes SiLU activation functions to introduce non-linearity, capturing more intricate patterns. These three parts compose the Bottleneck blocks, comprising of sequential layers to facilitate information [13]. Additionally, C2F blocks, consist of two convolutional layers, additionally enhancing feature extraction. The Spatial Pyramid Pooling block aggregates features at different scales (levels), aiding in contextual understanding. Upsampling layers increase feature map dimensionality, while concatenation layers concatenate information from multiple stages, resulting in more learned information. Finally, the detection head, incorporates these convolution operations to predict object bounding boxes and class probabilities. Together, these components form YOLOv8m model architecture for object detection [13].

We also employed ensemble methods to combine predictions from multiple object detection models, leveraging their complementary strengths and improving overall localization accuracy. Techniques like non-maximum suppression (NMS) and soft-NMS were employed to handle overlapping bounding box predictions and improve the final object localization results. Additionally, data augmentation techniques increased the diversity of the training data by applying various transformations (e.g., translation, sheering) to the existing images, improving the model's robustness.

Due to the large number of layers in the model architecture, our model was able to handle some of the specific challenges of object detection, such as varying lighting conditions, surface reflection, and water turbidity. Along with the large amounts of layers, the high dimensionality of our model required a lot of computing power when training our model. This is where we relied on computational resources from Modal Labs to efficiently train the model across multiple GPUs (H100s), significantly reducing the training time and resource requirements.

4 Experiments, Results, & Analysis

The system was evaluated using existing datasets, such as the Hugging Face dataset mentioned in the proposal, as well as additional labeled data generated through manual annotation efforts. Performance metrics such as precision, recall, F1-score, mean Average Precision (mAP), and Intersection over Union (IoU) were reported for both the object detection and classification components.

Performance Evaluations:

Accuracy: In the context of our multi-class classification model, accuracy represents the proportion of water pollutant samples that were correctly classified into their respective categories (plastic, aluminum, oil, sewage, etc.). It provides an overall measure of the model's ability to make correct predictions across all classes. However, it should be noted that accuracy alone may not be a reliable metric when dealing with imbalanced class distributions, as it can be heavily influenced by the majority class. Mathematically, accuracy is formulated as[12]:

$$\text{Accuracy} = \frac{TP + TN}{(TP + FN + FP + TN)}$$

Precision: For our multi-class classification model, precision represents the ability to correctly identify a specific pollutant type out of all the instances predicted as belonging to that class. It is particularly useful for assessing the model's performance in minimizing false positive predictions, which is crucial in environmental monitoring applications where misclassifying a pollutant could lead to incorrect remediation efforts. Precision is calculated for each class independently, and a high precision score indicates that when the model predicts a certain pollutant type, it is likely to be correct. Mathematically, precision for a given class is formulated as[12]:

$$\text{Precision} = \frac{TP}{FP + TP}$$

Recall: In the context of our research, recall represents the model's ability to correctly identify all instances of a particular pollutant type from the entire dataset. It is a crucial metric for ensuring that the system does not miss important instances of pollutants, as overlooking them could hinder effective

monitoring and remediation efforts. Like precision, recall is calculated for each class independently, and a high recall score indicates that the model is capable of detecting most instances of a given pollutant type. Mathematically, recall for a given class is formulated as[12]:

$$\text{Recall} = \frac{TP}{FN + TP}$$

F1 Score: The F1 score is a harmonic mean of precision and recall, providing a balanced evaluation metric that considers both false positives and false negatives. It is particularly useful when dealing with imbalanced class distributions, as it accounts for the trade-off between precision and recall. In our research, the F1 score is calculated for each pollutant class, allowing us to assess the model's overall performance across different types of water pollutants. A high F1 score indicates that the model achieves a good balance between precision and recall for a given class. Mathematically, the F1 score for a given class is formulated as [12]:

$$\text{F1 Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Recall} + \text{Precision}}$$

4.1 Multi-class Classification Results

4.1.1 Experiments

We trained our first model, 18-layer ResNet with Dropout, for nearly 20 hours or 90 epochs. We employed a learning rate scheduler that reduced on plateau with the following hyperparameters: patience=2, reduce=0.5. We found the model to plateau at the end of the training period, at 88-90% accuracy. This coincided with the learning rate approaching 0, due to the plateaus observed in the test accuracy.

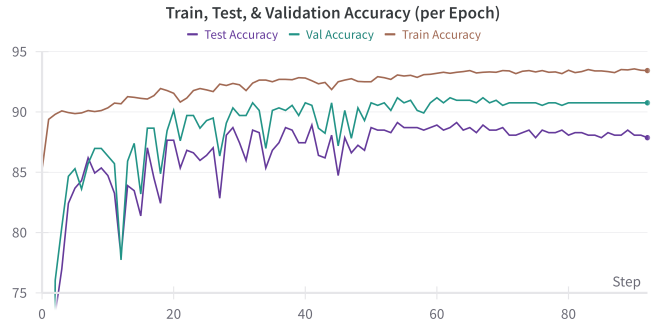

Figure 3: Training, validation, and test accuracy curves for the model one of multi-class classification component (per epoch), demonstrating the model's convergence and generalization performance.

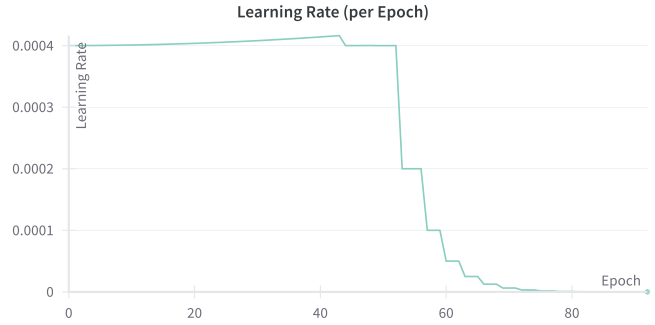

Figure 4: Graph visualizing the learning Rate per epoch

On the other hand, the second model, aimed to provide a more balanced classification, trained for around 10 hours. In an attempt to prevent plateaus at the end of the training period, we used a One Cycle learning rate scheduler such that the learning rate would increase with more momentum, then quickly decrease, and repeat in cycles. While we found this method of scheduling to sometimes result in drops in both test and train accuracy and loss, we did not experience the same zeroing at the end of the training period.

4.1.2 Results and Analysis

The high accuracy demonstrated by the first model embodies success in being able to identify trash in bodies of water and a failure in being able to distinguish between clean water and water with trash. It should be noted that, when observed in the larger model pipeline where the classification output becomes the input for a routing decision, false positives are more agreeable than false negatives.

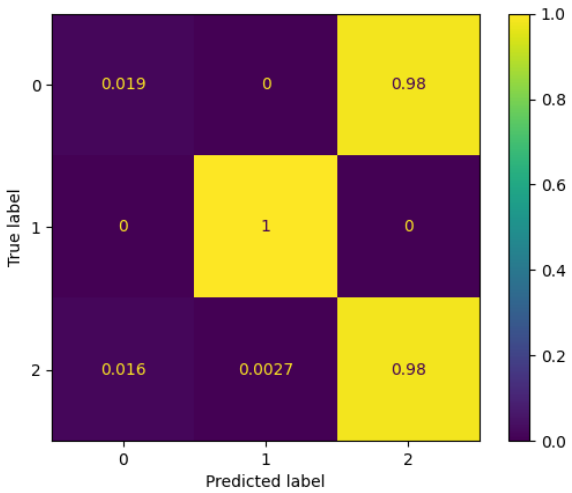

Figure 5: Model 1's confusion matrix and class-wise performance metrics for the multi-class classification component.

Observing the confusion matrix of the first model 5 on the test set, it becomes clear that the model adapted to the skewed dataset by predicting the label "trash", or 2, much more frequently. The qualitative similarities between images of clean and dirty water were learned by the model, which then predicted "trash" as, due to the underlying distribution of the dataset, the label was more likely to be the ground truth. The high false positive rate of model 1 was the impetus to train another model with the balanced dataset.

Class	Precision	Recall	F1-Score	Support
0 (Clean)	0.14	0.02	0.03	52
1 (Oil)	0.98	1.00	0.99	51
2 (Trash)	0.88	0.98	0.93	375
Accuracy		0.88		478
Macro Avg	0.67	0.67	0.65	478
Weighted Avg	0.81	0.88	0.84	478

Table 1: Model 1's statistics on Precision, Recall, and F1 Score for Classes

We can observe that based on Table 1, the class labeled 0 has very low precision, recall, and F1-score, indicating that the model performs poorly in identifying this class correctly. The precision of 0.14 means that only 14% of the instances predicted as class 0 are true positives, while the recall of 0.02 means that the model only identifies 2% of the actual instances of class 0 according to Table 1.

On the other hand, the class labeled 1 has extremely high precision, recall, and F1-score, suggesting that the model is highly accurate in identifying this class. The precision of 0.98 implies that 98% of the instances predicted as class 1 are true positives, and the recall of 1.00 means that the model correctly identifies all instances of class 1 according to Table 1.

The class labeled 2 also has relatively high precision, recall, and F1-score, with values around 0.90, indicating that the model performs reasonably well in identifying this class, but not as accurately as for class 1 according to Table 1. The overall accuracy of the model is 0.88, which is a decent value, but the macro average and weighted average metrics show lower scores, suggesting that the model's performance is not consistent across all classes according to Table 1.

The second model saw its train accuracy reach nearly 98% in Figure 6. At about 100 epochs, the train and test accuracy, which was tightly coupled, diverged and separated by around 20% for nearly another 100 epochs in Figure 6. This indicates that even stronger regularization might be needed, or a larger, more robust dataset to allow more generalizable learned relationships. A higher dropout rate might be sufficient or the use of a different loss function with a stronger regularization term, but the inclusion of another regularization method might prove most fruitful. Random crops of the dataset are not preferred due to the potential exclusion of areas of interest, such as pollutants, resulting in incorrect inference.

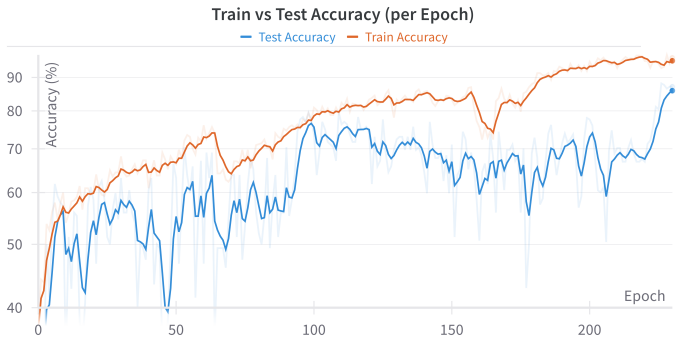

Figure 6: Model 2's training and test accuracy curves for the multi-class classification component.

The training was stopped when the test accuracy reached above 85% according to Figure 6; the final upward trajectory was due to hyperparameter tuning on the validation set, resulting in a much lower learning rate. This is corroborated by the training accuracy decreasing–the model becoming more general–and the test accuracy increasing dramatically, a reversal of potential overfitting.

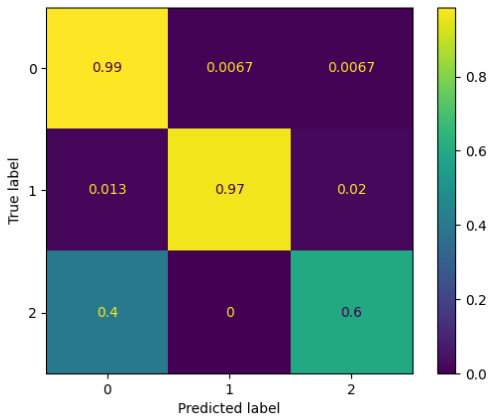

Figure 7: Model 2's confusion matrix and class-wise performance metrics for the multi-class classification component.

The confusion matrix shows the model's strong predictive capability on classes 0 and 1, with a less accurate prediction for class 2, the "trash" class according to Figure 7. In comparison to the first model's confusion matrix 5, model 2 would have more false negatives, but a greater overall accuracy. At their current state, both models provide benefits and drawbacks in the larger detection pipeline; false positives would allow the decision routing process to forward the image to the object detection model to make its inference where false negatives could have detection stop prematurely. Balancing computing and time is of the utmost importance and depends on the real-world requirements beyond the scope of this paper and the creation of this detection pipeline. Model 2 is more promising for further tuning, demonstrating the importance of a balanced dataset. However, real-world applications might prove the first model more useful as it experienced a larger dataset and might generalize better to unseen scenarios.

Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0.73	0.99	0.84	150
1	0.99	0.97	0.98	150
2	0.95	0.60	0.74	130
Accuracy		0.86		430
Macro Avg	0.89	0.85	0.85	430
Weighted Avg	0.89	0.86	0.86	430

Table 2: Model 2's statistics on Precision, Recall, and F1 Score for Classes

The class labeled 0 has a high recall of 0.99, indicating that the model correctly identifies most instances of this class according to Figure 2. However, its precision is relatively lower at 0.73, suggesting that the model may have a higher number of false positives for this class. The F1-score of 0.84 represents a balance between the precision and recall values and the support value of 150 indicates the number of instances of this class in the dataset according to Figure 2.

For class 1, the precision of 0.99 and recall of 0.97 are both very high, demonstrating that the model is highly accurate in identifying instances of this class according Table 2. The F1-score of 0.98 also reflects this excellent performance. The support value of 150 is the same as for class 0 according to Figure 2. Class 2 has a high precision of 0.95, but a lower recall of 0.60 according to Table 2. This means that while the model is quite precise in identifying instances of this class, it misses a significant portion of the actual instances. The F1-score of 0.74 reflects the trade-off between precision and recall. The support value of 130 is slightly lower than the other two classes.

The overall accuracy of the model is 0.86, which is a decent value according to Figure 2. The macro average and weighted average metrics also show higher scores, suggesting that the model's performance is more consistent across all classes.

4.2 Object Detection

4.2.1 Experiments

When experimenting with the hyperparameters during training, the parameter we kept constant was the optimizer (Adam), and the parameters we changed were the learning rate, batch size, image size, and epochs. The following are the results with each parameter change.

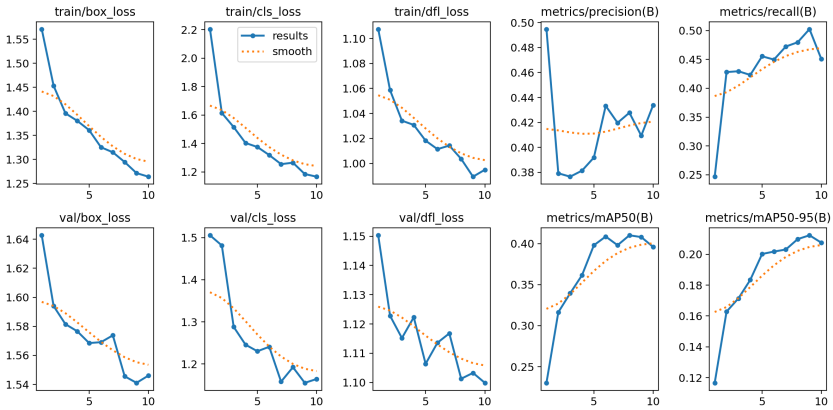

Figure 8: Results of trained models with a learning rate of 1e-4, batch size of 4, image size of 1280x720, and 10 epochs

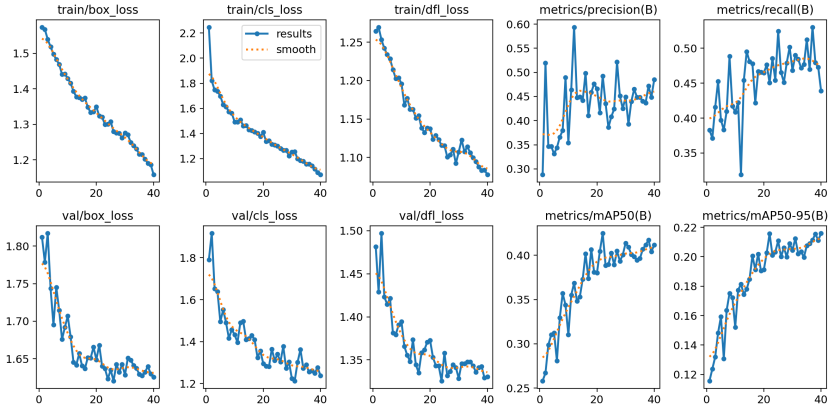

Figure 9: Results of trained models with a learning rate of 1e-3, batch size of 4, image size of 1280x720, and 40 epochs

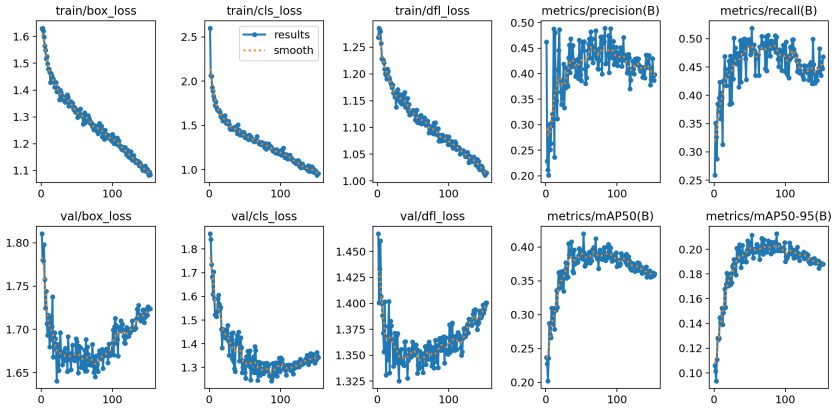

Figure 10: Results of trained models with a learning rate of 1e-3, batch size of 2, image size of 1920x1080, and 155 epochs

From our experimentation when we trained our model, we realized that a lower learning rate, a higher batch size, lower detailed images (not too high resolution), and medium epoch actually resulted in better results. This may be because our model was overfitting to the training data, as seen in the

losses plateauing or increasing in Figure 10 So, for the results of the paper, we will be basing the results from our second model, or the model from Figure 9.

4.2.2 Results and Analysis

As explained previously, we will represent our results through metrics such as accuracy, precision, recall, and F1 scores. The advantages of each of these evaluation metrics were explained in Section 4, Experiments, Results, & Analysis.

The following is the Figure for the accuracies of our best-performing model.

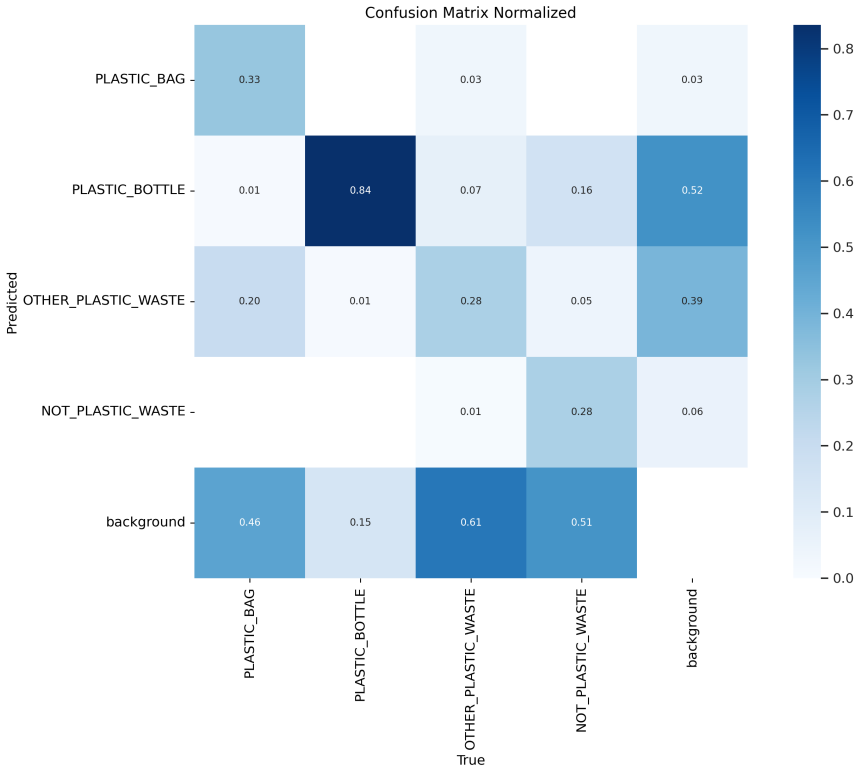

Figure 11: Normalized confusion matrix for the object detection component, illustrating the accuracies between predicted classes

The confusion matrix in Figure 11 illustrates the model's performance in distinguishing different objects, with high true positive rates for primarily "PLASTIC_BOTTLE". The main misidentification is "PLASTIC_BAG", being predominantly misidentified with "OTHER_PLASTIC_WASTE", but generally, when an object is detected they all have greater than a 50% chance of properly detecting the right object, where other than "PLASTIC_BAG" ranges from 57% to 98% accuracy, with plastic bottle identification being the upper bound and not plastic waste being the lower bound.

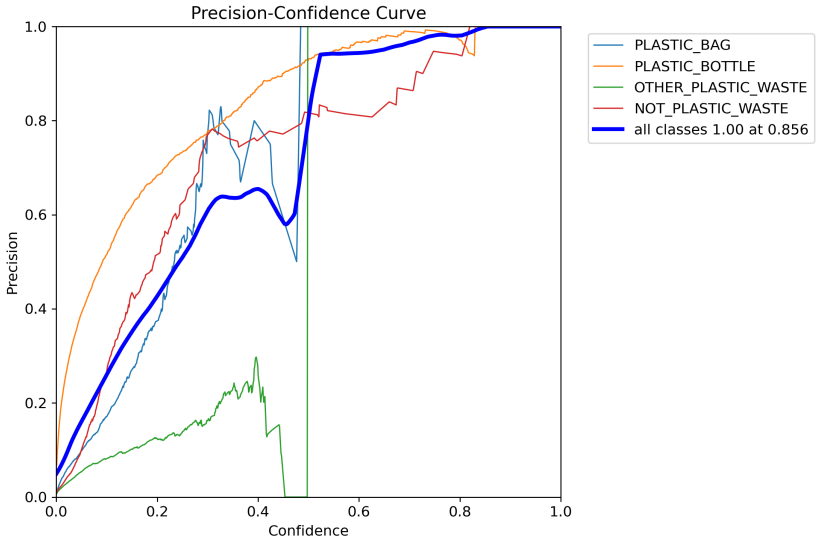

Figure 12: Precision for the best-trained object detection model

This precision curve indicates the ability to correctly identify a specific object out of all instances predicted within this object class. As seen in this, all classes, other than other plastic waste have really good precision.

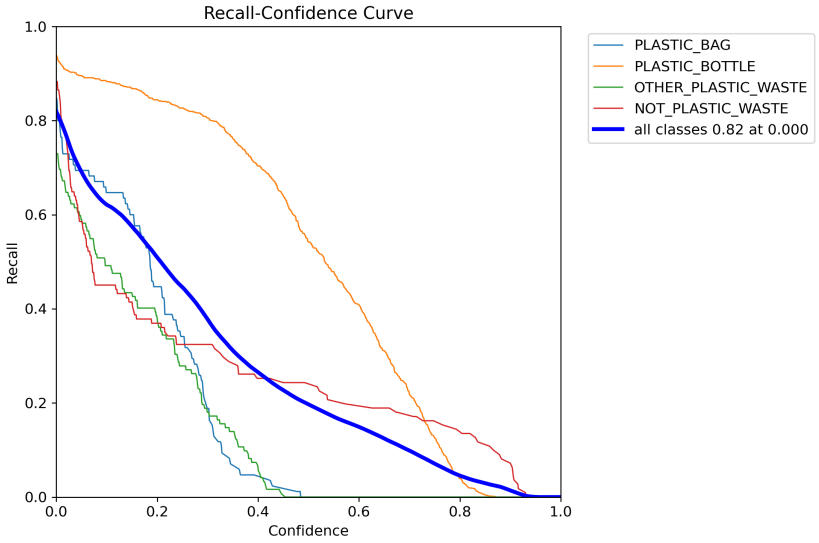

Figure 13: Recall for the best-trained object detection model

The recall curve represents the model's ability to identify a particular object from the entire dataset. In this, the recall does extremely well when identifying plastic bottles within the data set, and decent on the rest. This may be because plastic bottles are the most predominant part of this dataset.

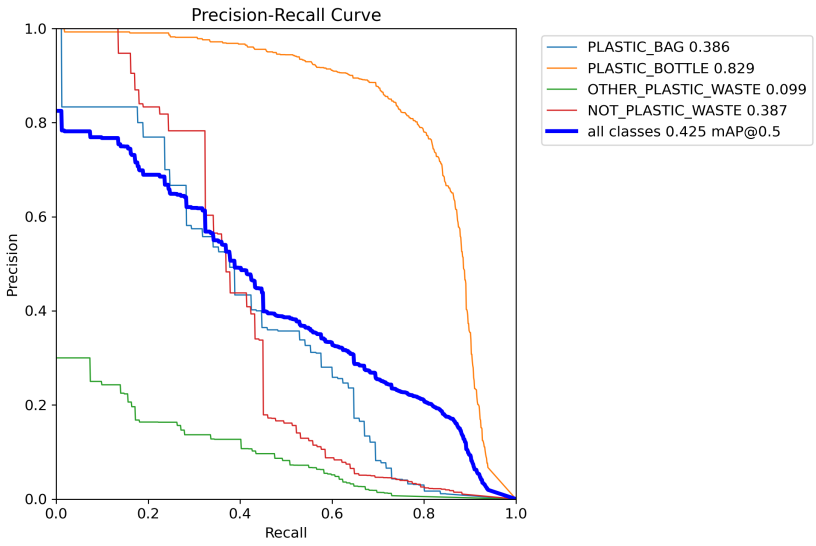

Figure 14: Precision-Recall for the best-trained object detection model

The curve in Figure 14 shows the tradeoff between precision and recall for different thresholds. A high area under the curve represents a high recall and precision. Plastic bottle has a good mix of both, other plastic waste and plastic bag has a medium mix, and other plastic waste has the lowest.

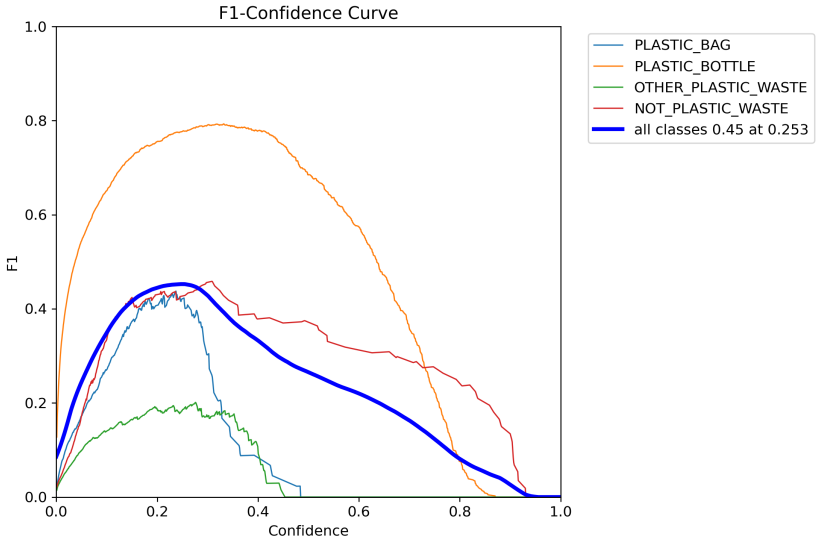

Figure 15: F1-Scores for the best-trained object detection model

The F1 Score provides a balanced metric that considers both false positives and false negatives. Indicative in Figure 15, the plastic bottle has a high F1 score, which means that the model achieves a good balance between precision and recall for identifying plastic bottles. The model has a low F1 Score for other plastic waste, meaning that it lacks precision and recall for identifying other plastic waste. As for identifying plastic bags or not plastic waste, it has a medium score, hence it is okay to identify plastic waste, beyond plastic bottles. Thus, our model does relatively well on plastic waste, performs extremely well when detecting plastic bottles, and is relatively poor on other plastic waste.

To increase the odds of identifying other plastic waste more properly, more classes can be added to mitigate these issues, so it can specifically choose what can be identified as "other plastic waste".

Figure 16: The validation image with the correct labels.

Figure 17: The validation image with bounding boxes predicted by the model.

Examples of successful object detection on the validation dataset can be seen in Figure 17 demonstrating the model's generalization capabilities across diverse scenarios.

Now, when directly comparing Figure 16 to Figure 17, we can see that our model predicts the majority if not all the waste in the images. One of the limitations of our model is that it misidentifies some of the objects, classifying them more regularly as plastic bottles when it isn't. Additionally, another issue is when the model detects the reflection in the water as potential trash, which is another limitation of this model.

Proven from our quantitative and qualitative evidence, this model performs especially well when detecting plastic bottles, mediocrely when identifying plastic bags and not plastic waste, and relatively poorly when identifying other plastic waste.

Thus, this model is generally good at identifying where trash is in an image (creating binding boxes), but it lacks in its classification, as it sometimes does not properly classify objects according to their label.

4.3 Challenges

4.3.1 General

One of the primary challenges faced in this research was the limited availability of large-scale, diverse, and accurately labeled datasets for training and evaluating the models. This posed a significant bottleneck in developing robust and generalizable models capable of handling the varying environmental conditions encountered in real-world scenarios. Such conditions, including varying lighting, water turbidity, surface reflections, and occlusions, impacted the model's performance and highlighted the need for techniques to enhance robustness and generalization.

4.3.2 Multi-class classification

For the multi-class classification component, the class imbalance in the initial dataset proved to be a significant hurdle. The first model struggled to distinguish between clean water and water with trash due to the skewed distribution, leading to a high false positive rate for the "trash" class. This issue prompted the creation of a more balanced dataset for training the second model. However, even with the balanced dataset, the second model exhibited signs of overfitting, with a significant gap between train and test accuracy. This observation highlighted the need for stronger regularization techniques, such as higher dropout rates or different loss functions with regularization terms, to improve the model's generalization ability and prevent overfitting.

The high accuracy demonstrated by the first model embodies a success in being able to identify trash in bodies of water and a failure in being able to distinguish between clean water and water with trash. It should be noted that, when observed in the larger model pipeline where the classification output becomes the input for a routing decision, false positives are more agreeable than false negatives.

4.3.3 Object Detection

The object detection component faced challenges in accurately classifying objects, with the model misidentifying some objects more regularly as plastic bottles when they were not. Additionally, the model sometimes detected reflections in the water as potential trash, leading to false positive detections. Occlusions in the images also posed challenges for accurate object localization, suggesting the need for techniques like attention mechanisms or multi-scale feature extraction to improve performance.

During the training process for the object detection model, finding the optimal hyperparameters and avoiding overfitting were significant challenges. The paper indicates that lower learning rates, higher batch sizes, and medium epochs resulted in better performance, highlighting the complexities involved in tuning the model architecture and training parameters.

4.4 General Results

Overall, the developed computer vision system demonstrated promising performance in detecting and classifying various types of water pollutants, contributing to environmental monitoring and remediation efforts.

5 Conclusion

The research successfully developed a robust and accurate computer vision system capable of detecting and classifying various types of water pollutants and waste materials present in images captured from rivers, lakes, and coastal waters. By leveraging state-of-the-art machine learning techniques like deep residual learning [6] and ensemble deep learning [5], and addressing challenges such as varying environmental conditions and class imbalances, this system contributed significantly to environmental monitoring and remediation efforts. By combining the classification and object detection models, we can enhance the practical applications of this technology in real-world scenarios. After classifying the dataset to identify if the water is clean, has oil, or has trash, object detection can precisely locate, create boundary boxes, and quantify these trash pollutants within the images. This combined approach not only facilitates cleanup efforts by selectively pinpointing areas of contamination but also provides valuable insights into locations that require significant assistance.

Furthermore, the scalability of this system allows for its deployment in geographical areas across the world, enabling comprehensive monitoring of water bodies. Along with the increasing availability of high-resolution satellite imagery and drone footage, technology has further highlighted the potential of such computer vision techniques for environmental monitoring applications [7, 15]. Automated systems like drones and boats equipped with these computer vision capabilities can play a crucial role in the fight against plastic pollution [15, 16].

However, developing robust and accurate models for detecting and classifying water pollutants posed several challenges. One limitation was the reliance on a limited dataset, which may not fully capture real-world diversity. Future work could focus on curating larger and more diverse datasets, potentially leveraging crowdsourcing or collaboration with environmental organizations [1]. Additionally, the system's performance may be impacted by adverse weather conditions or occlusions, prompting the need for techniques like attention mechanisms, multi-scale feature extraction, or instance segmentation [2, 12].

Furthermore, the current system focuses on a predefined set of pollutant types. Future research could explore open-set recognition capabilities, allowing the system to identify and adapt to new or unseen types of pollutants without extensive retraining [4]. Real-world deployment may also require considerations like computational resource optimization, real-time processing, and seamless data integration with existing systems [2, 3].

Despite these limitations, the research represents a significant step toward leveraging machine learning and computer vision for environmental monitoring and remediation efforts. By addressing the challenges identified and incorporating suggested improvements, future iterations could contribute to more comprehensive, accurate, and sustainable monitoring of water bodies, supporting the preservation of aquatic ecosystems and the well-being of communities relying on these vital resources [7, 15].

References

[1] Zunaira Asif, Zhi Chen, Chunjiang An, and Jinxin Dong. Environmental impacts and challenges associated with oil spills on shorelines. *Journal of Marine Science and Engineering*, 10(6), 2022.

[2] Luya Chen and Jianping Zhu. Water surface garbage detection based on lightweight yolov5. *Scientific Reports*, 14(1):6133, 2024.

[3] Jinhao Fan, Lizhi Cui, and Shumin Fei. Waste detection system based on data augmentation and yolo_ec. *Sensors*, 23(7), 2023.

[4] Sara Freitas, Hugo Silva, and Eduardo Silva. Hyperspectral imaging zero-shot learning for remote marine litter detection and classification. *Remote Sensing*, 14(21), 2022.

[5] Mudasir Ahmad Ganaie, Minghui Hu, Mohammad Tanveer, and Ponnuthurai N. Suganthan. Ensemble deep learning: A review. *CoRR*, abs/2104.02395, 2021.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[7] M. Hino, E. Benami, and N. Brooks. Machine learning for environmental monitoring. *Nature Sustainability*, 1(10):583–588, 2018.

[8] Jenna R. Jambeck, Roland Geyer, Chris Wilcox, Theodore R. Siegler, Miriam Perryman, Anthony Andrady, Ramani Narayan, and Kara Lavender Law. Plastic waste inputs from land into the ocean. *Science*, 347(6223):768–771, 2015.

[9] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLO, January 2023.

[10] Periyadan Krishnakumar. *Environmental impacts of marine pollution- effects, challenges and approaches.*, pages 731–739. 01 2017.

[11] Wanqi Ma, Hong Chen, Wenkang Zhang, Han Huang, Jian Wu, Xu Peng, and Qingqing Sun. Dsyolo-trash: An attention mechanism-integrated and object tracking algorithm for solid waste detection. *Waste Management*, 178:46–56, 2024.

[12] Mariano Morell, Pedro Portau, Antoni Perelló, Manuel Espino, Manel Grifoll, and Carlos Garau. Use of neural networks and computer vision for spill and waste detection in port waters: An application in the port of palma (majorca, spain). *Applied Sciences*, 13(1), 2023.

[13] Juan Terven, Diana-Margarita Córdova-Esparza, and Julio-Alejandro Romero-González. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Machine Learning and Knowledge Extraction*, 5(4):1680–1716, November 2023.

[14] G. G. N. Thushari and J. D. M. Senevirathna. Plastic pollution in the marine environment. *Heliyon*, 6(8):e04709, 2020.

[15] Jingbo Wang, Kaiwen Zhou, Wenbin Xing, Huanhuan Li, and Zaili Yang. Applications, evolutions, and challenges of drones in maritime transport. *Journal of Marine Science and Engineering*, 11(11), 2023.

[16] Christopher F Wooldridge, Christopher McMullen, and Vicki Howe. Environmental management of ports and harbours—implementation of policy through scientific monitoring. *Marine Policy*, 23(4-5):413–425, 1999.

[17] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *CoRR*, abs/1911.02685, 2019.

[18] Khaled Ziani, Corina-Bianca Ioniță-Mîndrican, Magdalena Mititelu, Sorinel Marius Neacșu, Carolina Negrei, Elena Moroșan, Doina Drăgănescu, and Olivia-Teodora Preda. Microplastics: A real global threat for environment and food safety: A state of the art review. *Nutrients*, 15(3), 2023.