Research & Reviews: Journal of
Space Science & Technology
https://journals.stmjournals.com/rrjosst

ISSN: 2321-2837 (Online)
ISSN: 2321-6506 (Print)
Volume 14, Issue 1, 2025
January–April
DOI (Journal): 10.37591/RRJoSST

Review RRJoSST

From Space to Sea: Leveraging Satellite Technology for Monitoring Marine Debris

K. Anupriya[1,*], Dhivyaa S.[2], Hemalatha M.[2], Mahaa Poorani S.[2]

Abstract

Marine trash endangers ecosystems, so efficient detection is critical. This article describes a novel strategy for improving detection accuracy by integrating YOLOv7 instance segmentation with attention processes. Three models are evaluated: lightweight coordinate attention, the convolutional block attention module (CBAM) for spatial-channel focus, and the bottle neck transformer, which relies on self-attention. On an annotated satellite image dataset, CBAM has the greatest F1 scores in box recognition (77%) and mask evaluation (73%), outperforming coordinate attention and YOLOv7, which score about 71% and 68% to 69%, respectively. Although the bottle neck transformer has lower overall accuracy, it excels at detecting larger trash that human annotators miss, indicating its promise for applications that require great precision for large particles. These findings highlight the necessity of adopting models that are tailored to specific environmental monitoring objectives. Broken bottles, plastic toys, food wrappers during a walk along the coast one finds any of these items, and more. In all that litter, there is one item more common than any other: cigarette butts. Cigarette butts are a pervasive, long-lasting, and a toxic form of marine debris. The Marine Debris Act requires the program to "identify, determine sources of, assess, prevent, reduce, and remove marine debris and address the adverse impacts of marine debris on the economy of the United States, marine environment, and navigation safety".

Keywords: Marine debris, YOLO (you look only once) technology, coordinate, spatial and channel attention, convolutional block attention module (CBAM)

INTRODUCTION

Marine debris, primarily plastic, is defined as any persistent solid item dumped into marine habitats, often resulting in significant pollution zones. Plastic's longevity and lightweight characteristics cause it to amass over time, resulting in places such as the Great Pacific Garbage Patch (GPGP), which spans around 1.6 million km^2 and contains hundreds of metric tons of waste. This accumulation harms marine ecosystems; for example, black-footed albatrosses in Midway Atoll, which is adjacent to the GPGP, have high levels of hazardous chemicals in their livers. The contamination extends to human health, with plastics detected in drinking water, salt, and soil, posing developmental, reproductive, and carcinogenic concerns [1].

Traditionally, identifying maritime trash has relied on manual surveys, which are expensive and limited. However, advances in machine learning technology have opened up new opportunities. While drones have showed promise, satellite imagery provides a broader perspective, albeit resolution limits continue to present issues. Our study overcomes these issues by utilizing the

***Author for Correspondence**
K. Anupriya
E-mail: anupriyaaiml@mvit.edu.in

[1]Assistant Professor, Department of Artificial Intelligence and Machine Learning, Manakula Vinayagar Institute of Technology, Kalitheerthalkuppam, Puducherry, India
[2]Student, Department of Artificial Intelligence and Machine Learning, Manakula Vinayagar Institute of Technology, Kalitheerthalkuppam, Puducherry, India

Received Date: January 10, 2025
Accepted Date: January 17, 2025
Published Date: January 25, 2025

Citation: K. Anupriya, Dhivyaa S., Hemalatha M., Mahaa Poorani S. From Space to Sea: Leveraging Satellite Technology for Monitoring Marine Debris. Research & Reviews: Journal of Space Science & Technology. 2025; 14(1): 1–9p.

YOLOv7 instance segmentation model for exact detection of maritime trash in satellite pictures. Furthermore, we incorporate attention processes to improve detection accuracy, resulting in a system that is economical and accessible to even the smallest environmental organizations. Finally, we want to create a cost-effective and efficient instrument for monitoring and reducing marine pollution [2].

RELATED WORK

Current Methods

With increased emphasis to environmental preservation, much study has been conducted to understand the sources and locations of marine trash, notably plastics. Traditional approaches for estimating this detritus have relied on manual methods. The National Oceanic and Atmospheric Administration (NOAA) runs the Marine Debris Monitoring and Assessment Program (MDMAP), which works with partners and volunteers throughout the world to conduct regular coastline surveys. Participants are given standardized monitoring kits that specify procedures for collecting and recording data on the kind and quantity of debris. This uniformity ensures the quality and comparability of the data obtained. Since its inception in 2012, MDMAP has completed 4421 surveys at 335 monitoring sites in nine countries. Cigarette butts, food packaging, drinking bottle tops, and plastic bags are among the most common types of garbage detected near coastlines, according to research [3].

NOAA's research emphasizes the value of in situ monitoring for assessing plastic trash in marine habitats. On-site measurements allow for precise assessments of the spatial and temporal distribution of trash, revealing patterns influenced by ocean currents. Debris under ocean is shown in Figure 1. These findings also encourage research into the negative consequences of plastic garbage on marine organisms and ecosystems, which is critical for risk assessments. However, field work necessitates significant labor, equipment, and financial resources, restricting coverage. Differences in participant abilities and motivation can result in uneven data collection and procedure adherence, emphasizing the need for alternate alternatives. Recent technological advancements have introduced machine-learning methods, particularly convolutional neural networks (CNNs), to detect debris in visual data from drones, satellites, and vessels, demonstrating promising potential for addressing marine debris pollution using computer vision techniques [4].

Figure 1. Debris under ocean.

Drones and aerial photography are increasingly used to monitor marine debris accumulation along coastlines. For example, Ellipsis Earth employs drones equipped with cameras to map plastic pollution, using deep learning-based image recognition to identify the type and size of debris, and sometimes even the brand or source. Drones capture high-speed video in inaccessible areas, and artificial intelligence (AI) analyzes the recorded data for debris detection. Some advanced algorithms allow for real-time recognition of images transmitted from drones. Research by Song et al. [5] indicated that drone mapping can be more accurate than traditional methods, with a U-Net-based segmentation model achieving F1 scores of 0.97 for polypropylene foam and 0.93 for plastic. However, drones are limited to coastal areas and cannot detect microplastics—particles smaller than 5mm, of which an estimated 14 million tons may be on the ocean floor [6].

Another method for identifying larger debris uses onboard cameras as an alternative to traditional neuston trawls. This approach employs deep learning algorithms to detect plastic fragments from high-resolution images captured by cameras like Go Pro, tagging images with GPS (global positioning system) data to quantify surface debris in the ocean. This method proves to be able to accommodate perspective than drones, allowing for more precise estimates of debris density [7].

Current methods for identifying maritime debris, such as unmanned aerial vehicles, vessel-mounted cameras, and satellite photography, each have advantages and disadvantages. Unmanned aerial vehicles (UAVs) and ship-based cameras provide great precision and precise data, but the infrared range (IR) range and cost limit their ability to monitor huge areas. Satellite imaging, on the other hand, covers a larger area but has inferior resolution and data. Most existing algorithms for recognizing plastic particles in satellite photos focus on coastal locations, limiting the full potential of satellite photography. Thus, evaluating the YOLO (you only look once) algorithm's effectiveness on high-resolution, surface-oriented satellite photos is critical for creating a complete system capable of detecting marine debris across large areas while capitalizing on satellite technology's distinct benefits.

YOLO

YOLO is a CNN-based architecture that mimics a fully convolutional network. CNNs are artificial neural networks that extract local characteristics from preceding layers' fields.

After the feature extraction is completed, the algorithm maps the positional relationships between these local features [8].

In the YOLO framework, each neural network predicts multiple bounding boxes and class probabilities at the same time. This allows YOLO to process an image only once, yielding predictions in a single output. This approach improves detection speed and minimizes processing latency, allowing YOLO to identify and localize objects in near real time, making it perfect for applications such as object identification in streaming video.

The most recent version, YOLOv7, has a new model head that enables both object detection and instance segmentation. One of its key advantages is speed, since it can interpret images at an astonishing 155 frames per second, much outpacing other cutting-edge algorithms.

While its core idea is based on MaskR-CNN, YOLOv7 is more efficient and better suited to real-time applications [9].

During the COVID-19 pandemic, YOLO's efficiency was demonstrated in real-time mask detection tasks. A study by Liu and Ren [10] showed that YOLOv3 outperformed traditional methods like Faster R-CNN, achieving a higher F1 score while operating nearly 50% faster.

In the realm of marine debris detection, YOLO has been primarily used for object detection, notably in UAV aerial photography. Watanabe et al. used YOLOv3 to detect debris on beaches and nearby sea surfaces and achieved a mean average precision (MAP) of 77.2%, substantially higher than the 41.2% obtained with Faster R-CNN [11]. This research validated YOLO's effectiveness for marine debris detection, showcasing its superior accuracy and speed. Despite its advantages, the application of YOLO in instance segmentation remains limited. Most implementations rely on earlier versions like YOLOv3 and adaptations such as Poly-YOLO and Insta-YOLO which have not been thoroughly evaluated in real-world scenarios. This underscores the need for testing YOLOv7 for instance segmentation in practical applications to further establish capabilities in this domain [5].

METHODOLOGY

Gathering and Labeling Data

We need high-resolution photos to detect microscopic debris bits. We chose to use NASA satellite photos with a resolution of 3 meters [10], even though Sentinel-2 provides a resolution of more than 10 meters per pixel, which restricts the detection of smaller debris. Four spectral bands are shown in these 256 × 256 images: red, green, blue, and near-infrared. To personally inspect Planet Scope sceneries and verify the existence of marine trash, we used Planet Explorer, concentrating on optical channels (RGB). The majority of the photos show the Honduran Gulf Islands, where the trash is made up of wood, plastic, algae, and other human-made items. Pictures of coasts and surrounding were mostly left out because the goal of this study is ocean detection. The original dataset had 707 photos, as shown in Figure 2, but many of them were unusable because of unclear debris, too much cloud cover, or coastal areas. Images of beaches and surrounding areas were mostly left out of this study because its goal is ocean detection. Debris that only took up 2 to 9 pixels in the pictures was also momentarily eliminated. In the end, we used Robo flow in two rounds to manually choose and identify 321 photos (Figure 2). The dataset for this study was created by randomly selecting 249 photos from this pool for the training set (79%), and 68 images for the test set (21%).

The chosen image is on the left, and the manually annotated image of the polygon appropriate for instance segmentation is on the right.

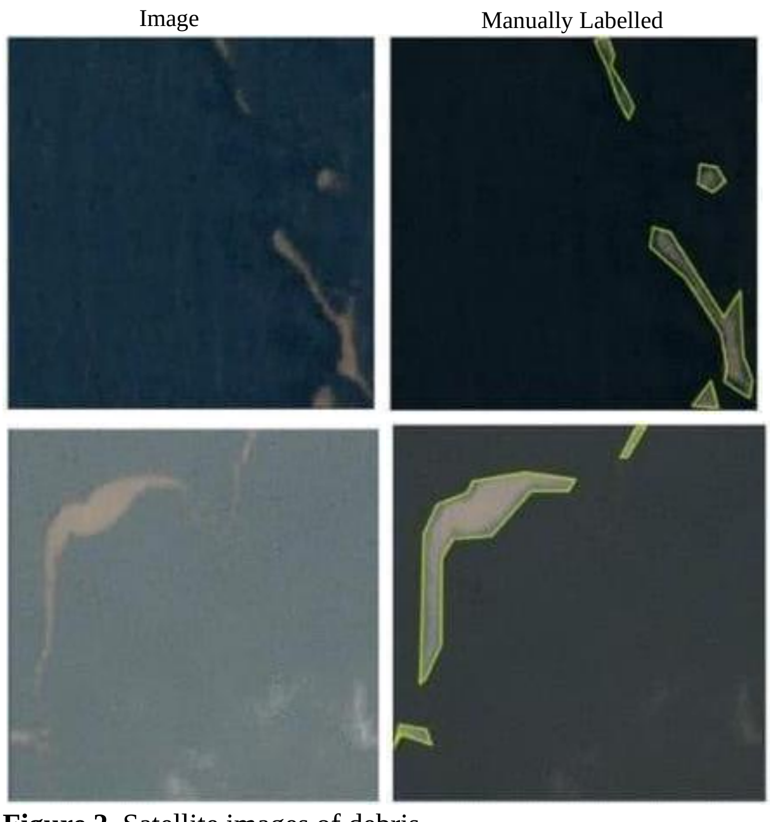

Figure 2. Satellite images of debris.

Segmenting YOLOv7 Instances

YOLOv7 is a single-stage instance segmentation technique that improves accuracy and efficiency by introducing a revolutionary Overlapping Double Structure (Overlapping Bilayers) to efficiently handle overlapping concerns. The four main parts of the YOLOv7 architecture are the input, backbone, neck, and head. Before being sent to the backbone, the image first goes through preprocessing in the input stage, which includes data augmentation. Features are then extracted from the processed image by the backbone. The neck module then fuses these features to produce multi-scale features for small, medium, and large objects.

Ultimately, the head receives the fused information and generates detection results [12]. Two-stage approaches and single-stage methods are the two categories into which this methodology can be divided. Architecture diagram of C2f is shown in Figure 3. Two-stage approaches create candidate boxes using a region proposal network (RPN) and then categorize and divide each one. Single-stage techniques, on the other hand, categorize and segment the entire image without first producing. Like YOLOv5, YOLOv7's neck component uses a feature pyramid network (FPN) and PANet. In the end, the PANet component maps bounding boxes and masks independently. Only two classes—debris and no debris—are used in this study because it focuses on debris identification; as a result, when mapping the feature layer for the bounding boxes, the outputs are 0 and 1. An instance segmentation version of YOLOv7 based on OpenCV and PyTorch is used in this work.

The figure shows the comparison with the C3 module used in YOLOv7, it is more light weight and has richer gradient.

Coordinate Attention Application

By adding position information, the Coordinate Attention Application developed a novel attention mechanism that improves channel attention. Coordinating attention splits the process into two 1-D feature encodings, capturing features along two spatial directions, in contrast to typical channel attention, which uses 2-D global pooling to form a single feature vector. This method allows the model to identify remote dependencies and preserve precise positional data. The target object's representation can be enhanced by successfully applying the direction-aware and location-sensitive attention maps that are produced to the input feature map. A coordinate attention block's structure is shown in Figure 4.

Convolutional Block Attention Module Application

The attention mechanism module of the convolutional module, which integrates spatial and channel attention, is known as the convolutional block attention module (CBAM). It can produce greater outcomes than the Senet's attention mechanism, which solely concentrates on channels. Prior to operation, the channel attention module obtains a 1-D vector by compressing the feature map in the spatial dimension. Combination of spatial and channel attention is shown in Figure 5. When compressing in spatial dimension, both the average pooling and the maximum pooling are considered. Maximum pooling only gives feedback to the gradient in the feature map with the largest response, whereas average pooling gives feedback to every pixel in the gradient backpropagation computation.

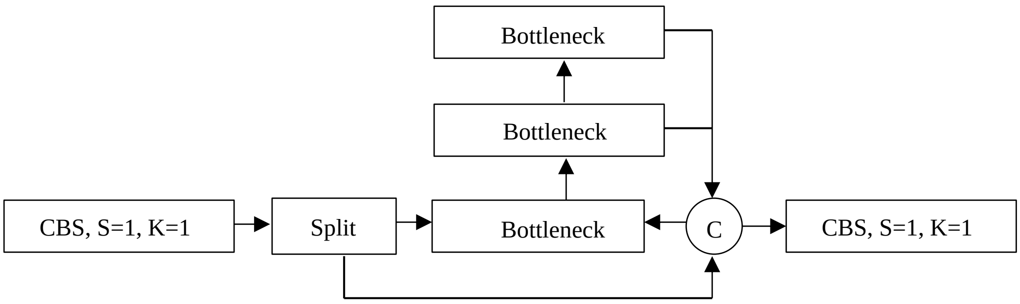

Figure 3. Architecture diagram of C2f.

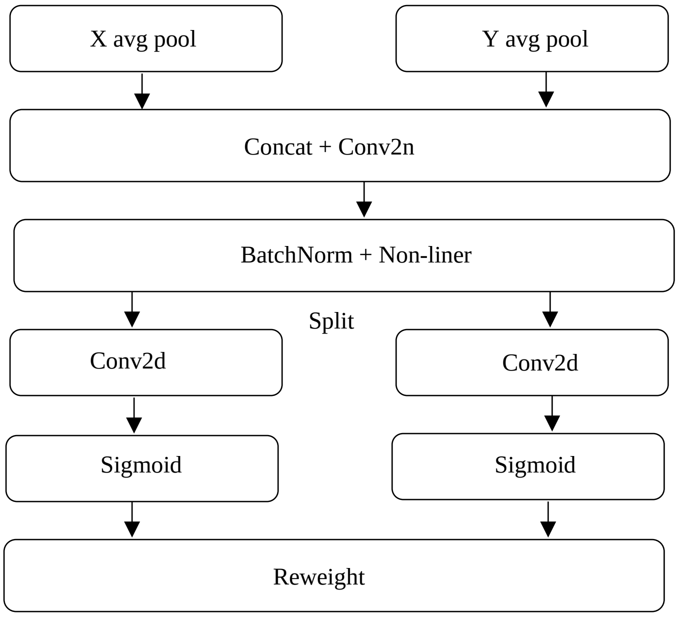

Figure 4. Schematic expression of coordinate attention.

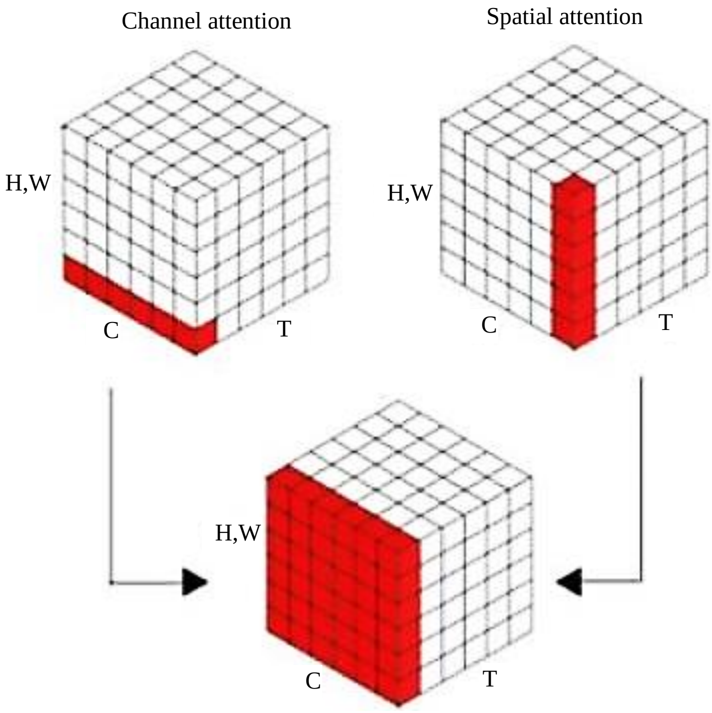

Figure 5. Combination of spatial and channel attention.

In the image, T stands for the temporal domain, H and W for the geographical domain, and C for the channel domain.

The Google Colab Python environment was used to implement all of the enhanced models, including YOLOv7 and YOLOv8. Testing showed only little modifications over the last 100 epochs, despite YOLOv7's typical run time of 299 epochs. To maximize efficiency within the video memory constraints, all models were trained for 199 epochs with a batch size of 4. hyp.scratch-high was the hyperparameter setting that was employed. All other training settings were maintained in order to guarantee uniform training duration among models, which is crucial for evaluating performance.

Research & Reviews: Journal of Space Science & Technology
Volume 14, Issue 1
ISSN: 2321-2837 (Online), ISSN: 2321-6506 (Print)

Interestingly, the CBAM model took about 4 hours to train, which is roughly twice as long as the other models. Structure diagram of CBAM is shown in Figure 6.

Self-Attention Application

One type of attention mechanism that reduces dependence on external input and is particularly good at identifying internal correlations between features or data is the self-attention mechanism, also known as the transformer. By determining the interaction between words, the self-attention mechanism has been used in text to overcome the long-distance dependence problem. Three vectors—query, key, and value—are the fundamental components of a text processing self-attention method. A SoftMax calculation will determine each word's relevance to the current word once the query is dot multiplied by the key to determine the word's score. The value of self-attention at the current point can be found by multiplying Value by Softmax and is shown in Figure 7.

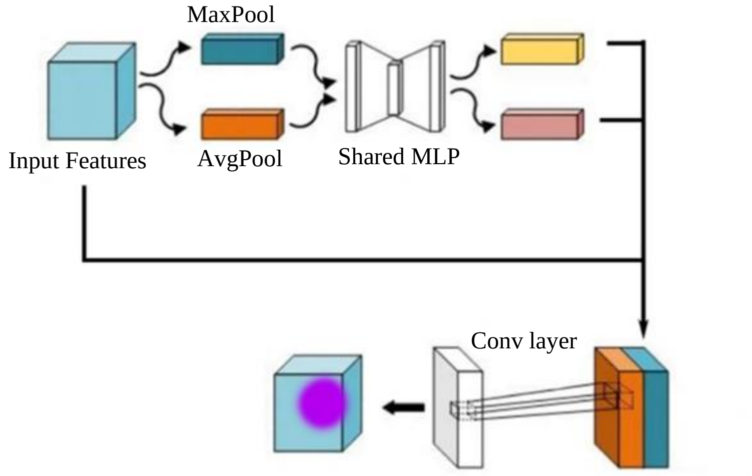

Figure 6. Structure diagram of convolutional block attention module (CBAM).

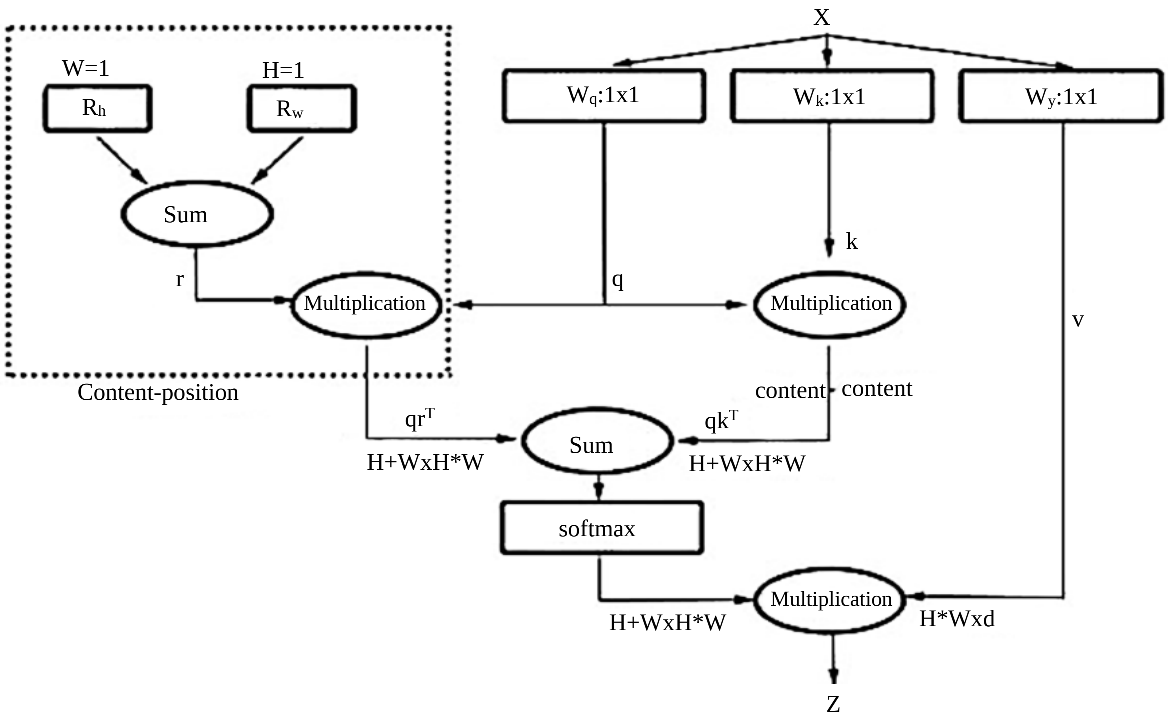

Figure 7. Self-attention block used in BoTNet.

CBAM Self-Attention

The self-attention model exhibits good plasticity, and the BAM model performs well. As a result, an effort was undertaken to integrate the two into a more effective framework. This design complements the two attention mechanisms by considering their respective capacities for internal and exterior information adoption. In order to pass the attention block based on external information, the input feature map will first proceed to CBAM. After that, it will return to self-attention to pass the attention block based on internal information.

In this study, this tentative design will be called the dual-attention model. For comparison, a version of this structure based on C2f modules is also available.

RESULTS

Setup for Experiments

The Google Colab Python environment was used to implement all the enhanced models, including YOLOv7 and YOLOv8. Testing showed only a few modifications over the last 100 epochs, despite YOLOv7's typical run time of 299 epochs. To maximize efficiency within the video memory constraints, all models were trained for 199 epochs with a batch size of 4. hyp.scratch-high was the hyperparameter setting that was employed. All other training settings were maintained to guarantee uniform training duration among models, which is crucial for evaluating performance. Interestingly, the CBAM model took about four hours to train, which is roughly twice as long as the other models.

Assessment

Since debris only takes up a small percentage of satellite images, high accuracy (over 95%) is maintained in this instance segmentation task, however false negatives (FNs) arise when the ocean background is mis identified. Both accuracy and recall are used, as well as the intersection over union (IoU) metric for object identification, to objectively evaluate model performance. Predictions are divided into true positive (TP), true negative (TN), false positive (FP), and false negative (FN) categories by confusion matrices. The ratio of matching pixels between the ground truth and predicted maps, or precision, indicates the percentage of accurately predicted samples. The CBAM model performed exceptionally well in the MASK aspect with a precision of 0.787, while the self-attention model had the maximum precision of 0.832 for the BOX aspect.

The percentage of true positive samples that were accurately recognized is known as recall. The best recall scores of 0.787 for BOX and 0.773 for MASK were obtained by the coordinate attention model with C2f, demonstrating its efficacy in detecting marine trash.

Since there is only one class in this study, mean average precision (MAP), which averages precision across classes, is equivalent to average precision (AP). The coordinate attention model received as core of 0.743 for MASK, while the dual-attention model with C2f achieved the maximum MAP of 0.773 for BOX.

With F1 values of 0.77 for BOX and 0.73 for MASK, the F-score—the harmonic mean of precision and recall—showed that the CBAM model performed well. Lastly, the overlap between the predicted and ground truth zones is measured by the intersection over union (IoU).

Analysis of the Results

Eight models, including those with the C2f module, were analyzed, and the results indicate that C2f mainly improves some metrics with little effect on overall performance. Asa result, models with both attention and C2f processes are seen as being on par with models that simply have attention. Dual-attention models include, for instance, the dual- attention model and its C2f variation. Even though the numerical differences might not seem like much, especially for MASK, the CBAM model performs the best overall, followed by the dual-attention model, coordinate attention model, and self-attention model.

Because of the self-attention barrier, the dual-attention model with CBAM performs worse and more closely resembles the coordinate attention model. It provides balanced recall and precision performance, but when MASK and BOX are taken into account simultaneously, bias is evident.

Although it takes around 25% longer to train than the self-attention model, it is comparable to the coordinated attention model, which is better when training expenses are an issue.

Although it performs worse than a number of enhanced models in this study, the YOLOv7 instance segmentation model is nevertheless useful for identifying trash in satellite photos of the sea surface, in contrast toYOLOv8. There is a need for more study because the self-attention model has serious flaws that affect its assessment, producing data that is not trustworthy and having less practical use.

CONCLUSION

To sum up, our study shows how important attention processes are for improving the precision of satellite imagery's ability to detect maritime debris. While acknowledging the usefulness of the bottleneck transformer in particular situations, we find CBAM to be the most efficient model. In addition to highlighting immediate applications, this study sets the stage for future research, highlighting the necessity of integrating spectrograms, oceanographic knowledge, and varied datasets. Our work supports international efforts to fight ocean pollution by helping to create reliable marine debris recognition technologies.

REFERENCES

1. Shen A, Zhu Y, Angelov P, Jiang R. Marine debris detection in satellite surveillance using attention mechanisms. IEEE J Select Top Appl Earth Observ Remote Sensing. 2024; 17: 4320–4330.
2. McCauley DJ, Andrzejaczek S, Block BA, Cavanaugh KC, Cubaynes HC, Hazen EL, Hu C, Kroodsma D, Li J, Young HS. Improving ocean management using insights from space. Annu Rev Marine Sci. 2024; 17: 381–408.
3. Mahrad BE, Newton A, Icely JD, Kacimi I, Abalansa S, Snoussi M. Contribution of remote sensing technologies to a holistic coastal and marine environmental management framework: a review. Remote Sensing. 2020; 12 (14): 2313.
4. Barry K. Space debris mitigation and remediation: historical best practices and lessons learned for economically preserving and utilizing common areas. New Space. 2022; 10 (3): 246–258.
5. Song K, Jung JY, Lee SH, Park S, Yang Y. Assessment of marine debris on hard-to-reach places using unmanned aerial vehicles and segmentation models based on a deep learning approach. Sustainability. 2022; 14 (14): 8311.
6. Merlino S, Calabrò V, Giannelli C, Marini L, Pagliai M, Sacco L, Bianucci M. The Smart Drifter Cluster: monitoring sea currents and marine litter transport using consumer IoT technologies. Sensors. 2023; 23 (12): 5467.
7. Karakuş O. On advances, challenges and potentials of remote sensing image analysis in marine debris and suspected plastics monitoring. Front Remote Sensing. 2023; 4: 1302384.
8. Biermann L, Clewley D, Martinez-Vicente V, Topouzelis K. Finding plastic patches in coastal waters using optical satellite data. Sci Rep. 2020; 10 (1): 5364.
9. Cheng HD, Jiang XH, Sun Y, Wang J. Color image segmentation: advances and prospects. Pattern Recogn. 2001; 34 (12): 2259–2281.
10. Liu R, Ren Z. Application of Yolo on mask detection task. In2021 IEEE 13th International Conference on Computer Research and Development (ICCRD) 2021 Jan 5 (pp. 130-136). IEEE.
11. Watanabe JI, Shao Y, Miura N. Underwater and airborne monitoring of marine ecosystems and debris. Journal of Applied Remote Sensing. 2019 Oct;13(4):044509.
12. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell. 2016; 39 (6): 1137–1149.

