Underwater Coral Morphology Classification And Tracking: Spatio-Temporal And IoU

Anupriya Gupta[1], Sakthivel Murugan S[2]
[1]*Dept. of Electronics and VLSI Design and Technology, Bangalore Institute of Technology, Bengaluru, India*
[2]*Dept. of Electronics and Communication Engineering, National Institute of Technical Teachers, Training and Research, Chennai, India*

***Abstract*—Automated analysis of underwater coral imagery is challenging due to variations in illumination, complex backgrounds, partial visibility, and the small size of some coral regions. This paper presents a coral-focused spatio-temporal pipeline for detecting and classifying coral morphology in underwater video. The approach uses an existing YOLO-based detector to identify coral regions, followed by overlapping tiled inference to improve spatial coverage. Detected coral regions are classified using a MobileNetV3-Small morphology classifier trained on 10,605 coral image patches across nine morphology categories. To improve temporal consistency, detections across consecutive frames are associated using Intersection over Union (IoU)-based tracking, and a stable morphology label is assigned at the track level. Additional filtering is applied to remove unrealistic, short-lived, and highly overlapping detections. The system was evaluated on a 911-frame underwater video and produced 2,867 candidate detections during tiled processing. After temporal association and subsequent cleanup, 1,443 detection records across 137 tracks were retained, with eight morphology categories represented in the final output. The best validation accuracy of the morphology classifier was 63.22%. The work demonstrates a practical adaptation of existing underwater computer-vision methods for coral morphology analysis while identifying the need for controlled baseline experiments to quantitatively evaluate the effect of spatial tiling and temporal stabilization.**

***Index Terms*—Coral morphology, underwater computer vision, YOLO, MobileNetV3-Small, tiled inference, object tracking, temporal consistency.**

I. INTRODUCTION

Underwater imagery is increasingly used for coral reef monitoring, ecological assessment, and marine biodiversity analysis. Coral reefs contain diverse organisms and structures whose identification is important for understanding reef condition and changes over time. However, manually analysing large amounts of underwater imagery and video is time-consuming and requires considerable effort. Automated computer-vision techniques can assist this process by detecting and classifying coral regions from underwater images and video.

Automated analysis of underwater scenes is challenging because of several environmental and visual factors. Variations in illumination, colour attenuation, low contrast, complex backgrounds, partial visibility, and overlapping objects can affect the performance of computer-vision models [7], [8]. Coral colonies may also appear small within a video frame, while different coral forms can have visually similar characteristics. These challenges make reliable coral detection and classification difficult, particularly when the objective is to analyse continuous underwater video rather than individual images.

Deep-learning-based object detection and image classification have been increasingly applied to underwater image analysis. Previous work has demonstrated the use of a two-stage approach in which an object detector first identifies regions of interest and a separate classification model performs further classification. Such a modular approach allows the detection and classification stages to be developed independently. Other underwater video classification studies have also highlighted the problem of frame-to-frame prediction fluctuations, where the predicted class of the same object may change across consecutive frames.

Building upon these existing approaches, the present work focuses on coral morphology classification in underwater video. The system uses a YOLO-based detector to identify coral regions and a MobileNetV3-Small classifier to classify the detected coral crops according to their morphological forms. To improve the detection of smaller or difficult coral regions, the video frames are processed using overlapping image tiles. Detections obtained from different tiles are converted into the original frame coordinates and overlapping detections are handled before classification.

Since the input is a video sequence, the same physical coral may be observed across multiple consecutive frames. Treating every frame independently can therefore lead to inconsistent morphology predictions. To address this issue, the proposed pipeline associates' detections across frames using Intersection over Union (IoU)-based tracking and derives a stable morphology label at the track level. Additional filtering is applied to remove unrealistic, short-lived, and highly overlapping detections.

The objective of this work is to investigate the practical adaptation of established underwater computer-vision techniques for coral morphology analysis and to examine whether combining spatial tiled inference with temporal tracking can provide more consistent and useful predictions in underwater video. The work does not propose a new YOLO or MobileNet architecture; instead, it integrates and adapts established components for the coral-specific video analysis problem.

II. LITERATURE SURVEY

Previous research in underwater computer vision has explored deep-learning-based object detection and classification for analysing underwater biological environments. These studies have demonstrated the usefulness of separating object localization from classification, while also identifying challenges caused by underwater image quality, complex backgrounds, overlapping objects, and variations in predictions across video frames. Deep-learning-based approaches have also been investigated specifically for automated coral classification, demonstrating the potential of deep neural networks for underwater coral image analysis [6], [10].

A two-stage underwater classification approach has been developed in which a YOLO-based detector first localizes objects and a MobileNet-based stage performs subsequent classification [1]. The separation of detection and classification provides a modular framework in which the classification component can be extended independently. The study also identifies temporal consistency in video as a possible direction for further improvement. This modular approach provides a useful foundation for adapting detection and classification methods to other underwater biological targets.

Another underwater classification study investigated fish and coral classification using deep-learning-based detection and classification [2], [9]. The work considered practical challenges such as poor illumination, overlapping objects, and interference from background coral structures. It also highlighted frame-to-frame prediction fluctuations and suggested the integration of tracking and temporal information to improve consistency in video-based classification. These observations are particularly relevant when the same biological object is observed across multiple consecutive frames.

Lightweight transfer-learning approaches have also been explored for underwater species classification under limited-data conditions [3]. MobileNetV3 provides a lightweight architecture suitable for computationally efficient image classification [5]. Such approaches demonstrate the practical use of pretrained and computationally efficient models for underwater image analysis. The use of lightweight architectures is also useful when developing practical classification pipelines where computational resources and training data may be limited.

Based on these studies, the present work adapts and integrates established detection and classification approaches for coral morphology classification in underwater video. Overlapping tiled inference is incorporated to improve spatial coverage, while IoU-based temporal tracking is used to associate detections across consecutive frames. Track-level morphology stabilization is further applied to reduce fluctuations in the predicted morphology of the same coral throughout the video. The resulting approach is therefore positioned as a coral-specific adaptation and integration of existing underwater computer-vision methods, rather than as a new detection or classification architecture.

III. PROPOSED SYSTEM

The proposed system is designed to automatically detect and classify coral regions in underwater video according to their morphological characteristics. The system follows a modular detection-classification framework in which coral localization and morphology classification are performed as separate stages. This design allows the detection and classification components to be developed and evaluated independently.

The input to the system is an underwater video sequence. Each video frame is processed using overlapping image tiles to improve the spatial coverage of coral detection, particularly for smaller coral regions. A YOLO-based detector is applied to the individual tiles, and the resulting bounding boxes are transformed into the coordinates of the original video frame. Overlapping detections are then handled to reduce duplicate coral regions.

The detected coral regions are cropped from the original frame and provided to a MobileNetV3-Small classifier trained to recognize coral morphology. The classifier considers nine morphology categories during training: Branching, Dead coral, Digitate, Encrusting, Foliose, Massive, Soft coral, Submassive, and Tabular. The classification output is subsequently associated with detections across consecutive video frames using IoU-based tracking.

To improve consistency throughout the video, detections belonging to the same tracked coral are grouped using their spatial overlap across frames. A stable morphology label is assigned at the track level based on the accumulated classification results. Additional filtering is applied to remove unrealistic bounding boxes, short-lived tracks, and highly overlapping detections. The final system produces both an annotated video and a structured CSV containing frame-level detection, tracking, and morphology information.

The overall processing flow of the proposed system is illustrated in Fig. 1. The pipeline combines spatial processing and temporal analysis to obtain more reliable coral morphology information from underwater video. Spatial processing is performed through overlapping tiled inference and coordinate-level duplicate removal, while temporal analysis is achieved through IoU-based tracking and track-level morphology stabilization. The final outputs provide both a visually annotated video and structured detection records for further analysis.

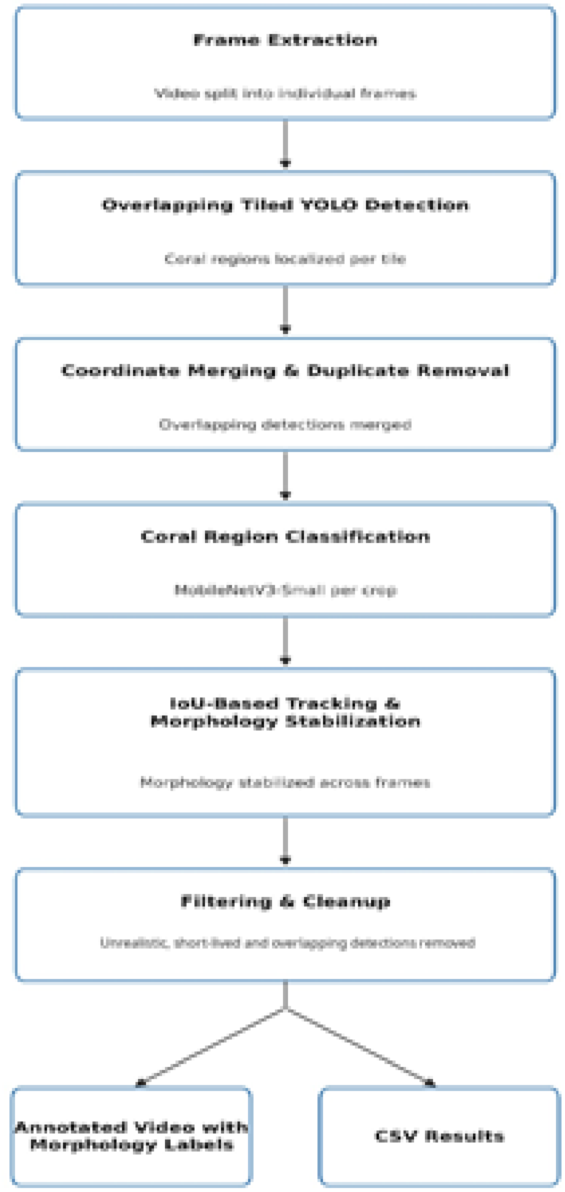


Fig. 1. Overall architecture of the proposed coral morphology classification system.

IV. METHODOLOGY

A. Overlapping Tiled Inference

Underwater video frames may contain coral regions that occupy relatively small portions of the complete image. Direct full-frame detection can therefore make it difficult to identify smaller or partially visible coral regions. To improve spatial coverage, each video frame is divided into overlapping tiles before detection.

In the implemented system, four overlapping tiles are used to process a frame of size 1280 × 720 pixels. The tile regions are defined as (0,0,760,460), (520,0,1280,460), (0,260,760,720), and (520,260,1280,720). The overlap between adjacent tiles allows coral regions near tile boundaries to be observed in more than one tile.

A YOLO-based detector is applied independently to each tile to obtain candidate coral bounding boxes. Since the detections are initially expressed in tile coordinates, the corresponding tile offsets are added to convert them into the coordinate system of the original video frame. This produces a common set of bounding boxes that can be processed across the complete frame. Because overlapping tiles can produce multiple detections for the same coral region, duplicate detections are subsequently identified and removed using Intersection over Union (IoU)-based overlap processing. The resulting detections are then used for coral-region cropping and morphology classification.

B. Coral Region Classification Using MobileNetV3-Small

The detected coral regions are cropped from the original video frames and processed using a separate image classification model. A MobileNetV3-Small network with pretrained ImageNet weights is used as the backbone for coral morphology classification. The classification head is modified to produce predictions for nine morphology categories: Branching, Dead coral, Digitate, Encrusting, Foliose, Massive, Soft coral, Submassive, and Tabular.

A dataset of 10,605 coral image patches was used for training the morphology classifier. The dataset was divided into 8,484 training samples and 2,121 validation samples. Data augmentation was applied during training using random resized cropping, horizontal and vertical flipping, rotation, and colour-based augmentation to improve the robustness of the classifier to variations in underwater imagery.

The classifier was trained using the AdamW optimizer with a learning rate of 3×10^{-4}, weight decay of 1×10^{-4}, a batch size of 64, and five training epochs. The best validation accuracy obtained during training was 63.22%.

During video processing, each detected coral region is cropped and resized to the input dimensions required by MobileNetV3-Small. The classifier then assigns a morphology label and corresponding confidence to each detected region. These frame-level classification results are subsequently passed to the temporal tracking stage to improve classification consistency across consecutive frames. This stage provides the input for subsequent temporal analysis.

C. IoU-Based Temporal Tracking

Since the input is a continuous underwater video, the same coral may appear in multiple consecutive frames. Processing each detection independently can result in repeated detections and inconsistent association of the same coral across frames. To address this, an IoU-based tracking procedure is used to associate spatially overlapping coral detections between consecutive frames.

For each new detection, its bounding box is compared with the detections from the previous frames using the Intersection over Union (IoU) measure. A detection is associated with an existing track when the spatial overlap satisfies the selected matching criterion. Each associated detection is assigned a common track ID, allowing observations of the same coral across multiple frames to be grouped together.

The implemented tracking stage uses an IoU matching threshold of 0.25 and permits a maximum gap of 8 frames when maintaining a track. Tracks with insufficient temporal duration are subsequently removed during the filtering stage. The resulting track information is stored along with the frame number, bounding-box coordinates, morphology label, and classification confidence.

This temporal association provides the basis for assigning a stable morphology label to each tracked coral in the subsequent stage.

D. Track-Level Morphology Stabilization

Frame-level morphology predictions may fluctuate when the same coral is observed across consecutive frames because of changes in viewpoint, illumination, partial visibility, or detection variations. To reduce such fluctuations, the morphology predictions associated with each track are combined over the observed frames.

For each tracked coral, the morphology labels obtained from the individual detections are accumulated throughout the track. A stable morphology label is then assigned based on the most frequently occurring morphology prediction within that track. This track-level assignment provides a consistent label for the same coral rather than allowing its morphology to change independently from frame to frame.

The stabilized morphology label is stored as stable_morphology in the tracking results and is used during the final filtering and video-rendering stages. This approach improves temporal consistency while retaining the original frame-level detections and classification information.

E. Filtering and Final Output Generation

Following detection, classification, and tracking, additional filtering is applied to remove detections that may reduce the reliability of the final results. Bounding boxes that are unrealistically large, tracks with insufficient temporal duration, and highly overlapping duplicate detections are removed during this stage.

The remaining detections are used to generate the final annotated underwater video, where each retained coral is displayed with its corresponding morphology label. The associated information, including frame number, track ID, bounding-box coordinates, morphology, classification confidence, and stable morphology label, is stored in a structured CSV file for further analysis.

This final stage produces the cleaned outputs used for the experimental evaluation and discussion presented in the following section.

V. RESULTS AND DISCUSSION

A. Morphology Classifier Results

The MobileNetV3-Small morphology classifier was trained using 10,605 coral image patches representing nine morphology categories. The dataset was divided into 8,484 training samples and 2,121 validation samples. During training, the validation accuracy increased from 50.40% in the first epoch to a maximum of 63.22% in the fourth epoch, while the training accuracy reached 70.13% by the fifth epoch.

The best validation accuracy of 63.22% indicates that the classifier was able to learn meaningful visual characteristics associated with the different coral morphology categories.

However, the result also indicates that further improvement is required for reliable morphology classification under challenging underwater conditions.

TABLE I. TRAINING AND VALIDATION PERFORMANCE OF THE MORPHOLOGY CLASSIFIER

Epoch	Training Acc. (%)	Validation Acc. (%)
1	45.76	50.40
2	58.75	58.60
3	63.37	61.72
4	67.43	63.22
5	70.13	63.13

B. Tiled Detection Results

The overlapping tiled detection stage was applied to the underwater video to improve the spatial coverage of coral detection, particularly for smaller or partially visible coral regions. Instead of processing the complete frame as a single input, each frame was divided into four overlapping regions. The overlapping arrangement allowed coral regions near tile boundaries to be observed in more than one tile and reduced the possibility of missing small coral regions.

The video consisted of 911 frames, and four overlapping tiles were processed for each frame using the YOLO-based detector. The detections obtained from individual tiles were transformed back to the original frame coordinate system, after which overlapping detections were handled to reduce duplicate predictions.

The tiled processing generated 2,867 candidate detections across 851 frames. These candidate regions were subsequently processed for morphology classification and temporal association.

TABLE II. MORPHOLOGY DISTRIBUTION ACROSS PIPELINE STAGES

Morphology	After Detection	After Tracking	After Filtering
Encrusting	1,098	1,012	662
Soft coral	822	663	474
Massive	296	162	133
Tabular	292	67	42
Foliose	140	74	47
Branching	105	75	68
Digitate	62	25	13
Submassive	35	11	4

Morphology	After Detection	After Tracking	After Filtering
Dead coral	17	—	—
Total	2,867	2,089	1,443

C. Temporal Tracking Results

Since the input is a continuous underwater video, the same coral may appear across multiple consecutive frames. Processing each detection independently can result in repeated detections and inconsistent association of the same coral over time. To address this issue, an IoU-based tracking procedure was used to associate spatially overlapping coral detections between consecutive frames.

For each new detection, the bounding box was compared with existing detections using the Intersection over Union (IoU) measure. A detection was associated with an existing track when the overlap satisfied the selected matching criterion. Each associated detection was assigned a common track ID, allowing observations of the same coral across multiple frames to be grouped together.

The tracking stage used an IoU matching threshold of 0.25 and allowed a maximum gap of 8 frames when maintaining a track. This process produced 2,089 detection records across 221 unique tracks, with detections present in 801 of the 911 video frames. The associated morphology predictions were subsequently used for track-level morphology stabilization.

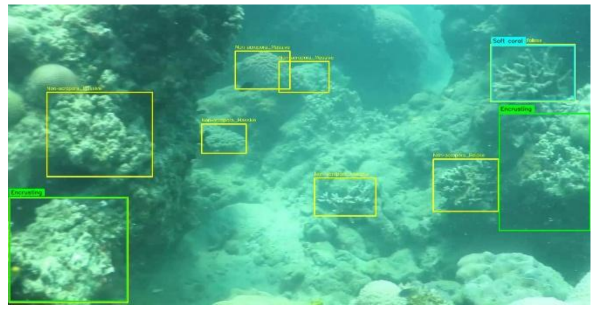
The morphology distribution at the detection, tracking, and final filtering stages is summarized in Table II.

D. Final System Results

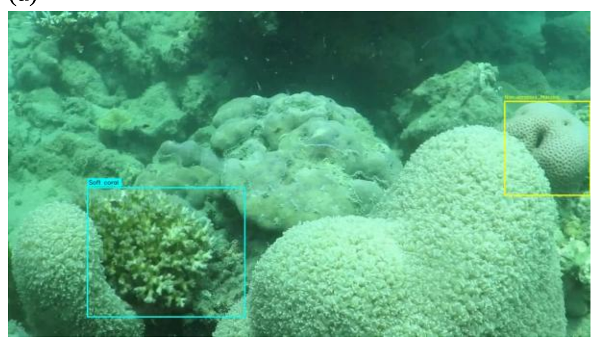
Following temporal tracking and morphology stabilization, an additional filtering and cleanup stage was applied to improve the reliability of the final detections. Bounding boxes that were unrealistically large, tracks with insufficient temporal duration, and highly overlapping duplicate detections were removed during this stage.

The final cleaned output contained 1,443 detection records distributed across 137 tracks. A minimum temporal duration of 3 frames was used when retaining tracks. The final output represented eight morphology categories: Encrusting, Soft coral, Massive, Branching, Foliose, Tabular, Digitate, and Submassive.

The resulting detections were used to generate the final annotated video, while the corresponding frame number, track ID, bounding-box coordinates, morphology, classification confidence, and stable morphology label were retained in the CSV results. The morphology distribution across the detection, tracking, and final filtering stages is presented in Table II.



(a)



(b)

(c)

Fig. 2. Representative frame from the final annotated coral morphology classification video: (a) 0:11, (b) 0:15, and (c) 0:30.

E. Discussion

The experimental results demonstrate the feasibility of combining object detection, morphology classification, and temporal analysis for coral-focused underwater video processing. The morphology classifier achieved a best validation accuracy of

63.22%, indicating that the trained model was able to learn discriminative visual characteristics from the coral image patches. However, the moderate validation performance also shows that morphology classification remains challenging under variations in illumination, viewpoint, partial visibility, and background conditions.

The tiled detection stage produced 2,867 candidate detections across 851 frames, providing a larger set of coral regions for subsequent classification. The use of overlapping tiles also introduced duplicate detections that required coordinate-level overlap handling, demonstrating the trade-off between spatial coverage and redundant predictions.

Temporal association grouped detections into tracks and provided a mechanism for maintaining the identity of coral regions across consecutive frames. The use of track-level morphology stabilization addresses the problem of frame-to-frame label fluctuations by assigning a consistent morphology label based on the predictions accumulated within each track. This is particularly relevant for underwater video, where the appearance of the same coral can change between frames.

After the final filtering and cleanup stage, 1,443 detection records across 137 tracks were retained. The reduction in detections indicates that spatial and temporal filtering removed a portion of the initial candidate outputs, including unrealistic bounding boxes, short-lived tracks, and highly overlapping detections. The final results therefore represent a cleaned subset of the system output rather than a direct measure of detection accuracy.

Although the results demonstrate a practical end-to-end pipeline, the present evaluation does not establish quantitative improvement over a simpler frame-wise baseline.

In particular, the effects of tiled inference and temporal stabilization have not yet been isolated through controlled comparative experiments. Future evaluation should therefore compare the proposed pipeline with a full-frame, frame-wise baseline using manually verified annotations and metrics such as detection recall, false detections, and morphology classification accuracy. Such comparison would provide stronger evidence for the effectiveness of the proposed adaptations.

VI. CONCLUSION AND FUTURE WORK

This paper presented a spatio-temporal pipeline for coral morphology classification in underwater video by integrating established object detection, image classification, and temporal tracking techniques. The system combines overlapping tiled inference for coral detection, MobileNetV3-Small-based morphology classification, IoU-based temporal tracking, track-level morphology stabilization, and post-processing filters for obtaining a cleaned video output.

The morphology classifier was trained using 10,605 coral image patches and achieved a best validation accuracy of 63.22%. Tiled processing generated 2,867 candidate detections across 851 frames, while temporal association produced 2,089 detection records across 221 tracks. Following filtering and cleanup, 1,443 detection records across 137 tracks were retained in the final output. These results demonstrate the feasibility of applying a modular deep-learning pipeline to coral-focused underwater video analysis.

The present work primarily demonstrates the adaptation and integration of existing computer-vision techniques rather than proposing a new detection or classification architecture.

A key limitation is that the individual contributions of tiled inference and temporal morphology stabilization have not yet been evaluated through a controlled baseline comparison. Future work will therefore focus on comparing the proposed pipeline with a simpler frame-wise approach using manually verified annotations and quantitative metrics.

Further improvements may also investigate larger and more diverse coral datasets, stronger morphology classification models, and more robust multi-object tracking methods for challenging underwater conditions.

ACKNOWLEDGMENT

This work was carried out as part of an internship at the Aazhi Research Laboratory, Department of Electronics and Communication Engineering, National Institute of Technical Teachers Training and Research (NITTTR), Chennai. The authors gratefully acknowledge the support of NITTTR Chennai.

REFERENCES

[1] V. Sundaram, *Open-Set Underwater Species Classification Using a Two-Stage YOLO–MobileNet Embedding Pipeline*, Internship Report, SSN College of Engineering, Kalavakkam, Chennai.

[2] T. Sibiram and K. Kiruba, Underwater Fish & Coral Species Classification Using Large Dataset, Internship Report, 2026.

[3] T. Sibiram and K. Kiruba, Image Classification Using Large Dataset, KPRIET presentation/report material.

[4] E. Arjun, Underwater Species Classification, Internship Report, 2026.

[5] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for MobileNetV3," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), 2019, pp. 1314–1324.

[6] A. Mahmood, M. Bennamoun, S. An, F. Sohel, F. Boussaid, R. Hovey, G. Kendrick, and R. B. Fisher, "Deep learning for coral classification," in Handbook of Neural Computation, pp. 383–401. Academic Press, 2017.

[7] A. Salman, A. Jalal, F. Shafait, A. Mian, M. Shortis, J. S. Lee, and E. W. K. Wong, "Fish species classification in underwater videos using deep convolutional neural network and data augmentation," in Proc. 2016 IEEE Int. Conf. Image Process. (ICIP), 2016, pp. 3483–3487.

[8] B. V. Deep and R. Dash, "Underwater fish species recognition using deep learning techniques," in Proc. 2019 6th Int. Conf. Signal Process. Integrated Netw. (SPIN), 2019, pp. 665–669, doi: 10.1109/SPIN.2019.8711657.

[9] V. Pagire and A. Phadke, "Underwater fish detection and classification using deep learning," in Proc. 2022 Int. Conf. Intelligent Controller and Computing for Smart Power (ICICCSP), 2022, doi: 10.1109/ICICCSP53532.2022.9862410.

[10] H. Zhang, M. Li, J. Zhong, and J. Qin, "CNet: A novel seabed coral reef image segmentation approach based on deep learning," in Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops, 2024, pp. 767–775, doi: 10.1109/WACVW60836.2024.00090.