

## Review

# Coral Visual Recognition for Marine Environmental Monitoring: A Systematic Review of Progress, Challenges, and Future Directions

Hu Liu <sup>1</sup>, Yinwei Luo <sup>2,\*</sup> , Qianyu Luo <sup>3,4</sup> , Yuelin Xu <sup>1</sup>, Xiuhai Wang <sup>5,6,\*</sup> and Xingsen Guo <sup>2,4,5,6,7,8,9,10</sup> 

<sup>1</sup> R&D Department, Qingdao Robotfish Marine Technology Co., Ltd., Qingdao 266100, China; lbfliuh@163.com (H.L.); xuy15665@163.com (Y.X.)

<sup>2</sup> Shandong Provincial Key Laboratory of Marine Engineering Geology and the Environment, Ocean University of China, Qingdao 266100, China; xingsen.guo@ouc.edu.cn

<sup>3</sup> School of Architecture, Technology and Engineering, University of Brighton, Brighton BN2 4GJ, UK; q.luo2@brighton.ac.uk

<sup>4</sup> Department of Civil, Environmental and Geomatic Engineering, University College London, London WC1E 6BT, UK

<sup>5</sup> Key Laboratory of Marine Environmental Science and Ecology, Ministry of Education, Ocean University of China, Qingdao 266100, China

<sup>6</sup> College of Environmental Science & Engineering, Ocean University of China, Qingdao 266100, China

<sup>7</sup> Sanya Oceanographic Institution, Ocean University of China, Sanya 572024, China

<sup>8</sup> Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK

<sup>9</sup> State Key Laboratory of Intelligent Construction and Healthy Operation and Maintenance of Deep Underground Engineering, Sichuan University, Chengdu 610065, China

<sup>10</sup> State Key Laboratory of Ocean Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

\* Correspondence: haut\_lyw@163.com (Y.L.); showseas@ouc.edu.cn (X.W.)

## Abstract

Coral reefs are among the most biodiverse marine ecosystems, playing irreplaceable roles in maintaining marine ecological balance and coastal services. Under dual pressures of global climate change and human activities, coral bleaching and degradation have become increasingly frequent, creating an urgent need for large-scale, long-term, and highly automated monitoring technologies. In recent years, advances in underwater imaging and deep learning have made visual recognition a core approach for coral classification and health assessment. However, most studies only focus on isolated model accuracy optimization, lacking systematic full-chain analysis integrating datasets, model evolution, cross-domain generalization, engineering constraints, and ecological adaptation, which severely hinders large-scale cross-regional and long-term application. This paper systematically reviews coral visual recognition technologies. It summarizes underwater image acquisition, public dataset characteristics, and annotation system evolution, then compares traditional feature engineering and deep learning in key tasks, highlighting their differences in feature representation and generalization. Four core challenges are identified: class imbalance, poor underwater image quality, weak cross-device/region generalization, and mismatched algorithm metrics with ecological needs. Finally, feasible solutions based on self-supervised pre-training, domain adaptation, and multimodal fusion are discussed to enhance model robustness and ecological interpretability, providing methodological support for intelligent coral reef monitoring systems.



Academic Editors: Tom Spencer and Azizur Rahman

Received: 11 March 2026

Revised: 29 March 2026

Accepted: 10 April 2026

Published: 13 April 2026

**Copyright:** © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

**Keywords:** coral reef monitoring; deep learning; semantic segmentation; underwater visual recognition; ecological informatics

## 1. Introduction

Coral reef ecosystems cover only 0.2% of the global ocean area, yet are home to more than 25% of all marine species [1]. As the core carrier of marine biodiversity, they play an irreplaceable role in sustaining fishery resources, providing coastal protection, regulating carbon cycles, and supporting ecotourism development [2]. However, under the context of intensifying global climate change and anthropogenic activities, marine heatwaves (MHWs) have shown a significant long-term increasing trend over the past four decades (1982–2023): the global occurrence frequency of MHWs has risen by 9.03% per decade, the mean intensity has increased by 0.12 °C per decade, and the average event duration has extended by 16.98 days per decade [3,4], which have significantly exacerbated coral bleaching and degradation (Figure 1). Global coral reef ecosystems are now facing the dual pressures of structural degradation and functional decline [5]. Statistics show that the 2023–2025 global coral bleaching event, the most severe on record, has impacted 84% of coral reefs worldwide [6]. Over 50% of coral reefs remain at moderate to high risk, and warm-water coral reefs have crossed the climate survival tipping point. There is an urgent need to establish a monitoring system with large-scale coverage, long-term stability, and automated processing capacity, so as to provide data support for coral reef conservation and management decision-making.

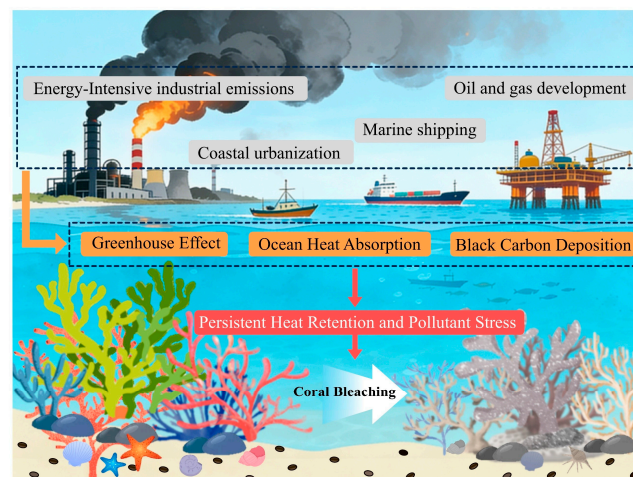


**Figure 1.** The phenomenon of coral bleaching and degradation under the influence of human activities. Anthropogenic marine production and transportation activities amplify the oceanic greenhouse effect, which elevates the ambient temperature around corals, leads to the loss of symbiotic zooxanthellae in coral colonies, and prolonged thermal stress in turn triggers coral bleaching and further reef degradation.

Traditional coral reef monitoring mainly relies on manual scuba diving transect surveys and manual image interpretation. While this method can ensure high interpretation accuracy, it has inherent limitations including high labor costs, limited spatial coverage, poor temporal continuity, and strong subjective bias from expert judgment, making it unable to meet the demands of large-scale, long-term dynamic monitoring of coral reefs [7]. In recent years, the rapid development of technologies including underwater high-resolution imaging devices, autonomous underwater vehicles (AUVs), and Structure from Motion (SfM) photogrammetry has enabled efficient and standardized acquisition of coral reef benthic image data, laying a solid data foundation for automated monitoring. Meanwhile, computer vision technologies based on deep learning have achieved major breakthroughs. From the advent of AlexNet [8] in 2012 to the launch of the improved lightweight model YOLOv2 [9] in 2016, deep learning has been systematically applied to the classification of coral and benthic images over the subsequent nearly two years, driving the transformation of coral reef monitoring from manual interpretation to automated and intelligent analysis.

The development of coral visual recognition technology has undergone a transformation from traditional handcrafted feature engineering to deep learning-driven approaches. Early studies mainly relied on manually designed features (color, texture, morphology, etc.) combined with traditional classifiers to achieve category discrimination, with obvious limitations in feature representation capacity and cross-scene generalization performance. With the widespread application of convolutional neural networks (CNNs), coral visual recognition tasks have gradually expanded from image-level classification to pixel-level semantic segmentation, multi-label health status recognition, and real-time object detection, with significant improvements in model accuracy and scene adaptability [10]. To date, several reviews [11,12] have provided periodic summaries of coral image segmentation and classification algorithms. However, these earlier reviews have notable content gaps across three core dimensions of the full technical chain. In terms of data gathering, they lack a systematic analysis of the evolution of underwater image acquisition technologies, dataset heterogeneity, and the standardization of annotation systems. For model development, existing studies largely overlook the holistic evolutionary framework of coral visual recognition technologies, cross-domain generalization mechanisms, and the mismatch between algorithm evaluation metrics and the practical requirements of ecological monitoring. With respect to engineering deployment, they barely cover the translation pathway from model outputs to ecological management decision-making. Furthermore, the in-depth discussion on the core bottlenecks restricting the large-scale deployment of the technology and corresponding solutions remains insufficient.

Based on the above background, this paper systematically reviews the research progress of visual recognition technology in the field of coral classification and reef monitoring, with core contributions as follows. (1) The evolutionary characteristics of underwater coral image acquisition methods, public dataset categories, and annotation strategies are comprehensively analyzed. (2) A holistic evolutionary framework of coral visual recognition technologies spanning traditional methods to deep learning approaches is established, and the merits and limitations of distinct technical pathways are clarified. (3) An in-depth dissection of the four core bottlenecks restricting the large-scale application of such technologies is performed, and the underlying logic and applicable boundaries of different learning paradigms for addressing domain shift issues are defined. (4) Integrated with cutting-edge technological trends and practical requirements of ecological monitoring, the core development directions for future research in this field are proposed.

## 2. Underwater Coral Image Datasets and Data Characteristics

Datasets form the fundamental cornerstone for the training and field deployment of coral visual recognition models. Data acquisition protocols, annotation frameworks, and data distribution characteristics directly define the maximum achievable performance and practical application boundaries of these models. This section systematically synthesizes the data infrastructure underpinning coral visual recognition research across four core dimensions: data acquisition technologies, the full dataset ecosystem, annotation specifications, and inherent challenges of coral image data.

### 2.1. Evolution of Data Acquisition Technologies

The development of coral image data acquisition technologies has directly driven the evolution of recognition tasks from static two-dimensional (2D) classification to three-dimensional (3D) structural quantification and multimodal comprehensive sensing. Based on key technological milestones, this progression can be broadly divided into five developmental phases (Figure 2).

The first phase is the manual diving photography stage, the mainstream method for early coral reef surveys. Divers typically capture images along parallel and perpendicular transects 10–50 m in length using standardized 1 m × 1 m quadrats with handheld cameras, acquiring images at 12–24 megapixels to produce high-resolution close-range records that support precise species identification and cover estimation [13–16]. However, constrained by dive duration, operational safety, and labor costs, this method fails to achieve large-scale and high-frequency monitoring. In addition, data acquisition standards are highly susceptible to operator variability, leading to poor data consistency across different survey batches.

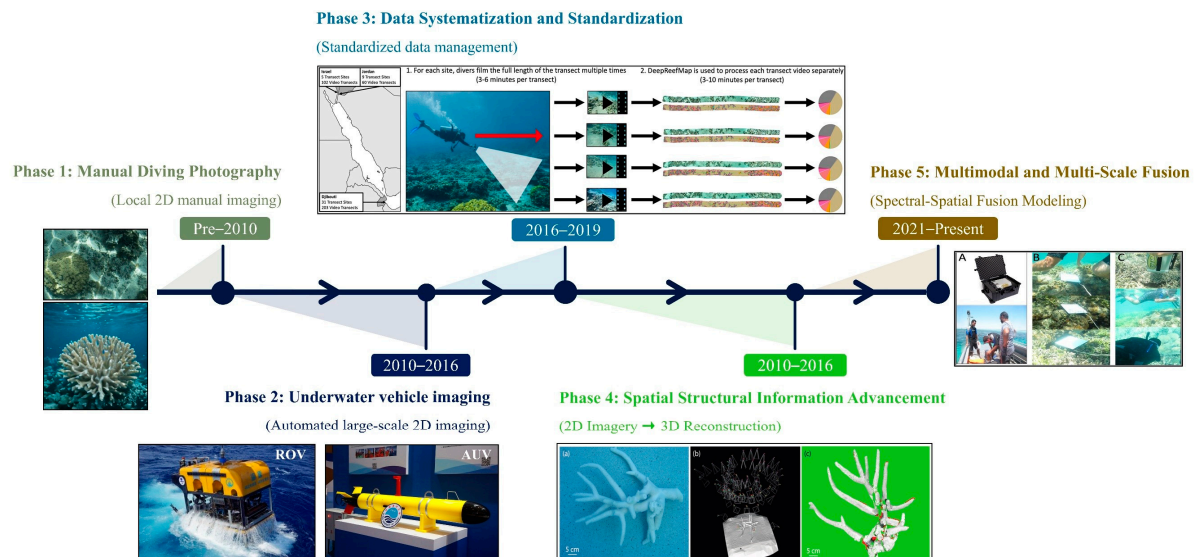


**Figure 2.** Framework for the evolution of underwater coral data acquisition technologies and information dimensions.

The second phase is the automated underwater vehicle acquisition stage. The application of autonomous underwater vehicles (AUVs) and remotely operated vehicles (ROVs) breaks through the depth and duration limitations of manual diving, enabling efficient and continuous imaging of the coral reef benthic environment. Typically, AUVs/ROVs work at depths of 50–500 m [17] with a single deployment endurance of 4–10 h [18–20], and are equipped with high-resolution digital cameras, LED lighting systems, and inertial navigation sensors to guarantee stable and consistent data acquisition. Compared with manual diving photography, this method cuts labor costs by over 60%, boosts data collection efficiency by 3–8 times, and achieves wider survey coverage and better data consistency. This has greatly improved the data acquisition efficiency and spatial coverage, fueling the rapid growth of coral image data volume [21]. In recent years, the development of miniaturized mini-ROVs has further lowered the technical threshold for data acquisition [22]. Such vehicles can also be combined with environmental DNA (eDNA) sampling technology to realize the collaborative acquisition of image data and biomolecular data [23], opening up a new pathway for multimodal monitoring. Meanwhile, the platform-based data system DeepReefMap (v1.1.0) [24] enhances the consistency and reusability of cross-regional data through standardized acquisition workflows and data organization protocols, making the “data system” a core component of the monitoring framework in the third phase.

The fourth phase is the 3D structural data acquisition stage. SfM photogrammetry can reconstruct 2D sequential images into high-precision 3D point cloud models [25], breaking through the spatial information limitations of 2D images. It provides spatial structural information for the quantification of key ecological indicators of coral reefs, including structural complexity, surface area, and roughness [26,27]. The combination of low-cost camera systems and point cloud semantic segmentation technology has further reduced

the technical threshold for 3D monitoring [28], facilitating the large-scale deployment of 2D image and 3D data fusion.

The fifth phase is the multimodal data fusion acquisition stage. Hyperspectral imaging technology captures changes in reflectance caused by differences in the pigments of coral symbiotic zooxanthellae and tissue structure through continuous narrow-band spectral information. It provides physical dimensional support for species-level differentiation and health status assessment [29], making up for the limitations of RGB imagery in fine-grained recognition. In situ spectral measurement studies have confirmed significant differences in spectral characteristics in the blue-green bands between different coral species, as well as between live and dead corals [30], laying the foundation for spectral-spatial fusion modeling.

## 2.2. Dataset Types and Scale Characteristics

Public coral datasets serve as the core foundation for the development of coral visual recognition technologies. Based on disparities in annotation granularity, spatial dimensionality, and task objectives, existing public data resources can be categorized into five classes, with their core characteristics and applicable scenarios summarized in Table 1. Point-level annotated datasets (represented by CoralNet, with over 15 publicly available global datasets) support coral cover estimation via random point sampling, boasting the advantages of low annotation cost and large sample size, yet suffer from limitations in spatial information retention. Patch-level data (represented by ReefNet, with approximately 12 publicly available datasets) are suitable for training early CNNs, but the cropping operation impairs the integrity of contextual information. Pixel-level semantic segmentation datasets (represented by Coralscapes, with around 8 publicly available standardized datasets) can provide fine-grained spatial structural information, yet are constrained by high annotation labor costs and limited dataset scale. 3D point cloud datasets (represented by the DeepReefMap 3D Dataset, with approximately 6 publicly available datasets), constructed based on SfM or low-cost 3D reconstruction techniques, enable the quantitative analysis of coral reef structural complexity, though no unified and standardized benchmark has been established for such data to date. Specifically, the missing benchmarking elements include unified coordinate calibration, consistent 3D annotation rules, standardized training–test splits, and shared evaluation metrics. In addition, multi-label health classification datasets (represented by the Coral-health-classification Dataset, with approximately 4 publicly available datasets) incorporate ecological stressors and coral health status into the annotation system, allowing visual recognition results to map more directly to ecological risk assessment. Overall, existing coral datasets exhibit significant heterogeneity in sea area geographic distribution, annotation protocols, and scale structure, resulting in prominent distribution shift issues across different datasets.

**Table 1.** Comparison of data structural characteristics and ecological application capability across different coral dataset types.

| Data Type                     | Spatial Dimension | Annotation Granularity | Typical Task                          | Ecological Indicator Support       | Key Limitations                                          |
|-------------------------------|-------------------|------------------------|---------------------------------------|------------------------------------|----------------------------------------------------------|
| Point-Level Annotation [31]   | 2D                | Random Points          | Coral Cover Estimation                | Cover Statistics                   | Lack of complete spatial structural information          |
| Patch-Level [32]              | 2D                | Image Patch            | Image Classification                  | Taxa Group Proportion Estimation   | Weak contextual information integrity                    |
| Pixel-Level Segmentation [33] | 2D                | Dense Pixel Annotation | Fine-Grained Semantic Segmentation    | Cover and Boundary Analysis        | High annotation labor cost                               |
| 3D Point Cloud [34]           | 3D                | Point/Face Annotation  | Structural Modeling                   | Structural Complexity and Rugosity | Limited dataset scale                                    |
| Multimodal Data [35]          | 2D + Spectral     | Multi-Label Annotation | Species and Health Status Recognition | Health and Species Identification  | High acquisition cost and lack of standardized benchmark |

### 2.3. Annotation Strategies and Taxonomic Classification Systems

Annotation format directly governs the spatial representation capacity and ecological interpretability of recognition models, and its evolution has consistently revolved around two core objectives: reducing the annotation costs and aligning with the practical requirements of ecological monitoring. In terms of annotation strategies, point-level annotation incurs low labor costs but cannot support fine-grained structural analysis [36]. In contrast, pixel-level annotation provides complete spatial contour information, yet relies heavily on expert participation and demands extensive time and labor input. To balance annotation costs and model performance, researchers have proposed a strategy for training dense segmentation models based on sparse labels [37]. This framework combines sparse point annotation, image-level weak supervision, and pseudo-label generation, maintaining state-of-the-art segmentation accuracy while drastically reducing annotation workload.

Regarding the taxonomic classification system, the classification hierarchy of coral recognition has displayed distinct evolutionary characteristics. Early studies mainly focused on binary or multi-class classification of functional groups, including corals, algae, and benthic substrates [38]. With the expansion of dataset scale and the growing demands for refined ecological monitoring, research has gradually extended to fine-grained species identification at the genus and species levels [39]. In recent years, the classification system has further expanded to multi-label health status recognition, including the differentiation of live and dead corals, as well as the identification of bleaching, disease, and bioerosion. This advancement enables the direct mapping of visual recognition results to the standardized coral reef health risk assessment framework. Meanwhile, to address the limitations of the traditional intersection over union (IoU) metric in fine-grained boundary evaluation for coral imagery, Boundary IoU (BIOU)—a boundary-sensitive metric—has been increasingly adopted [40]. Unlike IoU, which only measures pixel-set overlap and ignores subtle contour deviations critical to coral morphology, BIOU quantifies the alignment of predicted and ground-truth coral boundary regions, thus making model performance evaluation more consistent with the analytical demands of coral morphological structure analysis. For instance, in semantic segmentation of branching *Acropora* corals from underwater photogrammetry imagery, BIOU accurately captures under-segmentation of delicate branch tips that IoU fails to detect [41].

### 2.4. Data Distribution Characteristics and Inherent Challenges

Two inherent characteristics of underwater coral image data constitute the core data-level bottlenecks restricting the performance of recognition models, and represent the primary challenges to be addressed by preprocessing techniques. The first core challenge is severe class imbalance. In natural coral reef ecosystems, background taxa including benthic substrates and macroalgae account for an overwhelming proportion of samples, while target coral species, especially rare coral taxa, occupy an extremely low proportion. This pattern is well-exemplified by the representative Coralscapes dataset (Figure 3), which shares common traits with other coral datasets. Collected from 35 dive sites across five Red Sea countries, this dataset comprises 2075 images annotated with 39 benthic categories. A rough calculation based on segmentation mask counts reveals that non-coral background masks account for approximately 78% of the total, coral-related segmentation masks make up around 21%, and rare coral masks account for less than 1%. This natural long-tailed distribution causes model training to be dominated by majority classes, which severely impairs the recognition accuracy for rare groups and further undermines the reliability of ecological statistical results. To mitigate this issue, loss functions widely adopted in the computer vision domain, namely focal loss [42] and boundary loss [43], have been extensively introduced. Focal loss enhances the model's ability to identify hard-to-classify rare

samples by down-weighting the contribution of easily classified samples, while boundary loss addresses pixel-level segmentation in highly imbalanced scenarios via distance metric optimization. Both methods have yielded significant performance improvements.

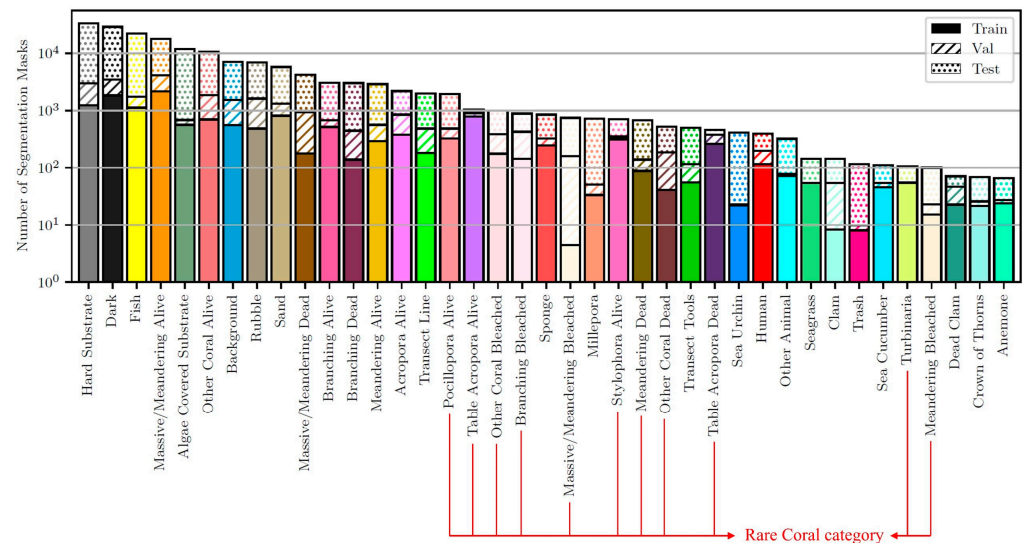


**Figure 3.** Number of annotated segmentation masks per class in the Coralscapes dataset splits for each of the 39 classes (adapted from [44]).

The second core challenge is the quality degradation of underwater imaging. The selective absorption and scattering effects of water on light lead to severe color shift, reduced contrast, blurred details, and noise interference in underwater images. The degree of degradation varies drastically with water turbidity, imaging depth, and illumination conditions, which greatly increases the difficulty of robust feature extraction [45]. To tackle this problem, researchers have developed a series of underwater image enhancement and restoration methods. Physics-based methods built on the dark channel prior achieve image dehazing and color correction by modeling the underwater light transmission process [46]. Deep learning-based real-time enhancement algorithms realize image quality optimization in complex underwater environments through end-to-end learning while satisfying the real-time requirements of in situ monitoring via AUVs/ROVs [47]. Furthermore, improved Retinex optimization [48] and hybrid generative adversarial network (GAN) [49] methods have further boosted image enhancement performance in extreme low-illumination and high-turbidity scenarios. When applied to downstream tasks such as semantic segmentation and object detection, physics-based and learnable enhancement methods exhibit distinct trade-offs. Physics-based methods offer strong interpretability and require no task-specific training data, yet their performance may degrade under non-ideal water conditions due to the constraints of model assumptions. In contrast, learnable enhancement methods can adaptively suppress task-irrelevant degradation while preserving features critical for segmentation or detection. Thus, the choice between the two approaches depends on the trade-off among interpretability, data availability, and task-specific performance requirements.

### 3. Coral Classification Methods Based on Traditional Visual Features

Prior to the widespread application of deep learning techniques, automated coral image classification mainly relied on handcrafted visual features combined with classic machine learning models. This approach represents a typical feature engineering-driven paradigm, which laid the early technical foundation for the automated analysis of coral imagery.

### 3.1. Handcrafted Feature Construction Framework

The core of traditional methods is to extract discriminative low-level visual features from coral images via manually designed feature extraction operators, which are mainly divided into three categories: color features, texture features, and morphological structural features.

Color features are the earliest visual features applied to coral classification. Researchers typically extract statistical features such as color histograms and color moments in color spaces including RGB, HSV, and Lab, and achieve classification by leveraging the color differences among different coral taxa [50–52]. However, the selective absorption of light in different wavebands by water causes severe color shift, making color features highly sensitive to imaging depth and illumination conditions, and thus extremely unstable in cross-scene applications.

Texture features, which characterize the microstructural differences of coral surfaces, are the most widely used feature type in traditional coral classification. Methods including the Gray Level Co-occurrence Matrix (GLCM), Local Binary Pattern (LBP), and Gabor filters can effectively describe the surface texture differences among branching, massive, and plate corals, and exhibit stronger scale and illumination robustness compared with color features. Nevertheless, texture features are highly sensitive to image blurring and noise interference, with their discriminative power reduced significantly in complex underwater backgrounds. Quantitatively, when Gaussian noise with a variance of 0.02–0.08 is added to underwater coral images, the classification accuracy of classic LBP descriptors drops by 28–46% compared with noise-free conditions [53]; under  $3 \times 3$  to  $7 \times 7$  Gaussian blur (simulating underwater light scattering and imaging defocus), the classification accuracy of GLCM and LBP for coral taxa decreases by 24–43% [54]. Even the relatively more robust Gabor filters still exhibit an 15–28% accuracy drop under the above blurring conditions [55].

Morphological features and local descriptors are applied to enhance the characterization of macro-morphological differences among corals. Researchers have quantified the morphological and structural variations of corals using features including edge density, contour descriptors, and fractal dimension [56]. Meanwhile, classic local invariant feature descriptors such as SIFT, SURF, and HOG have been introduced to improve the scale and rotation invariance of extracted features. Nevertheless, the inherent degradation of underwater imaging, including light scattering, color cast, low contrast, and backscattering noise [57], leads to significant limitations for these methods in underwater scenarios. For instance, SIFT and SURF suffer from a drastic reduction in stable feature points in turbid and low-contrast underwater images and exhibit poor robustness to non-uniform illumination and caustic effects [58]. HOG features, which rely on clear edge gradient information, show severely degraded discriminative power for coral targets due to scattering-induced edge blurring and gradient distortion [59]. Fundamentally, such methods still depend on local gradient information and are highly vulnerable to noise interference in complex underwater backgrounds.

Overall, although the traditional handcrafted feature system has strong interpretability, its representation capacity is mainly limited to low-level visual information, making it difficult to effectively capture the complex multi-scale structures and spatial contextual relationships of corals.

### 3.2. Classic Classification Models

After handcrafted feature extraction, traditional coral classification studies mainly employ classic machine learning models, including support vector machine (SVM), random forest, and k-nearest neighbor (k-NN), to achieve category discrimination. Among these models, SVM maps high-dimensional features into a linearly separable space via

kernel functions, delivering superior classification performance in small-sample and high-dimensional feature scenarios [60]. When applied to binary or multi-class classification of functional groups including corals, it typically achieves an overall classification accuracy of 80–92%, and thus has become the most widely used classifier in traditional methods. As an ensemble learning model, random forest mitigates overfitting risk through the voting mechanism of multiple decision trees, and exhibits more robust stability when addressing feature noise interference and multi-class classification problems [61]. On early coral image datasets, it generally reaches an overall classification accuracy of 70–85%; despite its slightly inferior accuracy compared with SVM, it presents a smaller magnitude of performance degradation when processing noisy underwater image data. Under the condition of limited data scale in the early research stage, methods combining handcrafted features and classic classifiers achieved promising classification performance in controlled scenarios and single-region datasets, accomplishing the preliminary transition of coral image analysis from manual interpretation to automated processing.

### 3.3. Method Limitations and Ecological Applicability Analysis

Although methods based on traditional visual features can achieve relatively stable recognition performance in controlled environments or with survey data from a single region, their inherent limitations have gradually become evident. First, handcrafted features cannot effectively represent the complex three-dimensional structures and spatial contextual relationships of corals, resulting in insufficient capacity for fine-grained taxonomic differentiation and health status assessment. Second, such methods usually output discrete category labels, which can hardly directly support the extraction of key ecological indicators, including coral cover estimation, spatial structure analysis, and long-term dynamic change evaluation. In addition, in special scenarios such as cold-water coral reefs or mesophotic coral ecosystems, two-dimensional texture and color features are insufficient to characterize structural complexity comprehensively [62], nor can they realize comprehensive discrimination of multi-dimensional information such as health status and stressors. Therefore, with the expansion of data scale and the improvement of model representation capacity, the research paradigm has gradually shifted from manual feature construction to data-driven end-to-end learning methods. This transition constitutes a key technical node for the development of subsequent methods.

## 4. Deep Learning-Based Coral Visual Classification and Segmentation Methods

### 4.1. Image-Level Classification: Establishment of the Deep Learning Paradigm

Deep learning automatically learns high-level semantic features of data through multi-layer nonlinear mapping, which completely breaks through the representation capacity bottleneck of traditional handcrafted features, and has become the mainstream technical approach for current coral visual recognition. As the earliest application form of deep learning in coral recognition research, image-level classification marks the formal entry of coral recognition into the data-driven deep learning era. Early studies mainly adopted a patch-based classification strategy, which splits large-scale underwater images into fixed-size image patches and feeds them into the CNN model for training and prediction. This strategy greatly reduces the computational burden of the model while solving the classification challenge caused by the mixing of multiple categories in full-field images. Mahmood et al. conducted the first systematic comparison of the performance of handcrafted features and deep convolutional features in coral classification tasks [10], and confirmed that CNN models exhibit stronger feature representation capacity and generalization stability in

cross-dataset testing, thus consolidating the core position of deep learning in the field of coral recognition.

Subsequently, classic CNN architectures (as shown in Figure 4) including AlexNet, VGG, and ResNet were widely introduced into coral classification tasks [63,64]. Among them, the residual connection mechanism effectively alleviates the gradient degradation problem of deep networks, and greatly improves the training stability and classification accuracy of deep models. To address the limitation of limited labeled coral samples, transfer learning has become the core technical approach. By pre-training the model on large-scale natural image datasets such as ImageNet and then performing parameter fine-tuning on coral datasets, the convergence speed of the model and classification performance in small-sample scenarios can be significantly improved. For example, Gómez-Ríos et al. [65] systematically verified this transfer learning paradigm on classic small-sample coral texture datasets (EILAT, RSMAS and EILAT2). Compared with traditional handcrafted feature methods, the ResNet series models pre-trained on ImageNet achieved an absolute accuracy improvement of 5.87 percentage points on the EILAT2 dataset (from 93.1% to 98.97%). This paradigm only needs to fine-tune the top fully connected layers of the pre-trained model to complete domain adaptation from natural images to coral images, which significantly accelerates the model convergence speed and effectively alleviates the underfitting problem caused by the scarcity of high-quality annotated coral samples. Therefore, it has been widely adopted in subsequent coral classification research.

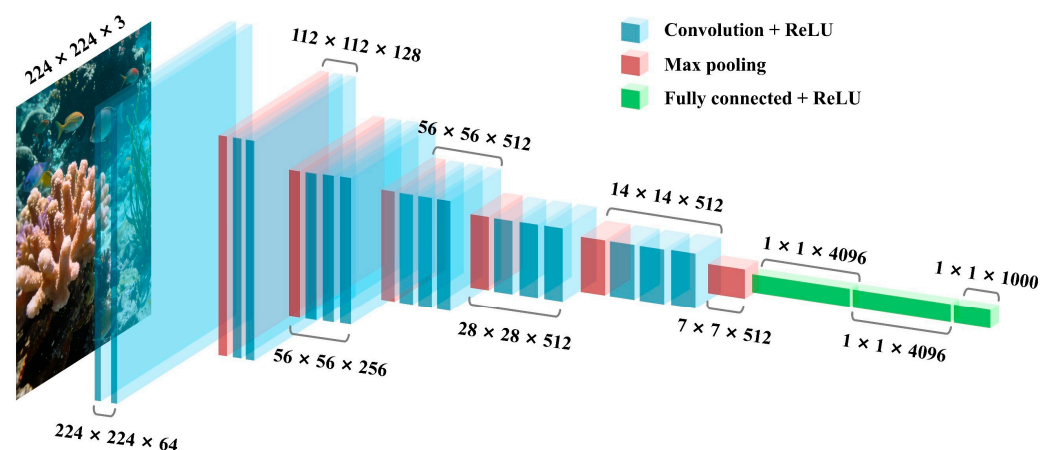


**Figure 4.** Schematic diagram of the coral classification CNN architecture (taking VGG as an example).

In recent years, with the increasing demand for in situ real-time monitoring via AUVs/ROVs, lightweight networks and edge deployment have become a key research direction for image-level classification. The application of lightweight models such as InceptionV3 and MobileNet [66–68] greatly reduces the parameter volume and computational cost of the model while maintaining high classification accuracy, enabling deployment on underwater embedded devices. However, image-level classification can only output the category label of a single image patch, and cannot provide fine-grained spatial distribution information, making it difficult to meet the core ecological monitoring requirements such as live coral cover quantification and spatial pattern analysis. For this reason, the research focus has gradually expanded to the field of pixel-level semantic segmentation.

#### 4.2. Semantic Segmentation: From Category Recognition to Ecological Quantification

Semantic segmentation, which performs pixel-wise category prediction on images, has enabled the transition from image-level classification to pixel-level fine-grained recognition. Its outputs can be translated directly into core ecological indicators such as live coral cover, spatial distribution of benthic taxa, and patch morphological features, establishing

it as the central research focus in contemporary coral visual recognition. According to the performance evaluation of coral semantic segmentation architectures in the existing literature, mainstream models exhibit significant performance variations across different datasets. Among the classical convolutional architectures, DeepLabV3+ achieved 72.97% mean intersection over union (mIoU) on the HKCoral dataset [69]. Transformer-based architectures demonstrate superior segmentation capability: Swin-UNet achieved 85.7% mIoU and 78.9% Boundary IoU on the Coralscapes dataset [33]. In the domain of weakly supervised learning, Sparse-Label U-Net achieved 72.4% mIoU on the CoralSeg dataset using only 50 sparse point annotations per image [70]. In addition, the specially designed RGB-D fusion network CNet achieved 81.83% mIoU on the Moorea dataset [71]. These results indicate that Transformer architectures hold distinct advantages in coral segmentation tasks, while weakly supervised approaches offer a feasible pathway for reducing annotation costs.

Fully convolutional networks (FCNs) first extended convolutional architectures to dense prediction tasks [72], and this was followed by the widespread adoption of U-Net and the DeepLab series in coral segmentation studies [73,74]. U-Net preserves fine-grained spatial details through an encoder–decoder structure with skip connections, while the DeepLab series enhances multi-scale modeling by expanding the receptive field via atrous convolution, delivering superior performance in segmenting large-scale remote sensing coral reef imagery.

To address the recognition challenges posed by complex underwater environments, researchers have optimized segmentation models along multiple dimensions: attention mechanisms, including channel and spatial attention modules, are incorporated to strengthen target features, suppress background and noise interference, and improve feature extraction in turbid water conditions [75,76]; physics-guided image enhancement modules are integrated to jointly optimize the underwater light transmission model and the segmentation network in an end-to-end manner, achieving synergistic adaptation between image enhancement and feature extraction that boosts segmentation accuracy significantly in high-turbidity and low-illumination scenarios [77]; and Transformer architectures, such as the Swin-Transformer [78], are employed to capture long-range spatial dependencies of corals through global modeling, thereby enhancing the representation of multi-scale coral structures.

#### *4.3. Object Detection and Real-Time Recognition: Engineering Applications for In-Situ Monitoring*

Object detection technology enables the simultaneous localization and classification of coral targets while balancing recognition accuracy and inference efficiency, making it the core enabling technology for real-time in situ monitoring via ROVs/AUVs. Early research on coral object detection was mainly based on two-stage detection models represented by Faster R-CNN, which first generate region proposals and then perform category discrimination. While these models deliver high detection accuracy, their slow inference speed cannot meet the requirements of real-time monitoring. Subsequently, one-stage detection models such as YOLO and SSD have been widely introduced. These models achieve simultaneous target localization and classification through end-to-end regression, greatly improving inference speed while maintaining high detection accuracy. Among them, models including YOLOv8 and the improved YOLOv11-UOS integrate attention mechanisms and multi-scale feature fusion, which further optimize detection performance for small coral targets in complex underwater backgrounds. Meanwhile, they enable the unified output of detection and segmentation, becoming the mainstream architecture for current real-time coral detection [79].

Regarding practical deployment, quantitative comparisons on coral and coral reef fish detection tasks reveal distinct trade-offs. On coral health assessment tasks, an enhanced YOLOv8 model integrating CBAM and BiFPN achieved a mean average precision (mAP) of 88.4% at IoU = 0.5, with a processing time of 11.2 ms and a compact model size of 7.2 MB [80]. For lightweight coral detection, YOLOv8-small achieved a mAP of 53.7%, while its Ghost module-enhanced variant improved to 55.9% with a compact model size [81]. Regarding inference speed on edge devices, benchmarking studies show that YOLOv8-Nano achieved 24.1 FPS on the NVIDIA Jetson Nano with 7.2 GFLOPs [82], while YOLOv7 optimized with TensorRT (v8.6.1) reached 85 ms per image (approximately 11.8 FPS) on the same platform [83]. For coral reef fish detection, SSD-MobileNet demonstrated superior computational speed with a mAP of approximately 92.21%, achieving real-time detection at 30 FPS when accelerated with a Coral USB Accelerator on a Raspberry Pi 4 [84]. For small coral target detection, YOLOv5 enhanced with attention mechanisms achieved a 3.2% increase in average detection accuracy over baseline methods [85].

In recent years, Transformer architectures have demonstrated tremendous potential in the object detection field. The RT-DETR-based Transformer detection model achieves a well-balanced trade-off between high real-time performance and detection accuracy via the cross-attention mechanism of the decoder. It outperforms traditional CNN-based detection models in hard coral detection tasks using coral quadrat imagery [86], exhibiting significant advantages in global modeling and small target recognition. To address the deployment requirements on underwater embedded devices, researchers have adopted lightweighting techniques including model quantization, pruning, and knowledge distillation. These methods further reduce model size and computational overhead under the premise of controllable accuracy loss, enabling edge-side real-time detection and recognition on ROV/AUV platforms.

#### *4.4. Data Augmentation, Robustness and Generalization Capacity: Toward Reliable Cross-Scene Model Deployment*

Light absorption, scattering, and particulate interference in underwater optical environments degrade image quality significantly, thus image enhancement techniques are commonly integrated into the model training pipeline. In the field of underwater image enhancement, numerous studies have identified geometric transformations as the most effective category, as such techniques effectively simulate the natural variations in coral orientation and scale encountered in underwater surveys. Furthermore, RGB histogram equalization, color correction, dehazing enhancement, and generative augmentation methods based on GAN frameworks each offer distinct advantages in matching specific imaging conditions, thereby delivering optimal performance for recognition tasks [87–89].

Building upon these augmentation strategies, recent advances have further integrated physical models with deep neural networks to achieve more robust performance under complex water conditions. In recent years, a series of methods have jointly optimized the physical radiative transfer model of underwater light propagation with deep neural networks, achieving more robust performance under complex water conditions. Meanwhile, weakly supervised and semi-supervised learning strategies have been adopted to alleviate the challenge of high annotation costs. Through consistency regularization or pseudo-labeling mechanisms, these methods can leverage unlabeled data to improve the generalization capacity of models, yet their stability in highly complex underwater scenarios still requires systematic validation. When there are geographic variations across sea areas or discrepancies in imaging equipment between training and test datasets, model performance degrades markedly. Throughout the phased evolution of coral visual recognition technologies (Table 2), developing model architectures and training strategies that are ro-

bust to distribution shift has become the core bottleneck restricting the further engineering deployment of deep learning methods.

**Table 2.** The phased evolution of deep learning in coral visual recognition.

| Phase                                                | Primary Task Form                                             | Technical Characteristics                                                 | Advantages                                                                                     | Limitations                                                                                               |
|------------------------------------------------------|---------------------------------------------------------------|---------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|
| Image-Level Classification Phase                     | Image or Patch-Level Category Discrimination [7]              | Automatic CNN Feature Learning; Transfer Learning                         | Replaces handcrafted features; Well-established training pipeline                              | Lack of complete spatial structural information; Inability to support quantitative coral cover estimation |
| Pixel-Level Semantic Segmentation Phase              | Pixel-Level Category Prediction [90]                          | Encoder–Decoder Architecture; Multi-scale Feature Fusion                  | Enables direct quantification of core ecological indicators (coral cover, taxa proportion)     | High annotation labor cost; High sensitivity to class imbalance                                           |
| Structural Optimization and Enhancement Phase        | High-Precision Segmentation and Complex Scene Adaptation [91] | Attention Mechanism; Physics-Guided Enhancement; Transformer Architecture | Improved robustness and multi-scale representation capacity in complex underwater environments | High model complexity; Increased computational resource requirements                                      |
| Real-Time Detection and Engineering Deployment Phase | Real-Time Detection and Edge Deployment [86]                  | Transformer Detection Architecture; Lightweight Network                   | Enables real-time in situ monitoring via AUVs/ROVs                                             | Inherent trade-off between detection accuracy and inference speed                                         |
| Generalization and Robustness Improvement Phase      | Cross-Sea Area and Cross-Device Application [5]               | Domain Shift Analysis; Data Augmentation; Semi-Supervised Learning        | Facilitates cross-regional model deployment and generalization                                 | Theoretical and engineering solutions for field application are not yet fully established                 |

## 5. Core Challenges for the Large-Scale Application of Coral Visual Recognition

### 5.1. Domain Shift and Bottlenecks in Cross-Scene Generalization Performance

Domain offset is the core bottleneck that restricts the cross-regional and cross-device deployment of coral recognition models. The spectral attenuation laws of water bodies, suspended particle concentrations, seabed compositions, and coral species compositions in different sea areas exhibit systematic differences. Even within the same sea area, variations in imaging depth, lighting conditions, and acquisition equipment parameters can lead to significant changes in image data distribution. For instance, quantitative analysis shows that when photographing a DGK color calibration card (X-Rite, Grand Rapids, MI, USA) underwater, the red channel intensity exhibits an approximately linear decrease from 1.0 m to 2.0 m relative to the air baseline [92]. Such distance-dependent attenuation directly alters the RGB feature distributions that models rely on. These factors collectively violate the core assumption of supervised learning that “training and testing data are independently and identically distributed”, ultimately resulting in a substantial decline in the model’s performance in unknown scenarios.

It is important to clarify that structural optimization strategies, such as attention mechanisms and multi-scale feature fusion, can only improve the feature discriminative ability of the model within the training data distribution, and cannot fundamentally solve the out-of-distribution generalization problem. To address domain shift, existing studies have formed a progressive solution framework ranging from transfer learning and domain adaptation to domain generalization, yet their application boundaries and inherent limitations remain prominent.

Transfer learning is currently the most widely used fundamental method in coral recognition [93]. By pre-training the model on large-scale natural images and then fine-tuning it on coral imagery, it can improve the convergence efficiency and initial performance of the model when labeled samples are limited, and its effectiveness has been verified in lightweight networks and Transformer models. However, this method relies on the

distribution correlation between the source domain and the target domain. The inherent differences in spectral and texture features between natural images and underwater coral imagery tend to limit the effectiveness of feature transfer. Therefore, it is only suitable for scenarios with minor environmental differences, and lacks stability when deployed across unknown sea areas [94].

Domain adaptation addresses the shift problem by explicitly reducing feature discrepancies between domains. It introduces target domain data during the training phase, and achieves feature distribution matching between the source and target domains through mechanisms such as adversarial training and statistical alignment. Unsupervised domain adaptation (UDA) is the mainstream research direction at present [95]. The source-free domain adaptation framework can complete adaptation only by retaining the parameters of the source model, which breaks through the constraints of source data storage and privacy protection, and expands its engineering applicability [96]. Nevertheless, it still requires the acquisition of target domain data before deployment, and cannot solve the generalization problem in completely unknown environments. Regarding practical requirements, effective UDA can be achieved with as few as several hundred unlabeled images, provided that they represent the target domain's variability. Source-free adaptation requires only the trained source model, eliminating the need for original source data. For coral-specific applications, UDA with synthetic data enabled sea urchin detection at 84.3% AP50 without manually annotated real images [97]. Similarly, a vision foundation model (DINOv2) adapted via Low-Rank Adaptation (LoRA) achieved a 64.77% match ratio for multi-label coral condition classification across seasons and sites, using only 5.91 M trainable parameters [98]. These metrics confirm that UDA and source-free adaptation can deliver high performance when unlabeled target data capture relevant environmental variability.

Domain generalization completely eliminates the dependence on target domain data. It enables the model to maintain stable performance in unseen environments by learning cross-domain invariant feature representations, which highly aligns with the “train-first, deploy-later” requirement of ecological monitoring. Current research has evolved from simple data mixing to multi-dimensional structured augmentation methods. To address open-category and class imbalance problems, meta-learning [99] has been introduced to coordinate cross-domain gradient updates, which improves the model's ability to recognize unseen categories and abnormal states, and provides a new path for the identification of rare coral taxa.

### 5.2. Data Constraints and Bottlenecks in Sample Utilization Efficiency

Data-level constraints are the fundamental bottleneck restricting the performance improvement and large-scale application of models, which are mainly reflected in three aspects.

First, there is an extreme scarcity of high-quality annotated data. Pixel-level annotation and species-level classification of coral imagery require experts with solid coral taxonomic expertise, resulting in a high threshold and extremely time-consuming annotation process, and thus a severe shortage of standardized, high-quality annotated datasets [100]. Existing public datasets suffer from uneven geographic distribution across sea areas, inconsistent annotation protocols, and limited species coverage, which can hardly support the training of generalizable coral recognition models.

Second, the inherent class imbalance problem cannot be fundamentally resolved. The natural species distribution of coral reef ecosystems presents a significant long-tailed characteristic, with extremely small sample sizes for rare coral taxa. Even if alleviated by methods such as loss function optimization and sample resampling, it cannot compensate for the insufficient feature learning caused by limited sample size, resulting in consistently low recognition accuracy of the model for rare groups [101], which are often the key targets of ecological conservation. To address this, targeted data collection strategies

may require additional human and material resources to actively acquire imagery from underrepresented habitats or to select specific seasonal windows during [102] which rare taxa are more observable, thereby enhancing sample diversity for minority classes.

Third, the level of standardization across datasets is low. Coral image data collected by different research teams have significant differences in imaging equipment, shooting parameters, quadrat design, annotation specifications, and taxonomic classification systems [103]. The poor data compatibility makes it difficult to integrate multi-source data and conduct cross-dataset validation of models, which severely restricts the development and iteration of generalizable models.

### *5.3. Bottlenecks in the Alignment Between Algorithm Evaluation and Ecological Monitoring Requirements*

There is a key misconception in current coral visual recognition research: model training and evaluation are centered on computer vision metrics, which are severely disconnected from the practical requirements of coral reef ecological monitoring, making it difficult to directly convert model outputs into a basis for management decision-making.

In terms of model evaluation, existing studies generally adopt classic computer vision metrics, which can only reflect the pixel-level classification accuracy of the model, but cannot measure the model's ability to estimate core ecological indicators. The core reason is that coral reef ecological monitoring focuses on the overall proportion of taxa, spatial patterns, and dynamic changes, rather than the classification correctness of a single pixel. There is a lack of direct mapping between traditional visual metrics and ecological statistical indicators.

Regarding the ecological interpretability of the model, the black-box nature of deep learning models [104] leads to poor interpretability of their classification decisions. It is impossible to explain the basis of classification results to ecologists and managers, nor to locate the source of model errors, resulting in insufficient credibility of model results in ecological surveys and management decision-making. Meanwhile, most existing models can only output the category and spatial distribution of corals, and cannot conduct a comprehensive assessment of the health status, degradation risk, and change trend of coral reefs combined with ecological knowledge, making it difficult to achieve the leap from image recognition to decision support.

### *5.4. Bottlenecks in Engineering Deployment and Long-Term Monitoring Implementation*

The practical field implementation of coral visual recognition technology requires full adaptation to engineering application scenarios such as ROV/AUV in situ monitoring and long-term continuous observation. However, most existing relevant studies rely on high-performance servers for offline model training and testing, which is significantly disconnected from the actual requirements of engineering deployment. The embedded devices of ROV/AUVs carried by underwater platforms have strict constraints on computing power, memory, and power consumption. High-precision, large-parameter recognition models are difficult to deploy directly on the edge side, while models modified for lightweighting are highly prone to loss of recognition accuracy [105]. How to achieve the optimal trade-off between recognition accuracy, inference speed, and hardware resource consumption has become the core challenge for the engineering deployment of models. Meanwhile, most existing recognition models adopt a static offline training paradigm, where model parameters are fixed after deployment. During the long-term monitoring of coral reefs, factors such as the expansion of monitoring areas, seasonal environmental changes, and natural succession of coral communities will cause continuous changes in data distribution, and the recognition performance of static models will continue to degrade accordingly. Furthermore, the continuous introduction of new monitoring data can easily lead to the

catastrophic forgetting problem of the model, making it difficult to effectively retain the knowledge accumulated from historical learning.

To address catastrophic forgetting in long-term underwater monitoring, several continual learning strategies have been explored in coral detection contexts. Knowledge distillation has been successfully applied to coral reef monitoring, enabling the transfer of fine-scale classification knowledge from underwater imagery to drone-based aerial imagery through a teacher–student framework, thereby allowing scalable and cost-effective large-scale reef assessment [106]. Pseudo-labeling has also been employed in instance segmentation frameworks to iteratively refine coral detection models using unlabeled data, as demonstrated by Picek et al., who integrated pseudo-labeling with Mask R-CNN (v1.0) to achieve state-of-the-art performance in the ImageCLEFcoral competition [107]. Although direct applications of rehearsal and regularization strategies to coral visual recognition remain limited, these distillation-based approaches demonstrate promising pathways for enabling dynamic model updates in long-term coral reef monitoring deployments.

At present, the application of coral visual recognition technology is still in the stage of fragmented research and exploration, and no unified technical specifications have been established in the field. There is a lack of unified requirements from data acquisition standards, data preprocessing pipelines, to model evaluation specifications and result output formats. This situation not only makes it difficult to reuse technical achievements among different research teams, but also makes it impossible to carry out effective cross-comparison of coral reef monitoring results from different regions. It not only severely restricts the large-scale promotion of the technology itself, but also fundamentally hinders the construction and implementation of a cross-regional collaborative monitoring system, becoming an important obstacle for coral visual recognition technology to move from laboratory research to practical ecological monitoring applications. In fact, realizing the transition of coral visual recognition from experimental validation to large-scale ecological application does not only rely on the improvement of a single model architecture, but requires the formation of a collaborative framework in terms of distribution modeling, domain generalization strategies, data governance systems, and continual learning mechanisms (as shown in Figure 5). This systematic perspective will also become the key direction for the technological development in this field in the future.

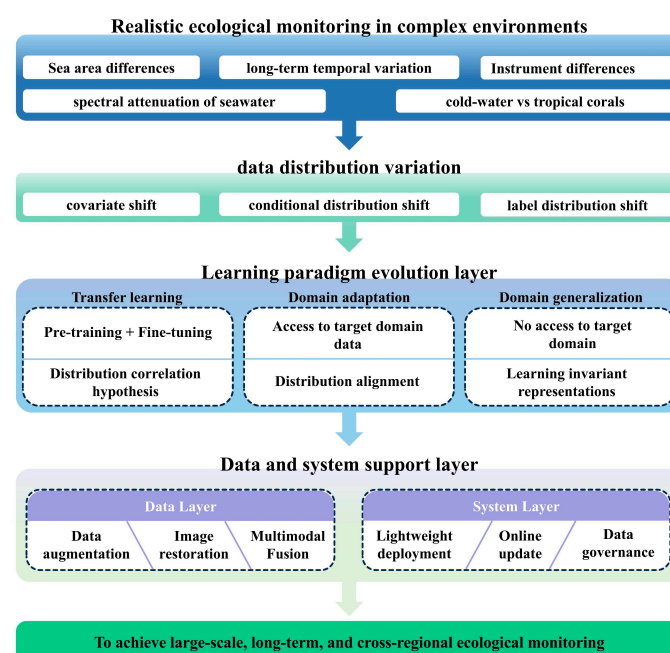


**Figure 5.** Theoretical framework of cross-domain generalization and engineering deployment.

## 6. Future Directions and Technological Evolution Trends

Building on the incremental advances in deep learning methodologies, coral visual recognition research is undergoing a paradigm shift from the optimization of individual model performance to the integration of multimodal fusion, advances in representation learning, and the construction of system-level intelligent monitoring architectures. Future developments will not only focus on improving recognition accuracy, but also address the systematic reconstruction of cross-modal information integration, data efficiency optimization, and the coupling mechanism between technical outputs and ecological decision-making.

### 6.1. Underwater Scene-Oriented General Visual Representation and Foundation Model Construction

Existing transfer learning pipelines rely heavily on pre-trained models developed from natural images. However, there are fundamental differences in imaging mechanisms and feature distributions between natural images and underwater imagery, which constitutes a key bottleneck limiting the cross-domain generalization capability of existing models. Accordingly, one of the core future directions is to establish a large-scale self-supervised pre-training system and underwater visual foundation models tailored for underwater scenarios.

Leveraging the massive volume of unannotated underwater coral imagery acquired by global ROV/AUV fleets, generalizable and robust visual feature representations in underwater environments can be learned via self-supervised pre-training tasks, including contrastive learning and masked image modeling. Compared to models pre-trained on natural images, models pre-trained on underwater data via self-supervision can better adapt to the inherent characteristics of underwater imaging and learn domain-invariant features that are insensitive to illumination degradation, color shift, and variations in water turbidity, thereby fundamentally enhancing the cross-scene generalization capability of the models. Meanwhile, integrating multimodal large language models (MLLMs) to construct vision-language multimodal foundation models for coral reef monitoring will enable the natural language interpretation of coral recognition results, ecological knowledge integration, and intelligent question answering, addressing the critical limitations of poor ecological interpretability and low usability of existing models.

### 6.2. Multimodal Information Fusion and Fine-Grained Modeling of 3D Ecological Structure

Reliance on sole 2D RGB imagery makes it difficult to comprehensively characterize the 3D structural features and health status of coral reef ecosystems, and multimodal information fusion represents a critical developmental direction for intelligent coral reef monitoring. Future research should prioritize the development of an integrated multimodal fusion and recognition framework. First, the fusion of hyperspectral data and RGB imagery can leverage the fine physical spectral features of hyperspectral data to improve the accuracy of fine-grained coral species classification and health status recognition, addressing the key limitations of 2D RGB imagery in live/dead coral discrimination and early-stage bleaching detection. Second, integrating 2D semantic segmentation with 3D point cloud modeling enables the accurate mapping of pixel-level recognition results to 3D point cloud models, supporting the precise quantification of key ecological metrics including coral coverage, surface area, volume, structural complexity, and reef rugosity, while mitigating the quantification errors caused by the projection of 2D imagery. Furthermore, the integration of multi-source data such as sonar, water quality parameters, and eDNA will facilitate the construction of a cross-modal integrated characterization system for coral reef ecology, enabling the transition from single visual recognition to comprehensive multi-dimensional perception of ecological health status.

### *6.3. Low-Data Efficient Learning Paradigm and Construction of the Ecological Monitoring Closed-Loop System*

The scarcity of annotated data and inherent class imbalance represent long-standing fundamental bottlenecks restricting the generalization improvement and large-scale application of coral visual recognition models. Developing data-efficient learning paradigms with low annotation dependency is the core pathway to overcome these constraints. Future research can focus on advancing weakly-supervised and semi-supervised learning frameworks: leveraging low-cost annotation information such as image-level labels and sparse point annotations, combined with techniques including pseudo-label generation and consistency regularization, to fully exploit the value of massive unannotated underwater imagery and substantially reduce the reliance of models on expert-intensive dense annotations. For the challenges of limited samples of rare coral taxa and difficult adaptation to cross-sea scenarios, few-shot learning methods including meta-learning, metric learning, and prompt learning can be applied to achieve rapid model adaptation to new species and novel environments. In addition, generative AI can be employed to build sample synthesis models for rare taxa, which, coupled with the optimization of ecology-oriented loss functions, can further alleviate model bias caused by class imbalance and improve the recognition reliability for priority taxa in ecological conservation.

The core value of coral visual recognition technologies lies in providing reliable support for coral reef ecological conservation and management decision-making. Future research must break through the functional boundary of single image recognition and establish a full-chain intelligent monitoring system aligned with the needs of ecological management. First, an ecology-oriented model evaluation system should be established, where core metrics for ecological management (including estimation error of live coral cover, statistical accuracy of species abundance, and accuracy of degradation trend identification) are incorporated into model optimization and performance validation, to achieve the in-depth coupling of algorithm development and ecological monitoring requirements. On this basis, efforts should be made to promote the formulation of standardized technical specifications for intelligent coral reef monitoring. Ultimately, the development of an integrated intelligent coral reef monitoring platform will enable the standardized deployment of models and one-stop management of monitoring outputs, facilitating the translation of technologies from laboratory-based algorithm validation into practical decision-support tools for coral reef conservation and management.

## **7. Conclusions and Future Perspectives**

This paper systematically reviews the developmental trajectory of visual recognition technologies for coral reefs. From early traditional vision methods based on color, texture and morphological features, to deep learning models centered on CNNs, and further to the development stage focusing on cross-domain generalization, multimodal fusion and ecological decision coupling, the research paradigms have undergone a multi-stage evolution: from local feature engineering to data-driven representation learning, and then to system-level intelligent monitoring construction.

Technically, deep learning has significantly improved the accuracy of coral classification and segmentation, enabling pixel-level estimation of coral cover and health status recognition. However, the large-scale field application of these technologies is still constrained by key bottlenecks, including insufficient cross-regional generalization capability of models, high cost of data annotation, and unstable conversion from model outputs to ecological indicators. Relying solely on the optimization of model structure can hardly address the challenges of distribution shift caused by environmental variations. Future research

should therefore pay more attention to the construction of general visual representations, multi-source data fusion, and the establishment of continuous learning mechanisms.

In terms of data modality, the scope of coral visual recognition has gradually expanded from 2D optical imagery to the integration of 3D structure reconstruction and hyperspectral visual sensing. 3D point cloud models provide a spatial basis for the quantification of key ecological indicators such as reef structural complexity, surface area, and rugosity, while hyperspectral information offers a physical foundation for fine-grained species differentiation and health status discrimination. The integration of multimodal data helps alleviate the sensitivity of single visual information to illumination conditions and water turbidity, providing more robust data support for cross-regional ecological monitoring.

From the application perspective, future research on coral visual recognition should break through the boundary of mere image recognition and transform toward a holistic intelligent ecological monitoring system. Model outputs should not be limited to pixel-level classification results, but should establish a robust mapping relationship with ecological statistical indicators, temporal change trends, and environmental driving factors. By constructing a closed-loop system covering data acquisition, model inference, performance evaluation, and decision support, dynamic model updating and ecological risk early warning under long-term monitoring can be realized, enabling visual recognition technologies to truly serve the conservation and management decision-making of coral reefs.

Overall, coral visual recognition technology is in a critical transition phase from model performance-driven development to system integration-driven innovation. Future research should make sustained advances in the following three aspects: (1) constructing a self-supervised pre-training system for underwater environments to improve the cross-scene generalizability of models; (2) strengthening the capability of multimodal fusion and 3D ecological structure characterization to build a feature space more consistent with ecological reality; (3) promoting the in-depth coupling between visual recognition outputs and ecological indicator systems to establish a long-term, large-scale, and intelligent framework for coral reef monitoring.

Through the coordinated development of algorithm innovation, data system construction and ecological application, coral visual recognition technology is expected to evolve from an auxiliary analysis tool to a critical decision support system for coral reef ecological management, providing a sustained and reliable technical foundation for global coral reef conservation.

**Author Contributions:** Conceptualization, H.L. and X.G.; investigation, Y.L. and X.W.; writing—original draft preparation, H.L. and Y.L.; writing—review and editing, Q.L., Y.X., X.W. and X.G.; visualization, Y.L. and Y.X.; supervision, X.G.; project administration, Q.L. and X.G.; funding acquisition, X.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Key Research and Development Program of China (Grant No. 2024YFC2815400), the Young Taishan Scholars Program of Shandong Province (Grant No. TSQN202507107), the Shandong Provincial Natural Science Foundation (Grant No. ZR2025MS647), the European Commission (HORIZON MSCA-2024-PF-01, Grant No. 101200637), the Opening Fund of the State Key Laboratory of Intelligent Construction and Healthy Operation and Maintenance of Deep Underground Engineering (Grant No. SDGZ2529), and the Opening Fund of the State Key Laboratory of Ocean Engineering (Grant No. GKZD010090).

**Data Availability Statement:** The original contributions presented in this study are included in the article material. Further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** Hu Liu and Yuelin Xu are employees of R&D Department, Qingdao Robotfish Marine Technology Co., Ltd. The other authors declare no conflicts of interest. R&D Department, Qingdao Robotfish Marine Technology Co., Ltd. had no role in the design of the study; in the

collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Bhuyan, S.; Jenzri, M.; Islam, T.; Adikari, D.; Hoque, M. Climate change impacts on coral reefs and emerging resilience pathways: A systematic review. *Ocean Coast. Manag.* **2026**, *276*, 108134. [\[CrossRef\]](#)
2. Bellwood, D.R.; Hughes, T.P.; Folke, C.; Nyström, M. Confronting the coral reef crisis. *Nature* **2004**, *429*, 827–833. [\[CrossRef\]](#)
3. Capotondi, A.; Rodrigues, R.R.; Gupta, A.S.; Benthuisen, J.A.; Deser, C.; Frölicher, T.L.; Lovenduski, N.S.; Amaya, D.J.; Le Grix, N.; Xu, T.; et al. A global overview of marine heatwaves in a changing climate. *Commun. Earth Environ.* **2024**, *5*, 701. [\[CrossRef\]](#)
4. Dong, T.; Zeng, Z.; Pan, M.; Wang, D.; Chen, Y.; Liang, L.; Yang, S.; Jin, Y.; Luo, S.; Liang, S.; et al. Record-breaking 2023 marine heatwaves. *Science* **2025**, *389*, 369–374. [\[CrossRef\]](#)
5. Walker, J.L.; Zeng, Z.; Wu, C.L.; Jaffe, J.S.; Frasier, K.E.; Sandin, S.S. Underwater object detection under domain shift. *IEEE J. Ocean. Eng.* **2024**, *49*, 1209–1219. [\[CrossRef\]](#)
6. Emslie, M.J.; Ceccarelli, D.M.; Logan, M.; Blandford, M.I.; Bray, P.; Campili, A.R.; Cantin, N.; Choukroun, S.; Cole, A.; Jonker, M.J.; et al. Anthropogenic climate change causes substantial loss of coral on the northern Great Barrier Reef during the 2024 bleaching event. *Coral Reefs* **2025**, 1–17. [\[CrossRef\]](#)
7. González-Rivero, M.; Beijbom, O.; Rodriguez-Ramirez, A.; Holtrop, T.; González-Marrero, Y.; Ganase, A.; Roelfsema, C.; Phinn, S.; Hoegh-Guldberg, O. Scaling up ecological measurements of coral reefs using semi-automated field image collection and analysis. *Remote Sens.* **2016**, *8*, 30. [\[CrossRef\]](#)
8. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *60*, 84–90. [\[CrossRef\]](#)
9. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2017; pp. 7263–7271.
10. Mahmood, A.; Bennamoun, M.; An, S.; Sohel, F.; Boussaid, F.; Hovey, R.; Kendrick G.Fisher, R.B. Deep learning for coral classification. In *Handbook of Neural Computation*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 383–401.
11. Li, M.; Zhang, H.; Gruen, A.; Li, D. A survey on underwater coral image segmentation based on deep learning. *Geo-Spat. Inf. Sci.* **2025**, *28*, 472–496. [\[CrossRef\]](#)
12. Malik, H.; Hanapiah, M.F.M.; Toha, S.F.; Mustapa, M.Z.; Azmi, A.H.; Johari, H.A.; Idris, A.S.; Ibrahim, A.M.; De Wilde, P.; Alqedra, A.I.A. Advancing coral reef monitoring: A deep learning perspective on automated segmentation and classification. *Discov. Appl. Sci.* **2026**, *8*, 217. [\[CrossRef\]](#)
13. Jonker, M.; Johns, K.; Osborne, K. Surveys of benthic reef communities using underwater digital photography and counts of juvenile corals. *Long-Term Monit. Great Barrier Reef. Stand. Oper. Proced.* **2008**, *10*, 36.
14. Nocerino, E.; Menna, F.; Gruen, A.; Troyer, M.; Capra, A.; Castagnetti, C.; Rossi, P.; Brooks, A.J.; Schmitt, R.J.; Holbrook, S.J. Coral reef monitoring by scuba divers using underwater photogrammetry and geodetic surveying. *Remote Sens.* **2020**, *12*, 3036. [\[CrossRef\]](#)
15. Monk, J.; Barrett, N.; Bridge, T.; Carroll, A.; Friedman, A.; Ierodiaconou, D.; Jordan, A.; Kendrick, G.; Lucieer, V. *Marine Sampling Field Manual for AUV's (Autonomous Underwater Vehicles)*; NESP Marine Biodiversity Hub: Canberra, Australia, 2018.
16. Liu, X.; Ho, L.; Bruneel, S.; Goethals, P. Applications of unmanned vehicle systems for multi-spatial scale monitoring and management of aquatic ecosystems: A review. *Ecol. Inform.* **2025**, *85*, 102926. [\[CrossRef\]](#)
17. Desa, E.; Madhan, R.; Maurya, P.; Navelkar, G.; Mascarenhas, A.; Prabhudesai, S.; Afzulpurkar, S.; Desa, E.; Pascoal, A.; Nambiar, M. The detection of annual hypoxia in a low latitude freshwater reservoir in Kerala, India, using the small AUV Maya. *Mar. Technol. Soc. J.* **2009**, *43*, 60–70. [\[CrossRef\]](#)
18. Llewellyn, L.E.; Bainbridge, S.J. Getting up close and personal: The need to immerse autonomous vehicles in coral reefs. In *OCEANS 2015-MTS/IEEE Washington*; IEEE: New York, NY, USA, 2015; pp. 1–9.
19. Maurya, P.; Balakrishnan, M.; Raj, R.; Naik, L.; Fernandes, L.; Dabholkar, N.; Prabhudesai, S.; Ravindran, J.; Agarwadekar, Y.; Navelkar, G. Augmented coral reef monitoring using a stationary reef monitoring system. *Ecol. Inform.* **2023**, *74*, 101972. [\[CrossRef\]](#)
20. Mallet, D.; Pelletier, D. Underwater video techniques for observing coastal marine biodiversity: A review of sixty years of publications (1952–2012). *Fish. Res.* **2014**, *154*, 44–62. [\[CrossRef\]](#)
21. Singh, H.; Armstrong, R.; Gilbes, F.; Eustice, R.; Roman, C.; Pizarro, O.; Torres, J. Imaging coral I: Imaging coral habitats with the SeaBED AUV. *Subsurf. Sens. Technol. Appl.* **2004**, *5*, 25–42. [\[CrossRef\]](#)
22. Satoh, N.; Sinniger, F.; Narisoko, H.; Nagahama, S.; Okada, N.; Shimizu, Y.; Yoshioka, Y.; Hisata, K.; Harii, S. Using underwater mini-ROV for coral eDNA survey: A case study in Okinawan mesophotic ecosystems. *Coral Reefs* **2025**, *44*, 209–219. [\[CrossRef\]](#)

23. Alexander, J.B.; Bunce, M.; White, N.; Wilkinson, S.P.; Adam, A.A.S.; Berry, T.; Stat, M.; Thomas, L.; Newman, S.J.; Dugal, L.; et al. Development of a multi-assay approach for monitoring coral diversity using eDNA metabarcoding. *Coral Reefs* **2020**, *39*, 159–171. [\[CrossRef\]](#)
24. Sauder, J.; Banc-Prandi, G.; Meibom, A.; Tuia, D. Scalable semantic 3D mapping of coral reefs with deep learning. *Methods Ecol. Evol.* **2024**, *15*, 916–934. [\[CrossRef\]](#)
25. Ferrari, R.; McKinnon, D.; He, H.; Smith, R.N.; Corke, P.; González-Rivero, M.; Mumby, P.J.; Upcroft, B. Quantifying multiscale habitat structural complexity: A cost-effective framework for underwater 3D modelling. *Remote Sens.* **2016**, *8*, 113. [\[CrossRef\]](#)
26. de Oliveira, L.M.C.; Lim, A.; Conti, L.A.; Wheeler, A.J. High-resolution 3D mapping of cold-water coral reefs using machine learning. *Front. Environ. Sci.* **2022**, *10*, 1044706. [\[CrossRef\]](#)
27. Price, D.M.; Robert, K.; Callaway, A.; Lo Lacono, C.; Hall, R.A.; Huvenne, V.A. Using 3D photogrammetry from ROV video to quantify cold-water coral reef structural complexity and investigate its influence on biodiversity and community assemblage. *Coral Reefs* **2019**, *38*, 1007–1021. [\[CrossRef\]](#)
28. Ma, B.; Zhao, F.; Xi, D.; Wang, J.; Shao, X.; Wang, S.; Tabeta, S.; Mizuno, K. A new coral classification method using speed sea scanner-portable and deep learning-based point cloud semantic segmentation. In *OCEANS 2024-Halifax*; IEEE: New York, NY, USA, 2024; pp. 1–4.
29. Smart, J.N.; Cowley, D.; Rivera, D.E.C.; Hafizt, M.; Lyons, M.; Phinn, S.; Roelfsema, C. Hyperspectral Remote Sensing for Monitoring Coral Reef and Seagrass Ecosystems: Capabilities, Limitations and Future Directions. *Mar. Environ. Res.* **2026**, *217*, 107928. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Khaled, M.A.; Abdelsalam, A.A. Discrimination of some red sea coral reef species based on hyperspectral signature field data. *Sci. Afr.* **2025**, *28*, e02696. [\[CrossRef\]](#)
31. Battach, Y.; Felemban, A.; Khan, F.F.; Radwan, Y.A.; Li, X.; Marchese, F.; Beery, S.; Jones, B.H.; Benzoni, F.; Elhoseiny, M. ReefNet: A Large scale, Taxonomically Enriched Dataset and Benchmark for Hard Coral Classification. *arXiv* **2025**, arXiv:2510.16822. [\[CrossRef\]](#)
32. Beijbom, O.; Edmunds, P.J.; Kline, D.I.; Mitchell, B.G.; Kriegman, D. Automated annotation of coral reef survey images. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2012; pp. 1170–1177.
33. Sauder, J.; Domazetoski, V.; Banc-Prandi, G.; Perna, G.; Meibom, A.; Tuia, D. The Coralscapes Dataset: Semantic scene understanding in coral reefs. *arXiv* **2025**, arXiv:2503.20000. [\[CrossRef\]](#)
34. Morsy, S.; Yáñez-Suárez, A.B.; Robert, K. 3D colored point cloud classification of a deep-sea cold-water coral and sponge habitat using geometric features and machine learning algorithms. *Front. Remote Sens.* **2025**, *6*, 1680353. [\[CrossRef\]](#)
35. Shao, X.; Chen, H.; Magson, K.; Wang, J.; Song, J.; Chen, J.; Sasaki, J. Deep learning for multilabel classification of coral reef conditions in the indo-pacific using underwater photo transect method. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **2024**, *34*, e4241. [\[CrossRef\]](#)
36. Chowdhury, A.; Jahan, M.; Kaisar, S.; Khoda, M.E.; Rajin, S.A.K.; Naha, R. Coral Reef Surveillance with Machine Learning: A Review of Datasets, Techniques, and Challenges. *Electronics* **2024**, *13*, 5027. [\[CrossRef\]](#)
37. Piñeros, V.J.; Reveles-Espinoza, A.M.; Monroy, J.A. From remote sensing to artificial intelligence in coral reef monitoring. *Machines* **2024**, *12*, 693. [\[CrossRef\]](#)
38. Marcos, M.S.A.C.; Soriano, M.N.; Saloma, C.A. Classification of coral reef images from underwater video using neural networks. *Opt. Express* **2005**, *13*, 8766–8771. [\[CrossRef\]](#)
39. Han, H.; Wang, W.; Zhang, G.; Li, M.; Wang, Y. Enhancing vision-language models with morphological and taxonomic knowledge: Towards coral recognition for ocean health. *Proc. AAAI Conf. Artif. Intell.* **2025**, *39*, 28052–28060. [\[CrossRef\]](#)
40. Cheng, B.; Girshick, R.; Dollár, P.; Berg, A.C.; Kirillov, A. Boundary IoU: Improving object-centric image segmentation evaluation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2021; pp. 15334–15342.
41. Zhang, H.; Grün, A.; Li, M. Deep learning for semantic segmentation of coral images in underwater photogrammetry. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2022**, *2*, 343–350. [\[CrossRef\]](#)
42. Lin, T.-Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In *Proceedings of the IEEE International Conference on Computer Vision*; IEEE: New York, NY, USA, 2017.
43. Kervadec, H.; Bouchtiba, J.; Desrosiers, C.; Granger, E.; Dolz, J.; Ayed, I.B. Boundary loss for highly unbalanced segmentation. *Med. Image Anal.* **2021**, *67*, 101851. [\[CrossRef\]](#)
44. Coralscapes Dataset: Semantic Scene Understanding in Coral Reefs. Available online: <https://josauder.github.io/coralscapes/> (accessed on 28 March 2026).
45. Kaur, A.; Rani, S.; Shabaz, M. Underwater image dehazing using a hybrid GAN with bottleneck attention and improved Retinex-based optimization. *Sci. Rep.* **2025**, *15*, 26132. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Fayaz, S.; Parah, S.A.; Qureshi, G. Efficient underwater image restoration utilizing modified dark channel prior. *Multimed. Tools Appl.* **2023**, *82*, 14731–14753. [\[CrossRef\]](#)

47. Islam, M.J.; Xia, Y.; Sattar, J. Fast underwater image enhancement for improved visual perception. *IEEE Robot. Autom. Lett.* **2020**, *5*, 3227–3234. [\[CrossRef\]](#)
48. Oladi, M.; Ghazilou, A.; Rouzbehani, S.; Polgardani, N.Z.; Kor, K.; Ershadifar, H. Photographic application of the Coral Health Chart in turbid environments: The efficiency of image enhancement and restoration methods. *J. Exp. Mar. Biol. Ecol.* **2022**, *547*, 151676. [\[CrossRef\]](#)
49. Yao, F.; Zhang, H.; Gong, Y.; Zhang, Q.; Xiao, P. A study of enhanced visual perception of marine biology images based on diffusion-GAN. *Complex Intell. Syst.* **2025**, *11*, 227. [\[CrossRef\]](#)
50. Soriano, M.; Marcos, S.; Saloma, C.; Quibilan, M.; Alino, P. Image classification of coral reef components from underwater color video. In *MTS/IEEE Oceans 2001. An Ocean Odyssey. Conference Proceedings (IEEE Cat. No. 01CH37295)*; IEEE: New York, NY, USA, 2001; Volume 2, pp. 1008–1013.
51. Stough, J.; Greer, L.; Matt, B. Texture and color distribution-based classification for live coral detection. In *Proceedings of the 12th International Coral Reef Symposium*; James Cook University: Townsville, Australia, 2012; pp. 9–13.
52. Mahmood, A.; Bennamoun, M.; An, S.; Sohel, F.; Boussaid, F.; Hovey, R.; Kendrick, G.; Fisher, R.B. Coral classification with hybrid feature representations. In *2016 IEEE International Conference on Image Processing (ICIP)*; IEEE: New York, NY, USA, 2016; pp. 519–523.
53. Shakoor, M.H.; Boostani, R. A novel advanced local binary pattern for image-based coral reef classification. *Multimed. Tools Appl.* **2018**, *77*, 2561–2591. [\[CrossRef\]](#)
54. Gómez-Ríos, A.; Tabik, S.; Luengo, J.; Shihavuddin, A.; Herrera, F. Coral species identification with texture or structure images using a two-level classifier based on Convolutional Neural Networks. *Knowl.-Based Syst.* **2019**, *184*, 104891. [\[CrossRef\]](#)
55. Tang, Q.; Li, N.; Zhang, Y.; Dong, Z.; Zheng, Y.; Bao, J.; Zhang, J. High-precision classification of benthic habitat sediments in shallow waters of islands by multi-source data. *J. Oceanol. Limnol.* **2025**, *44*, 99–108. [\[CrossRef\]](#)
56. Bhanu, S.; Saravanan, K.S. Coral reef image classification using deep CNNs, transformer features, and YOLOv8. *Mar. Syst. Ocean Technol.* **2026**, *21*, 6. [\[CrossRef\]](#)
57. Anwar, S.; Li, C. Diving deeper into underwater image enhancement: A survey. *Signal Process. Image Commun.* **2020**, *89*, 115978. [\[CrossRef\]](#)
58. Hidalgo, F.; Bräunl, T. Evaluation of several feature detectors/extractors on underwater images towards vSLAM. *Sensors* **2020**, *20*, 4343. [\[CrossRef\]](#)
59. Zhang, T.; Zhang, X.; Ke, X.; Liu, C.; Xu, X.; Zhan, X.; Wang, C.; Ahmad, I.; Zhou, Y.; Pan, D.; et al. HOG-ShipCLSNet: A novel deep learning network with hog feature fusion for SAR ship classification. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5210322. [\[CrossRef\]](#)
60. Zhang, L.; Lin, X. Some considerations of classification for high dimension low-sample size data. *Stat. Methods Med. Res.* **2013**, *22*, 537–550. [\[CrossRef\]](#)
61. Katuwal, R.; Suganthan, P.N. Enhancing multi-class classification of random forest using random vector functional neural network and oblique decision surfaces. In *2018 International Joint Conference on Neural Networks (IJCNN)*; IEEE: New York, NY, USA, 2018; Volume 42, pp. 1–8.
62. Xu, G.; Zhou, D.; Yuan, L.; Guo, W.; Huang, Z.; Zhang, Y. Vision-based underwater target real-time detection for autonomous underwater vehicle subsea exploration. *Front. Mar. Sci.* **2023**, *10*, 1112310. [\[CrossRef\]](#)
63. Sulistianingsih, N.; Martono, G.H. Coral Reefs Health Detection Using Pretrained Deep Learning Models and Feature Importance Visualization on Underwater Images. In *2025 7th International Conference on Cybernetics and Intelligent System (ICORIS)*; IEEE: New York, NY, USA, 2025; pp. 1–6.
64. Ogidi, M.S.; Sah, M. Binary Classification of Coral Reefs Using Deep Learning for Enhanced Monitoring. In *2025 9th International Symposium on Innovative Approaches in Smart Technologies (ISAS)*; IEEE: New York, NY, USA, 2025; pp. 1–7.
65. Gómez-Ríos, A.; Tabik, S.; Luengo, J.; Shihavuddin, A.S.M.; Krawczyk, B.; Herrera, F. Towards highly accurate coral texture images classification using deep convolutional neural networks and data augmentation. *Expert Syst. Appl.* **2019**, *118*, 315–328. [\[CrossRef\]](#)
66. Kaur, A.; Kukreja, V.; Upadhyay, D.; Aeri, M.; Sharma, R. An efficient deep learning-based InceptionV3 model for coral reef classification. In *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*; IEEE: New York, NY, USA, 2024; pp. 1–6.
67. Artates, J.S.; Taylar, J.V.; Medina, R.P. Deep Learning-Based Coral Reef Health Assessment Using Modified Inception V3. In *2024 15th International Conference on Information and Communication Technology Convergence (ICTC)*; IEEE: New York, NY, USA, 2024; pp. 1857–1862.
68. Hadi, H.P.; Rachmawanto, E.H.; Ali, R.R. Comparison of densenet-121 and mobilenet for coral reef classification. *MATRIK J. Manaj. Tek. Inform. Rekayasa Komput.* **2024**, *23*, 333–342. [\[CrossRef\]](#)
69. Zheng, Z.; Liang, H.; Wut, F.H.; Wong, Y.H.; Chui, A.P.Y.; Yeung, S.K. HKCoral: Benchmark for dense coral growth form segmentation in the wild. *IEEE J. Ocean. Eng.* **2025**, *50*, 697–713. [\[CrossRef\]](#)

70. Contini, M.; Illien, V.; Poulain, S.; Bernard, S.; Barde, J.; Bonhommeau, S.; Joly, A. The point is the mask: Scaling coral reef segmentation with weak supervision. In *ICCV 2025-Joint Workshop on Marine Vision*; The Computer Vision Foundation: New York, NY, USA, 2025.
71. Zhang, H.; Li, M.; Zhong, J.; Qin, J. CNet: A novel seabed coral reef image segmentation approach based on deep learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*; IEEE: New York, NY, USA, 2024; pp. 767–775.
72. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2015; pp. 3431–3440.
73. Patro, B.; Namboodiri, V.P. Differential attention for visual question answering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: New York, NY, USA, 2018.
74. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; MICCAI: Abu Dhabi, United Arab Emirates, 2015.
75. Pochamreddy, M.R.; Reddy, K.A.; Verma, S.S.; Sharma, R. Enhancing U-Net for Semantic Segmentation with Integrated Attention Mechanisms. In *2024 International Conference on Augmented Reality, Intelligent Systems, and Industrial Automation (ARIIA)*; IEEE: New York, NY, USA, 2024; pp. 1–6.
76. Li, X.; Lu, Q.; Li, Y.; Li, M.; Qi, Y. Optimized Unet with attention mechanism for multi-scale semantic segmentation. In *2025 5th International Conference on Consumer Electronics and Computer Engineering (ICCECE)*; IEEE: New York, NY, USA, 2025; pp. 535–539.
77. Qin, J.; Li, M.; Li, D.; Gruen, A.; Gong, J.; Liao, X. Causal learning-driven semantic segmentation for robust coral health status identification. *ISPRS J. Photogramm. Remote Sens.* **2025**, *229*, 78–91. [[CrossRef](#)]
78. Pavithra, S. An efficient approach to detect and segment underwater images using Swin Transformer. *Results Eng.* **2024**, *23*, 102460. [[CrossRef](#)]
79. Bo, A.H.B.; Yongsiriwit, K. Performance Evaluation of YOLO-based Models for Coral Species Detection. In *2025 9th International Conference on Information Technology (InCIT)*; IEEE: New York, NY, USA, 2025; pp. 605–610.
80. Das, D.; Mukherjee, K. Advanced YOLOv8 with CBAM and BiFPN for Assessing Coral Health in Underwater Environments. In *2024 IEEE Silchar Subsection Conference (SILCON 2024)*; IEEE: New York, NY, USA, 2024; pp. 1–6.
81. Saragih, R.E.; Husin, H.S.; Mursalin, M.K.N. Coral Detection based on Optimised Lightweight YOLO Model. *Indones. J. Inf. Syst.* **2025**, *8*, 10–20. [[CrossRef](#)]
82. Narang, G.; Berardini, D.; Pietrini, R.; Tasseti, A.N.; Mancini, A.; Galdelli, A. Edge-AI for buoy detection and mussel farming: A comparative study of YOLO frameworks. In *2024 20th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)*; IEEE: New York, NY, USA, 2024; pp. 1–8.
83. Shi, X.; Liu, M.; Zhou, S.; Tang, X.; Cai, K.; Song, J.; Zhao, Y. Sea treasure target detection based on improved Yolov7 and TensorRT deployments. In *IET Conference Proceedings CP842*; The Institution of Engineering and Technology: Stevenage, UK, 2023; Volume 2023, pp. 69–73.
84. Santoso, S.A.; Jaya, I.; Priandana, K. Optimizing coral fish detection: Faster r-cnn, ssd mobilenet, yolov5 comparison. *Indones. J. Comput. Cybern. Syst.* **2024**, *18*, 2. [[CrossRef](#)]
85. Liu, Y.; An, B.; Chen, S.; Zhao, D. Multi-target detection and tracking of shallow marine organisms based on improved YOLO v5 and DeepSORT. *IET Image Process.* **2024**, *18*, 2273–2290. [[CrossRef](#)]
86. Oraño, J.F.V.; Feliscuzo, L.S.; Aliac, C.J.G.; Napala, J.J.O. Transformer-Based Coral Detection: Applying RT-DETR for the Identification of Scleractinian Corals in Quadrat Images. In *Novel & Intelligent Digital Systems Conferences*; Springer Nature: Cham, Switzerland, 2025; pp. 373–384.
87. Jeothi, R.P.; Raj, S.H.; Anil, P.; Vaishnavi, R.J. Integrated model for underwater image enhancement and coral health classification. In *2024 International Conference on Intelligent Computing and Emerging Communication Technologies (ICEC)*; IEEE: New York, NY, USA, 2024; pp. 1–7.
88. Wang, J.; Li, P.; Deng, J.; Du, Y.; Zhuang, J.; Liang, P.; Liu, P. CA-GAN: Class-condition attention GAN for underwater image enhancement. *IEEE Access* **2020**, *8*, 130719–130728. [[CrossRef](#)]
89. Hambarde, P.; Murala, S.; Dhall, A. UW-GAN: Single-image depth estimation and image enhancement for underwater images. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 5018412. [[CrossRef](#)]
90. Li, Z.; Zhao, S.; Lu, Y.; Song, C.; Huang, R.; Yu, K. Deep learning-based automatic estimation of live coral cover from underwater video for coral reef health monitoring. *J. Mar. Sci. Eng.* **2024**, *12*, 1980. [[CrossRef](#)]
91. Lu, Z.; Liao, L.; Xie, X.; Yuan, H. SCoralDet: Efficient real-time underwater soft coral detection with YOLO. *Ecol. Inform.* **2025**, *85*, 102937. [[CrossRef](#)]
92. Džaja, B.; Turić, H.; Pleština, V. Fading Colours: Investigating Spectral Attenuation in Underwater Photography with Distance. In *2025 10th International Conference on Smart and Sustainable Technologies (SpliTech)*; IEEE: New York, NY, USA, 2025; pp. 1–7.
93. Indriani, R.; Adiwijaya, R.; Jarmawijaya, N.; Yusuf, M.; Agoes, A.S. Applying Transfer Learning ResNet-50 for Tracking and Classification of A Coral Reef in Development The Mobile Application with Scrum Framework. *J. Inf. Technol.* **2023**, *4*, 24–29. [[CrossRef](#)]

94. Han, H.; Wang, W.; Zhang, G.; Li, M.; Wang, Y. Cross-Domain Coral Image Classification Using Dual-Stream Hierarchical Neural Networks. In *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*; ACM: New York, NY, USA, 2024; pp. 945–952.
95. Liu, X.; Yoo, C.; Xing, F.; Oh, H.; Fakhri, G.E.; Kang, J.W.; Woo, J. Deep unsupervised domain adaptation: A review of recent advances and perspectives. *arXiv* **2022**, arXiv:2208.07422. [[CrossRef](#)]
96. Long, R.; Zhou, J.; Liang, N.; Yang, Y.; Shen, H. Deep unsupervised adversarial domain adaptation for underwater source range estimation. *J. Acoust. Soc. Am.* **2023**, *154*, 3125–3144. [[CrossRef](#)]
97. Doig, H.; Pizarro, O.; Williams, S. Training marine species object detectors with synthetic images and unsupervised domain adaptation. *Front. Mar. Sci.* **2025**, *12*, 1581778. [[CrossRef](#)]
98. Shao, X.; Chen, H.; Zhao, F.; Magson, K.; Chen, J.; Li, P.; Wang, J.; Sasaki, J. Multi-label classification for multi-temporal, multi-spatial coral reef condition monitoring using vision foundation model with adapter learning. *Mar. Pollut. Bull.* **2026**, *223*, 119054. [[CrossRef](#)] [[PubMed](#)]
99. Moniz, N.; Cerqueira, V. Automated imbalanced classification via meta-learning. *Expert Syst. Appl.* **2021**, *178*, 115011. [[CrossRef](#)]
100. Furtado, D.P.; Vieira, E.A.; Nascimento, W.F.; Inagaki, K.Y.; Bleuel, J.; Alves, M.A.Z.; Longo, G.O.; Oliveira, L.S. #DeOlhoNosCorais: A polygonal annotated dataset to optimize coral monitoring. *PeerJ* **2023**, *11*, e16219.
101. Ovaskainen, O.; Winter, S.; Tikhonov, G.; Abrego, N.; Anslan, S.; Dewaard, J.R.; Dewaard, S.L.; Fisher, B.L.; Furneaux, B.; Hardwick, B.; et al. Common to rare transfer learning (CORAL) enables inference and prediction for a quarter million rare Malagasy arthropods. *Nat. Methods* **2025**, *22*, 2074–2082. [[CrossRef](#)] [[PubMed](#)]
102. Couëdel, M.; Dettai, A.; Guillaume, M.M.; Bonillo, C.; Frattini, B.; Bruggemann, J.H. Settlement patterns and temporal successions of coral reef cryptic communities affect diversity assessments using autonomous reef monitoring structures (ARMS). *Sci. Rep.* **2024**, *14*, 27061. [[CrossRef](#)] [[PubMed](#)]
103. Bryant, D.E.P.; Rodriguez-Ramirez, A.; Phinn, S.; González-Rivero, M.; Brown, K.T.; Neal, B.P.; Hoegh-Guldberg, O.; Dove, S. Comparison of two photographic methodologies for collecting and analyzing the condition of coral reef ecosystems. *Ecosphere* **2017**, *8*, e01971. [[CrossRef](#)]
104. ŞAHİN, E.; Arslan, N.N.; Özdemir, D. Unlocking the black box: An in-depth review on interpretability, explainability, and reliability in deep learning. *Neural Comput. Appl.* **2025**, *37*, 859–965. [[CrossRef](#)]
105. Mohammadi, M.; Huang, S.E.; Barua, T.; Rekleitis, I.; Islam, M.J.; Zand, R. Caveline detection at the edge for autonomous underwater cave exploration and mapping. In *2023 International Conference on Machine Learning and Applications (ICMLA)*; IEEE: New York, NY, USA, 2023; pp. 1392–1398.
106. Contini, M.; Illien, V.; Barde, J.; Poulain, S.; Bernard, S.; Joly, A.; Bonhommeau, S. From underwater to drone: A novel multi-scale knowledge distillation approach for coral reef monitoring. *Ecol. Inform.* **2025**, *89*, 103149. [[CrossRef](#)]
107. Picek, L. *Coral Reef Annotation, Localisation and Pixel-Wise Classification Using Mask R-CNN and Bag of Tricks*; CEUR: Munich, Germany, 2020.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.