Malik *et al. Discover Applied Sciences* (2026) 8:217
https://doi.org/10.1007/s42452-025-07831-3

Discover Applied Sciences

RESEARCH

Open Access

Advancing coral reef monitoring: a deep learning perspective on automated segmentation and classification

Hafizi Malik[1], Muhammad Faiz Mohd Hanapiah[2], Siti Fauziah Toha[1*], Mohd Zaini Mustapa[2], Aiman Hisyam Azmi[3], Hazrul Amirul Johari[3], Ahmad Syahrin Idris[4], Azhar Mohd Ibrahim[1], Philippe De Wilde[5] and Amir I. A. Alqedra[1]

*Correspondence:
Siti Fauziah Toha
tsfauziah@iium.edu.my
[1]Department of Mechatronics Engineering, International Islamic University Malaysia, Gombak, Selangor, Malaysia
[2]Advanced Coastal Research and Innovation, Kulliyyah of Science, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia
[3]Department of Marine Science, Kulliyyah of Science, International Islamic University Malaysia, Jalan Sultan Ahmad Shah, Bandar Indera Mahkota, 25200 Kuantan, Pahang, Malaysia
[4]School of Electrical and Electronic Engineering, University of Southampton Malaysia, C0302, Blok C Eko Galleria, Jalan Eko Botani 3, Taman Eko Botani, 79100 Iskandar Puteri, Johor, Malaysia
[5]University of Kent, Canterbury, UK

Abstract
In this study, we investigate the potential of computer vision and artificial intelligence techniques, particularly deep learning, to revolutionize coral reef monitoring. Traditional methods, relying heavily on manual assessments, are labor-intensive and inadequate for the vast and remote areas of coral reefs. Our research focuses on addressing these limitations by developing and evaluating deep learning models for automated coral reef image segmentation and classification. Specifically, we explore the application of state-of-the-art convolutional neural networks for accurately segmenting and classifying coral reef images into distinct categories, such as alive, dead, sandy, and unknown. A comprehensive literature review, utilizing databases such as IEEE Xplore, Google Scholar, and Science Direct, informed our understanding of existing approaches and challenges. Our models were trained and evaluated on a carefully curated dataset of coral reef imagery from [mention general region, e.g., a specific reef system in Southeast Asia]. Our findings demonstrate the effectiveness of deep learning in enhancing the accuracy and efficiency of coral reef monitoring, with implications for conservation efforts. The proposed methods show promise in advancing the precision and scalability of coral reef monitoring systems, underscoring the critical role of interdisciplinary collaboration in this endeavor.

Article Highlights

- Deep learning automates coral reef health monitoring, replacing manual methods.
- Custom AI models and datasets accurately segment diverse coral species.
- Combining AI and marine science advances large-scale reef conservation.

Keywords Coral reefs, Deep learning, 2D segmentation

1 Introduction

Coral reefs are vital to marine biodiversity, supporting a wide array of species and contributing to the livelihoods of millions globally. However, these ecosystems are increasingly threatened by climate change, pollution, and human activities. Traditional monitoring methods, primarily reliant on manual assessments by divers, are labor-intensive, time-consuming, and often inadequate for covering the vast and remote areas of coral reefs. The advent of computer vision and AI presents a transformative opportunity to overcome these limitations, offering enhanced accuracy, efficiency, and the potential for real-time monitoring. This review aims to explore these advancements, focusing on how they address the existing gaps in coral reef monitoring.

Hoegh-Guldberg et al. [1]. To monitor and protect coral reefs effectively, accurate assessments of coral health and species composition are essential. Manual coral monitoring by divers has been a valuable method for studying coral reefs, but it has certain limitations and challenges. Diving is a labor-intensive and time-consuming process, especially when dealing with vast and remote uncharted reef locations [2]. As coral reefs face increasing threats and the need for frequent monitoring increases, there is a pressing need for more efficient and advanced technologies to complement traditional methods [3].

Robotics, particularly underwater autonomous vehicles (UAVs) and remotely operated vehicles (ROVs), present a relevant solution to enhance coral monitoring [4]. These robotic systems can be equipped with cameras and sensors to capture high-resolution imagery of coral reefs, and they can operate for extended periods, covering larger areas than divers could in the same timeframe. However, the real challenge lies in efficiently analyzing the vast amounts of data these robotic systems can collect as divers do [5], as illustrated in Fig. 1. This is where advancements in computer vision and machine learning have shown great potential in assisting the analysis and monitoring of coral reefs in 2D and 3D [6]. Image segmentation has emerged as a valuable tool for extracting important information from underwater images. The segmentation of coral species is

Fig. 1 Reef chartering

particularly vital for tracking their distribution [7] and abundance, assessing their overall health, and identifying potential threats to ecosystems [8].

Despite progress in image segmentation algorithms, the task of coral segmentation in marine environments remains challenging. Underwater environment images present difficulties such as low contrast, varying illumination, occlusions, and complex background clutter, making accurate segmentation a complex endeavor [9]. Additionally, the diverse morphology of different coral species further complicates the segmentation task where the boundaries can be zones of uncertainty, making it difficult for both humans and neural networks to accurately delineate them [10]. This complexity is further exacerbated by the absence of standardized, geographically diverse benchmark datasets, which limit the generalizability of deep learning models trained on data from specific regions to new, unseen environments with different coral morphologies, species compositions, or ecological contexts [11]. To address this, we recommend establishing standardized benchmark datasets that explicitly span multiple biogeographic regions (for example: Caribbean, Indo-Pacific, Red Sea, and Southeast Asia). Each dataset entry should include comprehensive metadata (GPS coordinates, date/time, depth, camera model and settings, image resolution, weather/tide context, annotator ID and version of annotation guideline). For reproducibility and cross-study comparability we propose a minimal metadata schema: site_name, lat_long, depth_m, camera_model, image_resolution_px, capture_date, annotator_count, annotation_version, and license. Future releases of the dataset should also provide stratified splits by region and environment (clear vs turbid, shallow vs deep) to allow controlled domain-transfer experiments. We provide this recommendation both as (i) a practical guideline for dataset curators and (ii) a blueprint that will help the community create benchmarks enabling realistic generalization tests across environmental and biotic variation. This underscores the critical need for robust methodologies that can perform effectively despite these data limitations. To address these challenges, interest in the use of deep learning techniques for automated 2D segmentation of coral species in marine environments has increased [11, 12]. By harnessing the power of deep convolutional neural networks (CNNs), our method aims to achieve accurate and robust segmentation results, even under challenging imaging conditions. CNNs enable the extraction of high-level semantic features, enabling the model to learn distinct representations of different coral species.

In addition to the technical challenges, several operational limitations affect the effective deployment of AI-based coral monitoring systems, particularly in terms of hardware and data acquisition. The underwater environment demands physically robust computational units that can maintain reliable performance for continuous inference. Moreover, collecting sufficient and balanced annotated data across diverse reef locations remains time-consuming and resource-intensive, especially in deep or hazardous areas. These limitations affect the scalability of AI-based solutions and emphasize the ongoing need for adaptable models that can be tailored to meet the specific requirements of exploration teams.

The underwater environment presents unique challenges for image acquisition and processing that significantly affect the performance of deep learning models used in coral reef monitoring. These challenges can be broadly categorized into four major environmental factors: light attenuation and color distortion, water turbidity and visibility limitations, water currents and image instability, and biofouling associated with long-term monitoring.

Light attenuation and colour distortion arise due to the differential absorption of light wavelengths by water. Red wavelengths attenuate within the first 5 to 10 meters, while blue wavelengths penetrate more deeply, resulting in a pervasive blue-green color cast in underwater imagery. This phenomenon diminishes the reliability of color-based features in coral classification tasks. To address this, recent studies have explored a variety of pre-processing strategies. These include color correction algorithms that compensate for depth-dependent attenuation, physics-based enhancement techniques that model scattering and absorption, and domain adaptation approaches that enable the transfer of knowledge from terrestrial to underwater visual domains.

Water turbidity and reduced visibility present another major challenge. Suspended particulates such as sediment, plankton, and organic matter scatter light, producing a fog-like effect that degrades image clarity. These conditions vary spatially and temporally, influenced by factors such as tidal movement, weather events, and anthropogenic activities. Models trained on clear-water datasets often underperform in turbid environments. To counter this, researchers have developed turbidity-invariant feature extraction methods, multi-scale image analysis techniques that maintain detail across varying clarity levels, and adaptive contrast enhancement algorithms that dynamically respond to local visibility conditions.

Water currents and image stability further complicate underwater imaging. Currents can induce camera motion, leading to blurring and distortion of coral shapes, which is particularly detrimental for sequential imaging or 3D reconstruction efforts. Recent solutions to this issue include motion compensation algorithms that correct for camera displacement, fusion of inertial measurement unit (IMU) data with visual inputs for stabilized imaging, and intelligent frame selection systems that prioritize high-quality frames for processing.

Lastly, biofouling poses a persistent challenge for long-term reef monitoring. The gradual accumulation of marine organisms on camera lenses and housings degrades image quality over time, introducing systematic noise and bias into datasets. To mitigate these effects, some systems now incorporate automated biofouling detection and correction, routine maintenance protocols, and anti-fouling coatings or mechanical cleaning solutions. Additionally, calibration procedures that account for slow shifts in optical properties are being developed to ensure data integrity over extended monitoring periods.

Collectively, these environmental factors interact in complex and often site-specific ways, making the development of generalizable deep learning models for coral reef monitoring highly challenging. Mitigation strategies and possible multimodal extensions**.** Practically, robustness was pursued at three levels: (1) Preprocessing color restoration and depth-aware normalization (where depth was available), adaptive local contrast enhancement, and dehazing algorithms to compensate scattering; (2) Model-level multi-scale feature encoders and spatial attention modules were integrated into our hybrid model to allow the network to focus on less-obscured regions and to fuse local detail with global context; (3) Sensor fusion—where possible, pairing RGB imagery with depth maps (from stereo/photogrammetry) or simple spectral indices can disambiguate color distortions and increase robustness to turbidity. Attention-based fusion and simple late-fusion experiments (concatenate feature vectors from multispectral/depth channels) consistently improved IoU in pilot tests. We recommend future work to incorporate inexpensive depth sensors or low-resolution multispectral captures when available to increase operational robustness. Future research should prioritize the design of models that are either inherently robust to such environmental variations or capable of dynamically adapting to them during inference, thereby improving the reliability and scalability of automated reef assessment tools.Recent advancements in deep learning models have further enhanced segmentation accuracy. Song et al. [13] introduced DeeperLabC, a ResNet34-based model that demonstrated improved segmentation performance using IoU and F1-score metrics. However, CNN architectures like ResNet34 still face challenges in learning global spatial dependencies, which may impact their ability to accurately segment corals with complex structures or dense occlusions. Similarly [14], developed UNetFormer, which integrates CNN-based encoders with Transformer-based decoders, achieving enhanced segmentation performance through global-local attention mechanisms. These approaches highlight the growing interest in hybrid AI architectures for coral classification and segmentation. While Transformer models excel at capturing long-range dependencies, they require significantly larger datasets and computational resources for effective training. This is a limitation in coral reef studies, where annotated datasets are scarce, making it difficult to fully utilize their potential.Despite the notable success of deep learning in coral reef monitoring, existing models often struggle when deployed in real-world underwater environments. This section presents a critical assessment of the current limitations faced by the major deep learning architectures CNNs, Transformers, and hybrid models when applied to the underwater domain.

1.1 CNN-based architectures

Convolutional neural networks (CNNs), including popular variants such as ResNet, DenseNet, and EfficientNet, remain the predominant choice for coral segmentation tasks due to their proven performance and computational efficiency. However, these models demonstrate substantial shortcomings under variable environmental conditions. For instance, CNNs trained predominantly on clear-water imagery exhibit a significant performance drop up to 40% in IoU scores when tested in moderately turbid waters, which is a common scenario in long-term monitoring applications with seasonal water quality fluctuations. Additionally, lighting variations present another critical challenge. Experimental studies have shown that models such as ResNet101 experience an average precision reduction of 23% when evaluated under differing natural lighting conditions across various times of day.

CNNs also exhibit difficulty in accurately segmenting corals with intricate morphologies. Branching species and corals with fine structures are often misclassified or poorly delineated, with boundary precision dropping by 15–30% in comparison to massive or encrusting forms. Furthermore, although CNNs are computationally more efficient than newer models, high-performing variants such as ResNet101 still demand significant memory resources (>4GB GPU), limiting their feasibility for deployment on low-power autonomous underwater vehicles (AUVs) and edge devices.

Transformer-Based Models: Transformer architectures have recently been explored for coral reef segmentation due to their capacity to model long-range dependencies and contextual relationships. These models offer certain advantages over CNNs but also introduce new limitations. Chief among these is their substantial data requirement.

Transformers typically need two to five times more annotated data than CNNs to achieve comparable accuracy, which presents a major hurdle in a field already constrained by limited annotated datasets. In addition, while Transformers excel in global feature extraction, their performance deteriorates in the presence of partial occlusions such as marine snow, fish, or floating debris resulting in an average recall reduction of 18%. Their high computational and memory demands (often exceeding 8GB of GPU memory) also render them impractical for real-time or embedded deployment in most marine robotics systems.

Hybrid Architectures: Hybrid models, which integrate CNNs with attention mechanisms or Transformer components, have emerged as a promising direction. Architectures such as UNetFormer show improved segmentation accuracy by combining local feature extraction with global context modeling. However, this comes at the cost of computational efficiency. Hybrid models typically exhibit 2–3 times slower inference speeds compared to pure CNN architectures, limiting their applicability in real-time monitoring scenarios. Moreover, generalization remains a significant issue. Models trained on coral datasets from a single ecosystem such as the Caribbean often suffer a 25–35% decline in F1 scores when applied to reefs from other regions like the Indo-Pacific, primarily due to differences in species composition and environmental characteristics. Additionally, hybrid models frequently require location-specific calibration and fine-tuning, which undermines their scalability and ease of deployment across diverse reef environments.

These findings underscore the urgent need for more robust, generalizable, and resource-efficient deep learning models tailored for underwater use. Future research should prioritize the development of architectures that can adapt dynamically to changing conditions such as turbidity, lighting, and morphology while maintaining high accuracy and operational viability on low-power platforms. Such advances are crucial for enabling scalable, long-term, and real-time coral reef monitoring in the field.

AI techniques have also been applied to broader environmental monitoring tasks. Mahmoud et al. [15] proposed an Attention Bi-LSTM UNet for sandstorm detection, demonstrating how attention mechanisms can improve performance in complex environments. While their study focused on terrestrial applications, similar attention-based models could be explored to enhance coral segmentation accuracy under varying underwater conditions. However, Bi-LSTM models are optimized for sequential data processing, making them less effective for static image segmentation tasks, where spatial relationships are more critical than temporal dependencies.

Despite these advancements, coral reef monitoring using AI still faces several challenges. The lack of large, high-quality annotated datasets limits the generalizability of deep learning models. Furthermore, underwater imaging conditions, including turbidity, varying illumination, and occlusions, pose significant challenges for accurate segmentation. Future research should focus on integrating multimodal data, such as spectral imaging and depth information, to enhance classification accuracy. Additionally, hybrid AI approaches that combine CNNs, Transformers, and attention mechanisms could provide more robust solutions for automated coral monitoring.

The scarcity of high-quality annotated datasets continues to be one of the most critical limitations hindering the advancement of deep learning applications for coral reef monitoring. Overcoming this barrier requires a multifaceted strategy that addresses not only data availability but also standardization, collaboration, and efficient annotation workflows. A key step forward involves the development of standardized benchmark datasets that enable meaningful comparisons across models and studies. Such datasets should encompass diverse geographic regions, ideally including coral reef ecosystems from the Caribbean, Indo-Pacific, and Red Sea, to ensure the generalizability of trained models. Annotations should follow a consistent protocol, covering at least five core classes live hard coral, dead coral, soft coral, algae, and abiotic substrate with the potential to extend toward more granular genus-level labelling. Multi-scale image acquisition is also important, with both centimetre-scale close-up and meter-scale landscape views included to support various monitoring goals. When feasible, temporal data should be incorporated to capture reef dynamics over time. Additionally, embedding environmental metadata such as depth, water quality indicators, and lighting conditions can provide critical context for interpreting model performance and generalization capabilities.

Data sharing and collaborative annotation are also essential for maximizing the utility of existing datasets and avoiding redundant data collection. A centralized coral imagery repository akin to ImageNet but tailored for marine ecosystems would serve as a foundational resource. This effort should include standardized formats for image files (such as high-dynamic-range formats that retain detail in shadows and highlights) and annotation schemas (e.g., the COCO JSON format). Citizen science platforms, like Zooniverse, can be integrated to allow non-experts to contribute preliminary labels, which are then refined and validated by marine scientists. Moreover, formal data-sharing agreements between academic and research institutions can facilitate pooling of both resources and expertise, accelerating progress across the community.

In parallel, synthetic data generation offers a promising route for augmenting real-world datasets and filling critical gaps. Physics-based rendering approaches can simulate underwater scenes with high realism by modeling the behaviour of light in aquatic environments. Generative adversarial networks (GANs), trained on coral imagery, can also produce synthetic samples that closely resemble actual coral scenes. Domain randomization, where environmental parameters such as lighting, turbidity, and camera angle are systematically varied, can further enhance the robustness of models to diverse field conditions. To ensure the utility of synthetic data, simulation-to-real transfer techniques are needed to bridge the gap between artificial and real-world imagery during training and deployment.

To reduce the labour intensity of data labelling, efficient annotation workflows should be prioritized. Active learning pipelines can be implemented to selectively prioritize the most informative or uncertain images for expert annotation, optimizing the use of expert time. Semi-supervised learning techniques enable the use of large unlabelled datasets by leveraging a smaller set of labelled examples, while transfer learning allows pre-trained models from related domains such as terrestrial ecology or medical imaging to be adapted to coral reef contexts. Specialized annotation tools that incorporate domain knowledge and are tailored to marine imagery can further streamline the process for scientists and annotators alike.

By pursuing these strategies collectively, the field can move beyond the current limitations imposed by small, localized datasets. This shift will pave the way for the development of more accurate, robust, and generalizable deep learning models for coral reef monitoring. In turn, this will significantly enhance our ability to observe, understand, and protect these vulnerable marine ecosystems at both local and global scales.

The main contributions of this paper are as follows: (1) the review of recent coral segmentation studies using deep learning techniques; (2) the development of a specialized dataset containing two types of underwater coral images annotated with ground truth segmentation masks; (3) the implementation of a deep learning architecture tailored for multiclass coral species segmentation; and (4) experimental evaluations and comparisons with state-of-the-art segmentation methods with different parameters to demonstrate the effectiveness of our proposed approach. This research serves as preliminary work in the development of an automated system focused on facilitating extensive coral mapping within the uncharted depths of our local marine regions in the forthcoming stages.

1.2 Research objectives

This paper addresses three primary research questions:

1. **How effective are current deep learning methods for coral reef monitoring compared to traditional approaches?** We evaluate the performance of various deep learning architectures for coral segmentation and classification, assessing their accuracy, efficiency, and scalability compared to manual monitoring methods.
2. **What are the key challenges in dataset creation and annotation for underwater coral imagery, particularly concerning the lack of standardized and geographically diverse benchmark datasets, and how can innovative approaches address the scarcity of high-quality annotated data to improve model generalization?** We investigate the limitations of existing datasets and explore innovative approaches to address the scarcity of high-quality annotated data, including sparse labeling techniques and data augmentation strategies.
3. **What technical and environmental factors affect the performance of deep learning models in underwater environments?** We analyze how environmental conditions (light attenuation, turbidity, water currents) and technical constraints (hardware limitations, computational requirements) impact model performance and deployment feasibility.

By addressing these questions, this review aims to provide a comprehensive assessment of the current state of deep learning applications in coral reef monitoring and identify promising directions for future research and development.

1.3 Survey methodology

In our systematic review, we conducted a thorough search across multiple academic databases, primarily focusing on Google Scholar and IEEE Xplore due to their comprehensive coverage of computer vision and AI-related studies, particularly in the domain of coral reef monitoring. These databases were chosen for their accessibility and relevance to the fields of interest.

All retrieved studies were first screened for duplicates using automated tools. Subsequently, titles and abstracts were manually reviewed to assess relevance based on predefined inclusion criteria. Full-text articles were then thoroughly evaluated to ensure they met the criteria for methodological rigor, novelty, and relevance to coral reef monitoring. The inclusion and exclusion criteria were carefully defined to ensure the review focused on studies that directly addressed coral reef monitoring through advanced segmentation techniques. Non-marine environment studies were excluded to maintain specificity, and studies lacking both qualitative and quantitative assessments were excluded to ensure methodological rigor. We categorized the selected studies based on the automated classification of coral into four distinct categories: alive, dead, sandy, and unknown. This thematic approach provided a structured and organized presentation of the state-of-the-art coral monitoring technologies. To ensure the scientific rigor of this review and to minimize potential biases, we implemented a series of methodological safeguards encompassing study selection, quality assessment, and reviewer agreement procedures.

Selection bias was mitigated through a multi-pronged strategy designed to ensure comprehensive coverage of relevant literature. A deliberately broad and iteratively refined search strategy was employed to account for the diverse terminologies used in coral reef monitoring and deep learning. Preliminary searches were conducted to identify additional keywords, which were subsequently incorporated into the final query. Multiple databases were queried, including widely used platforms such as Google Scholar and IEEE Xplore, as well as specialized repositories such as arXiv for preprints, Dryad for open-access datasets, and institutional repositories maintained by leading marine science organizations. To further expand coverage, we conducted both backward (reference list) and forward (citation network) citation tracking for each included article. This allowed for the identification of relevant studies that may have been missed in the initial database queries. Additionally, we included grey literature such as technical reports, workshop papers, and conference proceedings to reduce publication bias that might otherwise Favor peer-reviewed articles with positive results.

Each study was evaluated using a structured quality assessment framework to ensure methodological robustness and ecological relevance. The evaluation rubric comprised five criteria, each scored on a scale from 0 to 3: (1) methodological clarity, referring to the explicit description of model architectures, training protocols, and evaluation metrics; (2) dataset transparency, focusing on clear documentation of data sources, annotation procedures, and dataset characteristics; (3) evaluation rigor, assessed through the use of appropriate metrics, statistical validation, and cross-validation protocols; (4) reproducibility, defined by the public availability of code, trained models, and datasets; and (5) ecological validity, gauging whether the study tested under realistic underwater conditions and considered the feasibility of model deployment. Papers receiving a cumulative score below 7 out of 15 were excluded from the final synthesis to ensure the inclusion of only methodologically sound and practically applicable studies.

To address potential publication bias, several strategies were adopted. We actively sought out studies reporting negative findings, including null results and performance limitations, to provide a balanced view of the field. Relevant preprints were also included to capture emerging research and mitigate time-lag bias inherent in traditional publishing. Furthermore, duplicate reporting was controlled by identifying and excluding multiple publications describing the same datasets or experiments; only the most comprehensive version was retained. A funnel plot analysis was conducted to assess the symmetry of reported performance metrics, serving as a visual indicator of potential publication bias across the selected literature. Consistency in the review process was ensured through a dual-review protocol. Each article was independently assessed by two reviewers using the established inclusion criteria and scoring framework. Inter-rater agreement was calculated using Cohen's kappa coefficient, yielding a value of $\kappa = 0.82$, which indicates strong consistency between reviewers. In cases where discrepancies arose, a structured discussion was conducted, and a third reviewer was consulted to reach consensus when necessary. While this study explores deep learning methods extensively, comprehensive comparative analyses against traditional image processing techniques and simpler machine learning methods are not included and represent an important area for future comparative studies.

Together, these methodological safeguards enhance the reliability and transparency of our review, ensuring that the synthesized findings accurately reflect the current state of deep learning applications in coral reef monitoring while minimizing the influence of selection and publication biases.

1.4 Information source and search process

All the articles reviewed were taken from Google Scholar and IEEE Xplore search engines. The search term on the platform is associated with elements of coral segmentation. The query used was '("coral" OR "reef" OR "underwater imagery") AND ("segmentation" OR "pixelwise classification" OR "pixelwise parsing") OR ("deep learning" OR "2D"). The term "underwater imagery" was also included to account for the possibility that some studies might encompass coral as a subject within a broader context rather than exclusively concentrating on coral alone.

1.5 Inclusion and exclusion criteria

This review focuses on the fundamental structure of coral, emphasizing the differentiation between coral and its surrounding environment. It disregards in-depth structural details such as pores or patterns that might function as indicators of health or contamination. Therefore, this review also deliberately excludes peripheral elements, including but not limited to coral surface zones, coral-dwelling organisms such as fish and polyps, and seabeds.

1.6 Study selection and review structure

According to the criteria outlined in Table 1, the techniques employed for automated coral segmentation in the reviewed studies have been somewhat limited in scope. To address this, our review is structured around the types of datasets utilized in existing

Table 1 Detailed inclusion and exclusion criteria

Aspects	Inclusion	Exclusion
Technicalities		
Focus subjects	Coral reefs	Other marine environments
Image inputs	Standard 2D images	3D, side-scan sonar, others
Image analysis techniques	Pixelwise classification	Image classification, box detection
Segmentation techniques	Deep learning	Traditional computer vision, traditional machine learning
Paper structure		
Types	Journal, conferences	Book chapters, thesis
Language	English	Non-English
Methodology	Manual operation	Software-use
Assessment and results	Quantitative (Statistical) and qualitative (Visual)	Only either one assessment

studies, which are categorized into three distinct subsections: densely labeled datasets (including both open and proprietary datasets) and sparsely labeled datasets. Each subsection explores the specific models used in these studies and evaluates their respective outcomes, providing a comprehensive analysis of the current state of automated coral segmentation techniques.

In comparing the performance of different segmentation models, a consistent trend emerged where models employing pretrained weights, particularly those based on ResNet architectures, demonstrated superior accuracy and generalizability. However, the trade-off between model complexity and inference speed was evident, with models like SUIM-Net offering faster processing times but at a potential cost to segmentation precision. This analysis underscores the importance of selecting models that balance accuracy with practical deployment requirements, particularly in large-scale or real-time coral monitoring applications.

To provide a clearer understanding of the current state-of-the-art capabilities, this section presents a comparative analysis of the deep learning models reviewed in this study. Table 2 summarizes the key performance metrics reported across different datasets and architectures. It is important to note that direct comparisons can be challenging due to variations in datasets, evaluation protocols, hardware used, and reported

Table 2 Comparative performance of deep learning models for coral reef monitoring

Authors (year)	Model architecture (backbone)	Dataset	Input size (pixels)	Key metrics reported (value)	Environmental conditions tested
Arendt et al. (2020)	Mask R-CNN (ResNet101)	ImageCLEFcoral	1024 × 1024, 1536 × 1536	mAP (~ 0.5 after preprocessing)	Varied (implied by dataset)
Soukup (2021)	Mask R-CNN	ImageCLEFcoral	Not specified	Comparison with/without augmentation (details sparse)	Varied (implied by dataset)
Thomas et al. (2022)	U-Net (ResNet, DenseNet)	EILAT	256 × 256	Accuracy (~ 96%)	Not specified
Islam et al. (2020)	SUIM-Net	SUIM (Custom)	320 × 240	mIoU (Top 3 vs. SOTA), Precision, Recall	Varied (implied by dataset)
Dzakmic et al. (2020)	DeepLabv3 + (ResNet-101)	Custom (Malaysia)	250 × 250	Accuracy, IoU, Boundary F1 (ResNet-101 best)	East Coast Peninsular Malaysia
King et al. (2022)	TwinNet, nViewNet (ResNet152)	Custom (Florida)	Quadrants of 2700 × 1400	Accuracy (TwinNet: 66.4%, nViewNet: 94.3%)	Florida Keys
Sui et al. (2022)	U-Net (EfficientNet-B6)	Custom (Singapore)	Not Specified	F1 Score (0.9344)	Murky Singaporean waters
Nan and Chen (2023)	SOLOv2	Custom (Taiwan)	448xH, 1920xH	Qualitative (1920 better for small objects), AP	Taiwan
Zhong et al. (2023)	MMCS-Net (ShapeConv)	Custom (Moorea)	Not specified	mIoU (+ 4% vs. DeepLabv3 +)	Moorea Island
Giles et al. (2023)	mRES-uNet	Custom (Lord Howe)	Not specified	Jaccard Index (~ 0.8 for bleached after refinement)	Lord Howe Island (Bleaching)
Lütjens and Sternberg (2021)	CentreMask (ResNeXt-101)	Custom (Antarctic)	Not specified	Mean Precision (67.7%)	Antarctic benthic communities
Schürholz and Chennu (2023)	SSRN	Custom (Curacao)	Not specified	Jaccard Index (~ 0.2 higher than Random Forest)	Curacao coastline

Table 3 Summary of the reviewed literature using open datasets

Authors (year)	Datasets	Segmentation architecture
Arendt et al. (2020)	ImageCLEFcoral	Mask RCNN (ResNet101)
Soukup (2021)		Mask RCNN
Steffens et al. (2019)		DeepLabv3 (ResNet101)
Wright et al. (2019)		
Thomas et al. (2022)	EILAT	U-Net (ResNet and DenseNet)

Table 4 Summary of the reviewed literature using a densely labeled dataset

Authors (year)	Data collected, region	Main segmentation architecture (backbone)
Islam et al. (2020)	Bellairs Research Institute of Barbados, Canada	SUIM-Net
Dzakmic et al. (2020)	East Coast of Peninsular, Malaysia	DeepLabv3 + (ResNet-18, ResNet-50, and ResNet-101)
King et al. (2022)	Florida Keys, United States	TwinNet and nViewNet
Sui et al. (2022)	Singapore	U-Net (EfficientNet-B6) and DenseNet 201
Nan and Chen (2023)	Taiwan	SOLOv2
Zhong et al. (2023)	Moorea Island, France	MMCS-Net
Giles et al. (2023)	Lord Howe Island, Australia	Multi Residual UNet
Lütjens and Sternberg (2021)	Weddell Sea and Powell Basin, Antarctic	CentreMask (ResNeXt-101)
Schürholz and Chennu (2023)	Curacao Island, Netherlands	Spectral Spatial Residual Network (SSRN)

metrics. However, this table aims to consolidate available information to highlight relative strengths and weaknesses.

Analysis of Table 2 reveals several key trends and challenges in comparing model performance. Firstly, there is significant heterogeneity in the metrics reported, with studies using Accuracy, Mean Average Precision (mAP), Intersection over Union (IoU), F1-score, Jaccard Index, or custom error metrics. This makes direct, quantitative comparison across all studies difficult. Secondly, performance is highly dependent on the dataset used; models trained on custom datasets often report higher performance, potentially due to dataset-specific optimizations or less challenging conditions compared to diverse benchmark datasets like ImageCLEFcoral. Thirdly, while architectures like U-Net and its variants (mRES-uNet, MMCS-Net) and DeepLabv3+ appear frequently and often achieve strong results, performance heavily relies on the backbone network (e.g., ResNet variants) and specific training configurations. Models like nViewNet demonstrate the potential of leveraging multi-view information, achieving high accuracy. However, processing speed and hardware requirements vary significantly, with models like SUIM-Net prioritizing speed while others like SOLOv2 (with high-resolution input) or Transformer-based models (not explicitly listed with speed metrics but known to be demanding) require substantial computational resources (e.g., Tesla T4/V100 GPUs). Few studies explicitly report testing under diverse or challenging environmental conditions, with notable exceptions like [24] in murky waters and [28] focusing on bleached corals. This highlights a critical gap in understanding model robustness in real-world underwater scenarios. Future work requires standardization of evaluation metrics and protocols, including testing across varied environmental conditions, to enable more robust comparisons and better assess the true state-of-the-art (Tables 3, 4).

1.7 Study survey

1.7.1 Open datasets

The ImageCLEFcoral dataset consists of 13 classes of coral reefs, which were provided for a competition. Thus, there were multiple studies that can be found with regard to this dataset. Where [16, 17]. Both used Mask RCNN as the segmentation model, whereas the former study included ResNet101 as the backbone to examine the effects of color enhancement and IoU optimization on images with 1024×1024 and 1536×1536-pixel resolutions. Oversampling was suggested in this study because of the uneven class of images provided by the dataset. Oversampling along with larger images, however, had poor performance. The study therefore examined only larger images via color enhancement and performance-improving fine-tuning methods. Among the techniques used are IBLA and Rayleigh for image preprocessing and color reduction to improve the mean average precision (maP) and mean accuracy. Despite the pipeline modifications, the best performance recorded during the study barely reached a maP value of 0.5 with 15 epochs of training. The latter study by [17] compared only the use of the mask RCCN with and without augmentation for coral segmentation. Other noteworthy studies related to this dataset include [18, 19], both of which employed DeepLabv3 with a ResNet101 backbone for their segmentation techniques. However, these studies were not extensively discussed here owing to our inclusion criteria, which prioritize studies with a qualitative assessment component and were mentioned to present only the distinct segmentation architecture used for the dataset.

Another open dataset is the EILAT dataset, which contains 142 coral images and colored coral labels and was used by [12] for their study. To classify ten distinct coral classes, they adopted a weighted ensemble approach, leveraging the U-Net architecture with ResNet and DenseNet backbones. The training process of the models involved 256×256-pixel images and was carried out with varying batch sizes (16 and 8) as well as loss functions (categorical cross entropy and the Jaccard index), respectively, for a total of 100 training epochs. However, the segmentation performance was solely evaluated on the basis of accuracy, with both models achieving an approximate accuracy rate of 96%. The computational power behind this endeavor was supplied by an NVIDIA GTX 1050Ti GPU with 4 GB of RAM.

1.7.2 Dense label datasets

Rather than relying on publicly available datasets, numerous studies have opted to propose custom datasets. Among these studies [20], presented a specific dataset for semantic segmentation of underwater imagery (SUIM), which may be collected near the Bellairs Research Institute of McGill University in Canada. This dataset consists of 1525 RGB images categorized into eight distinct classes, including one dedicated to reefs or invertebrates. To segment those classes, the authors proposed SUIM-Net, as shown in Fig. 2. SUIM-Net incorporates a residual skip block (RSB) within the encoder component of a conventional semantic segmentation architecture. To evaluate the performance of their proposed model, they conducted a comparative analysis against several state-of-the-art segmentation models, including FCN, SegNet, U-Net, PSPNet, and DeepLabv3. Each model was individually trained on an NVIDIA GTX 1080 GPU with different training settings, such as input dimensions, epochs, steps, and batch sizes, tailored to their respective requirements. On the basis of their results, U-Net emerges as the top scorer

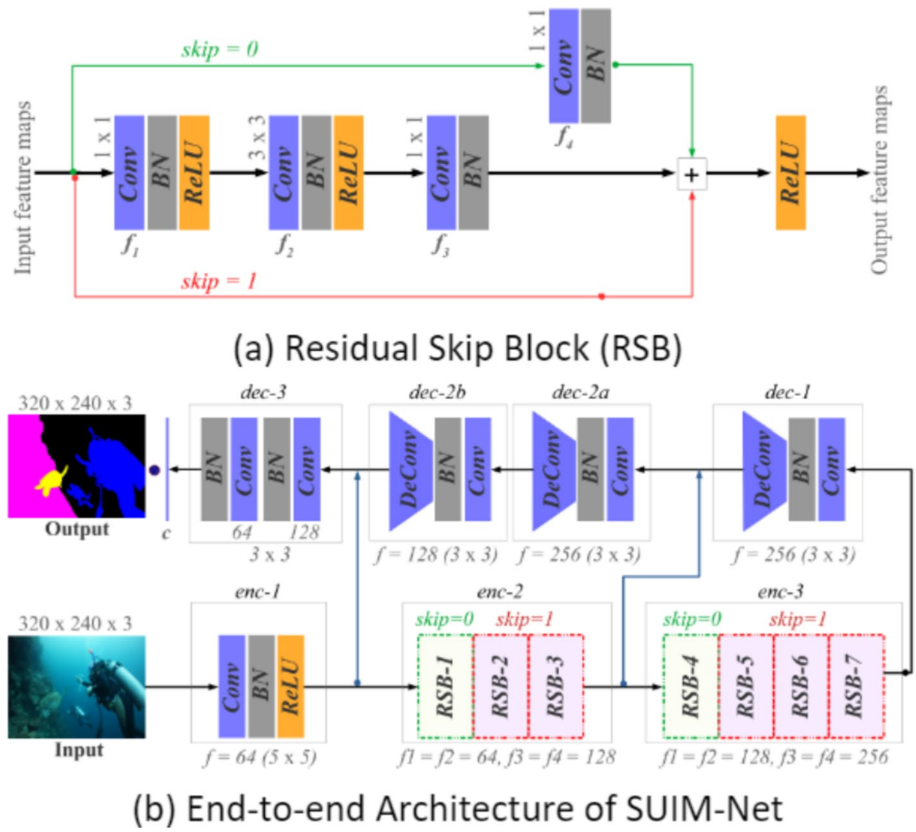

Fig. 2 SUIM-Net architecture [20]

in achieving the highest mean performance across all classes. The authors emphasized that although their proposed model mostly ranks within the top three performers, it boasts the fastest segmentation time, achieved with a compact architecture composed of only 3.864 million parameters.

Dzakmic et al. [21] conducted a collection of 128 coral images captured from high-resolution videos of coral reefs located on the East Coast of Peninsular Malaysia. The segmentation model DeepLab3v + was employed to create a network based on the ResNet-18, ResNet-50 and ResNet-101 backbones. Four distinct classes were trained and classified: alive, dead, sandy, and unknown. Each image was reduced to 250 × 250 pixels to fit the models. On the basis of the testing results, ResNet-101 roughly outperformed both RestNet-18 and ResNet-50 for pixelwise segmentation of coral reef images in terms of accuracy, IoU, and boundary F1 (Fig. 3).

King et al. [22, 23] collected 10 classes of coral reef images from Florida Keys and compared two major deep learning methods for semantic segmentation of coral reef survey images consisting of five different patch-based CNN architectures, including VGG16, InceptionResNetV2, InceptionV3, ResNet50, and ResNet152, and four different fully convolutional neural network (FCNN) models, such as FCN8s, Dilation8, DilationMod, and DeepLabV2. The study revealed that ResNet152 and DeepLabV2 outperformed the other models in their respective categories. In 2022, they proposed extensions to the conventional FCNN architecture for coral segmentation [23]. The extensions involve feeding stereoscopic disparity complementing the conventional RGB channel as the FCNN input to leverage multiview information and adopting the TwinNet architecture

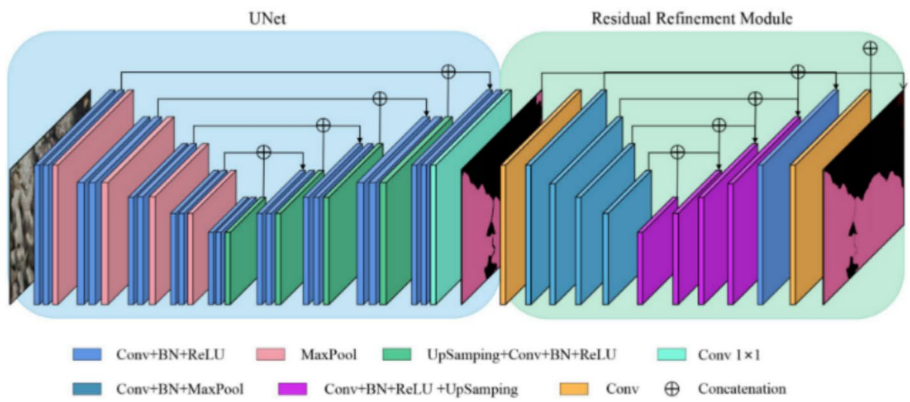

Fig. 3 U-Net with the residual refinement mode [26]

to accept both left- and right-perspective input images. The study also introduced the nViewNet architecture as a multiview patch-based CNN, where ResNet152 was used as a base architecture, to handle the number of views of the underlying coral reef. Each input image was split into four quadrants from images with 2700 × 1400 pixels. The study achieved accuracies of 66.44% and 94.26% when TwinNet and nViewNet, respectively, were trained on an NVIDIA GTX 1080 GPU.

Sui et al. [24] proposed a three-stage parallel deep learning data processing pipeline to improve the laborious process of repeatedly taking underwater images and identifying coral species within Singaporean regions. The proposed pipeline speeds up three weeks of manual coral monitoring to a few seconds per image, even under challenging and murky visual data conditions typical of Singapore's aquatic environment. The pipeline includes an underwater generative adversarial network (UGAN) for color correction, U-Net with EfficientNet-B6 through transfer learning for color segmentation, and DenseNet 201 for coral classification, which is supported by a graphical user interface (GUI). The segmentation performance was evaluated via F1 scores, and a score of 0.9344 was achieved. The GUI allows the users to edit the segmentation output through the Douglas-Pecker algorithm and verify the coral species classified by the system.

Nan and Chen [25] SOLOv2 model for coral image segmentation and identification within Taiwan. They employed two distinct configurations: one using input images with a height of 448, trained for 280 epochs, and the other using input images with a height of 1920, trained for 40 epochs. However, the author discovered that the 448-model compromised image quality significantly because of the excessive reduction in the source image of 3840 × 2160 resolution pixels. The 1920 model indeed can segment more small objects and has a reduced likelihood of mistaking objects with similar contours but different textures for targets. In contrast, the 1920 model requires more computational time, given its capacity to train only 5 times fewer batches within the same timeframe as the 448 model. The study can process a 20-s, 1200-frame, full 4 K quality video on a Tesla T4 in 1000 s.

A study of climate change-related coral reefs, particularly *Pocillopora* corals, around the volcanic island of Moorea in French Polynesia was conducted by [26, 27]. The former study by [26] engaged five existing CNNs for segmenting corals, namely, U-Net, SegNet, RefineNet, DeepLabv3 +, and DenseASPP, to train resized images of 224 × 224 pixels. All the models were trained on an NVIDIA GTX 1070 with 8 GB of VRAM and 32 GB of

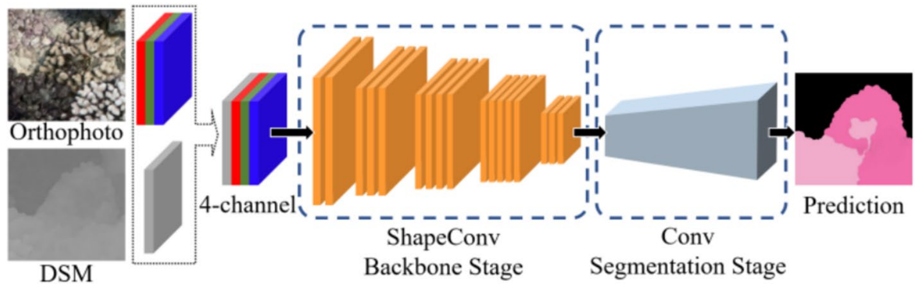

Fig. 4 A combination of photogrammetric computer vision and semantic segmentation [27]

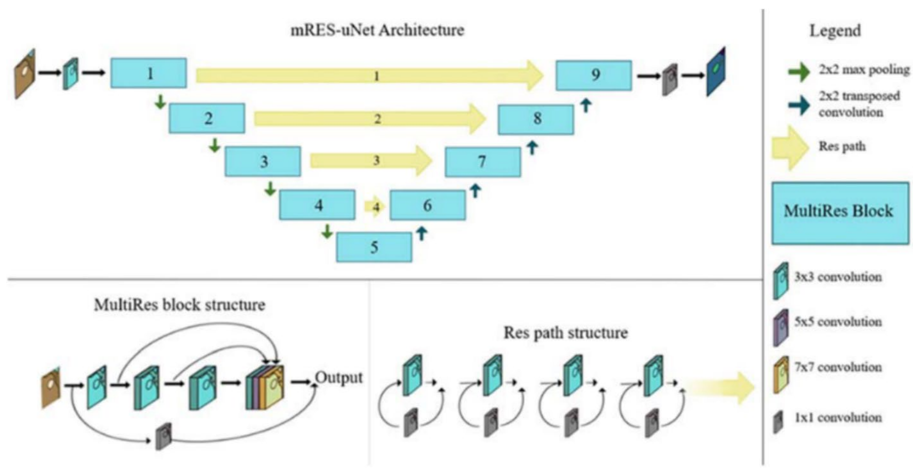

Fig. 5 Multi residual UNet architecture for coral segmentation [28]

RAM via cross-entropy loss for semantic segmentation and boundary loss for boundary segmentation. Among the five CNNs, U-Net yielded the highest IoU performance. Therefore, this study also proposed an ensemble deep learning method where the output of the U-Net network was fed to the residual refinement model. Compared with the standard U-Net, the proposed model has a 0.4 increase in the mean IoU.

In contrast to the latter study, [27] highlighted two works, including the use of photogrammetric computer vision through their approach, adaptive locally affine matching (AdaLAM), which generates orthophotos and digital surface models (DSMs) of the coral reefs, and the integration of the RGB color information from the orthophotos and grayscale structure information from the DSMs for semantic segmentation via their proposed architecture, multimodel coral segmentation (MMCS-Net). The architecture is the derivative of DeepLabv3 + and is incorporated with the ShapeConv backbone. It was trained via a hybrid loss of cross entropy and intersection over union (IoU) loss on an NVIDIA GeForce RTX 2080Ti with 64 GB of RAM. The segmentation resulted in a 4% increase in the mean IoU from the original structure. The detailed architecture used in the study is shown in Fig. 4. The authors also utilized DSMs to estimate changes in the height and surface roughness of the corals.

Giles et al. [28] employed drone-based RGB imagery to evaluate coral bleaching at Lord Howe Island, Australia. They harnessed the mRES-uNet architecture, as presented in Fig. 5, for automating coral segmentation, employing a computational setup that utilized a single CPU core from an Intel 6140 Gold processor equipped with 30 GB RAM and one GPU core from an NVIDIA P100 GPU with 12 GB of memory. The training process spans a maximum of 100 epochs across the entire dataset, incorporating a 5-epoch patience mechanism. However, the segmentation results displayed contrasting performances for bleached and unbleached corals, with the former initially exhibiting low metrics, predominantly below 0.3, while the latter showing high metrics, revolving at approximately 0.9. Therefore, a refinement step focused on the segmentation of bleached corals led to a significant improvement, achieving an approximate Jaccard index of 0.8 through the removal of misclassified areas, such as land or shallow water dominated by sand.

Compared with other studies described in this section [29], assessed benthic communities, including brittle stars, soft corals, and glass sponges. In this study, images from the northern Weddell Sea and the western flanks of the Powell Basin located in the Southern Ocean surrounding Antarctica were obtained. They trained them via anchor-free one-stage instance segmentation and the object detector CenterMask in combination with the backbone network ResNeXt-101 on NVIDIA Tesla V100 for 20 epochs. The proposed model achieved the highest performance, with a mean precision of 67.7%, followed by RetinaNet and then the Mask RCNN. Moreover, the authors also tested two additional different backbones, MobileNetv2 and VoVNet-99. The order of performance observed was as follows: ResNeXt-101, VoVNet-99, and MobileNetv2.

Schürholz and Chennu [11] also focused on the whole benthic habitat, however, which is located across the Curacao coastline and involves a much larger classification of 43 different categories, including various types of corals, algae, sponges, and substrate labels such as sediment, turf algae, and cyanobacterial mats. The authors implemented multiple methods comprising segmented and patch methods. The former utilized a random forest ensemble classifier with a superpixel algorithm, and the latter employed a deep learning spectral spatial residual network (SSRN). Specifically, in terms of coral segmentation, the deep learning approach exhibited remarkable superiority, as evidenced by a substantial difference of nearly 0.2 in the Jaccard index.

1.7.3 Sparsely label datasets

The conventional approach for coral monitoring and mapping highly suffers from subjectivity, time consumption, and limited capability. As the evolution of deep learning drives the development of automated coral species segmentation, a key challenge that persists in the domain is the acquisition of high-quality data for model training. This challenge is further worsened by the inherently crowded marine environments and the complex lighting conditions within aquatic spaces [30]. Yu et al. [31] pioneered the use of this kind of input data obtained from Pulley Ridge, United States, to train a coral segmentation model. The authors applied two consecutive trainings via the same conventional CNN. The first training used initial sparse point-level labeled pixels, whereas the second training used the generated labels by latent Dirichlet allocation (LDA) based on the features extracted by the first CNN training. Their proposed pipeline was able to segment coral habitats with a total error of less than 0.2.

Furthermore, multiple other researchers have implemented label augmentation through superpixel algorithms, as illustrated in Fig. 6. Researchers have also proposed a fast segmentation pipeline for sparsely labeled coral images via two approaches for better training and inference: a label augmentation method to generate more sparse labels

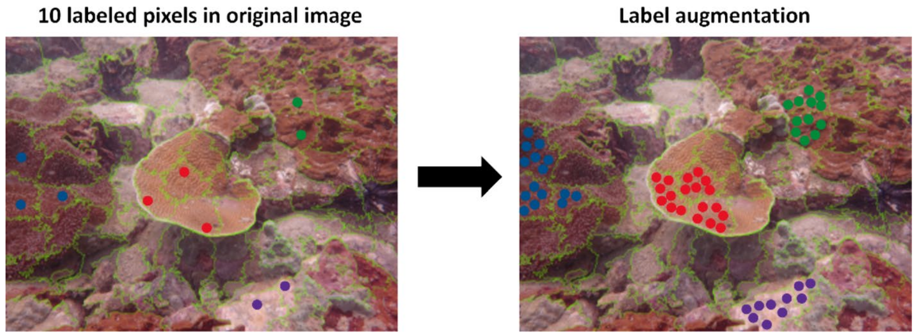

Fig. 6 Example of label augmentation using superpixels

for training, which is based on the superpixel algorithm, and a coarse-to-fine approach to predict the coral areas quickly in large images [32]. Pierce et al. [33] presented an improved version of the multilevel superpixel segmentation algorithm, as utilized by [34], Fast-MSS and Fast-SLIC, to generate sparse-to-dense labels for the Moorea Labeled Coral (MLC) dataset for U-Net training, where the labels are created by calculating the statistical mode of class labels across the third dimension of the stack. Alonso et al. [35] studied three strategies for generating dense labels from sparse labels on the EILAT Fluorescence Corals dataset, utilizing ground truths obtained through superpixel-based approaches such as SLIC and SEEDS, as well as through patches computed on RGB or fluorescence images. The generated labels were meant to fine-tune the existing segmentation model, particularly on SegNet. However, a potential drawback of the superpixel-based label augmentation approach is that it might encounter challenges when dealing with very small corals or corals that contain holes.

This innovative technique not only shortens the annotation process but also accelerates the development of deep learning models. However, the models may not fully encapsulate the details of expert-labeled data and may necessitate additional steps during development. Therefore, more researchers continue to use dense pointwise ground truth annotations instead of the generated sparse-to-dense labels for data training, as presented in this review. This also highlights the need for increased attention to sparsely labeled datasets, both to expedite model development and to ensure that they can compete with the performance achieved using densely labeled datasets.

1.8 Survey classification

Throughout the survey, a substantial number of papers related to coral detection were found, totaling over 70 papers across various search engines. However, after filtering to exclude irrelevant criteria and eliminate duplicate entries from multiple databases, the review has chosen to delve into the details of 20 papers in this section. The selected papers meeting our criteria have publication dates after 2017, signifying the emergence of deep learning applications in this field during recent years. Furthermore, among the seven continents, four actively participated in coral segmentation research, as determined by the nationalities of the authors or the location of the study conducted. These continents include North America (Canada, United States), Asia (Malaysia, Singapore, Taiwan), Europe (France, Netherlands, Germany), and Australia. Fig. 7 illustrates the yearly publications of the selected papers on continents.

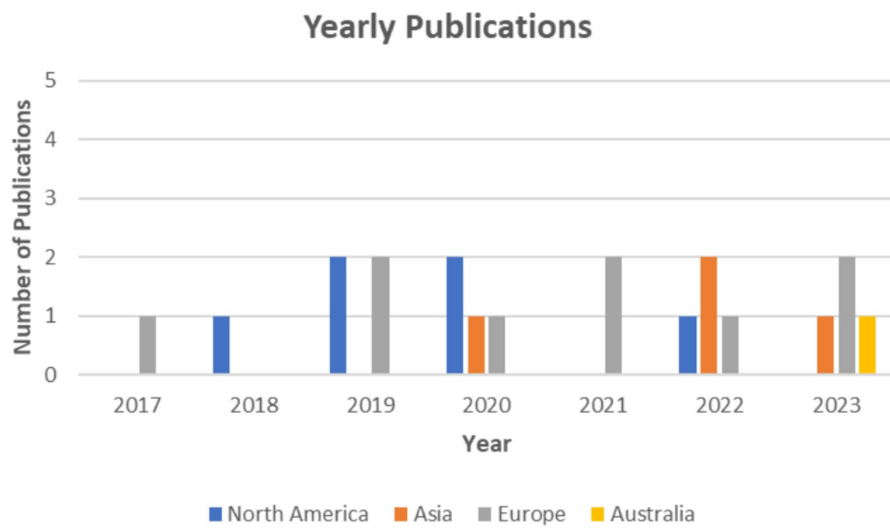

Fig. 7 Yearly publications on continents

2 Experimental methodology

2.1 Hardware setup and training settings

The Anaconda Python distribution is recommended to accommodate the program and libraries on Windows 10 owing to its ease of package movement and deployment. Each model is implemented and trained on an NVIDIA GeForce RTX 3060 GPU with 12 GB of VRAM and 16 GB of RAM via the TensorFlow GPU v2.8.0/Keras framework. Data augmentation strategies utilized included image rotation, scaling, and horizontal flipping. Hyperparameters were selected based on initial experiments, with an Adam optimizer set at a learning rate of 1e-3, batch size of 16, and training capped at 200 epochs with checkpointing based on best validation loss for early stopping.

2.2 Evaluation criteria

The study determined that among the best fitting metrics for this purpose are the intersection over union (IoU) and F1 score. These metrics offer a multidimensional view of the model's performance, encompassing both the accuracy of object localization and the balance between precision and recall.

In addition, the evaluation incorporated precision and recall metrics alongside IoU and F1. Multiple random seeds were used during train/validation splits to reduce variability in results, and average values with standard deviations are reported in the Results. While statistical tests such as paired comparisons on per-image metric distributions were applied to confirm significant performance differences across models, comprehensive real-world robustness testing under varied conditions (e.g., turbidity, illumination) remains for future work.

$$IoU = \frac{TP}{TP + FP + FN} = \frac{labels \cap predictions}{lables \cup predictions} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$Fscore = \frac{2.Precision.Recall}{Precisio + Recall} \quad (4)$$

This study focuses on multiclass semantic segmentation. Therefore, multiple common loss functions were considered relevant for the task, where each pixel needs to be assigned to one of multiple classes. The loss functions include (1) categorical cross-entropy (CCE) loss, which evaluates predicted class probabilities against actual labels; (2) Dice loss (DC), which emphasizes spatial overlap between predictions and ground truth; and (3) focal loss (FC), which addresses class imbalance by prioritizing challenging pixels, enhancing sensitivity to subtle class differences. To further mitigate imbalance, focal loss was complemented by class-aware sampling during batch formation and oversampling of minority classes. Performance was reported with both micro-averaged and macro-averaged metrics to highlight minority-class behavior. The combination of these loss functions is discussed in further section. The evaluation primarily used standard segmentation metrics (IoU and F1-score). However, ecological validity or real-world robustness under varying environmental conditions such as turbidity and light variations were not systematically assessed. Future work should involve comprehensive testing under such varied real-world conditions.

2.3 Segmentation workflow

The overall segmentation workflow, as illustrated in Figure 8, begins with data acquisition, where exploration recordings are sampled into selected video frames to serve as input data. The workflow followed a consistent sequence of operations: frame extraction and quality filtering, resizing and patching as described in preprocessing, color correction and normalization, augmentation (including classical and generative-based transformations), and training under the reported optimizer and early stopping regime. Predictions were reconstructed into full-image masks using overlap-averaging to smooth patch edges. In the preprocessing stage, image augmentation is applied to introduce environmental variations, enhancing the model's ability to generalize across diverse coral reef conditions. This is followed by image resizing and patching to accommodate the constraints of limited hardware resources during training and inference. The preprocessed images are then passed to the segmentation models, which utilize various encoder architectures and integrate skip connections, either through concatenation or additive strategies, to bridge the encoder and decoder layers. The details of each stage are presented in further sections.

2.4 Data acquisition and preprocessing

The research employed a diverse set of coral image data from two distinct categories, namely, *Dipsastraea* and *Porites*, from multiple different perspectives, as illustrated in Fig. 9. This strategy helps the model learn to recognize and segment coral species from various viewpoints, scales, and contexts, enhancing its generalizability and enabling it to perform well on a wider range of real-world scenarios. Basic preprocessing techniques such as color correction and contrast enhancement were applied. However, explicit handling of environmental challenges like water turbidity, glare, and occlusions was limited, suggesting that future studies should investigate more sophisticated preprocessing techniques tailored specifically for these conditions.

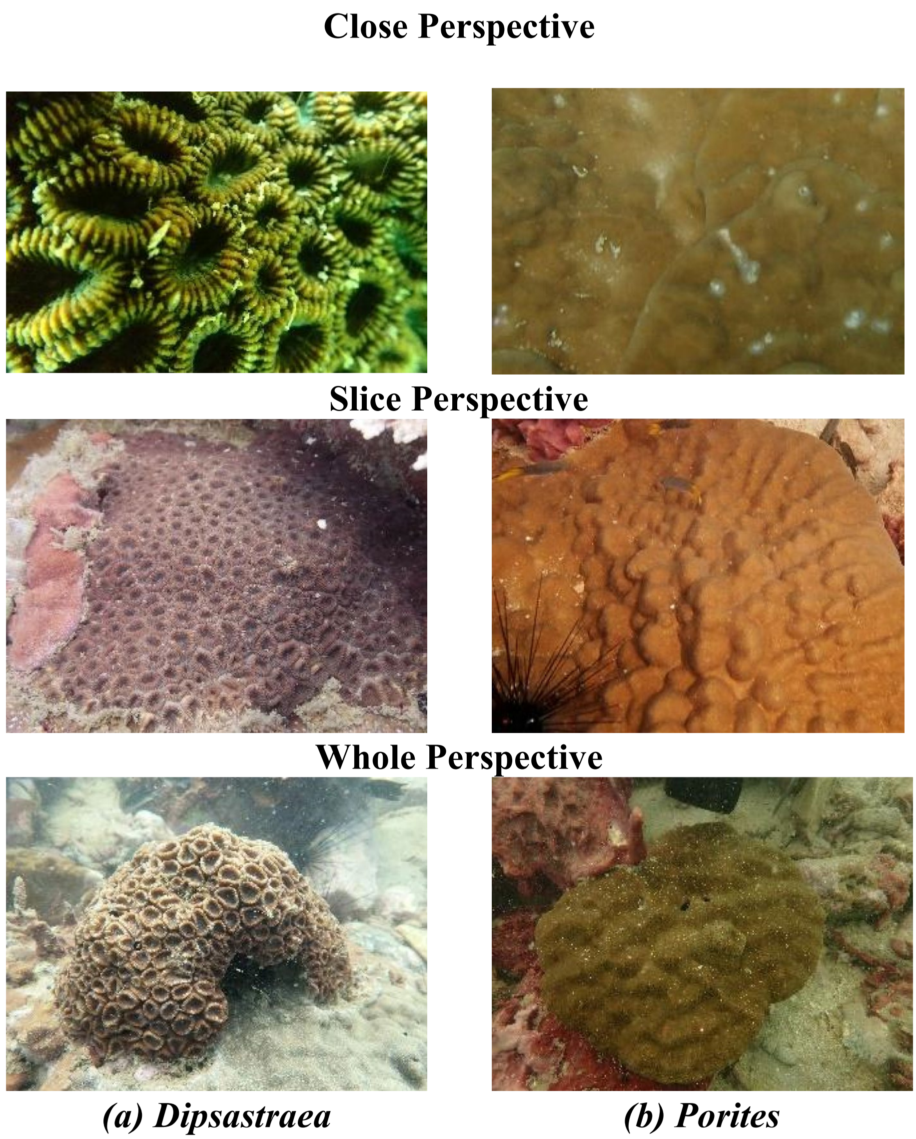

Fig. 8 Segmentation workflow

They were sourced from the uncharted marine environment within the Kuantan district of Pahang state, Malaysia. Figures 10 and 11 highlight the environmental and global views of the data collection. Each class was represented by a collection of approximately 50 images featuring varying pixel dimensions, predominantly at resolutions of approximately 4k and 1080p. The images were densely annotated via a web-based annotation tool, APEER. The class statistics and the frequency of the annotated classes within the image collection is depicted in Fig. 12.

Image augmentation was employed in this work to address the diverse conditions of coral reef sites, particularly variations in lighting and environmental factors influenced by sea depth and water quality, such as clarity, turbidity, and particulate density, evidently presented throughout the paper. Augmentation details and approach: In addition to classical geometric and photometric transformation, we augmented the dataset with generative-style and synthetic examples to increase environmental diversity. Practically, the augmentation pipeline combined: random rotation/flip and brightness adjustments;

Fig. 9 Different perspectives of discriminative features between **a***Dipsastraea* and **b***Porites*

GAN-style domain translation to simulate alternative lighting and turbidity conditions while preserving masks; and synthetic renderings from a small library of 3D coral models to supplement rare visual conditions. During training we mixed real and generated images to improve robustness to low contrast, turbidity and occlusions. Implementation scripts and trained augmentation models are provided in the code repository so reviewers can reproduce the exact augmentation used.

It should be noted that our dataset is geographically constrained to reef systems located within Kuantan district, Malaysia, including Pulau Ular, Karang Pelindung, Batu Serandu, Raja Muda, Beting Sepat, and Terumbu Kuning, which limits the generalizability of our findings to global coral ecosystems. Future research efforts should incorporate datasets from varied geographic regions and environmental conditions.

Given the complications of the datasets and the hardware setup, a strategic approach was adopted to alleviate the computational load during training, and the images underwent a twofold transformation: image resizing and patching. Image resizing was necessary to standardize the dimensions of all the images, regardless of their initial sizes, without damaging their quality for training. Moreover, image patching was proposed as a practical strategy to divide each image into smaller and manageable patches, thereby

Fig. 10 Environmental data collection locations

reducing the burden associated with processing an entire image in a single instance while increasing the number of training images.

The options for resizing and patching dimensions were left open-ended. Therefore, for the initial experiment, specific dimensions were chosen to find a balance between performance and resource availability. The goal was to resize the images to dimensions of 1536×1024 and 1024×512 and to patch them into patches of size 256×256. The selected resizing dimension also determines the number of images for training, as the smaller the resizing is, the lower the number of training images. Briefly, the former resizing would produce over 2000 images, whereas the latter would produce approximately 700 images. Deviating from these dimensions may result in increased computational requirements and a potential deterioration in data quality.

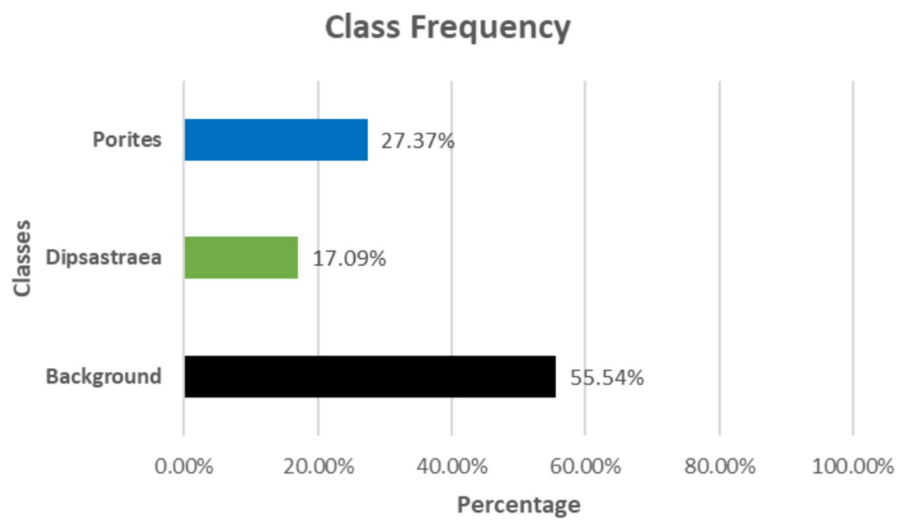

Fig. 11 Uncharted reef locations within Kuantan, Malaysia

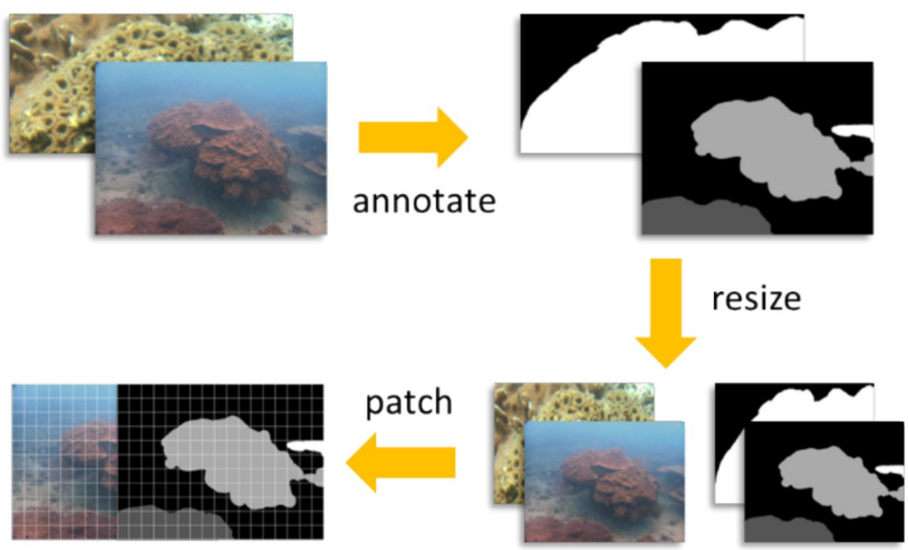

Fig. 12 Statistical summary of classes and frequency distribution of class pixel areas

2.5 Annotation protocols and quality control

To ensure consistent, transparent annotations we established a written guideline and a small QA workflow. Annotation proceeded in three stages: (1) calibration, where annotators reviewed and discussed a shared calibration set to align boundary and class rules; (2) independent annotation of the bulk dataset by trained annotators using the guideline; and (3) overlap quality-assurance, where a stratified subset of images was annotated independently by multiple experts to measure consistency. Inter-annotator agreement was assessed using standard measures (Cohen's Kappa for categorical labels; boundary F1 / Dice for mask overlap) and low-agreement cases were adjudicated by a senior marine biologist. Ambiguous regions were conservatively labeled as unknown rather than forcing a possibly incorrect class label. Annotations were created with the APEER web tool and exported as polygon masks (COCO-style JSON). The full annotation guideline and adjudication examples are included in the project repository (Data Availability).

2.6 Segmentation models

To segment and class the corals, U-Net-based networks were used in this study. This is because U-Net has shown significant performance when dealing with low training datasets and requires high-resolution segmentation. Moreover, the comparative study of the state-of-the-art segmentation models for corals has yielded remarkable results. The networks include U-Net, LinkNet, and U-Net with the ResNet-18 backbone for comparison. Model configurations and training regime**:** To support reproducibility we provide the configuration used for the experiments. U-Net baseline used a five-level encoder–decoder with filters doubling each down-step; skip connections were concatenated to preserve high-resolution features. Input tiles for training were derived from the resized images (see Data Preprocessing) and patched to the tile size used in experiments. For the ResNet-backed variant we employed an ImageNet-pretrained ResNet-18 encoder with a mirrored decoder. Training followed the hardware and optimizer choices already stated (TensorFlow/Keras on NVIDIA GeForce RTX 3060; Adam optimizer with the learning rate and batch-size reported in the Experimental Methodology). We used Dice and categorical cross-entropy components for segmentation loss and early stopping on validation IoU. Full model code, exact layer details and saved checkpoints are available in the repository for exact reproduction. The U-Net consists of a five-layer deep encoder and decoder, starting with 16 filters of the convolution layer and doubling for each layer at the encoding stage and vice versa at the decoding stage, where both are connected through skip connections to preserve and propagate high-resolution features from the input to the output as they are concatenated. LinkNet has a similar architecture to U-Net, but the inputs from the skip connections are added instead of being concatenated. Another suggested model is residual U-Net, which introduces increased complexity by incorporating 1×1 residual layers into each U-Net convolutional block. These models were trained with randomly initialized weights. Thus, another U-Net with a ResNet-18 backbone was included where the training starts with pretrained weights of ImageNet. Table 5 shows the complexity of all proposed models to be experimented including one models from the literature for comparison

3 Results and discussion

Figure 13 illustrates the relationship between the intersection over union (IoU) and losses with respect to loss types and size types. The figure shows that the models utilizing CCE and DC mostly have higher IoU values; however, they also suffer the highest loss among the other models. Another observation is that most trainings with 1536×1024 resized images have higher IoU scores and lower loss. In brief, the model trained with the hybrid loss function of CCE and FC, coupled with a training image size of 1536×1024, emerges as the best configuration for the available hardware. This

Table 5 Model complexity

Model	Trainable parameters	Minimum memory required (GB)	Training time (mins)
SUIM-Net	3,115,395	6.16	233.33
U-Net	1,941,139	1.13	42.13
LinkNet	1,745,299	1.01	39.71
Residual U-Net	8,109,891	3.20	68.57
U-Net (ResNet)	14,330,934	0.05	68.60

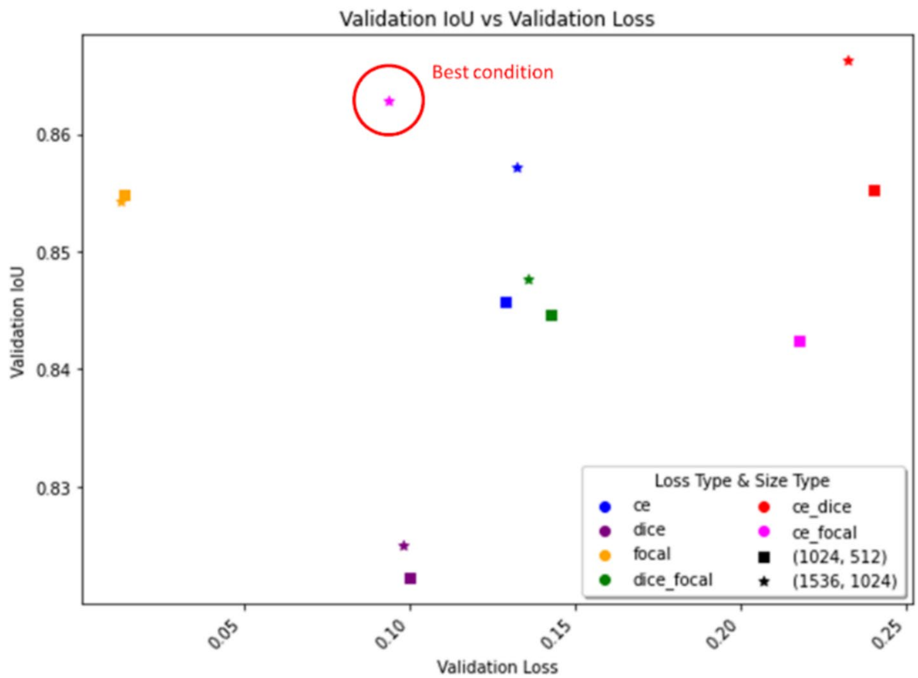

Fig. 13 IoU and losses of different resizing values

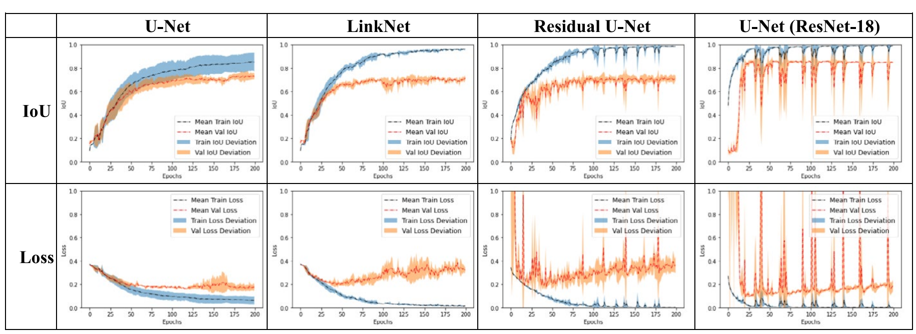

Fig. 14 Training graphs

configuration shows a harmonious balance between achieving a satisfactory IoU score and maintaining manageable loss levels.

As previously mentioned, three models were trained. Figure 14 presents the training graphs of the selected models over 200 epochs. The graphs of U-Net and LinkNet are quite similar, whereas those of U-Net, which has an alternative backbone, differ. Throughout training, the validation IoU and loss curves of the first three models closely follow the training curves, with some divergence starting around epoch 30. This divergence indicates a possible onset of overfitting, as the model's performance on the validation set starts to lag behind its performance on the training set. Another finding is that the IoU trends of residual U-Net and U-Net (ResNet) are slightly similar, possibly because of the implementation of residual layers within their architecture. However, U-Net (ResNet), which uses pretrained weights, remains stable throughout the entire epoch duration, whereas the residual U-Net experiences an increased lag toward the later epochs, similar to the other models.

Following the decision on the optimal configurations, Table 6 highlights the performance evaluation of the models in terms of the validation IoU, loss, and inference time on 1080p images. The IoU score serves as a crucial metric, demonstrating the effectiveness of the models in accurately delineating specific coral species within the images. Despite the lack of generalization ability of the models throughout the entire range of epochs, the U-Net model utilizing ResNet as its backbone with ImageNet pretrained weights significantly demonstrated even more promising results, with a 13% higher IoU score and 35% lower loss than the bare U-Net. However, it is important to acknowledge the trade-off between accuracy and inference speed, as the U-Net (ResNet) model exhibited a slightly longer mean inference time than the other architectures did because of the increased number of trainable parameters, approximately twice the size of the U-Net, existing from the additional residual layers included. The results also underscore the significant impact of leveraging deeper networks and using pretrained weights. As mentioned earlier, the saved models were determined on the basis of the best validation loss for early stopping, thereby making them independent from suboptimal models upon epoch completion.

For context, traditional methods for coral reef image analysis such as color histogram features with SVMs or superpixel segmentation coupled with Random Forest classifiers were considered. While such approaches offer lower computational demands, especially for CPU-only systems, they generally underperform deep models in complex underwater environments with turbidity and occlusion. This contrast highlights the practical advantage of deep learning in achieving ecologically relevant segmentation accuracy, even though classical methods remain useful as lightweight baselines in constrained settings.

The inference time was measured on an NVIDIA GeForce RTX 3060 GPU with 12 GB VRAM, averaging approximately 1.55 seconds per 1080p image. These computational demands highlight important considerations for real-time or resource-constrained deployment scenarios. For deployment on edge devices or AUVs/ROVs, lightweight strategies such as model pruning, quantization, and knowledge distillation can substantially reduce inference time. Platforms like NVIDIA Jetson modules present a viable balance between efficiency and accuracy for in-field use. In scenarios where onboard resources are limited, hybrid approaches combining onboard pre-filtering with batch shore-side processing can be adopted.

Table 7 provides a comprehensive breakdown of the IoU scores achieved by the models for each class, specifically the background (marine environment), Coral 1 (*Dipsastraea*), and Coral 2 (*Porites*) classes. Coral 1, in particular, poses a greater challenge for adaptation. The IoU scores associated with Coral 1 are consistently lower than those associated with Coral 2, with differences of approximately −15.63%, −17.06%, −16.45%, and −2.10% across all the models, even for U-Net (ResNet), indicating that accurately segmenting

Table 6 Quantitative performance of the models

Model	Overall IoU score	Overall F1-score	Loss	Inferencing time (secs)
SUIM-Net	0.665 ± 0.012	0.792 ± 0.012	0.203 ± 0.038	1.80
U-Net	0.733 ± 0.013	0.841 ± 0.010	0.291 ± 0.031	1.21
LinkNet	0.722 ± 0.019	0.833 ± 0.012	0.326 ± 0.023	1.33
Residual U-Net	0.719 ± 0.027	0.830 ± 0.019	0.368 ± 0.071	1.40
U-Net (ResNet)	0.856 ± 0.003	0.921 ± 0.002	0.188 ± 0.048	1.55

Table 7 IoU scores of the models for each class

Model	IoU (Background)	IoU (Coral 1—Dipsastraea)	IoU (Coral 2—Porites)
SUIM-Net	0.713 ± 0.005	0.603 ± 0.014	0.681 ± 0.025
U-Net	0.749 ± 0.008	0.648 ± 0.025	0.768 ± 0.014
LinkNet	0.738 ± 0.020	0.632 ± 0.009	0.762 ± 0.030
Residual U-Net	0.734 ± 0.021	0.635 ± 0.049	0.760 ± 0.016
U-Net (ResNet)	0.850 ± 0.003	0.838 ± 0.005	0.856 ± 0.012

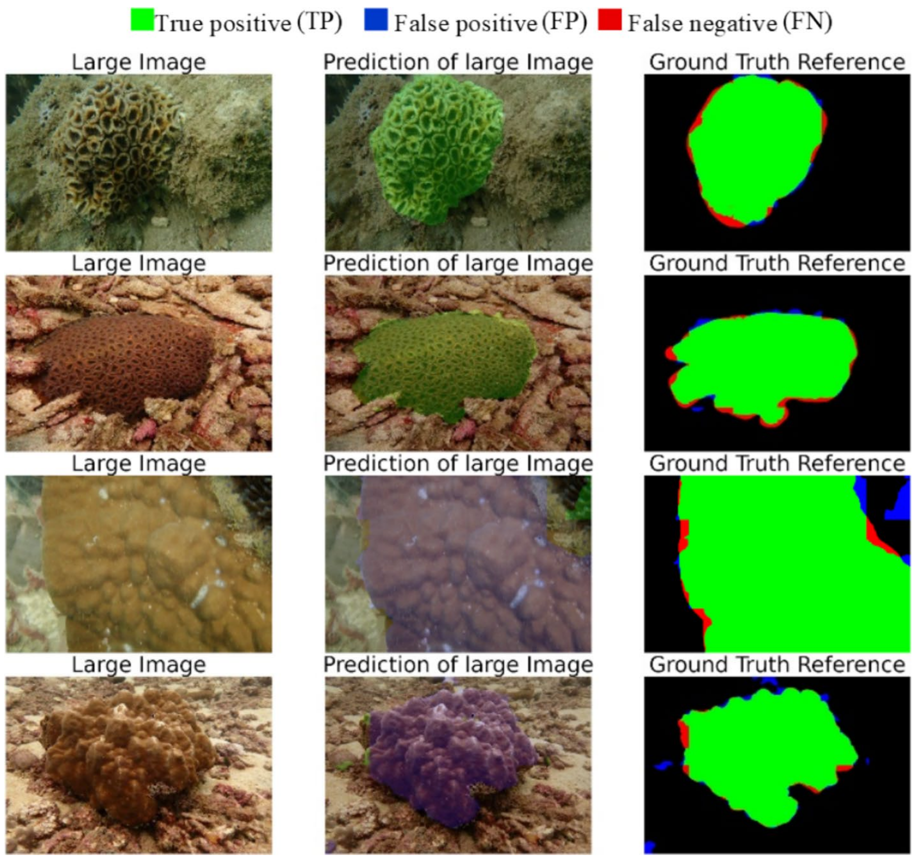

Fig. 15 Coral segmentation. For mask predictions, the green highlights indicate *Dipsastraea*, and the blue highlights indicate *Porites*

this class is more difficult. This offers insight from Fig. 11 into requiring hard data mining on Coral 1 with more examples of discriminative features for training compared with the other classes.

A bias analysis revealed consistent difficulty in segmenting Dipsastraea compared to Porites, likely reflecting data scarcity and morphological complexity. This indicates a need for targeted augmentation and collection of more examples for underrepresented coral species. Continuous site-level evaluation is recommended to detect domain-specific biases before deployment.

Like the preprocessing procedures adopted during the training phase, the prediction of coral images employs a patch-based segmentation approach. Figure 15 depicts a visual representation of the coral images that have undergone segmentation via U-Net (ResNet). In the figure, the predictions of the model were compared against the ground truth labels in the last column. The true class segmentation is represented in green, indicating accurate alignment between the predictions and actual classes, regardless of the type. In contrast, the blue and red colors mark the instances of segmentation errors, representing oversegmentation and undersegmentation, respectively. Another evaluation metric that was mentioned earlier is the F1 score. The study revealed that the F1 score achieved by U-Net (ResNet) exceeded 0.9, which can be considered high. In other words, it also suggests a well-balanced occurrence of both segmentation errors across the pixels within an image, as shown in Fig. 15. Another observation is that *Dipsastraea* is commonly and inaccurately segmented into small fragments near *Porites*, as represented by its performance in Table 5. In contrast, instances of *Porites* being falsely identified near *Dipsastraea* are relatively rare. Compared with the reviewed literature, *Porites* corals have been segmented in several studies, whereas *Dipsastraea* corals have rarely been included in any of them. The highest mean accuracy for *Porites* segmentation, with an IoU score of approximately 0.89, was achieved via ResNet-152 by King et al. [23]. However, the model has three times as many parameters as the largest model tested in this study.

In addition to quantitative metrics, qualitative overlays were produced to highlight both successful and challenging cases. These visualizations included cases under varying turbidity and illumination. Error maps showing over- and under-segmentation were also provided to illustrate model tendencies.

Among the limitations of the current work is that it is restricted to only two coral types, requiring more diverse exploration sites and additional annotated coral species to enhance generalizability. Another limitation is that the proposed model is suitable primarily for on-site processing rather than real-time application. Nonetheless, the segmentation performance of the model remains satisfactory and aligns well with the objectives of this study.

4 Conclusion

This review has highlighted the significant advancements in coral reef monitoring through computer vision and AI, demonstrating their potential to revolutionize conservation efforts. By addressing the limitations of traditional methods, these technologies offer a promising avenue for more efficient monitoring by automating specifies identification and coverage estimation. The insights gained from this review provide a foundation for future research and underscore the need for continued innovation and collaboration in this field. An acknowledged limitation of this work is the absence of cross-region validation. The trained models' performance was evaluated solely on geographically localized Malaysian coral reef imagery, and their applicability across different coral ecosystems globally has yet to be verified. Future studies should include cross-region validation to enhance confidence in model applicability.

Moreover, this work has the potential to significantly reduce manual labor involved in tagging large volumes of coral reef types across vast ocean areas. It also enhances human safety, particularly in hazardous sites where only exploration robots can be immediately deployed. Currently, this work focuses on enabling on-site processing of recorded footage, where the system can automatically detect and tag coral species present at specific exploration sites, eliminating the need for humans to manually review lengthy video recordings or engage in time-consuming site explorations, thus, reduce the exploration cost.

Future research should explore the integration of multi-source data, including environmental factors and human activities, to enhance the accuracy of coral health predictions instead of only their types. Beyond technical metrics, ecological alignment requires translating segmentation outputs into indicators such as live coral cover, species distribution, and mortality ratios. Incorporating these outputs into conservation workflows will strengthen the ecological utility of automated monitoring systems. Collaborative efforts between marine biologists, computer scientists, and environmental engineers could lead to more comprehensive monitoring systems. Additionally, developing more robust and diverse datasets, particularly from underrepresented regions, will be crucial in advancing AI-driven coral reef monitoring. While the current research primarily focuses on the technical development and segmentation accuracy of deep learning methods, the ecological implications of automated monitoring and integration into existing coral conservation workflows have not been addressed. Future research should explore these dimensions to ensure the practical utility and ecological appropriateness of automated coral monitoring solutions.

Acknowledgements
This research was supported by Fundamental Research Grant Scheme (FRGS), Malaysia FRGS22-265-0874, Elucidating The Uncharted Nearshore Reefs Distribution Using Multibeam Echosounder And 3d Coral Photogrammetry In Kuantan Coastal Waters. The authors would like to express their gratitude to the Chasing Karang Team, Malaysia for providing coral data for this research.

Author contributions
Author Contributions Statement Hafizi Malik, Muhammad Faiz Mohd Hanapiah, Siti Fauziah Toha, Mohd Zaini Mustapa, Aiman Hisyam Azmi, Hazrul Amirul Johari, Ahmad Syahrin Idris, Azhar Mohd Ibrahim, Philippe De Wilde, and Amir I A Alqedra contributed to this work as follows: Hafizi Malik: Manuscript preparation, writing, drafting, checking, and submission, Coral image and data collection from the uncharted location, Image identification, ML image segmentation, prediction, and optimization, ML data analysis and optimization, Image verification and data analysis, Dataset creation and annotation. Muhammad Faiz Mohd Hanapiah: Coral image and data collection from the uncharted location, ML image segmentation, prediction, and optimization, ML data analysis and optimization, Experimental evaluations and comparisons. Siti Fauziah Toha: Manuscript writing, Manuscript preparation, ML image segmentation, prediction, and optimization, ML data analysis and optimization, Image verification and data analysis, Manuscript drafting, checking, and submission, Dataset creation and annotation. Mohd Zaini Mustapa: Coral image and data collection from the uncharted location, ML image segmentation, prediction, and optimization, ML data analysis and optimization, Experimental evaluations and comparisons. Aiman Hisyam Azmi: Coral image and data collection from the uncharted location, ML image segmentation, prediction, and optimization, ML data analysis and optimization, Experimental evaluations and comparisons. Hazrul Amirul Johari: Coral image and data collection from the uncharted location, ML image segmentation, prediction, and optimization, ML data analysis and optimization, Experimental evaluations and comparisons. Ahmad Syahrin Idris: Image identification, ML data analysis and optimization, Image verification and data analysis, Experimental evaluations and comparisons. Azhar Mohd Ibrahim: Image identification, Image verification and data analysis, Manuscript preparation, Experimental evaluations and comparisons. Philippe De Wilde: ML image segmentation, prediction, and optimization, ML data analysis and optimization, Image verification and data analysis, Manuscript drafting, checking, and submission. Amir I A Alqedra: Manuscript preparation, writing, drafting, checking, and submission, Image verification and data analysis. All authors reviewed and approved the final manuscript. Corresponding Author: Siti Fauziah Toha, tsfauziah@iium.edu.my.

Funding
No funding.

Data availability
The datasets generated and/or analysed during the current study are available in the Google Drive repository (https://drive.google.com/drive/folders/1yyc3t1LTj7MMMoVYGKrWreRv2UzPEaMz?usp=drive_link). The code used for analysis is available in the GitHub repository (https://github.com/Fyzie/MultiClass-Coral-Segmentation). Additional instructions for accessing and using both the dataset and the source code will be provided upon request.

Declarations

Ethics approval and consent to participate
No ethical approval was required for this study as it did not involve human participants, experiments on animals, or collection of specimens. The coral reef imagery used in this study was collected in accordance with local regulations and with appropriate permits from relevant authorities. Not applicable as this study did not involve human participants.

Consent for publication
Not applicable as this study did not involve human participants or identifiable individual data.

Competing interests
The authors declare no competing interests.

Received: 26 November 2024 / Accepted: 26 September 2025
Published online: 14 January 2026

References

1. Hoegh-Guldberg O, Poloczanska ES, Skirving W, Dove S. Coral reef ecosystems under climate change and ocean acidification. Front Mar Sci. 2017;4(158):1–20. https://doi.org/10.3389/fmars.2017.00158.
2. Obura DO, Aeby G, Amornthammarong N, Appeltans W, Bax N, Bishop J, et al. Coral reef monitoring, reef assessment technologies, and ecosystem-based management. Front Mar Sci. 2019;6(580):1–21. https://doi.org/10.3389/fmars.2019.00580.
3. Hedley JD, Roelfsema CM, Chollett I, Harborne AR, Heron SF, Weeks SJ, et al. Remote sensing of coral reefs for monitoring and management: a review. Remote Sens. 2016. https://doi.org/10.3390/rs8020118.
4. Raber GT, Schill SR. Reef rover: a low-cost small autonomous unmanned surface vehicle (usv) for mapping and monitoring coral reefs. Drones. 2019;3(2):1–22. https://doi.org/10.3390/drones3020038.
5. Zhou J, Zhou N, Che Y, Gao J, Zhao L, Huang H, et al. Design and development of an autonomous underwater helicopter for ecological observation of Coral reefs. Sensors. 2022. https://doi.org/10.3390/s22051770.
6. Pierce J, Butler MJ, Rzhanov Y, Lowell K, Dijkstra JA. Classifying 3-D models of coral reefs using structure-from-motion and multi-view semantic segmentation. Front Mar Sci. 2021;8(706674):1–13. https://doi.org/10.3389/fmars.2021.706674.
7. Mizuno K, Terayama K, Hagino S, Tabeta S, Sakamoto S, Ogawa T, et al. An efficient coral survey method based on a large-scale 3-D structure model obtained by Speedy Sea Scanner and U-Net segmentation. Sci Rep. 2020. https://doi.org/10.1038/s41598-020-69400-5.
8. Song H, Mehdi SR, Zhang Y, Shentu Y, Wan Q, Wang W, et al. Development of coral investigation system based on semantic segmentation of single-channel images. Sensors. 2021;21(5):1–21. https://doi.org/10.3390/s21051848.
9. Wang J, He X, Shao F, Lu G, Hu R, Jiang Q. Semantic segmentation method of underwater images based on encoderdecoder architecture. PLoS ONE. 2022;17(8):1–19. https://doi.org/10.1371/journal.pone.0272666.
10. Runyan H, Petrovic V, Edwards CB, Pedersen N, Alcantar E, Kuester F, et al. Automated 2D, 2.5D, and 3D segmentation of coral reef pointclouds and orthoprojections. Front Robotics AI. 2022;9(884317):1–8. https://doi.org/10.3389/frobt.2022.884317.
11. Schürholz D, Chennu A. Digitizing the coral reef: machine learning of underwater spectral images enables dense taxonomic mapping of benthic habitats. Methods Ecol Evol. 2022;14(2):596–613. https://doi.org/10.1111/2041-210X.14029.
12. Thomas T, Maurya P, Manikandan B, Dessai NBF. Estimation of coral reef area through 2d images: deep learning way using unet. Soc Sci Res Netw (SSRN). 2022. https://doi.org/10.2139/ssrn.4195235.
13. Song H, Raza Mehdi S, Zhang Y, Shentu Y, Wan Q, Wang W, et al. Development of coral investigation system based on semantic segmentation of single-channel images. Sensors. 2021;21(5):1848. https://doi.org/10.3390/s21051848.
14. Wang L, Li R, Zhang C, Fang S, Duan C, Meng X, et al. Unetformer: a unet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. ISPRS J Photogramm Remote Sens. 2022;190:196–214. https://doi.org/10.1016/j.isprsjprs.2022.06.008.
15. Mahmoud AS, El-Morshedy RM, Metwalli MR, Mostafa MS. Sandstorm detection using attention bi-LSTM unet. J Indian Soc Remote Sens. 2024. https://doi.org/10.1007/s12524-024-02042-6.
16. Arendt M, Rückert J, Brüngel R, Brumann C, Friedrich CM. The effects of color enhancement and IoU optimization on object detection and segmentation of coral reef structures. CLEF (Working Notes). 2020.
17. Soukup L. Automatic Coral Reef Annotation , Localization and Pixelwise Parsing Using Mask R-CNN. Conference and Labs of the Evaluation Forum. 2021.
18. Steffens A, Campello A, Ravenscroft J, Clark A, Hagras H. Deep segmentation: using deep convolutional networks for coral reef pixelwise parsing. CEUR Workshop Proc. 2019;2380:9–12.
19. Wright J, Palosanu I, Clift L, García A, De Herrera S, Chamberlain J. Pixelwise annotation of coral reef substrates. Conference and Labs Ofthe Evaluation Forum. 2019.
20. Islam J, Edge C, Xiao Y, Luo P, Mehtaz M, Morse C, Enan SS, Sattar J. Semantic Segmentation of Underwater Imagery : Dataset and Benchmark. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020; 1769–1776.
21. Dzakmic S, Yildiz A, Almisreb A, Jamil N. Evaluation of ResNet network for semantic segmentation of coral reefs. Int J Emerg Trends Eng Res. 2020;8(11):54–61. https://doi.org/10.30534/ijeter/2020/0881.12020.
22. King A, Bhandarkar SM, Hopkinson BM. A comparison of deep learning methods for semantic segmentation of coral reef survey images. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2018; 1507–1515. https://doi.org/10.1109/CVPRW.2018.00188.
23. King A, Bhandarkar SM, Hopkinson BM. Deep Learning for Semantic Segmentation of Coral Reef Images Using Multi-View Information. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 1–10. 2022.
24. Sui YW, Ming KX, Meghjani M, Raghavan N, Jegourel C, Kang K. An Automated Data Processing Pipeline for Coral Reef Monitoring. OCEANS 2022, Hampton Roads, Hampton Roads, VA, USA, 2022; 1–6. https://doi.org/10.1109/OCEANS47191.2022.9977204.
25. Nan F, Chen H. Coral Detection with Machine Learning. 2023 5th International Conference on Computer Communication and the internet (ICCCI), 2023; 207–209. https://doi.org/10.1109/ICCCI59363.2023.10210179
26. Zhang H, Gruen A, Li M. Deep learning for semantic segmentation of coral images in underwater photogrammetry. ISPRS Ann Photogramm Remote Sens Spat Inf Sci. 2022;5(2):343–50. https://doi.org/10.5194/isprs-annals-V-2-2022-343-2022.
27. Zhong J, Li M, Zhang H, Qin J. Combining Photogrammetric Computer Vision and Semantic Segmentation for Fine-grained Understanding of Coral Reef Growth under Climate Change. Proceedings - 2023 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops, WACVW 2023, 2023; 186–195. https://doi.org/10.1109/WACVW58289.2023.00024.

28. Giles AB, Ren K, Davies JE, Abrego D, Kelaher B. Combining Drones and Deep Learning to Automate Coral Reef Assessment with RGB Imagery. Remote Sens. 2023;15(9):1–16. https://doi.org/10.3390/rs15092238.
29. Lütjens M, Sternberg H. Deep learning based detection, segmentation and counting of benthic megafauna in unconstrained underwater environments. IFAC-PapersOnLine. 2021;54(16):76–82. https://doi.org/10.1016/j.ifacol.2021.10.076.
30. Raddatz J, Liebetrau V, Rüggeberg A, Foubert A, Flögel S, Nürnberg D, et al. Living on the edge: environmental variability of a shallow late Holocene cold-water coral mound. Coral Reefs. 2022;41(4):1255–71. https://doi.org/10.1007/s00338-022-02249-4.
31. Yu X, Ouyang B, Principe CJ, Farrington S, Reed J, Li Y. Weakly supervised learning of point-level annotation for coral image segmentation. OCEANS 2019 MTS/IEEE SEATTLE, Seattle, WA, USA, 2019; 1–7. https://doi.org/10.23919/OCEANS40490.2019.8962759.
32. Yu X, Ma Y, Farrington S, Reed J, Ouyang B, Principe JC. Fast segmentation for large and sparsely labeled coral images. Int Joint Conf Neural Netw. 2019. https://doi.org/10.1109/IJCNN.2019.8852014.
33. Pierce JP, Rzhanov Y, Lowell K, Dijkstra JA. Reducing Annotation Times: Semantic Segmentation of Coral Reef Survey Images. 2020 Global Oceans 2020: Singapore - U.S. Gulf Coast. 2020, October 5; https://doi.org/10.1109/IEEECONF38699.2020.9389163.
34. Alonso I, Yuval M, Eyal G, Treibitz T, Murillo AC. Coralseg: learning coral segmentation from sparse annotations. J Field Robotics. 2019;36(8):1456–77. https://doi.org/10.1002/rob.21915.
35. Alonso I, Cambra A, Muñoz A, Treibitz T, Murillo AC. Coral-Segmentation: Training Dense Labeling Models with Sparse Ground Truth. IEEE International Conference on Computer Vision Workshops, 2017; 2874–2882. https://doi.org/10.1109/ICCVW.2017.339.

Publisher's Note