



ARTICLE

A Self-Supervised Hybrid Similarity Framework for Underwater Coral Species Classification

Yu-Shiuan Tsai*, Zhen-Rong Wu and Jian-Zhi Liu

Department of Computer Science and Engineering, National Taiwan Ocean University, Keelung City, 202, Taiwan

*Corresponding Author: Yu-Shiuan Tsai. Email: ystsai@mail.ntou.edu.tw

Received: 10 April 2025; Accepted: 04 June 2025; Published: 03 July 2025

ABSTRACT: Few-shot learning has emerged as a crucial technique for coral species classification, addressing the challenge of limited labeled data in underwater environments. This study introduces an optimized few-shot learning model that enhances classification accuracy while minimizing reliance on extensive data collection. The proposed model integrates a hybrid similarity measure combining Euclidean distance and cosine similarity, effectively capturing both feature magnitude and directional relationships. This approach achieves a notable accuracy of 71.8% under a 5-way 5-shot evaluation, outperforming state-of-the-art models such as Prototypical Networks, FEAT, and ESPT by up to 10%. Notably, the model demonstrates high precision in classifying Siderastreidae (87.52%) and Fungiidae (88.95%), underscoring its effectiveness in distinguishing subtle morphological differences. To further enhance performance, we incorporate a self-supervised learning mechanism based on contrastive learning, enabling the model to extract robust representations by leveraging local structural patterns in corals. This enhancement significantly improves classification accuracy, particularly for species with high intra-class variation, leading to an overall accuracy of 76.52% under a 5-way 10-shot evaluation. Additionally, the model exploits the repetitive structures inherent in corals, introducing a local feature aggregation strategy that refines classification through spatial information integration. Beyond its technical contributions, this study presents a scalable and efficient approach for automated coral reef monitoring, reducing annotation costs while maintaining high classification accuracy. By improving few-shot learning performance in underwater environments, our model enhances monitoring accuracy by up to 15% compared to traditional methods, offering a practical solution for large-scale coral conservation efforts.

KEYWORDS: Few-shot learning; self-supervised learning; contrastive representation learning; hybrid similarity measures; local feature aggregation; voting-based classification; marine species recognition; underwater computer vision

1 Introduction

Coral reefs are among the most biologically diverse and structurally complex ecosystems, playing a crucial role in maintaining marine biodiversity, supporting human livelihoods, and regulating global ecological balance. Dating back to before the Cambrian period [1], coral polyps have coexisted with symbiotic algae for over 500 million years, forming extensive reef structures through calcium carbonate deposition. Modern taxonomy classifies corals under the phylum Cnidaria, primarily within the Anthozoa class, which is further divided into Hexacorallia and Octocorallia [2]. Taiwan's waters, due to their unique ocean topography and currents, host one of the richest coral ecosystems, encompassing a vast array of reef-building and non-reef-building species. Fringing reefs dominate its coastline, while barrier reefs and atolls are present in other regions, supporting diverse marine habitats.



Despite their ecological importance, coral reefs are increasingly threatened by climate change, pollution, and human activities. Masson-Delmotte et al. [3] highlighted that rising ocean temperatures, marine heatwaves, and ocean acidification severely impact coral health, leading to widespread bleaching events. If global ocean temperatures rise by 2°C, coral reefs could face near-total extinction by 2050. Additional stressors, including overfishing, coastal development, and chemical pollution, further accelerate reef degradation [4]. These threats emphasize the urgency of large-scale, automated coral reef monitoring and classification to support conservation efforts.

Deep learning has emerged as a powerful tool for coral classification, significantly reducing reliance on labor-intensive manual annotation. Early methods, such as those by Beijbom et al. [5], utilized texture and color descriptors to estimate coral coverage. More recent approaches leverage convolutional neural networks (CNNs) and deep feature representations for improved classification accuracy. King et al. [6] introduced TwinNet, which utilizes multi-view stereoscopic data to enhance coral segmentation. Gapper et al. [7] applied remote sensing and support vector machines (SVMs) to analyze 14 years of satellite data, demonstrating the potential of large-scale automated monitoring. Raine et al. [8] further improved annotation efficiency by integrating foundation models like DINOv2 with a human-in-the-loop labeling strategy.

However, despite these advancements, a fundamental challenge remains: the high cost of labeled data in deep learning models. Supervised deep learning techniques typically require large datasets with extensive annotations, which are difficult and time-consuming to acquire in underwater environments. For example, Raphael et al. [9] trained ResNet-50 with over 5000 labeled coral images, while Reshma et al. [10] required 115,000 data points for CNN-based coral classification. The demand for such massive datasets limits the scalability of traditional deep learning models, particularly in remote coral reef ecosystems where data collection is restricted.

Few-Shot Learning (FSL) has emerged as a promising solution to reduce dependency on large labeled datasets, allowing models to generalize from only a handful of annotated examples. FSL techniques, such as metric learning and meta-learning, have demonstrated success in adapting models to new classes with minimal supervision. Li et al. [11] reviewed deep metric learning approaches for FSL, categorizing them into feature embedding, prototype learning, and similarity-based classification. Zhang et al. [12] further explored meta-transfer learning, where knowledge is transferred across domains to improve generalization in novel categories. These studies highlight the potential of FSL for coral classification, particularly in scenarios where acquiring extensive labeled data is impractical. To further substantiate the effectiveness and contemporary relevance of our proposed model, we have integrated insights from several recent studies published after 2019 that employ advanced few-shot learning algorithms. For example, Liu et al. [13] introduced a hybrid representation framework that balances class-specific and shared features through residual-based refinement, highlighting the importance of enriched feature representations—an approach that aligns closely with our use of combined Euclidean and cosine similarity. Similarly, Sim and Kim [14] developed a cross-attention-based dual similarity network that captures both channel-wise and spatial correlations, providing a strong comparative reference for our emphasis on local feature modeling in coral imagery. In addition, recent methods such as ESPT [15] leverage self-supervised learning to enhance feature quality under limited supervision, which directly informed our integration of contrastive learning into our model. Collectively, these studies affirm the technical soundness and competitiveness of our framework within the evolving landscape of few-shot learning research.

Recent advancements have pushed the boundaries of FSL applications in environmental monitoring. Contini et al. (2025) [16] introduced a multi-scale knowledge distillation model that integrates fine-scale underwater imagery with aerial data, demonstrating how multi-resolution data fusion can enhance FSL

performance in large-scale ecological assessments. Additionally, Shao et al. (2024) [17] developed a multi-label classification framework to automate the detection of coral health conditions, improving FSL models' adaptability to variable underwater environments. Blondin et al. (2024) [18] further contributed to this field by proposing a hierarchical classification system, which aligns coral classification with ecological taxonomies, allowing FSL models to leverage hierarchical dependencies for improved recognition accuracy.

Despite these advancements, several key challenges persist in applying FSL to coral classification. First, underwater image variability, including lighting distortions, turbidity, and depth-induced color shifts, complicates feature extraction and domain adaptation [19]. The quality and clarity of underwater imagery are further affected by environmental factors such as water current speed, particle density, and algae distribution, particularly in deep-sea regions where traditional optical imaging becomes unreliable. To address these limitations, Sture et al. [20] explored synthetic aperture sonar (SAS) imaging techniques to enhance coral detection in low-visibility and deep-water environments. Their findings suggest that non-optical imaging modalities, such as SAS, could complement deep learning models by providing alternative feature representations that are less affected by water column distortions. However, integrating these diverse sensing modalities into FSL frameworks remains a challenge, as sonar-based imagery often differs significantly from standard RGB image datasets used in supervised learning. Second, morphological diversity within coral species presents a challenge, as intra-species variability leads to shape variations that traditional FSL models may struggle to distinguish [2]. Third, cross-domain generalization remains an open problem, as FSL models trained on one dataset often fail to adapt to different geographic regions due to domain shifts in environmental conditions and imaging techniques. Zhou et al. [21] proposed the Meta-Collaborative Comparison Network (MeCo-Net), which integrates both global and local representations through a collaborative comparison mechanism. Their model is designed to enhance cross-domain generalization in few-shot scenarios and demonstrates competitive performance across benchmark datasets. The architecture addresses the challenge of representation collapse by incorporating a similarity-guided collaborative fusion module, which aligns well with our motivation to combine complementary distance metrics. By leveraging both global semantics and localized variations, their method exemplifies the trend toward more robust, hybrid strategies in modern few-shot learning models—further validating the direction of our proposed hybrid similarity framework, while Suchendra et al. [22] explored 3D coral reconstructions to enhance model robustness through structure-from-motion (SfM) techniques.

To address these challenges, we propose the Optimized Coral Evaluation and Assessment Network (OCEAN), a novel few-shot learning framework for underwater coral classification. Our approach incorporates several key innovations. First, we exploit the repetitive morphological structures of corals to enhance feature extraction, enabling FSL models to better leverage local patterns for classification. Second, we integrate self-supervised contrastive learning to generate pseudo-labels, improving feature representations without requiring additional human annotations. Third, we implement a hybrid similarity measure combining Euclidean and cosine similarity, which refines prototype-based classification and enhances robustness to intra-species variability. Lastly, we introduce a voting-based evaluation mechanism, reducing misclassification errors by aggregating local predictions across coral images.

Experimental results demonstrate that OCEAN significantly outperforms existing FSL models in underwater coral classification tasks. By leveraging self-supervised learning and improved similarity measures, our model achieves higher accuracy with fewer labeled samples, making it a practical solution for large-scale coral monitoring. Furthermore, we introduce a benchmark dataset of ten coral species, providing a standardized evaluation platform for future research in FSL-based coral classification.

2 Related Work

Our research primarily addresses the few-shot classification problem, with additional relevance to self-supervised learning. This section reviews key developments in these areas and discusses their significance to our approach.

Few-shot learning (FSL) aims to enable machines to generalize from a limited number of training samples, mimicking the cognitive flexibility of humans when encountering new concepts with minimal prior exposure. One of the foundational models in FSL is Prototypical Networks (ProtoNets), introduced by Snell et al. in 2018 [23]. ProtoNets formalized the concept of a prototype, representing the central tendency of each class in the feature space. The model employs a meta-learning framework, where training is conducted in an episodic manner. Each episode consists of a Support Set (S) and a Query Set (Q), where the Support Set is used to compute class prototypes, and the Query Set is used to evaluate the model's classification performance.

During prototype computation, the model determines the prototype of each class by averaging all sample features within that class, as expressed in Eq. (1):

$$\mathcal{P}_n = \frac{1}{|\mathcal{S}_n|} \sum_{x_i \in \mathcal{S}_n} f_\theta(x_i) \quad (1)$$

where \mathcal{P}_n denotes the prototype of the n -th class, \mathcal{S}_n represents all sample features belonging to that class, x_i is the feature representation of the i -th sample, and f_θ is the feature extraction function of the model.

During classification, query samples are mapped into the same feature space using f_θ , and their similarity to each class prototype is computed. The class with the highest similarity score is selected as the predicted label. The similarity measurement between a query sample and a class prototype is formulated as Eq. (2):

$$sim_score = d(f_\theta(x), \mathcal{P}_k) \quad (2)$$

where d represents the function for computing similarity between the query sample x and the prototype \mathcal{P}_k . ProtoNets employ Euclidean distance as the similarity measure, classifying each query sample based on its closest class prototype in the feature space.

Prototypical Networks (ProtoNets) effectively integrate comprehensive feature representations from a limited number of samples, reducing over-reliance on individual instances and making them particularly well-suited for few-shot learning scenarios. However, despite their effectiveness, ProtoNets still leave room for improvement in both feature extraction and prototype evaluation. To address these limitations, Ye et al. (2020) introduced FEAT [24], a few-shot learning framework that extends the prototype-based approach. A key contribution of their study was recognizing that different classification tasks require task-specific feature representations, rather than relying on a fixed, generic feature space. To achieve this, FEAT employs a set-to-set adaptation strategy, where sample features are adjusted dynamically to optimize classification performance. By incorporating a Transformer-based feature refinement mechanism [25], FEAT learns to enhance feature representations for unseen classes, improving task adaptability. Their experimental results demonstrated that FEAT consistently outperformed conventional few-shot learning models across various datasets and settings, highlighting the effectiveness of prototype-based approaches with adaptive feature extraction.

Building upon these advancements, Sim and Kim (2024) proposed the Cross-Attention based Dual-Similarity Network (DSN) [14], which introduces a dual-similarity mechanism combining channel-similarity and map-similarity to improve few-shot learning. Their use of cross-attention mechanisms aligns with

the broader trend of employing Transformer-style architectures for image recognition. While DSN has demonstrated strong performance in N-way K-shot and 1-shot classification settings, its generalization capabilities in complex ecological domains remain unexplored. The Batch Enhanced Contrastive Few-Shot Learning (BECLR) model [26] integrates a Dynamic Clustered mEmory (DyCE) module and an Optimal Transport-based distribution Alignment (OpTA) mechanism to enhance representation quality and inference robustness. BECLR achieves strong performance in low-shot settings by refining both pretraining and downstream classification stages through self-supervised strategies. In contrast, our proposed model, OCEAN, leverages self-supervised contrastive learning and repetitive coral structures, making it particularly effective for coral classification tasks. By incorporating local feature aggregation techniques, OCEAN enhances model robustness in marine environments, where species exhibit high intra-class variation and morphological diversity.

Another notable contribution to few-shot learning is Liu et al.'s hybrid representation approach [13], which balances specific and shared feature representations. Their method introduces residual feature learning, allowing for finer adjustments in the feature extraction process. While Liu's model focuses on refining the base learner for classification, OCEAN differentiates itself by integrating a voting-based evaluation mechanism, which enhances classification reliability for highly diverse coral species. Moreover, OCEAN's application in ecological conservation offers practical benefits, significantly reducing the need for extensive coral image collection and annotation, thus making large-scale marine biodiversity monitoring more feasible.

Self-supervised learning (SSL) is a powerful paradigm that enables neural networks to learn meaningful feature representations by generating pseudo-labels without requiring explicit human annotations. The core principle of SSL is to design pretext tasks that automatically extract supervisory signals from the data, allowing the model to develop discriminative feature representations that enhance performance in subsequent downstream tasks. Various SSL strategies have been explored, with one of the earliest approaches involving prediction-based self-supervision. These methods include tasks such as predicting the relative positions of image patches [27], solving shuffled jigsaw puzzles [28], or reconstructing occluded or missing image parts to simulate artifact restoration [29]. However, several studies have highlighted a major drawback of these approaches—models tend to exploit low-level statistical correlations or shortcut learning, rather than developing robust semantic representations. This limitation has led to a shift toward contrastive learning, which has demonstrated superior performance in self-supervised learning tasks.

One of the most influential contrastive learning frameworks is Momentum Contrast (MoCo), introduced by He et al. in 2020 [30]. MoCo incorporates a dynamic dictionary queue with a momentum update mechanism, allowing contrastive learning to leverage a larger and more diverse set of negative samples. Unlike standard contrastive learning approaches, where negative samples are drawn from the same batch, MoCo maintains a queue of previously encountered samples and continuously updates it during training. Specifically, MoCo initializes the dictionary queue randomly, and at each training step, it removes the oldest batch of key values while adding the current batch into the queue. This approach ensures that the dictionary captures a broad data distribution while preserving consistency with the latest representations. By leveraging this momentum-based design, MoCo effectively mitigates the challenge of limited negative samples, thereby improving the model's ability to learn generalizable feature representations.

In addition to MoCo, the SimCLR framework, developed by the Google Brain team [31], provides a simpler yet highly effective approach to contrastive learning. Unlike MoCo, SimCLR does not rely on momentum encoders or memory banks. Instead, it achieves self-supervised learning through a carefully designed data augmentation pipeline combined with a contrastive loss function. The SimCLR pipeline consists of four key steps: (1) Data Augmentation, where each input image is transformed into two augmented views using random cropping, rotation, noise addition, and color distortion; (2) Feature Extraction, where

both augmented views are processed by a shared backbone network; (3) Feature Mapping, where representations are projected into a lower-dimensional contrastive space; and (4) Contrastive Learning, where a contrastive loss function minimizes the distance between positive pairs (augmented views of the same image) while maximizing the distance between negative pairs (augmented views of different images). A key advantage of SimCLR lies in its flexibility, as the data augmentation techniques can be tailored to specific tasks. For instance, in tasks requiring models to learn geometric structures, augmentations such as random cropping and rotation can be emphasized. In contrast, when robustness to noise is critical, noise-based augmentations can be incorporated to improve model performance.

While MoCo and SimCLR have demonstrated state-of-the-art performance in contrastive learning, their design principles also provide a foundation for enhancing few-shot learning methods. By leveraging self-supervised learning (SSL) techniques, our proposed OCEAN framework integrates contrastive learning into few-shot learning to improve feature representation quality, reduce reliance on labeled data, and enhance classification accuracy in challenging underwater environments. The ability to generate high-quality embeddings through SSL is particularly beneficial in marine ecosystem monitoring, where labeled datasets are often scarce and expensive to obtain. This integration highlights the potential of self-supervised learning as a key enabler for scalable, data-efficient coral classification systems.

3 Methodology

3.1 Overview of the OCEAN Model

The Optimized Coral Evaluation and Assessment Network (OCEAN) is designed to address the challenges of few-shot coral classification in real-world underwater environments. Traditional few-shot learning methods often struggle to generalize due to the high intra-class variability and morphological diversity of coral species, as well as the difficulty of obtaining large annotated datasets. To overcome these limitations, OCEAN integrates self-supervised learning (SSL) to enhance feature representation, a hybrid similarity measure to improve classification accuracy, and local patch-based feature extraction to capture fine-grained coral structures.

OCEAN operates in two primary stages: the training stage and the testing stage. In the training stage, the model builds on the prototype-based learning framework from Prototypical Networks (ProtoNets) while incorporating self-supervised contrastive learning to generate additional pseudo-labels, reducing the reliance on manually annotated data. Furthermore, a dual-metric similarity approach, combining Euclidean distance and cosine similarity, is employed to refine prototype computation, ensuring better class separation and improved generalization. In the testing stage, OCEAN enhances classification accuracy by extracting local image patches in addition to global image representations. This approach leverages the repetitive structures and local patterns inherent in coral species, which provide additional discriminative features for classification. The final prediction is determined through a voting-based strategy, where the classification results from different patches are aggregated, ensuring robustness against variations in lighting, occlusions, and environmental distortions.

The conceptual framework of OCEAN is illustrated in [Fig. 1](#), which outlines the training and testing processes. The integration of self-supervised learning, a hybrid similarity measure, and local patch evaluation enables OCEAN to achieve higher classification accuracy with fewer labeled samples, making it a practical solution for large-scale marine biodiversity monitoring.

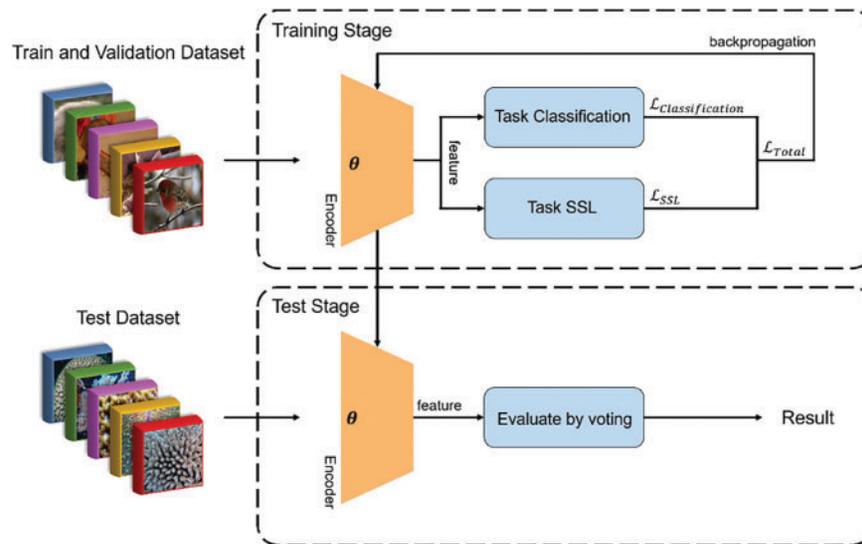


Figure 1: Structural overview of the optimized coral evaluation and assessment network (OCEAN) model for few-shot coral classification

The training stage of OCEAN is designed to learn discriminative feature representations while maintaining adaptability to novel coral species. The model follows the prototype-based learning approach, where each class is represented by a class prototype, computed as the mean embedding of all samples within that class. This approach allows the model to generalize effectively to unseen classes using only a small number of labeled examples.

To further enhance the model's learning capability, a self-supervised learning task is incorporated into the training process. This self-supervision is implemented using a contrastive learning pretext task, where the model generates pseudo-labels from the data itself, rather than relying solely on human-labeled annotations. By encouraging the model to learn both global and local feature representations, self-supervised learning improves classification robustness, particularly in cases where labeled coral datasets are limited or imbalanced. Additionally, OCEAN introduces a dual-metric loss function that refines the classification process. While ProtoNets traditionally use Euclidean distance as the similarity measure, OCEAN integrates cosine similarity to complement Euclidean-based distance calculations. Euclidean distance captures feature magnitude differences, while cosine similarity focuses on directional alignment between feature vectors. By combining these two metrics, the model improves class discrimination and reduces misclassification errors, especially for visually similar coral species.

In the testing stage, OCEAN incorporates local feature extraction to enhance classification accuracy. Given that corals exhibit distinctive yet repetitive structural patterns, evaluating local image patches alongside global features allows the model to leverage fine-grained morphological details for improved recognition. This is particularly useful in real-world underwater environments, where coral images may suffer from blur, occlusions, and varying lighting conditions. To implement this, each test image is divided into multiple local patches, and each patch undergoes individual feature extraction and classification using the trained OCEAN model. The classification results from both the global image and local patches are then aggregated using a voting-based strategy, where the final decision is determined based on the most frequently predicted class. This approach ensures higher classification reliability by minimizing the impact of environmental distortions and enabling the model to utilize multiple perspectives of the same coral specimen.

By combining few-shot learning, self-supervised contrastive learning, and a dual-metric loss function, OCEAN significantly enhances coral classification performance in data-scarce environments. The integration of both global and local feature representations allows the model to adapt to diverse coral morphologies, improving classification accuracy while reducing dependency on large manually labeled datasets. Furthermore, the voting-based evaluation mechanism provides an additional layer of robustness, making OCEAN a practical and scalable solution for real-world marine ecosystem monitoring and conservation efforts.

3.2 Hybrid Similarity-Based Classification in Few-Shot Learning

Few-shot learning methods, such as Prototypical Networks (ProtoNets), commonly rely on Euclidean distance as a similarity metric to compare feature representations. Euclidean distance is computed by taking the square root of the sum of squared differences across each feature dimension, thereby quantifying the absolute positional difference between feature vectors. While this metric effectively captures magnitude-based differences, it does not account for directionality, which is crucial in high-dimensional feature spaces where feature vectors of the same class may have different magnitudes but similar orientations.

To address this limitation, OCEAN introduces a hybrid classification loss function, $\mathcal{L}_{Classification}$, which integrates Euclidean distance and cosine similarity to provide a more comprehensive assessment of feature relationships. This approach ensures that both absolute differences (magnitude variations) and angular similarities (directional alignment) are considered, leading to improved feature discrimination and classification stability. The classification loss function is defined as follows (Eq. (3)):

$$\mathcal{L}_{Classification} = -\frac{1}{K} \sum_{i=1}^K \text{LogSoftmax}(-\mathcal{D}_{i, y_i} \times \hat{\mathcal{S}}_{i, y_i}) \quad (3)$$

where K represents the number of query samples, and y_i denotes the true class label of the query sample Q_i . The LogSoftmax function normalizes the computed distance scores into a logarithmic probability distribution, which penalizes incorrect classifications by assigning higher loss values to lower confidence predictions.

A key component of this loss function is the Euclidean distance measure, denoted as $\mathcal{D}_{i,j}$, which quantifies the feature space distance between a query sample Q_i and a class prototype \mathcal{P}_j . This is formulated as (Eq. (4)):

$$\mathcal{D}_{i,j} = \sqrt{\sum_{c=1}^C (Q_i^c - \mathcal{P}_j^c)^2} \quad (4)$$

where C represents the dimensionality of the feature space, and Q_i^c and \mathcal{P}_j^c denote the feature values of the query sample and the class prototype in the c -th dimension, respectively. While Euclidean distance is effective for magnitude-based separation, it does not consider feature orientation, which can lead to misclassification in cases where the magnitude differs but the feature direction remains similar.

To complement Euclidean distance, OCEAN incorporates cosine similarity, denoted as $\mathcal{S}_{i,j}$, to measure the angular relationship between feature vectors. Cosine similarity is computed as Eq. (5):

$$\mathcal{S}_{i,j} = \frac{Q_i \cdot \mathcal{P}_j}{\|Q_i\|_2 \|\mathcal{P}_j\|_2} \quad (5)$$

where $Q_i \cdot \mathcal{P}_j$ represents the dot product of the feature vectors, and $\|Q_i\|_2 \|\mathcal{P}_j\|_2$ denotes the product of their L2 norms. The cosine similarity score ranges from -1 (opposite directions) to 1 (perfect alignment), with higher values indicating greater semantic similarity between the query sample and the class prototype.

To further refine classification decisions, the cosine similarity probability, denoted as $\hat{\mathcal{S}}_{i,j}$, is derived by applying a Softmax function to the similarity scores, ensuring a probability-based weighting of similarities across classes (Eq. (6)):

$$\hat{\mathcal{S}}_{i,j} = 1 - \text{Softmax}(\mathcal{S}_{i,j}) \quad (6)$$

where the Softmax function normalizes similarity scores, ensuring that similarity scores across different classes are proportionally weighted in the final classification process.

The final classification loss term, $\mathcal{D}_{i,y_i} \times \hat{\mathcal{S}}_{i,y_i}$, serves as a hybrid distance score that integrates both magnitude-based and angular similarity-based features. This term is incorporated into the LogSoftmax function, which normalizes the hybrid distance score into a probability distribution, ensuring that larger distances correspond to smaller classification probabilities. The final loss formulation is expressed as Eq. (7):

$$\text{LogSoftmax}(-\mathcal{D}_{i,j} \times \hat{\mathcal{S}}_{i,j}) = \log \left(\frac{e^{-\mathcal{D}_{i,j} \times \hat{\mathcal{S}}_{i,j}}}{\sum_{n=1}^N e^{-\mathcal{D}_{i,n} \times \hat{\mathcal{S}}_{i,n}}} \right) \quad (7)$$

where N represents the total number of class prototypes. The numerator exponentiates the negative hybrid distance score, ensuring that larger distances correspond to smaller probability values, while the denominator sums these values across all prototypes, producing a normalized probability distribution.

By integrating both Euclidean distance and cosine similarity, the $\mathcal{L}_{\text{Classification}}$ loss function provides a richer and more stable representation of feature relationships by capturing both absolute positional differences and directional alignment in the feature space. This hybrid approach ensures that the model effectively learns both magnitude-based and orientation-based relationships, leading to more robust and accurate classification performance.

One key advantage of this approach is its ability to reduce misclassification risk. Since Euclidean distance alone measures only feature magnitude differences, it may incorrectly classify scaled but semantically similar features as distinct. By incorporating cosine similarity, the model preserves feature orientation, ensuring that samples belonging to the same class remain closely aligned in the embedding space, even if their magnitudes differ. Additionally, the hybrid metric helps enhance class separation. Coral species often exhibit high intra-class variability, making it challenging for traditional similarity measures to differentiate visually similar but distinct species. By jointly considering both magnitude and directional relationships, the model is better equipped to distinguish subtle morphological differences, improving classification accuracy, particularly in few-shot settings where only a limited number of labeled examples are available. Another significant benefit of this approach is its ability to improve generalization in few-shot learning scenarios. Few-shot learning models typically suffer from overfitting due to limited labeled data, which can result in unstable feature representations. By incorporating cosine similarity, the model gains an additional learning signal that helps stabilize the feature space, reducing variance and improving classification robustness across different coral species. This refined classification loss function significantly enhances the performance of OCEAN in few-shot coral classification tasks, particularly in challenging underwater environments where lighting conditions, occlusions, and intra-species variability can negatively impact recognition accuracy. By leveraging both Euclidean distance and cosine similarity, OCEAN achieves more reliable, stable, and scalable classification performance, making it an effective solution for marine biodiversity monitoring and conservation efforts.

3.3 Self-Supervised Learning Tasks

The OCEAN model is designed to enhance classification accuracy by incorporating self-supervised learning (SSL) to address the limitations of conventional few-shot learning (FSL) approaches. In standard FSL frameworks such as Prototypical Networks, the model primarily captures global relationships between images, often failing to recognize fine-grained local feature structures. This limitation prevents the encoder from effectively utilizing local feature correspondences, making classification less robust when dealing with high intra-class variability, lighting variations, or occlusions.

To overcome this challenge, OCEAN integrates self-supervised learning as an auxiliary training task, reinforcing the model's ability to capture local feature relationships. SSL provides two key benefits: it eliminates the need for additional labeled data, significantly reducing annotation effort, and it allows modular adjustment of learning tasks to optimize feature extraction. By leveraging contrastive learning, OCEAN improves both global and local feature representations, making it more effective in real-world coral classification tasks where labeled data is scarce.

As illustrated in Fig. 2, the self-supervised learning process in OCEAN consists of three main steps. First, the model extracts features from three sets of samples using the backbone network. Next, it applies similarity distance loss to measure relationships among these feature sets. The three input sets include the original image dataset (Q), a locally cropped patch (Q^{sub1}), and another locally cropped patch (Q^{sub2}). These additional patches are generated using a random cropping function (φ_r), which ensures that local features are preserved while introducing spatial diversity into the training process. The self-supervised loss function (\mathcal{L}_{SSL}) optimizes feature similarity learning by encouraging local patches to maintain structural consistency with the original image.

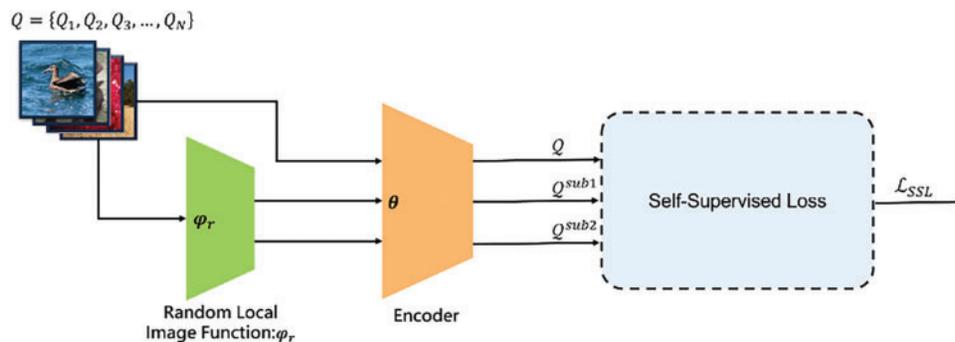


Figure 2: Overview of the self-supervised learning (SSL) process in OCEAN

Before initiating the self-supervised training process, all input images are resized to a fixed resolution of 160×160 pixels to ensure consistency across episodes. This uniform preprocessing step allows the model to process both global and local image views at the same spatial scale. During the self-supervised task, we adopt a two-stage cropping method to extract local patches. First, a center cropping operation removes the peripheral region of the image, which often contains sparse or ambiguous visual cues. Then, two quarter-sized patches are randomly sampled from within the center-cropped region to form localized views, ensuring that the cropped patches are likely to contain meaningful coral structures. Unlike traditional pipelines, we deliberately avoid using visual augmentations such as color jittering or normalization, preserving the raw image characteristics that are essential in underwater visual analysis.

To further refine local feature learning, OCEAN employs a two-stage cropping strategy, as depicted in Fig. 3. The first stage, center cropping, removes the outer 25% of the image to ensure that critical coral structures remain in focus. This step prevents the model from learning non-object regions that could introduce noise into feature extraction. In the second stage, two additional patches are randomly cropped from the remaining image, generating Q^{sub1} and Q^{sub2} . These local patches help the model establish correspondences between global and local features, allowing for more fine-grained classification.

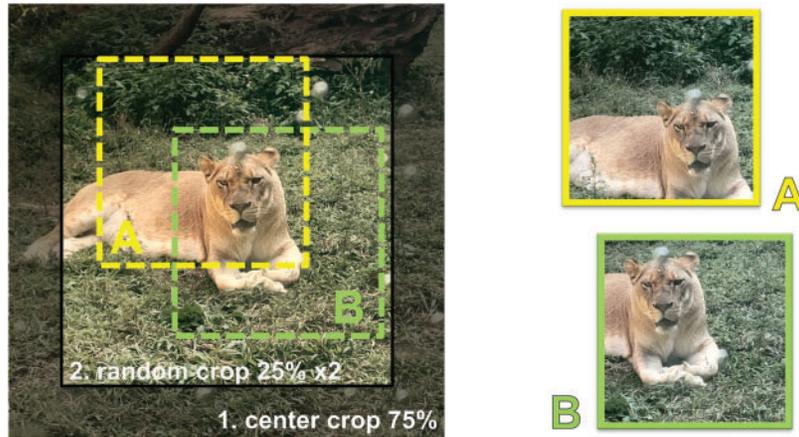


Figure 3: Data transformation for OCEAN's self-supervised task (Random local patch selection)

The OCEAN model's self-supervised learning task is designed to enhance feature representation by leveraging contrastive learning, which ensures that local and global features remain consistent within the feature space. To achieve this, OCEAN computes similarity distances between the original feature representation Q and two randomly cropped local patch features Q^{sub1} and Q^{sub2} . These similarity measurements serve as the foundation for the self-supervised learning loss L_{SSL} , which is formulated as the average of three pairwise distance losses, defined as follows (Eq. (8)):

$$\mathcal{L}_{SSL} = \frac{\mathcal{L}_{src_sub1} + \mathcal{L}_{src_sub2} + \mathcal{L}_{sub1_sub2}}{3} \quad (8)$$

The term \mathcal{L}_{src_sub1} represents the loss between the original image Q and the first local patch Q^{sub1} . Similarly, \mathcal{L}_{src_sub2} denotes the loss between the original image Q and the second local patch Q^{sub2} . Finally, \mathcal{L}_{sub1_sub2} corresponds to the loss between the two local patches, Q^{sub1} and Q^{sub2} , ensuring consistency between different regions of the same image.

To calculate these losses, OCEAN utilizes the *SimDistance* function, which evaluates similarity distances between two feature sets. As illustrated in Fig. 4, the *SimDistance* function takes three sets of features as inputs: the original image feature Q and the two local patch features Q^{sub1} and Q^{sub2} . It then computes a pairwise similarity distance matrix, allowing the model to assess how well the extracted local patches align with the global image representation.

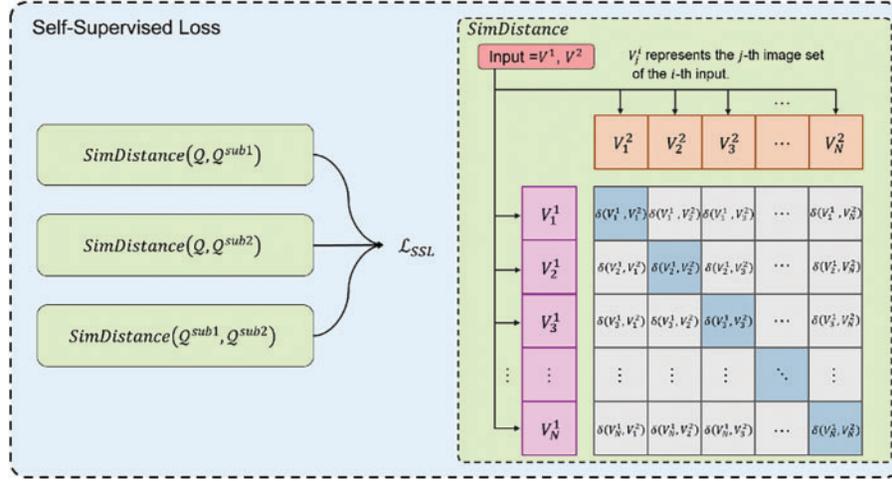


Figure 4: Loss computation and similarity distance measurement in self-supervised learning

The *SimDistance* function measures feature similarity between two sets of features V^1 and V^2 by computing a similarity distance matrix for all samples in the dataset. The core of this function focuses on whether features from the same source are structurally similar, as indicated by the values highlighted in blue in Fig. 4. Mathematically, the *SimDistance* loss function is formulated as follows (Eq. (9)):

$$\text{SimDistance}(V^1, V^2) = -\frac{1}{K} \sum_{i=1}^K \log \text{Softmax}(-\mathcal{D}_{i,i} \times \hat{\mathcal{S}}_{i,i}) \quad (9)$$

The variables V^1 and V^2 represent two sets of features, where $V^1, V^2 \in \mathbb{R}^{K \times C}$, with K denoting the number of samples and C representing the feature dimension. The matrix $D \in \mathbb{R}^{K \times K}$ corresponds to the Euclidean distance matrix, which is computed using the extracted feature embeddings. Additionally, $\hat{\mathcal{S}} \in \mathbb{R}^{K \times K}$ represents the cosine similarity matrix, capturing the directional alignment between feature vectors in the feature space.

To fully define the loss relationships between different feature sets, the three loss terms in Eq. (8) are computed as follows:

$$\mathcal{L}_{\text{src_sub1}} = \text{SimDistance}(Q, Q^{\text{sub1}}) \quad (10)$$

$$\mathcal{L}_{\text{src_sub2}} = \text{SimDistance}(Q, Q^{\text{sub2}}) \quad (11)$$

$$\mathcal{L}_{\text{sub1_sub2}} = \text{SimDistance}(Q^{\text{sub1}}, Q^{\text{sub2}}) \quad (12)$$

Here, Q represents the query sample features, and $Q^{\text{sub1}}, Q^{\text{sub2}} \in \mathbb{R}^{K \times C}$ are the features of the derived samples from Q through the random local image function φ_r in the self-supervised task. The loss $\mathcal{L}_{\text{src_sub1}}$ is calculated using the query sample feature Q and the local patch Q^{sub1} through the *SimDistance* function (Eq. (10)). The loss $\mathcal{L}_{\text{src_sub2}}$ is calculated using the query sample feature Q and the local patch Q^{sub2} (Eq. (11)). The loss $\mathcal{L}_{\text{sub1_sub2}}$ is calculated using the two local patches Q^{sub1} and Q^{sub2} (Eq. (12)). To unify the few-shot classification loss and self-supervised learning loss, OCEAN integrates these two objectives into a total training loss function (Eq. (13)):

$$\mathcal{L}_{\text{Total}} = \alpha \cdot \mathcal{L}_{\text{Classification}} + \beta \cdot \mathcal{L}_{\text{SSL}} \quad (13)$$

where $\mathcal{L}_{Classification}$ represents the loss from the few-shot classification task, and \mathcal{L}_{SSL} represents the loss from the self-supervised learning task. The coefficients α and β represent the weights of the two losses in the overall loss \mathcal{L}_{Total} . By combining contrastive self-supervised learning with few-shot learning, OCEAN achieves more robust feature extraction, ensuring that the model can generalize effectively even with limited labeled data. This fusion of global and local feature learning is particularly critical for coral classification, where species exhibit high intra-class variability and limited labeled datasets are available.

3.4 Voting-Based Evaluation

Corals exhibit repetitive structural patterns, meaning that different local patches from the same coral should share significant semantic similarity. Building upon this property, OCEAN enhances its classification capability by integrating local semantics into the evaluation process. This strategy enables the model to better capture subtle morphological differences among coral species, leading to improved classification accuracy. To achieve this, we introduce an innovative few-shot voting-based evaluation method, which modifies the evaluation framework to allow independent predictions for both the original image and its corresponding local patches.

As illustrated in Fig. 5, the voting-based evaluation process in OCEAN involves multiple stages of feature extraction and classification. During evaluation, OCEAN applies two distinct local selection functions, φ_s and φ_q , to extract local regions from the support set and the query set, respectively. These extracted patches, along with the original images, are then passed through the feature extractor to obtain multiple representations per sample. Notably, φ_s and φ_q operate independently, allowing for flexible patch selection strategies that can include either multiple fixed-size local patches or the use of the entire image without cropping. This process ensures that both global context and local feature variations are incorporated into the classification decision.

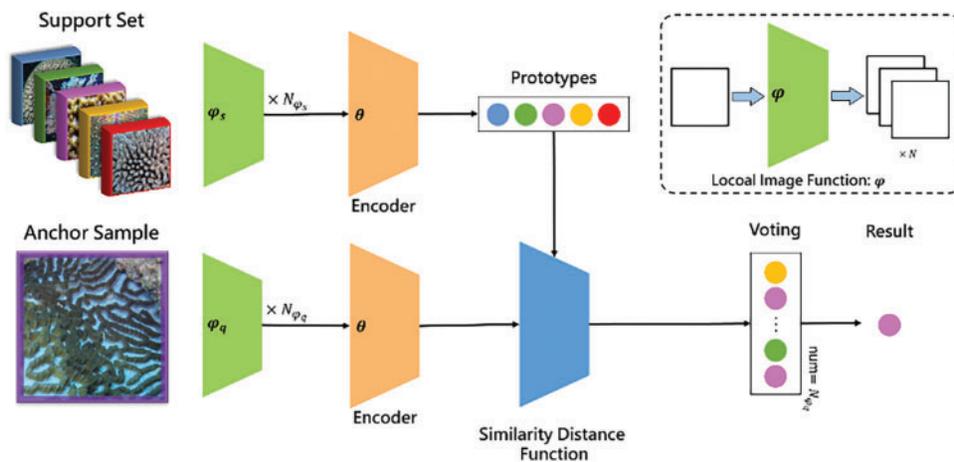


Figure 5: Workflow of the voting-based evaluation process in OCEAN

To further improve classification robustness, various patch selection strategies are employed, as demonstrated in Fig. 6. The selection mechanism for multiple local patches follows different approaches, including fixed-position cropping, random selection, and adaptive patch generation. These strategies enable OCEAN to dynamically adjust its evaluation process, leveraging spatial consistency across multiple local regions. By aggregating predictions from different patches through a voting mechanism, OCEAN reinforces classification reliability and mitigates potential errors caused by local feature variations or occlusions.

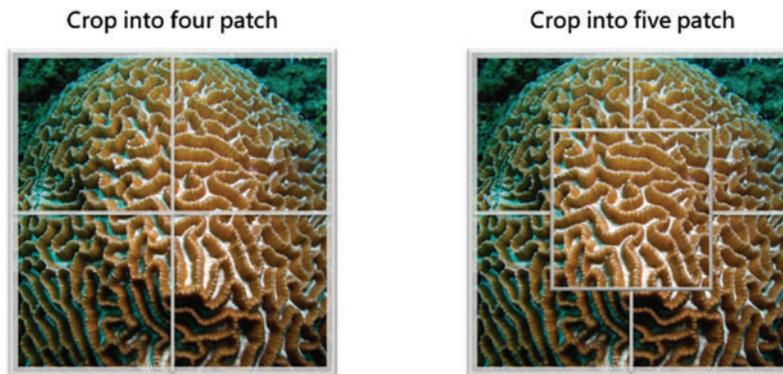


Figure 6: Patch selection strategies for the voting-based evaluation in OCEAN

This voting-based approach significantly enhances OCEAN’s performance in few-shot learning scenarios, particularly in cases where training data is limited and corals exhibit high intra-species variability. The combination of global image evaluation and local feature aggregation ensures a more stable and accurate classification process, making OCEAN well-suited for real-world coral monitoring applications.

4 Experiments and Results

4.1 Datasets Used in This Study for Few-Shot Learning and Classification

To validate the effectiveness of the proposed OCEAN model, this study utilizes the Tiered-ImageNet dataset for training, validation, and testing. Tiered-ImageNet is a subset of the widely used ImageNet dataset, specifically designed to enhance cross-category few-shot learning by grouping categories based on their hierarchical structure in WordNet. This organization introduces an additional level of complexity, making it well-suited for evaluating generalization across novel categories.

Table 1 provides an overview of the dataset composition used in this study. The training set of Tiered-ImageNet comprises 351 categories with 448,695 images, while the test set includes 160 categories with 206,209 samples. Additionally, a validation set containing 97 categories with 124,261 images is used to assess the model’s performance during training.

Table 1: Utilization of tiered-ImageNet and other datasets in this study

Dataset	Number		Usage	
	Categories	Samples		
Tiered-ImageNet [32]	Training set	351	448,695	Training
	Test set	160	206,209	Training
	Validation set	97	124,261	Validation
CUB-200-2011 [33]		200	11,788	Ablation study
FGVC-Aircraft [34]		102	10,200	Ablation study
Coral dataset of this study		10	1224	Testing

In addition to Tiered-ImageNet, two additional fine-grained classification datasets—CUB-200-2011 [33] and FGVC-Aircraft [34]—are incorporated to evaluate the adaptability of the model across different domains.

CUB-200-2011 consists of 200 bird species and is widely used for fine-grained biological classification. Similarly, FGVC-Aircraft is a dataset specializing in aircraft model recognition, making it an ideal benchmark for assessing fine-grained classification performance in few-shot learning scenarios. Both datasets contribute to the ablation studies conducted in this research, allowing for a comprehensive analysis of the model's ability to generalize across diverse categories.

For the final testing phase, we introduce a custom coral dataset, specifically compiled for this study. The dataset construction follows the coral classification systems of Dai (2023) [2] and the World Register of Marine Species (WoRMS) [35]. It consists of 10 distinct coral genera, comprising a total of 1224 images, primarily sourced from online repositories [36] and underwater photography screenshots. To minimize classification inconsistencies, images containing multiple coral species were excluded.

Table 2 coral dataset used for model evaluation, listing 10 distinct coral families and their corresponding genera, classification labels, and sample counts. This dataset was carefully compiled to represent a diverse range of coral morphologies, ensuring that the model is exposed to varied structural patterns commonly found in real-world underwater environments. The Acroporidae family, comprising 259 samples, is the most well-represented, as *Acropora* species are among the most abundant reef-building corals. Meanwhile, Melithaeidae (210 samples) and Merulinidae (136 samples) introduce additional morphological diversity, with branching and encrusting growth patterns.

Table 2: Composition of the coral dataset used for model evaluation

Coral species	Genus	Coral label	Number of samples
Acroporidae	<i>Acropora</i>	acro	259
Agariciidae	<i>Leptoseris</i> , <i>Pavona</i>	agar	107
Euphylliidae	<i>Euphyllia</i> , <i>Fimbriaphyllia</i>	euph	82
Fungiidae	<i>Fungia</i> , <i>Cycloseris</i> , <i>Lithophyllon</i> , <i>Pleuractis</i>	fung	87
Lobophylliidae	<i>Lobophyllia</i> , <i>Acanthastrea</i>	lobo	99
Melithaeidae	<i>Melithaea</i>	meli	210
Merulinidae	<i>Favites</i> , <i>Paragoniastrea</i> , <i>Platygyra</i> , <i>Oulophyllia</i>	meru	136
Pocilloporidae	<i>Pocillopora</i>	poci	77
Sarcophytidae	<i>Sarcophyton</i>	sarc	62
Siderastreidae	<i>Pseudosiderastrea</i>	sidr	105

One of the key challenges in few-shot coral classification is high intra-class variability, where species within the same genus exhibit significant morphological differences. For example, Merulinidae corals can appear in spherical, flattened, or encrusting forms, making their classification particularly difficult. Additionally, corals such as Sarcophytidae (soft corals) and Fungiidae (mushroom corals) further increase dataset complexity, as they lack rigid skeletons or grow in unconventional shapes. The inclusion of Pocilloporidae and Agariciidae, known for their delicate branching and plate-like structures, adds another level of classification difficulty.

Given these challenges, this dataset serves as a valuable benchmark for testing few-shot learning models, as it reflects the visual complexity and environmental conditions encountered in real-world marine ecosystems. The dataset's wide range of coral structures ensures that the model must learn to distinguish subtle morphological differences, making it a robust evaluation tool for developing automated coral classification systems. While our current dataset was curated to include only single-species images to ensure controlled and focused evaluation, we acknowledge that in real-world reef environments, coral images often

contain multiple co-occurring species. This poses an additional challenge for fine-grained classification. Although our study does not yet integrate instance-level segmentation or multi-object detection techniques, we recognize the importance of such solutions for practical deployment. As a preliminary step, our exclusion of multi-species images aims to establish a robust baseline. In future work, we plan to expand the dataset to include annotated multi-species coral scenes and investigate hybrid approaches that combine few-shot learning with instance segmentation or weakly supervised localization to enable the recognition of individual coral taxa within complex compositions. This direction holds potential to further enhance OCEAN's applicability in unconstrained underwater monitoring tasks.

4.2 Meta-Learning-Based Optimization in OCEAN

OCEAN follows a meta-learning training paradigm, where the model is trained over 300 epochs, with each epoch comprising a training stage and a validation stage. In each training stage, 100 episodes are randomly sampled from the Tiered-ImageNet training set (Table 1), and model parameters are adjusted based on the computed loss. During the validation stage, another 100 episodes are randomly selected from the Tiered-ImageNet validation set, and the validation accuracy is recorded to monitor performance.

Each episode follows the same structured setup for both training and validation. The dataset is randomly split into support sets and query sets, ensuring no overlap between them. Each episode consists of 5 classes, with 5 support samples per class and 15 query samples per class, resulting in a total of 100 samples per episode. This consistent episode-based learning structure enables the model to develop robust representations suitable for few-shot classification. After completing training, the best-performing model weights, based on the highest validation accuracy, are selected for final evaluation on the coral dataset.

The backbone network used in OCEAN is ResNet-18, a well-established feature extractor in deep learning. The input resolution is set to 160×160 pixels, and the final pooling layer and fully connected layer are replaced with a global average pooling layer that outputs a 1×1 feature representation. The initial model parameters are pre-trained on ImageNet, leveraging its large-scale image features to accelerate convergence and improve feature extraction capabilities.

For model optimization, we adopt Eq. (13) as the primary loss function, balancing contributions from few-shot classification loss and self-supervised learning loss. The hyperparameters are set as $\alpha = 0.5$ and $\beta = 0.5$, ensuring equal weighting between the two loss components. The initial learning rate is set to 0.001, following a gradual decay strategy to stabilize training.

To further enhance classification robustness, we introduce a hybrid similarity metric, termed *mix_dist*, to evaluate vector distances in the OCEAN model (Eq. (14)):

$$mix_dist = Sim_{cos} \times Sim_{euc} \quad (14)$$

Here, Sim_{cos} represents cosine similarity probability values for non-target classes, while Sim_{euc} corresponds to the Euclidean distance between sample features and class prototypes. This dual-metric approach enhances the classification process by considering both magnitude (Euclidean distance) and direction (cosine similarity) within the feature space.

The cosine similarity component is first transformed into a probability distribution using the *softmax* function and then inverted to quantify the probability of non-target classes (Eq. (15)):

$$Sim_{cos} = 1 - softmax(cosine_similarity(\mathcal{P}, F)) \quad (15)$$

where \mathcal{P} represents the prototypes of each class, and F denotes the sample features. By integrating these two complementary similarity measures, OCEAN ensures a more stable and discriminative classification

process, leading to improved performance in few-shot coral classification tasks. For a standard 5-Way 5-Shot setting: the total training time was approximately 4 h and 39 min (16,848.70 s), with peak memory usage measured at approximately 2697 MB. The use of a lightweight ResNet-18 backbone, fixed input resolution, and no auxiliary model branches ensures that the added complexity from hybrid similarity computation and contrastive learning is offset by efficient convergence and moderate resource demands. These properties make OCEAN deployable in real-world marine applications, especially for scenarios where computational resources are constrained and periodic monitoring is essential.

4.3 Coral Classification Results Using Few-Shot Learning

The coral dataset used in this study consists of ten distinct coral species, each with varying levels of classification difficulty. Among them, Fungiidae, Melithaeidae, and Siderastreidae exhibit unique morphological characteristics, making them relatively easier to identify. Conversely, Lobophylliidae and Merulinidae frequently get misclassified due to their similar structural features, while Acroporidae and Pocilloporidae share visually similar appearances, making them particularly challenging to distinguish.

Table 3 presents the experimental results of the OCEAN model on the coral dataset under 5-Way 5-Shot and 5-Way 10-Shot settings, with 5000 evaluations conducted in each setup. The table reports overall accuracy, as well as the precision, recall, and F1-score for each coral species. The results demonstrate that OCEAN achieves 71.83% accuracy in the 5-Way 5-Shot setting and 76.52% accuracy in the 5-Way 10-Shot setting. Notably, Fungiidae and Siderastreidae achieved the highest classification scores, with F1-scores exceeding 90%, suggesting that their distinctive morphological features contribute to easier identification. Melithaeidae also performed well, reaching an F1-score of 88.35% in the 10-Shot scenario.

Table 3: Few-shot classification results on the coral dataset (5-Way 5-Shot & 5-Way 10-Shot)

OCEAN (Ours) Coral species	5-Way 5-Shot Accuracy: 71.83%			5-Way 10-Shot Accuracy: 76.52%		
	Precision	Recall	F1	Precision	Recall	F1
Acroporidae	63.07%	63.81%	63.44%	68.42%	68.85%	68.63%
Agariciidae	67.46%	73.74%	70.46%	72.99%	75.64%	74.29%
Euphylliidae	69.08%	70.61%	69.83%	72.93%	76.20%	74.53%
Fungiidae	88.95%	89.33%	89.14%	90.71%	90.48%	90.60%
Lobophylliidae	64.20%	56.05%	59.85%	69.81%	61.30%	65.28%
Melithaeidae	82.30%	85.85%	84.03%	87.63%	89.08%	88.35%
Merulinidae	64.50%	59.57%	61.94%	69.26%	66.29%	67.74%
Pocilloporidae	68.39%	64.66%	66.48%	72.04%	72.09%	72.06%
Sarcophytidae	62.81%	64.26%	63.52%	70.38%	70.98%	70.68%
Siderastreidae	87.52%	97.28%	92.14%	90.89%	98.36%	94.48%

However, the model struggled with Lobophylliidae and Merulinidae, both of which exhibited lower recall scores due to high inter-class similarity with other coral species. Interestingly, Acroporidae and Pocilloporidae, despite their similar appearances, performed better than expected, suggesting that the model effectively captured subtle differentiating features between these two genera.

Fig. 7 illustrates the accuracy trend of OCEAN as the number of samples per class increases. When trained with only one sample per class (1-Shot), accuracy remains around 50%. However, as the number of support samples increases to 5-Shot, accuracy surpasses 70%. Beyond 20-Shot, accuracy improvements plateau around 80%, indicating that additional samples provide diminishing returns in model performance. To further examine the influence of support sample size on classification accuracy, we analyzed the model's

performance across a range of N-shot settings. As shown in Fig. 7, the accuracy increases steadily from 1-Shot to 10-Shot, and begins to plateau around 20-Shot. This trend suggests that OCEAN effectively benefits from additional labeled examples, though its performance remains competitive even with minimal supervision. Moreover, we observe that certain coral categories such as Siderastreidae and Fungiidae consistently achieve high F1-scores, even under lower-shot settings, indicating the robustness of the learned feature representations.

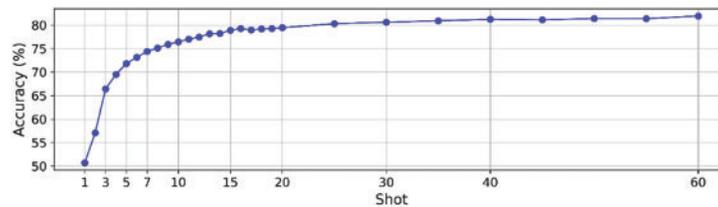


Figure 7: OCEAN accuracy progression with increasing support samples

In addition to aggregate metrics, we present a detailed error analysis through the confusion matrix shown in Fig. 8. This analysis reveals that misclassifications often occur between morphologically similar coral species, such as Merulinidae and Lobophylliidae. Notably, the proposed hybrid similarity strategy enables OCEAN to distinguish between these classes more effectively than baseline models. This qualitative insight complements our quantitative results and reinforces the model's strength in handling high intra-class variability—a critical challenge in underwater ecological classification.

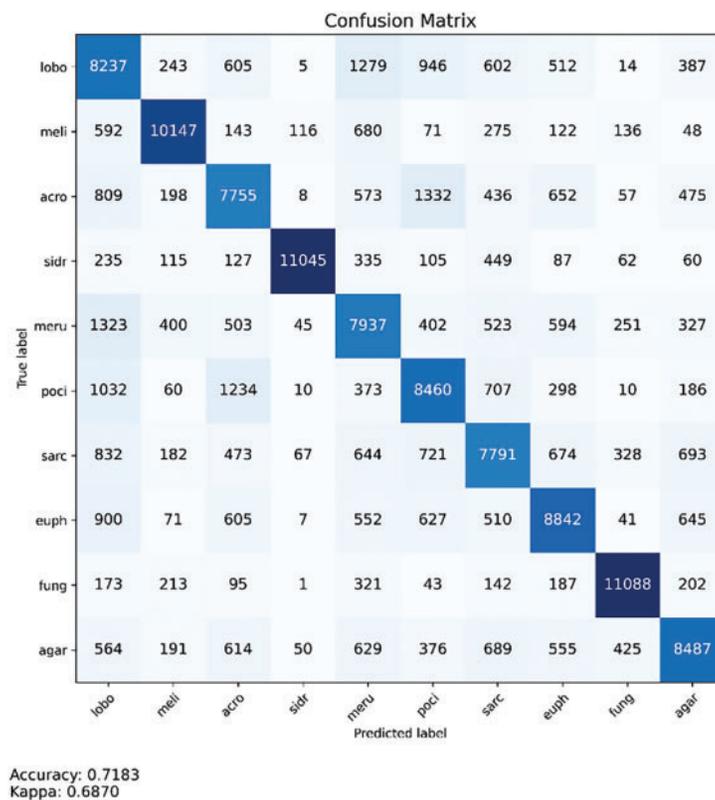


Figure 8: OCEAN confusion matrix: performance analysis on coral species recognition

However, one notable limitation emerged with Sarcophytidae, a soft coral species characterized by irregular, non-repetitive morphological structures. Unlike hard corals that exhibit more predictable and repetitive patterns, Sarcophytidae’s diverse and amorphous appearance challenges the assumptions of OCEAN’s voting-based classification mechanism. The current approach assumes structural repetition within each coral image, which enhances classification for most stony corals but may underperform when applied to species lacking such consistency. This is reflected in Sarcophytidae’s lower precision score compared to other species. To address this issue, future versions of OCEAN could incorporate adaptive voting mechanisms that evaluate patch-level feature reliability or adopt alternative strategies such as attention-guided aggregation. These enhancements would enable the model to better accommodate species with less predictable structures, thus improving generalization across a broader range of coral types.

This confusion matrix highlights the complexity of coral classification, where morphological variations, intra-species diversity, and environmental factors influence classification accuracy. The OCEAN model’s ability to differentiate visually similar corals demonstrates its effectiveness, particularly in low-data scenarios where few-shot learning is required.

4.4 Evaluating Loss Function Combinations for Few-Shot Learning

To assess the effect of loss function weighting on classification performance, OCEAN incorporates a linear combination of two loss terms: classification loss $\mathcal{L}_{Classification}$ and self-supervised learning loss \mathcal{L}_{SSL} . The weights α and β control their relative contributions, where α governs the classification objective, and β regulates the auxiliary self-supervised learning task. To determine the optimal balance, we conducted experiments varying these weights, with results summarized in [Table 4](#).

Table 4: Effect of classification and self-supervised loss weights on few-shot accuracy

α (Classification loss weight)	β (Self-supervised loss weight)	Accuracy (%)
0.01	0.99	60.83
0.1	0.9	66.48
0.3	0.7	69.75
0.5	0.5	71.83
0.7	0.3	69.62
0.9	0.1	70.37
0.99	0.01	70.31

As observed in [Table 4](#), the model performs better when the classification loss is given a higher weight. The highest accuracy of 71.83% was achieved when both losses were balanced equally at $\alpha = 0.5$ and $\beta = 0.5$. Notably, reducing the classification loss weight ($\alpha < 0.5$) significantly decreases accuracy, as self-supervised learning alone is insufficient to guide the model toward optimal classification. Conversely, increasing α beyond 0.7 does not yield further performance gains, suggesting that a purely classification-driven approach may overlook valuable feature representations learned through self-supervision. These findings indicate that an equal contribution of classification and self-supervised loss is optimal for OCEAN’s performance in 5-Way 5-Shot learning.

Beyond weight balancing, the formulation of the loss function itself is critical to model effectiveness. OCEAN employs a dual-metric approach, integrating Euclidean distance and cosine similarity into the

classification loss. To examine the impact of different loss function formulations, we conducted additional experiments comparing four distinct loss strategies, detailed in [Table 5](#).

Table 5: Evaluating loss metrics: Euclidean, cosine, and combined approaches

Loss calculation method	Formula	Accuracy (%)
Euclidean distance only	$\mathcal{L}_{method1}$	70.83
Independent calculation of two metrics	$\mathcal{L}_{method2}$	71.01
Additive fusion of two metrics	$\mathcal{L}_{method3}$	70.48
Multiplicative fusion of two metrics	$\mathcal{L}_{Classification}$	71.83

[Table 5](#) presents accuracy results across different loss formulations. The baseline approach using only Euclidean distance ($\mathcal{L}_{method1}$, [Eq. \(16\)](#)) achieved 70.83% accuracy. A naive approach of independently computing Euclidean and cosine similarity losses ($\mathcal{L}_{method2}$, [Eq. \(17\)](#)) slightly improved performance to 71.01%, demonstrating the added benefit of incorporating similarity-based learning. However, applying additive fusion of the two metrics ($\mathcal{L}_{method3}$, [Eq. \(18\)](#)) led to a drop in accuracy to 70.48%, likely due to improper scaling of the two terms. In contrast, OCEAN’s proposed multiplicative fusion approach ($\mathcal{L}_{Classification}$, [Eq. \(3\)](#)) achieved the highest accuracy of 71.83%, outperforming all other formulations.

$$\mathcal{L}_{method1} = -\frac{1}{K} \sum_{i=1}^K \text{LogSoftmax}(-\mathcal{D}_{i,y_i}) \quad (16)$$

$$\mathcal{L}_{method2} = \mathcal{L}_{method1} + \frac{1}{K} \sum_{i=1}^K \text{LogSoftmax}(\hat{\mathcal{S}}_{i,y_i}) \quad (17)$$

$$\mathcal{L}_{method3} = -\frac{1}{K} \sum_{i=1}^K \text{LogSoftmax}(-\mathcal{D}_{i,y_i} - \hat{\mathcal{S}}_{i,y_i}) \quad (18)$$

These results confirm that multiplicative fusion of Euclidean distance and cosine similarity provides a more robust feature representation, improving few-shot classification accuracy. Unlike additive approaches, multiplicative integration ensures that the model accounts for both feature magnitude and directional similarity, leading to more precise class differentiation. The improved performance suggests that this formulation effectively enhances class separability, particularly in low-data scenarios common in few-shot learning.

4.5 Effectiveness of Local Feature Voting in Few-Shot Learning Evaluation

The proposed voting-based evaluation method is designed to leverage the repetitive structures of corals by incorporating local feature information into the classification process. To investigate its effectiveness, we conducted experiments evaluating different local patch integration strategies within the support and query sets. The results, presented in [Table 6](#), illustrate the impact of various evaluation configurations on model accuracy in the 5-Way 5-Shot task.

Table 6: Comparison of local patch integration strategies in few-shot classification

Support set configuration	Query set configuration	Accuracy (%)
No action	No action	65.70

(Continued)

Table 6 (continued)

Support set configuration	Query set configuration	Accuracy (%)
No action	Add local blocks	63.75
Add local blocks	No action	60.76
Add local blocks	Add local blocks	71.83

As observed in [Table 6](#), when neither the support set nor the query set includes local patches (No Action/No Action configuration), the model achieves a baseline accuracy of 65.70%. However, modifying only one of the two sets—either adding local patches to the query set (No Action/Add Local Blocks) or to the support set (Add Local Blocks/No Action)—leads to a decrease in accuracy, dropping to 63.75% and 60.76%, respectively. This decline suggests that mismatched feature representations between the support and query sets introduce inconsistencies in classification.

Conversely, when both the support and query sets incorporate local patch information (Add Local Blocks/Add Local Blocks configuration), the model achieves a notable accuracy improvement to 71.83%, marking a 6.13% increase over the baseline. This result underscores the importance of maintaining feature consistency across both sets, reinforcing that the integration of local feature voting significantly enhances OCEAN’s classification performance in the 5-Way 5-Shot setting.

4.6 Impact of Training Data Composition on Coral Classification Performance

To assess the impact of different dataset configurations on OCEAN’s classification performance, we conducted additional experiments beyond the initial Tiered-ImageNet [32] training and validation setup. Specifically, we incorporated two additional datasets commonly used in few-shot learning—CUB-200-2011 and FGVC-Aircraft—to analyze how dataset diversity influences model generalization. The results of these evaluations, conducted under the 5-Way 5-Shot setting, are presented in [Table 7](#).

Table 7: Effect of training and validation dataset combinations on 5-Way 5-Shot accuracy

Training set	Validation set	Accuracy (%)
Tiered-ImageNet training set	Tiered-ImageNet validation set	70.83
Tiered-ImageNet training and testing sets	Tiered-ImageNet validation set	71.83
Tiered-ImageNet training and testing sets + CUB-200-2011 + FGVC-Aircraft	Tiered-ImageNet validation set	71.11
CUB-200-2011	FGVC-Aircraft	63.00
FGVC-Aircraft	CUB-200-2011	59.19

The results indicate that using only Tiered-ImageNet as the training and validation dataset yields high accuracy, with 70.83% achieved when training on the Tiered-ImageNet Training Set and validating on the Tiered-ImageNet Validation Set. Expanding the training dataset to include both the Tiered-ImageNet Training and Testing Sets further improves accuracy to 71.83%, suggesting that within-domain augmentation enhances model generalization.

However, incorporating CUB-200-2011 and FGVC-Aircraft alongside Tiered-ImageNet results in a slight accuracy decline to 71.11%, indicating that the introduction of fine-grained datasets adds variability that may alter feature distributions, affecting classification performance. Notably, when training exclusively on CUB-200-2011 and validating on FGVC-Aircraft, accuracy drops significantly to 63.00%, while the reverse configuration (training on FGVC-Aircraft and validating on CUB-200-2011) results in an even lower 59.19%. These findings highlight the increased challenge of fine-grained classification, where subtle intra-class variations require more advanced feature representations and adaptation strategies. While Tiered-ImageNet has proven effective for generalizable feature extraction, we acknowledge that its terrestrial image content does not fully reflect the characteristics of underwater environments. Moreover, the custom coral dataset used for testing comprises only 1224 images, which may not sufficiently represent the diversity of marine ecosystems across different regions or imaging conditions. As a result, the cross-domain performance of the current OCEAN model remains untested. To address this limitation, future research should explore domain adaptation strategies to improve the robustness of the model in marine-specific scenarios. In addition, expanding the coral dataset to include images from multiple locations and environmental conditions—or incorporating alternative data modalities such as sonar or hyperspectral imaging—could enhance the model’s adaptability and ecological applicability. These directions are critical for transitioning OCEAN from a controlled few-shot setting toward broader deployment in real-world coral monitoring systems.

4.7 Comparative Analysis of Few-Shot Learning Models for Coral Classification

To evaluate the effectiveness of the OCEAN model in few-shot coral classification, we compare its performance with several state-of-the-art few-shot learning models. All models are trained under identical experimental conditions to ensure a fair comparison. The same dataset is used for training and validation, and all models employ ResNet-18 as the backbone network with an input resolution of 160×160 . This consistency eliminates variations due to differences in network architecture. However, it is important to note that FEAT [24] incorporates a Transformer architecture in addition to ResNet-18, which distinguishes it from other models that rely solely on convolutional neural networks.

During training, all models follow a standardized 5-Way 5-Shot setup. After training, the models are evaluated on the same coral dataset in the testing phase. While each model retains its unique characteristics, such as different loss functions and similarity metrics, the evaluation criteria remain consistent across all experiments. The model achieving the highest validation accuracy is selected for testing, and its accuracy is averaged over 5000 test episodes to ensure statistical reliability. The comparative results are presented in Table 8.

Table 8: Performance comparison of few-shot learning models on the coral dataset (5-way classification)

Model	1-Shot accuracy (%)	5-Shot accuracy (%)	10-Shot accuracy (%)
SiameseNet [37]	32.00%	35.05%	37.23%
ProtoNets [23]	46.00%	62.13%	66.63%
RelationNetwork [38]	40.49%	55.33%	59.83%
BSNet [39]	39.98%	52.70%	56.68%
FEAT (Transformer) [24]	37.81%	52.55%	57.10%
ESPT(SSL) [15]	48.84%	67.51%	71.54%
BECLR [26]	49.82%	61.54%	65.72%
OCEAN (Ours)	50.71%	71.83%	76.52%

The results show that OCEAN consistently outperforms prior few-shot learning models across all evaluated settings. In the 5-Way 5-Shot classification task, OCEAN achieved an accuracy of 71.83%, surpassing traditional metric-based models such as Prototypical Networks [23] (62.13%) and Relation Network [38] (55.33%), as well as advanced models such as FEAT [24] (52.55%) and BSNet [39] (52.70%). Notably, even when compared to more recent self-supervised frameworks like ESPT [15] and BECLR [26], which achieved 67.51% and 61.54%, respectively, OCEAN demonstrates a clear advantage—outperforming the strongest of these by more than 4%. Moreover, this performance superiority persists across different evaluation scenarios. In the 1-Shot setting, where training samples are extremely limited, OCEAN attained 50.71% accuracy, exceeding BECLR’s 49.82% and ESPT’s 48.84%. Similarly, in the 10-Shot configuration, OCEAN reached 76.52%, the highest among all compared models, indicating its robustness when additional support samples are available. These results collectively confirm that the integration of hybrid similarity and contrastive self-supervision in OCEAN offers significant benefits for generalization in data-scarce ecological classification tasks, setting a new benchmark for few-shot coral species recognition.

4.8 Enhancing Coral Classification: A Comparative Study of OCEAN and ProtoNets

To further assess the effectiveness of OCEAN in coral classification, we conducted a direct performance comparison with Prototypical Networks (ProtoNets) [23], as both models leverage class prototypes for few-shot learning. The evaluation was performed under 5-Way 5-Shot and 5-Way 10-Shot settings using the coral dataset. This comparison focuses on precision across different coral species to highlight how OCEAN improves upon existing few-shot learning methods. The results are summarized in Table 9.

Table 9: Precision comparison of OCEAN and ProtoNets in few-shot coral classification

Coral species	5-Shot precision		10-Shot precision	
	ProtoNets	OCEAN (Ours)	ProtoNets	OCEAN (Ours)
Acroporidae	55.47%	63.07%	61.32%	68.42%
Agariciidae	66.82%	67.46%	73.50%	72.99%
Euphylliidae	50.79%	69.08%	55.90%	72.93%
Fungiidae	85.43%	88.95%	86.41%	90.71%
Lobophylliidae	43.22%	64.20%	50.43%	69.81%
Melithaeidae	79.88%	82.30%	81.21%	87.63%
Merulinidae	47.17%	64.50%	52.69%	69.26%
Pocilloporidae	55.10%	68.39%	61.82%	72.04%
Sarcophytidae	53.84%	62.81%	58.44%	70.38%
Siderastreidae	83.37%	87.52%	84.28%	90.89%
Accuracy	62.13%	71.83%	66.63%	76.52%

Table 9 presents the precision scores of OCEAN and ProtoNets across ten different coral species. In the 5-Way 5-Shot setup, OCEAN outperforms ProtoNets across all coral categories, achieving an overall accuracy improvement from 62.13% to 71.83%. Similarly, in the 5-Way 10-Shot setting, OCEAN’s accuracy increases further, reaching 76.52%, compared to 66.63% for ProtoNets. These results demonstrate that OCEAN provides a significant enhancement in classification performance, particularly for visually similar coral species.

A deeper examination of the species-wise precision reveals substantial improvements in certain challenging coral categories. Lobophylliidae and Merulinidae, which share overlapping structural characteristics, exhibited the most noticeable gains, with precision improvements of 21% and 17%, respectively. ProtoNets struggled to differentiate these species, likely due to their similar textures and morphological features. Euphylliidae, another challenging category, showed a precision increase of nearly 20%, reinforcing OCEAN's ability to extract fine-grained distinguishing features. For species with distinct appearances, such as Fungiidae, Melithaeidae, and Siderastreidae, both ProtoNets and OCEAN achieved relatively high precision. However, OCEAN further improved classification performance, pushing the precision of Fungiidae and Siderastreidae close to 90% in the 10-Shot setting. These results indicate that OCEAN enhances classification not only for visually ambiguous species but also for those with clearer morphological differentiation.

Despite the significant accuracy improvements, some coral families, such as Sarcophytidae, experienced relatively smaller gains. The precision of Sarcophytidae increased from 53.84% to 62.81%, which, while an improvement, was less pronounced than for other species. This may be attributed to the higher intra-class variability of Sarcophytidae, making it more difficult for the model to define a stable prototype. Additionally, feature overlap with other corals could contribute to increased classification ambiguity.

To further enhance classification accuracy, future research could explore several potential improvements. Expanding the training dataset to include a more diverse range of coral images could help the model learn richer feature variations and improve generalization. Additionally, refining feature extraction techniques, such as multi-scale feature learning, could enable better differentiation of visually similar species. Another possible direction is the integration of hybrid learning models, including transformer-based architectures, which could enhance feature representation and improve cross-category classification accuracy. The comparison highlights that OCEAN successfully overcomes key limitations of ProtoNets, making it a more robust and effective solution for few-shot coral classification in underwater environments. The combination of self-supervised learning and voting-based evaluation strategies allows OCEAN to achieve superior performance, particularly in challenging classification scenarios where species exhibit high morphological similarity.

5 Conclusion

This study proposed OCEAN, an optimized few-shot learning model for coral species classification in underwater environments where annotated data are limited. By integrating Euclidean distance and cosine similarity into a hybrid similarity framework, OCEAN captures both magnitude and directional relationships among feature vectors, enabling more accurate classification of morphologically similar corals. The model achieved 71.8% accuracy under a 5-Way 5-Shot setup—outperforming state-of-the-art models such as Prototypical Networks, FEAT, and ESPT by over 4%. In particular, it demonstrated high precision in classifying Siderastreidae (87.52%) and Fungiidae (88.95%). These results are further strengthened by the inclusion of a self-supervised contrastive learning task, which enhances the model's ability to learn discriminative features without requiring additional labels. A voting-based classification mechanism was also introduced to leverage the repetitive structures of corals, boosting accuracy to 76.5% in the 5-Way 10-Shot scenario. Together, these innovations make OCEAN an effective and scalable solution for coral monitoring and biodiversity conservation, with up to 15% improvement in classification performance compared to traditional methods. While this study focuses on single-species coral images for controlled few-shot evaluation, we acknowledge that real-world underwater imagery frequently contains multiple coral species within the same frame. Addressing this limitation remains an important direction for future work. Specifically, we plan to explore the integration of instance segmentation or weakly supervised localization techniques to enable OCEAN to distinguish and classify multiple coral taxa from a single image. Such

an extension would significantly enhance the model's applicability in practical reef survey and ecosystem monitoring scenarios.

However, OCEAN still faces several limitations that warrant future research. First, the model does not yet handle multi-species coral imagery well—a common challenge in real-world reef datasets. Incorporating instance segmentation or attention-based feature separation could help address this limitation. Second, while Tiered-ImageNet and a curated coral dataset were used for training, evaluating the model's generalizability to other marine environments and data modalities remains an open question. Future work should consider domain adaptation strategies and the inclusion of non-RGB inputs (e.g., sonar or hyperspectral data) to improve robustness in diverse underwater conditions. In particular, multi-modal fusion strategies that combine RGB images with complementary modalities like sonar or synthetic aperture sonar (SAS) may help mitigate the effects of turbidity and lighting variability, thereby enhancing classification performance in challenging aquatic environments.

Acknowledgement: Not applicable.

Funding Statement: This research was funded by the National Science and Technology Council (NSTC), Taiwan, under grant numbers NSTC 112-2634-F-019-001 and NSTC 113-2634-F-A49-007.

Author Contributions: Yu-Shiuan Tsai was responsible for conceptualization, methodology development, supervision, project administration, and manuscript preparation, including writing the original draft, reviewing, and editing. Zhen-Rong Wu contributed to investigation, conceptualization, methodology, and resource acquisition, as well as validation, visualization, experimentation, and writing the original draft. Jian-Zhi Liu was involved in experimental implementation. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The datasets and source code used in this study have been made publicly available at the following GitHub repository: <https://github.com/yystsai-lab/OCEAN> (accessed on 03 June 2025).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. McFadden CS, Quattrini AM, Brugler MR, Cowman PF, Dueñas LF, Kitahara MV, et al. Phylogenomics, origin, and diversification of Anthozoans (Phylum Cnidaria). *Syst Biol.* 2021;70(4):635–47. doi:10.1093/sysbio/syaa103.
2. Dai CF. One hundred marine livings in Taiwan-corals: ocean conservation administration. Kaohsiung, Taiwan: Ocean Affairs Council Press; 2023.
3. Masson-Delmotte V, Zhai P, Pirani S, Connors C, Péan S, Berger N, et al. Summary for policymakers. In: *Climate change 2021: The physical science basis. IPCC Sixth Assessment Report.* Geneva, Switzerland: IPCC; 2021. doi:10.25071/6sw6za31.
4. Thirukanthan CS, Azra MN, Lananan F, Sara' G, Grinfelde I, Rudovica V, et al. The evolution of coral reef under changing climate: a scientometric review. *Animals.* 2023;13(5):949. doi:10.3390/ani13050949.
5. Beijbom O, Edmunds PJ, Kline DI, Mitchell BG, Kriegman D. Automated annotation of coral reef survey images. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition; 2012 Jun 16–21; Providence, RI, USA.* p. 1170–7.
6. King A, Bhandarkar SM, Hopkinson BM. Deep learning for semantic segmentation of coral reef images using multi-view information. In: *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops; 2019 Jun 15–20; Long Beach, CA, USA.* p. 1–10.
7. Gapper JJ, El-Askary H, Linstead E, Piechota T. Coral reef change detection in remote Pacific islands using support vector machine classifiers. *Remote Sens.* 2019;11(13):1525. doi:10.3390/rs11131525.

8. Raine S, Marchant R, Kusy B, Maire F, Sunderhauf N, Fischer T. Human-in-the-loop segmentation of multi-species coral imagery. In: Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2024 Jun 17–18; Seattle, WA, USA. p. 2723–32.
9. Raphael A, Dubinsky Z, Iluz D, Benichou JI, Netanyahu NS. Deep neural network recognition of shallow water corals in the Gulf of Eilat (Aqaba). *Sci Rep.* 2020;10(1):12959. doi:10.1038/s41598-020-69201-w.
10. Reshma B, Rahul B, Sreenath K, Joshi K, Grinson G. Taxonomic resolution of coral image classification with convolutional neural network. *Aquat Ecol.* 2023;57(4):845–61. doi:10.1007/s10452-022-09988-0.
11. Li X, Yang X, Ma Z, Xue J-H. Deep metric learning for few-shot image classification: a review of recent developments. *Pattern Recognit.* 2023;138(4):109381. doi:10.1016/j.patcog.2023.109381.
12. Zhang Y, Guo X, Leung H, Li L. Cross-task and cross-domain SAR target recognition: a meta-transfer learning approach. *Pattern Recognit.* 2023;138(1):109402. doi:10.1016/j.patcog.2023.109402.
13. Liu B-D, Shao S, Zhao C, Xing L, Liu W, Cao W, et al. Few-shot image classification via hybrid representation. *Pattern Recognit.* 2024;155:110640. doi:10.1016/j.patcog.2024.110640.
14. Sim C, Kim G. Cross-attention based dual-similarity network for few-shot learning. *Pattern Recognit Lett.* 2024;186(2):1–6. doi:10.1016/j.patrec.2024.08.019.
15. Rong Y, Lu X, Sun Z, Chen Y, Xiong S. ESPT: a self-supervised episodic spatial pretext task for improving few-shot learning. In: Proceedings of the 37 AAAI Conference on Artificial Intelligence; 2023 Feb 7–14; Washington, DC, USA. p. 9596–605.
16. Contini M, Illien V, Barde J, Poulain S, Bernard S, Joly A, et al. From underwater to aerial: a novel multi-scale knowledge distillation approach for coral reef monitoring. *arXiv:2502.17883.* 2025.
17. Shao X, Chen H, Magson K, Wang J, Song J, Chen J, et al. Deep learning for multilabel classification of coral reef conditions in the Indo-Pacific using underwater photo transect method. *Aquat Conserv: Mar Freshw Ecosyst.* 2024;34(9):e4241. doi:10.1002/aqc.4241.
18. Blondin C, Guérin J, Inagaki K, Longo G, Berti-Équille L. Hierarchical classification for automated image annotation of coral reef benthic structures. *arXiv:2412.08228.* 2024.
19. Akkaynak D, Treibitz T. A revised underwater image formation model. In: Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 6723–32.
20. Sture Ø, Ludvigsen M, Scheide MS, Thorsnes T. Recognition of cold-water corals in synthetic aperture sonar imagery. In: 2018 IEEE/OES Autonomous Underwater Vehicle Workshop (AUV); 2018 Nov 6–9; Porto, Portugal. p. 1–6. doi:10.1109/auv.2018.8729718.
21. Zhou F, Wang P, Zhang L, Wei W, Zhang Y. Meta-collaborative comparison for effective cross-domain few-shot learning. *Pattern Recognit.* 2024;156(3):110790. doi:10.1016/j.patcog.2024.110790.
22. Bhandarkar SM, Kathirvelu S, Hopkinson BM. Object detection in 3D coral ecosystem maps from multiple image sequences. In: 2022 26th International Conference on Pattern Recognition (ICPR); 2022 Aug 21–25; Montreal, QC, Canada. p. 4637–43.
23. Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning. In: *Advances in neural information processing systems.* Vol. 30; 2017.
24. Ye H-J, Hu H, Zhan D-C, Sha F. Few-shot learning via embedding adaptation with set-to-set functions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2020 Jun 13–19; Seattle, WA, USA. p. 8805–14.
25. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst.* 2017;30:5998–6008.
26. Poulakakis-Daktylidis S, Jamali-Rad H. BECLR: batch enhanced contrastive few-shot learning. *arXiv:2402.02444.* 2024.
27. Doersch C, Gupta A, Efros AA. Unsupervised visual representation learning by context prediction. In: Proceedings of the 2015 IEEE International Conference on Computer Vision; 2015 Dec 7–13; Santiago, Chile. p. 1422–30.
28. Noroozi M, Favaro P. Unsupervised learning of visual representations by solving jigsaw puzzles. In: *Computer Vision—ECCV, 2016 (ECCV 2016).* Cham, Switzerland: Springer; 2016. p. 69–84 doi:10.1007/978-3-319-46466-4_5.

29. Paumard M-M, Picard D, Tabia H. Deepzzle: solving visual jigsaw puzzles with deep learning and shortest path optimization. *IEEE Trans Image Process.* 2020;29:3569–81. doi:10.1109/tip.2019.2963378.
30. He K, Fan H, Wu Y, Xie S, Girshick R. Momentum contrast for unsupervised visual representation learning. In: *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2020 Jun 13-19; Seattle, WA, USA. p. 9729–38.
31. Chen T, Kornblith S, Norouzi M, Hinton G. A simple framework for contrastive learning of visual representations. In: *The 37th International Conference on Machine Learning*; 2020; Vienna, Austria.
32. Triantafillou E, Larochelle H, Snell J, Tenenbaum J, Swersky KJ, Ren M, et al. Meta-learning for semi-supervised few-shot classification. In: *The 6th International Conference on Learning Representations (ICLR 2018)*; 2018 Apr 30–May 3; Vancouver, BC, Canada.
33. Wah C, Branson S, Welinder P, Perona P, Belongie S. *The caltech-ucsd birds-200-2011 dataset*. Pasadena, CA, USA: California Institute of Technology; 2011.
34. Maji S, Rahtu E, Kannala J, Blaschko M, Vedaldi A. Fine-grained visual classification of aircraft. arXiv:1306.5151. 2013.
35. Board WE. World register of marine species; 2024 [Internet]. [cited 2025 Jun 3]. Available from: <https://www.marinespecies.org>.
36. Veron JEN, Stafford-Smith MG, Turak E, DeVantier LM. Corals of the world 2016 [Internet]. [cited 2025 Jun 3]. Available from: http://www.coralsoftheworld.org/species_factsheets/.
37. Koch G, Zemel R, Salakhutdinov R. Siamese neural networks for one-shot image recognition. In: *ICML Deep Learning Workshop*. Lille, France; 2015.
38. Sung F, Yang Y, Zhang L, Xiang T, Torr PH, Hospedales TM. Learning to compare: relation network for few-shot learning. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 1199–208.
39. Li X, Wu J, Sun Z, Ma Z, Cao J, Xue J-H. BSNet: bi-similarity network for few-shot fine-grained image classification. *IEEE Trans Image Process.* 2020;30:1318–31. doi:10.1109/tip.2020.3043128.