E3S Web of Conferences 711, 01001 (2026) https://doi.org/10.1051/e3sconf/202671101001
EMER 2026

Hybrid Framework for Underwater Coral Image Recognition Using Generative Adversarial Networks and Semi-Supervised Segmentation

Yuqi Yang[1*], Chenxi Xin[2]

[1]School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Bengbu, China
[2]School of Artificial Intelligence, Tianjin University of Science and Technology, Tianjin, China

Abstract. Underwater environments are highly complex, where light attenuation, color shifts, and suspended particles degrade image quality and hinder reliable coral recognition. To address this, we propose an underwater coral image recognition framework that combines generative adversarial networks with semi-supervised segmentation. In this framework, the generative model enhances low-quality images and reduces domain discrepancies, while the semi-supervised strategy enables accurate semantic segmentation and species recognition with limited labeled data. Experiments on publicly available underwater coral datasets show that the proposed model outperforms traditional methods in terms of mean Intersection over Union (mIoU) and recognition accuracy. These results demonstrate a promising approach for underwater ecological monitoring and coral reef conservation.

1 Introduction

Coral reefs are a key component of marine ecosystems and are essential for maintaining biodiversity and ecological balance [1]. However, coral reefs are increasingly threatened by ocean acidification, climate change, and human activities, leading to widespread degradation. To support precise monitoring and automated assessment, coral image recognition based on computer vision has become an important research direction. Traditional monitoring methods rely on divers for manual photography and visual inspection, which is costly, time-consuming, and subjective, making large-scale and high-frequency monitoring difficult [2-3].

In recent years, deep learning methods [4-6], especially semantic segmentation networks such as U-Net, Fully Convolutional Networks (FCN), and DeepLab, have shown strong potential for underwater image analysis [7]. By assigning a semantic label to each pixel (e.g., hard coral, soft coral, sponge, background), these models enable quantitative estimation of coral coverage and community composition [15].

However, underwater images are often affected by uneven illumination, scattering, and color distortion, which reduce the effectiveness of conventional convolutional neural networks in feature extraction and classification. At the same time, underwater coral datasets are usually small and difficult to annotate, which limits the generalization of fully supervised models. To address these challenges, a variety of image enhancement methods have been proposed to restore more realistic color, contrast, and visibility. Traditional approaches are mostly based on physical imaging models or image priors, such as DCP and UDCP for modeling light attenuation, and Retinex-based correction or white balancing for contrast enhancement. More recently, learning-based models such as WaterNet and UWCNN have achieved better performance across different underwater conditions by learning end-to-end mappings from degraded to high-quality images [8-9].

Generative Adversarial Networks (GANs) extend this idea by introducing adversarial learning between a generator and a discriminator, providing a data-driven solution for unpaired image enhancement and style adaptation. Models such as CycleGAN, WaterGAN, and Ucolor have been successfully applied to underwater image restoration and color correction [10-12]. Building on this line of research, StyleGAN3 introduces an improved generator architecture with alias-free design and more stable style control, allowing it to synthesize high-fidelity images with consistent geometric structure across different resolutions. These properties make StyleGAN3 particularly suitable for generating realistic underwater coral images, supporting both visual enhancement and domain adaptation for downstream recognition tasks [19].

In this study, we propose an underwater coral image recognition framework that combines StyleGAN3-based image synthesis, super-resolution reconstruction, and semi-supervised segmentation. The StyleGAN3 module is used to generate diverse and realistic coral scenes and to improve the quality of low-resolution or degraded images, thereby enriching the training domain. A super-resolution reconstruction network further refines spatial details and texture, providing clearer inputs for segmentation. On this basis, a semi-supervised segmentation model is employed to leverage both labeled and unlabeled underwater images. By using consistency regularization and pseudo-labeling

* Corresponding author: yangyq_aufe22@163.com

[16, 20-21], the framework improves segmentation accuracy under limited annotation [13]. Among semi-supervised strategies, Teacher–Student architectures have shown strong performance: the teacher network produces pseudo-labels, while the student network learns under consistency constraints, and the teacher parameters are updated via exponential moving average to realize stable knowledge transfer [14][18].

Within the proposed framework, semi-supervised learning not only improves segmentation performance but also helps filter and refine StyleGAN3-generated samples, reducing noise and low-quality synthesis. The joint optimization of generative enhancement, super-resolution, and semi-supervised segmentation mitigates data scarcity and label noise, ultimately enhancing the robustness and accuracy of underwater coral recognition in practical monitoring scenarios.

2 Materials and methods

2.1 Dataset description

All images used in this study were collected during scuba surveys at 35 coral reef sites in Djibouti, Eritrea, Sudan, Jordan, and Israel, using GoPro Hero 10 cameras. The annotated data consist of video frames extracted from footage recorded at a resolution of 1080 × 1920 pixels and 30 frames per second with a linear lens setting. The videos cover a wide range of reef conditions, from healthy reefs with high live coral cover to degraded areas affected by severe bleaching and coral mortality. They also include scenes captured under varying water clarity and lighting, from clear, well-lit environments to turbid, low-light conditions. This diversity provides a representative basis for training and evaluating underwater coral recognition models. **Fig. 1** illustrates representative coral images and their corresponding pixel-level annotations, highlighting the visual complexity of underwater scenes and the detailed labeling used for model training and evaluation

2.2 Generative adversarial network method

To address the problems of limited sample size and uneven class distribution in underwater coral image datasets, this study uses a Generative Adversarial Network (GAN) to synthesize high-quality coral images that preserve realistic underwater characteristics. This approach increases the diversity of training data and improves model generalization, while reducing the need for extensive field data collection.

During training, the discriminator is first updated to improve its classification ability. The generator is then optimized based on the discriminator's feedback so that it can create more realistic images. Through iterative adversarial training, the generator gradually learns the complex distribution of real underwater coral images, including light attenuation, color shifts, texture patterns, and colony structures.

To further enhance structural fidelity and semantic consistency, this study adopts an advanced GAN architecture based on StyleGAN3 [23]. StyleGAN3 is used to generate high-quality coral images with fine textures and stable geometric structures. Unlike conventional convolution-based GANs, StyleGAN3 employs an alias-free design that reduces spatial artifacts caused by sampling, maintaining smooth transitions and preserving overall coral morphology. Its style-based modulation mechanism allows more explicit control over multi-scale visual attributes, such as color tone, texture richness, and colony layout. As a result, the images generated are more realistic and topologically consistent and better reflect underwater optical characteristics.

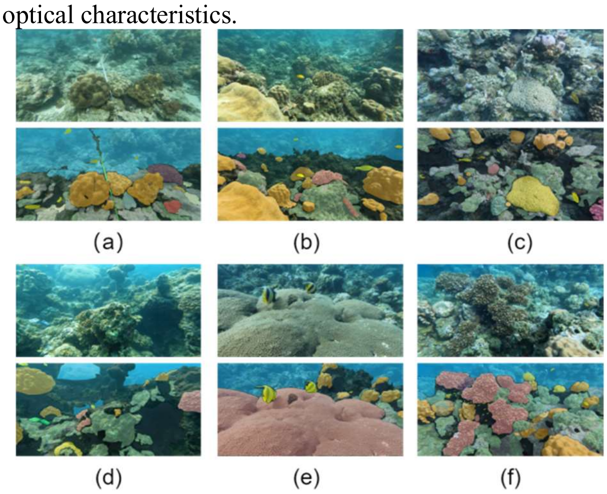

Fig. 1. Coral images and corresponding pixel-level annotations.

As illustrated in **Fig. 2**, high-resolution images are first degraded by blurring, down sampling, noise addition, and JPEG compression to create corresponding low-resolution inputs. Filters are used to simulate common artifacts such as ringing and overshooting. These degraded images are then passed into the generator to produce 1-, 2-, and 4-times super-resolved outputs, which are jointly trained with real high-resolution images in an adversarial manner.

Overall, the GAN-based data augmentation pipeline effectively alleviates the constraints of underwater image acquisition, data scarcity, and class imbalance. It enables the generation of visually convincing and statistically consistent synthetic coral images, thereby improving the robustness and generalization of subsequent underwater coral recognition and segmentation models.

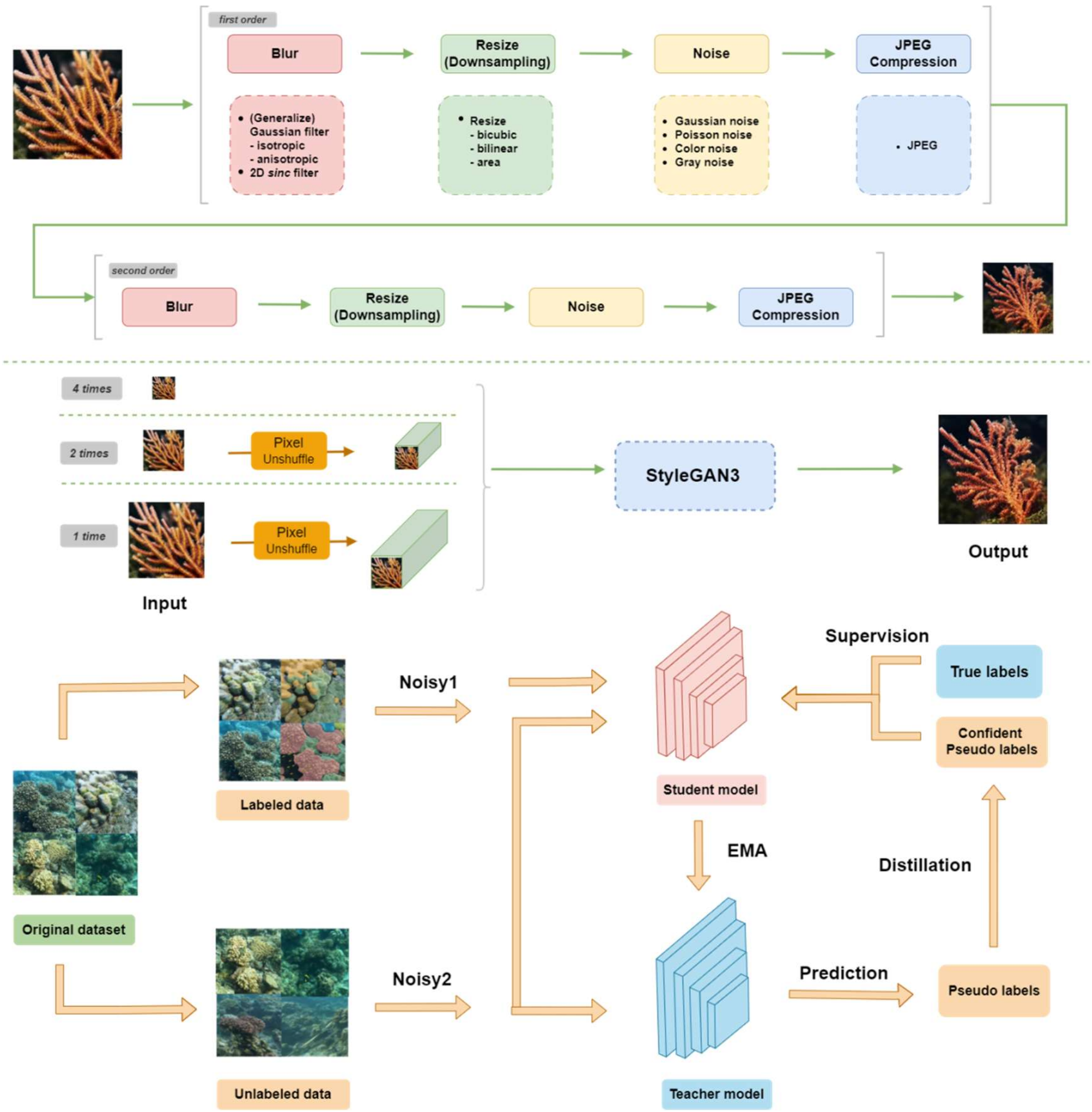

Fig. 2. Pipeline of GAN-based super-resolution and semi-supervised segmentation for underwater coral recognition.

2.3 Semi-supervised segmentation method

To address the limited number of samples and the high cost of annotation in underwater coral image datasets, this study adopts a semi-supervised segmentation strategy that makes joint use of labeled and unlabeled data. In this framework, the model learns discriminative features from a small set of manually annotated coral images, while simultaneously exploiting the larger pool of unlabeled images to improve generalization and segmentation performance [22].

As shown in **Fig. 2,** the original training set is evenly divided into two subsets, where one portion is manually annotated to form the labeled dataset, while the other is retained as the unlabeled dataset. For dataset construction, the control group applies conventional physical augmentation strategies such as 1-time and 3-time scaling, whereas the experimental group is further expanded using synthetic coral images generated by multiple GAN-based augmentation methods. These generative models produce diverse coral textures under varying illumination, turbidity, and color-attenuation conditions, thereby enhancing the segmentation model's ability to generalize across underwater environments. Among them, StyleGAN3 serves as the primary image generator. Its alias-free architecture effectively alleviates feature entanglement and suppresses sampling artifacts in synthesized images. The style-based modulation mechanism and progressive feature refinement enable stable adversarial training and accurate reconstruction of fine-grained coral structures, maintaining geometric consistency across different resolutions and transformations. By integrating degradation modeling with residual-dense generation, the produced synthetic coral images exhibit high visual and structural realism, ensuring that the augmented dataset closely approximates real underwater scenes and contributes to improved model robustness and generalization.

The training process follows a student – teacher paradigm, consisting of two identical segmentation networks: a student network and a teacher network. The student network is updated through gradient backpropagation using both supervised and unsupervised loss terms, while the teacher network is updated by applying an exponential moving average (EMA) to the

 https://doi.org/10.1051/e3sconf/202671101001

student's parameters. The teacher thus acts as a stable pseudo-label generator that evolves smoothly during training.

The supervised loss is computed on the labeled subset using standard pixel-wise cross-entropy between the predicted segmentation maps and the ground truth annotations. For the unlabeled subset, the unsupervised loss enforces consistency between the predictions of the student and teacher networks under different perturbations, including Gaussian noise, color jitter, random cropping, and spatial transformations. This consistency regularization encourages the model to produce stable and reliable segmentation outputs even without explicit labels.

By combining labeled and unlabeled data in this way, the semi-supervised framework improves the model's ability to capture fine coral structures and adapt to variations in underwater conditions, such as illumination changes, turbidity, and color attenuation. As a result, segmentation accuracy and robustness are enhanced, while the dependence on large-scale manual annotation is substantially reduced [17].

2.4 Experimental environment

All training and evaluation were performed on an Ubuntu 22.04 LTS operating system. The experiments were performed on a high-performance deep learning workstation equipped with an AMD Ryzen™ 9 7950X CPU and an NVIDIA RTX A6000 GPU (48 GB).

3 Results and discussions

3.1 Analysis of dataset expansion using Style GAN3

To assess the visual quality and realism of the StyleGAN3-generated coral images, we employ three widely used quantitative metrics: Peak Signal-to-Noise Ratio (PSNR) [24], Fréchet Inception Distance (FID) [25], Kernel Inception Distance (KID) [26], and Structural Similarity Index (SSIM) [24].

PSNR measures the pixel-level similarity between a generated image and its corresponding reference image, and is commonly used to evaluate reconstruction fidelity. A higher PSNR indicates lower distortion and closer agreement with the ground truth. In the context of coral image synthesis, PSNR is particularly relevant for assessing the preservation of fine structural details, including branch geometry, surface texture, and subtle color gradients. The metric therefore reflects the generator's ability to maintain local structural integrity while producing visually plausible underwater coral scenes. The formulas are given by:

$$MSE = \frac{1}{\mathrm{n}}\sum_{i=1}^{n}(I_{SR}^{i} - I_{HR}^{i})^2 \quad (1)$$

$$PSNR = 10 \cdot \log_{10}(\frac{MAX_I^2}{MSE}) \quad (2)$$

Here, n denotes the complete number of pixels, $\mathrm{I_{HR}^{i}}$ and $\mathrm{I_{SR}^{i}}$ refer to the intensity of the $\mathrm{i-th}$ pixel in HR and SR images. For 8-bit images, the $\mathrm{MAX_I}$ is typically set to 255. Higher PSNR values indicate better reconstruction quality, as they correspond to lower distortion and closer pixel-wise similarity to the reference images.

FID evaluates the perceptual quality of generated images by comparing the feature distributions of real and synthetic images, as extracted by a pre-trained Inception network. A lower FID value means that the generated samples are more realistic and their distribution is more consistent with that of real images. KID is conceptually similar to FID but is computed using the squared Maximum Mean Discrepancy with polynomial kernels. Unlike FID, KID provides an unbiased estimate even when the number of samples is limited. This makes it particularly suitable for underwater coral datasets, where high-quality annotated images are relatively scarce. SSIM is used to measure the structural fidelity of GAN-generated coral images. By jointly considering luminance, contrast, and structural information, SSIM offers a perceptual evaluation that aligns more closely with human visual judgment than simple pixel-wise error metrics. Compared with the baseline model, the proposed StyleGAN3-based framework achieves higher SSIM values, indicating better preservation of coral branch geometry, surface texture, and local contrast. These results suggest that the synthesized images more closely approximate real underwater coral scenes.

As shown in **Table 1**, several generative models are compared on the task of underwater coral image synthesis. For pixel-level reconstruction, PSNR and SSIM are metrics where higher values indicate better quality. DDPM achieves the highest PSNR (28.89 dB) and SSIM (81.38%), demonstrating strong ability to preserve coral branch textures, fine structural details, and overall structural consistency. StyleGAN3 ranks second, with a PSNR of 27.49 dB and an SSIM of 79.08% and clearly outperforms traditional GAN variants and the super-resolution model ESRGAN.

From the perspective of perceptual quality and distribution alignment, FID and KID are metrics where lower values are better. DDPM obtains FID and KID scores of 22.78 and 0.0077, indicating that its generated images are closest to real coral images in the high-level feature space, with strong visual realism and diversity. StyleGAN3 achieves FID and KID values of 20.96 and 0.0079, which also reflect good distribution matching and a clear advantage over earlier GAN-based methods in terms of perceptual quality.

Table 1. Metrics of generated images for different models

Methods	PSNR (dB)	SSIM (%)	FID	KID	Parameter (MB)
DCGAN	19.45	71.45	28.93	0.0621	3.5
WGAN-GP	20.56	72.65	29.64	0.0485	15.4
CycleGAN	22.78	75.69	26.07	0.0383	11.4
CGAN	22.62	73.35	25.17	0.0184	54.4
StyleGAN 2	24.06	78.01	24.65	0.0284	30.0

ESRGAN	24.64	77.96	23.25	0.0186	27.6
DDPM	28.89	81.38	22.78	0.0077	87.2
StyleGAN3	**27.49**	**79.08**	**20.96**	**0.0079**	**16.7**

However, these gains must be considered together with model complexity. The parameter size of DDPM is 87.2 MB, far exceeding that of StyleGAN3 (16.7 MB) and the other models. Although DDPM is slightly superior to StyleGAN3 on most quality metrics, it requires significantly more computation and storage, making it less suitable for resource-limited underwater monitoring platforms or edge devices. In contrast, StyleGAN3 maintains high PSNR and SSIM and achieves FID and KID scores close to those of DDPM, while using only about one-fifth of its parameter size. This yields a more favorable trade-off between performance and complexity.

Overall, when generation quality, model complexity, and deployment constraints are considered together, StyleGAN3 is the preferred model for underwater coral image generation in this study. DDPM can be viewed as an alternative choice in scenarios where computational resources are sufficient and the highest possible generation quality is the primary objective.

3.2 Analysis of semi-supervised segmentation results

To assess the semi-supervised segmentation results, we employ two evaluation metrics: IoU (Intersection over Union) [27], and FPS (Frames Per Second) [28]. The formulas are as follows:

$$IoU = \frac{TP}{TP + FN + FP} \tag{3}$$

where True Positive (TP), False Negative (FN), and False Positive (FP) serve as foundational indicators of performance. The FPS metric measures the model's inference speed and is defined as:

$$FPS = \frac{N}{T} \tag{4}$$

where N represents the amount of test images processed, and T is the total inference time. A higher FPS indicates better real-time performance.

To further evaluate the proposed semi-supervised segmentation model for underwater coral recognition, Precision and Recall [29] were employed as key metrics, defined as:

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

TP, FP, and FN denote true positives, false positives, and false negatives, individually. Experimental results indicate that the semi-supervised model outperforms the fully supervised baseline in both Precision and Recall, efficiently reducing inaccurate positives in complex backgrounds while enhancing the detection of small and low-contrast coral structures.

Table 2 presents the performance comparison of different segmentation algorithms. In this experiment, all models were fine-tuned on the Coral dataset and evaluated on its test set.

The results show a clear performance hierarchy. Traditional CNN-based models, such as FPN and U-Net, provide a reasonable baseline, with mIoU scores of 71.28% and 72.54%, respectively. However, they are surpassed by more recent transformer-based architectures. Among the supervised methods, SegFormer achieves the best performance with an mIoU of 84.65%, highlighting its stronger capability to model long-range context and handle complex underwater coral scenes.

Table 2. Comparison of Segmentation Methods

Methods	MIoU	F1	Precision	Recall	FPS
FPN	71.28	80.49	80.96	81.41	145
U-Net	72.54	81.64	81.78	83.57	151
MANet	73.87	81.67	83.87	83.53	96
HRNet	75.72	83.32	84.95	84.14	57
SFPN	78.42	84.15	85.14	84.25	120
Mask2 Former	80.65	85.87	86.45	86.69	39
DPT	82.02	86.98	87.25	87.96	48
Segformer	**84.65**	**89.01**	**88.69**	**89.36**	**59**

3.3 Analysis of the effectiveness of semi-supervised segmentation

Table 3 highlights these advantages. The fully supervised baseline C1 achieves 84.65% mIoU and 89.01% F1. When the same labeled budget is combined with 1,000 unlabeled images (C2-1k), performance already surpasses the baseline, with gains of +1.47% mIoU and +1.90% F1. As the amount of unlabeled data increases, improvements become more substantial—for example, C2-3k yields roughly +6.4% mIoU and +5.2% F1 over C1. The best configuration, C2-5k, reaches 92.97% mIoU and 95.57% F1, representing significant boosts in Precision and Recall. These findings confirm that leveraging large-scale unlabeled data effectively enhances feature learning and segmentation robustness.

In the semi-supervised segmentation framework, the quality of the teacher model and its pseudo-labels directly influence the learning of the student model. Our experiments show a consistent performance gain as the teacher model is trained with more data. Since pseudo-labels set the performance ceiling of semi-supervised learning, expanding the teacher dataset improves accuracy and strengthens generalization. The inclusion of diverse GAN-generated coral images further enriches visual variability, enabling the teacher model to learn more discriminative features and produce more reliable pseudo-labels.

4 Conclusions

This study proposes an underwater coral image recognition framework that integrates StyleGAN3-based data augmentation, super-resolution reconstruction, and semi-supervised segmentation. By expanding the training set with realistic synthetic coral images, enhancing degraded inputs, and leveraging unlabeled data through a teacher–student strategy, the framework significantly improves segmentation accuracy and robustness under limited annotations. Experiments on the coral dataset demonstrate superior accuracy and generalization compared with existing segmentation models. Future work will extend the framework to more coral species, incorporate diverse environmental conditions, and explore real-time deployment for long-term coral reef monitoring and conservation.

Table 3. Evaluation of semi-supervised algorithms

Model	MIoU	F1	Precision	Recall
C1	84.65	89.01	88.69	89.36
C2-1k	86.12	90.91	89.89	90.87
C2-2k	89.36	92.05	91.52	92.71
C2-3k	91.05	94.21	93.14	94.64
C2-4k	92.14	95.09	94.78	95.88
C2-5k	92.97	95.57	95.12	96.24

References

1. Zhang, G., Zhao, F., Wang, X., et al. (2025). Deep learning-driven super-resolution and DenseCLIP-based semantic segmentation for coral reef mapping. In OCEANS 2025 Brest (pp. 1–4). IEEE.
2. Chiang, J. Y., & Chen, Y. C. (2012). Underwater image enhancement by wavelength compensation and dehazing. IEEE Transactions on Image Processing, 21(4), 1756–1769.
3. Shao, X., Chen, H., Zhao, F., et al. (2026). Multi-label classification for multi-temporal, multi-spatial coral reef condition monitoring using vision foundation model with adapter learning. Marine Pollution Bulletin, 223, 119054.
4. Zhao, F., He, Y., Song, J., et al. (2025). Smart UAV-assisted blueberry maturity monitoring with Mamba-based computer vision. Precision Agriculture, 26(4), 56.
5. Chen, J., Li, P., Lee, J., et al. (2025). Optimization of oblique drone photogrammetry for avoiding sun glint in submerged seagrass mapping. Estuarine, Coastal and Shelf Science, 109356.
6. Zhao, F., Shao, X., Wang, J., et al. (2025). A novel underwater Holothurians monitoring system using consumer-grade amphibious UAV with Mamba-based super-resolution reconstruction and enhanced YOLOv10. Marine Environmental Research, 107510.
7. González-Rivero, M., et al. (2020). Monitoring of coral reefs using artificial intelligence: A feasible and scalable approach. Remote Sensing in Ecology and Conservation, 6(4), 472–482.
8. Li, C., Guo, J., Guo, C., & Cong, R. (2019). WaterNet: An adaptive deep network for underwater image enhancement. IEEE Transactions on Image Processing, 28(11), 5580–5595.
9. Li, C., Anwar, S., & Porikli, F. (2021). Underwater scene prior inspired deep underwater image and video enhancement. Pattern Recognition, 114, 107858.
10. Li, C., Guo, C., Ren, W., Cong, R., & Hou, J. (2020). Underwater image enhancement via medium transmission-guided multi-color space embedding. IEEE Transactions on Image Processing, 29, 9380–9395.
11. Goodfellow, I., Pouget-Abadie, J., Mirza, S., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems (NeurIPS), 27.
12. Li, M., Zhang, H., Gruen, A., et al. (2025). A survey on underwater coral image segmentation based on deep learning. Geo-spatial Information Science, 28(2), 472–496.
13. Mittal, S., Tatarchenko, M., & Brox, T. (2021). Semi-supervised semantic segmentation with high- and low-level consistency. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 43(12), 4311–4324.
14. Oliver, A., et al. (2018). Realistic evaluation of deep semi-supervised learning algorithms. Advances in Neural Information Processing Systems (NeurIPS), 31.
15. Raine, S., Marchant, R., Kusy, B., Maire, F., & Fischer, T. (2022). Point label aware superpixels for multi-species segmentation of underwater imagery. arXiv preprint, arXiv:2201.09997.
16. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 (pp. 234–241).
17. Sohn, K., Berthelot, D., Li, C. L., Zhang, Z., Carlini, N., Cubuk, E. D., Kurakin, A., Zhang, H., & Raffel, C. (2020). FixMatch: Simplifying semi-supervised learning with consistency and confidence. Advances in Neural Information Processing Systems (NeurIPS), 33.
18. Tarvainen, A., & Valpola, H. (2017). Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in Neural Information Processing Systems (NeurIPS), 30, 1195–1204.
19. Valdebenito Maturana, C. N., Sandoval Orozco, A. L., & García Villalba, L. J. (2023). Exploration of metrics and datasets to assess the fidelity of images generated by generative adversarial networks. Applied Sciences, 13(3), 1842.
20. Wang, J., Mizuno, K., Tabeta, S., et al. (2025). Multi-dataset-integrated Coral-Lab segmentation with enhanced towed camera array for rapid large-scale coral reef monitoring and mapping. International

Journal of Applied Earth Observation and Geoinformation, 143, 104819.

21. Williams, I. D., Couch, C. S., Beijbom, O., et al. (2019). Leveraging automated image analysis to monitor coral reef ecosystems. Environmental Monitoring and Assessment, 191(8), 466.
22. Yu, X., Ouyang, B., & Principe, J. C. (2021). Coral image segmentation with point-supervision via latent Dirichlet allocation with spatial coherence. Journal of Marine Science and Engineering, 9(2), 157.
23. Alaluf, Y., Patashnik, O., Wu, Z., Zamir, A., Shechtman, E., Lischinski, D., & Cohen-Or, D. (2022). Third time's the charm? Image and video editing with StyleGAN3. In European Conference on Computer Vision (ECCV) (pp. 204–220). Springer Nature Switzerland.
24. Hore, A., & Ziou, D. (2010). Image quality metrics: PSNR vs. SSIM. In 2010 20th International Conference on Pattern Recognition (pp. 2366–2369). IEEE.
25. Yu, Y., Zhang, W., & Deng, Y. (2021). Frechet inception distance (FID) for evaluating GANs. China University of Mining Technology Beijing Graduate School, 3(11).
26. Bińkowski, M., Sutherland, D. J., Arbel, M., & Gretton, A. (2018). Demystifying MMD GANs. arXiv preprint, arXiv:1801.01401.
27. Rezatofighi, H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 658–666).
28. Benenson, R., Mathias, M., Timofte, R., & Van Gool, L. (2012). Pedestrian detection at 100 frames per second. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (pp. 2903–2910). IEEE.
29. Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., & Aila, T. (2019). Improved precision and recall metric for assessing generative models. Advances in Neural Information Processing Systems (NeurIPS), 32.