Revista Eletrônica de Iniciação Científica em Computação, 2026, 24:1
ISSN: 3085-8461 • doi: 10.5753/reic.2026.8511

Research Paper

Feature-Centric Underwater Vision Models

Emanuel da Costa Silva [Universidade Federal do Rio Grande | *emanuel_silva@furg.br*]
Tatiana Taís Schein [Universidade Federal do Rio Grande |*tatischein@furg.br*]
Paulo Lilles Jorge Drews Junior [Universidade Federal do Rio Grande |*paulodrews@furg.br*]

Computer Science Center (C3), Universidade Federal do Rio Grande,Campus Carreiros - Av. Itália Km 8, bairro Carreiros, Rio Grande, RS, 96203-900, Brazil.

Abstract. Underwater image enhancement (UIE) models are often optimized for human visual perception, which does not necessarily translate to improved performance in automated vision tasks. This paper describes a consolidated research journey consisting of the AquaFeat ecosystem: a series of plug-and-play modules designed to enhance hierarchical features for downstream robotics tasks. We present the transition from AquaFeat, a detection-focused module, to AquaFeat+, which introduces global-scale attention for classification and tracking, and finally C-Feat. C-Feat integrates these feature-centric backbones into a Compact Model framework, achieving a remarkable inference speed of 280 FPS while maintaining state-of-the-art performance in complex underwater environments.

Keywords: Underwater Computer Vision; Feature Enhancement; Object Detection; Object Classification; Object Tracking; Image Restoration; Autonomous Underwater Vehicles

Received: 17 June 2026 • **Accepted:** 17 June 2026 • **Published:** 10 July 2026

1 Introduction

Autonomous underwater robots rely on visual perception for critical operations such as species identification and infrastructure inspection Dawkins *et al*. [2024]. However, the underwater environment presents severe challenges, including color distortion and light scattering Schein *et al*. [2024]. While deep learning-based enhancement has shown promise, many models focus on visual aesthetics that may obscure key features for computer vision Nathan *et al*. [2024].

This work consolidates a three-stage undergraduate research project aimed at bridging the gap between enhancement quality and real-time robotic efficiency. The progression moves from modular enhancement (AquaFeat) Silva *et al*. [2025b] to multi-task versatility (AquaFeat+) Silva *et al*. [2025a] and culminates in high-speed generative efficiency (C-Feat) Silva *et al*. [2026]. The source code for AquaFeat and Aquafeat+ are available **here**, C-Feat's code is available **here**.

2 The Evolution of the AquaFeat Ecosystem

The core philosophy of this research is that enhancement should be explicitly guided by the requirements of the downstream task rather than generic metrics Hashmi *et al*. [2023].

2.1 Phase I: Task-Driven Modular Enhancement (AquaFeat)

The initial stage addressed the fidelity-efficiency trade-off. Traditional Visual-based UIE methods often prioritize human aesthetics Hashmi *et al*. [2023]. However, our findings demonstrated that visually appealing images do not always correlate with improved detection metrics if key features are distorted. AquaFeat (**Figure 1**) was thus developed as a plug-and-play module specifically designed to optimize features for machine perception.

2.1.1 Preprocessing and Color Correction

Before feature extraction, the input image undergoes a non-trainable color correction step Liu *et al*. [2023]. This module analyzes channel intensities to stabilize the input by mitigating chromatic distortions, allowing the network to focus on high-level features rather than low-level noise.

2.1.2 Hierarchical Underwater-Feature Enhancement Network (U-FEN)

The U-FEN core uses a shared-weight architecture across three resolutions (original, quarter I_q, and one-eighth I_o), combining six standard convolutions with an initial SpecialConv layer that introduces a content-aware mechanism to adjust contrast dynamically by computing local statistics (μ and σ) to apply adaptive feature scaling. Within the primary full-resolution pathway, all convolutional operations, including this initial SpecialConv, utilize 3×3 kernels with a stride of 1 and padding of 1, strictly preserving the native spatial dimensions ($H \times W$) of the input image. Spatial downsampling is restricted exclusively to the parallel multi-scale pipelines (I_q and I_o).

2.1.3 Scale-Aware Fusion and Residual Learning

To integrate these multi-scale features, we utilize the SAFA module Hashmi *et al*. [2023]. SAFA projects feature maps into a common embedding space, employing a multi-head attention mechanism to calculate normalized weights for a contextually rich representation. To resolve the spatial discrepancies introduced by the downsampled pathways, the aggregated lower-resolution features are upsampled back to the native dimensions ($H \times W$) via bilinear interpolation. This combined full-resolution tensor is then processed by a final SpecialConv layer and a $tanh$ activation function to generate the Adaptive Residual Output. Because spatial boundaries are preserved along the primary path and restored for the multi-scale paths, the output of this final SpecialConv layer matches the exact dimensions of the input image, ensuring

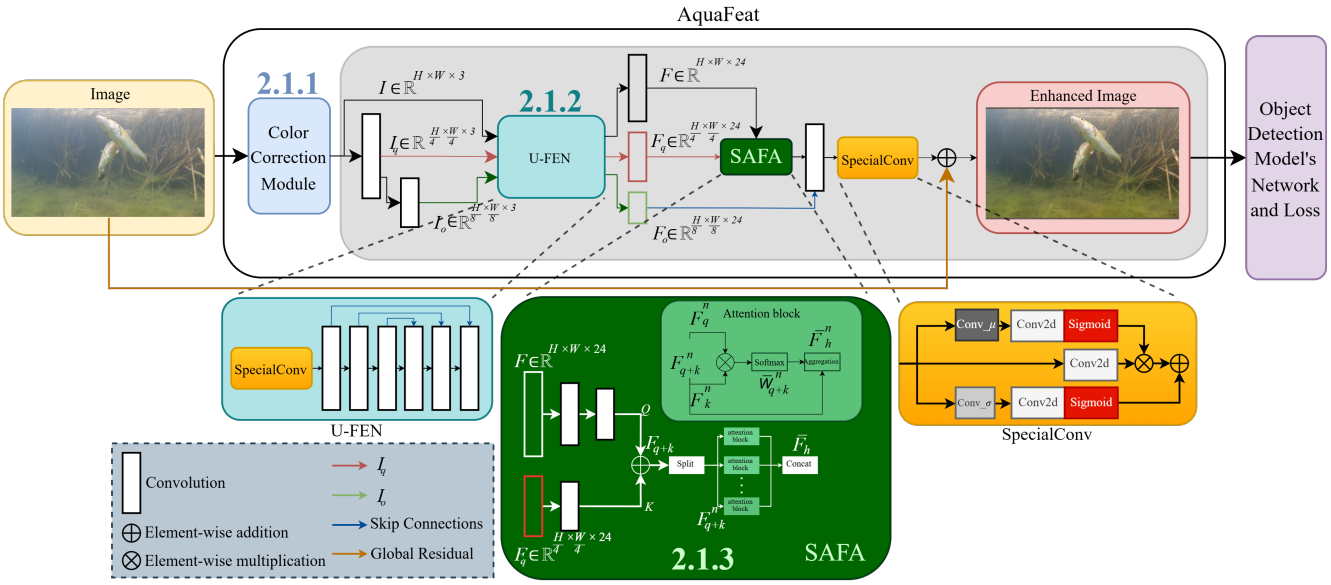

Figure 1. Overview of the AquaFeat architecture Silva *et al*. [2025b]. The pipeline begins with the color correction module, followed by the U-FEN for feature extraction. The SAFA module merges these features to produce an enhanced image for YOLOv8m Ultralytics [2023] Silva *et al*. [2025b].

mathematical consistency during the element-wise addition.

2.2 Phase II: Multi-Task Versatility (AquaFeat+)

AquaFeat+ (**Figure 2**) transitioned to a comprehensive underwater vision suite supporting detection, classification, and tracking. This phase addressed robust feature representations in dynamic videos where targets suffer from long-range dependency challenges Hashmi *et al*. [2023].

2.2.1 *The Global-Scale Attention Module (GSAM)*

The technical advancement of AquaFeat+ is the **GSAM** Silva *et al*. [2025a]. While the original SAFA was efficient at combining different sized features Hashmi *et al*. [2023], GSAM concurrently performs intra-scale refinement and inter-scale fusion through a dual-pathway design, and can be defined in two main stages:

- **Global Feature-Aware (GFA)** Zhou *et al*. [2024]: This module models long-range spatial dependencies. By generating attention maps, it amplifies salient regions, which helps identifying small targets in turbid waters.
- **Scale-Aware Feature Aggregation (SAFA)**: This component remains the same as described in Subsection 2.1.3, performing inter-scale fusion.

2.3 Phase III: Compact Enhancement Efficiency (C-Feat)

The final stage shifts focus toward image enhancement for human perception, targeting maximum efficiency in both training and inference stages. Most state-of-the-art Underwater Image Enhancement methods struggle with the trade-off between efficiency and enhancement quality, often relying on scarce paired datasets. Our goal with C-Feat was to address high-fidelity enhancement without the need for ground-truth pairs, minimizing this gap through a self-supervised strategy. This is achieved by leveraging the compact feature-centric backbone developed in prior phases and further upgraded in this work.

C-Feat (**Figure 3**) works using three images: the original, a low-light version of it Schein *et al*. [2024] and a version enhanced by the previous Color Correction Module Liu *et al*. [2023]. The low-light version is fed to the U-FEN and GSAM, then, the result and original images are processed by the Gated Restoration Head (GRH), which produces the resulting enhanced image. The color-corrected version serves as a target to optimize model weights via specific loss functions. This approach leverages a pseudo-target, thereby eliminating the need for manually labeled ground-truth pairs in training.

AquaFeat models do not include a loss function of their own, as they are plugged into different backbones and utilize their native objectives. Differently, C-Feat employs a task-specific composite loss $\mathcal{L}_{\text{total}} = W_{L1}\mathcal{L}_{\text{Huber}} + W_{per}\mathcal{L}_{\text{LPIPS}} + W_{Lab}\mathcal{L}_{\text{Lab}}$ to jointly enforce structural fidelity, perceptual quality, and chromatic consistency. We employ a Pseudo-Huber loss ($W_{L1} = 1.0$) for luminance, AlexNet-backed LPIPS ($W_{per} = 0.82$) for detail, and Chromaticity-Consistent loss ($W_{Lab} = 1.1$) for color cast correction.

2.3.1 *The Gated Restoration Head (GRH)*

The main technical innovation is the **Gated Restoration Head (GRH)**, which bridges physical formation models with data-driven refinement. The GRH processes aggregated features through two parallel paradigms:

- **Physics-Guided Dehazing Path**: This pathway recovers visibility by explicitly estimating the Atmospheric Scattering Model (ASM) parameters. It predicts a transmission map $t(x)$ and global atmospheric light A to perform restoration:

$$J_{phy} = \frac{I - A}{t(x) + \epsilon} + A, \tag{1}$$

  where I is the input and ϵ is a stability constant.
- **Residual Enhancement Path**: To address complex non-linear color distortions that physical models often fail to capture, a parallel head learns a data-driven residual map R, generating restoration $J_{res} = I + R$.

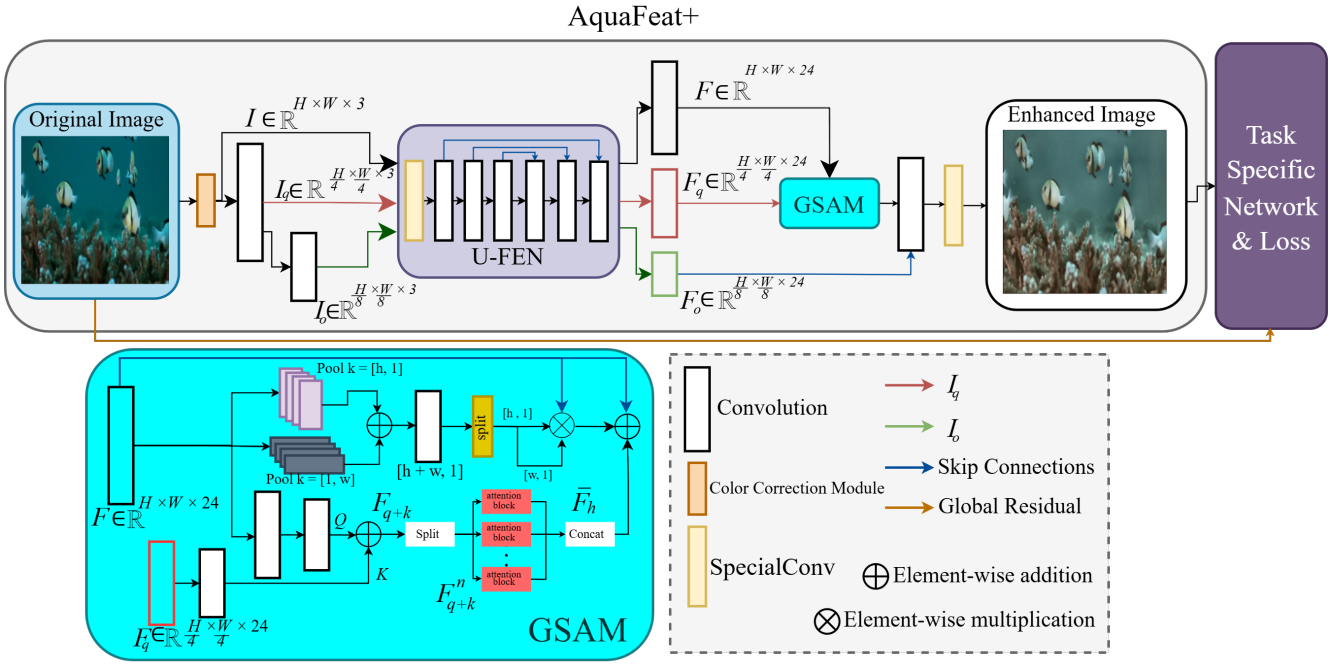

Figure 2. Overview of the AquaFeat+ architecture Silva *et al*. [2025a]. The pipeline begins with the color correction module, followed by the U-FEN for feature extraction. The Global-Scale Attention Module (GSAM) aggregates global features and merges them with those from the original 1/8th-sized image to produce an enhanced image, which is then fed into the chosen task backbone Silva *et al*. [2025a].

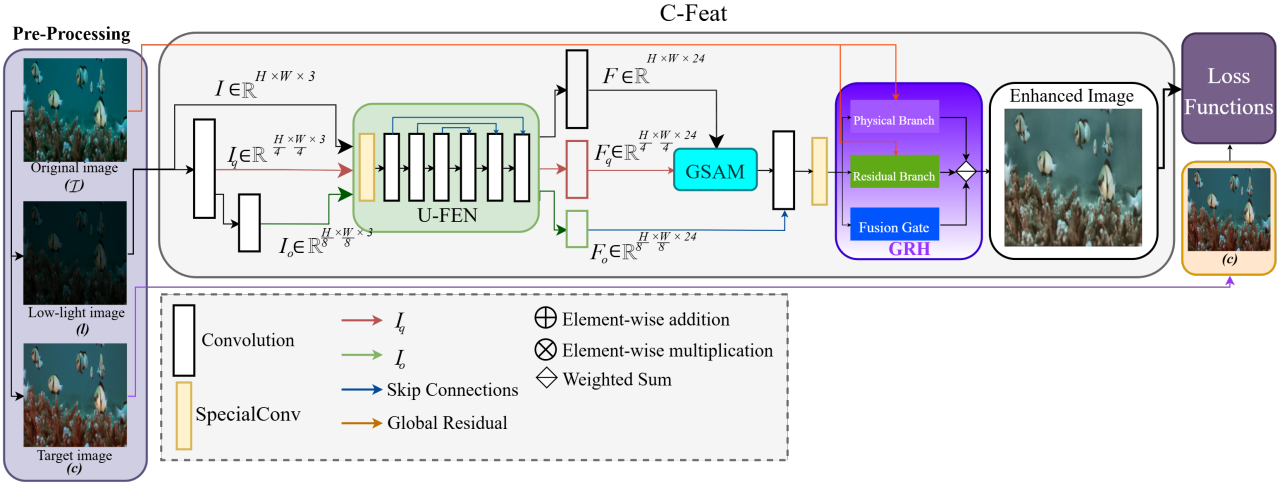

Figure 3. Overview of the C-Feat architecture Silva *et al*. [2026]. The pipeline begins with the U-FEN for feature extraction. The GSAM then aggregates global features and merges them with those from the original 1/8th-sized image to produce an enhanced image, that is once again enhanced through a SpecialConv, the result is fed into the GRH for final dehazing and adjustments Silva *et al*. [2026].

An **Uncertainty-Aware Gating** mechanism predicts a pixel-wise weight map $G \in [0, 1]$ to dynamically fuse these paths:

$$\hat{J} = G \odot J_{phy} + (1 - G) \odot J_{res}, \tag{2}$$

where $\odot$ denotes the Hadamard product (element-wise multiplication). This ensures that the gating weights are applied independently to each pixel across the spatial dimensions of the feature maps.

This allows C-Feat to adaptively rely on physical dehazing for turbid regions while preserving fine textures through residual learning.

3 Experimental Results and Impact

The proposed ecosystem was evaluated across benchmarks to validate its effectiveness in real-world underwater robotics scenarios.

3.1 Experimental Setup and Datasets

Validation utilized the **FishTrack23** dataset Dawkins *et al*. [2024] for detection, classification, and tracking. The first AquaFeat (Phase I) dataset was composed of $6,392$ images split into $4,474$ for training, $1,278$ for testing, and 640 for validation. AquaFeat+ (Phase II) required tracking and classification, thus, a new sample of the FishTrack23 was developed, the dataset resulting from this stage totaled $5,149$ images for

training, 1,098 for validation, and $14,575$ sequential images for testing. For image enhancement (Phase III), we employed the **UIEB** Li *et al*. [2020], **SUIM** Islam *et al*. [2020] and **NUID** Hou *et al*. [2023] dataset. Experiments were conducted on NVIDIA GeForce RTX 4070 Ti and RTX 5090 GPUs.

3.2 Quantitative Multi-Task Performance

As shown in **Tables 1–2**, the evolution from AquaFeat to AquaFeat+ demonstrates a consistent balance across multiple vision tasks, including detection, classification, and tracking. In object detection, AquaFeat+ maintains competitive mAP and F1 scores while improving overall stability compared to the original AquaFeat configuration. For classification, AquaFeat+ achieves the best overall performance, obtaining the highest precision (0.816), recall (0.791), and F1-score (0.791). In tracking, both variants remain highly competitive, with AquaFeat+ (v10s) reaching the best HOTA (55.20) and association accuracy (AssA = 60.19), indicating improved identity consistency across frames.

Table 1. Object Detection Metrics comparison on AquaFeat and AquaFeat+ datasets.

(a) AquaFeat Results

Method	$\text{mAP}_{0.50}\uparrow$	$\text{mAP}_{0.50:0.95}\uparrow$	Prec.↑	Rec.↑	FPS↑
YOLOv8m Ultralytics [2023]	0.647	0.387	0.847	0.584	321.54
YOLOv10 Wang *et al*. [2024]	0.592	0.325	0.777	0.549	**700.18**
FeatEnhancer Hashmi *et al*. [2023]	0.649	0.384	0.838	0.593	60.24
AMSP-UOD Zhou *et al*. [2024]	**0.724**	**0.46**	0.866	0.578	41.84
Osmosis Nathan *et al*. [2024]	0.023	0.005	0.133	0.005	0.0037
UDNET Saleh *et al*. [2022]	0.566	0.336	0.810	0.505	15.92
UDBE Schein *et al*. [2024]	0.379	0.195	0.691	0.345	0.14
AquaFeat (v8m)	0.677	0.421	**0.877**	**0.624**	46.51
AquaFeat (v10s)	0.676	0.421	0.859	0.621	49.24

(b) AquaFeat+ Results

Method	Prec.↑	Rec.↑	F1↑	mAP50↑	mAP95↑
YOLOv8m	**0.792**	0.582	0.677	0.528	0.319
YOLOv10s	0.787	0.542	0.642	0.486	0.277
FeatEnHancer	0.753	0.582	0.657	0.515	0.293
AquaFeat (v8m)	0.746	0.624	0.680	0.554	0.332
AquaFeat (v10s)	0.754	**0.625**	0.684	**0.557**	**0.336**
AquaFeat+ (v8m)	0.767	0.624	**0.688**	0.556	0.332
AquaFeat+ (v10s)	0.770	0.619	0.687	0.553	0.333

Table 2. Classification and Tracking Results comparison on AquaFeat+.

(a) Classification Results

Method	Prec.↑	Rec.↑	Acc.↑	F1↑
YOLOv11s-cls	0.723	0.764	0.764	0.737
FeatEnHancer	0.746	0.779	0.779	0.752
ResNeXt Xie *et al*. [2017]	0.681	0.605	0.834	0.618
ConvNeXt Liu *et al*. [2022b]	0.716	0.619	**0.862**	0.646
WildFish Zhuang *et al*. [2018]	0.642	0.448	0.773	0.440
AquaFeat	0.798	0.765	0.765	0.766
AquaFeat+	**0.816**	**0.791**	0.791	**0.791**

(b) Tracking Results

Method	HOTA↑	MOTA↑	DetA↑	AssA↑	IDF1↑
YOLOv8m	52.75	53.78	**51.42**	54.40	65.09
YOLOv10s	49.15	51.80	47.32	51.46	64.22
FeatEnHancer (v8m)	47.47	37.22	41.14	54.96	59.42
FeatEnHancer (v10s)	49.13	44.19	45.11	53.68	60.84
AquaFeat (v8m)	54.71	**55.76**	50.93	59.14	**68.41**
AquaFeat (v10s)	52.14	55.20	50.48	54.09	63.35
AquaFeat+ (v8m)	54.19	54.97	50.10	58.90	67.63
AquaFeat+ (v10s)	**55.20**	55.01	50.89	**60.19**	68.09

3.3 Enhancement Fidelity and Temporal Efficiency

The introduction of C-Feat addressed high-fidelity enhancement without the need for paired datasets Silva *et al*. [2026].

The quantitative results across multiple datasets, summarized in **Tables 3 and 4**, highlight C-Feat's effectiveness in balancing visual fidelity with computational efficiency. Regarding fidelity, C-Feat leads the unsupervised/self-supervised category in no-reference metrics, achieving a UIQM of 3.28 and a BRISQUE score of 53.61. This performance is coupled with a significant speed breakthrough, reaching a record 281.45 FPS, which ensures suitability for real-time robotic deployment. The substantial increase in efficiency compared to previous phases is attributed to C-Feat operating as a standalone, compact model, whereas AquaFeat+ metrics include the computational overhead of an integrated YOLO detector. Furthermore, the model exhibits low training latency, completing its full training cycle in under one hour with an average epoch time of 5.01s.

3.4 Qualitative Results

In the qualitative analysis, **Figure 4** illustrates the comparison of the ground truth, AMSP-UOD, OSMOSIS, UDBE and AquaFeat. Our model consistently improved detection by enhancing image visibility and enabling more precise object localization. Most notably, as demonstrated in the scenario of the last column, AquaFeat was the only method capable of correctly detecting a partially occluded object at the image border, a task that all other techniques failed to accomplish.

Figure 5 compares the ground truth, YOLOv8m, FeatEnHancer, AquaFeat and AquaFeat+ in detection. AquaFeat+ was the only model to detect all objects in the three scenarios, with the previous AquaFeat being right behind as the second best result. In classification, **Figure 6**, the results were similar, AquaFeat+ was the best one correctly identifying $4/6$ fishes, while other models failed to do that.

Figure 7 compares C-Feat against several models on the UIEB dataset. Our approach effectively mitigates haze and color casts, preserving structural detail without artifacts. Conversely, UDBE introduces noise, while supervised methods like U-Shape exhibit grayish tones and residual haze. Regarding color, OSMOSIS shows a blue bias, and LANet displays inconsistent red tones and local shadowing. While FiveAPlus offers decent restoration, it fails to brighten key regions. Overall, C-Feat achieves superior brightness and color consistency across all scenarios.

3.5 Research Impact and Scientific Contributions

The modular AquaFeat was published at **SIBGRAPI 2025** (h5-index: 22) Silva *et al*. [2025b], its successor AquaFeat+ was published at **ICAR 2025** (h5-index: 18) Silva *et al*. [2025a], and C-Feat was accepted at **ICPR 2026** (h5-index: 68) Silva *et al*. [2026].

4 Conclusion

This work presents the AquaFeat ecosystem, demonstrating that task-driven, feature-centric enhancement is more effective for underwater robotics. Through modular evolution, we

Table 3. Full and non-reference image quality on UIEB. Higher ↑ is better, lower ↓ is better.

	Method	Venue&Year	UIQM ↑	UISM ↑	UICONM ↑	BRISQUE ↓	CLIP-IQA ↑	PSNR ↑	SSIM ↑	T/Ep ↓	Ep ↓	FPS ↑
Un/Self-sup.	OSMOSIS Nathan *et al.* [2024]	ECCV '24	2.9376	6.5336	0.2088	107.36	0.5097	12.72	0.235	–	–	0.002
	UESAM Mello *et al.* [2022]	C&G '22	2.9243	**7.2268**	0.1785	71.6013	0.5247	20.13	0.767	40.78	200	58.48
	UDBE Schein *et al.* [2024]	ICMLA '24	2.1383	5.0806	0.1285	66.7676	0.5191	11.95	0.286	66.94	1000	0.09
	USUIR Fu *et al.* [2022a]	AAAI '22	2.5608	5.4566	0.2238	58.0498	0.5438	17.31	0.764	4.21	**50**	23.98
	C-Feat	ICPR '26	3.2811	7.0401	0.3111	**53.6175**	**0.5617**	17.16	0.706	5.01	**50**	**281.45**
Supervised	U-shape Peng *et al.* [2023]	TIP '23	**3.3096**	6.9012	0.3162	62.5456	0.5569	17.33	0.733	–	–	*49.15*
	FiveAPlus Jiang *et al.* [2023]	BMVC '23	2.9151	5.4935	**0.3193**	87.9160	0.5379	21.60	0.882	**4.12**	**50**	38.93
	LANet Liu *et al.* [2022a]	RA-L '22	3.0852	6.9255	0.2405	64.6383	0.5457	18.47	0.705	149.58	**50**	20.07
	PUIE NET Fu *et al.* [2022b]	ECCV '22	2.9748	5.9249	0.3007	83.4976	0.5433	22.44	**0.899**	8.41	500	2.00
	Phaseformer Khan *et al.* [2025]	WACV '25	2.9547	5.7699	0.3072	75.5437	0.5482	**22.66**	0.886	61.37	400	8.34

Table 4. Full and non-reference image quality on SUIM (left) and NUID (right).

Method	UIQM ↑	UISM ↑	UICONM ↑	BRISQUE ↓	CLIP-IQA ↑
OSMOSIS	2.5708	5.5326	0.1716	110.9335	0.4897
UESAM	2.8748	6.6696	0.1963	64.0301	0.5252
UDBE	2.3912	5.8147	0.1321	69.4953	0.5085
USUIR	2.4995	4.7089	0.2348	87.3981	0.5064
C-Feat	**3.0022**	**6.8385**	**0.2432**	**54.7392**	**0.5550**

Method	UIQM ↑	UISM ↑	UICONM ↑	BRISQUE ↓	CLIP-IQA ↑
OSMOSIS	2.3647	5.1275	0.1873	129.2642	0.5189
UESAM	2.0137	5.0330	0.1191	84.0558	0.5426
UDBE	2.1434	4.6485	0.1672	88.0729	0.5333
USUIR	**2.4504**	4.3982	**0.2751**	98.8329	0.5280
C-Feat	2.3629	**5.1352**	0.2163	**79.1992**	**0.5585**

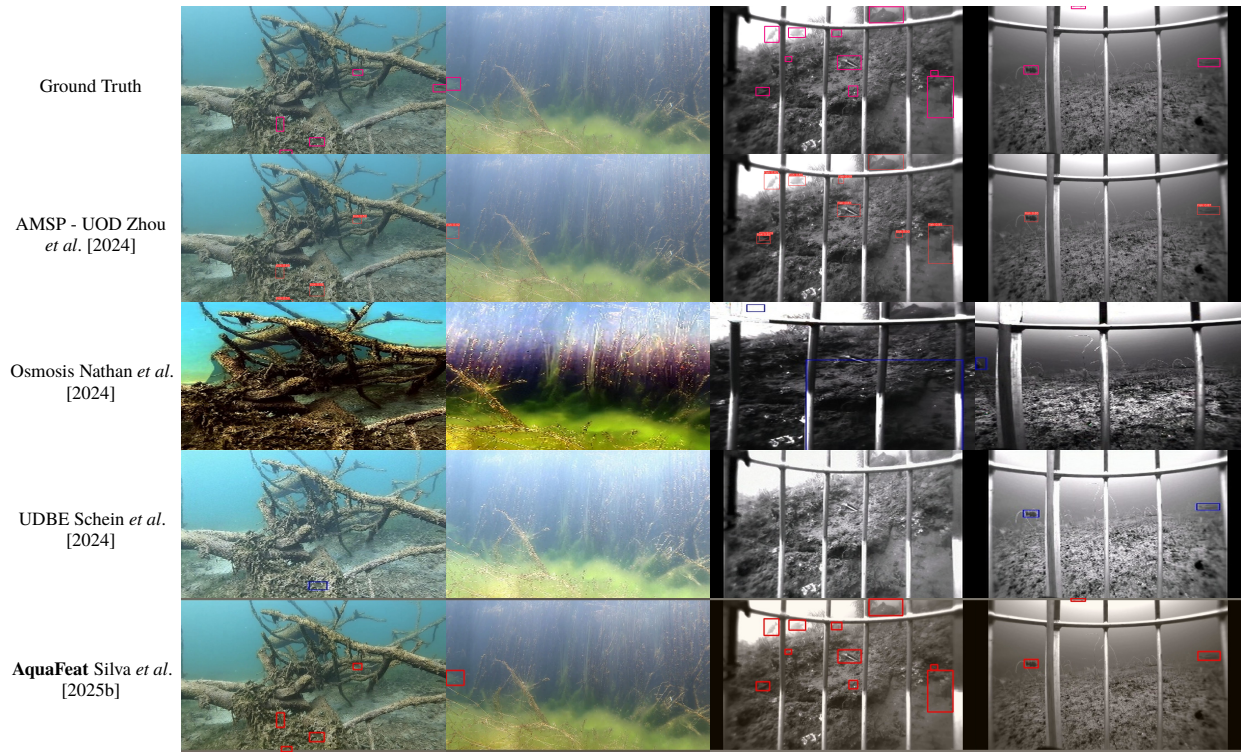

Figure 4. Qualitative object detection comparison on FishTrack23 dataset. Each column shows a different scene. The rows display the ground truth (top), results from competing methods (middle rows), and the first AquaFeat model (bottom).

have addressed the trade-off between image fidelity and computational efficiency. Notably, the C-Feat framework achieves an inference speed of 281.45 FPS. This research trajectory, validated through SIBGRAPI and ICAR and ICPR, provides a robust foundation for underwater perception.

Declarations

Acknowledgements

The authors would also like to thank FAPERGS, FAURG, FINEP and CNPq for their research support and financial assistance. The authors also acknowledge the financial support from the Human Resources Training Program of the Brazilian National Agency for Petroleum, Natural Gas and Biofuels (PRH-ANP), managed by the São Paulo Research Foundation (FAPESP), Brazil, under Process No. 2025/03736-5 – PRH-ANP 22/FURG Program.

Authors' Contributions

ES contributed to the conception of the study, software development, experimental design, data analysis, and manuscript writing. TS contributed to the experiments, manuscript writing, review, data analysis, and supervision. PD contributed to the funding, supervision and review of the whole research. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The source code for AquaFeat and AquaFeat+ is publicly available **here**, while the source code for C-Feat can be accessed **here**. The datasets generated in this work are also publicly available: the on used fot the AquaFeat+ paper **here** and AquaFeat **here**.

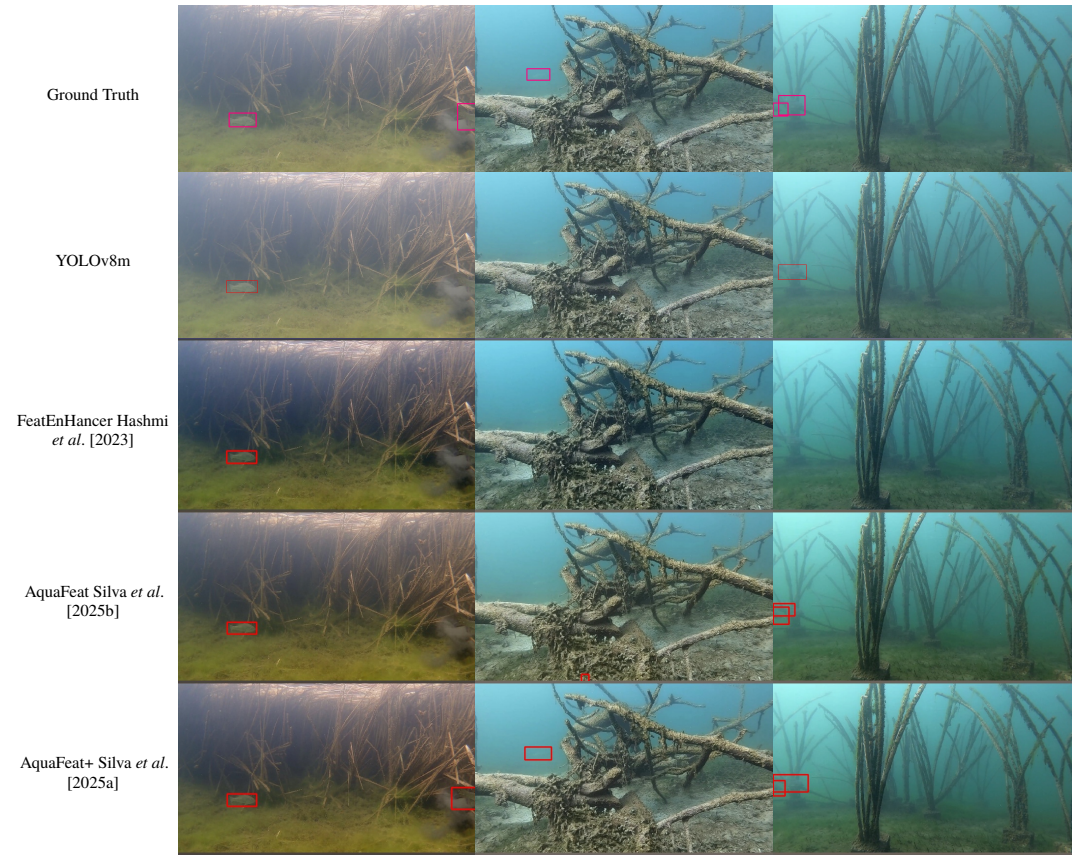

Figure 5. Qualitative comparison of model performance on the FishTrack23 dataset for the object detection task. The rows display bounding box predictions across different architectures compared against the Ground Truth using YOLOv8m as the baseline tracker.

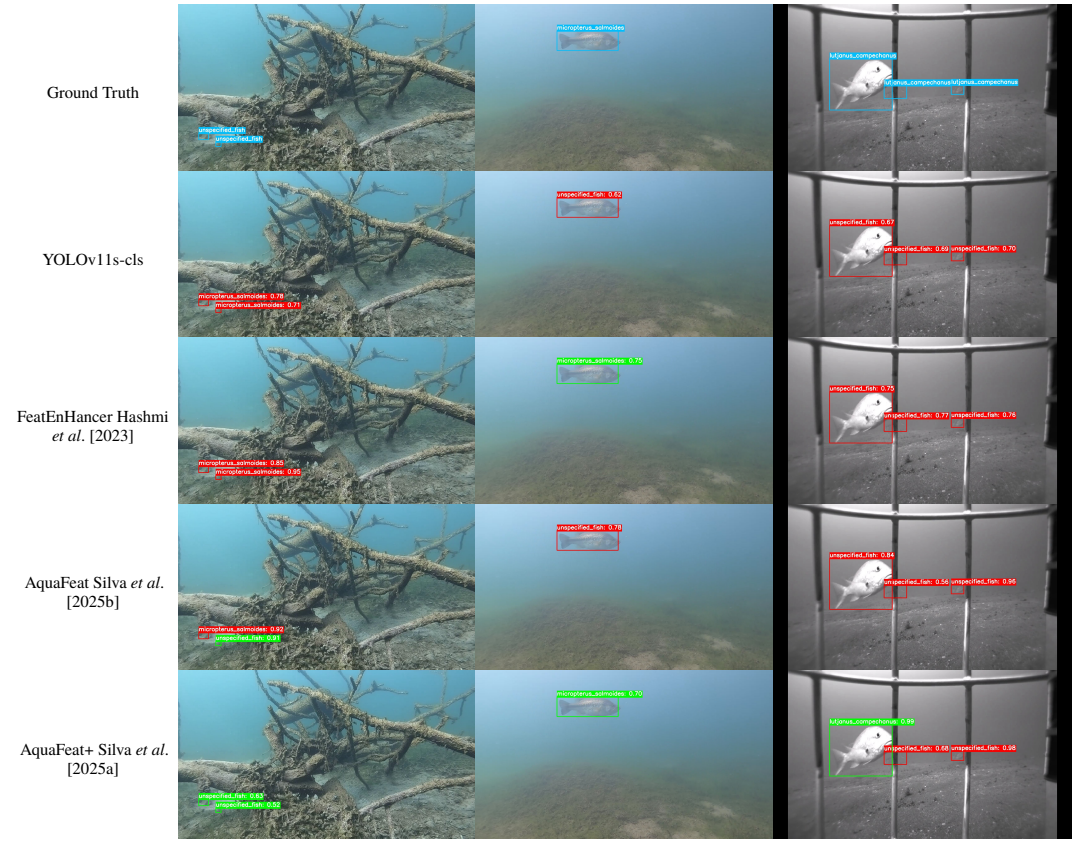

Figure 6. Qualitative comparison of model performance on the FishTrack23 dataset for the classification task using YOLOv11s-cls. Blue bounding boxes denote the ground truth, green boxes indicate a correct classification by the model, and red boxes signify a misclassification.

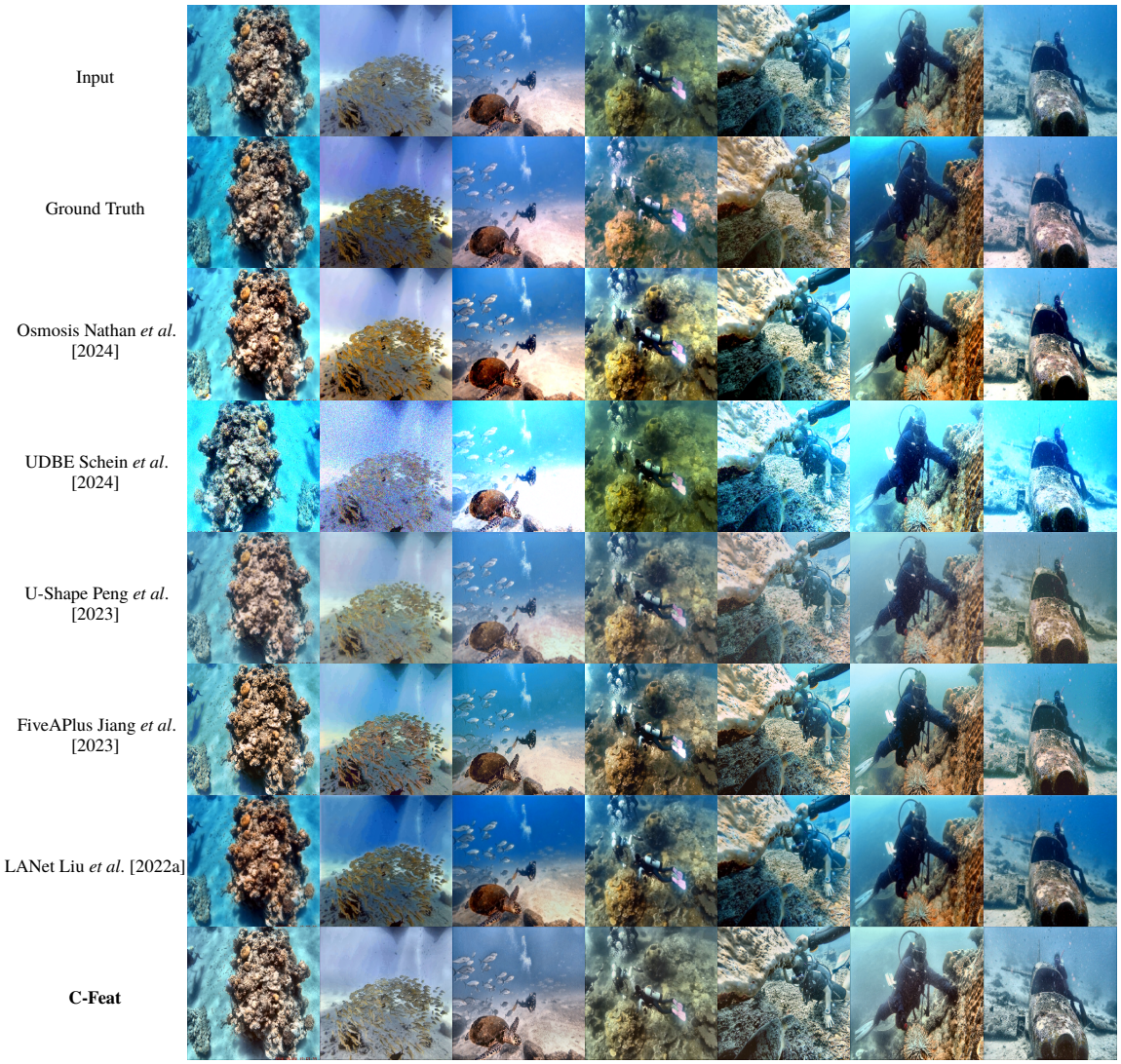

Figure 7. Qualitative results on UIEB dataset.

References

Dawkins, M., Prior, J., Lewis, B., *et al.* (2024). Fishtrack23: An ensemble underwater dataset for multi-object tracking. In *WACV*, pages 7167–7176.

Fu, Z., Lin, H., Yang, Y., Chai, S. b., Sun, L., Huang, Y., and Ding, X. (2022a). Unsupervised underwater image restoration: From a homology perspective. In *AAAI*, volume 36, pages 643–651. DOI: 10.1609/aaai.v36i1.19944.

Fu, Z., Wang, W., Huang, Y., Ding, X., and Ma, K.-K. (2022b). Uncertainty inspired underwater image enhancement. In *ECCV*, pages 465–482. Springer. DOI: 10.1007/978-3-031-19797-0_27.

Hashmi, K. A., Kallempudi, G., Stricker, D., and Afzal, M. Z. (2023). Featenhancer: Enhancing hierarchical features for object detection and beyond under low-light vision. In *ICCV*, pages 6725–6735. DOI: 10.48550/arXiv.2308.03594.

Hou, G., Li, N., Zhuang, P. b., Li, K., Sun, H., and Li, C. (2023). Non-uniform illumination underwater image restoration via illumination channel sparsity prior. *IEEE Trans. Circuits Syst. Video Technol.*, 34(2):799–814. DOI: 10.1109/TCSVT.2023.3290363.

Islam, M. J., Edge, C., Xiao, Y., *et al.* (2020). Semantic Segmentation of Underwater Imagery: Dataset and Benchmark. In *IROS*. DOI: 10.48550/arXiv.2004.01241.

Jiang, J., Ye, T., Chen, S., Chen, E., Liu, Y., Jun, S., Bai, J., and Chai, W. (2023). Five a+ network: You only need 9k parameters for underwater image enhancement. In *BMVC*. DOI: 10.48550/arXiv.2305.08824.

Khan, M. R., Negi, A., Kulkarni, A., Phutke, S. S., Vipparthi, S. K., and Murala, S. (2025). Phaseformer: Phase-based attention mechanism for underwater image restoration and beyond. In *WACV*, pages 9618–9629. IEEE. DOI: 10.1109/WACV61041.2025.00931.

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., and Tao, D. (2020). An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.*, 29:4376–4389. DOI: 10.1109/TIP.2019.2955241.

Liu, C., Shu, X., Pan, L., Shi, J., and Han, B. (2023). Multiscale underwater image enhancement in RGB and HSV color spaces. *IEEE Trans. Instrum. Meas.*, 72:1–14. DOI: 10.1109/TIM.2023.3298395.

Liu, S., Fan, H., Lin, S., Wang, Q., Ding, N., and Tang, Y. (2022a). Adaptive learning attention network for underwater image enhancement. *IEEE Robot. Autom. Lett.*,

7(2):5326–5333. DOI: 10.1109/LRA.2022.3156176.

Liu, Z., Mao, H., Wu, C.-Y., *et al*. (2022b). A convnet for the 2020s. In *CVPR*, pages 11976–11986. DOI: 10.1109/CVPR52688.2022.01167.

Mello, C. D., Moreira, B. U., de Oliveira Evald, P. J. D., Drews, P. J. L., and da Costa Botelho, S. S. (2022). Underwater enhancement based on a self-learning strategy and attention mechanism for high-intensity regions. *Comput. Graph.*, 107:264–276. DOI: 10.1016/j.cag.2022.08.003.

Nathan, O. B., Levy, D., Treibitz, T., and Rosenbaum, D. (2024). Osmosis: Rgbd diffusion prior for underwater image restoration. In *ECCV*, pages 302–319. DOI: 10.1007/978-3-031-73033-7_17.

Peng, L., Zhu, C., and Bian, L. (2023). U-shape transformer for underwater image enhancement. *IEEE Trans. Image Process.*. DOI: 10.1109/TIP.2023.3276332.

Saleh, A., Sheaves, M., Jerry, D., and Rahimi Azghadi, M. (2022). Adaptive uncertainty distribution in deep learning for unsupervised underwater image enhancement. *Available at SSRN 4742854*. DOI: 10.2139/ssrn.4362438.

Schein, T. T., De Almeira, G. P., Brião, S. L., De Bem, R. A., De Oliveira, F. G., and Drews-Jr, P. L. (2024). Udbe: Unsupervised diffusion-based brightness enhancement in underwater images. In *ICMLA*, pages 664–670. IEEE. DOI: 10.1109/ICMLA61862.2024.00096.

Silva, E. C., Schein, T. T., Ramos, J. D. G., Oliveira, F. G., and Drews-Jr, P. L. J. (2026). C-feat: A compact feature-centric network shattering training and inference latency in underwater vision. *ICPR*.

Silva, E. C., Schein, T. T., Ramos, J. D. G., Silva, E. L., Brião, S. L., Oliveira, F. G., and Drews, P. L. J. (2025a). Aquafeat+: an underwater vision learning-based enhancement method for object detection, classification, and tracking. In *ICAR*, pages 432–437. DOI: 10.1109/ICAR65334.2025.11338647.

Silva, E. d. C., Schein, T. T., Brião, S. L., Costa, G. L. M., Oliveira, F. G., Almeida, G. P., Silva, E. L., Devincenzi, S. d. S., Machado, K. d. S., and Drews-Jr, P. L. J. (2025b). Aquafeat: A features-based image enhancement model for underwater object detection. In *SIBGRAPI*. DOI: 10.1109/SIBGRAPI67909.2025.11223382.

Ultralytics (2023). YOLO-v8. GitHub repository.

Wang, A., Chen, H., Liu, L., *et al*. (2024). Yolov10: Real-time end-to-end object detection. *NeurIPS*, 37:107984–108011. DOI: 10.48550/arXiv.2405.14458.

Xie, S., Girshick, R., Dollár, P., Tu, Z., and He, K. (2017). Aggregated residual transformations for deep neural networks. In *CVPR*. DOI: 10.48550/arXiv.1611.05431.

Zhou, J., He, Z., Lam, K.-M., Wang, Y., Zhang, W., Guo, C., and Li, C. (2024). AMSP-UOD: When vortex convolution and stochastic perturbation meet underwater object detection. In *AAAI*, volume 38. DOI: 10.48550/arXiv.2308.11918.

Zhuang, P., Wang, Y., and Qiao, Y. (2018). Wildfish: A large benchmark for fish recognition in the wild. In *26th ACM international conference on Multimedia*, pages 1301–1309. DOI: 10.1145/3240508.324061.