ClearDive: Conquering the Challenges of Underwater Photography

Leen Said
Hacettepe University
Ankara, Türkiye
leen.said@hacettepe.edu.tr

Sümeyra Koç
Hacettepe University
Ankara, Türkiye
sumeyrakoc@hacettepe.edu.tr

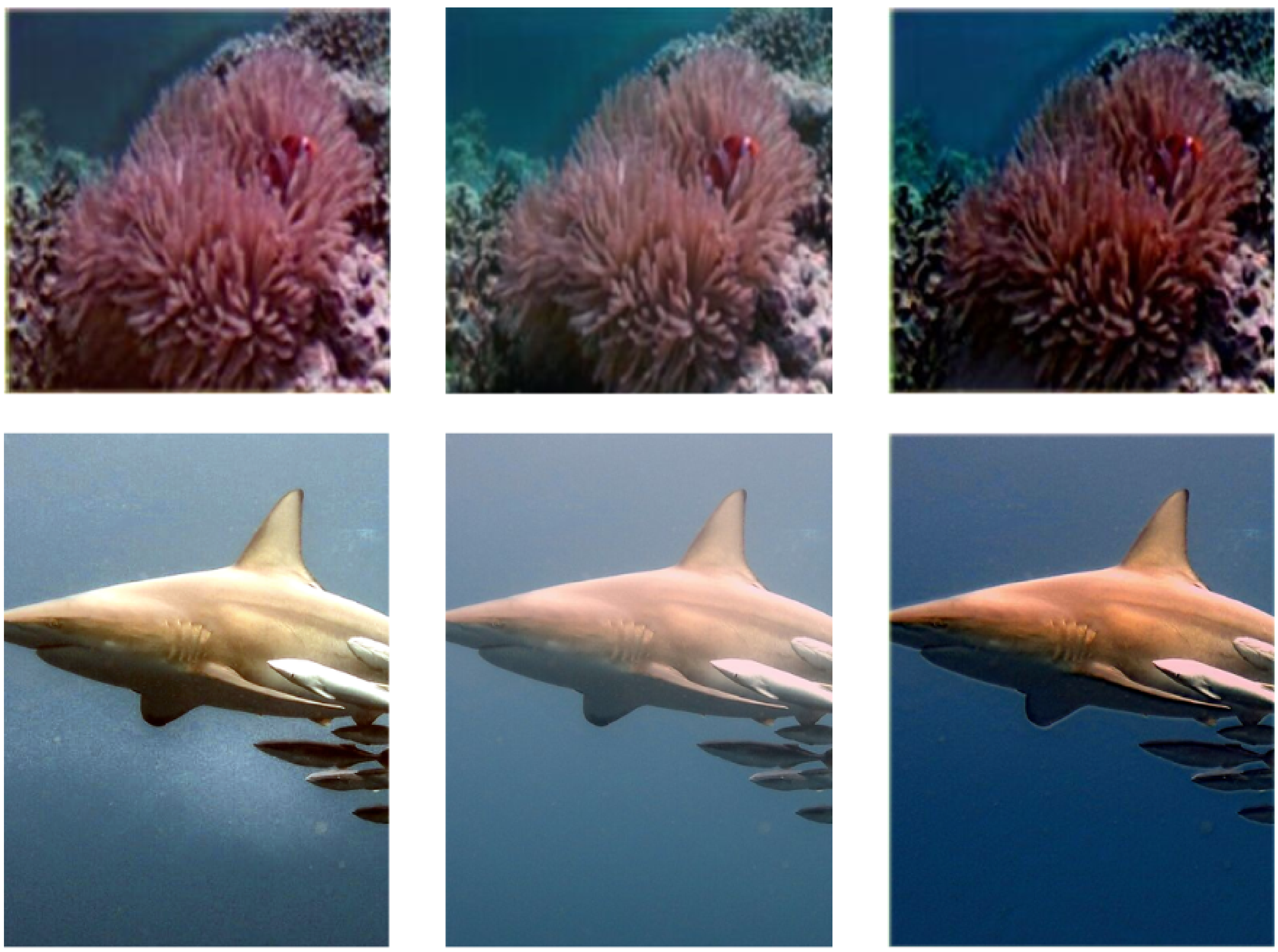

Figure 1: AOD-Net dehazing results. Upper (EUVP): fine-tuned (left), raw input (middle), before fine-tuning (right). Lower (UIEB): reference (left), raw input (middle), before fine-tuning (right).

ABSTRACT

Underwater imaging suffers from unique and complex degradation factors, including light scattering, wavelength-dependent absorption, and non-uniform illumination, all of which significantly impair visibility and color fidelity. This paper introduces ClearDive, a modular underwater image enhancement framework that integrates two core strategies: (1) a structured two-stage pipeline that combines deep learning-based dehazing with pretrained color correction models, and (2) a lightweight exploratory extension that utilizes event camera data to restore structural detail in motion-degraded scenes. We conduct a series of controlled experiments across benchmark datasets (UIEB, EUVP, DAVIS-NUIUIED) using both full-reference and non-reference quality metrics (PSNR, SSIM, UIQM, UCIQE). Our findings highlight the strong influence of input preprocessing on downstream enhancement quality, and demonstrate the potential of event-based vision for motion-aware sharpening. The results

, ,
.

underscore the value of modularity and multi-modal input fusion in advancing practical underwater vision solutions.

KEYWORDS

Underwater Imaging, Event-Guided Enhancement, Non-Uniform Illumination, Image Sharpening, Color Restoration, Multi-Modal Fusion, UIEB Benchmark, DAVIS-NUIUIED Dataset, Deep Learning for Underwater Vision, Perceptual Image Quality

ACM Reference Format:
Leen Said and Sümeyra Koç. . ClearDive: Conquering the Challenges of Underwater Photography. In *Proceedings of* . ACM, New York, NY, USA, 8 pages.

1 INTRODUCTION

Underwater image enhancement presents persistent challenges due to the complex and variable nature of aquatic environments. Factors such as wavelength-dependent absorption, light scattering, and non-uniform illumination can significantly degrade visual quality, introducing artifacts that hinder both human interpretation and computer vision tasks. These forms of degradation often co-occur and interact in unpredictable ways, making universal solutions difficult to achieve.

In this project, we adopt a modular and exploratory approach to investigate how different input preparation strategies and sensor modalities affect enhancement performance. Our work begins with a structured two-stage pipeline, where images are first dehazed using a lightweight deep learning model and then processed through pretrained color correction networks. This design allows us to examine whether structural enhancement prior to color correction improves the final output.

To complement this pipeline, we introduce an exploratory extension that investigates the use of event camera data for edge reconstruction in motion-degraded scenes. Event streams offer temporally precise, edge-focused structural cues that are not available in conventional RGB inputs, making them a promising modality for enhancing underwater imagery in dynamic conditions. Rather than developing end-to-end models, our focus lies in assessing the potential of lightweight, interpretable techniques for guiding enhancement.

Together, these experiments reflect a broader goal: to understand how preprocessing decisions and alternative sensing strategies can address the multifaceted nature of underwater degradation. By evaluating both conventional and novel methods across different datasets and metrics, we aim to provide insight into practical, modular solutions for underwater image restoration.

2 RELATED WORK

Underwater image enhancement has been extensively studied due to the challenging visual conditions caused by light scattering, absorption, and wavelength-dependent color distortion. Traditional approaches, such as histogram equalization and model-based dehazing, often struggle to generalize across different underwater scenes.

2.1 Underwater Image Dehazing Approaches

Underwater image dehazing has been a widely explored preprocessing step for enhancing image quality by addressing the effects of light scattering and turbidity. Traditional dehazing techniques often rely on physical image formation models, such as the dark channel prior or wavelength compensation models. However, these approaches typically assume simplified water conditions and tend to perform inconsistently in real-world underwater scenes. Recent studies have shifted toward deep learning-based dehazing methods due to their ability to learn complex degradation patterns directly from data. For example, [7] incorporate multiple color spaces and fusion strategies to enhance robustness across diverse underwater environments.

2.2 Deep Learning-Based Underwater Enhancement

In recent years, deep learning-based approaches have become the dominant paradigm for underwater image enhancement. CNN architectures have been widely adopted for their ability to extract multi-scale features and perform end-to-end restoration under various degradation conditions. Models such as Water-Net [8] and UWCNN [9] focus on learning pixel-level corrections directly from data, often incorporating domain-specific priors or multi-branch designs. GAN-based frameworks, including models like UGAN [3] and UWGAN [1], employ adversarial loss to improve perceptual realism and generate visually pleasing outputs. Some methods combine adversarial training with content or structure-aware loss functions to better preserve edges and global color consistency. These data-driven methods significantly outperform traditional techniques, particularly when large-scale annotated datasets are available.

2.3 Event-Based Image Enhancement and Deblurring

Event cameras offer high temporal resolution and are well-suited for motion-rich or low-light environments where conventional cameras struggle. Unlike standard frame-based sensors, they asynchronously capture pixel-level brightness changes, making them effective for tasks like motion deblurring and edge detection [4].

Works such as E2VID [12] and EVDeblurNet [10] demonstrate how event streams can be fused with or used to reconstruct frames with improved temporal sharpness. These methods recover fine details in dynamic scenes using event-driven priors.

Recently, Zhang et al. [2] explored RGB/event fusion for underwater image enhancement, using simulated event data to address issues like motion blur and lighting inconsistency. Their results suggest that event signals can provide valuable structural cues under degraded underwater conditions.

In contrast to end-to-end deep fusion models, our method focuses on lightweight, interpretable enhancement by injecting event-derived edge information into RGB frames. This modular design aligns with recent efforts toward explainable, resource-efficient event-based vision [11].

3 THE APPROACH

3.1 Deep Learning-based Enhancement Pipeline

3.1.1 Dehazing Process. To address structural degradation such as scattering-induced blur and reduced contrast, we implemented and trained the AOD-Net (All-in-One Dehazing Network) from scratch. AOD-Net is a lightweight end-to-end model designed to recover clean images from hazy input by learning a direct mapping, without the need to explicitly estimate transmission maps or atmospheric light [7]. Its architecture consists of five convolutional layers, and it is widely recognized for its efficiency and strong performance on single-image dehazing tasks.

We used a PyTorch reimplementation of AOD-Net [5] and initially trained the model on the RESIDE dataset, which includes synthetically generated hazy images paired with corresponding ground truth clean images. The training setup employed mean squared error (MSE) as the loss function, optimized using Adam with a learning rate of 1×10^{-4}, a batch size of 4, and 10 training epochs. To enhance generalization and structural accuracy, we applied standard resizing and normalization preprocessing during training, and validated on a held-out split of the same dataset.

Following initial training, we fine-tuned the model for 5 additional epochs using the UIEB dataset, which contains real-world underwater images exhibiting natural degradation patterns like scattering, color cast, and veiling light. This fine-tuning step was designed to adapt the model to the unique characteristics of underwater scenes, bridging the domain gap between terrestrial haze (RESIDE) and underwater visibility conditions (UIEB).

After training, the AOD-Net model was used to preprocess underwater images prior to applying color correction.

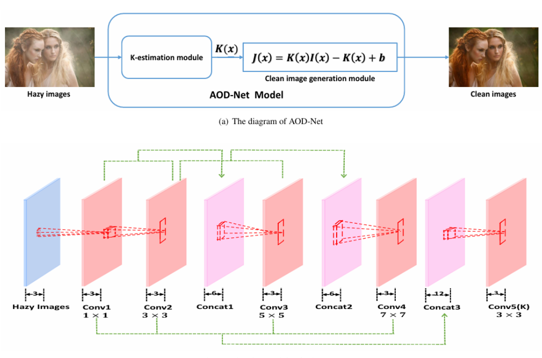

Figure 2: The AOD-Net architecture used for dehazing and sharpening, as proposed by Li et al. [7].

3.1.2 Color Correction Process. To restore the natural appearance of underwater images after sharpening, we utilized deep learning-based color correction models. In this study, we selected FUnIE-GAN and UIEC²-Net, two widely recognized and cited architectures for underwater image enhancement. The use of two different models enables us to evaluate the consistency of the sharpening effect across different color correction strategies, thereby improving the reliability and generalizability of our experimental results.

FUnIE-GAN (Fast Underwater Image Enhancement GAN) is a lightweight generative adversarial network specifically designed for real-time underwater image enhancement. It is based on the classical GAN framework, but is tailored to efficiently handle the color distortions and reduced visibility commonly found in underwater environments [6].

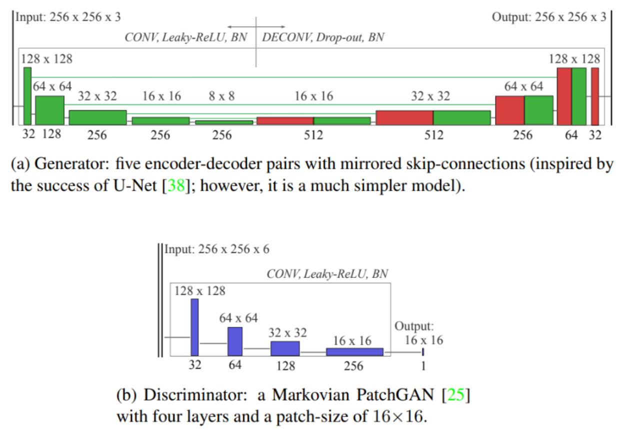

Figure 3: Network architecture of the proposed model: FUnIE-GAN[6]

UIEC²-Net is a lightweight deep convolutional neural network designed for real-time underwater image enhancement. The model is built around the idea of attentive feature fusion, integrating both local pixel-level detail and global color correction through specialized branches. [13]. FUnIE-GAN performs enhancement using

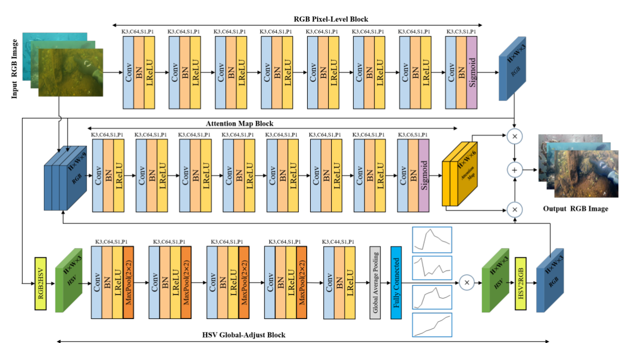

Figure 4: An overview of the proposed UIEC^2-Net architecture.[13]

a GAN-based encoder-decoder architecture that is optimized for real-time applications and perceptual quality, particularly in visually degraded underwater scenes. In contrast, UIEC²-Net adopts a fusion-based strategy that combines local structural features with global color cues through attention-guided processing in both RGB and HSV domains. Together, these two models represent distinct enhancement philosophies—one focusing on generative realism, the other on structure-preserving correction. Evaluating them side by side enables a more robust analysis of how input preparation, such as dehazing, interacts with different architectural assumptions in underwater image enhancement.

For both FUnIE-GAN and UIEC²-Net, we utilized available pretrained weights provided by their respective research communities or train the models briefly on a small subset of underwater images if pretrained models were not directly available.

3.1.3 Experimental Pipeline. The dataset is divided into two groups: Group A consists of original images without sharpening, while Group B includes images enhanced using the pretrained AOD-Net model, which directly predicts a clean image by removing haze and blur from the input, Figure 5. This end-to-end dehazing approach enables simultaneous sharpening and structural enhancement of underwater images. While the current experimental pipeline relies solely on conventional RGB underwater images. Subsequently, both

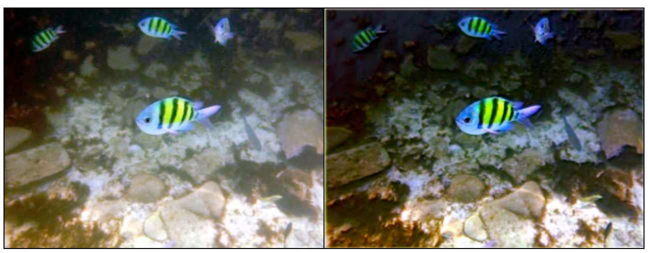

Figure 5: Comparison of two input types. The left is raw underwater input without any preprocessing, the right is dehazed one.

groups are processed by two color correction models, FUnIE-GAN and UIEC²-Net, resulting in four enhanced outputs per original image. The first set consists of outputs from FUnIE-GAN and UIEC²-Net without sharpening, and the second set consists of outputs with sharpening preprocessing applied. Finally, the quality of all enhanced images is assessed using evaluation metrics (see Experimental Evaluation), and a comparative analysis is conducted to investigate the effects of sharpening on different color correction strategies.

3.2 Event-Based Enhancement

In the second part of our study, we investigate the potential of leveraging **event-based vision** to enhance underwater images, particularly under conditions where conventional frame-based cameras fail, such as motion blur and low-light environments. Unlike standard cameras that capture frames at fixed intervals, **event cameras** like the DAVIS sensor operate asynchronously, recording per-pixel brightness changes with microsecond precision [4]. This produces a high-temporal-resolution stream of events encoding edge and motion information that traditional RGB inputs cannot capture.

These characteristics make event cameras highly suitable for underwater vision. As depth increases, ambient light diminishes, and scattering from suspended particles leads to severe degradation in RGB image quality. Event cameras, however, are more resilient to such conditions due to their high dynamic range, noise tolerance, and edge-focused sensing.

To explore this potential, we employed the **DAVIS-NUIUIED** dataset [2], which provides synchronized underwater RGB and event streams captured using a DAVIS sensor. Unlike prior deep fusion models, our approach emphasizes *modular and interpretable* enhancement techniques—specifically evaluating whether lightweight event-guided operations can recover structural detail in motion-degraded RGB frames using event-derived edge information.

3.2.1 Preprocessing. We extracted RGB frames and aligned event streams from the dataset. For each RGB frame, events were aggregated over a fixed **200 ms window** preceding the frame to create a 2D **grayscale event map**. These maps highlight regions of recent intensity change, effectively capturing motion edges.

Due to minor timestamp inconsistencies between RGB and event streams, we manually synchronized the data by referencing the first RGB frame and applying a global offset correction. This produced temporally aligned RGB frames and corresponding event maps, ready for enhancement.

3.2.2 Event-Guided Enhancement Techniques. We developed two lightweight fusion methods that inject event-based edge information into the RGB stream:

1. Event-Weighted Temporal Fusion. This adaptive approach assigns **pixel-wise blending weights** to the previous, current, and next frames using the normalized event map:

- High-motion pixels favor the next frame, anticipating temporal lag.
- Static pixels favor the previous frame, likely less distorted.
- The current frame is retained neutrally for continuity.

This weighted blending approach adapts to local scene dynamics. After fusion, a Laplacian edge map is injected into **motion-active regions**, gated by a **binarized event mask**. The sharpening strength is also **scaled dynamically** based on overall event activity, allowing the method to handle non-uniform blur, preserve **temporal coherence**, and minimize artifacts.

2. Event-Guided Static Fusion. This method recovers structural details by sharpening motion-blurred regions based on event activity. We compute a **temporal average** of three consecutive RGB frames (previous, current, next) to suppress noise and jitter, though this blurs fine details. A **Laplacian edge map** of the averaged image is then blended back using the **normalized event map** as an attention mask—sharpening motion-active regions while leaving static areas untouched.

4 DATASETS

We employ three publicly available underwater datasets, each contributing distinct visual challenges and modalities to our study.

4.1 UIEB

The UIEB dataset includes 950 real underwater images, with 890 paired with high-quality references and 60 classified as challenging, unpaired cases. It covers typical degradations like scattering, color casts, and low contrast. We used UIEB to fine-tune AOD-Net and assess perceptual and structural restoration.

4.2 EUVP

EUVP offers over 11,000 underwater images, split into paired (CycleGAN-generated) and unpaired subsets collected from diverse cameras and environments. It spans conditions such as turbidity and low light. We used EUVP to evaluate the generalization of our enhancement pipeline.

4.3 DAVIS-NUIUIED

DAVIS-NUIUIED provides synchronized underwater RGB and event data from a DAVIS sensor, capturing real-world degradations including motion blur and illumination shifts. Its dual-modality design enabled our exploration of event-guided enhancement in motion-degraded scenes.

5 EXPERIMENTAL RESULTS

5.1 Deep Learning-based Enhancement Experiments Results

We conducted a series of experiments to evaluate the impact of different input preprocessing strategies on the performance of underwater image enhancement models. Specifically, we tested two models—FUnIE-GAN and UIEC^2-Net—using two input variants: raw underwater images and dehazed versions of the same images produced by AOD-Net dehazing model. For both models, the outputs were evaluated against ground truth references using full-reference image quality metrics: Peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) and non-reference image quality metrics such as the Underwater Image Quality Measure (UIQM) and Underwater Color Image Quality Evaluation (UCIQE).

5.1.1 Dehazing with AOD-Net. We evaluated the performance of AOD-Net both before and after fine-tuning, using two different underwater datasets to measure generalization and avoid evaluating on the fine-tuning set. UIEB was used to test the model prior to fine-tuning, while EUVP served as a distinct test set for post-finetuning evaluation. This section focuses exclusively on the results of the dehazing process, prior to color correction.

Quantitative Results. Table 1 summarizes both full-reference and non-reference metrics obtained from these experiments. The model showed consistent perceptual improvements across both datasets, even though traditional full-reference metrics like PSNR and SSIM remained relatively modest.

Table 1: Dehazing performance on UIEB and EUVP datasets. Full-reference metrics are only available for UIEB.

Dataset	Stage	PSNR	SSIM	UIQM	UCIQE
UIEB	Before FT	12.66	0.5456	67.34	4.61
EUVP	Before FT	–	–	67.38 ± 15.01	6.41 ± 2.12
EUVP	After FT	–	–	67.25 ± 15.86	4.98 ± 2.21

Unlike the UIEB dataset, which includes paired reference images, the EUVP dataset contains only raw underwater scenes without corresponding ground truth. As a result, full-reference metrics such as PSNR and SSIM could not be computed on EUVP. This is a common constraint in underwater imaging, where capturing perfect reference data is often infeasible due to dynamic illumination, motion, and environmental variability.

Visual Results. Figure 6 shows a visual comparison from the UIEB dataset, featuring the raw input, the dehazed output from AOD-Net, and the corresponding reference image. This illustrates how dehazing improved structural clarity and local contrast while slightly shifting color balance.

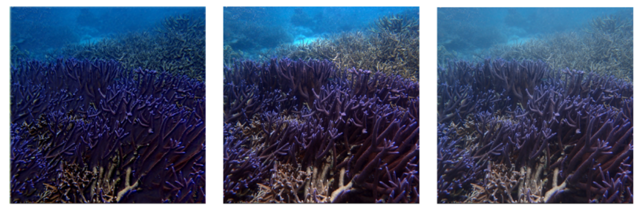

Figure 6: Qualitative comparison on UIEB: (left) Dehazed with AOD-Net, (middle) Reference image, (right) Raw input.

Figure 7 illustrates the same input image processed by AOD-Net before and after fine-tuning, using a sample from the EUVP dataset. The fine-tuned model produced a more visually balanced result, with reduced over-saturation and improved tone consistency.

Figure 7: Dehazing results on EUVP dataset: (left) Raw input, (middle) AOD-Net output before fine-tuning, (right) Output after fine-tuning.

Qualitative Observations. While the numerical scores remained stable before and after fine-tuning, we observed meaningful differences in the visual appearance of the outputs. The fine-tuned model tended to produce images that were less saturated and visually more natural.

Interestingly, although the UCIQE score decreased slightly after fine-tuning, we observed that the fine-tuned model produced images with more balanced color and less overcorrection. In particular, the dehazed outputs exhibited reduced saturation and less aggressive tonal shifts, which may explain the lower UCIQE values despite visual improvement.

This behavior is desirable in our pipeline, as these outputs are intended to serve as inputs to downstream color correction models. By avoiding overly vivid or distorted enhancements at the dehazing stage, the fine-tuned model preserves more natural image statistics and structure, allowing subsequent stages to operate on more neutral inputs. Thus, the observed drop in metric performance reflects a shift in enhancement strategy rather than a degradation in visual quality.

5.1.2 Color Correction. UIEC²-Net demonstrated significant sensitivity to input type. When raw images were used, the model achieved a mean PSNR of 23.68 dB and SSIM of 0.9099, outperforming its performance on dehazed inputs (PSNR of 16.84 dB, SSIM of 0.7590). These results suggest that UIEC²-Net benefits from retaining the raw input structure, possibly due to its fusion-based architecture, which leverages low-level spatial features more effectively.

FUnIE-GAN yielded lower overall enhancement performance. The PSNR scores remained nearly unchanged between raw and dehazed inputs (11.80 dB vs. 11.79 dB), while SSIM showed a modest improvement with raw inputs (from 0.4792 to 0.5820). This suggests that although the model's output fidelity remains consistent in terms of signal, structural information is better preserved when operating on raw inputs.

Model	Input Type	PSNR	SSIM
UIEC2Net	Dehazed	16.84	0.7590
UIEC2Net	Raw	23.68	0.9099
FUnIE-GAN	Dehazed	11.79	0.4792
FUnIE-GAN	Raw	11.80	0.5820

Table 2: Comparison of model performance on different input types with full-reference metrics.

In addition, we observed that both models internally resize input images to 256×256 resolution, which introduces a mismatch when comparing outputs to high-resolution ground truths; see Figure 8. This resizing may partially explain the relatively modest metric values, especially for PSNR, which is sensitive to pixel-level differences. Despite this limitation, the observed trends remain robust and provide insight into the interaction between input preparation and model behavior. To further validate our findings, we evaluated the enhanced outputs using non-reference quality metrics, namely UIQM and UCIQE, which assess perceptual aspects of underwater images without requiring ground truth. Interestingly, results were

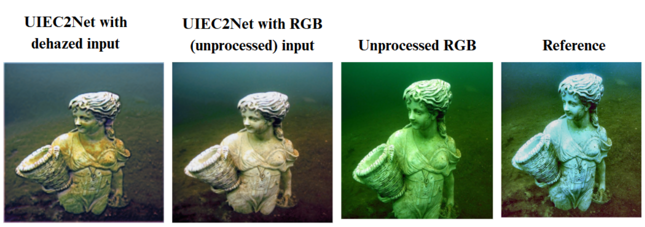

Figure 8: Visual comparison of UIEC²-Net outputs under different input conditions.

mixed Table ??: the raw input set achieved a higher UIQM score (78.57 vs. 73.88), indicating better overall perceptual quality, while the dehazed input set obtained a slightly higher UCIQE score (5.02 vs. 4.59), suggesting improved colorfulness and contrast in localized regions. These results suggest that different preprocessing strategies may benefit different aspects of visual quality, and highlight the importance of using complementary evaluation metrics when assessing enhancement performance.

	Model	Input Type	UIQM	UCIQE
1	UIEC2Net	Dehazed	73.8887	5.0206
2	UIEC2Net	Raw	78.5729	4.5883
3	FUnIE-GAN	Dehazed	57.3688	6.1911
4	FUnIE-GAN	Raw	57.7676	6.0041

Table 3: Comparison of model performance on different input types with non-reference metric results

UIEC²-Net achieved higher UIQM scores when raw inputs were used (78.57 compared to 73.88 with dehazed inputs), indicating that raw images lead to overall better perceptual enhancement. However, the UCIQE score was slightly higher for the dehazed input (5.02 vs. 4.59), suggesting that dehazing may improve localized color contrast, even if it slightly compromises global perceptual quality. These results imply that UIEC²-Net benefits from preserving the natural structure of raw inputs, likely due to its attention-based fusion design that utilizes both local and global features.

FUnIE-GAN, on the other hand, produced much lower UIQM scores overall (57 for both input types), Figure 9, suggesting that its generative framework is less effective at improving perceptual quality in this context. Interestingly, it achieved the highest UCIQE scores among all settings (6.19 with dehazed and 6.00 with raw), which indicates stronger enhancement in terms of color contrast and distribution. However, the minimal UIQM difference between raw and dehazed inputs (57.77 vs. 57.36) suggests that FUnIE-GAN is less sensitive to input variations.

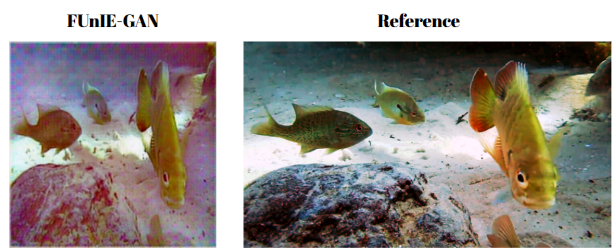

Figure 9: A comparison of the result produced by the FUnIE-GAN model

These findings highlight the importance of evaluating enhancement models across multiple quality dimensions. While raw inputs generally led to better perceptual scores (UIQM), dehazed inputs were sometimes favored by color-based metrics (UCIQE). This divergence underscores the fact that preprocessing steps may benefit certain visual aspects (e.g., color contrast) at the expense of others (e.g., structure preservation), and that model performance should be interpreted accordingly.

To assess the effect of improved dehazing quality on overall enhancement performance, we fine-tuned the dehazing model used to generate inputs for both UIEC²-Net and FUnIE-GAN. The outputs from the enhancement models were then re-evaluated using non-reference metrics—UIQM and UCIQE, see Table 4. The new results were compared against the original scores shown in Table **??**. For

	Model	Input Type	UIQM	UCIQE
1	UIEC2Net	Dehazed (Fine-tuned)	78.696	4.428
2	FUnIE-GAN	Dehazed (Fine-tuned)	60.273	6.73

Table 4: Non-reference metric results on fine tuned dehazed image.

UIEC²-Net, the average UIQM increased significantly from 73.89 to 78.70, indicating a substantial gain in overall perceptual quality after feeding in higher-quality dehazed inputs. However, the UCIQE score dropped slightly from 5.02 to 4.43, suggesting a possible trade-off in local color contrast or chroma consistency. This may indicate that the fine-tuned dehazer produced smoother but less vibrant images, which benefited the overall structure-aware fusion in UIEC²-Net but slightly penalized color-focused metrics. For FUnIE-GAN, the UIQM improved from 57.37 to 60.27, confirming a moderate enhancement in visual quality. More strikingly, the UCIQE score increased from 6.19 to 6.73, reflecting improved color balance and contrast. Given that FUnIE-GAN is a GAN-based model focusing on perceptual realism, it appears more responsive to the color consistency introduced by the fine-tuned dehazed inputs.

Overall, these results demonstrate that improving the dehazing stage through fine-tuning leads to consistent gains in non-reference quality metrics across both models. The effect is especially pronounced for FUnIE-GAN in UCIQE, and for UIEC²-Net in UIQM. This supports the hypothesis that preprocessing—when optimized—can significantly influence downstream enhancement quality, even without altering the enhancement models themselves.

5.2 Event-Guided Edge Reconstruction

We evaluated the proposed event-based enhancement methods using no-reference image quality metrics across three variants: raw RGB, Event-Guided Static Fusion, and Event-Weighted Temporal Fusion.

The following metrics were selected to assess perceptual and structural quality:

- **Sharpness (Laplacian Variance)** – Measures edge clarity and contrast.
- **Shannon Entropy** – Reflects image information content.
- **Edge Density** – Indicates preservation of fine structures.
- **RMS Contrast** – Approximates perceptual contrast.

Table 5: No-reference metrics for Raw, Static, and Temporal Fusion outputs.

Metric	Raw	Static	Temporal
Sharpness (Laplacian)	65.9	**70.3**	44.5
Entropy (Shannon)	7.30	7.35	**7.39**
Edge Density	0.029	**0.032**	0.024
RMS Contrast	**80.0**	78.5	76.6

Quantitative Results.

Discussion. The **Static Fusion** method delivered the best overall performance, with a +6.7% gain in sharpness and a +10.3% increase in edge density over the raw input, while maintaining comparable entropy and contrast.

The raw RGB served as a solid baseline, achieving the highest RMS contrast despite minor degradation.

By contrast, **Temporal Fusion** slightly improved entropy but significantly reduced sharpness (−32%) and edge density (−17%). Its adaptive blending introduced oversmoothing in motion regions, suggesting that motion-aware strategies require further tuning in noisy conditions.

Figure 10: Qualitative comparison: Raw RGB (left), Static Fusion (middle), Temporal Fusion (right). Static fusion clearly restores edge sharpness in degraded regions.

6 CONCLUSION

In this work, we presented ClearDive, a modular framework for addressing the multifaceted challenges of underwater photography. Our approach combined a structured deep learning-based pipeline with an exploratory investigation of event-guided enhancement methods, focusing on two main goals: (1) evaluating how input preprocessing—specifically dehazing with AOD-Net—affects the performance of pretrained color correction models, and (2) exploring the potential of event-based vision for restoring structural detail in motion-degraded underwater scenes.

Our results demonstrate that careful input preparation significantly impacts the perceptual quality of enhanced images. In particular, fine-tuning the dehazing model led to consistent improvements in both UIQM and UCIQE metrics across FUnIE-GAN and UIEC2-Net. UIEC2-Net benefited most from cleaner inputs in terms of perceptual quality (UIQM), while FUnIE-GAN showed notable gains in color fidelity (UCIQE). These findings validate the importance of preprocessing choices in underwater image enhancement workflows.

For future work, several promising directions emerge from our findings. One potential extension is the joint training of dehazing and color correction models in a single, end-to-end pipeline. This could allow the system to learn optimal preprocessing and enhancement strategies collaboratively, rather than treating them as separate stages.

Additionally, our lightweight event-guided enhancement methods revealed the potential of leveraging high-temporal-resolution edge information to recover structural detail in low-visibility, motion-heavy scenarios. The static fusion approach proved especially effective, outperforming both raw RGB input and more complex temporal fusion in sharpness and edge density metrics.

Our results highlight the potential of event-guided methods for enhancing underwater images, especially under motion blur and low-light conditions.

While promising, the temporal method showed limitations in high-motion regions, suggesting room for improvement. Future work might explore more robust motion-aware fusion, learning-based event integration, and broader testing across diverse underwater conditions to improve adaptability and effectiveness.

We also see value in deploying the pipeline on real-time, embedded systems, particularly for autonomous underwater vehicles (AUVs) where on-board image enhancement could improve navigation and perception.

Overall, ClearDive illustrates the value of modular, interpretable enhancement strategies and multi-modal sensing in underwater vision. Rather than relying on end-to-end black-box models, our work highlights how lightweight and targeted interventions—both at the input level and during enhancement—can produce significant quality gains.

REFERENCES

[1] Kancharagunta Kishan Babu, Ashreen Tabassum, Bommakanti Navaneeth, Tenneti Jahnavi, and Yenka Akshaya. 2025. Underwater Image Enhancement using Generative Adversarial Networks: A Survey. *arXiv preprint arXiv:2501.06273* (2025). https://doi.org/10.48550/arXiv.2501.06273

[2] Xiuwen Bi, Pengfei Wang, Wei Guo, Fusheng Zha, and Lining Sun. 2024. RGB/Event signal fusion framework for multi-degraded underwater image enhancement. *Frontiers in Marine Science* 11 (2024), 1366815. https://doi.org/10.3389/fmars.2024.1366815

[3] Claudio Fabbri, Md Jahidul Islam, and Junaed Sattar. 2018. Enhancing underwater imagery using generative adversarial networks. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 7159–7165. https://doi.org/10.1109/ICRA.2018.8460962

[4] Guillermo Gallego, Tobi Delbruck, Garrick Orchard, Chiara Bartolozzi, Brian Taba, Andrea Censi, Stefan Leutenegger, Andrew J Davison, Jorg Conradt, Kostas Daniilidis, and Davide Scaramuzza. 2022. Event-based vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 1 (2022), 154–180.

[5] Ren-Wu Hong. 2020. AOD-Net PyTorch Implementation. https://github.com/walsvid/AOD-Net-PyTorch. Accessed: 2024-05-28.

[6] Md Jahidul Islam, Youya Xia, and Junaed Sattar. 2020. Fast Underwater Image Enhancement for Improved Visual Perception. *arXiv preprint arXiv:1903.09766* (2020). https://doi.org/10.48550/arXiv.1903.09766

[7] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. 2017. AOD-Net: All-in-One Dehazing Network. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (2017), 4770–4778.

[8] Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren. 2021. Underwater Image Enhancement via Medium Transmission-Guided Multi-Color Space Embedding. *arXiv preprint arXiv:2104.13015* (2021). https://doi.org/10.48550/arXiv.2104.13015

[9] Chongyi Li, Jiandong Guo, Runmin Cong, Yonghong Pang, and Bo Wang. 2018. Underwater image enhancement by dehazing and color correction. *IEEE Transactions on Image Processing* 27, 12 (2018), 3797–3812. https://doi.org/10.1109/TIP.2018.2839701

[10] Longguang Pan, Yulun Jiang, Cheng Wang, Liang Zhang, Xuming Mei, and Ming-Hsuan Yang. 2020. High frame rate video reconstruction based on an event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[11] Priya Ramesh, Ethan Heller, and Ken Sugihara. 2023. Event-based Vision meets Explainability: Interpretable Techniques for Motion-Aware Enhancement. *arXiv preprint arXiv:2305.12345* (2023).

[12] Henri Rebecq, Rene Ranftl, Vladlen Koltun, and Davide Scaramuzza. 2019. High speed and high dynamic range video with an event camera. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

[13] Yudong Wang, Jichang Guo, Huan Gao, and Huihui Yue. 2021. UIEC^2-Net: CNN-based Underwater Image Enhancement Using Two Color Space. *arXiv preprint arXiv:2103.07138* (2021). https://doi.org/10.48550/arXiv.2103.07138