UWHFORMER: A SPECTRALLY COHERENT TRANSFORMER FOR UNDERWATER HYPERSPECTRAL IMAGE RESTORATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Hyperspectral imaging captures dense spatial–spectral information, enabling fine-grained material identification and scene understanding. Underwater acquisition, however, is degraded by wavelength-dependent attenuation, non-uniform scattering, insufficient illumination, and sensor noise, which jointly corrupt spatial structures and spectral signatures. We formulate underwater hyperspectral image (HSI) restoration as a physics-guided spatial–spectral reconstruction problem and propose UWHFormer, a spectrally coherent Transformer for restoring real-world UnderWater HSIs. UWHFormer first employs a water-conditioned state modeling module to estimate physically interpretable, wavelength-dependent degradation states. A physics-guided restoration transformer then integrates these states with image features to condition residual reconstruction, enabling adaptive compensation for attenuation and scattering while retaining reliable information from the observations. We further introduce a spectral–spatial refinement module that jointly aggregates spectral and spatial neighborhoods to suppress local artifacts without distorting spectral profiles. For systematic evaluation, we construct a paired real-world underwater HSI dataset spanning diverse objects, imaging distances, and degradation levels. Extensive experiments on this dataset show that UWHFormer achieves state-of-the-art performance, with a PSNR of 39.70, SSIM of 0.860, SAM of 1.339 and ERGAS of 4.34, which indicates improvements in both spatial reconstruction and spectral fidelity across a range of underwater conditions.

1 INTRODUCTION

Underwater hyperspectral imaging provides both spatial appearance and dense spectral signatures, making it valuable for marine perception, material recognition, and ecological monitoring. Unlike RGB imaging, underwater hyperspectral image (HSI) enables fine-grained analysis of object reflectance and wavelength-dependent absorption. However, real underwater HSI is highly degraded due to low illumination, sensor noise, scattering, and wavelength-dependent attenuation, which jointly corrupt spatial structures and spectral signatures.

The key challenge is that underwater HSI restoration is not merely a spatial denoising problem. Water selectively absorbs and scatters light at different wavelengths, causing band-dependent intensity loss and spectral distortion. As a result, restoring each band independently may improve local image quality but break spectral consistency, while enforcing spectral smoothness alone may over-smooth spatial details. This creates a coupled restoration problem where spatial fidelity and spectral fidelity should be optimized jointly, as illustrated in Figure 1.

Existing underwater enhancement methods mainly focus on RGB images and cannot preserve dense spectral signatures (Pang et al., 2024). Generic HSI denoising methods exploit spectral-spatial correlations, but they are usually designed for synthetic Gaussian, stripe, or mixed noise, and do not explicitly model underwater wavelength-dependent degradation (Wei et al., 2020). Recent Transformer-based restoration models capture long-range spatial–spectral dependencies but often process bands uniformly, leaving band-specific degradation and coherent spectral recovery underexplored in complex underwater environments (Li et al., 2023a). Progress is further constrained by

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

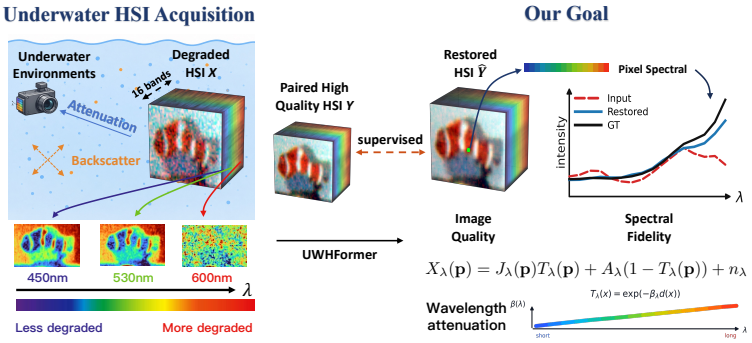

Figure 1: **Illustration of the underwater HSI restoration problem.** Underwater imaging introduces wavelength-dependent attenuation and backscatter, causing degraded spatial appearance and distorted spectral signatures across bands. Our goal is to recover a high-quality HSI with faithful spectral structure while exploiting physical cues such as wavelength-dependent attenuation.

the lack of paired real-world benchmarks for underwater HSI restoration (Li et al., 2020). Consequently, existing evaluations rely on synthetic degradations or limited controlled samples that may not faithfully represent real underwater imaging conditions (Li et al., 2024).

To address these limitations, we construct a paired underwater HSI dataset of short-exposure degraded observations and long-exposure references, capturing realistic optical degradation and sensor noise. It includes diverse objects, object-disjoint evaluation splits, and distance-shift subsets for testing generalization. Building on this dataset, we further introduce UWHFormer, a spectrally coherent Transformer that corrects band-dependent degradation in UnderWater HSIs while preserving spatial details and meaningful spectral relationships. Our main contributions are summarized as follows:

- **Underwater HSI Restoration Task Formulation**: To the best of our knowledge, we formulate underwater HSI restoration for the first time as a distinct spatial–spectral reconstruction task that jointly recovers fine spatial details and faithful spectral signatures under wavelength-dependent degradation. Beyond conventional RGB enhancement and generic HSI restoration, our formulation explicitly addresses coupled underwater spatial-spectral distortions and provides a paired real-world benchmark with object-level splits and distance-shift subsets.
- **Spectrally Coherent Restoration Transformer**: We propose UWHFormer, a spectrally coherent restoration transformer, comprising water-conditioned state modeling, a physics-guided restoration transformer, and spectral–spatial refinement. UWHFormer estimates physically interpretable, wavelength-dependent degradation states and integrates them with image features to condition residual reconstruction. The resulting spectral–spatial interactions correct spectral distortions while preserving fine-scale spatial structures, yielding spectrally coherent and spatially faithful restoration.
- **Strong Performance and Generalization**: Extensive experiments demonstrate that our method consistently outperforms strong underwater HSI restoration baselines while maintaining robust performance under imaging-distance shifts.

Codes and data will be made publicly available upon acceptance.

2 Related Work

Our work is related to three research areas: general image restoration, hyperspectral image restoration, and underwater image restoration. We briefly review representative studies in these areas and distinguish them from our proposed underwater hyperspectral restoration framework.

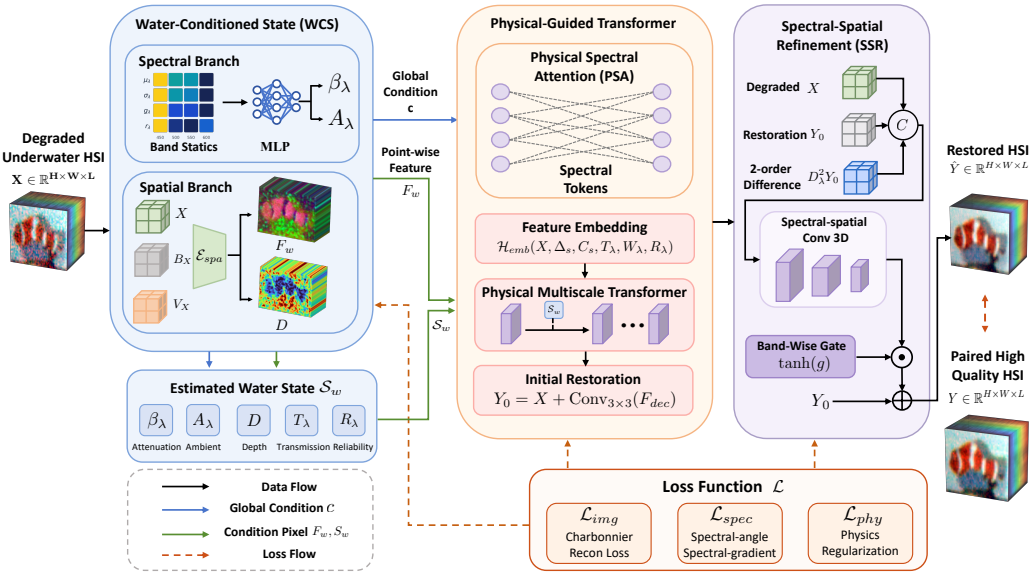

Figure 2: **Overall pipeline of UWHFormer.** Given a degraded underwater HSI X, WCS estimates latent water states $\mathcal{S}_w$ and point-wise features to guide the physics-guided Transformer restoration. The restored result is further refined by SSR to obtain the final HSI $\hat{Y}$.

2.1 General Image Restoration

Image restoration aims to recover high-quality images from degraded observations. Classical methods generally rely on hand-crafted priors or physical models, such as the dark channel prior for estimating transmission and atmospheric light (He et al., 2011). Deep networks subsequently shifted restoration toward data-driven feature learning, represented by the multi-stage MPRNet (Zamir et al., 2021) and the simplified NAFNet baseline (Chen et al., 2022). Following the development of self-attention and visual Transformers (Vaswani et al., 2017; Dosovitskiy et al., 2021; Liu et al., 2021), Uformer employs a U-shape Transformer (Wang et al., 2022), while Restormer develops efficient attention for high-resolution restoration (Zamir et al., 2022). Although these architectures provide strong spatial modeling capabilities, they are mainly designed for RGB images and do not explicitly address dense spectral correlations or wavelength-dependent underwater degradation.

2.2 Hyperspectral Image Restoration

Hyperspectral image restoration requires simultaneous recovery of spatial details and spectral signatures. Traditional approaches exploit priors such as total variation (Rudin et al., 1992), non-local similarity, and low-rank spectral representation, as exemplified by NGMeet (He et al., 2019). Deep methods instead learn nonlinear spatial-spectral mappings: QRNN3D employs quasi-recurrent 3D modeling (Wei et al., 2020), while MAN uses mixed attention to capture spectral dependencies (Lai & Fu, 2023). Transformer-based methods further improve spatial-spectral dependency modeling (Li et al., 2023a;b; Lai et al., 2023; Fu et al., 2024b). Diffusion and state-space architectures have also recently been introduced into HSI restoration (Pang et al., 2024; Liu et al., 2024; Fu et al., 2024a). Nevertheless, most existing methods are developed for in-air scenarios, targeting generic noise or synthetic degradation, and do not explicitly model underwater absorption, scattering, and non-uniform band attenuation.

2.3 Underwater Image Restoration

Underwater imaging is affected by wavelength-dependent absorption and scattering, leading to color cast, contrast loss, and backscatter. Early studies describe this degradation through physical image formation models that separate direct transmission from scattered light (Schechner & Karpel, 2004), with the revised model further distinguishing underwater imaging from atmospheric haze forma-

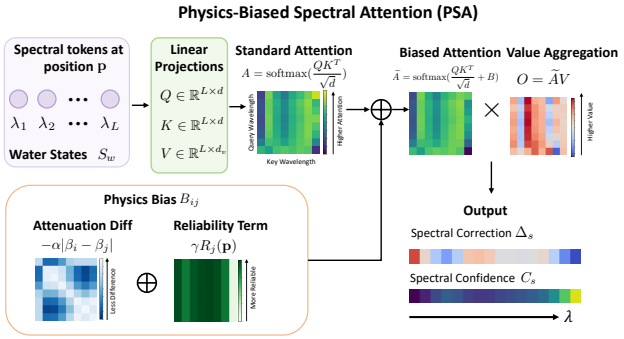

Figure 3: **Illustration of the Physics-Biased Spectral Attention (PSA) module.** For each spatial position $\mathbf{p}$, PSA represents spectral bands as tokens and introduces a physics bias B_{ij} based on attenuation differences and observation reliability to guide spectral attention. The biased attention produces spectrally coherent features, including the spectral correction map Δ_s and confidence representation C_s.

tion (Akkaynak & Treibitz, 2018). Physical-prior methods estimate transmission, background light, or color attenuation using UDCP (Drews Jr et al., 2013), Haze-Lines (Berman et al., 2021), statistical background-light modeling (Song et al., 2020), and Sea-Thru (Akkaynak & Treibitz, 2019). Learning-based studies include the UIEB benchmark (Li et al., 2020) and transmission-guided multicolor embedding in Ucolor (Li et al., 2021).

However, these methods mainly address three-channel RGB images. Although conditional diffusion has been explored for joint underwater HSI restoration and target detection (Li et al., 2024), jointly recovering image quality and spectral structure under interpretable water guidance remains insufficiently studied. Our method therefore combines explicit water-condition modeling, physics-guided spectral interaction, and spectral-spatial refinement for underwater hyperspectral image restoration.

3 Method

3.1 Underwater HSI Restoration Task Formulation

Given a degraded underwater hyperspectral image $X \in \mathbb{R}^{H\times W\times L}$, our objective is to recover a high-quality HSI $\hat{Y}$ with improved spatial quality and spectral fidelity. Unlike conventional hyperspectral restoration methods that learn a fixed degradation-to-reconstruction mapping, underwater HSI restoration must account for wavelength-dependent attenuation, spatially varying transmission, and non-uniform observation reliability.

To address these challenges, we propose UWHFormer, a water-conditioned spectral-spatial restoration framework. It first estimates latent underwater degradation states from the degraded observation and then uses them to dynamically guide the restoration process. Specifically, UWHFormer consists of Water-Conditioned State modeling (WCS), a physics-guided multiscale restoration transformer equipped with Physics-biased Spectral Attention (PSA), and Spectral-Spatial Refinement (SSR).

The overall restoration process of UWHFormer is illustrated in Figure 2. Given a degraded underwater HSI X, the overall restoration process is then formulated as

$$\hat{Y} = \mathcal{R}\left(X; \mathbf{c}, \mathbf{F}_w, \mathcal{S}_w\right), \tag{1}$$

where $\mathcal{R}(\cdot)$ represents the overall restoration process consisting of water-conditioned spectral-spatial reconstruction and residual refinement. Here, $\mathbf{c}$, $\mathbf{F}_w$, and $\mathcal{S}_w$ represent the global condition, pixel-wise physical feature, and latent water states extracted from the degraded input, respectively.

3.2 Water-Conditioned State Modeling

Underwater hyperspectral degradation varies across wavelengths and spatial locations. Following the underwater image formation model (Akkaynak & Treibitz, 2018; 2019), the degraded observa-

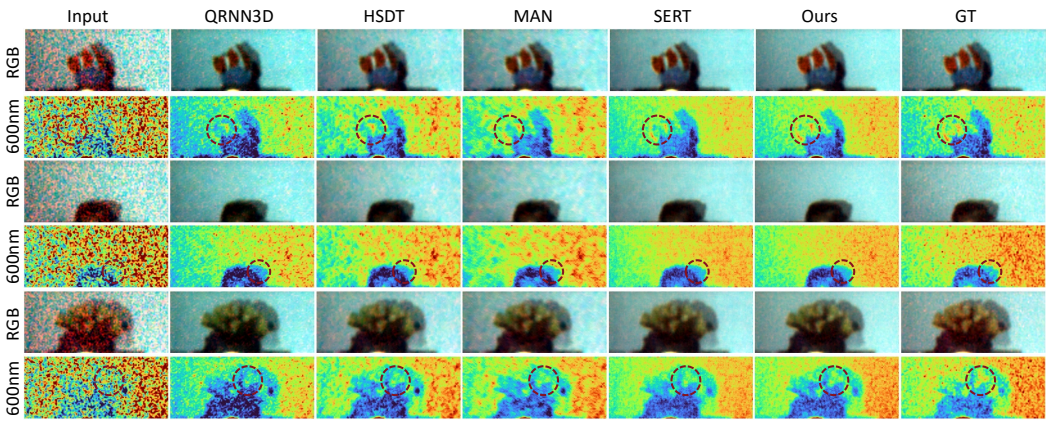

Figure 4: **Qualitative comparison of different methods on representative test scenes.** For each scene, the first row shows the RGB image, while the second row presents the corresponding long-wavelength band at 600 nm. The circles highlight regions with noticeable restoration differences. Compared with existing methods, UWHFormer better suppresses degradation and preserves object structures, particularly in the severely degraded 600 nm band.

tion is written as

$$X_\lambda(\mathbf{p}) = J_\lambda(\mathbf{p}) \cdot T_\lambda(\mathbf{p}) + A_\lambda \cdot (1 - T_\lambda(\mathbf{p})) + n_\lambda, \tag{2}$$

where $J_\lambda(\cdot)$ denotes the latent scene radiance, A_λ represents the wavelength-dependent ambient component, $T_\lambda(\mathbf{p})$ is the transmission term, and n_λ denotes noise. The transmission can be expressed as

$$T_\lambda(\mathbf{p}) = \exp[-\tau \cdot \beta_\lambda \cdot D(\mathbf{p})], \tag{3}$$

where β_λ and $D(\mathbf{p})$ describe the wavelength-dependent attenuation tendency and spatial path-length variation, respectively, and τ is a learnable global optical scale controlling the overall attenuation strength. Since our reference is still acquired underwater, these quantities are treated as latent guidance rather than calibrated physical measurements.

Based on this prior, we design WCS as a lightweight dual-branch encoder. Its spectral branch describes each band using

$$z_\lambda = [\mu_\lambda, \sigma_\lambda, g_\lambda, r_\lambda], \tag{4}$$

where μ_λ, σ_λ, g_λ, and r_λ denote the band intensity, variation, gradient response, and spectral residual, respectively. These descriptors are encoded by a shared multilayer perceptron to estimate the relative attenuation spectrum β_λ and ambient-light state A_λ.

In parallel, the spatial branch models pixel-wise degradation variation. It processes the degraded HSI X together with its brightness B_X, spectral variation V_X through a lightweight convolutional encoder $\mathcal{E}_{\text{spa}}$:

$$(\mathbf{F}_w(\mathbf{p}), D(\mathbf{p})) = \mathcal{E}_{\text{spa}}(X, B_X, V_X). \tag{5}$$

The outputs of the two branches are then combined according to Eq. (3) to obtain $T_\lambda(\mathbf{p})$. We further define the degradation strength as $W_\lambda(\mathbf{p}) = 1 - T_\lambda(\mathbf{p})$ and compute the observation reliability as

$$R_\lambda(\mathbf{p}) = \frac{|\mu_\lambda(\mathbf{p})|}{|\mu_\lambda(\mathbf{p})| + |X_\lambda(\mathbf{p}) - \mu_\lambda(\mathbf{p})| + \epsilon} \cdot T_\lambda(\mathbf{p}). \tag{6}$$

The estimated attenuation, ambient light, depth, transmission, degradation, and reliability states constitute the latent water condition $\mathcal{S}_w$, which provides physical guidance for subsequent spectral interaction and Transformer restoration.

3.3 Physics-Guided Restoration Transformer

The estimated water states provide explicit descriptions of underwater degradation, but they cannot directly recover the lost spectral and spatial information. Therefore, we design a physics-guided

restoration Transformer to adaptively exploit spectral correlation and spatial structures in different underwater scenes.

Physics-Guided Spectral Interaction. Underwater degradation is highly wavelength-dependent, where different bands suffer from different attenuation and reliability. Conventional spectral attention treats all wavelength interactions equally and may introduce information from severely degraded bands. To address this issue, we introduce a Physics-biased Spectral Attention (PSA) mechanism, where the attention distribution is explicitly guided by the estimated water states.

For each spatial position $\mathbf{p}$, the spectral dimension is regarded as a sequence of wavelength tokens. Each token combines the observed intensity, spectral residual, wavelength coordinate, and water-conditioned states from WCS. Given the query $\mathbf{Q_p}$, key $\mathbf{K_p}$, and value projections, the spectral attention is computed as

$$\mathbf{A_p} = \mathrm{Softmax}\left(\frac{\mathbf{Q_p}\mathbf{K_p}^\top}{\sqrt{d}} + \mathbf{B_p}\right), \tag{7}$$

where d denotes the feature dimension. The physical bias term is defined as

$$\mathbf{B}_{ij}(\mathbf{p}) = -\alpha|\beta_i - \beta_j| + \gamma R_j(\mathbf{p}), \tag{8}$$

where α and γ are learnable positive parameters. The first term reduces information exchange between wavelength bands with different attenuation characteristics, while the second term encourages attention toward more reliable observations. Consequently, PSA enables the network to selectively aggregate spectrally compatible information instead of performing wavelength-independent feature mixing.

The resulting spectral features are projected into a spectral correction map Δ_s and a spectral confidence representation C_s. Together with the estimated transmission, degradation, and reliability maps, they are embedded as the initial feature representation:

$$F_0 = \mathcal{H}_{emb}(X, \Delta_s, C_s, T_\lambda, W_\lambda, R_\lambda), \tag{9}$$

where $\mathcal{H}_{emb}$ denotes a convolutional embedding layer.

Water-Conditioned Multiscale Transformer The embedded features are processed by a multiscale U-shaped restoration Transformer. The backbone adopts a hybrid spectral-spatial attention design, where channel-wise attention is used to capture global feature interaction, while rectangular spatial attention models directional structures in the spatial domain. In addition, a gated depth-wise feed-forward network is employed to enhance local feature transformation.

Different from generic restoration Transformers, the proposed backbone is continuously conditioned on the water states estimated by WCS. For each Transformer block, the global water-condition vector $\mathbf{c}$ generates channel-wise modulation parameters, while the pixel-wise feature map F_w produces spatially varying modulation:

$$\widetilde{F}_i = \gamma_i(\mathbf{c}, F_{w,i}) \odot \mathrm{LN}(F_i) + \delta_i(\mathbf{c}, F_{w,i}), \tag{10}$$

where γ_i and δ_i denote adaptive scale and shift parameters. This design allows different spatial regions and wavelength responses to receive different restoration strategies according to their estimated degradation conditions. Therefore, the water information is deeply integrated into the restoration process rather than being used as an additional input constraint.

The decoder predicts the residual degradation with respect to the observed HSI:

$$Y_0 = X + \mathrm{Conv}_{3\times3}(F_{dec}), \tag{11}$$

where F_{dec} denotes the decoded feature representation. The residual learning strategy preserves reliable observations in the degraded input while focusing the network capacity on correcting wavelength-dependent degradation.

3.4 Spectral-Spatial Refinement

Although the Transformer effectively restores global spectral-spatial structures, local spectral fluctuations and spatial artifacts may remain due to complex underwater degradation. To further improve spectral continuity, we introduce a lightweight Spectral-Spatial Refinement (SSR) module.

Given the initial restoration Y_0, SSR constructs a spectral-spatial volume from the degraded input, the restored image, and the second-order spectral difference:

$$V = \mathrm{Stack}\left(X, Y_0, D_\lambda^2 Y_0\right), \tag{12}$$

where D_λ^2 denotes the second-order spectral difference operator. The refinement residual is predicted by a lightweight 3D convolutional network:

$$\hat{Y} = Y_0 + \tanh(g) \odot Conv_{3D}(V), \tag{13}$$

where g is a learnable band-wise gate initialized to zero. This initialization makes SSR behave as an identity mapping at the beginning of training and allows it to gradually learn wavelength-dependent refinement. By jointly modeling spectral and spatial neighborhoods, SSR reduces local artifacts while preserving the spectral shape of the restored HSI.

3.5 Loss Function

The proposed network is trained in an end-to-end manner using paired degraded and reference underwater HSIs. The loss function combines reconstruction, spectral consistency, and physical regularization:

$$\mathcal{L} = \mathcal{L}_{img} + \lambda_s \mathcal{L}_{spec} + \lambda_p \mathcal{L}_{phy}, \tag{14}$$

where $\mathcal{L}_{img}$ is the Charbonnier reconstruction loss, $\mathcal{L}_{spec}$ preserves spectral fidelity through spectral-angle and spectral-gradient constraints, and $\mathcal{L}_{phy}$ regularizes the latent water states estimated by WCS.

Specifically, $\mathcal{L}_{phy}$ constrains the spectral smoothness of the attenuation and transmission states, the continuity of the ambient-light response, and the spatial regularity of the estimated depth representation. Different from enforcing an exact physical decomposition, these constraints only provide soft physical priors to improve the plausibility and stability of the learned water states.

4 Experiment

4.1 Dataset

As publicly available underwater hyperspectral datasets remain scarce, we evaluate UWHFormer on our collected real-world underwater HSI dataset, which, to the best of our knowledge, is the first exposure-paired dataset for this task. Each pair comprises a short-exposure observation (10 ms) and a long-exposure reference (50 ms) captured from the same viewpoint. The former exhibits low signal-to-noise ratios and weak spectral responses, while the latter provides a higher-quality underwater reference. We use an object-disjoint split of 85 training, 18 validation, and 24 test samples. Each HSI contains 16 bands spanning 450–600 nm at 10 nm intervals, with a spatial resolution of 96×224. Additional acquisition and dataset details are provided in the supplementary material.

4.2 Evaluation Metrics

Following standard hyperspectral image restoration protocols, we employ PSNR, SSIM, SAM, and ERGAS for quantitative evaluation. PSNR is computed from the mean squared error over the complete hyperspectral cube, while SSIM is calculated independently for each band and then averaged over all wavelengths. SAM measures the angular discrepancy between the restored and reference spectral signatures at each pixel. ERGAS evaluates the relative global spectral reconstruction error by considering the normalized differences across all spectral bands. Higher PSNR and SSIM indicate better spatial reconstruction quality, whereas lower SAM indicates better spectral fidelity, and lower ERGAS value indicates better overall spectral reconstruction accuracy. All metrics are first computed for each test scene and subsequently averaged over the entire test set.

4.3 Implementation Details

The proposed model is implemented in PyTorch and trained from scratch on a single NVIDIA RTX 3090 GPU. Training is conducted for 400 epochs using $64 \times 64 \times 16$ patches with a batch size of

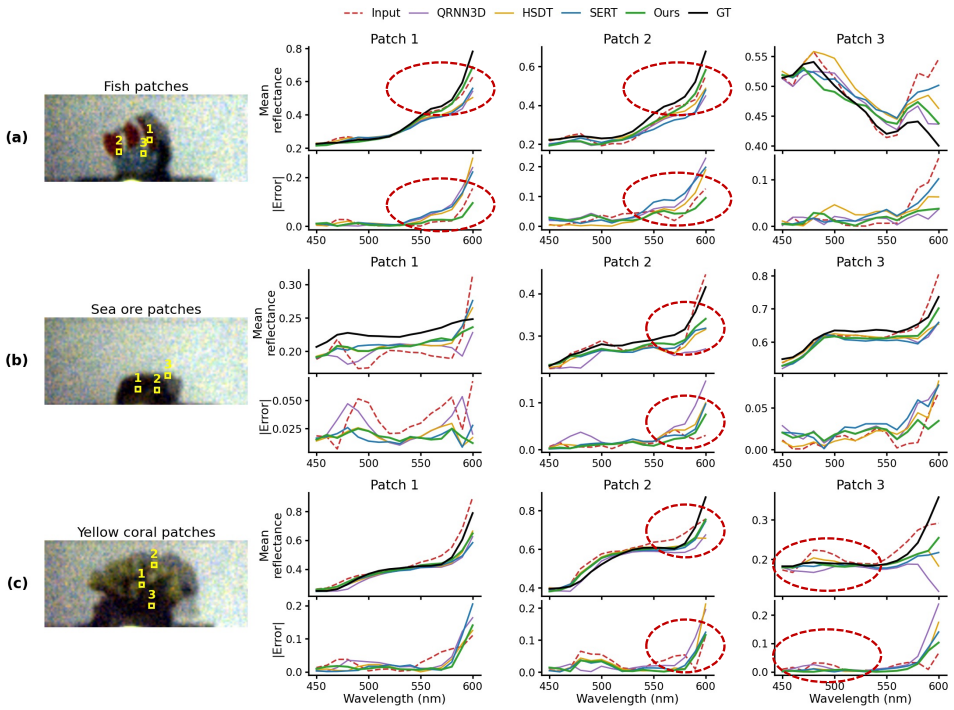

Figure 5: **Spectral comparison on selected object patches.** The left panels indicate the sampled 5×5 patches, and the right panels present their mean spectral signatures and absolute errors to the ground truth. UWHFormer achieves more accurate spectral recovery, particularly at longer wavelengths.

4. We adopt AdamW with an initial learning rate of 1×10^{-4} and a weight decay of 1×10^{-4}. The learning rate is linearly warmed up for 10 epochs and then decayed using a cosine schedule. The base feature dimension is 48, and the numbers of Transformer blocks are set to $[3, 3, 6, 3, 3]$ with $[2, 4, 8]$ attention heads across three scales. PSA uses two spectral-attention layers with a token dimension of 24 and three heads. Mixed-precision training, gradient clipping at 1.0, and exponential moving average with a decay of 0.995 are employed. The loss weights in Eq. (14) are set to $\lambda_{\text{spec}} = 0.05$, and $\lambda_{\text{p}} = 5 \times 10^{-4}$. Additional hyperparameter analyses are provided in the supplementary.

4.4 Comparative Evaluation

We compare UWHFormer with four representative hyperspectral image denoising methods on the standard test set. Specifically, QRNN3D (Wei et al., 2020) is selected as a recurrent 3D convolutional baseline, HSDT (Lai et al., 2023) and SERT (Li et al., 2023b) as representative Transformer-based methods, and MAN (Lai & Fu, 2023) as a state-space-model-based approach. All methods are retrained from scratch using the same object-disjoint split and evaluated on identical test images for a fair comparison.

For visual comparison, Figure 4 shows that all restoration methods improve the degraded observations, but noticeable differences remain in structural recovery and long-wavelength reconstruction. In the RGB results, competing methods may produce blurred object boundaries and residual color distortion, while UWHFormer preserves clearer object details. These differences become more evident in the 600 nm band, where the degraded input suffers from severe attenuation and noise. Compared with other methods, UWHFormer better restores object structures and suppresses residual artifacts in strongly degraded regions.

For spectral comparison, Figure 5 further evaluates the recovered spectral signatures on selected object patches. Existing methods recover the overall spectral trends but still exhibit larger deviations, especially in the long-wavelength range with stronger attenuation. In contrast, UWHFormer

Table 1: Quantitative comparison on the standard test set. Best results in every column are in bold.

Method	PSNR↑	SSIM↑	SAM↓	ERGAS↓
Input	31.79	0.58	3.401	12.51
QRNN3D	38.64	0.849	1.576	5.87
MAN	37.61	0.844	1.623	5.47
HSDT	37.95	**0.860**	1.536	5.42
SERT	39.28	0.857	1.391	4.52
Ours	**39.70**	**0.860**	**1.339**	**4.34**

Table 2: Component ablation on the standard test set.

Configuration	Params	PSNR↑	SSIM↑	SAM↓	ERGAS↓
Baseline	2.870M	39.31	0.839	1.455	4.57
+ WCS	2.926M	39.41	0.843	1.433	4.49
+ PSA	2.935M	39.59	0.857	1.372	4.36
+ SSR (Ours)	2.944M	**39.70**	**0.865**	**1.339**	**4.34**

maintains stable spectral responses in shorter wavelengths and achieves more accurate recovery toward longer wavelengths, with lower absolute errors across different object regions. These results demonstrate that UWHFormer effectively compensates for wavelength-dependent degradation while preserving inter-band spectral consistency.

As shown in Table 1, UWHFormer achieves the best overall performance. Compared with SERT, it improves PSNR from 39.28 dB to 39.70 dB and reduces SAM and ERGAS from 1.391 and 4.52 to 1.339 and 4.34, respectively. Although HSDT obtains the same reported SSIM of 0.860, UWHFormer performs better on the remaining metrics, indicating a more balanced improvement in spatial quality and spectral fidelity. These quantitative gains are also consistent with the clearer structures and lower spectral errors shown in the qualitative comparisons.

4.5 Ablation Study

We conduct component ablation experiments to evaluate the effectiveness of WCS, PSA, and SSR. The baseline retains the same multiscale restoration Transformer but removes the three proposed components. We then progressively introduce WCS, PSA, and SSR to examine their individual and complementary contributions and number of parameters. Further analysis of the estimated physical states and their role in guiding restoration is provided in Appendix.

As shown in Table 2, each proposed component consistently improves the restoration performance. Introducing WCS increases PSNR and reduces SAM and ERGAS, respectively, indicating that the estimated water-condition states provide useful degradation guidance. Adding PSA further improves all four metrics, with PSNR reaching 39.59 dB and SAM decreasing to 1.372. This demonstrates that incorporating attenuation and reliability cues into spectral interaction enables more effective information exchange across wavelength bands. Finally, SSR improves the full model, confirming its effectiveness in correcting residual spatial artifacts and local spectral fluctuations. Overall, the progressive gains show that WCS, PSA, and SSR are complementary rather than redundant.

5 Conclusion

In this paper, we proposed UWHFormer for underwater hyperspectral image restoration, combining water-conditioned state modeling, physics-guided Transformer restoration, and spectral-spatial refinement. Experiments on our paired real-world dataset demonstrate improved spatial reconstruction and spectral fidelity, while ablations support the contributions of the proposed components. Future work will investigate more diverse underwater conditions and physically calibrated acquisitions to improve generalization and interpretability.

AI USE STATEMENT

In this work, we used generative AI tools to provide feedback on research methodology and experimental design and to assist with the interpretation and presentation of experimental results. We did not use generative AI tools for other tasks requiring disclosure under the ICLR policy that are applicable to this work. Additionally, we used generative AI tools for literature search, manuscript structuring, language polishing, and suggestions for scientific figure design. All AI-assisted content and suggestions were reviewed and verified by the authors. We take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The source code for UWHFormer and the paired real-world underwater hyperspectral dataset will be made publicly available upon acceptance. Section 3 describes the proposed framework, its constituent modules, and the training objective. Section 4.1 details the dataset, object-disjoint splits, and spectral configuration; Section 4.2 describes the evaluation metrics; and Section 4.3 reports the implementation details, including hardware, model configuration, and training settings. Section 4.4 describes the baseline retraining and comparison setup. The supplementary material provides additional information on data acquisition, robustness evaluation under increased imaging distance, hyperparameter settings and sensitivity analysis, and computational complexity and efficiency.

REFERENCES

Derya Akkaynak and Tali Treibitz. A revised underwater image formation model. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6723–6732, 2018. doi: 10.1109/CVPR.2018.00703.

Derya Akkaynak and Tali Treibitz. Sea-thru: A method for removing water from underwater images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1682–1691, 2019. doi: 10.1109/CVPR.2019.00178.

Dana Berman, Deborah Levy, Shai Avidan, and Tali Treibitz. Underwater single image color restoration using haze-lines and a new quantitative dataset. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8):2822–2837, 2021. doi: 10.1109/TPAMI.2020.2977624.

Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pp. 17–33. Springer, 2022.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR*, 2021.

P. Drews Jr, E. do Nascimento, F. Moraes, S. Botelho, and M. Campos. Transmission estimation in underwater single images. In *2013 IEEE International Conference on Computer Vision Workshops*, pp. 825–830, 2013. doi: 10.1109/ICCVW.2013.113.

Guanyiman Fu, Fengchao Xiong, Jianfeng Lu, and Jun Zhou. Ssumamba: Spatial-spectral selective state space model for hyperspectral image denoising. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024a. doi: 10.1109/TGRS.2024.3446812.

Guanyiman Fu, Fengchao Xiong, Jianfeng Lu, Jun Zhou, Jiantao Zhou, and Yuntao Qian. Hyperspectral image denoising via spatial–spectral recurrent transformer. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–14, 2024b. ISSN 1558-0644. doi: 10.1109/tgrs.2024.3374953. URL http://dx.doi.org/10.1109/TGRS.2024.3374953.

Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2011. doi: 10.1109/TPAMI.2010.168.

Wei He, Quanming Yao, Chao Li, Naoto Yokoya, and Qibin Zhao. Non-local meets global: An integrated paradigm for hyperspectral denoising. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6861–6870, 2019. doi: 10.1109/CVPR.2019.00703.

Zeqiang Lai and Ying Fu. Mixed attention network for hyperspectral image denoising. *arXiv preprint arXiv:2301.11525*, 2023.

Zeqiang Lai, Yan Chenggang, and Ying Fu. Hybrid spectral denoising transformer with guided attention. In *Proceedings of the IEEE International Conference on Computer Vision*, 2023.

Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29:4376–4389, 2020. doi: 10.1109/TIP.2019.2955241.

Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Transactions on Image Processing*, 30:4985–5000, 2021.

Miaoyu Li, Ying Fu, and Yulun Zhang. Spatial-spectral transformer for hyperspectral image denoising. In *AAAI*, 2023a.

Miaoyu Li, Ji Liu, Ying Fu, Yulun Zhang, and Dejing Dou. Spectral enhanced rectangle transformer for hyperspectral image denoising. In *CVPR*, 2023b.

Qi Li, Jinghua Li, Tong Li, and Yan Feng. A joint framework for underwater hyperspectral image restoration and target detection with conditional diffusion model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:17263–17277, 2024. doi: 10.1109/JSTARS.2024.3464557.

Yang Liu, Jiahua Xiao, Yu Guo, Peilin Jiang, Haiwei Yang, and Fei Wang. Hsidmamba: Exploring bidirectional state-space models for hyperspectral denoising. *ArXiv*, abs/2404.09697, 2024. URL `https://api.semanticscholar.org/CorpusID:269149609`.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012–10022, 2021.

Li Pang, Xiangyu Rui, Long Cui, Hongzhong Wang, Deyu Meng, and Xiangyong Cao. Hir-diff: Unsupervised hyperspectral image restoration via improved diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3005–3014, June 2024.

Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

Y.Y. Schechner and N. Karpel. Clear underwater vision. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pp. I–I, 2004. doi: 10.1109/CVPR.2004.1315078.

Wei Song, Yan Wang, Dongmei Huang, Antonio Liotta, and Cristian Perra. Enhancement of underwater images with statistical model of background light and optimization of transmission map. *IEEE Transactions on Broadcasting*, 66(1):153–169, 2020. doi: 10.1109/TBC.2019.2960942.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17683–17693, June 2022.

Kaixuan Wei, Ying Fu, and Hua Huang. 3-d quasi-recurrent neural network for hyperspectral image denoising. *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *CVPR*, 2021.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*, 2022.

SUPPLEMENTARY MATERIAL

A OVERVIEW

This supplementary material provides additional details and experimental analyses that complement the main paper. We first describe the acquisition procedure and preprocessing pipeline of the proposed underwater hyperspectral dataset. We then present additional experimental evaluations, including robustness analysis under long-distance imaging conditions and hyperparameter sensitivity studies. Finally, we report the computational complexity and efficiency analysis of UWHFormer to further demonstrate its practical applicability.

B DATA ACQUISITION

We construct a paired underwater hyperspectral dataset to evaluate restoration performance under realistic underwater imaging conditions. The dataset contains diverse foreground targets, including coral-like objects, sponges, fabrics, stones, bottles, cans, and other materials with different shapes, colors, textures, and spectral reflectance characteristics. As illustrated in Fig. 6, the dataset includes both representative hyperspectral observations and diverse training scenes. The collected samples cover substantial variations in object appearance and spectral responses, providing a challenging benchmark for evaluating spatial reconstruction and spectral fidelity.

For each scene, a degraded underwater hyperspectral observation and its corresponding reference image are captured under the same imaging configuration, providing spatially aligned supervision for paired restoration. Specifically, the degraded inputs are acquired with a 10 ms exposure time, while the reference images are captured with a longer 50 ms exposure time to obtain higher-quality measurements with improved signal-to-noise ratio. Figure 6(a) presents representative input-reference pairs together with normalized intensity maps at several wavelengths. The selected wavelengths highlight the wavelength-dependent attenuation characteristics of underwater imaging. According to the estimated transmission model $T(\lambda) = \exp(-K_d(\lambda)d)$ at the imaging distance of 4.2 m, short wavelengths around 450 nm maintain high transmission (approximately 93%), whereas longer wavelengths such as 600 nm experience substantial attenuation with only about 36% signal remaining. Therefore, the selected spectral range captures both relatively well-preserved and severely degraded bands, providing meaningful spectral variations for evaluating restoration performance.

To balance spectral fidelity and computational efficiency, the original hyperspectral measurements are sampled at a 10 nm interval within the selected range. This sampling density is sufficient to preserve typical underwater material spectral characteristics, whose absorption and reflection variations generally occur at scales of 20–30 nm, while avoiding excessive redundancy caused by highly correlated adjacent 5 nm measurements. Such a spectral configuration enables the model to recover meaningful spectral structures while maintaining a manageable computational cost.

C ROBUSTNESS TO IMAGING DISTANCE

To further evaluate the robustness ability of UWHFormer under challenging imaging conditions, we conduct an additional robustness experiment on a reserved 6.3 m subset. This subset contains four scenes, including can, coral, stone, and tube. Compared with the training conditions, the increased imaging distance introduces stronger light attenuation and noise accumulation, providing a more challenging test for restoration under distribution shifts. The trained model is directly applied to these samples without retraining, fine-tuning, or scene-specific intensity adjustment.

As shown in Fig. 7, we compare the degraded inputs, restored results, and ground-truth references in both RGB visualization and representative spectral bands at 460, 520, and 600 nm. The short- and middle-wavelength bands preserve relatively clear structural information, while the 600 nm band suffers from substantial signal degradation due to stronger underwater attenuation and reduced transmission over long propagation distances. Despite these severe degradations, our method effectively suppresses noise fluctuations and reconstructs consistent spatial structures across different wavelengths. The restored results recover object boundaries and spectral responses that are closer to

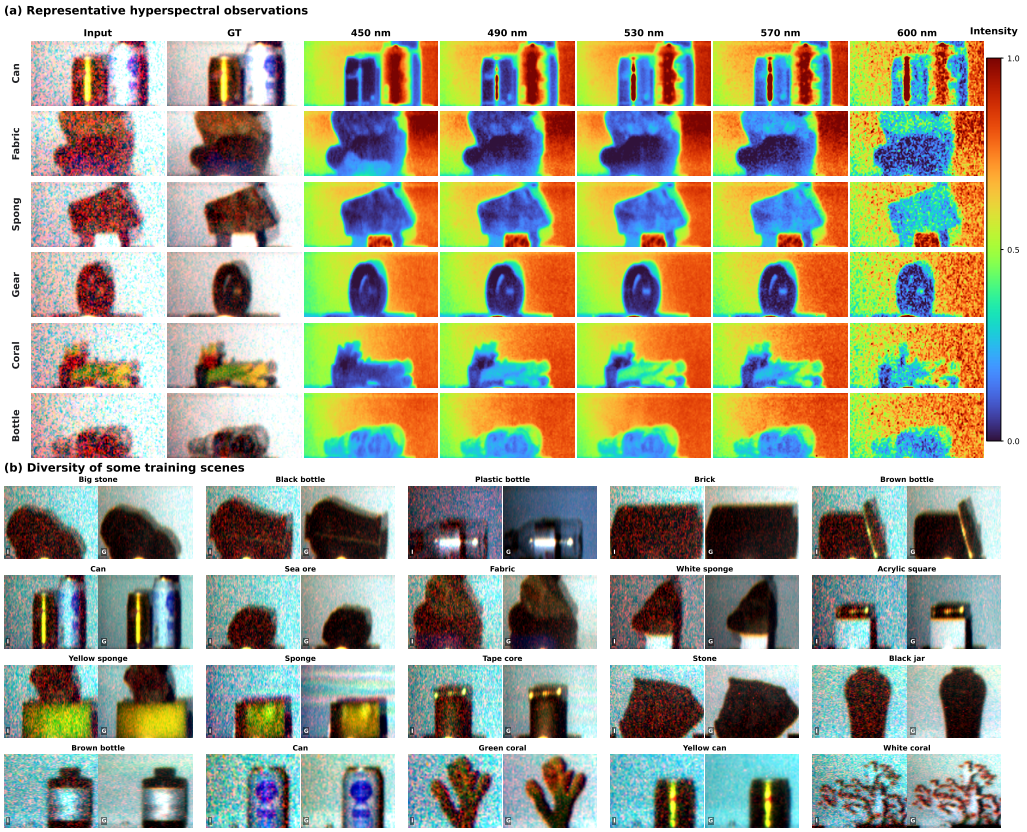

Figure 6: **Overview and diversity of the proposed underwater hyperspectral dataset.** (a) Representative paired observations, including degraded inputs, corresponding reference measurements, and normalized intensity maps at 450, 490, 530, 570, and 600 nm. The selected wavelengths reveal wavelength-dependent variations in signal attenuation and material appearance. (b) Representative training scenes containing targets with diverse shapes, colors, textures, and spectral characteristics. Each example shows the degraded input and its paired reference measurement. The colors and normalized intensity scales are used only for visualization.

the ground truth, demonstrating the effectiveness of the proposed water-conditioned spectral-spatial restoration strategy.

Quantitatively, UWHFormer achieves 28.93 dB PSNR, 0.702 SSIM, 5.397 SAM, and 28.05 ERGAS on the 6.3 m subset. Compared with the degraded input, it improves PSNR by 4.51 dB and SSIM by 0.291, while reducing SAM and ERGAS by 5.964 and 24.53, respectively. These improvements indicate that the proposed method maintains stable spatial reconstruction and spectral consistency even under unseen long-distance underwater imaging conditions.

D Analysis of Physics-Guided Restoration

Figure 8 illustrates how estimated water states guide reliability, spectral information exchange, and reconstruction at 600 nm.

Reliability-guided spectral information exchange. The clownfish body exhibits relatively high R_{600}. Its reddish surface can support a stronger 600 nm return under the available illumination, making this observation a potentially useful source of spectral information. This reliability enters PSA explicitly through $B_{ij} = -\alpha|\beta_i - \beta_j| + \gamma R_j$: higher source-band reliability increases its attention bias, while attenuation differences penalize incompatible interactions. PSA therefore favors reliable, attenuation-compatible information.

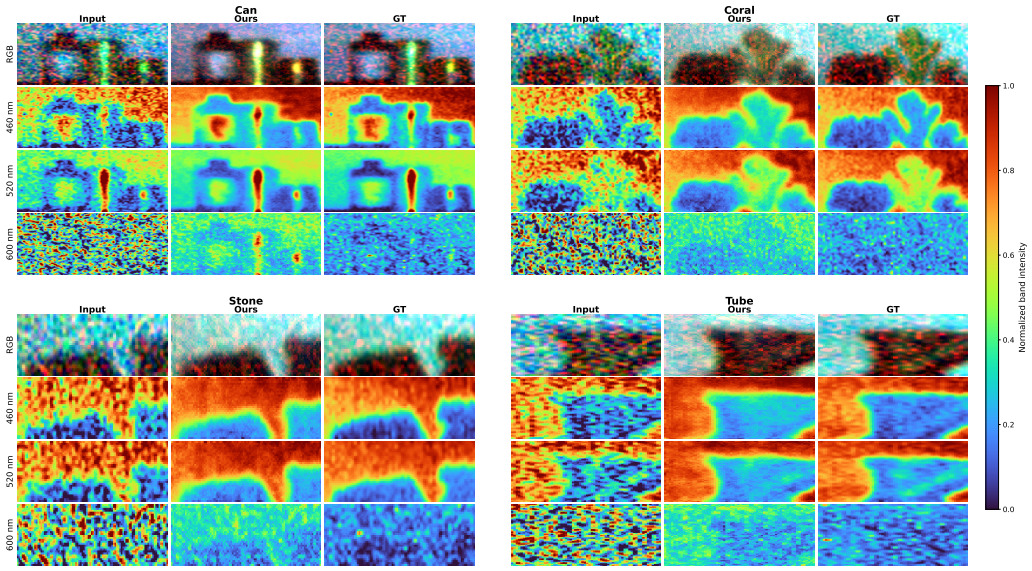

Figure 7: **Robustness evaluation on the reserved 6.3 m subset.** The results are presented on four scenes under increased imaging distance. For each scene, the columns show the degraded input, restored result, and reference measurement, while the rows contain RGB visualization and representative spectral bands at 460, 520, and 600 nm. This experiment evaluates the generalization capability of UWHFormer without retraining or scene-specific adjustment.

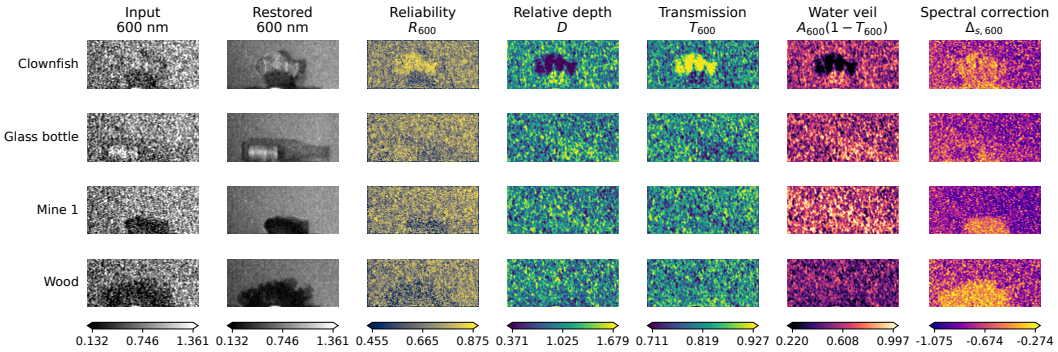

Figure 8: **Physics-guided restoration.** Rows show views of the clownfish, glass bottle, Mine 1, and wood. Columns show the degraded input, restored image, reliability R_{600}, relative path state D, transmission T_{600}, derived water veil $A_{600}(1 - T_{600})$, and spectral correction $\Delta_{s,600}$. Input and output share one intensity scale. Guidance maps use quantity-specific scales shared across rows, based on pooled interior-pixel 1st–99th percentiles over all 24 test acquisitions. A two-pixel boundary is excluded; clipping affects display only.

Conversely, the dark Mine 1 and wood regions exhibit lower R_{600}. Their weaker responses motivate downweighting the 600 nm observation as a source of information, reducing the propagation of unreliable features to other bands. Restoration at these positions can instead draw on more reliable, attenuation-compatible wavelengths at the same spatial location. The heterogeneous bottle response further motivates spatially varying guidance rather than a uniform correction across the object.

Propagation-conditioned reconstruction. The clownfish body also exhibits lower effective D, higher T_{600}, and a weaker water veil. These states are coupled by $T_{600} = \exp(-\tau\beta_{600}D)$ and $A_{600}(1 - T_{600})$. Higher transmission directly increases the reliability estimate, linking propagation guidance to spectral attention. Meanwhile, transmission, degradation, and reliability states enter

feature embedding, and water-conditioned features modulate the restoration blocks. This conditions reconstruction on estimated attenuation and observation quality.

The spectral correction $\Delta_{s,600}$ is an internal feature. These maps illustrate physics-guided conditioning without establishing calibrated physical parameters or causal performance gains.

E Hyperparameter Analysis

The loss weights are selected to keep image reconstruction as the dominant objective while providing sufficient spectral and physics-inspired regularization. We set the reconstruction weight to 1, the spectral-angle and spectral-gradient weights to 0.05 and 0.15, respectively, and the physics regularization weight λ_{phy} to 5×10^{-4}. The relatively small physical weight is used because the water-related variables represent latent degradation cues rather than calibrated physical measurements. We evaluate the sensitivity by halving and doubling the spectral and physics-related loss weights while keeping the data split, random seed, and training schedule unchanged. Across these settings, PSNR varies only from 39.66 to 39.71 dB, while SAM and ERGAS remain within $[1.335, 1.351]$ and $[4.323, 4.359]$, respectively. These minor variations demonstrate that UWHFormer maintains stable restoration performance over a reasonable range of loss configurations.

For the proposed PSA module, the spectral token dimension, number of attention heads, and number of layers are set to 24, 3, and 2, respectively. This configuration provides an 8-dimensional representation per head, enabling sufficient capacity for inter-band interaction without introducing excessive computational overhead. The attenuation and reliability bias strengths are initialized to -1.5 before the softplus transformation and are subsequently optimized during training. Therefore, their effective contributions remain positive and adaptively learned rather than manually fixed. Overall, the adopted configuration achieves a practical balance among restoration accuracy, spectral fidelity, physics-inspired regularization, and computational efficiency.

F Model Complexity and Efficiency

Table 3: Computational complexity comparison. Runtime and peak memory are measured in FP32 on an NVIDIA RTX 3090 using an input of size $96 \times 224 \times 16$.

Method	Params (M)	FLOPs (G)	Time (ms)	Memory (MB)
QRNN3D	0.860	0.149	13.901	242.98
HSDT	0.515	16.809	26.486	211.07
MAN	1.394	0.280	35.028	330.29
SERT	1.879	87.353	35.221	165.36
Baseline	2.870	21.057	22.646	227.41
Baseline + WCS	2.926	21.057	22.569	224.08
Baseline + WCS + PSA	2.935	28.765	38.187	251.15
Ours Full	2.944	28.766	38.946	251.15

Table 3 reports the computational complexity of different methods. All inference times are measured in FP32 on an NVIDIA RTX 3090 using an input HSI of size $96 \times 224 \times 16$. Our complete model contains 2.944M parameters and requires 28.766G FLOPs, maintaining the overall model size below 3M parameters. Compared with SERT, our method reduces the computational cost from 87.353G to 28.766G FLOPs while incorporating explicit water-conditioned and spectral refinement mechanisms.

The component-wise results show that the proposed modules introduce limited parameter overhead. WCS increases the parameter count from 2.870M to 2.926M, while adding PSA results in 2.935M parameters. The complete model with SSR contains 2.944M parameters. These results indicate that WCS, PSA, and SSR improve the model's physical and spectral modeling capabilities without substantially increasing its parameter count. Overall, the proposed method achieves a practical balance between restoration performance and computational complexity.