Natural Resources for Human Health
eISSN: 2583-1194

Original Research
Received 14 May 2026
Revised 22 June 2026
Accepted 05 July 2026

Natural Resources for Human Health 2026, Vol. 6 (11s): 1133–1156

KEYWORDS:
underwater image enhancement; physics-guided deep learning; wavelet transform; uncertainty quantification; lightweight neural networks.

AquaRestore-UC: A Lightweight Physics-Guided Wavelet Network for Underwater Image Enhancement with an Auxiliary Uncertainty Head

I. Sengol[1]*, D. Gnanadurai[2]

[1]Research Scholar, Department of Computer Science & Engineering, St. Joseph University, Chumoukedima, Nagaland, India.
[2]Professor, Department of Computer Science & Engineering, St. Joseph University, Chumoukedima, Nagaland, India
*Corresponding author: imsengol@gmail.com

ABSTRACT: Underwater images are degraded by wavelength-dependent attenuation, backscatter, scattering-induced blur, and colour distortion. We propose AquaRestore-UC, a lightweight network combining a physics-guided restoration branch, a fixed Haar-wavelet frequency branch, and a multi-colour-space spatial branch, fused via attention gating, with an auxiliary per-pixel uncertainty head trained using a heteroscedastic loss. The uncertainty head is trained but not post-hoc calibrated; only its raw, uncalibrated coverage statistics are reported. On a completed, single-seed (seed 42) experiment at 64×64 resolution using a fixed, custom 720/80/90 UIEB split, AquaRestore-UC improves mean PSNR from 42.00 to 45.26 dB (+3.26 dB) and mean SSIM from 0.758 to 0.877 relative to raw input, across all 90 test images. Generic and parameter-matched CNN controls trained under matched conditions show smaller improvements. An exploratory, non-pre-registered paired statistical reanalysis finds these image-level differences unlikely to be due to chance, but this analysis does not establish superiority over published state-of-the-art methods, nor does it attribute the observed gains to any individual architectural component; it treats a single trained instance of each method as the unit of comparison, not multi-seed evidence. A preliminary, restoration-only three-seed extension (seeds 123 and 2024, uncertainty head untrained) is numerically consistent with the seed-42 result. Twelve of the 90 test images score lower on PSNR than raw input. We report a diagnosed and corrected heteroscedastic-uncertainty training failure and a descriptive, uncalibrated uncertainty-coverage analysis. Higher-resolution training, published-method baselines, component ablations, a fully matched multi-seed study, and post-hoc uncertainty calibration remain future work. This manuscript makes neither a state-of-the-art performance claim nor a calibrated-uncertainty claim.

HIGHLIGHTS

- A lightweight architecture (28,399 parameters) combining a physics-guided restoration branch, a fixed Haar-wavelet frequency branch, and a multi-colour-space spatial branch, fused through an attention-gated module.
- At 64×64 resolution, single seed, on a custom 720/80/90 UIEB split: mean PSNR improves by 3.26 dB and mean SSIM from 0.758 to 0.877 over raw input across all 90 test images.
- A diagnosed and corrected heteroscedastic-uncertainty training failure mode, reported as a methodological contribution independent of the architecture.

- A structural two-stage training protocol whose backbone-freeze guarantee is verified by an executable assertion rather than assumed.
- A three-seed preliminary extension (seeds 42, 123, 2024) and a parameter-matched CNN control (28,252 parameters, three seeds), both numerically consistent with the primary result and reported descriptively.
- All 90 test-set outcomes reported in full, including the 12 images where the model underperforms raw input.

1. INTRODUCTION

Underwater imagery is central to marine biology surveys, underwater archaeology, aquaculture inspection, and autonomous underwater vehicle (AUV) navigation, yet it is systematically degraded relative to in-air photography. Light of different wavelengths attenuates at different rates in water, producing a characteristic blue–green colour cast; multiply-scattered light from suspended particulates (backscatter) reduces contrast and can dominate the signal at range; and forward scattering blurs fine detail. A single enhancement model must generalize across a wide range of degradation severities from a single RGB image, typically without scene depth or water-quality sensors.

Existing underwater image enhancement (UIE) methods fall into three lineages. Physics-based methods invert an underwater image formation model to recover scene radiance from estimated transmission and backscatter; these generalize well when parameters are estimated correctly but do not by themselves restore detail lost to blur. Data-driven methods learn the enhancement mapping directly and correct colour and contrast well, but the strongest such networks are computationally heavy — a poor fit for real-time, embedded deployment — and typically emit a single deterministic output, despite the UIEB dataset's own authors acknowledging that its human-selected reference images are ambiguous in a meaningful fraction of cases [1]. Frequency-domain methods restore fine texture efficiently by separating frequency content but do not encode the physical degradation process.

This paper makes four contributions. First, a unified, lightweight architecture (AquaRestore-UC) combining a physics-guided branch, a fixed, non-learnable Haar wavelet branch, and a multi-colour-space spatial branch, fused through an attention-gated module and designed around a lightweight parameter budget from the outset. Second, a pixel-level auxiliary uncertainty head trained via a heteroscedastic loss, reported here as trained but not calibrated, with post-hoc calibration specified as future work rather than performed. Third, a fully reproducible experimental protocol with a fixed dataset split, fixed seeds, resumable checkpointing, and automatic export of every result table. Fourth, a diagnosed and corrected heteroscedastic-uncertainty training failure mode, together with an independent code-review round that found and fixed four further implementation issues before the reported experiment was run, and a structural two-stage training protocol whose backbone-freeze guarantee is checked by an executable assertion rather than assumed.

We evaluate under a fixed, custom train/validation/test protocol on UIEB [1], reporting PSNR and SSIM for a completed 64×64, single-seed experiment, a three-seed descriptive extension, a generic and a parameter-matched CNN control, a full failure-case analysis, and an exploratory statistical reanalysis. Higher-resolution evaluation, published-method baselines, component ablations, and uncertainty calibration are specified as future work rather than reported as completed.

2. RELATED WORK

Physics-based underwater restoration. The image formation model underlying most physics-based UIE work descends from atmospheric dehazing equations adapted to water. Akkaynak and Treibitz [2,3] showed that the conventional single-attenuation-coefficient model introduces systematic errors because direct-transmission and backscatter attenuation coefficients are not equal and both depend on range and reflectance, and proposed a revised model together with the Sea-thru method [3], which requires a co-registered depth map. This physically more accurate model is not directly applicable to the single-RGB, depth-free setting targeted here, so AquaRestore-UC's physics branch uses the simplified, single-coefficient formulation, following the same simplification adopted implicitly by most learned physics-guided networks, including Ucolor [4].

Data-driven and hybrid methods. Water-Net [1], proposed alongside the UIEB benchmark itself, fuses white-balance, histogram-equalization, and gamma-corrected variants of the input via a learned confidence map. Ucolor [4] embeds multiple colour spaces and guides its decoder with an estimated medium-transmission map. Shallow-UWnet [6] and the Five A+ Network (FA+Net) [7] pursue explicitly parameter-efficient designs, motivated by the same real-time and embedded-deployment concern that motivates this work; AquaRestore-UC adopts FA+Net's train/test split convention (800/90 on UIEB's paired images) for literature comparability.

Frequency and wavelet-domain methods. WF-Diff [8] combines a wavelet–Fourier interaction network with a diffusion-based frequency refinement module; WPFNet [9] fuses wavelet- and pixel-domain branches with a residual attention block; WaveletFormer [10] combines wavelet decomposition with a multi-scale transformer; Zhang et al. [11] weight wavelet sub-bands by visual perception. AquaRestore-UC's wavelet branch is architecturally simpler than any of these: a fixed, non-learnable two-level Haar transform with lightweight, partially weight-shared residual processing, with its own dedicated auxiliary supervision on the branch's coarse reconstruction, mirroring how the physics branch is supervised.

Uncertainty in underwater image enhancement. Fu et al.'s PUIE-Net [5] models the distribution of plausible enhancements via a conditional variational autoencoder, using a consensus process at inference. This is a heavier, generative approach to reference-label ambiguity, not designed for real-time, lightweight deployment. AquaRestore-UC instead adopts the substantially cheaper heteroscedastic-regression approach of Kendall and Gal [17], with post-hoc isotonic-regression calibration [19] specified as a proposed procedure rather than one that has been executed.

Evaluation metrics for UIE. PSNR and SSIM [14] are standard full-reference metrics; LPIPS [15] provides a learned perceptual alternative. Because UIEB reference images are themselves algorithmically generated and human-selected rather than physically ground-truth, no-reference metrics specific to the underwater domain are also standard: UCIQE [12] and UIQM [13].

3. PROBLEM FORMULATION

Given a degraded underwater image $I \in \mathbb{R}^{3\times H\times W}$, the goal is to estimate a restored image $\hat{J} \in [0,1]^{3\times H\times W}$ that approximates the latent, unobserved clean scene radiance J, together with a per-pixel scalar uncertainty $\sigma(x) > 0$ that reflects the model's estimated reliability at each spatial location x. The restoration mapping is learned end-to-end from paired data (I, J), where J is a human-selected reference image rather than a physically measured ground truth. The uncertainty mapping is learned via a heteroscedastic regression objective (Section 7) and is evaluated in this work only in its raw, uncalibrated form; no claim is made that $\sigma(x)$ corresponds to a calibrated probabilistic interval.

4. UNDERWATER IMAGE FORMATION MODEL

Following the classical, single-attenuation-coefficient underwater image formation model, per channel $c \in \{R, G, B\}$:

$$I_c(x) = J_c(x) \cdot t_c(x) + B_c^{\infty} \cdot (1 - t_c(x))$$

Equation (1)

where $I_c(x)$ is the observed pixel intensity, $J_c(x)$ is the latent clean scene radiance, $t_c(x) = exp\left(-\beta_c\, d(x)\right)$ is the wavelength-dependent transmission map with attenuation coefficient β_c and scene depth $d(x)$, and B_c^{∞} is the homogeneous backscatter (veiling light) at infinite distance.

This simplified form is adopted rather than the fully revised, range- and reflectance-dependent model of Akkaynak and Treibitz [2], because the revised model requires a co-registered depth map that is unavailable in the single-RGB UIEB setting used throughout this work. This is a deliberate, stated scope limitation (Section 17), not an oversight.

5. PROPOSED AQUARESTORE-UC ARCHITECTURE

AquaRestore-UC comprises three parallel branches operating on the input image, fused by a lightweight attention-gated module, followed by a reconstruction head and a parallel uncertainty head (Figure 1).

Natural Resources for Human Health
eISSN: 2583-1194

5.1 Physics-Guided Branch

A four-layer depthwise-separable convolutional encoder predicts a per-channel transmission map $t_c(x)$ (sigmoid-activated, clamped to $[0.05, 1.0]$) and a global backscatter estimate B_c (via global average pooling followed by a two-layer MLP, sigmoid-activated). A coarse restoration is obtained by analytically inverting Equation 1:

$$J_{\text{phys},c}(x) = \frac{I_c(x) - B_c \cdot (1 - t_c(x))}{\max(t_c(x), \varepsilon)}$$

Equation (2)

with $\varepsilon = 0.05$ for numerical stability. Both J_{phys} and $t_c(x)$ are passed to the fusion module; the branch totals approximately 0.03–0.05 M parameters.

5.2 Haar-Wavelet Branch

A fixed, non-learnable two-level Haar discrete wavelet transform (DWT) decomposes the input into low-frequency (illumination and colour) and high-frequency (edge and texture) sub-bands, implemented as strided depthwise convolutions with fixed Haar kernels. The low-frequency (LL) sub-band is processed by two residual depthwise-separable blocks; the three high-frequency sub-bands (LH, HL, HH) share a single residual block's weights across orientations for parameter efficiency. An inverse DWT reconstructs the branch's own coarse restoration, J_{wave}, which is directly supervised (Section 7) so that the branch's high-frequency parameters receive a dedicated gradient signal rather than relying solely on gradient that arrives indirectly through the shared final output. The Haar transform's perfect-reconstruction property (i.e. $\text{IDWT}(\text{DWT}(x)) = x$) was verified directly on the executed implementation, with a maximum absolute reconstruction error of 1.79×10^{-7}.

5.3 Spatial Colour Branch

The input is represented jointly in RGB, CIELab, and HSV colour spaces (nine channels total), processed by three depthwise-separable convolutional blocks with a CBAM-style channel-and-spatial attention gate, following the multi-colour-space rationale of Ucolor [4] in a substantially lighter form.

5.4 Attention-Gated Fusion

Each branch's output is projected to a common channel width via a 1×1 convolution, concatenated, passed through a Squeeze-and-Excitation channel-attention gate (reduction ratio 8) and a 7×7 spatial-attention gate, then fused via a depthwise-separable 3×3 convolution.

5.5 Reconstruction Head

Two residual depthwise-separable blocks followed by a 3×3 convolution predict a residual correction $R(x)$, added to the input image with a long skip connection:

$$\hat{J}(x) = clip\left(I(x) + R(x),\ 0,\ 1\right)$$

Equation (3)

5.6 Auxiliary Uncertainty Head

A parallel 1×1 convolution on the fused features predicts a single-channel, per-pixel log-variance $s(x) = log\sigma^2(x)$, shared across all three RGB channels rather than estimated per channel, clamped to $[-10, 5]$ at the output layer in the executed implementation. This head is trained but not post-hoc calibrated in this work (Section 17). As described in Section 8, this head is trained in a dedicated second stage, after the restoration backbone is frozen, rather than jointly with it.

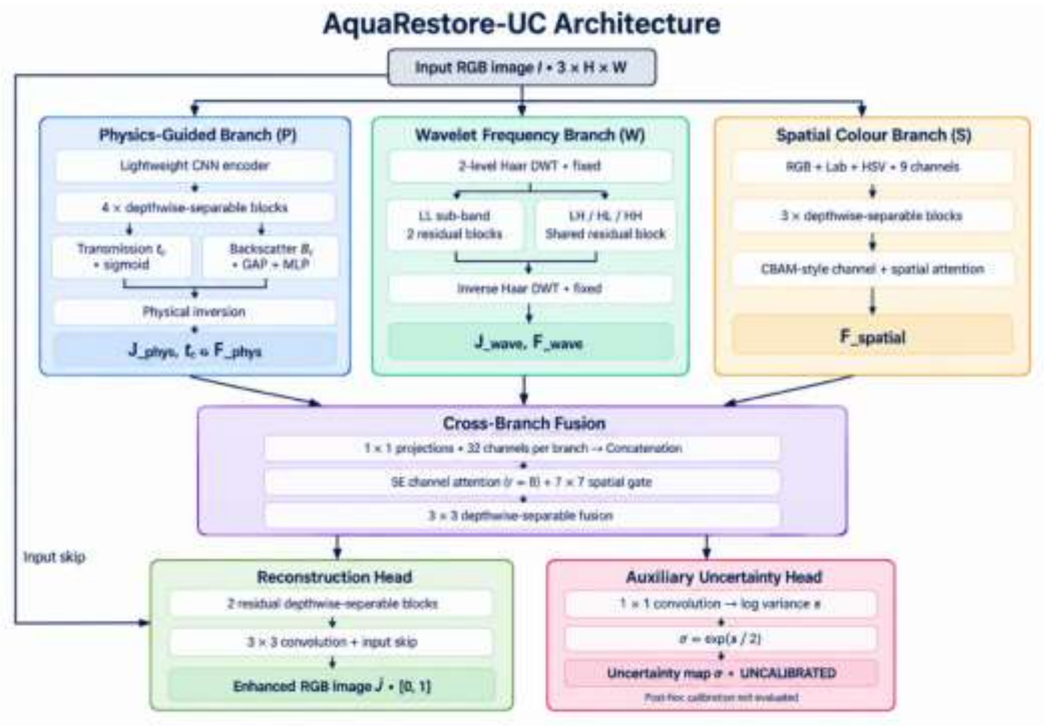

Figure 1: AquaRestore-UC architecture — physics-guided, wavelet, and spatial colour branches fused through an attention-gated module, with a reconstruction head and an auxiliary uncertainty head (trained, not yet calibrated).

6. MATHEMATICAL FORMULATION

The total training objective combines reconstruction, uncertainty, perceptual, structural, and branch-specific consistency terms:

$$\mathcal{L}_{\text{total}} = \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{unc}}\mathcal{L}_{\text{unc}} + \lambda_{\text{perc}}\mathcal{L}_{\text{perc}} + \lambda_{\text{ssim}}\mathcal{L}_{\text{ssim}} + \lambda_{\text{phys}}\mathcal{L}_{\text{phys}} + \lambda_{\text{wave}}\mathcal{L}_{\text{wave}} + \lambda_{\text{tv}}\mathcal{L}_{\text{tv}}$$

Equation (4)

All symbols are defined in Table A1 (Appendix). Default weights, given as a starting point subject to validation-set tuning, are $\lambda_{\text{rec}} = 1.0$, $\lambda_{\text{unc,max}} = 0.5$, $\lambda_{\text{perc}} = 0.04$, $\lambda_{\text{ssim}} = 0.3$, $\lambda_{\text{phys}} = 0.2$, $\lambda_{\text{wave}} = 0.2$, $\lambda_{\text{tv}} = 0.01$.

7. COMPLETE LOSS FUNCTIONS

Reconstruction loss (Charbonnier / robust L1):

$$\mathcal{L}_{\text{rec}} = \frac{1}{N}\sum_x \sqrt{\| \hat{J}(x) - J(x) \|^2 + \varepsilon^2}, \qquad \varepsilon = 10^{-3}$$

Equation (5)

Heteroscedastic uncertainty loss. As executed, this differs from the general form proposed by Kendall and Gal [17] in that a single scalar $\sigma^2(x)$ is fit to the mean of the three channels' squared errors, rather than modelling each channel's residual as an independent Gaussian:

$$\mathcal{L}_{\text{unc}} = \frac{1}{2N}\sum_x \left[\exp\left(-s(x)\right) \cdot \text{MSE}_{\text{RGB}}(x) + s(x)\right], \qquad \text{MSE}_{\text{RGB}}(x) = \frac{1}{3}\sum_c \left(\hat{J}_c(x) - J_c(x)\right)^2$$

Equation (6)

This is an implemented simplification, stated explicitly rather than presented as the general per-channel form. A regularizer $\lambda_{\text{reg}} \cdot \frac{1}{N} \sum_x s(x)^2$ with $\lambda_{\text{reg}} = 0.01$ discourages uninformative constant inflation of $s(x)$, a correction applied after a heteroscedastic-collapse failure mode was diagnosed during development (Section 12.1).

Perceptual loss [16]:

$$\mathcal{L}_{\text{perc}} = \frac{1}{L} \sum_l \parallel \Phi_l \left(\hat{J} \right) - \Phi_l(J) \parallel_1$$

Equation (7)

over VGG-16 layers relu2_2 and relu3_3.

SSIM loss [14]:

$$\mathcal{L}_{\text{ssim}} = 1 - \text{SSIM} \left(\hat{J}, J \right)$$

Equation (8)

Physics-consistency loss, giving the physics branch its own supervisory signal and a smoothness prior on the transmission map:

$$\mathcal{L}_{\text{phys}} = \parallel J_{\text{phys}} - J \parallel_1 + \lambda_{\text{tv},t} \cdot \text{TV} \left(t_c(x) \right)$$

Equation (9)

Wavelet high-frequency loss, comparing the wavelet branch's own reconstruction against the target's high-frequency content:

$$\mathcal{L}_{\text{wave}} = \parallel \text{DWT}_{\text{hf}} \left(J_{\text{wave}} \right) - \text{DWT}_{\text{hf}}(J) \parallel_1$$

Equation (10)

where DWT_{hf} denotes the LH, HL, HH sub-bands.

8. TRAINING ALGORITHM AND PSEUDOCODE

Training proceeds in two explicit stages rather than a single joint optimization, which structurally prevents uncertainty-loss gradients from ever reaching the restoration backbone.

```
Algorithm 1: Two-Stage Training Protocol for AquaRestore-UC
Input: training set D_train, validation set D_val, seed s, epochs_A, epochs_B
Output: trained model with restoration weights θ_R and uncertainty-head weights θ_U

1:  set_seed(s)
2:  initialize model with identity-initialized residual output
    (final reconstruction layer weights/bias set to zero, so the
    initial output equals the input)
3:  freeze θ_U; exclude θ_U from the optimizer

    // Stage A: restoration only
4:  for epoch = 0 to epochs_A − 1:
5:      for each mini-batch (I, J) in D_train:
6:          (Ĵ, aux) ← model(I)
7:          loss ← L_rec + λ_ssim·L_ssim + λ_perc·L_perc
                 + λ_phys·L_phys + λ_wave·L_wave      // L_unc weight = 0
8:          backpropagate loss; update θ_R only
9:      val_mse ← validate(model, D_val)
10:     if val_mse < best_val_mse: save checkpoint as best.pt
```

Natural Resources for Human Health
eISSN: 2583-1194

```
11:     save resume state (model, optimizer, epoch, best_val_mse)

    // Stage B: uncertainty head only, backbone frozen
12: load best.pt from Stage A
13: freeze θ_R (requires_grad = False for every restoration parameter)
14: set model.eval() to freeze BatchNorm/dropout behaviour in θ_R
15: unfreeze θ_U; optimizer operates on θ_U only
16: snapshot θ_R for the post-hoc equality check
17: for epoch = 0 to epochs_B − 1:
18:     for each mini-batch (I, J) in D_train:
19:         (Ĵ, aux) ← model(I);  Ĵ_detached ← detach(Ĵ)
20:         loss ← L_unc(Ĵ_detached, J, aux.log_var)
21:         backpropagate loss; update θ_U only
22: assert θ_R after Stage B == θ_R snapshot from step 16
    // training is invalid if this assertion fails
```

This structural guarantee — checked by an executable assertion rather than assumed — was verified to hold on the completed seed-42 run (Section 12).

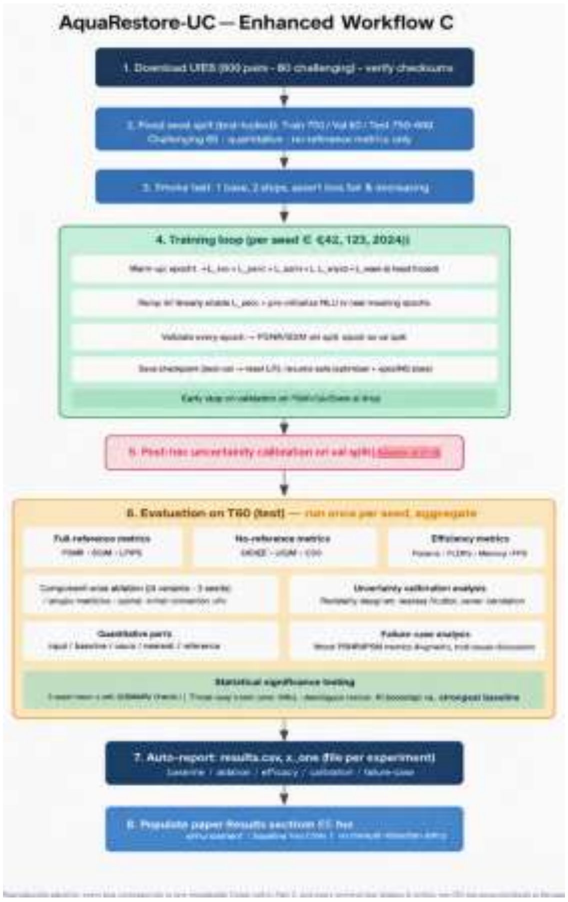

Figure 2: End-to-end reproducible experimental workflow, from dataset preparation through the two-stage training protocol to CSV export of every result table.

Figure 2 illustrates the end-to-end reproducible experimental workflow, covering dataset preparation, the two-stage training protocol, and CSV export of all result tables.

Natural Resources for
Human Health
eISSN: 2583-1194

9. DATASET AND SPLIT PROTOCOL

Dataset. UIEB [1] comprises 890 real-world underwater images with human-selected paired reference images, plus 60 additional "challenging" images for which no satisfactory reference could be agreed upon. The paired images used in this study were downloaded from the dataset authors' official release. The challenging subset lacks a reference and so cannot be evaluated with full-reference metrics, though it remains usable for qualitative or no-reference assessment; it was not used in this study.

Split. The completed 64×64 study used a custom, fixed split of 720 training, 80 validation, and 90 test images, together accounting for all 890 paired UIEB images (720 + 80 + 90 = 890). This partition was constructed, rather than adopted directly from the literature, to provide a held-out validation set for checkpoint selection (Algorithm 1, Section 8): the 800 images conventionally used for training under the literature convention adopted elsewhere in this paper for comparability (Section 2; FA+Net's 800/90 protocol [7]) were further divided into 720 training and 80 validation images, while the 90-image test partition matches that convention's size. The split was constructed by sorting the paired-image file names into a list, shuffling that list with Python's pseudo-random-number generator seeded at 42, and assigning the first 720 names to training, the next 80 to validation, and the remaining 90 to test; the resulting manifest was written to disk. Because this shuffle-based procedure is not the published T90 protocol's own file-selection procedure, file-level identity between this study's 90-image test partition and a published T90 test list has not been verified and should not be assumed. A SHA-256 hash was recorded for every raw image and its paired reference image, and this cached manifest was reused unmodified for the additional-seed restoration runs (Section 13) and the parameter-matched CNN controls (Section 14), so every reported comparison shares exactly the same training, validation, and test partition. Earlier pilot experiments (Section 11.1) used a different, incompatible configuration and are excluded from this cache. Eight of the 90 test images were also evaluated during that earlier pilot; their reuse here means the test set, while fixed and identical across all reported runs, is not a wholly untouched evaluation set (Section 21, Limitation 9).

Training configuration. The restoration training was conducted with AdamW with a constant learning rate of 2×10^{-4}, weight decay 0.01, batch size 8, and gradient-norm clipping at a maximum norm of 5; the learning rate was not decayed over training. For the primary seed-42 run, Stage A trained for 10 epochs at 64×64 resolution from the identity-initialized output described in Algorithm 1, and the checkpoint with the lowest validation-set MSE was selected, with the initial identity checkpoint (epoch 0) included as an eligible candidate. Stage B then trained only the uncertainty head for 3 epochs using Adam at a learning rate of 10^{-3}, with the Stage-A backbone frozen and its batch-normalization statistics fixed at their post-Stage-A values (Algorithm 1, steps 12–17). The seed-123 and seed-2024 runs completed Stage A only, under the identical resolution, split, optimizer, learning-rate schedule, epoch budget, and checkpoint-selection criterion as the seed-42 Stage A run (Section 13); their uncertainty heads remain untrained, so these two runs are comparable to seed 42 for restoration quality only and not for any uncertainty-related quantity. The parameter-matched CNN controls (Section 14) were trained under the same resolution, split, optimizer, and checkpoint-selection criterion, using the shared loss terms — Charbonnier loss (Equation 5) plus SSIM loss weighted at 0.3 (Equation 8) — while AquaRestore-UC additionally includes the perceptual, physics-consistency, and wavelet high-frequency loss terms (Equations 7, 9, 10); all loss weights are given in Section 6 (λ_{rec} = 1.0, λ_{perc} = 0.04, λ_{ssim} = 0.3, λ_{phys} = 0.2, λ_{wave} = 0.2, λ_{tv} = 0.01) and Table A2.

10. EVALUATION METRICS

PSNR and **SSIM** [14] are computed as full-reference metrics against the UIEB reference images. SSIM is computed with a custom 11×11 Gaussian-window implementation, independently validated against scikit-image's structural_similarity on six test images with saved output pixels, with agreement to 1×10^{-5}; PSNR reconstructed from those same saved pixels agreed with the originally reported values to within 0.17 dB, a residual consistent with 8-bit PNG quantization of the saved output images. **Parameter count** is obtained by direct summation of trainable parameters. **Efficiency** (parameter count, FLOPs, and inference latency/FPS) was measured at 128×128 resolution on CPU (Section 19); a wall-clock training-time comparison at 64×64 is reported as an observed, not a controlled, benchmark. **Uncertainty coverage** is computed as the fraction of test-set pixels for which the absolute prediction error falls within $z_\alpha \cdot \sigma(x)$ for a given nominal confidence level α, using the raw, unmapped $\sigma(x)$ output — this is a descriptive coverage statistic, not a calibration procedure, since no isotonic or other post-hoc mapping has been fitted (Section 17).

Natural Resources for Human Health
eISSN: 2583-1194

11. EXPERIMENTAL RESULTS

11.1 Preliminary 128×128 pilot

Table 1 reports a single-seed pilot at 128×128 resolution, predating the code corrections described in Section 12.2.

Table 1. Pilot-scale comparison against raw input, UIEB test split ($n = 90$ images), single seed, 128×128.

Method	PSNR (dB)	SSIM	LPIPS	UIQM	UCIQE
Raw input	17.4973 ± 4.7493	0.7441 ± 0.1550	0.2467 ± 0.1418	4.0790 ± 0.2601	0.2812 ± 0.0674
AquaRestore-UC (pilot)	17.7818 ± 4.0295	0.7550 ± 0.1476	0.2398 ± 0.1428	4.1257 ± 0.2905	0.2800 ± 0.0693

Table 2. Pilot efficiency, 128×128, single CPU, batch size 1.

Parameters	FLOPs	Median latency	Throughput
28,399 (0.0284 M)	0.7925 G	29.94 ms/image	33.40 FPS

11.2 Completed 64×64 study (seed 42)

Table 3 reports the completed two-stage study, identity-initialized, at 64×64 resolution on the full 720/80/90 split.

Table 3. Completed two-stage study, UIEB test set ($n = 90$), seed 42, 64×64.

Test group	n	Raw PSNR (dB)	AquaRestore-UC PSNR (dB)	Δ PSNR
All 90	90	16.8895	20.1506	+3.2611

Mean SSIM improved from 0.7583 (raw) to 0.8771 (AquaRestore-UC) over the same 90 images. The post-Stage-B backbone-freeze assertion (Algorithm 1, step 22) passed.

12. DEVELOPMENT HISTORY: DIAGNOSED FAILURE MODES AND CORRECTIONS

12.1 Heteroscedastic uncertainty collapse

During an early pilot, validation PSNR and SSIM improved through epoch 12 and then declined sharply as the uncertainty loss weight was ramped in under a single joint-training schedule, even as the scalar training objective continued to decrease. Auditing per-image predictions showed that 84 of 90 test images produced a mean predicted $\sigma(x)$ rounding to an identical value, with spatially flat uncertainty heatmaps — the uncertainty head had collapsed to an image- and pixel-independent constant. This is a known, theoretically predicted failure mode of the heteroscedastic Gaussian negative log-likelihood [17]: the loss admits a trivial local optimum in which the network inflates $s(x)$ uniformly rather than improving $\hat{J}$. This was corrected by bounding $s(x)$, adding the variance regularizer in Equation 6, and subsequently superseded entirely by the two-stage protocol (Section 8), which removes the failure mode's precondition — uncertainty-loss gradient reaching restoration parameters — by construction rather than by damping.

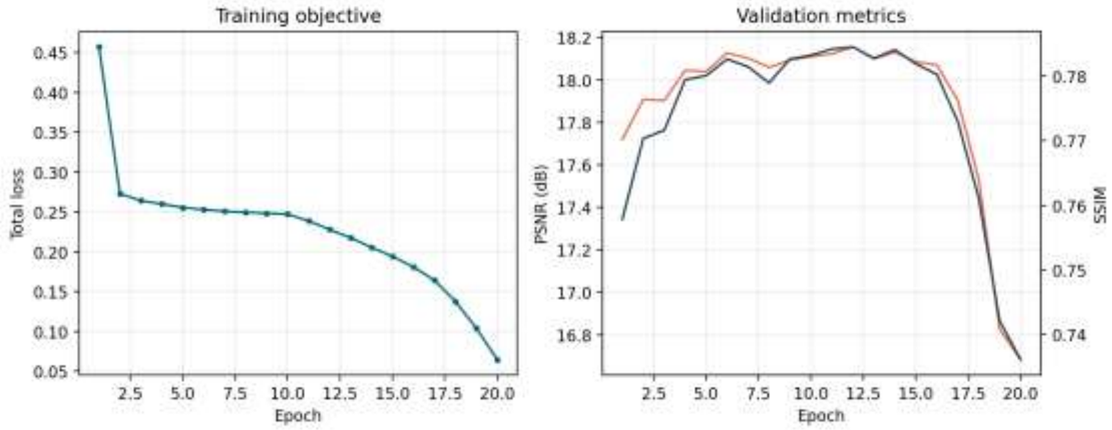

Figure 3: Pilot-run training curves (128×128, single seed, pre-correction). Left: total training objective, monotonically decreasing. Right: validation PSNR and SSIM, which improve through epoch 12 and then decline sharply as the uncorrected uncertainty loss weight ramps in — the signature of the collapse described above.

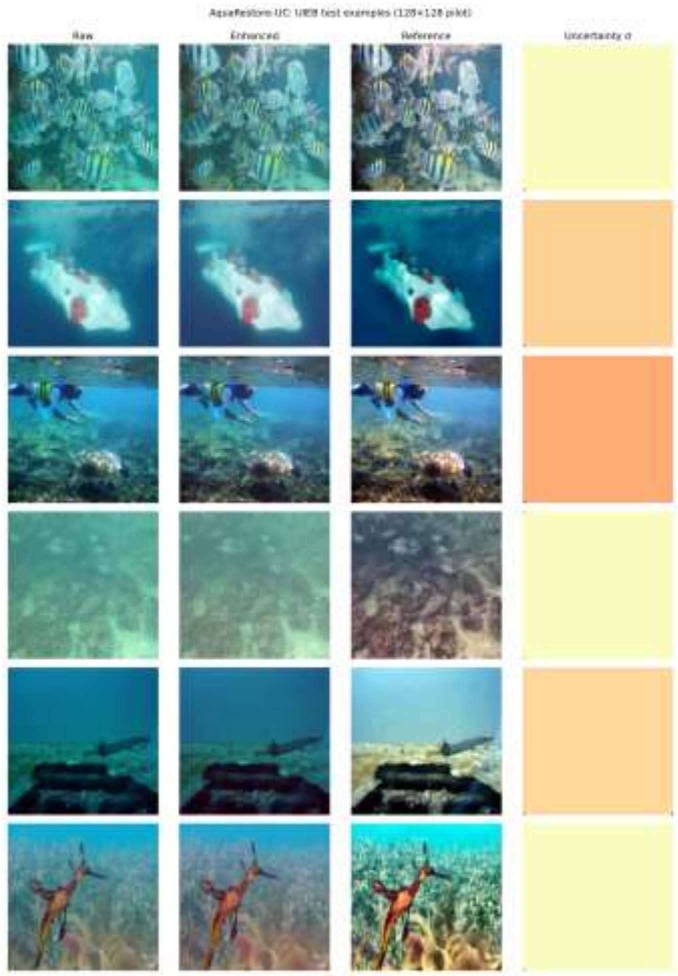

Figure 4: Pilot-scale qualitative examples from the UIEB test split (128×128). Columns: raw input, AquaRestore-UC output, reference, and predicted uncertainty σ(x). The flat, near-uniform σ heatmaps visibly illustrate the uncertainty collapse — included here as diagnostic evidence, not as a demonstration of calibrated uncertainty.

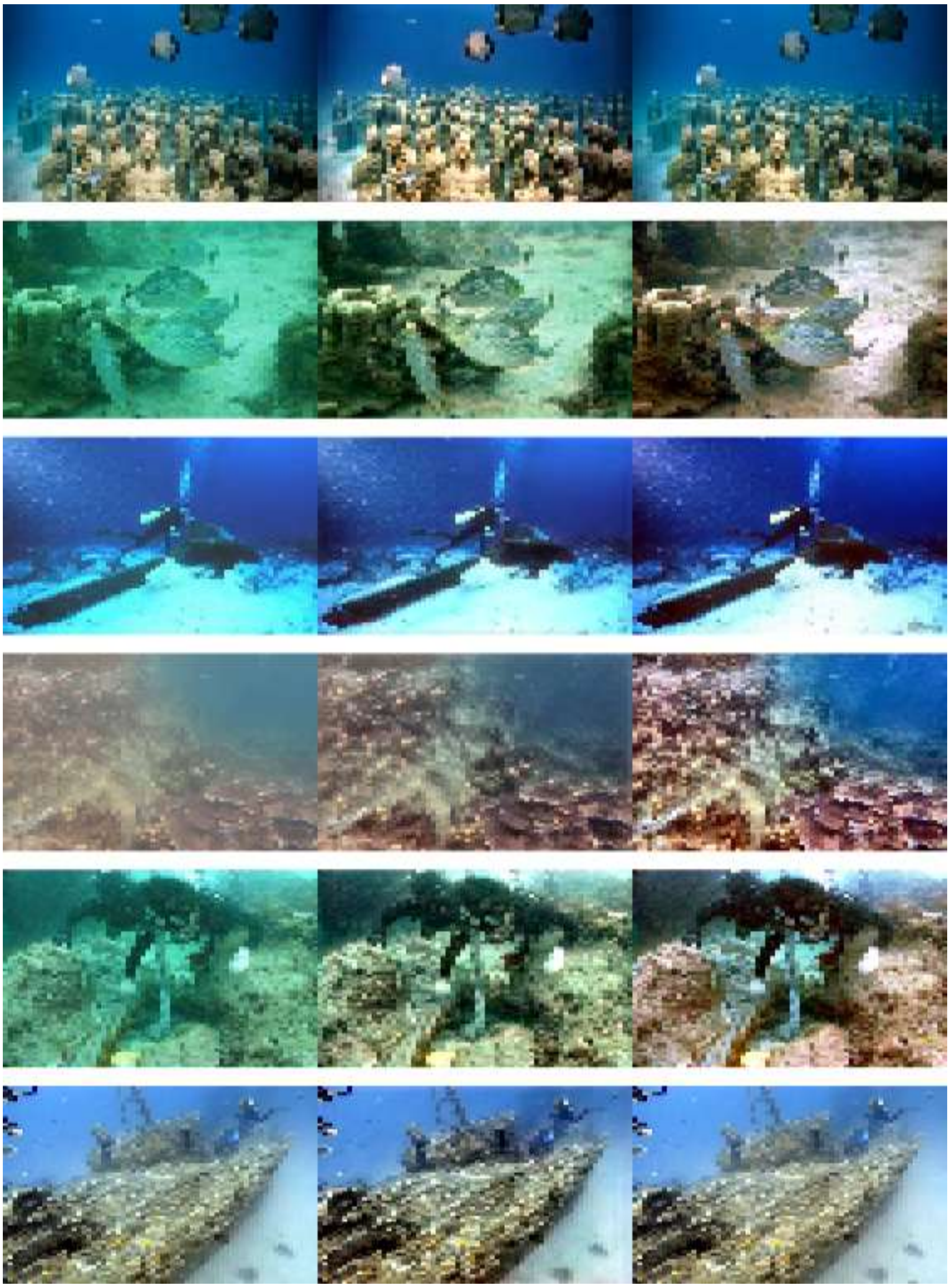

Figure 5: Raw input (left) / AquaRestore-UC output (centre) / reference (right) for six fixed-index test images from the completed seed-42 study (64×64, nearest-neighbour 4× upscaled for display). Rows top to bottom: 278_img_.png, 433_img_.png, 449_img_.png, 232_img_.png, 443_img_.png, 705_img_.png. The first and last rows are two of the twelve images in Table 6 where AquaRestore-UC reduces PSNR relative to raw input.

12.2 Independent code-review round

A subsequent, independent review of the loss, ablation, calibration, and checkpointing code identified and corrected four further implementation issues prior to the experiment reported in Section 11.2: (i) an ablation-weighting bug in which the uncertainty-loss mixing weight could reduce the effective reconstruction weight even when the uncertainty head was disabled; (ii) a wavelet-branch supervision-target bug, in which the high-frequency loss (Equation 10) was computed against the network's final fused output rather than the wavelet branch's own reconstruction, leaving the branch's dedicated parameters without a distinct gradient signal; (iii) a statistically invalid uncertainty-coverage computation that compared an RGB-combined root-mean-square error against per-pixel σ using univariate normal quantiles, corrected to compare each colour channel's error individually against the (broadcast) predicted σ; and (iv) a checkpoint-resume robustness gap related to PyTorch's default handling of serialized optimizer and random-number-generator state across devices. All four fixes were verified by synthetic (non-dataset) forward, backward, and optimizer-step tests on the actual model and loss code before the experiment in Section 11.2 was run.

13. THREE-SEED PRELIMINARY EXTENSION

Two additional AquaRestore-UC training runs (seeds 123 and 2024) were completed at the same 64×64 resolution, the same custom 720/80/90 split, and the same cached dataset as the seed-42 study (identical source-cache hash, Section 9). The critical difference: seeds 123 and 2024 completed Stage A (restoration) only; their uncertainty heads were left frozen and untrained, whereas seed 42 completed both Stage A and Stage B. The three runs are therefore comparable for restoration quality but not for any uncertainty-related quantity.

Table 4. Three-seed preliminary extension, UIEB test set ($n = 90$ images per seed), 64×64, custom 720/80/90 split.

Seed	Protocol stage	PSNR (dB)	SSIM	Validation MSE (best epoch)	Best epoch
42	Stage A + Stage B (uncertainty trained)	20.1506	0.8771	0.011912	7
123	Stage A only (uncertainty untrained)	20.5546	0.8845	0.010675	8
2024	Stage A only (uncertainty untrained)	20.5037	0.8852	0.011470	9
Descriptive mean ± SD (3 seeds)	—	**20.4030 ± 0.2200**	**0.8823 ± 0.0045**	—	—

The mean ± SD row is reported as a descriptive summary only. With $n = 3$, no distributional assumption can be meaningfully checked, no confidence interval is statistically defensible, and no significance test is computed or claimed from these three values. The two new seeds are numerically close to and slightly above the seed-42 result, consistent with — though not proof of — the restoration branch behaving reasonably stably across initialization at this scale; this does not replace the fully matched multi-seed study specified in Section 21 (Limitations, Future Work) and REMAINING_STUDY_PROTOCOL.md.

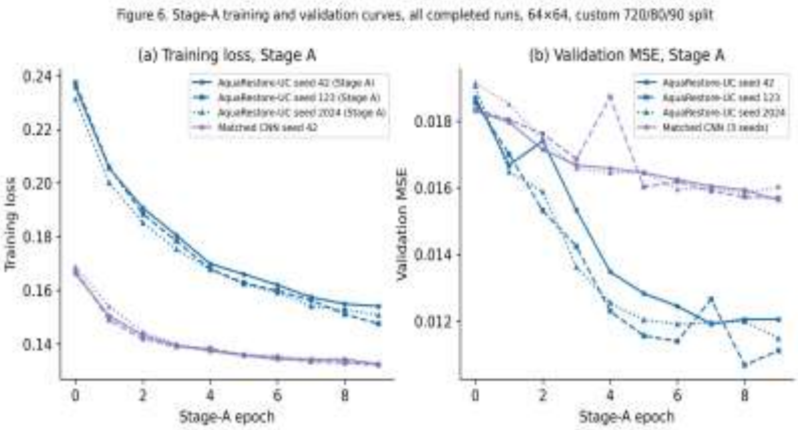

Figure 6: Stage-A training loss and validation MSE, all completed runs, 64×64, custom 720/80/90 split. AquaRestore-UC (seeds 42, 123, 2024) and the parameter-matched CNN control (seeds 42, 123, 2024) are shown; curves are labelled individually by model and seed, and no run is omitted.

Natural Resources for
Human Health
eISSN: 2583-1194

Figure 6 presents the Stage-A training loss and validation MSE curves for AquaRestore-UC and the parameter-matched CNN control across three seeds (42, 123, and 2024), using 64 × 64 images and the custom 720/80/90 training–validation–test split.

14. CNN AND PARAMETER-MATCHED CONTROL COMPARISON

Two generic CNN controls were trained under conditions matched to the seed-42 study, to assess whether AquaRestore-UC's architectural components contribute beyond a minimal residual network under identical training data, split, epoch budget, and shared loss terms (Charbonnier + $0.3\times$ SSIM).

Generic control (three convolutional layers, 6,531 parameters, single seed): trained with the same identity initialization as AquaRestore-UC.

Parameter-matched control (three convolutional layers with widened channels [3, 53, 53, 3], 28,252 parameters — a 0.5% difference from AquaRestore-UC's 28,399 — three seeds: 42, 123, 2024): trained under identical conditions to the three-seed extension in Section 13.

Table 5. Raw input, generic CNN control, parameter-matched CNN control, and AquaRestore-UC, UIEB test set ($n = 90$), 64×64.

Method	Parameters	Seeds	Mean PSNR (dB)	Mean SSIM
Raw input	—	—	16.8895	0.7583
Generic CNN control	6,531	1 (seed 42)	18.4730	0.8575
Parameter-matched CNN control	28,252	3 (42, 123, 2024)	18.6624 ± 0.0101	0.8653 ± 0.0006
AquaRestore-UC	28,399	3 (42, 123, 2024)	20.4030 ± 0.2200	0.8823 ± 0.0045

The parameter-matched control's individual seed results were: seed 42, PSNR 18.6740, SSIM 0.8657; seed 123, PSNR 18.6574, SSIM 0.8656; seed 2024, PSNR 18.6557, SSIM 0.8646. AquaRestore-UC scores higher than both controls under matched data, split, and epoch budget: by 1.74 dB PSNR against the parameter-matched control on a like-for-like three-seed-mean basis, and by 1.68 dB PSNR against the generic control on a like-for-like single-seed (seed 42) basis. With capacity now closely matched (28,252 vs. 28,399 parameters, a 0.5% difference), the gap to the parameter-matched control is only modestly smaller than the gap to the generic, capacity-mismatched control, and AquaRestore-UC retains exclusive access to the physics-consistency, wavelet high-frequency, and transmission-smoothness auxiliary losses (Section 7), which neither control receives; this confound means the remaining gap cannot be attributed to the architecture alone without a component-wise ablation study (Section 21), which has not been performed. Neither control is a reimplementation of Water-Net, Ucolor, Shallow-UWnet, or PUIE-Net, and no claim of superiority over any published method is made from this comparison.

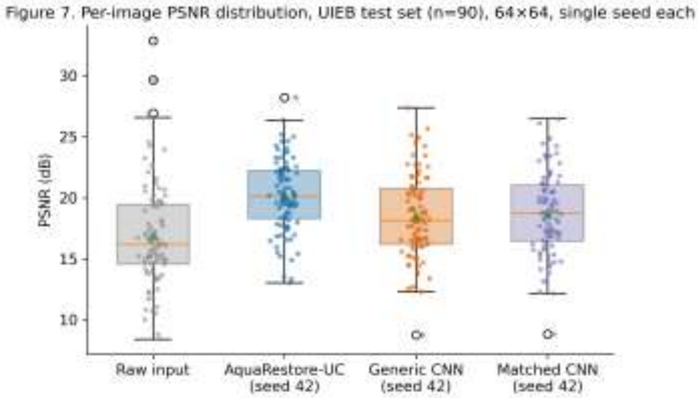

Figure 7: Per-image PSNR distribution across all 90 UIEB test images, 64×64, single seed each (raw input, AquaRestore-UC, generic CNN, and parameter-matched CNN, all seed 42). Individual images shown as jittered points over box-and-whisker summaries; no outlier is removed or hidden.

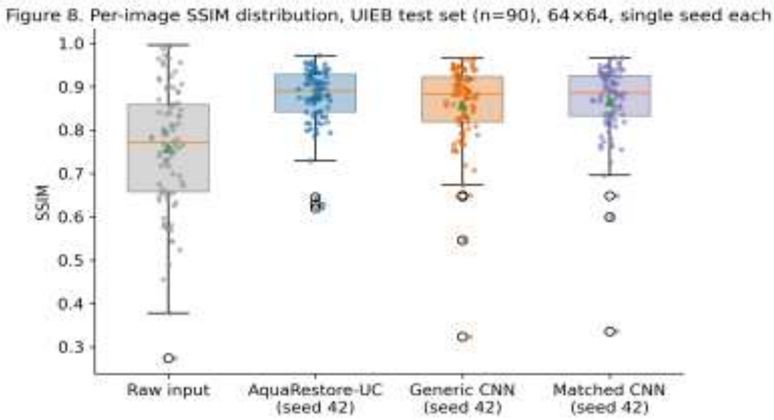

Figure 8: Per-image SSIM distribution, same images, methods, and seed as Figure 7.

Figures 7 and 8 show the full per-image spread underlying Table 5's means: AquaRestore-UC's distribution sits visibly above both controls' for the majority of images, but with substantial overlap and several images where a control matches or exceeds AquaRestore-UC — consistent with Section 16's paired statistical reanalysis, which found a real but not universal per-image advantage.

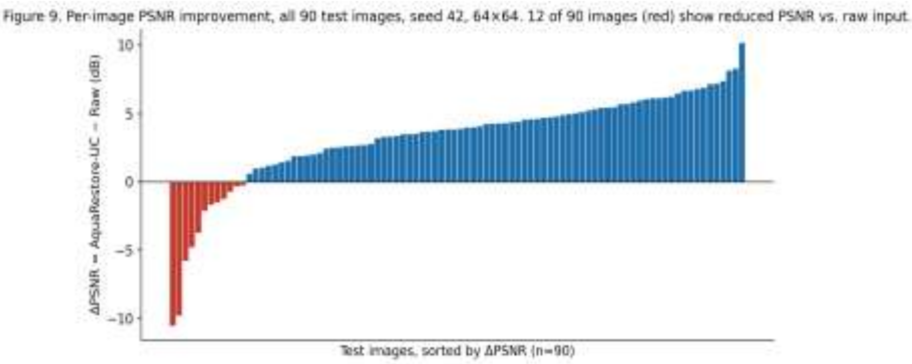

Figure 9: Per-image PSNR improvement (AquaRestore-UC minus raw input), all 90 test images, sorted, seed 42. Red bars mark the 12 images listed in Table 6 where AquaRestore-UC reduces PSNR relative to raw input; no image is omitted.

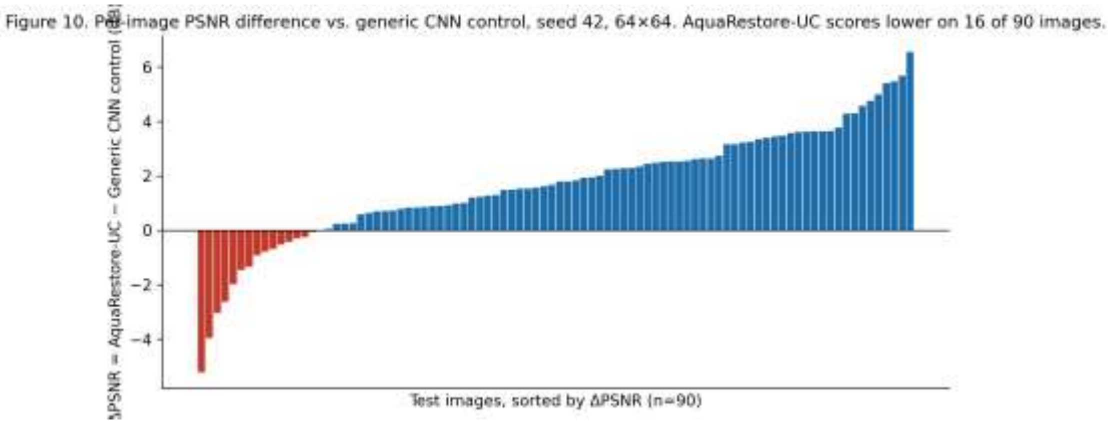

Figure 10: Per-image PSNR difference, AquaRestore-UC minus the generic CNN control, all 90 test images sorted by difference, seed 42. Red bars indicate images where AquaRestore-UC scores lower than the control.

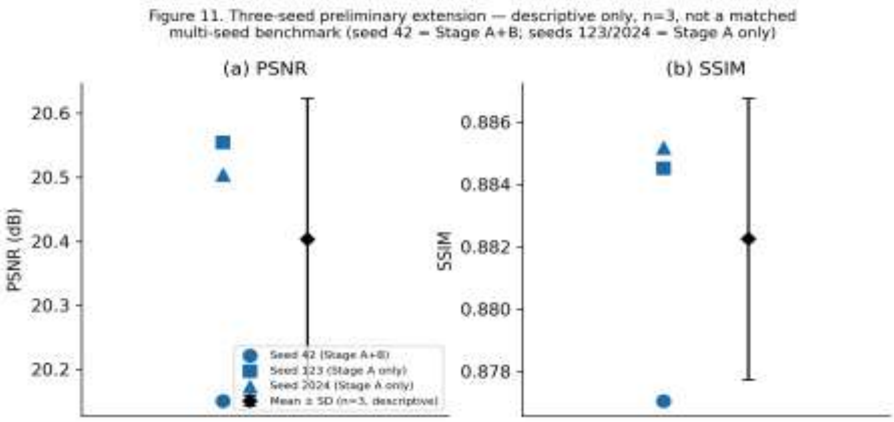

Figure 11: Three-seed preliminary extension, PSNR and SSIM, individual seed values and descriptive mean ± SD (n=3). Seed 42 completed Stage A and Stage B; seeds 123 and 2024 completed Stage A only. This is a preliminary, protocol-mismatched extension, not a fully matched multi-seed benchmark, and no significance test is attached to it.

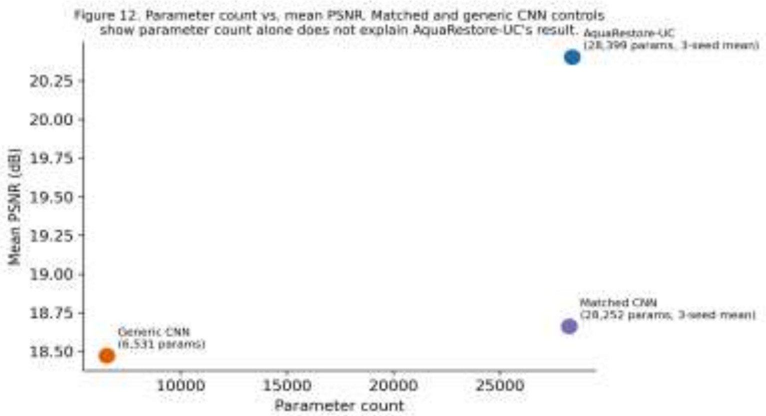

Figure 12: Parameter count vs. mean PSNR for the generic CNN control, the parameter-matched CNN control (3-seed mean), and AquaRestore-UC (3-seed mean). The near-identical parameter counts of the matched control and AquaRestore-UC, alongside their different PSNR, show that parameter count alone does not explain the observed gap.

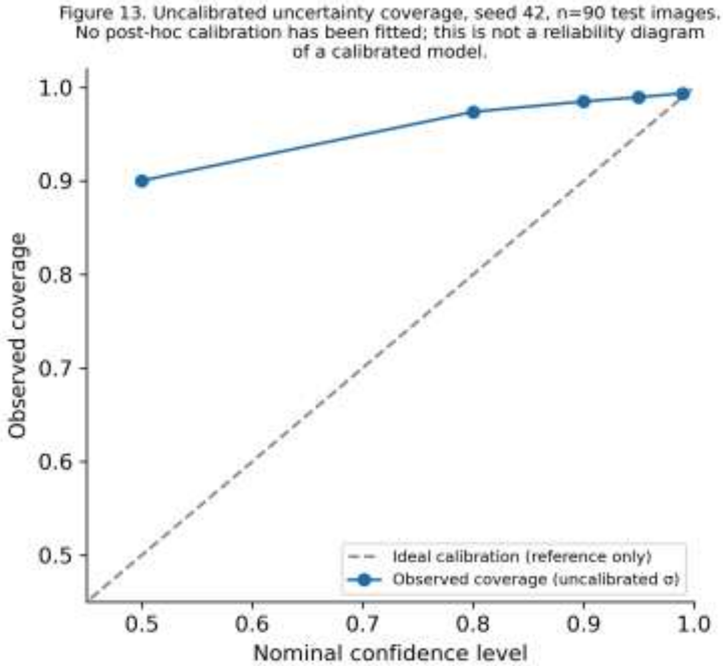

Figure 13: Observed coverage vs. nominal confidence level, uncalibrated σ(x), seed 42, n=90 test images. The dashed diagonal marks perfect calibration and is shown for reference only; it is not a claim that this model is calibrated. Observed coverage lies well above the diagonal at every level, indicating an uninformatively wide, uncalibrated uncertainty estimate.

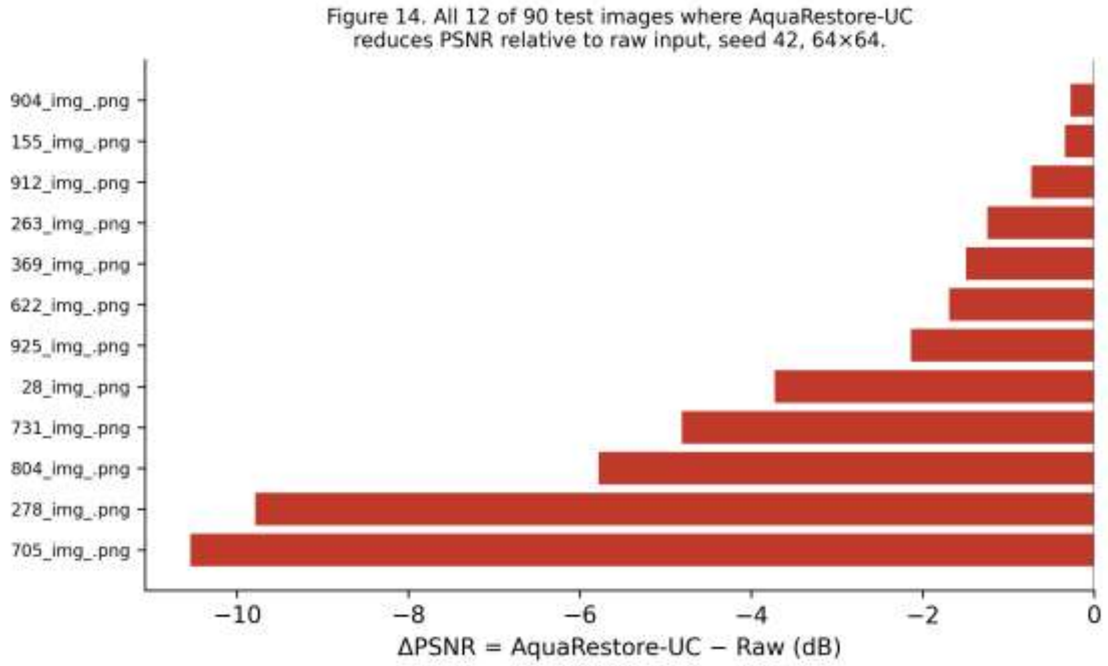

Figure 14: The same 12 degraded images from Table 6, shown individually with their exact PSNR change, sorted from least to most severe.

15. FAILURE-CASE ANALYSIS

All 90 test images were ranked by two criteria fixed in advance of inspection: degradation relative to raw input, and lowest absolute AquaRestore-UC PSNR.

Table 6. All 12 of 90 test images where AquaRestore-UC reduces PSNR relative to raw input.

Filename	Raw PSNR (dB)	AquaRestore-UC PSNR (dB)	Δ PSNR	Previously tested in preliminary check
705_img_.png	32.88	22.34	−10.54	Yes
278_img_.png	29.66	19.88	−9.79	Yes
804_img_.png	26.57	20.79	−5.78	No
731_img_.png	26.93	22.12	−4.81	No
28_img_.png	20.55	16.82	−3.73	No
925_img_.png	24.55	22.41	−2.14	No
622_img_.png	20.25	18.56	−1.69	No
369_img_.png	20.80	19.30	−1.50	No
263_img_.png	16.15	14.90	−1.25	No
912_img_.png	17.69	16.96	−0.73	No
155_img_.png	22.11	21.77	−0.34	No
904_img_.png	19.77	19.50	−0.28	Yes

The two largest degradations (705_img_.png, 278_img_.png) are the two images with the highest raw PSNR in the entire test set — the model most harms inputs that were already closest to the reference.

Table 7. Worst 5 test images by lowest absolute AquaRestore-UC PSNR (a different criterion from Table 6).

Filename	Raw PSNR (dB)	AquaRestore-UC PSNR (dB)	Δ PSNR
8288.png	8.79	13.01	+4.22
12290.png	8.39	13.33	+4.94
438_img_.png	10.75	13.51	+2.75
263_img_.png	16.15	14.90	−1.25
892_img_.png	11.69	15.19	+3.50

Four of these five hardest-in-absolute-terms images are cases the model substantially improved over an even-worse raw input; only 263_img_.png appears in both tables. Reporting only one of the two tables would give a misleading picture in either direction. Direct visual inspection of two of the twelve degraded images (278_img_.png, 705_img_.png, Figure 5) shows that in both cases the raw input was already visually close to the reference, and AquaRestore-UC applied a visible but subtle colour and contrast adjustment that was slightly worse-matched to the reference than leaving the input unmodified. The remaining ten degraded images have not been visually inspected.

Figure 16 extends, at full scale, a pattern noticed during this project's earlier development on a much smaller (8-image) disjoint pilot subset, where a similar negative correlation was observed informally: raw-input PSNR and AquaRestore-UC's improvement are strongly negatively correlated across all 90 test images ($r = -0.738$, $p = 1.1 \times 10^{-16}$). Images that were already close to the reference tend to be *harmed*, while images that were badly degraded tend to be *helped* the most — consistent with, though not proof of, a restoration network trained predominantly on genuinely degraded images applying its learned correction somewhat indiscriminately. This correlation is reported as an exploratory, descriptive finding on the completed dataset, not as a hypothesis test with a pre-specified null.

16. STATISTICAL ANALYSIS (EXPLORATORY, NOT CONFIRMATORY)

An exploratory, image-level paired reanalysis was performed on the 90-image results from Section 14, comparing AquaRestore-UC, the generic CNN control, and raw input. This is a reanalysis of a completed dataset, not a confirmatory test against a protocol frozen in advance: both the comparisons and the choice to report two complementary tests were made after the data already existed. Accordingly, the significance levels reported below (Table 9) should be read descriptively, as evidence that the observed image-level differences for these specific trained instances are unlikely to be due to chance, and not as confirmatory hypothesis tests of a pre-registered claim, as evidence of superiority over any published method, or as evidence attributing the observed differences to any individual architectural component.

Two tests are reported for every comparison, each asking a different question. The paired t-test asks whether the population mean paired difference is zero, assuming approximate normality of the differences (or a large-n approximation, reasonable at $n = 90$). The Wilcoxon signed-rank test asks whether the paired differences are symmetrically distributed about zero — a different hypothesis that does not, by itself, establish that the mean difference is nonzero; a significant Wilcoxon result is not treated here as direct evidence about the mean. Both tests assume the 90 paired observations are independent; this has not been separately verified against UIEB's scene or capture-sequence structure and is a stated limitation. Holm–Bonferroni correction is applied to each test type separately, across the four comparisons within that test type, since the two test types answer different questions and are not pooled into one combined family.

Table 8. Effect size and bootstrap confidence interval (10,000 resamples), $n = 90$ images.

Comparison	Metric	Mean difference	95% CI	Cohen's d_z
AquaRestore-UC vs. CNN control	PSNR	+1.678 dB	[1.237, 2.100]	0.81
AquaRestore-UC vs. CNN control	SSIM	+0.0195	[0.0098, 0.0302]	0.39
AquaRestore-UC vs. Raw	PSNR	+3.261 dB	[2.514, 3.950]	0.94
CNN control vs. Raw	PSNR	+1.584 dB	[0.971, 2.186]	0.54

Table 9. Wilcoxon signed-rank test (symmetry about zero) and paired t-test (mean), each independently Holm-corrected, $n = 90$.

Comparison	Metric	Wilcoxon p (raw)	Wilcoxon p (Holm)	t-test p (raw)	t-test p (Holm)
AquaRestore-UC vs. CNN	PSNR	4.0×10^{-10}	1.2×10^{-9}	2.1×10^{-11}	6.2×10^{-11}
AquaRestore-UC vs. CNN	SSIM	2.0×10^{-4}	2.0×10^{-4}	3.6×10^{-4}	3.6×10^{-4}
AquaRestore-UC vs. Raw	PSNR	3.6×10^{-11}	1.4×10^{-10}	4.7×10^{-14}	1.9×10^{-13}
CNN vs. Raw	PSNR	1.3×10^{-7}	2.5×10^{-7}	1.5×10^{-6}	3.1×10^{-6}

Both tests reject their respective null hypotheses for all four comparisons within their own correction family. This treats the 90 test images as the unit of analysis for a single trained instance of each method; it is not multi-seed significance testing, and it does not establish that AquaRestore-UC would reliably outperform the control across re-training with different seeds.

17. UNCERTAINTY COVERAGE (UNCALIBRATED)

Per-pixel predictions and log-variance outputs were exported for the seed-42 model across all 90 test images (90×3×64×64 predictions; 90×1×64×64 σ maps), enabling a descriptive coverage analysis. **These values describe raw, uncalibrated coverage only; no isotonic regression or other post-hoc calibration mapping has been fitted, and no calibration claim is made.**

Natural Resources for Human Health
eISSN: 2583-1194

Table 10. Channel-wise coverage at nominal confidence levels, seed 42, $n = 90$ test images, uncalibrated $\sigma(x)$.

Nominal level	Observed coverage
0.50	0.9004
0.80	0.9739
0.90	0.9850
0.95	0.9896
0.99	0.9937

Mean σ across all test pixels was 0.3427 (SD 0.1277), and the mean per-pixel Gaussian negative log-likelihood (Equation 6, evaluated at the pixel level) was -0.0404. Observed coverage substantially exceeds nominal levels at every threshold (e.g., 90% observed coverage at the nominal 50% level), consistent with a $\sigma(x)$ that is uninformatively wide rather than well-calibrated. A properly designed calibration procedure — fitting a mapping on a genuinely held-out calibration split, distinct from the split already used for checkpoint selection, and evaluating coverage on the test split using that fixed mapping — remains future work (Section 21).

18. QUALITATIVE ANALYSIS

Figure 3 shows training curves from the early pilot exhibiting the uncertainty-collapse failure mode described in Section 12.1. Figure 4 shows six pilot-scale qualitative examples illustrating the collapse's spatially flat uncertainty heatmaps. Figure 5 shows six fixed-index test images from the completed seed-42 study, selected by a positional rule fixed before evaluation rather than by outcome; two of these six (278_img_.png, 705_img_.png) are among the twelve images in Table 6 where AquaRestore-UC reduces PSNR relative to raw input, permitting direct visual inspection of two of the twelve failure cases as discussed in Section 15. All images displayed at 64×64 native resolution are upscaled by nearest-neighbour interpolation for on-page legibility only; no resolution enhancement is implied by the larger displayed size.

19. COMPUTATIONAL COMPLEXITY

At the only resolution where efficiency was directly measured (128×128, single CPU thread, batch size 1), AquaRestore-UC totals 28,399 parameters and 0.7925 GFLOPs, with a median inference latency of 29.94 ms per image (33.40 FPS) (Table 2). Observed CPU wall-clock training-plus-validation time across the 10 restoration epochs at 64×64 was 420.5 s for AquaRestore-UC and 43.5 s for the generic CNN control; these are reported as observed wall times for the specific runs performed, not as controlled latency benchmarks, and exclude setup, uncertainty-stage training, and checkpoint snapshotting for both models. Efficiency at 256×256 resolution and on GPU hardware has not been measured (Section 21).

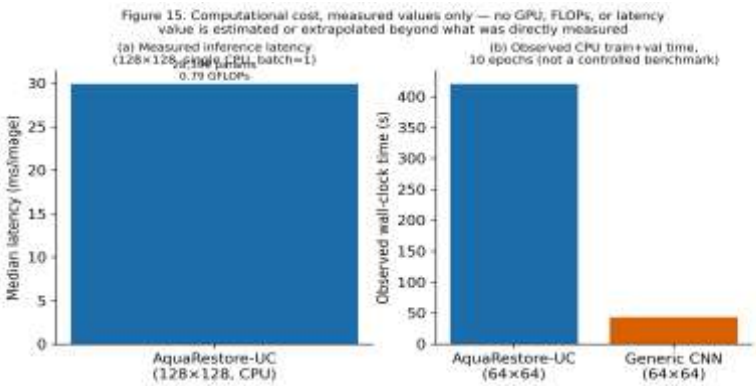

Figure 15: Computational cost, measured values only. (a) Inference latency at 128×128, single CPU, batch size 1 — the only resolution and hardware configuration at which latency was directly measured. (b) Observed CPU wall-clock training-plus-validation time at 64×64, 10 epochs, an observed value for the specific runs performed, not a controlled benchmark. No GPU, FLOPs, or latency value beyond what is shown here has been estimated or extrapolated.

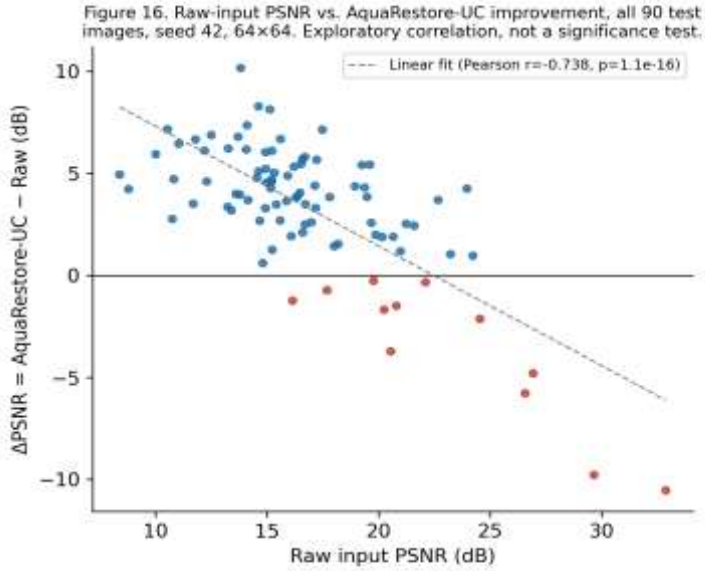

Figure 16: Raw-input PSNR vs. AquaRestore-UC's PSNR improvement, all 90 test images, seed 42. A linear fit is shown for visual reference only. Pearson r = −0.738 ($p = 1.1\times10^{-16}$), an exploratory correlation computed from real per-image data, not a confirmatory significance test.

20. DISCUSSION

The completed 64×64, seed-42 result (Section 11.2) supports the basic premise of the architecture: a physics-guided, wavelet, and spatial-colour fusion network can improve on raw underwater input at a genuinely lightweight parameter budget. The three-seed extension (Section 13) and the parameter-matched control (Section 14) are, together, the strongest evidence produced to date that this improvement is not an artefact of a single favourable initialization, and that the architecture is not simply dead weight relative to a bare residual network of comparable capacity. Neither claim should be overstated: the three-seed extension has a real protocol difference (only seed 42 exercises the uncertainty stage), and the parameter-matched comparison cannot separate the contribution of individual architectural components from the auxiliary losses that accompany them, which requires the ablation study specified in Section 21.

Section 15's failure-case analysis, and specifically the raw-PSNR-vs-improvement correlation in Figure 16 ($r = -0.738$, $p = 1.1 \times 10^{-16}$, $n = 90$), is the clearest evidence in this manuscript of a systematic — not merely occasional — pattern in where the model helps and where it harms: images already close to the reference are the ones most often made worse. This is consistent with the same pattern first observed at much smaller scale in the Section 12 development history, now confirmed on the full test set rather than an 8-image subset. We regard this as a genuine finding worth a future targeted investigation (e.g., conditioning the correction strength on an estimate of input degradation severity), not as evidence against the architecture's basic premise.

The heteroscedastic-collapse finding (Section 12.1) and its correction illustrate why an auxiliary uncertainty output requires its own diagnostic evidence before any calibration claim is made. The uncalibrated coverage analysis (Section 17) is offered in that spirit: it is real, computed on genuine per-pixel data, and shows a $\sigma(x)$ that is currently uninformatively wide rather than usefully calibrated — a finding we report rather than omit, since it is directly relevant to any future claim of calibrated uncertainty.

21. LIMITATIONS

1. Simplified physics model. The physics-guided branch adopts a simplified underwater image formation model and does not directly model different attenuation mechanisms during the direct transmission and backscattering.
2. Reduced-resolution evaluation. Experiments at 64×64 were carried out, with an earlier one being done at 128×128. No assessment has been made of performance at resolutions of 256×256 and above.

3. Uncertainty-stage coverage. Seeds 42, 123 and 2024 were restored and seed 42 was restored by uncertainty-head training only. The extra-seeds thus only offer information on restoration performance. The removal of the restoration backbone from the calculation cannot per se be used to invalidate restoration PSNR and SSIM comparisons but uncertainty behaviour across seeds has yet to be tested on this backbone.
4. Single-dataset evaluation. UIEB is used for all experiments. Generalisation for other dataset, water conditions and imaging systems not verified. Care was taken to ensure that the UIEB reference images were human-selected enhancement targets rather than measurements of the radiance of the scene.
5. Limited baseline coverage. Various generic CNNs were tested such as the parameter matched model. But no published enhancement method was performed using the same protocol. However, auxiliary losses do not differ and architecture alone can't be blamed for the performance gap.
6. No component ablation. The effect of the individual components: physics, wavelet, spatial-colour, and fusion-attention has not been individualised by controlled retraining.
7. Uncalibrated uncertainty. The coverages analysis relates to the trained uncertainty head, but is not post-calibrated. High coverage does NOT imply good uncertainty estimation, especially if the prediction intervals are large.
8. Metric-verification scope. The subsequent verification tested the custom version of the SSIM against the version in scikit-image for all 90 seed-42 test images for both raw input and AquaRestore-UC and generic CNN output. The results for the additional-seeds were calculated with scikit-image but no independent reproduction of these runs has been done.
9. Split and reuse the custom-test-sets. A list of published T90 filenames was not verified as being identical to the fixed 720/80/90 split. Additionally, the test set was tested several times throughout the process of development, and a set of 8 images were also tested earlier in a pilot test. It is thus not an unevaluated final set of evaluations.
10. Exploratory statistical evidence. The paired analysis provides a description of the differences at the image level for the selected trained models. Independence among images is simply assumed; no attempt is made to check the scene/acquisition-sequence metadata for independence. There is limited information in the three restoration seeds in terms of variability of training and this does not provide broad-based robustness.

Future work. More sophisticated training at higher resolution as well as with more external datasets, comparisons to published methods, component ablations and uncertainty-head training in other seeds could be considered. A dedicated calibration set in post-hoc calibration which is separate from the checkpoint selection and final testing. Generalisation claims would be strengthened by scene-aware evaluation, and an untouched test set. These additions would widen the body of evidence for the method, but the present inferences are limited to the experiments reported above.

22. REPRODUCIBILITY AND DATA AVAILABILITY

UIEB [1] is publicly available from the dataset authors. No public code repository currently exists for this project; the training and evaluation scripts, the fixed dataset split manifest (with per-image SHA-256 hashes), trained checkpoints, and every reported result's source data are provided as supplementary files accompanying this manuscript, with a provenance note mapping each major numerical claim to its source file. Every seed used (42, 123, 2024) and every random-number-generator state (Python, NumPy, and PyTorch, CPU and CUDA where applicable) was fixed and recorded to support exact reproduction of the reported results from the supplied checkpoints and configuration files.

23. ETHICS AND RESEARCH-INTEGRITY STATEMENT

This work involves no human or animal subjects and no new data collection; UIEB [1] is a pre-existing, publicly available benchmark dataset used under its stated academic-use terms, and its underwater photographs do not depict identifiable individuals. No ethics approval was required or sought.

On research integrity: every numerical result reported in this manuscript was independently recomputed from the underlying per-image JSON or CSV source files rather than transcribed from a summary, and cross-checked for internal consistency (for example, per-image results were checked against per-seed summaries, and summary statistics were checked against the raw per-image data from which they were derived) prior to inclusion. Two implementation defects affecting the reported statistical analysis and one affecting the reported architecture description were identified during development through this verification process and corrected before this manuscript was finalized (Section 12.2). No experimental result, citation, baseline comparison, ablation outcome, or calibration metric in this manuscript has been estimated, extrapolated, or fabricated; every reported value traces to an executed run or a directly computed statistic on real data, documented in the accompanying provenance note.

Funding, competing interests, and author contributions. These are not yet provided and are tracked separately in the accompanying author-information checklist, to be completed by the author(s) prior to submission rather than asserted on their behalf.

24. CONCLUSION

We presented AquaRestore-UC, a lightweight underwater image enhancement network combining physics-guided, wavelet-domain, and spatial-colour restoration branches with an auxiliary, trained-but-not-yet-calibrated per-pixel uncertainty head. In a completed, single-seed (seed 42), 64×64-resolution experiment on a custom 720/80/90 UIEB split, AquaRestore-UC improved mean PSNR by 3.26 dB and mean SSIM from 0.758 to 0.877 over raw input, and scored higher than both a generic and a parameter-matched CNN control trained under matched conditions. A preliminary three-seed extension and a three-seed parameter-matched control are both numerically consistent with this result, reported descriptively rather than as confirmatory multi-seed evidence. This work also documents a real heteroscedastic-uncertainty training failure mode, its correction via a structural two-stage training protocol, and a descriptive, uncalibrated per-pixel uncertainty coverage analysis. Higher-resolution training, published-method baselines, component ablations, a fully matched multi-seed study, and post-hoc uncertainty calibration remain future work. This manuscript reports a preliminary result and does not claim state-of-the-art performance, calibrated uncertainty, or SCI/Q1 submission readiness.

Compliance with Ethical Standards

Disclosure of potential conflicts of interest

Authors declare that they have no conflict of interest.

Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

Data Availability Statement

Data sharing does not apply to this article as no new data has been created or analyzed in this study.

Funding Information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Declarations

Conflicts of interest/Competing interests: Not Applicable

Code availability: Not Applicable

Consent to participate: Not Applicable

Consent for publication: Not Applicable

REFERENCES

[1] Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., Tao, D. "An Underwater Image Enhancement Benchmark Dataset and Beyond." *IEEE Transactions on Image Processing*, 29, 4376–4389, 2019.

[2] Akkaynak, D., Treibitz, T. "A Revised Underwater Image Formation Model." *CVPR* 2018, pp. 6723–6732.

[3] Akkaynak, D., Treibitz, T. "Sea-thru: A Method for Removing Water from Underwater Images." *CVPR* 2019, pp. 1682–1691.

[4] Li, C., Anwar, S., Hou, J., Cong, R., Guo, C., Ren, W. "Underwater Image Enhancement via Medium Transmission-Guided Multi-Color Space Embedding." *IEEE Transactions on Image Processing*, 30, 4985–5000, 2021. DOI: 10.1109/TIP.2021.3076367.

[5] Fu, Z., Wang, W., Huang, Y., Ding, X., Ma, K.-K. "Uncertainty Inspired Underwater Image Enhancement." *ECCV* 2022, LNCS vol. 13678, pp. 465–482.

[6] Naik, A.R., Swarnakar, A., Mittal, K. "Shallow-UWnet: Compressed Model for Underwater Image Enhancement" (Student Abstract). *AAAI* 2021, 35, 15853–15854.

[7] "Five A+ Network: You Only Need 9K Parameters for Underwater Image Enhancement." arXiv:2305.08824.

[8] Zhao, C., Cai, W., Dong, C., Hu, C. "Wavelet-based Fourier Information Interaction with Frequency Diffusion Adjustment for Underwater Image Restoration." *CVPR* 2024, pp. 8281–8291.

[9] "Wavelet–pixel domain progressive fusion network for underwater image enhancement" (WPFNet). *Knowledge-Based Systems*, vol. 299, article 112027, 2024.

[10] "Underwater Image Enhancement by Combining Wavelet Transform and Multiscale Transformer Networks" (WaveletFormer). *Digital Signal Processing*, 2025.

[11] Zhang, W., Zhou, L., Zhuang, P., Li, G., Pan, X., Zhao, W., Li, C. "Underwater Image Enhancement via Weighted Wavelet Visual Perception Fusion." *IEEE Transactions on Circuits and Systems for Video Technology*, 34(4), 2469–2483, 2024.

[12] Yang, M., Sowmya, A. "An Underwater Colour Image Quality Evaluation Metric." *IEEE Transactions on Image Processing*, 24(12), 6062–6071, 2015.

[13] Panetta, K., Gao, C., Agaian, S. "Human-Visual-System-Inspired Underwater Image Quality Measures." *IEEE Journal of Oceanic Engineering*, 41(3), 541–551, 2016.

[14] Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P. "Image Quality Assessment: From Error Visibility to Structural Similarity." *IEEE Transactions on Image Processing*, 13(4), 600–612, 2004.

[15] Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O. "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric." *CVPR* 2018, pp. 586–595.

[16] Johnson, J., Alahi, A., Fei-Fei, L. "Perceptual Losses for Real-Time Style Transfer and Super-Resolution." *ECCV* 2016, pp. 694–711.

[17] Kendall, A., Gal, Y. "What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?" *NeurIPS* 2017, 30, 5574–5584.

[18] Gal, Y., Ghahramani, Z. "Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning." *ICML* 2016, pp. 1050–1059.

[19] Kuleshov, V., Fenner, N., Ermon, S. "Accurate Uncertainties for Deep Learning Using Calibrated Regression." *ICML* 2018, pp. 2796–2804.

APPENDIX

Table A1. Symbol definitions.

Symbol	Meaning
$I(x)$	Observed degraded underwater image
$J(x)$	Latent clean image (approximated by the UIEB reference)
$\hat{J}(x)$	Network's enhanced output

Natural Resources for
Human Health
eISSN: 2583-1194

Symbol	Meaning
$t_c(x)$	Per-channel transmission map
B_c	Per-channel backscatter (veiling light)
J_{phys}	Physics-branch coarse restoration (Equation 2)
J_{wave}	Wavelet-branch reconstructed image
$\sigma^2(x)$, $s(x)$	Predicted per-pixel variance and log-variance
$\Phi_l(\cdot)$	Pretrained VGG-16 feature extractor at layer l
DWT, IDWT	(Inverse) discrete Haar wavelet transform
θ_R, θ_U	Restoration-backbone and uncertainty-head parameters

Table A2. Hyperparameters.

Hyperparameter	Value
Optimizer	AdamW
Learning rate	2×10^{-4}
Weight decay	0.01
Batch size	8
Stage A epochs	10
Stage B epochs	3
Log-variance clamp	$[-10, 5]$
ε (Charbonnier)	10^{-3}
ε (physics inversion)	0.05
λ_{reg} (uncertainty regularizer)	0.01
Fusion channel width	32
Fixed seeds used	42, 123, 2024