Additional Material

1. Loss Formulation

Score Distillation Sampling (SDS). During distillation, the frozen Diffusion Transformer (DiT) guides the Retinex-UNet to align with its denoising prior. The SDS gradient is computed as:

$$\nabla_\theta \mathcal{L}_{\text{SDS}} = \mathbb{E}_{t,\epsilon}\left[w(t)\,(\epsilon_\phi(z_t;t,y)-\epsilon)\,\frac{\partial \hat{I}}{\partial \theta}\right], \tag{1}$$

where z_t is the noisy latent of $\hat{I}$ at timestep t. This enables direct color-prior transfer without iterative diffusion.

Inter-Component Residual (ICR) Regularization. To reduce coupling between reflectance (R) and illumination (L), we apply two lightweight penalties that require no architectural change. The first penalizes cross-component similarity:

$$\mathcal{L}_{\text{icr-cos}} = \mathbb{E}\left[\frac{\langle \text{vec}(R), \text{vec}(L)\rangle}{\|\text{vec}(R)\|_2\,\|\text{vec}(L)\|_2+\varepsilon}\right], \tag{2}$$

where $\varepsilon>0$ ensures numerical stability. The second discourages co-occurring strong gradients:

$$\mathcal{L}_{\text{icr-grad}} = \|\nabla R \odot \nabla L\|_1. \tag{3}$$

Together, these losses suppress halos, ghosting, and color drift while preserving structural fidelity.

Overall Objective. The Retinex-UNet is optimized with the joint loss:

$$\mathcal{L}_{\text{UNet}} = \mathcal{L}_{\text{SDS}} + \lambda_{\text{icr}}\big(\mathcal{L}_{\text{icr-cos}} + \mathcal{L}_{\text{icr-grad}}\big), \tag{4}$$

where λ_{icr} controls the regularization strength. Inference requires only a single UNet forward pass to produce (R,L) and $\hat{I}$, achieving stable, artifact-free color reconstruction.

2. Training Data Construction via Physical Prior

To provide the Diffusion Transformer (DiT) with a robust, physics-guided prior, we synthesize underwater images from the NYU-Depth V2 dataset [SHKF12] using a physical image-formation model, as illustrated in Figure 1. Specifically, for each RGB channel $c\in\{R,G,B\}$, we apply:

$$I_c^{\text{uw}} = I_c\, e^{-\eta_c d} + B_c\big(1-e^{-\eta_c d}\big), \tag{5}$$

where $\eta_c d\in[0,5]$ and $B_c\in[0,1]$ denote the wavelength-dependent attenuation and background light, respectively. As shown in Figure 1, this process generates synthetic underwater training images from terrestrial data (top), while the model is evaluated on real underwater scenes (bottom). Several visual examples of the synthesized data are further illustrated in Figure 2, demonstrating the diverse color casts and visibility degradations captured by the physical prior.

To verify the fidelity of the synthetic dataset, we analyze its color statistics in the CIELAB space. Following [PZB23, QT09], we randomly sample 2000 images each from LSUI (real underwater), MIT67 (terrestrial), and our synthesized dataset, computing the mean and standard deviation of the a and b channels. As summarized in Table 2, our synthesized data exhibit close chromatic alignment with real underwater images while remaining distinct from the terrestrial domain, confirming that the physical-prior formulation effectively models underwater color characteristics.

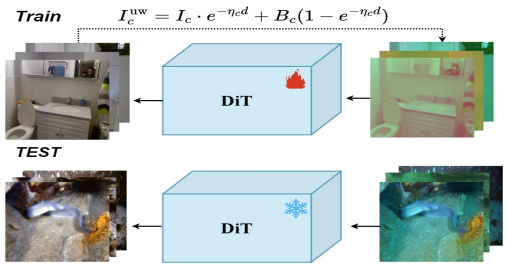

Figure 1: ***Training and testing pipeline.*** *The Train set is synthesized from terrestrial images using a physical prior, while testing is performed on real underwater images.*

Figure 2: ***Examples of synthetic underwater images*** *generated from terrestrial images using the physical prior. The results exhibit diverse color casts and illumination degradations, enriching the training domain.*

Table 1: *PSNR and SSIM results on the synthetic underwater dataset. Higher is better.*

Metric	PCDE [ZJZ*23]	WFAC [ZLL*25]	WWPF [ZZZ*24]	U-shape [PZB23]	Spectroformer [KMM*24]	Phaseformer [KNK*25]	CE-VAE [PM25]	SS-UIE [PB25]	SMDR [ZZG*24]	PhISH-Net [CSB24]	WF-Diff [ZCDH24]	UIEDP [DDSX*]	DM-underwater [TKI23]	Diffwater [GXJ*23]	**RetiDiff (Ours)**
PSNR ↑	21.02	23.70	23.29	26.16	25.71	30.92	27.82	28.16	27.25	26.48	27.02	26.08	23.69	23.12	**39.63**
SSIM ↑	0.71	0.73	0.81	0.79	0.81	0.85	0.86	0.86	0.83	0.81	0.78	0.78	0.72	0.69	**0.97**

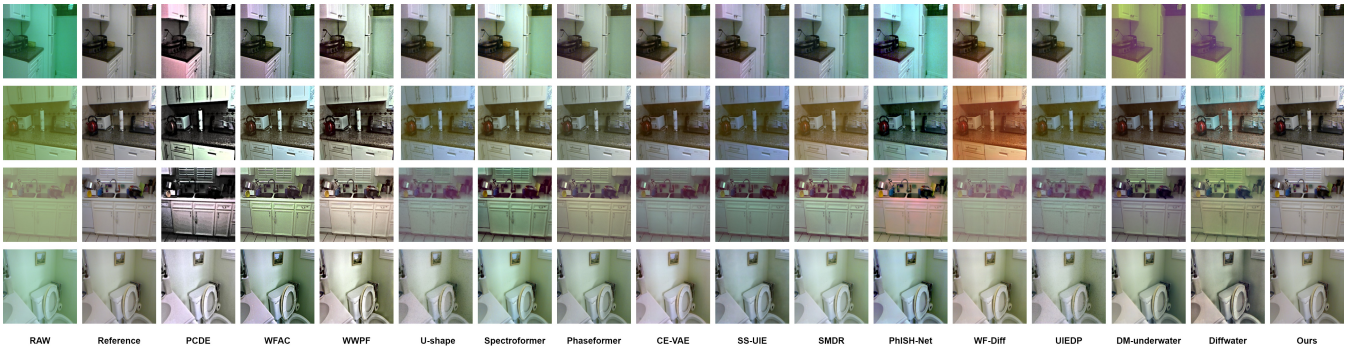

Figure 3: *Qualitative comparison on the synthetic underwater dataset.* ***RetiDiff*** *effectively restores color distributions to match terrestrial references, while other methods retain domain-specific color bias or structural artifacts.*

Table 2: *Statistical validation of synthesized images in CIELAB space (a, b channels). The close alignment between our synthesized dataset and real underwater images validates the chromatic fidelity achieved through the physics-guided data generation process.*

Categories	a-mean	a-std	b-mean	b-std
Underwater Images	-19.54	12.58	-1.64	14.54
Land Images	2.61	3.05	4.06	4.95
Synthesized Images	-17.15	10.12	0.79	11.95

3. Results on the Synthetic Dataset

To further verify the color reconstruction ability of **RetiDiff**, we evaluate all methods on our physics-based synthetic underwater dataset, where the ground truth terrestrial counterparts are available. This setting enables direct observation of how well each model restores the color distribution of underwater images toward that of land scenes.

As shown in Figure 3, traditional CNN-, GAN-, and transformer-based approaches exhibit residual color casts or contrast imbalance, failing to recover the natural color statistics of terrestrial domains. In contrast, **RetiDiff** produces balanced illumination and chromatic consistency closely aligned with the reference land images, demonstrating its ability to reconstruct the true color distribution rather than merely enhancing visual contrast.

We also provide quantitative PSNR and SSIM results in Table 1 to complement perceptual analysis. **RetiDiff** achieves the highest fidelity scores, confirming its capacity to recover both accurate color distributions and structural consistency under synthetic-to-real generalization.

4. SIFT-Based Structural Consistency

To further validate the structural fidelity of **RetiDiff**, we employ SIFT-based keypoint consistency analysis [Low04, AAHB12]. The SIFT algorithm extracts and matches local features between the input underwater image I^{uw} and the reconstructed output $\hat{I}$. As illustrated in Figure 4, RetiDiff achieves dense and accurate feature correspondences, where high inlier ratios and strong match similarity indicate that spatial structures are well preserved during enhancement rather than distorted by color correction. Quantitatively, Table 3 reports average matches, inliers, and inlier ratios across UIEB, LSUI, and TEST-U45, showing $\sim 97\%$ keypoint similarity, which corroborates the model's strong structural consistency across diverse scenes.

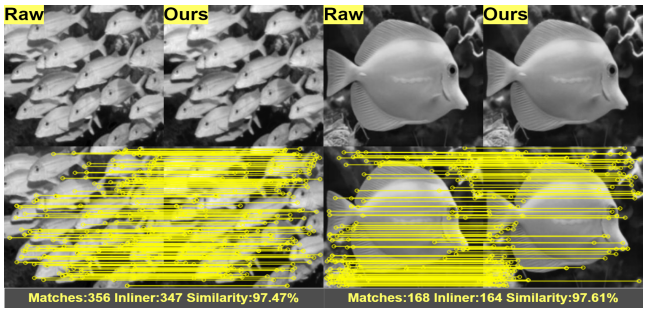

Figure 4: *SIFT-based keypoint matching between grayscale maps of the input (left) and reconstructed image (right). Yellow lines denote correspondences; overlays report match, inlier, and similarity statistics. RetiDiff achieves $\sim 97\%$ keypoint similarity, evidencing structural preservation.*

The results show that across UIEB, LSUI, and TEST-U45

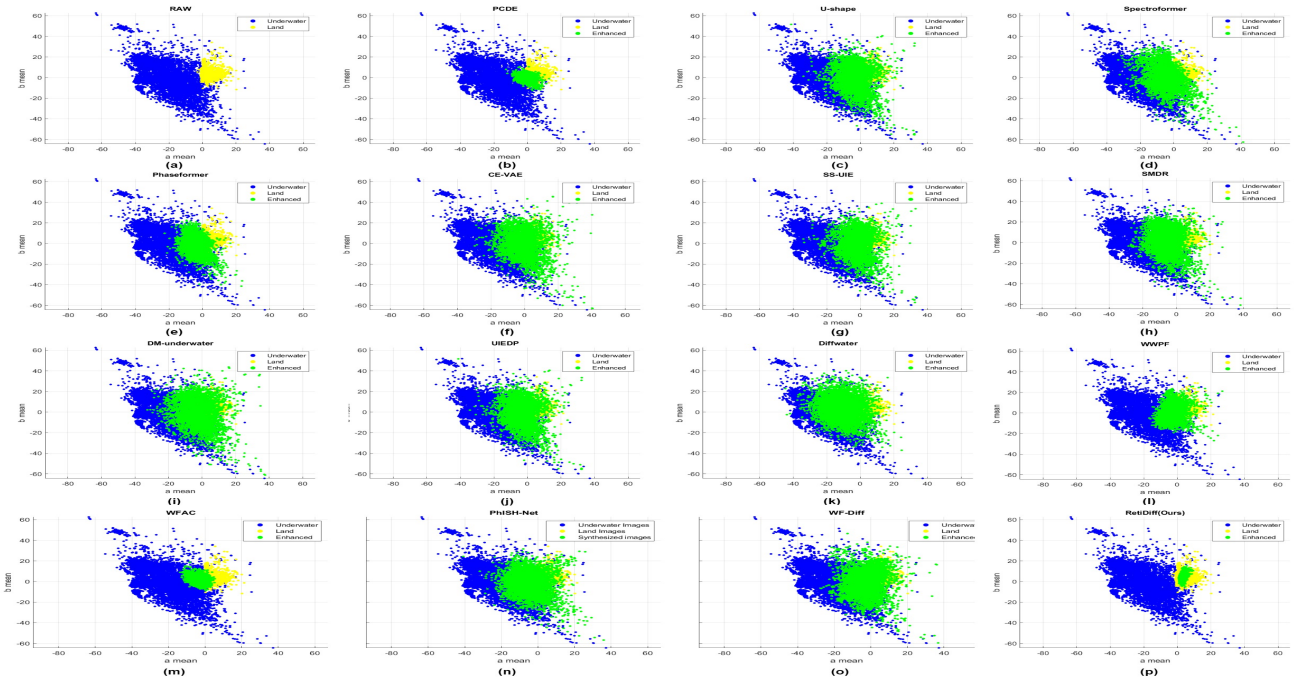

Figure 5: *Distributions of underwater (blue), enhanced (green), and terrestrial (yellow) images in the CIELAB color space. Each point denotes one image based on the mean a–b values. RetiDiff shifts the chromatic distribution of enhanced images toward the terrestrial cluster, reducing color bias caused by underwater absorption.*

Table 3: *SIFT-based keypoint consistency between input I^{uw} and output $\hat{I}$. We report the average number of inliers, total matches, and inlier ratio (Average similarity).*

Metrics	UIEB	LSUI	TEST-U45
Average inliers	278.86	327.45	292.42
Average matches	287.10	336.27	301.22
Average similarity (%)	97.13	97.38	97.08

datasets, **RetiDiff** maintains $\sim 97\%$ keypoint similarity, confirming that its color reconstruction process preserves geometric structures and fine spatial details.

5. CIELAB Color Distribution Analysis

To visually evaluate the effectiveness of **RetiDiff** in restoring natural color tones, we analyze the distribution of enhanced images in the CIELAB color space. Each image is represented by the mean values of its a and b channels, which describe chromaticity along the red–green and yellow–blue axes. As shown in Figure 5, underwater images (blue) cluster toward negative a and b values due to wavelength-dependent attenuation, while the enhanced results (green) shift markedly toward the terrestrial domain (yellow). This movement indicates that **RetiDiff** effectively corrects underwater color casts and reconstructs color distributions consistent with those of land images, demonstrating its capacity for physically grounded and visually natural color restoration.

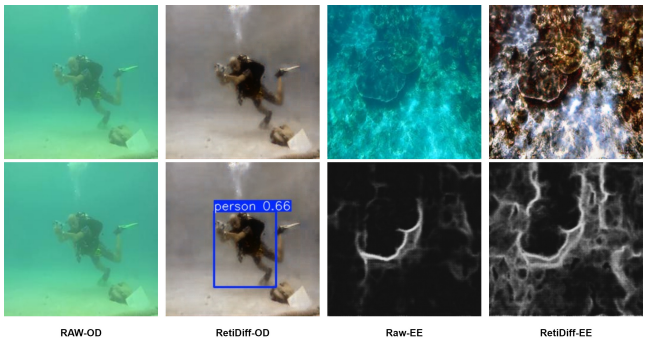

Figure 6: *Applicability of **RetiDiff** in downstream tasks. Left: Object detection (OD) under YOLOv7 [WBL23], where the human target undetected in the raw image becomes detectable after RetiDiff enhancement. Right: Sobel edge extraction (EE), where RetiDiff restores fine details and continuous contours in reef textures.*

6. Applicability Analysis

To further demonstrate the practical value of **RetiDiff**, we evaluate its impact on downstream visual tasks, including object detec-

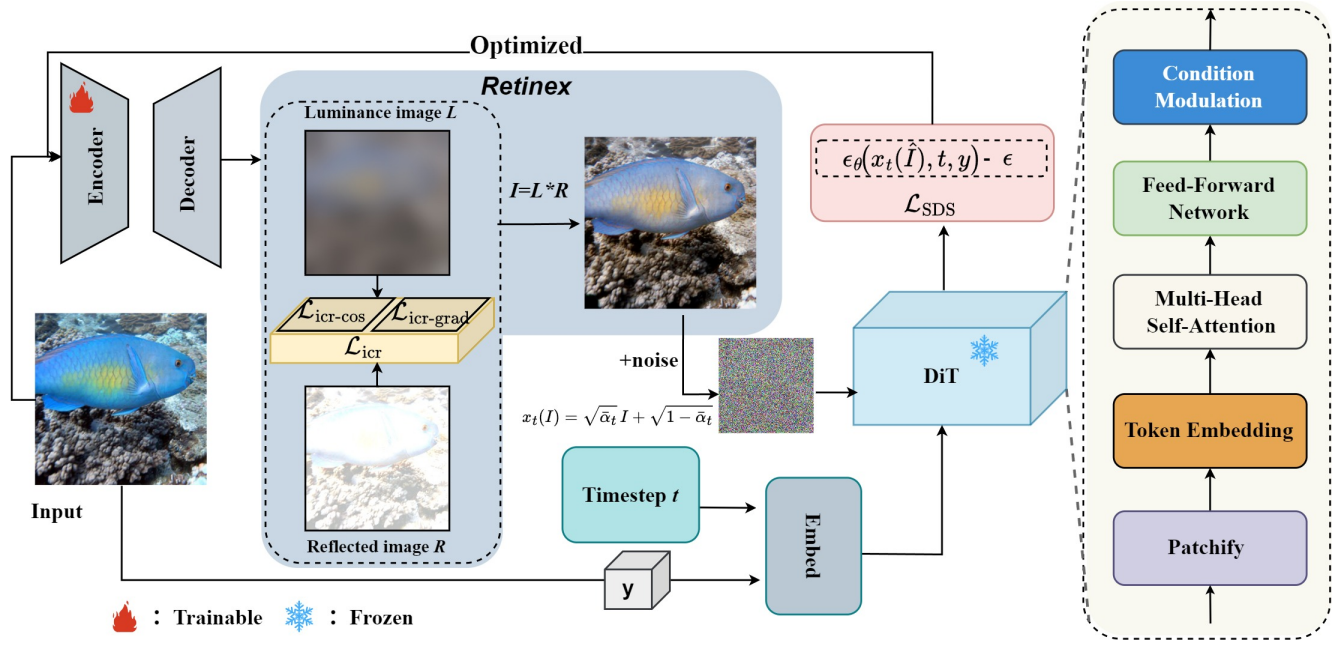

Figure 7: *Architecture of RetiDiff.*

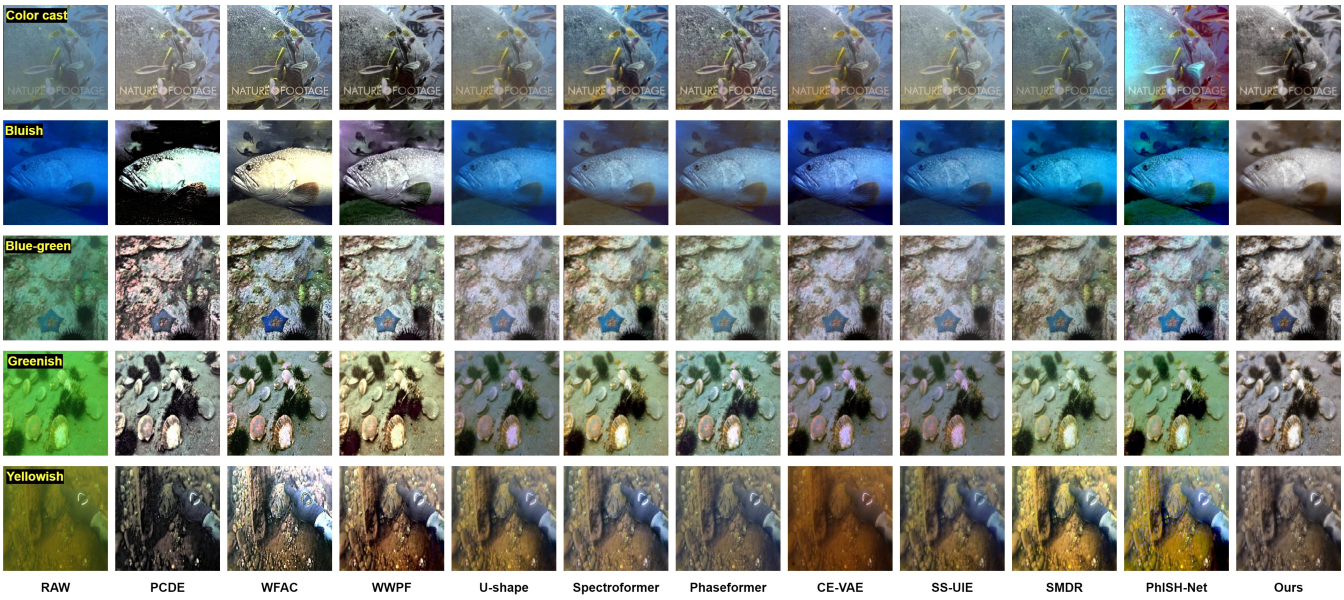

Figure 8: *Qualitative comparison of the proposed method with state-of-the-art approaches on real underwater images.*

tion and edge extraction. As illustrated in Figure 6, the left pair (RAW–OD vs. RetiDiff–OD) presents object detection results under YOLOv7 [WBL23]. In the raw underwater image, the human target is barely visible and fails to be detected due to low contrast and severe color degradation. After enhancement by **RetiDiff**, the restored image exhibits balanced chromaticity and higher local contrast, enabling successful detection with tighter bounding boxes and higher confidence.

The right pair (RAW–EE vs. RetiDiff–EE) shows Sobel-based edge extraction on a reef scene. While the raw image produces weak and discontinuous edge responses, the RetiDiff-enhanced result reveals finer structural details and more continuous contours, especially around coral textures and object boundaries. These results confirm that **RetiDiff** not only restores perceptual realism but also provides feature-consistent enhancements beneficial to downstream vision tasks in underwater environments.

References

[AAHB12] ANCUTI C., ANCUTI C. O., HABER T., BEKAERT P.: Enhancing underwater images and videos by fusion. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (2012), pp. 81–88. doi:10.1109/CVPR.2012.6247661. 2

[CSB24] CHANDRASEKAR A., SREENIVAS M., BISWAS S.: Phishnet: Physics inspired system for high resolution underwater image enhancement. In *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (2024), pp. 1495–1505. doi:10.1109/WACV57701.2024.00153. 2

[DDSX*] DU D L. E., SI L., XU F., NIU J., SUN F.: Uiedp: underwater image enhancement with diffusion prior (2023). *arXiv preprint arXiv:2312.06240*. 2

[GXJ*23] GUAN M., XU H., JIANG G., YU M., CHEN Y., LUO T., ZHANG X.: Diffwater: Underwater image enhancement based on conditional denoising diffusion probabilistic model. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 17* (2023), 2319–2335. 2

[KMM*24] KHAN R., MISHRA P., MEHTA N., PHUTKE S. S., VIPPARTHI S. K., NANDI S., MURALA S.: Spectroformer: Multi-domain query cascaded transformer network for underwater image enhancement. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (2024), pp. 1454–1463. 2

[KNK*25] KHAN M. R., NEGI A., KULKARNI A., PHUTKE S. S., VIPPARTHI S. K., MURALA S.: Phaseformer: Phase-based attention mechanism for underwater image restoration and beyond. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (2025), IEEE, pp. 9618–9629. 2

[Low04] LOWE D. G.: Distinctive image features from scale-invariant keypoints. *International journal of computer vision 60*, 2 (2004), 91–110. 2

[PB25] PENG L., BIAN L.: Adaptive dual-domain learning for underwater image enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence* (2025), vol. 39, pp. 6461–6469. 2

[PM25] PUCCI R., MARTINEL N.: Ce-vae: Capsule enhanced variational autoencoder for underwater image enhancement. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (2025), IEEE, pp. 2113–2123. 2

[PZB23] PENG L., ZHU C., BIAN L.: U-shape transformer for underwater image enhancement. *IEEE transactions on image processing 32* (2023), 3066–3079. 1, 2

[QT09] QUATTONI A., TORRALBA A.: Recognizing indoor scenes. In *2009 IEEE conference on computer vision and pattern recognition* (2009), IEEE, pp. 413–420. 1

[SHKF12] SILBERMAN N., HOIEM D., KOHLI P., FERGUS R.: Indoor segmentation and support inference from rgbd images. In *European conference on computer vision* (2012), Springer, pp. 746–760. 1

[TKI23] TANG Y., KAWASAKI H., IWAGUCHI T.: Underwater image enhancement by transformer-based diffusion model with non-uniform sampling for skip strategy. In *Proceedings of the 31st ACM International Conference on Multimedia* (2023), pp. 5419–5427. 2

[WBL23] WANG C.-Y., BOCHKOVSKIY A., LIAO H.-Y. M.: Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2023), pp. 7464–7475. 3, 5

[ZCDH24] ZHAO C., CAI W., DONG C., HU C.: Wavelet-based fourier information interaction with frequency diffusion adjustment for underwater image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), pp. 8281–8291. 2

[ZJZ*23] ZHANG W., JIN S., ZHUANG P., LIANG Z., LI C.: Underwater image enhancement via piecewise color correction and dual prior optimized contrast enhancement. *IEEE Signal Processing Letters 30* (2023), 229–233. doi:10.1109/LSP.2023.3255005. 2

[ZLL*25] ZHANG W., LIU Q., LU H., WANG J., LIANG J.: Underwater image enhancement via wavelet decomposition fusion of advantage contrast. *IEEE Transactions on Circuits and Systems for Video Technology 35*, 8 (2025), 7807–7820. doi:10.1109/TCSVT.2025.3545595. 2

[ZZG*24] ZHANG D., ZHOU J., GUO C., ZHANG W., LI C.: Synergistic multiscale detail refinement via intrinsic supervision for underwater image enhancement. In *Proceedings of the AAAI conference on artificial intelligence* (2024), vol. 38, pp. 7033–7041. 2

[ZZZ*24] 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 (2024), 2469–2483. doi:10.1109/TCSVT.2023.3299314. 2