Spectroformer: Multi-Domain Query Cascaded Transformer Network For Underwater Image Enhancement

MD Raqib Khan, Priyanka Mishra, Nancy Mehta, Shruti S. Phutke
Santosh Kumar Vipparthi, Sukumar Nandi, Subrahmanyam Murala

Supplementary Material

Overview

The Supplementary material includes:

- Details on loss function used for training the network.
- Quantitative results comparison of various losses settings on the UIEB dataset Table S 1
- Qualitative comparison of results obtained with various loss settings in Figure S 1.
- Qualitative results of the proposed Spectroformer and existing state-of-the-arts for underwater image restoration on real-world UCCS dataset in Figure S 2.
- Qualitative results comparison on real-world SQUID dataset in Figure S 3.
- Qualitative comparison of the proposed Spectroformer and existing state-of-the-arts for underwater image restoration on underwater real-world U45 dataset in Figure S 4.
- Application of the proposed Spectroformer and existing state-of-the-arts as pre-processing step for depth on underwater U45 dataset in Figure S 5.

Training Losses

$$L_{Total} = \lambda_1 * L_1 + \lambda_2 * L_2 + \lambda_3 * L_3 + \lambda_4 * L_4 \tag{1}$$

Where, $\lambda_{1,2,3,4} \in (0.03, 0.02, 0.01, 0.025)$, Perceptual loss ($L_1$), Charbonnier loss ($L_2$), Multiscale Structural Similarity Index (MS-SSIM) loss (L_3), and Gradient loss (L_4).

Perceptual Loss (L_1):
Perceptual loss provides a way to measure the perceptual similarity between generated and target images using the feature representations of a pre-trained neural network. It has proven effective in enhancing the quality of generated images in various image generation tasks. Let O be the target image and G_t be the generated image. We use a pre-trained VGG19 network (ϕ_i) to extract feature maps at different layers. The perceptual loss, L_1, is then computed as the difference between the feature maps of the target and restored images:

$$L_1 = \sum_{i=1}^{N=4} \|\phi_i(O) - \phi_i(G_t)\|_2^2 \tag{2}$$

Here, ϕ_i represents the feature extraction function at layer i of the CNN, and $(N = 4)$ is the total number of layers considered for perceptual loss calculation.

Charbonnier loss (L_2):

Using the MSE loss to train the network generally causes blurry reconstruction because it maximizes the log-likelihood of a Gaussian distribution. We opted for the Charbonnier loss, a differentiable version of the L_1 norm,, to avoid this issue. The Charbonnier loss is determined between the restored image images (O) and their corresponding ground-truth image (G_t), and it is defined as follows:

$$L_2 = \mathbb{E}_{O\sim Q(O), G_t\sim Q(G_t)} \sqrt{(O - G_t)^2 + \epsilon} \tag{3}$$

Where, $Q(O)$ and $Q(G_t)$ are the distributions of the restored image (O) and the ground-truth image (G_t), respectively. Additionally, the value of ϵ is empirically set to 1×10^{-3}.

MS-SSIM loss (L_3):

The Structural Similarity (SSIM) loss primarily deals with a single input resolution. In contrast, the Multiscale SSIM (MS-SSIM) loss offers greater flexibility by considering different input resolutions.

$$L_3 = 1 - (MSSSIM(O, G_t)) \tag{4}$$

Gradient loss (L_4):

Generally, the Charbonnier loss prioritizes low-frequency components. However, when training the network to incorporate high-frequency details, the gradient loss plays an important role. It is a second-order loss function that enhances the sharpness of edges in the output [1]. $\hat{G_O}$ and $\hat{G_{G_t}}$ represents distribution of $Q(O)$ and $Q(G_t)$ respectively.

$$L_4 = \mathbb{E}_{\hat{G_O} \sim Q(O), \hat{G}_O \sim Q(G_t)} \left\| \hat{G}_{G_t} - \hat{G}_O \right\|_1 \tag{5}$$

Table S 1: Quantitative results comparison of various loss settings on the UIEB dataset *(L_1: Perceptual loss, L_2: Charbonnier loss, L_3: Multiscale Structural Similarity Index (MS-SSIM) loss and L_4: Gradient loss)*.

Loss settings	PSNR	SSIM
L_1	21.99	0.866
$L_1 + L_2$	22.04	0.886
$L_1 + L_2 + L_3$	24.61	0.901
$L_1 + L_2 + L_3 + L_4$	**24.96**	**0.917**

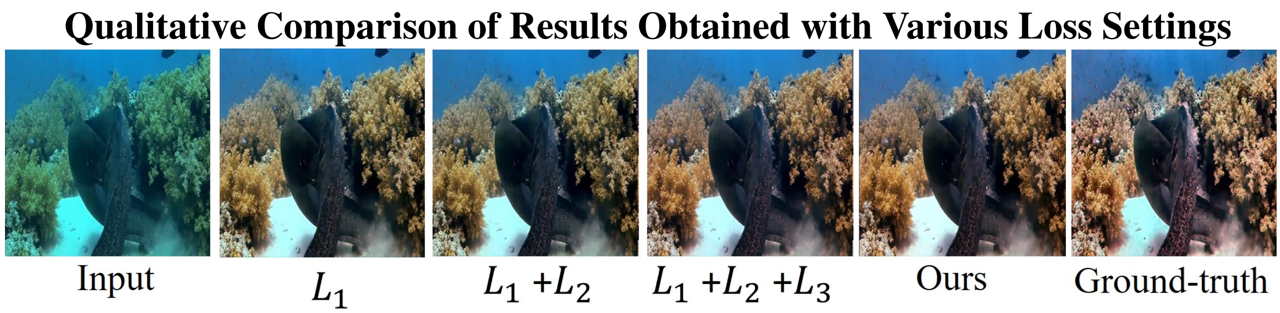

Figure S 1: Qualitative comparison of results obtained with various loss settings.

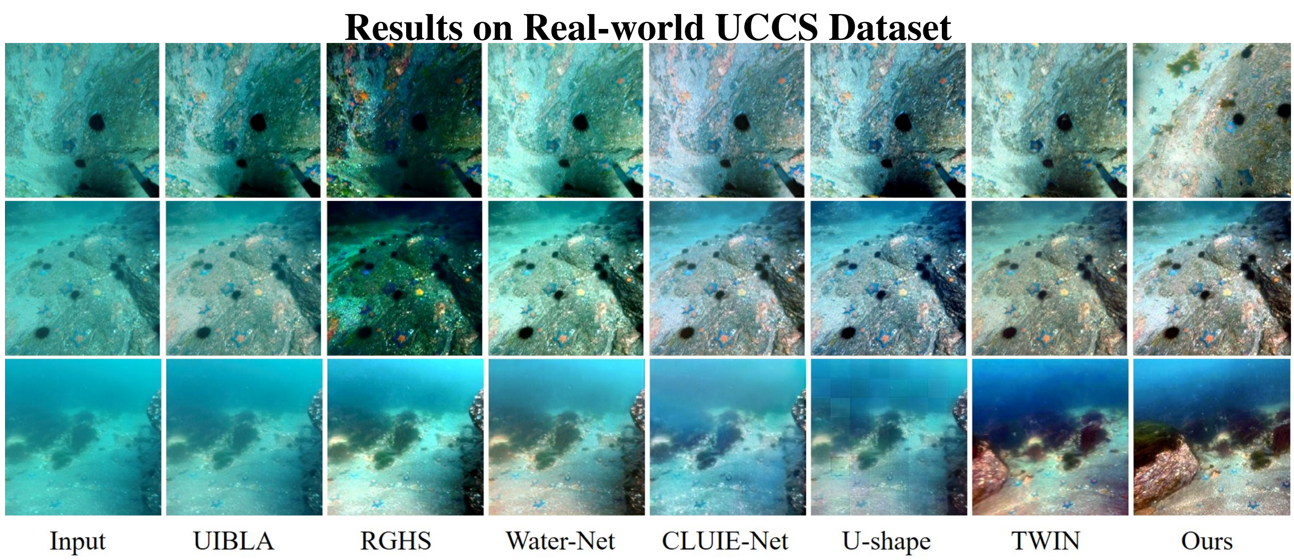

Figure S 2: Qualitative comparison of the proposed method (Ours) with existing state-of-the-art methods (UIBLA [2], RGHS [3], Water-Net [4], CLUIE-Net [5], U-shape [6], TWIN [7]) for underwater image restoration on real-world UCCS dataset [8].

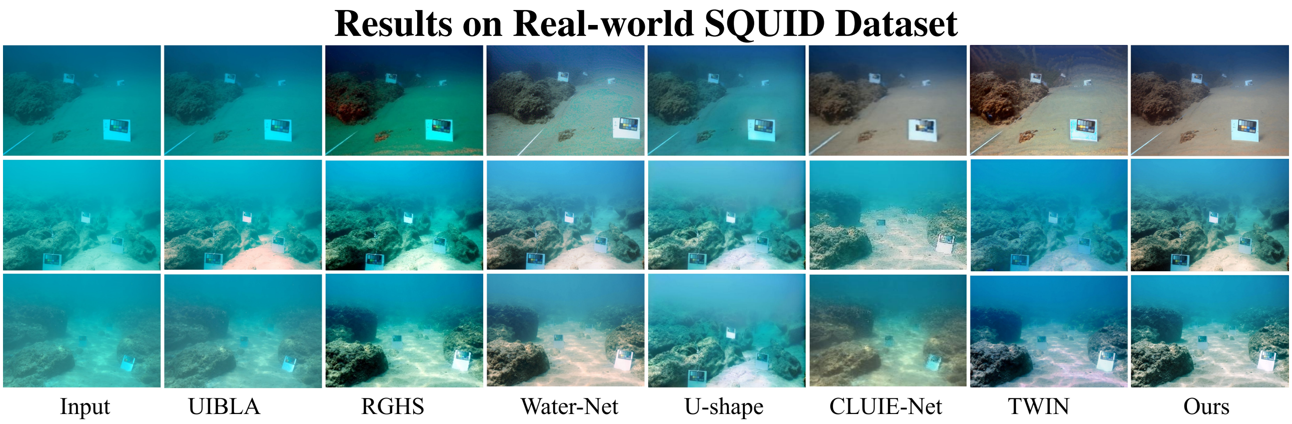

Figure S 3: Qualitative results comparison of state-of-the-art (UIBLA [2], RGHS [3], Water-Net [4], CLUIE-Net [5], U-shape [6], TWIN [7]) and the proposed method (Ours) on SQUID dataset [9] for underwater image restoration.

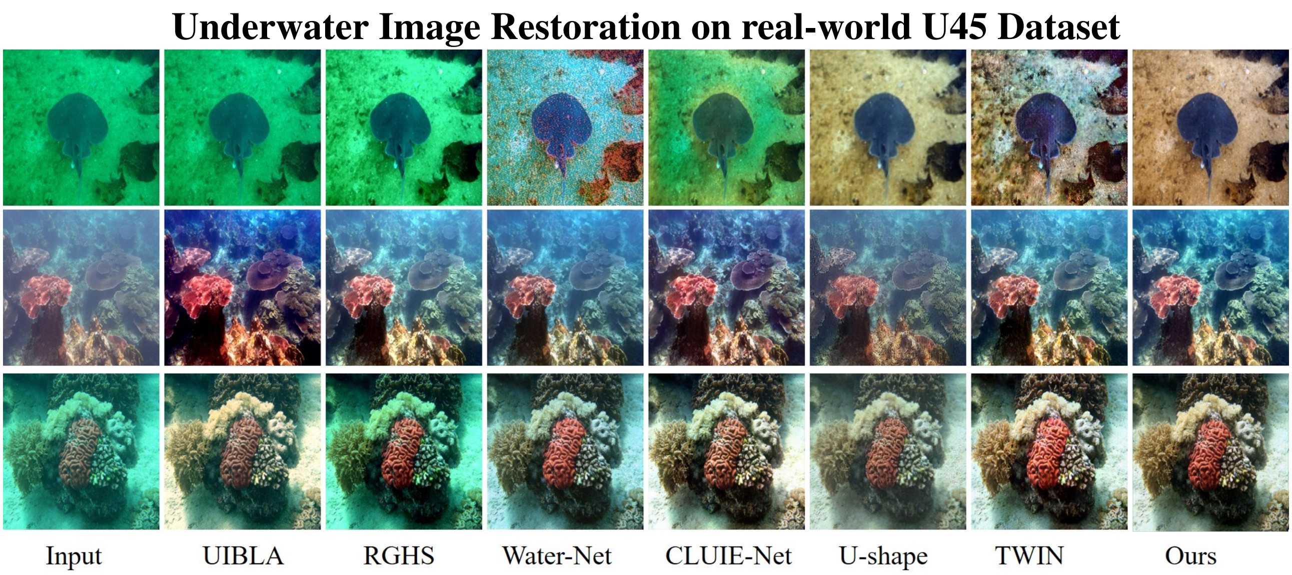

Figure S 4: Qualitative comparison of the proposed method (Ours) with existing state-of-the-art methods (UIBLA [2], RGHS [3], Water-Net [4], CLUIE-Net [5], U-shape [6], TWIN [7]) for underwater image restoration on real-world U45 dataset [8].

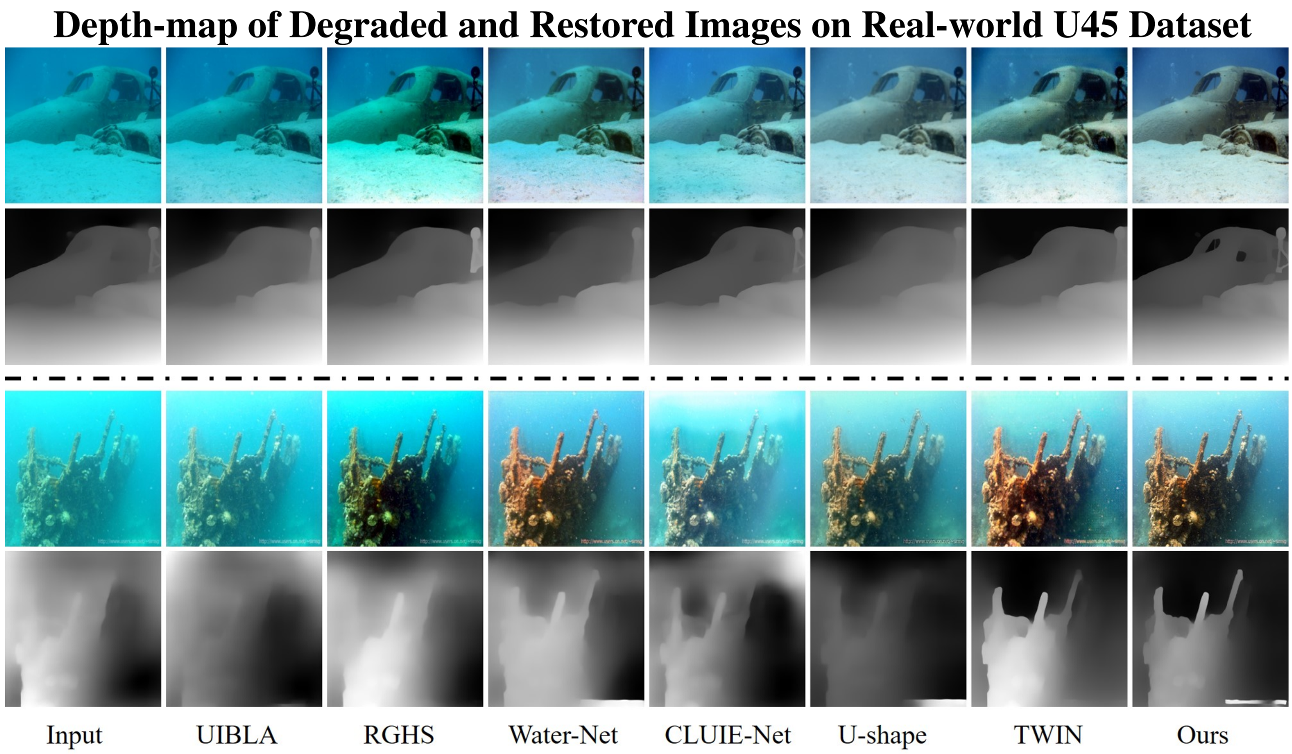

Figure S 5: Application of the proposed Spectroformer and existing state-of-the-arts (UIBLA [2], RGHS [3], Water-Net [4], CLUIE-Net [5], U-shape [6], TWIN [7]) as a pre-processing step for depth-estimation on underwater U45 dataset [10].

References

[1] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," *arXiv preprint arXiv:1511.05440*, 2015.

[2] Y.-T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE transactions on image processing*, vol. 26, no. 4, pp. 1579–1594, 2017.

[3] D. Huang, Y. Wang, W. Song, J. Sequeira, and S. Mavromatis, "Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition," in *MultiMedia Modeling: 24th International Conference, MMM 2018, Bangkok, Thailand, February 5-7, 2018, Proceedings, Part I 24*. Springer, 2018, pp. 453–465.

[4] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE Transactions on Image Processing*, vol. 29, pp. 4376–4389, 2019.

[5] K. Li, L. Wu, Q. Qi, W. Liu, X. Gao, L. Zhou, and D. Song, "Beyond single reference for training: underwater image enhancement via comparative learning," *IEEE Transactions on Circuits and Systems for Video Technology*, 2022.

[6] L. Peng, C. Zhu, and L. Bian, "U-shape transformer for underwater image enhancement," in *Computer Vision–ECCV 2022 Workshops: Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part II*. Springer, 2023, pp. 290–307.

[7] R. Liu, Z. Jiang, S. Yang, and X. Fan, "Twin adversarial contrastive learning for underwater image enhancement and beyond," *IEEE Transactions on Image Processing*, vol. 31, pp. 4922–4936, 2022.

[8] H. Li, J. Li, and W. Wang, "A fusion adversarial underwater image enhancement network with a public test dataset," *arXiv preprint arXiv:1906.06819*, 2019.

[9] D. Berman, T. Treibitz, and S. Avidan, "Single image dehazing using haze-lines," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 3, pp. 720–734, 2018.

[10] M. J. Islam, P. Luo, and J. Sattar, "Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception," *arXiv preprint arXiv:2002.01155*, 2020.