

# Supplementary Materials: Enhancing Underwater Images via Asymmetric Multi-Scale Invertible Networks

## 1 INTERPRETATION FROM DECOMPOSITION

Recall that in our proposed approach, an input underwater image  $Y$  is perceived as the composition of a distortion-free image  $X$  and a distortion map  $D$  through a non-linear process  $f(\cdot, \cdot)$ . The map  $D$  encapsulates all necessary degradation information of  $X$ , enabling perfect reconstruction of  $Y$  from  $X$  and  $D$ . Our method can be envisioned as learning to simultaneously predict both  $X$  and  $D$  from  $Y$ , i.e.,  $\mathcal{G}(Y) \rightarrow (X, D)$ . In this context, the composition process  $f = \mathcal{G}^{-1}$  naturally arise as an integrated outcome. In other words, our approach can be conceptually interpreted as the integration of three inter-connected components:

$$\mathcal{G}_{cp}(\cdot; \theta_{cp})[\text{implicit}], \mathcal{G}_{dm}(\cdot; \theta_{dm}), \mathcal{G}_{im}(\cdot; \theta_{im}),$$

which accounts for the composite function, distortion map predictor, and distortion-free image predictor, parametrized by  $\theta_{cp}$ ,  $\theta_{dm}$ ,  $\theta_{im}$ , respectively. Then, the interactions among these components across stages in the NN can be conceptually described as follows: for  $t = 1, 2, \dots, T$ :

$$\begin{aligned} D^t &:= \mathcal{G}_{dm}(X^{t-1}, D^{t-1}; \theta_{dm}^t), \\ X^t &:= \mathcal{G}_{im}(X^{t-1}, D^{t-1}; \theta_{im}^t), \\ \text{s.t. } \mathcal{G}_{cp}(D^t, X^t; \{\theta_{cp}^1, \dots, \theta_{cp}^t\}) &= Y, \end{aligned}$$

where  $D^t, X^t$  denote the features of distortion map and the distortion-free image at the  $t$ -stage, respectively.

## 2 DETAILS OF COMPARED METHODS

Training data varies in exiting literature of UIE. Following TACL [2], the training set of UIEBD with GTs is used in our experiment. Therefore, for TACL, we directly quote the results in PSNR and SSIM from its original paper [2]. Note that the programs for calculating the results of UIQM and UCIQE are not included in the released codes of [2], and different implementations may lead to different results. Therefore, we utilize the code of UIQM provided by [1] and the code of UCIQE provided by [5] to re-calculate the UIQM and UCIQE results of TACL as well as all other compared methods. As a result, the UIQM and UCIQE results differ from the ones reported in [2]. In addition, we also note that TACL does not perform very well on SUD. The reason is probably that TACL utilizes a detector model trained on real-world underwater images, which does not work well on the synthetic images of SUD.

Regarding Waternet and SemiUIR which need to call traditional methods before inference, we omit their computational time of running those traditional methods. Hence, the actual test time of these two methods is longer than that reported in the main paper.

## 3 SUPPLEMENTAL EXPERIMENTAL RESULTS

### 3.1 Comparison on LSUI dataset

We also perform evaluation on the LSUI (large-scale underwater image) dataset [3] which contains 3879 paired underwater images for training and 400 for test. See the quantitative results on Table 1 and the visual results in Fig. 1.

Table 1: PSNR/SSIM on LSUI. Bold: best; Underline: 2nd-best

NU <sup>2</sup> Net	Semi-UIR	EUIVF	AACP
24.33/0.915	24.29/0.898	17.50/0.663	17.71/0.711
USFormer	DM-Water	GUPDM	AMSIN
24.16/ <b>0.930</b>	<u>27.65</u> /0.886	25.35/0.922	<b>28.29</b> / <u>0.926</u>

### 3.2 Comparison to General Transformer Models

Table 2 compares AMSIN to two recent Transformer-based methods designed for general image restoration: Uformer [4] and Restormer [6]. For Restormer, in addition to its original model, we also evaluate a reduced-size version (denoted by Restormer\*) whose channel number is reduced to have a similar model size as AMSIN. From the results in the table, we can see that AMSIN outperforms those Transformer-based method. We can also see that Uformer performs worse than Restormer. The reason is probably that the channel self-attention of Restormer is more efficient for recovering chrominance information in underwater images, compared to the spatial self-attention used in Uformer.

### 3.3 More Visual Comparison

**Demo on underwater video enhancement:** See Fig. 2 for the frame-by-frame processing using our method on underwater videos.

**Visualization on residuals:** See Fig. 3.

**Additional visual results:** See Figs. 4, 5, 6, 7, 8, 9 for more visual results. It can be observed that our method better enhances the brightness and contrast in almost all examples. Moreover, in comparison to other methods, it achieves more accurate color restoration (as demonstrated in UIEBD-UP, the second example from UCCS, SUIM, and U45), along with better detail restoration (as exemplified in the second example from UIEBD-P).

## REFERENCES

- [1] Zhenqi Fu, Huangxing Lin, Yan Yang, Shu Chai, Liyan Sun, Yue Huang, and Xinghao Ding. 2022. Unsupervised underwater image restoration: From a homology perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 643–651.
- [2] Zhiying Jiang, Zhuoxiao Li, Shuzhou Yang, Xin Fan, and Risheng Liu. 2022. Target oriented perceptual adversarial fusion network for underwater image enhancement. *IEEE Transactions on Circuits and Systems for Video Technology* 32, 10 (2022), 6584–6598.
- [3] Lintao Peng, Chunli Zhu, and Liheng Bian. 2023. U-shape transformer for underwater image enhancement. *IEEE Transactions on Image Processing* (2023).
- [4] Zhendong Wang, Xiaodong Cun, Jianmin Bao, Wengang Zhou, Jianzhuang Liu, and Houqiang Li. 2022. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 17683–17693.
- [5] Jieyu Yuan, Wei Cao, Zhanchuan Cai, and Binghua Su. 2020. An underwater image vision enhancement algorithm based on contour bougie morphology. *IEEE Transactions on Geoscience and Remote Sensing* 59, 10 (2020), 8117–8128.
- [6] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5728–5739.

**Table 2: Quantitative comparison on six datasets. Bold: best.**

Method	Source	UIEBD-P		UIEBD-UP		EUVP		UCCS		SUIM		U45	
		PSNR	SSIM	UIQM	UCIQE	UIQM	UCIQE	UIQM	UCIQE	UIQM	UCIQE	UIQM	UCIQE
Uformer [4]	CVPR2022	19.86	0.869	2.0902	0.5153	2.9343	0.5210	3.5921	0.4533	2.1004	0.5810	4.0898	0.5577
Restormer [6]	CVPR2022	22.40	0.904	2.2341	0.5580	3.0812	0.5726	4.0659	0.5251	2.2972	0.6014	4.0584	0.5924
Restormer* [6]	CVPR2022	22.21	0.903	2.1298	0.5548	3.0870	0.5763	3.9523	0.5213	2.2633	0.6019	4.0760	0.5942
AMSIN	Proposed	<b>24.16</b>	<b>0.918</b>	<b>2.2843</b>	<b>0.6081</b>	<b>3.2115</b>	<b>0.6141</b>	<b>4.1493</b>	<b>0.5780</b>	<b>2.3758</b>	<b>0.6348</b>	<b>4.4598</b>	<b>0.6216</b>

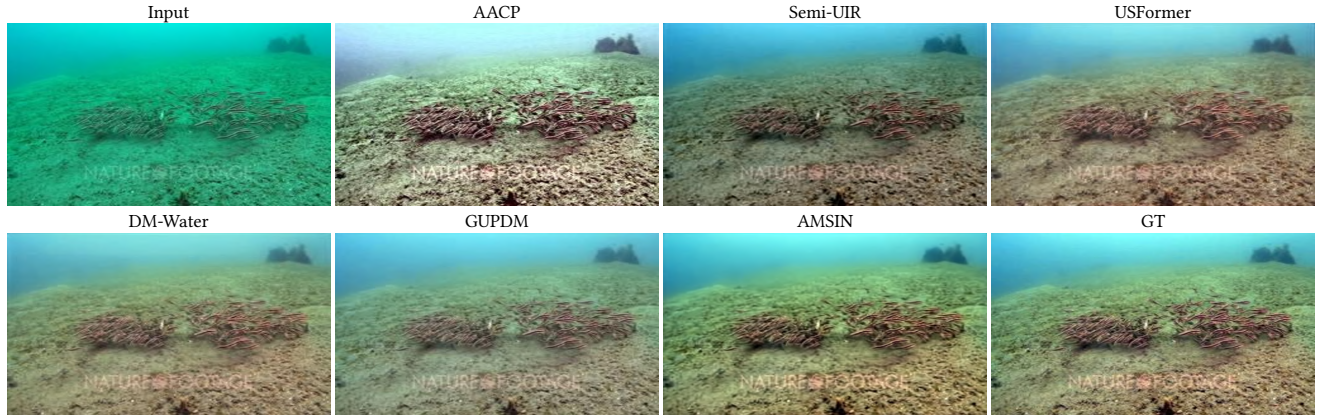
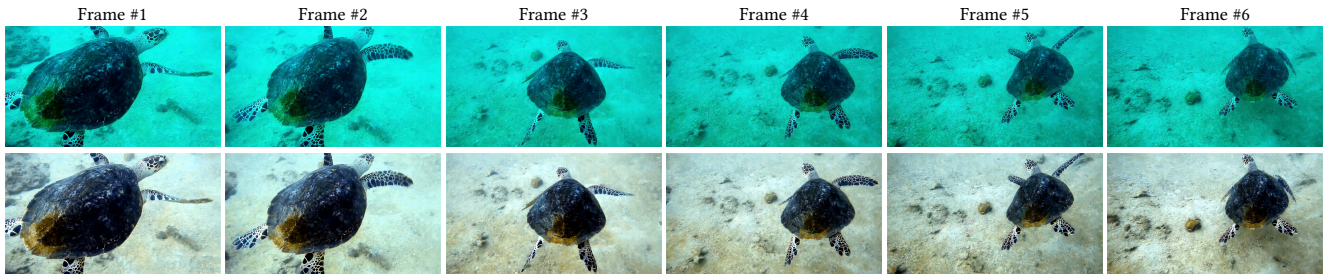
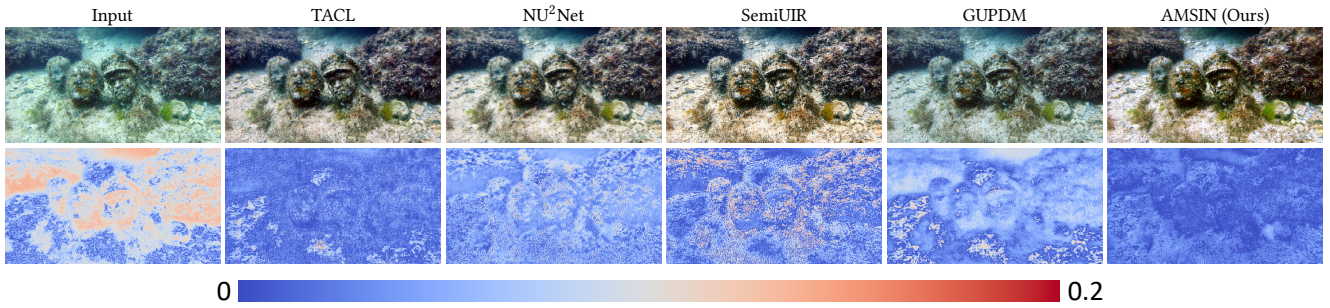
**Figure 1: Visual comparison on LUSI.****Figure 2: Visual results on frame-by-frame underwater video enhancement. Upper row: Input; Bottom row: Enhanced results.****Figure 3: Visual comparison on absolute residuals between enhanced images with GTs.**





Figure 4: Visual comparison of UIEBD-P. Images at even rows correspond to the cropped regions of the images at odd rows.



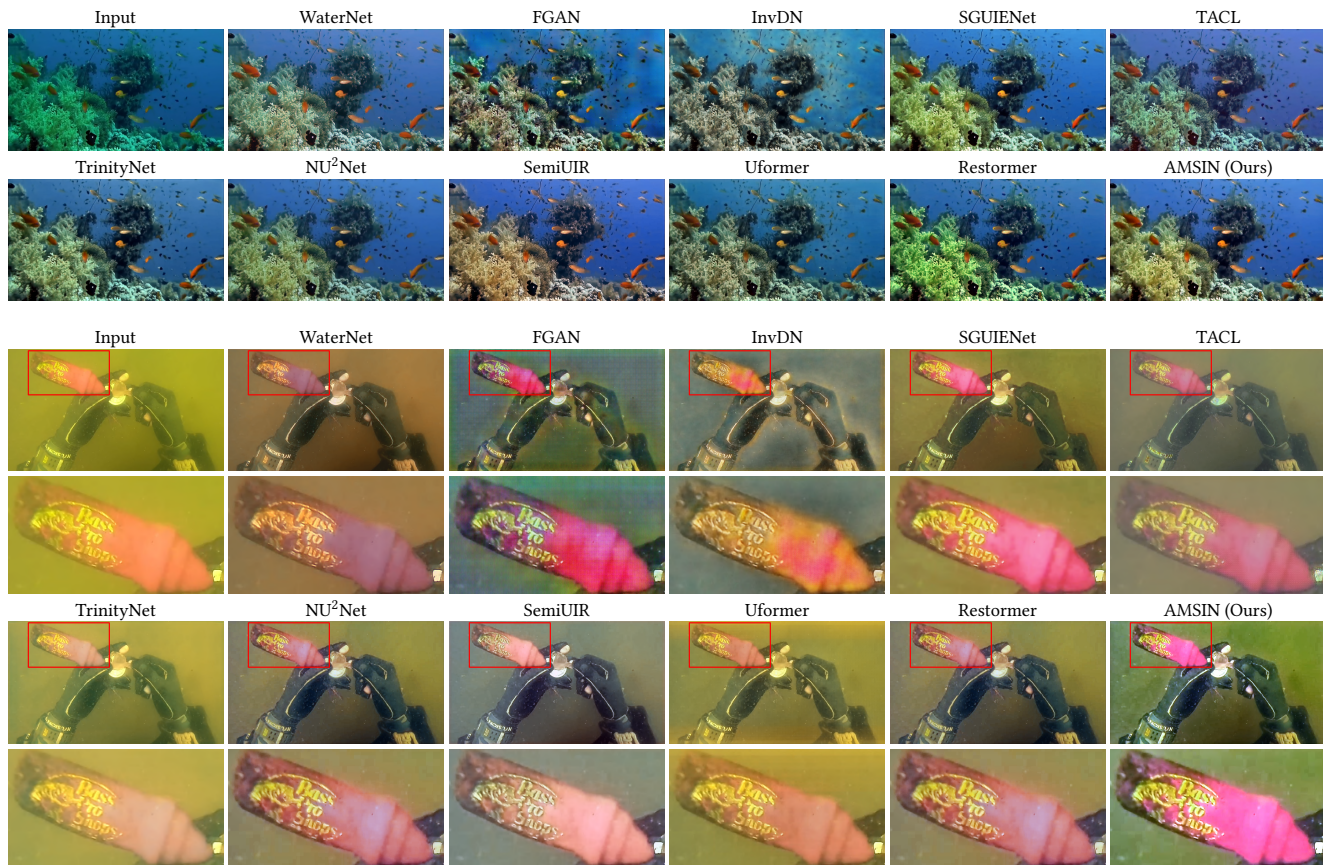


Figure 5: Visual comparison of UIEBD-UP.

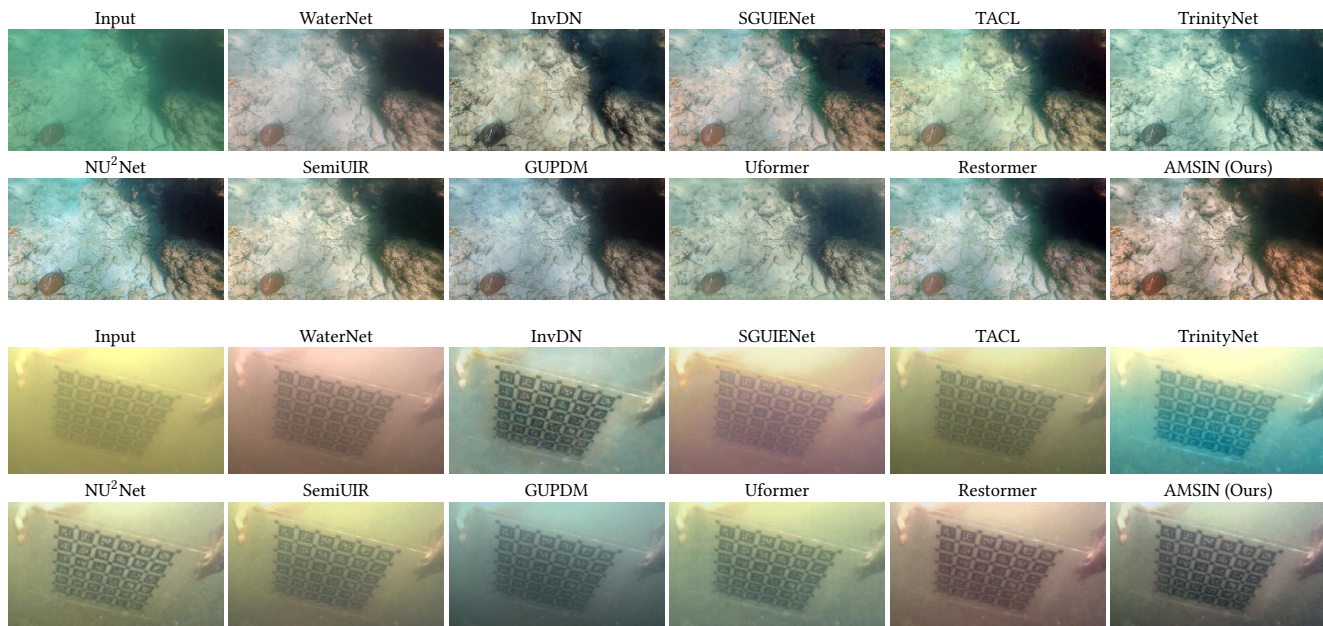


Figure 6: Visual comparison of EUVP.



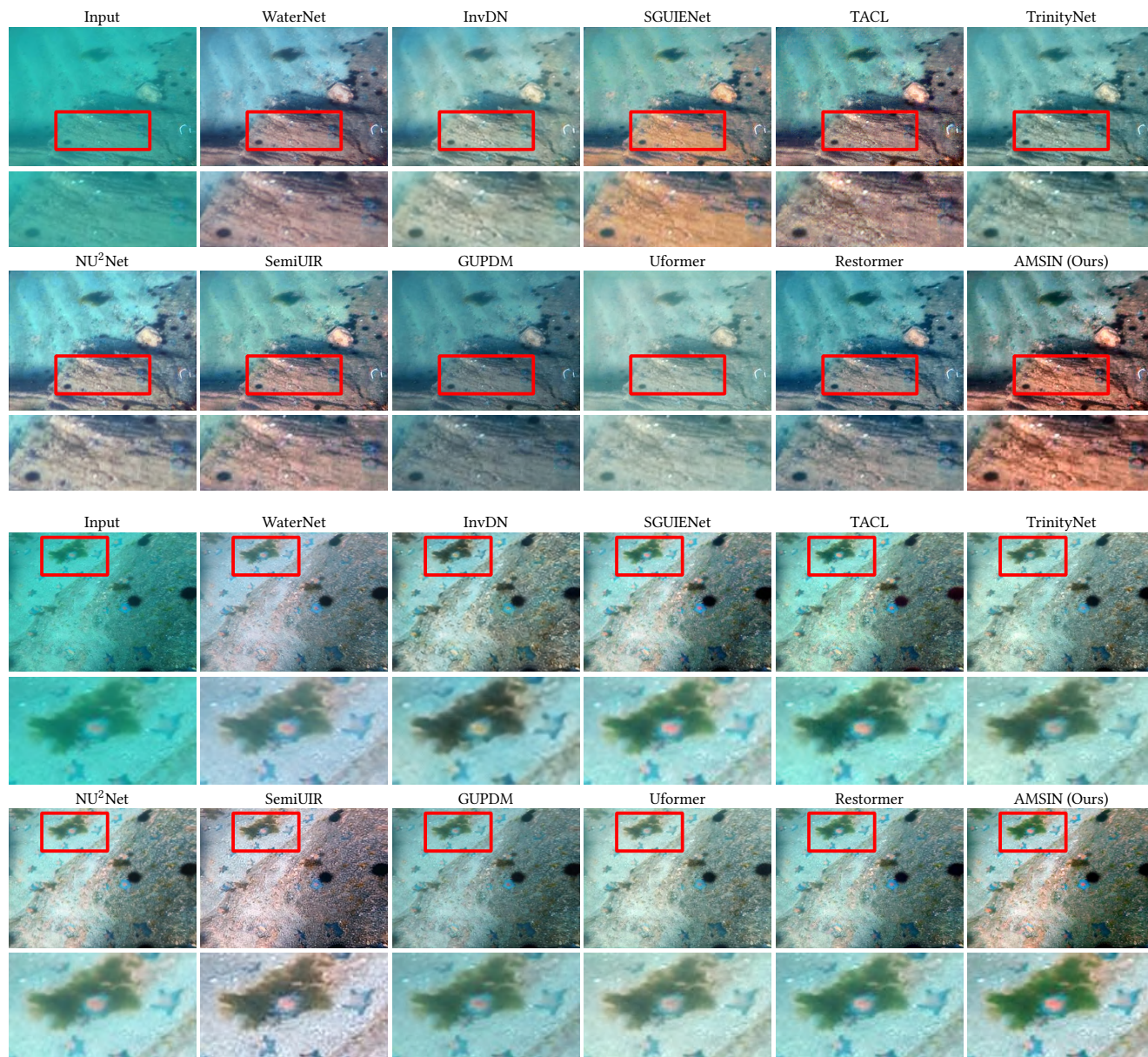


Figure 7: Visual comparison of UCCS. Images at even rows correspond to the cropped regions of the images at odd rows.



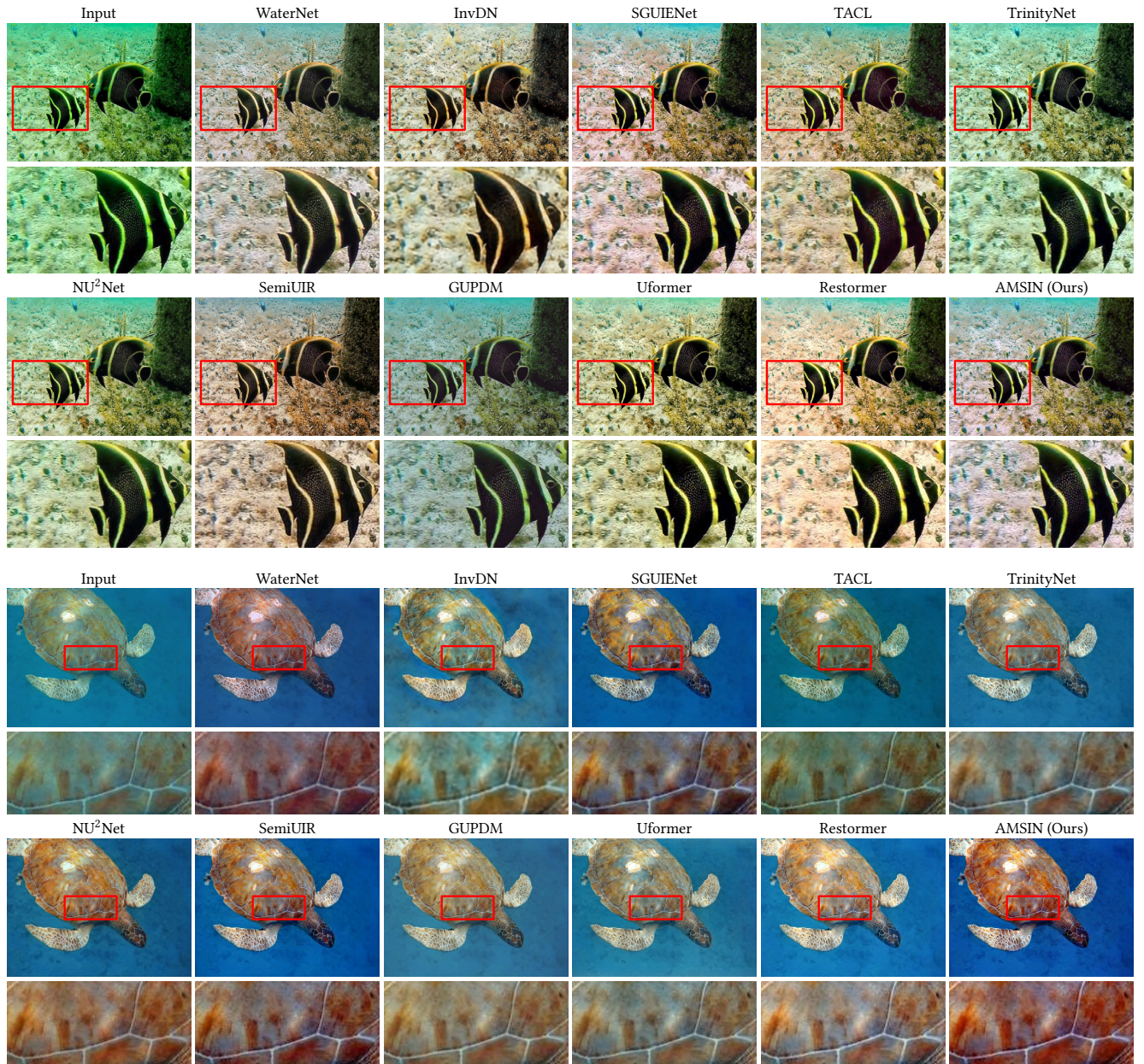


Figure 8: Visual comparison of SUIM. Images at even rows correspond to the cropped regions of the images at odd rows.



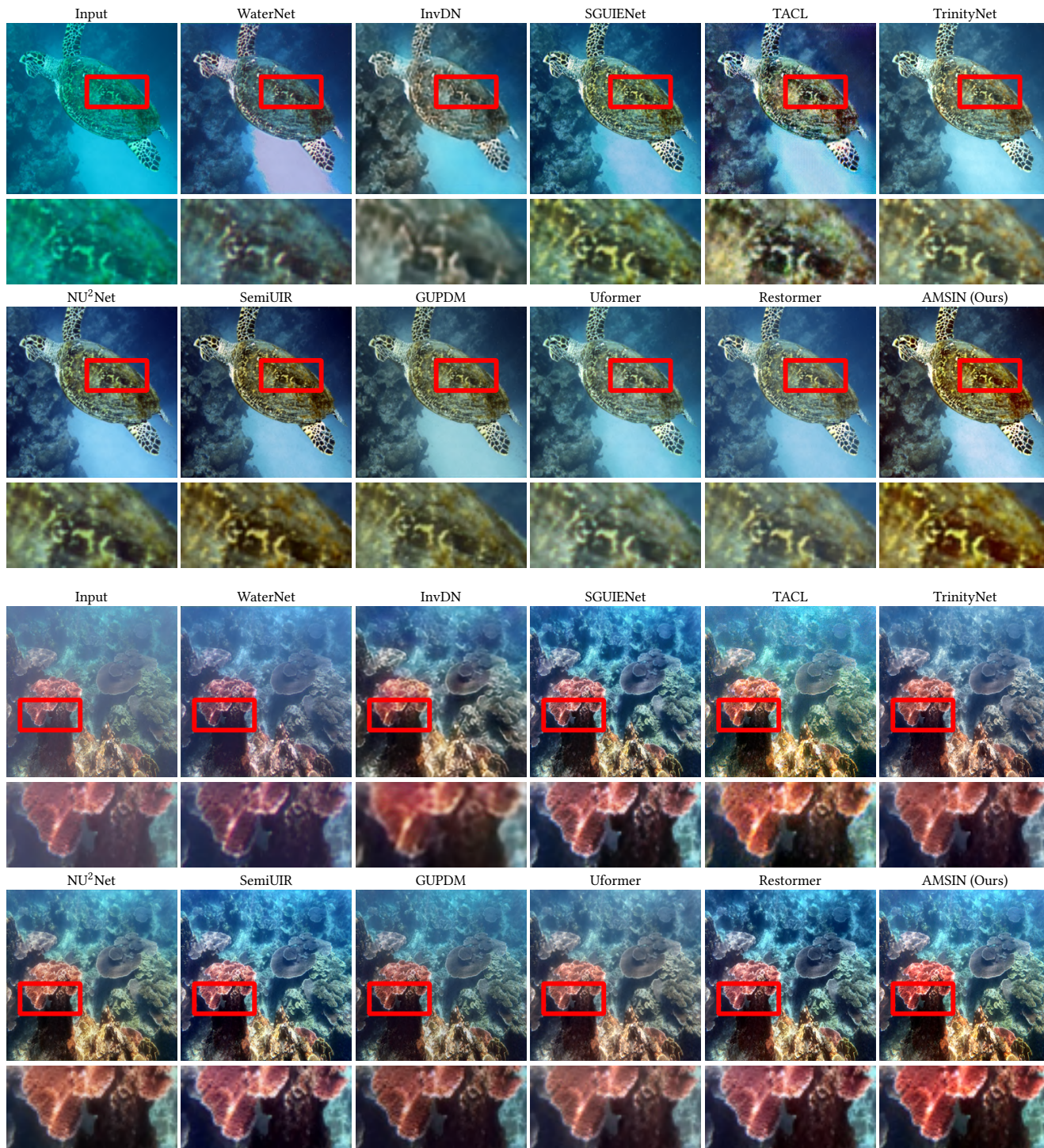


Figure 9: Visual comparison of U45. Images at even rows correspond to the cropped regions of the images at odd rows.