Supplementary for PhISH-Net: Physics Inspired System for High Resolution Underwater Image Enhancement

Aditya Chandrasekar, Manogna Sreenivas, Soma Biswas
Indian Institute of Science, Bangalore
{caditya, manognas, somabiswas}@iisc.ac.in

1. Ablation Studies

Apart from evaluating the importance of each loss component in our proposed approach, we conducted evaluations covering several key aspects. These experiments explored the impacts of photofinishing, variations due to changes in model intrinsics, and the results of degrading both image and depth information. The following section provides a detailed overview of the comprehensive ablation studies performed for each of these factors.

1.1. Photofinishing

The image quality metrics applied to the proposed model both with and without photofinishing, including metrics such as PSNR, PSNR-L, SSIM, PCQI, UIQM, UCIQE, and UIConM are presented in Table 1. In this context, I_{out} denotes the result derived from PhISH-Net. Notably, the metrics show a consistent enhancement in image quality across the majority of categories after the application of photofinishing, evident in both datasets.

For a concrete illustration of this improvement, refer to Figure 1, which showcases an instance of enhanced lighting and a more natural appearance in the output. However, it's important to acknowledge that certain scenarios exhibit a marginal decline in metrics following photofinishing. This can be attributed to the color correction step inherent in the photofinishing process, aimed at mitigating red artifacts stemming from significant red light attenuation underwater. It is worth noting that such correction may not universally apply, and an overcompensating the red channel can induce reddish artifacts post-photofinishing, as seen in Figure 2, leading to a slight drop in the metrics.

In practical terms, since the post-processing step doesn't impose a substantial computational burden, it's advisable to generate both processed images. This approach allows for a choice between selecting the visually more appealing image or applying a blind or reference-based image quality metric to guide the final decision-making process.

Metric		UIEB		EUVP	
		I_{out}	I_{out} + PF	I_{out}	I_{out} + PF
PSNR	(↑)	21.103	**21.139**	**21.313**	20.919
$PSNR_L$	(↑)	**23.545**	23.431	**27.764**	27.472
SSIM	(↑)	0.8362	**0.8686**	0.8500	**0.8559**
PCQI	(↑)	0.9009	**0.9294**	1.0177	**1.0378**
UIQM	(↑)	1.5123	**1.5968**	1.5322	**1.5925**
UCIQE	(↑)	0.6405	**0.6405**	**0.5928**	0.5918
UIConM	(↑)	1.0825	**1.1513**	1.1232	**1.1512**

Table 1. Image quality metrics using PhISH-Net for UIEB and EUVP datasets before and after photofinishing (PF)

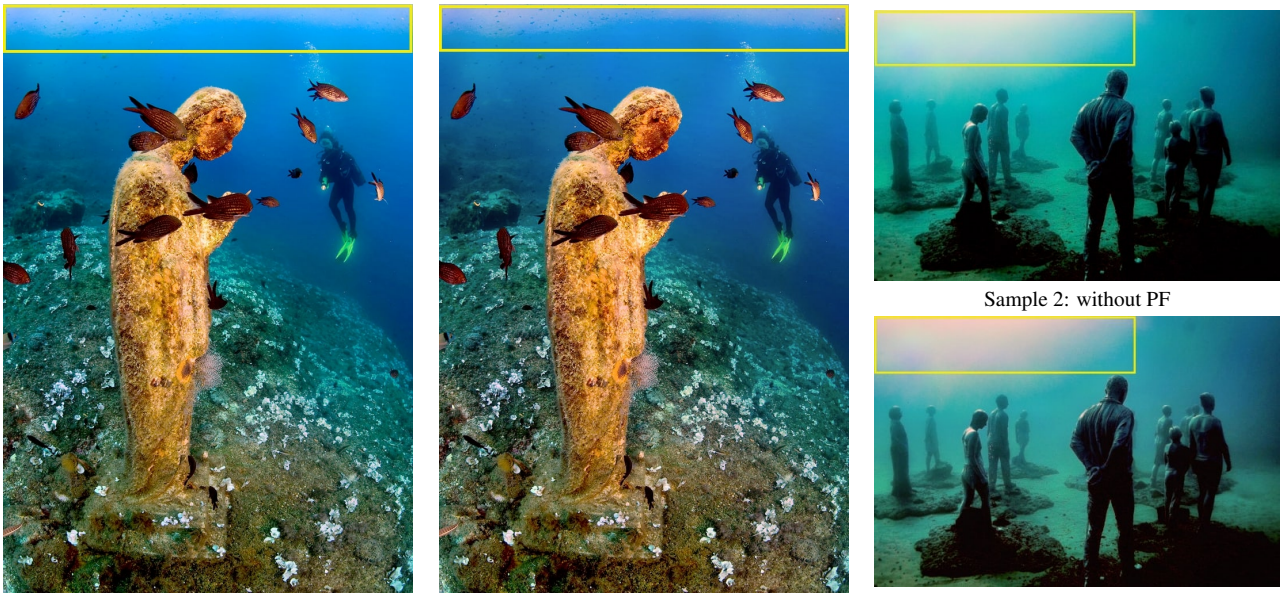
1.2. Model Intrinsics

Considering the typically high-resolution nature of underwater images, PhISH-Net performs most network computations at a reduced resolution (D_{lr}). This approach ensures efficient real-time processing for input high-resolution images (D_{hr}). As elucidated in Section 3.3 of the main paper, this methodology not only enhances computational efficiency but also empowers the model to accommodate images of varying dimensions. To delve deeper, our experimentation encompasses variations in both D_{lr} and D_{hr} sizes, aiming to comprehend their influence on diverse image quality metrics and runtime performance. The results, obtained after training all models for 50 epochs without applying photofinishing, are presented in Table 2. Given the square nature of the images, the table includes only the width dimension.

For our experiments pertaining to D_{lr} size, the D_{hr} size remains constant at 512 pixels. Conversely, when investigating the impact of D_{hr} size, we maintain a fixed D_{lr} size of 256 pixels. Our observations indicate that a majority of the reference-based metrics, including PSNR, $PSNR_L$, SSIM and PCQI demonstrate an upward trend as the size increases. In contrast, a notable portion of the non-reference-based metrics showcase either a slight decline or an oscillatory pattern. Concurrently, the average runtime across

Raw Underwater Image | Output from PhISH-Net | Output from PhISH-Net + PF

Figure 1. Enhanced Image Clarity through photofinishing (PF) on a sample from the UIEB Dataset.



Sample 1: without PF | Sample 1: with PF | Sample 2: without PF | Sample 2: with PF

Figure 2. Excessive color correction during photofinishing (PF) leads to diminished image quality in some cases. This is evident in increased redness within illuminated regions, as highlighted in yellow on samples from the UIEB Dataset.

Metric		D_{lr} size					D_{hr} size		
		64	128	256	384	512	512	768	1024
PSNR	(↑)	19.2187	19.3523	20.0477	20.3986	20.6985	20.0477	20.9997	20.9443
$PSNR_L$	(↑)	21.9342	21.8004	22.7988	22.9073	23.3791	22.7988	23.6187	23.4659
SSIM	(↑)	0.81042	0.81168	0.82146	0.82409	0.82923	0.82146	0.83528	0.83609
UIQM	(↑)	1.52350	1.50390	1.50210	1.51080	1.49090	1.50210	1.49870	1.49470
UICM	(↑)	9.61920	9.17650	9.68150	9.17860	9.32350	9.68150	9.02290	8.80380
UIConM	(↑)	1.08170	1.07820	1.07730	1.07440	1.07230	1.07730	1.07810	1.07660
CCF	(↑)	37.2623	36.5588	37.3866	37.3605	37.0942	37.3866	37.1043	36.7817
PCQI	(↑)	0.88290	0.86040	0.88340	0.88700	0.87650	0.88340	0.89020	0.88170
UCIQE	(↑)	0.64750	0.64009	0.64949	0.64421	0.64592	0.64949	0.64205	0.63945
Runtime (s)	(↓)	0.00779	0.00795	0.00817	0.00838	0.00843	0.00817	0.00824	0.00844

Table 2. Metric Variability Across D_{lr} and D_{hr} Size Variations

Type	Reference-based					Non Reference-based			
	PSNR (↑)	$PSNR_L$ (↑)	SSIM (↑)	UCIQE (↑)	PCQI (↑)	CCF (↑)	UIQM (↑)	UICM (↑)	UIConM (↑)
Original	**21.2488**	**26.1750**	**0.9281**	**0.6785**	**1.0596**	36.7361	1.5161	8.6105	1.0340
75% Darker Image	19.5019	24.7553	0.8796	0.6270	0.9668	**38.8549**	1.4699	9.3525	1.0549
Noisy Image (S&P)	15.2499	20.6743	0.6252	0.6232	0.3484	31.9978	1.4657	9.3169	0.9918
Noisy Image (Poisson)	18.2395	21.5353	0.4366	0.6469	0.5088	26.3174	**1.7046**	7.1597	**1.2764**
Noisy Image (Gaussian)	15.0410	18.8314	0.2128	0.6683	0.2970	31.4644	1.6153	**11.485**	1.0751
Weaker Depth Model	18.7892	22.0227	0.9173	0.6766	0.9955	29.1331	1.4037	9.1041	0.9652
Manually Altered Depth	18.0412	21.1915	0.9112	0.6724	0.9648	26.6382	1.3363	9.0291	0.9374

Table 3. Effects of Image and Depth Degradation on various Quality Metrics.

the dataset displays a proportional increase as size is augmented.

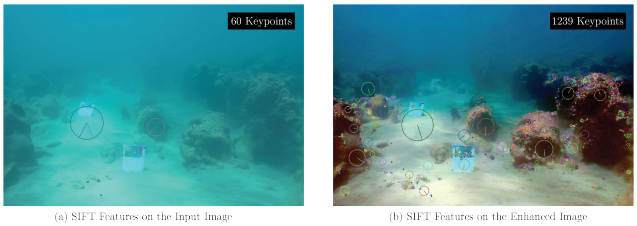
In alignment with Section 4.1 of the main paper, we conducted these experiments using an NVIDIA RTX A5000 GPU with a batch size of 64. However, it's important to note that for D_{hr} sizes exceeding 512 pixels, memory overflow errors necessitated a reduction in the batch size.

1.3. Image and Depth Degradation

In this section, we delve into an analysis of the model's performance by scrutinizing its metrics in response to image and depth degradation. For image degradation, we manually generate degraded images through techniques such as darkening the raw image (Fig. 4b) and introducing diverse forms of noise (Figs. 4c to 4e), while simultaneously obtaining depth information through depth boosting [3]. Shifting the focus to depth-based degradation, we employ the original raw image but introduce a weaker depth model (Fig. 4f) or manually manipulate the produced depth map after depth boosting (Fig. 4g), aiming to study their respective impacts.

The outcomes are visually depicted in Figure 4, and a comprehensive overview of the metrics is presented in Table 3. Our analysis can be bifurcated into discussions around reference-based metrics and non-reference-based metrics.

For reference-based metrics, we observe optimal performance when the original image and depth combination is maintained. Intriguingly, the model showcases considerable resilience to alterations in depth maps, performing commendably even with modified depth information. However, when confronted with image-based modifications, we note exceptionally high scores in non-reference-based metrics, surpassing those of the original image-depth pairing. Yet, the SSIM and PCQI values raise concerns, as evident in Figure 4, where residual noise remains present, leading to improvements in parameters like contrast that boost the non-reference metrics. This underlines the necessity of evaluating performance from both reference and non-reference metrics perspectives, as they offer nuanced insights into various facets of the model's behavior. Notably, it's worth noting that instances affected by noise can be improved using a simple filter-based pre-processing step, which can further refine the model's output quality.



(a) SIFT Features on the Input Image (b) SIFT Features on the Enhanced Image

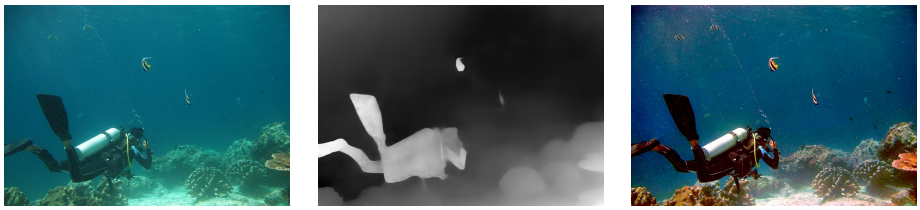
Figure 3. SIFT features for the input and the enhanced image.

1.4. SIFT-based Qualitative Analysis

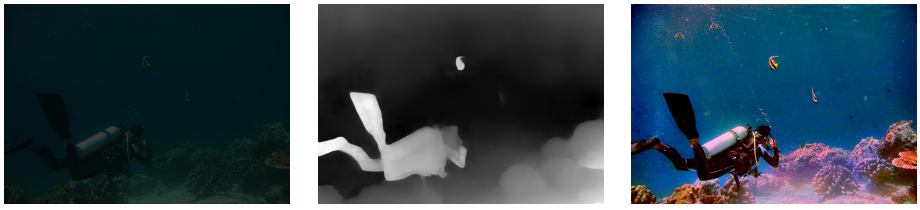
In Section 4.2 of the main paper, we demonstrated our framework's efficacy on two datasets and assessed it generalization capability via a cross-dataset scenario. Here, we compute SIFT features before and after enhancement to understand the impact on downstream tasks. The input image yielded 60 keypoints, whereas the enhanced image detected 1239 keypoints (Fig 3). This increase in SIFT keypoints holds promise for improved feature extraction, enabling more robust and accurate analyses in various downstream tasks.

2. Depth Estimation

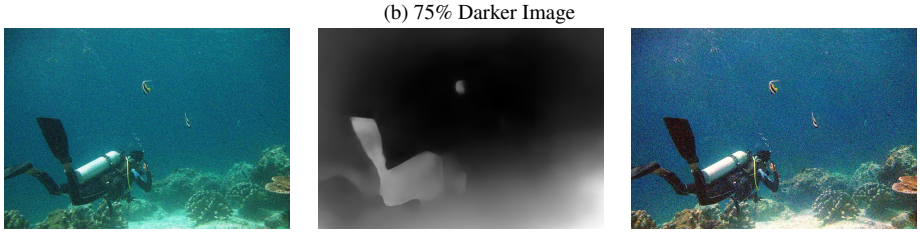
As outlined in Section 3.2 of the main paper, we adopt a boosted version of monocular depth estimation proposed by [3], building upon the foundation of MiDaS [2]. This approach integrates two key techniques: double estimation and patch selection. Through iterative refinement, the double estimation process refines the initial depth estimation, while patch selection selectively incorporates local details. By fusing depth estimates at multiple resolutions, this method produces high-quality depth maps without re-



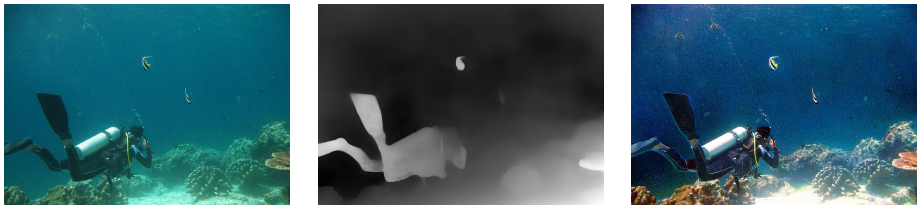
(a) Original



(b) 75% Darker Image

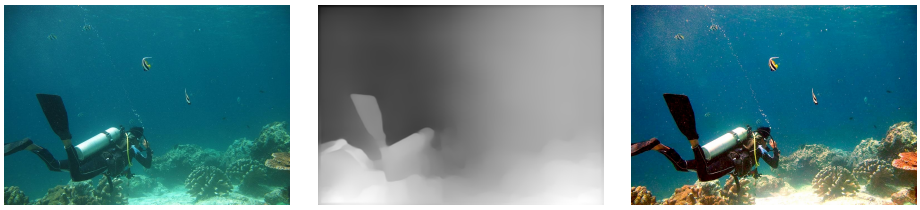


(c) Noisy Image (Gaussian)

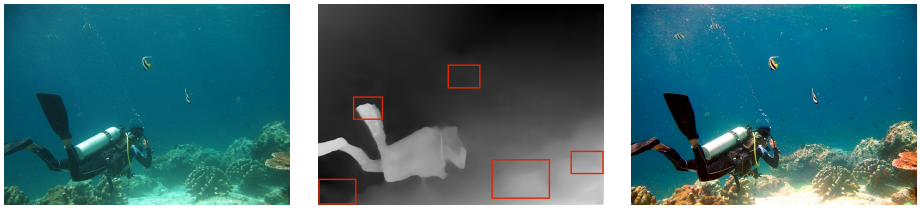


(d) Noisy Image (Poisson)

(e) Noisy Image (Salt & Pepper)

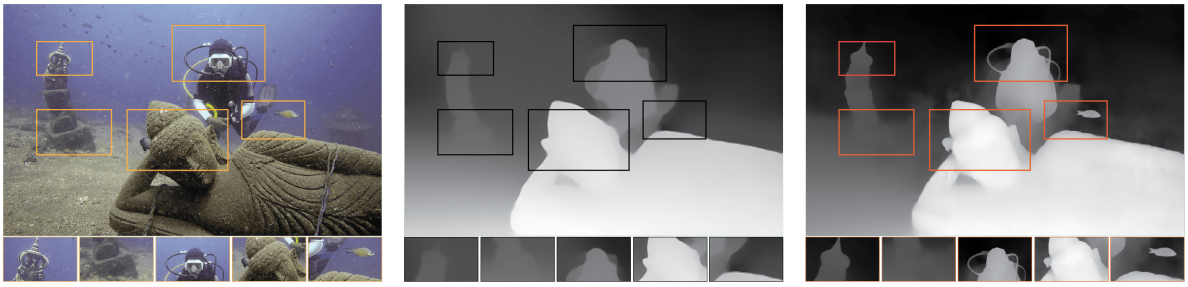


(f) Weaker Depth Model



(g) Manually Altered Depth

Figure 4. Sample from the UIEB dataset depicting diverse image and depth degradation scenarios and it's impact on visual quality.



(a) Sample Image from the UIEB Dataset (b) Depth Estimate from the base MiDaS model (c) Depth Estimate after Boosting

Figure 5. Impact of depthmap quality via Depth Boosting.

	Module	Type	Kernel Size	Stride	Channels	Padding	Activation
Encoder	EN^1	*c*	3	2	8	1	*relu*
	EN^2	*c*	3	2	16	1	*relu*
	EN^3	*c*	3	2	32	1	*relu*
	EN^4	*c*	3	2	64	1	*relu*
Local & Global Features	LF^1	*c*	3	1	64	1	*relu*
	LF^2	*c*	3	1	64	1	*none*
	GF^1	*c*	3	2	64	1	*relu*
	GF^2	*c*	3	2	64	1	*relu*
	GF^3	*fc*	-	-	256	-	*relu*
	GF^4	*fc*	-	-	128	-	*relu*
	GF^5	*fc*	-	-	64	-	*none*
	FS	*c*	1	1	96	0	*none*
	CP	*fc*	-	-	4	-	*relu*
Guide	GN^1	*c*	1	1	16	0	*relu*
	GN^2	*c*	1	1	1	0	*sigmoid*

Table 4. Model Architecture (convention as per Section 3)

training the base model.

By incorporating this boosted depth estimation technique, our study achieves improved detail and accuracy in depth maps, contributing to more dependable and resilient underwater image enhancement, as visualized in Figure 5. We further observe from Figure 4 that the depth estimation technique demonstrates effective performance even for dark or low-lit images (Fig. 4b) and yields satisfactory results for noisy images (Figs. 4c and 4d) without any pre-processing.

3. Model Architecture

In this section, we delve into the architectural specifics of PhISH-Net, illustrated in Figure 2 in the main paper. The overall flow unfolds as follows: the low-resolution D_{lr} is fed into the Encoder (EN), and its output further flows through the Local Feature (LF) and Global Feature (GF) extractors. These outputs are then fused (FS) to derive the bilateral grid coefficients. Simultaneously, the fused features are directed to the Coefficient Predictor (CP), which computes coefficients a, b, c, and d for the wideband attenuation prior, as outlined in Section 3.3 of the main paper.

The high-resolution D_{hr} is channeled into the Guide Network (GN) to acquire a guidance map. This map then interacts with the previously determined coefficients, facilitating efficient output upsampling [1] to obtain the high resolution illumination map $\mathcal{S}_{hr}$. Detailed specifications of individual units, including layer type, kernel size, stride, channels, padding, and activation function, are meticulously documented in Table 4. In this context, *c* signifies a convolution layer, while *fc* denotes a fully connected layer.

As noted in Section 4 of the main paper, our training procedure involves maintaining fixed dimensions for D_{lr} and D_{hr} at 256x256 and 512x512, respectively. This design choice (specifically D_{lr} renders the entire network adaptable to images of varying sizes. During the testing phase, we retain the original size of D_{hr} without any resizing.

This ensures that the output maintains the same dimensions as the input, effectively catering to high-resolution inputs while simultaneously reducing computational costs associated with low-resolution processing.

References

[1] Michaël Gharbi, Jiawen Chen, Jonathan T Barron, Samuel W Hasinoff, and Frédo Durand. Deep bilateral learning for real-time image enhancement. *ACM Transactions on Graphics*, 36(4):118, 2017. 5

[2] Katrin Lasinger, René Ranftl, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *CoRR*, abs/1907.01341, 2019. 3

[3] S. Mahdi H. Miangoleh, Sebastian Dille, Long Mai, Sylvain Paris, and Yağız Aksoy. Boosting monocular depth estimation models to high-resolution via content-adaptive multi-resolution merging. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9680–9689, 2021. 3