WaterFlow: Heuristic Normalizing Flow for Underwater Image Enhancement and Beyond

Zengxi Zhang Dalian University of Technology cyouzoukyuu@gmail.com Zhiying Jiang Dalian University of Technology zyjiang0630@gmail.com

Jinyuan Liu Dalian University of Technology atlantis918@hotmail.com

Xin Fan Dalian University of Technology xin.fan@dlut.edu.cn Risheng Liu* Dalian University of Technology Peng Cheng Laboratory rsliu@dlut.edu.cn

ABSTRACT

Underwater images suffer from light refraction and absorption, which impairs visibility and interferes the subsequent applications. Existing underwater image enhancement methods mainly focus on image quality improvement, ignoring the effect on practice. To balance the visual quality and application, we propose a heuristic normalizing flow for detection-driven underwater image enhancement, dubbed WaterFlow. Specifically, we first develop an invertible mapping to achieve the translation between the degraded image and its clear counterpart. Considering the differentiability and interpretability, we incorporate the heuristic prior into the data-driven mapping procedure, where the ambient light and medium transmission coefficient benefit credible generation. Furthermore, we introduce a detection perception module to transmit the implicit semantic guidance into the enhancement procedure, where the enhanced images hold more detection-favorable features and are able to promote the detection performance. Extensive experiments prove the superiority of our WaterFlow, against state-of-the-art methods quantitatively and qualitatively.

CCS CONCEPTS

• Computing methodologies \rightarrow Computer vision.

KEYWORDS

underwater image enhancement, normalizing flow, underwater object detection

ACM Reference Format:

Zengxi Zhang, Zhiying Jiang, Jinyuan Liu, Xin Fan, and Risheng Liu. 2023. WaterFlow: Heuristic Normalizing Flow for Underwater Image Enhancement and Beyond. In *Proceedings of the 31st ACM International Conference on Multimedia (MM '23), October 29–November 3, 2023, Ottawa, ON, Canada.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3581783.3611727

MM '23, October 29-November 3, 2023, Ottawa, ON, Canada.



Figure 1: The invertible translation process between underwater and enhanced images. Heuristic priors are embedded to enhance the interpretability of the enhanced network. Highlevel perceptual information is introduced into the model by backpropagation to help the enhanced image potentially contain more semantic information conducive to object detection. We illustrate the k-means clustering results along with the procedure of the invertible translation.

1 INTRODUCTION

In recent years, there have been significant advances in underwater robots for exploration in various fields [31, 58]. However, underwater object detection, a critical component of underwater exploration tasks, still faces considerable difficulties. The images captured underwater often suffer from severe distortion due to the complex and variable underwater environment, leading to a significant reduction in image visibility. The performance of subsequent object detection applications is also significantly affected.

To mitigate the impact of image distortion on underwater object detection, existing methods typically employ underwater image enhancement as a preprocessing step for object detection [5]. They feed the enhanced underwater images as prospective guidance into the detection module to achieve better detection accuracy. Traditional underwater enhancement methods estimate imaging parameters through the underwater degradation formula to obtain enhanced images. However, fixed imaging parameters may not fully simulate the diverse and complex nature of real underwater environments. In recent years, deep learning has been widely concerned by researchers. They [19, 25] have achieved good restoration effects through the manually crafted network by end-to-end training. However, image enhancement and object detection are usually regarded as two parallel independent tasks. Enhancement

^{*}Corresponding author: Risheng Liu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

^{© 2023} Copyright held by the owner/author(s). Publication rights licensed to ACM. ACM ISBN 979-8-4007-0108-5/23/10...\$15.00 https://doi.org/10.1145/3581783.3611727

results that solely focus on visual perception are insufficient to fully capture the scene information required by subsequent detection algorithms. Moreover, the enhancement network may introduce uncertain interferences that affect the detection performance.

In this paper, we propose a detection-driven heuristic normalizing flow for underwater image enhancement. we first develop a reversible translation framework based on normalizing flow to facilitate domain translation between degraded underwater images and clear counterpart. Specifically, the reversible translation between them is established by synchronous optimization with shared parameters and bilateral constraints of forward and reverse processes. The forward process aims to map the degraded image to the clear restored image by learning the nonlinear function $F_n(.)$, and the reverse process aims to map the clear image back to the degraded image through the nonlinear function $F_n^{-1}(.)$. Then, considering the differentiability and interpretability, we estimate the underwater imaging parameters using different modalities of knowledge and incorporate them as heuristic priors into the data-driven mapping process. Guided by the underwater imaging physical model, the proposed method effectively prevents the introduction of the noise interference and undesirable artifacts, thereby avoiding adverse impacts on subsequent detection tasks. In order to improve the adaptability of the enhanced results for subsequent detection tasks, we further introduce a Detection Perception Module. By propagating the high-level perceptual features to the enhancement module, the generated enhanced images can implicitly possess more semantic information beneficial for subsequent detection tasks while achieving visually pleasing enhancement effects.

In summary, contributions can be concluded as follows:

- We apply the normalizing flow to the underwater image enhancement task, which realizes the invertible mapping between the degradation image and its clear counterpart.
- We incorporate the heuristic prior into the data-driven mapping process, which can be widely applied in a variety of real underwater scenes by improving the interpretability of the whole enhancement framework.
- We propose a Detection Perception Module, which transmits high-level latent perceptual information to retain and extract detection-oriented semantic information.
- Qualitative and quantitative results demonstrate that the proposed WaterFlow recovers the intrinsic scene clearly and is more conducive to the subsequent detection.

2 RELATED WORK

2.1 Underwater Image Enhancement

In recent years, numerous underwater image enhancement methods have been proposed. The early traditional methods often adjust the pixel distribution of different color channels to weaken the degradation of natural light underwater. Ghani *et al.* [13] extended the histogram of the red channel and the blue channel of the underwater image upward and downward respectively according to the law of Rayleigh distribution. Li *et al.* [28] corrected the red channel according to the Gray-World assumption theory after the color restoration of the blue-green image through the dark channel. Although these methods can weaken the fading of light on different color channels, it is easier to introduce underwater images into artificial colors due to the lack of guidance from physical models. Therefore, model-based methods [27, 45, 52, 57] are also widely used to improve the interpretability of networks by incorporating domain-specific prior knowledge.

In recent years, many methods based on deep learning have been proposed: Li *et al.* [25] introduced the multi-color space into the transmission-guided network to reduce the influence of color casts on underwater images. Mu *et al.* [43] introduced a bi-level model that hierarchically incorporates various knowledge modalities to enhance the quality of underwater images. Jiang *et al.* [19] proposed a perceptual adversarial mechanism and introduced the global module to narrow the gap between enhanced images and reference images. Huang *et al.* [17] proposed a semi-supervised framework based on the mean-teacher approach to enhance the generalization ability of the enhance model on real-world data.

2.2 Normalizing Flow

Normalizing flow facilitates the transformation between complicated probability distribution and Gaussian distribution through bijection functions and differentiable mappings. The normalizing flow has received far less attention than GANs and VAEs because it requires a significant quantity of video memory with a fine structural design to handle a large amount of reversible computing. Many methods for accelerating the calculation have been proposed in recent years. Dinh et al. [10] proposed the additive coupling layer, which makes the flow easier to calculate the determinant of Jacobian matrix. In order to increase the log-likelihood, Kingma et al. [24] developed the actnorm to normalize each channel of input features and 1×1 convolution in place of the permutation [10]. Conditional normalizing flow was introduced for increasing the expressiveness and flexibility of normalizing flows, and has been widely used in vision tasks [1, 2, 21, 30, 41, 53]. However, there is little precedent dedicated to adopting flow-based methods to solve the problem of ill-posed underwater image enhancement tasks.

2.3 Object Detection

In recent years, due to the rapid development of deep learning, the effect of object detection has also been significantly improved. The existing methods are generally divided into one-stage method and two-stage method. The two-stage method [7, 14, 16, 48, 51] first needs to extract the region of interest proposals area, and then classify it through the classification of neural network. By contrast, the classification label and boundary information are often directly obtained by the one-stage method [12, 23, 47, 55].

The current detectors can achieve good performance when facing in-air images. However, the problems of scene blur and light imbalance faced by underwater images will greatly affect the effectiveness of the detectors. Existing studies [33, 38–40, 42] have used low-level visual enhancement modules as preprocessing steps for detection tasks. However, the existing enhanced networks for the underwater image often do not play a significant role in the detector, which ignores the position and semantic information while restoring the contrast of the underwater image. WaterFlow: Heuristic Normalizing Flow for Underwater Image Enhancement and Beyond

MM '23, October 29-November 3, 2023, Ottawa, ON, Canada.



Figure 2: Workflow of the proposed WaterFlow. The underwater imaging physical model is embedded into the normalizing flow as a heuristic prior to better simulate the mapping of the underwater image and its clear counterpart. At the same time, the detection module is combined to improve the detection effect of underwater object detection by transmitting the high-level latent feature to the enhancement module.

3 THE PROPOSED METHOD

Existing deep learning based underwater image enhancement methods rely on the enriched training data for a more robust and effective performance. However, the ignorance of physical priors in the learning process introduces the noise interference and undesirable artifacts, degrading the visibility further and undermining the subsequent detection performance. To overcome the above limitations, we propose a detection-driven heuristic normalizing flow network for underwater image enhancement. Specifically, to reduce the dependence on training data, we develop a normalizing flow based reversible translation framework to achieve domain translation between degraded underwater images and clear counterpart. The heuristic model constraints incorporated with gradient and depth information are embedded into the reversible procedure, guiding the reversible results to be more accurate and reliable. Moreover, to improve the applicability of the enhanced results for subsequent object detection task, we propose a Detection Perception Module that feeds back the performance of the enhanced results in detection tasks, allowing the enhancement module to learn more detection-favorable implicit features. Next, we provide the detailed description of each module.

3.1 Hybrid Invertible Block

Hybrid Invertible Block (HIB) was introduced as the main part of the heuristic normalizing flow. It ensembles the heuristic prior into the data-driven network to build the invertible mapping relationship between the underwater image and its clear counterpart. As illustrated in Fig. 2, the underwater image restoration and the underwater image degradation process are combined in a invertible mapping manner during the training process. In the forward process, the underwater image is squeezed and input into multiple invertible blocks to generate enhanced images. In the reverse process, clear images are inversely input to invertible blocks with shared parameters to generate degraded images.

The Hybrid Invertible Block consists of the following parts: Actnorm [24], Conv 1 \times 1 [24], Heuristic Prior Injector, Conditional Affine Coupling [30] and squeeze operation. Actnorm [24] is a normalizing operation, which can change the input tensor into the zero mean and unit variance tensors. Conv 1×1 [24] can convolve the input tensor of $c \times h \times w$ by the weight matrix of $c \times c$ to make the calculation of Jacobian matrix and network inversion easier.

Heuristic Prior Injector was designed to better characterize the underwater imaging model by embedding a heuristic prior, which will be introduced in detail in section 3.2.

Conditional Affine Coupling was proposed by [30], which realizes the mutual conversion with conditional information between output and input through reversible transformation. We modified the previous formulation for adjusting the underwater image enhancement framework, which is described as:

$$\begin{aligned} \mathbf{u}_{i+1}^{1} &= \mathbf{u}_{i}^{1}, \\ \mathbf{u}_{i+1}^{2} &= \left(\phi_{i}\left(C(\mathbf{u}_{i}^{1}, f_{i}^{A}\left(I_{u}, I_{g}, I_{d}\right)\right)\right) \odot \mathbf{u}_{i}^{2} \\ &+ \rho_{i}\left(C(\mathbf{u}_{i}^{1}, f_{i}^{B}\left(I_{u}, I_{g}, I_{d}\right))\right), \end{aligned}$$
(1)

where $\mathbf{u}_{i+1} = C(\mathbf{u}_{i+1}^1, \mathbf{u}_{i+1}^2)$, *C* represents the concatenation operation. ϕ_i , ρ_i , f_i^A , and f_i^B denote convolutional networks. $i \in \{1, ..., n-1\}$ represents the *i*-th flow steps, I_u , I_g and I_d will be explained in detail in section 3.2. Squeeze operation can convert the image of $c \times h \times w$ to $4c \times \frac{h}{2} \times \frac{w}{2}$. Unsqueeze is the reverse operation of squeeze. By the combination of these operations, the proposed architecture can attain accurate mapping between the underwater image and it clear counterpart while maintaining reversibility.

3.2 Heuristic Prior Injector

The Heuristic Prior guided Injector (HPI) is designed to incorporate the estimated physical imaging parameters into the invertible block. According to the underwater image formation model [8, 11], the enhanced image can be represented as:

$$J^{c}(x) = \frac{1}{t(x)}I^{c}(x) + \frac{1}{t(x)}B^{c}(t(x) - 1), c \in \{r, g, b\},$$
(2)

where *I* and *J* denote the underwater image captured by sensors and the enhanced image at pixel point *x* respectively. $c \in \{r, g, b\}$ represents the corresponding color channels. *B* and *t* indicates the

Zengxi Zhang, Zhiying Jiang, Jinyuan Liu, Xin Fan, and Risheng Liu

ambient light and the medium transmission coefficient. *t* can also be described as $t(x) = e^{\beta d(x)}$ with the scene depth d(x) and the attenuation coefficient β according to the Beer-Lambert law [4].

Considering that the gradient map provides information about the edges and contrast in the image, which can be used to estimate the scatteration of light by the atmosphere and fine impurities in the water. Furthermore, the depth map provides information about the attenuation of the light propagation, which is also a key factor in the Beer-Lambert law. Inspired by [45], we first estimate the depth map I_d and the gradient map I_g as auxiliary information from the underwater image.

Then, we concatenate them with underwater image as input to the Heuristic Prior guided Encoder (HPE) \mathcal{H} to progressively estimate the imaging parameters by extracting the depth, gradient and color information of the underwater image. The equation can be illustrated as:

$$B_i, T_i = \mathcal{S}\left(\mathcal{H}_i\left(C\left(I_d, I_q, I_u\right)\right)\right),\tag{3}$$

where I_u represents underwater images. *C* and *S* denote the concatenate and split operation respectively. *T* is the reciprocal of *t*. The detailed architecture of \mathcal{H} is in the appendix due to limited space. After estimating *T* and *B*, we insert them into the HIB as heuristic information. The operation can be formulated as:

$$\mathbf{u}_{i+1} = T_i \odot \mathbf{u}_i + B_i \odot (1 - T_i),$$

$$\mathbf{u}_i = (\mathbf{u}_{i+1} - B_i \odot (1 - T_i)) / T_i,$$
(4)

where \mathbf{u}_i represents the middle feature of the *i*-th flow steps, \odot denotes dot product. By introducing the underwater image formation model based heuristic prior to the invertible blocks, the proposed method can generate enhanced images, which are more suitable for realistic underwater scenes.

3.3 Detection Perception Module

Object detection aims to acquire the location and category of each object. Hence, the performance of object detection will be improved if the enhanced image holds a more profound level of semantic feature and object localization information. Inspired by the perceptual loss [22], deep neural networks designed for high-level visual tasks can retain semantic information and own the ability to describe potential features. Therefore, in order to make the enhanced images potentially improve the effectiveness of the underwater object detection, we propose a Detection Perception Module (DPM) to improve the performance of the enhancement module in preserving and extracting detection-oriented perceptual features.

In the whole training process, we first use the enhancement module as the data preprocessing stage of underwater object detection. After that, the enhanced image was directly fed into the DPM and conducted separate detection training. Then, we jointly optimize the pre-trained DPM and enhancement module through different benchmarks. Specifically, the underwater image is first input to the enhancement module to obtain enhanced data. We input it into the detection module to transmit the high-level implicit perceptual features extracted from the network to the enhancement module, so as to guide the visual improvement effect to be more conducive to subsequent object detection tasks.

3.4 Loss Function

Contrastive learning has been applied to multiple low-level visual tasks [20, 34, 35, 37]. We introduce contrastive learning to improve the quality of enhanced images by making them closer to the inair images and more isolated from the underwater images. We have designated the reference image as a positive sample and the original underwater image as a negative sample. VGG19 [50] was used to extract the implicit characteristic of enhanced images. The contrastive loss is expressed as follows:

$$\mathcal{L}_{c} = \sum_{i=1}^{N} \rho_{i} \cdot \frac{\|\mathcal{V}_{i}(I_{r}) - \mathcal{V}_{i}(G_{E}(I_{u}))\|_{1}}{\|\mathcal{V}_{i}(I_{u}) - \mathcal{V}_{i}(G_{E}(I_{u}))\|_{1}},$$
(5)

where V_i represents the *i*-th layer of VGG19 [50], and ρ_i denotes the *i*-th weight of each layer. G_E represents the proposed enhancement network. I_u and I_r represent the underwater image and the reference image respectively.

We additionally incorporate the style loss [9] to make the generated output closer to the style pattern of the reference image. The definition of style loss function \mathcal{L}_s is described as follows:

$$\mathcal{L}_{s} = \frac{1}{N} \sum_{i=0}^{N} \|\mu \left(\mathcal{V}_{i} \left(G_{E}(I_{u}) \right) \right) - \mu \left(\mathcal{V}_{i} \left(I_{r} \right) \right) \|_{2} + \|\nu \left(\mathcal{V}_{i} \left(G_{E}(I_{u}) \right) \right) - \nu \left(\mathcal{V}_{i} \left(I_{r} \right) \right) \|_{2},$$
(6)

where μ and ν represent the mean and variance of the image. *N* is the number of layers.

Localization loss \mathcal{L}_{loc} and classification loss \mathcal{L}_{cla} are used as detection-driven loss \mathcal{L}_{det} , which is illustrated as:

$$\mathcal{L}_{det} = \mathcal{L}_{cla} + \mathcal{L}_{loc} , \qquad (7)$$

where classification loss is designed to minimize the discrepancy between the prediction category and the ground truth category. The localization loss is adopted to reduce the position difference between the prediction box and the ground truth box. We employ Focal loss [32] and GIoU loss [49] as classification loss and localization loss respectively.

Ideally, the incorporation of the reverse mapping process could impose a regular constraint on the degraded image, thereby enhancing the performance of the forward mapping [15]. Therefore, L1 loss is used as a bilateral constraint to make the output of the forward and the reverse process closer to the reference and underwater image respectively. The formula can be described as:

$$\mathcal{L}_1 = \|G_E(I_u) - I_r\|_1 + \|G_E^{-1}(I_r) - I_u\|_1,$$
(8)

where G_E^{-1} denotes the reverse transformation of G_E . Therefore, the total loss of network training is expressed as follows:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_c + \lambda_2 \mathcal{L}_s + \lambda_3 \mathcal{L}_{det} + \lambda_4 \mathcal{L}_1.$$
⁽⁹⁾

4 EXPERIMENTS

In this section, we evaluate the effectiveness of the proposed method through qualitative and quantitative comparison. Specifically, widely used UIEBD [26], EUVP [18], U45 [29] and UCCS [36] datasets are used to evaluate the performance of WaterFlow for underwater image enhancement. Six representative methods Water-Net [26], DLIFM [6], Ucolor [25], TOPAL [19], TACL [37] and SemiUIR [17] in

WaterFlow: Heuristic Normalizing Flow for Underwater Image Enhancement and Beyond

MM '23, October 29-November 3, 2023, Ottawa, ON, Canada.



Figure 3: Enhancement results on UIEBD dataset. We compared the pixel distribution of the reference images and enhanced images. Obviously, the distribution of our result is the closest to the distribution of reference image.



Figure 4: Enhancement results on EUVP dataset. We calculate the histogram distribution of RGB color space in the dataset. The x and y axis of the diagram represents the pixel intensity and the probability distribution. It can be found that the distribution of RGB color space in our results fits to the distribution of in-air images best.



Figure 5: Enhancement results on U45 dataset. We calculate the dispersion of RGB color space. The red, green and blue boxs respectively represent the corresponding color channel. It can be observed that the proposed results solve the problem of varying degrees of light decay underwater.



Figure 6: Enhancement results on UCCS dataset. Our method is obviously superior to other methods in correcting color attenuation and enhancing contrast, especially in the regions of red and green frames.

Table 1: Ç	Quantitative compariso	on for underwater image	enhancement in terms	of UCIQE(\uparrow), UIQM(\uparrow)	,UICM(↑) ,PSNR(↑) and
SSIM(↑). 1	denotes that large val	ues mean better results. 7	The best and second res	alts are marked in bol	d and <u>underline</u> .

Method	UIEBD			EUVP		U45			UCCS					
	UCIQE	UIQM	UICM	PSNR	SSIM	UCIQE	UIQM	UICM	UCIQE	UIQM	UICM	UCIQE	UIQM	UICM
Water-Net	0.5856	1.8681	-37.9812	18.8156	0.8257	0.5836	4.4419	-17.2934	0.5680	4.3612	-23.7057	0.5559	3.5007	-14.0720
DLIFM	0.6095	1.9553	-38.0444	20.0894	0.8565	0.6205	4.1051	-26.4611	0.5880	4.2763	-25.4130	0.5200	2.9137	-25.6534
Ucolor	0.5542	1.6799	-40.9628	19.7202	0.8273	0.5846	4.3043	-19.4697	0.5641	4.3708	-21.6621	0.5201	3.4145	-34.9304
TOPAL	0.5646	2.2013	-48.1846	19.1631	0.8043	0.6010	4.2689	-21.6535	0.5524	3.9625	-32.6012	0.4801	3.7760	-23.3429
TACL	0.6128	2.2848	-39.8192	21.4864	0.8398	0.5842	3.6547	-42.8193	0.6223	4.4013	-20.0128	0.5935	3.6249	-18.3285
SemiUIR	0.6202	1.9916	-40.0554	<u>21.6818</u>	0.8753	0.6175	4.4360	-15.3805	0.6110	4.5201	-18.7781	0.5538	3.7847	-5.0835
Ours	<u>0.6157</u>	2.5968	-35.1569	21.7420	<u>0.8584</u>	0.6397	<u>4.4400</u>	-12.6861	0.6229	<u>4.4129</u>	-13.6104	0.5569	3.9842	-4.1614



Figure 7: Evaluation of object detection on UCCS dataset. It can find that the proposed method achieves good performance in both accuracy and quantity.

recent years are compared for evaluating the performance. Both subjective and objective results are adopted for analysis. Underwater Color Image Quality Evaluation (UCIQE) [56], Underwater Image Quality Measurement (UIQM) [44] and Underwater Image Contrast Measure (UICM) [44] are used as non-reference evaluation metrics. A higher UCIQE, UIQM or UICM score indicates a better human visual perception. Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) [54] are adopted as the fully reference evaluations. Higher values of these metrics indicate a better similarity between the resulting image and the reference image in terms of both content and structure. Furthermore, UCCS [36] and Aquarium [46] datasets are used to evaluate the ability of the proposed method to adapt to subsequent detection tasks. The commonly utilized Average Precision (AP) is adopted as the detection-driven evaluation metric, which is positive for the performance of underwater object detection.

4.1 Implement Details

Our network is implemented by the pytorch and trained on NVIDIA RTX 3090 GPU. We first randomly crop the UIEBD [26] dataset into the size of 384×384 and use the cropped dataset as the training dataset for underwater image enhancement. UCCS [36] and Aquarium [46] datasets are used as the training dataset for underwater object detection. We first train the heuristic normalizing flow and Detection Perception Module separately for $5 \times e^5$ iterations. Then, we jointly train the pre-trained modules for $3 \times e^5$ iterations. During the training process, Adam was used as our optimizer with the

uniform learning rate of 1e-6 and the batch size of 2. λ_1 , λ_2 , λ_3 , λ_4 are set to 1, 100, 0.1, 1.

4.2 Qualitative Results

For visual comparison, Fig. 3 shows the enhancement results of all methods on UIEBD dataset. Water-Net and Ucolor not only fail to recover the original reflection, but also suffer from low contrast and saturation. DLIFM, TOPAL, TACL and SemiUIR have limited effects on underwater image enhancement, which still remain conspicuous color deviation at close scenes. Compared with the other methods, the proposed method recovers the degradation of light and rectifies the unexpected coloration. In addition, we demonstrate the pixel distribution similarity diagram of a particular region between the corresponding enhanced and reference images. The red line and the blue line in the diagram indicate the pixel distribution of the reference image and the enhanced results closest to the reference image.

The comparison on the EUVP dataset is shown in Fig. 4. Ucolor and TOPAL introduce artifacts and unnatural colors. Water-Net, TACL and SemiUIR fail to alleviate the scattering of light underwater. Although DLIFM produces pleasing images, the instructive details of the image have not been recovered well. Compared with the other methods, the proposed method not only effectively corrects the color deviation, but also restores the scene radiance and contrast of the image. Fig. 4 also shows the distribution of RGB color space for the images obtained by different methods. In general, the proposed method more effectively addresses the swift attenuation of red wavelengths underwater, with its color space distribution closely approximating that of in-air images.

The qualitative comparison on the U45 datasets is shown in Fig. 5. It is obvious that the red light and blue light decay significantly more than the green light underwater in this acquisition environment. However, the other compared methods exhibit limited effectiveness in restoring the decay of blue light. The boxplot for the intensity distribution of the RGB color channels is set below each image, with the horizontal axis representing pixel intensity. Additionally, semi-transparent rectangles are introduced to expound the relationship between distinct color spaces more lucidly, with the gray rectangle signifying the region where outliers emerged. In comparison with the proposed method, the other methods not only have a noticeably inferior ability to revive the attenuation of



Figure 8: Evaluation of object detection on Aquarium dataset with different enhancement methods. It can find that the proposed method is most suitable for underwater object detection.



Figure 9: Detection accuracy on UCCS (a) and Aquarium (b) datasets with the enhanced images generating from the representative methods. The x-axis represents the category of the object.



Figure 10: Ablation study of loss function. \mathcal{L}_1 , \mathcal{L}_s and \mathcal{L}_c denote L1 loss, style loss, and contrasive loss.

blue light but also introduce certain aberrant outliers, adversely impacting the efficacy of the enhanced images.

The visual comparison on the UCCS datasets is shown in Fig. 6. TOPAL fails to restore the color deviation of the image, which suffers from low saturation. Unreasonable color artifacts are introduced by Water-Net, TACL and SemiUIR. Informative details are ignored by DLIFM and Ucolor, resulting in obscure reflections and low contrast. Compared with other methods, our method significantly restores the brightness and contrast of underwater scenes without the interference introduction.

We utilize the same comparison methods to assess whether the proposed enhancement results implicitly contain more semantic features that are applicable to subsequent detection tasks. The visual comparison of UCCS dataset is shown in Fig. 7. The proposed method achieves a higher detection confidence than other methods with more detected trepangs and urchins. The qualitative comparison of the Aquarium dataset is shown in Fig. 8, which clearly shows the remarkable advantages of the proposed method in adapting to subsequent detection tasks.

4.3 Quantitative Results

The quantitative results of representative methods for underwater image enhancement are shown in Tab. 1. It can be observed that the proposed method exceeds all other methods in UICM, and achieves competitive results in UCIQE and UIQM. However, as noted by [3, 17], although both UCIQE and UIQM metrics can represent the degree of restoration of underwater images, they are somewhat heuristic and have limited applicability on evaluating the performance of underwater image enhancement. We further compared fully-reference evaluation on the UIEBD dataset which contains reference images. The proposed method achieves a significant advantage on PSNR and is only lower than SemiUIR on SSIM. However, as noted by [25], the reference images of UIEBD are synthetic by multiple enhancement methods, which indicates that some unreliable samples may affect the evaluation of the performance. Combining with NR-IQA and full-reference benchmarks, our method still achieves great advantages in underwater image enhancement, consistent with the qualitative results. The quantitative comparison for underwater object detection is shown in Fig. 9. It can be observed that the proposed method outperforms the other competitive methods in terms of detection accuracy in both UCCS and Aquarium datasets. It can be proven that the proposed results can not only achieve good visual enhancement effects, but also potentially contain more perceptual information which is beneficial for subsequent detection tasks.

4.4 Ablation Study

4.4.1 Study on Loss Function. We discuss the effectiveness of different loss functions. The visual results are shown in Fig. 10. w/o \mathcal{L}_1 and w/o \mathcal{L}_c make the images deviate from their intrinsic colors and introduce unpleasant artifacts. w/o \mathcal{L}_s still suffers from low contrast and clarity although it has improved the color balance of



Figure 11: Quantitative results of the ablation experiment. (a) denotes the evaluation on the loss function. (b) represents the evaluation for the input of Heuristic Prior guided Encoder. (c) represents the evaluation for the number of Hybrid Invertible Blocks. (d) denotes the Precision-Recall curve of ablation study for underwater object detection, the x and y axis respectively represent recall and precision. (e) denotes the F1-confidence curve, the x and y axis represent the confidence and F1 score.



Input w/o Color w/o Grad w/o Depth Ours Figure 12: Visualization for ablation study on the input of Heuristic Prior guided Encoder.



Figure 13: Visualization for ablation study on the number of Hybrid Invertible Blocks.

the image to a certain extent. Combined with quantitative comparison in (a) of Fig. 11, the proposed method achieves the best enhancement effect, which verifies the effectiveness of all functions.

4.4.2 Study on Heuristic Prior guided Encoder. We conduct ablation experiments on the input of HPE to verify the effectiveness of each input to the proposed method. The qualitative comparison is shown in Fig. 12. In the first example, when the color image or depth map is not adopted as input, unpleasant shadows are introduced. The effect of image enhancement is obviously limited when the gradient map is not used. In the second example, the proposed method achieved the best enhancement effect, which not only restore the color of the ground but also significantly improved the clarity of the content in the red and green frames. The quantitative comparison is shown in (b) of Fig. 11, which shows that the proposed method achieves the best results on all benchmarks. Any input removed significantly attenuates the enhanced image.

4.4.3 Study on Number of Hybrid Invertible Blocks. We conduct ablation experiments on the number of Hybrid Invertible Blocks

N. The visual comparison is illustrated in Fig. 13. It is obvious that when the number of blocks is less than 3, the enhanced image suffers from local over enhancement and the image artifacts introduction. Quantitative comparison is shown in (c) of Fig. 11. Considering the limitations of network parameters and computational costs, we finally employ N as 3 in our framework.

4.4.4 Study on underwater object detection. In order to verify the effectiveness of the proposed method for underwater object detection. We directly train the existing detection network without enhancement on underwater images to verify the effect of enhanced data on underwater object detection. At the same time, the enhancement module and the Detection Perception Module are not trained jointly, so as to test whether DPM can make the enhanced data more conducive to subsequent detection tasks. Quantitative comparison is shown in (d) and (e) of Fig. 11. w/o Enhancement and w/o Joint respectively indicate that the enhanced image is not used as a preprocessing stage and the latent features are not transmitted to the enhancement module from the DPM. Evidently, the proposed method achieves the highest detection performance, which proves the effectiveness of the DPM.

5 CONCLUSION

In this paper, we propose a detection-driven heuristic normalizing flow underwater image enhancement. We first adopt the underwater image restoration process into the normalizing flow to establish the invertible mapping of the degraded image and the its clear counterpart through bilateral constraints. Then we further introduce the heuristic prior guided injector to improve the representation ability of the enhancement network by progressively estimating the underwater imaging parameters. At last, we introduce the detection perception module, which propagates high-level perceptual features into the enhancement model as gradient information to help generate the detection-driven enhanced image. Extensive experiments on multiple benchmarks show that the proposed method not only achieves good visual enhancement effects, but is also more suitable for subsequent detection tasks.

ACKNOWLEDGMENTS

This work is partially supported by the National Key R&D Program of China (No. 2022YFA1004101), the National Natural Science Foundation of China (No. U22B2052). WaterFlow: Heuristic Normalizing Flow for Underwater Image Enhancement and Beyond

MM '23, October 29-November 3, 2023, Ottawa, ON, Canada.

REFERENCES

- Rameen Abdal, Peihao Zhu, Niloy J Mitra, and Peter Wonka. 2021. Styleflow: Attribute-conditioned exploration of stylegan-generated images using conditional continuous normalizing flows. ACM Transactions on Graphics (ToG) 40, 3 (2021), 1–21.
- [2] Jie An, Siyu Huang, Yibing Song, Dejing Dou, Wei Liu, and Jiebo Luo. 2021. Artflow: Unbiased image style transfer via reversible neural flows. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 862–871.
- [3] Dana Berman, Deborah Levy, Shai Avidan, and Tali Treibitz. 2021. Underwater Single Image Color Restoration Using Haze-Lines and a New Quantitative Dataset. IEEE Transactions on Pattern Analysis and Machine Intelligence 43, 8 (2021), 2822– 2837. https://doi.org/10.1109/TPAMI.2020.2977624
- [4] Pierre Bouguer. 1729. Essai d'optique sur la gradation de la lumière. Claude Jombert.
- [5] Long Chen, Zheheng Jiang, Lei Tong, Zhihua Liu, Aite Zhao, Qianni Zhang, Junyu Dong, and Huiyu Zhou. 2020. Perceptual underwater image enhancement with deep learning and physical priors. *IEEE Transactions on Circuits and Systems* for Video Technology 31, 8 (2020), 3078–3092.
- [6] Xuelei Chen, Pin Zhang, Lingwei Quan, Chao Yi, and Cunyue Lu. 2021. Underwater image enhancement based on deep learning and image formation model. arXiv preprint arXiv:2101.00991 (2021).
- [7] Xuelian Cheng, Huan Xiong, Deng-Ping Fan, Yiran Zhong, Mehrtash Harandi, Tom Drummond, and Zongyuan Ge. 2022. Implicit motion handling for video camouflaged object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 13864–13873.
- [8] John Y Chiang and Ying-Ching Chen. 2011. Underwater image enhancement by wavelength compensation and dehazing. *IEEE transactions on image processing* 21, 4 (2011), 1756–1769.
- [9] Yingying Deng, Fan Tang, Weiming Dong, Chongyang Ma, Xingjia Pan, Lei Wang, and Changsheng Xu. 2022. StyTr2: Image Style Transfer with Transformers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 11326–11336.
- [10] Laurent Dinh, David Krueger, and Yoshua Bengio. 2014. Nice: Non-linear independent components estimation. arXiv preprint arXiv:1410.8516 (2014).
- [11] Paul Drews, Erickson Nascimento, Filipe Moraes, Silvia Botelho, and Mario Campos. 2013. Transmission estimation in underwater single images. In Proceedings of the IEEE international conference on computer vision workshops. 825–830.
- [12] Chengjian Feng, Yujie Zhong, Yu Gao, Matthew R Scott, and Weilin Huang. 2021. Tood: Task-aligned one-stage object detection. In 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE Computer Society, 3490–3499.
- [13] Ahmad Shahrizan Abdul Ghani and Nor Ashidi Mat Isa. 2015. Underwater image quality enhancement through integrated color model with Rayleigh distribution. *Applied soft computing* 27 (2015), 219–230.
- [14] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition. 580-587.
- [15] Yong Guo, Jian Chen, Jingdong Wang, Qi Chen, Jiezhang Cao, Zeshuai Deng, Yanwu Xu, and Mingkui Tan. 2020. Closed-loop matters: Dual regression networks for single image super-resolution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 5407–5416.
- [16] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. 2017. Mask r-cnn. In Proceedings of the IEEE international conference on computer vision. 2961–2969.
- [17] Shirui Huang, Keyan Wang, Huan Liu, Jun Chen, and Yunsong Li. 2023. Contrastive semi-supervised learning for underwater image restoration via reliable bank. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 18145–18155.
- [18] Md Jahidul Islam, Youya Xia, and Junaed Sattar. 2020. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters* 5, 2 (2020), 3227–3234.
- [19] 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* (2022).
- [20] Zhiying Jiang, Risheng Liu, Shuzhou Yang, Zengxi Zhang, and Xin Fan. 2023. Contrastive Learning Based Recursive Dynamic Multi-Scale Network for Image Deraining. arXiv preprint arXiv:2305.18092 (2023).
- [21] Zhiying Jiang, Zengxi Zhang, Jinyuan Liu, Xin Fan, and Risheng Liu. 2023. Modality-Invariant Representation for Infrared and Visible Image Registration. arXiv preprint arXiv:2304.05646 (2023).
- [22] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for realtime style transfer and super-resolution. In *European conference on computer* vision. Springer, 694–711.
- [23] Kang Kim and Hee Seok Lee. 2020. Probabilistic anchor assignment with iou prediction for object detection. In *European Conference on Computer Vision*. Springer, 355–371.

- [24] Durk P Kingma and Prafulla Dhariwal. 2018. Glow: Generative flow with invertible 1x1 convolutions. Advances in neural information processing systems 31 (2018).
- [25] Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren. 2021. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Transactions on Image Processing* 30 (2021), 4985–5000.
- [26] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. 2019. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing* 29 (2019), 4376–4389.
- [27] Chongyi Li, Jichang Guo, Chunle Guo, Runmin Cong, and Jiachang Gong. 2017. A hybrid method for underwater image correction. *Pattern Recognition Letters* 94 (2017), 62–67.
- [28] Chongyi Li, Jichang Quo, Yanwei Pang, Shanji Chen, and Jian Wang. 2016. Single underwater image restoration by blue-green channels dehazing and red channel correction. In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 1731–1735.
- [29] Hanyu Li, Jingjing Li, and Wei Wang. 2019. A fusion adversarial underwater image enhancement network with a public test dataset. arXiv preprint arXiv:1906.06819 (2019).
- [30] Hongyu Li, Jia Li, Dong Zhao, and Long Xu. 2021. DehazeFlow: Multi-scale Conditional Flow Network for Single Image Dehazing. In Proceedings of the 29th ACM International Conference on Multimedia. 2577–2585.
- [31] Chuan Lin, Guangjie Han, Mohsen Guizani, Yuanguo Bi, Jiaxin Du, and Lei Shu. 2020. An SDN architecture for AUV-based underwater wireless networks to enable cooperative underwater search. *IEEE Wireless Communications* 27, 3 (2020), 132–139.
- [32] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. 2017. Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision. 2980–2988.
- [33] Jinyuan Liu, Xin Fan, Zhanbo Huang, Guanyao Wu, Risheng Liu, Wei Zhong, and Zhongxuan Luo. 2022. Target-aware dual adversarial learning and a multiscenario multi-modality benchmark to fuse infrared and visible for object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 5802–5811.
- [34] Jinyuan Liu, Runjia Lin, Guanyao Wu, Risheng Liu, Zhongxuan Luo, and Xin Fan. 2022. CoCoNet: Coupled Contrastive Learning Network with Multi-level Feature Ensemble for Multi-modality Image Fusion. arXiv preprint arXiv:2211.10960 (2022).
- [35] Jinyuan Liu, Guanyao Wu, Junsheng Luan, Zhiying Jiang, Risheng Liu, and Xin Fan. 2023. HoLoCo: Holistic and local contrastive learning network for multi-exposure image fusion. *Information Fusion* 95 (2023), 237–249.
- [36] Risheng Liu, Xin Fan, Ming Zhu, Minjun Hou, and Zhongxuan Luo. 2020. Realworld underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE Transactions on Circuits and Systems for Video Technology* 30, 12 (2020), 4861–4875.
- [37] Risheng Liu, Zhiying Jiang, Shuzhou Yang, and Xin Fan. 2022. Twin adversarial contrastive learning for underwater image enhancement and beyond. *IEEE Transactions on Image Processing* 31 (2022), 4922–4936.
- [38] Risheng Liu, Zhu Liu, Jinyuan Liu, and Xin Fan. 2021. Searching a hierarchically aggregated fusion architecture for fast multi-modality image fusion. In Proceedings of the 29th ACM International Conference on Multimedia. 1600–1608.
- [39] Risheng Liu, Zhu Liu, Jinyuan Liu, Xin Fan, and Zhongxuan Luo. 2023. A Taskguided, Implicitly-searched and Meta-initialized Deep Model for Image Fusion. arXiv preprint arXiv:2305.15862 (2023).
- [40] Zhu Liu, Jinyuan Liu, Guanyao Wu, Long Ma, Xin Fan, and Risheng Liu. 2023. Bi-level Dynamic Learning for Jointly Multi-modality Image Fusion and Beyond. arXiv preprint arXiv:2305.06720 (2023).
- [41] Andreas Lugmayr, Martin Danelljan, Luc Van Gool, and Radu Timofte. 2020. Srflow: Learning the super-resolution space with normalizing flow. In *European* conference on computer vision. Springer, 715–732.
- [42] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. 2022. Toward Fast, Flexible, and Robust Low-Light Image Enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 5637– 5646.
- [43] Pan Mu, Haotian Qian, and Cong Bai. 2022. Structure-Inferred Bi-level Model for Underwater Image Enhancement. In Proceedings of the 30th ACM International Conference on Multimedia. 2286–2295.
- [44] Karen Panetta, Chen Gao, and Sos Agaian. 2015. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering* 41, 3 (2015), 541–551.
- [45] Yan-Tsung Peng, Keming Cao, and Pamela C Cosman. 2018. Generalization of the dark channel prior for single image restoration. *IEEE Transactions on Image Processing* 27, 6 (2018), 2856–2868.
- [46] Slavko Prytula. 2020. Underwater Object Detection Dataset. https://public.roboflow.com/object-detection/aquarium.
- [47] Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767 (2018).

Zengxi Zhang, Zhiying Jiang, Jinyuan Liu, Xin Fan, and Risheng Liu

- [48] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems 28 (2015).
- [49] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. 2019. Generalized intersection over union: A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 658–666.
- [50] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).
- [51] Peize Sun, Rufeng Zhang, Yi Jiang, Tao Kong, Chenfeng Xu, Wei Zhan, Masayoshi Tomizuka, Lei Li, Zehuan Yuan, Changhu Wang, et al. 2021. Sparse r-cnn: Endto-end object detection with learnable proposals. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 14454–14463.
- [52] Yi Wang, Hui Liu, and Lap-Pui Chau. 2017. Single underwater image restoration using adaptive attenuation-curve prior. *IEEE Transactions on Circuits and Systems* I: Regular Papers 65, 3 (2017), 992–1002.
- [53] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. 2022. Low-light image enhancement with normalizing flow. In Proceedings

of the AAAI Conference on Artificial Intelligence, Vol. 36. 2604–2612.

- [54] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions* on image processing 13, 4 (2004), 600–612.
- [55] Chenxi Xie, Changqun Xia, Mingcan Ma, Zhirui Zhao, Xiaowu Chen, and Jia Li. 2022. Pyramid Grafting Network for One-Stage High Resolution Saliency Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 11717–11726.
- [56] Miao Yang and Arcot Sowmya. 2015. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing* 24, 12 (2015), 6062–6071.
- [57] Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. 2023. Implicit Neural Representation for Cooperative Low-light Image Enhancement. arXiv:2303.11722 [cs.CV]
- [58] Dana R Yoerger, Annette F Govindarajan, Jonathan C Howland, Joel K Llopiz, Peter H Wiebe, Molly Curran, Justin Fujii, Daniel Gomez-Ibanez, Kakani Katija, Bruce H Robison, et al. 2021. A hybrid underwater robot for multidisciplinary investigation of the ocean twilight zone. *Science Robotics* 6, 55 (2021), eabe1901.