Semantic-aware Texture-Structure Feature Collaboration for Underwater Image Enhancement

Di Wang[1], Long Ma[1], Risheng Liu[2] and Xin Fan[2,*]

Abstract— **Underwater image enhancement has become an attractive topic as a significant technology in marine engineering and aquatic robotics. However, the limited number of datasets and imperfect hand-crafted ground truth weaken its robustness to unseen scenarios, and hamper the application to high-level vision tasks. To address the above limitations, we develop an efficient and compact enhancement network in collaboration with a high-level semantic-aware pretrained model, aiming to exploit its hierarchical feature representation as an auxiliary for the low-level underwater image enhancement. Specifically, we tend to characterize the shallow layer features as textures while the deep layer features as structures in the semantic-aware model, and propose a multi-path Contextual Feature Refinement Module (CFRM) to refine features in multiple scales and model the correlation between different features. In addition, a feature dominative network is devised to perform channel-wise modulation on the aggregated texture and structure features for the adaptation to different feature patterns of the enhancement network. Extensive experiments on benchmarks demonstrate that the proposed algorithm achieves more appealing results and outperforms state-of-the-art methods by large margins. We also apply the proposed algorithm to the underwater salient object detection task to reveal the favorable semantic-aware ability for high-level vision tasks. This code is available at STSC.**

I. INTRODUCTION

Underwater image enhancement is a practical but challenging technology in the field of underwater vision, which is widely contributed to many applications such as aquatic robotics [21], underwater path planning [2], and underwater object real-time tracking [3], etc. Over the past few decades, a series of underwater enhancement methods also have been explored, ranging from traditional model-free methods [1], [6], [7], [16] to physical model-based methods [10], [17], [20], [25], [31].

In recent years, significant progress has been witnessed on underwater image enhancement tasks due to the use of deep CNNs [13]–[15] and GANs [?], [5], [9]. These kinds of methods are either employed to estimate the parameters of the physical models or directly generate the enhanced images, which have shown impressive advantages in improving image contrast and alleviating color cast. However, they fail to preserve the pleasing texture and structure information

*Corresponding Author: xin.fan@dlut.edu.cn

[1]School of Software Technology, Dalian University of Technology, Dalian, 116024, China.

[2]DUT-RU International School of Information Science & Engineering, Dalian University of Technology, Dalian, 116024, China.

(a) WaterNet [15] (b) UWCNN [14] (c) FGAN [9] (d) Ours

Fig. 1. Saliency object detection on enhanced underwater images. WaterNet and UWCNN tend to miss detection, and FGAN tends to incomplete detection and blurry edges, while our method obtains a pleasing saliency map.

due to the limited number of datasets with imperfect hand-crafted ground truth, which hinders their application in high-level visual tasks. As shown in Fig. 1, the human outlines in (a) and (b) cannot be detected due to the loss of structure information, and (c) shows blurred edges due to the loss of texture details. Thus, how to develop an effective algorithm for learning complete image structures and precise textures is very urgent for underwater enhancement tasks.

Recent researches reveal that semantic clues of the high-level vision tasks [?], [18], [29], [30] offer guidance in low-level vision tasks, such as super-resolution [33], [34], dehazing [26], deraining [32], and deblurring [27]. A common theme is that the semantic labels are exploited as global priors to guide networks to generate photo-realistic results. To be specific, one idea is to directly concatenate the semantic probability map with the corrupted image as the inputs of the networks [27], [33], and the other idea is to extract features from the semantic labels to guide the decoding of features from inputs [26], [34]. Considering the extreme complexity of underwater scenes, semantic probability maps, the final products of segmentation networks, can only resolve the ambiguity in category-related object boundary but are not enough to restore accurate textures and structures required by underwater image enhancement. Thus, we characterize the semantic-aware features of high-level visual tasks from texture and structure perspectives, and build our algorithm to exploit more informative features to facilitate underwater image enhancement.

Based on the above motivation, we develop a novel underwater image enhancement algorithm in cooperation with a classical pretrained semantic-aware model. Our algorithm

Fig. 2. The overview architecture of the proposed semantic-aware texture-structure feature collaboration network for underwater image enhancement.

framework is designed into three parts: 1) With the semantic-aware model as a powerful auxiliary, we present an efficient and compact UNet-like framework as the base network. 2) Since the generation of near-lossless image content depends on the collaboration of textures and structures, we represent features from the semantic-aware model via two independent branches according to the consensus that the shallow features imply textures while the deep features imply structures. Besides, to refine texture and structure features in multiple scales and model correlation between features, we design and deploy a multi-path Contextual Feature Refinement Module (CFRM) on the two branches. 3) Aiming at the characteristics of the non-uniform illumination and various degradation in underwater images, we present a feature dominative network to perform channel-wise modulation for the aggregated features, to adapt to different feature patterns of the base network. Given that the modulated features have gone through an encoding-like process, we determine to embed them into the decoder of the base network to facilitate feature reconstruction. Our algorithm is a general framework, in which any semantic-aware model from high-level vision tasks (i.e., classification, semantic segmentation, and object detection) can provide important guidance for the enhancement models to restore the photo-realistic images. We further apply the proposed algorithm to the underwater salient object detection task. The results in Fig. 1 reveal that our algorithm has the semantic-aware ability.

The main contributions of this paper are three-fold:

- We exploit the hierarchical feature representation of the high-level semantic-aware model as an auxiliary and propose an efficient and compact underwater enhancement framework.
- We characterize semantic-aware features as textures and structures separately, and propose a multi-path contextual feature refinement module to model the correlation

between features for formulating near-lossless image content. Besides, a feature dominative network is developed to perform channel-wise feature modulation to adapt to the base enhancement network.
- We conduct extensive experiments and demonstrate that the proposed algorithm performs favorably against state-of-the-art methods. The application on underwater salient detection reveals its semantic-aware ability.

II. PROPOSED ALGORITHM

A. Network Architecture

As shown in Fig. 2, the architecture of the proposed algorithm consists of three main components, a UNet-like base network, a semantic-aware feature aggregation network, and a feature dominative network.

UNet-like Base Network. It serves as a base network for underwater image enhancement, which takes an underwater image x as input and is suppose to reconstruct an enhanced image y with complete structures and precise textures. Let E and G denote the encoder and decoder of the base network, respectively. The input image x is first fed into the E to capture hierarchical multi-scale features denoted by $f^i = E(x; \Theta_E)$, where $i \in \{1, 2, \ldots, w\}$ and w is scale factor. To enlarge the receptive field and maximize the information utilization, we employ a pyramid context block $\mathcal{P}(\cdot)$ at the bottom of the UNet-like network. Thus, the features to be decoded are noted as $\mathcal{P}(f^w)$, and multi-scale reconstructed features $g^i = G(\mathcal{P}(f^w); \Theta_G)$ are generated by the decoder. Θ_E and Θ_G are model parameters of the E and G.

Semantic-aware Feature Aggregation Network. In this network, we exploit a high-level semantic-aware model as the feature extractor, which is a widely-used model VGG16 [28] on classification tasks. Thanks to the training on an extra-large scale benchmark ImageNet [4], this model embraces powerful feature representation ability and does not need to

be retrained on the underwater datasets. We directly feed x into the semantic-aware model to extract hierarchical multi-scale features denoted by F_i, and also $i \in \{1, 2, \ldots, w\}$. Due to the influence of too deep semantic-aware features on low-level tasks is negligible, the value of w is set to 4 in all the experiments. The core of the network is to build texture branch $\mathcal{B}_{te}$ and structure branch $\mathcal{B}_{st}$ to tackle the extracted features separately. We represent the shallow features from the first two scales as textures and represent the deep features from the last two scales as structures, and reorganize them via reshaping operation and concatenation. For the reorganized texture features $\mathcal{F}_{te}$ and structure features $\mathcal{F}_{st}$, we present a multi-path Contextual Feature Refinement Module (CFRM) to implement two refinement processes $\mathcal{F}_{te} \rightarrow \tilde{\mathcal{F}}_{te}$ and $\mathcal{F}_{st} \rightarrow \tilde{\mathcal{F}}_{st}$. As shown in Fig. 2, there are three parallel convolution paths with progressively increased kernel sizes in the CFRM. The refined texture and structure features by each path are denoted as $\phi_{te}^k(\mathcal{F}_{te})$ and $\phi_{st}^k(\mathcal{F}_{st})$, where $k \in K$ and $K = [3, 5, 7]$. Note that both $\phi_{te}^k(\cdot)$ and $\phi_{st}^k(\cdot)$ denote the convolution paths with kernel size of $k \times k$ in $\mathcal{B}_{te}$ and $\mathcal{B}_{st}$, respectively. Thus, the textures and structures refined by the CFRMs are represented as

$$\begin{aligned} \tilde{\mathcal{F}}_{te} &= \mathcal{C}\left(\phi_{te}^k(\mathcal{F}_{te}), k \in K\right), \\ \tilde{\mathcal{F}}_{st} &= \mathcal{C}\left(\phi_{st}^k(\mathcal{F}_{st}), k \in K\right). \end{aligned} \quad (1)$$

Subsequently, we aggregate the texture and structure features by

$$\mathcal{F}_{ste} = \mathcal{C}\left(\tilde{\mathcal{F}}_{te}, \tilde{\mathcal{F}}_{st}\right), \quad (2)$$

where $\mathcal{C}$ denotes the concatenation and a followed 1×1 convolution layer to reduce the feature channels.

Feature Dominative Network. Due to the characteristics of non-uniform illumination and various degradation in underwater images, it is not applicable to directly embed the aggregated features into the base network. Hence, we present a feature dominative network to learn image-specific and region-specific features through performing channel-wise feature modulation. To match features with different scales of the base network, the aggregated features $\mathcal{F}_{ste}$ are first reshaped into those with specific sizes, which are then fed to the Channel Transformation Layer (CTL) to obtain applicable feature embeddings matching the base network. As shown in Fig. 2, a convolution layer of 1×1 is fist used to conduct a transformation $\mathcal{F}_{ste} \rightarrow f_t^i$, also $i \in \{1, 2, \ldots, w\}$. And then an adaptive weight vector $\mathbf{w}_t^i$ is obtained by global average pooling, down-upscaling operations, and sigmoid function. The final modulated features f_{ste}^i are obtained by $f_t^i \otimes \mathbf{w}_t^i$, where $\otimes$ represents the element-wise multiplication.

Finally, we embed modulated features into the decoder of the UNet-like base network to generate an enhanced underwater image by

$$y = G_{i \rightarrow w}\left(f^{i \rightarrow w}, f_{ste}^{i \rightarrow w}, g^{i \rightarrow w}\right), \quad (3)$$

where $i \rightarrow w$ is the scale range of the features. $G_{i \rightarrow w}$ denotes the progressive feature reconstruction by the decoder under the scale range.

Datasets	Training (#)	Testing (#)	Paired/Unpaired
UIEB [15]	712	238	Paired
EUVP [9]	7,200	4,284	Paired
RUIE [19]	0	300	Unpaired
USOD [8]	0	300	Paired

B. Loss Functions

Considering that human visual perception often pays more attention to image details and textures, we use multi-scale structure similarity (MS-SSIM [35]) loss function $\mathcal{L}_{ssim}^{sm}$ to optimize our network, thus generate more realistic image content. However, doing so leads to image contrast change and color distortion. In view of this, we use a widely-used pixel-wise loss function $\mathcal{L}_1$ to ensure the insensitivity of the network to image contrast and color, so as to achieve a good trade-off between content restoration and picture fidelity. Therefore, the loss function is defined as:

$$\mathcal{L} = \lambda \cdot \mathcal{L}_{ssim}^{ms} + (1 - \lambda) \cdot \mathcal{L}_1, \quad (4)$$

where, λ is a balance parameter. It is set to 0.8 in our work.

III. EXPERIMENTAL RESULTS

A. Dataset and Implementation Details

Datasets. We evaluate the proposed algorithm using two labeled underwater datasets UIEB [15] and EUVP [9], and an unlabeled real-world underwater dataset RUIE [19], respectively. Moreover, we also utilize a new challenging underwater salient object detection dataset USOD [8] to verify the semantic-aware ability of our algorithm on the high-level vision tasks. Table I provides a detailed description of these datasets.

Training Settings. We randomly crop 8 image patches of size 224×224 to form a batch. The Adam optimizer [12] ($\beta_1 = 0.9$, and $\beta_2 = 0.999$) is used to optimize our model. The initial learning rate is 5×10^{-4}, and decreasing to 0.2 times every $8,000$ iterations during training. The whole training phase goes through $100,000$ iterations. We implement the proposed network using the PyTorch framework with an NVIDIA 1080Ti GPU.

Evaluation Metrics. For the sake of the comprehensive and fair assessment, we employ four metrics involving reference and non-reference approaches. For the UIEB [15] and EUVP [9] dataset with reference images, we mainly adopt two widely-used metrics (i.e., PSNR (dB) and SSIM) for evaluation. For the RUIE [19] dataset without reference images, we mainly use the other two non-reference evaluation metrics (i.e., Underwater Image Quality Measurement UIQM [23] and Natural Image Quality Evaluator NIQE [22]).

B. Comparison with State-of-the-art Methods

We conduct extensive experiments to quantitatively and qualitatively evaluate our algorithm against several state-of-the-art methods including conventional methods [1], [6], [16], physical model-based methods [10], [31], and data-driven deep learning-based methods [5], [9], [13]–[15].

Dataset	Metric	EUIVF [1]	OCM [16]	UDCP [10]	TSA [6]	UGAN [5]	WaterNet [15]	AIO [31]	FGAN [9]	UWCNN [14]	Ucolor [13]	Ours
UIEB	PSNR	**21.93**	16.19	11.73	14.32	17.73	19.65	12.69	18.16	13.35	20.62	**22.45**
	SSIM	0.823	0.759	0.509	0.763	0.765	0.824	0.466	0.597	0.773	**0.921**	**0.902**
EUVP	PSNR	17.06	15.62	14.53	13.21	19.31	18.68	16.25	**19.49**	18.37	––	**23.23**
	SSIM	0.894	0.843	0.888	0.672	0.890	0.952	0.881	**0.963**	0.948	––	**0.987**

Quantitative Evaluation. We report the quantitative results on UIEB, EUVP, and RUIE benchmarks in Table II and Table III. It can be seen that our algorithm numerically outperforms most existing methods by large margins and ranks first or second in the four evaluation metrics. Especially, compared with Ucolor [13], a recent research, our algorithm gains **1.83dB** and **0.982** improvements in PSNR and UIQM on the UIEB dataset. Due to the fact that transmission maps are required by Ucolor and the code [24] for estimating them has not been released, we cannot make comparisons with it on EUVP and RUIE datasets. In addition, our algorithm performs better than current prevalent data-driven methods in all four metrics, such as WaterNet [15] and UWCNN [14].

Qualitative Evaluation. We first show the qualitative evaluations on the UIEB benchmark in Fig. 4. By observing the local enlarged areas, we note that some methods such as TSA [6], AIO [31], WaterNet [15], and FGAN [9], can not effectively alleviate the underwater haze-effect, while the UDCP [10] and UWCNN [14] cause severe contrast reduction and color cast. More seriously, almost all the comparison methods fail to restore the complete structures and precise textures. In contrast, the underwater image enhanced by our algorithm has a sharper structure and richer texture and achieves a balance of contrast and color cast simultaneously. Then, we also show qualitative results on the EUVP benchmark in Fig. 5. It can be observed that our algorithm performs well in dealing with structure characteristics such as contrast and color, and the precise textures are also clearly displayed incidentally. Moreover, we also show some visualization examples on the RUIE benchmark in Fig. 6. The greenish color cast weakens structure information and hides texture details of the underwater scenes as shown in (a). According to (b)-(k), we can observe that these visual results are either under-enhanced or introduce the reddish and brownish color cast, while our algorithm shows relatively more realistic textures.

Time Complexity Evaluation. For the efficiency of our algorithm, we compare time complexity against the state-of-the-art models on a single 1080Ti GPU. As shown in Fig. 3, our model performs faster, especially compared to the latest underwater enhancement method Ucolor [13]. Combined with ahead experimental results, it can be illustrated that the proposed algorithm can obtain desirable underwater enhancement results at a low computational cost.

C. Application for other High-level Tasks

To further verify that the improvement can be provided by our algorithm for high-level visual tasks, we apply an underwater salient object detection algorithm [11] to evaluate

Methods	NIQE ↓			UIQM ↑		
	UIEB	EUVP	REIU	UIEB	EUVP	REIU
EUIVF [1]	4.059	4.358	4.542	2.679	2.763	3.073
OCM [16]	**3.877**	4.628	4.538	2.545	2.776	2.912
UDCP [10]	4.303	4.398	5.131	1.772	2.079	2.099
TSA [6]	4.165	5.623	4.842	1.996	2.869	2.512
UGAN [5]	7.057	6.467	6.680	2.528	3.254	3.043
WaterNet [15]	4.484	4.375	4.544	2.857	3.065	**3.150**
AIO [31]	3.994	4.892	5.637	3.078	**3.346**	3.137
FGAN [9]	6.364	5.175	5.696	2.512	3.211	2.982
UWCNN [14]	4.441	**4.251**	**4.411**	**3.078**	2.231	2.781
Ucolor [13]	3.772	––	––	2.871	––	––
Ours	**3.451**	**4.165**	**3.694**	**3.763**	**3.297**	**3.966**

Fig. 3. Running time comparisons against state-of-the-art methods on a color image of size 648×480.

it on the benchmark USOD [8] dataset. Fig. 7 shows that the saliency maps generated by our algorithm have a more integrated structure and precise boundary, even though in the dark underwater environment. By contrast, some state-of-the-art methods [1], [6], [10] even fail to capture the rough outline of the objects. We implement quantitative evaluation on the USOD dataset. As summarized in Table IV, the proposed algorithm performs favorably against the state-of-the-art methods in three common evaluation metrics (i.e., F-measure, S-measure, and MAE). Application examples and quantitative results illustrate that the proposed underwater enhancement algorithm can make further effects on the implementation of relevant high-level vision tasks in the underwater environment.

D. Ablation Studies

In Table V, $M_0, M_1, \ldots, M_3$ refer to algorithms implemented for ablation analysis. M_0 is the UNet-like base network called BaseNet for underwater image enhancement. M_1 refers to that the multi-scale features F from the semantic-aware Feature Aggregation network (SFANet) are directly embedded into the decoder of the BaseNet. M_2 is a variant of

Fig. 4. Qualitative comparisons on the UIEB dataset. The enhanced result by our algorithm has more pleasing contrast and more precise textures.

(a) Input   (b) EUIVF   (c) OCM   (d) UDCP   (e) TSA   (f) UGAN

(g) WaterNet   (h) AIO   (i) FGAN   (j) UWCNN   (k) Ucolor   (l) Ours

EUIVF   OCM   UDCP   TSA   UGAN   WaterNet   AIO   FGAN   UWCNN   Ours

Fig. 5. Qualitative comparisons on the EUVP dataset. The enhanced results by our algorithm have better textures and colors.

(a) Input

(b) EUIVF   (c) OCM   (d) UDCP   (e) TSA   (f) UGAN

(g) WaterNet   (h) AIO   (i) FGAN   (j) UWCNN   (k) Ours

Fig. 6. Visualization of the underwater enhancement results by different methods on the RUIE dataset.

Input   EUIVF   OCM   UDCP   TSA   UGAN   WaterNet   FGAN   UWCNN   Ours   GT

Fig. 7. Application examples of different methods on the salient object detection task in real-world USOD underwater dataset.

our algorithm with only structure branch $\mathcal{B}_{st}$ in the SFANet, and M_3 is another variant with only texture branch $\mathcal{B}_{te}$. M_4 denotes that the final aggregated features of the SFANet are directly embedded into the decoder of the base network without the feature dominative network (FDNet).

Effectiveness of Texture and Structure Branches. We can observe from Table V that: i) On the whole, compared with M_0, M_1 only obtains a gain of 0.70dB, while our algorithm outperforms M_1 by a large margin (**1.51dB**). It can be explained that texture and structure branches are indeed conducive to maximizing feature utilization for underwater image enhancement. ii) For texture branch $\mathcal{B}_{te}$, the performance of the algorithm with it is improved by 1.96dB via comparing M_3 with M_0. To more explicitly demonstrate the effect of texture branch, a Gaussian filter is applied to the enhanced images to obtain texture layers for analysis. As shown

TABLE IV

QUANTITATIVE COMPARISONS ON SALIENT OBJECT DETECTION TASK WITH STATE-OF-THE-ART UNDERWATER ENHANCEMENT METHODS.

Metric	EUIVF [1]	OCM [16]	UDCP [10]	TSA [6]	UGAN [5]	WaterNet [15]	AIO [31]	FGAN [9]	UWCNN [14]	Ours
F-measure ↑	0.850	0.840	0.835	0.715	0.836	0.852	0.813	**0.851**	0.834	**0.854**
S-measure ↑	**0.833**	0.831	0.811	0.705	0.822	**0.833**	0.797	0.830	0.808	**0.837**
MAE ↓	**0.081**	**0.081**	0.088	0.124	0.084	**0.080**	0.096	0.082	0.092	**0.080**

TABLE V

ABLATION ANALYSIS ON THE UIEB DATASET.

Method	BaseNet	SFANet			FDNet	UIEB	
		F	$\mathcal{B}_{st}$	$\mathcal{B}_{te}$		PSNR	SSIM
M_0	✓					20.244	0.8772
M_1	✓	✓				20.943	0.8932
M_2	✓		✓		✓	20.879	0.8821
M_3	✓			✓	✓	22.202	0.8893
M_4	✓		✓	✓		21.846	0.8918
Ours	✓		✓	✓	✓	**22.448**	**0.9019**

(a)   (b)   (c)

Fig. 8. Ablation analysis of the texture branch. (a) Texture layer of the enhanced image of the algorithm without $\mathcal{B}_{te}$. (b) Texture layer of the enhanced image of the algorithm with $\mathcal{B}_{te}$. (c) Histogram comparison of (a) and (b).

in Fig. 8(a) and (b), the enhanced image of the algorithm with $\mathcal{B}_{te}$ has more textures. Besides, the statistical results in Fig. 8(c) show that the texture branch contributes to the restoration of image textures. iii) For structure branch $\mathcal{B}_{st}$, the algorithm with it gains 0.64dB improvement compared with M_0. The corresponding qualitative results in Fig. 9 show the content of the structure layer enhanced by the algorithm without $\mathcal{B}_{st}$ is seriously blurred. Therefore, we argue that texture and structure branches can model the correlation between features, and both of them are significant to generate near-lossless underwater image content.

Effectiveness of Feature Dominative Network. To illustrate the effectiveness of the feature dominative network (FDNet), we also train the model without the feature dominating process. As shown in Table V, using the FDNet can obtain a gain of 0.58dB, which indicates that the FDNet can reasonably dominate aggregated features from the SFANet to make them more applicable to underwater image enhancement.

Effectiveness of Multi-path CFRM. To show the effectiveness of the multi-path Contextual Feature Refinement Module (CFRM), we remove it as our comparison model. As shown in Table VI, an improvement of 0.879dB is obtained using CFRM, suggesting that the feature refinement module can highlight more informative features by modeling the correlation between features.

E. Additional Analysis

Encoder embedding *vs.* Decoder embedding. We conduct feature embedding investigations on the encoder and de-

(a) w/o structure branch   (b) w/ structure branch

Fig. 9. Ablation analysis of the structure branch. (a) and (b) are structure layers through gaussian filtering corresponding to the enhanced results without and with structure branch, respectively.

TABLE VI

INVESTIGATION OF THE MULTI-PATH CFRM ON THE UIEB DATASET.

	w/o CFRM	w/ CFRM
PSNR	21.569	$22.448_{\uparrow 0.879}$
SSIM	0.8902	$0.9019_{\uparrow 0.012}$

coder of the UNet-like base network, respectively. As shown in Table VII, we note that the performance of the proposed algorithm is improved by a large margin by embedding the aggregated features from the SFANet into the decoder rather than the encoder, on the UIEB and EUVP datasets. It indicates that embedding features from the semantic-aware model into the decoder of the base network for feature reconstruction can maximize feature utilization and achieve better enhancement performance.

TABLE VII

ANALYSIS OF THE FEATURE EMBEDDING LOCATION.

Dataset	Encoder		Decoder	
	PNSR	SSIM	PNSR	SSIM
UIEB [15]	21.703	0.8836	$22.448_{\uparrow 0.745}$	0.9019
EUVP [9]	22.858	0.9007	$23.079_{\uparrow 0.221}$	0.9033

IV. CONCLUSIONS

In this paper, we proposed an efficient underwater image enhancement algorithm based on semantic-aware texture and structure feature collaboration. The main novel points of the proposed algorithm lie in that the hierarchical features of the high-level semantic-aware model were exploited as an auxiliary and were characterized via structure and texture branches. A multi-path contextual feature refinement module was deployed on both branches to model the correlation between features resulting in near-lossless image content. We presented a feature dominative network to perform channel-wise feature modulation to adapt to the different feature patterns of the enhancement network. Experiments implemented on four widely-used underwater benchmarks demonstrated the superiority of our algorithm and its semantic-aware ability for high-level vision tasks. In future work, we plan to explore the potential of the proposed algorithm in the field of underwater machine vision, such as underwater robotics and underwater object detection.

REFERENCES

[1] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 81–88.

[2] D. Cagara, M. Dunbabin, and P. Rigby, "A feature-based underwater path planning approach using multiple perspective prior maps," in *IEEE International Conference on Robotics and Automation*, 2020, pp. 8573–8579.

[3] K. De Langis and J. Sattar, "Realtime multi-diver tracking and re-identification for underwater human-robot collaboration," in *IEEE International Conference on Robotics and Automation*, 2020, pp. 11 140–11 146.

[4] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and F. Li, "Imagenet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[5] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *IEEE International Conference on Robotics and Automation*, 2018, pp. 7159–7165.

[6] X. Fu, Z. Fan, M. Ling, Y. Huang, and X. Ding, "Two-step approach for single underwater image enhancement," in *International Symposium on Intelligent Signal Processing and Communication Systems*, 2017, pp. 789–794.

[7] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X. S. Zhang, and X. Ding, "A retinex-based enhancing approach for single underwater image," in *IEEE International Conference on Image Processing*, 2014, pp. 4572–4576.

[8] M. J. Islam, R. Wang, K. de Langis, and J. Sattar, "Svam: Saliency-guided visual attention modeling by autonomous underwater robots," *arXiv preprint arXiv:2011.06252*, 2020.

[9] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robotics Automation Letters*, vol. 5, no. 2, pp. 3227–3234, 2020.

[10] P. D. Jr., E. R. Nascimento, S. S. C. Botelho, and M. F. M. Campos, "Underwater depth estimation and image restoration based on single images," *IEEE Computer Graphics and Applications*, vol. 36, no. 2, pp. 24–35, 2016.

[11] W. Jun, W. Shuhui, and H. Qingming, "F³net: Fusion, feedback and focus for salient object detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, 2020, pp. 12 321–12 328.

[12] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," Y. Bengio and Y. LeCun, Eds., 2015.

[13] C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, and W. Ren, "Underwater image enhancement via medium transmission-guided multi-color space embedding," *IEEE Transactions on Image Processing*, vol. 30, pp. 4985–5000, 2021.

[14] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognition*, vol. 98, 2020.

[15] 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, 2020.

[16] C. Li, J. Guo, R. Cong, Y. Pang, and B. Wang, "Underwater image enhancement by dehazing with minimum information loss and histogram distribution prior," *IEEE Transactions on Image Processing*, vol. 25, no. 12, pp. 5664–5677, 2016.

[17] C. Li, J. Guo, C. Guo, R. Cong, and J. Gong, "A hybrid method for underwater image correction," *Pattern Recognition Letter*, vol. 94, pp. 62–67, 2017.

[19] R. Liu, X. Fan, M. Zhu, M. Hou, and Z. Luo, "Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 12, pp. 4861–4875, 2020.

[20] B. McGlamery, "A computer model for underwater camera systems," in *Ocean Optics VI*, 1980, pp. 221–231.

[21] J. McMahon and E. Plaku, "Autonomous data collection with timed communication constraints for unmanned underwater vehicles," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1832–1839, 2021.

[22] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.

[23] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541–551, 2016.

[24] Y. Peng, K. Cao, and P. C. Cosman, "Generalization of the dark channel prior for single image restoration," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2856–2868, 2018.

[25] Y. Peng, X. Zhao, and P. C. Cosman, "Single underwater image enhancement using depth estimation based on blurriness," in *IEEE International Conference on Image Processing*, 2015, pp. 4952–4956.

[26] W. Ren, J. Zhang, X. Xu, L. Ma, X. Cao, G. Meng, and W. Liu, "Deep video dehazing with semantic segmentation," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1895–1908, 2019.

[27] Z. Shen, W. Lai, T. Xu, J. Kautz, and M. Yang, "Deep semantic face deblurring," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8260–8269.

[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, Y. Bengio and Y. LeCun, Eds., 2015.

[29] H. Tang, Z. Li, Z. Peng, and J. Tang, "Blockmix: meta regularization and self-calibrated inference for metric-based meta-learning," in *ACM International Conference on Multimedia*, 2020, pp. 610–618.

[30] S. Tian, H. Tang, and L. Dai, "Coupled patch similarity network for one-shot fine-grained image recognition," in *IEEE International Conference on Image Processing*, 2021, pp. 2478–2482.

[31] P. M. Uplavikar, Z. Wu, and Z. Wang, "All-in-one underwater image enhancement using domain-adversarial learning," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 1–8.

[32] D. Wang, H. Tang, J. Pan, and J. Tang, "Learning a tree-structured channel-wise refinement network for efficient image deraining," in *IEEE International Conference on Multimedia and Expo*, 2021, pp. 1–6.

[33] T. Wang, M. Liu, J. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8798–8807.

[34] X. Wang, K. Yu, C. Dong, and C. C. Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 606–615.

[35] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.