RFD-ECNet: Extreme Underwater Image Compression with Reference to Feature Dictionary

This supplementary material is organized as follows:

- *Sec. A shows the effect of the cluster number K on our compression performance.*
- *Sec. B provides some visual examples of underwater priors of UWIs.*
- *Sec. C provides additional experimental results on underwater videos.*
- *Sec. D further validates the comprehensiveness of our dictionary.*
- *Sec. E provides the preliminary experimental results of our framework expanded to Martian images.*
- *Sec. F provides the visualization of the intermediate features in our RFD-ECNet.*
- *Sec. G provides the detailed network architecture of our RFD-ECNet.*

A. Effect of K on the compression performance

Considering the compactness of our underwater dictionary, we adopt K-means to select K centrist features for the dictionary. To verify the effect of K value on compression performance, we conduct multiple experiments of different K values, where K is set to $\{0, 400, 800, 1200, 1600\}$. The BD-rate at each K value is shown in Table 5. Within the certain range ($K \leq 1200$), the compression performance gets better as K increases. When K increases from 1200 to 1600, the improvement of BD-rate saving is small, indicating that the number of features in dictionary reaches saturation. Since the model($K = 1200$) performs almost on par with model($K = 1600$) but with a smaller dictionary, K is set to 1200 in our RFD-ECNet.

Table 5: BD-rate (↓) at different K values.

K values	0	400	800	1200	1600
BD-rate ↓	0%	-16%	-26%	-46%	-47%

B. Visual examples of underwater priors

In Fig. 9, we provide some visual examples of underwater priors of different UWIs. As shown, different UWIs present multifarious underwater styles of color shift and distance-dependent clarity, which are caused by the unique underwater imaging. In practice, these underwater styles can be described by underwater physical priors extracted from the underwater physical imaging model, *i.e.*, the ambient light and transmission map.

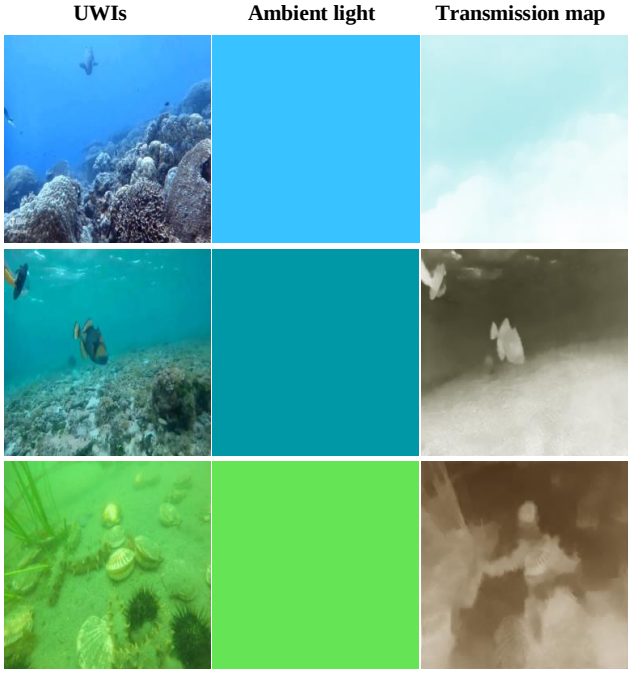

Figure 9: Visual examples of UWIs and their underwater physical priors of ambient light and transmission map.

C. Performance on underwater video

This section tests the performance of our RFD-ECNet on underwater video sequences. Some visual comparisons between RFD-ECNet with other SOTA image compression methods on underwater videos are provided in https://github.com/lilala0/RFD-ECNet. It can be clearly seen that our RFD-ECNet also achieves better video compression performance than other SOTA image compression methods.

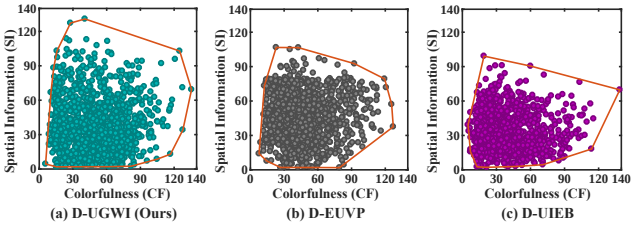

Figure 10: Diversity comparison of different dictionary. D-UGWI, D-EUVP and D-UIEB denote dictionaries built by UGWI, EUVP and UIEB datasets.

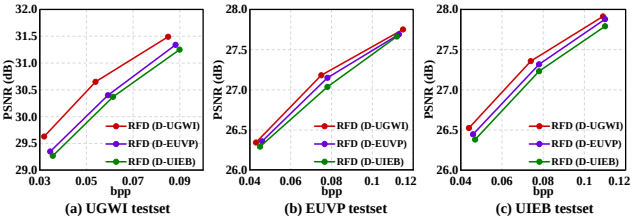

Figure 11: R-D performance of RFD-ECNet using different dictionaries of D-UGWI, D-EUVP and D-UIEB.

D. Validation of dictionary comprehensiveness

As demonstrated in Sec 3.1, the diversity of our dictionary has been verified by analyzing the spatial information and colorfulness. This part further compares our dictionary with two other dictionaries built from popular UWI datasets including EUVP and UIEB, to validate the diversity and efficiency of our dictionary. First, we compare the diversity of three dictionary by colorfulness and spatial information. As shown in Fig. 10, our dictionary has wider CF-SI distribution, indicating ours higher diversity. After that, the performance of RFD-ECNet using different dictionaries is test on three UWI testsets shown in Fig. 11. It can be clearly seen that our dictionary (D-UGWI) brings better performance than two other dictionaries built from EUVP and UIEB, indicating the efficiency of our dictionary.

E. Performance on Martian images [36]

As discussed in the main paper (Sec. 6), it is feasible to expand our framework to other image domains because the redundancy between images exists not only in UWIs but also in other image domains where independent images contain some common objects specific to the image domain. To validate this hypothesis, we preliminarily conduct experiments on Martian images [36] by adjusting our network, and compare with the most advanced VVC, the widely used BPG, and the martian image compression network VCIP'22 [36]. The comparison of R-D curves is shown in Fig. 12. As shown, our method still achieves significantly better compression performance than the other methods, illustrating the efficiency, feasibility, and great potential of removing redundancy between images in other image domains.

F. Visualization of the intermediate features

To illustrate how our RFD-ECNet works, we visualize the intermediate features in our RFD-ECNet in Fig. 13. For a clear view, we only depict the channel with the highest entropy. Specifically, the 1-st and 2-nd columns, denoted as (a) and (b), respectively are the input UWI and the features of input. Columns (c) and (d) respectively are the reference feature matched by out USN-FMM module and the variant reference feature morphed by our RFVM. As shown, the reference feature matched by our USN-FMM has certain similarity with the input feature. Moreover, after being mor-

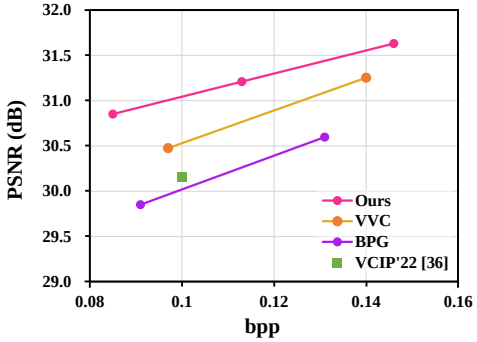

Figure 12: R-D curves of different image compression methods on Martian images.

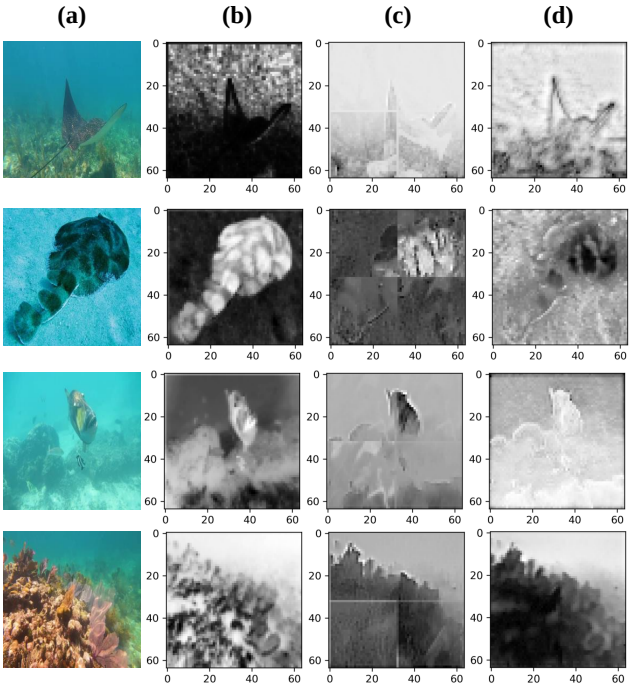

Figure 13: Visualization of the intermediate features in our RFD-ECNet. (a) Input UWI. (b) Input feature of the input UWI. (c) Reference feature matched by our USN-FMM module. (d) Variant reference feature after our RFVM.

phed by our RFVM, the variant reference feature has higher similarity with the input feature. These visualized features show that our approach does match the reference features with similarity to the input and further improve their similarity, which directly results to a sparser latent feature of removing redundancy between UWIs and fewer coding bits.

G. Detailed architecture of our RFD-ECNet

The overall framework of our RFD-ECNet has been provided in the main paper. Here, we provide the detailed network architectures of our encoder, USN-FMM block, RFVM block, and decoder, which are respectively presented in Table 6, Table 7, Table 8, Table 9.

Table 6: Detailed network structure of our encoder

Input	UWI
layer 1	Conv [$3 \times 192 \times 3 \times 3$, s=2], GDN
layer 2-1	Conv [$192 \times 192 \times 3 \times 3$, s=2], GDN
layer 2-2	USN-FMM block
layer 2-3	RFVM block
layer 2-4	Residual operation
layer 3-1	Conv [$192 \times 192 \times 3 \times 3$, s=2], GDN
layer 3-2	USN-FMM block
layer 3-3	RFVM block
layer 3-4	Residual operation
layer 4-1	Conv [$192 \times 192 \times 3 \times 3$, s=2], GDN
layer 4-2	USN-FMM block
layer 4-3	RFVM block
layer 4-4	Residual operation
layer 5	Conv [$192 \times 64 \times 3 \times 3$, s=2]
Output	Latent feature

Conv [input channels $\times$ output channels $\times$ kernel size, stride] indicates the convolution layer. GDN is the normalization layer [12]

Table 7: Detailed network structure of our USN-FMM

Input	Underwater priors (UPPs)
layer 1	Conv [$3 \times 192 \times 3 \times 3$, s=2], LRelu
layer 2	Conv [$192 \times 192 \times 3 \times 3$, s=2], LRelu
layer 3	Conv [$192 \times 1 \times 3 \times 3$, s=1]
Output	UPPs feature
$\Downarrow$	
Input	UPPs feature and our Dictionary (D)
layer 1	Linear flow in Eq. (4)
Output	Normalized Dictionary (ND)
$\Downarrow$	
Input	Input feature and ND
layer 1	Calculation of similarity by Eq. (5)
Output	Matched reference feature

Table 8: Detailed network structure of our RFVM

Input	Reference and input features
layer 1	Concatenation
layer 2	Conv [$384 \times 192 \times 3 \times 3$], GDN
layer 3-1	Conv [$192 \times 192 \times 3 \times 3$, dilation=1], GDN
layer 3-2	Conv [$192 \times 192 \times 3 \times 3$, dilation=2], GDN
layer 3-2	Conv [$192 \times 192 \times 3 \times 3$, dilation=3], GDN
layer 4	Concatenation
layer 5	Conv [$576 \times 192 \times 3 \times 3$], GDN
layer 6	Residual layer
layer 7	Conv [$192 \times 4 \times 3 \times 3$], Sigmoid
Output	Dependency map (W)
$\Downarrow$	
Input	Reference feature and W
layer 1-1	SVRConv [34]
layer 1-2	Conv [$192 \times 192 \times 3 \times 3$], GDN
layer 2-1	SVRConv
layer 2-2	Conv [$192 \times 192 \times 3 \times 3$], GDN
layer 3-1	SVRConv
layer 3-2	Conv [$192 \times 192 \times 3 \times 3$], GDN
layer 4-1	SVRConv
layer 4-2	Conv [$192 \times 192 \times 3 \times 3$], GDN
Output	Variant reference feature

Table 9: Detailed network structure of our decoder

Input	Latent feature, W, and indexes
layer 1-1	DeConv [$64 \times 192 \times 3 \times 3$, s=2], GDN
layer 1-2	RFVM
layer 1-3	Addition operation
layer 1-4	Residual blocks (192) $\times$ 2
layer 2-1	DeConv [$192 \times 192 \times 3 \times 3$, s=2], GDN
layer 2-2	RFVM
layer 2-3	Addition operation
layer 2-4	Residual blocks (192) $\times$ 2
layer 3-1	DeConv [$192 \times 192 \times 3 \times 3$, s=2], GDN
layer 3-2	RFVM
layer 3-3	Addition operation
layer 3-4	Residual blocks (192) $\times$ 2
layer 4	DeConv [$192 \times 192 \times 3 \times 3$, s=2], GDN
layer 4-2	Residual blocks (192) $\times$ 2
layer 5	DeConv [$192 \times 3 \times 3 \times 3$, s=2]
Output	Decoded UWI

DeConv [input channels $\times$ output channels $\times$ kernel size, stride] indicates the nn.ConvTranspose2d layer. Residual blocks (192) indicate the channel number of the residual block is 192.