Five A^{+} Network: You Only Need 9K Parameters for Underwater Image Enhancement

Jingxia Jiang[1†]
202021114006@jmu.edu.cn
Tian Ye[2†]
owentianye@hkust-gz.edu.cn
Jinbin Bai[3†]
jinbin.bai@u.nus.edu
Sixiang Chen[2]
sixiangchen@hkust-gz.edu.cn
Wenhao Chai[4]
reselfchai@gmail.com
Shi Jun[5]
junshi2022@gmail.com
Yun Liu[6]
yunliu@swu.edu.cn
Erkang Chen[1,7*]
ekchen@jmu.edu.cn

[1] School of Ocean Information Engineering, Jimei University, Xiamen, China
[2] Hong Kong University of Science and Technology (GZ), Guangzhou, China
[3] Department of Computer Science, National University of Singapore, Singapore
[4] University of Washington, State of Washington, USA
[5] the School of Information Science and Engineering, Xinjiang University, Uruqmi, China
[6] College of Artificial Intelligence, Southwest University, Chongqing, China
[7] Fujian Provincial Key Laboratory of Oceanic Information Perception and Intelligent Processing, China

Abstract

A lightweight underwater image enhancement network is of great significance for resource-constrained platforms, but balancing model size, computational efficiency, and enhancement performance has proven difficult for previous approaches. In this work, we propose the Five A^{+} Network (FA^{+}Net), a highly efficient and lightweight real-time underwater image enhancement network with only $\sim$ **9k** parameters and $\sim$ **0.01s** processing time. The FA^{+}Net employs a two-stage enhancement structure. The powerful prior stage aims to decompose challenging underwater degradations into sub-problems, while the fine-grained stage incorporates multi-branch color enhancement module and pixel attention module to amplify the network's perception of details. To the best of our knowledge, FA^{+}Net is the only network with the capability of real-time enhancement of 1080P images. Through extensive experiments and comprehensive visual comparisons, we show

* :Corresponding Author. † :Equal Contribution.

that FA$^+$Net outperforms previous approaches by obtaining state-of-the-art performance on multiple datasets while significantly reducing both the number of parameters and computational complexity. The code is available at https://github.com/Owen718/FiveAPlus-Network.

1 Introduction

Underwater images are often plagued by severe blurring and color distortion, making it difficult to meet the demands of practical applications. With the rise of underwater archaeologies [15, 58] and marine ecological researches [27, 34, 49], some researchers have begun to explore how to embed underwater image enhancement algorithms into platforms such as underwater robots. However, due to the limited resources of underwater robots, underwater cameras, and other equipment, traditionally gigantic learning-based models are challenging to achieve efficient enhancement on these platforms [18, 29, 31, 60].

One potential solution to this issue is to design lightweight networks with fewer parameters and computations. For example, the Shallow-uwnet [37] constructed by introducing lightweight network components and residual convolution blocks. However, this purely resource-driven approach are not necessarily result in lower computational complexity. Additionally, no specific design has been proposed to target certain degradation phenomena in underwater enhancement tasks, resulting in unsatisfactory visual effects and performance metrics for the restored images.

Constructing a real-time underwater image enhancement framework that simultaneously possesses ultra-lightweight parameters and powerful enhancement ability has been a longstanding challenge in the field.

To overcome this longstanding challenge, we decompose the underwater degradations into sub-problems based on the characteristics of the Underwater Image Enhancement (UIE) task and design a lightweight and embedded real-time UIE network called the Five A$^+$ Network. The FA$^+$ symbolizes that our network achieves superb performance in terms of PSNR, SSIM, FPS, GFLOPs and Parameters. As demonstrated by Fig. 1, FA$^+$Net achieves state-of-the-art performance while saliently reducing both the number of parameters and computational complexity by an order of magnitude compared to previous methods by 10-100 times, with a total number of parameters of less than 9K.

To reduce model complexity, some computationally expensive operators and operations such as large-kernel convolutions [8, 10, 13] and self-attention [3, 4, 5, 43, 50, 62] were discarded while channel dimensions are precisely restricted to control the number of parameters. For UIE, the problem can be addressed by breaking it down into sub-problems, effectively solving the mixed degradation problem by separately correcting color distortion and restoring details of the degraded image. Therefore, two complementary components are proposed: Multi-Branch Color Enhancement Module(MCEM) and Multi-scale Pyramid Module (MPM) in the strong prior stage. MCEM is an effective module for serious color distortion of underwater images. and MPM enables processing of input feature maps at multiple scales, thus capturing detail information at varying scales to enhance the model's perception of image details. In this way, our strong prior-based designs endow the network with highly effective restoring capabilities for underwater degradations.

Recently, some researchers have proposed the separation of global background light and texture in the Fourier domain [32, 47]. Specifically, global background light is represented by amplitude, while texture is intertwined with phase. By separating them in the Fourier domain, Gaussian noise can be avoided when enhancing color, whilst providing abundant

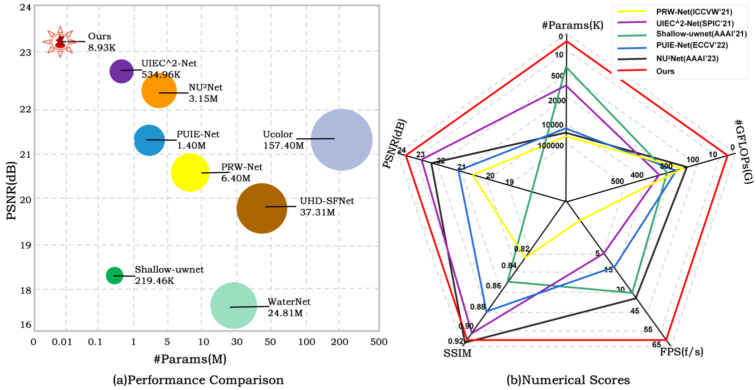

Figure 1: Comparison of recent state-of-the-art methods and our method: We report the computational efficiency (#*Params*, GFLOPs, and FPS) and numerical scores for two types of resto ration quality measurement metrics including PSNR and SSIM, it can be easily observed that our method is remarkably superior to others.

global information. To better extract valuable feature information from different components' outputs and capture global contextual information, we design a Spatial-frequency Domain Feature Interaction Module(SDFIM). We utilize adjustable hyperparameter α to control the fusion of spatial domain and frequency domain information. Moreover, Fast Fourier Convolution (FFC) [6] is adopted to enlarge the receptive field of our network to entire resolution, significantly amplifying the network's perception ability. Although these operations have shown basic success in addressing the mixed degradation issue, the complex underwater environments are often impacted by multiple factors causing some challenging detail problems. In these cases, typical single-stage networks may struggle to accurately capture tiny objects, intricate colors, and textures. To further enhance the model's performance, we introduce a fine-grained stage for more in-depth image analysis, aiming to better manage these intricate detail issues. MCEM and Pixel Attention module [42] are incorporated to assist the model in comprehending each image element and detail more effectively, thereby improving the model's performance and generalization capability. By introducing the fine-grained stage, substantial progress is made in the model's ability to tackle complex underwater images, as it is better equipped for handling intricate detail issues.

The combination of these two stages not only proposes novel design ideas for underwater image enhancement, but also expands the horizons of potential research in this field. Notably, the ultra-lightweight parameters allow FA$^+$Net to be embedded into edge devices, and we are the only ones able to enhance 1080P images in real time on RTX 3090. Our model also has high throughput, allowing for faster inference and processing of input data to meet the requirements of mobile platforms such as underwater filming rigs and robotic platforms.

The main contributions of this paper are as follows:

- We introduce FA$^+$Net, that reduces the the number of parameters of an enhancement model to 8.9K, which is approximately $10-100\times$ fewer than previous methods.

- We propose a two-stage architecture that provides novel designs and directions for image enhancement. The strong prior stage decomposes mixed degradation into subproblems, while the fine-grained stage focuses on enhancing the network's perception of intricate details.

- FA^{+}Net is the only model capable of real-time enhancement for 1080P images, efficiently running on an RTX 3090 GPU. It demonstrates exceptional performance across multiple datasets, making it a viable choice for deployment on mobile platforms.

2 Related Work

2.1 Learning-based Underwater Image Enhancement Method

With the successful application of deep learning in high-level computer vision tasks [65], an increasing number of researchers have begun to apply it to low-level computer vision tasks [23, 24, 25, 26, 59, 61], such as underwater image enhancement [12, 20, 30, 60]. For instance, Jiang et al. [22] designed a novel domain adaptation framework based on transfer learning to transform aerial image deblurring into realistic underwater image enhancement. Despite their varying degrees of success in terms of performance metrics, these approaches fail to incorporate dedicated modules for addressing color shift and texture loss of degraded images. Li et al. [31]presented an underwater image enhancement network via medium transmission-guided multi-color space embedding, named Ucolor. Huo et al. [18] employed wavelet-enhanced learning units to decompose hierarchical features into high-frequency and low-frequency components, and then strengthen them with normalization and attention mechanisms. Although this approach has shown excellent visual effects, its extensive network parameters (6.30M) and computational requirements (223.37G) make it unsuitable for existing underwater devices. Moreover, it cannot effectively address the issue of color distortion.

2.2 Efficient Neural Network For Image Restoration

Efficient neural network for image restoration [7, 55, 59] is a recent development in deep learning-based image restoration and has been demonstrated to achieve state-of-the-art performance while requiring fewer computational resources than other methods. For example, Song et al. [45] proposed an efficient residual dense block search algorithm with multiple objectives to identify fast, lightweight, and accurate networks for image super-resolution. Guo et al. [14] presented an effective low-light image enhancement method (LIME) that estimated the illumination of each pixel individually and refined it using a structure prior. Naik et al. [37] proposed a lightweight underwater enhancement framework by introducing lightweight components and residual blocks.

2.3 Fast Fourier Convolution

In order to address the low efficacy in connecting two distant locations in the network. Chi et al. [6] proposed a novel convolutional operator dubbed as Fast Fourier Convolution (FFC), which has the characteristics of non-local receptive fields and cross-scale fusion within the convolutional unit. Furthermore, modern image inpainting systems commonly struggle with large missing areas, complex geometric structures, and high-resolution images. To alleviate this issue, Suvorov et al. [47] proposed a new method termed large mask inpainting that is based on a new inpainting network architecture relying on FFCs. When dealing with the challenging task of joint luminance enhancement and noise removal whilst remaining efficient. Li et al. [32] devised a new solution, UHDFour, which differs from existing approaches that take a spatial domain-oriented approach. Specifically, UHDFour is motivated by a few unique characteristics of the Fourier domain, such as the fact that most luminance information is concentrated in amplitudes while noise is closely related to phases.

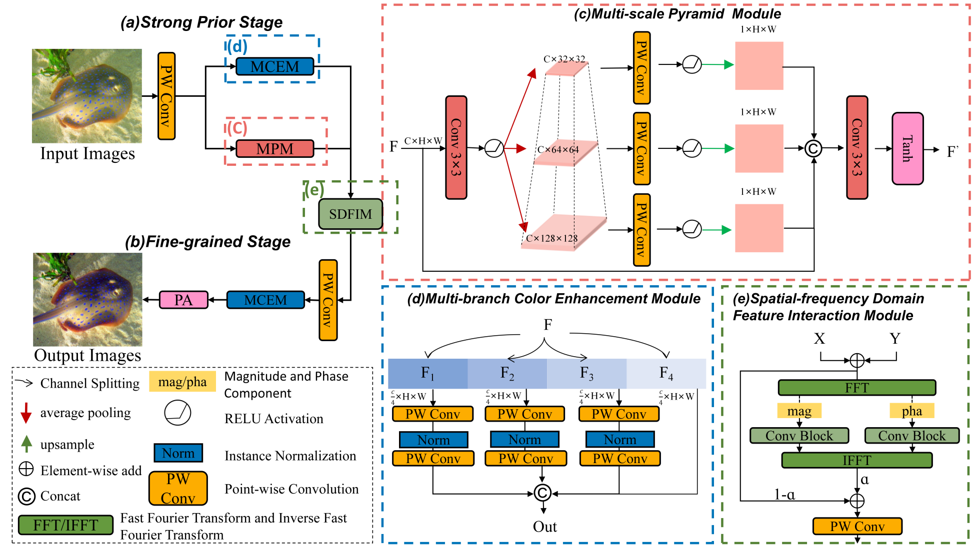

Figure 2: The overall architecture of Five A$^+$ Network: FA$^+$Net is composed of a strong prior stage and a fine-grained stage, augmented by the efficient SDFIM. The core components of the network comprise: (c) MPM captures granular details across various scales, endowing the model with potent detail perception; (d) MCEM perform consecutive processing of individual image pixels, thereby enabling our network to achieve precise color restoration; and (e) SDFIM aids the network in sifting valuable feature information from the outputs of diverse components and acquiring global contextual features, and α is a hyperparameter that controls the fusion ratio of spatial-frequency domain information.

3 FA$^+$Net: An Ultra-lightweight Real-time Enhancement Network

3.1 Motivation

Limited computing resources on embedded platforms, such as underwater robots, have posed a significant challenge in achieving high-quality image enhancement using traditional deep learning models. Consequently, recent methods [7, 14, 45] have prompted the development of lightweight yet powerful models. In this context, FA$^+$Net emerges as a noteworthy contribution, which has been demonstrated to be an efficient and innovative solution, as illustrated in Fig. 2.

To ensure the computational efficiency of our model, we initially removed several computationally expensive operators and operations, such as large kernel convolutions and self-attention mechanisms. We also imposed constraints on the channel dimension to ensure precise parameter control. Furthermore, to effectively address the mixed degradation challenge, we adopt a divide-and-conquer strategy to separately enhance color and restore details from degraded images. Additionally, to improve our model's overall performance, we introduced a fine-grained stage for comprehensive image analysis. In combination, our approach allows for effective color enhancement and detail restoration, even under extreme underwater conditions.

3.2 Model Structure

3.2.1 Multi-Scale Pyramid Module

To recover fine details in degraded underwater images, we propose Multi-scale Pyramid Module(MPM) in the strong prior stage. By downsampling the input image to different sizes, the network can capture features at multiple scales and resolutions, which is critical for improving the appearance of objects with different sizes and shapes in challenging underwater scenarios. To achieve real-time performance, we designed the MPM as a three-branch structure with down-sampled target size of 32×32, 64×64, and 128×128. The selection of this structure is based on a series of careful ablation experiments reported in supplementary material, which ensured a good trade-off between performance and effectiveness.

3.2.2 Multi-branch Color Enhancement Module

The attenuation rates of different wavelengths of light in underwater environments vary, with red light experiencing the fastest attenuation and blue and green light experiencing the slowest [21]. This results in conspicuous differences in the R, G, and B channels, leading to poor contrast and color distortion in underwater images, which has been a largely unaddressed issue in previous methods [1, 19, 31, 51].

To overcome this limitation, we propose the MCEM, which employs a branch enhancement strategy to better capture the color feature distribution across R, G, and B channels. Each pixel carries color information, and the 1×1 convolutional operations in MCEM can be viewed as enhancing the color information on a per-pixel basis. Given the varying color emphasis in feature extraction across different channels, MCEM specifically utilizes instance normalization to normalize each channel separately, thereby mitigating the potential risk of color border blurring associated with batch normalization.This approach is similar to the underlying operation of a multi-layer perception [48], allowing our network to achieve accurate color reproduction, which is particularly crucial for color-sensitive underwater image enhancement tasks. Additionally, we opt not to use 3×3 convolutions due to their increased parameter burden. As shown in Fig. 2(d), weights are not shared between each branch. The effectiveness of this module is demonstrated in the supplementary material.

3.2.3 Spatial-frequency Domain Feature Interaction Module

Recent studies have shown that global background lighting and textures in underwater images can be partially decomposed in the Fourier domain, as evidenced by recent works such as [6, 17, 32, 47, 64]. However, current methods for restoring degraded images mostly rely on spatial domain processing, and traditional convolutional approaches tend to overlook the rich global information present in the Fourier domain. To address this issue, we propose the cross-domain design component called Spatial-Frequency Domain Interaction Module. By fusing feature information in the Fourier domain, SDFIM achieves receptive field coverage of the entire image, which improves the network's perceptual quality and parameter efficiency. The hyperparameter α in SDFIM controls the fusion ratio of spatial-frequency domain information, and its varying values generate different visual effects. Furthermore, the induction bias of FFC enhances the network's generalization performance, thereby reducing the requirements for extensive training data and computation.

The key operations of SDFIM are as follows, given the features $X \in R^{C \times H \times W}$ from MCEM and $Y \in R^{C \times H \times W}$ from MPM:

$$F' = X + Y \tag{1}$$

$$F_{MAG}, F_{PHA} = \mathrm{f}_{FFT}(F') \tag{2}$$

$$F_{OUT} = \alpha[f_{IFFT}(f_{FC}(F_{MAG}), f_{FC}(F_{PHA}))] + (1-\alpha)F' \tag{3}$$

where F_{MAG} and F_{PHA} represent the magnitude component and phase component of the feature, respectively. $\mathrm{f}_{FFT}(\cdot)$ denotes the fast Fourier transform, $f_{FC}(\cdot)$ represents the Fourier domain convolution operation, and $f_{IFFT}(\cdot)$ denotes the inverse fast Fourier transform. The hyperparameter α controls the fusion ratio of spatial-frequency domain information.

3.3 Loss Function

We introduce the Charbonnier Loss [2] as our basic reconstruction loss:

$$\mathcal{L}_{rec} = \mathcal{L}_c(\mathcal{I}(X), \mathcal{J}_{gt}) \tag{4}$$

where the $\mathcal{I}(\cdot)$ is our FA$^+$ Net, X and $\mathcal{J}_{gt}$ stand for the input and the corresponding ground-truth, respectively. $\mathcal{L}_c$ denotes the Charbonnier loss, which can be express as:

$$\mathcal{L}_c = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\|X^i - Y^i\|^2 + \varepsilon^2} \tag{5}$$

where constant ε is empirically set to 1×10^{-3} for all experiments of ours. In addition, the perceptual level of the restored image is also critical. We apply a perceptual loss to improve the restoration performance. To further enhance the restoration of degraded images by preserving their intrinsic texture, we also introduce perceptual loss at the strong prior stage, which is enforced as a form of deep supervision [16, 52, 57]. The perceptual loss can be formulated as follows:

$$\mathcal{L}_{perceptual} = \sum_{j=1}^{2}\frac{1}{C_j H_j W_j}||\phi_j(\mathcal{I}(x)) - \phi_j(\mathcal{Y})||_1 \tag{6}$$

where in the ϕ_j represents the 1-th and the 3-th layers of VGG19 [44]. C_j , H_j , W_j represent the channel number, height, and width of the feature map, respectively.

Many underwater image enhancement tasks use $L2$ loss for training. As shown by [30, 36, 41], the $L2$ loss produces over-smoothed backgrounds and ghost artifacts, which is detrimental to the semantic information. In order to better reflect the human visual system's perception of image quality, we adopt negative SSIM loss to focus on luminance, contrast, and structure. The negative SSIM loss is:

$$\mathcal{L}_{ssim} = -SSIM(\mathcal{I}(X), \mathcal{J}_{gt}) \tag{7}$$

Where the $\mathcal{I}(\cdot)$ is our Five A$^+$ Net, X and $\mathcal{J}_{gt}$ stand for the input and the corresponding ground-truth, respectively.

Overall loss function can be expressed as:

$$\mathcal{L} = \lambda_1\mathcal{L}_c + \lambda_2\mathcal{L}_{perceptual} + \lambda_3\mathcal{L}_{ssim} + \lambda_4\mathcal{L}_{perceptual_{SPS}} \tag{8}$$

where the λ_1, λ_2, λ_3 and λ_4 are set to 1, 0.2, 0.5 and 0.2, respectively.

4 Experiments

4.1 Experiment details

All experiments are implemented using the PyTorch [39] framework with a single NVIDIA A100 Tensor Core GPU (40GB). During training, the training epochs are set to 400, and the total batch size is 100. We use Adam optimizer as the optimization algorithm. The learning rate is set to 4×10^{-4} at first, and the default values of β_1 and β_2 are 0.9 and 0.999, respectively. We used CyclicLR to adjust the learning rate, with an initial momentum of 0.9

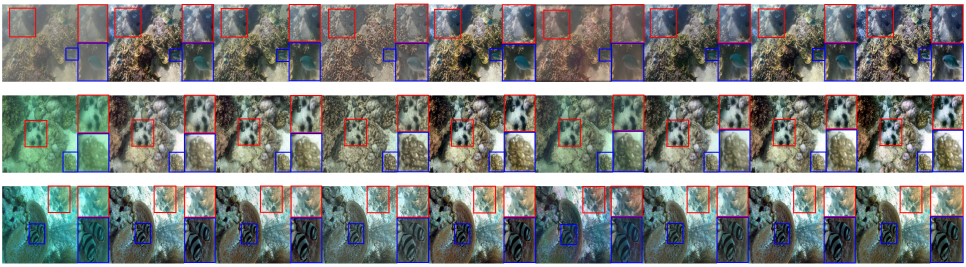

Figure 3: Visual comparison of UIE networks on T90.

and 0.999. Data augmentation included horizontal flipping, and randomly rotating the image to $90, 180$, and 270 degrees.

During the training process, the input data was randomly cropped as 256×256 patches from original images. Underwater Image Enhancement Benchmark (UIEB) datasets [29] contains 890 high-resolution raw underwater images and corresponding high-quality reference images, and 60 challenge images (C60) for which no corresponding reference images were obtained. Despite consisting of only 890 high-quality images, it is still a well regarded benchmark dataset for its superb qualities, receiving broad acceptance and appreciation within the research community. Li et al. carefully selected 45 authentic underwater images, named U45 [33]. It is partitioned into three subsets according to the color cast of underwater degradation, low contrast, and blur effects: green, blue, and haze. Then, 800 pairs of original images and clear images were extracted from UIEB to train the model. The remaining 90 images in UIEB named T90 were used to test the effect of our method on degraded images. In order to evaluate the generalization performance of FA$^+$Net, we used the C60 and U45 datasets for testing.

4.2 Evaluation metrics

In order to acquire quantitative measurements, we use Peak Signal-to-Noise Ratio (PSNR) [28], Structural Similarity Index (SSIM) [54], the Mean Squared Error (MSE) [35], Underwater Color Image Quality Evaluation (UCIQE) [56], and Underwater Image Quality Metric (UIQM) [38] as performance metrics for image quality. PSNR is a full-reference image quality evaluation metric based on errors between corresponding pixels. The higher the PSNR score, the better the image quality. SSIM measures the visual quality of three features of an image: brightness, contrast, and structure. A higher SSIM value indicates a higher similarity between the enhanced and reference images. UCIQE mainly measures the degree of detail and color recovery of distorted images. UCIQE is one of the most comprehensive image evaluation standards. UIQM is required to evaluate color, sharpness, and contrast.

4.3 Comparison with SOTA methods

We compared FA$^+$Net with several state-of-the-art methods, including traditional methods and deep learning methods. Traditional methods included UDCP [9], IBLA [40], SMBL [46] and MLLE [63], and deep learning methods included UWCNN [30], Water-Net [29], PRW-Net [18], Shallow-uwnet [37], Ucolor [31], UIEC^2-Net [53], UHD-SFNet [55], PUIE-Net [11] and the latest NU2Net [12] for underwater image enhancement. We present the objective metrics comparison with previous SOTA methods in Table 1. From that, we can observe that our method achieves the best results on PSNR, MSE and UCIQE metrics, proving that

Table 1: Experimental results on T90 [29], C60 [29], and U45 [33] datasets, the best and second-best results are marked in red and blue respectively. ↑ represents the higher is the better as well as ↓ represents the lower is the better. The efficiency evaluation uses 720P images as input on RTX 3090.

	T90					C60		U45		Efficiency			
Methods	PSNR↑	SSIM↑	MSE↓	UCIQE↑	UIQM↑	UCIQE↑	UIQM↑	UCIQE↑	UIQM↑	GFLOPs(G)↓	#Params(M)↓	#Runtime(s)↓	FPS(f/s)↑
UDCP(ICCVW'13)[[illegible]]	13.415	0.749	0.228	0.572	2.755	0.560	1.859	0.574	2.275	-	-	42.13s	-
IBLA(TIP'17)[[illegible]]	18.054	0.808	0.142	0.582	2.557	0.584	1.662	0.565	2.387	-	-	-	-
WaterNet(TIP'19)[[illegible]]	16.305	0.797	0.161	0.564	2.916	0.550	2.113	0.576	2.957	193.70G	24.81M	0.680s	-
SMBL(TB'20)[[illegible]]	16.681	0.801	0.158	0.589	2.598	0.571	1.643	0.571	2.387	-	-	-	-
UWCNN(PR'20)[[illegible]]	17.949	0.847	0.221	0.517	3.011	0.492	2.222	0.527	3.063	-	-	-	-
PRW-Net(ICCVW'21)[[illegible]]	20.787	0.823	0.099	0.603	3.062	0.572	2.717	0.625	3.026	223.4G	6.30M	0.216s	4.624
Shallow-uwnet(AAAI'21)[[illegible]]	18.278	0.855	0.131	0.544	2.942	0.521	2.212	0.545	3.109	304.75G	0.22M	0.031s	31.836
Ucolor(TIP'21)[[illegible]]	21.093	0.872	0.096	0.555	3.049	0.530	2.167	0.554	3.148	443.85G	157.42M	2.758s	-
UIEC^2-Net(SPIC'21)[[illegible]]	22.958	0.907	0.078	0.599	2.999	0.580	2.228	0.604	3.125	367.53G	0.53M	0.174s	5.742
MLLE(TIP'22)[[illegible]]	19.561	0.845	0.115	0.592	2.624	0.581	1.977	0.597	2.454	-	-	-	-
UHD-SFNet(ACCV'22)[[illegible]]	18.877	0.810	0.144	0.559	2.551	0.528	1.741	0.585	2.826	15.24G	37.31M	0.059s	16.769
PUIE-Net(ECCV'22)[[illegible]]	21.382	0.882	0.093	0.566	3.021	0.543	2.155	0.563	3.192	423.05G	1.40M	0.071s	14.194
NU2Net(AAAI'23,Oral)[[illegible]]	22.419	0.923	0.086	0.587	2.936	0.555	2.222	0.593	3.185	146.64G	3.15M	0.024s	42.345
Ours	23.061	0.911	0.076	0.616	2.828	0.593	2.088	0.609	3.174	8.33G	0.009M	0.016s	60.724

Table 2: Performance comparison is tested on RTX 3090 using 1080P resolution (1920×1080) images, best and second-best results are marked in red and blue respectively.

	For 1080P Real-time Test			
Methods	GFLOPs(G)↓	#Params(M)↓	#Runtime(s)↓	FPS(f/s)↑
Shallow-uwnet(AAAI'21)[[illegible]]	685.70G	0.22M	0.0745s	13.4188
UHD-SFNet(ACCV'22)[[illegible]]	15.42G	37.31M	0.0666s	15.0045
NU2Net(AAAI'23,Oral)[[illegible]]	329.95G	3.15M	0.0516s	19.3495
Ours	18.74G	0.009M	0.0333s	29.9431

the proposed architecture has remarkable effects with detailed textures, restoring promising contrast and color of images. Compared with the last method NU2Net on T90, we exceed 0.642dB, 0.01 and 0.029 on PSNR, MSE and UCIQE respectively.

According to the data in Table 1, our proposed FA$^+$Net outperforms all other designs in terms of efficiency. The amount of parameters possessed by FA$^+$Net is only **1/17500** of that of Ucolor, yet manifests a considerable qualitative improvement. In comparison with Shallow-uwnet, the quantity of parameters has reduced by more than **1/20**, but the PSNR index is 4.783dB higher, which clearly shows the superiority and viability of our method. What's even more noteworthy is that from Table 2, we can observe that FA$^+$Net is the **sole** network able to carry out real-time enhancement for 1080P size images.

Additionally, we also gave an intuitive comparison with previous SOTA methods in terms of visual effects. As seen in Fig. 3, Shallow-net was incapable of sufficiently restoring underwater images due to its straightforward network structure; the intensified image saturation was low, and the edge processing effect was subpar; the enhancement result presented by PRW-Net appeared layered; the image processed by UHD-SFNet still contains some local patches; PRW-Net and PUIE-Net demonstrated poor perception of details, resulting in significant erosion of texture details in the augmented photographs. On the other hand, NU2Net lacked precise color control, hence leading to visible chromatic deviations that can be noticed by the human eye. Our method exhibited quite compatible color and detail recovery, enhancing the entire degraded image, and making its contrast and texture details meet the sensory requirements of the human eye. That is credited to our carefully designed MCEM, MPM, and SDFIM. More visual comparisons are available in the supplementary material.

4.4 Ablation Study of Model Structures

The detailed ablation experiments on the model's structure are presented in Table 3, the data demonstrates the effectiveness of each component in proposed method.

Based on the experimental outcomes of the control groups **(a)** and **(b)**, it is evident that the performance benefits brought by SDFIM are rather limited. However, in the supplementary material, we have made additional noteworthy findings regarding the influence of different hyperparameters α on enhancing image hues. Moreover, experiments **(d)** and **(f)** clearly demonstrate that MCEM leads to significant performance enhancements. Addition-

Table 3: Ablation study on structures of model. The best results are underlined. The second-best results are in **bold**. ↑ represents that higher is better, and ↓ represents that lower is better. *O* means the component is selected during experiments.

Setting	MCEM_1	MPM	SDFIM	MCEM_2	PA	PSNR↑	SSIM↑	UCIQE↑	#Params(K)↓
a)	*O*	*O*				21.767	0.890	0.597	**5.72K**
b)	*O*	*O*	*O*			21.890	0.899	0.602	8.15K
c)	*O*			*O*	*O*	21.747	0.893	0.612	1.44K
d)	*O*	*O*	*O*		*O*	22.528	0.897	0.606	8.51K
e)	*O*	*O*	*O*	*O*		**22.889**	**0.908**	**0.613**	8.95K
f)	*O*	*O*	*O*	*O*	*O*	23.061	0.911	0.616	8.99K

ally, the experiments conducted in groups **(c)** and **(f)** validate the feasibility of MPM. Despite the relatively higher GFlops associated with the multi-scale pyramid, it makes exceptional contributions to capturing and perceiving fine-grained details within the network.

5 Limitation

Although FA$^+$Net has exhibited its effectiveness and exceptional performance in underwater image enhancement tasks through experiments on multiple datasets, it is still restricted owing to insufficient numbers of training data and unrefined model optimization. Specifically, FA$^+$Net may require more model design and optimization to improve its performance in handling complex underwater image detail problems, such as those containing small objects, complex colors, and textures. Additionally, even though FA$^+$Net exhibits efficiency and flexibility on resource-constrained mobile platforms, further experiments should be conducted to validate its performance and dependability in practical applications.

In the future, we may consider appending more adaptive settings to transform FA$^+$Net into a universal enhancement framework, thus enhancing its applicability and scalability.

6 Conclusion

This paper presents a highly lightweight model with fewer than 9K parameters, resulting in a significant reduction in complexity. Notably, FA$^+$Net stands as the sole network capable of real-time enhancement of 1080P images. Additionally, we propose a two-stage structure to address mixed degradation, enhance color, and restore details for real-time underwater image enhancement. Our motivation includes to develop a lightweight network for underwater devices. Given the challenges posed by mixed degradations in underwater images, generic solutions prove ineffective. To address this, we propose the Multi-Channel Enhancement Module (MCEM) for per-pixel color processing. Additionally, the Multi-Pixel Module (MPM) enhances image detail perception, while a parallel structure improves spatial modeling. The Spatially Differentiable Fine-Grained Image Manipulation Module (SDFIM) effectively mitigates noise and blurring during color enhancement. Moreover, a fine-grained stage is introduced to capture fine details. The integrated approach presented in this paper opens new avenues for underwater image enhancement research.

7 Acknowledgement

This work was supported by Natural Science Foundation of Fujian Province, China (2021J018 67), Xiamen Ocean and Fisheries Development Special Funds (22CZB013HJ04), the Youth Science and Technology Innovation Program of Xiamen Ocean and Fisheries Development Special Funds (23ZHZB039QCB24).

References

[1] Codruta O Ancuti, Cosmin Ancuti, Christophe De Vleeschouwer, and Philippe Bekaert. Color balance and fusion for underwater image enhancement. *IEEE Transactions on image processing*, 27(1):379–393, 2017.

[2] Pierre Charbonnier, Laure Blanc-Feraud, Gilles Aubert, and Michel Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proceedings of 1st international conference on image processing*, volume 2, pages 168–172. IEEE, 1994.

[3] Sixiang Chen, Tian Ye, Yun Liu, Erkang Chen, Jun Shi, and Jingchun Zhou. Snowformer: Scale-aware transformer via context interaction for single image desnowing. *arXiv preprint arXiv:2208.09703*, 2022.

[4] Sixiang Chen, Tian Ye, Yun Liu, Taodong Liao, Jingxia Jiang, Erkang Chen, and Peng Chen. Msp-former: Multi-scale projection transformer for single image desnowing. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2023. doi: 10.1109/ICASSP49357.2023.10095605.

[5] Sixiang Chen, Tian Ye, Jun Shi, Yun Liu, JingXia Jiang, Erkang Chen, and Peng Chen. Dehrformer: Real-time transformer for depth estimation and haze removal from varicolored haze scenes. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5, 2023. doi: 10.1109/ICASSP49357.2023.10096828.

[6] Lu Chi, Borui Jiang, and Yadong Mu. Fast fourier convolution. *Advances in Neural Information Processing Systems*, 33:4479–4488, 2020.

[7] Ziteng Cui, Kunchang Li, Lin Gu, Shenghan Su, Peng Gao, Zhengkai Jiang, Yu Qiao, and Tatsuya Harada. Illumination adaptive transformer. *arXiv preprint arXiv:2205.14871*, 2022.

[8] Xiaohan Ding, Xiangyu Zhang, Jungong Han, and Guiguang Ding. Scaling up your kernels to 31x31: Revisiting large kernel design in cnns. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11963–11975, 2022.

[9] Paul Drews, Erickson Nascimento, Filipe Moraes, Silvia Botelho, and Mario Campos. Transmission estimation in underwater single images. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 825–830, 2013.

[10] Hao Feng, Liejun Wang, Yongming Li, and Anyu Du. Lkasr: Large kernel attention for lightweight image super-resolution. *Knowledge-Based Systems*, 252:109376, 2022.

[11] Zhenqi Fu, Wu Wang, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Uncertainty inspired underwater image enhancement. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVIII*, pages 465–482. Springer, 2022.

[12] Chunle Guo, Ruiqi Wu, Xin Jin, Linghao Han, Zhi Chai, Weidong Zhang, and Chongyi Li. Underwater ranker: Learn which is better and how to be better. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023.

[13] Meng-Hao Guo, Cheng-Ze Lu, Zheng-Ning Liu, Ming-Ming Cheng, and Shi-Min Hu. Visual attention network. *arXiv preprint arXiv:2202.09741*, 2022.

[14] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017. doi: 10.1109/TIP.2016.2639450.

[15] Nik Nurhalida Binti Nik Hariry, Fahirah Syaliza binti Mokhtar, and Mohd Norazmi bin Nordin. Enforcement of maritime archaeology in malaysia: A review. *Journal of Contemporary Issues in Business and Government Vol*, 27(2), 2021.

[16] Kai Hu, Yingjie Huang, Wei Huang, Hui Tan, Zhineng Chen, Zheng Zhong, Xuanya Li, Yuan Zhang, and Xieping Gao. Deep supervised learning using self-adaptive auxiliary loss for covid-19 diagnosis from imbalanced ct images. *Neurocomputing*, 458:232–245, 2021.

[17] Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep fourier-based exposure correction network with spatial-frequency interaction. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX*, pages 163–180. Springer, 2022.

[18] Fushuo Huo, Bingheng Li, and Xuegui Zhu. Efficient wavelet boost learning-based multi-stage progressive refinement network for underwater image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1944–1952, 2021.

[19] Md Jahidul Islam, Youya Xia, and Junaed Sattar. Fast underwater image enhancement for improved visual perception. *IEEE Robotics and Automation Letters*, 5(2):3227–3234, 2020. doi: 10.1109/LRA.2020.2974710.

[20] Jingxia Jiang, Jinbin Bai, Yun Liu, Junjie Yin, Sixiang Chen, Tian Ye, and Erkang Chen. Rsfdm-net: Real-time spatial and frequency domains modulation network for underwater image enhancement. *arXiv preprint arXiv:2302.12186*, 2023.

[21] Qiuping Jiang, Yuese Gu, Chongyi Li, Runmin Cong, and Feng Shao. Underwater image enhancement quality evaluation: Benchmark dataset and objective metric. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(9):5959–5974, 2022.

[22] Qun Jiang, Yunfeng Zhang, Fangxun Bao, Xiuyang Zhao, Caiming Zhang, and Peide Liu. Two-step domain adaptation for underwater image enhancement. *Pattern Recognition*, 122:108324, 2022.

[23] Yeying Jin, Ruoteng Li, Wenhan Yang, and Robby T Tan. Estimating reflectance layer from a single image: Integrating reflectance guidance and shadow/specular aware learning. *arXiv preprint arXiv:2211.14751*, 2022.

[24] Yeying Jin, Wenhan Yang, and Robby T Tan. Unsupervised night image enhancement: When layer decomposition meets light-effects suppression. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXVII*, pages 404–421. Springer, 2022.

[25] Yeying Jin, Wenhan Yang, Wei Ye, Yuan Yuan, and Robby T Tan. Shadowdiffusion: Diffusion-based shadow removal using classifier-driven attention and structure preservation. *arXiv preprint arXiv:2211.08089*, 2022.

[26] Yeying Jin, Wending Yan, Wenhan Yang, and Robby T Tan. Structure representation network and uncertainty feedback learning for dense non-uniform fog removal. In *Computer Vision–ACCV 2022: 16th Asian Conference on Computer Vision, Macao, China, December 4–8, 2022, Proceedings, Part III*, pages 155–172. Springer, 2023.

[27] Michel J Kaiser, Simon Jennings, David N Thomas, David KA Barnes, et al. *Marine ecology: processes, systems, and impacts*. Oxford University Press, 2011.

[28] Jari Korhonen and Junyong You. Peak signal-to-noise ratio revisited: Is simple beautiful? In *2012 Fourth International Workshop on Quality of Multimedia Experience*, pages 37–38. IEEE, 2012.

[29] Chongyi Li, Chunle Guo, Wenqi Ren, Runmin Cong, Junhui Hou, Sam Kwong, and Dacheng Tao. An underwater image enhancement benchmark dataset and beyond. *IEEE Transactions on Image Processing*, 29:4376–4389, 2019.

[30] Chongyi Li, Saeed Anwar, and Fatih Porikli. Underwater scene prior inspired deep underwater image and video enhancement. *Pattern Recognition*, 98:107038, 2020.

[31] Chongyi Li, Saeed Anwar, Junhui Hou, Runmin Cong, Chunle Guo, and Wenqi Ren. Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Transactions on Image Processing*, 30:4985–5000, 2021.

[32] Chongyi Li, Chun-Le Guo, Man Zhou, Zhexin Liang, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Embedding fourier for ultra-high-definition low-light image enhancement. *arXiv preprint arXiv:2302.11831*, 2023.

[33] Hanyu Li, Jingjing Li, and Wei Wang. A fusion adversarial underwater image enhancement network with a public test dataset. *arXiv preprint arXiv:1906.06819*, 2019.

[34] Ronán Long. The marine strategy framework directive: a new european approach to the regulation of the marine environment, marine natural resources and marine ecological services. *Journal of Energy & Natural Resources Law*, 29(1):1–44, 2011.

[35] Hans Marmolin. Subjective mse measures. *IEEE Transactions on Systems, Man, and Cybernetics*, 16(3):486–489, 1986. doi: 10.1109/TSMC.1986.4308985.

[36] Pan Mu, Haotian Qian, and Cong Bai. Structure-inferred bi-level model for underwater image enhancement. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 2286–2295, 2022.

[37] Ankita Naik, Apurva Swarnakar, and Kartik Mittal. Shallow-uwnet: Compressed model for underwater image enhancement (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 15853–15854, 2021.

[38] Karen Panetta, Chen Gao, and Sos Agaian. Human-visual-system-inspired underwater image quality measures. *IEEE Journal of Oceanic Engineering*, 41(3):541–551, 2016. doi: 10.1109/JOE.2015.2469915.

[39] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf.

[40] Yan-Tsung Peng and Pamela C Cosman. Underwater image restoration based on image blurriness and light absorption. *IEEE transactions on image processing*, 26(4):1579–1594, 2017.

[41] Qi Qi, Kunqian Li, Haiyong Zheng, Xiang Gao, Guojia Hou, and Kun Sun. Sguie-net: Semantic attention guided underwater image enhancement with multi-scale perception. *IEEE Transactions on Image Processing*, 31:6816–6830, 2022.

[42] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 11908–11915, 2020.

[43] Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. Self-attention with relative position representations. *arXiv preprint arXiv:1803.02155*, 2018.

[44] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[45] Dehua Song, Chang Xu, Xu Jia, Yiyi Chen, Chunjing Xu, and Yunhe Wang. Efficient residual dense block search for image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 12007–12014, 2020.

[46] Wei Song, Yan Wang, Dongmei Huang, Antonio Liotta, and Cristian Perra. Enhancement of underwater images with statistical model of background light and optimization of transmission map. *IEEE Transactions on Broadcasting*, 66(1):153–169, 2020.

[47] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 2149–2159, 2022.

[48] Hind Taud and JF Mas. Multilayer perceptron (mlp). *Geomatic approaches for modeling land change scenarios*, pages 451–455, 2018.

[49] Peter A Todd, Eliza C Heery, Lynette HL Loke, Ruth H Thurstan, D Johan Kotze, and Christopher Swan. Towards an urban marine ecology: characterizing the drivers, patterns and processes of marine ecosystems in coastal cities. *Oikos*, 128(9):1215–1242, 2019.

[50] Ashish Vaswani, Prajit Ramachandran, Aravind Srinivas, Niki Parmar, Blake Hechtman, and Jonathon Shlens. Scaling local self-attention for parameter efficient visual backbones. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12894–12904, 2021.

[51] Yang Wang, Jing Zhang, Yang Cao, and Zengfu Wang. A deep cnn method for underwater image enhancement. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 1382–1386, 2017. doi: 10.1109/ICIP.2017.8296508.

[52] Yi Wang, Na Wang, Min Xu, Junxiong Yu, Chenchen Qin, Xiao Luo, Xin Yang, Tianfu Wang, Anhua Li, and Dong Ni. Deeply-supervised networks with threshold loss for cancer detection in automated breast ultrasound. *IEEE transactions on medical imaging*, 39(4):866–876, 2019.

[53] Yudong Wang, Jichang Guo, Huan Gao, and Huihui Yue. Uiec^ 2-net: Cnn-based underwater image enhancement using two color space. *Signal Processing: Image Communication*, 96:116250, 2021.

[54] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

[55] Yiwen Wei, Zhuoran Zheng, and Xiuyi Jia. Uhd underwater image enhancement via frequency-spatial domain aware network. In *Proceedings of the Asian Conference on Computer Vision*, pages 299–314, 2022.

[56] Miao Yang and Arcot Sowmya. An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing*, 24(12):6062–6071, 2015. doi: 10.1109/TIP.2015.2491020.

[57] Xiangli Yang, Zixing Song, Irwin King, and Zenglin Xu. A survey on deep semi-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[58] Lan Yao, Longguan Zhu, and Bingjian Zhang. Application of 4-dihydrochromone as a temporary consolidant in underwater archeology. *Journal of Cultural Heritage*, 57:235–242, 2022.

[59] Tian Ye, Sixiang Chen, Yun Liu, Yi Ye, Jinbin Bai, and Erkang Chen. Towards real-time high-definition image snow removal: Efficient pyramid network with asymmetrical encoder-decoder architecture. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pages 366–381, December 2022.

[60] Tian Ye, Sixiang Chen, Yun Liu, Yi Ye, Erkang Chen, and Yuche Li. Underwater light field retention: Neural rendering for underwater imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 488–497, 2022.

[61] Tian Ye, Yunchen Zhang, Mingchao Jiang, Liang Chen, Yun Liu, Sixiang Chen, and Erkang Chen. Perceiving and modeling density for image dehazing. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision – ECCV 2022*, pages 130–145, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-19800-7.

[62] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-attention generative adversarial networks. In *International conference on machine learning*, pages 7354–7363. PMLR, 2019.

[63] Weidong Zhang, Peixian Zhuang, Hai-Han Sun, Guohou Li, Sam Kwong, and Chongyi Li. Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Transactions on Image Processing*, 31:3997–4010, 2022.

[64] Man Zhou, Jie Huang, Chongyi Li, Hu Yu, Keyu Yan, Naishan Zheng, and Feng Zhao. Adaptively learning low-high frequency information integration for pan-sharpening. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 3375–3384, 2022.

[65] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years: A survey. *Proceedings of the IEEE*, 2023.