Proceedings of Machine Learning Research vol 278:1–7, 2025 2025 2nd International Conference on Machine Learning and Intelligent Computing

Study on Wavelet Convolution-Based Underwater Image Denoising

Wei Li
Dalian Maritime University, Dalian 116000, Liaoning, China

Haiyan Xie* 59991625@QQ.COM
Dalian Maritime University, Dalian 116000, Liaoning, China

Jiaxi Li
Dalian Maritime University, Dalian 116000, Liaoning, China

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

The processing of underwater images is critical for marine science, seafloor mapping, and underwater rescue operations. However, underwater optical images often suffer from poor quality due to light absorption and scattering caused by suspended particles. Additionally, due to technical limitations and environmental interference, underwater robots often capture images where light has been reflected and refracted multiple times before reaching the camera, further exacerbating noise. To address these challenges, this paper proposes an innovative underwater image processing model that combines wavelet convolution and dilated convolution for noise reduction. The model employs wavelet transformation to decompose images into high- and low-frequency components for preliminary processing, followed by the use of dilated convolution to extract noise and image features. This approach effectively removes noise from underwater images. Experimental results demonstrate that this method can adaptively handle illumination and detail information across different scales, addressing challenges such as uneven lighting, low contrast color distortion, and suspended particle noise. The processed images exhibit significantly improved clarity and contrast, even in complex underwater environments.

Keywords: underwater image processing, wavelet transform, wavelet convolution, dilated convolution, convolutional neural network (CNN)

1. Introduction

Recent studies have employed traditional CNN-based models for underwater image denoising. Liu et al. (2023) proposed a multiscale dual-color space underwater image enhancement network, leveraging the asymmetric multiscale architecture of deep learning to extract and fuse features from RGB and HSV color spaces, resulting in more natural underwater images. Wang et al. (2021) addressed light scattering and particulate absorption by integrating CNN and Transformer networks into a U-Net structure. This model uses CNN to extract local image features and employs Transformers to capture contextual relationships, thereby optimizing underwater image details. However, the U-Net model lacks generalization capability for images captured under diverse underwater environments and lighting conditions. To address this, Lu et al. (2024) utilized the DiffWater method, which applies diffusion probabilistic models (DDPM) to underwater image processing. By gradually adding noise to images until they resemble pure noise and then learning to reverse the process, this method effectively restores original images. The conditional diffusion model's strong adaptability allows it to maintain image clarity while highlighting object features. Yin et al. (2024) further improved upon this approach by designing a dual-branch wavelet diffusion model with dual-prior refinement. This model processes low- and high-frequency features of underwater images separately using wavelet decomposition and dual-branch diffusion, while enhancing overall image quality through dual-prior optimization. However, due to its high computational demands and reliance on prior knowledge, this model's practical application is limited by hardware requirements and the need for further refinement.

Overall, existing models face limitations in practical underwater image processing, often requiring significant computational resources to achieve performance improvements. To enhance the efficiency of underwater image denoising, this study integrates the wavelet-convolution (WTConv) into the ECNDNet architecture. The contributions of this work are as follows:

1. We incorporate wavelet convolution into the ECNDNet architecture to improve the efficiency of underwater image denoising. This allows multi-scale feature extraction while reducing computational complexity.

2. Innovatively combining wavelet convolution with dilated convolution for underwater image denoising, this approach addresses the challenge of preserving structural details in noisy underwater environments.

2. Methodology

The wavelet transform is capable of processing signals across different frequency domains, whereas convolutional neural networks (CNNs) exhibit robust feature learning abilities. Wavelet convolution, integrating these strengths, substantially improves the denoising capability of the model in image processing. This subsection elaborates on the theoretical principles of wavelet convolution and the quantitative metrics for assessing denoising performance.

2.1. WaveletTransform

The primary function of the wavelet transform (Sifuzzaman et al., 2009) lies in signal decomposition and reconstruction. A signal can be decomposed into a series of components across different frequency bands, facilitating a deeper understanding of its key characteristics. During decomposition, the signal is divided into low-frequency and high-frequency components. The low-frequency components reflect the overall trend of the signal, while the high-frequency components play a critical role in preserving detailed features.The core principle of wavelet transform involves using a wavelet function as the basis function, which can be scaled (dilated/compressed) and translated (shifted) before being multiplied with the signal. Different scaling factors correspond to different frequencies: a narrower wavelet (compressed basis) covers a smaller range on the time axis, enabling precise capture of rapid signal variations (high-frequency components), whereas a wider wavelet (dilated basis) is suited for analyzing low-frequency components.By scaling and translating the basis function, the wavelet transform dynamically aligns with local features of the signal at arbitrary positions, achieving a balance between time-domain and frequency-domain resolution. This ensures high-frequency precision while maintaining robustness for low-frequency trends. During reconstruction, the processed low-frequency and high-frequency components are recombined via the inverse wavelet transform, resulting in a denoised signal.

2.2. Integration of Wavelet Transform and Convolution

Convolutional neural networks (CNNs) utilize convolutional kernels to extract features from input images, capturing fine-grained details through their powerful learning capabilities. In contrast, the wavelet transform employs distinct wavelet bases to decompose an image into multi-scale low-frequency (LF) and high-frequency (HF) components. By integrating wavelet transforms with convolutional operations—performing convolution on each sub-frequency band to extract features, followed by reconstructing the processed bands via the inverse wavelet transform—the original image is restored. This fusion strategy separately optimizes global information (LF) and detailed features (HF), enabling robust feature extraction and enhanced denoising performance (Tian et al., 2023).

2.3. Evaluation indicators

Image denoising evaluation metrics serve as critical criteria for assessing the performance of denoising algorithms. This study employs the following metrics: Peak Signal-to-Noise Ratio (PSNR) (Huynh-Thu and Ghanbari, 2008), Structural Similarity Index (SSIM) (Wang et al., 2004), and Normalized Mean Squared Error (NMSE) (Jakub and Grzegorz, 2012). The scope and evaluation effectiveness of these metrics are summarized in Table 1.

Table 1: Objective Evaluation Metrics for Image Denoising Quality

Evaluation indicators	Scope of Evaluation	Evaluation Effectiveness
PSNR	[20dB,40dB]	The higher the value, the richer the semantic information of the denoised image, and the better the denoising effect.
SSIM	[0,1]	The closer the value converges to 1, the more similar the denoised image is to the original image and the more similar the structure is.
NMSE	[0,1]	The closer the value is to 0, the smaller the deviation of the denoised image from the original image.

3. Basic Theory

This study integrates ECNDNet with wavelet convolution to develop a lightweight denoising network. The proposed framework holds significant potential for subsea engineering and marine exploration, offering broad application prospects.

3.1. Experimental data

In all experiments in this study, the training of the denoising model utilized a subset of images from the publicly available underwater dataset UFO-120 (Islam et al., 2020).

Boundary artifacts are common during model training. To mitigate this issue, splitting images into small patches effectively avoids such artifacts. This approach addresses the distortion caused by the convolutional kernel's inability to capture contextual relationships near image boundaries during sliding operations. Therefore, images were cropped into small patches to enhance the model's ability to learn boundary features during training.

The detailed procedure is as follows: First, 20 clear underwater images with a size of 200×200 pixels were selected from the public seabed dataset UF0-120 as original images. During data preprocessing, the sliding window size was set to 41×41, and the image was cropped with a stride of 10 to serve as the ground truth training images. Then, Gaussian noise was added to the images. To enhance the diversity of the training dataset, each cropped image was randomly transformed four times using the following operations: original cropped image; horizontal flip; 90° counterclockwise rotation; horizontal flip after 90° counterclockwise rotation; 180° counterclockwise rotation; horizontal flip after 180° counterclockwise rotation; 270° counterclockwise rotation; horizontal flip after 270° counterclockwise rotation. Through the above operations, 55,680 color synthetic noise images of 41×41 were obtained and used as the training dataset.

3.2. Wavelet Convolution Underwater Processing Model

The underwater processing model is composed of wavelet convolution (WTConv) and ECNDNet, as shown in Figures 1 and 2.

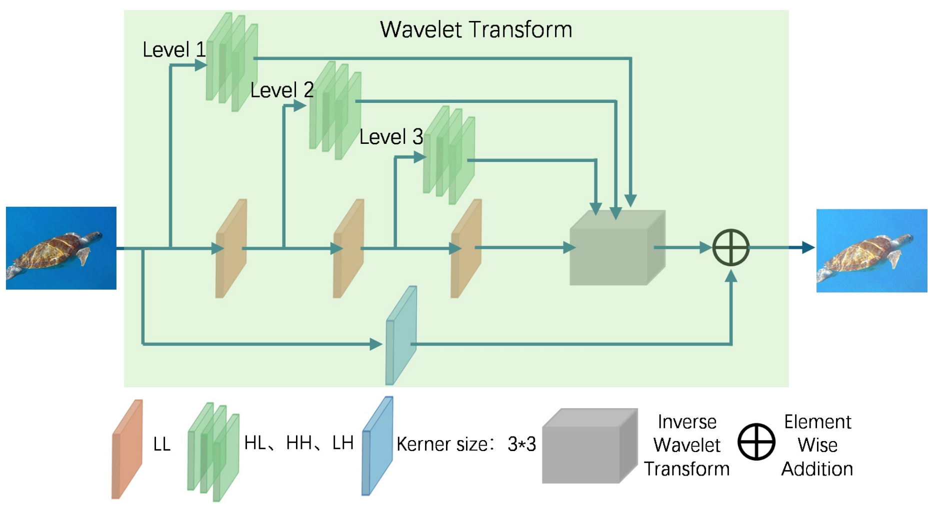

Figure 1: The network architecture of the proposed WECNDnet.

The input image is first processed by the wavelet convolution layer (WTConv), where it is decomposed into high-frequency subbands (HL, HH, LH) and a low-frequency subband (LL). The high-frequency subbands represent the horizontal, vertical, and diagonal high-frequency components of the image, while the low-frequency subband contains most of the image's outline information. The WTConv layer employs three layers of wavelet decomposition, enabling feature analysis at different scales and effectively separating noise from the signal across these scales. By combining wavelet transforms with depthwise separable convolutions, the model performs multi-resolution analysis and denoising. Finally, the image is reconstructed through an inverse wavelet transform.

To enhance underwater image processing without increasing computational overhead, WECNDNet employs the Haar wavelet basis (Finder et al., 2022). The Haar wavelet basis offers high

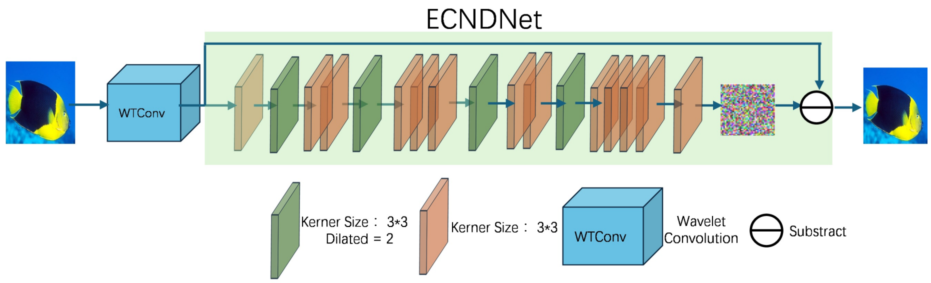

Figure 2: Architecture of Wavelet Convolution .

computational efficiency, facilitates edge detection and feature extraction, and effectively captures local image features. Subsequently, the ECNDNet framework further enhances feature extraction, improving the model's generalization capability.

4. Experimental Results

This study employs an underwater wavelet convolutional network for image denoising. A subset of images from the publicly available underwater dataset UFO-120 was used for training and testing, with experiments conducted under a noise level of 15, 25 and 50. Additionally, the proposed method is compared with two representative CNN-based approaches: DnCNN (Zhang et al., 2017) and ECNDNet (Tian et al., 2019), evaluating their performance through Peak Signal-to-Noise Ratio (PSNR) on the same dataset.

Table 2: Average PSNR/SSIM/NMSE Values of Denoising Methods on UFO-120 Dataset

Evaluation indicators	**Noise level**	**DnCNN**	**ECNDNet**	**WECNDNet**
PSNR	15	30.5071	26.0794	31.0818
	25	28.6121	21.8864	28.7455
	50	26.0105	16.4035	26.212
SSIM	15	0.8448	0.5124	0.9118
	25	0.8097	0.3159	0.8598
	50	0.7067	0.136	0.7898
NMSE	15	0.0044	0.0118	0.0038
	25	0.0070	0.0310	0.0066
	50	0.0119	0.11	0.0118

As shown in Table 2, the WECNDNet model achieves the highest denoising performance. Notably, under a noise level of 50 dB, WECNDNet outperforms DnCNN by approximately 0.2 dB in PSNR, 0.0831 in SSIM, and 0.0001 in NMSE (see Table 1 for metric definition). These results confirm the robust denoising capability of WECNDNet.

5. Conclusion

The WECNDNet proposed in this study demonstrates promising denoising performance for underwater image processing. By integrating a WTconv module, the denoising model optimizes the architecture of the baseline ECNDNet. Experimental results indicate that wavelet convolution effectively suppresses noise in underwater images while reducing computational complexity. WECNDNet exhibits significant potential for applications in complex underwater environments, streamlining workflows for subsea operations. Future work will focus on further refining WECNDNet or training alternative models to enhance the clarity of underwater image restoration.

References

Shahaf E. Finder, Yair Zohav, Maor Ashkenazi, and Eran Treister. Wavelet feature maps compression for image-to-image cnns, 2022.

Q. Huynh-Thu and M. Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics Letters*, 44:800–801, 2008. doi: 10.1049/el:20080522.

Md Jahidul Islam, Peigen Luo, and Junaed Sattar. Simultaneous enhancement and super-resolution of underwater imagery for improved visual perception. *CoRR*, abs/2002.01155, 2020.

Peksinski Jakub and Mikolajczak Grzegorz. Based on spectral analysis of voice controllable surveillance system using normalized mean square error. In *Mechanical Properties of Materials and Information Technology*, volume 340 of *Advanced Materials Research*, pages 156–160. Trans Tech Publications Ltd, 1 2012. doi: 10.4028/www.scientific.net/AMR.340.156.

Chufan Liu, Xin Shu, Lei Pan, Jinlong Shi, and Bin Han. Multiscale underwater image enhancement in rgb and hsv color spaces. *IEEE Transactions on Instrumentation and Measurement*, PP:1–1, 01 2023. doi: 10.1109/TIM.2023.3298395.

Siqi Lu, Fengxu Guan, Hanyu Zhang, and Haitao Lai. Speed-up ddpm for real-time underwater image enhancement. *IEEE Trans. Cir. and Sys. for Video Technol.*, 34(5):3576–3588, May 2024. doi: 10.1109/TCSVT.2023.3314767.

M Sifuzzaman, M R Islam, and M Z Ali. Application of wavelet transform and its advantages compared to fourier transform. Journal of Physical Science;Vol 13 [2009], 2009. Accessed: Mar. 05, 2025. [Online]. Available: https://ir.vidyasagar.ac.in/jspui/handle/123456789/779.

Chunwei Tian, Yong Xu, Lunke Fei, Junqian Wang, Jie Wen, and Nan Luo. Enhanced cnn for image denoising. *CAAI Transactions on Intelligence Technology*, 4(1):17–23, March 2019. doi: 10.1049/trit.2018.1054.

Chunwei Tian, Menghua Zheng, Wangmeng Zuo, Bob Zhang, Yanning Zhang, and David Zhang. Multi-stage image denoising with the wavelet transform. *Pattern Recogn.*, 134(C), February 2023. doi: 10.1016/j.patcog.2022.109050.

Yudong Wang, Jichang Guo, Huan Gao, and HuiHui Yue. Uiec^2-net: Cnn-based underwater image enhancement using two color space. *CoRR*, abs/2103.07138, 2021.

Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861.

S. Yin, W. Wang, Y. Wang, L. Ma, and Y. Yang. Dual-branch wavelet diffusion models with dual-prior refinement for underwater image enhancement. *Social Science Research Network*, 4766632, Mar 2024. doi: 10.2139/ssrn.4766632.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26 (7):3142–3155, 2017. doi: 10.1109/TIP.2017.2662206.