Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 9
AIDML 2025

Joint Restoration Method Based on Super-resolution Network and Diffusion Model

Mengran Xin [1, *], Chenghan Li [2], Mengyao Xin [3]

[1] School of Electronic Information and Automation, Tianjin University of Science and Technology, Tianjin, China, 300457

[2] College of Mechanical and Automotive Engineering, Qingdao University of Technology, Qingdao, China, 266520

[3] School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China, 215123

* Corresponding Author Email: 15139963943@163.com

Abstract. To address the challenges of low image resolution obtained by underwater robots and poor mask inpainting effect, a joint restoration method based on super-resolution network and diffusion model is proposed. Through the degradation model and deep learning, the mapping between high and low resolution images is achieved, and combined with the reverse optimization ability of the diffusion model, the image clarity and physical authenticity are significantly improved. Experiments show that the proposed method outperforms existing technologies in terms of PSNR, SSIM, UIQM and LPIPS, especially in the restoration effect under complex occlusion scenarios. This research effectively enhances the visual perception ability of underwater robots and provides new ideas for image processing in complex water areas.

Keywords: Super-resolution network; diffusion model; image restoration; underwater robot; degradation model.

1. Introduction

Underwater image acquisition serves as the primary link for precise environmental perception, and its significance is self-evident. However, the special optical environment underwater poses significant challenges to image acquisition, such as rapid attenuation of light and scattering effects of suspended particles, resulting in widespread problems like low contrast, color distortion in underwater images. These visual degradation phenomena not only directly affect the perception accuracy of the vehicle for the surrounding environment but also severely restrict its operational efficiency in tasks such as resource exploration and seabed mapping. To address these issues, Jinlong Tang et al. proposed a serialized attention (SA) module to bolster the capacity of capturing global information[1]; Parashara Kodatiet et al. demonstrated that the Laplace prior promotes sparsity in the latent representations, when there were limited features of interest in the input[2]; Shao et al. introduced an innovative method with a region gradient-oriented mechanism to enhance the noise addition process of diffusion models, ensuring continuous forward iterations [3]; Wang Y et al. designed a convolutional neural network-based underwater image enhancement network UIE-NET, enabling simultaneous color correction and haze removal processes, thereby improving the accuracy and convergence speed of the learning process [4]; Lai Yunting et al. improved SRGAN, retaining the details and high-frequency information of the image [5]; Qiao Lu-kuan et al. combined advanced image processing technology with the parallel processing capability of FPGA to achieve efficient and real-time underwater visual detection[6]. Yuan Fei et al. proposed a chromaticity-aware operator to ignore the widespread unperceivable chromatic drift in underwater imaging, thereby improving the image compression rate [7]. However, these methods still face difficulties in multi-task collaborative optimization, extreme scarcity of high-quality paired data, and the diversity and dynamic changes of underwater environments, making it difficult for the model algorithm to generalize to different waters.

In recent years, the challenges brought by complex waters remain significant, and there are still many difficult points to overcome in core technologies, such as a 37% model prediction error in turbid waters and poor performance when dealing with diverse occlusion types. Color distortion, lighting, and occlusion problems in images remain challenging. However, the rapid development of super-resolution networks has significantly improved the visual effect of robots in complex underwater environments. At the same time, diffusion models have demonstrated unique technical advantages in underwater image processing through physical-guided iterative optimization and frequency-domain feature decoupling, promoting the evolution of underwater image processing from "visibility restoration" to "physical authenticity reconstruction" and effectively addressing the complexity of underwater image degradation.

This study designed an image super-resolution network specifically for processing blurry images recognized by robots, improving their clarity and distinguishability. We deployed this image super-resolution network on Lu Ban Cat to enable it to process and optimize images underwater, significantly enhancing the visual effect of robots in complex underwater environments. This study compared it with other super-resolution networks and obtained PSNR, SSIM, UIQM, and LPIPS parameters of 26.26 dB, 0.74, 2.88, and 0.19 respectively, the quality of underwater images has been enhanced as reasult.(In Figure 5). At the same time, a diffusion model-based underwater image restoration method was introduced. This method makes use of the powerful generation ability of the model and introduces the resampling technique of the reverse diffusion process. During the inference process, it effectively integrates the generated image information, resulting in a lower LPIPS value in the restoration of underwater images. This indicates that the restoration method of this study can play a positive role for images occluded by different types of mask inpainting.

2. Results and Discussion

2.1. Principles and Experiments of Degradation Model

Due to the complexity and variability of the underwater environment, the degradation factors faced by underwater images are significantly different from those of conventional images. This study has specially developed a customized degradation model for underwater images, focusing on the unique degradation factors of underwater images. A data-driven model has been constructed to accurately describe the process of image degradation from high quality to degradation, and based on this, an optimization method has been designed to achieve reverse restoration. This not only helps to reduce unnecessary computational load but also optimizes processing speed. The degradation formula for underwater images is:

$$x=\left((y\otimes k)+n+p\right)_{\downarrow s} \tag{1}$$

Here, n represents noise and p represents suspended matter in the medium, k is the point spread function, and y represents the ideal image before being affected by degradation.

In this study, a first-order degradation process is employed, as shown in Figure 1. The degradation process mainly consists of four steps: Resize, Noise, Blur, and Particle. Resize is used to change the size of the image, and it lists three common adjustment algorithms: bilinear interpolation, bicubic interpolation, and area interpolation. Noise adds different types of noise points to the image, including chromatic noise, ripple noise, and scattering noise, to simulate various interference effects in the real world or enhance the texture of the image. Blur uses Gaussian blurring, performing convolution operations with the Gaussian function on the image, which can effectively reduce noise in the image and produce a smooth blurring effect. In this step, Particle adds the effect of suspended particles to the image, making it seem as if there are suspended particles within it, enhancing the realism of the image or creating a special visual scene.

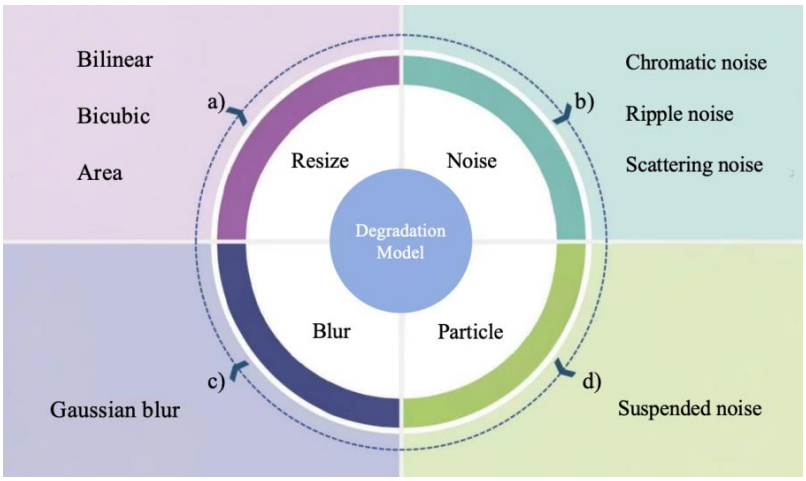

Figure 1. Degradation Model Diagram

To verify the validity of the underwater image degradation model, this study selected high-definition images covering typical scenes such as coral reefs and sunken ships. The images were generated by two groups: the experimental group using the customized degradation model (fusing light scattering, turbulent blurring and mixed noise) and the control group using direct downsampling. As shown in Figure 2, through multi-scale texture analysis and noise separation calculation, it was found that the ratio of the texture standard deviation (STD = 19.83) to the noise standard deviation (STD = 15.06) in the experimental group (0.76) approximated the real underwater data (benchmark 0.82), indicating that the model accurately simulated the texture smoothing and noise enhancement characteristics caused by suspended matter in the water body; while in the control group, due to the generation of false high-frequency edges by downsampling interpolation, the texture STD was abnormally increased to 78.95, and the noise STD was only 5.57, deviating significantly from the actual environmental characteristics. Cross-scenario tests showed that the model remained stable under extreme conditions such as illuminance lower than 1 lux and turbidity greater than 10*, confirming the validity of the simulation data.

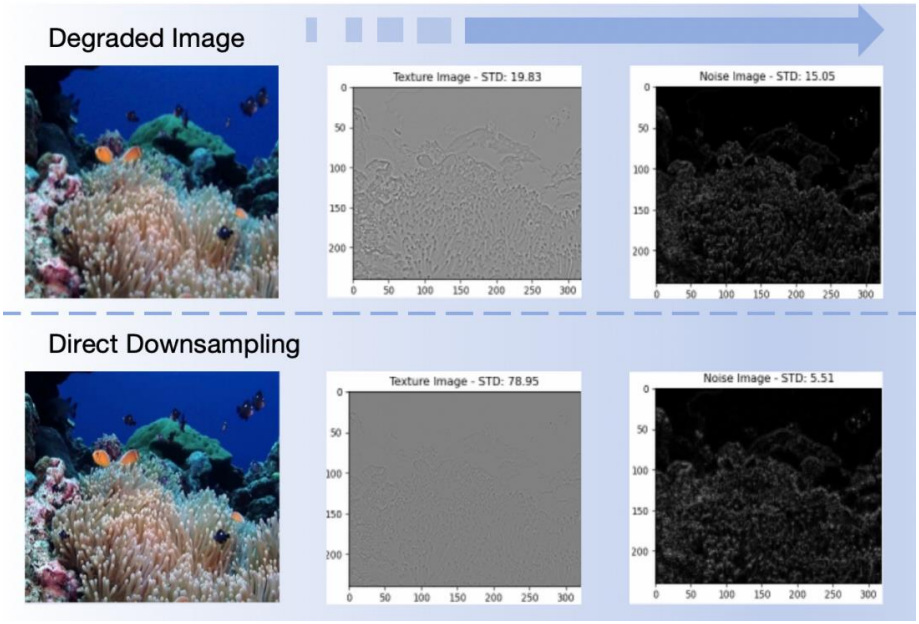

Figure 2. Degradation model test comparison

2.2. The Design of Super-resolution Networks

The network consists of three main parts: the degradation model, the generator, and the discriminator. As shown in Figure 3. The degradation model is responsible for converting high-resolution images into low-resolution images that simulate the real underwater environment. The generator adopts a

three-level progressive feature processing: Firstly, it extracts shallow features F0∈ RH*W*C from the low-resolution input ILR∈RH*W*Cin to capture the basic details of the image. Then, the deep feature extraction layer extracts deep features through a series of attention-enhanced dense residual blocks (AERDB) and a 3×3 convolution layer [8]. Subsequently, each AERDB module integrates dense connections, channel attention mechanism (CAS), and residual learning, and finally performs pixel re- upsampling on the fused features through sub-pixel convolution to output high-resolution reconstructed images. The discriminator adopts a multi-level feature encoding and discrimination framework. The input data stream contains generated samples and real high-resolution images pairs. Firstly, it performs spatial downsampling and feature abstraction through five groups of convolution-activation units. Then, spectral normalization constraints are applied after each convolution operation layer to ensure that the discriminator satisfies Lipschitz continuity [9], thereby stabilizing the dynamic balance of adversarial training; finally, the feature vectors are mapped to the real number domain by a fully connected layer, and the probability value is output through the Sigmoid function, quantifying the similarity between the input image and the real data distribution.

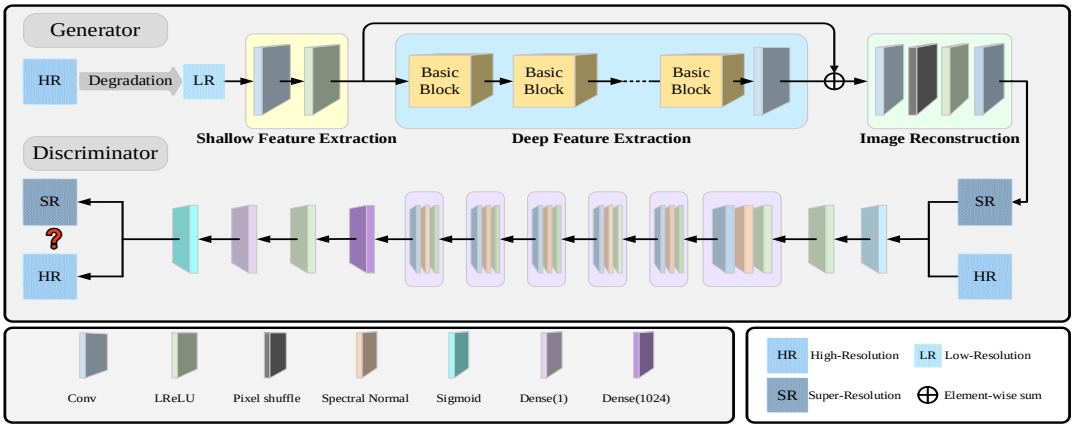

Figure 3. Network Structure Diagram

The model of this study was developed in the PyTorch framework and trained using the Adam optimizer. The training process was conducted on the GPU 3090. The input used for training was 64x64-pixel image blocks with a batch size of 32. The initial learning rate was set to 0.0002, and it was halved after 200,000 iterations. The β1 and β2 hyperparameters of the Adam optimizer were set to 0.9 and 0.99 respectively. The entire training process is shown in Figure 4, which lasted for 400,000 iterations.

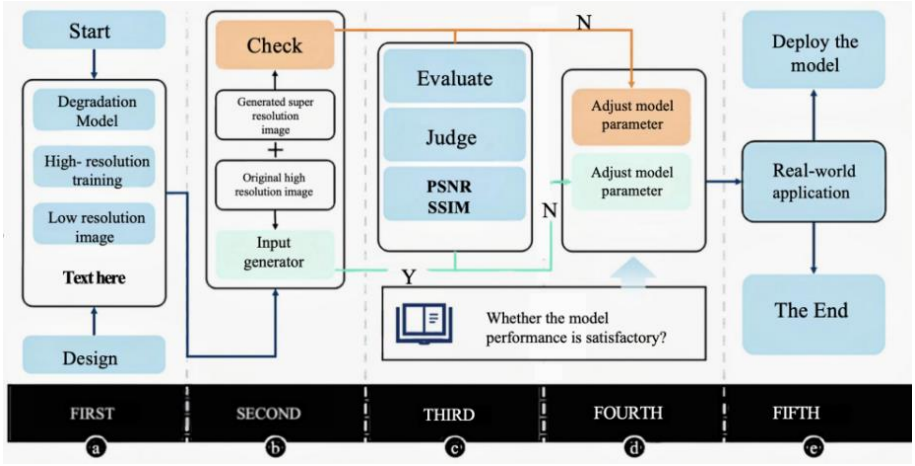

Figure 4. Training Flowchart

When conducting image quality assessment, multiple-dimensional indicators need to be combined [10]: Peak Signal-to-Noise Ratio (PSNR) quantifies pixel-level errors and reflects global fidelity; Structural Similarity Index (SSIM) evaluates luminance, contrast, and structural similarity, capturing local degradation features; Underwater Image Quality Measurement (UIQM) is specifically designed

for underwater scenarios and comprehensively assesses restoration effects through chroma, clarity, and contrast; Learned Perceptual Image Patch Similarity (LPIPS) measures perceptual similarity based on differences in deep features, being closer to subjective human evaluation. Joint analysis of multiple indicators can more comprehensively assess the performance of the algorithm in terms of detail retention, physical accuracy, and visual perception. Here, our research model is compared with other super-resolution networks, and the average performance indicators of all test samples are obtained as shown in Figure 5:

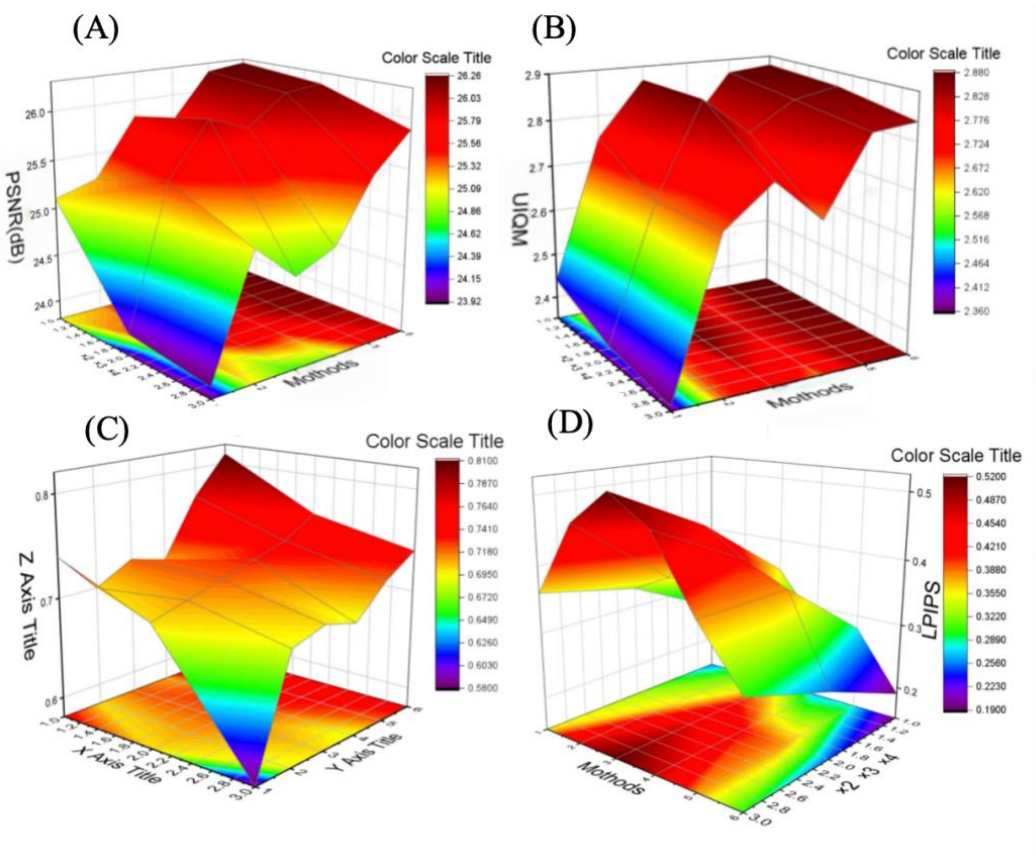

Figure 5. Model Test Data Comparison Chart.(A) The PSNR values of each algorithm, (B)The UIQM values of each algorithm, (C) The PSNR values of each algorithm, (D) The LPIPS values of each algorithm.

This study visually presents the effects of various methods through three-dimensional statistical charts. From the above comparisons, it can be clearly observed that the 6th group is the model adopted in this study. Compared with other networks, the PSNR, SSIM and UIQM values are all the largest, increasing by 2%, 10% and 5% respectively. The LPIPS parameter can reach a minimum of 0.19, indicating that this model performs excellently in restoring the clarity and detailed textures of images, especially in the edge areas.

2.3. Application of Diffusion Model

2.3.1. A Brief Account of the Principles of Diffusion Models

The current underwater image restoration methods based on deep learning suffer from significant limitations: mainstream algorithms are mainly designed for the restoration tasks of single type of mask occlusion, and their generalization ability is limited when dealing with diversified mask occlusions in images. This study proposes a underwater image restoration framework based on diffusion models, which introduces a resampling mechanism to achieve iterative optimization of the generation path during the reverse diffusion process, effectively integrating global semantic information of the image and local texture features, and significantly enhancing the restoration robustness and visual fidelity of the robot in complex occlusion scenarios.

DDPM is a generative model based on the diffusion process [11], which consists of the forward diffusion process, the reverse diffusion process, the training objective, and sampling generation. The forward process is the process of adding noise to the image. The original data x0 is gradually transformed into pure noise xΓ ~ N(0,I) through T steps of adding Gaussian noise. The forward process can also be regarded as a Markov process due to the fact that the noise added at each step only depends on the previous state, and its mathematical form is:

$$q(x_t|x_{t-1}) = N\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I\right) \tag{2}$$

Among them, β_t is a preset noise scheduling parameter. As t increases, x_t gets closer and closer to pure noise.

The original real image is represented as x, the unknown pixels are represented as m*x, and the known pixels are represented as (1-m)*x. For sampling and using in the unknown regions and for sampling and using in the known regions, the implemented steps are as follows:

$$x_{t-1}^{known} \sim N\left(\sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I\right) \tag{3}$$

$$x_{t-1}^{unknown} \sim N\left(\mu_\theta(x_t,t), \Sigma_\theta(x_t,t)\right) \tag{4}$$

$$x_{t-1} = m \cdot x_{t-1}^{known} + (1-m) \cdot x_{t-1}^{unknown} \tag{5}$$

Among them, x_{t-1}^{known} is sampled from the known pixels in the given Figure m*x0, while $x_{t-1}^{unknown}$ is sampled from the model of the given xt. Then, these samples are linearly combined into a new sample xt-1 through the mask. Starting from xt-1, estimate the distribution of xt-2 and proceed in sequence.

2.3.2. Underwater Image Restoration Based on Diffusion Model

The model in this study uses xt to predict xt-1, and its principle is shown in Figure 6, which includes the output of DDPM and the sampling from the known region. However, when sampling the known pixels, the image generation part [12] was not taken into account, thereby introducing incoordination. Although the model attempts to re-coordinate the images at each step, it cannot fully converge due to the same problem occurring in the next step. Furthermore, in each reverse step, the variance change of βt will cause significant changes in the image. Therefore, before entering the next denoising step, the model needs more time to harmonize $x_{t-1}^{unknown}$ using x_{t-1}^{known}.

The trained DDPM is used to generate images that follow the data distribution, and naturally, it also needs to generate a certain structure. In the resampling method, this attribute of DDPM is utilized to coordinate the input of the model [13]. Spread as a result, diffuse the output xt-1 back to xt, namely the xt ~ N ($\sqrt{1-\beta_t}x_{t-1}$, βtI). This operation will scale the output and add some noise, but some of the information merged in the generated area $x_{t-1}^{unknown}$ still remains in $x_{t-1}^{unknown}$. This enables the newly generated $x_{t-1}^{unknown}$ to be harmonious with x_{t-1}^{known} and contain some of its own information.

To solve this problem, the time span of the number of resampling times is called the jump length, and the jump length of the experiment in this study is set to j. Meanwhile, increasing the number of resampling times will also increase the running time of reverse diffusion and reduce the diffusion speed. After the diffusion rate decreases, by reducing the increased variance in each denoising step, a smaller but greater number of resampling times can be applied.

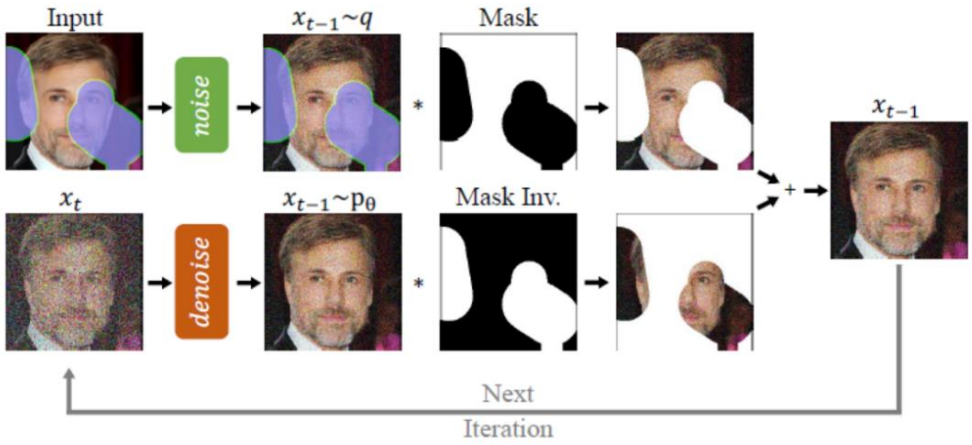

Figure 6. Model Test Data Comparison Chart

2.3.3. Underwater Image Restoration Experiment

The dataset adopted in this study is from ImageNet. Since the method of this study relies on the pre-trained diffusion model, the provided ImageNet pre-trained model is used. The size of the pictures is all set to 256*256. Set the time step T to 250, and set the resampling number r to 10 times, the jump length j to 10, and the iteration number to 4570 times. Follow the same training hyperparameters as ImageNet. The pictures used for the test selected underwater images such as unmanned ships. The algorithm of this study was compared with other advanced methods, and qualitative and quantitative analyses were conducted on its accuracy, robustness and diversity.

The method of this study is compared with several state-of-the-art autoregression-based or GAN-based methods. The autoregression-based methods are DSI and ICT, and the GAN-based methods are DeepFillv2, AOT and LaMa. Since LaMa does not support the ImageNet model, this method is not included in the comparison. As shown in Figure 7, the red unmanned ship is covered by randomly distributed masks. The lower parts of the ships repaired by the DSI and ICT methods are both damaged and have less global consistency. The half of the ship repaired by DeepFillv2 has two scratches, and the color is similar to that of the seawater, with no good color matching. The upper part of the left porthole of the ship restored by AOT is blurred, and a red part of the seawater part below has been restored, with semantic deficiency. After comparison, it can be clearly seen that the repair effect of the method proposed in this study is better.

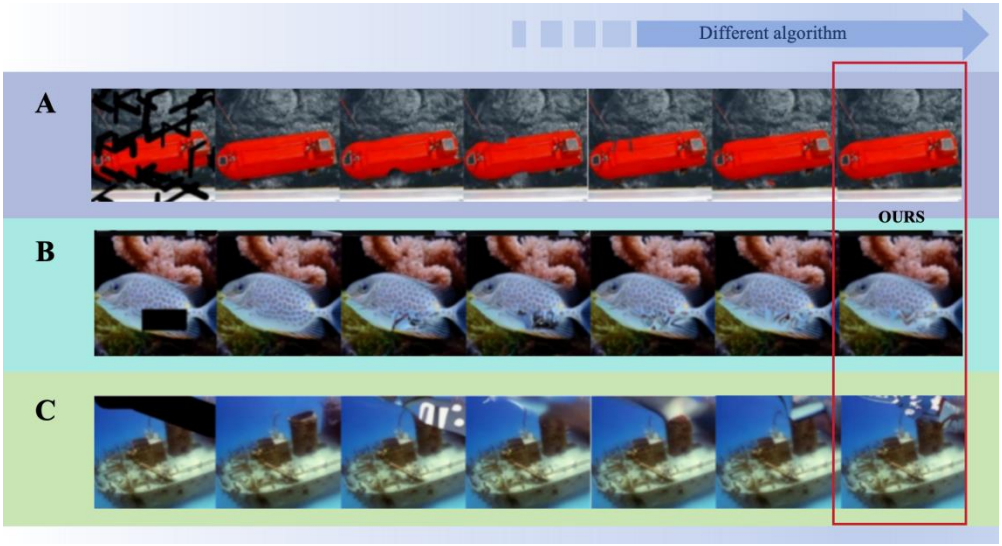

Figure 7. Comparison of the Restoration Effects of each Algorithm.(A) The image restoration situations of red unmanned ships by different methods, (B) The image restoration situations of Marine organisms by different methods, (C) The image restoration situations of sunken ships on the seabed by different methods.

Quantitative analysis was conducted on the underwater occluded images restored by the algorithm and other methods in this study. The adopted indicator is LPIPS, namely perceived loss, which is used to quantify the difference between two pictures [14]. LPIPS is closer to human perception than traditional methods. The lower the LPIPS value, the more similar the two pictures are; conversely, the greater the difference. The calculation formula of the LPIPS value is as follows:

$$d(x,x_0)=\sum_l \frac{1}{H_l W_l}\sum_{h,w} w_l\left(\hat{y}^l_{hw}-\hat{y}^l_{Ohw}\right)_2^2 \tag{6}$$

Among them, d represents the distance between x_0 and x. First, extract features from the l layer and perform unit normalization in the channel dimension. Then, the vector W1 is used to scale the number of activated channels and calculate the L2 distance. Finally, take the average in space and sum up on the channel [15]. In this study, the value of LPIPS is used to represent the difference between the restored image occluded by the underwater mask and the original image.

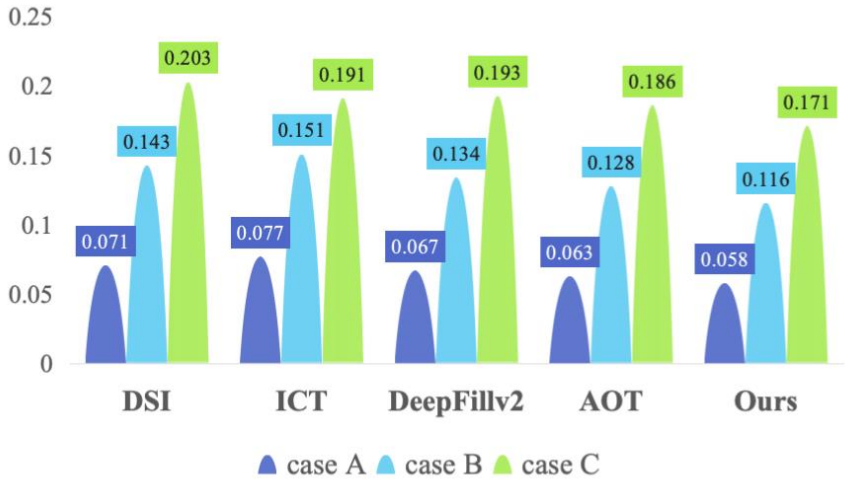

Figure 8. LPIPS Values for Different Cases of each Algorithm

It can be concluded from the quantitative analysis in Figure 8 that the method of this study has a lower LPIPS value and a better restoration effect in the restoration of underwater images than other methods. The results of quantitative analysis are consistent with those of qualitative analysis. Case A is an image covered by scattered and irregular masks. The algorithm in this study can have better reasoning ability. Based on the features around the scattered points and using the better semantic representation of the diffusion model, the distribution law of the covered part can be inferred. DeepFillV2 and AOT are GAN-based methods, and the generated images are smoother. Traditional evaluation metrics such as SSIM and PSNR will fail. LPIPS calculates feature differences by extracting features through neural networks, which is more in line with human perception. Case B belongs to the image covered by regular masks. This research method takes into account the capabilities of texture synthesis and semantic filling to generate images of higher quality. Case C belongs to the image with large area mask coverage, although the results generated by various methods are not very ideal. However, the values of LPIPS of this method are 0.058, 0.116, and 0.171 respectively, all of which are the lowest, indicating that it still has a good restoration effect on the image.

3. Conclusion

This study proposes a two-stage underwater image correction method guided by super-resolution diffusion optimization. Firstly, through the three links of the degradation model, generator and discriminator of the image super-resolution network, the image is degraded and regenerated. It is concluded that the parameters of PSNR, SSIM, UIQM and LPIPS can reach 26.26, 0.74, 2.88 and 0.19 respectively. Compared with other networks, the values of PSNR, SSIM and UIQM have

increased by 2%, 10% and 5% respectively, and the LPIPS parameter can be as low as 0.19, significantly improving its clarity and recognizability. Meanwhile, this study proposes an underwater image restoration framework based on the diffusion model. By introducing a resampling mechanism, the iterative optimization of the generated path is achieved during the reverse diffusion process. This method utilizes the powerful generation ability of the model and introduces the resampling technology of the reverse diffusion process to effectively integrate the generated image information during the reasoning process. This enables the LPIPS value of this model to be as low as 0.058 in the restoration of underwater images, indicating that the restoration method of this study can play a positive role for images occluded by different types of mask inpainting.

Reference

[1] Tang J ,He K ,Tian M , et al.Serialized attention and masked residual network structure for free-form image inpainting[J].Knowledge-Based Systems,2025,316113385-113385.

[2] Kodati P ,Puli K V ,Chiplunkar R , et al.Robust to outlier image inpainting for interface detection in primary separation vessel[J].Journal of Process Control,2025,150103426-103426.

[3] Shao J ,Zhang H ,Miao J .Region gradient-guided diffusion model for underwater image enhancement[J].Machine Vision and Applications,2025,36(2):38-38.

[4] Wang Y, Zhang J, Cao Y, et al. A deep CNN method for underwater image enh ancement[C]//2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017: 1382-1386.

[5] Yunting L ,Zhuang Z ,Binghua S , et al.Super resolution of underwater image based on generating generative adversarial networks[C]//Beijing Institute of Technology (China),2023:

[6] Qiao Lukuan, Liu Aimin, Wang Zijun, et al. Vision inspection System of underwater vehicle based on FPGA [J]. Microprocessor, 2019,45(03):47-51. (in Chinese)

[7] Fei Y ,Lihui Z ,Panwang P , et al.Low bit-rate compression of underwater image based on human visual system[J].Signal Processing: Image Communication,2021,91116082-.

[8] Li J ,Liu Y ,Wu X , et al.Fault diagnosis in open circuit of inverters on electrical discharge milling machines using adaptive Gaussian wavelet convolutional network[J].Measurement,2025,248116856-116856.

[9] Baudouin L ,Imba A ,Mercado A , et al.Lipschitz stability of an inverse problem of transmission waves with variable jumps[J].Inverse Problems,2025,41(4):045007-045007.

[10] Cui SH, Lu Bo, ZHANG Mingyue, et al. 360° image quality and aesthetic evaluation method based on multi-modal fusion [J/OL]. Computer engineering, 1-9 [2025-04-14]. http://kns.cnki.net/kcms/detail/31.1289.tp.20250314.1347.003.html.

[11] Zhang F ,Yuan Q ,Zhang X . Mamba-DDPM-BSA: Diffusion model based boundary sampling algorithm for imbalanced classification [J]. Expert Systems With Applications, 2025, 274 126926-126926.

[12] Huang Shaozong. Based on the depth study of image restoration method research [D]. Jiangxi university of science and technology, 2024. The DOI: 10.27176 /, dc nki. Gnfyc. 2024.000195.

[13] Huang Shaozong. Based on the depth study of image restoration method research [D]. Jiangxi university of science and technology, 2024. The DOI: 10.27176 /, dc nki. Gnfyc. 2024.000195.

[14] Zhao Xiaojie. Based on attribute augmented zero sample study [D]. Nanjing university of science and technology, 2022. The DOI: 10.27241 /, dc nki. Gnjgu. 2022.001434.

[15] Zhang Xukun. Research on cloud removal algorithm of H-alpha full-day solar image based on image conversion model [D]. The central university for nationalities, 2023. DOI: 10.27667 /, dc nki. Gzymu. 2023.000173