A Joint Framework for Underwater Hyperspectral Image Restoration and Target Detection With Conditional Diffusion Model

Qi Li, *Student Member, IEEE*, Jinghua Li, Tong Li, and Yan Feng, *Member, IEEE*

***Abstract*—Underwater hyperspectral imaging is crucial for various marine applications, with underwater hyperspectral target detection (HTD) holding significant importance. However, existing research on underwater HTD is limited, as most methods fail to adequately consider the impact of underwater target spectral variability and image quality degradation. To address these critical issues, we propose a novel joint framework for underwater hyperspectral image restoration and target detection, which is based on a conditional diffusion model. Our proposed framework consists of two main modules: the variable spectral group extraction module, and the joint underwater hyperspectral image restoration and target detection (JURTD) module. The variable spectral group extraction module leverages the conditional diffusion model to extract variable spectral image groups, thereby simulating the diverse range of underwater target spectra. Subsequently, the JURTD module extracts deep features from intrinsic images and the group of variable spectral images. Operating under the dual constraints of image restoration and target detection, this module achieves high-quality restored images and superior detection performance concurrently. Experimental evaluations conducted on both real-world and synthetic datasets demonstrate the effectiveness of our proposed framework in enhancing image quality and improving target detection performance. Moreover, the results indicate that our framework outperforms state-of-the-art methods, underscoring its practical utility and superiority in underwater hyperspectral imaging applications.**

***Index Terms*—Deep learning, diffusion model, hyperspectral imagery (HSI), underwater image restoration, underwater target detection.**

I. Introduction

SIGNIFICANT advancements in underwater target detection technology have been witnessed, playing a pivotal role across diverse domains including national defense security, marine resource development, scientific research, and environmental monitoring [1], [2]. Hyperspectral data exhibit characteristics such as multichannel capacity, high accuracy, and extensive information content. Through the utilization of hyperspectral imaging technology, it is possible not only to observe the morphology of targets, but also to identify and analyze them across spectral dimensions, presenting a distinct advantage over methods reliant on single-wavelength bands [3], [4], [5]. This technology has demonstrated superior performance across various domains, including military management [6], [7], scientific agriculture [8], [9], environmental stewardship [10], geographic surveying [11], urban planning [12], and oceanic remote sensing [13], [14]. However, underwater hyperspectral imageries (HSIs) typically suffer from diminished image quality due to the absorption and radiation properties of water bodies, as well as particle interference, thereby impacting the accuracy of underwater target detection. Addressing this critical challenge will significantly enhance the accuracy and reliability of the underwater target detection method. Therefore, it is necessary to develop the joint problem of underwater HSI restoration and target detection.

Current research primarily focuses on restoring HSIs on land or RGB underwater images, leaving a notable gap in the restoration of underwater HSIs. Çelebi and Ertürk [15] combined wavelet noise reduction with empirical model decomposition to develop an underwater degrading image denoising algorithm. He et al. [16] introduced the concept of the dark channel prior for addressing foggy weather acquisition image defogging problems. Chang et al. [17] further enhanced this approach by combining the dark-channel prior theory with wavelength compensation, considering the nonuniform illumination in underwater images. In addition, Land et al. [18] proposed the Retinex theory, serving as the foundation for achieving color uniformity. Jobsan et al. [19] extended this theory by presenting a multiscale Retinex image enhancement method with color recovery. Yuan et al. [20] involved learning a nonlinear end-to-end mapping between noise and clean HSIs through a combined spatial–spectral deep convolutional neural network. Mao et al. [21] and Zhang et al. [22] employed convolutional neural networks to extract intrinsic and disparity image features, aiming to circumvent complex a priori constraints. Their approach demonstrated advanced performance on natural images. Furthermore, Yu et al. [23] proposed a novel dual-stream transformer for HSI restoration. The dual-stream feed-forward network operates to extract global signals and local details in parallel branches. In conclusion, both physical and deep learning methods have demonstrated

Received 19 March 2024; revised 10 June 2024; accepted 17 September 2024. Date of publication 19 September 2024; date of current version 2 October 2024. This work was supported by the Science and Technology on Electromechanical Dynamic Control Laboratory under Grant 614260124020202. *(Corresponding author: Jinghua Li.)*

Qi Li, Tong Li, and Yan Feng are with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China (e-mail: lq132069@163.com; LeeTongStudy@163.com; sycfy@nwpu.edu.cn).

Jinghua Li is with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, ChinaScience and Technology on Electromechanical Dynamic Control Laboratory,, China (e-mail: ljhy6331@nwpu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3464557

remarkable efficacy in HSI restoration. However, research dedicated specifically to underwater HSI restoration remains absent.

Significant advancements have been made in deep generative models, encompassing autoregressive models, variational autoencoders, generative adversarial networks, and diffusion models. Notably, diffusion models have emerged as prominent tools in image generation [24] and have found utility in various image restoration tasks [25], including superresolution [26], painting [27], and denoising [28]. A notable contribution by Wu et al. [29] involves the proposal of an HSI superresolution method utilizing a conditional diffusion model. This method integrates high-resolution multispectral images with corresponding low-resolution HSIs, generating high-resolution HSIs through iterative refinement initiated with pure Gaussian noise. Similarly, Miao et al. [25] introduced a self-supervised diffusion model for HSI restoration. Furthermore, several studies have leveraged the diffusion model for HSI classification. Chen et al. [30] proposed SpectralDiff, a generative framework for HSI classification utilizing the diffusion model. The diffusion model demonstrates significant potential holds promise in the realm of HSI processing.

Hyperspectral target detection (HTD) methods have reached a high level of maturity [31], [32], [33], [34], [35], a novel weighted Cauchy distance graph and local adaptive collaborative representation detection method has been proposed, which adaptively utilizes spatial–spectral information to adjust the detection probability [36]. A hyperspectral time-series target detection method based on spectral perception and spatial–temporal tensor (SPSTT) decomposition has been introduced, effectively utilizing prior spectral information and spatial-temporal overall structural information [37]. However, their application in underwater target detection often yields unsatisfactory results. In reality, the spectral profile of a target can undergo significant changes as its depth increases due to the water column, thereby impacting detector performance. Garaba et al. [38] conducted laboratory-based reflectance measurements on submerged large plastics across various water clarity conditions and submergence depths. Papakonstantinou et al. [39] performed experiments to generate hyperspectral datasets for eight submerged plastics. Jay et al. [40] incorporated a bathymetric model to correct the spectral aberrations induced by water, subsequently employing traditional land-based algorithms for underwater target detection. Gillis [41] utilized both physical and nonlinear mathematical models to estimate the underwater target space within a given scene. Several studies have also employed deep learning methodologies to facilitate underwater HTD. Qi et al. [42] introduced depth information into their methodology by developing an underwater target detection network that integrates a bathymetric model with an autoencoder. Qi et al. [43] proposed the self-improving underwater target detection framework. Li et al. [44] introduced the transfer-based underwater target detection framework. Then, Li et al. [45] proposed the TDSS-UTD framework, where the spectral and spatial features are extracted, and the spatial–spectral features are harnessed for target detection. These methods have achieved good performance in underwater HTD.

Despite the favorable outcomes achieved with the aforementioned methods in underwater HTD, they still face some limitations and drawbacks. First, during the process of underwater hyperspectral imaging, factors such as water turbidity and noise interference from water particles contribute to the presence of noise interference, fogging, blurring, and low contrast in the acquired underwater HSIs. Regrettably, existing methods predominantly overlook the processing of low-quality underwater HSIs. Second, the existing underwater HTD techniques fail to account for the influence of underwater HSI quality. Research focusing on terrestrial imagery has demonstrated that image enhancement techniques can effectively enhance image quality, thereby enhancing the performance of target detection. Therefore, the lack of consideration for underwater HSI quality represents a notable gap in current research efforts.

In this article, we aim to overcome the aforementioned limitations by proposing a joint framework for underwater HSI restoration and target detection based on the conditional diffusion model. This framework primarily comprises a variable spectral group extraction module and a joint underwater hyperspectral image restoration and target detection (JURTD) module. In the variable spectral group extraction module, we leverage the conditional diffusion model to learn the structural features of underwater HSIs. During the sampling process, a priori variable spectral image groups are extracted to emulate the diverse spectral changes of underwater targets in real underwater environments. Furthermore, the JURTD module is designed as a two-stream network. This module outputs restored underwater HSIs and target detection results simultaneously. By jointly optimizing image restoration and target detection, our network can produce images suitable for both visual inspection and computer perception, thereby achieving detection-oriented restoration. The main contributions of this article are as follows.

1) We introduce the variable spectral group extraction module to mitigate the variability of underwater target spectra. To the best of our knowledge, this is the first time that conditional diffusion models have been applied to underwater HTD.
2) To achieve high-quality HSIs and superior detection performance simultaneously, we propose the JURTD module. We devise a hybrid loss function based on the dual constraints of target detection and image restoration. To the best of our knowledge, this is the first time that the issue of underwater HSI restoration is jointly optimized with underwater HTD.
3) Numerous experiments conducted on two real-world datasets and one synthetic dataset demonstrate the superiority of the proposed method. Moreover, we also validate the excellent performance by ablation experiments and robustness analysis.

The rest of this article is organized as follows. In Section II, a brief overview of background and related work is provided. Section III presents the detailed framework proposed in this article, with its key components and methodologies. Section IV discuss the experimental results and related analysis. In addition, we take the ablation experiments and robustness analysis of the framework. Finally, Section V concludes this article.

II. Related Work

A. Underwater HSI

There are two primary distinctions between underwater HSI and land HSI in the imaging process. First, in land HSI, it is typically assumed that the observed spectral signals of the target and those of the background are independent of each other. However, this assumption does not hold true for underwater targets. Second, in underwater HSI, any light reflected by the underwater target and captured by the sensor must traverse through the surrounding water column. Consequently, the measured underwater target spectra are significantly influenced by the optical properties of the water column and the depth of the target. The radiative transfer equation for an underwater target can be expressed as follows:

$$r(\lambda) = r^{\mathrm{dp}}(\lambda)\left[1 - e^{-K(\lambda)H}\right] + \frac{1}{\pi}\rho(\lambda)e^{-K(\lambda)H} \tag{1}$$

where $r^{\mathrm{dp}}(\lambda)$ represents the optical deep-water remote sensing reflectance ratio, while $K(\lambda)$ denotes the effective attenuation coefficient of the water body. The variable H signifies the depth at which the target is situated, and $\rho(\lambda)$ indicates the reflectance of the land target unaffected by water interference. It can be seen that the actual collection of underwater target spectra is jointly determined by the two parts: the water body and the target. The main influencing factors are the effective attenuation coefficient of the water body, the depth of the target, the reflectance ratio of the optical deep-water remote sensing, and the reflectance of the land target.

The mechanism underlying the influence of the absorption and scattering characteristics of the water body on collected underwater target spectra directly impacts the accuracy of underwater target detection and the recovery effect of underwater HSIs. The absorption and scattering effects of the water body lead to significant distortion of the target spectral curve, resulting in a substantial deviation from the spectral curve of the same target on land. In addition, underwater targets composed of different materials exhibit varying spectral curves depending on their depth within the water environment. This discrepancy often results in different spectra for the same object in an HSI compared to those observed in a marine environment. To mitigate the variability of underwater target spectra, which is heavily dependent on depth and water optical properties, this article adopts the conditional diffusion model. This model facilitates the acquisition of spectral variation groups, aiming to suppress the influence of water body particle interference and unknown spectral variations, thereby enhancing the quality of underwater HSIs.

B. Denoising Diffusion Probability Model

The denoising diffusion probabilistic model belongs to a class of generative models that primarily learn a Markov chain, gradually transitioning the Gaussian noise distribution to the training data distribution. The diffusion process comprises both a forward process and a reverse process. Following a variational schedule $\beta_1, \ldots, \beta_T$, Gaussian noise is initially added to the HSI, and then, in the backward process, the added noise is gradually eliminated using a noise estimation network. In this context, H and W represent the height and width of the hyperspectral image HSI $\in \mathbb{R}^{H\times W\times C}$, respectively, while C denotes the number of spectral channels. Multiple image patches, denoted as $X \in \mathbb{R}^{k\times k\times C}$, are extracted from the entire HSI.

1) Forward Diffusion Process: The forward diffusion process operates as a stationary Markov chain. Over the total T steps, in accordance with the variational schedule $\beta_1, \ldots, \beta_T$, it progressively corrupts $X \sim p(X)$ by adding Gaussian noise

$$p(X_t|X_{t-1}) = N(X_t; \sqrt{1-\beta_t}X_{t-1}, \beta_t I) \tag{2}$$

$$p(X_{1:T}|X_0) = \prod_{t=1}^{T} p(X_t|X_{t-1}) \tag{3}$$

where X_{t-1} and X_t denote the states at step $t-1$ and step t, and I is the standard normal distribution. $\sqrt{1-\beta_t}X_{t-1}$ and $\beta_t I$ are the mean and variance of the conditional distribution $p(X_t|X_{t-1})$, respectively.

The Gaussian diffusion process can be marginalized by directly sampling the intermediate term X_t from the original data X_0 via the following:

$$p(X_t|X_0) = N(X_t; \sqrt{\overline{\alpha}_t}X_0, (1-\overline{\alpha}_t)I) \tag{4}$$

which can also be expressed in closed form

$$X_t = \sqrt{\overline{\alpha_t}}X_0 + \sqrt{1-\overline{\alpha_t}}z_t \tag{5}$$

where $\alpha_t = 1-\beta_t$, $\overline{\alpha_t} = \prod_{i=1}^{t}\alpha_i$, and $z_t \sim N(0, I)$ has the same data dimension as the original data X_0 and the intermediate state X_t.

2) Reverse Process: The reverse process concentrates on the joint distribution $q_\theta(X_{0:T})$, which also operates as a Markov chain aimed at learning Gaussian noise removal. It starts with a priori standard normal distribution $q(X_T) = N(X_T; 0, I)$

$$q_\theta(X_{0:T}) = q(X_T)\prod_{t=1}^{T} q_\theta(X_{t-1}|X_t) \tag{6}$$

$$q_\theta(X_{t-1}|X_t) = N(X_{t-1}; \mu_\theta(X_t, t), \widetilde{\beta}_t I) \tag{7}$$

where $\mu_\theta(X_t, t)$ represents the mean of the conditional distribution $q_\theta(X_{t-1}|X_t)$, while $\widetilde{\beta}_t$ denotes the variance term. The parametric distribution can be expressed as follows:

$$\mu_\theta(X_t, t) = \frac{1}{\sqrt{\alpha_t}}\left(X_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(X_t, t)\right) \tag{8}$$

$$\widetilde{\beta}_t = \frac{1-\overline{\alpha_{t-1}}}{1-\overline{\alpha_t}}\beta_t \tag{9}$$

where $\epsilon_\theta(X_t, t)$ is a noise estimation network with inputs of time step t and intermediate state X_t. The reverse process can be specifically represented as

$$X_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(X_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha_t}}}\epsilon_\theta(X_t, t)\right) + \widetilde{\beta}_t z \tag{10}$$

where $z \sim N(0, I)$. To accurately estimate the noise added in the forward diffusion process, the noise estimation network $\epsilon_\theta(X_t, t)$ needs to be well-trained.

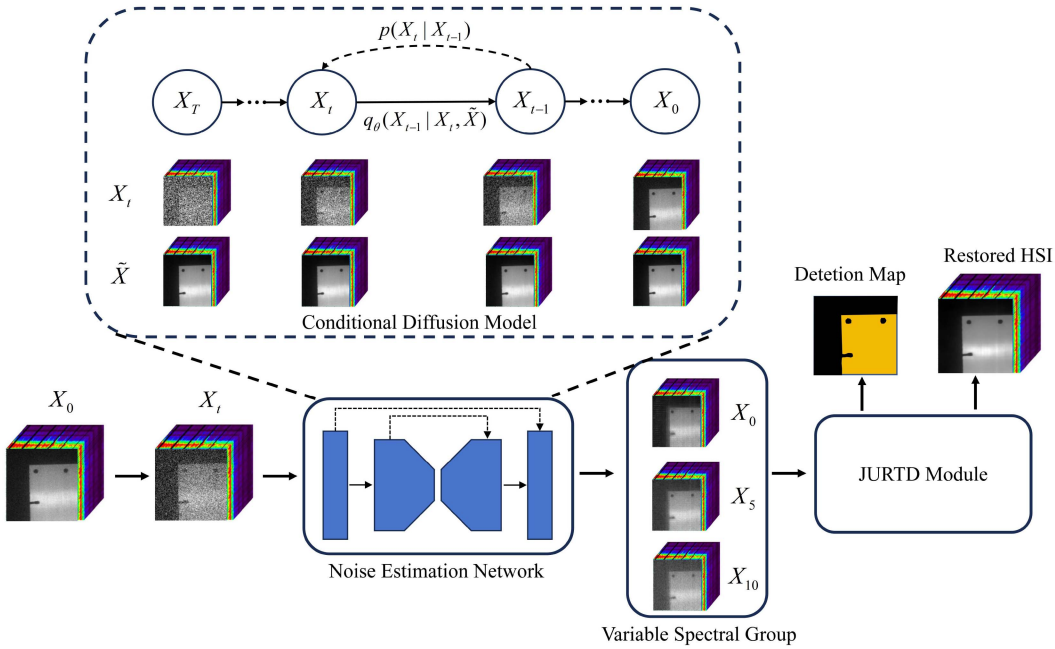


Fig. 1. Overall architecture of the joint framework for underwater hyperspectral image restoration and target detection with the conditional diffusion model. The framework comprises the variable spectral group extraction module and the JURTD module. First, the principle component image of the underwater HSI is fed into the conditional diffusion model to obtain the variable spectral group. After that, the variable spectral group is input to the JURTD module. Finally, we can obtain the detection map and restored HSI from the JURTD module.

III. Methodology

In this section, we will provide a detailed description of our proposed joint framework for underwater HSI restoration and target detection with the conditional diffusion model, as shown in Fig. 1. In the variable spectral group extraction module, the full potential of the conditional diffusion model to discern structural features of the image is harnessed. Its guiding condition is the principal component HSI (PHSI) derived from the original underwater HSI collected during the experiment. The PHSI encapsulates the primary information of the original HSI while preliminarily filtering out noise interference. Utilizing this component, the module guides image denoising, generates diverse sample data, and acquires the a priori variable spectral group, simulating the variations in underwater target spectral curves. The JURTD module adopts a two-stream network design, leveraging the shared features of intrinsic images and variable spectral images within the deep feature. The input to the JURTD module comprises the variable spectral image group, while the output includes both the detection probability map and the restored HSI. Furthermore, a hybrid loss function is devised based on the dual constraints of target detection and image restoration. This function is instrumental in guiding the training of JURTD, facilitating the attainment of high-quality images and superior detection performance concurrently. The detailed descriptions of these two modules are provided separately in the subsequent sections.

A. Variable Spectral Group Extraction Module

Within the conditional diffusion model, the PHSI of the experimentally acquired underwater HSI serves as the guiding condition. This component aids in guiding image denoising and generating diverse sample data. The fundamental concept revolves around learning the inference process of the original underwater HSI under the PHSI condition, along with acquiring a series of state transitions to transform noise. This process is primarily divided into the forward diffusion process, involving noise addition, and the reverse process, focused on noise removal. In the subsequent part, we will delve into the intricacies of the forward diffusion process, the reverse process, the loss function of the noise estimation network, and the training procedure in meticulous detail.

1) Forward Diffusion Process: Under the variance schedule $\beta_1, \ldots, \beta_T$, the forward diffusion process operates as a stationary Markov chain, progressively adding Gaussian noise to X_0. Mathematically, this process can be expressed as

$$X_t = \sqrt{\overline{\alpha_t}} X_0 + \sqrt{1 - \overline{\alpha_t}} z_t \tag{11}$$

where X_0 represents the original HSI, and X_t denotes the state at time step t forward.

2) Reverse Process: To initiate the reverse process, we extract matched pairs of data distributions $(X, \widetilde{X})$ from the original underwater HSI along with its PHSI. These instances are then utilized as inputs to the reverse process

$$q_\theta(X_{0:T}|\widetilde{X}) = q(X_T) \prod_{t=1}^{T} q_\theta(X_{t-1}|X_t, \widetilde{X}) \tag{12}$$

$$q_\theta(X_{t-1}|X_t, \widetilde{X}) = N(X_{t-1}; \mu_\theta(X_t, \widetilde{X}, t), \widetilde{\beta}_t I). \tag{13}$$

The mean and variance of the conditional distribution $q_\theta(X_{t-1}|X_t, \widetilde{X})$ are determined as follows:

$$\mu_\theta(X_t, \widetilde{X}, t) = \frac{1}{\sqrt{\alpha_t}}\left(X_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha_t}}}\epsilon_\theta(X_t, \widetilde{X}, t)\right) \quad (14)$$

$$\widetilde{\beta}_t = \frac{1-\overline{\alpha_{t-1}}}{1-\overline{\alpha_t}}\beta_t. \quad (15)$$

Integrating the aforementioned formulas, one obtains the sampling process

$$X_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(X_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha_t}}}\epsilon_\theta(X_t, \widetilde{X}, t)\right) + \widetilde{\beta}_t z \quad (16)$$

where X_t represents the state at t steps in reverse, $\epsilon_\theta(X_t, \widetilde{X}, t)$ is the noise estimation network, and $\widetilde{X}$ denotes the PHSI of the original HSI. It is noteworthy that both the inputs X_t and $\widetilde{X}$ of the noise estimation network are combined via convolutional blocks. Moreover, the input t is utilized to generate high-level features via positional coding, which guides the feature extraction of the model. The architecture of the noise estimation network is referred from the literature [28].

3) Hybrid Loss Function: The performance of the noise estimation network directly impacts the reverse process of the diffusion model, determining whether the appropriate X_0 can be inferred. Hence, it is crucial to train an effective noise estimation network. To achieve this, we devise a hybrid loss function to quantify the disparity between the actual added noise in the forward diffusion process and the predicted noise outputted by the noise estimation network. This hybrid loss function can be expressed as follows:

$$L_{\text{diff}} = L_1 + L_2 + L_S \quad (17)$$

where L_1 represents a magnitude loss term aimed at minimizing the magnitude difference between the ith actual additive noise $\epsilon^{(i)}$ from the forward diffusion process and the predicted noise $\epsilon_\theta(X_t^{(i)}, \widetilde{X}, t)$ generated by the noise estimation network. The term L_1 is defined as

$$L_1 = \sum_i^N ||\epsilon^{(i)} - \epsilon_\theta(X_t^{(i)}, \widetilde{X}, t)||_1. \quad (18)$$

The term L_2 denotes the mean square error between the actual added noise and the predicted noise. This component serves to expedite the training process and mitigate the risk of vanishing gradients

$$L_2 = \frac{1}{2}\sum_i^N ||\epsilon^{(i)} - \epsilon_\theta(X_t^{(i)}, \widetilde{X}, t)||_2. \quad (19)$$

The term L_S represents the spectral angular distance between the actual added noise and the predicted noise. This metric captures the distinction in trend between the two curves

$$L_S = \sum_i^N \text{SAM}(\epsilon^{(i)}, \epsilon_\theta(X_t^{(i)}, \widetilde{X}, t)) \quad (20)$$

where

$$\text{SAM}(a, b) = \arccos\left(\frac{< a, b >}{||a||, ||b||}\right). \quad (21)$$

Algorithm 1: Conditional Diffusion Model Training.

Data: HSI and PHSI instance pairs $(X, \widetilde{X})$
Result: Well-trained noise estimation network ϵ_θ
while *epoch* $\leq 10,000$ *and not converged* **do**
  $(X, \widetilde{X}) \sim p(X, \widetilde{X})$;
  $t \sim Uniform(\{1, \ldots, T\})$;
  $\epsilon \in N(0, I)$;
  Take gradient descent step on
    $\nabla_\theta||\epsilon - \epsilon_\theta(\sqrt{\overline{\alpha_t}}X_0 + \sqrt{1-\overline{\alpha_t}}z_t, \widetilde{X}, t)||^2$
end

The training algorithm for the conditional diffusion model is outlined in Algorithm 1. Upon achieving a well-trained noise estimation network, optimization of the underwater HSI proceeds by reverse sampling under the guidance of the PHSI condition $\widetilde{X}$. When the sampling steps t in the conditional diffusion model are set as 10, 5, and 0, we consider these feature images to contain information resistant to interference. These states exhibit a certain degree of change compared to the original underwater HSI, retaining the original information while enhancing resistance to noise. Consequently, we aggregate them into the group of variable spectral images, utilizing them as input for the JURTD module. This approach simulates the spectral variation phenomenon of target spectra in the underwater environment, thereby mitigating the influence of spectral variance on underwater target acquisition during the actual acquisition process.

B. JURTD Module

The block diagram of the JURTD module is depicted in Fig. 2. The JURTD module comprises dual-branch networks, upper and lower, encompassing the encoder module, feature fusion module, and decoder module. The input to the JURTD module is the group of variable spectral images, representing the variable images. Through a convolution operation, the two images $X_{10}, X_5 \in \mathbb{R}^{H\times W\times C}$ are fused, and then, fed into the lower branch of the JURTD module to comprehensively explore high-level features of the changing spectra. Meanwhile, the intrinsic image X_0 is directly input to the upper branch. This JURTD module maximizes the complementary information between the intrinsic image and the variable spectral images to identify common features and distinguish characteristics between the two types of images. To extract critical features while minimizing computation, a dimensionality reduction module is initially incorporated into the dual-branch network, comprising

$$I_1 = f(X_0) \quad (22)$$

$$I_2 = f(\text{Conv}(X_5, X_{10})) \quad (23)$$

where $f(\cdot)$ represents the dimensionality reduction module, which comprises a convolutional layer, a batch normalization layer, and a LeakyReLU layer. $\text{Conv}(X_5, X_{10})$ denotes the convolution operation after concatenating X_5 and X_{10}. I_1 and I_2 are the feature maps obtained after dimensionality reduction.

1) Dual-Branch Encoder: The features extracted in the encoder module primarily serve to enhance both spectral and

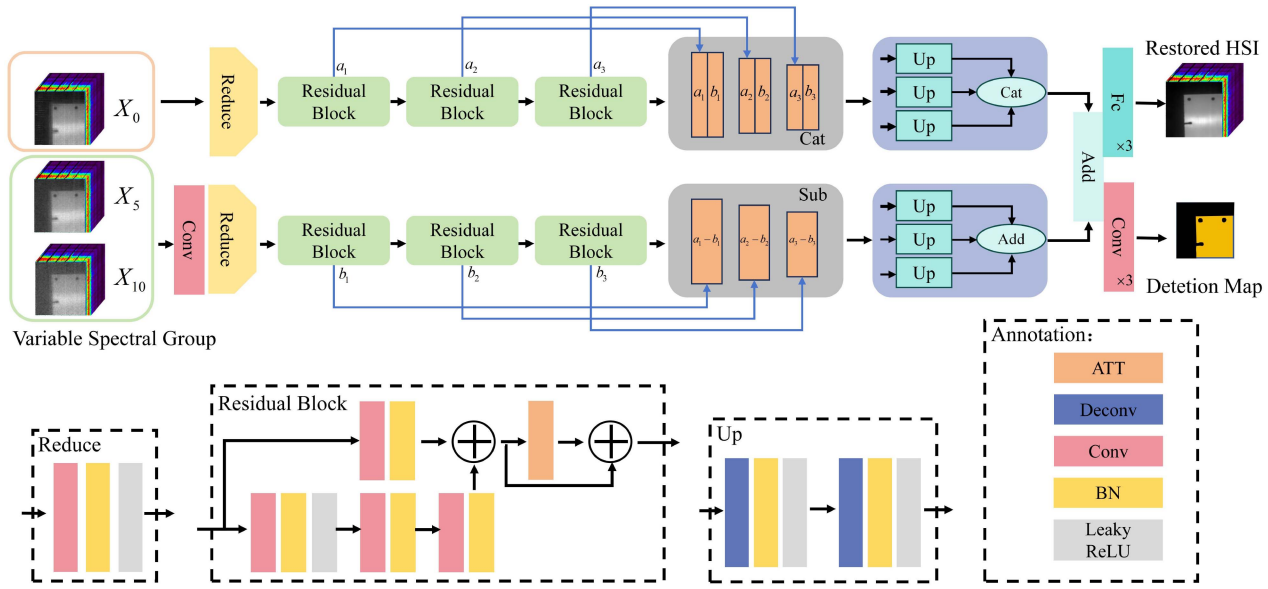


Fig. 2. Block diagram of the JURTD module. The JURTD module takes the variable spectral group as input and generates the restored HSI and detection map as its outputs.

spatial correlations, allowing neurons to flexibly capture texture, spatial structure, and spectral features in the input data. To extract multiscale information, the encoder is designed with three residual blocks, each comprising convolutional layers, batch normalization layers, LeakyReLU layers, and an attention mechanism

$$O'_j = f_0(f(I_i)) + f_0(I_i) \tag{24}$$

$$O_j = O'_j + \mathrm{ATT}(O'_j) \tag{25}$$

where $f_0(\cdot)$ represents the dimensionality reduction module without the LeakyReLU layer, while $\mathrm{ATT}(\cdot)$ denotes the attention mechanism. Sequentially, the attention mechanism includes the spatial attention mechanism, spectral attention mechanism, and spatial–spectral joint attention mechanism within the three residual blocks. To facilitate representation and computation in the feature fusion module, the output O_j of residual block j in the upper branch network is denoted as a_j, and the output O_j of the residual block j in the lower branch network is denoted as b_j.

2) Feature Fusion: Small-scale feature maps are adept at capturing detailed information, whereas large-scale feature maps excel at capturing macroscopic structural and semantic information. Therefore, the multiscale features extracted by the dual-branch encoder are fused to enhance the robustness and generalization of the subsequent dual-branch decoder in data processing. To achieve optimal fusion of the extracted multiscale information, concatenation and differencing operations are employed to create a richer and integrated feature representation. The concatenation operation merges the multiscale feature images a_j and b_j from the upper and lower branches of the network in the encoder, thereby concatenating the features of the variable spectral image and the intrinsic image. Conversely, the differencing operation discerns the multiscale feature images a_j and b_j to highlight the feature disparities between the variable spectral image and the intrinsic image.

The feature fusion approach preserves the integrity of the information within individual features while acknowledging the interrelationships and interactions between the intrinsic features and the variable spectral features. The fused features of the variable spectral features and the intrinsic features can be expressed as follows:

$$c_j = Cat(a_j, b_j), \quad j = 1, 2, 3. \tag{26}$$

The multiscale feature image differencing operation extracts the information of the difference between the intrinsic feature and the variable spectral feature. This operation accentuates the substantial differences between them, leading to a more comprehensive representation of the features

$$s_j = a_j - b_j, \quad j = 1, 2, 3. \tag{27}$$

The characteristic assists the dual-branch decoder in capturing both detailed information and global context. In addition, the concatenation and differencing strategy helps compensate for information loss to some extent and enhances feature representation capability.

3) Dual-Branch Decoder: The multilevel fusion features and difference features obtained from the feature fusion module are concatenated or summed using the up-sampling method. With the concatenation method, the channel dimension of the fused data triples compared to the original dimension

$$f_c = \mathrm{Cat}(Up(c_1), \mathrm{Up}(c_2), \mathrm{Up}(c_3)) \tag{28}$$

$$f_a = \mathrm{Up}(s_1) + \mathrm{Up}(s_2) + \mathrm{Up}(s_3) \tag{29}$$

where f_c denotes the concatenated feature after up-sampling the multilevel fusion features to the same dimension, and f_a represents the summation feature after up-sampling the multilevel difference features to the same dimension.

Finally, a complete feature map is formed by summing the concatenated feature f_c with the summation feature f_a via the summation operation. This represents the advanced feature map after learning the variation spectral features and intrinsic features. The advanced feature image is then mapped into X_{rec}, a reconstructed underwater HSI by three convolution operations (this branch is called the image restoration branch subnetwork). In addition, the advanced features are represented as target detection probability values Y by three fully connected layers (this branch is called the target detection branch subnetwork)

$$X_{\text{rec}} = \text{MConv}(f_c + f_a) \tag{30}$$

$$Y = \text{MFc}(f_c + f_a) \tag{31}$$

where $\text{MConv}(\cdot)$ denotes three consecutive convolution operations, and $\text{MFc}(\cdot)$ denotes three consecutive fully connected layers.

C. Double Constraint Loss Function

In this article, we propose an optimization formulation for the joint problem of underwater HSI restoration and target detection, aiming to find the optimal model by training the network under the double constraints of image restoration and target detection. We formulate the model as a dual optimization problem

$$\begin{aligned} &\min_{w_d} L^d(\varphi(R^*; w_d)) \\ &\text{s.t. } R^* \in \text{argmin}_{w_r} L^r(\phi(X_0, X_5, X_{10}; w_r)) \end{aligned} \tag{32}$$

where L^d is the training loss function specific to target detection, φ represents the target detection branch network with learnable parameters w_d, L^r is the training loss function specific to image restoration, ϕ denotes the image restoration branch network with learnable parameters w_r, and R^* is the restored underwater HSI in the optimal model.

To train the JURTD module, we design the hybrid loss function, which can be expressed as

$$L = \alpha_1 L^d + \alpha_2 L^r, \quad \alpha_1 + \alpha_2 = 1 \tag{33}$$

where α_1 and α_2 are adaptive weight coefficients, and the design L^r is consistent with the loss function L_{diff} in the conditional diffusion model.

The target detection loss function L^d adopts the classical cross-entropy loss function, expressed as

$$L^d = \frac{1}{N} \sum_i -[y_i \cdot \log(p_i) + (1 - y_i) \cdot \log(1 - p_i)] \tag{34}$$

where y_i denotes the label of the sample i, and p_i is the probability that sample i is predicted to be the target.

With this strategy, we can successfully tackle the joint problem of underwater HSI restoration and target detection. This approach not only produces visually appealing underwater HSIs but also yields accurate underwater target detection results given the trained network parameters. The training process of the JURTD module is outlined in Algorithm 2.

Algorithm 2: JURTD Model Training.

Data: Noisy underwater HSI, well-trained noise estimation network ϵ_θ
Result: Restored HSI and Detection Map

HSI and PHSI instance pairs $(X, \widetilde{X})$;
while $epoch \leq 100$ **do**
  $(X, \widetilde{X}) \sim p(X, \widetilde{X})$;
  for $t = T, \ldots, 1$ **do**
    $z_t \sim N(0, I)$ if $t > 1, else$ $z_t = 0$;
    $X_{t-1} = \frac{1}{\sqrt{\alpha_t}}(X_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(X_t, \widetilde{X}, t)) + \widetilde{\beta}_t z$;
  end
  fed (X_0, X_5, X_{10}) into JURTD;
  perform a single gradient descent step for $\nabla_\vartheta[\alpha_1 L^d(X, X_{rec}) + \alpha_2 L^r(Label, Y)]$
end

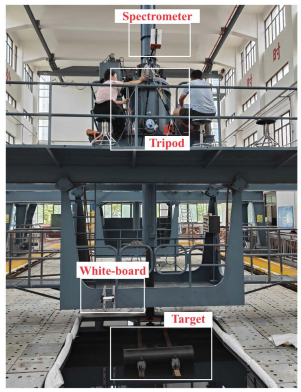


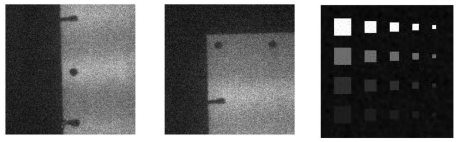
Fig. 3. Experimental environment for indoor pool dataset acquisition.

IV. Experiment

In this section, we will validate the effectiveness and associated analysis of the joint framework for underwater HSI restoration and target detection based on the conditional diffusion model proposed in this article. We begin by describing the experimental setup, including the experimental dataset, evaluation metrics, and training details information. Then, we analyze the effectiveness of the JURTD module through image restoration analysis and detection analysis. Next, we conduct ablation experiments to verify the impact of each module, and finally, we analyze the performance of the conditional diffusion model and discuss the robustness of the proposed framework under different signal-to-noise ratio (SNR) noise interference.

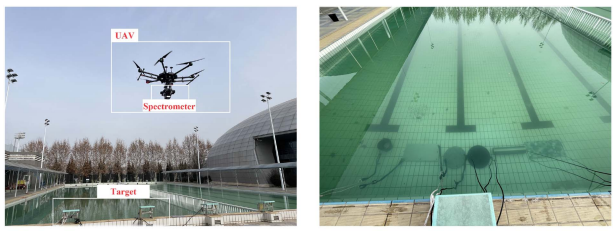
A. Experimental Dataset

1) Indoor Pool Dataset: The experimental dataset was acquired in an anechoic pool at the School of Navigation, Northwestern Polytechnical University, Shaanxi Province, China. The experimental environment, depicted in Fig. 3, features a pool measuring 20 m in length, 8 m in width, and 7 m in depth.



(a) (b) (c)

Fig. 4. Gray-scale image of the acquired indoor pool dataset and synthetic HSI. (a) InPool1. (b) InPool2. (c) SynHSI.



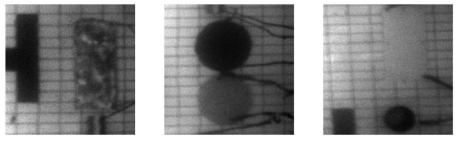
(a) (b)

Fig. 5. Experimental environment for outdoor pool dataset acquisition. (a) UAV carrying the spectrometer above the pool. (b) Target placed at 0.89-m depth.

Hyperspectral data were captured using a GaiaField Portable spectral imager with a spectral resolution of 2.8 nm and a range spanning from 400 to 1000 nm. Specifically, the GaiaField hyperspectrometer operates within a spectral band range of 400–1000 nm with a spectral resolution of 2.8 nm. Data collection involved various depths and targeted three different materials: iron, wood, and stone. To precisely control the depth of the underwater targets, a lift table was employed, allowing for fine adjustments of the target position underwater. The spectrometer recorded digital number (DN) values, which was later corrected by whiteboard calibration to enhance applicability. Since the spectrometer was positioned only 4.2-m away from the water surface during indoor experiments, atmospheric correction was unnecessary. In this article, two HSIs sized 200×200 with 120 spectral bands were extracted from the experimental dataset, captured at a depth of 1 m, and named InPool1 and InPool2, as illustrated in Fig. 4.

2) Outdoor Pool Dataset: To explore the development of underwater HTD models in real outdoor settings, we collected underwater multitarget HSIs at the outdoor pool of Northwestern Polytechnical University, as depicted in Fig. 5(a). Utilizing a hyperspectral camera mounted on an UAV, the pool was photographed, with each target positioned at a water depth of 0.89 m, as depicted in Fig. 5(b). The UAV was configured to capture HSIs from varying flight altitudes 20, 50, and 100 m, to obtain data from different viewpoints and altitudes, enhancing the comprehensive analysis of underwater target spectral characteristics. In this study, three pairs of HSIs were extracted from the data captured at a flight altitude of 50 m. Each image is sized 200×200 and comprises 120 spectral bands. These images, named OutPool1, OutPool2, and OutPool3, are illustrated in Fig. 6.

3) Synthetic Hyperspectral Image: Prior to actual data collection, it is customary to initially assess the performance of the proposed model by synthesizing data. The synthetic HSI generated has dimensions of 100×100 with 120 spectral bands, spanning a band range of 400–740 nm, as depicted in Fig. 4(c). The water region in the image is extracted from the pure water region of the 1.9-m indoor pool dataset. Leveraging the underwater target spectral radiative transfer equation introduced in our previous work, we designated iron as the experimental target. By inputting the assumed depth of the target, the effective attenuation coefficient of the water body, and the land target spectral curve of iron into this model, we obtained the theoretical underwater target spectral curve. Varying the sizes of target patches (13×13, 9×9, 7×7, 5×5, 3×3), and different depths at which the target is positioned (0.1 m, 0.5 m, 1.0 m, 1.6 m), the HSI is denoted as SynHSI.



(a) (b) (c)

Fig. 6. Gray-scale image of the acquired outdoor pool dataset. (a) OutPool1. (b) OutPool2. (c) OutPool3.

B. Evaluation Metrics

1) Restoration Evaluation Metrics: In this article, both subjective and objective evaluations are employed to assess the image restoration effectiveness. Subjective evaluation involves observers rating the restoration quality based on predefined criteria and their own expertise through visual inspection. Objective evaluation, on the other hand, relies on mathematical calculations. The objective evaluation metrics utilized are peak signal-to-noise ratio (PSNR), structure similarity (SSIM), and spectral angle mapper (SAM).

2) Detection Evaluation Metrics: When evaluating the target detection effect, we employ the classical receiver operating characteristic (ROC) curve and calculate the area under the ROC curve (AUC) as evaluation metrics. These metrics effectively assess the accuracy of the target detection model. To construct the ROC curve [46], we vary the classification threshold from 0 to 1 based on the output of the detection network. At each threshold value, we compute the probability of detection (PD) and the false alarm rate (PF)

$$\mathrm{PF} = \frac{N_f}{N_b} \quad \mathrm{PD} = \frac{N_c}{N_t} \tag{35}$$

where N_f is the number of false alarm pixels, N_b is the total number of background pixels, N_c is the number of correct detection target pixels, and N_t is the number of total true target pixels.

C. Training Details

1) Experimental Equipment: Our experiments were conducted on a system equipped with an Intel Core i9-12900KF CPU, GeForce RTX 3080 Ti GPU, 64-GB RAM, and the

Windows 11 operating system. The code was developed using Python 3.9 and the deep learning framework PyTorch 1.12.1.

2) Training Setup: HSIs are preprocessed with normalization before being fed into the model to expedite model convergence. In training the conditional diffusion model, we set the diffusion time step T to 400, the data patch size to 16×16 with 120 bands, the batch size to 32, and the total number of training epochs to 10 000. The initial learning rate is 0.0001, and it decreases by a factor of 0.8 after every 2000 epochs. The network optimizer employed is the Adam algorithm.

For training the JURTD module, the data patch size remains 16×16 with 120 bands. The total dataset sizes used for the experiments are 80 000 pixels from the indoor pool dataset, 120 000 pixels from the Outdoor Pool dataset, and 10 000 pixels from the synthetic hyperspectral image. The number of training datasets is 5% of all pixel points, with a batch size of 64. The total training epoch is 100, with a learning rate set to 0.0002. The optimizer used is also the Adam algorithm.

3) Detection Comparison Method: To further validate the effectiveness of our proposed JURTD module, we compared our method in this article with five detection methods, including three classical methods (ACE [34], RX [31], and MF [32]), and two improved novel methods (HSI-CNN [33] and MSAM [35]).

D. Performance Analysis

1) Image Restoration Analysis: In Section III, we introduce the variable spectral image extraction module, which restores the input underwater HSI based on the conditional diffusion model. As the sampling step approaches 0, the sampled image converges toward the original image during the reverse process. Since the model is trained to diffuse under the guidance of the PHSI of the original underwater HSI, image restoration can occur to some extent during the reverse process, resulting in a cleaner diffusion image. We define the state of the image when t is 0 as the diffusion underwater HSI. One of the outputs of the JURTD module is the restored underwater HSI, which further refines the input diffused underwater HSI. Therefore, to assess the effectiveness of our model image restoration, we compare the original underwater HSI, the noisy underwater HSI, the diffusion underwater HSI, and the JURTD restored underwater HSI.

We added mixed noise, combining Gaussian noise and strip noise, into the experimentally acquired original underwater HSIs. These images were then processed through the proposed framework outlined in this article to produce the diffusion underwater HSI output by the conditional diffusion model, followed by the restored underwater HSI output by the JURTD module. Figs. 7–9 illustrate the process of underwater HSI restoration for the outdoor pool dataset, the indoor pool dataset, and the synthetic HSI. Figs. 7(a), 8(a), and 9(a) show the original underwater HSIs of the three datasets, characterized by some blurring and weak noise attributed to electronic or environmental factors during acquisition. Figs. 7(b), 8(b), and 9(b) depict the noisy underwater HSIs of the three datasets, featuring Gaussian noise and strip noise. From these images, we extract PHSIs, which along with the noisy underwater HSIs, serve as inputs to the conditional diffusion model. This model undergoes a forward

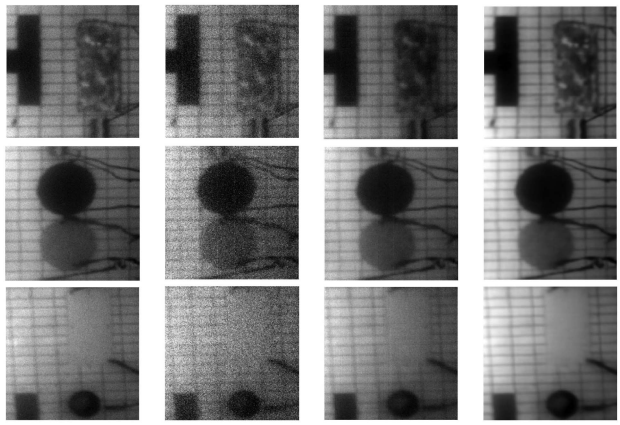


Fig. 7. Process of underwater HSI restoration for the outdoor pool dataset. (a) Original HSI. (b) Noisy underwater HSI. (c) Diffusion underwater HSI. (d) JURTD restored underwater HSI, the three rows from the top to the bottom are OutPool1, OutPool2, and OutPool3 in order. Our proposed method eliminates weak noise and enhances image clarity, resulting in a cleaner image. It exhibits excellent capability in underwater HSI restoration.

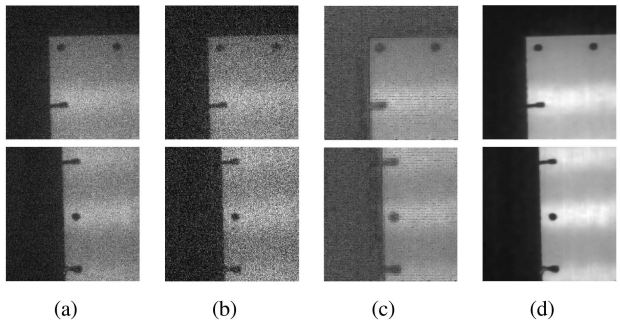


Fig. 8. Process of underwater HSI restoration for the indoor pool dataset. (a) Original underwater HSI. (b) Noisy underwater HSI. (c) Diffusion underwater HSI. (d) JURTD restored underwater HSI, the two rows from the top to the bottom are InPool1 and InPool2 in order.

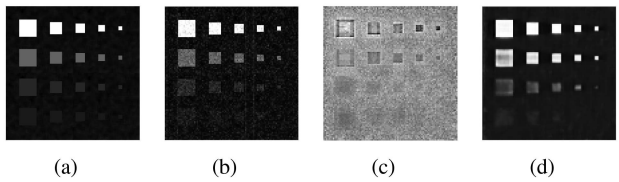


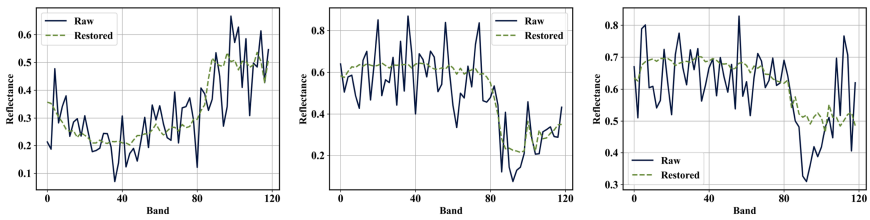
Fig. 9. Synthetic HSI restoration process. (a) Original underwater HSI. (b) Noisy underwater HSI. (c) Diffusion underwater HSI. (d) JURTD restored underwater HSI.

diffusion process with a time step of T initially, followed by a reverse process utilizing a well-trained noise estimation network. We define the state with a time step t of 0 as the diffusion underwater HSI, achieving preliminary image restoration, as depicted in Figs. 7(c), 8(c), and 9(c). Observing the figures, we note that the diffusion underwater HSI outcomes are most favorable for the indoor pool dataset, while outcomes for the outdoor pool dataset and the synthetic HSI are comparatively inferior. This discrepancy arises because the training dataset for

TABLE I
UNDERWATER HSI RESTORATION RESULTS ON THREE DATASETS

Data	OutPool1		OutPool2		OutPool3		InPool1		InPool2		SynHSI	
Metric	Diff	JURTD	Diff	JURTD	Diff	JURTD	Diff	JURTD	Diff	JURTD	Diff	JURTD
PSNR ↑	14.299	**20.243**	**20.341**	19.662	15.567	**22.254**	9.772	**19.843**	10.367	**21.134**	10.556	**17.285**
SSIM ↑	**0.272**	0.226	**0.309**	0.169	**0.341**	0.289	0.137	**0.134**	0.140	**0.168**	0.032	**0.395**
SAM ↓	0.461	**0.228**	**0.275**	0.296	0.351	**0.163**	0.526	**0.337**	0.476	**0.274**	0.565	**0.446**

The best values are Bolded.



(a) (b) (c)

Fig. 10. Spectral curves comparison for the outdoor pool dataset between the restored HSI and raw HSI. (a) OutPool1. (b) OutPool2. (c) OutPool3.

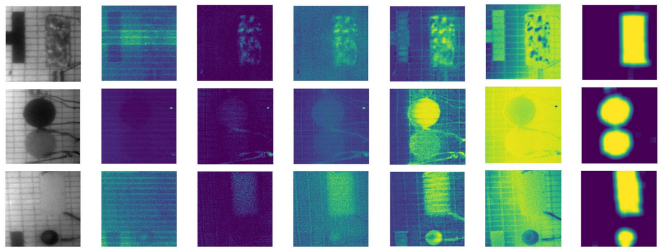
our conditional diffusion model primarily consists of the indoor pool dataset. We did not train on the remaining two datasets due to resource constraints, acknowledging the necessity to enhance the model generalization performance.

The quantitative results of underwater HSI restoration on the three datasets are illustrated in Table I, where "JURTD" represents the underwater HSI restored by JURTD, and "Diff" represents the diffusion underwater HSI. Intuitively, the JURTD restored underwater HSI demonstrates a significant denoising effect compared to the noisy underwater HSI and the original underwater HSI. It not only eliminates weak electronic noise but also enhances image clarity, resulting in a cleaner image overall, as shown in Figs. 7(d), 8(d), and 9(d).

Fig. 10 displays the spectral curve comparison for the outdoor pool dataset between the restored HSI and raw HSI, where the solid line represents the raw HSI and the dashed line represents the restored HSI. It can be observed that the restored spectral curves are relatively smooth with significantly reduced noise, indicating that the restoration method effectively suppresses high-frequency noise. In addition, the restored spectral curves retain the main trends and features of the original spectrum, especially in the blue-green wavelength bands. The framework proposed in this article exhibits excellent capability in underwater HSI restoration, and the restored underwater HSIs will facilitate subsequent visual analysis and target detection.

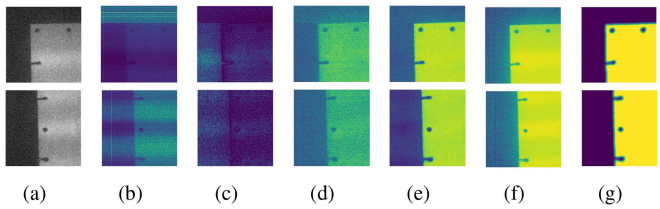
2) Detection Analysis: In this section, we show the results of underwater HTD and analyze them. To further demonstrate the superior performance of our method, JURTD is compared with the previously mentioned classical methods ACE, RX, MF, and the improved methods MSAM, and HSI-CNN. We evaluate the performance of these models in terms of both qualitative and quantitative analysis.

Figs. 11–13 show the detection maps of each compared method on the indoor pool dataset, the outdoor pool dataset, and the synthetic HSI. JURTD demonstrates excellent resistance to water body interference, and its results are very close to the real label. The traditional methods RX, ACE, and MF perform poorly on the three datasets and cannot effectively detect underwater



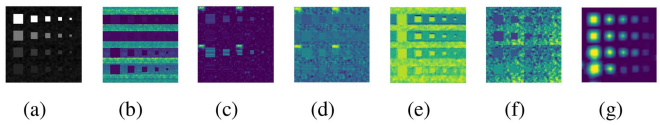
(a) (b) (c) (d) (e) (f) (g)

Fig. 11. Underwater HTD maps for the outdoor pool dataset. (a) Original underwater HSI. (b) RX. (c) ACE. (d) MF. (e) HSI-CNN. (f) MSAM. (g) JURTD.



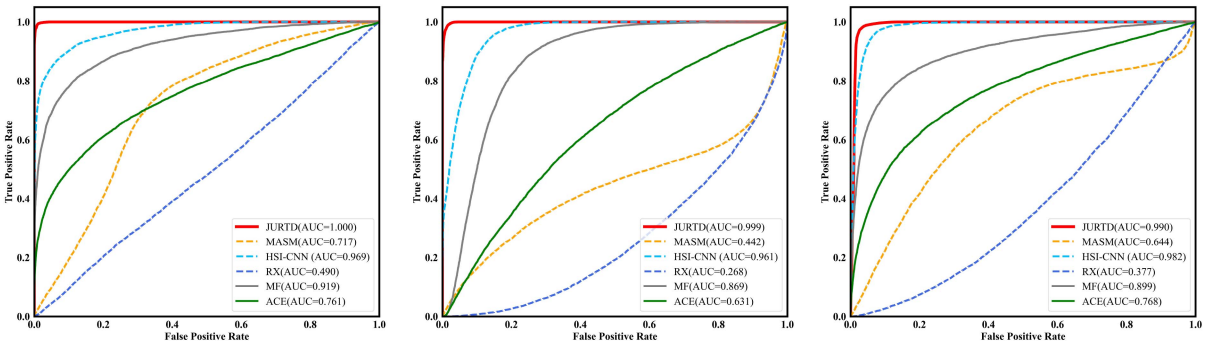
(a) (b) (c) (d) (e) (f) (g)

Fig. 12. Underwater HTD maps for the indoor pool dataset. (a) Original underwater HSI. (b) RX. (c) ACE. (d) MF. (e) HSI-CNN. (f) MSAM. (g) JURTD.



(a) (b) (c) (d) (e) (f) (g)

Fig. 13. Underwater HTD maps for the synthetic HSI. (a) Original underwater HSI. (b) RX. (c) ACE. (d) MF. (e) HSI-CNN. (f) MSAM. (g) JURTD.



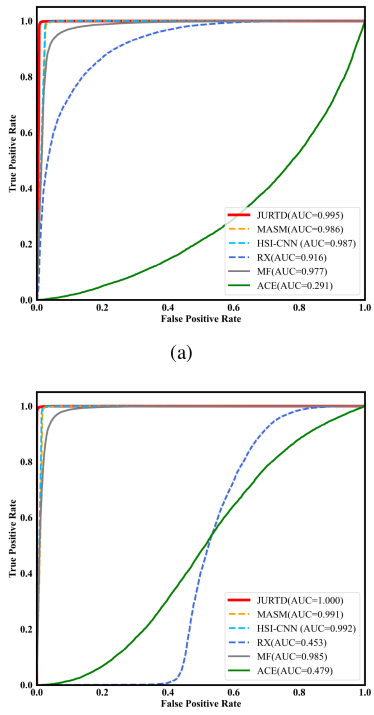
(a) (b) (c)

Fig. 14. ROC curves for comparing methods on the outdoor pool dataset. (a) OutPool1. (b) OutPool2. (c) OutPool3, the ROC curve of our method is closest to the upper left corner, indicating excellent target detection performance.

targets. In underwater environments, light scattering and absorption can blur the target spectral features and introduce additional spectral noise, causing the covariance estimation of ACE to fail, making it difficult to effectively separate the target from the background. On the indoor pool dataset, the spectral curves of the water body and the target on the captured images show a large difference because the depth of the current target in the water is 1 m and the depth of the indoor anechoic pool is 7 m. Therefore, the MF can roughly detect the underwater target. The HSI-CNN model takes 5% of the data to train the model in the outdoor pool dataset and the indoor pool dataset, and it performs better on the two datasets, while it performs poorly on synthetic HSIs, which indicates that its generalization ability is poor. MSAM is a physical method that does not require training, though it needs a priori target spectral information, and it performs better on the indoor pool dataset. The JURTD method proposed in this article shows excellent underwater target detection on all three datasets, which can effectively distinguish underwater targets, and on synthetic HSIs, even if the target size is 3×3, JURTD can distinguish it from the surrounding water bodies. Our proposed method is superior to these comparative methods.

To quantitatively demonstrate the superiority of the framework proposed in this article, the ROC curves and AUC values of each detection method are used as quantitative evaluation metrics. Fig. 14 illustrates the ROC curves of the detection results of our proposed method versus the other methods on the outdoor pool dataset. In addition, Figs. 15 and 16 also show the ROC curves for each detection method on the indoor pool dataset and the synthetic HSI, respectively. For the ROC curves, when the corresponding ROC curve of a model is closer to the upper left corner, it means that the detection performance of this model is better. As can be seen, the ROCs of our method are all in the upper leftmost corner, which indicates that our method demonstrates excellent underwater target detection performance on all three datasets.

We also compare the AUC values of these methods on each experimental dataset, as shown in Table II. We highlight the largest value in the table in red, the second in blue, and the third in underlined so that the comparative performance of each



(a) (b)

Fig. 15. ROC curves for comparing methods on the indoor pool dataset. (a) InPool1. (b) InPool2.

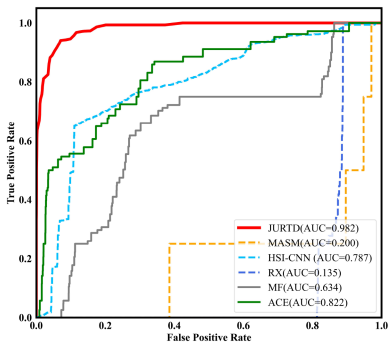


Fig. 16. ROC curves for comparing methods on the synthetic HSI.

TABLE II
AUC VALUES FOR COMPARING METHODS ON THE THREE EXPERIMENTAL DATASET

Dataset	RX	ACE	MF	HSI-CNN	MSAM	JURTD
OutPool1	0.490	0.761	0.919	0.969	0.717	1.000
OutPool2	0.268	0.631	0.869	0.961	0.442	0.999
OutPool3	0.377	0.768	0.899	0.982	0.644	0.990
InPool1	0.453	0.479	0.985	0.992	0.991	1.000
InPool2	0.916	0.291	0.977	0.987	0.986	0.995
SynHSI	0.135	0.822	0.634	0.787	0.200	0.982
Average	0.439	0.625	0.881	0.946	0.663	0.994

The best value is shown in red, the second in blue, and the third value is underlined.

method can be clearly and intuitively analyzed. Table II reveals that the AUC values of the methods in this article are the highest on all three datasets, and the HSI-CNN method has the second highest AUC value on the indoor pool dataset and the outdoor pool dataset, but it has the third highest AUC value on the synthetic HSI, which indicates that the generalization ability of this method is poor and unsuitable for practical application. MF performs the third best on the outdoor pool dataset, and MSAM. The performance is also the third. For the various types of comparison methods, we average the AUC on the three datasets, and JURTD is ranked first, HSI-CNN is the second one, and MF is ranked third. Based on the aforementioned analysis, it is easy to find that our proposed method has great advantages in hyperspectral underwater target detection.

E. Model Analysis

1) Ablation Experiment: To validate the key performance of each module in our proposed framework, we conducted a series of ablation experiments. These experiments aimed to analyze the performance of the model under different configurations and assess the extent to which each module contributes to the overall framework performance. In this section, we will delve into the roles and interactions of the three key modules: diffusion, restoration, and detection, in the framework by comparing the results of the ablation experiments, which are presented in Table III. Since the final application scenario of this article is underwater HTD, we have retained the detection module in the ablation experiments. When the conditional diffusion model is not employed, we take three same original underwater HSIs as the input of the JURTD module. The output of the JURTD module will be only the detection map if we omit the use of the restore module.

By comparing and analyzing the four cases, we conclude that both the conditional diffusion model and restoration techniques play pivotal roles in underwater HSI restoration and target detection tasks. The application of either the conditional diffusion model or restoration model alone leads to improvements in image quality or target detection performance. Furthermore, our method achieves optimal results in both image restoration and target detection by jointly applying these modules, as highlighted in bold in Table III.

2) Effectiveness of the Conditional Diffusion Model: In the preceding section, we verified the significance of the conditional diffusion model for JURTD in detecting underwater targets. To further validate the effectiveness of the conditional diffusion model, we applied the well-trained model to conduct forward diffusion and reverse processes. Using OutPool2 from the outdoor pool dataset as a case study, Fig. 17 shows the 500-nm grayscale image of the conditional diffusion model at different sampling times during the reverse process. We propagated the original underwater HSI forward by T steps, and then, reversed from this state, showcasing the HSIs at sampling times t of 100, 60, 40, 10, 5, and 0, respectively. From a visual perspective, it is evident that the conditional diffusion model effectively reconstructs the original underwater HSI. Moreover, the restored diffusion image exhibits improved clarity and visibility compared to the original image under the conditional guidance of the PHSI.

Fig. 18 depicts the spectral curves of the conditional diffusion model at different sampling times during the reverse process. In Fig. 18, the green curve represents the spectral curve of the noise HSI at point $(40, 40)$ after T steps of forward diffusion, the red curve depicts the spectral curve of the sampled image at the same point, and the blue curve signifies the spectral curve of the PHSI of the original underwater HSI at the same point. It is evident that as the sampling time t decreases, the sampled spectral curve gradually aligns with the spectral curve of the PHSI. When the sampling times are 10, 5, and 0, the spectral curves closely resemble the spectral curve of the PHSI, although slight differences persist in terms of amplitude and curve trend. These variations simulate the interference of the absorption and scattering properties of the water column on the underwater HSIs. Therefore, we aggregate these three sampled images into the group of variable spectral images and employ them as the input of the subsequent JURTD module.

3) Robustness Analysis: To further assess the robustness of our proposed framework, we examine the impact of different SNRs on target detection accuracy and image restoration effectiveness. SNR variations are common in real underwater scenarios and can significantly influence underwater image processing and target detection tasks.

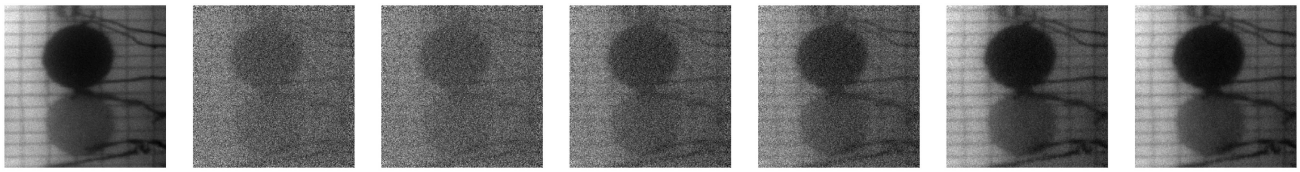
We depict the relationship between target detection AUC values and different SNRs in Fig. 19. The AUC value serves as a quantitative metric to evaluate the performance of underwater target detection. As SNR increases, we observe a general improvement in the performance of underwater target detection. This trend arises because low SNR environments typically result in more severe interference and occlusion of target information in underwater images, leading to decreased clarity and contrast of target edges, thereby increasing detection difficulty. It is noteworthy that target detection results align closely with those observed in noiseless interference scenarios when SNR exceeds 9 in the indoor pool dataset and when SNR exceeds 18 in the outdoor pool dataset. This suggests that our framework maintains consistent target detection performance under varying SNR conditions, particularly in higher SNR environments.

In addition, we conducted a study on the image restoration effectiveness of the framework under varying SNR conditions, using the quantitative metrics PSNR and SAM values to evaluate

TABLE III
AUC VALUES OF DETECTION PERFORMANCE OF THE ABLATION EXPERIMENT ON EACH DATASET

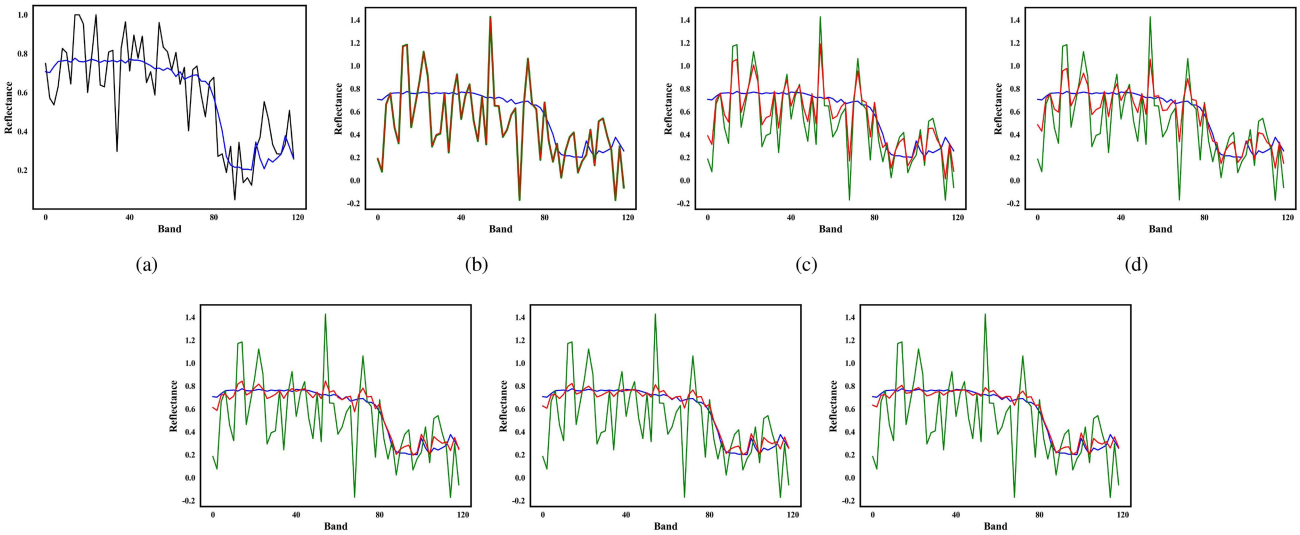
Model	Diff	Restore	Detect	OutPool1	OutPool2	OutPool3	InPool1	InPool2	SynHSI
Model1	✗	✗	✓	0.955	0.785	0.966	0.981	0.975	0.873
Model2	✓	✗	✓	0.975	0.985	0.965	0.979	0.977	0.873
Model3	✗	✓	✓	0.876	0.976	0.920	0.980	0.976	0.881
Model4	✓	✓	✓	**1.000**	**0.999**	**0.990**	**1.000**	**0.995**	**0.982**

The best value is shown in Bold.



(a) (b) (c) (d) (e) (f) (g)

Fig. 17. Grayscale image of 500 nm for the conditional diffusion model on OutPool2 in the outdoor pool dataset at different sampling times during the reverse process. (a) Original underwater HSI. (b) Sampling $t = 100$. (c) $t = 60$. (d) $t = 40$. (e) $t = 10$. (f) $t = 5$. (g) $t = 0$. The conditional diffusion model effectively and gradually reconstructs the underwater HSI.



(a) (b) (c) (d) (e) (f) (g)

Fig. 18. Spectral curves of the conditional diffusion model on OutPool2 in the outdoor pool dataset at different sampling times during the reverse process. (a) Original underwater hyperspectral image. (b) Sampling $t = 100$. (c) $t = 60$. (d) $t = 40$. (e) $t = 10$. (f) $t = 5$. (g) $t = 0$. As the sampling time t decreases, the sampled spectral curve gradually aligns with the spectral curve of the PHSI.

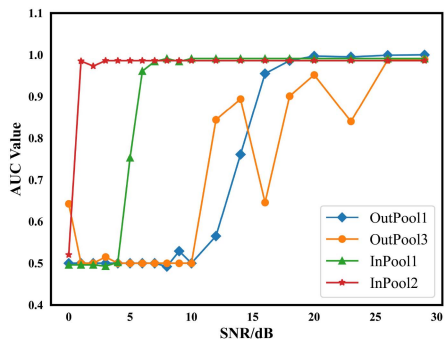


Fig. 19. Relationship between AUC values and SNR on four data scenarios.

performance. To present the experimental findings clearly, we averaged the evaluations across four different data scenes to establish the relationship between average PSNR and SAM with SNR, as illustrated in Fig. 20. The results indicate that as SNR increases, PSNR rises, while SAM values decrease, suggesting a gradual improvement in image restoration quality. Specifically, higher SNR levels correspond to better preserved spectral fidelity and reduced spectral distortions, leading to clearer and more accurate restoration of the underwater hyperspectral images. In summary, our experimental findings demonstrate that our proposed framework exhibits exceptional robustness in underwater HSI restoration and target detection across various

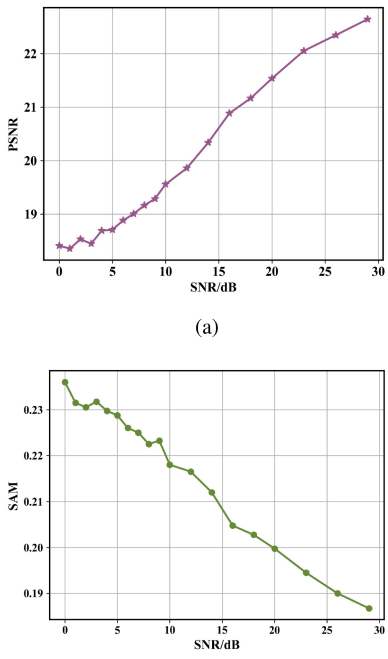


Fig. 20. Relationship between average PSNR, SAM values, and SNR on four data scenarios. (a) PSNR. (b) SAM.

SNR environments, consistently maintaining high performance and reliability despite varying noise conditions.

V. Conclusion

In this article, we propose a novel joint framework for underwater HSI restoration and target detection based on a conditional diffusion model. The proposed framework effectively overcomes the challenges of underwater hyperspectral imaging, and the limitations of existing methods by considering underwater target spectral variability and image quality degradation.

It comprises two key modules: the variable spectral group extraction module and the JURTD module. The variable spectral group extraction module utilizes the conditional diffusion model to simulate spectral variance in underwater environments, generating variable spectral groups for a comprehensive representation of underwater target spectra. Then, the JURTD module leverages these variable spectral groups to optimize image restoration and target detection tasks simultaneously. By integrating these modules, our framework effectively tackles the challenges in underwater HSI processing, providing a novel solution for enhancing image quality and improving target detection accuracy in underwater environments. Experimental evaluations on both real-world and synthetic datasets demonstrate the superior performance of our framework in enhancing image quality and improving target detection accuracy.

In terms of future development, the trajectory of research will focus on probing underwater HTD within real marine environments. Considerable attention will be directed toward assessing the influence of sea surface fluctuations and water turbidity on the spectral radiative transfer of submerged targets, along with its consequential implications for underwater target detection. Furthermore, attaining target detection at greater depths will emerge as a pivotal area of investigation. We hold the belief that the field of underwater HTD will mature and become more comprehensive as more researchers are involved in this area of study.

References

[1] H. Ghafoor and Y. Noh, "An overview of next-generation underwater target detection and tracking: An integrated underwater architecture," *IEEE Access*, vol. 7, pp. 98841–98853, 2019.

[2] F. Lei, F. Tang, and S. Li, "Underwater target detection algorithm based on improved YOLOv5," *J. Mar. Sci. Eng.*, vol. 10, no. 3, 2022, Art. no. 310.

[3] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[4] A. Plaza et al., "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, pp. S110–S122, 2009.

[5] M. J. Khan, H. S. Khan, A. Yousaf, K. Khurshid, and A. Abbas, "Modern trends in hyperspectral image analysis: A review," *IEEE Access*, vol. 6, pp. 14118–14129, 2018.

[6] W. Gross et al., "A multi-temporal hyperspectral target detection experiment: Evaluation of military setups," in *Proc. Target Background Signatures VII*, 2021, pp. 38–48.

[7] M. Shimoni, R. Haelterman, and C. Perneel, "Hypersectral imaging for military and security applications: Combining myriad processing and sensing techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 101–117, Jun. 2019.

[8] B. Lu, P. D. Dao, J. Liu, Y. He, and J. Shang, "Recent advances of hyperspectral imaging technology and applications in agriculture," *Remote Sens.*, vol. 12, no. 16, 2020, Art. no. 2659.

[9] P. Singh et al., "Hyperspectral remote sensing in precision agriculture: Present status, challenges, and future trends," in *Hyperspectral Remote Sensing*. Amsterdam, The Netherlands: Elsevier, 2020, pp. 121–146.

[10] M. B. Stuart, A. J. McGonigle, and J. R. Willmott, "Hyperspectral imaging in environmental monitoring: A review of recent developments and technological advances in compact field deployable systems," *Sensors*, vol. 19, no. 14, 2019, Art. no. 3071.

[11] T. Bucher and F. Lehmann, "Fusion of hymap hyperspectral with HRSC-A multispectral and DEM data for geoscientific and environmental applications," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. Taking Pulse Planet, Role Remote Sens. Manag. Environ.*, 2000, pp. 3234–3236.

[12] G. Dobler, M. Ghandehari, S. E. Koonin, and M. S. Sharma, "A hyperspectral survey of New York city lighting technology," *Sensors*, vol. 16, no. 12, 2016, Art. no. 2047.

[13] A. Ibrahim et al., "Atmospheric correction for hyperspectral ocean color retrieval with application to the hyperspectral imager for the coastal ocean (HICO)," *Remote Sens. Environ.*, vol. 204, pp. 60–75, 2018.

[14] S. P. Garaba et al., "Sensing ocean plastics with an airborne hyperspectral shortwave infrared imager," *Environ. Sci. Technol.*, vol. 52, no. 20, pp. 11699–11707, 2018.

[15] A. T. Çelebi and S. Ertürk, "Visual enhancement of underwater images using empirical mode decomposition," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 800–805, 2012.

[16] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[17] J. Y. Chiang and Y.-C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1756–1769, Apr. 2012.

[18] E. H. Land, "The Retinex theory of color vision," *Sci. Amer.*, vol. 237, no. 6, pp. 108–129, 1977.

[19] D. J. Jobson, Z.-u. Rahman, and G. A. Woodell, "A multiscale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, Jul. 1997.

[20] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1205–1218, Feb. 2019.

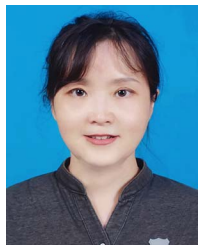
[21] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," 2016, *arXiv:1606.08921*.

[22] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[23] D. Yu, Q. Li, X. Wang, Z. Zhang, Y. Qian, and C. Xu, "DSTrans: Dual-stream transformer for hyperspectral image restoration," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2023, pp. 3739–3749.
[24] C. Zhang, C. Zhang, M. Zhang, and I. S. Kweon, "Text-to-image diffusion model in generative AI: A survey," 2023, *arXiv:2303.07909*.
[25] Y. Miao, L. Zhang, L. Zhang, and D. Tao, "DDS2M: Self-supervised denoising diffusion spatio-spectral model for hyperspectral image restoration," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 12086–12096.
[26] H. Li et al., "SRDiff: Single image super-resolution with diffusion probabilistic models," *Neurocomputing*, vol. 479, pp. 47–59, 2022.
[27] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. V. Gool, "RePaint: Inpainting using denoising diffusion probabilistic models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 11461–11471.
[28] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 6840–6851.
[29] C. Wu, D. Wang, Y. Bai, H. Mao, Y. Li, and Q. Shen, "HSR-Diff: Hyperspectral image super-resolution via conditional diffusion models," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2023, pp. 7083–7093.
[30] N. Chen, J. Yue, L. Fang, and S. Xia, "SpectralDiff: A generative framework for hyperspectral image classification with diffusion models," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, no. 5522416, pp. 1–16, Aug. 2023.
[31] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
[32] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.
[33] Y. Luo, J. Zou, C. Yao, X. Zhao, T. Li, and G. Bai, "HSI-CNN: A novel convolution neural network for hyperspectral image," in *Proc. Int. Conf. Audio, Lang. Image Process.*, 2018, pp. 464–469.
[34] S. Kraut and L. L. Scharf, "The CFAR adaptive subspace detector is a scale-invariant GLRT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2538–2541, Sep. 1999.
[35] S. Oshigami et al., "Mineralogical mapping of southern Namibia by application of continuum-removal MSAM method to the hymap data," *Int. J. Remote Sens.*, vol. 34, no. 15, pp. 5282–5295, 2013.
[36] X. Zhao, W. Li, C. Zhao, and R. Tao, "Hyperspectral target detection based on weighted cauchy distance graph and local adaptive collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, no. 5527313, pp. 1–13, Apr. 2022.
[37] X. Zhao, K. Liu, K. Gao, and W. Li, "Hyperspectral time-series target detection based on spectral perception and spatial-temporal tensor decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, no. 5520812, pp. 1–12, Aug. 2023.
[38] S. P. Garaba and T. Harmel, "Top-of-atmosphere hyper and multispectral signatures of submerged plastic litter with changing water clarity and depth," *Opt. Exp.*, vol. 30, no. 10, pp. 16553–16571, 2022.
[39] A. Papakonstantinou, A. Moustakas, P. Kolokoussis, D. Papageorgiou, R. de Vries, and K. Topouzelis, "Airborne spectral reflectance dataset of submerged plastic targets in a coastal environment," *Data*, vol. 8, no. 1, 2023, Art. no. 19.
[40] S. Jay, M. Guillaume, and J. Blanc-Talon, "Underwater target detection with hyperspectral data: Solutions for both known and unknown water quality," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1213–1221, Aug. 2012.
[41] D. B. Gillis, "An underwater target detection framework for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 1798–1810, Apr. 2020.
[42] J. Qi, Z. Gong, W. Xue, X. Liu, A. Yao, and P. Zhong, "An unmixing-based network for underwater target detection from hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 5470–5487, May 2021.
[43] J. Qi et al., "A self-improving framework for joint depth estimation and underwater target detection from hyperspectral imagery," *Remote Sens.*, vol. 13, no. 9, 2021, Art. no. 1721.
[44] Z. Li et al., "A transfer-based framework for underwater target detection from hyperspectral imagery," *Remote Sens.*, vol. 15, no. 4, 2023, Art. no. 1023.
[45] Q. Li, J. Li, T. Li, Z. Li, and P. Zhang, "Spectral-spatial depth-based framework for hyperspectral underwater target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, no. 4204615, pp. 1–15, May 2023.
[46] Z. Zou and Z. Shi, "Hierarchical suppression method for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 330–342, Jan. 2016.

Qi Li (Student Member, IEEE) received the B.S. degree in communication engineering from the Xi'an University of Technology, Xi'an, China, in 2022. He is currently working toward the Ph.D. degree in information and communication engineering with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an.

His research interests include image processing, machine learning, and remote sensing.



Jinghua Li was born in Shanxi, China, in 1964. She received the B.S. and the Ph.D. degrees in signal and information processing from the Northwestern Polytechnical University, Xi'an, China, in 1987 and 2008, respectively.

She is currently a Full Professor with the School of Electronics and Information, Northwestern Polytechnical University. Her research interests interests include image processing, digital signal processing, and pattern recognition, with applications in the field of remote sensing.

Tong Li received the B.S. degree in electronic information science and technology from the Hefei University, Hefei, China, in 2022. She is currently working toward the M.S. degree in electronics and information with the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China.

Her research interests include deep learning, hyperspectral image processing, and hyperspectral unmixing.

Yan Feng (Member, IEEE) received the B.S. degree in electronics engineering and the M.S. and Ph.D. degrees in signal and information processing from Northwestern Polytechnical University, Xi'an, China, in 1984, 1989, and 2007, respectively.

She is currently a Professor with the School of Electronics and Information, Northwestern Polytechnical University. Her research interests include image processing, compression and classification of hyperspectral images, compressed sensing theory, and intelligent information processing.