

OPEN ACCESS

EDITED BY

An-An Liu,
Tianjin University, China

REVIEWED BY

Xuebo Zhang,
Northwest Normal University, China
Xiaoxue Qian,
University of Texas Southwestern Medical
Center, United States
Huo Shumeng,
Tianjin University, China

*CORRESPONDENCE

Lei Cai
✉ cailei2014@126.com

RECEIVED 24 September 2023

ACCEPTED 06 March 2024

PUBLISHED 27 March 2024

CITATION

Zhou M, Cai L, Jia J and Gao Y (2024)
Multi-scale aware turbulence network
for underwater object recognition.
Front. Mar. Sci. 11:1301072.
doi: 10.3389/fmars.2024.1301072

COPYRIGHT

© 2024 Zhou, Cai, Jia and Gao. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Multi-scale aware turbulence network for underwater object recognition

Meng Zhou¹, Lei Cai^{2*}, Jishen Jia¹ and Yuhe Gao¹

¹School of Mathematical Sciences, Henan Institute of Science and Technology, Xinxiang, China,

²School of Artificial Intelligence, Henan Institute of Science and Technology, Xinxiang, China

Underwater imagery is subject to distortion, and the presence of turbulence in the fluid medium poses difficulties in accurately discerning objects. To tackle these challenges pertaining to feature extraction, this research paper presents a novel approach called the multi-scale aware turbulence network (MATNet) method for underwater object identification. More specifically, the paper introduces a module known as the multi-scale feature extraction pyramid network module, which incorporates dense linking strategies and position learning strategies to preprocess object contour features and texture features. This module facilitates the efficient extraction of multi-scale features, thereby enhancing the effectiveness of the identification process. Following that, the extracted features undergo refinement through comparison with positive and negative samples. Ultimately, the study introduces multi-scale object recognition techniques and establishes a multi-scale object recognition network for the precise identification of underwater objects, utilizing the enhanced multi-scale features. This process entails rectifying the distorted image and subsequently recognizing the rectified object. Extensive experiments conducted on an underwater distorted image enhancement dataset demonstrate that the proposed method surpasses state-of-the-art approaches in both qualitative and quantitative evaluations.

KEYWORDS

underwater image distortion, distortion correction, multi-scale feature, turbulence, underwater objective recognition

1 Introduction

With regard to the exploration and exploitation of marine resources, underwater objective recognition serves as a crucial medium and representation method for comprehending and perceiving the underwater realm. However, underwater environments are complex and variable, and underwater objectives can suffer from image degradation, distortion, or aberrations due to turbulence. More specifically, turbulence is a common phenomenon in the underwater environment that can lead to degradation, distortion, or deformation of underwater target images, causing features to

deform, be lost, or become distorted, thereby significantly increasing the difficulty of recognition. Therefore, addressing the challenges brought by turbulence in the underwater environment to enhance the accuracy and efficiency of underwater target recognition is of paramount importance for in-depth exploration of the ocean, conservation of marine ecosystems, and advancement of sustainable development and utilization of marine resources.

To address these issues, many methods for underwater image detection have been investigated. However, amphibious robots suffer from feature loss when recognizing underwater objectives. The current algorithms were usually solved using multi-scale feature extraction (Zhao et al., 2021). In addition, underwater turbulence leads to incomplete objective feature information. There are limited features available for use in the underwater detection and identification process. Meanwhile, there are large differences in the scale of objective features of various types and different distances underwater. It is impossible to accurately extract the effective features of the objective for identification. Some scholars used contextual feature learning modeling to enhance the recognition of objectives with inconspicuous appearance features (Pato et al., 2020). Other scholars used generative adversarial learning to map low-resolution objective features into features that are equivalent to high-resolution objectives (Deng et al., 2021). These strategies can improve the algorithm recognition performance to some extent, but they are less effective in objective recognition caused by underwater turbulence.

In our work, we propose a multi-scale aware turbulence network (MATNet) method to solve the aberrant object recognition problem. It consists of two main phases: multi-scale feature extraction and corrective identification of distorted objectives. First, we perform feature extraction of objective contour features and positional features for the distorted objective feature information. Subsequently, correction of distorted images is achieved by fusion processing based on the extracted features. Finally, a loss function is introduced to accurately recognize the distorted objective.

In summary, the key contributions of this article can be highlighted as follows:

1. In this paper, we propose to construct a density linking strategy and a location learning strategy for multi-scale feature extraction using different learning strategies to extract object texture and contour features from underwater environments. Utilizing the dual strategy learning module enables our module to be more perfect for object feature extraction.
2. This paper proposes a contrast correction module, which realizes object distortion correction by means of positive and negative samples of multi-scale features, and the module utilizes the extracted object features by using the contrast method to complete the correction of the distorted object.
3. This paper proposes a loss function to solve the object recognition problem. This loss function effectively solves the small sample category misclassification problem by

changing the sample category weights to accomplish the accurate object recognition problem.

2 Related work

2.1 Multi-scale feature extraction

The accuracy of the extracted features directly affects the results of network localization and recognition during the objective recognition process. Huang et al. (2023) introduced a new hybrid attention model (S-CA), a compact channel attention module (C-ECA), and a streamlined target feature extraction network (S-FE) to enhance the capture of positional feature information. Ye et al. (2022) introduced a fusion multi-scale attention mechanism network to address boundary ambiguity, utilizing a feature refinement compensation module to minimize inter-class disparities. Cai et al. (2022a) proposed a dynamic multi-scale feature fusion method for underwater target recognition. Zhang et al. (2022) introduced MLE, an effective method for enhancing underwater images. In this, Zhang and Dong (2022) proposed a novel method that was introduced for enhancing underwater images by combining a color correction technique inspired by Retime with a fusion technique that maintains fine details. Zhang et al. (2024) proposed a cascaded visual attention network (CVANet) for single image superresolution, which is used for feature extraction and detail reconstruction. Zhou et al. (2022) proposed a restoration method that utilizes backscattered pixel prior and color bias removal to enhance the contrast of underwater images, effectively correcting color distortions and preserving crucial image details through a fusion process. Yu et al. (2022) proposed a dual predictive feature pyramid module and a spatial channel attention mechanism module, which can obtain multi-scale contextual information on a large scale and improve the objective recognition rate. A dual prior optimization contrast enhancement method was proposed by researchers. This method employs distinct enhancement strategies for each layer, aiming to enhance both the contrast and texture details of the underwater image (Zhang et al., 2023a). Cai et al. (2022b) proposed a multi-bit pose feature generation mapping network M-PFGMNet for visual object tracking. Zhou et al. (2023) proposed a novel multi-feature underwater image enhancement method based on the embedded fusion mechanism (MFEF), which uses its decoder and encoder to recover the underwater scene and thus complete the image enhancement. Li et al. (2019) proposed an underwater image enhancement algorithm, which constructs an underwater image enhancement benchmark and trains an underwater image enhancement network based on this benchmark. Hyun et al. (2021) proposed to introduce the correlation region suggestion network to analyze the effect of regional convolutional neural network (R-CNN) in image feature extraction applications. Zhang et al. (2023c) proposed a novel imaging algorithm for multi-receiver synthetic aperture sonar (SAS), which rephrases the range change stage and range invariant stage.

2.2 Image contrast restoration

In aberrant objective image processing research, underwater images face quality degradation challenges in complex underwater environments. Zhang et al. (2023b) proposed an underwater image enhancement method that utilizes a weighted wavelet visual perception fusion technique. Eigel et al. (2022) proposed a framework using computational homogenization to enhance the shape uncertainty to have fuzzy properties and, finally, to compute its mean displacement boundary. Wang et al. (2023c) proposed an intelligent protocol for underwater image enhancement, which carried out intelligent configuration through protocol reinforcement learning, and finally produced underwater image enhancement results. Li et al. (2022c) proposed binocular structured light to measure the geometric parameters of internal threads as a vision system. It can achieve high-precision recovery from 2D virtual image to actual image. Sun et al. (Sun et al., 2022) construct a novel bidirectional recursive VSR architecture to recover fine details by dividing the task into two subtasks and directing the attention to a motion compensation module that eliminates the effect of inter-frame misalignment. Cheng et al. (Cheng et al., 2022) propose a dual generative adversarial network patch model (DGPM) based image recovery for structural defects detection. Jiang et al. (Jiang et al., 2022b) proposed two restart nonlinear conjugate gradient method (CGM) with different restart degrees to solve the problem of unconstrained optimization and image restoration. Liu et al. (2023) proposed a multi-purpose haze removal framework for nighttime hazy images. Kim et al. (2022) propose a vehicle localization method that fuses aerial maps and LiDAR measurements in an urban canyon environment. Image restoration is accomplished by correcting the contours by correcting the scale distortion of the projections. He et al. (2022) investigate a new setup that aims to modulate the output effects across multiple degradation types and levels. Li et al. (2022e) proposed image Laplacian dark channel attenuation defogging method, which can reduce the transmission value deviation in different regions. Li et al. (2022d) introduced a novel multi-scale feature representation and interaction network for underwater object detection. While the methods mentioned above have shown promising results in their respective fields, they mainly focused on aberration correction or model optimization. However, characterizing turbulence-induced aberration images poses significant challenges due to their complex and dynamic nature.

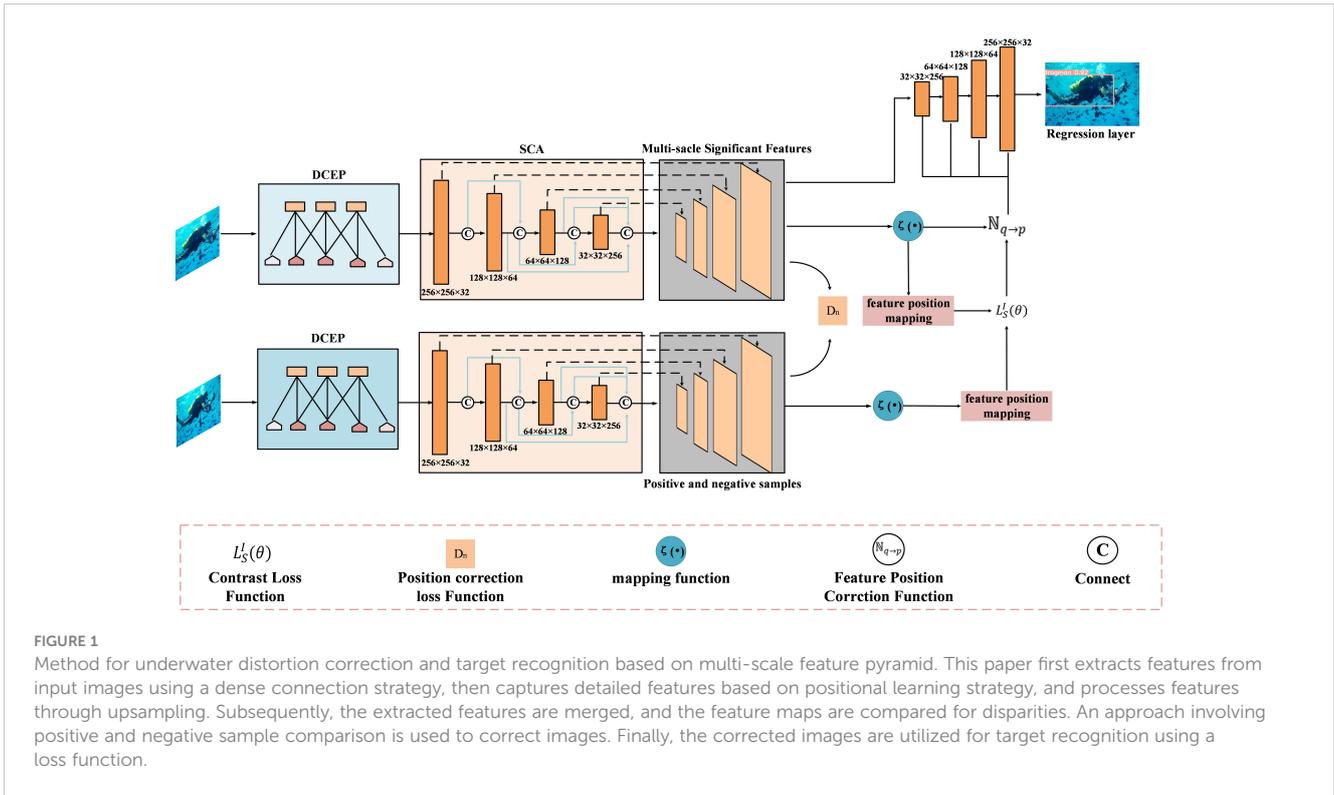
2.3 Image object recognition

The existing body of research has made significant progress in the field of object recognition. Xu et al. (2020) introduced the SA-FPN architecture, designed to extract underwater image features and enhance the detection performance of marine objects. Chen et al. (2023) proposed an adaptive hybrid attention convolutional neural network (AHA-CNN) framework. Zhang et al. (2023) proposed a generative adversarial-driven cross-perception

network (GACNet) for wheat variety identification and authentication. Wang et al. (2023b) proposed reinforcement learning with visual enhancement for object detection in underwater scenes to gradually enhance visual images to improve detection results. Guo et al. (2022) proposed a fully automated model compression framework called 3D-Pruning (3DP), which aims to achieve efficient 3D action recognition. Yang et al. (Yang, 2023) proposed a multi receiver synthetic aperture sonar (SAS) that generates high resolution by coherently stacking continuous echo signals. Lin et al. (2022b) proposed a system tailored for recognizing and tracking underwater target objects. Li et al. (Li et al., 2021) proposed an underwater image enhancement network guided by mid projection, which utilizes multi-color space embedding and physical model learning methods to effectively improve visual quality. Palomeras et al. (2022) proposed what was called ATR, which combines detectors and classifiers using a convolutional neural network model. Li et al. (2022a) investigated first-person hand movement recognition for RGB-D sequences with eight classical pre-trained networks and one pre-trained network designed to extract RGB-D features. Cai et al. (2022c) proposed an enhanced dilated convolution framework for underwater blurred target recognition. Yamada et al. (2021) proposed a novel self-supervised representation learning method. The method allows deep learning convolutional autoencoders to utilize multiple metadata sourced to normalize their learning. Wang et al. (2023d) proposed an adaptive attenuation channel compensation method for optimal channel precorrection and a guided fusion method for eliminating color deviation in RGB color space. Miao et al. (2021) proposed a method for hull modeling and identification of operational objectives based on 3D point clouds collected by a laser measurement system mounted on a ship loader. Wang et al. (2022) proposed an effective method for inertial feature recognition of conical spatial objectives based on deep learning. Xu et al. (2022) and others used YOLOv5s-improved algorithm to add a CA attention module by suppressing complex background and negative sample interference in the image. In traditional aberration recognition, direct recognition of the objective without correction can lead to reduced accuracy. Aberration correction often involves constructing correlation aberrations, but characterizing turbulence-induced aberrations with a fixed model is challenging, especially in correcting distortion in underwater images. The dynamic nature of turbulence makes establishing a static model difficult, complicating the correction process. This paper proposes a comparison correction method for distorted objectives to enhance recognition accuracy.

3 Methodology

To solve the problem of objective aberration recognition caused by turbulence, this paper proposes a multi-scale feature perception module; the model architecture is shown in Figure 1. This paper proposes a density linking module and a position learning module, which are used for objective multi-scale feature extraction and in the contrast correction module for the aberration correction of the



aberration image. Our approach can be summarized into three processes: multi-scale feature extraction, distortion correction, and objective recognition.

3.1 Multi-scale feature extraction

Due to the complex underwater environment, feature data in images may be difficult to extract. Object recognition poses a significant challenge. Therefore, this article constructs a MATNet network that integrates the DCEP module and SCA module. The former uses density linking to cluster the dataset, ensuring multi-scale feature extraction with high feature resolution, and can extract a wider range of features at different scales. The latter adopts a position learning strategy, utilizing the location information of the data to enhance the learning ability and performance of the model, which can better extract the detailed features of the object. The MATNet network uses a combination of two modules to better extract object features at different scales, which helps the network achieve higher accuracy of the object.

The paper proposes a method of utilizing pyramid feature transfer to better utilize target features, improve information transmission rate between different layers, and facilitate the correction and recognition of distorted images. The entire backbone network includes multiple encoder branches, and both encoder branches employ an identical structure for the restoration and recognition of distorted images. The transition layer, serving as a convolutional connector, is used to sample between each module. Subsequently, the extracted features are fed as inputs into the

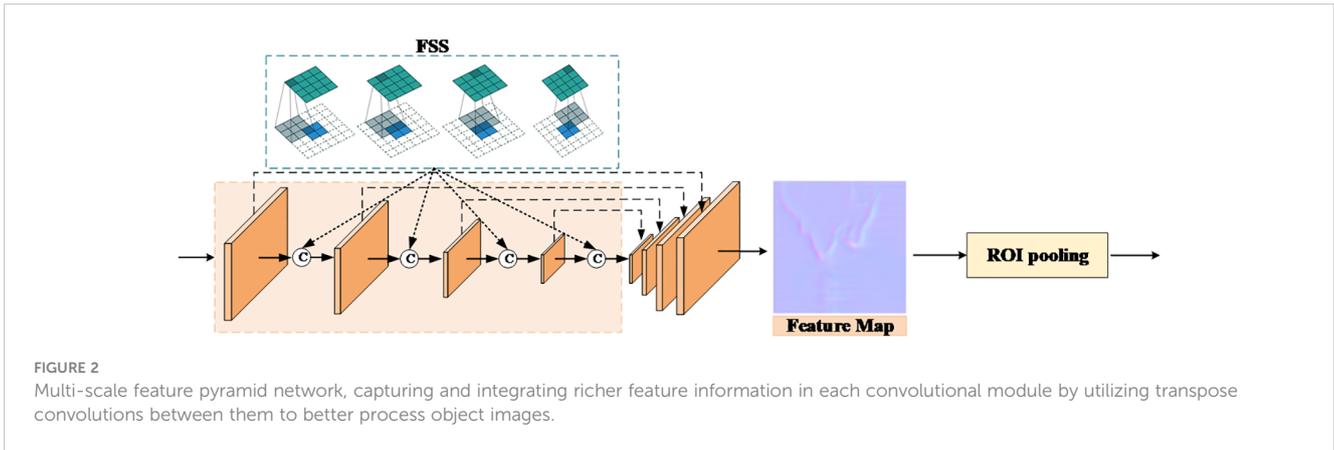
subsequent module, yielding multi-scale features of the object. The process of the multi-scale perception pyramid model is illustrated in Figure 2.

In this paper, we utilize the DCEP module to obtain multi-scale features over a large area. Density links are utilized to extract objective features in a larger range to obtain objective contour feature information. A smaller size filter is set. Thus, the generated features are discriminative for objective detection. Since the input image is distorted and deformed due to turbulence, in the process of feature extraction, in addition to considering the positional information of the data, it is also necessary to store the relative positions between features in a computational manner.

After extracting features through convolutional neural networks, the size of the feature map often decreases, which requires upsampling of the image. Using deconvolution for image feature processing increases the receptive field of view, allowing subsequent convolution kernels to learn more global information and obtain more accurate sampling results. The feature map of the target is restored to its original resolution. The reverse convolution kernel is applied to the input feature map, and high-resolution output feature maps are generated through convolution operations. The upsampling ratio is controlled by adjusting the step size of the reverse convolution kernel. The step size is set to 1 without applying padding according to Equation (1)

$$output = \frac{i + 2p - k}{s} + 1 \tag{1}$$

There, s is the stream, p is the padding, i is the input, and k is the number of filters. The calculated output image size is $2 * 2$. Next, the input image is set to $2 * 2$ and filled to 0, and a $3 * 3$ convolution



kernel is used with a step size of 1 to perform upsampling to an output size of 4×4 .

The multi-scale perception pyramid $f_{\tau}(\cdot)$ utilizes the MATNet network framework consisting of three DCEP modules and SCA modules. By adding FSS, appropriate image features are transmitted from shallow nodes to deep nodes, thereby avoiding inconsistency in gradient calculation. The FSS is designed as Equation (2):

$$P'_{k-1} = (A_{k-1} \Theta f_{nu}(A_k)) \Theta f_{nu}(P'_k) + C_{k-1} \quad (2)$$

where the intersection of A_{k-1} and A_k can be computed by Θ , f_{nu} is the upsampling operation, P'_k is the merged mapping of the k -th layer, C_{k-1} is the residual fast output of $k - 1$, and the FSS serves as the scaling selection. FSS acts as a scaling option.

Based on semantic and multi-scale feature extraction, explore the loss of image features caused by turbulence, and use deep semantics to evaluate the quality of the affected images. This chapter adopts the semantic contextual approach, using semantic perception for different semantics-extracted feature parameters as image restoration parameters, to repair the distorted image more accurately. Semantic feature extraction is added to the convolutional kernel network. The feature stream is generated through the convolutional layer, and the DC feature stream reshapes the output features through the convolutional layer to get the deviation between the distorted features and the original image features. The process can be represented as Equation (3):

$$\beta_i = W(A(i), \rho) \quad (3)$$

where W denotes the semantic-aware mapping function, $A(i)$ denotes the semantic features of the image, and ρ denotes the semantic parameters.

Affected by the complex underwater environment, the image feature differentiation is low, the feature data of the objective is less, the features are not obvious, which causes the objective recognition to be difficult, and the recognition accuracy is not high. In this chapter, the original sample image is enhanced by twisting and distorting the data, and the enhancement database is constructed to store and put the sample image after data enhancement. The original image of the same type of image and the different enhanced image are set as a positive sample x_i , and the image

affected by turbulence and its enhanced image is set as a negative sample x_j^- . A multi-scale perception pyramid network, denoted as $f_{\tau}(\cdot)$, was trained using a set of samples $\mathbb{X} = \{x_1, x_2, x_3, \dots, x_n\}$. The trained $f_{\tau}(\cdot)$ can extract multi-scale salient features from images.

The effect of the mapping function $\xi(\cdot)$ for the turbulence-induced distorted image features is verified by recoding the features of each input image. Let the relative position-distance relationship of the multi-scale features be D_n . The image multi-scale feature in-plane distance difference can be expressed by the equation Equation (4):

$$D_n(f_q^i, f_p^j) = \sum_i (\xi_n(\|f_q^i, f_p^j\|) \log(\xi_n(x_s^i))) + (1 - f_p^j(x_s^i) \log(1 - f_q^i(x_s^i))) \quad (4)$$

where $D_n(f_q^i, f_p^j)$ denotes the error of feature mapping to the plane f_q^i with respect to f_p^j . The autonomous repair of aberrations is achieved by reducing the spatial location of features in the objective samples and the objective library. To improve the repair accuracy of the features, find the way to reduce the error for optimal distortion image feature correction. The computational Equation (5) is:

$$\mathbb{R}_{q \rightarrow p}(f_q^i, f_p^j) = \underset{f_q^i}{\operatorname{argmin}} (D_n(f_q^i, f_p^j)) \quad (5)$$

The multi-scale perception model obtains a series of feature sub-vectors by processing the input image at different scales. These sub-vectors can be used for feature comparison and extraction in target detection tasks and for target detection and classification through contrast learning. The features of the multi-scale perceptual model of the object image can be represented by Equation (6):

$$F = \{f_{q0}(x), f_{q1}(x), f_{q2}(x), f_{q3}(x), f_{q4}(x)\} \quad (6)$$

where x is the test image that is provided as input, and $f_{qi}(x)$ is the multi-scale feature of the output of the input test image capture target, which is obtained by training the encoder f_q . Additionally, during the objective feature extraction process, the relative positional relationships among various features are simultaneously preserved.

3.2 Underwater distorted image correction

During objective detection, distorted images can pose challenges to recognition by causing the loss of crucial feature information. The extracted object features are fused so that the extracted features interact with each other with information to obtain more information about the object features under the influence of turbulence, which is more conducive to the aberration repair of the object. In this chapter, the multi-scale comparison correction method is used to correct the similar feature positions by comparing the input objective image features with the clear image feature information. The distorted image is fused with the input features to get an output image with richer feature information, which is then compared with the samples, and feature encoding and decoding are used to achieve the repair of distorted images. The process of constructing a distortion correction model requires adjusting the image features of the distorted image. This can be done by using the following Equation (7):

$$N_{a \rightarrow \gamma} = \|\zeta_a^+, \zeta_\gamma^-, \zeta_a^+ \text{ and } \zeta_\gamma^- \in \mathbb{N}^d \quad (7)$$

where ζ_a^+ is the feature vector of positive samples under the influence of turbulence, and ζ_γ^- is the feature vector of image under the influence of turbulence of negative samples with similar distance. $\|\cdot\|$ denotes the correction calculation of the positive and negative samples, which achieves the autonomous repair of the aberration by reducing the position of the feature space between the positive and negative samples.

Positive and negative samples based on task objectives are determined: the target to be recognized as positive samples and set categories that do not belong to the target category as negative samples is set. The data was preprocessed and organized to ensure its quality and validity. Positive and negative sample data were selected for annotation of the original data, ensuring the comparability and consistency of the samples throughout the entire dataset. During data augmentation, sufficient consideration

and processing of the data are carried out to ensure the accuracy and reliability of the data results.

The model learns by comparing positive and negative samples. The obtained features are gradually feature-biased toward positive samples with the expression Equation (8) as follows:

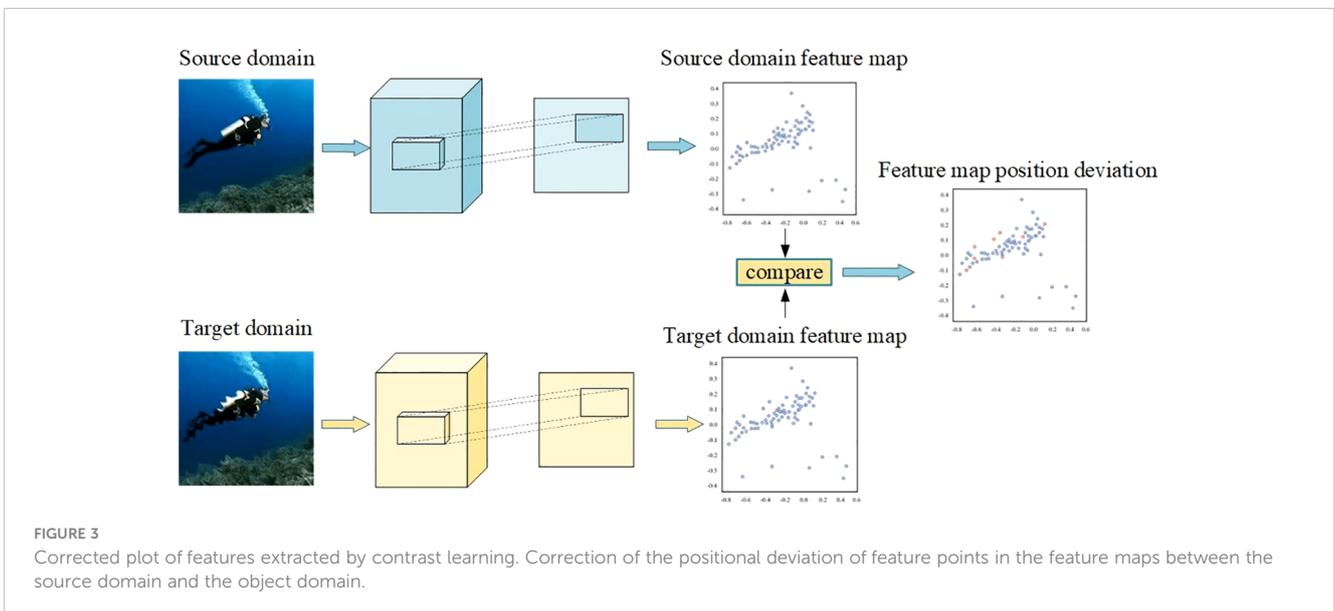
$$Y(x, x_i^+) \gg Y(x, x_i^-) \quad (8)$$

where $Y(x, x_i)$ represents the degree of similarity between the samples, and the difference between the positive and negative samples is changed by contrast learning, and the difference between feature similar samples is gradually reduced. The contrast module is used to correct the distorted image so that the image is recovered close to the original image. The positional deviation of the feature map is calculated by comparing the multi-scale object feature distance difference between the source and object feature domain maps. Comparative learning for feature extraction and correction is shown in Figure 3:

By introducing a combination of distance and pre-selected similarity to improve the similarity between samples, the performance and stability of the model can be improved, while also better adapting to the characteristics of different types of data. Let $x_{s \rightarrow t}^i(I, \theta)$ denote the learned training features for a batch of samples in the source domain S and distorted image samples in the objective domain T . Construct N positive sample pairs $(x_s^i(I, \theta), x_{s \rightarrow t}^i(I, \theta))$ and $N - 1$ negative sample pairs $(x_s^i(I, \theta), x_{s \rightarrow t}^i(I, \theta))$. The loss of domain contrast from the source domain to the abstract feature space is Equation (9):

$$L_S^I(\theta) = \frac{1}{N} \sum_i \log(\text{sim}(x_s^i(I, \theta), x_{s \rightarrow t}^i(I, \theta)) * D_n) + \frac{1}{N} \sum_j \log(\text{sim}(x_s^j(I, \theta), x_{s \rightarrow t}^j(I, \theta)) - D_n) + (1 - \text{sim}(x_s^i(I, \theta), x_{s \rightarrow t}^i(I, \theta))) \quad (9)$$

where $\text{sim}()$ denotes the corresponding cosine similarity relationship between the global feature vectors of the two views. $x_s^i(I, \theta)$ denotes the feature of a convolutional layer of the sample image, and τ denotes the temperature parameter.



To improve the restoration accuracy of the features, find the way to reduce the error for optimal distorted image feature correction. The computational Equation (10) is as follows:

$$\mathbb{R}_{a \rightarrow \gamma}(\zeta_a^+, \zeta_a^-) \text{ar } \text{gmin}(\mathbb{N}_{a \rightarrow \gamma}) \quad (10)$$

After correction, the calibrated image is obtained. The corrected image features are represented as $F_{q \rightarrow p}$, and the modified image is reconstructed using a decoder. The calibration process is illustrated in Figure 4.

3.3 Loss function

In this paper, we propose an underwater aberration objective identification method with multi-scale features to construct a multi-target scale monitoring network. The method solves the object aberrations caused by turbulence by training the multi-scale loss function to accomplish the accurate detection and identification of the object. In order to reduce the impact caused by redundancy on model detection, orthogonality loss is utilized to impose orthogonality in the feature space to maintain feature separation between different categories and feature aggregation in the same category and thus to the repaired aberrant object recognition.

Non-maximum suppression (NMS) for eliminating duplicate boxes is used to select the enclosing box for objective detection and suppress its neighboring boxes. Anchor points are utilized to generate bounding boxes, and redundancy is eliminated by suppression method. When establishing the bounding box, the predicted box is utilized as a reference.

In the process of network feature extraction construction, the image features of the distorted region are set as relevant semantics and their encoder counterparts, and the network classification loss

function is calculated as shown in Equation (11):

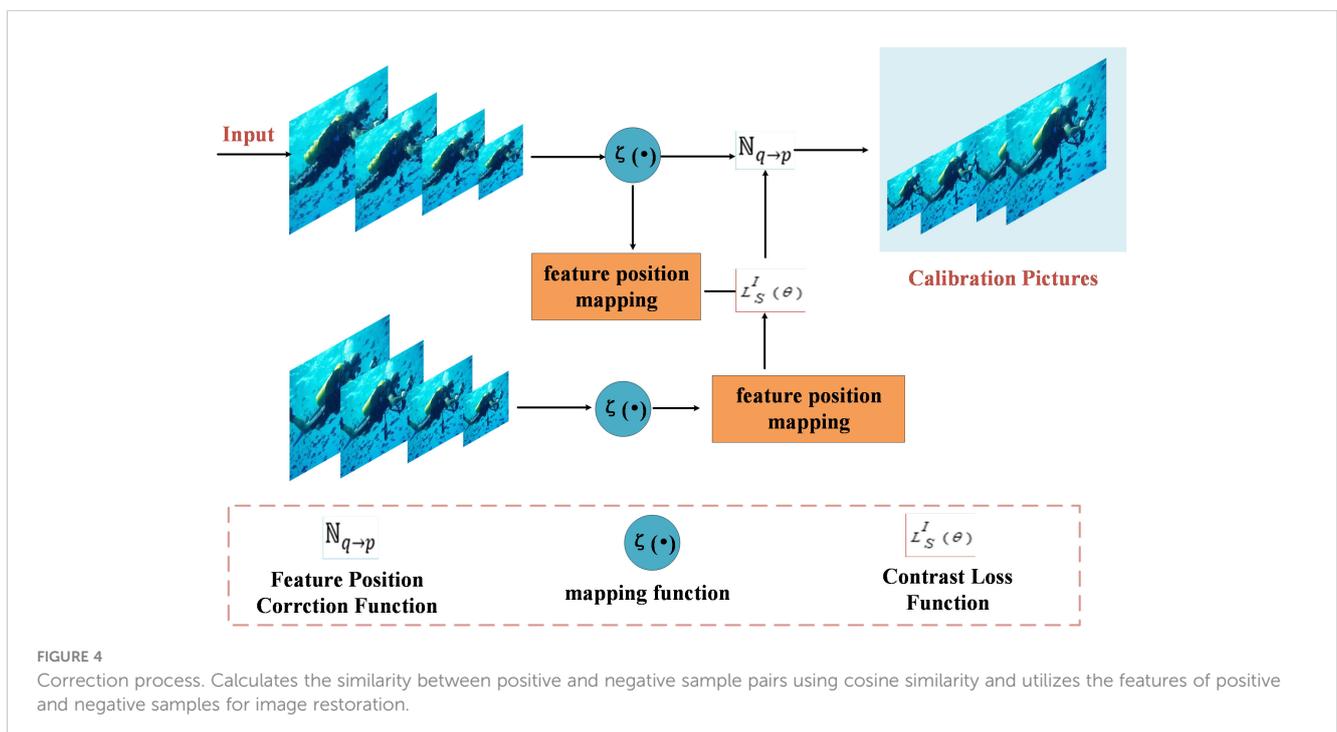
$$L_{cls}(p_i, p_i^*) = \|\beta_i - W_i\|^2 + \|D_n - W_d\|^2 \quad (11)$$

where W_i denotes the number of relevant features, and W_d denotes the feature space corresponding to the relevant features.

During the network training process, a loss function is used to measure the error between the predicted values and the true values of the objective detection. This Equation (12) can be defined as follows:

$$LOSS = \lambda_c \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda_n \frac{1}{N_{rpu}} \sum_i L_{rpu}(c_i, c_i^*) + \lambda_l \frac{1}{N_{loc}} \sum_i L_{loc}(t_i, t_i^*) \quad (12)$$

where i is the anchor, p_i is the predicted probability of the anchor, p_i^* is the true anchor probability, L_{loc} is the loss for predicting the coordinates, and t_i and t_i^* denote the predicted bounding box and the vector associated with the anchor, respectively. L_{rpu} is the loss of distinguishing between foreground background and fine-tuned anchors. c_i and c_i^* , the true confidence interval and the prediction confidence interval, are denoted to represent different aspects. L_{cls} is the category loss. The three terms N_{loc} , N_{rpu} , and N_{cls} are the normalization parameters. The model in this chapter sets $\lambda_c = 0.3$, $\lambda_n = 0.5$, and $\lambda_l = 0.2$. Among λ_n is the weight loss of the confidence interval $\lambda_c + \lambda_n + \lambda_l = 1$. The weight of the entire loss is equivalent to one unit. The objective of the loss function is to minimize the difference between the predicted and true values of the model, and the coefficient parameters determine the importance of each loss term in the overall loss. For specific loss items that require more attention and attention, their weight in the overall loss can be increased or decreased by adjusting the coefficient parameters. Different loss terms may have different measurement units and magnitudes, and their importance may also vary. A coefficient parameter of 1 in the loss function indicates that



each loss term has different weights, which is useful for balancing the importance of different loss terms. Adjusting the coefficient parameters can better reflect the importance and contribution of each loss term, thereby achieving better model performance.

To ensure its robustness, an appropriate convergence interval is defined during the iterative calculation process, and a threshold is set as the convergence condition. When the parameter changes during the iteration process do not affect the objective function value, the algorithm has converged. By controlling the step size of each iteration, the stability of the algorithm is ensured and instability caused by parameter updates that are too fast or too slow is avoided. By setting an appropriate learning rate to enable the model to converge quickly, optimizing the prediction results by changing the learning rate, and gradually reducing the learning rate as training progresses, the model can converge quickly during training and update parameters more stably in the later stages of training. It also helps the model to output more stable probability values during prediction, thereby improving the accuracy and reliability of the model. The Equation (13) is as follows:

$$\varepsilon = \varepsilon_0 \cdot \left(1 - \frac{t}{T}\right)^p \quad (13)$$

where ε_0 represents the initial learning rate, p is power, t represents the number of training epochs, and T is the total number of training epochs.

4 Experimental results and analysis

4.1 Experimental setup

4.1.1 Experimental environment

In this experiment, training, validation, and testing are performed on a small server with Intel(R) Core (TM) i7-1165G7 CPU, RTX 3090 GPU, and 64G RAM. In the comparison experiments, to reflect the objectivity of the proposed method, it is implemented using PyCharm deep learning tool.

4.1.2 Datasets

The main research problem addressed in this paper is the challenges posed by distorted underwater object images and the approach to address them using a multi-scale feature attention method. The datasets used in this paper are two datasets categorized by the publicly available underwater target identification datasets CADDY, NATURE Central. These datasets contain a total of 3,215 images, which are used to train the multi-scale salient feature extraction model. To ensure the effective training of the proposed objective aberration image correction recognition network under turbulence, a data augmentation technique is employed to expand the dataset by a factor of four. Consequently, the dataset is increased to 12,860 images, with 9,002 images used for training, 2,572 for testing, and 1,286 for the validation set, maintaining a ratio of 7:2:1.

4.1.3 Learning rate and training settings

In this chapter, the recognition means accuracy rate (MAP) and frame rate (FPS) are chosen as the evaluation metrics for the

proposed algorithm as well as the comparative performance of the algorithms. The parameters of the proposed algorithm are set as follows: momentum is set to 0.9, weight decay is set to 0.0005, initial learning rate is set to 0.01, batch size is set to 24, and the algorithm iterates 200 epochs throughout the training process.

4.1.4 Assessment of indicators

Experimental validation was performed to evaluate the object recognition performance of the model when subjected to turbulence, the mean accuracy (MAP), distortion parameter, frames per second (FPS), and evaluation metrics for object detection under the influence of turbulence including MAP and FPS. MAP is used to evaluate the accuracy of object detection, while FPS is used to evaluate the real-time processing speed of the model. Higher MAP and FPS values represent better detection and faster processing. The calculation Equations (14)–(16) is as follows:

$$MSE = \frac{1}{ml} \sum_{k=0}^{m-1} \sum_{u=0}^{l-1} [I(k, u) - K(k, u)]^2 \quad (14)$$

I is the image under the influence of turbulence, the K is the corrected image, and ml is the input image size.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (15)$$

μ is the mean, σ is the variance, and c is a constant.

$$MAP = \frac{\sum_{i=1}^{n-1} (r_{i+1} - r_i) p_{inter} p(r_{i+1})}{K} \quad (16)$$

K is the category, and r_i is the identification value of each point.

4.2 Results and analysis

4.2.1 Objective distortion correction

In this paper, six sets of experiments are conducted to verify the effectiveness and accuracy of the proposed method in repairing turbulence-induced image distortion and performing multi-scale perceptual object recognition. To showcase the effectiveness of the method proposed in this paper, several algorithms including CEEMDAN-Fast (Lin et al., 2022a), COCO (Mensink et al., 2021), yolov4 (Xu et al., 2022), and Scaled-yolov5 (Scoulding et al., 2022), Yolov7 (Wang et al., 2023a), TOOD (Feng et al., 2021), Boosting R-CNN (Song et al., 2023) as well as comparative validation with the algorithms proposed in this paper have been employed. This comprehensive approach enables a thorough evaluation and comparison of the proposed method with existing algorithms to highlight its performance and efficacy.

In the turbulence distortion image correction experiments, we focus on correcting and reconstructing input images that have turbulence-induced aberrations. Specifically, we select submarine, fish, diver, and AUV as the aberration objects. The results of these experiments are illustrated in Figure 5. The first column in Figure 5 displays the original aberration images. The subsequent four columns represent the results after different correction iterations. Each column corresponds to a specific correction iteration. Finally,

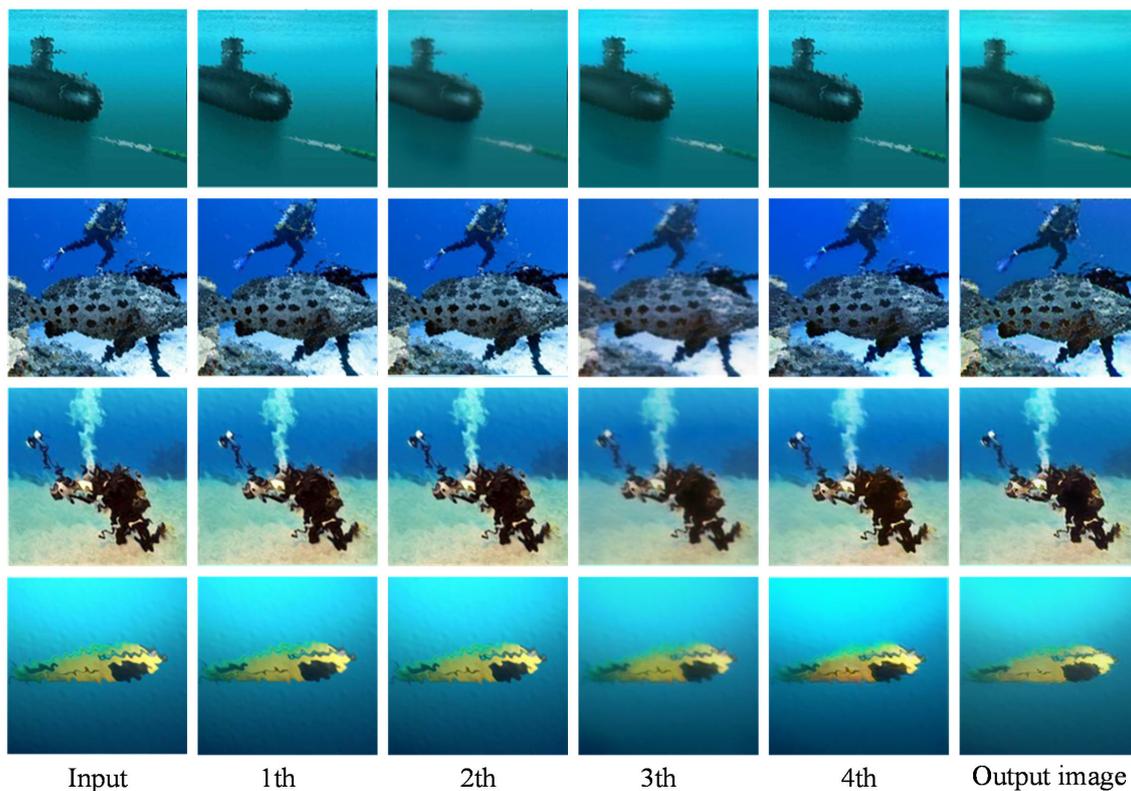


FIGURE 5 Correction and reconstruction results of distorted images.

the last column shows the output of the corrected image, which can be used for subsequent object detection.

This paper demonstrates the evolution of the reference frame during the distortion process. The results of the image quality evaluation of the reference frame at each iteration of the algorithm are shown in Table 1, where the gap between the estimated and real values of the parameter frames is decreasing as the number of iterations increases. The quality difference between the reference image and the parameter frame is gradually decreasing. The structural similarity between the parameter frame and the reference image gradually increases. As the number of iterations increases, the image quality of the parameter frames gradually improves. The MSE continues to decrease, and the PSNR and SSIM continue to increase, indicating that the gap between the estimated and true values of the parameter frames is decreasing, the quality difference between the reference image and the parameter frames is gradually decreasing, and the structural similarity is

gradually improving. The image quality of the parameter frame is improved. By comparing the input images with the output images, it can be observed that the algorithm proposed in this paper effectively corrects underwater distorted images with specific distortion characteristics. The correction of these distorted images helps reduce the impact of environmental factors on objective recognition. In other words, the algorithm mitigates the negative effects of distortions, resulting in improved image quality and more accurate recognition of objects in underwater environments.

To evaluate the detection performance of the method proposed in this paper, several experiments were conducted. We recognize the original image and compare the recognition results with the above-cited literature. The specific recognition results are shown in Figure 6. The results show that the recognition accuracy is improved compared with the traditional recognition, and each method shows a relatively stable detection confidence.

From Table 2, this paper’s algorithm has good recognition accuracy that is highest for frogmen and submarines, which is 0.9310 and 0.9027, respectively. CEEMDAN-Fast performs better on most of the categories, especially on the fish and frogmen categories with high detection accuracy. However, its MAP is low and the frame rate is slow. The COCO method achieved high detection accuracy on all categories and had high MAP values. Yolov4 showed high detection accuracy on most categories, especially on fish. It has a high MAP and a fast frame rate relative to the other methods. Scaled-yolov5 exhibits high detection accuracy on AUV categories but slightly decreases on other

TABLE 1 Image quality evaluation results of the parameter frames.

Iteration	MSE	PSRN	SSIM
1st	0.01421	20.1934	0.4864
2nd	0.0122	21.4257	0.6170
3rd	0.0120	22.4323	0.6854
4th	0.0010	22.5986	0.7475

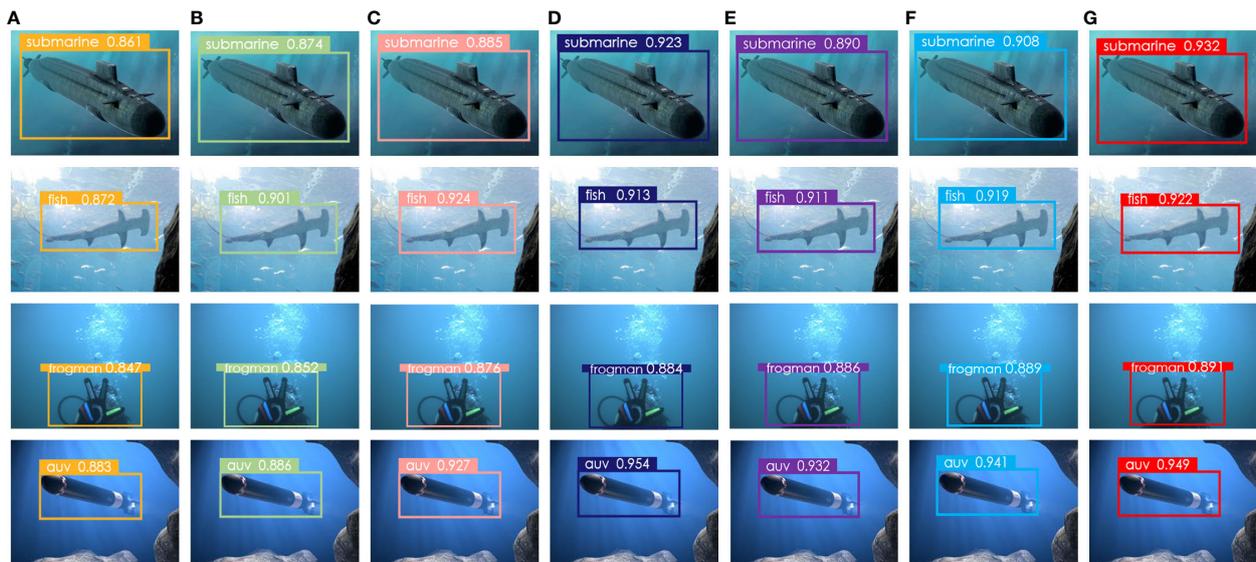


FIGURE 6 Underwater multi-scale objective image recognition results. (A) CEEMDAN-Fast, (B) Coco, (C) Yolov4, (D) Scaled-yolov5, (E) Yolov7, (F) TOOD, and (G) ours.

categories. It has a relatively high MAP and certain frame rate. Yolov7 achieved relatively high performance scores in the fish, diver, and AUV missions but performed slightly lower in the submarine mission. The TOOD achieved relatively high performance scores in the diver task but performed poorly in the other tasks. The method in this paper shows high detection accuracy on most of the categories, especially on the diver and submarine categories. It also shows relatively good MAP and frame rate. Overall, in terms of MAP and FPS metrics, our method performs best in submarines and frogmen, while the CEEMDAN-Fast method performs relatively poorly. The other methods have their own advantages and disadvantages, and their performance varies in different tasks.

4.2.2 Image distortion correction recognition

Part of the effect of the algorithm proposed in this paper is shown in Figure 7, from which the effect of objective recognition under the influence of turbulence can be seen. This paper's algorithm is for divers and submarines in several categories of

recognition of high confidence. This paper's algorithm for the recognition of the objective aberration caused by the turbulence confidence remains relatively stable, and in the recognition of the aberration of underwater objectives it is excellent.

This algorithm recognizes the objective image without distortion correction, so the average accuracy and number of frames recognized by the algorithms in this chapter and the comparison algorithms are shown in Table 3, where the entries in bold black font are the best data.

From Table 3, the highest recognition accuracies for submarines and divers in this chapter are 0.9201 and 0.8905. COCO shows high detection accuracy on most categories, which has high MAP values and fast frame rates, showing a strong objective detection. Yolov4 shows high detection accuracy on most categories, especially on the fish category with high accuracy. It has a high MAP and a certain frame rate. Scaled-yolov5 shows high detection accuracy on most categories, with high accuracy on the AUV category. It also shows relatively good MAP and frame rate. Our method shows high detection accuracy on most categories, especially on the

TABLE2 Results of underwater distortion original image identification.

Method	Submarine	Fish	Frogman	AUV	MAP	FPS
CEEMDAN- Fast	0.8314	0.8572	0.8347	0.8716	0.8487	25.3
COCO	0.9011	0.8539	0.9013	0.8916	0.8869	41.7
Yolov4	0.8851	0.9327	0.8651	0.8739	0.8892	38.1
Scaled-yolov5	0.9213	0.9013	0.8534	0.9015	0.8943	30.4
Yolov7	0.8724	0.9261	0.8425	0.8364	0.8693	35.1
TOOD	0.8135	0.8341	0.8371	0.8211	0.8264	30.2
Ours	0.9310	0.9014	0.9027	0.8891	0.9060	32.7

The bold part represents the optimal value.

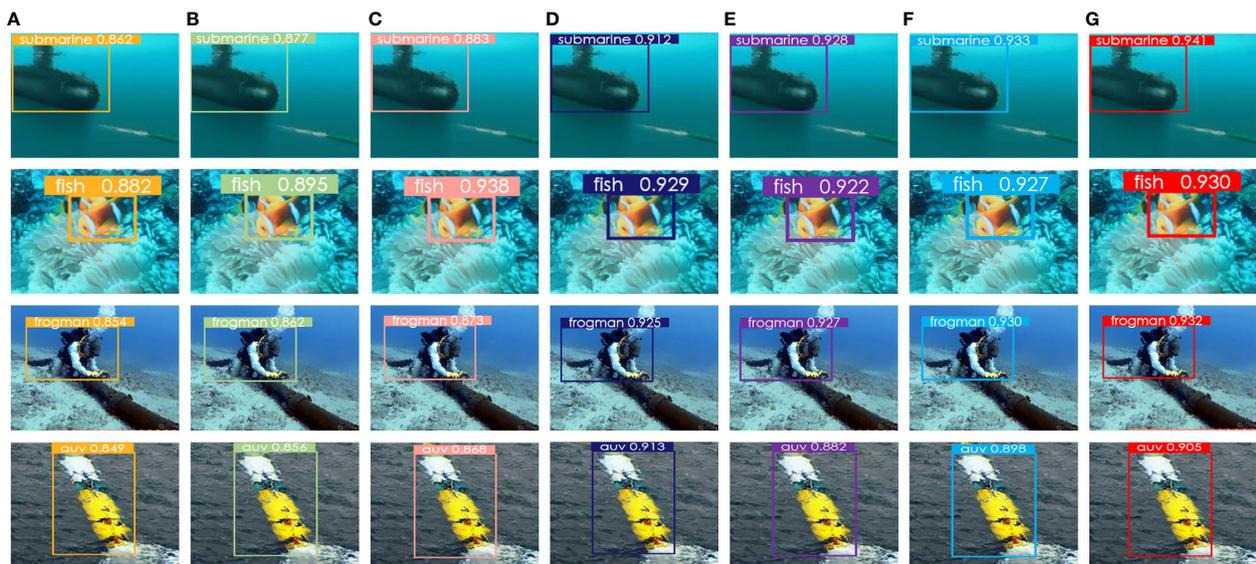


FIGURE 7 Underwater distortion uncorrected objective image recognition results. (A) CEEMDAN-Fast, (B) Coco, (C) Yolov4, (D) Scaled-yolov5, (E) Yolov7, (F) TOOD, and (G) ours.

submarine and frogman categories. It also performs relatively well in terms of MAP and frame rate. Yolov7 achieved high performance scores in the fish task but performed poorly in the other tasks. TOOD is more stable across task recognition. In summary, based on the MAP and FPS metrics, the methods in this paper perform relatively well in all tasks, especially in submarines and frogmen with high recognition rates. The other methods have their own strengths and weaknesses, and their performance varies from task to task.

4.2.3 Recognition result of distorted objective

Figure 8 demonstrates some of the effect diagrams of the algorithms in this chapter for aberration recognition. From Figure 8, the proposed algorithms in this chapter have the highest confidence level of 0.8523 and 0.8517 for the recognition of frogman and submarines, which is higher than other algorithms, and has a better recognition ability compared to other algorithms.

From Table 4, it can be seen that our method demonstrates a certain level of detection accuracy in most categories, especially

excelling in the submarine and diver categories, achieving accuracy values of 0.8517 and 0.8523, respectively. In comparison, CEEMDAN-Fast performs well in most categories, particularly showing high detection accuracy in the fish and diver categories. The COCO method achieves high detection accuracy in all categories and has a higher mAP value. Yolov4 shows high detection accuracy in most categories while also having higher mAP and faster frame rates compared to other methods. Scaled-yolov5 performs well in the AUV category but slightly decreases in other categories. Boosting R-CNN has high detection accuracy, especially in the fish category, but has the lowest frame rate. Our method demonstrates high detection accuracy in most categories, particularly excelling in the diver and submarine categories, while also performing relatively well in terms of mAP and frame rate.

4.2.4 Ablation study

Ablation experiments are performed to verify the effectiveness of the feature extraction module and distortion correction module of the target image in the method proposed in this paper. The

TABLE 3 Results of underwater distorted images.

Method	Submarine	Fish	Frogman	AUV	MAP	FPS
CEEMDAN- Fast	0.7921	0.8716	0.8244	0.8371	0.8313	27.6
COCO	0.8847	0.8342	0.8901	0.8611	0.8675	45.3
Yolov4	0.8842	0.9214	0.8231	0.8514	0.8699	40.8
Scaled-yolov5	0.8914	0.8724	0.8264	0.8623	0.8631	33.4
Yolov7	0.8513	0.9017	0.8375	0.8102	0.8501	34.1
TOOD	0.8003	0.8235	0.8251	0.8024	0.8128	31.8
Ours	0.9201	0.8736	0.8905	0.8357	0.8799	32.5

The bold part represents the optimal value.

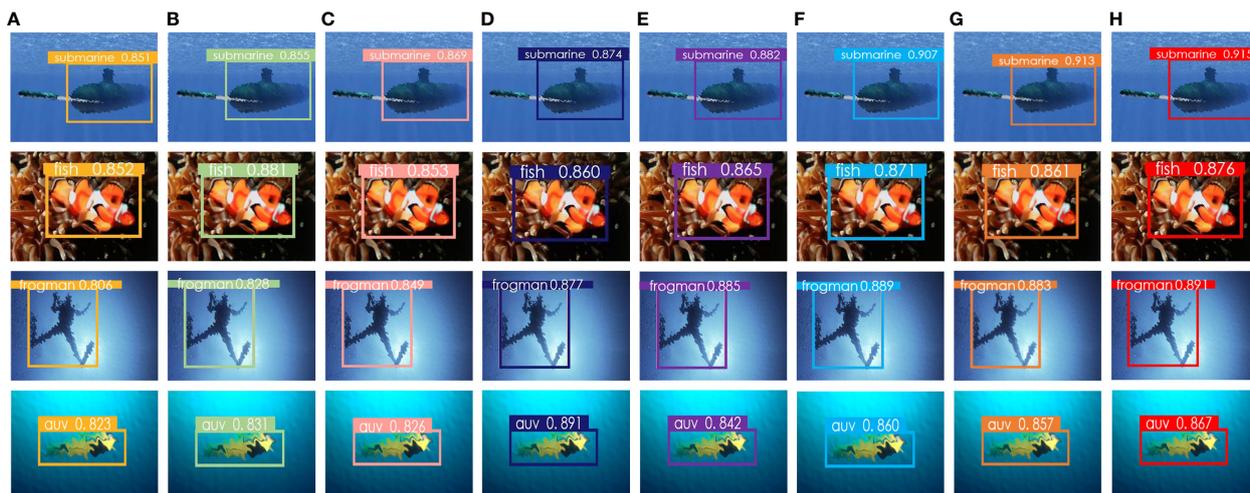


FIGURE 8 Objective distortion recognition diagram. (A) CEEMDAN-Fast, (B) Coco, (C) Yolov4, (D) Scaledyolov5, (E) Yolov7, (F) TOOD, (G) Boosting R-CNN, and (H) ours.

experiment was performed on a publicly available dataset. The densest network was used as the baseline for the ablation experiment. The results of the experiment are shown in Table 5, and the images are shown in Figure 9.

The feature extraction pyramid module and the distortion correction module correct and recognize the image object by making the object features more visible, i.e., highlighting the contours and textures of the object. The image object can be made clearer and thus easier for object detection and recognition. The object feature information is extracted by multi-scale features of the object image, and the dense link and position strategy is used to compare to realize the distorted image repair. The combination of the two modules not only repairs the distorted image but also provides help in the recognition of the distorted image.

From Table 5, it can be observed that the recognition accuracies for submarine, fish, diver, and AUV in the network are 0.7199, 0.7537, 0.7741, and 0.7368, respectively. With the addition of the multi-feature extraction module to the network, the recognition accuracies for submarine, fish, diver, and AUV all improved to 0.8341, 0.7954, 0.8363, and 0.7961, respectively. Similarly,

incorporating the distortion correction module into the network also resulted in improved recognition accuracies for submarine, fish, diver, and AUV, which were 0.8372, 0.7911, 0.8325, and 0.7905, respectively. When both the multi-feature extraction module and distortion correction module were added to the network simultaneously, significant improvements were observed in all categories. The recognition accuracies for submarine, fish, diver, and AUV were 0.8517, 0.8037, 0.8523, and 0.8014, respectively. Overall, the performance of the network improved when both modules were added compared to when only one of them was included.

To prove the effectiveness of each module for our MATNet for distorted object identification, we performed ablation studies on the proposed dataset, namely: our MATNet without autoencoder(w/a), our MATNet without pyramid model (w/PM), our MATNet without feature extraction (w/FE), and our MATNet without distortion correction (w/DC).

Table 6 exhibits the submarine fish frogman and AUV scores corresponding to the ablated models. Which can be shown that our MATNet (full model) has the best score compared with other

TABLE 4 Results of underwater distorted objective identification.

Method	Submarine	Fish	Frogman	AUV	MAP	FPS
CEEMDAN-Fast	0.7372	0.8572	0.7864	0.8216	0.8006	22.3
COCO	0.8510	0.7467	0.8470	0.8418	0.8216	42.5
Yolov4	0.8307	0.8824	0.8154	0.8039	0.8251	39.2
Scaled-yolov5	0.8213	0.8143	0.7965	0.8451	0.8193	32.7
Yolov7	0.8341	0.8712	0.8113	0.7942	0.8249	31.4
TOOD	0.7871	0.8031	0.8023	0.7835	0.7940	32.7
Boosting R-CNN	0.8124	0.8835	0.8012	0.7911	0.8221	20.1
Ours	0.8517	0.8037	0.8523	0.8014	0.8272	31.8

The bold part represents the optimal value.

TABLE 5 Quantitative evaluation of turbulence-induced distorted image ablation experiments.

MFM	DC	Submarine	Fish	Frogman	AUV	MAP	FPS
		0.7199	0.7537	0.7741	0.7368	0.7461	30.7
√		0.8341	0.7954	0.8363	0.7961	0.8154	30.5
	√	0.8372	0.7911	0.8325	0.7905	0.8128	31.4
√	√	0.8517	0.8037	0.8523	0.8014	0.8272	32.5

The bold part represents the optimal value.

modules. Additionally, it also proves that each module has a positive effect on our MATNet.

5 Discussion

With the increasing research on the ocean, the complexity and variability of the marine environment make it challenging to explore the ocean. Light produces color-bias degradation during transmission and reconstruction of images to achieve image enhancement (Li et al., 2022b). This study investigates the effect of turbulence on the performance of marine object recognition, where feature extraction and recognition of objects become challenging in turbulent environments. Underwater image features are usually affected by contrast degradation, low illumination, color bias, and noise (Mishra et al., 2022). Marine deep learning methods still have high recognition accuracy and

robustness in turbulent environments. Although turbulence causes distortions and warping of images and data, deep learning can better overcome these disturbances. A better understanding of the impact on object recognition under the influence of turbulence is essential to improve the performance of marine object detection and recognition systems. These findings are useful for improving object recognition in marine environments and optimization of ruthless learning algorithms to adapt to turbulence disturbances. A deeper understanding of the effects of turbulence on object recognition have important deed applications as well.

The research field of object recognition based on underwater images does face many challenges, but many scholars have actively conducted research and proposed various solutions. One of the novel imaging algorithms is the use of multi-receiver synthetic aperture sonar (MSSA) technology, which aims to provide high-resolution images of underwater objects. This algorithm utilizes coherently superimposed continuous echo signals to generate high-

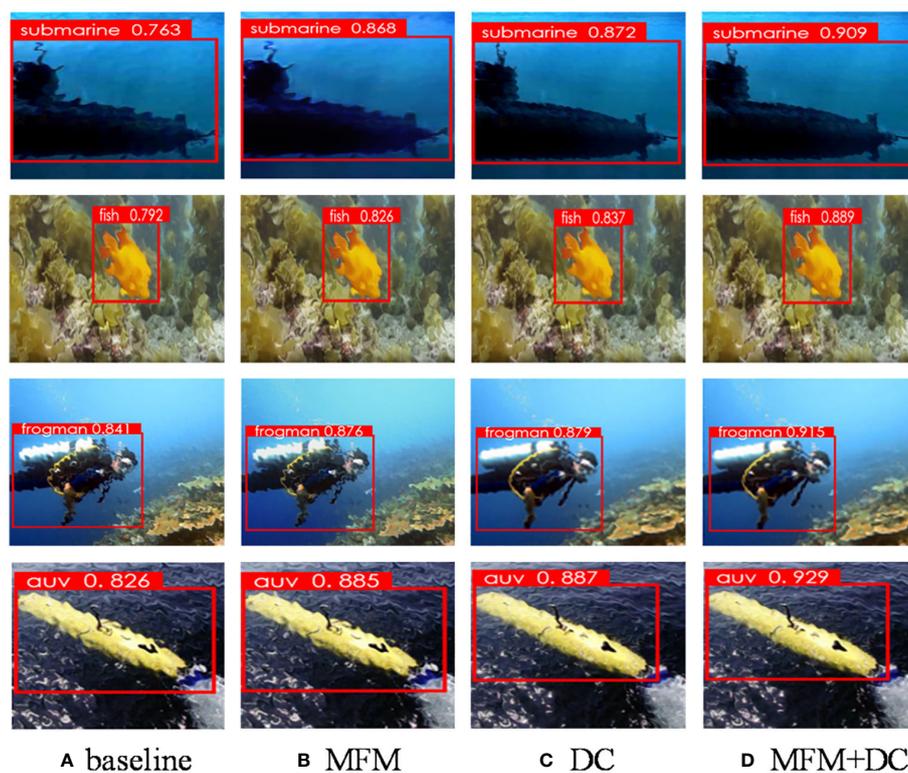


FIGURE 9 Experimental effect of turbulence-induced ablation of distorted images: (A) no module added, (B) MFM added, (C) DC added, and (D) both modules added.

TABLE 6 Results of ablation studies of different modules (optimal: red; suboptimal: blue).

Method	Submarine	Fish	Frogman	AUV	MAP	FPS
-w/a	0.8374	0.7921	0.8331	0.7923	0.8136	32
-w/PM	0.8380	0.7947	0.8369	0.7957	0.8163	33
-w/FE	0.8372	0.7911	0.8325	0.7905	0.8128	31.4
-w/DC	0.8341	0.7954	0.8363	0.7961	0.8154	30.5
MATNet (full model)	0.8517	0.8037	0.8523	0.8014	0.8272	32.5

resolution images. The signals are fused and processed by multiple receivers, and a signal scene model is constructed. At the same time, single–multiple interactions are introduced for sampling, and ultimately multiple signals are processed to obtain high-resolution images of underwater objects. This technique can improve the resolution and clarity of object recognition in underwater images, allowing researchers to analyze and identify underwater targets more accurately. However, it should be noted that the complexity and specificity of underwater environments still pose some challenges, such as light attenuation, noise interference, and changes in the shape of objects. Therefore, researchers are still working to improve the algorithms and techniques and to explore more efficient methods for object recognition in underwater images.

Prior research on this topic has carried out some significant work in marine object recognition and deep learning, providing an appropriate context for the development of this thesis. Many previous studies have proposed a variety of object recognition in marine environments, such as traditional methods combining feature extraction and classifiers (Jiang et al., 2022a), model-based methods, and deep learning methods (Abeyasinghe et al., 2022). These studies have provided a certain foundation for marine object recognition in terms of marine organisms, marine environment, and seabed targets. In the research on turbulence, turbulence is one of the important disturbances in the marine environment, but previous studies have focused on understanding the effects of turbulence on distortion and distortion of images and data. Through simulations and experimental analyses, the researchers discuss the effects of turbulence on the performance of object recognition algorithms and propose corresponding processing methods and improvement strategies. Previous studies have made some progress in the field of marine object recognition and deep learning, but relatively limited research has been conducted on object recognition in turbulent environments. Therefore, the findings in this thesis further explore the impact of turbulence on marine object recognition and propose improvement strategies for deep learning methods in turbulent environments. This research fills the research gap in the related field while providing new understanding and solutions for object recognition in turbulent environments with significant novelty. Although the research in this thesis has made important findings, there are still potential drawbacks and limitations in terms of experimental condition limitations, dataset selection, and comparison between deep learning models and real scenes. Future research should further remedy these limitations to improve the credibility and applicability of the study.

The findings of this thesis suggest that deep learning models are robust in turbulent environments and can achieve high recognition accuracy. Based on this observation, future research can further improve the deep learning model to better adapt to different types and intensities of turbulent disturbances—for example, new network structures or optimization algorithms can be explored to improve model robustness and object recognition performance. Consider the fusion of multimodal information: in addition to image data, marine object recognition may involve other sensors or data sources, such as sonar, LiDAR, etc. Future research could consider fusing multimodal information for object recognition, considering the effect of turbulence on multiple sensor data, and investigating how to optimize the fusion algorithm to improve the performance of object recognition. With regards integration with scene perception, object recognition is often affected by complex background disturbances in real marine environments, such as waves, sea spray, and so on. Future research can combine object recognition with scene perception technology to improve the performance of object recognition through the perception and understanding of environmental features. The interrelationship between turbulence and background features may become an important direction for future research to promote the application of object recognition in complex marine environments. In future research, the following hypothesis can be tested: it is assumed that the effect of turbulence on object recognition is influenced by the scale and shape of the target object. Smaller-scale and irregularly shaped objects may be more susceptible to turbulence, whereas larger-scale and regularly shaped objects may have better robustness. To test this hypothesis, experiments can be designed and collected on the recognition accuracy of objects with different scales and shapes under different turbulence intensities. By comparing the performance of different object types, the relationship between object properties and the effect of turbulence on object recognition can be further explored to provide more in-depth theoretical support for the optimization of object recognition algorithms.

6 Conclusions

This paper presents a MATNet aberration correction recognition method. The method in this paper considers the loss of objective features and the difficulty of feature extraction caused by the underwater environment in image recognition. In this paper, two learning strategies are used for multi-scale feature extraction,

location learning strategy and density linking strategy, which are invoked to efficiently extract underwater objective features. Contrast correction of the distorted objective by the extracted multi-scale features restores a sharper image. The corrected image is subsequently employed to train the objective network for recognition. During the training process, a loss function is utilized to optimize the network parameters and ensure precise recognition of distorted objects. Extensive evaluation experiments on a variety of scenes show that our method is effective for image recognition due to aberrations and achieves good results in image restoration and object detection in complex scenes. Despite the superior performance of the method proposed in this paper in recognizing distorted images, it is still unsatisfactory for recognizing underwater objective images in dim and blurred environments, and we take this challenge as a problem that needs to be solved in the future.

Data availability statement

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author.

Author contributions

MZ: Writing – original draft. LC: Conceptualization, Writing – review & editing. JJ: Data curation, Writing – review & editing. WZ: Writing – review & editing. YG: Investigation, Writing – review & editing.

References

- Abeysinghe, R., Zheng, F., Bernstam, E. V., Shi, J., Bodenreider, O., and Cui, L. (2022). A deep learning approach to identify missing is-a relations in snomed ct. *J. Am. Med. Inf. Assoc.* 30, 475–484. doi: 10.1093/jamia/ocac248
- Cai, L., Li, Y., Chen, C., and Chai, H. (2022a). Dynamic multiscale feature fusion method for underwater target recognition. *J. Sensors* 2022, 1–10. doi: 10.1155/2022/8110695
- Cai, L., Luo, P., Xu, T., and Chen, Z. (2022b). M-pfgmnet: multi-pose feature generation mapping network for visual object tracking. *Multimed. Tools Appl.* doi: 10.1007/s11042-022-12875-3
- Cai, L., Qin, X., and Xu, T. (2022c). Ehdc: Enhanced dilated convolution framework for underwater blurred target recognition. *Robotica*, 1–12. doi: 10.1017/S0263574722001059
- Chen, Q., Pan, G., Zhao, L., Fan, J., Chen, W., and Zhang, A. (2023). An adaptive hybrid attention based convolutional neural net for intelligent transportation object recognition. *IEEE Trans. Intell. Transport. Syst.* 24, 7791–7801. doi: 10.1109/TITS.2022.3227245
- Cheng, H., Liang, J., and Liu, H. (2022). Image restoration fabric defect detection based on the dual generative adversarial network patch model. *Text. Res. J.* 93, 2859–2876. doi: 10.1177/00405175221144777
- Deng, C., Wang, M., Liu, L., Liu, Y., and Jiang, Y. (2021). Extended feature pyramid network for small object detection. *IEEE Trans. Multimed.*, 1968–1979. doi: 10.1109/tmm.2021.3074273
- Eigel, M., Gruhlke, R., Moser, D., and Grasedyck, L. (2022). Numerical upscaling of parametric microstructures in a possibilistic uncertainty framework with tensor trains. *Comput. Mech.* 71, 615–636. doi: 10.1007/s00466-022-02261-z
- Feng, C., Zhong, Y., Gao, Y., Scott, M. R., and Huang, W. (2021). “Tood: Task-aligned one-stage object detection,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*. (Los Alamitos, CA, USA: IEEE Computer Society). 3490–3499. doi: 10.1109/iccv48922.2021.00349
- Guo, J., Liu, J., and Xu, D. (2022). 3d-pruning: a model compression framework for efficient 3d action recognition. *IEEE Trans. Circuits Syst. Vid. Technol.* 32, 8717–8729. doi: 10.1109/TCSVT.2022.3197395
- He, J., Dong, C., Liu, Y., and Qiao, Y. (2022). Interactive multi-dimension modulation for image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 9363–9379. doi: 10.1109/TPAMI.2021.3129345
- Huang, M., Wen-hui, D., Yan, W., and Wang, J. (2023). High-resolution remote sensing image segmentation algorithm based on improved feature extraction and hybrid attention mechanism. *Electronics* 12, 3660–3660. doi: 10.3390/electronics12173660
- Hyun, J., Seong, H., Kim, S., and Kim, E. (2021). Adjacent feature propagation network (afpnet) for real-time semantic segmentation. *IEEE Trans. Sys. Man Cybernet.: Syst.*, 1–12. doi: 10.1109/tsmc.2021.3132026
- Jiang, L., Quan, H., Xie, T., and Qian, J. (2022a). Fish recognition in complex underwater scenes based on targeted sample transfer learning. *Multimed. Tools Appl.* 81, 25303–25317. doi: 10.1007/s11042-022-12525-8
- Jiang, X., Zhu, Y., and Jian, J. (2022b). Two efficient nonlinear conjugate gradient methods with restart procedures and their applications in image restoration. *Nonl. Dynam.* 111, 5469–5498. doi: 10.1007/s11071-022-08013-1
- Kim, J., Cho, Y., and Kim, J. (2022). Urban localization based on aerial imagery by correcting projection distortion. *Autonomous Robots*. doi: 10.1007/s10514-022-10082-5
- Li, C., Anwar, S., Hou, J., Cong, R., Guo, C., and Ren, W. (2021). Underwater image enhancement via medium transmission-guided multi-color space embedding. *IEEE Trans. Image Process.* 30, 4985–5000. doi: 10.1109/TIP.2021.3076367
- Li, Y., Cai, L., and Jia, J. (2022d). “Msf: Multi-scale significant feature correction method for distorted underwater target recognition,” in *2022 International Conference on Advanced Robotics and Mechatronics (ICARM)*. (Guilin, China: IEEE). doi: 10.1109/ICARM54641.2022.9959229

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Henan Provincial focus on research and development Project (23111220700), Henan Provincial Major Project (221100110500), Henan Provincial Major Project (232102320338), the Natural Science Foundation of Henan Province under Grant (232300420428).

Acknowledgments

We thank the editors and reviewers for providing valuable comments that have helped enhance this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., et al. (2019). An underwater image enhancement benchmark dataset and beyond. *IEEE Trans. Image Process.* 29, 4376–4389. doi: 10.1109/TIP.83
- Li, Z.-X., Wang, Y.-L., Chen, P., and Peng, Y. (2022e). Laplace dark channel attenuation-based single image defogging in ocean scenes. *Multimed. Tools Appl.* 82, 21535–21559. doi: 10.1007/s11042-022-14103-4
- Li, R., Wang, H., Liu, Z., Cheng, N., and Xie, H. (2022a). First-person hand action recognition using multimodal data. *IEEE Trans. Cogn. Dev. Syst.* 14, 1449–1464. doi: 10.1109/TCDS.2021.3108136
- Li, W., Yang, X. J., Liu, Y., and Ou, X. (2022b). Single underwater image enhancement based on the reconstruction from gradients. *Multimed. Tools Appl.* 82, 16973–16983. doi: 10.1007/s11042-022-14158-3
- Li, X., Zhou, J., Xin, H., Li, W., Yin, X., Yuan, X., et al. (2022c). Vision measurement system for geometric parameters of tubing internal thread based on double-mirrored structured light. *Opt. Express* 30, 47701–47701. doi: 10.1364/OE.479067
- Lin, X., Yang, S., and Liao, Y. (2022a). Backward scattering suppression in an underwater lidar signal processing based on ceemdan-fastica algorithm. *Opt. Express* 30, 23270–23270. doi: 10.1364/OE.461007
- Lin, Y., Yu, C.-M., Huang, J. Y.-T., and Wu, C.-Y. (2022b). The fuzzy-based visual intelligent guidance system of an autonomous underwater vehicle: Realization of identifying and tracking underwater target objects. *Int. J. Fuzzy Syst.* 24, 3118–3133. doi: 10.1007/s40815-022-01327-7
- Liu, Y., Yan, Z., Tan, J., and Li, Y. (2023). Multi-purpose oriented single nighttime image haze removal based on unified variational retinex model. *IEEE Trans. Circuits Syst. Vid. Technol.* 33, 1643–1657. doi: 10.1109/TCSVT.2022.3214430
- Mensink, T., Uijlings, J., Kuznetsova, A., Gygli, M., and Ferrari, V. (2021). Factors of influence for transfer learning across diverse appearance domains and task types. *Nonl. Dynam.* 44, 9298–9314. doi: 10.1109/tpami.2021.3129870
- Miao, Y., Li, C., Li, Z., Yang, Y., and Yu, X. (2021). A novel algorithm of ship structure modeling and target identification based on point cloud for automation in bulk cargo terminals. *Measure. Control* 54, 155–163. doi: 10.1177/0020294021992804
- Mishra, A. K., Choudhry, M. S., and Kumar, M. (2022). Underwater image enhancement using multiscale decomposition and gamma correction. *Multimed. Tools Appl.* 82, 15715–15733. doi: 10.1007/s11042-022-14008-2
- Palomeras, N., Furfaro, T., Williams, D. P., Carreras, M., and Dugelay, S. (2022). Automatic target recognition for mine countermeasure missions using forward-looking sonar data. *IEEE J. Ocean. Eng.* 47, 141–161. doi: 10.1109/JOE.2021.3103269
- Pato, L., Negrinho, R., and Aguiar, M. P. (2020). “Seeing without looking: Contextual rescoring of object detections for ap maximization,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 14610–14618 (New York).
- Scouling, B., Maguire, K., and Orenstein, E. C. (2022). Evaluating automated benthic fish detection under variable conditions. *ICES J. Mar. Sci.* 79, 2204–2216. doi: 10.1093/icesjms/fsac166
- Song, P., Li, P., Dai, L., Wang, T., and Chen, Z. (2023). Boosting R-CNN: Reweighting R-CNN samples by RPN’s error for underwater object detection. *Neurocomputing* 530, 150–164. doi: 10.1016/j.neucom.2023.01.088
- Sun, W., Kong, X., and Zhang, Y. (2022). Attention-guided video super-resolution with recurrent 646 multi-scale spatial-temporal transformer. *Complex Intell. Syst.* 9, 3989–4002. doi: 10.1007/s40747-022-00944-x
- Wang, C.-Y., Alexey, B., and Liao, H.-Y. M. (2023a). “Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver (CVPR). 7464–7475. doi: 10.1109/CVPR52729.2023.00721
- Wang, S., Li, M., Tian, Y., Ai, X., Liu, J., Andriulli, F. P., et al. (2022). Cone-shaped space target inertia characteristics identification by deep learning with compressed dataset. *IEEE Trans. Antennas Propagation* 70, 5217–5226. doi: 10.1109/TAP.2022.3172759
- Wang, H., Sun, S., Bai, X., Wang, J., and Ren, P. (2023b). A reinforcement learning paradigm of configuring visual enhancement for object detection in underwater scenes. *IEEE J. Ocean. Eng.* 48, 443–461. doi: 10.1109/JOE.2022.3226202
- Wang, H., Sun, S., and Ren, P. (2023c). Meta underwater camera: a smart protocol for underwater image enhancement. *ISPRS J. Photogram. Remote Sens.* 195, 462–481. doi: 10.1016/j.isprsjprs.2022.12.007
- Wang, H., Sun, S., and Ren, P. (2023d). Underwater color disparities: Cues for enhancing underwater images toward natural color consistencies. *IEEE Trans. Circuits Syst. Vid. Technol.* doi: 10.1109/TCSVT.2023.3289566
- Xu, F., Wang, H., Peng, J., and Fu, X. (2020). Scale-aware feature pyramid architecture for marine object detection. *Neural Comput. Appl.* 33, 3637–3653. doi: 10.1007/s00521-020-05217-7
- Xu, P., Zheng, J., Wang, X., Wang, S., Liu, J., Liu, X., et al. (2022). Design and implementation of lightweight auv with multisensor aided for underwater intervention tasks. *IEEE Trans. Circuits Syst. II-express Briefs* 69, 5009–5013. doi: 10.1109/TCSIL.2022.3193300
- Yamada, T., Massot-Campos, M., Prugel-Bennett, A., Williams, Pizarro, O., and Thornton, B. (2021). Leveraging metadata in representation learning with georeferenced seafloor imagery. *IEEE robot. auto. Lett.* 6, 7815–7822. doi: 10.1109/LRA.2021.3101881
- Yang, P. (2023). An imaging algorithm for high-resolution imaging sonar system. *Multimed. Tools Appl.* 1–17. doi: 10.1007/s11042-023-16757-0
- Ye, H., Zhou, R., Wang, J., and Huang, Z. (2022). Fmam-net: Fusion multi-scale attention mechanism network for building segmentation in remote sensing images. *IEEE Access* 10, 134241–134251. doi: 10.1109/ACCESS.2022.3231362
- Yu, M., Zhang, W., Chen, X., Xu, H., and Liu, Y. (2022). Sca-net: a multiscale building segmentation network incorporating a dual-attention mechanism. *IEEE Access* 10, 79890–79903. doi: 10.1109/ACCESS.2022.3194919
- Zhang, W., and Dong, L. (2022). Retinex-inspired color correction and detail preserved fusion for underwater image enhancement. *Comput. Electron. Agric.* 192, 106585–106585. doi: 10.1016/j.compag.2021.106585
- Zhang, W., Jin, S., Zhuang, P., Liang, Z., and Li, C. (2023a). Underwater image enhancement via piecewise color correction and dual prior optimized contrast enhancement. *IEEE Signal Process. Lett.* 30, 229–233. doi: 10.1109/LSP.2023.3255005
- Zhang, W., Li, Z., Li, G., Zhuang, P., Hou, G., Zhang, Q., et al. (2023). GACNet: generate adversarial-driven cross-aware network for hyperspectral wheat variety identification. *IEEE Trans. Geosci. Remote Sens.* doi: 10.1109/TGRS.2023.3347745
- Zhang, X., Yang, P., and Sun, H. (2023c). An omega-k algorithm for multireceiver synthetic aperture sonar. *Electron. Lett.* 59. doi: 10.1049/ell2.12859
- Zhang, W., Zhao, W., Li, J., Zhuang, P., Sun, H., Xu, Y., et al. (2024). CVANet: Cascaded visual attention network for single image super-resolution. *Neural Networks* 170, 622–634. doi: 10.1016/j.neunet.2023.11.049
- Zhang, W., Zhou, L., Zhuang, P., Li, G., Pan, X., Zhao, W., et al. (2023b). Underwater image enhancement via weighted wavelet visual perception fusion. *IEEE Trans. Circuits Syst. Vid. Technol.* 1–1. doi: 10.1109/TCSVT.2023.3299314
- Zhang, W., Zhuang, P., Sun, H.-H., Li, G., Kwong, S., and Li, C. (2022). Underwater image enhancement via minimal color loss and locally adaptive contrast enhancement. *IEEE Trans. Image Process.* 31, 3997–4010. doi: 10.1109/TIP.2022.3177129
- Zhao, S., Li, B., Xu, P., Yue, X., Ding, G., and Keutzer, K. (2021). Madan: Multi-source adversarial domain aggregation network for domain adaptation. *Int. J. Comput. Vision* 129, 2399–2424. doi: 10.1007/s11263-021-01479-3
- Zhou, J., Sun, J., Zhang, W., and Lin, Z. (2023). Multi-view underwater image enhancement method via embedded fusion mechanism. *Eng. Appl. Artif. Intell.* 121. doi: 10.1016/j.engappai.2023.105946
- Zhou, J., Yang, T., Chu, W., and Zhang, W. (2022). Underwater image restoration via backscatter pixel prior and color compensation. *Eng. Appl. Artif. Intell.* 111. doi: 10.1016/j.engappai.2022.104785