Hindawi
International Journal of Distributed Sensor Networks
Volume 2023, Article ID 5337454, 11 pages
https://doi.org/10.1155/2023/5337454

Research Article

Underwater Incomplete Target Recognition Network via Generating Feature Module

Qi Shen,[1] **Jishen Jia**,[1,2] **and Lei Cai**[3]

[1]*School of Mathematical Sciences, Henan Institute of Science and Technology, Xinxiang 453003, China*
[2]*Henan Digital Agriculture Engineering Technology Research Center, Xinxiang 453003, China*
[3]*School of Artificial Intelligence, Henan Institute of Science and Technology, Xinxiang 453003, China*

Correspondence should be addressed to Lei Cai; cailei2014@126.com

Received 9 November 2022; Revised 29 December 2022; Accepted 5 January 2023; Published 12 January 2023

Academic Editor: Salvatore Serrano

A complex and changeable underwater archaeological environment leads to the lack of target features in the collected images, affecting the accuracy of target detection. Meanwhile, the difficulty in obtaining underwater archaeological images leads to less training data, resulting in poor generalization performance of the recognition algorithm. For these practical issues, we propose an underwater incomplete target recognition network via generating feature module (UITRNet). Specifically, for targets that lack features, features are generated by dual discriminators and generators to improve target detection accuracy. Then, multilayer features are fused to extract regions of interest. Finally, supervised contrastive learning is introduced into few-shot learning to improve the intraclass similarity and interclass distance of the target and enhance the generalization of the algorithm. The UIFI dataset is produced to verify the effectiveness of the algorithm in this paper. The experimental results show that the mean average precision (mAP) of our algorithm was improved by 0.86% and 1.29% under insufficient light and semiburied interference, respectively. The mAP for ship identification reached the highest level under all four sets of experiments.

1. Introduction

Underwater cultural heritage is a kind of nonrenewable cultural resource. In recent years, the substantial and high-intensity development in coastal areas has seriously threatened the safety of underwater cultural heritage, making the situation of underwater cultural heritage protection increasingly serious. The current stage of underwater archaeology requires high physical and professional skills of the staff, and the underwater scenarios are complex and changeable, posing significant safety risks. Therefore, the use of the autonomous underwater vehicle (AUV) for underwater archaeology can effectively reduce the risk of underwater archaeology.

However, in underwater archaeological operations, the target images collected by AUV have the problem of missing features due to the harsh underwater environment, such as insufficient underwater light, targets are mostly buried in mud and sand, and relics are corroded into wreckage for a long time, which leads to low recognition accuracy. Generative Adversarial Networks (GAN) can generate features through discriminators and generators, which can effectively improve the accuracy of underwater target recognition when all features of the image cannot be extracted. However, since underwater images are generally blurred, classical GAN generates a lot of noise while generating features and requires a large number of iterations resulting in slow convergence. In addition, a large number of labeled samples are required in the target recognition algorithm to effectively improve the accuracy, and most algorithms learn from labeled training sets, focusing on the recognition of labeled samples that have already appeared in the training. However, in practical applications, the difficulty in obtaining underwater target samples results in the lack of a large number of samples to train the network, and the large differences between individual relics make the algorithm's generalization performance poor. Few-shot learning does not rely on large-scale training samples and can achieve low-cost and fast target recognition for an emerging task with few collectible samples. Its application effectively improves the generalization performance

of target recognition algorithms, but the algorithms usually cause severe overfitting.

In response to the above specific questions, we propose an underwater incomplete target recognition network via generating features module in this paper. The overview of our algorithm is shown in Figure 1.

The main contributions of this paper are as follows:

(1) Dual discriminators and generators are introduced to generate missing features in two submodules. The generated features retain semantic information while reducing noise generated by the generator and reduce the number of iterations, thus improving the accuracy of the algorithm

(2) Supervised contrastive learning is applied to few-shot learning for target detection using contrastive proposal encoding. The intraclass similarity and interclass variance of targets are improved by cpe loss. This module improves the generalization performance of the algorithm recognition

(3) The proposed algorithm was evaluated on a dataset UIFI with disturbances such as insufficient lighting, partially buried targets, and wreckage. The superior performance of the algorithm in this paper is verified by comparing it with state-of-the-art algorithms

The rest of this article is arranged as follows. The related work is discussed in detail in Section 2. Section 3 presents the feature generation model, the few-shot learning network model, and the training process of the proposed algorithm in three subsections. In Section 4, simulation experiments are conducted to verify the effectiveness of the proposed algorithm. Section 5 concludes the paper.

2. Related Work

In recent years, object recognition has been widely used in many fields. Nevertheless, incomplete features of the target images collected by AUVs in underwater archaeological target detection lead to difficulties and low accuracy in recognition. In terms of feature-missing image reconstruction, some scholars have conducted in-depth research. Wang et al. [1] proposed a DPNet dual-pyramid reconstruction framework to learn more different scale features and further proposed a pyramid attention mechanism (PAM) in the decoder to obtain finer patches directly from the learning layer. Some scholars also perform a phased reconstruction for the missing features' objectives [2, 3], achieving global rough results first, followed by local refinement. Attention is also paid to the texture information of the image [4, 5], which guides the reconstruction of the image by generating the texture of the image. Cai et al. [6] innovatively proposed a framework for transfer reinforcement learning for the reconstruction of multiview optical fields. Niu et al. [7] proposed a defect image generation method with controllable defect area and intensity. The generated defect area was controlled by using a defect mask.

The small number of underwater image samples leads to poor generalization performance of the recognition algorithm. Great progress has also been made in the field of few-shot object detection. Meta-learning [8–10] can learn classes that have never been trained, and introducing meta-learning into the framework can effectively improve the performance of small-sample recognition algorithms. Lu et al. [11] designed Decouple Representation Transformation (DRT) and image-level distance metric learning to eliminate the adverse effects of manually annotated prior knowledge by predicting the object and anchoring shape. Kim et al. [12] inferred the geometric correlation between the new category and the basic region of interest. Zhang et al. [13] proposed a Joint Adaptive Detection Framework (JADF), which matches marginal and conditional distributions between domains without introducing any additional hyperparameters. Kaul et al. [14] obtained high-quality pseudoannotations for each new category from the training set and removed candidate detections with incorrect class labels by introducing a validation technique. Hu et al. [15] stated that context-aware polymerization (DCNet) intensive relation extraction is features to capture the object using the annotation of new characteristics of fine-grained.

For underwater image target recognition, there are still huge challenges. The performance of the identification algorithm is low due to a variety of disturbances caused by an underwater complex environment. Ref. [16] can effectively improve underwater target identification performance by extracting salient features and spatial semantic information of targets and then fusing them. Cai et al. [17] improved the accuracy of target detection under glass interference by minimizing the abstract feature distance between the source and target domains. Cai et al. [18] proposed a framework based on transfer reinforcement learning that can improve the accuracy of cooperative multi-AUV target recognition. Wang et al. [19] treat deep CNN as different views to extract semantic representations of images, and visual and semantic representations of images are used to predict the categories of images. Ref. [20] innovatively proposed a fusion framework (SSFNet), which effectively mitigates the gap between features by means of a semantic modulation model and a resolution-aware model. There are still many deep learning-based models that are applied for different tasks. Ref. [21] comprehensively investigates microorganism biovolume algorithms and classifies them according to digital image analysis methods. For the problem of parameter explosion, Ref. [22] proposed low-cost U-Net, which significantly reduces the high memory cost of U-Net. The proposed algorithm in Ref. [23] is divided into two stages, which significantly improves the performance of the algorithm in colorectal histopathology image classification.

In summary, there are some studies in target detection and few-shot learning. However, there are few studies on target detection in the case of images with missing features and few training samples at the same time. The proposed method in this paper effectively solves this practical problem.

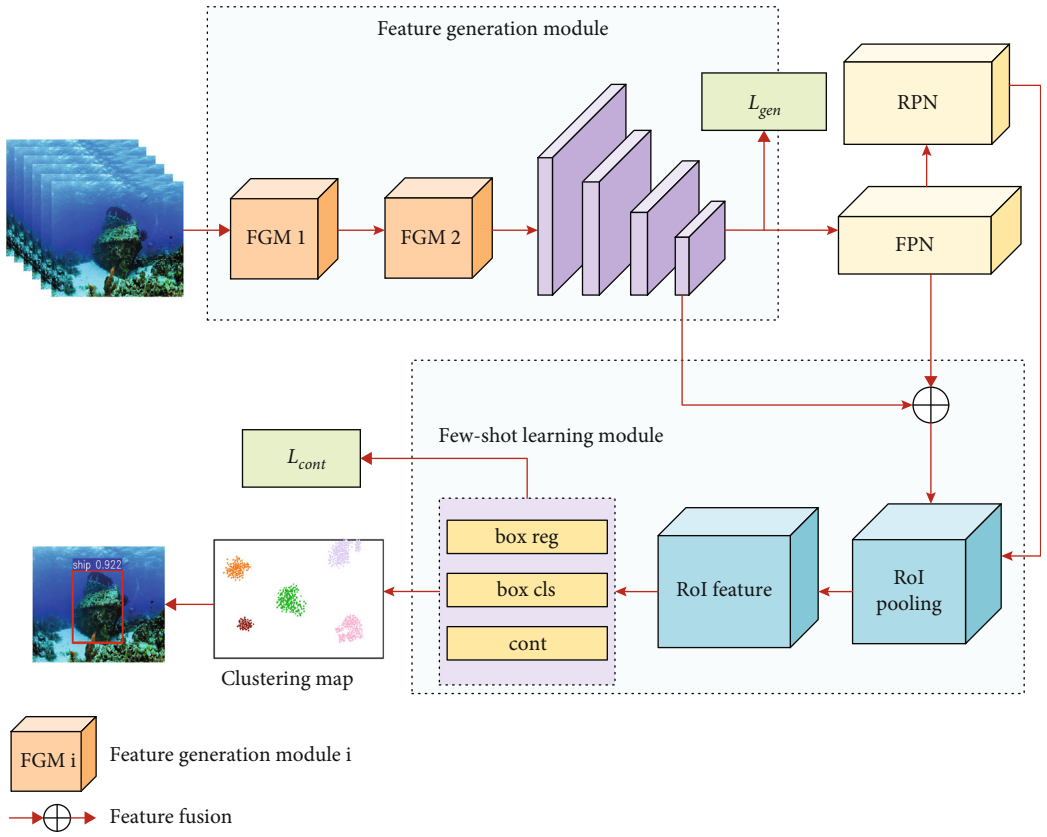

Figure 1: An overview of our proposed underwater incomplete target recognition network.

3. Proposed Method

In this section, for targets with missing features, we generate the missing features through the feature generation module (FGM), which consists of two submodules containing dual generators and discriminators, and the features of underwater archaeological target images are extracted by using RepVGG network. After FGM, the accuracy of target recognition can be significantly improved. The generalization ability of the algorithm for target recognition is improved by introducing contrastive learning into our algorithm framework. By applying the algorithm of this paper, the issue of low target identification accuracy under the interference of insufficient underwater light, target buried in mud and sand, and target wreckage has been effectively solved.

3.1. Underwater Image Missing Feature Generation Module. In the process of target recognition in underwater archaeology, the AUV usually fails to extract all the features from the collected underwater target images. This section utilizes dual discriminators and generators to generate the unextracted features by two generation submodules. This reduces the impact of missing features on underwater recognition algorithms. The feature generation module is shown in Figure 2.

However, classical GANs generate a lot of noise and require a large number of iterations. The feature generation model proposed in this section is divided into two submodules, submodule 1 for generating features while preserving the semantic information of the image and submodule 2 for noise reduction.

In real underwater archaeological scenes, the images taken by AUVs usually have some interference factors, resulting in low algorithm recognition efficiency, such as insufficient light, partial burial of target objects, and antique wreckage. In submodule 1, x_s represents the real image, and x_t represents the image with missing features. The generator G_1 is a deep network, and $G_1(x)$ represents the generation function of the input x. Take x_t as input, generate an intermediate image x_g with complete features by G_1 and then input x_g to the discriminator D_1. D_1 denotes a network of discriminators, and $D_1(x, t)$ denotes the discriminator function with input x and target label t . For discriminator D_1, we set x_s to 1 and x_t to 0 and use the complete real image x_r and the generated intermediate image x_g as inputs to discriminator D_1.

D_1 loss function is based on binary crossentropy loss, and the adversarial loss function of D_1 can be expressed as follows:

$$l_{\text{adv}}^{D_1}(x_s, x_t) = l_b(D_1(G_1(x_t), t), 0) + l_b(D_1(x_s, t), 1). \quad (1)$$

The adversarial loss function corresponding to G_1 can be written as follows:

$$l_{\text{adv}}^{G_1}(x_t, x_s) = l_{st}(x_s, G_1(x_t)) + l_b(D_1(G_1(x_t), t), 1), \quad (2)$$

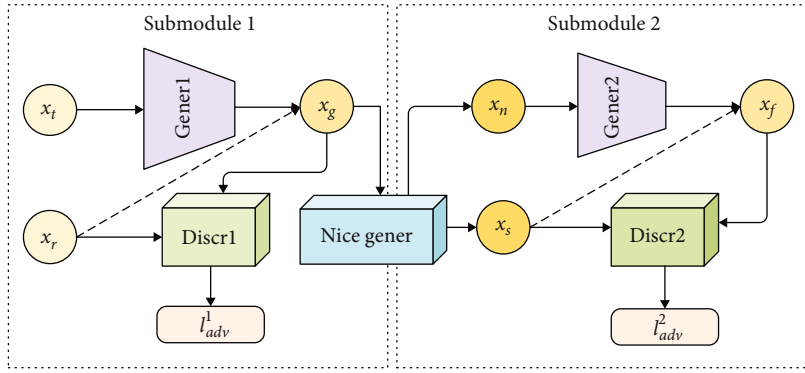

Figure 2: Feature generation module.

where l_b is the binary crossentropy loss function and l_{st} is the structural loss. The Nice Gener module acts as a link between the two submodules, it takes x_g as input, and set the loss function threshold to 0.01. It generates high-quality images x_n as input to submodule 2. In submodule 2, the image generated by submodule 1 is further denoised, and the generator G_2 is used as a denoising autoencoder. The adversarial loss function of D_2 is similar to D_1 and can be expressed as follows:

$$l_{\text{adv}}^{D_2}(x_s, x_n) = l_b(D_2(G_2(x_t), t), 0) + l_b(D_2(x_s, t), 1). \quad (3)$$

The main role of adversarial loss at this stage involves narrowing the gap between the distribution of the generation and true features. Similar to submodule 1, G_2 corresponding adversarial loss function is given by the following:

$$l_{\text{adv}}^{G_2}(x_n, x_s) = l_a(x_s, G_1(x_t)) + l_b(D_2(G_2(x_t), t), 1), \quad (4)$$

where $l_a = \triangle(l_{st}(x_s, G_1(x_t)), l_{st}(x_s, G_2(x_n)))$, and $\triangle$ is the differential operator. Submodules 1 and 2 can be trained independently, resulting in shorter training durations. Minimizing $l_{\text{adv}}^{G_2}$ means reducing the gap between the generated samples and the real samples and also reducing the noise in the generated samples. In the feature extraction process for images, the RepVGG-B2 network is used as the backbone.

Therefore, the final trained loss function of the feature generation module can be expressed by Equation (5), where λ_{adv} denotes the weight of adversarial loss in the loss function.:

$$\mathscr{L}_{\text{gen}} = l_{\text{adv}}^{D_1} + l_{\text{adv}}^{G_1} + \lambda_{\text{adv}} l_{\text{adv}}^{D_2} + \lambda_{\text{adv}} l_{\text{adv}}^{G_2}. \quad (5)$$

3.2. Few-Shot Learning Module. In the process of underwater target recognition, a large amount of data is usually needed to train the network. However, the small number of underwater image samples leads to the poor generalization performance of the algorithm. To address this problem, this section introduces contrastive proposal coding, where we perform few-shot target detection by supervised contrastive learning. The intraclass similarity and interclass distinction are increased to reduce the problem of low recognition accuracy of unknown images underwater due to small training samples, thus improving the generalization performance of the network. The flow of this module is shown in Figure 3.

In this module, we take the feature map of the backbone as input to the region proposal network (RPN) and generate region proposals. Then, each region proposal is classified by RoI head. In the classification results, the bounding box is returned through the loss function if the target is included. In the RoI head, the region of interest is first pooled to a fixed size by a feature extractor. Immediately afterwards, the features are encoded as RoI feature s_i. To obtain more significant target feature representations from a very small number of samples, this paper applies to batch contrastive learning to improve the intraclass similarity and interclass differentiation in the target suggested region.

This article introduces batch contrastive learning into our framework, adding a contrast branch to the RoI head. Then, the similarity between the targets is calculated on the RoI features and increase the intraclass similarity and interclass distinction. We use a bounding box classifier based on cosine similarity which denoted as Equation (6), where the sim is the scaled cosine similarity between RoI features s_i and category weights τ_j. By calculating sim, we can predict the i-th instance to be the class j, further improve the similarity of the same category, and expand the distinction between different categories.

$$\text{sim}_{\{i,j\}} = \delta \frac{s_i^T \tau_j}{\|s_i\| \cdot \|\tau_i\|}, \quad (6)$$

where δ is the scaling factor of the amplification gradient, which is usually set to $\delta = 20$. The contrastive learning head simplifies the distinction between different categories by learning the objects of contrast perception through RoI head. The embedding of contrastive learning makes the same class similarity higher in the classification task and the greater distance between the extended different classes. Therefore, the generalization performance of the algorithm is strengthened.

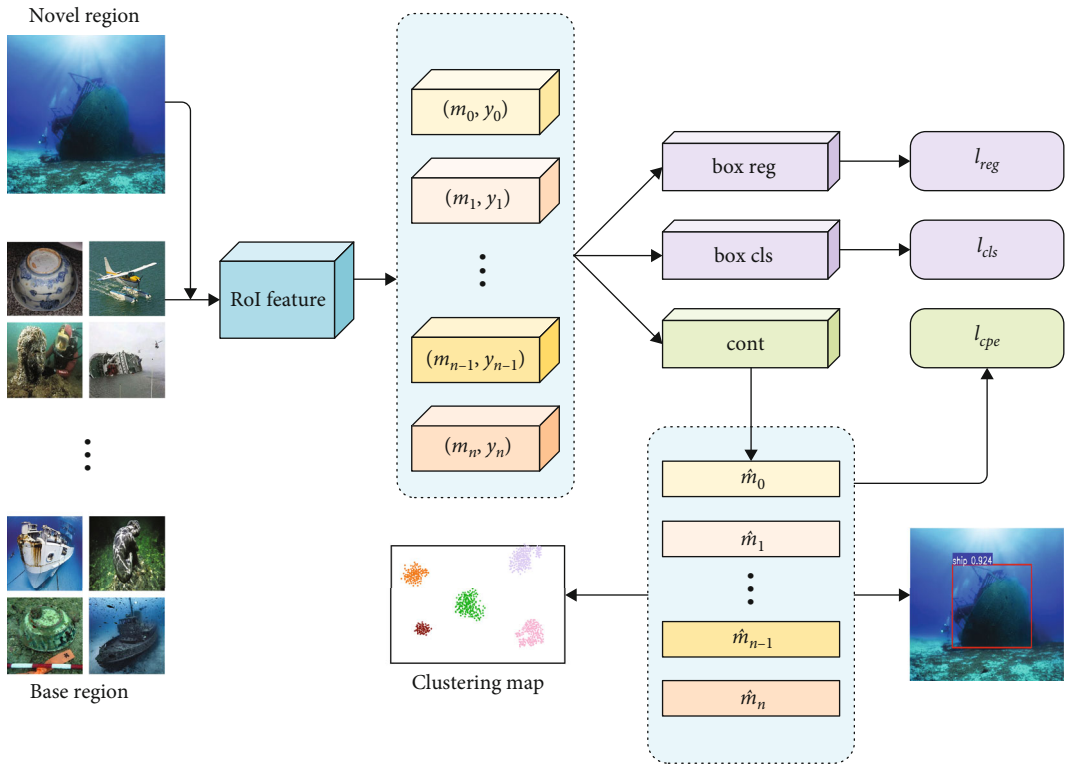

Figure 3: Few-shot learning module.

By introducing supervised comparative learning into the detect task, we can take advantage of the following cpe loss (Equation (7)). Specifically, in a small batch of N RoI head features $\{f_i, u_i, y_i\}_i^N$, we define f_i to be the i-th RoI feature, u_i represents the IoU of the bounding box to the ground truth, and y_i represents the ground truth.

$$l_{cpe} = \frac{1}{N}\sum_{i=1}^{N} g(u_i) \bullet E_{f_i}, \tag{7}$$

$$E_{f_i} = \frac{-1}{N_{y_i} - 1}\sum_{j=1, j\neq i}^{N} I\{y_i = y_j\} \bullet \log \frac{\exp\left(f_i \bullet \left(f_j/\sigma\right)\right)}{\sum_{k=1}^{N} I_{k\neq i} \bullet \exp\left(f_i \bullet (f_k/\sigma)\right)}, \tag{8}$$

where N_{y_i} represents the number of samples with the same label as y_i, σ is the hyperparameter, $g(u_i)$ controls the consistency of the proposals, $g(u_i) = I\{u_i \geq \omega\} \bullet k(u_i)$, and $k(\bullet)$ is the weight of the corresponding IoU score, setting the threshold ω to 0.7. Embedding the contrast learning head into the network, the loss function of the few-shot learning module can be expressed as

$$\mathscr{L}_{\text{cont}} = l_{\text{rpn}} + l_{\text{cls}} + l_{\text{reg}} + \lambda_{\text{cpe}} l_{\text{cpe}}, \tag{9}$$

where λ_{cpe} is the corresponding weight (usually set to 0.5), l_{cls} denotes the loss function of the bounding box classifier, l_{rpn} is set to the binary crossentropy loss, and the loss function l_{reg} is used for bounding box regression.

3.3. UITRNet and Training Process. UITRNet can effectively solve the problem of missing target features and few training samples in underwater archaeological scenes. First, the input image is generated through the feature generation module for missing features. In the feature generation module, the missing features are generated by two submodules while reducing the generation noise. Then, the features of different layers are fused through feature pyramid network (FPN), and the region of interest (RoI) is extracted and input into the few-shot learning module. Finally, by introducing contrast learning, RoI features are added to the detector by contrasting branches to increase intraclass similarity and interclass gaps. The training process of the algorithm is as follows.

For the adversarial loss $\mathscr{L}_{\text{adv}}$ of the two submodules of the feature generation module, in submodule 1, by minimizing Equation (1) and Equation (2), generator 1 performs better than discriminator 1 in reaching Nash equilibrium, where discriminator 1 considers that the feature images generated by generator 1 obey the true distribution. In submodule 2, the structure and pixel values of the generated features and the real samples are made more similar by Equation (4), thus reducing the noise in the generated samples. For the few-shot learning module, the loss function is given by Equation (9).

In summary, the loss function of the proposed framework in this paper can be expressed as follows:

$$\mathscr{L} = \mathscr{L}_{\text{gen}} + \lambda \mathscr{L}_{\text{cont}}, \tag{10}$$

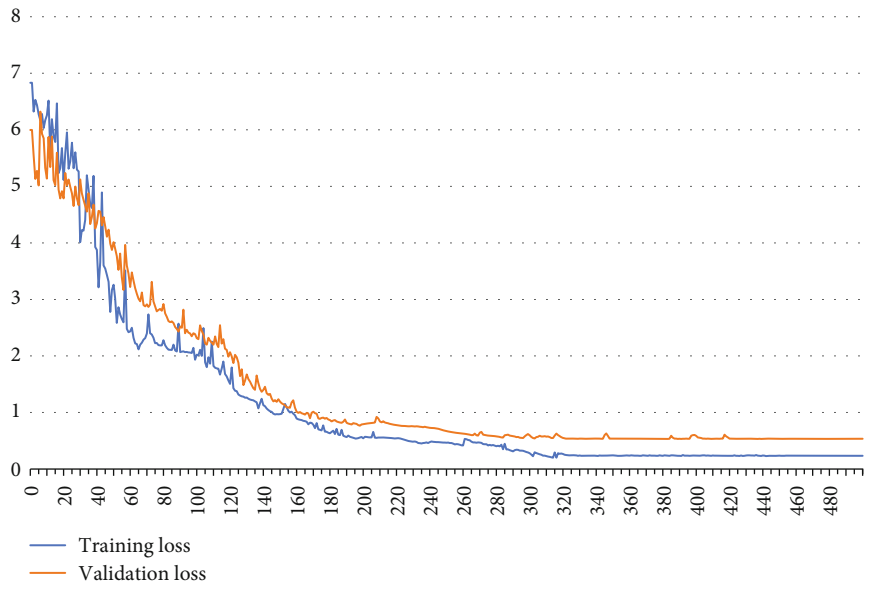

Figure 4: The loss curves of training and validation process.

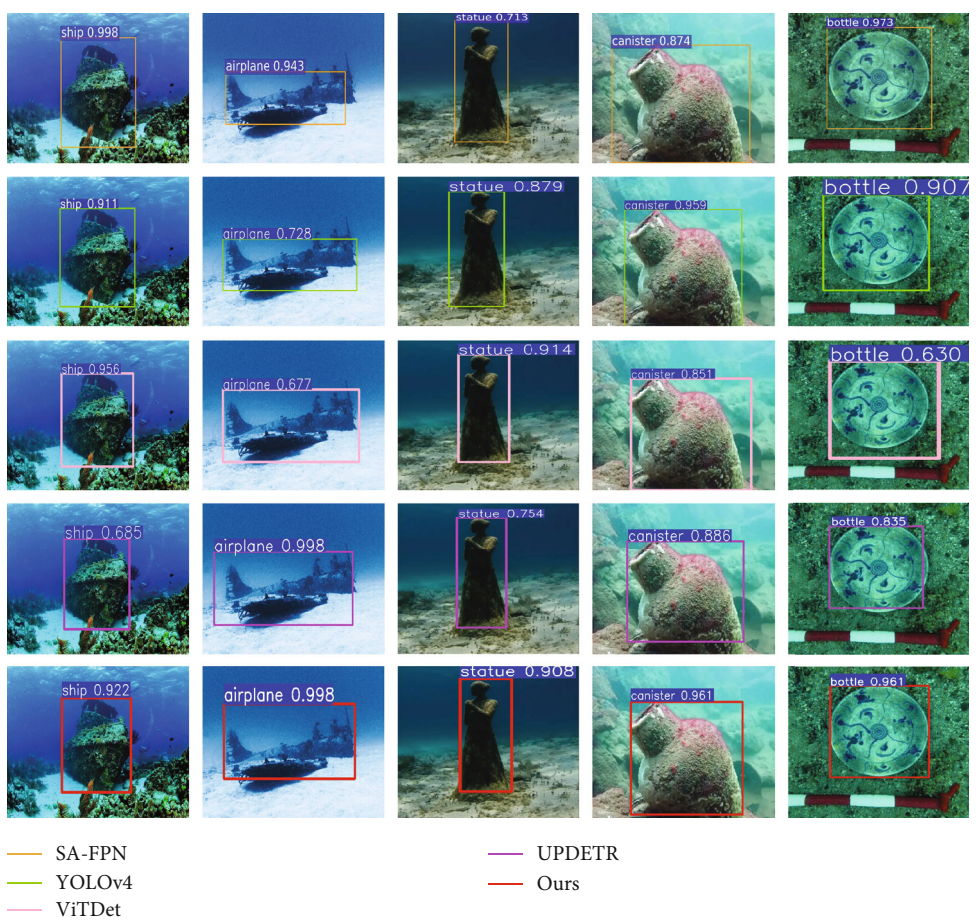

Figure 5: Comparison of recognition effects of advanced algorithms in the conventional underwater environment.

Table 1: Comparison of recognition results with advanced algorithms in the conventional underwater environment.

Method	Ship	Airplane	Statue	Canister	Bottle	mAP	Time
SA-FPN	0.8062	0.7528	0.6715	0.5512	**0.5695**	0.6702	0.201
YOLOv4	0.8344	0.7539	0.7483	0.5659	0.5184	0.6842	0.217
ViTDet	0.8631	**0.8973**	0.808	0.4862	0.5633	**0.7236**	0.326
UPDETR	0.8700	0.7055	0.5757	**0.5819**	0.5432	0.6553	**0.113**
Ours	**0.9032**	0.7592	**0.8156**	0.4681	0.5368	0.6966	0.229

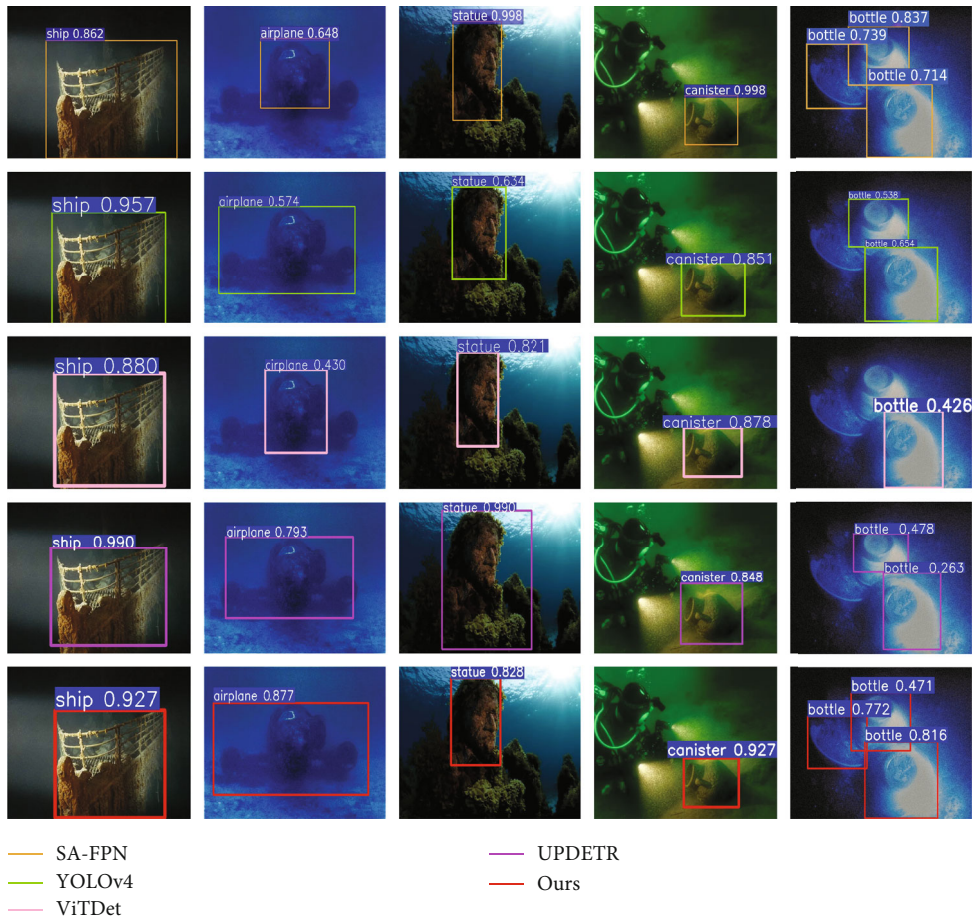

Figure 6: Comparison of recognition effects with advanced algorithms in the underwater insufficient light environment.

where the hyperparameter λ is used to balance the loss function, $\mathscr{L}_{\text{gen}}$ is the loss of the generation module mentioned above, and $\mathscr{L}_{\text{cont}}$ is the loss of the few-shot learning module.

4. Experiment

4.1. Dataset. This paper uses the homemade dataset UIFI (underwater incomplete feature images), the dataset includes 1112 images with targets, and the corresponding labels are given. The dataset is divided into 0.7 : 0.3 training and test sets. The dataset includes a variety of interfering factors, such as insufficient light, semiburial, and wreckage, which can be used to evaluate the performance of underwater missing features object detection algorithms.

4.2. Implementation Details. This article sets the batch size to 16 to train the model. Optimization is performed using stochastic gradient descent (SGD) with momentum set to 0.9, weight decay to 0.0005, initial learning rate set to 0.01, and 500 epochs throughout the training process. The number of trainable parameters for the proposed algorithm is 145,439,611, and the memory cost is 563 MB. We provide the loss curves of training and validation in Figure 4. In the algorithm operating environment, CPU is Intel(R) CORE i5 7200U, and GPU is RTX 3090 VENTUS 3X 24G.

4.3. Analysis of Results. In this section, simulation experiments are performed on the UIFI dataset. For the three different disturbances mentioned above, four sets of experiments containing different disturbances are designed

Table 2: Comparison of recognition results with advanced algorithms in the underwater insufficient light environment.

Method	Ship	Airplane	Statue	Canister	Bottle	mAP	Time
SA-FPN	0.6324	0.5812	0.5168	0.4337	**0.5093**	0.5347	0.209
YOLOv4	0.6864	0.6162	**0.5943**	0.4212	0.4489	0.5534	0.221
ViTDet	0.6327	0.6521	0.5916	0.4128	0.4849	0.5548	0.321
UPDETR	0.7013	**0.6962**	0.4571	**0.4419**	0.5027	0.5598	**0.112**
Ours	**0.7341**	0.6032	0.5857	0.4332	0.4859	**0.5684**	0.227

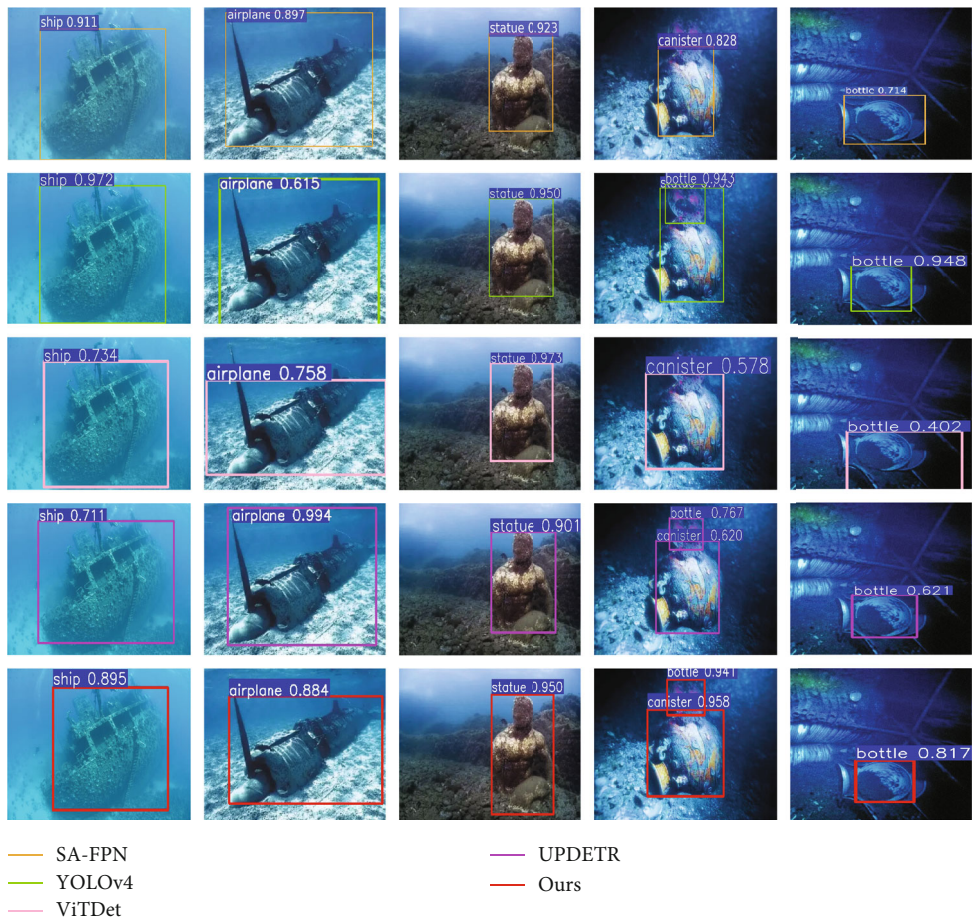

Figure 7: Comparison of recognition effects with advanced algorithms in underwater targets partially buried.

to verify the effectiveness of our algorithm. The evaluation indicators of the algorithm are mAP and time. First, we compare the methods proposed in this paper with advanced methods, such as SA-FPN [24], YOLOv4 [25], ViTDet [26], and UPDETR [27]. In each subsection, we analyze the experimental results, and the conclusions of the experiments provide a clear picture of the efficiency of the algorithm in this paper.

4.3.1. Results of Conventional Underwater Image Recognition. For the problem of target recognition in conventional underwater environment, this section compares the advanced algorithms SA-FPN, YOLOv4, ViTDet, and UPDETR with the algorithms in this paper. In Figure 5, we show some of the effect figure of different recognition algorithms. The mAP and recognition time of the algorithm in this paper and the advanced algorithm can be obtained from Table 1, where the black bold font is the best data. The comparison results can be seen visually from Table 1, and the mAP of the ViTDet in a conventional water environment is up to 0.7236, but its recognition speed is 0.326. The mAP of our algorithm is 2.7% lower than that of ViTDet, but the recognition speed in this article is 0.229, which is higher than ViTDet. The fastest recognition speed of UPDETR is 0.113, but compared with our algorithm, the recognition accuracy of this paper is improved by 4.13%. This paper has an excellent performance in the mAP of ship and stone statue.

4.3.2. Results of Underwater Insufficient Light Image Recognition. For the problem of target recognition in the underwater insufficient light environment, this section

Table 3: Comparison of recognition results with advanced algorithms in underwater targets partially buried.

Method	Ship	Airplane	Statue	Canister	Bottle	mAP	Time
SA-FPN	0.6846	0.5938	0.5332	0.4341	**0.4987**	0.5489	0.202
YOLOv4	0.6382	0.5504	0.6003	**0.4464**	0.4567	0.5384	0.218
ViTDet	0.6572	**0.7036**	**0.6291**	0.4028	0.4939	0.5773	0.326
UPDETR	0.6813	0.5892	0.4673	0.4026	0.4969	0.5275	**0.114**
Ours	**0.7278**	0.6987	0.6029	0.4339	0.4876	**0.5902**	0.229

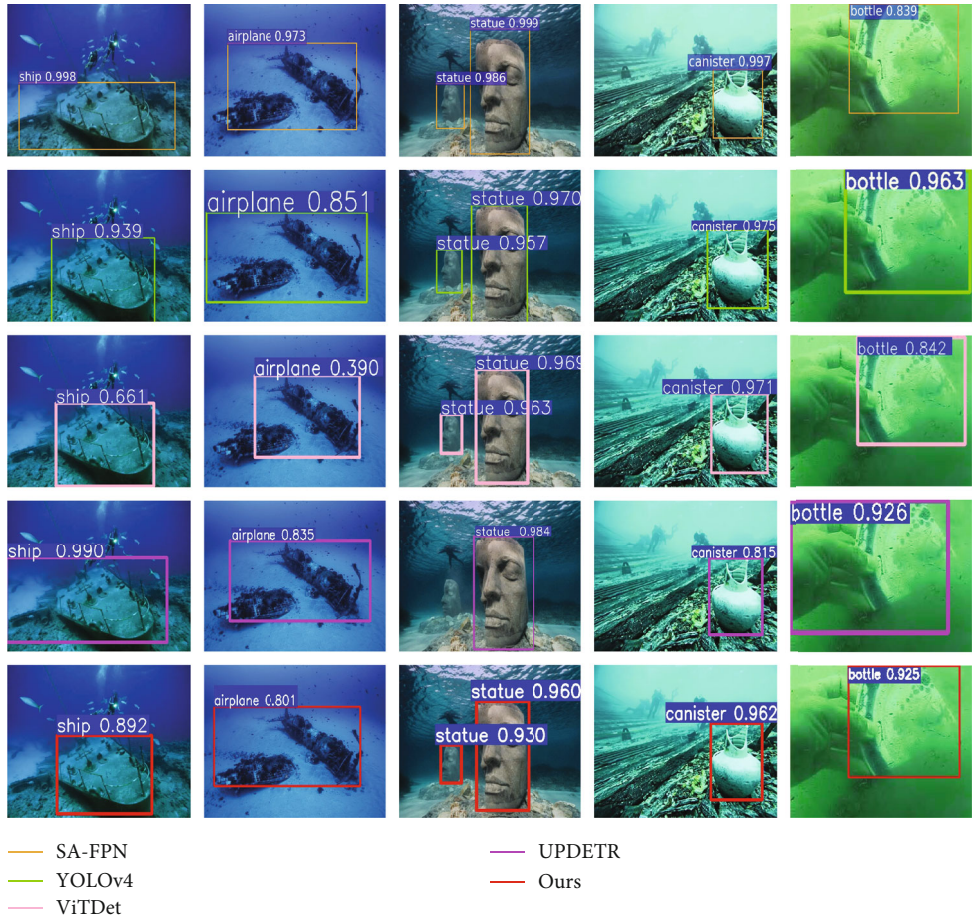

Figure 8: Comparison of recognition effects with advanced algorithms in underwater wreckage.

compares advanced algorithms such as SA-FPN, YOLOv4, ViTDet, and UPDETR with the algorithms in this paper. The effect of identification is shown in Figure 6. The mAP and recognition speed of each algorithm can be obtained from Table 2, and we bolded the best data. From Table 2, it can be seen that in a poorly lit underwater environment, the mAP of the proposed algorithm is up to 0.5684, and the recognition speed is 0.227. Compared with the excellent algorithm UPDETR, its recognition speed is 0.112, but compared with the method of this paper, this paper has improved the mAP by 0.86% over the UPDETR algorithm. This article achieves the highest level of accuracy for the identification of ship.

4.3.3. Results of Underwater Partially Buried Image Recognition. For the recognition problem of partially buried underwater targets, this section compares advanced algorithms such as SA-FPN, YOLOv4, ViTDet, and UPDETR with the algorithms we proposed. The identification results of different algorithms are presented in Figure 7. In Table 3, we summarize the mAP and identification time of our algorithm and advanced algorithms, where the black bolded font is the best data. From Table 3, we can see that the proposed algorithm has the highest mAP of 0.5902 and recognition speed of 0.229 under the partially buried condition of underwater targets. Compared to the advanced recognition algorithm ViTDet, the mAP is

Table 4: Comparison of recognition results with advanced algorithms in underwater wreckage.

Method	Ship	Airplane	Statue	Canister	Bottle	mAP	Time
SA-FPN	0.6582	0.6329	0.5452	0.4112	**0.4995**	0.5494	0.201
YOLOv4	0.6274	0.4146	0.4971	**0.4558**	0.4361	0.4862	0.217
ViTDet	0.6416	**0.6957**	**0.6243**	0.4165	0.4012	**0.5559**	0.331
UPDETR	0.6033	0.6046	0.4358	0.4516	0.4539	0.5098	**0.112**
Ours	**0.6597**	0.6195	0.5926	0.4029	0.4658	0.5481	0.228

improved by 1.29%. The fastest recognition speed of the UPDETR algorithm is 0.114, but compared with the method in this paper, this paper has higher mAP. For ship, our algorithm performs well in terms of recognition accuracy.

4.3.4. Results of Underwater Wreckage Image Recognition. For the target recognition problem of underwater wreckage, advanced algorithms such as SA-FPN, YOLOv4, ViTDet, and UPDETR are compared with the algorithms in this paper in this section. Figure 8 shows the recognition results. From Table 4, we can see the mAP and identifying time of the advanced algorithms, where the black bold font is the best data. Using Table 4, we can conclude that the ViTDet algorithm has a maximum mAP of 0.5559 for the wreckage of underwater targets, but its recognition speed is 0.331. The mAP of the proposed algorithm is 0.78% lower than that of ViTDet, but the speed of our algorithm is 0.229 higher than that of the ViTDet. The fastest identification speed of the UPDETR is 0.112, but compared with the method in this paper, the mAP of this paper is improved by 3.83%. This article has excellent performance in the mAP of the ship.

5. Conclusions

In a real underwater archaeological scene, AUV works under various disturbances causing difficulty in extracting the full features of the target. This paper proposes UITRNet, which can compensate for missing features in underwater images by generating features. In this paper, the algorithm is simulated on the UIFI of the self-made dataset, considering the detection under the conditions of conventional underwater images, insufficient light, buried targets, and wreckage. The mAP of the proposed algorithm in this paper is 56.84% under the interference of insufficient light, which is 0.86% better than the advanced algorithm UPDETR, and 59.02% under the interference of buried targets, which is 1.29% better than the advanced algorithm ViTDet. For target (ship) recognition with insufficient training data, the mAP is higher than the advanced algorithms SA-FPN, YOLOv4, ViTDet, and UPDETR under four different disturbances. The above experimental data show that our algorithm has excellent detection ability and position labeling ability in target recognition of missing feature images.

However, the performance of the algorithm needs to be improved in the situation of wreckage images. In addition, artifacts for the extracted features also affect the accuracy of the algorithm, and there is a need to continue to improve the algorithm to achieve better performance in the case of different image types.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

On behalf of my co-authors, the authors declare that there is no conflict of interests regarding the publication of this article.

Acknowledgments

This work was supported by the Science and Technology Project of Henan Province (222102110194, 222102320380).

References

[1] C. Wang, M. Shao, D. Meng, and W. Zuo, "Dual-Pyramidal image inpainting with dynamic normalization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 9, pp. 5975–5988, 2022.

[2] Z. Wan, J. Zhang, D. Chen, and J. Liao, "High-fidelity pluralistic image completion with transformers," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4692–4701, Montreal, Canada, 2021.

[3] W. Quan, R. Zhang, Y. Zhang, Z. Li, J. Wang, and D. M. Yan, "Image inpainting with local and global refinement," *IEEE Transactions on Image Processing*, vol. 31, no. 9, pp. 2405–2420, 2022.

[4] X. Guo, H. Yang, and D. Huang, "Image inpainting via conditional texture and structure dual generation," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 14134–14143, Montreal, Canada, 2021.

[5] R. Xu, M. Guo, J. Wang, X. Li, B. Zhou, and C. C. Loy, "Texture memory-augmented deep patch-based image inpainting," *IEEE Transactions on Image Processing*, vol. 30, pp. 9112–9124, 2021.

[6] L. Cai, P. Luo, G. Zhou, T. Xu, and Z. Chen, "Multiperspective light field reconstruction method via transfer reinforcement learning," *Computational Intelligence and Neuroscience*, vol. 2020, Article ID 8989752, 14 pages, 2020.

[7] S. Niu, B. Li, X. Wang, and Y. Peng, "Region-and strength-controllable GAN for defect generation and segmentation in industrial images," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 7, pp. 4531–4541, 2022.

[8] L. Liu, B. Wang, Z. Kuang et al., "Gendet: meta learning to generate detectors from few shots," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 8, pp. 3448–3460, 2021.

[9] H. Song, B. Deng, M. Pound, E. Özcan, and I. Triguero, "A fusion spatial attention approach for few-shot learning," *Information Fusion*, vol. 81, pp. 187–202, 2022.

[10] M. Cheng, H. Wang, and Y. Long, "Meta-learning-based incremental few-shot object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 4, pp. 2158–2169, 2022.

[11] Y. Lu, X. Chen, Z. Wu, and J. Yu, "Decoupled metric network for single-stage few-shot object detection," *IEEE Transactions on Cybernetics*, vol. 53, no. 1, pp. 514–525, 2023.

[12] G. Kim, G. H. Jung, and S. W. Lee, "Spatial reasoning for few-shot object detection," *Pattern Recognition*, vol. 120, article 108118, 2021.

[13] B. Zhang, T. Chen, B. Wang, and R. Li, "Joint distribution alignment via adversarial learning for domain adaptive object detection," *IEEE Transactions on Multimedia*, vol. 24, pp. 4102–4112, 2022.

[14] P. Kaul, W. Xie, and A. Zisserman, "Label, verify, correct: a simple few shot object detection method," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 14237–14247, New Orleans, LA, USA, 2022.

[15] H. Hu, S. Bai, A. Li, J. Cui, and L. Wang, "Dense relation distillation with context-aware aggregation for few-shot object detection," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10185–10194, Nashville, TN, USA, 2021.

[16] L. Cai, C. Chen, and H. Chai, "Underwater distortion target recognition network (UDTRNet) via enhanced image features," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 4193625, 10 pages, 2021.

[17] L. Cai, C. Chen, Q. Sun, and H. Chai, "Glass refraction distortion object detection via abstract features," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 5456818, 15 pages, 2022.

[18] L. Cai, Q. Sun, T. Xu, Y. Ma, and Z. Chen, "Multi-AUV collaborative target recognition based on transfer-reinforcement learning," *IEEE Access*, vol. 8, pp. 39273–39284, 2020.

[19] H. D. Wang, J. Li, and S. Zhu, "Few-labeled visual recognition for self-driving using multi-view visual- semantic representation," *Neurocomputing*, vol. 428, pp. 361–367, 2021.

[20] X. Zhang, Y. Chen, B. Zhu, J. Wang, and M. Tang, "Semantic-spatial fusion network for human parsing," *Neurocomputing*, vol. 402, pp. 375–383, 2020.

[21] J. Zhang, C. Li, M. M. Rahaman et al., "A Comprehensive survey with quantitative comparison of image analysis methods for microorganism biovolume measurements," 2022, https://arxiv.org/abs/2202.09020.

[22] J. Zhang, C. Li, S. Kosov et al., "LCU-Net: A novel low-cost U-Net for environmental microorganism image segmentation," *Pattern Recognition*, vol. 115, Article ID 107885, 2021.

[23] H. Chen, C. Li, X. Li et al., "IL-MCAM: An interactive learning and multi-channel attention mechanism-based weakly supervised colorectal histopathology image classification approach," *Computers in Biology and Medicine*, vol. 143, Article ID 105265, 2022.

[24] F. Xu, H. Wang, J. Peng, and X. Fu, "Scale-aware feature pyramid architecture for marine object detection," *Neural Computing and Applications*, vol. 33, no. 8, pp. 3637–3653, 2021.

[25] A. Bochkovskiy, Y. C. Wang, and M. Y. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020, https://arxiv.org/abs/2004.10934.

[26] Y. Li, H. Mao, R. Girshick, and K. He, "Exploring plain vision transformer backbones for object detection," 2022, https://arxiv.org/abs/2203.16527.

[27] Z. Dai, B. Cai, and Y. Lin, "Up-detr: unsupervised pre-training for object detection with transformers," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1601–1610, Nashville, TN, USA, 2021.