Robotica (2023), **41**, pp. 900–911
doi:10.1017/S0263574722001059

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

EHDC: enhanced dilated convolution framework for underwater blurred target recognition

Lei Cai[1,*], Xiaochen Qin[2] and Tao Xu[1]

[1]School of Artificial Intelligence, Henan Institute of Science and Technology, Xinxiang, China and [2]School of Information Engineering, Henan Institute of Science and Technology, Xinxiang, China
*Corresponding author. E-mail: cailei2014@126.com

Received: 13 May 2022; **Revised:** 18 June 2022; **Accepted:** 27 June 2022; **First published online:** 26 July 2022

Keywords: blurred small target, hybrid dilated convolution, spatial semantic features, low light conditions, target recognition

Abstract

The autonomous underwater vehicle (AUV) has a problem with feature loss when recognizing small targets underwater. At present, algorithms usually use multi-scale feature extraction to solve the problem, but this method increases the computational effort of the algorithm. In addition, low underwater light and turbid water result in incomplete information on target features. This paper proposes an enhanced dilated convolution framework (EHDC) for underwater blurred target recognition. Firstly, this paper extracts small target features through hybrid dilated convolution networks, increasing the perceptive field of the algorithm without increasing the computational power of the algorithm. Secondly, the proposed algorithm learns spatial semantic features through an adaptive correlation matrix and compensates for the missing features of the target. Finally, this paper fuses spatial semantic features and visual features for the recognition of small underwater blurred targets. Experiments show that the proposed method improves the recognition accuracy by 1.04% compared to existing methods when recognizing small underwater blurred targets.

1. Introduction

There are limited features available for small targets during underwater detection and recognition. The feature extraction network downsampling process suffers from the problem of disappearing feature gradients. This problem seriously affects the recognition accuracy of small targets underwater. Existing algorithms usually extract multi-scale features of the target to solve the problem of disappearing feature gradients. However, this approach significantly increases the computational effort of the algorithm [1]. By adding voids to the convolution kernel through dilated convolution, the resolution of the feature map can be increased without increasing the computational effort. However, extended convolution also suffers from the problem of "gridding," which can cause a partial loss of adjacent information [2].

The underwater environment is poorly lit and the water is murky. This reduces the clarity of the underwater image and prevents the autonomous underwater vehicle (AUV) from obtaining complete information about the target features in the acquired [3]. The graph convolutional neural network (CNN) can effectively learn the spatial semantic features of the targets by capturing the inter-target dependencies through information transfer between nodes. The lack of features in the target can be compensated by spatial semantic features, improving the recognition accuracy of underwater blurred images [4]. A suitable correlation matrix is the key to extracting spatial semantic features accurately. However, existing algorithms usually construct correlation matrices from the label co-occurrence relationships, which are weak in generalization.

To address the difficulties of underwater small target recognition, we propose an enhanced dilated convolution framework for underwater blurred target recognition, as shown in Fig. 1. The method enables the recognition of small targets in the presence of blurred images.

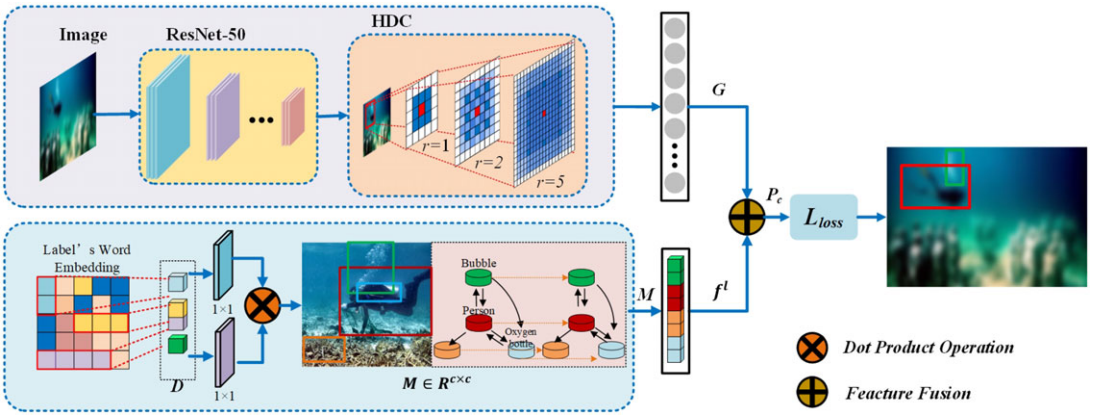

Figure 1. *Enhanced dilated convolution framework for underwater blurred target recognition.*

The main contributions of the methodology in this paper are as follows:

1. The proposed method improves the features extraction network. The introduction of hybrid dilated convolution with different expansion rates improves the resolution of small target features. The method effectively solves the problem of small target feature disappearance without increasing the computational effort.
2. The algorithm proposed in this paper constructs an adaptive correlation matrix through two 1×1 convolutional layers and a dot product operation and learns the spatial semantic relations of the targets through this matrix. The algorithm solves the problem of incomplete target feature information in underwater blurred images.
3. The proposed algorithm fuses the visual features and spatial semantic relations of the target and trains the network with a focal loss function. The algorithm effectively improves the recognition accuracy of small underwater blurred targets.

2. Related work

Rapid recognition of underwater targets is a key issue for autonomous AUV recognition. There is a tendency to miss detection during the recognition of small underwater targets. The paper [5] presents a new framework for underwater image saliency detection. The algorithm combines both a quaternion number system and principal components analysis to achieve superior performance. Kong et al. [6] proposed an efficient feature extraction method, which effectively improves the real-time performance of the algorithm. For small targets in complex environments that are easily masked by other objects or noise, Wu et al. [7] proposed an open-closed transformation algorithm to eliminate or weaken the background and noise. This algorithm extracts the weakened features by eliminating noise to achieve the recognition of small targets, which effectively improves the recognition efficiency of small targets. This paper presents a three-stage FCA algorithm for HR. It is used to extract face features [8]. In terms of CNNs, Li et al. [9] extracted features from high-resolution range profile (HRRP) and classified targets to achieve the detection of small targets. This target recognition has good generalization capability and stable performance. Cao et al. [10] proposed a wavelet neural network (WNN) to detect small low-altitude targets. The algorithm can detect multiple small targets at the same time. Wu et al. [11] proposed a new deep convolutional network for small targets in infrared images. The problem of small target detection is transformed into the classification of small target position distribution. Excellent results are obtained in different scenarios. In response to the fact that targets and backgrounds differ in some areas, He et al. [12] proposed a multi-scale local gray dynamic range (MLGDR) method, which achieves a high signal-to-noise ratio and low detection rate in different scenes. The paper reports on a new multi-view

https://doi.org/10.1017/S0263574722001059 Published online by Cambridge University Press

algorithm that combines information from multiple images of a single target object. This algorithm is used for the binary classification of underwater images [13]. Deng et al. [14] embeds multi-scale fuzzy metric detection in complex backgrounds, and the algorithm eliminates a large amount of background folding and noise. Cheng et al. [15] presents a method to improve the speed and accuracy rate for space robot visual target recognition based on illumination and affine invariant feature extraction and to reduce the effect of light and occlusion on target recognition. Li et al. [16] proposed a network framework (DMNet) incorporating dilation convolution and multi-scale mechanisms. The algorithm extracts image contextual information using the multi-scale mechanism and extracts small detail features using dilation convolution, which effectively improves the recognition performance of the algorithm. Wang et al. [17] used a single static image density estimation method for CNNs. A multi-scale expanded convolutional module was used to integrate the underlying detailed information into high-level semantic features to enhance the recognition capability of the network. The algorithm has excellent robustness. Fang et al. [18] constructs a multi-scale feature pyramidal fusion neural network based on dilated convolution, and the algorithm achieves faster recognition and tracking of targets. Experiments show that the algorithm has good convergence speed and generalization ability. These methods have effectively solved the problem of feature disappearance and improved the recognition ability for small targets. However, the recognition capability needs to be improved for targets with incomplete features.

Images captured in underwater environments often exhibit complex lighting and severe water turbidity [19, 20]. Also complex terrain obscures the image. These factors make it difficult for AUVs to acquire image information. In solving the occlusion problem, Shen et al. [21] used graph neural networks to mine graph node relationships and CNNs to construct body part maps. The algorithm implements target detection of obscured pedestrians. Wei et al. [22] used curvilinear signal processing (GSP) to characterize the representation space of a graphical neural network (GNN), giving the GNN better observability. Fu et al. [23] designed guided graph CNNs with a new residual shunt structure to investigate the relationship between skeletal data and human actions. Lu et al. [24] proposed a new model for converting semantic segmentation into graphical nodes. The model extends the receptive field and combines structure with feature extraction without losing location information. The approach validates the idea of combining graph structure with deep learning. Zhang et al. [25] extracted spatial and semantic convolutional features using CNNs to keep the spatial features at a high resolution, thus improving the accuracy of visual tracking effectively. The novel algorithm model of a hybrid network model based on CNN and long-short-term memory (LSTM) model is constructed [26]. To mitigate the data sparse problem, the paper [27] combines object proposal with attentional networks for efficiently capturing salient objects and human attention regions in dynamic video scenes. Their proposed framework runs better than existing deep models on saliency detection databases. Tian et al. [28] designed a new contrast loss function. The SGEN architecture was used to train the contrast loss for spatial and semantic similarity. The algorithm effectively improves the object detection performance. Li et al. [29] proposed a new end-to-end semantic segmentation network that integrates lightweight space and channel attention modules, which can refine features to adaptively improve the lightweight space and channel attention modules. The experiments show that the algorithm could achieve better semantic segmentation results. Yin et al. [30] designed an enhanced global attention decoder (EGAUD) that replies to detailed semantic information and makes predictions by enhancing the feature aggregation module for attention and semantic segmentation. A model of gated spaces and semantic attention headings is proposed in the literature [31], and experiments show that the algorithm is effective in terms of quantitative and qualitative results. The above methods have performed well in some tests but fall short for the recognition of small underwater blurred targets.

3. Proposed method

The paper extracts small underwater target features through a hybrid dilated convolution network, increasing the algorithm's perceptual field without increasing the algorithm's computational effort.

https://doi.org/10.1017/S0263574722001059 Published online by Cambridge University Press

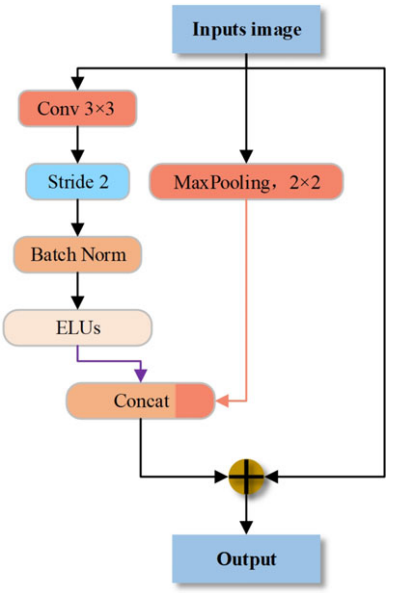

Figure 2. *Basic network module.*

The missing underwater target features of the target are compensated by learning the spatial semantic features of the target through an adaptive correlation matrix. Finally, the proposed algorithm fuses spatial semantic features and visual features for underwater blurred small targets recognition.

3.1. Network model

The underwater environment has problems such as low light and turbid water. These phenomena result in the loss of small target features underwater. The micro-target feature extraction network uses an optimized ResNet as the base network [32, 33]. The network input is a 256 × 256 three-channel image. Thirteen convolution kernels are used to convolution the input image. The convolution kernels are 3 × 3. The step size is 2. A 13-channel feature map is obtained. At the same time, maximum pooling of the input image can effectively preserve the original information of the image and speed up the training. The output result of maximum pooling is a three-channel feature map. The above two results are fused to obtain a 16-channel feature map, as shown in Fig. 2.

In order to maintain the resolution and perceptual field of the network, "holes" are inserted into the convolution kernel. This is called dilated convolution. An expansion filter of size $k_d \times k_d$ is obtained. The convolution kernel is $k \times k$. where $k_d = k + (k-1) \bullet (r-1)$. The expansion module can obtain more fields of view with fewer network layers, effectively speeding up the training while keeping the feature maps of the output layer at the same resolution as the input layer. In this paper, we use hybrid dilated convolution to extract image features to solve the problem of incomplete local information and irrelevant information. The expansion rate of the convolution kernel is set to 1, 2, and 5, as shown in Fig. 3. The method improves the resolution of small target features.

The convolution layer is followed by batch normalization and exponential linear units (ELUs) [34]. The ELUs activation function speeds up learning and avoids gradient disappearance. The activation function is as follows:

$$f(x) = \begin{cases} a\,(e^x - 1)\,x & \text{for } x < 0 \\ 0 & \text{for } x \geq 0 \end{cases} \tag{1}$$

The image x is the input to the feature extraction network Q, and Q output features as $G = Q(x)$.

https://doi.org/10.1017/S0263574722001059 Published online by Cambridge University Press

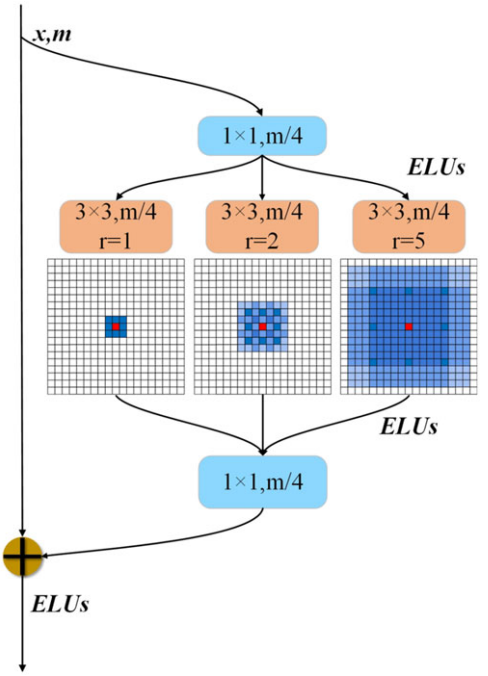

Figure 3. *Hybrid dilated convolution module.*

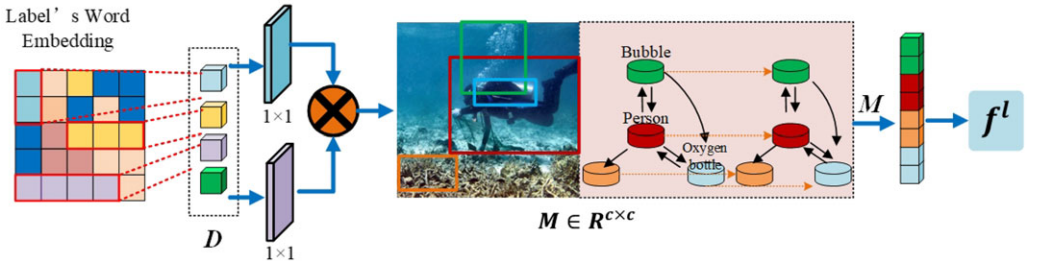

Figure 4. *Spatial semantic feature extraction model.*

3.2. Semantic space feature extraction

Underwater targets are blurred by turbidity and low light levels, resulting in a lack of information on target features. The correlation matrix represents the spatial semantic relationships between different targets. Existing correlation matrices are usually constructed from the label co-occurrence relations of the training set, and their generalization ability is weak. This paper designs adaptive correlation matrices to represent the semantic correlation between targets, as shown in Fig. 4. The adaptive correlation matrix module consists of two 1×1 convolutional layers and a dot product operation [35]. The output learned label correlation matrix M is follows:

$$M_{ij} = \frac{1}{C}\left(W_{\emptyset} * D\right)^{T}\left(W_{\theta} * D\right) = \begin{Bmatrix} a_{00} & \cdots & a_{0(C-1)} \\ \vdots & \ddots & \vdots \\ a_{(C-1)0} & \cdots & a_{(C-1)(C-1)} \end{Bmatrix} \tag{2}$$

where $W_{\emptyset}$ and W_{θ} are denoted as convolution kernels, $*$ is the convolution operation, D is the labeled word embedding vector, and C is the category. As some rare co-occurrence relations may be noise,

https://doi.org/10.1017/S0263574722001059 Published online by Cambridge University Press

a probability threshold τ is set to filter the noise in this paper. The filtered matrix is as follows:

$$B_{ij} = \begin{cases} 0 & M_{ij} < \tau \\ M_{ij} & M_{ij} > \tau \end{cases} \tag{3}$$

In this paper, a spatial semantic feature extraction network is constructed. The network uses an adaptive correlation matrix to represent the semantic correlation between targets and updates the feature representation through information transfer between nodes. Spatial semantic feature extraction networks can all be represented as:

$$f^{l+1} = L\left(B \bullet f^l \bullet W^l\right) \tag{4}$$

Initialize the spatial semantic features, denoted as $f^l = G.f^{l+1}$ is the updated spatial semantic features. B is the normalized adaptive correlation matrix. W^l is the transformation matrix to be learned. $l(\bullet)$ is a nonlinear Leaky ReLU activation function.

3.3. Enhanced dilated convolution framework for underwater blurred target recognition

After the visual features have been acquired, the target is recognized. First, the visual features and the spatial semantic features are fused. Due to the blurring of the underwater image, small targets cannot be effectively recognized by visual features alone. The graph neural network is used to capture the information of surrounding nodes, establish the connection relation between nodes, and extract the spatial semantic features of small targets. The extracted spatial semantic features are fused with visual features, which can effectively improve the accuracy of vision. Anchors are generated for the fused feature nodes. Each point is set with h anchors. Softmax function is used to obtain the determination of anchor frames and extract positive anchors. Bounding box regression regress positive anchors. The results are fed into the proposal layer to calculate the exact proposal. The Cls layer classifies the proposals. The reg layer regresses the proposals again to obtain the target anchor boxes.

On the basis of the acquired visual feature maps, this paper extracts the candidate regions of the target. The algorithm in this paper fuses spatial semantic features and visual features to achieve recognition of small targets. The candidate frame is considered as a node in the graph structure. The spatial semantic features f^{l+1} and visual features G of the nodes are fused, and the target type is predicted based on the fused results. The fused features are expressed as:

$$P_c = F_P\left(G, f^{l+1}\right) \tag{5}$$

where F_P is a feature fusion output function. F_P maps the spatial semantic features f^{l+1} and the feature set G into the feature vector P_c. P_c includes two types of information about the target, respectively, spatial semantic information and visual features. P_c is entered into the fully connected layer in this paper to successfully predict the target category scores.

Target classification is performed by the cls layer. The final output is a $C+1$ dimensional array Y, calculated by the SoftMax function. Y denotes the category confidence level of the target as:

$$Y = \left(y^0, y^1, \cdots, y^C\right) \tag{6}$$

Model training is carried out using a minimization loss function. The loss function includes regression loss and classification loss, which is shown in Eq. (7):

$$L = \frac{1}{N_{\text{cls}}} \sum_i \sum_{c=1}^{C} -(1-\hat{y}_i^c)^\gamma \log \hat{y}_i^c + \lambda \frac{1}{N_{\text{reg}}} \sum_i P_i^* R\left(T_i - T_i^*\right) \tag{7}$$

where c is the target category. R denotes the Smooth L2 function. i denotes the number of the candidate box. y_i^c denotes the confidence level of the category of anchor boxes i. T_i denotes the target anchor boxes coordinates, given by the regression layer. T_i^* is the target real area coordinates. N_{reg} denotes equal

https://doi.org/10.1017/S0263574722001059 Published online by Cambridge University Press

to the number of anchor boxes. N_{cls} denotes the minimum training batch size. N_{cls} and N_{reg} denote the normalization of the loss function. σ (•) is the sigmoid function. λ denotes balanced weights.

4. Experiment

The datasets used in this experiment are all from the Underwater Target dataset (UTD), Cognitive Autonomous Diving Buddy (CADDY) underwater dataset, and Underwater Image Enhancement Benchmark (UIEB) datasets. The images in the dataset contain frogmen, submarines, torpedoes, and AUV types. The dataset has 11,560 labeled images with a ratio of 7:3 between the training and test sets. The training set trained the extraction model, and the test set tested the recognition network. Training and testing were carried out in TensorFlow under Win10. The simulations were run on a small server with a GTX 2080 GPU and 64G RAM.

This paper proposes an enhanced dilated convolution framework for underwater blurred target recognition. Facing the problem of blurred underwater images, we designed two sets of target recognition simulation experiments in this paper. The two sets of simulation experiments were conventional underwater images and blurred images, and the compared algorithms were CRSNet [36], DMNet [37], Improved RetinaNet [38], and MobileNet-SSD [39]. The algorithms were evaluated in terms of recognition accuracy (mAP) and recognition time.

4.1. Clear underwater image recognition results

Figure 5 shows the target recognition results for clear underwater images. In Fig. 5, the rows indicate the recognition accuracy of the same algorithm for different targets. The six target types are torpedo, torpedowake, submarine, frogman, bubble, and AUV, respectively. Table I shows the recognition accuracy and recognition time of the algorithm for clear underwater image targets. The average recognition results of the five algorithms show our algorithm is the best with 0.7315. Our algorithm also has the highest recognition accuracy for frogmen and bubble targets with 0.7717 and 0.7477, respectively. However, the algorithm in this paper is slightly lower than MobileNet-SSD in recognition time with 0.208 s. CRSNet has the highest recognition accuracy for torpedoes and torpedo trails with 0.5641 and 0.6025, respectively. The DMNet algorithm had the highest accuracy in recognition of submarines at 0.9326. However, the average recognition accuracy of DMNet is weaker than the algorithm in this paper, and the algorithm in this paper is more advantageous for the recognition of underwater targets. The Improved RetinaNet algorithm is higher than this paper's algorithm in terms of AUV recognition accuracy, at 0.9470. In terms of recognition speed, the MobileNet-SSD algorithm works best at 0.108 s. The analysis above shows that the algorithm in this paper is the best in terms of average recognition accuracy but less so in terms of torpedo and torpedowake recognition. As can be seen from the first column on the left side of Fig. 5, the algorithm in this paper has no misses in the recognition of small-scale torpedowake and torpedo. The algorithm in this paper is optimal for the recognition of small targets underwater.

4.2. Blurred underwater image recognition results

Figure 6 shows the recognition results of the algorithms in this paper for underwater blurred images. The four columns in the figure represent the recognition accuracy for six target types: torpedo, torpedo wake, submarine, frogman, bubble, and AUV under different algorithms. Table II shows the target recognition accuracy and recognition time of underwater blurred images. From the table, it can be analyzed that the algorithm in this paper has the best recognition effect when facing blurred images, with an average recognition accuracy of 0.7063. It remains the highest recognition rate in recognizing frogmen and bubbles with 0.7588 and 0.7732, respectively. The algorithm in this paper is also the highest in recognizing torpedoes with 0.5149. CRSNet has the highest accuracy of 0.6136 for torpedo wake recognition.

https://doi.org/10.1017/S0263574722001059 Published online by Cambridge University Press

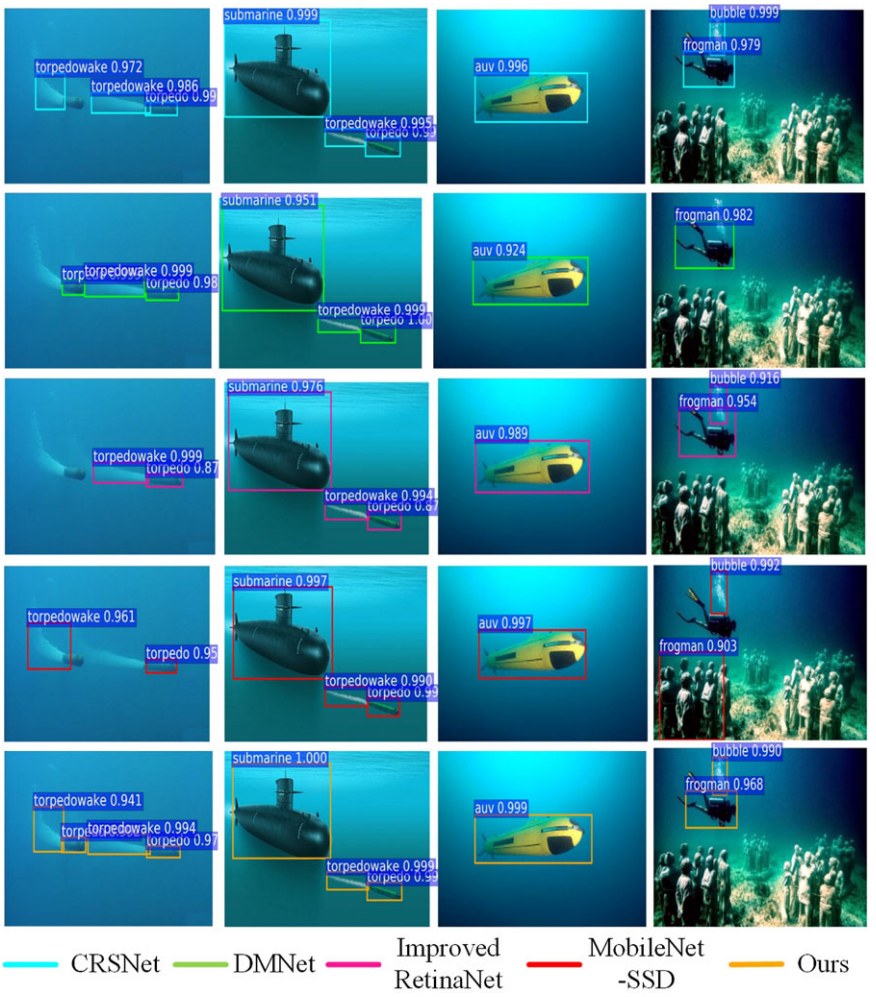

Figure 5. *Clear underwater image target recognition results.*

Improved RetinaNet has the highest accuracy for recognizing AUV and submarine targets, with 0.8420 and 0.9262, respectively. In terms of recognition time, MobileNet-SSD maintains the fastest recognition speed at 0.115 s. The above data show that the algorithm in this paper has the highest mAP when recognizing underwater blurred targets.

4.3. Low light conditions blurred underwater image recognition results

Figure 7 shows the recognition results of the algorithm in this paper for underwater blurred images in low light conditions. The six columns in the figure represent the recognition accuracy of the six target types torpedo, torpedowake, submarine, frogman, bubble, and AUV under different algorithms. Analysis of the graphs shows that the confidence levels shown by each algorithm are relatively good and have high values when performing the recognition of torpedo, submarine, and AUV. The second column of Fig. 7 was analyzed. For torpedowake recognition, CRSNet successfully recognized torpedowake in low light conditions, which is excellent among the algorithms. However, the algorithm has poor recognition results for small-scale torpedoes. The algorithm in this paper has the highest confidence level for small-scale torpedo recognition, at 0.974. For the analysis of the fourth column, CRSNet, Improved RetinaNet and

https://doi.org/10.1017/S0263574722001059 Published online by Cambridge University Press

Table I. *Target recognition accuracy and recognition time for clear underwater images.*

Method	Auv	Bubble	Frogman	Submarine	Torpedo	Torpedowake	mAP	Time
CRSNet	0.8077	0.7407	0.7580	0.8504	**0.5641**	**0.6025**	0.7205	0.372
DMNet	0.8733	0.7029	0.7638	**0.9326**	0.4468	0.5627	0.7136	0.214
Improved RetinaNet	**0.9470**	0.7084	0.7482	0.8751	0.4142	0.5542	0.7078	0.233
MobileNet-SSD	0.8996	0.7340	0.7514	0.8059	0.5637	0.5373	0.7153	**0.108**
Ours	0.9308	**0.7477**	**0.7717**	0.8594	0.5468	0.5330	**0.7315**	0.208

The black bolded font in the table indicates the excellence metrics for each algorithm.

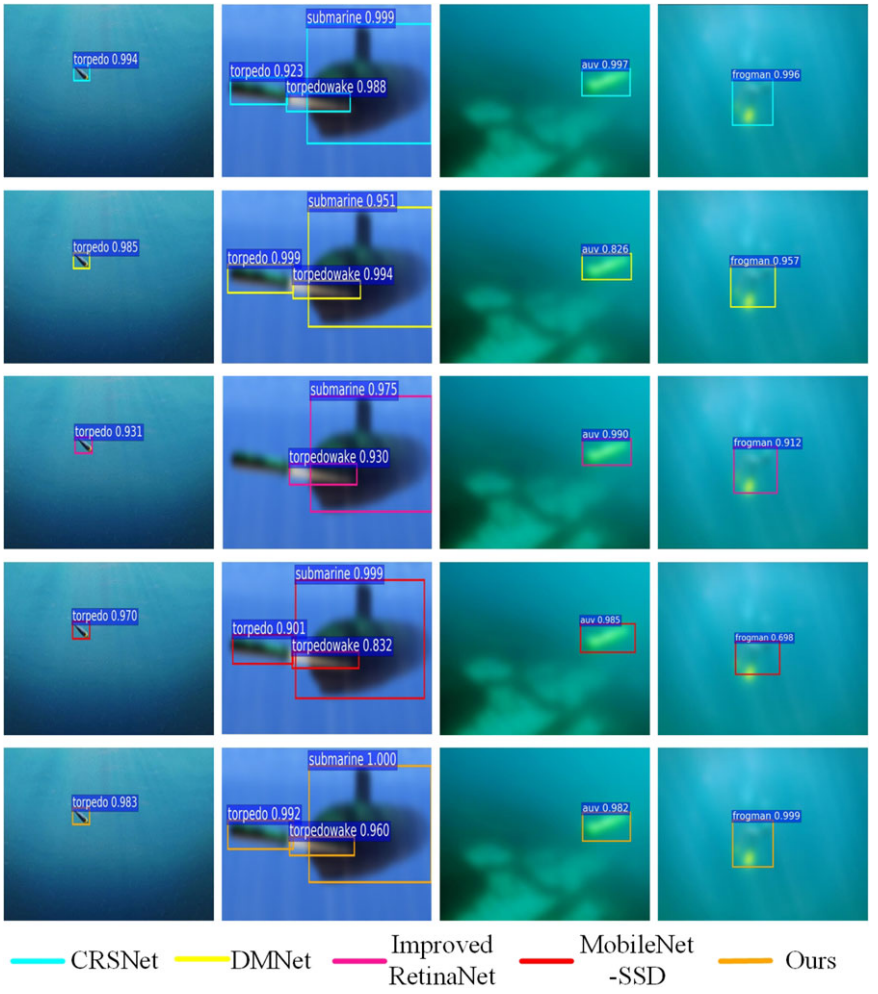

Figure 6. *Underwater blurred image recognition results.*

the algorithm in this paper recognized all the frogman, and bubble. The algorithm in this paper had the highest confidence level for the recognition of frogman and bubble, with 0.992 and 0.998, respectively. From the above analysis, the algorithm in this paper has better results in recognizing low light conditions for blurred underwater images.

https://doi.org/10.1017/S0263574722001059 Published online by Cambridge University Press

Table II. *Underwater blurred image target recognition accuracy and recognition time.*

Method	Auv	Bubble	Frogman	Submarine	Torpedo	Torpedowake	mAP	Time
CRSNet	0.8196	0.7351	0.7409	0.8102	0.4565	**0.6136**	0.6959	0.413
DMNet	0.8496	0.7227	0.7427	0.8311	0.4096	0.4670	0.6704	0.217
Improved RetinaNet	**0.9262**	0.6896	0.7402	**0.8420**	0.4583	0.4022	0.6764	0.248
MobileNet-SSD	0.8331	0.7047	0.6612	0.6941	0.4338	0.4101	0.6228	**0.115**
Ours	0.9022	**0.7732**	**0.7588**	0.7559	**0.5149**	0.5330	**0.7063**	0.215

The black bolded font in the table indicates the excellence metrics for each algorithm.

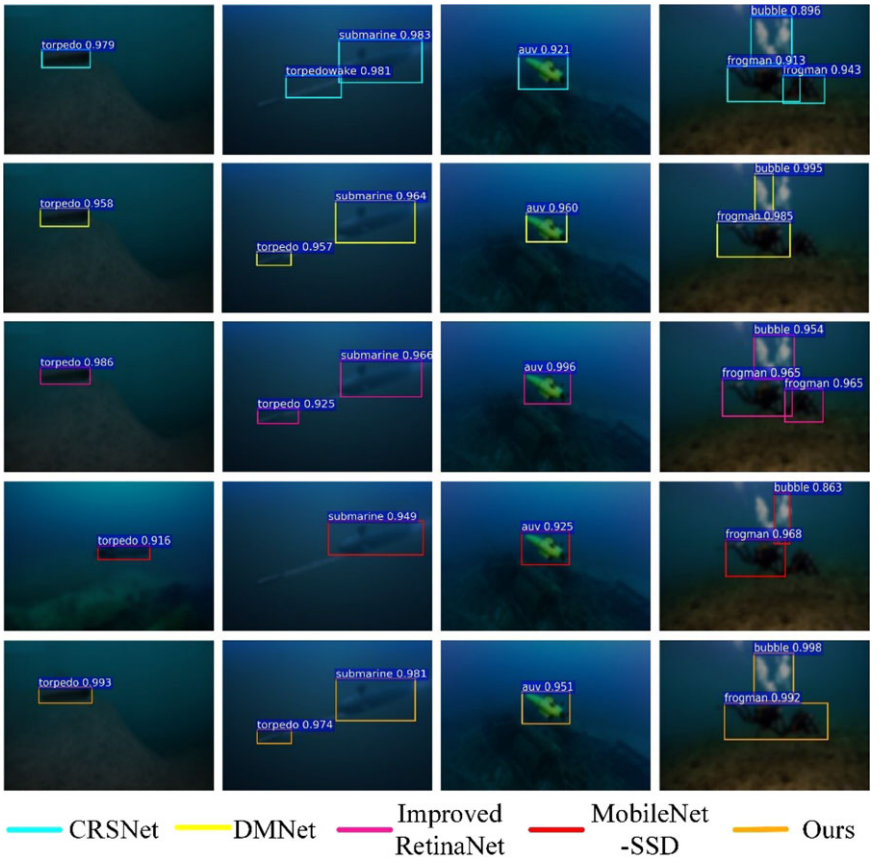

Figure 7. *Low light conditions underwater blurred image recognition results.*

4.4. Experience analysis

The experimental results are analyzed in terms of underwater blurred small target recognition. The algorithm in this paper has the highest average recognition accuracy among the compared algorithms. CRSNet has the longest recognition time, but the average recognition accuracy is only lower than the algorithm in this paper, and the recognition results are also very positive. The average recognition accuracy and recognition time results of DMNet and Improved RetinaNet for small underwater blurred targets are smaller than those of CRSNet. The difference between the average recognition accuracy and recognition time of DMNet and Improved RetinaNet is smaller. MobileNet-SSD has the best recognition speed, but the average recognition accuracy is less effective.

https://doi.org/10.1017/S0263574722001059 Published online by Cambridge University Press

5. Conclusions

Underwater images are blurred due to environmental and light disturbances, and AUVs are challenging for the recognition of small underwater targets. We propose an enhanced dilation convolution framework for underwater blurred target recognition. Firstly, the method extracts small target features through a hybrid dilated convolution feature extraction network, increasing the perceptive field of the algorithm without increasing its computational power. Secondly, this paper learns the spatial semantic features through an adaptive correlation matrix to compensate for the missing features of the target. Finally, this paper uses the fusion of node features and spatial semantic features to achieve the recognition of small blurred targets. The average recognition accuracy of the algorithm in this paper is 1.04% better than existing methods.

Funding. This work was supported by National Key R&D Program of China (2019YFB1311002), and Science and Technology Project of Henan Province (212102210161, 222102320380, 222102110194, and 222102110205).

References

[1] L. Lu, H. Li, Z. Ding and Q. Guo, "An improved target detection method based on multiscale features fusion," *Microw. Opt. Technol. Lett.* **62**(9), 1451–1460 (2020).
[2] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou and G. Cotrell, "Understanding Convolution for Semantic Segmentation," **In:** *2018 18th IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Tahoe, NV, USA (2018).
[3] Q. Sun and L. Cai, "Multi-AUV Target Recognition Method Based on GAN-meta Learning," **In:** *2020 5th International Conference On Advanced Robotics and Mechatronics (ICARM 2020)*, Shenzhen, China (2020) pp. 374–379.
[4] L. Cai, C. Chen and H. Chai, "Underwater distortion target recognition network (UDTRNet) via enhanced image features," *Comput. Intell. Neurosci.* **1**(9), 1–10 (2021).
[5] M. Jian, Q. Qi, J. Dong, Y. Yin and K. M. Lam, "Integrating QDWD with pattern distinctness and local contrast for underwater saliency detection," *J. Vis. Commun. Image Represent.* **53**, 31–41 (2018).
[6] W. Kong, J. Hong, M. Jia, J. Yao, W. Cong, H. Hu and H. Zhang, "YOLOv3-DPFIN: A dual-path feature fusion neural network for robust real-time sonar target detection," *IEEE Sens. J.* **20**(7), 3745–3756 (2019).
[7] Q. Wu, Z. An, H. Chen, X. Qian and L. Sun, "Small target recognition method on weak features," *Multimed. Tools Appl.* **80**(3), 4183–4201 (2021).
[8] F. Gongor and O. Tutsoy, "Design and implementation of a facial character analysis algorithm for humanoid robots," *Robotica* **37**(11), 1835–1849 (2019).
[9] J. Li, F. Zhang, Y. Xiang, S. Pan, "Towards small target recognition with photonics-based high resolution radar range profiles," *Opt. Express* **29**(20), 31574–31581 (2021).
[10] C. Cao, Q. Hou, T. A. Gulliver and Q. Lan, "A passive detection algorithm for low-altitude small target based on a wavelet neural network," *Soft Comput.* **24**(14), 10693–10703 (2020).
[11] W. Shuang-Chen and Z. Zheng-Rong, "Small target detection in infrared images using deep convolutional neural networks," *J. Infrared Millim. Waves* **38**(3), 371 (2019).
[12] Y. He, C. Zhang, T. Mu, T. Yan, Y. Wang and Z. Chen, "Multiscale local gray dynamic range method for infrared small-target detection," *IEEE Geosci. Remote Sens. Lett.* **18**(10), 1846–1850 (2020).
[13] P. Kannappan and H. G. Tanner, "Distance-based global descriptors for multi-view object recognition," *Robotica* **38**(1), 106–117 (2020).
[14] H. Deng, X. Sun and X. Zhou, "A multiscale fuzzy metric for detecting small infrared targets against chaotic cloudy/sea-sky backgrounds," *IEEE Trans. Cybern.* **49**(5), 1694–1707 (2018).
[15] L. B. Cheng, Z. H. Jiang, B. W. H.Li and Q. Huang, "Target-tools recognition method based on an image feature library for space station cabin service robots," *Robotica* **34**(4), 925–941 (2016).
[16] W. Li, X. Zhang, Y. Peng and M. Dong, "DMNet: A network architecture using dilated convolution and multiscale mechanisms for spatiotemporal fusion of remote sensing images," *IEEE Sens. J.* **20**(20), 12190–12202 (2020).
[17] Y. Wang, S. Hu, G. Wang, C. Chen and Z. Pan, "Pan "Multi-scale dilated convolution of convolutional neural network for crowd counting," *Multimed. Tools Appl.* **79**(1), 1057–1073 (2020).
[18] M. Jian, X. Liu, H. Luo, X. Lu, H. Yu and J. Dong, "Underwater image processing and analysis: A review," *Signal Process. Image Commun.* **91**, 116088 (2021).
[19] M. Jian, Q. Qi, H. Yu, J. Dong, C. Cui, X. Nie, H. Zhang, Y. Yin and K. M. Lam, "The extended marine underwater environment database and baseline evaluations," *Appl. Soft. Comput.* **80**, 425–437 (2019).
[20] J. Fang and G. Liu, "Visual object tracking based on mutual learning between cohort multiscale feature-fusion networks with weighted loss," *IEEE Trans. Circuits Syst. Video Technol.* **31**(3), 1055–1065 (2020).
[21] C. Shen, X. Zhao, X. Fan, X. Lian, F. Zhang, A. R. Kreidieh and Z. Liu, "Multi-receptive field graph convolutional neural networks for pedestrian detection," *IET Intell. Transp. Syst.* **13**(9), 1319–1328 (2019).

https://doi.org/10.1017/S0263574722001059 Published online by Cambridge University Press

[22] F. Gama, E. Isufi, G. Leus and A. Ribeiro, "Graphs, convolutions, and neural networks: From graph filters to graph neural networks," *IEEE Signal Process. Mag.* **37**(6), 128–138 (2020).
[23] B. Fu, S. Fu, L. Wang, Y. Dong and Y. Ren, "Deep residual split directed graph convolutional neural networks for action recognition," *IEEE Multimed.* **27**(4), 9–17 (2020).
[24] Y. Lu, Y. Chen, D. Zhao, B. Liu, Z. Lai and J. Chen, "CNN-G: Convolutional neural network combined with graph for image segmentation with theoretical analysis," *IEEE Trans. Cogn. Dev. Syst.* **13**(3), 631–644 (2020).
[25] J. Zhang, X. Jin, J. Sun, J. Wang, A. K. Sangaiah, "Spatial and semantic convolutional features for robust visual object tracking," *Multimed. Tools Appl.* **79**(21), 15095–15115 (2020).
[26] P. Zhang and J. X. Zhang, "Deep learning analysis based on multi-sensor fusion data for hemiplegia rehabilitation training system for stoke patients," *Robotica* **40**(3), 780–797 (2022).
[27] S. Tian, L. Kang, X. Xing, Z. Li, L. Zhao, C. Fan and Y. Zhang, "Siamese graph embedding network for object detection in remote sensing images," *IEEE Geosci. Remote Sens. Lett.* **2**(4), 602–606 (2020).
[28] H. Li, K. Qiu, L. Chen, X. Mei, L. Hong, C. Tao, "SCAttNet: Semantic segmentation network with spatial and channel attention mechanism for high-resolution remote sensing images," *IEEE Geosci. Remote Sens. Lett.* **18**(5), 905–909 (2020).
[29] L. Yin and H. Hu, "Enhanced global attention upsample decoder based on enhanced spatial attention and feature aggregation module for semantic segmentation," *Electron. Lett.* **56**(13), 659–661 (2020).
[30] S. Wang, L. Lan, X. Zhang and Z. Luo, "GateCap: Gated spatial and semantic attention model for image captioning," *Multimed. Tools Appl.* **79**(17), 11531–11549 (2020).
[31] M. Jian, J. Wang, H. Yu and G. G. Wang, "Integrating object proposal with attention networks for video saliency detection," *Inf. Sci.* **576**, 819–830 (2021).
[32] B. Avelin and K. Nyström, "Neural ODEs as the deep limit of ResNets with constant weights," *Anal. Appl.* **19**(3), 397–437 (2021).
[33] X. Zhang, Z. Chen, Q. J. Wu, L. Cai, D. Lu and X. Li, "Fast semantic segmentation for scene perception," *IEEE Trans. Ind. Inform.* **15**(2), 1183–1192 (2018).
[34] T. Yang, Y. Wei, Z. Tu, H. Zeng, M. A. Kinsy, N. Zheng and P. Ren, "Design space exploration of neural network activation function circuits," *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **38**(10), 1974–1978 (2018).
[35] Q. Li, X. Peng, Y. Qiao and Q. Peng, "Learning label correlations for multi-label image recognition with graph networks," *Pattern Recognit. Lett.* **138**(1), 378–384 (2020).
[36] Y. Li, X. Zhang and D. Chen, "Csrnet: Dilated Convolutional Neural Networks for Understanding the Highly Congested Scenes," **In:** *31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, pp. 1091–1100.
[37] J. Jiang, C. Lyu, S. Liu, Y. He and X. Hao, "RWSNet: A semantic segmentation network based on SegNet combined with random walk for remote sensing," *Int. J. Remote Sens.* **41**(2), 487–505 (2020).
[38] H. Tian, Y. Zheng and Z. Jin, "MobileNet-SSD MicroScope Using Adaptive Error Correction Algorithm: Real-Time Detection of License Plates on Mobile Devices," **In:** *6th International Conference on Energy, Environment and Materials Science (EEMS)*, Hulun Buir, China (2020) pp. 1091–1100.
[39] X. Hu, H. Li, X. Li and C. Wang, "MobileNet-SSD MicroScope using adaptive error correction algorithm: Real-time detection of license plates on mobile devices," *IET Intell.* **14**(2), 110–118 (2020).

Cite this article: L. Cai, X. Qin and T. Xu (2023). "EHDC: enhanced dilated convolution framework for underwater blurred target recognition", Robotica **41**, 900–911. https://doi.org/10.1017/S0263574722001059