MAGNet: A camouflaged object detection network simulating the observation effect of a magnifier

Xinhao Jiang
Xi'an High-Tech Institude https://orcid.org/0000-0002-4239-1738
Wei Cai (xhtu807@wanfeng.edu.bi)
Xi'an High-Tech Institude
Zhili Zhang
Xi'an High-Tech Institude
Bo Jiang
Xi'an High-Tech Institude
Zhiyong Yang
Xi'an High-Tech Institude
Xin Wang
Xi'an High-Tech Institude

Research Article

Keywords: Camouflaged object detection, Image segmentation, Deep learning, Human visual system, Computer vision

Posted Date: March 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1020529/v2

MAGNet: A camouflaged object detection network simulating the observation effect of a magnifier

Xin-hao Jiang, Wei Cai*, Zhi-li Zhang, Bo Jiang, Zhi-yong Yang, Xin Wang

Xi'an Research Institute of High Technology, Xi'an 710025, China

Abstract

In recent years, protecting important objects by simulating animal camouflage has been widely used in many fields. Therefore, camouflaged object detection (COD) technology has emerged. COD is more difficult than traditional object detection techniques because of the high degree of fusion of camouflaged objects with the background. In this paper, we strive to more accurately identify camouflaged objects. Inspired by the use of magnifiers to search for hidden objects in pictures, we propose a COD network that simulates the observation effect of a magnifier, called the MAGnifier Network (MAGNet). Specifically, our MAGNet contains two parallel modules, i.e., the ergodic magnification module (EMM) and the attention focus module (AFM). The EMM is designed to mimic the process of a magnifier enlarging an image, and AFM is used to simulate the observation process in which human attention is highly focused on a region. The two sets of output camouflaged object maps are merged to simulate the observation of an object by a magnifier. In addition, a weighted key point area perception loss function, which is more applicable to COD, is designed based on two modules to give higher attention to the camouflaged object. Extensive experiments demonstrate that compared with 14 cutting-edge detection models, MAGNet can achieve the best comprehensive effect on eight evaluation metrics on the public COD dataset, and the segmentation accuracy is significantly improved. We also validate the models' generalization ability on a military camouflaged object dataset constructed in-house. Finally, we experimentally explore some extended applications of COD.

Keywords: Camouflaged object detection; Image segmentation; Deep learning; Human visual system; Computer vision

1. Introduction

In nature, animals evolve according to the principle of survival of the fittest. They may be able to camouflage their shape or retain shape characteristics similar to those of their habitat to avoid being preyed on by attackers or to better ambush prey [1]. Animals that can achieve the former, such as chameleons, can adapt their skin coloring to match their external environment [2], while those that can achieve the latter, such as white moths and black moths, exhibit different survival probabilities in different habitats.

Currently, with the progress of science and technology, camouflage technology simulating animal

camouflage, such as camouflage clothing and camouflage nets [3], has been widely used in high-tech wars. With the use of camouflage technology, snipers can ambush an enemy's senior generals, and armored vehicles can deceive enemy reconnaissance in visible wavelengths. Therefore, research on accurate camouflaged object segmentation is of great significance in military fields.

In addition to its military value, camouflaged object detection (COD) can be applied to industrial detection (e.g., equipment defect detection [4]), medical diagnoses (e.g., testing whether lungs are infected by pneumonia [5, 6]), monitoring and protection (e.g., suspicious person or unmanned aerial vehicle intrusion detection [7, 8]) and unmanned driving (e.g., road obstacle detection [9]).

However, studies on camouflaged object segmentation are lacking. For example, in military fields, military camouflaged objects are often identified by means of infrared-, polarization-, and hyperspectral-based imaging and other technologies [10-12]. However, the scientific problem of how to accurately segment camouflaged objects in the visible light band has been ignored.

In this paper, we propose a camouflaged object segmentation network based on observation with magnifiers called MAGnifier Network (MAGNet). Fig. 1 is a schematic diagram demonstrating a search for camouflaged military objects based on observation with a magnifier. With the influence of camouflage coating, external camouflage materials, smoke barriers and ground object shielding, the soldier and tank in Fig. 1(a) achieve near-perfect integration with the background. However, Fig. 1(b) shows that the camouflaged objects in the picture can be simply and effectively observed with a magnifier. First, the magnifier visually enlarges the observation area, and then we can see edge information and key parts of camouflaged objects in the enlarged area, so we can focus on the key points for accurately identifying the camouflaged objects in the region.

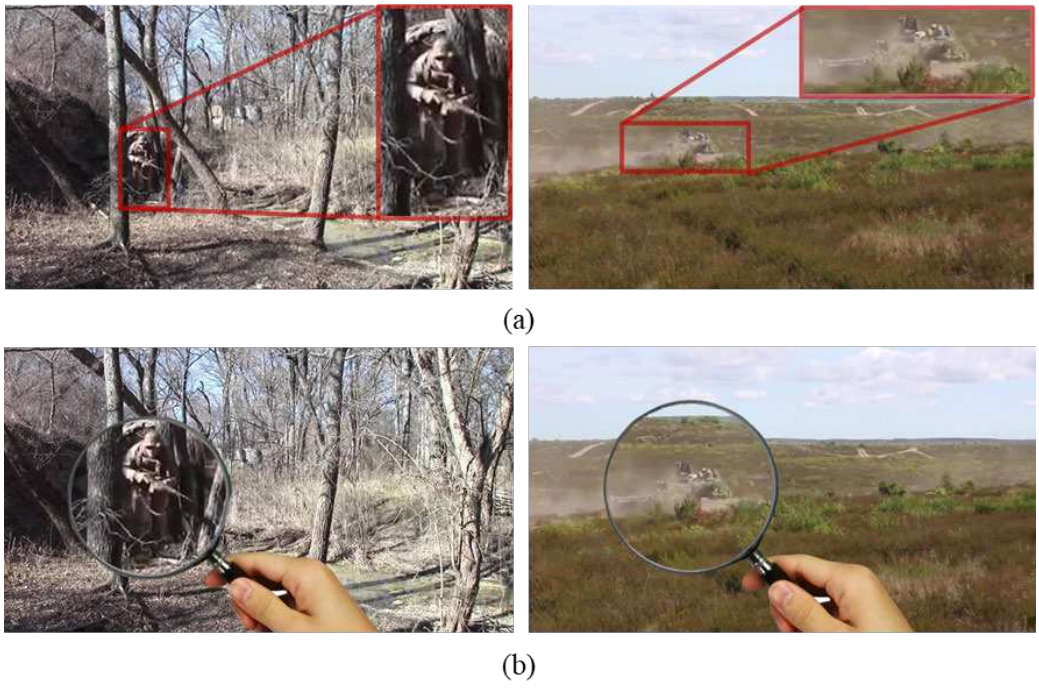

Fig.1. Schematic diagram of the observation of a camouflaged soldier and tank with a magnifier

In summary, the major contributions of this paper are threefold:

1. We apply the concept of observation with a magnifier to the COD problem and propose a novel camouflaged object segmentation network called MAGNet.
2. We design a parallel structure with the ergodic magnification module (EMM) and attention focus module (AFM) to simulate the functions of the magnifier. We propose a weighted key point area perception loss function to improve the segmentation performance.
3. We perform extensive experiments using public COD benchmark datasets and a camouflaged military object dataset constructed in-house. MAGNet has the best comprehensive effect in eight evaluation metrics in a comparison with 14 cutting-edge detection models. Finally, we experimentally explore several potential applications of camouflaged object segmentation.

This paper is organized as follows: Previous work similar to that of this study is introduced in Section 2. Section 3 provides detailed descriptions of our MAGNet and the associated modules. Section 4 presents comparative experiments and quantitative and qualitative analysis of the experimental results. Finally, Section 5 concludes the paper.

2. Related Works

2.1 Camouflaged Object Detection Based on Deep Learning

2020 can be regarded as the first year of research on COD based on deep learning. Fan et al. [13] constructed a complete camouflaged object dataset named COD10K and presented a corresponding camouflaged object segmentation network that promotes the rapid development of COD. In 2021, Mei et al. [14] simulated the predation process of animals and proposed PFNet, a camouflaged object segmentation network based on distraction mining. Lv et al. [15] proposed a joint learning network that can simultaneously localize, segment and rank camouflaged objects and proposed a new COD dataset called NC4K. However, the design principle and network structure of the existing COD models are relatively complex. This paper presents a bionic model based on observation with a magnifier. The principle is easy to understand, and the structure is simple and efficient.

2.2 COD Dataset

Because of the similarity between a camouflaged object and the background, the boundary between the foreground and the background is very difficult to distinguish, so the production of a camouflaged object dataset is very time-consuming [16]. Currently, three major published datasets are the most commonly used. The number of images in the CHAMELEON dataset is small, with only 76 published images collected from the internet [17]. The CAMO dataset contains 1250 images in eight categories [18]. In 2020, Fan et al. proposed the COD10K universal camouflaged object dataset, which has 78 subclasses of 10K images, and this dataset is very precise and challenging [14].

2.3 Semantic Segmentation Based on Deep Learning

In recent years, scene understanding technologies for use in autonomous driving [19], virtual reality [20] and

augmented reality [21] have developed rapidly. As the basic task of scene understanding, semantic segmentation technology based on pixel-by-pixel classification has been widely studied [22-24]. Many semantic segmentation methods based on deep learning have been proposed [25-28]. Currently, there are four main types of networks, namely, the fully convolutional network (FCN) [29], the convolutional neural network (CNN) [30], the recurrent neural network (RNN) [31], and the generative adversarial network (GAN) [32].

2.4 Salient Object Detection Based on Deep Learning

In contrast to camouflaged objects, salient objects are the most noticeable objects in an image. The research of salient object detection can promote image understanding [33], stereo matching [34, 35] and medical disease detection [36-38]. In recent years, salient object detection based on deep learning has been improved mainly by multiscale feature fusion [39], attention mechanisms [40] and edge information [41]. Research on salient object recognition can provide insight into camouflaged object recognition in terms of design principles.

3. MAGNet Detection Model

A magnifier can help an observer quickly find a camouflaged object in an image. This is because the magnifying effect of the magnifier makes it easier for the observer to spot the center, key points and miniscule details of the camouflaged object. Inspired by the magnifier, we apply the magnifier observation effect to the COD problem and design the EMM and the AFM. The EMM is designed to mimic the process of a magnifier enlarging an image, and the AFM is used to simulate the human visual system. Finally, we design a more applicable weighted key point area perception loss function for camouflaged object segmentation, which directs more attention to the camouflaged object in the region by weighting.

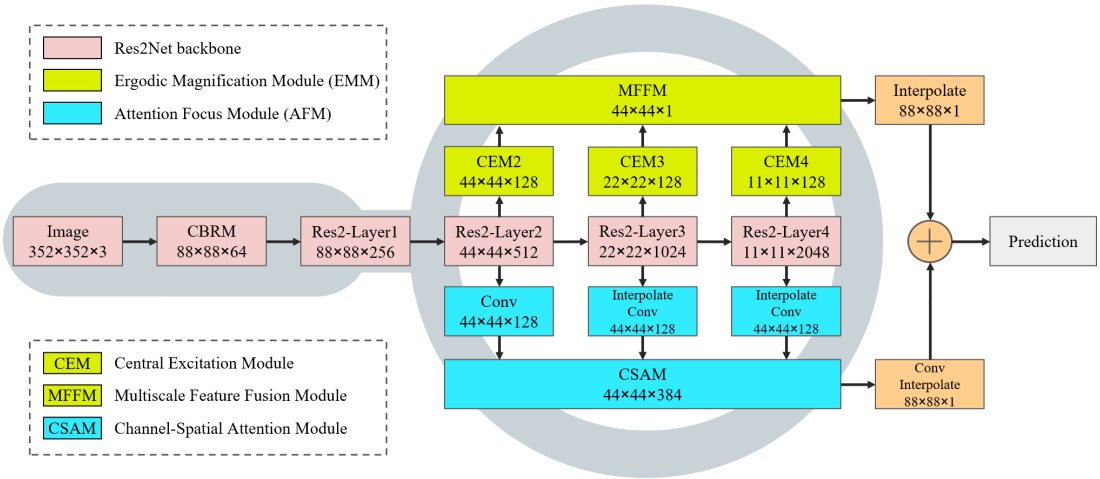

Fig.2. MAGNet structure

3.1 Network Overview

The network structure of MAGNet is shown in Fig. 2. We input a camouflaged object image into this network. MAGNet first extracts multiscale feature maps through a Res2Net-50 backbone [42], and then the latter three feature maps are fed to the EMM and the AFM in parallel. Finally, the output feature maps of the two modules

are fused to simulate observation with a magnifier.

3.2 Ergodic Magnification Module (EMM)

As shown in Fig. 2, the EMM consists of two parts, i.e., the central excitation module (CEM) and the multiscale feature fusion module (MFFM).

The CEM is used to traverse the feature maps of the different scales of output from the last three layers of the backbone to expand the receptive field and to intensify the central point and key points.

The MFFM is designed to fully integrate the multiscale feature maps after the CEM to realize the efficient utilization of high-level and low-level features.

3.2.1 Central Excitation Module (CEM)

We find that when observers use a magnifier to observe an object, they observe the central area of the magnifying glass more carefully than the edge areas. With the human visual receptive field mechanism, an observer is more attracted to the center of an object [43]. Then, we use the magnifier to traverse the whole picture until the center of the magnifier coincides with the center of the object.

To simulate the visual magnification and traversing of the magnifier, we design a simple and efficient CEM, as shown in Fig. 3. The realization of the above functions mainly depends on dilated convolution (DConv) with different sizes of convolution kernels [44].

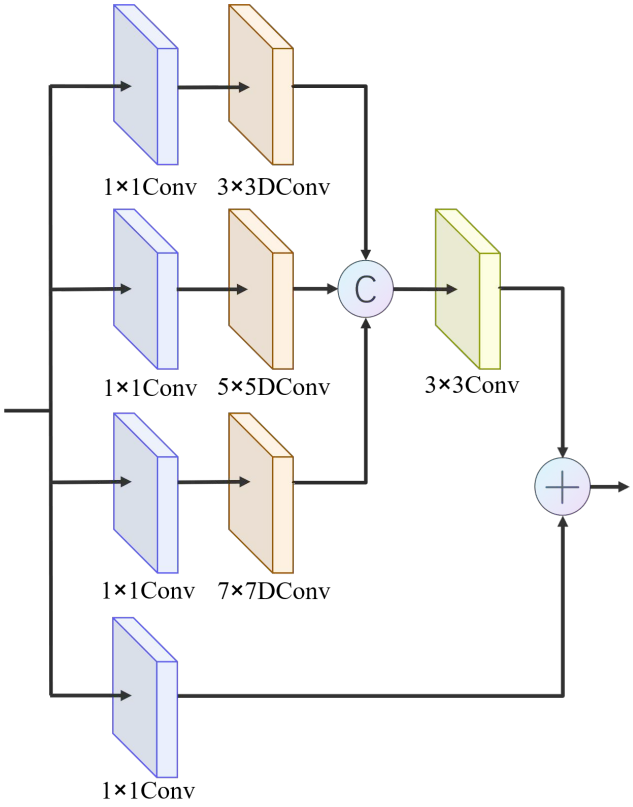

Fig.3. The structure of the CEM

Specifically, the CEM consists of four branches, and the input feature maps are simultaneously fed into all four branches. The four branches first use a 1×1 convolution to change the number of output channels. Then, to

achieve efficient multiscale visual amplification, three of the branches use 3×3, 5×5, and 7×7 DConvs with an expansion factor of 2. After the three sets of output feature maps are connected, a 3×3 convolutional layer is used for fusion between channels. The fourth branch is the residual connection module, which aims to retain part of the original features to reduce the feature loss due to convolution. The two sets of features are connected to obtain a centrally excited feature map. The multiscale centrally excited feature maps obtained from the last three layers of backbone input to the CEM have the same number of channels (128) to ensure a balanced utilization of information at each scale.

The connection of three sets of DConvs can increase the importance of the central features while increasing the receptive field. As shown in Fig. 4, central excitation can be achieved.

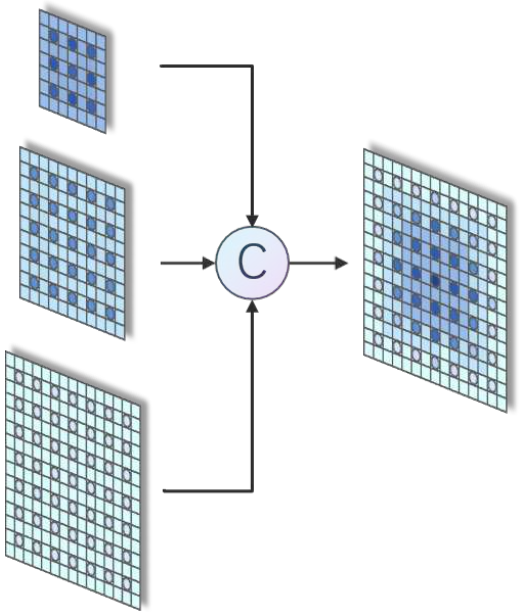

Fig.4. Schematic diagram of the central excitation effect of CEM

3.2.2 Multiscale Feature Fusion Module (MFFM)

The function of the MFFM is to fully integrate the feature maps after central excitation of different scales, thereby outputting a camouflaged object map that contains abundant high- and low-level features. The MFFM structure diagram is shown in Fig. 5. The small-scale excitation feature map transmits the feature information to the large-scale feature map through continuous upsampling and fusion and then generates an output feature map with a size of 44×44×1.

The front-end fusion method of the module adopts the Hadamard product ($\otimes$). The Hadamard product calculation method is pixel-by-pixel multiplication, which can better achieve feature crossover, eliminating the difference between the two groups of features and improving the feature fusion capabilities.

The back end of the module is fused by adding the channels, which can fuse the features of each layer to increase the feature dimension but does not increase the internal feature information, making full use of the semantic information of the high-level and low-level features.

The module output map is denoted Fout, the large-scale feature map in the module is denoted F_i, and the small-scale feature map is denoted F_{i-1}. In Fig. 5, the feature map output by the Hadamard convolution module in blue is F_h, and the feature map output by the green concat module is F_c. Then, we have the following formula:

$$F_h = F_i \times CBR(UP(F_{i-1})) \tag{1}$$

$$F_c = Concat(F_i, CBR(UP(F_{i-1}))) \tag{2}$$

$$F_{out} = CBR(F_c) \tag{3}$$

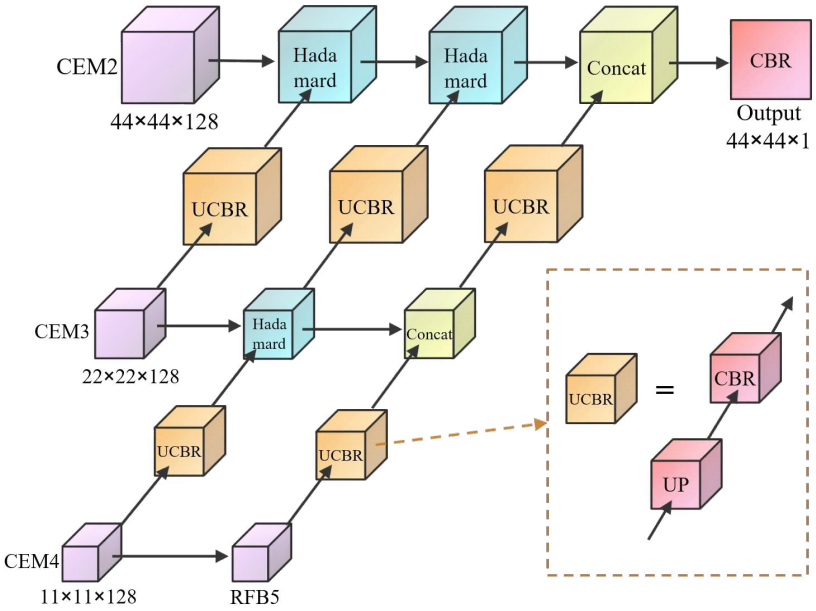

Fig.5. Structure of the MFFM module. (UP: upsample, CBR: Conv+BatchNorm+ReLU)

3.3 Attention Focus Module (AFM)

The AFM has two steps. First, through upsampling and convolution operations, the three sets of feature maps output by the backbone are processed into feature maps of the same size with the same number of channels. Then, the maps are input into the channel-spatial attention module (CSAM) to simulate the effect of human attention focused on observing objects in the field of view of a magnifier.

3.3.1 Channel-Spatial Attention Module (CSAM)

Attention mechanisms in deep learning can simulate the human visual attention mechanism, where the goal is to obtain more important information [43]. Attention mechanisms are mainly divided into two types: spatial attention mechanisms and channel attention mechanisms. A spatial attention mechanism module can extract the most important regional features in the spatial domain and retain locally important information by spatial transformation. A channel attention mechanism module can assign different weights according to the importance of each channel so that the model pays more attention to channels with more important information [45]. The two methods have advantages and disadvantages, and the CSAM we propose is a parallel fusion mechanism of spatial attention and channel attention, as shown in Fig. 6.

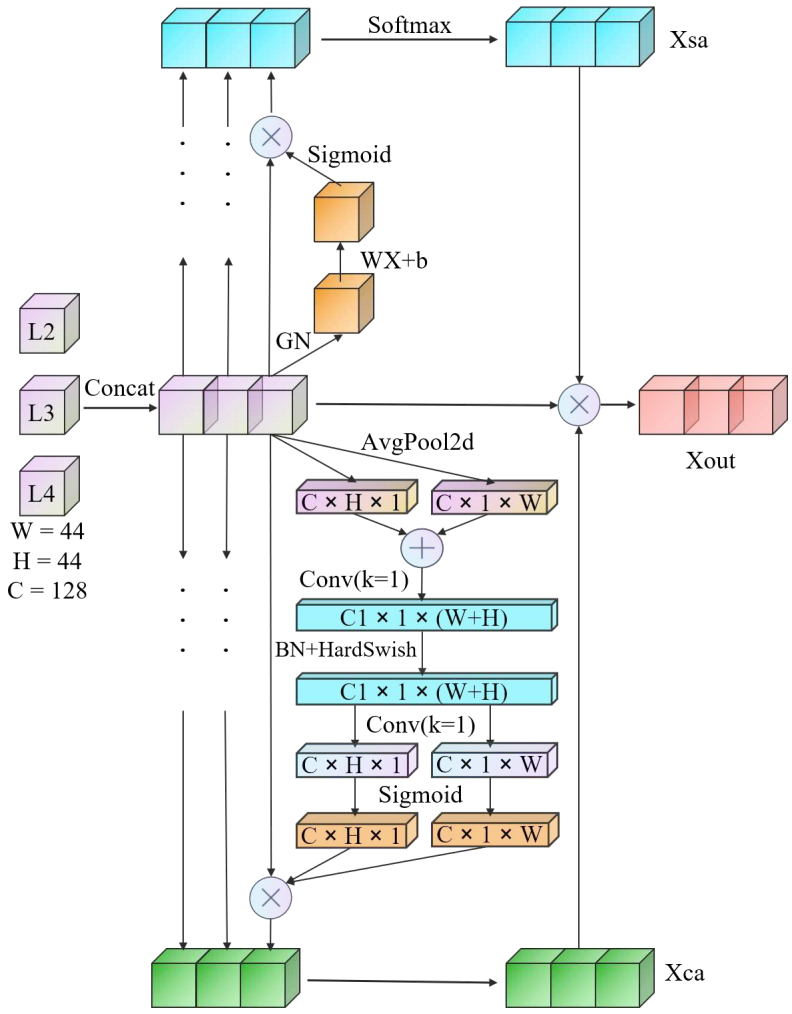

Fig.6. The structure of the CSAM. (L2, L3, and L4 refer to Res2-Layer2, Res2-Layer3, and Res2-Layer4, respectively)

As illustrated in Fig. 6, the CSAM is mainly implemented in four steps. The pseudocode of the CSAM is as follows:

Algorithm 1: CSAM Algorithm

Input: L2, L3, L4.

```
# 1. Feature Maps Concat
X-original = Concat(L2, L3, L4)
for i = 2, 3, 4:
    # 2. Spatial Attention
    xsa_i = GN(Li)
    xsa_i = Weight * xsa_i + bias
    xsa_i = Li * Sigmoid(xsa_i)
    # 3. Channel Attention
    xca_i = CAmodule(Li)
```

Xsa = Concat(xsa_3, xsa_4, xsa_5)

Xsa = Softmax(Xsa)

Xca = Concat(xca_3, xca_4, xca_5)

 # 4. Fusion Attention Maps

Xout = X-original * Xca * Xsa

Output: Xout.

Feature Maps Concat: Superimposing the feature maps of the same size with the same number of channels in the latter three layers of the backbone after processing can achieve the average utilization of feature maps of each scale and fully fuse the semantic information of high- and low-level features. Then, the feature maps of the three different layers are input into the channel attention mechanism branch and spatial attention mechanism branch to generate a channel attention map and a spatial attention map, respectively.

Channel Attention: The squeeze-and-excitation (SE) module is the most commonly used method of channel attention [46]. It can extract important features by assigning weights to each channel but does not learn the importance of location information. Therefore, we embed the coordinate attention (CA) module [47], which can fully perceive position information, into CSAM. The CA module first aggregates features near key points in the image into a pair of key point direction-aware feature maps $K_H(C,H,1)$ and $K_W(C,1,W)$ with different orientations using two 2D-average-pooling operations in the horizontal and vertical dimensions.

$$K_W(c,1,w)=\frac{1}{H}\sum_{0\le i<H}F_{input}(c,i,w),\quad 0\le c<C,0\le w<W \tag{4}$$

$$K_H(c,h,1)=\frac{1}{W}\sum_{0\le j<W}F_{input}(c,h,j),\quad 0\le c<C,0\le h<H \tag{5}$$

Where F_{input} denotes the input feature maps, the two direction-aware feature maps are fused by cascade and convolution operations, yielding

$$F(C1,1,W+H)=\xi(Conv_{C1}(\square K(C,H,1),K(C,1,W)\square)) \tag{6}$$

where $\square\cdot,\cdot\square$ denotes the concatenation operation along the spatial dimension, $Conv_{C1}(\cdot)$ denotes the 1×1 convolution with C1 convolution kernels, $\xi(\cdot)$ denotes BatchNorm and HardSwish operations on feature maps. And the fused feature maps are sliced and encoded into two attention maps storing location information.

$$\{F_H(C,H,1),F_W(C,1,W)\}=\delta(Conv_C(Slice[F(C1,1,W+H)])) \tag{7}$$

where $Slice[\cdot]$ denotes the slece operation along the spatial dimension, $Conv_C(\cdot)$ denotes the 1×1 convolution with C convolution kernels. $\delta(\cdot)$ denotes sigmoid activation function.

Finally, the new and old feature maps are multiplied pixel by pixel by Hadamard convolution to generate a channel attention map with location and direction information embedded.

$$F_{output}(c,i,j)=F_{input}(c,i,j)\times F_H(c,i,1)\times F_W(c,1,j),\quad 0\le c<C,0\le i<H,0\le j<W \tag{8}$$

Spatial Attention: The spatial attention mechanism is particularly important for finding special targets and

can retain important local information. We first use GroupNorm (GN) for group normalization. The second step is to use a set of trainable parameters, namely, weight (w) and bias (b), to assign spatial weights to enhance the representation abilities of the feature map. The third step is to use a sigmoid function for activation and then to multiply the original feature map pixel by pixel to obtain the spatial attention map. Finally, we use softmax to normalize again.

Fusion Channel and Spatial Attention Maps: We use the Hadamard product for the fusion of attention maps, that is, the method of pixel-by-pixel multiplication, which can better enhance the feature information to obtain a more accurate feature map.

3.4 Output Prediction and Loss Function

Finally, the feature maps output by the EMM and AFM are transformed into a single-channel camouflaged object map through an upsampling operation. The two feature maps are fused by pixel-by-pixel addition.

Of a large number of target segmentation algorithms, the binary cross entropy (BCE) loss function and the intersection over union (IOU) loss function are the most common [48]. However, BCE loss and IOU loss averaging of all pixel points cannot be applied to COD. In these images, camouflaged objects require more attention than other objects (especially salient objects) due to their indistinguishable characteristics.

Combining the designed pair of focusing and amplifying modules, we propose a weighted key point area perception loss based on the BCE loss and IOU loss (L_{kaa}^{w}), adding the key point area perception weight to jointly obtain the loss function:

$$L_{kaa}^{w} = L_{wbce}\left(P,GT\right) + L_{wiou}\left(P,GT\right) \tag{9}$$

$$L_{wbce}\left(P,GT\right) = -\frac{\sum_{i=1}^{H}\sum_{j=1}^{W} w_{ij} * L_{bce}\left(P,GT\right)}{\sum_{i=1}^{H}\sum_{j=1}^{W} w_{ij}} \tag{10}$$

$$L_{wiou}\left(P,GT\right) = 1 - \frac{\sum_{i=1}^{H}\sum_{j=1}^{W} L_{i,j}^{enter} * w_{ij}}{\sum_{i=1}^{H}\sum_{j=1}^{W} L_{i,j}^{union} * w_{ij}} \tag{11}$$

where *P* is the prediction map, *GT* is the ground truth map, *H* and *W* are the picture length and width, and $L_{bce}\left(P,GT\right)$ is the original BCE loss function. The expression for the key point area perception weight $w_{i,j}$ is as follows:

$$w_{i,j}=\begin{cases}\left|\dfrac{\sum_{h,w}GT_{h,w}}{hw}-GT_{i,j}\right| & ,\dfrac{\sum_{h,w}GT_{h,w}}{hw}<\dfrac{1}{2}\\ 1-\left|\dfrac{\sum_{h,w}GT_{h,w}}{hw}-GT_{i,j}\right| & ,\dfrac{\sum_{h,w}GT_{h,w}}{hw}>\dfrac{1}{2}\end{cases} \tag{12}$$

where h and w are the sizes of the regions around the pixel points in the GT map, $\sum_{h,w}GT_{h,w}$ denotes the sum of the values of all the pixel points within the region of $h\times w$ centered on the pixel point (i,j) in the GT map. h and w are taken as small as possible, because taking too large a value will affect the model efficiency. However, it should not be smaller than the maximum perceptual field of 32*32 for a single pixel (i.e., the maximum number of down-sampling multiples). So a region range of size 33 × 33 is selected in this experiment, and 33 being an odd number also avoids the case of where the weight is equal to 1/2, and $GT_{i,j}$ is the value of the pixel point (x,y) in the GT map. From equation (12), it can be seen that the key point area perception weight directs more attention to the camouflaged object regardless of the percentage of the camouflaged object in the region, thus making the model training favorable to segmenting camouflaged objects.

4. Experimental Results and Analysis

4.1 Preparation Work

The experimental platform system is Windows 10, the GPU of the platform is an NVIDIA Quadro GV100, and the video memory is 32 GB. The CPU is an Intel Xeon Silver 4210. The experiment uses the PyTorch deep learning development framework. The computing platform is CUDA11.0. We use the Adam optimizer for network optimization during training, the image input size is set to 352×352, and the learning rate is set to 0.0001.

4.1.1 Dataset Preprocessing

We select the CAMO [18] and COD10K [14] datasets, with relatively large data volumes, for evaluation. CAMO includes 1250 images, and COD10K includes 5066 camouflage images. The combined total of 6316 images are combined are divided into a training set, validation set and testing set according to a ratio of 6:2:2. In addition, we perform validation experiments on a military camouflaged object dataset that we constructed. The dataset contains 2700 images of camouflaged soldiers and tanks, and the division ratio is also 6:2:2.

4.1.2 Evaluation Metrics

At present, there are many evaluation metrics suitable for COD, and each metric focuses on different points. Based on previous scholars' research, we select 8 evaluation metrics. A brief introduction of the metrics is as follows: The structure measure (S_α) is a structural similarity evaluation metric focusing on evaluating the structural information of the prediction map [49]. The weighted F-measure (F_β^w) is a comprehensive evaluation of the accuracy and recall rate of the prediction map [50]. The mean absolute error (MAE) is the sum of the

absolute values of the differences between the pixels of the prediction map and the GT map [51]. The adaptive enhanced alignment measure (E_ϕ^{ad}) can evaluate the pixel-level similarity effect and obtain image-level statistics [52]. The mean Dice coefficient (meanDic) represents the percentage of correctly segmented area to true area in the GT image [53]. The mean intersection over union (meanIOU) is the ratio of the area of overlap and concatenation between the predicted and ground truth maps. The mean sensitivity (meanSen) measures the percentage of results predicted to be correct that are actually correct according to the GT image. The mean specificity (meanSpe) measures the percentage of results predicted to be incorrect that are actually incorrect according to the GT image.

4.1.3 Compared Methods

To prove the effectiveness of the MAGNet proposed in this paper, we compare it with 14 classical and state-of-the-art algorithms. These include medical image segmentation methods UNet++ [54], HarDNet [55], PraNet [6], SANet [28], Caranet [56] and UACANet-L [57]; salient object detection methods BASNet [58], SCRN [59], F3Net [48] and GCPANet [60]; and camouflaged object segmentation methods SINet-V1 [13], Rank-Net [15], PFNet [14] and SINet-V2 [61]. For a fair comparison of segmentation performance, all algorithms are trained, validated and tested using the partitioned dataset discussed in Section 4.1.1, and the input sizes are set to 352×352. In addition, the evaluation metrics are calculated using the same set of codes. The evaluation code uses the toolboxes disclosed by PFNet [14] and SINet-V2 [61].

4.2 Comparison with State-of-the-art Algorithms on Public Datasets

4.2.1 Quantitative Comparison

Table 1 comprehensively reports the quantitative results of MAGNet and the latest algorithms on the combined dataset. As seen from the table, MAGNet exhibits the best comprehensive performance according to the eight standard evaluation metrics, in particular, achieving the best performance in the S_α, F_β^w, meanDic and meanIOU metrics. The meanSpe MAGNet is basically equal to that of SINet-V2. MAGNet does not optimize this metric because it has a powerful ability to extract the features of camouflaged targets, which readily results in a certain number of false positives.

Fig. 7 shows the training loss value curves of MAGNet and the optimal COD detection algorithm SINet-V2 [61]. From the figure, we can see that the loss value of MAGNet decreases faster, leveling off at 20 epochs, and that the final loss value is lower.

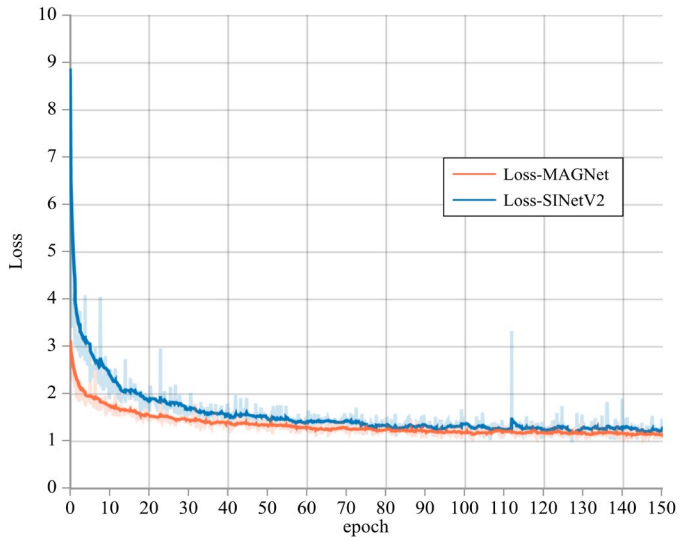

Fig.7. Loss value curves

Table 1 Comparison results of MAGNet and 14 algorithms on public datasets. (⁜ : medical image segmentation method, ◊: saliency object detection method, and ⦚: COD method).

Methods	Pub. 'Year	S_α	F_β^w	MAE	E_ϕ^{ad}	meanDic	meanIoU	meanSen	meanSpe
Unet++ ⁜	DLMIA '17	0.678	0.491	0.067	0.763	0.529	0.416	0.553	0.859
BASNet◊	CVPR '19	0.663	0.439	0.097	0.732	0.490	0.381	0.611	0.865
SCRN◊	ICCV '19	0.791	0.583	0.052	0.799	0.640	0.529	0.676	0.926
HarDNet ⁜	ICCV '2019	0.785	0.651	0.043	0.874	0.676	0.575	0.690	0.930
F3Net◊	AAAI '20	0.781	0.636	0.049	0.851	0.675	0.565	0.709	0.940
PraNet ⁜	MICCAI '20	0.799	0.665	0.045	0.866	0.700	0.595	0.737	0.939
GCPANet◊	AAAI '20	0.800	0.646	0.042	0.851	0.674	0.573	0.691	0.934
SINet-V1⦚	CVPR '20	0.806	0.684	0.039	0.883	0.714	0.608	0.737	0.948
SANet ⁜	MICCAI '21	0.791	0.659	0.046	0.862	0.702	0.593	0.766	0.938
CaraNet ⁜	arXiv '21	0.815	0.679	0.044	0.862	0.722	0.618	0.789	0.937
RankNet⦚	CVPR '21	0.799	0.661	0.043	0.860	0.696	0.588	0.723	0.947
PFNet⦚	CVPR '21	0.805	0.683	0.040	0.882	0.714	0.607	0.737	0.951
UACANet-L ⁜	ACM MM '21	0.816	0.724	0.034	0.901	0.745	0.646	0.763	0.945
SINet-V2⦚	TPAMI '21	0.822	0.700	0.038	0.883	0.735	0.627	0.767	0.955
MAGNet⦚	**Ours**	**0.829**	**0.727**	**0.034**	**0.901**	**0.757**	**0.656**	**0.789**	**0.954**

4.2.2 Qualitative Comparisons

Fig. 8 comprehensively shows the visualization results of all the algorithms in the comparison experiments. It can be observed that MAGNet can more accurately segment camouflaged targets. The EMM can better identify small targets hidden in complex backgrounds (e.g., the fourth column) by magnifying the receptive field and

fusing multiscale features. The AFM can acquire more important information in channels and space by simulating the human visual attention mechanism, so it can accurately segment the details of camouflaged objects (e.g., the seventh column). The use of the weighted key point area perception loss function causes the model to focus more on the regions near the key points of a camouflaged object, thus reducing the segmentation false positive rate (e.g., first column, second column).

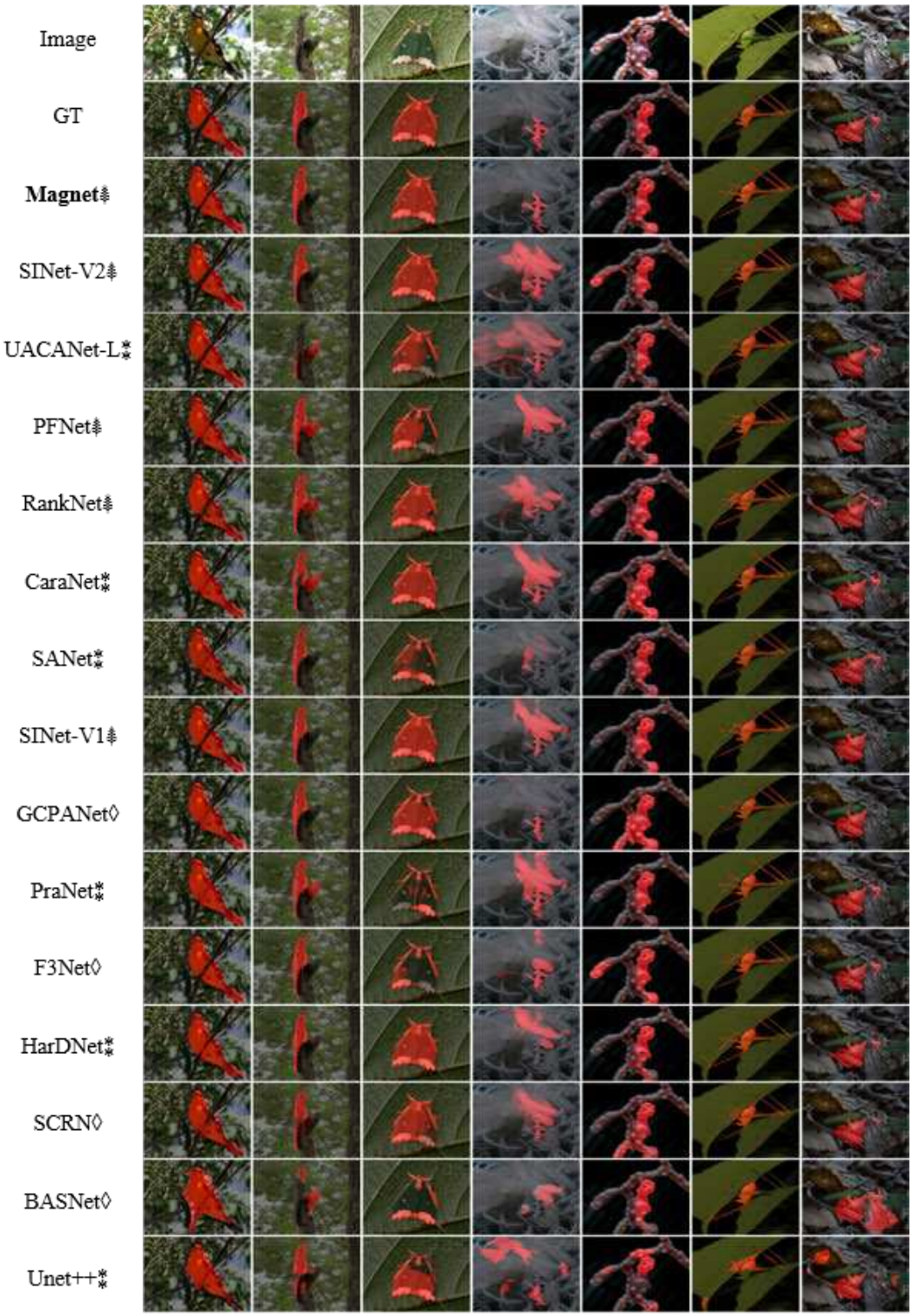

Fig.8. Visualization results for all algorithms on public datasets

4.3 Ablation Experiment

We conduct ablation experiments to verify the effectiveness of two specific modules designed for COD, namely, the EMM and AFM.

4.3.1 Quantitative Comparison

The results of the MAGNet ablation experiments are reported comprehensively in Table 2. Adding the two modules alone significantly improves the model performance. Adding the AFM optimizes meanSen due to the effect of the attention mechanism of the model, which reduces the probability of missed detection. The addition of the EMM optimizes meanSpe because the model's receptive field magnifying mechanism works to reduce the model's false positive probability. We also compare the results with the two key modules connected in series and parallel and ultimately find that the parallel structure better maximizes the effects of both modules.

Table 2 MAGNet ablation experiment results.

baseline	With AFM	With EMM	In series	In parallel	S_α	F_β^w	MAE	E_ϕ^{ad}	meanDic	meanIoU	meanSen	meanSpe
√					0.663	0.315	0.151	0.711	0.522	0.399	0.761	0.826
√	√				0.675	0.308	0.163	0.843	0.616	0.509	**0.824**	0.812
√		√			0.825	0.715	0.035	0.900	0.742	0.638	0.755	**0.956**
√	√	√	√		0.827	0.723	0.034	**0.902**	0.753	0.652	0.785	0.949
√	√	√		√	**0.829**	**0.727**	**0.034**	0.901	**0.757**	**0.656**	0.789	0.954

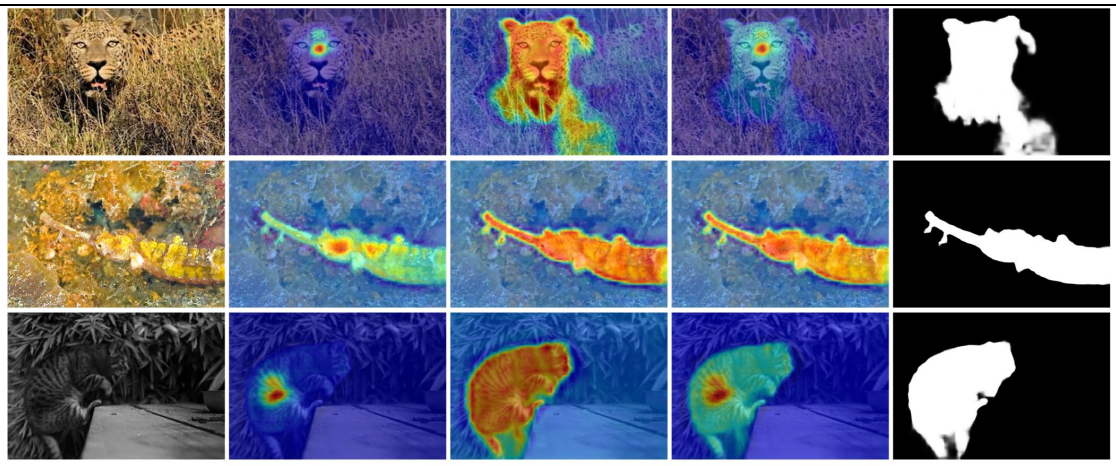

Fig.9. Visualization of MAGNet feature maps. (F_{EMM}: output by the EMM, F_{AFM}: output by the AFM, F_{fuse}: final fused camouflaged object map)

4.3.2 Qualitative Comparisons

We visualize the feature maps output by the EMM and AFM and compare them with the final fused camouflaged object map. The results are shown in Fig. 9. The feature map output by the EMM proves that this module focuses more on the center of a camouflaged object, while the AFM can retain more important information about the target itself. The fused output camouflage feature map combines the advantages of both

modules. The center of the camouflaged object is used as a key point to precisely find important information in the vicinity of the point, thus improving the accuracy of segmentation.

4.4 Comparison on In-house Military Camouflaged Object Dataset

Table 3 shows the experimental comparison results of the MAGNet method proposed in this paper and other methods on the military camouflaged object dataset built in-house; this dataset is not very challenging for various networks due to the small size of the dataset and mainly reflects the feature extraction ability and segmentation accuracy of the network itself. As seen from Table 3, MAGNet reaches the optimum in seven of the metrics and has the best comprehensive segmentation ability; in particular, the mean sensitivity (meanSen) is improved by 7.4% compared with the next-best method SINet-V2, which means that the MAGNet model has the lowest missing detection rate. Since each image contains camouflaged objects, the mean specificity (meanSpe) of each model is relatively high, while that of MAGNet is still 1% higher, which means that MAGNet has the lowest false positive rate at the same time. The balance of the missed detection rate and false positive rate represents the stability of the network model and is particularly important in practical military applications. Fig. 10 visualizes the results of the comparison experiments in this subsection on the in-house-built military camouflaged object dataset.

Table 3 Comparison results on the in-house military camouflaged object dataset. (⁂ : medical image segmentation method, ◊: saliency object detection method, ⋕: COD method).

Methods	Pub. 'Year	S_α	F_β^w	MAE	E_ϕ^{ad}	meanDic	meanIoU	meanSen	meanSpe
Unet++ ⁂	DLMIA '17	0.717	0.594	0.009	0.736	0.513	0.421	0.471	0.747
BASNet◊	CVPR '19	0.865	0.757	0.008	0.928	0.763	0.666	0.758	0.950
SCRN◊	ICCV '19	0.847	0.603	0.010	0.677	0.687	0.575	0.726	0.955
HarDNet ⁂	ICCV '2019	0.876	0.784	0.005	0.953	0.795	0.695	0.806	0.967
F3Net◊	AAAI '20	0.889	0.798	0.005	0.944	0.816	0.716	0.846	0.972
PraNet ⁂	MICCAI '20	0.887	0.781	0.006	0.915	0.802	0.696	0.834	0.977
GCPANet◊	AAAI '20	0.874	0.721	0.006	0.821	0.733	0.623	0.714	0.971
SINet-V1⋕	CVPR '20	0.876	0.800	0.005	0.965	0.810	0.706	0.842	0.977
SANet ⁂	MICCAI '21	0.804	0.647	0.010	0.853	0.673	0.563	0.720	0.917
CaraNet ⁂	arXiv '21	0.865	0.729	0.006	0.873	0.763	0.654	0.832	0.964
RankNet⋕	CVPR '21	0.847	0.693	0.008	0.825	0.737	0.622	0.840	0.960
PFNet⋕	CVPR '21	0.873	0.771	0.006	0.941	0.785	0.682	0.804	0.965
UACANet-L ⁂	ACM MM '21	0.880	0.823	0.004	0.963	0.817	0.715	0.853	0.979
SINet-V2⋕	TPAMI '21	0.884	0.788	0.004	0.926	0.806	0.699	0.843	0.982
MAGNet⋕	**Ours**	**0.924**	**0.864**	**0.003**	**0.946**	**0.868**	**0.779**	**0.917**	**0.992**

Fig.10. Visualization results for all algorithms on the in-house military camouflaged object dataset

4.5 Discussion

From the comparison with the latest methods in Section 4.2, we find that the results of several saliency object detection algorithms are unsatisfactory, which proves that it is not reasonable to apply saliency object detection algorithms to the detection of camouflaged objects. The results show that medical image segmentation methods can achieve better results in camouflaged object segmentation tasks because some medical image datasets (e.g., polyp datasets) have properties similar to those of camouflaged objects, i.e., inconspicuous edges and high integration with the surrounding environment [62-64]. Therefore, COD has a high potential application in the medical field. Fig. 11 shows the visualization results of MAGNET applied to polyp detection, where the

dataset used for the experiment is the Kvasir-SEG polyp dataset [62].

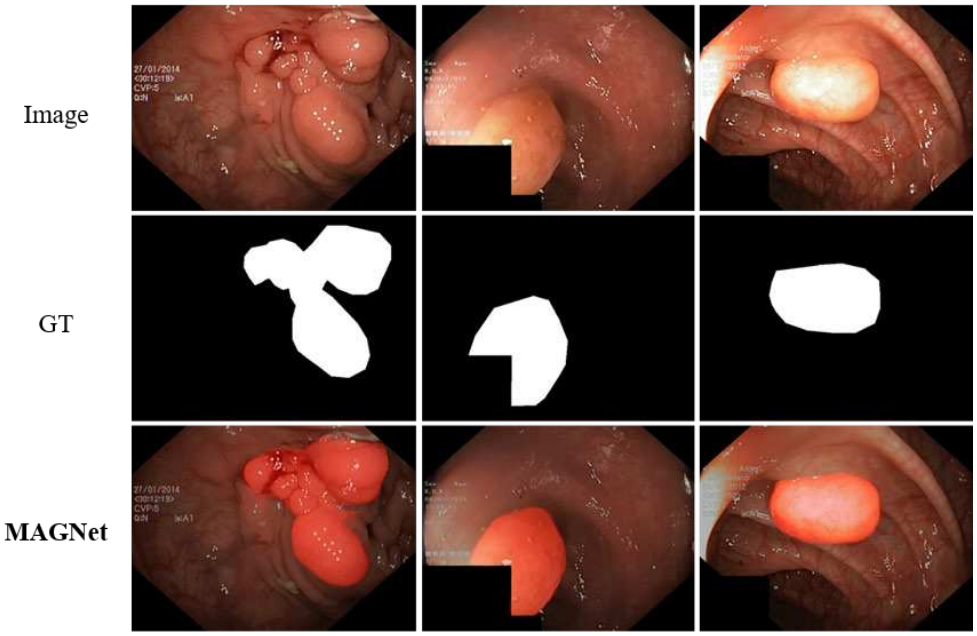

Fig.11. Visualization of detection results on the Kvasir-SEG polyp dataset

In addition, we explore other extended applications similar to COD. Fig. 12 shows the visualization results applied to defect detection in industry, where the dataset used is the magnetic tile defect dataset [65]. Fig. 13 shows the visualization results from applying MAGNET to infrared vehicle detection in rain and fog at night with the AAU-RainSnow dataset [66]. In these applications, similar to camouflaged objects, the object to be detected exhibits a high degree of fusion with the background, so the detection of camouflaged objects can be extended to similar applications.

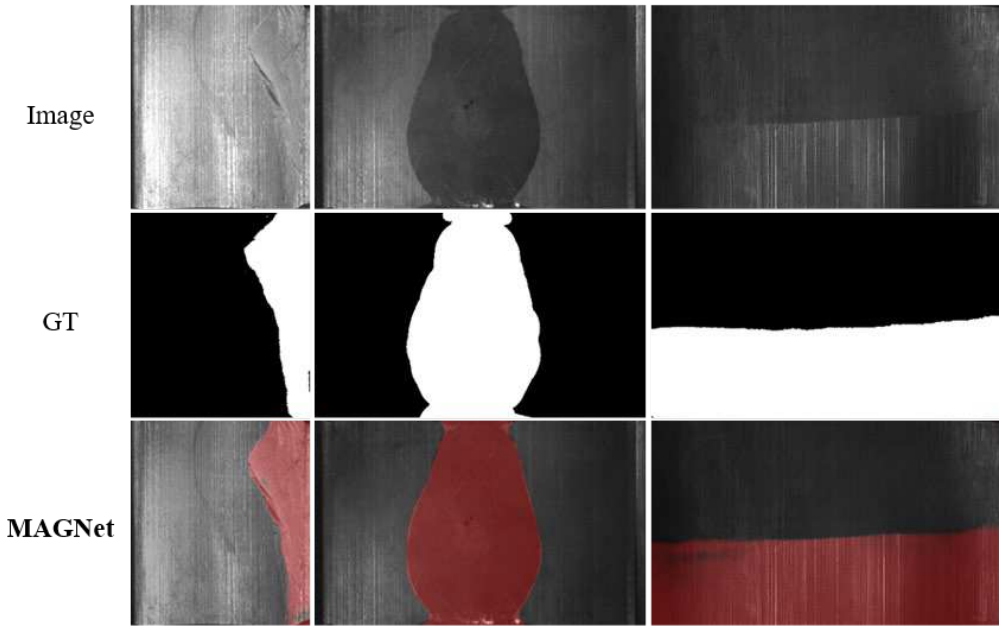

Fig.12. Visualization of the detection results on the magnetic tile defect dataset

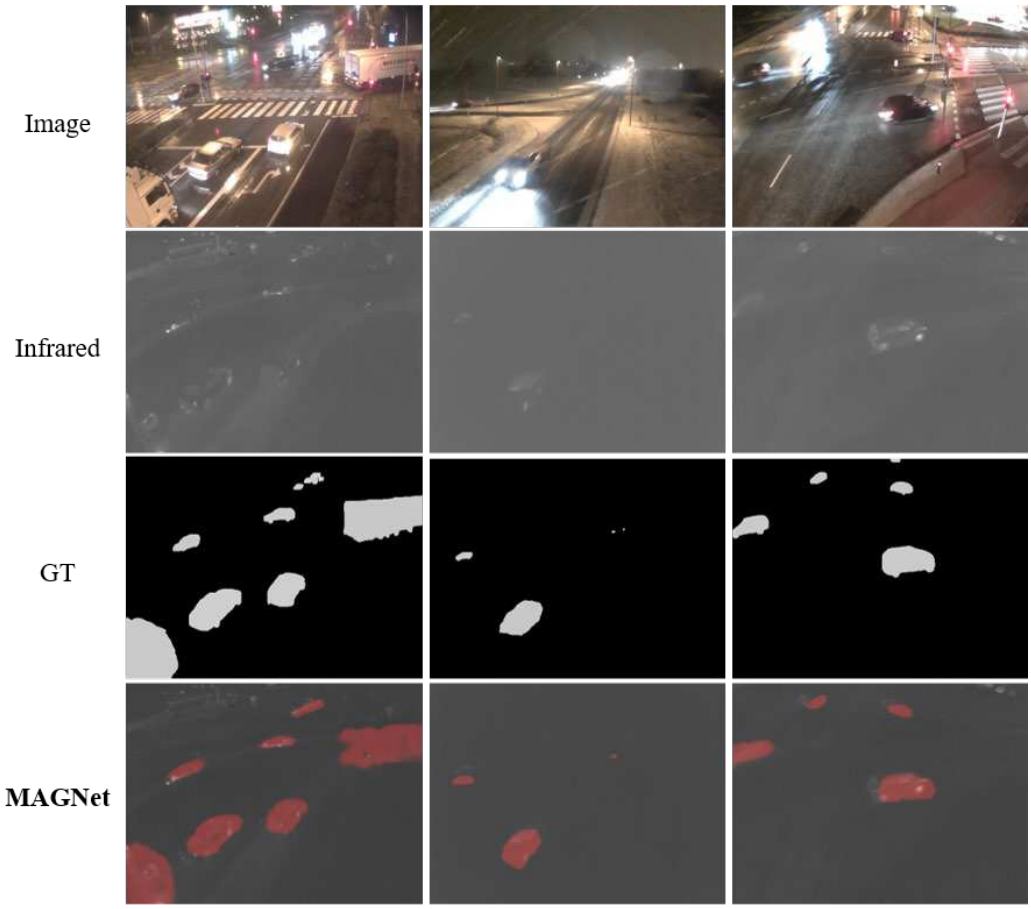

Fig.13. Visualization of detection results on the AAU-RainSnow dataset

5. Conclusion

This paper is dedicated to achieving more accurate detection of camouflaged objects. By simulating the search function of a magnifier, we propose a new network based on the observation effect of a magnifier named MAGNet. We design two bionic modules processed in parallel and propose a more applicable weighted key point area perception loss that allows the network to more fully exploit important information about an object, thus achieving an accurate search for camouflaged objects. The results demonstrate the accuracy advantages of MAGNet for COD through quantitative and qualitative evaluation on challenging public datasets and an in-house-built dataset. Additionally, MAGNet has potential value for application in other fields (e.g., medical image segmentation, nighttime vehicle detection, and industrial defect detection). In the future, we will continue to explore the accurate recognition of low-detectability objects.

Acknowledgements

The authors would like to acknowledge National Defense Science and Technology 173 Program Technical Field Fund Project (Grant No. 2021-JCJQ-JJ-0871) to provide fund for conducting experiments.

Declaration of competing interest

The authors declare that they have no conflicts of interest.

References

[1] Stevens M, Merilaita S. Animal camouflage: current issues and new perspectives. Philos Trans R Soc B Biol Sci 2009;364:423-7. https://doi.org/10.1098/rstb.2008.0217.

[2] Stuart-Fox D, Moussalli A, Whiting MJ. Predator-specific camouflage in chameleons. Biol Lett 2008;4:326-9. https://doi.org/10.1098/rsbl.2008.0173.

[3] Puzikova N, Uvarova E, Filyaev I, Yarovaya L. Principles of an approach coloring military camouflage. Fibre Chem 2008;40:155-9. https://doi.org/10.1007/s10692-008-9030-9.

[4] Li Y, Zhang D, Lee DJ. Automatic fabric defect detection with a wide-and-compact network. Neurocomputing 2019;329:329-38. https://doi.org/10.1016/j.neucom.2018.10.070.

[5] Zhang M, Li H, Pan S, Lyu J, Ling S, Su S. Convolutional neural networks-based lung nodule classification: a surrogate-assisted evolutionary algorithm for hyperparameter optimization. IEEE Trans Evol Comput 2021;25:869-82. https://doi.org/10.1109/TEVC.2021.3060833.

[6] Fan D-P, Ji G-P, Zhou T, Chen G, Fu H, Shen J, et al. PraNet: parallel reverse attention network for polyp segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI (2020). vol 12266. Springer International Publishing, Cham, 2020;pp: 263-73. https://doi.org/10.1007/978-3-030-59725-2_26.

[7] Zhou M, Li Y, Yuan H, Wang J, Pu Q. Indoor WLAN personnel intrusion detection using transfer learning-aided generative adversarial network with light-loaded database. Mob Netw Appl 2021;26:1024-42. https://doi.org/10.1007/s11036-020-01663-8.

[8] Jiang X, Cai W, Yang Z, Xu P, Jiang B. IARet: a lightweight multiscale infrared aerocraft recognition algorithm. Arab J Sci Eng 2022;47:2289-303. https://doi.org/10.1007/s13369-021-06181-7.

[9] Ghose D, Desai SM, Bhattacharya S, Chakraborty D, Fiterau M, Rahman T. Pedestrian detection in thermal images using saliency maps. In: IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW). Cornell University, Ithaca, NY, 2019;pp:988-97. https://doi.org/10.1109/CVPRW.2019.00130.

[10] Ding Y, Zhao X, Zhang Z, Cai W, Yang N, Zhan Y. Semi-supervised locality preserving dense graph neural network with ARMA filters and context-aware learning for hyperspectral image classification. IEEE Trans Geosci Remote Sens 2022;60:1-12. https://doi.org/10.1109/TGRS.2021.3100578.

[11] Zhang J, Zhang X, Li T, Zeng Y, Lv G, Nian F. Visible light polarization image desmogging via cycle convolutional neural network. Multimed Syst 2022;28:45-55. https://doi.org/10.1007/s00530-021-00802-9.

[12] Ding Y, Zhao X, Zhang Z, Cai W, Yang N. Graph sample and aggregate-attention network for hyperspectral image classification. IEEE Geosci Remote Sens Lett 2022;9:1-5. https://doi.org/10.1109/LGRS.2021.3062944.

[13] Fan DP, Ji GP, Sun G, Cheng MM, Shen J, Shao L. Camouflaged object detection. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Cornell University, Ithaca, NY, 2020;pp:2774-84. https://doi.org/10.1109/CVPR42600.2020.00285.

[14] Mei H, Ji GP, Wei Z, Yang X, Wei X, Fan DP. Camouflaged object segmentation with distraction mining. In: 2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Cornell University, Ithaca, NY, 2021;pp:8768-77. https://doi.org/10.1109/CVPR46437.2021.00866.

[15] Lv Y, Zhang J, Dai Y, Li A, Liu B, Barnes N, et al. Simultaneously localize, segment and rank the camouflaged objects. In: 2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Cornell University, Ithaca, NY, 2021;pp:11586-96. https://doi.org/10.1109/CVPR46437.2021.01142.

[16] Liang X, Lin H, Yang H, Xiao K, Quan J. Construction of semantic segmentation dataset of camouflage target image. Laser Optoelectron Prog 2021;58:0410015. https://doi.org/10.3788/LOP202158.0410015.

[17] Skurowski P, Abdulameer H, Błaszczyk J, Depta T, Kornacki A, Kozieł P. Animal camouflage analysis: chameleon database. Unpublished Manuscript. 2018. https://www.polsl.pl/rau6/chameleon-database-animal-camouflage-analysis/.

[18] Le T-N, Nguyen TV, Nie Z, Tran M-T, Sugimoto A. Anabranch network for camouflaged object segmentation. Comput Vis Image Underst 2019:184:45-56. https://doi.org/10.1016/j.cviu.2019.04.006.

[19] Caesar H, Bankiti V, Lang AH, Vora S, Liong VE, Xu Q, et al. nuScenes: a multimodal dataset for autonomous driving. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). IEEE, Seattle, WA, USA, 2020;pp:11618-28. https://doi.org/10.1109/cvpr42600.2020.01164.

[20] Chen YJ, Tu ZD, Kang D, Bao LC, Zhang Y, Zhe XF, et al. Model-based 3D hand reconstruction via self-supervised learning. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR). Cornell University, Ithca, NY, 2021;pp:10451-60. https://doi.org/10.1109/CVPR46437.2021.01031.

[21] An S, Che G, Guo J, Zhu H, Ye J, Zhou F, et al. ARShoe: real-time augmented reality shoe try-on system on smartphones. In: Proceedings of the 29th ACM international conference on multimedia. Association for Computing Machinery, New York, NY, 2021;pp:1111–9. https://doi.org/10.1145/3474085.3481537.

[22] Hou J, Graham B, Nießner M, Xie SN. Exploring data-efficient 3D scene understanding with contrastive scene

contexts. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR). Cornell University, Ithaca, NY, 2021;pp:15587-97. https://doi.org/10.1109/CVPR46437.2021.01533.

[23] Huang J, Wang H, Birdal T, Sung M, Arrigoni F, Hu SM, et al. MultiBodySync: multi-body segmentation and motion estimation via 3D scan synchronization. In: 2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Cornell University, Ithaca, NY, 2021;pp:7104-14. https://doi.org/10.1109/CVPR46437.2021.00703.

[24] Liu Z, Qi X, Fu CW. One thing one click: a self-training approach for weakly supervised 3D semantic segmentation. In: 2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Cornell University, Ithaca, NY, 2021;pp:1726-36. https://doi.org/10.1109/CVPR46437.2021.00177.

[25] Chen X, Yuan Y, Zeng G, Wang J. Semi-supervised semantic segmentation with cross pseudo supervision. In: 2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Cornell University, Ithaca, NY, 2021;pp:2613-22. https://doi.org/10.1109/CVPR46437.2021.00264.

[26] Yao Y, Chen T, Xie G-S, Zhang C, Shen F, Wu Q, et al. Non-salient region object mining for weakly supervised semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Cornell University, Ithaca, NY, 2021;pp:2623-32. https://doi.org/10.1109/CVPR46437.2021.00265.

[27] Fu Y, Yang LJ, Liu D, Huang TS, Shi H. CompFeat: comprehensive feature aggregation for video instance segmentation. arXiv preprint 2020. https://doi.org/10.48550/arXiv.2012.03400.

[28] Wei J, Hu Y, Zhang R, Li Z, Zhou SK, Cui S. Shallow attention network for polyp segmentation. In: Medical image computing and computer assisted intervention – MICCAI 2021. Springer International Publishing, Cham, 2021;pp:699-708. https://doi.org/10.1007/978-3-030-87193-2_66.

[29] Wang JF, Song L, Li ZM, Sun HB, Sun J, Zheng NN. End-to-end object detection with fully convolutional network. In: IEEE/CVF conference on computer vision and pattern recognition (CVPR). Cornell University, Ithaca, NY, 2021;pp:15849-58. https://doi.org/10.1109/CVPR46437.2021.01559.

[30] Patel K, Bur AM, Wang G. Enhanced U-Net: a feature enhancement network for polyp segmentation. In: 2021 18th conference on robots and vision (CRV). IEEE, Canada, 2021;pp:181-8. https://doi.org/10.1109/CRV52889.2021.00032.

[31] Fan H, Mei X, Prokhorov D, Ling H. RGB-D scene labeling with multimodal recurrent neural networks. In: 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW). IEEE, Honolulu, HI, USA, 2017;pp:203-11. https://doi.org/10.1109/CVPRW.2017.3.

[32] Liu K, Ye Z, Guo H, Cao D, Chen L, Wang FY. FISS GAN: a generative adversarial network for foggy image

semantic segmentation. IEEE/CAA J Autom Sin 2021;8:1428-39. https://doi.org/10.1109/JAS.2021.1004057.

[33] Tan W, Qin N, Ma L, Li Y, Du J, Cai G, et al. Toronto-3D: a large-scale mobile LiDAR dataset for semantic segmentation of urban roadways. In: 2020 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW). Cornell University, Ithaca, NY, 2020;pp:797-806. https://doi.org/10.1109/CVPRW50498.2020.00109.

[34] Dovesi PL, Poggi M, Andraghetti L, Martí M, Kjellström H, Pieropan A, et al. Real-time semantic stereo matching. In: 2020 IEEE international conference on robotics and automation (ICRA). IEEE, Paris, France, 2020;pp:10780-7. https://doi.org/10.1109/ICRA40945.2020.9196784.

[35] Gan W, Wong PK, Yu G, Zhao R, Vong CM. Light-weight network for real-time adaptive stereo depth estimation. Neurocomputing 2021;441:118-27. https://doi.org/10.1016/j.neucom.2021.02.014.

[36] Ahn E, Feng D, Kim J. A spatial guided self-supervised clustering network for medical image segmentation. In: Medical image computing and computer assisted intervention – MICCAI 2021. Springer International Publishing, Cham, 2021;pp:379-88. https://doi.org/10.1007/978-3-030-87193-2_36.

[37] Liu Z, Manh V, Yang X, Huang X, Lekadir K, Campello V, et al. Style curriculum learning for robust medical image segmentation. In: Medical image computing and computer assisted intervention – MICCAI 2021. Springer International Publishing, Cham, 2021;pp:451-60. https://doi.org/10.1007/978-3-030-87193-2_43.

[38] Hu X, Zeng D, Xu X, Shi Y. Semi-supervised contrastive learning for label-efficient medical image segmentation. In: Medical image computing and computer assisted intervention – MICCAI 2021. Springer International Publishing, Cham, 2021;pp:481-90. https://doi.org/10.1007/978-3-030-87196-3_45.

[39] Chen H, Li Y, Su D. Multi-modal fusion network with multi-scale multi-path and cross-modal interactions for RGB-D salient object detection. Pattern Recognit 2019;86:376-85. https://doi.org/10.1016/j.patcog.2018.08.007.

[40] Chen H, Li Y. Three-stream attention-aware network for RGB-D salient object detection. IEEE Trans Image Process 2019;28:2825-35. https://doi.org/10.1109/TIP.2019.2891104.

[41] Su J, Li J, Zhang Y, Xia C, Tian Y. Selectivity or invariance: boundary-aware salient object detection. In: 2019 IEEE/CVF international conference on computer vision (ICCV). Cornell University, Ithaca, NY, 2019;pp:3798-807. https://doi.org/10.1109/ICCV.2019.00390.

[42] Gao SH, Cheng MM, Zhao K, Zhang XY, Yang MH, Torr P. Res2Net: a new multi-scale backbone architecture. IEEE Trans Pattern Anal Mach Intell 2021;43:652-62. https://doi.org/10.1109/TPAMI.2019.2938758.

[43] Luo W, Li Y, Urtasun R, Zemel R. Understanding the effective receptive field in deep convolutional neural networks. In: Proceedings of the 30th international conference on neural information processing systems. Curran

Associates Inc., Barcelona, Spain, 2016;pp:4905–13. https://doi.org/10.48550/arXiv.1701.04128.

[44] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122. 2015. https://arxiv.org/abs/1511.07122.

[45] Zhu X, Cheng D, Zhang Z, Lin S, Dai J. An empirical study of spatial attention mechanisms in deep networks. In: 2019 IEEE/CVF international conference on computer vision (ICCV). Cornell University, Ithaca, NY, 2019;pp:6687-96. https://doi.org/10.1109/ICCV.2019.00679.

[46] Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. IEEE Trans Pattern Anal Mach Intell 2020;42:2011-23. https://doi.org/10.1109/TPAMI.2019.2913372.

[47] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. In: 2021 IEEE/CVF conference on computer vision and pattern recognition (CVPR). IEEE, Nashville, TN, 2021;pp:13708-17. https://doi.org/10.1109/CVPR46437.2021.01350.

[48] Wei J, Wang S. F³Net: fusion, feedback and focus for salient object detection. Proc AAAI Conf Artif Intell 2020;34:12321-8. https://doi.org/10.1609/aaai.v34i07.6916.

[49] Fan DP, Cheng MM, Liu Y, Li T, Borji A. Structure-measure: a new way to evaluate foreground maps. In: 2017 IEEE international conference on computer vision (ICCV). Cornell University, Ithaca, NY, 2017;pp:4558-67. https://doi.org/10.1109/ICCV.2017.487.

[50] Margolin R, Zelnik-Manor L, Tal A. How to evaluate foreground maps. In: 2014 IEEE conference on computer vision and pattern recognition. Cornell University, Ithaca, NY, 2014;pp:248-55. https://doi.org/10.1109/CVPR.2014.39.

[51] Perazzi F, Krähenbühl P, Pritch Y, Hornung A. Saliency filters: contrast based filtering for salient region detection. In: 2012 IEEE conference on computer vision and pattern recognition. IEEE, Providence, RI, 2012;pp:733-40. https://doi.org/10.1109/CVPR.2012.6247743.

[52] Fan DP, Gong C, Cao Y, Ren B, Cheng MM, Borji A. Enhanced-alignment measure for binary foreground map evaluation. In: Proceedings of the twenty-seventh international joint conference on artificial intelligence. Cornell University, Ithaca, NY, 2018;pp:698-704. https://doi.org/10.24963/ijcai.2018/97.

[53] Milletari F, Navab N, Ahmadi S. V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). IEEE, Stanford, CA, USA, 2016;pp:565-71. https://doi.org/10.1109/3DV.2016.79.

[54] Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. UNet++: redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Imaging 2020;39:1856-67.

https://doi.org/10.1109/TMI.2019.2959609.

[55] Chao P, Kao CY, Ruan Y, Huang CH, Lin YL. HarDNet: a low memory traffic network. In: 2019 IEEE/CVF international conference on computer vision (ICCV). IEEE, Seoul, Korea, 2019;pp:3551-60. https://doi.org/10.1109/ICCV.2019.00365.

[56] Lou AG, Guan SY, Loew M. CaraNet: context axial reverse attention network for segmentation of small medical objects. arXiv preprint arXiv:210807368. 2021. https://arxiv.org/abs/2108.07368.

[57] Kim T, Lee H, Kim D. UACANet: uncertainty augmented context attention for polyp segmentation. In: Proceedings of the 29th ACM international conference on multimedia. Association for Computing Machinery, New York, NY, 2021;pp:2167–75. https://doi.org/10.1145/3474085.3475375.

[58] Qin X, Zhang Z, Huang C, Gao C, Dehghan M, Jagersand M. BASNet: boundary-aware salient object detection. In: 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR). IEEE, Long Beach, CA, 2019;pp:7471-81. https://doi.org/10.1109/CVPR.2019.00766.

[59] Wu Z, Su L, Huang Q. Stacked cross refinement network for edge-aware salient object detection. In: 2019 IEEE/CVF international conference on computer vision (ICCV). IEEE, Seoul, Korea, 2019;pp:7263-72. https://doi.org/10.1109/ICCV.2019.00736.

[60] Chen Z, Xu Q, Cong R, Huang Q. Global context-aware progressive aggregation network for salient object detection. Proc AAAI Conf Artif Intell 2020;34:10599-606. https://doi.org/10.1609/aaai.v34i07.6633.

[61] Fan DP, Ji GP, Cheng MM, Shao L. Concealed object detection. IEEE Trans Pattern Anal Mach Intell. 2021. https://doi.org/10.1109/TPAMI.2021.3085766.

[62] Jha D, Smedsrud PH, Riegler MA, Halvorsen P, de Lange T, Johansen D, et al. Kvasir-SEG: a segmented polyp dataset. In: MultiMedia modeling. Springer International Publishing, Cham, 2020;pp:451-62. https://doi.org/10.1007/978-3-030-37734-2_37.

[63] Vázquez D, Bernal J, Sánchez FJ, Fernández-Esparrach G, López AM, Romero A et al. A benchmark for endoluminal scene segmentation of colonoscopy images. J Healthc Eng 2017;2017:1-9. https://doi.org/10.1155/2017/4037190.

[64] Tajbakhsh N, Gurudu SR, Liang J. Automatic polyp detection in colonoscopy videos using an ensemble of convolutional neural networks. In: 2015 IEEE 12th international symposium on biomedical imaging (ISBI). IEEE, Brooklyn, NY, USA, 2015;pp:79-83. https://doi.org/10.1109/ISBI.2015.7163821.

[65] Bahnsen CH, Moeslund TB. Rain removal in traffic surveillance: does it matter? IEEE Trans Intell Transp Syst 2019;20:2802-19. https://doi.org/10.1109/TITS.2018.2872502.

[66] Huang Y, Qiu C, Yuan K. Surface defect saliency of magnetic tile. Vis Comput 2020;36:85-96. https://doi.org/10.1007/s00371-018-1588-5.