MAGNet: A Camouflage Object Detection Network Simulating The Observation Effect of Magnifier

Xinhao Jiang (✉ jiangxinhao2020@outlook.com)

Xi'an High-Tech Institute <https://orcid.org/0000-0002-4239-1738>

Wei Cai

Xi'an High-Tech Institute

Bo Jiang

Xi'an High-Tech Institute

Zhiyong Yang

Xi'an High-Tech Institute

Xin Wang

Xi'an High-Tech Institute

Research Article

Keywords: Camouflage object detection, Image segmentation, Deep learning, Human visual system, Computer Vision

Posted Date: December 1st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1020529/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

In recent years, protecting important objects by simulating animal camouflage has been widely used in many fields. Therefore, the Camouflaged Object Detection (COD) technology has emerged. COD is more difficult than traditional target detection techniques because of the high degree of fusion of camouflaged objects with the background. In this paper, we strive to identify camouflaged objects more accurately. Inspired by humans using a magnifier to search for hidden objects in pictures, we propose a COD network that simulates the observation effect of a magnifier, termed Magnifier Network (MAGNet). Specifically, our MAGNet contains two parallel modules, i.e., Ergodic Magnify module (EMM) and Attention Focus module (AFM). The EMM is designed to mimic the magnifying process of a magnifier ergodizing an image, and the AFM is used to perform the observation process in which human attention is highly focused for focusing on a region. The two sets output camouflaged object maps are merged to achieve the effect of simulating the observation of the object by a magnifier. Extensive experiments demonstrate that compared with 14 cutting-edge detection models, the MAGNet can achieve the best comprehensive effect of 8 evaluation indicators on the public COD dataset, and the segmentation accuracy is significantly improved.

1 Introduction

Can you accurately find the tank in Fig. 1 in seconds? With the influence of camouflage coating, external camouflage materials, smoke barriers and ground object shielding, the tank in Fig. 1 achieves near-perfect integration with the background. If you notice it first, you win; if it spots you first, you lose more than just the victory.

In nature, animals follow the principle of survival of the fittest. They can camouflage their own shape characteristics, or retain the shape characteristics similar to their habitat, so as to avoid being preyed on by attackers and better ambush their prey [1]. The former, such as chameleons, can adapt their skin's coloring to match with the external environment [2], while the latter, such as white moth and black moth, have different survival numbers in different habitats.

Currently, with the progress of science and technology, visible light camouflage technology simulating animal camouflage has been widely used in high-tech war, such as continuously improved military camouflage clothing and camouflage net [3]. After using camouflage technology, snipers can ambush the enemy's senior generals, and tanks and armored vehicles can deceive the enemy's visible light reconnaissance. Therefore, the research of accurate visible light camouflage target segmentation is of great significance in the military field.

In addition to its military value, COD can also be applied to industrial detection (e.g., equipment defect detection [4]), medical diagnosis (e.g., testing whether the lung is infected by pneumonia virus [5, 6]), monitoring and protection (e.g., suspicious person or UAV intrusion detection [7, 8]) and unmanned driving (e.g., road obstacle detection [9]).

However, visible camouflage target segmentation is not well-studied at present. For example, in the military field, military camouflage targets are often identified by means of infrared, polarization, hyperspectral and other technologies [10–12]. Although the limitations of visible light target identification are reduced to a certain extent, the scientific problem of how to accurately segment camouflage targets in the visible light band is ignored.

In this paper, we propose a visible light camouflage target segmentation network based on the observation effect of magnifier and call it MAGNet. Fig. 1 is a schematic diagram of searching for military camouflage targets based on the observation effect of magnifier. From the Fig. 1, it can be found that it is simple and effective to observe the camouflage targets in the picture with magnifier. Firstly, the magnifier magnifies the observation area visually, and then we will be attracted by the edge information and key parts of camouflage targets in the enlarged area, in order to focus on the key points to accurately identify the camouflage targets in the region.

In a nutshell, the major contributions of this paper contain threefold:

1. We apply the observation effect of magnifier to the COD problem, and propose a novel camouflage target segmentation network which is called MAGNet.
2. We design a parallel structure of Ergodic Magnify module and Attention Focus module to simulate the function of magnifier.
3. We perform extensive experiments using public COD benchmark datasets. MAGNet has the best comprehensive effect of 8 evaluation indicators compared with 14 cutting-edge detection models.

This paper is organized as follows: The previous work similar to this study is introduced in Section 2. Section 3 provide detailed descriptions for our MAGNet and the associated modules. In Section 4, shows the comparative experiments made in this paper and analyzes the experimental results with quantitative and qualitative evaluation. Finally, Section 5 concludes the paper.

2 Related Works

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

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

Semantic Segmentation Based on Deep Learning: In recent years, scene understanding technologies such as autonomous driving [19], virtual reality [20] and augmented reality [21] are becoming more and more important for social development. As the basic task of scene understanding, semantic segmentation technology based on pixel by pixel classification has been widely studied [22-24]. A great deal of semantic segmentation based on deep learning has been proposed [25-28]. Currently, there are four main types of networks, namely, the Full Convolutional Network (FCN) [29], the Convolutional Neural Network (CNN) [30], the Recurrent Neural Network (RNN) [31], and the Generative Adversarial Network (GAN) [32].

Salient Object Detection Based on Deep Learning: In contrast to camouflaged objects, salient objects are the most noticeable objects in the 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, the salient object detection based on deep learning has been improved mainly by multi-scale feature fusion [39], attention mechanism [40] and edge information [41].

3 Magnet Detection Model

A magnifier can help the observer quickly find the camouflaged object from the image. This is because the magnifying effect of the magnifier makes it easier for the observer to spot the center, key points and tiny details of the camouflaged object. Inspired by the magnifier, we apply the observation effect to the COD problem, design the Ergodic Magnify module and the Attention Focus module. The ergodic magnification module is designed to mimic the magnifying process of a magnifier ergodicizing an image, and the attention focusing module is used to perform the observation process which human attention is highly focused for focusing on a region.

3.1 Network Overview

The network structure of MAGNet is shown in Figure 2. Input a camouflage object image into this network, MAGNet first extracts multi-scale feature maps through Res2net-50 backbone [42], and then inputs the latter three feature maps to the Ergodic Magnify module and the Attention Focus module respectively. Finally, the output feature maps of the two modules are fused to simulate the observation effect of magnifier on the object.

3.2 Ergodic Magnify Module (EMM)

As shown in Figure 2, the Ergodic Magnify module consists of two parts, i.e., the Central Excitation Module (CEM) and the Multi-scale Feature Fusion Module (MFFM).

The Central Excitation Module is used to traverse the feature maps of different scales output from the back three layers of the backbone to expand the receptive field and stimulate the central point and key points.

The Multi-scale Feature Fusion Module is designed to fully integrate the multi-scale feature maps after the Central Excitation Module to realize the efficient utilization of high-level and low-level features.

3.2.1 Central Excitation Module (CEM)

We find that when we use a magnifier to observe an object, we will observe the central area of the magnifier more carefully than the edge area. This is because the human visual receptive field mechanism determines that the observer will be more attracted to the center of the object [43]. Then, we will use the magnifier to traverse the whole picture until the center of the magnifier coincides with the center of the object.

In order to simulate the visual magnification and traversal function of the magnifier. We design a simple and efficient Central Excitation Module, as shown in Figure 3. The realization of the above functions mainly depends on the dilated convolution (DConv) with different sizes convolution kernels [44].

Specifically, the Central Excitation Module includes four branches. Input the feature map into four branches at the same time. The four branches first use 1×1 convolution to change the number of output channels, and then, three of them use 3×3 , 5×5 , and 7×7 dilated convolution with expansion coefficients of 2. After connecting the three sets of output feature maps, a 3×3 convolutional layer is used for fusion between channels. Finally, the residual connection is made with the fourth branch to obtain the feature map after the center excitation.

The connection of the three sets of dilated convolution can increase the importance of the central feature while increasing the receptive field. As shown in Figure 4, the function of central excitation can be realized. The multi-scale feature maps after central excitation have the same number of 128 channels to ensure the balanced utilization of information of each scale.

3.2.2 Multi-scale Feature Fusion Module(MFFM)

The function of the Multi-scale Feature Fusion Module is to fully integrate the excitation feature maps of different scales, thereby outputting a camouflage object map that fully contains high and low-level features. The MFFM structure diagram is shown in Figure 5. The small-scale excitation feature map transmits the feature information to the large-scale feature map through continuous upsample and fusion, and then generates an output feature map with a size of $44 \times 44 \times 1$.

The front-end fusion method of the module adopts Hadamard product ($\odot$). The Hadamard product calculation method is pixel-by-pixel multiplication, which can better achieve feature crossover, so as to eliminate the difference between the two groups of features and improve the ability of feature fusion.

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

The module output map is denoted as F_{out} , the large-scale feature map in the module is denoted as F_i , and the small-scale feature map is denoted as F_{i-1} . In the Figure 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 there is the following formula:

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

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

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

3.3 Attention Focus module

The attention focus module has two steps. First, through upsample and convolution operations, the three sets of feature maps output by Backbone are processed into feature maps of the same size and the same number of channels. Then input it into the Channel-

Spatial Attention Module to simulate the effect of human attention being focused on observing objects in the field of view of the magnifier.

3.3.1 Channel-Spatial Attention Module (CSAM)

The attention mechanism in deep learning simulates the human visual attention mechanism, and the goal is to obtain more important information [43]. It is mainly divided into two types: spatial attention mechanism and channel attention mechanism. The spatial attention mechanism can find the most important area in the space, and retain the important local information through spatial transformation. The channel attention mechanism 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 their own advantages and disadvantages, and the Channel-Spatial Attention Module we proposed is a parallel fusion mechanism of spatial attention and channel attention, as shown in Figure 6.

As illustrated by Figure 6, the Channel-Spatial Attention Module is mainly implemented in four steps. The pseudocode of the Channel-Spatial Attention Module 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 = CModule(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: Superimpose the three groups of input feature maps with the same size and number of channels on the channel dimension, so as to make average use of the feature maps of each scale and fully integrate the semantic information of high-level and low-level features. Then, input the feature maps of three different layers into the channel attention mechanism branch and spatial attention mechanism branch respectively to generate a channel attention map and a spatial attention map.

Channel Attention: 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] that can fully perceive position information into SAFM. The CA module first performs coordinate information embedding, uses 2D-average-pooling operations to aggregate the input features into a pair of direction-aware

feature maps. Then, the CA module performs coordinated attention generation. The first step is to process the direction-aware feature maps by the convolution layer. The second step is to segment and encode it into two attention maps that store position information. Finally, the feature maps are multiplied through Hadamard product to generate a channel attention map embedded with position and direction information.

Spatial Attention: The spatial attention mechanism is particularly important for finding special targets, and it can retain important local information. We first use GroupNorm (GN) for group normalization to eliminate the hardware platform's restrictions on BatchNorm. 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 ability of the feature map. The third step uses sigmoid to activate, and then multiplies 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 Hadamard product for the fusion of attention maps, that is, the method of pixel-by-pixel multiplication, which can obtain a more accurate feature map.

3.4 Output Prediction

Finally, the feature maps output by EMM and AFM are transformed into single-channel camouflaged object map through upsample operation. The two feature maps are fused by adding pixel by pixel. We select weighted BCE loss and weighted IOU loss [48] as the loss function. The overall loss function is:

$$L_{overall} = L(P_{EMM}, GT) + L(P_{AFM}, GT) \quad (4)$$

$$L(P, GT) = L_{wbce}(P, GT) + L_{wioU}(P, GT) \quad (5)$$

Where P_{EMM} and P_{AFM} are camouflaged object maps after upsample operation by the Ergodic Magnify module and Attention Focus module, and GT is the truth map.

4 Experimental Results And Analysis

4.1 Preparation Work

The experimental platform system is Win10, The GPU of the platform is NVIDIA Quadro GV100, video memory is 32G. The CPU is Inter Xeon Silver 4210. The experiment uses the Pytorch deep learning development framework. Computing platform is CUDA11.0. We use 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 Datasets Preprocessing

We select CAMO [18] and COD10K [14] datasets with relatively large data volume for evaluation. CAMO has 1250 images and COD10K has 5066 camouflage images. A total of 6316 images are combined, which are divided into training set, validation set and testing set according to the ratio of 6:4:4.

4.1.2 Evaluation Metrics

At present, there are many evaluation metrics suitable for camouflage target detection, and each metric focuses on different points. Based on previous scholars' research, we selected 8 evaluation metrics. Their brief introduction is as follows: Structure measure (S_α) is a structural similarity evaluation metric, focusing on evaluating the structural information of the prediction map [49]. Weighted F-measure (F_β^w) is a comprehensive evaluation method for the accuracy and recall rate of the prediction map [50]. Mean absolute error (MAE) calculates the sum of the absolute values of the differences between the pixels of the prediction map and the truth map [51]. The adaptive Enhanced alignment measure (E_ϕ^{ad}) can evaluate the pixel-level similarity effect and also obtain image-level statistics [52]. Mean dice coefficient (meanDic) is calculated for the percentage of the correctly segmented area to the true area in the GT image [53]. mean Intersection over Union (meanIOU) calculates the ratio of the area of overlap and concatenation between the predicted and ground truth maps. Mean Sensitivity (meanSen) measures the percentage of results that are actually correct in the GT image that are predicted to be correct. Mean Specificity (meanSpe) measures the percentage of results that are actually incorrect in the GT image that are predicted to be incorrect.

4.1.3 Compared Methods

To prove the effectiveness of the MAGNet proposed in this paper, we compared 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], SCRNet [59], F3Net [48] and GCPANet [60]; camouflaged object segmentation method 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 proposed 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 the State-of-the-art Algorithms

Quantitative Comparison: Table 1 comprehensively reports the quantitative comparison results of MAGNet with the latest algorithms on the combined dataset. As can be seen from the table, the MAGNet proposed in this paper has the best comprehensive performance under the eight standard evaluation metrics, especially achieving the leading performance in α , meanDic and meanIOU metrics. MAGNet is basically equal to SINet-V2 in this metric of meanSpe. The reason why this metric is not optimal is that MAGNet has a powerful feature extraction ability of camouflaged targets, which is easy to cause a certain degree of false alarm.

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

Qualitative Comparisons: Figure 8 comprehensively shows the visualization results of all the algorithms in the comparison experiments. It can be observed that the MAGNet proposed in this paper is able to segment the camouflaged targets more accurately. Ergodic Magnify Module can better identify small targets hidden in complex backgrounds (e.g., the fourth column) by magnifying the receptive field and fusing multi-scale features. Attention Focus module can acquire more important information on channel and space by simulating human visual attention mechanism, so that it can accurately segment the detailed areas of camouflaged objects (e.g., the seventh column).

4.3 Ablation Experiment

We conducted ablation experiments to verify the effectiveness of two specific modules designed for camouflage target detection, namely Ergodic Magnify Module and Attention Focus module.

Quantitative Comparison: The results of the MAGNet ablation experiments are reported comprehensively in Table 2. It can be noticed that adding the two modules alone improves the model performance significantly. Among them, adding Attention Focus module makes the meanSen optimal, which is due to the effect of the attention mechanism of the model, which reduces the probability of missing detection of the model. The addition of the Ergodic Magnify Module makes the meanSpe optimal, because the model's receptive field magnifying mechanism works to reduce the model's false alarm probability. We also compared the results of connecting two key modules in series and parallel, and ultimately found that the parallel structure better maximized the effect of both modules.

Qualitative Comparisons: We visualized the feature maps output by Ergodic Magnify Module and Attention Focus Module, and compared them with the final fused camouflaged object map. The results are shown in Figure 9. The feature map output by Ergodic Magnify Module proves that this module focuses more on the center of the camouflaged object, while Attention Focus Module 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 camouflage target is used as a key point to precisely find important information in the vicinity, thus improving the accuracy of segmentation.

4.4 Discussion

From the comparison with the latest methods in Section 4.2, we can find that the results of several saliency object detection algorithms are unsatisfactory, which proves that it is not reasonable to apply the SOD algorithm 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) also have similar properties to camouflaged objects, i.e., inconspicuous edges and high integration with the surrounding environment [62–64]. Therefore, camouflaged object detection has a high potential application in the medical field.

In addition, for the military application problem proposed in the beginning of this paper, we have done experiments on the self-constructed dataset and achieved better results. The visualization of the experimental results is shown in Figure 10.

5 Conclusion

This paper is dedicated to achieve more accurate detection of camouflaged objects. By simulating the search function of the magnifier, this paper proposes a new network based on the observation effect of the magnifier, named MAGNet. We designed two bionic modules processed in parallel, allowing the network to more fully exploit important information about the object, thus achieving the purpose of accurate search for camouflaged objects. The results demonstrate the accuracy advantage of MAGNet for camouflaged object detection through quantitative and qualitative evaluation on challenging public datasets. Also, MAGNet has potential value for application to other fields such as medical image segmentation. In the future, we will continue to explore the accurate recognition of low-detectable targets.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Stevens M, Merilaita S (2009) Animal camouflage: current issues and new perspectives. *Philosophical Transactions of The Royal Society B Biological Sciences* 364(1516):423–427. <https://doi.org/10.1098/rstb.2008.0217>
2. Stuart-Fox D, Moussalli A, Whiting MJ (2008) Predator-specific camouflage in chameleons. *Biol Lett* 4(4):326–329. <https://doi.org/10.1098/rsbl.2008.0173>
3. Puzikova NP, Uvarova EV, Filyaev IM, Yarovaya LA (2008) Principles of an approach for coloring military camouflage. *Fibre Chem* 40(2):155–159. <https://doi.org/10.1007/s10692-008-9030-9>
4. Li Y, Zhang D, Lee DJ (2018) Automatic fabric defect detection with a wide-and-compact network. *Neurocomputing* 329:329–338. <https://doi.org/10.1016/j.neucom.2018.10.070>
5. Zhang M, Li H, Pan S, Lyu J, Ling S, Su S (2021) Convolutional Neural Networks based Lung Nodule Classification: A Surrogate-Assisted Evolutionary Algorithm for Hyperparameter Optimization. *IEEE Transactions on Evolutionary Computation* PP 991–1. <https://doi.org/10.1109/TEVC.2021.3060833>
6. Fan DP, Ji GP, Zhou T, Chen G, Fu H, Shen J, Shao L (2020) Pranet: Parallel reverse attention network for polyp segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp 263–273. https://doi.org/10.1007/978-3-030-59725-2_26
7. Zhou M, Li Y, Yuan H, Wang J, Pu Q (2020) Indoor WLAN Personnel Intrusion Detection Using Transfer Learning-Aided Generative Adversarial Network with Light-Loaded Database. *Mobile Networks and Applications* 26:1024–1042. <https://doi.org/10.1007/s11036-020-01663-8>
8. Jiang XH, Cai W, Yang ZY, Xu PW, Jiang B (2021) IARet: A Lightweight Multiscale Infrared Aircraft Recognition Algorithm. *Arab J Sci Eng*. <https://doi.org/10.1007/s13369-021-06181-7>
9. Ghose D, Desai SM, Bhattacharya S, Chakraborty D, Fiterau M, Rahman T (2019) Pedestrian Detection in Thermal Images Using Saliency Maps. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 988–997. <http://arxiv.org/abs/1904.06859>
10. Ding Y, Zhao XF, Zhang ZL, Cai W, Yang NJ, Zhan Y (2021) Semi-Supervised Locality Preserving Dense Graph Neural Network With ARMA Filters and Context-Aware Learning for Hyperspectral Image Classification. *IEEE Trans Geosci Remote Sens*. <https://doi.org/10.1109/TGRS.2021.3100578>
11. Zhang J, Zhang X, Li T, Zeng Y, Lv G, Nian F (2021) Visible light polarization image desmogging via cycle convolutional neural network. *Multimedia Syst*. <https://doi.org/10.1007/s00530-021-00802-9>
12. Ding Y, Zhao XF, Zhang ZL, Cai W, Yang NJ (2021) Graph Sample and Aggregate-Attention Network for Hyperspectral Image Classification. *IEEE Geosci Remote Sens Lett*. <https://doi.org/10.1109/LGRS.2021.3062944>
13. Fan DP, Ji GP, Sun GL, Cheng MM, Shen JB, Shao L (2020) Camouflaged object detection. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* pp: 2774–2784. <https://doi.org/10.1109/CVPR42600.2020.00285>

14. Mei HY, Ji GP, Wei ZQ, Yang X, Wei XP, Fan DP (2021) Camouflaged Object Segmentation with Distraction Mining. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). <https://arxiv.org/abs/2104.10475>
15. Lyu YQ, Zhang J, Dai YC, Li AX, Liu BW, Barnes N, Fan DP (2021) Simultaneously Localize, Segment and Rank the Camouflaged Objects. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). <https://arxiv.org/abs/2103.04011>
16. Liang XY, Lin HK, Yang H, Xiao KH, Quan JC (2021) Construction of Semantic Segmentation Dataset of Camouflage Target Image. *Laser & Optoelectronics Progress* 58(4):0410015
17. Skurowski P, Abdulameer H, Błaszczuk J, Depta T, Kornacki A, Kozieł P (2018) Animal camouflage analysis: Chameleon database. Unpublished Manuscript
18. Le TN, Nguyen TV, Nie Z, Tran MT, Sugimoto A (2019) Anabranh network for camouflaged object segmentation. *Comput Vis Image Underst* 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, Krishnan A, Pan Y, Baldan G, Beijbom O (2020) nuScenes: A Multimodal Dataset for Autonomous Driving. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp: 11618-11628 <https://doi.org/10.1109/CVPR42600.2020.01164>
20. Chen YJ, Tu ZD, Kang D, Bao LC, Zhang Y, Zhe XF, Chen RZ, Yuan JS (2021) Model-based 3D Hand Reconstruction via Self-Supervised Learning. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). <http://arxiv.org/abs/2103.11703>
21. An S, Che GF, Guo JH, Zhu HG, Ye JJ, Zhou FR, Zhu ZQ, Wei D, Liu AS, Zhang W (2021) ARShoe: Real-Time Augmented Reality Shoe Try-on System on Smartphones. *Proceedings of the 29th ACM International Conference on Multimedia* 1111-1119. <https://doi.org/10.1145/3474085.3481537>
22. Hou Ji, Graham B, Nießner M, Xie SN (2021) Exploring Data-Efficient 3D Scene Understanding with Contrastive Scene Contexts. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). <http://arxiv.org/abs/2012.09165>
23. Huang JH, Wang HQ, Birdal T, Sung M, Arrigoni F, Hu SM, Leonidas JG (2021) MultiBodySync: Multi-Body Segmentation and Motion Estimation via 3D Scan Synchronization. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). <https://arxiv.org/abs/2101.06605>
24. Liu ZZ, Qi XJ, Fu CW (2021) One Thing One Click: A Self-Training Approach for Weakly Supervised 3D Semantic Segmentation. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). <https://arxiv.org/abs/2104.02246>
25. Chen XK, Yuan YH, Zeng G, Wang JD (2021) Semi-Supervised Semantic Segmentation with Cross Pseudo Supervision. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). <https://arxiv.org/abs/2106.01226>
26. Yao YZ, Chen T, Xie GS, Zhang CY, Shen FM, Wu Q, Tang ZM, Zhang J (2021) Non-Salient Region Object Mining for Weakly Supervised Semantic Segmentation. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). <https://arxiv.org/abs/2103.14581>
27. Fu Y, Yang LJ, Liu D, Huang TS, Shi H, Intelligence (2021) <https://arxiv.org/abs/2012.03400>
28. Wei J, Hu YW, Zhang RM, Li Z, Zhou SK, Cui SG (2021) Shallow Attention Network for Polyp Segmentation. *MICCAI* (2021). <https://arxiv.org/abs/2108.00882>
29. Wang JF, Song L, Li ZM, Sun HB, Sun J, Zheng NN (2021) End-to-End Object Detection with Fully Convolutional Network. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). <https://arxiv.org/abs/2012.03544>
30. Patel K, Bur AM, Wang G (2021) Enhanced U-Net: A Feature Enhancement Network for Polyp Segmentation. *2021 18th Conference on Robots and Vision (CRV)* 181-188. <https://arxiv.org/abs/2105.00999>
31. Fan H, Mei X, Prokhorov D, Ling HB (2017) RGB-D Scene Labeling with Multimodal Recurrent Neural Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* pp:203-211. <https://doi.org/10.1109/CVPRW.2017.31>
32. Liu KH, Ye ZH, Guo HY, Cao DP, Chen L, Wang FY (2021) FISS GAN: A Generative Adversarial Network for Foggy Image Semantic Segmentation. *IEEE/CAA Journal of Automatica Sinica* 8(8):1428–1439. <https://doi.org/10.1109/JAS.2021.1004057>
33. Tan WK, Qin NN, Ma LF, Li Y, Du J, Cai GR, Yang K, Li J (2020) Toronto-3D: A Large-scale Mobile LiDAR Dataset for Semantic Segmentation of Urban Roadways. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* pp:797-806. <http://arxiv.org/abs/2003.08284v3>
34. Dovesi PL, Poggi M, Andraghetti L, Marti M, Mattoccia S (2020) Real-Time Semantic Stereo Matching. *2020 IEEE International Conference on Robotics and Automation (ICRA)* 10780-10787. <https://arxiv.org/abs/1910.00541v2>

35. Gan W, Wong PK, Yu G, Zhao R, Chi MV (2021) Light-weight network for real-time adaptive stereo depth estimation. *Neurocomputing* 441(9):118–127. <https://doi.org/10.1016/j.neucom.2021.02.014>
36. Ahn E, Feng DG, Kim JM (2021) A Spatial Guided Self-supervised Clustering Network for Medical Image Segmentation. *Medical Image Computing and Computer Assisted Interventions (MICCAI)* 2021. <https://arxiv.org/abs/2107.04934>
37. Liu Z, Manh V, Yang X, Huang X, Lekadir K, Campello VM, Ravikumar N, Frangi AF, Ni D (2021) Style Curriculum Learning for Robust Medical Image Segmentation. *Medical Image Computing and Computer Assisted Interventions (MICCAI)* 2021. <https://arxiv.org/abs/2108.00402>
38. Hu XR, Zeng DW, Xu XW, Shi YY (2021) Semi-supervised Contrastive Learning for Label-efficient Medical Image Segmentation. *Medical Image Computing and Computer Assisted Interventions (MICCAI)* 2021. <https://arxiv.org/abs/2109.07407>
39. Chen H, Li Y, Su D (2019) Multi-modal fusion network with multi-scale multi-path and cross-modal interactions for rgb-d salient object detection. *Pattern Recogn* 86:376–385. <https://doi.org/10.1016/j.patcog.2018.08.007>
40. Chen H, Li Y (2019) Three-stream attention-aware network for rgb-d salient object detection. *IEEE Transactions on Image Processing* PP 62825–2835. <https://doi.org/10.1109/TIP.2019.2891104>
41. Su J, Li J, Zhang Y, Xia C, Tian Y (2019) Selectivity or Invariance: Boundary-Aware Salient Object Detection. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). pp: 3798-3807 <https://doi.org/10.1109/ICCV.2019.00390>
42. Gao S, Cheng MM, Zhao K, Zhang XY, Torr P (2019) Res2net: a new multi-scale backbone architecture. *IEEE Trans Pattern Anal Mach Intell* 43(2):652–662. <https://doi.org/10.1109/TPAMI.2019.2938758>
43. Luo WJ, Li YJ, Urtasun R, Zemel RS (2016) Understanding the Effective Receptive Field in Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*. pp:4898–4906 <https://arxiv.org/abs/1701.04128>
44. Yu F, Koltun V (2016) Multi-Scale Context Aggregation by Dilated Convolutions. *ICLR*. <https://arxiv.org/abs/1511.07122>
45. Zhu XZ, Cheng DZ, Zhang Z, Lin SCF, Dai JF (2019) An Empirical Study of Spatial Attention Mechanisms in Deep Networks. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) pp: 6687-6696. <https://doi.org/10.1109/ICCV.2019.00679>
46. Hu J, Shen L, Albanie S, Sun G, Wu E (2020) Squeeze-and-Excitation Networks. *IEEE Trans Pattern Anal Mach Intell* 42(8):2011–2023. <https://doi.org/10.1109/TPAMI.2019.2913372>
47. Hou Q, Zhou D, Feng J (2021) Coordinate Attention for Efficient Mobile Network Design. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. <https://arxiv.org/abs/2103.02907>
48. Wei JH, Wang SH, Huang QM (2020) F³Net: Fusion, Feedback and Focus for Salient Object Detection. *Proceedings of the AAAI Conference on Artificial Intelligence* 34(07): 12321-12328. <https://doi.org/10.1609/aaai.v34i07.6916>
49. Fan DP, Cheng MM, Liu Y, Li T, Borji A (2017) Structure-Measure: A New Way to Evaluate Foreground Maps. 2017 IEEE International Conference on Computer Vision (ICCV): 4558-4567. <https://doi.org/10.1109/ICCV.2017.487>
50. Margolin R, Zelnik-Manor L, Tal A (2014) How to Evaluate Foreground Maps. 2014 IEEE Conference on Computer Vision and Pattern Recognition pp:248-255. <https://doi.org/10.1109/CVPR.2014.39>
51. Perazzi F, Krähenbühl P, Pritch Y, Hornung A (2012) Saliency filters: Contrast based filtering for salient region detection. 2012 IEEE Conference on Computer Vision and Pattern Recognition pp: 733-740. <https://doi.org/10.1109/CVPR.2012.6247743>
52. Fan DP, Gong C, Cao Y, Ren B, Cheng MM, Borji A (2018) Enhanced-alignment Measure for Binary Foreground Map Evaluation. *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence* pp: 698–704. <https://doi.org/10.24963/ijcai.2018/97>
53. Milletari F, Navab N, Ahmadi SA (2016) V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. 2016 Fourth International Conference on 3D Vision (3DV) pp: 565-571. <https://doi.org/10.1109/3DV.2016.79>
54. Zhou ZW, Siddiquee M, Tajbakhsh N, Liang M (2020) Unet++: redesigning skip connections to exploit multiscale features in image segmentation. *IEEE Trans Med Imaging* 39(6):1856–1867. <http://arxiv.org/abs/1912.05074>
55. Chao P, Kao CY, Ruan YS, Huang CH, Lin Y (2019) HarDNet: A Low Memory Traffic Network. *IEEE/CVF International Conference on Computer Vision (ICCV)* pp: 3551-3560. <https://doi.org/10.1109/ICCV.2019.00365>
56. Lou AG, Guan SY, Loew M (2021) CaraNet: Context Axial Reverse Attention Network for Segmentation of Small Medical Objects. *ArXiv*. <https://arxiv.org/abs/2108.07368>
57. Kim T, Lee H, Kim D (2021) UACANet: Uncertainty Augmented Context Attention for Polyp Segmentation. *ArXiv*. <https://arxiv.org/abs/2107.02368>

58. Qin XB, Zhang ZV, Huang CY, Gao C, Dehghan M, Jägersand M (2019) BASNet: Boundary-Aware Salient Object Detection. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp: 7471-7481. <https://doi.org/10.1109/CVPR.2019.00766>
59. Wu Zhe, Su L, Huang QM (2019) Stacked Cross Refinement Network for Edge-Aware Salient Object Detection. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) pp: 7263-7272. <https://doi.org/10.1109/ICCV.2019.00736>
60. Chen ZY, Xu QQ, Cong RM, Huang QM (2020) Global Context-Aware Progressive Aggregation Network for Salient Object Detection. Proceedings of the AAAI Conference on Artificial Intelligence 34(07): 10599-10606. <https://doi.org/10.1609/aaai.v34i07.6633>
61. Fan DP, Ji GP, Cheng MM, Shao L (2021) Concealed Object Detection. IEEE transactions on pattern analysis and machine intelligence. <https://doi.org/10.1109/TPAMI.2021.3085766>
62. Jha D, Smedsrud PH, Riegler MA, Halvorsen P, de Lange T, Johansen D, Johansen HD (2020) Kvasir-seg: A segmented polyp dataset. International Conference on Multimedia Modeling. MMM. pp: 451–462. 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, Drozdal M, Courville A (2017) A benchmark for endoluminal scene segmentation of colonoscopy images. Journal of Healthcare Engineering 2017(2017):4037190. <https://doi.org/10.1155/2017/4037190>
64. Tajbakhsh N, Gurudu SR, Liang J (2015) Automated polyp detection in colonoscopy videos using shape and context information. 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI) pp: 79-83. <https://doi.org/10.1109/ISBI.2015.7163821>

Tables

Table 1

Comparison results of MAGNet with 14 comparison algorithms. (⊞: medical image segmentation method, ⋄: saliency object detection method, ⊞: camouflaged object detection method.)

Methods	Pub.'Year	S_{α}	F_{β}^w	MAE	E_{ϕ}^{ad}	meanDic	meanIoU	meanSen	meanSpe
Unet++⊞	DLIA'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
SINet-V2⊞	TPAMI'21	0.822	0.700	0.038	0.883	0.735	0.627	0.767	0.955
UACANet-L⊞	ACM MM'21	0.816	0.724	0.034	0.901	0.745	0.646	0.763	0.945
Magnet⊞	Ours	0.829	0.727	0.034	0.901	0.757	0.656	0.789	0.954

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

Figures

Figure 1

Schematic diagram of the effect of observing the camouflage tank with a magnifier

Figure 2

MAGNet network structure.

Figure 3

The structure of Central Excitation Module

Figure 4

Schematic diagram of the central excitation effect of CEM

Figure 5

The structure of MFFM module (UP: upsample, CBR: Conv+BatchNorm+ReLU)

Figure 6

The structure of Channel-Spatial Attention Module (L2, L3, L4 refer to Res2-Layer2, Res2-Layer3, Res2-Layer4)

Figure 7

Loss value decline comparison curve

Figure 8

Visualization results for all algorithms

Figure 9

Visualization of MAGNet feature maps. (FEMM: output by Ergodic Magnify Module, FAFM: output by Attention Focus Module, Ffuse: final fused camouflaged object map)

Figure 10

Visualization of detection results on self-constructed dataset