A Low-Contrast Saliency Object Detection Method for Camouflage Analysis Is Improved Based on DaCOD

Yuxi Wang*, Yang Xu

***Abstract*— Camouflage object detection in low-contrast scenes faces severe challenges due to the high similarity between the foreground and background in multiple dimensions, such as color, shape, and texture. To improve the detection accuracy and generalization ability, this paper proposes an improved TMLCOD model based on the DaCOD algorithm and optimizes it for RGB-D camouflage target detection tasks. Firstly, in the RGB modal feature extraction stage, a Dynamic Adaptive Triplet Attention (DATA) module is introduced, which adaptively adjusts the weight of each branch through the dynamic weight fusion mechanism, and combines with efficient linear attention to reduce the computational complexity. Thus, the global receptive field and computational efficiency of the model are improved. Secondly, in the stage of feature fusion and target localization, an Efficient Adaptive Multi-Scale Attention (EAMSA) module is introduced to enhance the perception ability of low-contrast targets by focusing on the foreground and background regions with multiple attention heads. Finally, the loss function design is optimized by combining the weighted BCE and IoU loss with the structure loss, and introducing a dynamic weight adjustment mechanism to adaptively balance the contribution of different loss terms in the training process, to effectively alleviate the problems of boundary blurring and background interference. The experimental results show that compared with the original DaCOD, the main evaluation index, weighted F-measure (F_β^w) of the improved TMLCOD model is increased by 0.75%, 1.65%, and 0.98%, respectively, on the three data sets of CAMO, COD10K, and NC4K. The ability to depict the target boundary and retain details in complex background scenes is significantly enhanced. In addition, compared with the current mainstream methods, TMLCOD has a more stable detection performance in low-contrast scenes and shows better generalization ability when detecting complex targets.**

***Index Terms*—Camouflage object detection, Computer vision, Three-channel attention mechanism, Cross-modal learning, Separable attention**

I. Introduction

SALIENT Object Detection (SOD) is a fundamental task in the field of computer vision, aiming to identify and localize the most visually distinctive objects within a scene. These salient regions typically exhibit strong contrast in color, texture, or brightness compared to their surroundings, making them easier to detect. This technique has been widely applied in practical scenarios such as automated surface defect detection [1].

In contrast, Camouflaged Object Detection (COD) presents a more challenging research direction. This task involves detecting targets that are highly integrated with the background, where visual characteristics such as color, texture, and edges closely resemble the surrounding environment. As a result, the detection process becomes significantly more difficult. Unlike conventional SOD methods that rely on strong visual disparities between foreground and background, COD requires more sophisticated techniques to identify subtle differences. Camouflage is prevalent in both natural and artificial environments. In nature, organisms use camouflage to blend into their surroundings in order to evade predators or ambush prey. Similarly, in military, surveillance, and industrial applications, camouflage techniques are employed to conceal objects effectively. Figure 1 illustrates the progression from easily recognizable salient objects to highly concealed camouflaged targets.

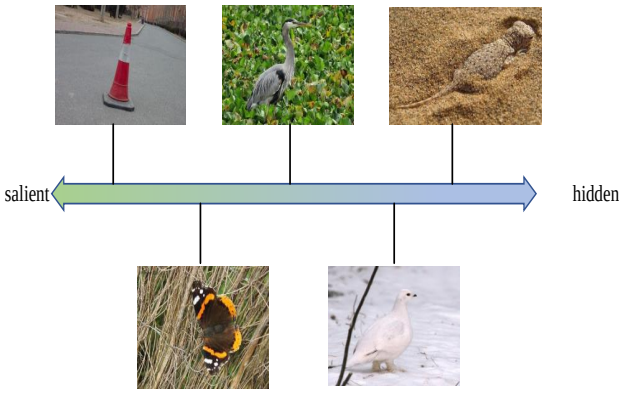

Fig. 1. Diagram of the transition from salient target to camouflaged (hidden) target.

The core challenge of camouflage object detection is how to accurately segment and locate the object from the complex background and interference factors (such as noise, occlusion, etc.). This task is an intensive prediction problem, which requires the model to have strong feature extraction ability and detailed identification ability. Early COD methods mainly rely on manual feature extraction or expert knowledge for prediction, but these methods often have low accuracy due to the difficulty of capturing small differences between the camouflaged target and the background. In recent years,

Manuscript received April 2, 2025; revised Jun 28, 2025.

This work was supported by the National Natural Science Foundation of China (61775169), the Education Department of Liaoning Province (LJKZ0310), the Excellent Young Talents Program of Liaoning University of Science and Technology (2021YQ04)

Yuxi Wang is a postgraduate student of Computer Science and Software Engineering, University of Science and Technology LiaoNing, Anshan 114051, China (corresponding author to provide phone: 86-18524332357; e-mail: 2692769200@qq.com).

Yang Xu is a Professor at the School of Computer Science and Software Engineering, University of Science and Technology LiaoNing, Anshan 114051, China (phone: 86-13889785726; e-mail: xuyang 1981@aliyun.com).

with the rapid development of deep learning technology and the open-source of large-scale datasets (such as CAMO [2] and COD10K [3]), COD methods based on deep learning have gradually become mainstream. Among them, some biological heuristics have made significant progress. For example, C2F-Net [4] proposes a Dual-branch Global Context (DGCM) module to optimize global information fusion; ZoomNeXt [5] designs a Scale Merging Subnetwork to simulate the visual strategy of human eyes when observing blurred objects, to enhance the ability to distinguish between foreground and background. R2CNet [6] introduces a Referring Feature Enrichment module to improve the recognition ability of specific camouflage objects.

However, most COD methods are still limited to RGB modality, while the human visual system not only relies on color, texture, and other information when recognizing objects, but also uses 3D depth information for perception and judgment. Biological and evolutionary studies have shown that depth information is crucial in recognizing the position, shape, and orientation of objects, so the introduction of RGB-D (color-depth) information can effectively enhance the detection ability of COD tasks. In recent years, RGB-D salient object detection has gradually become a research hotspot because depth information can provide rich geometric and spatial cues, which can help improve the detection accuracy. However, research on RGB-D COD is still in its infancy and has not been widely explored.

To solve the above problems, this paper proposes an improved method based on depth-assisted Camouflage Object Detection (DaCOD [7]), which further explores the contribution of depth modalities in salient object detection at low-contrast boundaries. Firstly, an improved ternary attention mechanism (DATA) is introduced in the RGB modal feature extraction stage, which improves the sensitivity of fine-grained features and reduces the interference of redundant information by capturing multi-directional feature interactions. Secondly, in the feature location and prediction stage, an improved Mask Separable Attention (EAMSA) module was added to focus on the foreground and background regions, respectively, to improve the detection ability of camouflaged objects. In addition, in the Loss function design, we define five Loss functions, from pixel-level classification (Loss 1), structure optimization of salient regions (Loss 2 and 4), difficult sample processing (loss 3), to final global optimization (loss 5), forming a hierarchical multi-stage loss design scheme. The Binary Cross Entropy (BCE) and weighted IOU loss function were combined, and the dynamic weight adjustment mechanism was used to adaptively balance the contribution of each loss term in the training process, to balance the contribution of term in the training process and alleviate the problems such as fuzzy boundary and complex target area. Through this design, the model can focus on different features at different stages, which not only improves the segmentation accuracy but also enhances the robustness of camouflage object detection.

II. The overall structure of the DaCOD algorithm framework

The DaCOD algorithm framework mainly studies how to make full use of depth information to assist the task of camouflage object detection. Since there is no data set specifically designed for RGB-D camouflage object detection, this method first generates the corresponding depth map from the RGB image by the monocular depth estimation technique, converts the depth map into a three-channel format after normalization, and concatenates it with the RGB image in the batch dimension for input into the network.

Inspired by the Salient Object Detection (SOD) task, the framework adopted a hybrid backbone network structure of Swin-L [8] and ResNet-50 [9]. This design fully combines the advantages of Transformer and Convolutional Neural Networks (CNNs) to achieve more efficient collaborative feature learning. In the feature extraction stage, the algorithm selected important features through the hierarchical feature selection mechanism, and used the Batch Segmentation (BSB) module to split these features, and extracted RGB features and depth features respectively.

In the feature fusion stage, the Cross-Modal Asymmetric Fusion (Cross-Modal Asymmetric Fusion, CAF) module was used in the framework. The CAF module combines RGB and depth features in an asymmetric manner, aiming to preserve the key details of depth information and effectively avoid possible depth ambiguity. This design idea fully reflects the adaptability to cross-modal feature differences and the attention to fusion quality, so as to improve the performance of camouflage target detection tasks. In the SWI-L part, the network will receive the input and generate a four-level feature representation, denoted as $\{S_i|i = 1, 2, 3, 4\}$. Since the Transformer encoder can capture global semantic information more accurately, the model chooses to use the highest-level features for object localization. By inputting the high-level prediction results of the RGB branch into the depth branch, the model can focus its attention on the regions with more accurate localization, thus effectively reducing interference.

For ResNet-50, the high-level feature output is discarded because these features are usually too small due to downsampling, thus missing rich edge detail information. Therefore, the output of ResNet can be represented as $\{R_i|i = 1, 2, 3\}$, and finally, the feature representation is generated by co-learning as R1, R2, R3, S4. These features are subsequently separated by the Batch Segmentation (BSB) module according to RGB and depth modalities, and the number of channels is compressed by the CBR (convolution, Batch Normalization, and ReLU) module. Finally, the modal features output by the model are $S_4^{D\text{epth}}$ for depth features, as well as joint features $\{R_i^{RGB}, R_i^{Depth}|i = 1, 2, 3\}$.

After multi-level feature generation, the model combines RGB and depth features through an asymmetric fusion mechanism (CAF module). RGB features and depth features are input into ABr (RGB attention module) and Abd (deep attention module), which contain channel attention and spatial attention, respectively, to further enhance the semantic information. Among them, semantically enhanced features S_f^{rgb}, S_f^{depth} and initial prediction Figure S_p^{depth}, can be obtained by the following formula:

$$S_f^{rgb} = SA(CA(S_A^{rgb})), S_p^{rgb} \quad = Conv(S_f^{rgb}) \tag{1}$$

$$S_f^{depth} = CA(S_4^{depth}), S_p^{depth} = ConV(S_f^{depth}) \tag{2}$$

Where CA denotes channel attention operation, and SA denotes spatial attention operation. Then, at each stage of feature decoding, the feature was gradually optimized by using the Feature Refinement (FR) module. The RGB prediction results are unidirectionally transferred to the depth branch by asymmetric fusion. For example, the operation in the RGB branch can be expressed as follows:

$$F_{fa3}^{rgb} = R_3^{rgb} \times sig(up(S_p^{rgb})) \tag{3}$$

Where Up represents the upsampling operation and sig represents the Sigmoid normalization operation. The cross-modal fusion is only passed from the RGB branch to the depth branch, where the depth features are refined by Up-sampling (Up) and Sigmoid normalization (sig) operations. This fusion method preferentially uses rich information of RGB images to enhance depth features, while reducing the interference caused by unreliable depth maps. Finally, the model combines the RGB features and the refined depth features to generate the final prediction result. Through this design, the model realizes the efficient fusion of RGB and depth information, and further improves the performance of camouflage object detection. The whole process with the final predicted image can be expressed as follows:

$$F_{fa}^{depth} = \begin{cases} R_i^{depth} \times sig(up(S_p^{rgb})), & i = 3 \\ R_i^{depth} \times sig(up(F_{p_{i+1}}^{rgb})), & i = 1, 2 \end{cases} \tag{4}$$

$$F_{ba}^{depth} = \begin{cases} R_i^{depth} \times (1 - sig(up(S_p^{rgb}))), & i = 3 \\ R_i^{depth} \times (1 - sig(up(F_{p_{i+1}}^{rgb}))), & i = 1, 2 \end{cases} \tag{5}$$

$$\mathrm{Pre}_{final} = CBR(F_{R1}^{rgb} + F_{R1}^{depth}) \tag{6}$$

III. Algorithm Model Improvement Scheme

In order to further improve the performance of the DaCOD model in RGB-D camouflage object detection tasks, two key innovative improvement modules are added: the Triplet Attention module and the Multi-Scale Attention (MSA) module. In the feature extraction stage, Swi-L is used to capture global semantic information, while ResNet-50 retains local edge details. Subsequently, the modal features are separated by the BSB module, and cross-modal fusion is performed by the Triplet Attention module. The fused features are processed by the MSA module for multi-scale enhancement to achieve more accurate detection and location of camouflage targets. This optimization method significantly optimizes the original method from the perspective of cross-modal feature fusion and multi-scale information capture, as shown in Figure 2.

A. Improve the Triplet Attention module

The main advantage of the Triplet Attention [10] module is its cross-dimensional attention mechanism. Traditional Attention mechanisms, such as SE [11] and CBAM [12], usually work only in the channel or spatial dimension, while Triplet Attention combines the attention calculation in three dimensions: width (C direction), height (H direction) and space (W direction), which can capture the interaction between features more comprehensively. Moreover, compared with some other complex Attention mechanisms, the design of the Triplet Attention module adds almost no additional computational overhead. The attention computation in each direction is achieved by a simple convolution operation with low computational complexity and does not require a significant increase in additional model parameters.

However, the traditional Triplet Attention module enhances the feature expression ability through the multi-branch attention mechanism of static fusion weights, but does not consider the importance of different input features. To solve this problem, a Dynamic Weight Generator module based on global statistics is designed and added to the Triplet Attention module. The branch fusion weights are generated by the mean and variance of the input features. The global statistics are calculated as shown in Equations (7) and (8).

$$\mathrm{mean}_b = \frac{1}{C \cdot H \cdot W} \sum_{c=1}^{C} \sum_{h=1}^{H} \sum_{w=1}^{W} x_{b,c,h,w} \tag{7}$$

$$\mathrm{var}_b = \frac{1}{C \cdot W \cdot H} \sum_{c=1}^{C} \sum_{h=1}^{H} \sum_{w=1}^{W} (x_{b,c,h,w} - mean_b)^2 \tag{8}$$

The mean value $mean_b$ reflects the overall feature strength of the BTH sample. The variance var_b reflects the dispersion of the feature distribution of the BTH sample. The input feature x is a four-dimensional tensor whose elements are all real numbers with dimension $B \times C \times H \times W$, and b is the sample index. The specific implementation details are that the global mean and variance of the BTH sample are obtained after summing and averaging the eigenvalues of all channels C, all heights H, and all widths W of the BTH sample.

Then, the statistics $mean_b$ and var_b are fed into the fully connected network to generate the fusion weights of each branch as shown in Equation (9) :

$$w_b = Softmax(W_2 \cdot ReLU(W_1 \cdot [mean_b; \mathrm{var}_b])) \tag{9}$$

Of these, $W_1 \in R^{4\times2}$, $W_2 \in R^{3\times4}$. Firstly, the statistics of each sample are concatenated into a 2D vector, and the final output 3D weight vector $w_b \in R^3$ is normalized by Softmax.

Finally, the output of each branch is weighted and dynamically fused according to the weight w_b. The specific calculation formula is shown in Equation (10):

$$Fused_b = \sum_{k=1}^{K} w_{b,k} \cdot Branch_k(x_b) \tag{10}$$

Where K = 3 (including spatial branches) or K = 2 (without spatial branches). By assigning score weight to each sample dynamically, the deviation of feature expression caused by a fixed weight is effectively avoided. The calculation of the global statistics does not depend on the number of channels C, and the calculation is efficient and only requires the input of two-dimensional statistics, and the number of parameters is very small (the parameter of the fully connected layer is $2 \times 4 + 4 \times 3 = 20$). In this way, the improved module can be adapted to any input dimension, which is more consistent with the idea of transpose on three dimensions of width (C

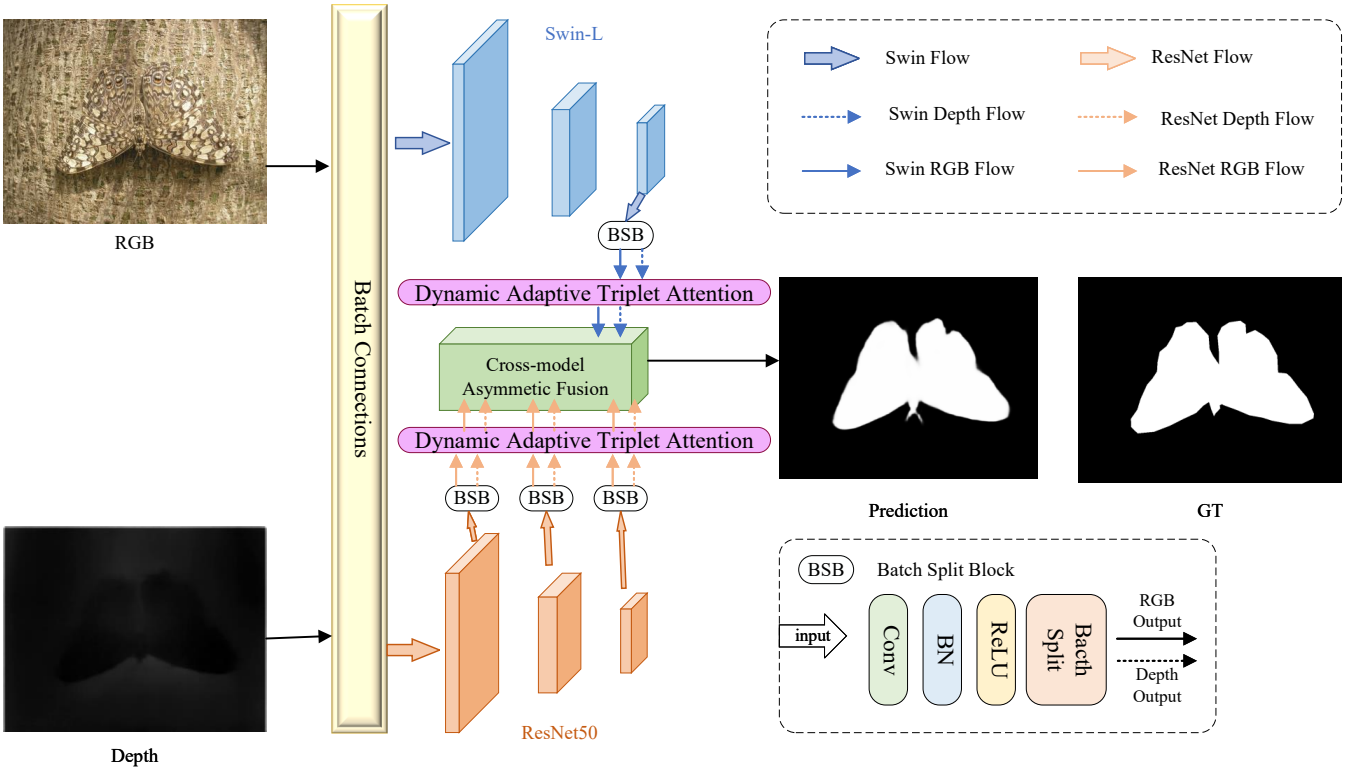

Fig. 2. The overall architecture of TMLCOD.

direction), height (H direction), and space (W direction) in the Triplet Attention module.

It is also noted that the space complexity of the standard attention mechanism is $O(N^2)(N = H \times W)$, which limits the application of the model in high-resolution scenarios. Therefore, the Linear Attention module [13] is introduced into the Attention Gate to reduce the computational complexity to $O(N)$. The standard attention calculation formula is Equation (11), and the linear attention calculation formula is Equation (12).

$$Attention(Q, K, V) = Softmax(\frac{QK^T}{\sqrt{d}})V \tag{11}$$

$$\widetilde{Q} = AvgPool(Q), \widetilde{K} = AvgPool(K), \tag{12}$$
$$LinearAttention(Q, K, V) = Softmax(\widetilde{Q}\widetilde{K}^T)V$$

Where Q, K, and V are the query, key, and value matrices, respectively. Finally, the cross-dimensional interaction was realized by Permute and Tensor Reshape in the forward propagation process. The structure of the improved DATA module is shown in FIG. 3, which mainly contains three core components: dynamic weight generator, linear attention gating mechanism, and channel-independent branch.

B. Masked Separable Attention module improvement

In the task of multi-scale camouflage object detection, how to effectively capture and fuse multi-scale information has become one of the key issues. In the literature [14], a Masked Separable Attention (MSA) module is proposed. Firstly, the spatial information is encoded by using the Multi-Dconv Head Transposed Attention (TA) module. The formula of TA is as follows:

$$TA(\mathbb{Q}, \mathbb{K}, \mathbb{V}) \quad = \mathrm{V} \cdot \mathrm{Softmax}(\frac{Q^T K}{\alpha}) \tag{13}$$

Whereα is a learnable scaling parameter and Q, K, and V are the query, key, and value matrices, respectively, which can be generated by three independent 1 * 1 convolutions followed by a 3 * 3 depth convolution.

Next, a prediction mask that can be generated at each feature level is introduced into the TA module as a front-background contrast prior, and all attention heads are divided into three groups: foreground head (F-TA), background head (B-TA), and normal TA. Finally, the feature aggregation is performed by a 3 * 3 convolution, the formula is as follows:

$$Z = Conv_{3\times3}([F - TA, B - TA, TA]) \tag{14}$$

In order to improve the performance of the multi-head self-attention (MSA) module in complex visual tasks, this study systematically optimized its computational efficiency, feature fusion strategy, and multi-scale feature extraction capability. The specific application flow chart of the optimized Efficient Adaptive Multi-Scale Attention (EAMSA) module in the cross-modal asymmetric fusion module is shown in Figure 4. Specific improvements include the following three aspects:

The first aspect is the optimization of the attention mechanism. The current MSA module adopts the standard self-attention mechanism, which is excellent at capturing global dependencies, but its computational complexity is high at $O(N^2)$(N is the sequence length), especially when dealing with high-resolution images. In order to reduce the computational complexity of the model and improve

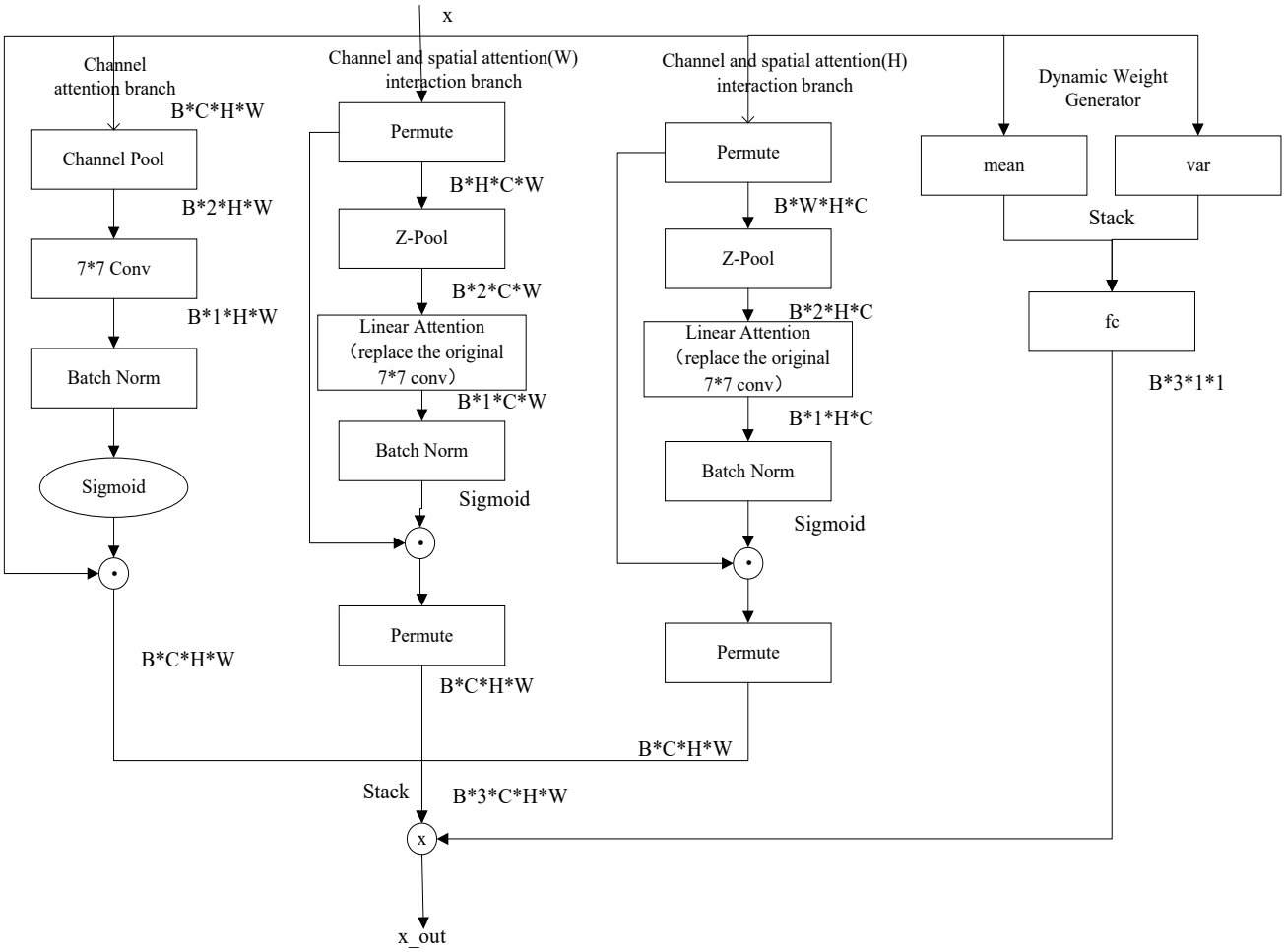

Fig. 3. The Triplet Attention Overall framework diagram.

the efficiency of the model, the Efficient Attention [15] mechanism is introduced in the research. The Efficient Attention mechanism generates the Query, Key, and Value matrix through the convolutional layer, and performs L2 normalization on the Q and K matrix to suppress the feature divergence problem in high-dimensional space. Its calculation process can be expressed as follows:

$$\begin{aligned} &Q, K, V = Conv_{1\times1}(X), \\ &\tilde{Q} = L2Norm(Q), \\ &\widetilde{K} = L2Norm(K) \end{aligned} \tag{15}$$

Then, the feature channel is divided into h attention heads, and the attention weight is calculated in blocks to reduce memory usage. The formula for calculating the attention score (16) is as follows:

$$\begin{aligned} &EfficientAttention(Q, K, V) = \\ &Softmax\left[\frac{\widetilde{Q}\widetilde{K}^T}{\sqrt{d}} \cdot \tau\right] \end{aligned} \tag{16}$$

Where Q, K, and V represent the query, key, and value matrices, respectively, d_k is the dimension of the key vector and is the learnable temperature coefficient, and τ is used to adjust the sharpness of the attention distribution. Reduce complexity from $O(N^2)$ to $O(N \log N)$ with channel grouping and parallel computing, while retaining global dependency capture. In this way, the consumption of computing resources is significantly reduced on the premise of maintaining the feature interaction capability, and it is suitable for high-resolution image processing.

The second aspect is the optimization of feature fusion. In the original MSA module, the feature fusion part relies on a simple convolution operation, which makes it difficult to dynamically weigh the importance of different branches, especially in the cross-modal task, which easily leads to the loss of key information. Therefore, to improve the effect of feature Fusion, the Adaptive Fusion strategy is introduced in this study. The Lightweight attention module is designed to automatically generate channel-level fusion weights based on input features. The weight generation process is defined as:

$$\begin{aligned} &W = \\ &Softmax(Comv_{1\times1}(ReLU(Comv_{1\times1}(X_1 + X_2 + X_3)))) \end{aligned} \tag{17}$$

Where X_1, X_2, X_3 is the branch features to be fused, and $W \in R^{3\times H\times W}$ are the three sets of spatial adaptive weights. Then, feature fusion is achieved by channel-by-channel multiplication and summation, as shown in formula (18) :

$$X_{fused} = \sum_{i=1}^{3} W_i \odot X_i \tag{18}$$

Where $\odot$ represents channel-by-channel multiplication,

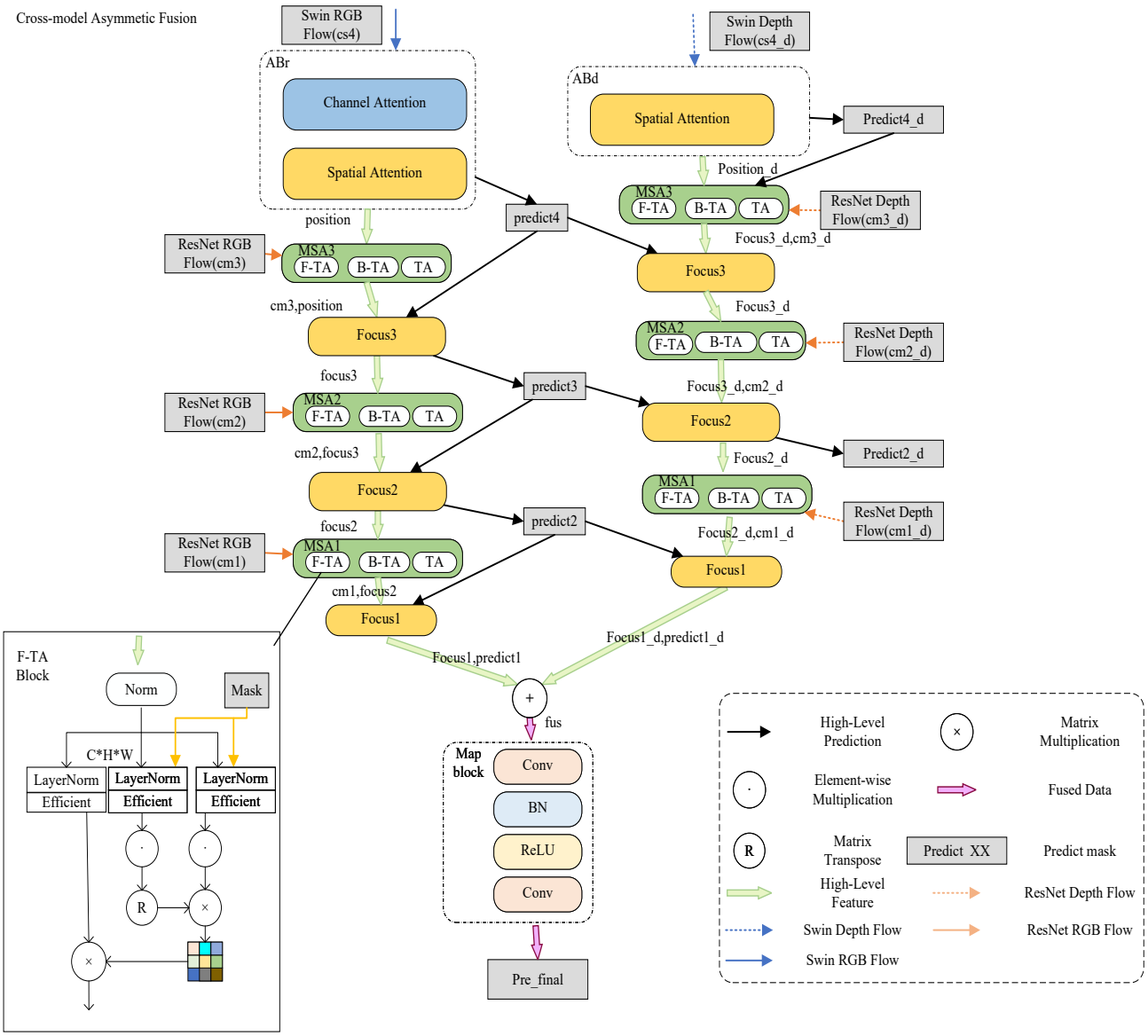

Fig. 4. Flowchart of the application of the improved Efficient Adaptive Multi-scale Attention Module (EAMSA) in cross-model non-pair fusion.

ensuring that key features are enhanced while redundant features are suppressed. It significantly improves the flexibility of the branch feature interaction and effectively enhances the expression ability of fine-grained features in the task of camouflage target detection.

Third, the single-scale convolution operation in the original module makes it difficult to capture the context information of different granularities in the image, especially the problem of blurred edges or insufficient sensitivity to small-scale targets. An improved scheme of dilation convolutional parallel structure is adopted, and multi-scale features are extracted in parallel by convolutional layers with multiple dilation rates to expand the receptive field:

$$\begin{aligned} X_1 &= Conv_{3\times3}(X; dilation = 1) \\ X_2 &= Conv_{3\times3}(X; dilation = 2) \\ X_3 &= Conv_{3\times3}(X; dilation = 4) \end{aligned} \tag{19}$$

After multi-scale features are concatenated, channel compression is carried out through 1×1 convolution to retain key information:

$$X_{multi-scale} = Conv_{1\times1}(Concat(X_1, X_2, X_3)) \tag{20}$$

On the whole, multi-scale context enhancement enables the improved model to capture both local details and global structure information at the same time and improves the accuracy of the model in dealing with complex backgrounds. Finally, the above module is integrated into the MSA module (as shown in Figure 5). After the normalization of the input features by LayerNorm, the input features are respectively input into the foreground attention head (F-TA), background attention head (B-TA), and ordinary attention head (TA) for pre-processing. Each branch then generates Attention weights through an Efficient Attention mechanism to suppress redundant calculations. The Adaptive Fusion module dynamically fuses the three-branch output, strengthens task-related features, and the Multi-Scale Feature module further extracts multi-granularity features to enhance the model's adaptability to scale changes.

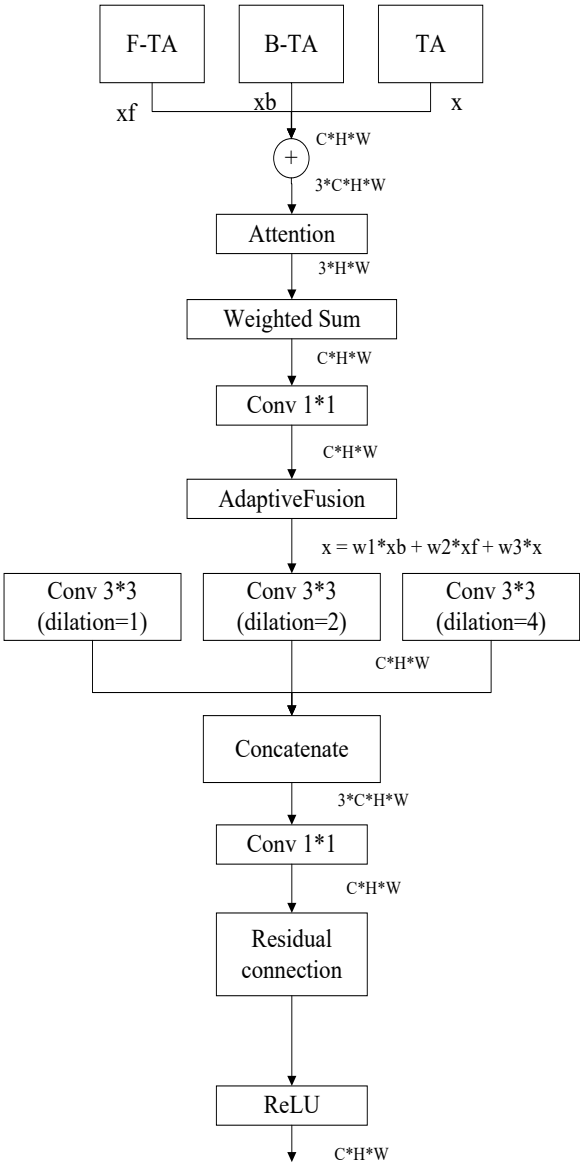

Fig. 5. Improved architecture diagram of Masked Separable Attention module.

C. Optimizing the loss function

To further enhance the performance and robustness of the model, an improved loss function is proposed, which integrates multiple loss components to take advantage of their complementary advantages. Similar to the original design, the loss function in the study consists of two main components: modal-specific attention loss and feature refinement, and asymmetric fusion loss. However, some improvements are introduced in the training process to improve the accuracy and optimize the gradient flow, further improve the model's learning ability for multi-scale features, and alleviate the problem of class imbalance.

For the RGB branch, the combination of binary cross-entropy (BCE) loss and IoU loss continues to be used to effectively balance classification and positioning. BCE losses ensure accurate pixel classification, while IoU losses enhance the spatial consistency of predictions, as shown in equation (21) :

$$L_{bce_iou} = L_{bce} + L_{iou} \tag{21}$$

For deep branches, the loss function is refined by combining weighted binary cross-entropy (WBCE) losses and weighted IoU (WiOU) losses. WBCE loss applies an adaptive weighting scheme based on the structural importance of different areas, emphasizing challenging areas. WiOU loss further improves accuracy by punishing misalignments in areas of interest. Thus, depth-specific losses are expressed as formula (22) :

$$L_{wbi} = L_{wbce} + L_{wiou} \tag{22}$$

The final attentional module loss is shown in equation (23) :

$$L_{attention} = L_{bcs_iou}(S_p^{rgb}, G) + L_{wbi}(S_p^d, G) \tag{23}$$

Where G represents the true value.

In the stage of feature refinement and asymmetric fusion, the loss formula is extended by combining multiple prediction scales. Specifically, the loss predicted by the RGB and depth branches is calculated before the final fusion, as well as the additional loss term for the final output. To achieve this, a mixed loss formula combining BCE, IoU, and WBCE losses is used. Intermediate forecast losses are given by:

$$L_{fision} = L_{wbi}(F_{p2}^{rgb}, G) + L_{wbi}(F_{P2}^d, G) + 2 \times (P_{final}, G) \tag{24}$$

To further improve performance, we introduced dynamically weighted structural loss ($L_{structure}$) to adapt the training period. Structure loss uses an edge-sensing weighting mechanism to assign higher importance to regions with complex structures, thereby improving fine-grained feature learning. The formula is:

$$L_{structure} = \frac{(W \cdot Lbce) + (W \cdot L_{iou})}{W_{sum}} \tag{25}$$

Where W represents the structural importance map obtained by the adaptive pooling operation.

In summary, the final loss function of this study dynamically integrates all the above loss functions and adjusts the weights of each period to balance the features from different levels.

$$\begin{aligned} L_{final} &= 1 \times L_{bce_iou}(P_1, G) + 1 \times L_{structure}(P_2, G) + \\ &2 \times L_{bce_iou}(P_3, G) + 1 \times L_{structure}(P_4, G) \\ &+ (1 + \lambda) \times L_{structure}(P_5, G) \end{aligned} \tag{26}$$

Among them, λ is a scaling factor related to the course of epochs, which increases as the training proceeds, enabling the model to pay more attention to fine-grained details in the later stage. By combining these improved loss functions, the loss function in this study can not only ensure stable optimization but also improve the overall detection accuracy by resolving structural inconsistencies and enhancing spatial consistency.

IV. Analysis and Discussion of Experimental Results

A. Experimental software and hardware environment

The hardware environment of this experiment is mainly equipped with a NVIDIA RTX 3090(24GB memory) GPU instance, Intel Xeon Platinum 8375C CPU, and 72G memory

to ensure the efficiency of large-scale model training. The software environment of the experiment is implemented based on the PyTorch framework, using Swin-L and ResNet-50 as dual backbone networks, and using ImageNet pre-trained weights for collaborative learning. The SGD optimizer with momentum 0.9 is used, the weight decay coefficient is, and the initial learning rate is set to 0.001 and dynamically adjusted through the poly strategy. The uniform input image size is pixels, the batch size is set to 6, and a total of 60 epochs are trained to fully converge the model. It provides reliable technical support for multi-scale feature learning and complex scene target recognition tasks.

B. Dataset and evaluation metrics

In this study, the proposed improved method is evaluated on three widely used benchmark datasets, namely CAMO [2], COD10K [3], and NC4K [16]. The CAMO dataset covers 1250 camouflage images of different categories, of which 1000 are used for training and 250 for testing. COD10K is currently the largest benchmark dataset. It collects 5066 camouflage images, including 3040 training images and 2026 test images, covering five main categories and 69 sub-categories. NC4K is a more recent camouflaged object detection dataset containing 4121 images, which is mainly used to evaluate the generalization ability of the model. In this experiment, the training sets of CAMO and COD10K and their corresponding depth images are mainly used as training data, while the remaining parts of the CAMO and COD10K datasets and the NC4K dataset are used for testing.

In terms of evaluation indicators, four commonly used standard indicators are used for quantitative evaluation, including structure measure (S_α) [17], adaptive E-measure (E_φ^{ad}) [18], weighted F-measure (F_β^ω) [19], and Mean Absolute Error (MAE) [20]. The structural metric mainly measures the structural similarity between the predicted results and the real results, and evaluates the ability of the model in capturing the structural information of the image by calculating the structural similarity index (SSIM) of the image. The adaptive E-measure is an evaluation metric based on edge and region information, which can adaptively adjust the weights to more accurately evaluate the performance of the model in edge detection and region segmentation. The weighted F-measure takes into account both Precision and Recall and evaluates the performance of the model on different classes more comprehensively using a weighted average, which is especially suitable for datasets with imbalanced classes. Mean absolute error quantifies the prediction error of a model by calculating the average of the absolute errors between the predicted results and the true results and is an intuitive and commonly used indicator to evaluate the performance of a model.

C. Analysis of experimental results

In order to fully verify the effectiveness of the Dynamic Adaptive Triplet Attention (DATA) module and the Efficient Adaptive Multi-Scale Attention (EAMSA) module in the proposed improved model, in this study, these modules are progressively introduced on the benchmark model, and their impact on model performance is evaluated. Multiple widely used Camouflage Object Detection (COD) datasets were used in the experiment, and quantitative analysis was carried out from multiple dimensions to ensure the scientific and fair nature of the experiment.

(A) Analysis of the influence of different improvement modules on the model

In order to verify the effectiveness of the proposed improved method, this study introduces the DATA module, the EAMSA module, and the improved Loss function (Loss) based on DaCOD, and makes a comprehensive comparative analysis of the three data sets CAMO, COD10K, and NC4K. Table 1 summarizes the performance of different models and evaluates the detection accuracy and error by four metrics S_α, E_φ^{ad} , F_β^ω, and MAE.

After adding the DATA module to DaCOD, the AUC is increased from 0.9224 to 0.9331 (+1.16%), which also verifies that the DATA module optimizes the generalization ability of the model by enhancing the feature interaction between channels. The evaluation index F_β^ω in the CAMO dataset is increased from 0.796 to 0.799, which is a relative increase of 0.38%, indicating that the detection accuracy of the target area has been improved. On the Cod10K dataset, F_β^ω is increased from 0.729 to 0.735, a relative increase of 0.82%. In the NC4K dataset, the index E_φ^{ad} is increased from 0.923 to 0.927 (+0.43%), indicating that the feature expression ability of the DATA module is enhanced in complex backgrounds.

When only EAMSA is added based on DaCOD, the improvement in four indicators is slightly smaller than that of the DATA module, but it still shows a good enhancement effect. The AUC increased from 0.9224 to 0.9356 (+1.43%), which was significantly better than the independent effect of the DATA module (+1.16%), which verified the advantage of EAMSA in global contrast modeling. EAMAS mainly improves the learning ability of global contrast information, especially on the NC4K dataset with complex background, S_α improves from 0.874 to 0.878 (+0.46%), but E_φ^{ad} slightly decreases from 0.923 to 0.922. This also suggests that the subsequent research direction of this study needs to combine other modules to balance global and local features.

After introducing DATA + EAMSA at the same time, the overall performance of the model is significantly improved. The proposed method achieves 0.741 and 0.822 on the COD10K and NC4K datasets, respectively, which are 1.65%and 0.98% higher than the baseline DaCOD model, indicating that the multi-scale feature fusion enhances the learning ability of the target region boundary. After the combination of the two, the model has a significant improvement in all indicators, which proves the effectiveness of feature fusion.

In summary, the experimental results fully show that the DATA module can enhance the feature expression through channel interaction, and the EAMSA module optimizes global contrast learning. The collaboration of the two modules significantly improves the overall ability of the model to detect multi-scale boundaries, which makes the model achieve the best performance on the three data sets of CAMO, COD10K, and NC4K. It is especially outstanding in complex backgrounds and provides an efficient solution for real-time applications in low-contrast environments, such as military reconnaissance and ecological monitoring.

TABLE I
Comparison of Different Models on CAMO, Cod10K, and NC4K Datasets

Setting	CAMO				Cod10K				NC4K			
	$S_\alpha \uparrow$	$E_\varphi^{ad} \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\varphi^{ad} \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\varphi^{ad} \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$
DaCOD	0.855	0.911	0.796	0.05	0.84	0.908	0.729	0.02	0.87	0.923	0.814	0.03
DaCOD+TA	0.851	0.91	0.799	0.05	0.84	0.912	0.735	0.02	0.87	0.927	0.818	0.03
DaCOD-MSA	0.855	0.905	0.795	0.05	0.84	0.896	0.733	0.02	0.87	0.922	0.818	0.03
DaCOD-TA-MSA	0.855	0.912	0.799	0.05	0.84	0.912	0.741	0.02	0.87	0.926	0.822	0.03
DaCOD-TA-MSA-Loss (Ours)	0.855	0.915	0.802	0.04	0.84	0.912	0.741	0.02	0.87	0.927	0.822	0.03

TABLE II
Comparison of Quantitative Results Between the Improved Method and Other 11 COD Algorithms on Three Benchmark Datasets (The Top Three Results Are Shown in Red, Blue, and Green, Respectively.)

Method	Venue	CAMO				Cod10K				NC4K			
		$S_\alpha \uparrow$	$E_\varphi^{ad} \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\varphi^{ad} \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$	$S_\alpha \uparrow$	$E_\varphi^{ad} \uparrow$	$F_\beta^\omega \uparrow$	$M \downarrow$
SINet [21]	CVPR 2020	0.751	0.771	0.606	0.100	0.771	0.806	0.551	0.051	0.810	0.873	0.772	0.057
MGL [22]	CVPR 2021	0.775	0.847	0.673	0.088	0.814	0.865	0.666	0.035	–	–	–	–
PFNet [23]	CVPR 2021	0.782	0.852	0.695	0.085	0.800	0.868	0.660	0.040	0.829	0.887	0.745	0.053
UGTR [24]	ICCV 2021	0.785	0.859	0.686	0.086	0.818	0.850	0.667	0.035	–	–	–	–
LSR [25]	CVPR 2021	0.793	0.826	0.725	0.085	0.793	0.868	0.685	0.041	0.839	0.883	0.779	0.053
SINet-V2 [26]	TPAMI 2021	0.820	0.882	0.743	0.070	0.815	0.887	0.680	0.037	0.847	0.903	0.769	0.048
PrerNet [27]	ACM 2022	0.790	0.854	0.708	0.077	0.813	0.894	0.697	0.034	0.834	0.897	0.763	0.050
ZoomNet [28]	CVPR 2022	0.820	0.878	0.752	0.066	0.838	0.892	0.729	0.029	0.853	0.904	0.784	0.043
SegMaR [29]	CVPR 2022	0.815	0.872	0.753	0.071	0.833	0.893	0.724	0.034	0.841	0.902	0.781	0.046
ZoomNetXt [5]	TPAMI 2024	0.821	0.885	0.760	0.069	0.848	0.910	0.738	0.026	0.869	0.925	0.808	0.038
DaCOD [7]	Am 2023	0.855	0.911	0.796	0.051	0.840	0.908	0.729	0.028	0.874	0.923	0.814	0.035
Ours	–	0.855	0.915	0.802	0.049	0.848	0.912	0.741	0.026	0.878	0.927	0.822	0.034

(B)Analysis of the influence of different improvement modules on the model

In the comparison experiment, this paper compares the improved model with the baseline model. And the comparison of the 11 most advanced COD model methods (SINet [21], MGL [22], PFNet [23], UGTR [24], LSR [25], SINet-V2 [26], PreyNet [27], ZoomNet [28], SegMaR [29], ZoomNetXt [5], and DaCOD [7]) Although they are among the more advanced methods, our method still exhibits performance advantages over recent methods. To ensure a fair comparison, this paper either uses the results reported in its counterpart paper or reproduces its model with the same recommendation Settings and training data.

As shown in Table 2, the proposed method achieves the best performance on all data sets and all evaluation indicators, which fully verifies the effectiveness of the proposed improvement strategy. Among them, compared with the current latest ZoomNetXt [5] method, the four indicators S_α, E_φ^{ad}, F_β^ω and MAE of the proposed method on CAMO, COD10K and NC4K datasets are increased by 3.37%, 0.44%, 0.44% and 8.7% on average. It shows that the proposed improvement not only improves the accuracy of salient objects but also effectively reduces the error.

At the same time, compared with UGTR [24] method based on Transformer structure, the four indicators S_α, E_φ^{ad}, F_β^ω and MAE of the proposed method on CAMO and COD10K datasets are increased by 7.0%, 6.6%, 9.3%, and 7.3% on average. This shows that the multi-scale feature fusion strategy proposed in this paper can effectively enhance the perception ability of the model to the target area without relying on the global self-attention mechanism, thereby improving the detection performance.

This study improved based on DaCOD [7] and achieved comprehensive surpasses in multiple indicators. As shown in Table 2, compared with the baseline model, the S_α index of the proposed method on CAMO, COD10K, and NC4K datasets is increased by 0.6%, 0.1%, and 0.4%, respectively, indicating that the improved model can obtain more stable target prediction results in different scenarios. In addition, the E_φ^{ad} index is increased by 0.4% (0.911 $\rightarrow$ 0.915) on the COD10K dataset and 0.4% (0.923 $\rightarrow$ 0.927) on the NC4K dataset, indicating that the proposed method has better robustness in the detection task of small targets and complex backgrounds. The absolute difference of F_β^ω index on the COD10K dataset is increased by 1.2% (0.729 $\rightarrow$ 0.741), and the absolute difference on the NC4K dataset is increased by 0.8% (0.814 $\rightarrow$ 0.822), indicating that the proposed method has better robustness in the detection task of small targets and complex backgrounds.

In terms of error index MAE, the proposed method achieves further reduction on all datasets, in which the MAE of the CAMO dataset is reduced from 0.051 to 0.049, the MAE of the COD10K dataset is reduced from 0.028 to 0.026, and the MAE of the NC4K dataset is reduced from 0.038 to 0.034. It is worth noting that on the NC4K dataset, the MAE reduction of the proposed method reaches 10.5%, indicating that the improved feature extraction and fusion strategy can effectively reduce false detections and improve the accurate localization ability to camouflaging targets.

According to the above comparison experiments, the DATA and EAMSA mechanisms proposed in this paper play an important role in multi-scale feature extraction and object localization. Compared with UGTR[23], which only relies on the Transformer structure, the method in this paper adopts a lightweight multi-scale fusion strategy to improve the detection accuracy and reduce the computational overhead so that the model can show strong competitiveness in different COD tasks. In addition, the performance improvement of the proposed method is particularly significant on the NC4K dataset, which indicates that the proposed improved mechanism has better generalization ability in more challenging complex background data.

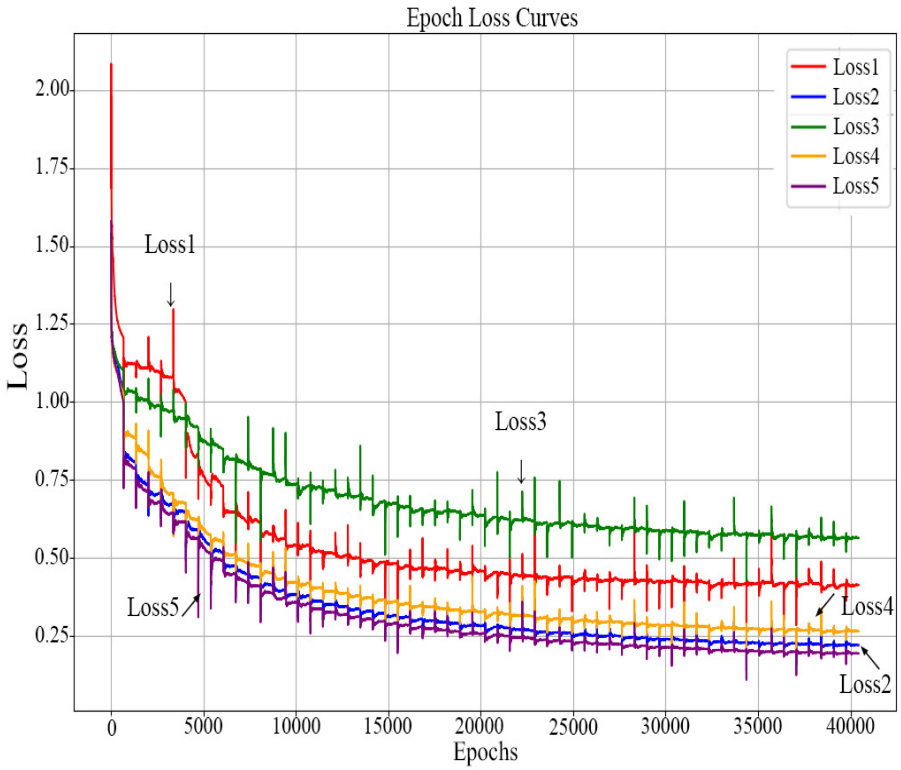

Fig. 6. Loss function plot.

In summary, the experimental results fully verify the effectiveness of the proposed method, which achieves the current optimal performance on multiple datasets and evaluation indicators, and provides a more accurate and efficient solution for the camouflage target detection task.

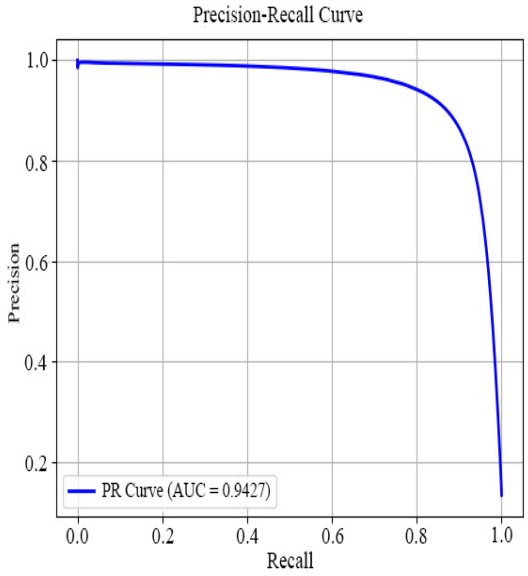

Fig. 7. PR plot.

Figure 6 shows how the model changes with different loss functions during training. It can be observed that with the increase in training rounds, all loss functions show a gradual decline, and the model starts to converge, which indicates

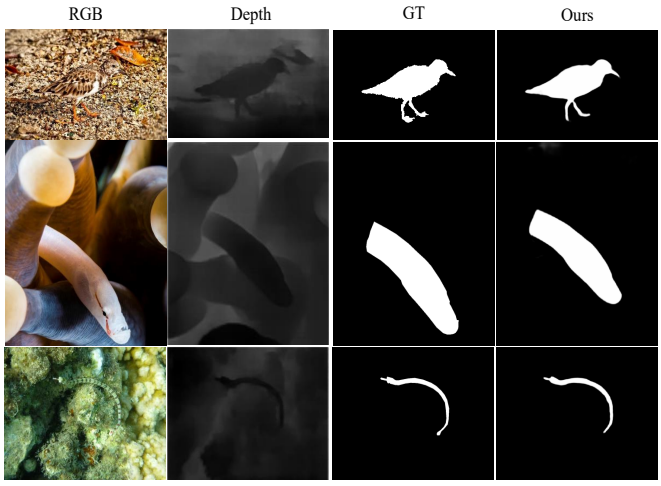

Fig. 8. Comparison of detection results of camouflaged targets assisted by depth information.

that the model is constantly learning and optimizing. In addition, the stability of the loss function in the later training period indicates that the model training process is stable and no overfitting occurs.

Figure 7 shows the precision-recall (PR) curve of the model, where the area under the curve (AUC) reaches 0.9427, indicating that the model has high accuracy in distinguishing between positive and negative samples. The tendency of the PR curve to approach the top right further confirms that the model maintains high precision while maintaining high recall, which is particularly important for class imbalance problems.

A visual comparison of improving the accuracy of camouflaged object detection with the help of depth

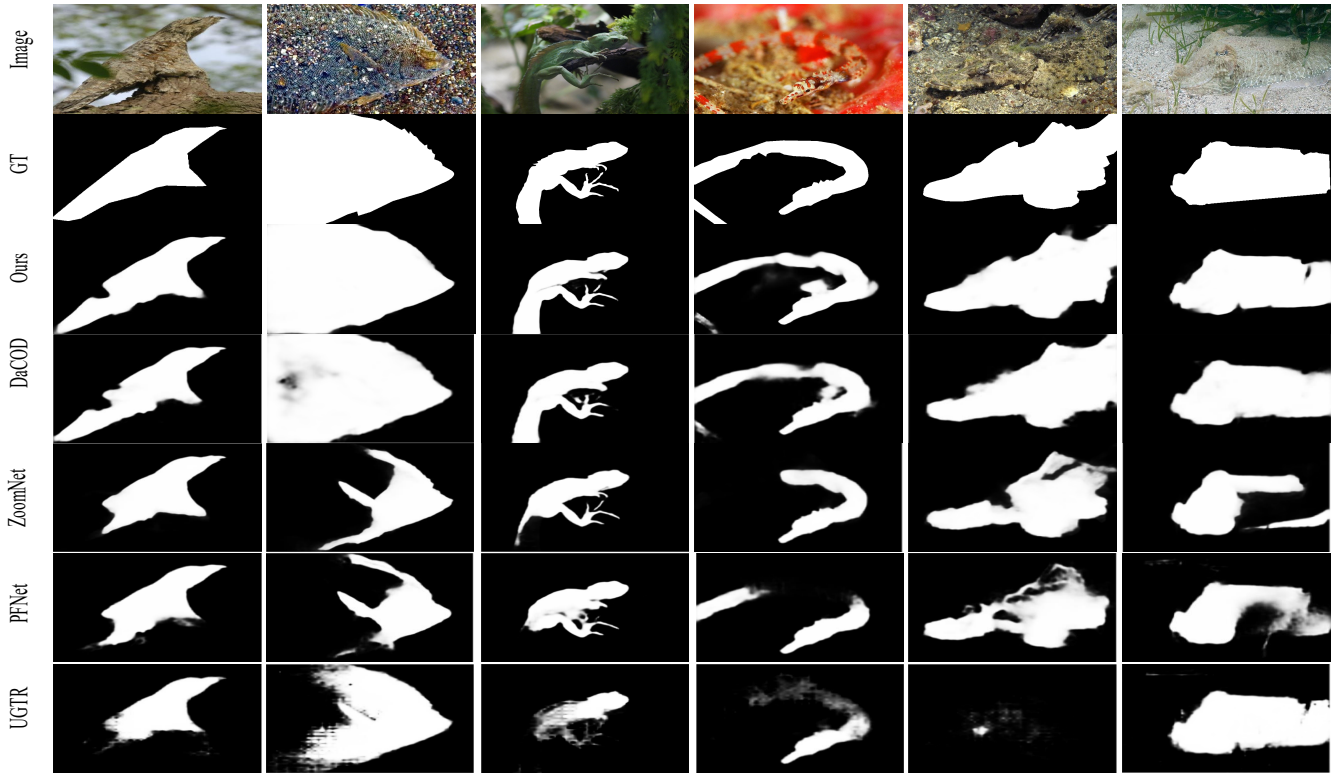

Fig. 9. Comparison of segmentation effect of different methods in camouflage target detection.

information is shown in FIG. 8. In Figure 9, there are three groups of comparison images, each group shows the original image, the corresponding depth map, the Ground Truth segmentation results (GT), and the segmentation results obtained by the TMLCOD model (Ours). In the second and third rows, the effectiveness of the TMLCOD method in the face of blurred background and complex texture is demonstrated, and it can be seen that the boundary of the detection result is also smoothed by improving the loss function. Therefore, by combining RGB and depth information, the TMLCOD model can more accurately identify and segment camouflage targets, and maintain high accuracy even when the target and the background have similar colors.

Synthesizing the analysis results in Figs. 7 and 8, the following conclusions can be drawn: Firstly, the proposed model shows excellent generalization ability and classification performance on the test set, which benefits from the reasonable design of the model structure and the effectiveness of the optimization algorithm. Secondly, the model training process is stable, and the parameters are adjusted properly, which provides a solid foundation for subsequent model optimization. Although the model already performs well, there is still room for further improvement. Future work can focus on exploring different model architectures and training strategies to further improve the performance of the model.

FIG. 9 shows the comparison of segmentation results of different methods in the COD task. The figure includes Ground Truth (GT), our method (Ours), and several other representative methods, such as PFNet, ZoomNet, SegMaR, UGTR, MGL, etc. Each row corresponds to a test sample, showing the original image, the truth map, and the segmentation results of different methods.

From the figure, it can be found that the TMLCOD model method can generate results closer to the true segmentation (GT) in multiple scenes, especially when dealing with samples with complex backgrounds and occlusions. For example, in the samples in the first and third columns, other methods lead to incomplete or inaccurate segmentation results because the background is too similar to the target, while the TMLCOD method can better identify the boundary and details of the target.

V. Conclusion

Based on the DaCOD network architecture, this paper systematically optimizes the perception ability and computational efficiency of the RGB-D camouflage object detection model. Firstly, by designing a dynamic weight fusion mechanism and introducing linear attention, the flexibility and pertinence of the DATA module for multi-dimensional feature interaction are enhanced, so that it can dynamically focus on key spatial regions and significantly improve the feature discrimination ability in complex scenes. Secondly, the Efficient Attention mechanism is introduced to replace the traditional self-attention calculation, and the strategy of using convolutions with different expansion rates to capture multi-scale information and adjusting the feature contributions of different branches through dynamic weights is combined to reduce the computational complexity of MSA module and strengthen the dynamic fusion ability of multi-scale features. Finally, the loss function design was improved, and the dynamic weight adjustment mechanism was introduced, which effectively alleviated the training bias and boundary ambiguity problems. The improved model shows broad application

prospects in the fields of military reconnaissance, ecological protection, medical image analysis, and security monitoring. For example, in military scenarios, this technology can accurately identify natural camouflaging targets (such as camouflage equipment or hidden fortifications) and improve battlefield situation awareness. In ecological monitoring, it can assist in tracking wild animals with high camouflage ability and promote non-intrusive biodiversity protection. In the medical field, its fine-grained feature extraction ability is helpful in detecting early lesions or tiny abnormal structures in medical images. In the future, this technology can be further combined with edge computing and multi-modal perception (such as infrared or radar data) to achieve real-time and lightweight embedded deployment and provide robust technical support for complex dynamic scenes such as autonomous driving and intelligent security. The experimental results show that the proposed method has the advantages of both performance and efficiency, which lays an important foundation for the practicality and generalization of camouflage target detection technology.

REFERENCES

[1] X. Zhang, W. Cui, Y. Tao, and T. Shi, "Steel surface defect detection algorithm based on s-yolov8," *IAENG International Journal of Computer Science*, vol. 52, no. 3, pp. 644–652, 2025.

[2] T.-N. Le, T. V. Nguyen, Z. Nie, M.-T. Tran, and A. Sugimoto, "Anabranch network for camouflaged object segmentation," *Computer Vision and Image Understanding*, vol. 184, pp. 45–56, 2019.

[3] D.-P. Fan, G.-P. Ji, G. Sun, M.-M. Cheng, J. Shen, and L. Shao, "Camouflaged object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2777–2787, 2020.

[4] Y. Sun, G. Chen, T. Zhou, Y. Zhang, and N. Liu, "Context-aware cross-level fusion network for camouflaged object detection," *arXiv preprint arXiv:2105.12555*, 2021.

[5] Y. Pang, X. Zhao, T.-Z. Xiang, L. Zhang, and H. Lu, "Zoomnext: A unified collaborative pyramid network for camouflaged object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[6] J. Zhang, Q. Liang, Q. Guo, J. Yang, Q. Zhang, and Y. Shi, "R2net: Residual refinement network for salient object detection," *Image and Vision Computing*, vol. 120, p. 104423, 2022.

[7] Q. Wang, J. Yang, X. Yu, F. Wang, P. Chen, and F. Zheng, "Depth-aided camouflaged object detection," in *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 3297–3306, 2023.

[8] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10012–10022, 2021.

[9] M. Shafiq and Z. Gu, "Deep residual learning for image recognition: A survey," *Applied Sciences*, vol. 12, no. 18, p. 8972, 2022.

[10] D. Misra, T. Nalamada, A. U. Arasanipalai, and Q. Hou, "Rotate to attend: Convolutional triplet attention module," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3139–3148, 2021.

[11] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, 2018.

[12] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 3–19, 2018.

[13] R. Li, J. Su, C. Duan, and S. Zheng, "Linear attention mechanism: An efficient attention for semantic segmentation," *arXiv preprint arXiv:2007.14902*, 2020.

[14] B. Yin, X. Zhang, D.-P. Fan, S. Jiao, M.-M. Cheng, L. Van Gool, and Q. Hou, "Camoformer: Masked separable attention for camouflaged object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[15] Z. Shen, M. Zhang, H. Zhao, S. Yi, and H. Li, "Efficient attention: Attention with linear complexities," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3531–3539, 2021.

[16] B. Dong, P. Wang, H. Luo, and F. Wang, "Adaptive query selection for camouflaged instance segmentation," in *Proceedings of the 32nd ACM International Conference on Multimedia*, pp. 6598–6606, 2024.

[17] M.-M. Cheng and D.-P. Fan, "Structure-measure: A new way to evaluate foreground maps," *International Journal of Computer Vision*, vol. 129, pp. 2622–2638, 2021.

[18] D.-P. Fan, G.-P. Ji, X. Qin, and M.-M. Cheng, "Cognitive vision inspired object segmentation metric and loss function," *Scientia Sinica Informationis*, vol. 6, no. 6, p. 5, 2021.

[19] R. Margolin, L. Zelnik-Manor, and A. Tal, "How to evaluate foreground maps?," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2014.

[20] T. O. Hodson, "Root mean square error (rmse) or mean absolute error (mae): When to use them or not," *Geoscientific Model Development Discussions*, vol. 2022, pp. 1–10, 2022.

[21] Q. Zhai, X. Li, F. Yang, C. Chen, H. Cheng, and D.-P. Fan, "Mutual graph learning for camouflaged object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12997–13007, 2021.

[22] Q. Zhai, X. Li, F. Yang, C. Chen, H. Cheng, and D.-P. Fan, "Mutual graph learning for camouflaged object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12997–13007, 2021.

[23] H. Mei, G.-P. Ji, Z. Wei, X. Yang, X. Wei, and D.-P. Fan, "Camouflaged object segmentation with distraction mining," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8772–8781, 2021.

[24] F. Yang, Q. Zhai, X. Li, R. Huang, A. Luo, H. Cheng, and D.-P. Fan, "Uncertainty-guided transformer reasoning for camouflaged object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4146–4155, 2021.

[25] Y. Lv, J. Zhang, Y. Dai, A. Li, B. Liu, N. Barnes, and D.-P. Fan, "Simultaneously localize, segment and rank the camouflaged objects," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11591–11601, 2021.

[26] D.-P. Fan, G.-P. Ji, M.-M. Cheng, and L. Shao, "Concealed object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 10, pp. 6024–6042, 2021.

[27] M. Zhang, S. Xu, Y. Piao, D. Shi, S. Lin, and H. Lu, "Preynet: Preying on camouflaged objects," in *Proceedings of the 30th ACM International Conference on Multimedia*, pp. 5323–5332, 2022.

[28] Y. Pang, X. Zhao, T.-Z. Xiang, L. Zhang, and H. Lu, "Zoom in and out: A mixed-scale triplet network for camouflaged object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2160–2170, 2022.

[29] Q. Jia, S. Yao, Y. Liu, X. Fan, R. Liu, and Z. Luo, "Segment, magnify and reiterate: Detecting camouflaged objects the hard way," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4713–4722, 2022.