Foreground Feature-Guided Camouflage Image Generation

Yuelin Chen[1], Yuefan An[2], Yonsen Huang[3], Xiaodong Cai[4]
School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin, China[1, 2]
School of Information and Communication, Guilin University of Electronic Technology, Guilin, China[3, 4]

***Abstract*—In the field of visual camouflage, generating a high-quality background image that seamlessly blends with complex foreground objects and diverse background environments is a critical task. When dealing with such complex scenes, the existing techniques have insufficient foreground feature extraction, resulting in insufficient fusion of the generated background image with the foreground objects, making it difficult to achieve the desired camouflage effect. In order to solve this problem and achieve the goal of higher quality visual camouflage effect, this paper proposes a new foreground feature-guided camouflage image generation method (Object Enhancement Module - Diffusion Refinement , OEM-DR), which generates camouflage images by enhancing the foreground features to guide the background. The method firstly designs a new object enhancement module to optimize the attention mechanism of the model, and eliminates the attention weights that have less influence on the output through pruning strategy, so that the model focuses more on the key features of the foreground objects, and thus guides the generation of the background more effectively. Second, a novel detail optimization framework based on diffusion strategy is constructed, which maintains the integrity of the global structure of the image while performing fine optimization processing on the local details of the image. In experiments on standard camouflaged image datasets, the proposed method in this study achieves significant improvement in both FID (Fréchet Inception Distance) and KID (Kullback-Leibler Divergence) evaluation metrics, which verifies the feasibility of the method. This suggests that by strengthening foreground features and detail optimization, the fusion between background images and foreground objects can be effectively improved to achieve higher quality visual camouflage effects.**

Keywords*—*Camouflage image; foreground features; object enhancement; detail optimization

I. Introduction

In the field of visual perception, camouflage image generation is a challenging task that aims at generating background images that can skillfully mask foreground objects for visual concealment. This technique plays an important role in several practical application areas such as pest detection, healthcare [2], and autonomous driving [8]. With the advancement of computer vision techniques, especially in the fields of style migration [9], image editing [10] and image generation [11], new ideas have been provided to address the challenges of camouflage image generation. The Poisson image editing method proposed by Di Martino et al. [10] brought innovations in the field of image editing by allowing researchers to work with the image in a natural and intuitive manner content. The pioneering work of Chu et al. [12] on camouflage image generation demonstrated how to generate hard-to-detect images by mimicking the camouflage mechanisms of natural organisms. The work of Huang and Belongie [9] further advanced the development of style-migration techniques, which allow us to change the style of an image to fit different contexts while keeping its content intact. Zhang et al. [13] proposed generating camouflaged images that can blend in with complex backgrounds by learning a large amount of natural image data. Li et al. [14] further proposed a camouflaged image generation network that does not require specific positional information. The work of Zheng et al. [15] provides a new solution for high fidelity image complementation by bridging global contextual interactions. Lugmayr et al. [16] provided a new idea for background complementation of camouflaged images by using denoising diffusion probabilistic model for image restoration.

However, despite the many advancements in existing technologies, several key issues remain. Firstly, most methods rely on manually selected backgrounds, which not only limits the diversity of generated samples but also significantly increases the cost of data collection. Secondly, these methods may perform poorly in complex and variable environments, as they often depend heavily on the precise extraction of background and foreground features. The LAKE-RED model [17], although innovative in generating camouflage images by fusing training backgrounds with extracted foreground features, may face challenges in complex or changing environments due to its reliance on precise feature extraction.

To address these issues, this study proposes a new object enhancement strategy. Inspired by the work of Dhariwal P et al. [19], this study utilizes weight sparsification and pruning to enhance the model's understanding and learning of deep features of target objects. This strategy aims to reduce dependence on precise feature extraction and improve the model's adaptability in complex environments.

Furthermore, camouflage images often lack detail optimization during generation, which can weaken their camouflage effect. Unnatural transitions in texture, color, or edge areas may reduce their concealment. To solve this, inspired by the work of Yang L et al. [18], this study designs a method to enhance the model's feature expression ability using non-zero weights after diffusion pruning, optimizing the connection between foreground and background to improve the quality of camouflage images.

In the following sections of this paper, the study will be explored in depth. The related work in Section II will review and analyze the technologies and methods involved in this study, clarifying its position and value in the field of camouflage image generation. The OEM-DR in Section III will detail the proposed new method, including the design and implementation of the object enhancement strategy, and how non-zero weights after diffusion pruning enhance the model's feature expression. The experiments in Section IV will design and conduct a series of experiments to verify the effectiveness and performance of the proposed method, including comparative and ablation experiments with existing methods. Finally, the conclusion in Section V will summarize the main achievements and contributions of this study, discuss existing limitations, and suggest directions for future work.

II. Related Work

A. Camouflage Image Generation

The goal of camouflage image generation is to create images of target objects that blend highly with the background and are difficult to detect. Early methods mainly relied on image processing techniques. For example, Chu et al. [12] hid the target by inserting similar textures and colors through image editing. Although these methods could achieve basic camouflage, the quality and diversity of the generated images were limited. With the development of deep learning, methods based on deep generative models have become mainstream. Zhang et al. [13] proposed a deep camouflage image generation model that learns complex camouflage patterns and generates more realistic images. Li et al. [14] developed a camouflage generation network without location information, which automatically learns the relationship between the target and the background, thereby improving the quality of the generated images. In the latest research, Zhao et al. [17] introduced the LAKE-RED method, which combines latent background knowledge retrieval with diffusion models to generate high-fidelity and diverse camouflage images. Yang et al. [18] provided a comprehensive review of diffusion models, covering their applications in camouflage image generation and other fields, offering valuable references for research. These advancements indicate that deep learning, especially diffusion models, holds great potential in camouflage image generation. Future research can further explore their combination with camouflage tasks to generate higher quality and more diverse camouflage images.

B. Data Augmentation

Data augmentation is crucial for enhancing the performance of deep learning models, especially when the number of samples is limited. It improves the model's generalization ability by generating more training samples. In the tasks of camouflage image detection and segmentation, data augmentation is particularly important because it is difficult to obtain camouflage images and their patterns are diverse. Fan et al. [1] enhanced the data by randomly inserting camouflage targets, which is a simple but effective way to improve detection performance. Le et al. [4] explored data augmentation in their Anabranch network, such as random cropping and rotation, but mainly focused on geometric transformations and simple editing, which had limited enhancement of the diversity of camouflage patterns. The development of diffusion models has brought new ideas to data augmentation. Dhariwal and Nichol [19] proposed an image synthesis method based on diffusion models, which generates high-quality images and performs excellently in multiple tasks, providing a new direction for camouflage image generation and data augmentation. Binkowski et al. [20] improved the MMD GAN training method, enhancing the quality and diversity of the generated images. Heusel et al. [21] introduced a two-time-scale update rule for GAN training, proving that it converges to a local Nash equilibrium, providing theoretical support for the training of generative models. These studies show that combining advanced generative models such as diffusion models with the task of camouflage image generation can generate richer and more diverse training data. Future research can further explore this combination to improve the performance of camouflage image detection and segmentation models.

III. OEM-DR Model

A. An Overview of the Proposed Framework

As shown in Fig. 1, the OEM-DR model starts from low-level image feature extraction and highlights important details in the image through the object enhancement module, which introduces sparsity to optimize the weight distribution, and then employs a pruning strategy to reduce the computational burden. The detail optimization module further enhances the image quality, which facilitates the fusion and generalization of the model to new data while retaining key information through the diffusion process. Finally, the model performance is tuned and optimized through the computation of the loss function. The specific details of these steps are elaborated in this section.

B. Object Enhancement Based on Sparse Pruning

In order to address the model's deficiency in object information learning, an object enhancement module is designed in this study. The purpose of this module is to deepen the model's understanding and learning of deep features of the target object through refined feature selection and enhancement. This study aims to further encode the object features through this enhancement module to achieve deeper understanding of foreground feature information in images. The working mechanism of this module is explained in detail in the following section, and its structure is detailed in Fig. 2.

In the forward propagation process of the model, the deep representation of foreground features f_g is extracted by a multilayer perceptron (MLP). the pre-trained background embedding matrix is subsequently transposed and the batch dimension is increased in order to obtain the background embedding vector $\boldsymbol{C}$. The query vector q is obtained by applying a linear transformation layer to f_g .$\boldsymbol{C}$ is used as the key k and the value v. The similarity matrix $\boldsymbol{S}_m$ is computed by using the einsum function for querying q and key k. matrix $\boldsymbol{S}_m$, which is computed as follows:

$$\boldsymbol{S}_m = einsum(q,k) \tag{1}$$

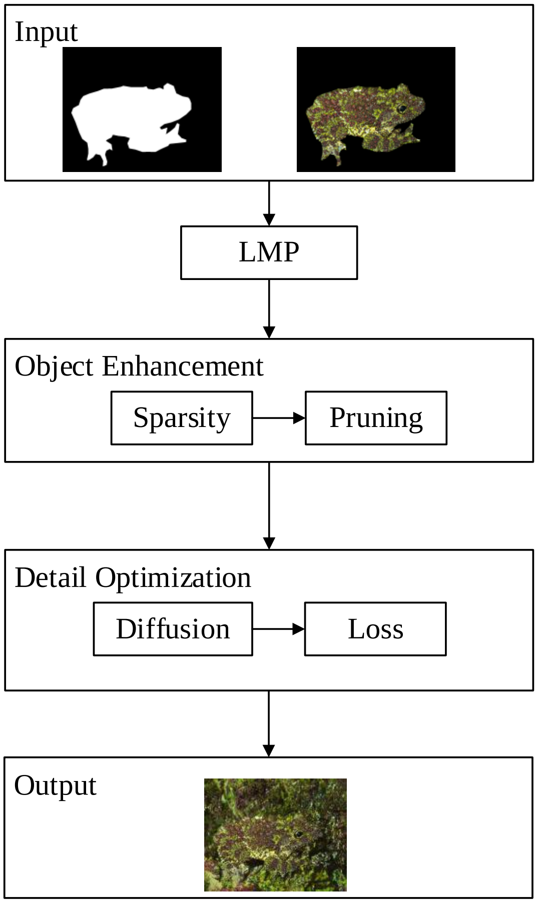

Fig. 1. OEM-DR model framework.

In this paper, we quantify sparsity by calculating the attentional weight of $\boldsymbol{S}_m$ and subsequently determining the pruning threshold. The $\boldsymbol{L}_1$ norm l_{n1} and $\boldsymbol{L}_2$ norm l_{n2} of the attention weight matrix $\boldsymbol{A}$ are computed with the following equations:

$$l_{n1} = \sum_{i=1}^{n} |a_i| \tag{2}$$

$$l_{n2} = \sqrt{\sum_{i=1}^{n} a_i^2} \tag{3}$$

Where a_i is the i-th element in the attention weight matrix $\boldsymbol{A}$ and n is the total number of elements.

The sparsity ratio $\boldsymbol{R}$ is defined as follows:

$$\boldsymbol{R} = \frac{l_{n1}}{l_{n2} + \varepsilon} \tag{4}$$

Where ε is a negligible positive number added to prevent division by zero.

A Boolean mask $\boldsymbol{P}$ is generated using the sparse ratio $\boldsymbol{R}$. Each element of this mask is determined by the condition:

$$\boldsymbol{P} = \begin{cases} \boldsymbol{T}, \boldsymbol{A} > \boldsymbol{R} \\ \boldsymbol{F}, \boldsymbol{A} \le \boldsymbol{R} \end{cases} \tag{5}$$

The initial pruning is accomplished by setting the elements of $\boldsymbol{A}$ for which $\boldsymbol{P}$ is $\boldsymbol{F}$ to zero through the product operation, as follows:

$$\boldsymbol{A}_p = \boldsymbol{A} \odot \boldsymbol{P} \tag{6}$$

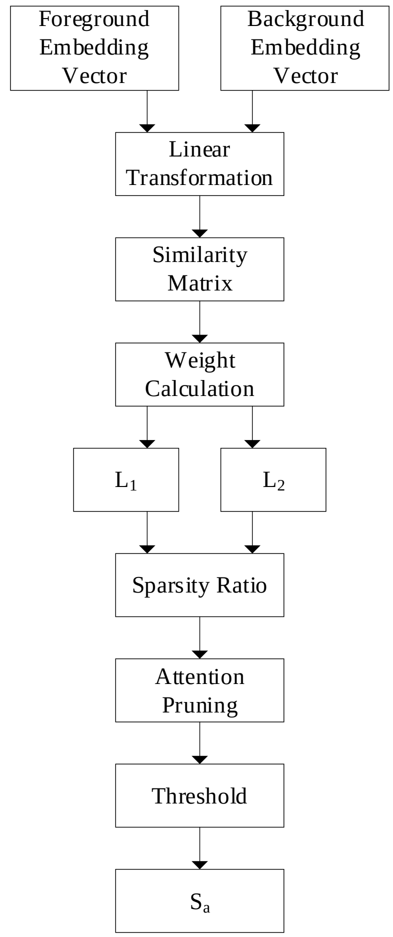

Fig. 2. Object enhancement module structure diagram.

Further, the $\boldsymbol{L}_1$ norm $\boldsymbol{A}_{p1}$ of the pruned attention weight $\boldsymbol{A}_p$ is calculated using Eq. (2) and the pruning threshold $\boldsymbol{T}_p$ for each attention head is determined by the pruning ratio $\boldsymbol{P}_r$ as follows:

$$\boldsymbol{M}_{AP} = \max(\boldsymbol{A}_{p1}) \tag{7}$$

$$\boldsymbol{T}_p = \boldsymbol{M}_{AP} \times \boldsymbol{P}_r \tag{8}$$

$\boldsymbol{M}_{AP}$ in Eq. (7) represents the maximum value of the computed $\boldsymbol{L}_1$ norm on the last dimension of the weight matrix $\boldsymbol{A}_p$.

The pruned weight matrix $\boldsymbol{S}_a$ is then obtained by generating a Boolean mask using the pruning threshold $\boldsymbol{T}_p$ through the operations of Eq. (5) and Eq. (6), and then the elements of $\boldsymbol{A}_p$ Ap below $\boldsymbol{T}_p$ are set to zero through the product operation in order to complete the pruning process.

C. Diffusion-based Detail Optimizer

In order to cope with the performance degradation caused by performing pruning, this study adds the DR module after pruning. the purpose of the DR module is to maintain or even improve the predictive performance of the model after pruning, and to enhance the generalization ability of the model to ensure the model's adaptability and accuracy to new data.

Referring to Fig. 3, the input to the DR module is the pruned weight matrix $\boldsymbol{S}_a$. The influence of the non-zero weights retained after pruning is spread to the adjacent zero-weight regions through an iterative process, with a view to maintaining or even enhancing the generalization capability of the model while reducing the complexity of the model. The details of the operation are described below.

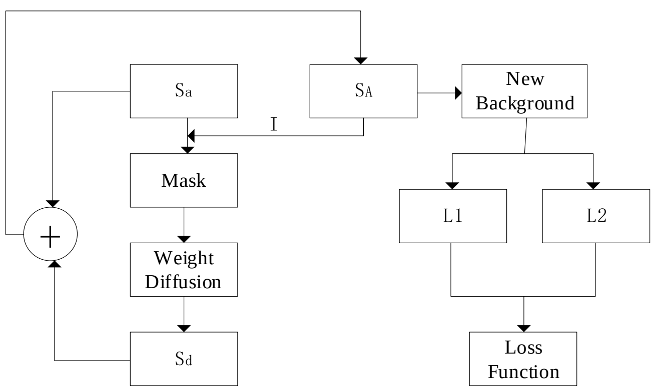

Fig. 3. Diffusion optimizer structure diagram.

First, a Boolean mask $\boldsymbol{M}$ is initialized to identify the location of the non-zero weights in $\boldsymbol{S}_a$. This mask is obtained by comparing $\boldsymbol{S}_a$ with zero, where the non-zero element corresponds to a mask value of Ture.

Next, $\boldsymbol{I}$ iterations are executed, and in each iteration the following steps are performed:

First $\boldsymbol{M}$ is shifted one position to the right along the last dimension (feature dimension) to obtain the new mask $\boldsymbol{M}_n$. This step simulates the process of diffusion of non-zero weights to the right along the feature dimension.

Next, the update of the weights is computed as follows:

$$\boldsymbol{S}_d = \boldsymbol{S}_a \times (1 - \boldsymbol{D} \times (1 - \boldsymbol{M}_n)) \tag{9}$$

$\boldsymbol{D}$ is a factor between 0 and 1 used to adjust the magnitude of the weight update.

The calculated update is then applied to $\boldsymbol{S}_a$ and the updated weights are:

$$\boldsymbol{S}_A = \boldsymbol{S}_d + \boldsymbol{S}_a \tag{10}$$

Finally, a non-negative constraint is imposed to ensure that all weight values are non-negative. After $\boldsymbol{I}$ iterations, the post-diffusion weight matrix $\boldsymbol{S}_A$ is obtained.

In the second step, $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ regularization losses are introduced to process the new background $\boldsymbol{P}_n$ obtained after training with the features $\boldsymbol{T}_a$ of the target image as a way to achieve fine tuning of the model complexity. The new foreground image $\boldsymbol{P}_n$ is generated in the following way: in the background region identified by the mask, this paper replaces the original foreground *f* with the background feature $\boldsymbol{S}_A$; while in the foreground region, f is kept unchanged. Thus, the updated foreground image $\boldsymbol{P}_n$ is obtained. The $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ regularization losses operate as follows:

The $\boldsymbol{L}_1$ loss measures the error by calculating the absolute value of the difference between the predicted value and the target value, and its mathematical expression is shown below:

$$\boldsymbol{L}_1 = \lambda_1 \sum_{i=1}^{n} | w_i - w_i^{\boldsymbol{T}_a} | \tag{11}$$

The $\boldsymbol{L}_2$ loss measures the error by calculating the square of the difference between the predicted value and the target value, the mathematical expression of which is shown below:

$$\boldsymbol{L}_2 = \lambda_2 \sum_{i=1}^{n} (w_i - w_i^{\boldsymbol{T}_a})^2 \tag{12}$$

The w_i denotes the actual value of the i-th image, $w_i^{\boldsymbol{T}_a}$ denotes the target value of the i-th image, and λ_1 and λ_2 denote the weights occupied by the two losses, respectively.

The $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ losses are combined into the total loss according to certain weights to realize the joint control of model complexity. The combined total loss function is defined as:

$$L = L_1 + L_2 \tag{13}$$

The main role of $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ losses is to prevent model overfitting and improve model generalization. Specifically, the $\boldsymbol{L}_1$ loss helps with feature selection, while the $\boldsymbol{L}_2$ loss helps with parameter stability.

IV. EXPERIMENT

A. Datasets

Following the research on covert object detection (COD) [3], this paper uses 4040 real images (3040 from the COD10K [1] dataset and 1000 from the CAMO [26] dataset) to train the model. To validate the generation performance, this study collects image-mask pairs from different domains and constructs a test dataset consisting of three subsets: covert objects (CO), salient objects (SO) and general objects (GO) [17]. In CO, there are 6473 image pairs from CAMO [4], COD10K [1] and NC4K [5]. We then randomly selected 6473 images from well-known salient target detection datasets (e.g., DUTS [6], DUT-OMRON [7], etc.) and segmentation datasets to evaluate the performance of the model on open domain data.

B. Evaluation Metrics and Parameter Setting

Following the good practices of AIGC and COD, the InceptionNet-based metrics FID [20] and KID [17] are chosen in this paper to measure the quality of the generated covert images. Once the image features have been extracted via InceptionNet, the distance between them is calculated to indicate the level of similarity between the model output and the target dataset. Unlike normal images, well-synthesized covert images should not contain easily recognizable objects, and extracting discriminative features is more challenging. A lower value of FID [2] indicates that the generated image is more similar to the real image in terms of visual features, which usually implies a better generation. KID [21] by kernel method is able to capture more sensitively the differences between the generated image and the real image, especially in terms of image detailed features. A lower value of KID [21] indicates that the quality of the generated image is closer to the real image.

In this paper, we use a latent diffusion model, pre-trained on a restoration task as an initialization. The model was designed to handle images and masks of size 512 × 512 and compressed to a potential space of 128 × 128 × 3 using the pre-trained VQ-VAE. During the training process, the focus is on improving the model's understanding and learning of the deep features of the target object through object augmentation and regularized diffusion strategies, and does not fine-tune the autoencoder and decoder components. The existing conditions are optimized and enhanced by the modules proposed in this paper. Parameter optimizations, such as initialization, data enhancement, and batch size, are set similar to the original paper. Finally, the model generates the artifact images and resizes them to align with the original input. This paper uses PyTorch for all experiments and a GeForce RTX 4060ti GPU for all experiments.

C. Comparison and Analysis of Model Results

To verify the effectiveness of the OEM-DR model, this paper selects the following nine models for comparison: The AB model [9] achieves seamless image region blending based on the Poisson equation, enabling the source image to naturally blend into the target background. The AdaIN model [12] rapidly accomplishes real-time arbitrary style transfer through adaptive instance normalization technology, converting the style while maintaining the image content unchanged. In the field of camouflage image generation, the CI model [11] imitates the camouflage mechanism of natural organisms to generate images that are difficult to detect; the DCI model [13] and LCGNet model [14] utilize deep learning to generate camouflage images that blend with complex backgrounds, with LCGNet being more flexible as it does not require specific location information; the LAKE-RED model [17] combines potential background knowledge retrieval with diffusion models to enhance the quality of camouflage images. In addition, the TFill model [15] achieves high-fidelity image completion by bridging global context interaction, while the RePaint-L model [16] performs image restoration based on denoising diffusion probabilistic models, generating content for missing areas. The LDM model [10] proposes a high-resolution image synthesis method based on latent diffusion models, promoting the development of high-quality image generation technology.

Table I shows the comparison of experimental results of various models on the dataset. Compared with the AB, CI, AdaIN, DCI, LCCNet models, which realize the camouflage effect by fusing the background with the foreground, EM-DR guides the background generation according to the features of the foreground, so that the generated image background is closer to the foreground and the foreground is more hidden.

TABLE I. COMPARISON OF EXPERIMENTAL RESULTS IN VARIOUS MODELS

Methods		Input	Camouflaged Objects		Salient Objects		General Objects		Overall	
			FID↓	KID↓	FID↓	KID↓	FID↓	KID↓	FID↓	KID↓
Image Blending	AB [9]	F +B	117.11	0.0645	126.78	0.0614	133.89	0.0645	120.2	0.0623
	CI [11]	F +B	124.49	0.0662	136.3	0.738	137.19	0.0713	128.5	0.0693
	AdaIN [12]	F +B	125.16	0.0721	133.2	0.0702	136.93	0.0714	126.9	0.0703
	DCI [13]	F +B	130.21	0.0689	134.92	0.0665	137.99	0.069	130.5	0.0673
	LCGNet [14]	F +B	129.8	0.0504	136.24	0.0597	132.64	0.0548	129.9	0.055
Image Inpainting	TFill [15]	F	63.74	0.0336	96.91	0.0453	122.44	0.0747	80.39	0.0438
	LDM [10]	F	58.65	0.038	107.38	0.0524	129.04	0.0748	84.48	0.0488
	RePaint-L [16]	F	76.8	0.0459	114.96	0.0497	136.18	0.0686	96.14	0.0498
	LAKE-RED[17]	F	39.55	0.0212	88.7	0.0428	102.67	0.0625	64.27	0.0355
	Ours	F	37.44	0.0181	86.9	0.0387	100.48	0.0581	61.52	0.0311

Compared with TFill and LDM models, EM-DR strengthens the feature screening of foreground features, mainly in its ability to refine and optimize, which gives it an advantage in preserving the details of important foreground objects in the image. TFill and LDM, on the other hand, while having their advantages in texture synthesis and learning-based approaches, may require additional tuning or optimization to better handle foreground features in specific cases. Compared to the benchmark model LAKE-RED, EM-DR achieves significant improvements in all metrics, further demonstrating its ability to adequately enhance foreground features when generating camouflaged images, thus providing foreground camouflage effects.

Compared with the baseline model on the CO dataset, the FID metrics improved by 5.3% and the KID metrics improved by 14.6%, on the SO model, the FID metrics improved by 2% and the KID metrics improved by 9.5%, on the GO model, the FID metrics improved by 2.1% and the KID metrics improved by 7%, and on the whole the FID metrics improved by 4.3% and the KID metrics improved by 12.3%. by 4.3% for FID and 12.3% for KID. These improvements indicate that the EM-DR model shows better performance in the task of generating camouflage images, and the improvements in FID and KID metrics indicate that the images generated by the new model are closer to the real images in terms of visual and statistical properties.

D. Ablation Study

In order to verify the impact of the object enhancement module and diffusion optimizer mentioned in this paper on the model performance, ablation experiments are conducted on datasets containing CO, SO and GO, and the results are summarized in Table II.

TABLE II. ABLATION STUDY RESULTS

Methods		**OEM**	**DR**
CO	FID↓	37.67	37.77
	KID↓	0.0183	0.0187
SO	FID↓	85.85	87.92
	KID↓	0.0381	0.0391
GO	FID↓	102.07	102.57
	KID↓	0.0599	0.0599
Overall	FID↓	61.73	62.51
	KID↓	0.0313	0.0317

First, the OEM module is individually validated in this paper. Compared with the LAKE-RED model, OEM achieves significant improvement in both FID and KID, especially on the SO dataset. This indicates that OEM not only makes the camouflage images generated on hidden objects as well as ordinary objects close to the real image in terms of enhanced foreground features, but also generates camouflage images on salient objects even closer to the real image.

Secondly, the DR module is individually validated in this paper. Compared to the LAKE-RED model, DR also achieves significant improvement in both FID and KID but lacks in SO dataset compared to the OEM module.

In summary, the OEM and DR modules show excellent advantages in the camouflage image generation task, effectively improving the model performance.

E. Example of Analysis

In order to visualize the improvement of the OEM-DR model mentioned in this paper compared to the benchmark model LAKE-RED, several examples are selected for visualization.

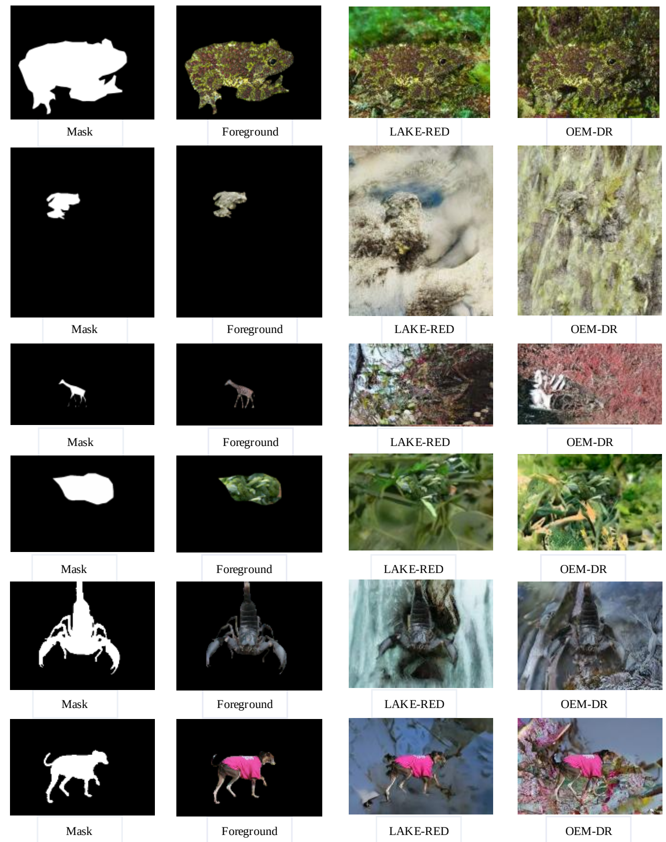

Fig. 4. OEM-DR Model visualization examples.

Fig. 4 visualizes the performance of the OEM-DR model proposed in this study compared with other benchmark models in generating camouflaged images. The figure groups the images according to different model sources: Fig. 4 (a) and 4 (b) are generated by the CO model, Fig. 4 (c) and 4 (d) by the SO model, while Fig. 4 (e) and 4 (f) are from the GO model. Taking the toad in Fig. 4 (a) as an example, the image generated by the OEM-DR model shows a more natural fusion between the background and the foreground, and the details of the background are optimized so that the foreground objects do not appear to be abrupt, which achieves a better camouflage effect visually. Similarly, in Fig. 4 (f), although the camouflaged objects are not completely hidden, the OEM-DR model still demonstrates its advantages in improving the overall image quality. The comprehensive comparison results show that the OEM-DR model has a significant advantage in generating images that contribute to effective camouflage, and is able to generate high-quality camouflage images that are more natural and rich in details.

Since both image generation quality and camouflage effectiveness need to be perceived by humans, this paper conducts a user study to obtain subjective judgments on the generation results. In this paper, 10 researchers in AI related fields are invited to judge which of the images in Fig. 4 are not easily detectable. According to the judgment results except Fig. 4 (b) all others are considered to be less detectable in the foreground in the images generated by OEM-DR, which fully demonstrates the quality of OEM-DR in image generation and the effectiveness of the camouflage effect.

F. Discussion

In this study, the OEM-DR model significantly enhanced the performance of camouflage image generation through the object enhancement module and the diffusion optimizer. Compared with various existing models, OEM-DR excelled in terms of FID and KID metrics, particularly when dealing with salient objects. Although the model has the limitation of high computational resource demands, this is expected to be overcome in the future through algorithm optimization and hardware acceleration. The OEM-DR model provides a new perspective and methodology for the field of camouflage image generation, holding broad prospects for application.

V. Conclusion

In this paper, an innovative camouflage image generation method based on sparse pruning and weight diffusion is proposed. First, a new object enhancement module is designed, which effectively improves the extraction of key features in the foreground by the model, and thus significantly improves the authenticity of the camouflage image generation. Second, a detail optimization module based on the weight diffusion strategy is constructed, which improves the quality of the generated images by optimizing the background image generation process. In the experiments on CO, SO and GO datasets, this paper demonstrates excellent performance. Future research in camouflage visual perception will be continued to further propose feasible solutions.

However, it is worth noting that although this study has achieved certain results in the field of camouflage image generation, it faces the challenge of balancing computational complexity and efficiency during the implementation of the weight diffusion strategy. In future research work, this study will be committed to optimizing the adaptive weight diffusion strategy, dynamically adjusting the diffusion parameters according to the training progress of the model and the unique characteristics of the dataset. The exploration and optimization in this direction are aimed at promoting the continuous progress and development of camouflage image generation technology.

Acknowledgment

National Natural Science Foundation of China（No.62001133），Fund Project of Guangxi Key Laboratory of Wireless Broadband Communication and Signal Processing（GXKL06200114).

References

[1] Fan D P, JI G P, SUN G, et al. Camouflaged Object Detection[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 2774-2784.

[2] JI G P, ZHANG J, CAMPBELL D L, et al. Rethinking polyp segmentation from an out-of-distribution perspective[J]. Machine Intelligence Research, 2024, 21(4): 631-639.

[3] FAN D P, JI G P, CHENG M M, et al. Concealed Object Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(10): 6024-6042.

[4] LE T N, NGUYEN T V, NIE Z L, et al. Anabranch network for camouflaged object segmentation[J]. Computer Vision and Image Understanding, 2019, 184(1): 45-56.

[5] LV Y Q, ZHANG J, DAI Y C, et al. Simultaneously Localize, Segment and Rank the Camouflaged Objects[C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE,2021: 11586-11596.

[6] WANG L J, LU H C, WANG Y F, et al. Learning to Detect Salient Objects with Image-Level Supervision[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI: IEEE, 2017: 3796-3805.

[7] YANG C, ZHANG L H, LU H C, et al. Saliency detection via graph-based manifold ranking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR: IEEE, 2013: 3166-3173.

[8] WANG J Y, QI Y. Multi-task learning and joint refinement between camera localization and object detection[J]. Computational Visual Media, 2024, doi: 10.1007/s41095-022-0319-z.

[9] DI M J M, FACCIOLO G, MEINHARDT E. Poisson Image Editing[J]. Image Processing On Line, 2016, doi: 10.5201/ipol.2016.163.

[10] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New Orleans,LA,USA:IEEE,2022: 10674-10685.

[11] CHU H K, HSU W H, MITRA N J, et al. Camouflage images[J]. ACM Transactions on Graph, 2010, 29(4): 1-8.

[12] HUANG X, BELONGIE S. Arbitrary style transfer in real-time with adaptive instance normalization[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 1510-1519.

[13] ZHANG Q, YIN G, NIE Y, et al. Deep camouflage images[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York, NY, USA:Association for the Advancement of Artificial Intelligence, 2020: 12845-12852.

[14] LI Y Y, ZHAI W, CAO Y, et al. Location-free camouflage generation network[J]. IEEE Transactions on Multimedia, 2022, 25(7): 5234-5247.

[15] ZHENG C X,SONG G X, CHAM T J, et al. Bridging global context interactions for high-fidelity image completion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE, 2022: 11502-11512.

[16] LUGMAYR A, DANELLJAN M, ROMERO A, et al. Repaint: Inpainting using denoising diffusion probabilistic models [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE, 2022: 11451-11461.

[17] ZHAO P C, XU P, QIN P D, et al. LAKE-RED: Camouflaged images generation by latent background knowledge retrieval-augmented diffusion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2024: 4092-4101.

[18] YANG L, ZHANG Z, SONG Y, et al. Diffusion models: A comprehensive survey of methods and applications[J]. ACM Computing Surveys, 2023, 56(4): 1-39.

[19] DHARIWAL P, NICHOL A. Diffusion models beat gans on image synthesis[EB/OL]. (2021-05-11) [2024-10-23] .https://arxiv.org/abs/2105.05233.

[20] BINKOWSKI M, SUTHERLAND D J, ARBEL M, et al. Demystifying mmd gans[EB/OL].(2018-01-04)[2024-10-23]. https://arxiv.org/abs/1801.01401

[21] HEUSEL M, RAMSAUER H, Unterthiner T, et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium[EB/OL].(2017-06-266)[2024-10-23]. https://arxiv.org/abs/1706.08500