Enhancing Camouflaged Object Detection via Diffusion Model Augmentation

Henry O. Velesaca[1,2] and Angel D. Sappa[1,3](✉)

[1] ESPOL Polytechnic University, Campus Gustavo Galindo, Guayaquil, Ecuador
hvelesac@espol.edu.ec
[2] Software Engineering Department, Research Center for Information and Communication Technologies (CITIC-UGR), University of Granada, 18071 Granada, Spain
[3] Computer Vision Center, 08193 Bellaterra, Spain
sappa@ieee.org

Abstract. This paper proposes a novel approach to improve camouflaged object detection (COD) techniques using a data augmentation strategy based on a diffusion model. COD represents a significant challenge in computer vision, being crucial in applications such as military surveillance, medical image analysis, and ecological studies. Although recent advances in deep learning have improved COD performance, they heavily rely on well-annotated datasets, which are particularly difficult to obtain in the context of camouflaged objects. The proposed method addresses the problem of data sparsity while maintaining crucial features of camouflaged objects and their relationships with the environment. Incorporating a diffusion model into the data augmentation pipeline shows improvement in the performance of the COD models, which varies between the different techniques and the proposed evaluation metrics.

Keywords: Camouflaged object detection · Computer vision · Diffusion models · Pest detection

1 Introduction

Camouflaged Object Detection (COD) represents a significant challenge in computer vision, where the goal is to identify objects that are deliberately or naturally blended with their surroundings [5]. This task is particularly crucial in various applications, including military surveillance, medical image analysis, and ecological studies [17]. Unlike traditional object detection, COD faces unique challenges due to the inherent similarity between target objects and their backgrounds, making it difficult for conventional detection methods to achieve satisfactory results [28]. Recent advances in deep learning have significantly improved COD performance [31]. However, these improvements are heavily dependent on large-scale and well-annotated datasets, which are particularly challenging to obtain in the context of camouflaged objects due to the time-consuming nature

E. Rodolà et al. (Eds.): ICIAP 2025 Workshops, LNCS 16169, pp. 352–364, 2026.
https://doi.org/10.1007/978-3-032-11317-7_30

of data collection and annotation [14]. This limitation often leads to models that underperform in real-world scenarios or fail to generalize across different camouflage patterns and environments.

To address these challenges, data augmentation has emerged as a promising solution [13]. Traditional augmentation techniques, such as geometric transformations and color adjustments, while useful, often fail to capture the complex nature of camouflage patterns and their interaction with the environment [30]. This limitation has motivated the exploration of more sophisticated approaches to data augmentation.

In recent years, diffusion models have demonstrated remarkable capabilities in generating high-quality, diverse images [19]. These models, which learn to gradually denoise images through an iterative process, have shown particular promise in maintaining structural coherence and semantic relationships in generated images [20]. However, their application to the specific challenge of COD data augmentation remains largely unexplored. This paper proposes a methodology to improve COD techniques using data augmentation with diffusion models. Our method addresses the data scarcity problem while maintaining the crucial characteristics of camouflaged objects and their relationships with their surroundings. By incorporating diffusion models into the data augmentation pipeline, we can generate synthetic yet realistic training samples that enhance the robustness and generalization capabilities of COD models.

The manuscript is organized as follows. Section 2 presents work related to camouflaged object detection and data augmentation using classical and diffusion model approaches. Section 3 presents the proposed methodology to improve camouflaged object detection using specialized techniques for this task. Experimental results, comparisons, and discussions with different COD approaches to validate the proposed methodology are presented in Sect. 4. Finally, conclusions are presented in Sect. 5.

2 Background

This section summarizes some of the most relevant techniques related to the camouflaged object detection techniques and data augmentation approaches.

2.1 Camouflaged Object Detection Approaches

Camouflaged Object Detection has emerged as a crucial research field in computer vision, with significant advances in recent years. The early work of Qin et al. [18] on salient object detection significantly influences the development of COD techniques. In 2021, Fan et al. [5] establish the fundamental foundations for concealed object detection, providing a robust theoretical framework for this challenging problem. In 2022, Chen et al. [3] propose a context-aware cross-level fusion-based approach, improving the accuracy of camouflaged object identification, while Chen et al. [4] introduce a boundary-guided network that enhances the precision in detecting edges and features of camouflaged objects. In the same year, Liu et al. [16] address the modeling of aleatoric uncertainty in camouflaged object detection. The field continues to evolve in 2023 with significant contributions such as the high-resolution iterative feedback network presented by Hu et al. [9], and the edge-aware network for COD presented by Sun et al. [24]. In the same year, Ji et al. [11] introduce an innovative deep gradient learning technique that significantly improves the efficiency of camouflaged object detection. More recently, in 2024, Yang et al. [29] expand the field towards specific applications with PlantCamo, focusing on camouflage detection in plants.

These advances contribute to creating more robust and accurate systems for camouflaged object detection, setting new standards in the field and paving the way for future research and practical applications.

2.2 Data Augmentation Approaches

Data augmentation has become a fundamental technique in deep learning, especially in computer vision tasks, where the availability of labeled data may be limited. Classic augmentation techniques include geometric transformations such as rotation, flipping, cropping, and rescaling, as well as color space modifications such as changes in brightness, contrast, and saturation [22]. These basic transformations help artificially increase the size of the training dataset and improve the generalization capabilities of models. A popular tool that has revolutionized the implementation of these techniques is Albumentations [1], a Python library that provides an efficient and optimized interface for applying data augmentation transformations. This library has become a standard in the computer vision community due to its superior performance and wide range of available transformations.

Recent research uses synthetic data and domain adaptation to combat data scarcity. Virtual environments effectively train deep learning models for camera pose estimation [2] and image dehazing [21]. Synthetic datasets aid instance segmentation [26] and agricultural applications like corn kernel classification [23], reducing annotation costs. This approach, combined with traditional augmentation, improves model performance while decreasing reliance on manual labeling.

In a significant advancement in the field of deep learning and data augmentation, Islam et al. [10] present DiffuseMix, a groundbreaking data augmentation technique that preserves labels using diffusion models. This method addresses one of the fundamental challenges in machine learning: generating synthetic data that maintains the integrity of the original labels. DiffuseMix leverages the power of diffusion models to create realistic and meaningful variations of the training data while ensuring that semantic labels remain consistent and accurate. This approach not only improves the diversity of the training dataset but also helps prevent overfitting and improves the robustness of deep learning models. The technique demonstrates promising results on several benchmark datasets, establishing itself as a valuable contribution to the computer vision and machine learning community.

3 Proposed Methodology

This section details the different stages followed to carry out the proposed methodology. Figure 1 shows the overall pipeline of the proposed methodology.

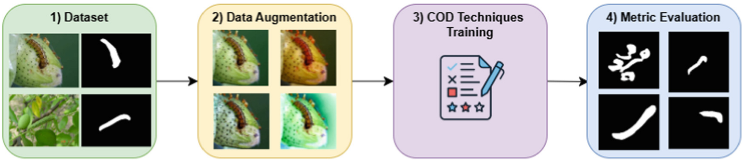

Fig. 1. Overall pipeline of the proposed methodology.

3.1 Data Augmentation

For the proposed work, two data augmentation approaches are used to enrich and diversify the training data set. These approaches are carefully selected to maximize the variability and quality of the training data.

The first approach corresponds to data augmentation using the classical method, specifically Albumentations [1], which include geometric and photometric transformations of the images, such as random rotations, horizontal and vertical flips, brightness and contrast adjustments, random cropping, scale changes, the addition of Gaussian noise and moderate Gaussian blur. These traditional transformations help improve the robustness of the model against basic variations that may occur in real scenarios. For the second data augmentation approach, a more advanced technique based on diffusion models is used, specifically, the work DiffuseMix proposed by Islam et al. [10]. This methodology leverages the generative power of diffusion models to create new synthetic samples that preserve important semantic features, generate realistic variations while maintaining the original labels, introduce significant diversity into the dataset, produce examples that cover cases underrepresented in the original dataset, and maintain coherence between the generated image and its corresponding segmentation mask.

3.2 SOTA COD Techniques

To validate the proposed methodology, SOTA COD techniques are first selected using various state-of-the-art surveys [6,15,28,31] as references. After reviewing the best-performing techniques and available codes, nine camouflaged object detection techniques are selected for fine-tuning using two data augmentation approaches: the classical method Albumentations [1] and the diffusion model-based technique DiffuseMix [10]. It is worth noting that DiffuseMix is not evaluated for segmentation tasks in its original paper.

Table 1. Distinctive features of the evaluated camouflage techniques.

Technique	Year	Image Size (px)	Backbone	#Param. (M)
BASNet [18]	2019	256 × 256	ResNet-34 [8]	87.06
SINet-v2 [5]	2021	352 × 352	Res2Net-50 [7]	24.93
BGNet [4]	2022	416 × 416	Res2Net-50 [7]	77.80
C^{2}F-Net [3]	2022	352 × 352	Res2Net-50 [7]	26.36
OCENet [16]	2022	352 × 352	ResNet-50 [8]	58.17
EAMNet [24]	2023	384 × 384	Res2Net-50 [7]	30.51
DGNet [11]	2023	352 × 352	EfficientNet [25]	8.30
HitNet [9]	2023	352 × 352	PVTv2 [27]	25.73
PCNet [29]	2024	352 × 352	PVTv2 [27]	27.66

The nine SOTA COD techniques evaluated are: BASNet [18], SINet-v2 [5], BGNet [4], C^{2}F-Net [3], OCENet [16], EAMNet [24], DGNet [11], HitNet [9], and PCNet [29]. Table 1 presents a comparative analysis of their architectural characteristics. Models from 2019 to 2024 are considered. Each model is trained for 150 epochs, and a batch size of 16 is considered. Techniques such as SINet-v2, C^{2}F-Net, OCENet, DGNet, HitNet, and PCNet use input images of 352 × 352 pixels, while BASNet, BGNet, and EAMNet use sizes of 256 × 256, 416 × 416, and 384 × 384 pixels, respectively. This dimensional consistency among most models indicates a standardized approach in the field.

The backbone architectures demonstrate significant diversity in their design choices. Res2Net50 serves as the backbone for four techniques (SINet-v2, BGNet, C^{2}F-Net, and EAMNet). Meanwhile, PCNet and HitNet implement PVT-V2 (Pyramid Vision Transformer V2), and DGNet utilizes EfficientNet, showcasing various architectural approaches to COD tasks. In addition, BASNet and OCENet use ResNet-34 and ResNet-50, respectively. The computational complexity, measured in terms of parameter count, varies substantially across these models. DGNet exhibits the most efficient architecture with 8.30 million parameters, while BASNet, BGNet, and OCENet represent the most complex design with 87.06, 77.80, and 58.17 million parameters, respectively. The remaining models (SINet-v2, C^{2}F-Net, EAMNet, HitNet, PCNet) maintain 24–31 million parameter counts, suggesting an optimal complexity range for COD applications.

3.3 Metric Evaluation

This study employs five widely recognized evaluation metrics to evaluate Camouflaged Object Detection (COD) performance. These metrics provide comprehensive assessment criteria for analyzing detection accuracy and effectiveness across different models. The Structure-measure (S_α), weighted F-measure (F_β^w), Mean Absolute Error (M), E-measure (E_ϕ), and F-measure (F_β). The S_α metric quantifies the structural similarity between prediction and ground truth maps.

The F_β^w represents an enhanced evaluation metric that extends the traditional F_β by incorporating spatial weights, providing a better assessment of segmentation quality with emphasis on boundary accuracy and location-based importance of detected pixels. The M metric focuses on pixel-level error evaluation between the normalized prediction and ground truth. The E_ϕ metric simultaneously evaluates the global and local accuracy of COD based on human visual perception mechanisms. The F_β provides a synthetic measure that considers both precision and recall components. For both F-measure and E-measure metrics, different scores can be obtained according to different precision-recall pairs. This leads to the computation of mean F-measure (F_β^{mean}) and maximum F-measure (F_β^{max}). Similarly, the E-measure utilizes maximum and mean variants, denoted as E_ϕ^{mean} and E_ϕ^{max}, which are also employed as evaluation metrics.

On the other hand, to evaluate the quality of the images obtained with both approaches, we use the work proposed by [12]. The evaluation metrics comprise three fundamental aspects: the Reconstruction Fidelity Score ($S_{R_f}^Q$), which measures the ability to preserve the original structure and content of the image, where a value close to 1 indicates excellent conservation of the original features; the Boundary Visibility Score (S_b^Q), which evaluates the quality and definition of the edges in the generated image, where higher values indicate better preservation of object contours; and finally, the Combined Score (S_α^Q), which acts as a comprehensive metric that balances both the reconstruction fidelity and the edge quality, providing an overall assessment of the quality camouflage, where a value close to 1 represents optimal performance in the three metrics.

4 Results and Discussion

This section presents experimental results of the proposed methodology. Model performance evaluation uses the assessment metrics detailed in Sect. 3.3.

4.1 Dataset

To carry out the different experiments in this work, a dataset on the cotton bollworm [17] is used, representing a comprehensive collection of specialized images focusing on one of the most destructive pests of agriculture that poses a significant threat to cotton crops worldwide. Created by Meng et al. [17], this dataset is a vital resource for researchers, specifically designed to address the complex challenges of detecting these elusive pests in various environmental conditions and varied scenes. The dataset, spanning a wide range of visual data capturing cotton bollworms in different contexts and environments, has been made publicly available to the scientific community through the Kaggle platform[1], enabling further research and development in pest detection and the agricultural context. The original dataset consists of 856 images for training, 161 images for validation, and 56 images for testing. While the dataset augmented with (AwA) [1] and

---

[1] https://www.kaggle.com/datasets/kexinmeng1/the-dataset-of-cotton-bollworms.

(AwD) [10] results in 4633 and 4406 images respectively. Figure 2 shows examples of the images that are part of the dataset augmented with the DiffuseMix technique [10].

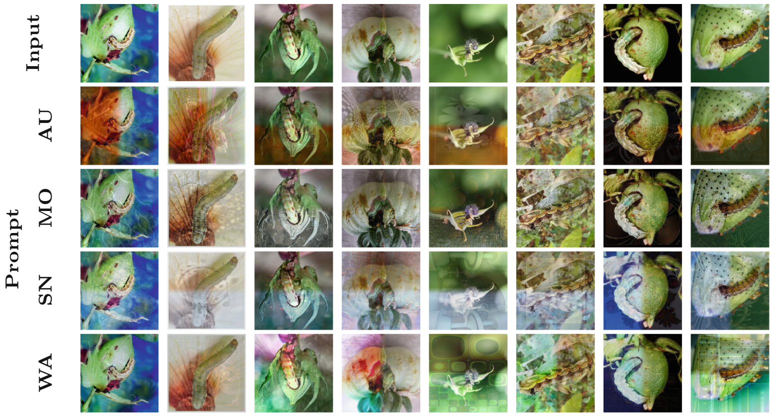

Fig. 2. Examples of images generated using DiffuseMix [10]. Prompt: Watercolor Art (WA), Snowy (SN), Autumn (AU), Mosaic (MO).

Table 2. Evaluation of the quality of camouflaged images according to [12] .The first two rows show the absolute median evaluation for three metrics for augmentation using Albumentations (AwA) and DiffuseMix (AwD) techniques. The last row shows the improvement (%) when using the DiffuseMix technique.

Data Augmentation	$S^Q_{R_f} \uparrow$	$S^Q_b \uparrow$	$S^Q_\alpha \uparrow$
AwA [1]	0.631	0.529	0.586
AwD [10]	0.637	0.539	0.603
Improv.(%)	+0.06%	+1.0%	+1.7%

4.2 Quantitative Results

Before evaluating each COD model's performance, the quality of the images generated with the augmentation techniques for the different data sets is measured as a first step. Table 2 shows in the first two rows the evaluation results of the metrics proposed by [12] on the images obtained with the data augmentation techniques. The last row shows the percentage difference between both data augmentation techniques, where it can be seen that the data augmentation technique using DiffuseMix presents better performance in all metrics. Based on

Table 3. Evaluation metrics for each COD technique according to the metrics described in Sect. 3.3. The best three performing results are highlighted in red (first), blue (second), and green (third) respectively.

	Technique	S_α ↑	F_β^w ↑	M ↓	E_ϕ^{adp} ↑	E_ϕ^{mean} ↑	E_ϕ^{max} ↑	F_β^{adp} ↑	F_β^{mean} ↑	F_β^{max} ↑
AwA [1]	BASNet [18]	0.7860	0.6839	0.0277	0.8938	0.8811	0.8890	0.7085	0.7073	0.7235
	SINet-v2 [5]	0.8588	0.7979	0.0184	0.9430	0.9571	0.9727	0.7837	0.8086	0.8392
	BGNet [4]	0.8573	0.7964	0.0185	0.9588	0.9532	0.9625	0.8264	0.8323	0.8527
	C^{2}F-Net [3]	0.8399	0.7184	0.0214	0.9368	0.9329	0.9441	0.7910	0.7989	0.8215
	OCENet [16]	0.8598	0.8042	0.0164	0.9421	0.9543	0.9647	0.7809	0.8115	0.8416
	EAMNet [24]	0.8627	0.8136	0.0174	0.9576	0.9611	0.9686	0.8164	0.8320	0.8570
	DGNet [11]	0.8565	0.8005	0.0163	0.9468	0.9585	0.9666	0.7827	0.8096	0.8377
	HitNet [9]	0.8422	0.7978	0.0178	0.9601	0.9663	0.9724	0.7991	0.8060	0.8244
	PCNet [29]	0.8463	0.7885	0.0184	0.9342	0.9471	0.9642	0.7691	0.7924	0.8311
AwD [10]	BASNet [18]	0.7761	0.6739	0.0269	0.9064	0.8726	0.9065	0.6992	0.7009	0.7113
	SINet-v2 [5]	0.8612	0.8123	0.0167	0.9493	0.9594	0.9689	0.7960	0.8199	0.8466
	BGNet [4]	0.8666	0.8237	0.0165	0.9698	0.9600	0.9708	0.9356	0.8431	0.8584
	C^{2}F-Net [3]	0.8487	0.7346	0.0212	0.9422	0.9509	0.9635	0.7931	0.8191	0.8450
	OCENet [16]	0.8548	0.7961	0.0179	0.9494	0.9528	0.9646	0.7831	0.8100	0.8352
	EAMNet [24]	0.8617	0.8180	0.0166	0.9524	0.9608	0.9674	0.8073	0.8303	0.8552
	DGNet [11]	0.8645	0.8068	0.0164	0.9454	0.9492	0.9597	0.7927	0.8196	0.8403
	HitNet [9]	0.8542	0.8206	0.0157	0.9704	0.9673	0.9713	0.8298	0.8318	0.8416
	PCNet [29]	0.8550	0.7949	0.0176	0.9425	0.9479	0.9535	0.7891	0.8057	0.8275

Table 4. Percentage difference between the results obtained for each COD technique and each data augmentation approach shown in Table 3. Positive values indicate improvement in the diffusion model-based approach, negative values indicate a deterioration in performance. The last row shows the improvement (%) of the DiffuseMix concerning the Albumentations technique.

Technique	ΔS_α	ΔF_β^w	ΔM	ΔE_ϕ^{adp}	ΔE_ϕ^{mean}	ΔE_ϕ^{max}	ΔF_β^{adp}	ΔF_β^{mean}	ΔF_β^{max}
BASNet [18]	-0.99	-1.00	-0.08	+1.26	-0.85	+1.75	-0.93	-0.64	-1.22
SINet-v2 [5]	+0.24	+1.44	-0.17	+0.63	+0.23	-0.38	+1.23	+1.13	+0.74
BGNet [4]	+0.93	+2.73	-0.20	+1.10	+0.68	+0.83	+10.92	+1.08	+0.57
C^{2}F-Net [3]	+1.62	-0.02	+0.54	+1.80	+1.94	+0.21	+0.27	+2.02	+2.35
OCENet [16]	-0.50	-0.81	+0.15	+0.73	-0.15	-0.01	+0.22	-0.15	-0.64
EAMNet [24]	-0.10	+0.44	-0.08	-0.08	-0.52	-0.03	-0.12	-0.17	-0.18
DGNet [11]	+0.80	+0.63	+0.01	+0.01	-0.14	-0.93	+0.69	+1.00	+0.26
HitNet [9]	+1.20	+2.28	-0.21	+1.03	+0.10	-0.11	+3.07	+2.58	+1.72
PCNet [29]	+0.87	+0.64	-0.08	+0.83	+0.08	+0.08	-1.07	+1.33	-0.36
Improv.(%)	+0.37	+0.54	0.00	+0.50	-0.27	+1.50	+0.31	+1.06	+0.24

these results, it is observed that the diffusion model achieves an improvement in the preservation of the structure and original content of the image ($S^Q_{R_f}$ +0.06%), a better definition of edges and contours (S^Q_b +1.0%) and consequently, a better overall balance between fidelity and visual quality of the image (S^Q_α +1.7%).

After evaluating the quality of the datasets, the next step is to compare the results between data augmentation using Albumentations [1] and DiffuseMix [10] techniques to improve the COD task. Table 3 shows the results of this evaluation. Overall, DiffuseMix exhibits superior performance across most metrics. BGNet with DiffuseMix achieves the best results in several key metrics, including S_α (0.8666), F^w_β (0.8237), F^{adp}_β (0.9356), F^{mean}_β (0.8431), and F^{max}_β (0.8584). HitNet with DiffuseMix also stood out, achieving the best performance in the M metric (0.0157, where lower values are better) and in the E^{adp}_ϕ (0.9704) and E^{mean}_ϕ (0.9673) metrics. Among the results obtained when the augmentation is performed with Albumentations, EAMNet and SINet-v2 show notable performance, with SINet-v2 reaching the highest E^{max}_ϕ (0.9727) across all comparisons. DGNet using Albumentations achieves the second-best result in the M metric (0.0163). Interestingly, implementing the diffusion model generally improves the performance of the base techniques, particularly in precision and edge detection metrics.

On the other hand, Table 4 shows the comparative analysis between augmentation with Albumentations [1] and augmentation with DiffuseMix [10]; it reveals interesting patterns in performance differences across various COD techniques. HitNet and BGNet demonstrate the most substantial improvements when implemented with the DiffuseMix technique, showing notable enhancements across all evaluation metrics. Interestingly, some techniques like BASNet show a slight deterioration in performance with the DiffuseMix approach, with decreases ranging from -0.64% to -1.22% across various metrics. The median difference performance comparison between both strategies indicates that DiffuseMix technique generally outperforms Albumentations technique, with positive differences across S_α, F^w_β, E^{adp}_ϕ, E^{max}_ϕ, F^{adp}_β, F^{mean}_β, and F^{max}_β metrics except for M, which remains neutral (0.00%) value and E^{mean}_ϕ with a negative difference of -0.27%. The most substantial improvements are observed in F^{mean}_β and E^{max}_ϕ, with gains of 1.50% and 0.91% respectively. The diffusion model approach improves performance, though its effectiveness varies across techniques and metrics.

4.3 Qualitative Results

Figure 3 compares ground truth (GT) masks and predicted masks across different techniques. In these visualizations, three distinct regions are highlighted: white areas represent successful matches between GT and predicted masks, indicating accurate detection; red areas denote false positive regions (over-segmentation), where the model incorrectly predicts camouflaged objects in areas not marked in the GT; and green areas indicate false negative regions (miss-segmentation), representing areas where the model fails to detect camouflaged objects present in

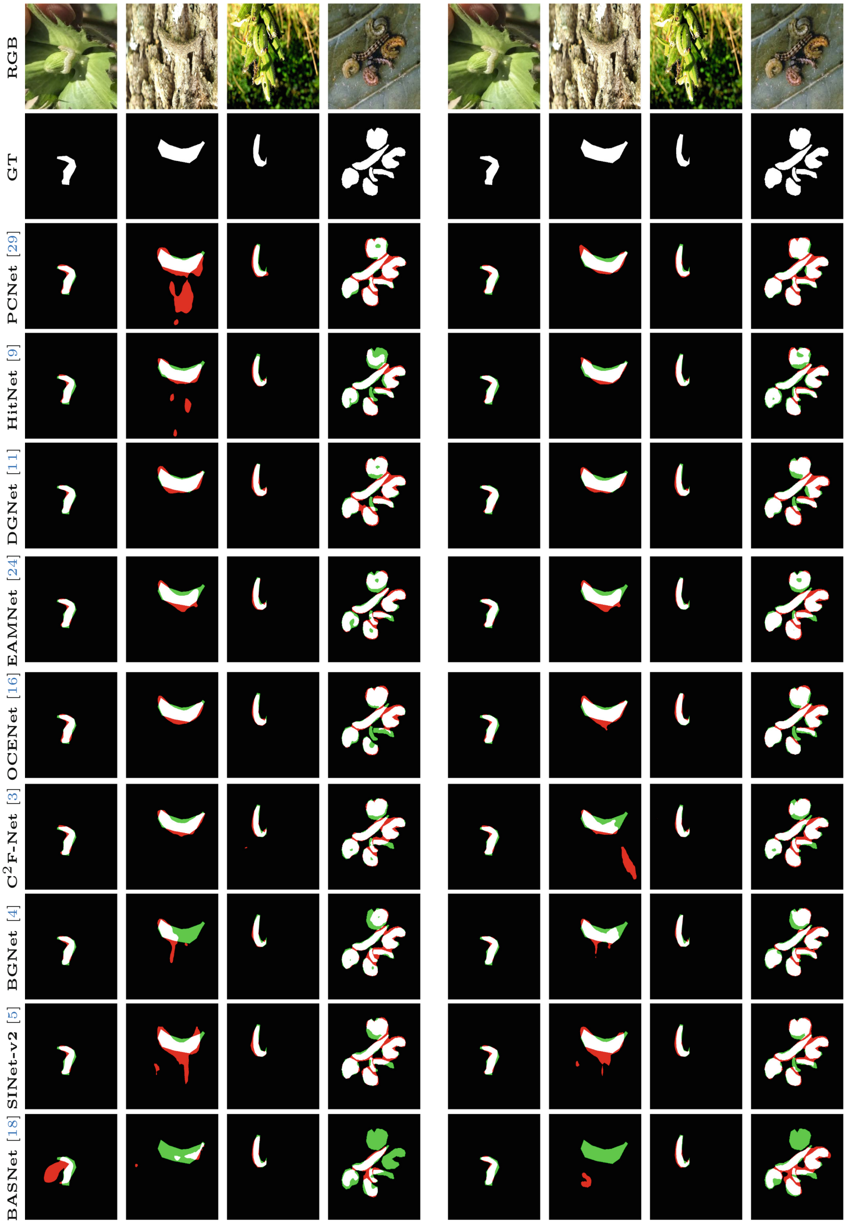

Fig. 3. Prediction results of nine state-of-the-art COD techniques trained with augmentation with Albumentations [1] (*col 1st to 4th*) and DiffuseMix [10] (*col 5th to 8th*). White areas represent successful matches between GT and predicted masks; red areas denote false positive regions (over-segmentation); and green areas indicate false negative regions (miss-segmentation).(Color figure online)

the GT. Visual analysis reveals that COD techniques trained with the Albumentations method tend to produce more false positives (red) and false negatives (green) compared to their counterparts trained with diffusion model-augmented datasets. This observation suggests that the diffusion model-based approach leads to more precise and accurate object detection, with fewer instances of both over-segmentation and missed detections.

5 Conclusions

Experimental results demonstrate that the diffusion model-based data augmentation technique outperforms the classical method, showing an improvement in the quality of the camouflaged image detection reflected in an increase in the S_α^Q metric (+1.7%), which shows an overall better balance between reconstruction fidelity score and boundary visibility score. On the other hand, a comparison shows an improvement in the metrics using DiffuseMix, which indicates that the diffusion model technique generally outperformes the classical method, with positive improvements across most metrics (S_α, F_β^w, E_ϕ^{adp}, E_ϕ^{max}, F_β^{adp}, F_β^{mean}, and F_β^{max}), the most substantial being in F_β^{mean} and E_ϕ^{max}, with gains of 1.50% and 0.91% respectively.

Acknowledgements. This work was supported in part by Grant PID2021-128945NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe", in part by the Air Force Office of Scientific Research Under Award FA9550-24-1-0206. The second author acknowledges the support of the Generalitat de Catalunya CERCA Program to CVC's general activities, and the Departament de Recerca i Universitats from Generalitat de Catalunya with reference 2021 SGR01499.

References

1. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: fast and flexible image augmentations. Information **11**(2), 125 (2020)
2. Charco, J.L., Sappa, A.D., Vintimilla, B.X., Velesaca, H.O.: Camera pose estimation in multi-view environments: from virtual scenarios to the real world. Image Vision Comput. **110**, 104182 (2021)
3. Chen, G., Liu, S.J., Sun, Y.J., Ji, G.P., Wu, Y.F., Zhou, T.: Camouflaged object detection via context-aware cross-level fusion. Trans. Circuits Syst. Video Technol. **32**(10), 6981–6993 (2022)
4. Chen, T., Xiao, J., Hu, X., Zhang, G., Wang, S.: Boundary-guided network for camouflaged object detection. Knowl. Based Syst. **248**, 108901 (2022)
5. Fan, D.P., Ji, G.P., Cheng, M.M., Shao, L.: Concealed object detection. Trans. Patt. Analy. Mach. Intell. **44**(10), 6024–6042 (2021)
6. Fan, D.P., Ji, G.P., Xu, P., Cheng, M.M., Sakaridis, C., Van Gool, L.: Advances in deep concealed scene understanding. Visual Intell. **1**(1), 16 (2023)
7. Gao, S.H., Cheng, M.M., Zhao, K., Zhang, X.Y., Yang, M.H., Torr, P.: Res2net: a new multi-scale backbone architecture. Trans. Patt. Analy. Mach. Intell. **43**(2), 652–662 (2019)

8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Conf. on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
9. Hu, X., et al.: High-resolution iterative feedback network for camouflaged object detection. In: Conf. on Artificial Intelligence. vol. 37, pp. 881–889 (2023)
10. Islam, K., Zaheer, M.Z., Mahmood, A., Nandakumar, K.: Diffusemix: label-preserving data augmentation with diffusion models. In: Conf. on Computer Vision and Pattern Recognition, pp. 27621–27630 (2024)
11. Ji, G.P., Fan, D.P., Chou, Y.C., Dai, D., Liniger, A., Van Gool, L.: Deep gradient learning for efficient camouflaged object detection. Mach. Intell. Res. **20**(1), 92–108 (2023)
12. Lamdouar, H., Xie, W., Zisserman, A.: The making and breaking of camouflage. In: Int. Conf. on Computer Vision, pp. 832–842 (2023)
13. Le, M.Q., Tran, M.T., Le, T.N., Nguyen, T.V., Do, T.T.: Unveiling camouflage: a learnable fourier-based augmentation for camouflaged object detection and instance segmentation. arXiv (2023)
14. Li, X., et al.: Camouflaged object detection with counterfactual intervention. Neurocomputing **553**, 126530 (2023)
15. Liang, Y., Qin, G., Sun, M., Wang, X., Yan, J., Zhang, Z.: A systematic review of image-level camouflaged object detection with deep learning. Neurocomputing **566**, 127050 (2024)
16. Liu, J., Zhang, J., Barnes, N.: Modeling aleatoric uncertainty for camouflaged object detection. In: Winter Conference on Applications of Computer Vision, pp. 1445–1454 (2022)
17. Meng, K., Xu, K., Cattani, P., Mei, S.: Camouflaged cotton bollworm instance segmentation based on pvt and mask r-cnn. Comput. Electron. Agricul. **226**, 109450 (2024)
18. Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., Jagersand, M.: Basnet: boundary-aware salient object detection. In: Conf. on Computer Vision and Pattern Recognition (2019)
19. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Conference on Computer Vision and Pattern Recognition, pp. 10684–10695 (2022)
20. Saharia, C., et al.: Photorealistic text-to-image diffusion models with deep language understanding. Adv. Neural. Inf. Process. Syst. **35**, 36479–36494 (2022)
21. Sappa, A.D., Suárez, P.L., Velesaca, H.O., Carpio, D.: Domain adaptation in image dehazing: Exploring the usage of images from virtual scenarios. In: Int. Conf. on Computer Graphics, Visualization, Computer Vision and Image Processing, pp. 85–92 (2022)
22. Shorten, C., Khoshgoftaar, T.M.: A survey on image data augmentation for deep learning. J. Big Data **6**(1), 1–48 (2019)
23. Suárez, P.L., Velesaca, H.O., Carpio, D., Sappa, A.D.: Corn kernel classification from few training samples. Artif. Intell. Agriculture **9**, 89–99 (2023)
24. Sun, D., Jiang, S., Qi, L.: Edge-aware mirror network for camouflaged object detection. In: Int. Conf. on Multimedia and Expo, pp. 2465–2470. IEEE (2023)
25. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: Int. Conf. on Machine Learning, pp. 6105–6114. PMLR (2019)
26. Velesaca, H.O., Suárez, P.L., Carpio, D., Sappa, A.D.: Synthesized image datasets: Towards an annotation-free instance segmentation strategy. In: International Symposium Advances in Visual Computing, pp. 131–143. Springer (2021)
27. Wang, W., et al.: Pvt v2: Improved baselines with pyramid vision transformer. Computational Visual Media **8**(3), 415–424 (2022)

28. Xiao, F., et al.: A survey of camouflaged object detection and beyond. arXiv (2024)
29. Yang, J., Wang, Q., Zheng, F., Chen, P., Leonardis, A., Fan, D.P.: Plantcamo: plant camouflage detection. arXiv (2024)
30. Yang, L., et al.: Diffusion models: a comprehensive survey of methods and applications. ACM Comput. Surv. **56**(4), 1–39 (2023)
31. Zhong, J., Wang, A., Ren, C., Wu, J.: A survey on deep learning-based camouflaged object detection. Multimedia Syst. **30**(5), 268 (2024)