A Multi-step Fusion Network Based on Environmental Knowledge Graph for Camouflaged Object Detection

Zheng Wang
wzheng@tju.edu.cn
Wenjun Huang
hwj0526@tju.edu.cn
Ruoxun Su
suruoxun@tju.edu.cn
Xinyu Yan
xinyuyan@tju.edu.cn
Meijun Sun
sunmeijun@tju.edu.cn

Machine Learning and
Systems Laboratory
Tianjin University,China

Abstract

Due to the high similarity in color and texture between camouflaged objects and noise backgrounds, existing single-step detection methods often fail especially when the camouflage level of objects is high. However, with prior knowledge of the environment, humans can effectively distinguish camouflaged objects, for example, when humans see snowy ground, they spontaneously associate that white rabbits might be concealed there. In this paper, we propose an Environmental Knowledge-guided Multi-step Network (EKNet) to simulate this mechanism. To extract prior knowledge of the background, we construct a knowledge graph with information extracted from the image and generate a relevance score matrix (RS) for prior knowledge and the camouflaged object with GCN as the correlation scoring matrix generation module (CSM). After that, we fuse the RS with Canny edge-enhanced features, which guides the model to detect camouflaged objects more accurately by observing the background information with edge semantics as the knowledge integration module (KIM). To our knowledge, this work is the first to introduce environmental knowledge to guiding camouflaged object detection (COD). Extensive experiments on three benchmark datasets show that our EKNet outperforms 15 existing state-of-the-art methods under four widely-used evaluation metrics.

1 Introduction

Camouflage is a unique method of concealment. A camouflaged object may disguise itself by mimicking the color or texture of another object, such as imitating the appearance of the surrounding environment or using disruptive coloring [7]. In the natural world, some animals utilize camouflage for predation or predator avoidance, which makes them blend in

seamlessly with the surroundings. In the medical field, detecting certain viruses, polyps, or tumors can be challenging because they blend in with the surrounding normal organ tissue. Research on camouflage object detection has the potential to advance developments in a wide range of fields, including species discovery and conservation [26], polyp segmentation [9], COVID-19 segmentation [6, 34], defect detection in the industrial field [10, 19], locust invasion detection, fruit ripeness detection in the agricultural field, as well as art collaging and body painting in the art field. Overall, research on camouflage object detection can significantly impact various tasks and has captured the attention of many researchers.

However, there are several challenges in camouflaged object detection (COD). One of these lies in the high consistency between the color and texture of the camouflaged object and its background, which makes it difficult to distinguish the foreground from the background and significantly increases the difficulty of the task. Recognizing and segmenting the camouflaged object simply in the visual feature space can lead to visual "traps" and often result in missed or false detections. Moreover, the high complexity resulting from the properties of camouflaged objects, such as the diversity of species, sizes, and shapes, often leads to instability in the results of methods that rely on single-step direct recognition. Overall, three major difficulty problems exist in camouflaged object detection: the wide variety of camouflaged objects, the obscured boundaries, and the obstruction in front of objects.

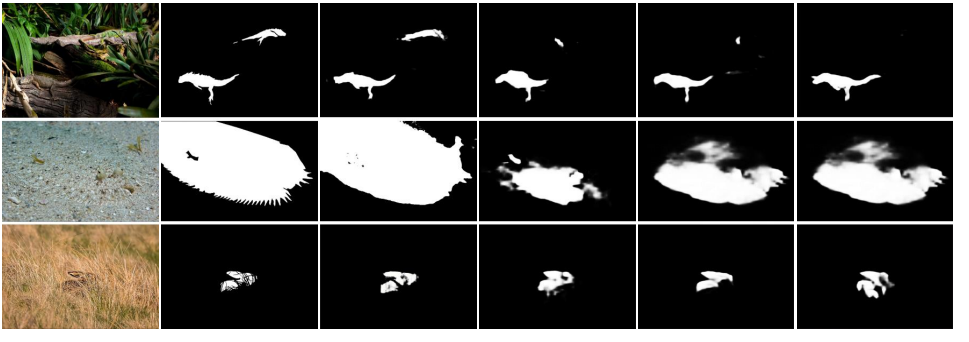

Image GT Ours SINet[8] PFNet[23] C2FNet[30]

Figure 1: From top to bottom, three challenging camouflage scenarios with multiple objects, indefinable boundaries, and occluded objects are listed. Our model outperforms SINet[8] ,PFNet[23] and CF2Net[30] under these challenging scenarios.

Recently, various solutions have been proposed by researchers to address these difficulties, including the application of visual attention. These methods can be broadly classified into four categories. The first category designs networks by simulating biological characteristics and human visual observation, such as SINet [8], SINet-V2 [6], PFNet [23], and C2FNet [30]. The second category is based on edge texture, with methods such as BGNet [31], TANet [29], and BASNet [39] enhancing segmentation by refining edge textures. The third category is based on multi-scale feature extraction networks, including ZoomNet[24]. The fourth category is based on a joint learning framework, such as JCSOD [18] and MGL [37], which can explore additional clues from shared features to enhance the feature representation of COD. Figure 1 provides a visualization result of our model and SINet[8], PFNet[23], and C2FNet[30] on these three types of problems. It can be seen that the above-mentioned challenges cannot be solved completely by these single-step recognition methods, thus, we mine semantic knowledge of the background and fuse it into a multi-step network.

In this paper, we propose an Environmental Knowledge-guided Multi-step Network(EK-Net), which leverages the inherent similarity between camouflaged objects and their backgrounds (*e.g.*, the probability of an object appearing in a snowy scene is higher when it has similar color and texture to snow). Our model utilizes the multi-dimensional feature fusion of environmental knowledge and edge feature enhancement. The contributions are as follows:

(1) We propose the correlation score matrix (CSM) generation module to simulate how the human brain utilizes background semantic knowledge to detect camouflaged objects. We construct a knowledge graph with explored semantic information and use a Graph Convolutional Network (GCN)[15] to generate a relevance score matrix (RS). To the best of our knowledge, this is the first study to introduce knowledge graphs to assist in recognizing camouflaged objects.

(2)We propose the knowledge integration module (KIM), which integrates environmental knowledge and Canny edge enhancement to capture visual feature clues. This module achieves multi-dimensional integration in both the knowledge space and visual feature space.

(3)We compare our proposed method with 15 state-of-the-art methods using three widely used datasets. Our method outperforms all others on four evaluation metrics, demonstrating its superior performance.

2 Related Work

2.1 Traditional Camouflaged Object detection.

Artifact object detection methods based on hand-crafted features primarily rely on color, texture, and optical features to distinguish between foreground and background and detect artifact targets. Galun et al [8] proposed a bottom-up aggregation framework called the B-to-Top model to detect objects by combining texture features with filter responses to identify the shape of camouflaged objects adaptively. Kavitha et al [13] proposed a UCT model that utilizes a local HSV (hue, saturation, value) color model and gray-level co-occurrence matrix(GLCM) features to identify camouflaged objects in images. Liu et al [20] proposed a model that integrates spatial, top-down, and spectral features of images to detect camouflaged objects. The detection accuracy of conventional methods is not stable in that these methods could be greatly affected by illumination changes.

2.2 Deep Learning-based Camouflaged Object Detection.

In recent years, methods for detecting camouflaged objects have advanced significantly. These methods can be broadly categorized into **four aspects**.

The first category is based on biological characteristics which simulate the human visual observation process. Representative works include SINet proposed by Fan [3], which gradually locates and searches for camouflaged objects by imitating the detection and recognition phases of predation. Mei [23] et al. proposed the PFNet[23] network consisting of a localization module and a focus module, which mimic the corresponding phase of predation. Sun et al [30] proposed C2FNet, which integrates cross-layer features and considers rich global contextual information.

The second category of approach is based on edge texture. Representative works include BGNet proposed by Sun [31] et al., which focuses on edge details when localizing the target region and uses edges to enhance detection accuracy. TANet, proposed by Ren[29]

et al., improves the accuracy of detection by amplifying the texture difference between the object and its background.

The third category is characterized by networks based on multi-scale feature extraction. The representative works include ZoomNet, a mixed-scale triad network proposed by Pang [24] et al., which constructs and unifies scale-specific features on different levels to reliably capture objects in complex scenes.

The fourth one, represented by JCSOD[18], **is based on joint learning frameworks,** Li[18] et al.'s purposed this network to enhance the detection of both salient targets and camouflaged objects by using contradictory information. Parts of simple positive samples are trained as hard positive samples of SOD to enhance the detection of both tasks. Zhai[37] et al.'s MGL network decomposes an image into two task-specific feature maps, one for locating the target roughly while another for precisely capturing its boundary details, inferring the higher-order relationships between them by graph theory.

However, the aforementioned methods are all limited by their reliance on single-step visual feature recognition approaches. To break through these limitations, we propose our own model.

3 Method

The overall model architecture of the proposed EKNet is shown in Figure 3. Firstly, in the correlation score matrix(CSM) module, with small-scale manual annotations, we extract the environmental knowledge by fine-tuning the object detection network. The information is leveraged to construct a knowledge graph. Then, we generate the relevance score(RS) matrix using GCN[15]. Secondly, in the knowledge integration module(KIM), the RS matrix is fused with Canny edge features to generate a more complete and detailed segmentation result. Each module will be described in later sections in more detail.

3.1 Correlation scoring matrix generation module(CSM)

Most of the existing research work chooses to feed images into the network directly, ignoring the semantic information contained in the pictures thriftlessly. However, this background prior knowledge is often leveraged by humans to capture clues about camouflaged objects. With the idea of simulating the human recognition process, we decide to treasure this overlooked information. Thus, in the pre-trained stage, extensive work is done to build a visual knowledge graph that illustrates the semantic relationship between environmental knowledge and camouflaged objects unambiguously.

Since camouflage roots in the similarity of edges, colors, and textures, which means there exist certain correlations between background attributes and camouflaged objects. This is the motivation behind our method. Taking Figure 2 as an example, the camouflaged object is a rabbit. With the object detection network, the background can be detected as the snowy ground and other visual semantic entities like reefs. A knowledge graph representing camouflage knowledge will be constructed for each image in the dataset. The relevance can be measured by the number of node interactions, that is, the more edge appearing between the two entities or feature attribute nodes in these graphs, the stronger the relevance is. Since the attribute nodes in the knowledge graph are entities extracted from visual image data, we name this graph an environment-based visual knowledge graph (EVKG).

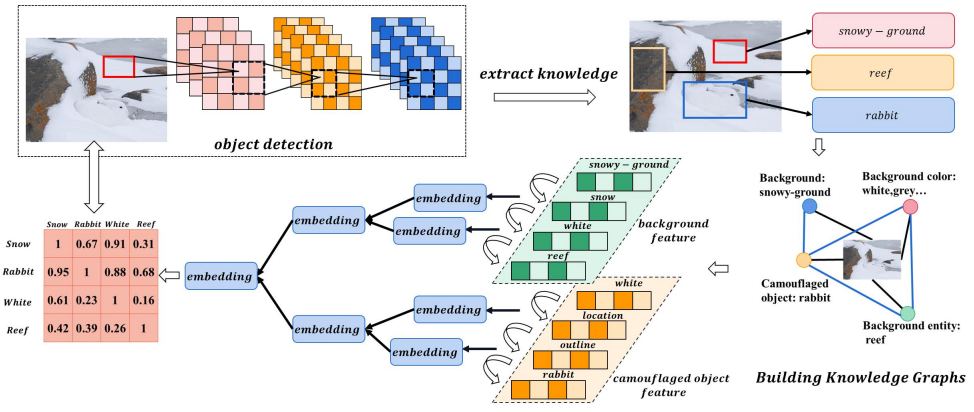

Figure 2: Correlation scoring matrix generation module (CSM). Firstly, extract the knowledge in the graph by the object detection algorithm. Secondly, construct the environment-based visual knowledge graph (EVKG) and generate the relevance score matrix (RS) by the graph convolution network embedding

The details are as follows. Firstly, a visual object detection method is used to detect semantic entities in the images. In this paper, YOLO[28] is applied to extract the semantic information of the objects and backgrounds in the dataset. To ensure that these object detection methods can effectively extract the features needed, we trained the object detection model using public datasets. However, directly transferring the model to camouflage scenes will result in low accuracy, as the information in camouflage scenes has not been annotated in previous training. Therefore, we manually labeled a small subset of the COD10K[6] dataset (A total of 14 scenes labeled in background categories, background colors, camouflaged object categories, etc.) and used these annotated datasets to fine-tune the object detection model above-mentioned so that it can effectively adapt to camouflage object detection tasks. Then, the fine-tuned object detection model is used to detect semantic entities in the images, these entities and their relationships are used to construct the EVKG.

After that, we need to incorporate the knowledge information in EVKG into the segmentation stage as prior knowledge. We use GCN[15] to generate vector representations of each node in the knowledge graph, which can be divided into three stages: node initialization, correlation information propagation, and network stacking. Firstly, during the node initialization process, the vector representations of each node in the visual hidden knowledge graph are initialized according to the normal distribution. Secondly, a message-passing mechanism is used to aggregate the relevant information of neighboring nodes for each node. Finally, relevant information propagation is executed through network stacking. The final vector representations of all nodes are output by the last layer of the network. Using these node vector representations and interaction frequency information, the parameters of models are updated.

We embedded the feature vectors of the camouflaged objects and backgrounds separately. After that, they are embedded and propagated layer by layer through the GCN network.

$$F^{n+1} = ReLU(AF^nW) \tag{1}$$

Where F^n represents the n-th layer of feature data, F^{n+1} represents the feature data of layer n+1, A is the Adjacency matrix, and W is the network parameter matrix.

The optimized model can output a relevance score matrix between any background information and camouflage object category to assist subsequent segmentation tasks.

3.2 Knowledge Integration Module (KIM)

In the segmentation stage, the RS is fused with the feature maps f_1, f_2, f_3, f_4, and f_5 obtained by convolving the backbone network. The network is constantly reasoning based on the knowledge information to consciously search for potential camouflage objects. The edge feature information of the ground truth is extracted using the canny operator[1]. Since f_5 loses some information near the edges, fusing the edge information extracted by the canny operator with f_5 features can help the network generate more complete and detailed segmentation results, which are then sent to the Knowledge Integration Module (KIM).

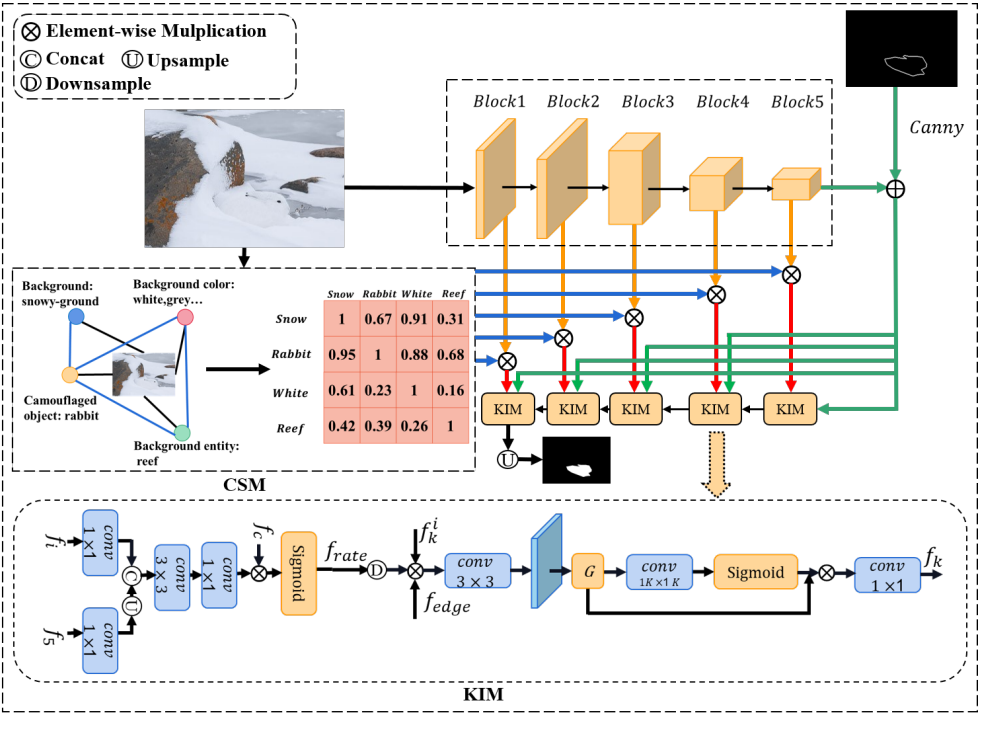

Figure 3: Complete model architecture(EKNet). The complete network design and the detailed design of the KIM module are shown here.

The design of KIM is shown in Figure 3. This module aims to integrate edge cue information and RS into representation learning to enhance the implicit knowledge reasoning ability and object-structured semantic feature representation. Therefore, we introduce a correlation attention mechanism to guide the network to explore key clues based on correlation information to search for the target disguised object. We fuse the relevance scores with the

features by element-wise multiplication and add the edge information to f_5 by element-wise addition. Then we perform 3×3 convolution to obtain the initial fused feature f_m, which can be represented as:

$$f_m = F_{conv3\times3}((f_i \otimes f_{rate}) \otimes (D(f_{edge}) \oplus f_5)) \tag{2}$$

where D denotes downsampling, $F_{conv3\times3}$ is a 3x3 convolution, $\otimes$ represents element-wise multiplication, $\oplus$ represents element-wise addition, and f_{rate} is the knowledge feature information for correlation score. Inspired by the work of Wang [32]et al., we utilize channel-wise global average pooling (GAP) to aggregate the convolutional feature f_m, and then obtain corresponding channel attention weights, *i.e.*, weight information, through one-dimensional convolution followed by the Sigmoid function, exploring local cross-channel interactions and learning background knowledge attention information. Afterward, we reduce the number of channels through 1×1 convolution and obtain f_k, *i.e.*, the feature map with reduced channels.

$$f_k = F_{conv1\times1}(Sigmoid(f^a_{1K}(G(f_m))) \otimes f_m \tag{3}$$

Where $F_{conv1\times1}$ is a 1×1 convolution, f^a_{1K} represents 1K convolution with kernel size a, where k is proportional to the channel size. Clearly, the KIM module can highlight key channel information rather than suppress noise or redundant channel information, thereby enhancing semantic representation.

3.3 Loss Function

Our model's loss function consists of two major parts: the knowledge score matrix generated loss L_G and the segmentation task loss L_S. In the segmentation task part, we introduce two types of supervision, namely, the camouflaged object mask S_m and the camouflaged object edge S_e. For S_m, we use weighted binary cross-entropy loss (L^{ω}_{BCE}) and weighted IOU loss (L^{ω}_{IOU}) [33], which focus more on hard pixels. For S_e, we use dice loss (L_{dice}) [35] to handle the strong imbalance between positive and negative samples. Since mask supervision is applied to the five KIM modules, the total loss function L_{sum} is defined as follows:

$$L_{sum} = \sum_{i=1}^{5}(L_G + L^{\omega}_{BCE}(R_i, S_m) + L^{\omega}_{IOU}(R_i, S_m)) + \varepsilon L_{dice}(R_e, S_e) \tag{4}$$

where R_i and R_e are the predictions for the overall model and camouflaged object edges, respectively. ε is a trade-off parameter, which is set to 5 in our experiments.

4 Experiments

4.1 Experimental Setup

Implemental Details. We employ GCN[15] pre-trained as the knowledge imformation model and Res2Net50[9] pre-trained on ImageNet as our backbone. Once the correlation score matrix is obtained, we resize all the input images to 416× 416 and enhance them with random horizontal flipping. During the training stage, the batch size is set to 12, and the Adam optimizer[14] is adopted. The learning rate is initialized to 1e-4 and adjusted by poly strategy with the power of 0.9. Accelerated by an RTX2080Ti 12G GPU, the whole training takes about 2.5 hours with 25 epochs.

Datasets. We evaluate our method on three public benchmark datasets: CAMO[17], COD10K[3] and NC4K[21]. We follow the previous works[3], which use the training set of

CAMO and COD10K as our training set, and use their testing set and NC4K as our testing sets.

Evaluation Metrics. We utilize four widely used metrics to evaluate our method,*i.e.*, mean absolute error (*MAE*, $\mathcal{M}$) [25], weighted F-measure (F_β^ω) [22], structure-measure (S_α) [2] and mean E-measure (E_ω) [7].

4.2 Comparative Studies

We compare EKNet with 15 state-of-the-art methods. The backbone of PraNet is also Res2Net50. Table 1 shows the comparison results. For a fair comparison, all the predictions of these methods are either provided by the authors or produced by models retrained with open-source codes. Our method outperforms all other models on three datasets under four evaluation metrics.

However, We also have some images that don't detect very well. There exist some camouflaged objects with a very high degree of camouflage and bad detection of the image. In these cases, detection results will have deletions, which indicates that the network is not enough to learn the local details. Therefore, we consider that for these special cases, we need the network to amplify these local texture details, and then apply these details in the segmentation.

Recently the latest SAM model in the field of CV segmentation has advanced the great progress of general-purpose segmentation, and we have also evaluated the SAM model on the task of camouflaged object detection, which can be seen from the experimental data record in Table 1 that it does not perform too well, due to this task having high requirements for the edge, semantic inference, etc. So it is very necessary to study the camouflaged object task as a specialized task.

Method	Pub./Year	CAMO-Test				COD10K-Test				NC4K			
		$S_\alpha\uparrow$	$E_\phi\uparrow$	$F_\beta^\omega\uparrow$	$\mathcal{M}\downarrow$	$S_\alpha\uparrow$	$E_\phi\uparrow$	$F_\beta^\omega\uparrow$	$\mathcal{M}\downarrow$	$S_\alpha\uparrow$	$E_\phi\uparrow$	$F_\beta^\omega\uparrow$	$\mathcal{M}\downarrow$
EGNet[33]	ICCV2019	0.732	0.796	0.601	0.107	0.736	0.802	0.515	0.059	0.777	0.842	0.639	0.078
PraNet[8]	MICCAI2020	0.769	0.824	0.676	0.094	0.784	0.863	0.642	0.056	0.797	0.889	0.685	0.073
F3Net[34]	AAAI2020	0.711	0.741	0.564	0.109	0.739	0.795	0.544	0.051	0.780	0.824	0.656	0.070
SINet[9]	CVPR2020	0.745	0.804	0.704	0.092	0.776	0.864	0.645	0.043	0.809	0.872	0.753	0.058
PFNet[24]	CVPR2021	0.782	0.841	0.695	0.085	0.800	0.868	0.660	0.040	0.829	0.887	0.745	0.053
R-MGL[36]	CVPR2021	0.775	0.812	0.673	0.088	0.814	0.851	0.666	0.035	0.833	0.867	0.739	0.053
TANet[10]	AAAI2021	0.781	0.847	0.678	0.087	0.793	0.848	0.635	0.043	-	-	-	-
C2FNet[30]	IJCAI2021	0.796	0.857	0.730	0.078	0.813	0.889	0.691	0.036	0.840	0.896	0.771	0.048
UGTR[35]	ICCV2021	0.785	0.822	0.685	0.086	0.818	0.852	0.667	0.035	0.839	0.876	0.746	0.052
JCSOD[28]	CVPR2021	0.800	0.859	0.728	0.073	0.809	0.884	0.684	0.035	0.841	0.898	0.771	0.047
OCENet[31]	WACV2022	0.807	0.866	0.744	0.075	0.829	0.890	0.721	0.034	0.848	0.899	0.785	0.046
SegMaR[11]	CVPR2022	0.811	0.868	0.749	0.073	0.831	0.899	0.722	0.033	0.841	0.896	0.781	0.046
CubeNet[40]	PR2022	0.788	0.838	0.682	0.085	0.795	0.865	0.643	0.041	-	-	-	-
ERRNet[12]	PR2022	0.779	0.842	0.679	0.085	0.786	0.867	0.630	0.043	0.827	0.887	0.737	0.054
SAM[14]	arXiv2023	0.684	0.687	0.606	0.132	0.783	0.798	0.701	0.050	0.767	0.776	0.696	0.078
EKNet(Ours)		0.821	0.879	0.749	0.073	0.833	0.900	0.727	0.032	0.850	0.904	0.785	0.044

Table 1: Quantitative comparison with state-of-the-art methods for COD on three benchmarks using four widely used evaluation metrics (*i.e.*, S_α, E_ϕ, F_β^ω, $\mathcal{M}$). "↑" / "↓" indicates that larger/smaller is better. The top three results are highlighted in red, green, and blue.

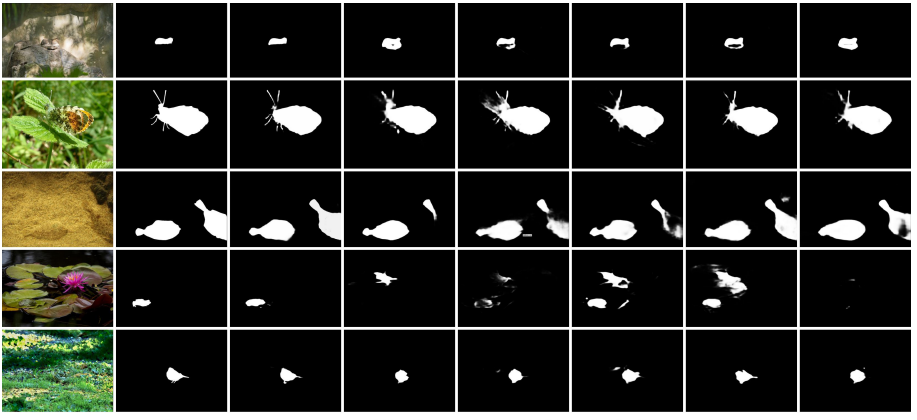

Figure 4: Visual comparison of the proposed model with five state-of-the-art COD methods. Obviously, our method is capable of accurately segmenting various camouflaged objects with more clear boundaries.

Figure 4 shows the qualitative comparisons of different COD methods on several typical samples from the COD10K dataset. It is obvious that our method provides accurate camouflaged object predictions with finer and more complete object structure and boundary details.

4.3 Ablation Study

In order to validate the effectiveness of each key component and make our analysis clear, we design several ablation experiments. For baseline model(B), we remove all the additional models (*i.e.*, CSM, KIM), and all the results are shown in Table 2.

Method	CAMO-Test				COD10K-Test				NC4K			
	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	$\mathcal{M} \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	$\mathcal{M} \downarrow$	$S_\alpha \uparrow$	$E_\phi \uparrow$	$F_\beta^\omega \uparrow$	$\mathcal{M} \downarrow$
B	0.799	0.858	0.726	0.080	0.823	0.893	0.702	0.035	0.845	0.897	0.773	0.048
B+CSM	0.807	0.864	0.739	0.075	0.830	0.896	0.715	0.034	0.848	0.902	0.781	0.046
B+KIM	0.809	0.868	0.742	**0.073**	**0.833**	0.898	0.718	0.033	**0.850**	0.903	0.783	0.044
Ours	**0.821**	**0.879**	**0.749**	**0.073**	**0.833**	**0.900**	**0.727**	**0.032**	**0.850**	**0.904**	**0.785**	**0.044**

Table 2: Quantitative evaluation for ablation studies on three datasets. The best results are highlighted in **Bold**. B: baseline.

Effectiveness of CSM. As can be seen in Table 2, compared with B model, the B+CSM model has more vantages on the metrics F_β^ω that shows 1.30% performance increases averagely.

Effectiveness of KIM. From Table 2, compared with B model, the B+KIM model provides better performance. Especially, the average performance gain with 1.08% on the metrics F_β^ω of our model for all datasets. Thus, the KIM is beneficial to boost detection performance.

Effectiveness of CSM and KIM. We also test the effectiveness of all the components. As shown in Table 2, the B+CSM+KIM model achieves obvious performance improvements and is also the best on all datasets, with the performance gains of 1.06%, 1.10% and 2.59% on average in terms of S_α, E_ϕ and F_β^ω, respectively.

5 Conclusion

In this paper, we address the limitations of visual feature single-step recognition methods by leveraging environmental knowledge relevance to camouflaged objects to enhance camouflaged object detection performance. We propose an effective Environmental Knowledge-guided Multi-step Network (EKNet), which includes a correlation scoring matrix generation module (CSM) and a knowledge integration module (KIM), to explore intrinsic semantic relevance between background and objects, guiding and improving representation learning for COD. We are the first to introduce knowledge atlas auxiliary information in camouflage object recognition and achieve multi-dimensional unification of knowledge space and visual feature space. Extensive experiments demonstrate that our approach outperforms 15 existing state-of-the-art methods on three benchmarks.

References

[1] John Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6):679–698, 1986.

[2] Deng-Ping Fan, Ming-Ming Cheng, Yun Liu, Tao Li, and Ali Borji. Structure-measure: A new way to evaluate foreground maps. In *Proceedings of the IEEE international conference on computer vision*, pages 4548–4557, 2017.

[3] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2777–2787, 2020.

[4] Deng-Ping Fan, Ge-Peng Ji, Tao Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Pranet: Parallel reverse attention network for polyp segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23*, pages 263–273. Springer, 2020.

[5] Deng-Ping Fan, Tao Zhou, Ge-Peng Ji, Yi Zhou, Geng Chen, Huazhu Fu, Jianbing Shen, and Ling Shao. Inf-net: Automatic covid-19 lung infection segmentation from ct images. *IEEE Transactions on Medical Imaging*, 39(8):2626–2637, 2020.

[6] Deng-Ping Fan, Ge-Peng Ji, Ming-Ming Cheng, and Ling Shao. Concealed object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10): 6024–6042, 2021.

[7] Deng-Ping Fan, Ge-Peng Ji, Xuebin Qin, and Ming-Ming Cheng. Cognitive vision inspired object segmentation metric and loss function. *Scientia Sinica Informationis*, 6 (6), 2021.

[8] Galun, Sharon, Basri, and Brandt. Texture segmentation by multiscale aggregation of filter responses and shape elements. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 716–723 vol.1, 2003. doi: 10.1109/ICCV.2003.1238418.

[9] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr. Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*, 43(2):652–662, 2019.

[10] Chuanfei Hu and Yongxiong Wang. An efficient convolutional neural network model based on object-level attention mechanism for casting defect detection on radiography images. *IEEE Transactions on Industrial Electronics*, 67(12):10922–10930, 2020.

[11] Ge-Peng Ji, Lei Zhu, Mingchen Zhuge, and Keren Fu. Fast camouflaged object detection via edge-based reversible re-calibration network. *Pattern Recognition*, 123: 108414, 2022.

[12] Qi Jia, Shuilian Yao, Yu Liu, Xin Fan, Risheng Liu, and Zhongxuan Luo. Segment, magnify and reiterate: Detecting camouflaged objects the hard way. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4713–4722, 2022.

[13] Ch Kavitha, B Prabhakara Rao, and A Govardhan. An efficient content based image retrieval using color and texture of image sub blocks. *International Journal of Engineering Science and Technology (IJEST)*, 3(2):1060–1068, 2011.

[14] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[15] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[16] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. *arXiv preprint arXiv:2304.02643*, 2023.

[17] Trung-Nghia Le, Tam V Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto. Anabranch network for camouflaged object segmentation. *Computer vision and image understanding*, 184:45–56, 2019.

[18] Aixuan Li, Jing Zhang, Yunqiu Lv, Bowen Liu, Tong Zhang, and Yuchao Dai. Uncertainty-aware joint salient object and camouflaged object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10071–10081, 2021.

[19] Jian Lian, Weikuan Jia, Masoumeh Zareapoor, Yuanjie Zheng, Rong Luo, Deepak Kumar Jain, and Neeraj Kumar. Deep-learning-based small surface defect detection via an exaggerated local variation-based generative adversarial network. *IEEE Transactions on Industrial Informatics*, 16(2):1343–1351, 2019.

[20] Zhou Liu, Kaiqi Huang, and Tieniu Tan. Foreground object detection using top-down information based on em framework. *IEEE Transactions on Image Processing*, 21(9): 4204–4217, 2012.

[21] Yunqiu Lv, Jing Zhang, Yuchao Dai, Aixuan Li, Bowen Liu, Nick Barnes, and Deng-Ping Fan. Simultaneously localize, segment and rank the camouflaged objects. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11591–11601, 2021.

[22] Ran Margolin, Lihi Zelnik-Manor, and Ayellet Tal. How to evaluate foreground maps? In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 248–255, 2014.

[23] Haiyang Mei, Ge-Peng Ji, Ziqi Wei, Xin Yang, Xiaopeng Wei, and Deng-Ping Fan. Camouflaged object segmentation with distraction mining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8772–8781, 2021.

[24] Youwei Pang, Xiaoqi Zhao, Tian-Zhu Xiang, Lihe Zhang, and Huchuan 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*, pages 2160–2170, 2022.

[25] Federico Perazzi, Philipp Krähenbühl, Yael Pritch, and Alexander Hornung. Saliency filters: Contrast based filtering for salient region detection. In *2012 IEEE conference on computer vision and pattern recognition*, pages 733–740. IEEE, 2012.

[26] Ricardo Pérez-de la Fuente, Xavier Delclòs, Enrique Peñalver, Mariela Speranza, Jacek Wierzchos, Carmen Ascaso, and Michael S Engel. Early evolution and ecology of camouflage in insects. *Proceedings of the National Academy of Sciences*, 109(52): 21414–21419, 2012.

[27] Natasha Price, Samuel Green, Jolyon Troscianko, Tom Tregenza, and Martin Stevens. Background matching and disruptive coloration as habitat-specific strategies for camouflage. *Scientific reports*, 9(1):7840, 2019.

[28] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[29] Jingjing Ren, Xiaowei Hu, Lei Zhu, Xuemiao Xu, Yangyang Xu, Weiming Wang, Zijun Deng, and Pheng-Ann Heng. Deep texture-aware features for camouflaged object detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[30] Yujia Sun, Geng Chen, Tao Zhou, Yi Zhang, and Nian Liu. Context-aware cross-level fusion network for camouflaged object detection. *arXiv preprint arXiv:2105.12555*, 2021.

[31] Yujia Sun, Shuo Wang, Chenglizhao Chen, and Tian-Zhu Xiang. Boundary-guided camouflaged object detection. *arXiv preprint arXiv:2207.00794*, 2022.

[32] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11534–11542, 2020.

[33] Jun Wei, Shuhui Wang, and Qingming Huang. F^3net: fusion, feedback and focus for salient object detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 12321–12328, 2020.

[34] Yu-Huan Wu, Shang-Hua Gao, Jie Mei, Jun Xu, Deng-Ping Fan, Rong-Guo Zhang, and Ming-Ming Cheng. Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation. *IEEE Transactions on Image Processing*, 30:3113–3126, 2021.

[35] Enze Xie, Wenjia Wang, Wenhai Wang, Mingyu Ding, Chunhua Shen, and Ping Luo. Segmenting transparent objects in the wild. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*, pages 696–711. Springer, 2020.

[36] Fan Yang, Qiang Zhai, Xin Li, Rui Huang, Ao Luo, Hong Cheng, and Deng-Ping Fan. Uncertainty-guided transformer reasoning for camouflaged object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4146–4155, 2021.

[37] Qiang Zhai, Xin Li, Fan Yang, Chenglizhao Chen, Hong Cheng, and Deng-Ping Fan. Mutual graph learning for camouflaged object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12997–13007, 2021.

[38] Jia-Xing Zhao, Jiang-Jiang Liu, Deng-Ping Fan, Yang Cao, Jufeng Yang, and Ming-Ming Cheng. Egnet: Edge guidance network for salient object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8779–8788, 2019.

[39] Hongwei Zhu, Peng Li, Haoran Xie, Xuefeng Yan, Dong Liang, Dapeng Chen, Mingqiang Wei, and Jing Qin. I can find you! boundary-guided separated attention network for camouflaged object detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 3608–3616, 2022.

[40] Jinchao Zhu, Xiaoyu Zhang, Shuo Zhang, and Junnan Liu. Inferring camouflaged objects by texture-aware interactive guidance network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 3599–3607, 2021.

[41] Mingchen Zhuge, Xiankai Lu, Yiyou Guo, Zhihua Cai, and Shuhan Chen. Cubenet: X-shape connection for camouflaged object detection. *Pattern Recognition*, 127:108644, 2022.