CamoTeacher: Dual-Rotation Consistency Learning for Semi-Supervised Camouflaged Object Detection

Xunfa Lai[1], Zhiyu Yang[1], Jie Hu[2], Shengchuan Zhang[1*], Liujuan Cao[1], Guannan Jiang[2], Zhiyu Wang[2], Songan Zhang[3], and Rongrong Ji[1]

[1] Key Laboratory of Multimedia Trusted Perception and Efficient Computing, Ministry of Education of China, Xiamen University
[2] Contemporary Amperex Technology Co. Limited
[3] Shanghai Jiao Tong University

Abstract. Existing camouflaged object detection (COD) methods depend heavily on large-scale pixel-level annotations. However, acquiring such annotations is laborious due to the inherent camouflage characteristics of the objects. Semi-supervised learning offers a promising solution to this challenge. Yet, its application in COD is hindered by significant pseudo-label noise, both pixel-level and instance-level. We introduce CamoTeacher, a novel semi-supervised COD framework, utilizing Dual-Rotation Consistency Learning (DRCL) to effectively address these noise issues. Specifically, DRCL minimizes pseudo-label noise by leveraging rotation views' consistency in pixel-level and instance-level. First, it employs Pixel-wise Consistency Learning (PCL) to deal with pixel-level noise by reweighting the different parts within the pseudo-label. Second, Instance-wise Consistency Learning (ICL) is used to adjust weights for pseudo-labels, which handles instance-level noise. Extensive experiments on four COD benchmark datasets demonstrate that the proposed CamoTeacher not only achieves state-of-the-art compared with semi-supervised learning methods, but also rivals established fully-supervised learning methods. Our code will be available soon.

Keywords: Camouflaged object detection · Semi-supervised learning

1 Introduction

Camouflaged object detection (COD) aims to identify objects that blend seamlessly into their environments including animals or artificial entities that possess protective coloring and have the ability to blend into their surroundings, a task complicated by low contrast, similar textures, and blurred boundaries [6, 7, 9, 14, 17, 24, 43]. Unlike general object detection, COD is challenged by these factors, making detection significantly more difficult. Existing methods in COD heavily depend on extensive pixel-level annotated datasets, the creation of which incurs substantial human effort and cost, thus limiting COD's advancement.

* Corresponding authors

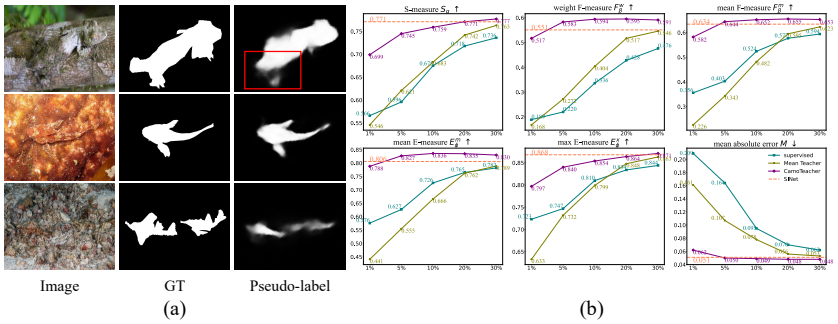


Fig. 1: (a) Illustration of pixel-level and instance-level noise. Pixel-level noise refers to varying degrees of noise in different parts within an instance. Instance-level noise pertains to varying degrees of noise among different instances. (b) Visualization of performance trends on COD10K [7] under different percentages of labeled data.

To mitigate that, semi-supervised learning [13, 27, 36, 47, 49] emerges as a promising approach by leveraging both labeled and unlabeled data, yet its application in COD is not straightforward due to complex backgrounds and subtle object boundaries. The efficacy of semi-supervised learning in COD is significantly compromised by the presence of substantial noise in pseudo-labels. Our investigation into pseudo-label noise reveals two primary types: pixel-level noise, indicating variation within a single pseudo-label, and instance-level noise, showing variation across different pseudo-labels. This distinction is critical, as it informs our approach to refining pseudo-label quality for improved model training. (1) Pixel-level noise is characterized by inconsistent labeling within various parts of a pseudo-label. As demonstrated in Fig. 1(a), in the first row, the tail of the gecko is visually more camouflaged than its head. The pseudo-label generated by the SINet [7] is less accurate in its tail region (highlighted by the red box). This observation underscores the inappropriateness of uniformly treating all parts within a pseudo-label. (2) Instance-level noise refers to the variability in noise levels across distinct pseudo-labels. As shown in Fig. 1(a), the pseudo-labels of the third row are less accurate compared to that in the second row because the camouflaged objects in the third row is more difficult to detect. These discrepancies indicate that each pseudo-label contributes variably to model training, emphasizing the need for a nuanced approach to integrating pseudo-label information.

To tackle the challenge of evaluating pseudo-label noise without ground truth (GT) for unlabeled data, we introduce two novel strategies based on pixel-wise inconsistency and instance-wise consistency across two rotation views. Specifically, for pixel-level noise, we observe that pixel-wise inconsistency—calculated by comparing pseudo-labels from two rotated views—mirrors the actual error relative to GT, as shown in Fig. 2(a). This relationship, illustrated by the polyline in Fig. 2(b), shows a positive correlation between mean pixel-wise inconsistency and the mean absolute error (MAE) across different parts. Consequently, areas

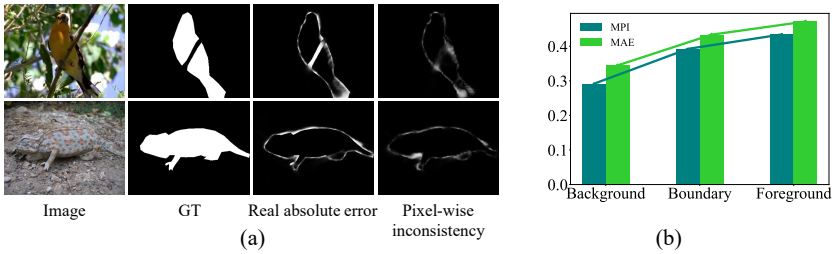


Fig. 2: (a) The real absolute error is obtained by subtracting the pseudo-label from the ground truth, reflecting the actual noisy level. On the other hand, pixel-wise inconsistency is derived from the subtraction of pseudo-labels from two rotated views. Visually, pixel-wise inconsistency and real absolute error appear very similar. (b) We define the area within 20 pixels from the edge as the boundary according to GT. We calculate mean pixel-wise inconsistency (MPI) and mean absolute error (MAE) in background, foreground, and boundary. The pseudo-labels are obtained by our proposed model on the test dataset COD10K [7] at a 10% semi-supervised setting.

with higher pixel-wise inconsistency are more prone to inaccuracies, suggesting a need for their de-emphasis during training. For instance-level noise, we determine that pseudo-labels with greater similarity across rotation views exhibit lower noise levels, as indicated in Fig. 3(a). A positive correlation between instance-wise consistency and the SSIM [39] computed by pseudo-labels and GT further supports this observation, depicted in Fig. 3(b). Thus, pseudo-labels demonstrating higher instance-wise consistency are likely of superior quality and should be prioritized in the learning process.

With the insights gained from these observations, we propose a semi-supervised camouflaged object detection framework called CamoTeacher, which incorporates a novel method known as Dual-Rotation Consistency Learning (DRCL). Specifically, DRCL operationalizes its strategy through two core components: Pixel-wise Consistency Learning (PCL) and Instance-wise Consistency Learning (ICL). PCL innovatively assigns variable weights to different parts within a pseudo-label, considering the pixel-wise inconsistency across diverse rotation view. Simultaneously, ICL adjusts the importance of individual pseudo-labels based on their instance-wise consistency, enabling a nuanced, noise-aware training process.

We adopt SINet [7] as the base model to implement CamoTeacher and apply it to more classic COD models, *i.e.* CNN-based model SINet-v2 [6] and SegMaR [17] and Transformed-based model DTINet [24] and FSPNet [14]. Extensive experiments on four COD benchmark datasets, *i.e.* CAMO [21], CHAMELEON [33], COD10K [7], and NC4K [25], demonstrate that the proposed CamoTeacher not only achieves state-of-the-art compared with semi-supervised learning methods, but also rivals established fully-supervised learning methods. Specifically, as shown on Fig. 1(b), with only 20% labeled data, it nearly reached the performance of fully supervised model on COD10K.

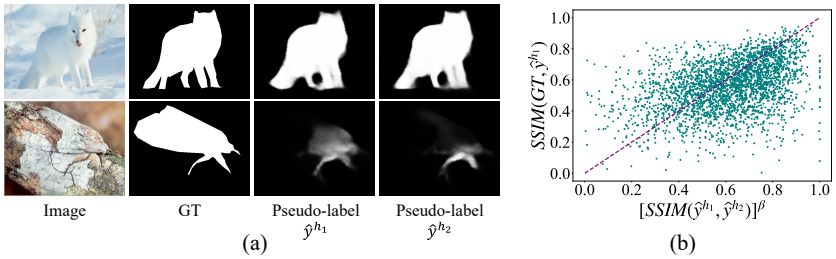


Fig. 3: (a) The more similar the pseudo-labels are between two rotated views, the lower the noise level and the higher the quality of the pseudo-labels. For example, in the first row, the pseudo-labels are more similar, indicating a lower noise level compared to the pseudo-labels in the second row. (b) The positive correlation between instance-wise consistency and the SSIM [39] of pseudo-labels and GT.

Our contributions can be summarized as follows:

- We introduce the first end-to-end semi-supervised camouflaged object detection framework CamoTeacher, offering a simple yet effective baseline for future research in semi-supervised camouflaged object detection.
- To address the problem of large amounts of noise in pseudo-labels in semi-supervised camouflaged object detection, we proposed the Dual-Rotation Consistency Learning (DRCL), which includes Pixel-wise Consistency Learning (PCL) and Instance-wise Consistency Learning (ICL), allows for the adaptive adjustment of the contribution of pseudo-labels with varying quality, enabling the effective utilization of pseudo-label information.
- We conducted extensive experiments on COD benchmark datasets, and achieved significant improvements compared to the fully supervised setting.

2 Related works

2.1 Camouflaged Object Detection

Camouflaged Object Detection(COD) is different from traditional visual detection tasks as objects and backgrounds in traditional detection tasks have significant visual distinctions, whereas camouflaged objects are typically concealed within their surrounding environments [6, 7, 14, 17, 24, 29, 43]. In recent years, many models have been proposed on this task, which can be divided into CNN-based and Transformed-based. i) CNN-based models: SINet [7] is a simple yet effective baseline that views camouflage detection as searching and identification process. SINet-v2 [6] significantly improves performance by introducing a neighbor connection decoder and cascaded group-reversal attention design. SegMaR [17] firstly performs initial detection of camouflaged objects and then progressively enlarges the detected regions to achieve precise detection. FEDER [9] achieves impressive results by conducting feature decomposition and

edge reconstruction. ii) Transformed-based models: DTINet [24] proposes a dual-task interactive transformer to detect accurate position and detailed boundary. CamoFormer [43] presents masked separable attention to separate foreground and background regions. FSPNet [14] designs a feature shrinkage decoder hierarchically decode locality-enhanced neighboring transformer features. These outstanding works have achieved excellent results in COD, but they rely on large-scale pixel-level annotations, which are time-consuming and labor-intensive to obtain.

In response to the reliance of full-supervised COD on expensive pixel-level annotations, some unsupervised/weakly-supervised approaches have emerged [10–12, 46]. UCOS-DA [46] formulates the unsupervised camouflaged object segmentation as a source-free unsupervised domain adaptation task and designs a foreground-background-contrastive self-adversarial pipeline. SCOD [11] proposes the first weakly-supervised COD method with relabeled scribbles. WS-SAM [10] uses the provided sparse annotations as prompts for SAM [18] to generate segmentation masks to train the model. These methods reduce annotation costs but often exhibit lower performance compared to fully-supervised methods. Is it possible to leverage existing annotated data without additional labeling while ensuring accuracy? To address this issue, we propose a new benchmark that combines semi-supervised learning with camouflaged object detection.

2.2 Semi-Supervised Learning

In traditional supervised learning, training a high-performing model requires a substantial amount of labeled data. However, obtaining labeled data in practical applications is often costly and time-consuming. Consequently, the effective utilization of unlabeled data to improve model performance and generalization has become a critical concern. Semi-supervised learning addresses this challenge by leveraging unlabeled data to enhance the model's learning process [2, 8, 22, 27, 28, 30, 34, 36, 48], thereby improving its performance and generalization capability. Pseudo-labeling [3, 35, 37, 38, 40, 42] and consistency regularization [13, 19, 20, 23, 41] are popular techniques in the field of semi-supervised learning. These methods have shown great efficacy when combined, producing impressive results.

To address the issue of heavy reliance on labeled data in camouflage object detection, we propose the first end-to-end semi-supervised camouflage object detection method. By integrating semi-supervised learning, our approach aims to enhance the performance of camouflage object detection models while significantly reducing the need for annotated data.

3 Methodology

3.1 Task Formulation

Semi-supervised camouflaged object detection aims to leverage a limited amount of annotated data to train a detector capable of identifying objects that seamlessly blend into their surroundings. This task is inherently challenging due to

the low contrast between the objects and their background. Given a camouflaged object detection dataset D for training, we represent the labeled subset with M labeled samples as $D_L = \{x_i^l, y_i\}_{i=1}^M$, and the unlabeled subset with N unlabeled samples as $D_U = \{x_i^u\}_{i=1}^N$, where x_i^l and x_i^u represent the input image, y_i represent the corresponding annotation mask of labeled data. Typically, D_L constitutes a small fraction of the entire dataset D, highlighting the semi-supervised learning scenario with $M \ll N$. The emphasis on $M \ll N$ underscores the challenge and opportunity within semi-supervised learning: to enhance detection capabilities by utilizing the untapped potential of the unlabeled data D_U, which far exceeds the labeled subset D_L.

3.2 Overall Framework

As shown in Fig. 4, we adopt Mean Teacher [36] as preliminary to achieve end-to-end semi-supervised camouflaged object detection framework, which contains two COD models with the same structure, *i.e.*, teacher model and student model, parameterized by Θ_t and Θ_s, respectively. The teacher model generates pseudo-labels, which are then used to optimize the student model. The overall loss function L can be defined as:

$$L = L_s + \lambda_u L_u, \tag{1}$$

where L_s and L_u denote supervised loss and unsupervised loss respectively, λ_u is the weight of unsupervised loss to balance the loss terms. Following the classic COD method, we use binary cross-entropy loss L_{bce} for training.

During the training process, we employ a combination of weak augmentation $\mathcal{A}^w(\cdot)$ and strong augmentation $\mathcal{A}^s(\cdot)$ strategies. Weak augmentation is applied to labeled data to mitigate overfitting, while unlabeled data undergo various data perturbations under strong augmentation to create diverse perspectives of the same image. The supervised loss L_s is defined by:

$$L_s = \frac{1}{M} \sum_{i=1}^{M} L_{bce}(F(\mathcal{A}^w(x_i^l); \Theta_s), y_i), \tag{2}$$

where $F(\mathcal{A}(x_i); \Theta)$ represents the detection results obtained by the model Θ for the i-th image with augmentations $\mathcal{A}(\cdot)$. For unlabeled images, we first apply weak augmentation, denoted as $\mathcal{A}^w(\cdot)$, before forwarding them to the teacher model. This initial step is crucial for generating reliable pseudo-labels $\widehat{y_i}$ under variations that do not significantly alter the core features of the images. These pseudo-labels serve as a form of soft supervision for the student model. Subsequently, the same images are subjected to strong augmentation, $\mathcal{A}^s(\cdot)$, and then passed to the student model. This process introduces a higher level of variability and complexity, simulating more challenging conditions for the student model to adapt to. The student model generates predictions p_i based on these strongly augmented images, leveraging the pseudo-labels $\widehat{y_i}$ for guidance in learning from the unlabeled data. It can be formulated as:

$$\widehat{y_i} = F(\mathcal{A}^w(x_i^u); \Theta_t), \ p_i = F(\mathcal{A}^s(\mathcal{A}^w(x_i^u)); \Theta_s). \tag{3}$$

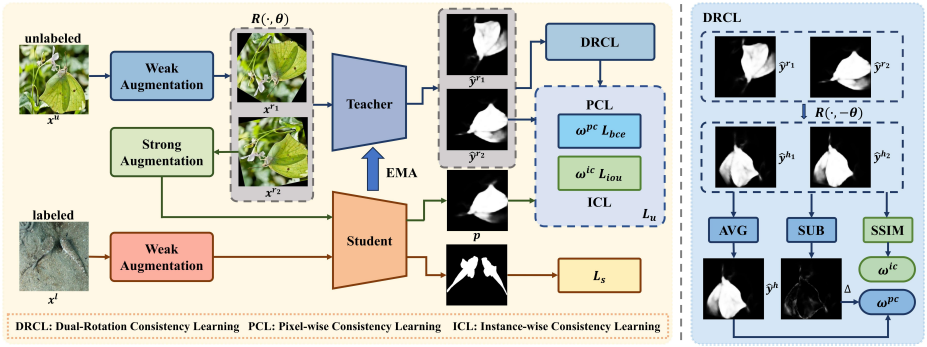


Fig. 4: The overall pipeline of our CamoTeacher. CamoTeacher consists of a teacher model and a student model. The teacher model generates pseudo-labels to optimize the student model, while the teacher model is updated through EMA [36] from the student. To mitigate pseudo-labels' noise, we introduced Dual-Rotation Consistency Learning (DRCL), which involves Pixel-wise Consistency Learning (PCL) and Instance-wise Consistency Learning (ICL). DRCL weights the loss function based on pixel-wise inconsistency and instance-wise consistency in pseudo-labels across different rotation views.

Hence, the unsupervised loss L_u can be expressed as:

$$L_u = \frac{1}{N}\sum_{i=1}^{N} L_{bce}(p_i, \widehat{y}_i). \tag{4}$$

Finally, the student model is intensively trained using the total loss L, which encapsulates both the supervised and unsupervised learning aspects of the semi-supervised framework. This approach ensures that the student model benefits from both labeled and pseudo-labeled data, enhancing its detection capabilities. Concurrently, the teacher model is methodically updated through the Exponential Moving Average (EMA) mechanism [36], which effectively distills students' knowledge while preventing noise, formulated as:

$$\Theta_t \leftarrow \eta\Theta_t + (1-\eta)\Theta_s, \tag{5}$$

where η is hyper-parameters denotes the keeping rate.

3.3 Dual-Rotation Consistency Learning

Due to the camouflaging nature of the objects, pseudo-labels contain a substantial amount of noise, and directly using them to optimize the student model potentially harm the model's performance. To address this problem, one of the most intuitive possible methods is to set a fixed high threshold to filter high-quality pseudo-labels, but this results in a low recall rate and makes it difficult to fully utilize the supervision information of pseudo-labels. To this end, we propose

Dual-Rotation Consistency Learning (DRCL) to dynamically adjust the weight of the pseudo-labels to reduce the impact of noise.

We randomly apply two separate rotations to the image x_i, which has previously been flipped and random resized, resulting in two distinct rotation views $x_i^{r_1}$ and $x_i^{r_2}$.

$$x_i^{r_1} = R(\mathcal{A}^w(x_i), \theta_1),\ x_i^{r_2} = R(\mathcal{A}^w(x_i), \theta_2), \tag{6}$$

where $x_i^r = R(x_i, \theta)$ represents rotate the input image x_i by θ degrees. We input the acquired rotation view into the teacher model to obtain the corresponding prediction, *i.e.* $\widehat{y}_i^r = F(x_i^r; \Theta_t)$. Subsequently, we apply an opposite rotation of $-\theta$ to return the prediction to its original horizontal orientation, resulting in $\widehat{y}_i^{h_1}$ and $\widehat{y}_i^{h_2}$, facilitating the calculation of prediction inconsistency under different rotation views, respectively,

$$\widehat{y}_i^{h_1} = R(\widehat{y}_i^{r_1}, -\theta_1),\ \widehat{y}_i^{h_2} = R(\widehat{y}_i^{r_2}, -\theta_2). \tag{7}$$

Note that rotation introduces black boundary areas, which are not involved in the implement of DRCL.

Since the noise levels vary across different regions of pseudo-labels and among different pseudo-labels, we introduce PCL and ICL dynamically adapt the contributions of different pixels within pseudo-labels and across various pseudo-labels.

Pixel-wise Consistency Learning We subtract the horizontal predictions $\widehat{y}_i^{h_1}$ and $\widehat{y}_i^{h_2}$ at the pixel level to obtain the pixel-wise inconsistency Δ_i,

$$\Delta_i = |\widehat{y}_i^{h_1} - \widehat{y}_i^{h_2}|. \tag{8}$$

The pixel-wise inconsistency Δ_i across different views reflects the reliability of pseudo-labels. However, in cases where both rotation views' prediction are close to 0.5, Δ_i cannot effectively distinguish them. These predictions exhibit high uncertainty, meaning they cannot be definitively classified as foreground or background, and are likely to represent noisy labels. Consequently, it becomes necessary to diminish their influence by lowering their weights. Hence, we calculate the average of horizontal predictions $\widehat{y}_i^h$,

$$\widehat{y}_i^h = avg(\widehat{y}_i^{h_1}, \widehat{y}_i^{h_2}), \tag{9}$$

where $avg(\cdot, \cdot)$ represents calculate the average of two inputs in pixel-level, and the use its L2 distance from 0.5 as a component to adjust the weights.

As a result, we derive the pixel-wise consistency weight ω_i^{pc} based on the pixel-wise inconsistency among different rotation views, as:

$$\omega_i^{pc} = (1 - \Delta_i^{\alpha})||\widehat{y}_i^h - \mu||_2^2, \tag{10}$$

where α is a hyper-parameter and $\mu = 0.5$. This dynamic pixel-wise consistency weight ω_i^{pc} assigns greater weight to regions with consistent predictions across various rotation views, while regions with inconsistent predictions receive a smaller weight.

In conclusion, we formulate the PCL loss L_u^{PC} as:

$$\begin{aligned} L_u^{PC} &= \frac{1}{N}\sum_{i=1}^{N}\omega_i^{pc}L_{bce}(p_i,\widehat{y}_i^{r_1}) \\ &= -\frac{1}{NHW}\sum_{i=1}^{N}\sum_{j=1}^{H\times W}\omega_{i,j}^{pc}[\widehat{y}_{i,j}^{r_1}\log p_{i,j} \\ &\qquad + (1-\widehat{y}_{i,j}^{r_1})\log(1-p_{i,j})], \end{aligned} \tag{11}$$

adaptively adjusts the weight of each pixel to ensure comprehensive supervision of student model while avoiding confirmation bias.

Instance-wise Consistency Learning The degree of camouflage varies between different images, leading to significant variations in the quality of pseudo-labels across images. Treating all pseudo-labels equally would be unreasonable. Unfortunately, for unlabeled images, assessing the quality of pseudo-labels is challenging because there are no GT labels available. We have observed a positive correlation between the instance-wise consistency of two rotation views and the quality of pseudo-labels, as quantified by SSIM [39]. Base on this, we introduce ICL to adjust the contributions of pseudo-labels with varying quality. We represent the instance-wise consistency weight ω_i^{ic} as follows:

$$\omega_i^{ic} = (SSIM(\widehat{y}_i^{h_1},\widehat{y}_i^{h_2}))^{\beta}, \tag{12}$$

where β is a hyper-parameter used to adjust the distribution relationship between instance-wise consistency and pseudo-label quality.

We use intersection-over-union (IoU) loss as the instance-wise restriction, therefore, the ICL loss can be expressed as:

$$\begin{aligned} L_u^{IC} &= \frac{1}{N}\sum_{i=1}^{N}\omega_i^{ic}L_{iou}(p_i,\widehat{y}_i^{r_1}) \\ &= \frac{1}{NHW}\sum_{i=1}^{N}\sum_{j=1}^{H\times W}\omega_i^{ic}\left(1-\frac{p_{i,j}\widehat{y}_{i,j}^{r_1}}{p_{i,j}+\widehat{y}_{i,j}^{r_1}-p_{i,j}\widehat{y}_{i,j}^{r_1}}\right). \end{aligned} \tag{13}$$

Therefore, the final total L loss comprises three components: supervised loss L_s, PCL loss L_u^{LC}, and ICL loss L_u^{GC} , which can be expressed as:

$$L = L_s + \lambda_u^{pc}L_u^{PC} + \lambda_u^{ic}L_u^{IC}, \tag{14}$$

where $\lambda_u^{pc}, \lambda_u^{ic}$ are the hyper-parameters.

4 Experiment

4.1 Experiment Settings

Dataset. We evaluate our CamoTeacher on four benchmark datasets, CAMO [21], CHAMELEON [33], COD10K [7], and NC4K [25]. In CAMO, there are

Table 1: Experimental results on four COD benchmark datasets includes CAMO, CHAMELEON, COD10K, NC4K. † denotes using the same data augmentation as semi-supervised training. ↑ indicates the higher the score the better, and ↓ indicates the lower the better. "n%" means that "n%" of the data was utilized as labeled data, with the remainder as unlabeled data. As the data partitioning is random, we conducted 5 fold experiments and the results are the **average of all 5 folds.**

CHAMELEON (76)

Methods	1% (41)						5% (202)						10% (404)					
	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	$E_\phi^m\uparrow$	$E_\phi^x\uparrow$	$F_\beta^m\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	$E_\phi^m\uparrow$	$E_\phi^x\uparrow$	$F_\beta^m\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	$E_\phi^m\uparrow$	$E_\phi^x\uparrow$	$F_\beta^m\uparrow$	$M\downarrow$
Supervised	0.605	0.268	0.582	0.741	0.435	0.216	0.613	0.283	0.601	0.732	0.439	0.185	0.714	0.434	0.725	0.829	0.592	0.110
Supervised†	0.635	0.301	0.630	0.766	0.479	0.196	0.659	0.331	0.655	0.787	0.499	0.159	0.728	0.454	0.744	0.844	0.607	0.106
Mean Teacher	0.537	0.199	0.418	0.636	0.229	0.204	0.611	0.309	0.524	0.745	0.353	0.137	0.679	0.450	0.650	0.812	0.512	0.102
CamoTeacher	**0.652**	**0.476**	**0.714**	**0.762**	**0.558**	**0.093**	**0.729**	**0.587**	**0.785**	**0.822**	**0.656**	**0.070**	**0.756**	**0.617**	**0.813**	**0.851**	**0.684**	**0.065**

CAMO (250)

Methods	1% (41)						5% (202)						10% (404)					
	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	$E_\phi^m\uparrow$	$E_\phi^x\uparrow$	$F_\beta^m\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	$E_\phi^m\uparrow$	$E_\phi^x\uparrow$	$F_\beta^m\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	$E_\phi^m\uparrow$	$E_\phi^x\uparrow$	$F_\beta^m\uparrow$	$M\downarrow$
Supervised	0.573	0.271	0.541	0.704	0.402	0.251	0.592	0.298	0.581	0.704	0.432	0.216	0.652	0.390	0.647	0.760	0.525	0.159
Supervised †	0.596	0.298	0.580	0.717	0.442	0.235	0.610	0.316	0.595	0.720	0.450	0.205	0.667	0.409	0.666	0.779	0.546	0.153
Mean Teacher	0.518	0.207	0.399	0.620	0.227	0.226	0.575	0.286	0.482	0.708	0.322	0.184	0.625	0.397	0.578	0.773	0.454	0.150
CamoTeacher	**0.621**	**0.456**	**0.669**	**0.736**	**0.545**	**0.136**	**0.669**	**0.523**	**0.711**	**0.775**	**0.601**	**0.122**	**0.701**	**0.560**	**0.742**	**0.795**	**0.635**	**0.112**

COD10K (2026)

Methods	1% (41)						5% (202)						10% (404)					
	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	$E_\phi^m\uparrow$	$E_\phi^x\uparrow$	$F_\beta^m\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	$E_\phi^m\uparrow$	$E_\phi^x\uparrow$	$F_\beta^m\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	$E_\phi^m\uparrow$	$E_\phi^x\uparrow$	$F_\beta^m\uparrow$	$M\downarrow$
Supervised	0.566	0.189	0.576	0.723	0.356	0.209	0.596	0.220	0.627	0.747	0.403	0.164	0.675	0.336	0.726	0.810	0.524	0.095
Supervised †	0.582	0.206	0.609	0.743	0.386	0.194	0.621	0.246	0.659	0.774	0.444	0.147	0.680	0.338	0.726	0.812	0.528	0.097
Mean Teacher	0.546	0.168	0.441	0.633	0.226	0.161	0.621	0.272	0.555	0.732	0.343	0.107	0.683	0.404	0.666	0.799	0.482	0.078
CamoTeacher	**0.699**	**0.517**	**0.788**	**0.797**	**0.582**	**0.062**	**0.745**	**0.583**	**0.827**	**0.840**	**0.644**	**0.050**	**0.759**	**0.594**	**0.836**	**0.854**	**0.652**	**0.049**

NC4K (4121)

Methods	1% (41)						5% (202)						10% (404)					
	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	$E_\phi^m\uparrow$	$E_\phi^x\uparrow$	$F_\beta^m\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	$E_\phi^m\uparrow$	$E_\phi^x\uparrow$	$F_\beta^m\uparrow$	$M\downarrow$	$S_\alpha\uparrow$	$F_\beta^w\uparrow$	$E_\phi^m\uparrow$	$E_\phi^x\uparrow$	$F_\beta^m\uparrow$	$M\downarrow$
Supervised	0.618	0.291	0.604	0.752	0.474	0.213	0.649	0.332	0.660	0.773	0.524	0.179	0.727	0.462	0.754	0.833	0.643	0.116
Supervised †	0.638	0.313	0.639	0.768	0.510	0.201	0.676	0.362	0.689	0.798	0.562	0.163	0.735	0.471	0.760	0.837	0.650	0.114
Mean Teacher	0.541	0.213	0.424	0.637	0.258	0.193	0.634	0.355	0.556	0.767	0.420	0.140	0.700	0.492	0.670	0.827	0.565	0.109
CamoTeacher	**0.718**	**0.599**	**0.779**	**0.814**	**0.675**	**0.090**	**0.777**	**0.677**	**0.834**	**0.859**	**0.739**	**0.071**	**0.791**	**0.687**	**0.842**	**0.868**	**0.746**	**0.068**

2500 images, consisting of 1250 camouflage images and 1250 non-camouflage images. CHAMELEON contains 76 manually annotated images. COD10K dataset consists of 5066 camouflage images, 3000 background images, and 1934 non-camouflage images. NC4K is another large-scale COD test dataset that comprises 4,121 images. Based on the data partitioning in previous works [7, 29, 45], we used 3040 images from COD10K and 1000 images from CAMO as the training set for our experiments. The remaining images from both datasets were used as the test set. During the training process, we followed a semi-supervised segmentation [15, 47] data partitioning approach. We sampled 1%, 5%, 10%, 20%, and 30% of the training set as labeled data, while the remaining portion was used as unlabeled data.

Evaluation Metrics. Referring to previous work [16, 17, 29], we used 6 common evaluation indicators in COD to evaluate our CamoTeacher, including S-measure(S_α) [4] , weight F-measure(F_β^w) [26] , mean E-measure(E_ϕ^m) [5] , max E-measure(E_ϕ^x), mean F-measure(F_β^m) [1] , and mean absolute error (M).

Implementation Details. The proposed CamoTeacher is implemented with PyTorch. We adopt SINet [7] as the baseline of COD model. We use an SGD optimizer with a momentum of 0.9 and a polynomial learning-rate decay with an initial value of 0.01 to train the student model. The training epoch is set to 40 with 10 epoch for burn-in phase. And the batch size is 20, with a ratio of

Table 2: Results of CamoTeacher on different COD models with 10% labeled data on COD10K.

Model	Setting	$S_\alpha \uparrow$	$F_\beta^w \uparrow$	$E_\phi^m \uparrow$	$E_\phi^x \uparrow$	$F_\beta^m \uparrow$	$M \downarrow$
CNN-Based Models							
SINet-v2 [6] (TPAMI2022)	supervised	0.732	0.544	0.819	0.835	0.586	0.056
	CamoTeacher	**0.760**	**0.605**	**0.835**	**0.843**	**0.652**	**0.051**
SegMaR [17] (CVPR2022)	supervised	0.755	0.583	0.837	0.851	0.623	0.050
	CamoTeacher	**0.775**	**0.629**	**0.856**	**0.864**	**0.675**	**0.045**
Transformer-Based Models							
DTINet [24] (ICPR2022)	supervised	0.771	0.626	0.857	0.865	0.670	0.043
	CamoTeacher	**0.794**	**0.661**	**0.873**	**0.882**	**0.695**	**0.040**
FSPNet [14] (CVPR2023)	supervised	0.789	0.622	0.823	0.867	0.675	0.042
	CamoTeacher	**0.817**	**0.684**	**0.884**	**0.903**	**0.720**	**0.037**

1:1 between labeled and unlabeled data, *i.e.* each batch consists of 10 labeled and 10 unlabeled images. During training and inference, each image is resized to 352×352. The teacher model is updated via EMA with a momentum η of 0.996. Weak data augmentation involves Random Flip and Random Scale, while strong data augmentation entails color space transformations including Identity, Autocontrast, Equalize, Gaussian blur, Contrast, Sharpness, Color, Brightness, Hue, Posterize, Solarize, with a random selection of up to 3 from this list.

4.2 Main Results

Comparison with Baselines. In Tab. 1, we compared CamoTeacher, the supervised baseline [7], and Mean Teacher [36]. Mean Teacher does not show improvements over the supervised in the semi-supervised setting due to treating all pseudo-labels and regions equally, which does not account for pseudo-labels noise problem. For instance, on the NC4K test set, under the 1% labeled data, the S_α dropped by 9.7% compared to supervised. Compared to Mean Teacher, CamoTeacher integrated with DRCL effectively mitigates pseudo-label noise issues and efficiently utilizes pseudo-label information. It demonstrates substantial improvements over both supervised and Mean Teacher across various proportions of labeled data, especially in scenarios with extremely limited. For instance, on 1% NC4K, CamoTeacher exhibited a substantial increase of 28.6% in the F_β^w and a 12.3% enhancement of S_α. Compared with Mean Teacher, the performance gains by CamoTeacher are more significant, which are 38.6% and 17.7%, espectively.

In Fig. 1(b), we visualize the performance changes of supervised, Mean teacher and our CamoTeacher on COD10K under different labeled data ratios of 1%, 5%, 10%, and further 20%, 30%, which intuitively reflects the effectiveness of CamoTeacher. It's evident that metrics like F_β^w, F_β^m, E_ϕ^m, and M surpass 100% of fully supervised performance with just 5% labeled data. Additionally, with 30% labeled data, all metrics significantly exceed 100% of fully supervised performance.

Table 3: Result of CamoTeacher on COD10K with additional data. We use 100% train dataset as labeled data and test dataset as the unlabeled data. "U" represents the unsupervised setting, "W" denotes the weakly-supervised setting, where weakly-supervision results from scribble annotations. "*L100%*" indicates using 100% of the training set as labeled data, "*U(CAMO)*" denotes utilizing CAMO as unlabeled data, and likewise for others.

Model	Setting	$S_\alpha \uparrow$	$F_\beta^w \uparrow$	$E_\phi^m \uparrow$	$E_\phi^x \uparrow$	$F_\beta^m \uparrow$	$M \downarrow$
Unsupervised							
SelfMask [31] (CVPRW2022)	U	0.645	0.440	0.687	0.728	0.478	0.125
FOUND [32] (CVPR2023)	U	0.670	0.482	0.751	0.753	0.520	0.085
UCOS-DA [46] (ICCVW2023)	U	0.689	0.513	0.740	0.741	0.546	0.086
Weakly-Supervised							
SCWSSOD [44] (AAAI2021)	W	0.710	0.546	0.805	-	-	0.055
SCOD [11](AAAI2023)	W	0.733	0.576	0.832	0.845	0.633	0.049
WS-SAM [10] (NeurIPS2023)	W	0.803	-	0.878	-	-	0.038
Semi-Supervised							
CamoTeacher $_{SINet}$	*L100%*	0.771	0.551	0.806	0.868	0.634	0.051
	L100%+U(CAMO)	0.779	0.554	0.799	0.875	0.632	0.051
	L100%+U(NC4K)	0.785	0.574	0.812	0.879	0.642	0.049
	L100%+U(CAMO,NC4K)	0.787	0.576	0.813	0.881	0.644	0.048

Application to Existing COD Methods. We applied the CamoTeacher approach to additional COD models, including CNN-based model SINet-v2 [6] and SegMaR [17] and Transformed-based model DTINet [24] and FSPNet [14], and the corresponding results are presented in Tab. 2. In terms of SINet-v2, CamoTeacher demonstrated a 2.8% improvement in the S_α metric and a 6.1% improvement in the F_β^w metric compared to the supervised baseline. Based on FSPNet, CamoTeacher achieved a 2.8% enhancement in the S_α metric and a 6.2% improvement in the F_β^w metric without utilizing additional discriminative annotation. These outcomes provide robust evidence of the effectiveness of CamoTeacher.

Addition unlabeled data. In Tab. 3, based on the base model SINet, we train CamoTeacher with 100% labeled data and check whether additional unlabeled data can further improve the performance. Compared with unsupervised and weakly-supervised methods, it shows the superiority of our method. We use the testset CAMO and NC4K as unlabeled data, and we find that adding more additional data CamoTeacher can make SINet continued to gain, such as S_α metric increased by 1.6% and F_β^w metric increased by 2.5%. We achieved significantly better performance than unsupervised methods [31,32,46] using existing labeled data. Compared with the weakly-supervision methods [10,11,44], we do not need additional annotation for the new unlabeled data, and to a certain extent, as the more unlabeled data, the performance will get better and better. And at the same time, replacing the base model with a stronger model will get better results.

Table 4: The ablation study of Pixel-wise Consistency Learning (PCL) and Instance-wise Consistency Learning (ICL) with 5% labeled data on COD10K.

PCL		ICL							
L_{bce}	ω^{pc}	L_{iou}	ω^{ic}	$S_\alpha \uparrow$	$F_\beta^w \uparrow$	$E_\phi^m \uparrow$	$E_\phi^x \uparrow$	$F_\beta^m \uparrow$	$M \downarrow$
✓				0.682	0.370	0.636	0.787	0.461	0.091
		✓		0.734	0.572	0.817	0.821	0.626	0.057
✓	✓			0.724	0.501	0.738	0.832	0.574	0.058
		✓	✓	0.742	0.585	0.828	0.834	0.639	0.052
✓	✓	✓		0.745	0.586	0.830	0.840	0.643	0.051
✓	✓	✓	✓	**0.746**	**0.586**	**0.830**	**0.843**	**0.645**	**0.050**

Table 5: The ablation study of various weight approach of Pixel-wise Consistency Learning (PCL) with 5% labeled data on COD10K. In this experiment α is set to 1/4.

weight approach	$S_\alpha \uparrow$	$F_\beta^w \uparrow$	$E_\phi^m \uparrow$	$E_\phi^x \uparrow$	$F_\beta^m \uparrow$	$M \downarrow$
$\omega^{pc} = \widehat{y}^r$	0.698	0.393	0.684	0.804	0.500	0.110
$\omega^{pc} = \|\|\widehat{y}^r - \mu\|\|_2^2$	0.717	0.485	0.734	0.822	0.559	0.061
$\omega^{pc} = 1 - \Delta^\alpha$	0.719	0.481	0.720	0.827	0.553	0.062
$\omega^{pc} = (1 - \Delta^\alpha)\widehat{y}^r$	0.695	0.405	0.696	0.795	0.506	0.114
$\omega^{pc} = (1 - \Delta^\alpha)\|\|\widehat{y}^r - \mu\|\|_2^2$	**0.724**	**0.501**	**0.738**	**0.832**	**0.574**	**0.058**

Table 6: The ablation study of hyper-parameters α in Pixel-wise Consistency Learning (PCL) with 5% labeled data on COD10K. In this experiment β is set to 4, λ_u^{PC} is set to 8, and λ_u^{IC} is set to 0.3.

α	$S_\alpha \uparrow$	$F_\beta^w \uparrow$	$E_\phi^m \uparrow$	$E_\phi^x \uparrow$	$F_\beta^m \uparrow$	$M \downarrow$
2	0.737	0.565	0.813	0.836	0.629	0.052
1	0.739	0.566	0.815	0.832	0.627	0.056
1/2	0.745	0.575	0.820	0.839	0.637	**0.051**
1/4	**0.747**	**0.579**	**0.825**	**0.842**	**0.638**	**0.051**
1/8	0.745	0.578	**0.825**	0.838	0.636	0.052

Table 7: The ablation study of hyper-parameters β in Instance-wise Consistency Learning (ICL) with 5% labeled data on COD10K. In this experiment α is set to 1/4, λ_u^{PC} is set to 8, and λ_u^{IC} is set to 0.3.

β	$S_\alpha \uparrow$	$F_\beta^w \uparrow$	$E_\phi^m \uparrow$	$E_\phi^x \uparrow$	$F_\beta^m \uparrow$	$M \downarrow$
1	0.743	0.576	0.824	0.839	0.636	**0.051**
2	0.742	0.573	0.823	0.837	0.632	0.052
4	**0.747**	**0.579**	**0.825**	**0.842**	0.638	**0.051**
5	0.744	0.578	0.824	0.841	**0.639**	**0.051**
6	0.745	0.575	0.822	0.839	0.635	0.052

4.3 Ablation Study

Ablation on PCL and ICL. We conducted ablation experiments on different combinations of weight and loss functions for PCL and ICL. From Tab. 4, increasing the weights ω^{pc} and ω^{ic} improves the model's performance. Specifically, with ω^{pc} resultes in a higher F_β^w score, increasing from 0.370 to 0.501. Similarly, with ω^{ic} improved the F_β^w metric from 0.572 to 0.585. Combining these two weighted loss functions led to a significant performance improvement, with the F_β^w score increasing from 0.370 to 0.586 (an increase of 21.6%). This demonstrates the effectiveness of PCL and ICL in enhancing the model's performance.

Ablation of weight approach in PCL. The choice of method for obtaining ω^{pc} significantly affects the performance of the model. Therefore, we conducted several experiments to select the most effective method for obtaining ω^{pc}. From Tab. 5, it can be observed that the best performance is achieved when ω^{pc} is obtained using Eq. (10).

Ablation of hyper-parameters in PCL and ICL. We perform ablation experiments on the hyper-parameters associated with the PCL and ICL modules. As shown in Tab. 6, the optimal results are achieved when setting the hyper-parameter α, which corresponds to the coefficient ω^{pc}, to 1/4. Similarly, according to Tab. 7, the best performance is observed when the hyper-parameter β, related to the coefficient ω^{ic}, is set to 4.

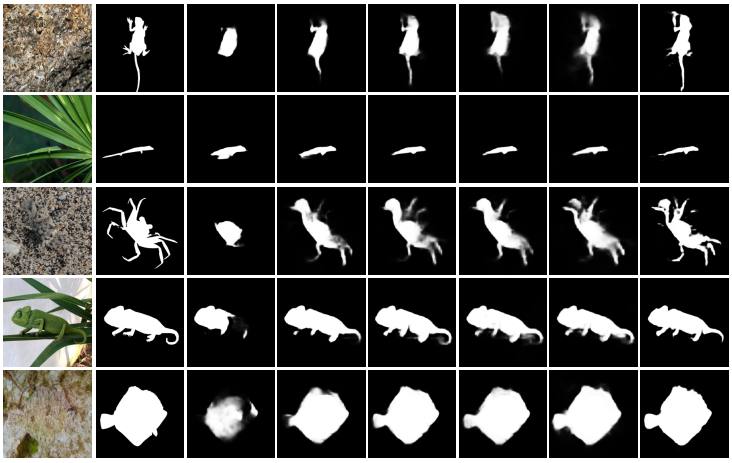


Fig. 5: Visualization of predictions in different proportions of labeled data.

4.4 Qualitative Analysis

Result visualization. We conducte visualizations of CamoTeacher's results under varying proportions of labeled data. The model is trained with 1%, 5%, 10%, 20% and 30% labeled data, and we compare its predicted masks with those produced by SINet under fully supervised conditions (as shown in Fig. 5). Notably, even with only 10% labeled data, CamoTeacher's predicted masks exhibit a remarkable similarity to the fully supervised SINet predictions. In fact, in certain detailed predictions, CamoTeacher even surpasses the performance of fully supervised SINet. These findings highlight the impressive capability of CamoTeacher in leveraging limited labeled data for effective camouflaged object detection.

5 Conclusions

In this paper, we present the first end-to-end semi-supervised camouflaged object detection model called CamoTeacher. To address the issue of substantial noise within pseudo-labels in semi-supervised COD, including both local noise and global noise, we've introduced a novel approach called Dual-Rotation Consistency Learning (DRCL), including Pixel-wise Consistency Learning (PCL) and Instance-wise Consistency Learning (ICL). DRCL assists the model in alleviating noise issues, effectively leveraging pseudo-labels information, so that the model can receive adequate supervision while avoiding confirmation bias. Extensive experiments validate CamoTeacher's superior performance while significantly reducing annotation expenses. We believe that CamoTeacher can serve as a strong baseline for future research on semi-supervised COD.

Acknowledgements

This work was supported by National Science and Technology Major Project (No. 2022ZD0118201), the National Science Fund for Distinguished Young Scholars (No.62025603), the National Natural Science Foundation of Chine a (No. U21B2037, No. U22B2051, No. U23A20383, No. 62176222, No. 62176223, No. 62176226, No. 62072386, No. 62072387, No. 62072389, No. 62002305 and No. 62272401), and the Natural Science Foundation of Fujian Province of China (No.2022J06001).

References

1. Achanta, R., Hemami, S., Estrada, F., Susstrunk, S.: Frequency-tuned salient region detection. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 1597–1604. IEEE (2009)
2. Chen, H., Tao, R., Fan, Y., Wang, Y., Wang, J., Schiele, B., Xie, X., Raj, B., Savvides, M.: Softmatch: Addressing the quantity-quality trade-off in semi-supervised learning. arXiv preprint arXiv:2301.10921 (2023)
3. Chen, X., Yuan, Y., Zeng, G., Wang, J.: Semi-supervised semantic segmentation with cross pseudo supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2613–2622 (2021)
4. Fan, D.P., Cheng, M.M., Liu, Y., Li, T., Borji, A.: Structure-measure: A new way to evaluate foreground maps. In: Proceedings of the IEEE international conference on computer vision. pp. 4548–4557 (2017)
5. Fan, D.P., Gong, C., Cao, Y., Ren, B., Cheng, M.M., Borji, A.: Enhanced-alignment measure for binary foreground map evaluation. arXiv preprint arXiv:1805.10421 (2018)
6. Fan, D.P., Ji, G.P., Cheng, M.M., Shao, L.: Concealed object detection. IEEE transactions on pattern analysis and machine intelligence **44**(10), 6024–6042 (2021)
7. Fan, D.P., Ji, G.P., Sun, G., Cheng, M.M., Shen, J., Shao, L.: Camouflaged object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2777–2787 (2020)
8. Grandvalet, Y., Bengio, Y.: Semi-supervised learning by entropy minimization. Advances in neural information processing systems **17** (2004)
9. He, C., Li, K., Zhang, Y., Tang, L., Zhang, Y., Guo, Z., Li, X.: Camouflaged object detection with feature decomposition and edge reconstruction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22046–22055 (2023)
10. He, C., Li, K., Zhang, Y., Xu, G., Tang, L., Zhang, Y., Guo, Z., Li, X.: Weakly-supervised concealed object segmentation with sam-based pseudo labeling and multi-scale feature grouping. Advances in Neural Information Processing Systems **36** (2024)
11. He, R., Dong, Q., Lin, J., Lau, R.W.: Weakly-supervised camouflaged object detection with scribble annotations. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 781–789 (2023)
12. Hu, J., Lin, J., Cai, W., Gong, S.: Relax image-specific prompt requirement in sam: A single generic prompt for segmenting camouflaged objects. arXiv preprint arXiv:2312.07374 (2023)

13. Hu, J., Chen, C., Cao, L., Zhang, S., Shu, A., Jiang, G., Ji, R.: Pseudo-label alignment for semi-supervised instance segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 16337–16347 (2023)
14. Huang, Z., Dai, H., Xiang, T.Z., Wang, S., Chen, H.X., Qin, J., Xiong, H.: Feature shrinkage pyramid for camouflaged object detection with transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5557–5566 (2023)
15. Hung, W.C., Tsai, Y.H., Liou, Y.T., Lin, Y.Y., Yang, M.H.: Adversarial learning for semi-supervised semantic segmentation. arXiv preprint arXiv:1802.07934 (2018)
16. Ji, G.P., Fan, D.P., Chou, Y.C., Dai, D., Liniger, A., Van Gool, L.: Deep gradient learning for efficient camouflaged object detection. Machine Intelligence Research **20**(1), 92–108 (2023)
17. Jia, Q., Yao, S., Liu, Y., Fan, X., Liu, R., Luo, Z.: Segment, magnify and reiterate: Detecting camouflaged objects the hard way. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4713–4722 (2022)
18. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. arXiv preprint arXiv:2304.02643 (2023)
19. Lai, X., Tian, Z., Jiang, L., Liu, S., Zhao, H., Wang, L., Jia, J.: Semi-supervised semantic segmentation with directional context-aware consistency. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1205–1214 (2021)
20. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. arXiv preprint arXiv:1610.02242 (2016)
21. Le, T.N., Nguyen, T.V., Nie, Z., Tran, M.T., Sugimoto, A.: Anabranch network for camouflaged object segmentation. Computer vision and image understanding **184**, 45–56 (2019)
22. Lee, D.H., et al.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on challenges in representation learning, ICML. vol. 3, p. 896. Atlanta (2013)
23. Liu, Y., Tian, Y., Chen, Y., Liu, F., Belagiannis, V., Carneiro, G.: Perturbed and strict mean teachers for semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4258–4267 (2022)
24. Liu, Z., Zhang, Z., Tan, Y., Wu, W.: Boosting camouflaged object detection with dual-task interactive transformer. In: 2022 26th International Conference on Pattern Recognition (ICPR). pp. 140–146. IEEE (2022)
25. Lv, Y., Zhang, J., Dai, Y., Li, A., Liu, B., Barnes, N., Fan, D.P.: Simultaneously localize, segment and rank the camouflaged objects. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11591–11601 (2021)
26. Margolin, R., Zelnik-Manor, L., Tal, A.: How to evaluate foreground maps? In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 248–255 (2014)
27. Mi, P., Lin, J., Zhou, Y., Shen, Y., Luo, G., Sun, X., Cao, L., Fu, R., Xu, Q., Ji, R.: Active teacher for semi-supervised object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14482–14491 (2022)
28. Oliver, A., Odena, A., Raffel, C.A., Cubuk, E.D., Goodfellow, I.: Realistic evaluation of deep semi-supervised learning algorithms. Advances in neural information processing systems **31** (2018)

29. Pang, Y., Zhao, X., Xiang, T.Z., Zhang, L., Lu, H.: Zoom in and out: A mixed-scale triplet network for camouflaged object detection. In: Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition. pp. 2160–2170 (2022)
30. Roelofs, B., Berthelot, D., Sohn, K., Carlini, N., Kurakin, A.: Adamatch: a unified approach to semi-supervised learning and domain adaptation. In: Proc. of the Int. Conf. on Learning Representations (ICLR) (2022)
31. Shin, G., Albanie, S., Xie, W.: Unsupervised salient object detection with spectral cluster voting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3971–3980 (2022)
32. Siméoni, O., Sekkat, C., Puy, G., Vobeckỳ, A., Zablocki, É., Pérez, P.: Unsupervised object localization: Observing the background to discover objects. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3176–3186 (2023)
33. Skurowski, P., Abdulameer, H., Błaszczyk, J., Depta, T., Kornacki, A., Kozieł, P.: Animal camouflage analysis: Chameleon database. Unpublished manuscript **2**(6), 7 (2018)
34. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in neural information processing systems **33**, 596–608 (2020)
35. Sohn, K., Zhang, Z., Li, C.L., Zhang, H., Lee, C.Y., Pfister, T.: A simple semi-supervised learning framework for object detection. arXiv preprint arXiv:2005.04757 (2020)
36. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems **30** (2017)
37. Wang, Y., Wang, H., Shen, Y., Fei, J., Li, W., Jin, G., Wu, L., Zhao, R., Le, X.: Semi-supervised semantic segmentation using unreliable pseudo-labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4248–4257 (2022)
38. Wang, Z., Li, Y., Guo, Y., Fang, L., Wang, S.: Data-uncertainty guided multi-phase learning for semi-supervised object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4568–4577 (2021)
39. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
40. Xu, M., Zhang, Z., Hu, H., Wang, J., Wang, L., Wei, F., Bai, X., Liu, Z.: End-to-end semi-supervised object detection with soft teacher. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3060–3069 (2021)
41. Yang, L., Qi, L., Feng, L., Zhang, W., Shi, Y.: Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7236–7246 (2023)
42. Yang, L., Zhuo, W., Qi, L., Shi, Y., Gao, Y.: St++: Make self-training work better for semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4268–4277 (2022)
43. Yin, B., Zhang, X., Hou, Q., Sun, B.Y., Fan, D.P., Van Gool, L.: Camoformer: Masked separable attention for camouflaged object detection. arXiv preprint arXiv:2212.06570 (2022)
44. Yu, S., Zhang, B., Xiao, J., Lim, E.G.: Structure-consistent weakly supervised salient object detection with local saliency coherence. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 3234–3242 (2021)

45. Zhang, M., Xu, S., Piao, Y., Shi, D., Lin, S., Lu, H.: Preynet: Preying on camouflaged objects. In: Proceedings of the 30th ACM International Conference on Multimedia. pp. 5323–5332 (2022)
46. Zhang, Y., Wu, C.: Unsupervised camouflaged object segmentation as domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4334–4344 (2023)
47. Zhao, Z., Yang, L., Long, S., Pi, J., Zhou, L., Wang, J.: Augmentation matters: A simple-yet-effective approach to semi-supervised semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11350–11359 (2023)
48. Zhao, Z., Zhou, L., Wang, L., Shi, Y., Gao, Y.: Lassl: Label-guided self-training for semi-supervised learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 9208–9216 (2022)
49. Zhou, Z.H., Zhou, Z.H.: Semi-supervised learning. Machine Learning pp. 315–341 (2021)