SF-CODnet: Spatial-Frequency Framework with Weak Sample Learning Strategy for Detecting Camouflaged Wildlife Objects

Margarita N. Favorskaya [1], Nazar V. Saidov [1]
[1] Reshetnev Siberian State University of Science and Technology, 31 Krasnoyarsky Rabochy ave., Krasnoyarsk, 660037 Russian Federation - favorskaya@sibsau.ru, nazar-saidov1999@mail.ru

Commission II, WG II/8

Keywords: Camouflaged Object Detection, Weak Examples, Wildlife Objects, Evaluation Metrics.

Abstract

The problem of camouflaged object detection (COD) is to identify hidden objects in their environment. This problem remained unsolved for many years until the advent of deep learning methods. The style of invisibility of the camouflaged object plays a significant role, which is difficult to capture only in the spatial area. At the same time, the structural properties of camouflaged patterns are characterized by a high discriminatory ability to distinguish camouflaged objects from the background. The spatial-frequency model called SF-CODnet achieves accurate structural segmentation, realizing not only semantic but also instance segmentation of wildlife camouflage objects. A learning strategy to synthesize weak samples also helped to generalize the COD ability of the baseline model. Unlike other popular learning strategies that improve samples as much as possible before training, our weak sample synthesis learning strategy helps to generalize the base model's COD ability. Such augmentation strategy and the proposed SF-CODnet model were tested using three publicly available datasets: CAMO, COD10K, and NC4K with good results, outperforming some COD models.

1. Introduction

There are many natural objects in the environment that have a protective camouflage mechanism developed by evolution. In recent years, camouflaged object detection (COD) has become a hot research topic with wide practical applications in ecology, agriculture, industry, military, medical imaging, and art. Camouflaged natural objects have a very similar texture, shape, colour to the background, resulting in hidden boundaries and deceptive textures. Object detection as a fundamental task in computer vision deals with various types of objects, which leads to different problem statements: generic object detection, salient object detection and camouflaged object detection. While salient objects stand out from the background, camouflaged objects, on the other hand, blend seamlessly into the environment, making COD an extremely challenging task. In contrast to COD, detection of general and salient objects has been well studied in recent decades and is actively used in practical applications nowadays. COD is classified into image-level and video-level. Image-level COD refers to object detection in still images, while video-level COD extends to the temporal domain with additional complexity due to dynamic scene changes and temporal continuity. This study focuses on the development of an image-level COD method.

The first traditional COD algorithms appeared in the 2010s and were based on hand-crafted features. They have proven to be ineffective and yielded poor detection results (Neider and Zelinsky, 2006; Hess et al., 2016). Traditional COD methods relied on low-level features specifically designed to capture nuances of textures, intensities, and colours. Texture-based methods aim to detect distinctive patterns in an image by analyzing the distribution of gray-scale pixels and their spatial neighbours. Intensity-based methods have evolved from simple intensity analysis to using 3D convexity, where the difference between 3D convex and 3D concave regions is assessed. This approach is similar to some deep learning COD methods based on depth estimation. In certain scenarios, colour contrast can also provide significant distinctiveness in detecting camouflaged objects. Often traditional algorithms used several features at once. However, their discriminatory ability was low and effective only in simple scenes.

The history of deep learning COD methods is short and began in the 2020s when the first large-scale datasets emerged. The first CNN called Anabrance network (Le et al., 2019) to solve the COD problem was proposed in 2019. Since then, several dozen deep learning models have been developed for the image-level COD task. Outstanding systematic surveys can be found in the literature (Liang et al., 2024; Xiao et al., 2024), where deep learning COD methods are broadly classified based on three main criteria: network architecture, learning paradigm, and supervision level. Systematic surveys show that CNNs are the most popular models, followed by transformer networks, diffusion networks, and capsule networks. At the same time, we see a gap in the application of generative adversarial networks (GANs), which are just beginning to be used to solve the COD problem (He et al., 2024a). The learning paradigm includes single-task learning and multi-task learning, and the supervision level is classified into four categories: fully supervised, weakly-supervised, semi-supervised, and unsupervised. The aim of this study is to develop an original weakly-supervised GAN-style learning strategy in the frequency domain. Our contribution is threefold:

1. We propose a learning strategy to synthesize weak samples in order to generalize the COD ability of the baseline model
2. The spatial-frequency model called SF-CODnet achieves accurate structural segmentation, realizing not only semantic but also instance segmentation of wildlife camouflage objects
3. We tested our augmentation strategy and the proposed SF-CODnet model using three publicly available datasets: CAMO (2,852 images), COD10K (6,000 images), and NC4K (4,121 images) with results superior to 3 models

The structure of this paper is following: Section 2 introduces the related work on four main categories of learning strategies

Publisher's note: Copernicus Publications has not received any payments from Russian or Belarusian institutions for this paper.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-2-W9-2025-79-2025 | © Author(s) 2025. CC BY 4.0 License.

aimed at solving the COD problem, augmentation methods and frequency-based COD methods. Background is presented in Section 3. The proposed augmentation approach using weak samples is discussed in Section 4. Our GAN-based detection of camouflaged wildlife objects in the spatial-frequency domain is described in Section 5. Section 6 contains the main experimental results. Section 7 concludes the paper.

2. Related Work

Four basic categories of learning strategy used to address the COD problem include:

1. Fully supervised strategy using complete ground truth data, typically ground truth masks of camouflaged objects created manually or using third party resources
2. Weakly-supervised strategy that applies limited or imprecise annotations (He et al., 2023)
3. Semi-supervised strategy that combines labeled and unlabeled data, generating pseudo labels for the latter (Fu et al., 2024)
4. Unsupervised strategy in which no explicit labels are specified (Zhang and Wu, 2023)

Besides these categories, we can mention aggregating multi-scale features, fusing multi-source data, multi-task learning, the "training-free" mode (Hu et al., 2024), among others.

Recently, many models have been proposed to solve the COD problem based on supervised strategy. Most COD models focus on the fully supervised strategy with original CNC and transformer-based architectures (Liang et al., 2024; Xiao et al., 2024). However, sometimes interesting learning ideas become the basis for implementation. Thus, a joint training method using camouflaged and salient objects was developed as an uncertainty-aware framework in (Li et al., 2021). First, the easy samples in COD dataset were applied to train a robust salient object detection model. Second, the similarity measure module explicitly modelled the contradicting attributes of the two tasks. Third, the uncertainty in salient and camouflaged object annotations was estimated using an adversarial learning network. This approach showed good COD results, especially on datasets with uncluttered images (DUTS, ECSSD, DUT, HKU-IS, and THUR).

Pre-training and fine-tuning are commonly used paradigms for training deep neural networks that detect objects, including camouflaged ones. The recent development of new pre-training methods, such as self-supervised training, has facilitated the transition from supervised strategies to unsupervised strategies, making pre-training easier for end users.

The first weakly supervised COD dataset with scribble annotations was presented in (He et al., 2023) with the aim of quickly labelling images with blue scribbles for background and red scribbles for foreground. This solution made it possible to speed up labelling by 360 times compared to pixel-wise annotation, but required the developing a new network that learned low-level contrast to expand scribbles to wider potential regions, and then analyzed logical semantic relations to determine the real foreground and background. Weakly-supervised concealed object segmentation (WSCOS) model was based on the "segment anything model" (SAM) to generate dense masks as pseudo labels (He et al., 2024b).

The first semi-supervised COD framework with a small amount of samples having noisy/incorrect annotations was developed in (Fu et al., 2024). To facilitate annotations, two techniques have been proposed, including loss re-weighting and ensemble learning. These authors calculated the loss weights specific to each pixel according to the neighbour information within the window (a window-based voting strategy) to check whether it belongs to a camouflaged object or not. Furthermore, the knowledge from different training moments provided by the momentum network allowed ensemble learning to be used to discard noise/outliers and generate relatively robust pseudo-labels for unlabeled images. A semi-supervised COD framework called CamoTeacher (Lai et al., 2025) incorporates dual-rotation consistency learning to effectively compensate for pseudo-label noise at both pixel and instance levels.

One of the main advantages of unsupervised methods is that they are able to perform pixel-wise classification at the instance level without using any manually created annotations. The unsupervised camouflaged object segmentation as domain adaptation (UCOS-DA) model proposed in (Zhang and Wu, 2023) consisted of three components: a self-supervised source model, a light-weighted target model and an adversarial domain adaptation module. This approach was based on the assumption that the blurriness of object boundaries is one of the main reasons for the large discrepancy between the camouflaged and the generic visual objects.

The three learning strategies mentioned above usually require augmentation as a way to increase the size and diversity of the training data. This is a challenging problem for camouflaged object detection and camouflaged instance segmentation that has not yet been thoroughly studied. Augmentation of images with camouflaged objects in the spatial domain may results in occlusions, deformations, or noise. Therefore, augmentation in the frequency domain is more preferable. An original augmentation method called CamoFA using the Fourier transform was proposed in (Le et al., 2024). This method involves mixing the low-frequency component of the reference image and the high-frequency component of the input image in order to transfer texture and colour information to the input image. A conditional GAN and cross-attention mechanism with adaptive parameters allowed to synthesize more visible camouflaged objects. Another strategy, the "prey-vs-predator" game, embodied in adversarial training framework called Camouflageator was implemented in (He et al., 2024a). The "prey" generates more deceptive camouflage objects while the "predator" should provide more precise detection results. Thus, the generator learns to generate more camouflaged objects, in other words, the generalizability of the model increases.

The main limitation of spatial-based COD methods is the inaccurate detection of object edges in complex natural scenes. Nowadays, the frequency-based COD methods do not predominate among the COD family. However, this branch of the COD family deserves attention, as does the development of hybrid models. In (Zhong et al., 2022), the frequency clues were embedded into a CNN model that had two separate flows – RGB flow and frequency flow. In this model, a frequency enhancement module based on the offline discrete cosine transform and a high-order relation module for handling the rich fusion features were developed. The two-stage frequency perception network (FPNet) included a frequency-guided coarse localization stage and a detail-preserving fine localization stage (Cong et al., 2023). The mechanism for separate frequency perception using offline discrete cosine transform was driven by a semantic hierarchy in the frequency domain. The feature decomposition and edge reconstruction (FEDER) model was developed to learn the intrinsic similarity of foreground and background using deep wavelet-like decomposition (He et al., 2023). The auxiliary task of edge reconstruction helped to generate precise masks of camouflaged objects with accurate boundaries. A COD model with edge perception in frequency domain (EPFDNet) was developed in (Fang et al., 2025). In this model, Res2Net-50 was used as the basis for initial feature extraction. The frequency components were obtained through deep wavelet-like transformation. The edge-target interaction and frequency-spatial fusion network (EFNet) proposed in (Guan et al., 2025) realized the idea of embedding frequency information into the spatial domain, facilitating frequency-spatial fusion. The frequency decomposition branch simulated the JPEG image compression process, using block-wise discrete cosine transformation, while the edge detection and object segmentation branches extracted spatial information.

A brief review of the literature indicates high activity in the field of COD. Although several dozen deep models have been developed in recent years, researchers continue to study increasingly complex cases in such a difficult problem, taking into account multimodality, depth, and various transforms.

3. Background

In general, the deep learning-based COD problem can be formulated as follows. Let I be an input RGB image, $I \in \mathrm{R}^{H\times W\times 3}$, where H and W represent the width and height of the image. The input image I needs to be transformed into a predicted camouflaged map C_{map} using a network F_{map} with trainable parameters θ_{map}:

$$C_{map} = F_{map}\left(I;\theta_{map}\right) \in \left[0,1\right]^{H\times W}, \quad (1)$$

where $C_{map} \in \mathrm{R}^{H\times W\times 1}$
F_{map} = the network
θ_{map} = the trainable parameters
I = the image
H, W = the width and height of the image

Let extend Equation 1 by adding the contour C_{con} of the camouflaged object:

$$C_{con} = C_{map}\left(\theta_{con}\right) \in \left[0,1\right], \quad (2)$$

where $C_{con} \in \mathrm{R}^{H\times W\times 1}$
θ_{con} = the trainable parameters

Let the dataset D contains N images, $D = \{I_n \in \mathrm{R}^{H\times W\times 3}\}$, $n = 1,...,$ N and corresponding ground-truth map labels $G_{map} \in \{0, 1\}^{H\times W}$ and ground-truth contour labels $G_{con} \in \{0, 1\}$. The goal is to find optimal parameters $\left(\hat{\theta}_{map},\hat{\theta}_{con}\right)$ that minimize the prediction error:

$$\left(\hat{\theta}_{map},\hat{\theta}_{con}\right) = \underset{\theta_{map},\theta_{con}}{\arg\min}\, Loss\left(\left(C_{map},G_{map}\right),\left(C_{con},G_{con}\right)\right), \quad (3)$$

where $Loss(\cdot)$ = the loss function

The ground-truth labels G_{map} and G_{con} can be collected by different techniques, including manually annotated pixel-level masks and contours.

4. Augmentation

The traditional augmentation techniques include stochastic augmentations randomly selected by flipping, rotation, and scaling within the given ranges of values. However, these types of data augmentation do not efficiently improve the model's ability to discriminate the foreground objects. We would like to discuss a special approach to augmentation related to the COD problem. The difficulty in distinguishing most camouflage objects is primarily due to their structure, texture or colour.

Detecting the boundaries of camouflaged objects is a more challenging task compared to detecting common and salient objects. Most COD-based augmentation techniques generate "good" samples for training with explicit boundaries of camouflaged objects. In contrast to this, we believe that generating weak samples with relatively blurred or distorted boundaries of camouflaged objects helps generalize the learning process. In other words, if we train the network to segment weak samples at the training stage, then at the testing stage the network will segment normal samples better.

Not all from possible types of blurring such as motion blur, defocus blur, lens blur, Gaussian blur, radial blur, zoom blur and natural blur (caused by environmental changes, such as heavy rain, snow, fog, dust, and so on) are suitable for our goal of controlled blurring of the boundaries of camouflaged objects. It is worth noting that we are talking about small local anisotropic or isotropic changes in boundaries. We used simple low-pass filters to smooth randomly selected areas of the boundaries as anisotropic blur and a slight Gaussian blur in the area of the camouflaged object as isotropic blur. Weak samples generated using these techniques with the OpenCV library are depicted in Figure 1.

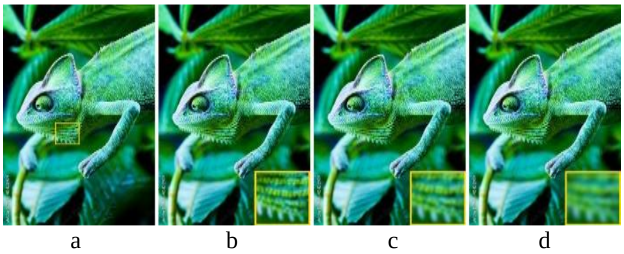

Figure 1. Visualization of generated weak samples: **a** original image, **b**, **c**, **d** generated samples with different Gaussian blur locally applied (**b** – 5% blur, **c** – 7% blur, **d** – 10% blur).

Figure 2 and Figure 3 show samples with randomly distorted texture and colour parameters, respectively.

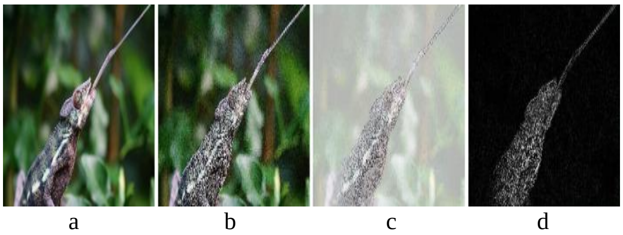

Figure 2. Visualization of generated samples with randomly distorted texture parameters: **a** original image, **b**, **c**, **d** generated samples.

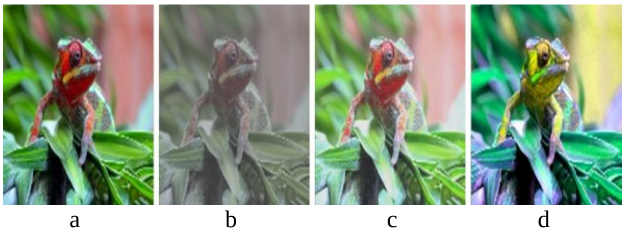

Figure 3. Visualization of generated samples with randomly distorted colour parameters: **a** original image, **b**, **c**, **d** generated samples.

Besides those mentioned above, there are other interesting augmentation methods, such as contrastive learning (Guo and Huang, 2025) or Fourier-based augmentation (Le et al., 2024). However, these augmentation methods are beyond the scope of the current study.

5. Proposed COD Method in Spatial-Frequency Domain

There are many strategies to solve the COD problem. One of the reasonable ideas is to develop strategies based on the natural camouflage functions of animals. We suppose that colour, texture and shape are more likely to deceive human vision, as are deep learning models built on the biological principals of the human brain, than general biological features of wildlife. In other words, the invisibility style of the camouflaged object plays a significant role, which is difficult to capture in the spatial domain. Our experiments show that the structural properties of camouflaged patterns are characterized by high discriminating ability to distinguish camouflaged objects from the background. One of the ways to extract the structural properties from an image is to move to the frequency domain and analyze high-frequency components.

The GAN-based architecture of the proposed COD network is presented in Figure 4. We offer architecture with two generators and three discriminators. The fusion of spatial and frequency masks allows to obtain a pseudo mask with subsequent visualization of the camouflaged object. Spatial generator analyzes semantic information, and frequency generator extracts structural information.

Since the COD task is very challenging, we decided to use one of the most effective segmentation models capable of capturing contextual information at multiple scales, i.e. DeepLabv3+ (Chen et al., 2018). The classical Encoder-Decoder architecture includes an encoding network to obtain rich semantic information and a decoding network to extract the detailed object boundaries. The outstanding segmentation capabilities of the DeepLabv3+ model are based on the use of depth-wise separable convolution and atrous spatial pyramid pooling, and continue to be improved with additional blocks. Note that to avoid analysis of channel components, the RGB image is first converted to a YUV colour space with the Y intensity component and U and V components that define the colour components. The frequency generator has a typical structure with, firstly, a two-level discrete wavelet transform of the gray image and, secondly, a CNN-based inverse wavelet transform architecture providing a so-called “frequency mask” with structured information.

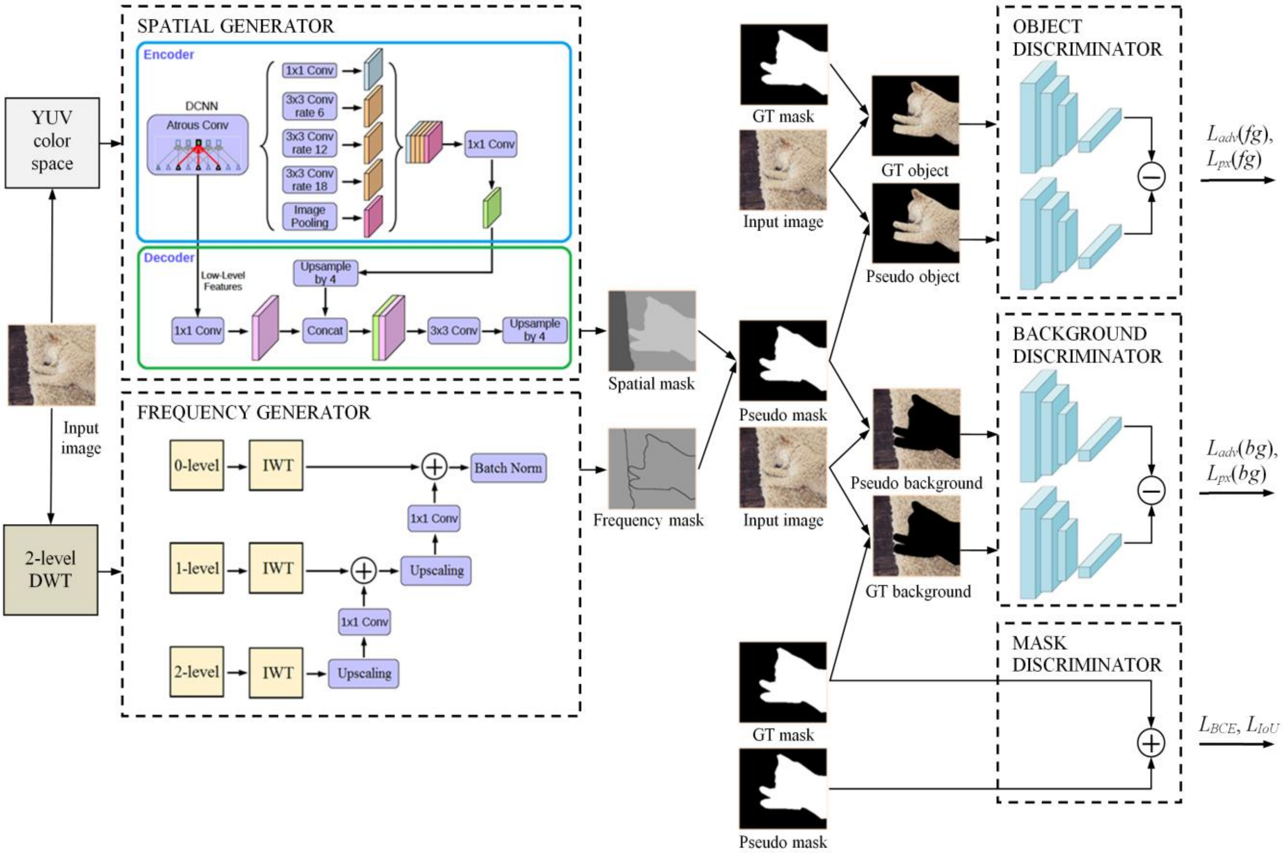

Figure 4. Architecture of the proposed SF-CODnet during training.

Let us consider the loss functions of the members of the proposed SF-CODnet. Adversarial loss L_{adv} can be reduced to relativistic loss according to (Naseer et al., 2020):

$$L_{adv} = -\log\left(\sigma\left(D_{fg}\left(G\left(I_{ps}\right)\right) - D_{fg}\left(I_{gt}\right)\right)\right), \quad (4)$$
$$-\log\left(\sigma\left(D_{bg}\left(G\left(I_{ps}\right)\right) - D_{bg}\left(I_{gt}\right)\right)\right)$$

where D_{fg}, D_{bg} = the object and background discriminators
G = the generator
I_{gt}, I_{ps} = the segmented ground truth image and segmented pseudo image
σ = the sigmoid layer

Object and background discriminators generate output pixel losses $L_{px}(fg)$ and $L_{px}(bg)$, respectively. Separate generation of output pixel losses is required since they have different degrees of influence on the total losses. Both discriminators have identical Siamese CNN-based architecture. The pixel losses $L_{px}(fg)$ and $L_{px}(bg)$ estimate the texture component based on the element-wise mean aggregation according to the following equation:

$$L_{px}\left(\cdot\right) = \frac{1}{n}\sum_{i}^{n}\left\|I_{gt(\cdot)} - \max\left(I_{gt(\cdot)}, I_{ps(\cdot)}\right)\right\|_1, \quad (5)$$

where $L_{px}(\cdot)$ = the pixel loss of foreground $L_{px}(fg)$ or pixel loss of background $L_{px}(bg)$
$I_{gt(\cdot)}$ = the segmented ground truth image of foreground or background
$I_{ps(\cdot)}$ = the segmented pseudo image of foreground or background
$\|\cdot\|_1$ = the L1-form
max(·) = the element-wise maximum selection
n = the number of pixels

Mask discriminator is trained to accurately segment the pseudo mask of the camouflaged object with summarized loss L_{pm} calculated as follows:

$$L_{pm} = L_{BCE}\left(D_m\left(M_{ps}\right), M_{gt}\right) + L_{IoU}\left(D_m\left(M_{ps}\right), M_{gt}\right), \quad (6)$$

where $L_{BCE}(\cdot)$ = the weighted binary cross-entropy loss (Jadon, 2020)
$L_{IoU}(\cdot)$ = the weighted intersection-over-union loss (Rahman and Wang, 2016)
D_m = the mask discriminator
M_{ps} = the pseudo mask
M_{gt} = the ground truth mask

The total loss function is a weighted sum of all sub-loss terms:

$$L_{total} = \alpha L_{adv} + \beta\left(L_{px}\left(fg\right), L_{px}\left(bg\right)\right) + \gamma L_{BCE} + \lambda I_{IoU}, \quad (7)$$

where α, β, γ, λ = the hyper-parameters tuned empirically

6. Experimental Results

Experiments were conducted to evaluate the performance of the proposed model on three benchmark datasets: CAMO, COD10K, and NC4K. These datasets were selected due to their diverse characteristics and their relevance to the task of camouflaged object detection (COD), which involves identifying objects that are partially or fully camouflaged within complex backgrounds. Below, we provide a brief description of each dataset used in our experiments:

1. The CAMO dataset (Le et al., 2019) contains more 1,250 images with camouflaged objects that are partially or fully masked by complex backgrounds, making detection particularly challenging. This dataset focuses on evaluating models in detecting objects under various levels of occlusion, noise, and background clutter, requiring advanced contextual understanding and multiscale detection abilities
2. The COD10K dataset (Fan et al., 2020) includes 10,000 images containing camouflaged objects, with paired “before” and “after” images from remote sensing and surveillance environments. The task involves detecting subtle changes in the scenes, where camouflaged objects may be revealed or altered. Ground truth annotations indicate the regions where changes occurred, challenging the model to detect camouflaged elements even in the presence of significant background variation
3. The NC4K dataset (Lv et al., 2021) consists of 4,121 images from natural environments, such as forests and wildlife habitats, containing camouflaged objects that blend seamlessly into their surroundings. These images are annotated with pixel-level segmentation masks highlighting the camouflaged objects

To enhance model robustness, we utilized a dual-sample training strategy, where the network was trained on both strong examples (original images) and weak examples (Gaussian blurred images). This approach allows the model to generalize better and improves its ability to detect camouflaged objects in varying conditions. The training process for SF-CODnet followed a structured pipeline to ensure robust camouflaged object detection.

First, data preprocessing was applied to all input images, resizing them to 513×513 pixels to match the input dimensions of DeepLabV3+ (ResNet-101 backbone). Standard normalization was performed using mean aligning the data distribution with ImageNet statistics. To improve the model's ability to detect camouflaged objects in challenging conditions, we incorporated weak sample generation by applying Gaussian blur to selected training images. This augmentation technique forces the model to focus on key structural features rather than fine-grained textures, improving generalization to low-contrast and low-texture scenarios. Feature maps from these modules are then combined and passed through the DeepLabV3+ classifier, which generates the final segmentation mask, accurately detecting camouflaged objects in both strong and weak examples. The network was trained using the Adam optimizer. Training was conducted for 60 epochs, alternating between strong and weak samples every batch to ensure balanced learning.

The model was evaluated on a separate validation set after every epoch. The best model was selected based on F-measure and Mean Absolute Error on the validation data. By training SF-CODnet with both high-quality (strong) and weak examples, the model learned to generalize across different levels of camouflage complexity, resulting in improved detection performance across various datasets.

To analyze the contribution of different components of SF-CODnet, we performed an ablation study by sequentially disabling key blocks and evaluating segmentation performance on strong and weak examples. We used for standard metrics:

1. Mean Absolute Error (M) measures the pixel-wise difference between the predicted and ground-truth masks (lower is better).
2. F-measure (F_β) assesses the balance between precision and recall (higher is better).
3. E-measure (E_ϕ) captures local and global similarity between prediction and ground truth (higher is better).
4. S-measure (S_α) evaluates the structural similarity of the segmentation (higher is better).

Experiments were conducted on both strong examples (source images) and weak examples (Gaussian blurred images). Table 1 and Table 2 show the results of the averaged values of segmentation metrics for strong and weak examples tested on different datasets.

Block Disabled	M (↓)	F_β (↑)	E_ϕ (↑)	S_α (↑)
None (Full Model)	0.065	0.572	0.752	0.788
Spatial Generator	0.065	0.000	0.741	0.789
Frequency Generator	0.066	0.000	0.741	0.792
Background Discriminator	0.065	0.183	0.742	0.789
Foreground Discriminator	0.074	0.000	0.741	0.791
Mask Discriminator	0.070	0.364	0.748	0.794

Table 1. Ablation study on SF-CODnet for strong examples

Block Disabled	M (↓)	F_β (↑)	E_ϕ (↑)	S_α (↑)
None (Full Model)	0.068	0.781	0.795	0.813
Spatial Generator	0.071	0.000	0.780	0.810
Frequency Generator	0.071	0.000	0.780	0.810
Background Discriminator	0.070	0.251	0.783	0.811
Foreground Discriminator	0.070	0.000	0.780	0.810
Mask Discriminator	0.069	0.428	0.790	0.812

Table 2. Ablation study on SF-CODnet for weak examples

Training on weak examples with stronger Gaussian blur significantly enhances segmentation performance. The F-measure (F_β) is notably higher for blurred images, indicating a better balance between precision and recall. Additionally, both the E-measure (E_ϕ) and S-measure (S_α) improve, confirming that the model retains object structure even under extreme blurring. The Mean Absolute Error (M) decreases, demonstrating that the model makes fewer pixel-wise mistakes when trained with weak examples. Figure 5 and Figure 6 show the results of ablation for the strong and weak examples respectively with sequential disconnection of the model blocks. Disabling Spatial Generator and Frequency Generator results in no segmentation being performed and the image matches the original image. This proves that the value of the metric F_β is zero.

As can be seen from the Figures 5 and 6, the image processed with Gaussian blur demonstrates better object boundary detection, as it effectively reduces background noise and prevents non-relevant elements, such as leaves, from being included in the contours. The Spatial and Frequency Generators play a crucial role in camouflaged object detection. Disabling either module results in F_β dropping to 0, meaning the model completely fails to segment objects. This highlights the importance of spatial and frequency-based features in detecting camouflaged objects. The Background Discriminator enhances object-background distinction, though its impact is less critical. When disabled, F_β and E_ϕ decrease, but the model is still able to segment objects. This suggests that while background differentiation improves segmentation, it is not the most crucial factor. Finally, the Mask Discriminator contributes to refining object boundaries. Disabling it leads to a decrease in structural similarity (S_α) and a slight reduction in F_β, indicating that this module helps define object edges more clearly, improving the overall segmentation quality.

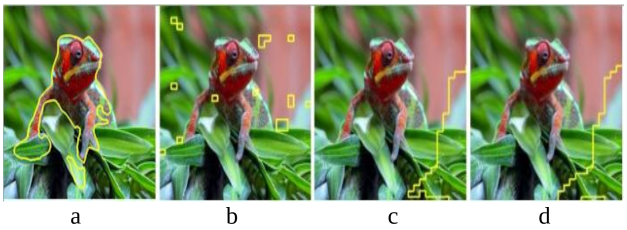

Figure 5. Contour visualization of generated samples from ablation study on strong example: **a** full model, **b**, **c**, **d** generated samples without background, mask and object discriminators, respectively.

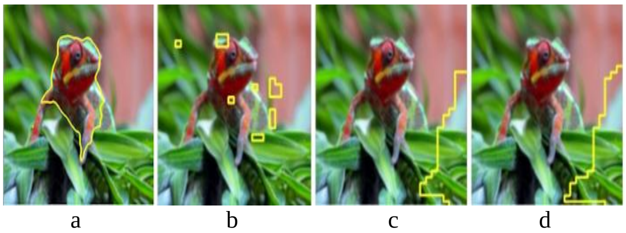

Figure 6. Contour visualization of generated samples from ablation study on weak example: **a** full model, **b**, **c**, **d** generated samples without background, mask and object discriminators respectively.

To further evaluate our approach, we compare SF-CODnet with state-of-the-art camouflaged object detection models, including SINet (Fan et al., 2022), MirrorNet (Yan et., 2021), and BasNet (Qin et al., 2019). The results are summarized in Table 3 (Camouflaged Object Segmentation on CAMO Open Source, 2025).

Model	M (↓)	F_β (↑)	E_ϕ (↑)	S_α (↑)
SF-CODnet (Ours)	0.068	0.781	0.795	0.813
SINet	0.082	0.750	0.606	0.751
MirrorNet	0.084	0.780	0.719	0.785
BasNet	0.056	0.618	0.413	0.618

Table 3. Ablation Study on SF-CODnet for weak examples

Figure 7 illustrates a qualitative comparison of these existing models with our proposed SF-CODnet. Unlike these models, SF-CODNet is trained using both strong and weak examples, which allows it to better generalize and improves its ability to preserve object boundaries even in low-contrast environments.

The SF-CODnet model presents several advantages and disadvantages in camouflaged object detection. Advantages include its improved performance across complex datasets, such as CAMO, COD10K, and NC4K, which contain images with varying levels of difficulty and noise. This demonstrates the model's versatility and ability to handle diverse camouflaged objects. A dual learning strategy that uses both strong and weak examples (with Gaussian blur) improves the model's generalization ability, allowing it to detect objects in low-contrast and low-texture conditions. Additionally, the model's performance is strongly influenced by the Spatial Generator and Frequency Generator, which are crucial for detecting camouflaged objects. If either of these components is disabled, the model fails completely, emphasizing their importance. The model also achieves high segmentation accuracy, with good precision-recall balance (F_β), structural similarity (S_α), and local-global similarity (E_ϕ) for both strong and weak examples. Moreover, when compared to other state-of-the-art models such as SINet, MirrorNet, and BasNet, SF-CODnet outperforms them in terms of both accuracy and error metrics, confirming its superiority in camouflaged object detection.

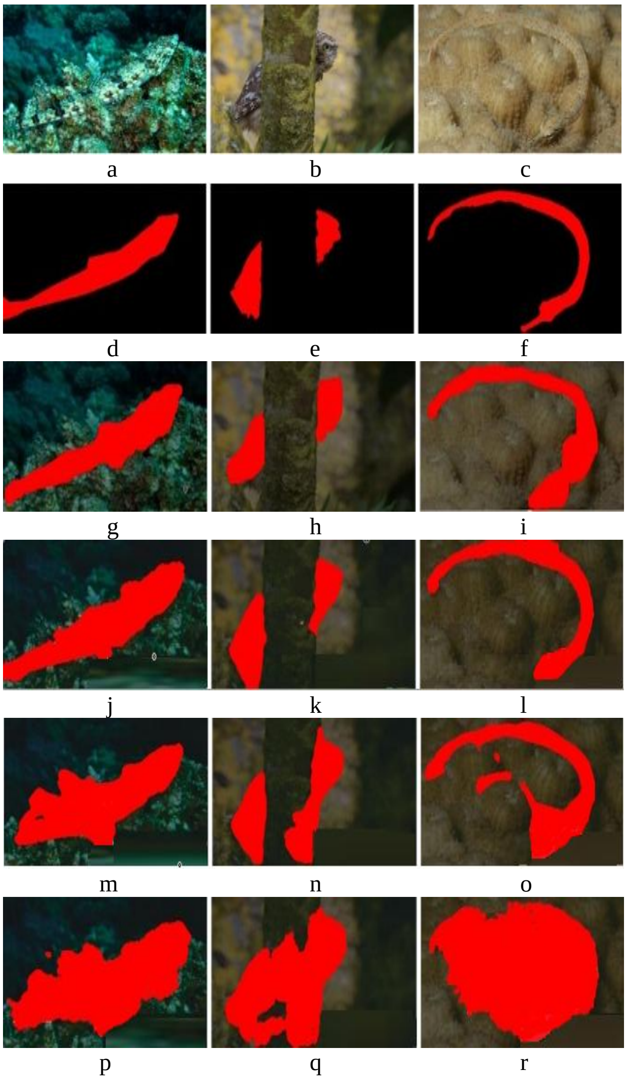

Figure 7. Comparison of detection models of masked objects in identical images: **a, b, c** original images, **d**, **e**, **f** ground truth masks, **g**, **h**, **i** SF-CODNet (Ours), **j, k, l** MirrorNet, **m, n, o** BasNet, **p**, **q**, **r** SINet.

Disadvantages, however, include its dependence on high-quality data. The model's performance may degrade when working with lower-quality or noisy images, requiring improvements in data pre-processing or training techniques. Furthermore, the model's computational complexity is high due to the use of multiple components, such as the Spatial and Frequency Generators, and the need to train on both strong and weak examples, resulting in increased computational demands and longer training times. While the Background Discriminator improves object-background differentiation, its impact is less critical, suggesting that the model could be simplified without significantly affecting performance, which could reduce computational costs. The model also poses challenges in interpretability, as it relies on complex components, making it difficult to understand which specific features are being leveraged for object detection. Lastly, achieving optimal results requires careful tuning of hyperparameters, such as the degree of Gaussian blur in weak examples, which may require additional effort during the training process.

7. Conclusions

In this study, we introduced SF-CODnet, a novel approach for camouflaged object detection that leverages spatial-frequency features and weak sample learning to improve segmentation performance. The model was trained on both strong examples (original images) and weak examples (Gaussian blurred images), allowing it to generalize better and detect camouflaged objects in diverse conditions. Compared to state-of-the-art camouflaged object detection models (SINet-, MirrorNet, BasNet), SF-CODnet achieved superior performance, particularly in F_β and E_ϕ, demonstrating better balance between precision and recall. Moreover, the model exhibited a lower mean absolute error (M), indicating improved pixel-wise segmentation accuracy. However, structural similarity (S_α) remains an area for potential refinement, and the model’s computational complexity could be optimized for real-time applications.

Overall, SF-CODnet provides a significant advancement in camouflaged object detection, proving that weak sample training with strong Gaussian blur is an effective strategy for improving segmentation robustness. Future work will focus on further enhancing structural refinement, reducing computational cost, and exploring additional augmentation techniques to improve model generalization across complex environments.

References

Camouflaged Object Segmentation on CAMO Open Source. https://paperswithcode.com/sota/camouflaged-object-segmentation-on-camo (5 February 2025).

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018: Encoder-decoder with atrous separable convolution for semantic image segmentation. *In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018.* Springer, Cham. LNCS, vol. 11211, pp. 833-851.

Cong, R., Sun, M., Zhang, S., Zhou, X., Zhang, W., Zhao, Y., 2023: Frequency perception network for camouflaged object detection. *The 31st ACM International Conference on Multimedia (ACM MM)*, Ottawa, ON, Canada, pp. 1179-1189.

Fang, X., Chen, J., Wang, Y., Jiang, M., Ma, J., Wang, X., 2025: EPFDNet: Camouflaged object detection with edge perception in frequency domain. *Image and Vision Computing*, 154, 105358.1-105358.10.

Fan, D.-P., Ji, G.-P., Sun, G., Cheng, M.-M., Shen, J., Shao, L., 2020: Camouflaged object detection. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Seattle, WA, USA, pp. 2777-2787.

Fan, D.-P., Ji, G.-P., Cheng, M.-M., Shao, L., 2022: Concealed Object Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44(10), 6024-6042.

Fu, Y., Ying, J., Lv, H., Xiaojie Guo, X., 2024: Semi-supervised camouflaged object detection from noisy data. *The 32nd ACM International Conference on Multimedia (MM ’24)*, Melbourne, VIC, Australia, pp. 4766-4775.

Guan, J., Qian, W., Zhu, T., Fang, X., 2025: Promoting camouflaged object detection through novel edge–target interaction and frequency-spatial fusion. *Neurocomputing* 617, 129064.1-129064.12.

Guo, C., Huang, H., 2025: Enhancing camouflaged object detection through contrastive learning and data augmentation techniques. *Engineering Applications of Artificial Intelligence* 141, 109703.1-109703.11.

He, C., Li, K., Zhang, Y., Tang, L., Zhang, Y., Guo, Z., Li, X., 2023: Camouflaged object detection with feature decomposition and edge reconstruction. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Vancouver, BC, Canada, pp. 22046-22055.

He, C., Li, K., Zhang, Y., Zhang, Y., You, C., Guo, Z., Li, X., Danelljan, M., Yu, F. 2024a: Strategic preys make acute predators: Enhancing camouflaged object detectors by generating camouflaged objects. *The Twelfth International Conference on Learning Representations (ICLR 2024)*, Vienna, Austria, pp.1-19.

He, C., Li, K., Zhang, Y., Xu, G., Tang, L., Zhang, Y., Guo, Z., Li, X.: 2024b. Weakly-supervised concealed object segmentation with SAM-based pseudo labeling and multi-scale feature grouping. *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*, pp. 1-12.

He, R., Dong, Q., Lin, J., Lau, R.W.H., 2023: Weakly-supervised camouflaged object detection with scribble annotations. *The Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-23)*, Washington, DC, USA, pp. 781-789.

Hess, A.S., Wismer, A.J., Bohil, C.J., Neider, M.B. 2016: On the hunt: Searching for poorly defined camouflaged targets. *PLoS One* 11(3), e0152502.1-e0152502.18.

Hu, J., Lin, J., Gong, S., Cai, W., 2024: Relax image-specific prompt requirement in SAM: A single generic prompt for segmenting camouflaged objects. *The AAAI Conference on Artificial Intelligence (AAAI-24)*, vol. 38, no. 11, pp. 12511-12518.

Jadon, S., 2020: A survey of loss functions for semantic segmentation. *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, IEEE, Via del Mar, Chile, pp. 1-7.

Lai, X., Yang, Z., Hu, J., Zhang, S., Cao, L., Jiang, G., Wang, Z., Zhang, S., Ji, R., 2025: CamoTeacher: Dual-rotation consistency learning for semi-supervised camouflaged object detection. *In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (eds) Computer Vision – ECCV 2024. ECCV 2024*, LNCS, vol, 15103. Springer, Cham, pp. 438-455.

Le, T.-N., Nguyen, T.V., Nie, Z., Tran, M.-T., Sugimoto, A., 2019: Anabranch network for camouflaged object segmentation. *Computer Vision and Image Understanding*, 184, 45-56.

Le, M.-Q., Tran, M.-T., Le, T.-N., Nguyen, T.V., Do, T.T., 2024: CamoFA: A learnable Fourier-based augmentation for camouflage segmentation. *arXiv:2308.15660v2*, 1-10.

Liang, Y., Qin, G., Sun, M., Wang, X., Yan, J., Zhang, Z. 2024: A systematic review of image-level camouflaged object detection with deep learning. *Neurocomputing* 566, 127050.1-127050.23.

Li, A., Zhang, J., Lv, Y., Liu, B., Zhang, T., Dai, Y., 2021: Uncertainty-aware joint salient object and camouflaged object detection. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Nashville, TN, USA, pp. 10071-10081.

Lv, Y., Zhang, J., Dai, Y., Li, A., Liu, B., Barnes, N., Fan, D.-P., 2021: Simultaneously localize, segment and rank the camouflaged objects. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Nashville, TN, USA, pp. 11591-11601.

Naseer, M., Khan, S., Hayat, M., Khan, F.S., Porikli, F., 2020: A self-supervised approach for adversarial robustness. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Seattle, WA, USA, pp. 259-268.

Neider, M., Zelinsky, G. 2006: Searching for camouflaged targets: Effects of target-background similarity on visual search. *Vision Research* 46(14), 2217-2235.

Rahman, M.A., Wang, Y., 2016: Optimizing intersection-over-union in deep neural networks for image segmentation. In: Bebis, G., et al. *Advances in Visual Computing. ISVC 2016*. LNCS, vol, 10072. Springer, Cham, pp. 234-244.

Qin, X., Zhang, Z., Huang, C., Gao, C., Dehghan, M., Jagersand, M., 2019: BASNet: Boundary-aware salient object detection. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Long Beach, CA, USA, pp. 7479-7489.

Xiao, F., Hu, S., Shen, Y., Fang, C., Huang, J., He, C., Tang, L., Yang, Z., Li, X. 2024: A survey of camouflaged object detection and beyond. *CAAI Artificial Intelligence Research* 3, 9150044.1-9150044.26.

Yan, J., Le, T.-N., Nguyen, K.-D., Tran, M.-T., Do, T.-T., Nguyen, T.V., 2021: MirrorNet: Bio-inspired camouflaged object segmentation. *IEEE Access* 9, 43290-43300.

Zhang, Y., Wu, C., 2023: Unsupervised camouflaged object segmentation as domain adaptation. *2023 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, IEEE, Paris, France, pp. 4334-4344, 2023.

Zhong, Y., Li, B., Tang, L., Kuang, S., Wu, S., Ding, S., 2022: Detecting camouflaged object in frequency domain. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, New Orleans, LA, USA, pp. 4504-4513.