Camouflaged Object Detection with Feature Decomposition and Edge Reconstruction

Chunming He[1] , Kai Li[2*], Yachao Zhang[1] , Longxiang Tang[1] , Yulun Zhang[3] , Zhenhua Guo[4] , Xiu Li[1*]
[1]Shenzhen International Graduate School, Tsinghua University,
[2]NEC Laboratories America, [3]ETH Zürich, [4]Tianyi Traffic Technology
{chunminghe19990224,li.gml.kai,lloong.x,yulun100,cszguo}@gmail.com
yachaozhang@stu.xmu.edu.cn, li.xiu@sz.tsinghua.edu.cn

Abstract

Camouflaged object detection (COD) aims to address the tough issue of identifying camouflaged objects visually blended into the surrounding backgrounds. COD is a challenging task due to the intrinsic similarity of camouflaged objects with the background, as well as their ambiguous boundaries. Existing approaches to this problem have developed various techniques to mimic the human visual system. Albeit effective in many cases, these methods still struggle when camouflaged objects are so deceptive to the vision system. In this paper, we propose the FEature Decomposition and Edge Reconstruction (FEDER) model for COD. The FEDER model addresses the intrinsic similarity of foreground and background by decomposing the features into different frequency bands using learnable wavelets. It then focuses on the most informative bands to mine subtle cues that differentiate foreground and background. To achieve this, a frequency attention module and a guidance-based feature aggregation module are developed. To combat the ambiguous boundary problem, we propose to learn an auxiliary edge reconstruction task alongside the COD task. We design an ordinary differential equation-inspired edge reconstruction module that generates exact edges. By learning the auxiliary task in conjunction with the COD task, the FEDER model can generate precise prediction maps with accurate object boundaries. Experiments show that our FEDER model significantly outperforms state-of-the-art methods with cheaper computational and memory costs. The code will be available at https://github.com/ChunmingHe/FEDER.

1. Introduction

Camouflaged object detection (COD) aims to detect and segment objects "seamlessly" integrated into surrounding environments. COD is a challenging task as it needs to combat against excellent camouflage strategies, including background matching [41], disruptive coloration [33], etc., and distinguish the subtle differences between candidate objects and their backgrounds. Research in COD can simultaneously facilitate the development of visual perception for nuance discrimination and promote various valuable real-life applications, ranging from concealed defect detection [17] in industry to pest monitoring [35] in agriculture.

COD faces two main challenges. The first is the intrinsic similarity (IS) challenge, which occurs when camouflaged objects share similar colors and patterns with their backgrounds. This makes it difficult to even roughly localize those camouflaged objects. The second is the edge disruption (ED) challenge, which arises from extremely ambiguous object boundaries. Even if a rough localization is achieved, precise segmentation can barely be obtained.

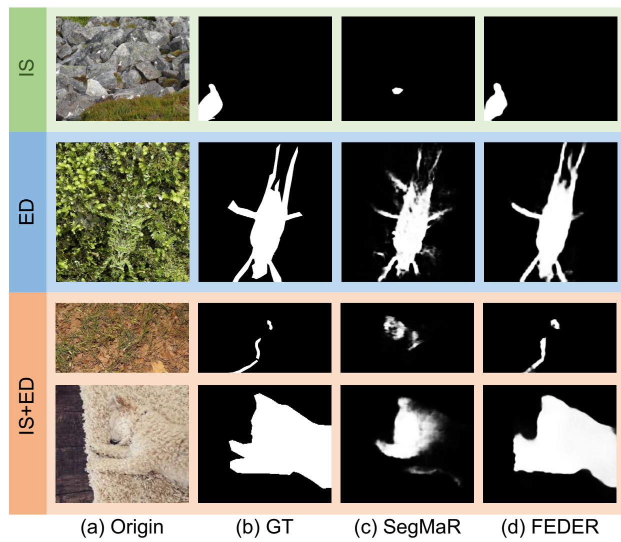

Figure 1. Results of SegMaR [14] and our method under the intrinsic similarity (IS) and edge disruption (ED) challenges. Our method better localizes the objects and produces clearer edges.

*Corresponding author.

To tackle these challenges, most existing works aim to develop models that mimic the human visual system [6,28]. However, since camouflage strategies are designed by prey to confuse the predator's visual system, and the intrinsic topological properties of candidate objects are not distinctive, such human perception-oriented attempts may struggle to identify subtle discriminative features and fail to effectively address the above challenges. For instance, as illustrated in Fig. 1, the state-of-the-art human perception-based COD method can only generate inaccurate prediction maps, such as the vague caddisfly and incomplete dog (Row 2 and 4), or even fail to detect camouflaged objects like the bird and snake (Row 1 and 3). Therefore, a better COD method should compensate for the "flaw" in human perception by emphasizing subtle discriminative features.

Based on the biological study [41], camouflaged objects often employ various camouflage strategies to conceal their discriminative differences, which mainly exist in texture details and global information distribution, within surrounding environments. Such a study inspires us to cope with the COD task by decomposing the camouflage scenario into different parts. This allows for the disentanglement of various intricate connections, enabling each part to be separately handled to fully excavate subtle discriminative cues.

With this inspiration, we propose the FEature Decomposition and Edge Reconstruction (FEDER) model for the COD task, which compensates for the deficiencies of human perception by emphasizing subtle discriminative features and effectively addresses the IS and ED challenges. Specifically, to combat the intractable localization problem caused by the IS challenge, we design the deep wavelet-like decomposition (DWD) strategy, which decomposes the extracted features into different frequency bands using learnable wavelet-like modules. Then, we focus on the most informative bands by filtering out noteworthy parts where discriminative cues are most likely to exist by a novel frequency attention (FA) module. Moreover, a guidance-based feature aggregation (GFA) module is proposed to aggregate the multi-scale decomposed features with attentional guidance to further emphasize discriminative information.

To address the ambiguous boundary problem of the ED challenge, we propose learning an auxiliary edge reconstruction task to encourage the network to excavate edge details. We design the ordinary differential equation (ODE)-inspired edge reconstruction (OER) module to reconstruct accurate and complete edge prediction maps using a high-order ODE solver, specifically, the second-order Runge-Kutta. Incorporating this auxiliary task with the COD task can facilitate the generation of precise segmentation results with accurate object boundaries.

Our contributions are summarized as follows:

- We propose the FEature Decomposition and Edge Reconstruction (FEDER) model for the COD task. To the best of our knowledge, we are the first to approach COD from a decomposition perspective.
- To highlight the subtle discriminative features, we propose frequency attention modules to filter out the noteworthy parts of corresponding features and design the Guidance-based Feature Aggregation module to aggregate the multi-scale features with attentional guidance.
- We propose to learn an auxiliary edge reconstruction task along with the COD task to help generate precise segmentation maps with accurate object boundaries and design the ODE-inspired edge reconstruction module for complete edge prediction.
- The proposed FEDER significantly outperforms the state-of-the-art methods on four datasets by a large margin with cheaper computational and memory costs.

2. Related Works

Camouflaged object detection. Unlike existing object detection tasks, camouflaged object detection (COD) poses new challenges for mining subtle discriminative features under complex camouflage strategies [6, 11]. Early techniques utilized the hand-crafted operators for COD [11,30], which were only applicable to camouflaged scenarios with simple backgrounds. Recent research has leveraged the huge capacity of deep learning to detect camouflaged objects in a learning manner [6,14,28]. Inspired by the hunting process of predators, SINet [6] designed a bio-inspired network to gradually search and locate the camouflaged object. PFNet [28] proposed the position module and focus module to imitate human identification with the distraction mining strategy. By simulating human behaviors in understanding complex scenarios, SegMaR [14] integrated segment, magnify and reiterate in a coarse-to-fine manner using the multi-stage strategy. However, these COD solutions mainly focus on mimicking biovision systems, which can be easily confused by complex camouflaged strategies and struggle to excavate the subtle discriminative features, thus failing to handle the IS and ED challenges (see Fig. 1). Unlike these human perception-oriented techniques, we first propose to address the COD task from a decomposition perspective by decomposing the extracted features into different frequency bands with learnable wavelets and filtering out the most informative bands to excavate those inconspicuous discriminative features, thus remedying the human visual deficiency and solving the IS challenge. To handle the ED challenge, we propose learning an auxiliary edge reconstruction task along with the COD task to facilitate the generation of precise segmentation results with clear object boundaries.

Deep wavelet decomposition. Deep wavelet decomposition is an effective tool to decompose image/feature into various frequency components and has gained immense popularity in many domains, such as image restoration [15]

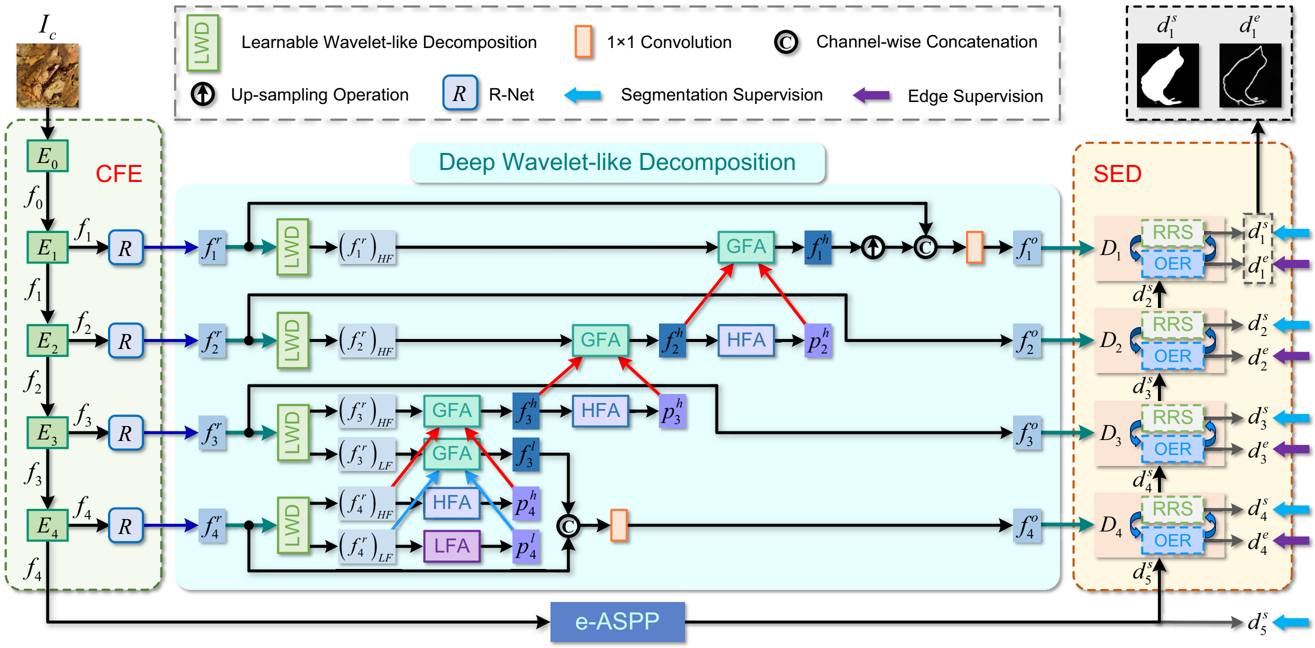

Figure 2. Overview of the proposed FEDER model. CFE and SED are short for camouflaged feature encoder and segmentation-oriented edge-assisted decoder. HFA/LFA and GFA denote high/low frequency attention modules and the guidance-based feature aggregation module. RRS and OER indicate the reversible re-calibration segmentation module and the ODE-inspired edge reconstruction module.

and style transfer [49]. To handle the IS challenge, we introduce deep wavelet decomposition into the COD task. Furthermore, to better accommodate the COD data, we employ the learnable wavelets for deep adaptive feature decomposition, whose coefficients are updated following AWD [8].

ODE-inspired network. Researchers have established a relationship between ODE and neural networks. [46] first analyzed ResNet from the perspective of discrete ODE and [1] further extended ResNet to an ODE-inspired network architecture with a more accurate transmission. Since then, ODE-inspired networks are widely utilized in many fields, such as image dehazing [37] and machine translation [19]. In this paper, to accommodate the fine-grained property of the edge, we propose an ODE-inspired edge reconstruction module with the second-order Runge-Kutta and a weighted gate mechanism, aiming to generate more accurate boundaries. Furthermore, we apply the Hamiltonian system to our OER module to ensure the stability of edge reconstruction.

3. Methodology

Given a camouflaged image, we first extract a cascade of features using the camouflaged feature encoder (CFE). We then perform a wavelet-like decomposition (DWD) on the features to decompose them into different frequency bands. We select the most informative bands, such as the high-frequency and low-frequency components, for further analysis. These informative bands are processed by the frequency attention (FA) module and guidance-based feature aggregation (GFA) module to highlight the inconspicuous discriminative features. With the aggregated features, the segmentation-oriented edge-assisted decoder (SED) outputs both the segmentation map and the edge prediction map. Fig. 2 presents the framework of our FEDER model.

3.1. Camouflaged Feature Encoder (CFE)

Following SINet V2 [4], the basic encoder E adopts ResNet50 [10]/Res2Net50 [7] as its backbone. Given an image I_c of size $W \times H$, the basic encoder E generates a set of feature maps $\{f_k\}_{k=0}^{4}$ with the resolution of $\frac{H}{2^{k+1}} \times \frac{W}{2^{k+1}}$, R-Net [6] is cascaded to transform $\{f_k\}_{k=1}^{4}$ into a more informative and compact output, i.e., a series of 64-channel feature maps $\{f_k^r\}_{k=1}^{4}$. Additionally, the last feature map f_4 from the basic encoder E is further fed into an efficient atrous spatial pyramid pooling (e-ASPP) A_e [16] to enlarge the receptive field and fuse the multi-context information, resulting in $d_5^s = A_e(f_4)$, where d_5^s is a coarse segmentation result with the same spatial resolution as f_4.

3.2. Deep Wavelet-like Decomposition

3.2.1 Learnable Wavelet-like Decomposition

Camouflaged objects share a high intrinsic similarity with the background, which poses challenges for common feature extractors to mine the inconspicuous discriminative features, ultimately resulting in the suppression of segmentation performance. Based on the biological study [41], the discriminative features of COD mainly exist in the high-

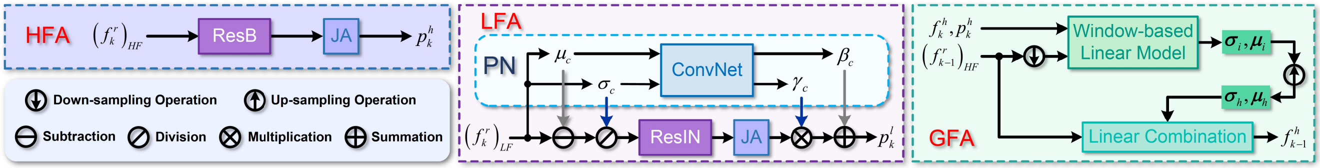

Figure 3. Details of the proposed GFA, HFA, and LFA

frequency (HF) component, e.g., texture and edge, and the low-frequency (LF) component, e.g., color and illumination. Inspired by this study, we propose to perform deep wavelet-like decomposition (DWD) on the extracted features $\{f_k^r\}_{k=1}^4$ and select the most informative HF and LF components for further refinement. We decompose f_k^r as

$$(f_k^r)_{HF} = W_{HF}(f_k^r), \; (f_k^r)_{LF} = W_{LF}(f_k^r), \quad (1)$$

where $(f_k^r)_{HF}$ and $(f_k^r)_{LF}$ denote the HF and LF components of f_k^r. W_{HF} and W_{LF} represent the learnable HF and LF filters with the coefficients updated following AWD [8] and initialized by Haar wavelet [40]. The learned wavelet-like transformer is expected to better cater to COD data than manually-designed wavelets [36, 42], thus further facilitating the extraction of inconspicuous discriminative features.

3.2.2 Frequency Attention Modules

To extract discriminative information from the decomposed features, we propose a high-frequency attention (HFA) module and a low-frequency attention (LFA) module, corresponding to HF and LF bands, respectively. The detailed structures of the two modules are illustrated in Fig. 3.

High-frequency attention module. We design the HFA module to accentuate those texture-rich regions for subtle discriminative feature extraction. Following [21, 26], we first apply a residual block for texture preservation, consisting of a 3×3 convolution layer, batch normalization (BN) [12], and ReLU. We then employ the joint attention module $JA(\cdot)$, which includes spatial attention [34] and channel attention [45], to highlight noteworthy parts in both spatial and channel domains. Therefore, given HF features $(f_k^r)_{HF}$, the HF attention map p_k^h is formulated as follows:

$$p_k^h = JA(ResB((f_k^r)_{HF})), \quad (2)$$

where $ResB(\cdot)$ denotes the residual block with BN.

Low-frequency attention module. Low-frequency components focus more on global information, such as color distribution and illumination, which inevitably leads to inevitably existing redundant components and slight perturbations [43]. To handle those problems, we design a comprehensive normalization strategy to suppress the undesired artifacts and provide cleaner global information for attention calculation both at the instance level and channel dimension, which can highlight those abnormal regions from a global perspective. Specifically, this module takes the decomposed LF features $(f_k^r)_{LF}$ as the input and outputs

$$p_k^l = JA(PN(ResIN((f_k^r)_{LF}))), \quad (3)$$

where $ResIN(\cdot)$, $PN(\cdot)$, and $JA(\cdot)$ denote the instance normalization [44] constrained residual block, positional normalization [20], and joint attention, respectively.

3.2.3 Guidance-based Feature Aggregation Module

As shown in Fig. 3, we propose a guidance-based feature aggregation (GFA) module to integrate the multi-scale decomposed features. Unlike existing heuristic-based feature aggregation strategies simply using concatenation [22, 24], GFA is specifically designed to address the key issue of COD, i.e., emphasizing the subtle discriminative features, by promoting inter-feature information interaction.

Taking HF bands as an example, GFA generates the aggregated feature $\{f_{k-1}^h\}_{k=2}^4$ that combines deep semantic information of the low-resolution feature $(f_k^r)_{HF}$ (at a higher level) and the abundant spatial details of the high-resolution feature $(f_{k-1}^r)_{HF}$ (at a lower level) with the guidance of the attention map p_k^h. Therefore, the aggregated feature f_{k-1}^h can better highlight the subtle discriminative features. To extract the attention-guided semantic information, we first generate the down-sampled aggregated feature f_{k-1}^{dh} with the window-based linear model [23]:

$$\left(f_{k-1}^{dh}\right)_i = \sigma_w down\left(\left(f_{k-1}^r\right)_{HF}\right)_i + \mu_w, \forall i \in s_w, \quad (4)$$

where $down(\cdot)$, s_w, and i are the down-sampling operation, local window, and pixel point i. $\{\sigma_w, \mu_w\}$ are linear aggregation coefficients for the pixels in window s_w, which can be acquired by optimizing the following objective function:

$$\min_{\sigma_w, \mu_w} \sum_{i \in s_w} \left[\left(p_k^h\right)_i^2 \left(\left(f_{k-1}^{dh}\right)_i - \left(\left(f_k^r\right)_{HF}\right)_i \right)^2 + \epsilon \sigma_w^2 \right], \quad (5)$$

where ϵ is a constraint value for σ_w. See Supplementary Material (**Supp**) for derivations and solutions of $\{\sigma_w, \mu_w\}$.

Considering pixel i covered by multiple windows, we average those window-wise coefficients and get the specific aggregation coefficients $\{\sigma_i, \mu_i\}$ for pixel i. By matrixing $\{\sigma_i, \mu_i\}$ into $\{\boldsymbol{\sigma}_i, \boldsymbol{\mu}_i\}$, Eq. (4) can be rewritten as follows:

$$f_{k-1}^{dh} = \boldsymbol{\sigma}_i \odot down((f_{k-1}^r)_{HF}) + \boldsymbol{\mu}_i, \quad (6)$$

where $\odot$ is the Hadamard product. We then up-sample $\{\boldsymbol{\sigma}_i, \boldsymbol{\mu}_i\}$ as $\{\boldsymbol{\sigma}_h, \boldsymbol{\mu}_h\}$ and acquire the high-resolution aggregated feature f_{k-1}^h for enriching spatial details:

$$\begin{aligned} f_{k-1}^h &= GFA\left((f_k^r)_{HF}, (f_{k-1}^r)_{HF}, p_k^h\right), \\ &= \boldsymbol{\sigma}_h \odot \left(f_{k-1}^r\right)_{HF} + \boldsymbol{\mu}_h. \end{aligned} \quad (7)$$

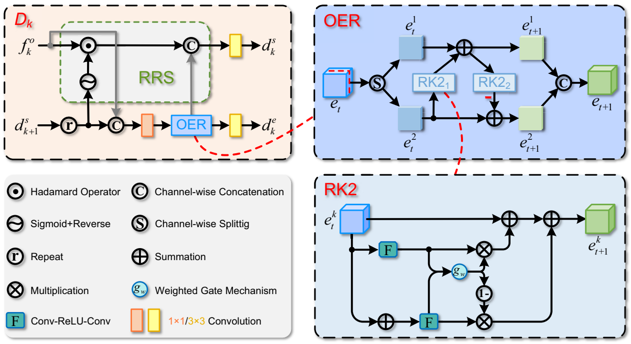

Figure 4. Details of SED with RRS and OER modules. OER is an RK2-based module guaranteed with the Hamiltonian system.

To iteratively acquire the aggregated features $\{f_{k-1}^h\}_{k=2}^4$, we redefine GFA module by replacing $(f_k^r)_{HF}$ with f_k^h:

$$f_{k-1}^h = GFA\left(f_k^h, \left(f_{k-1}^r\right)_{HF}, p_k^h\right), \tag{8}$$

where $p_k^h = JA(ResB(f_k^h))$ and $f_4^h = (f_4^r)_{HF}$. Guaranteed with frequency-specific attention, our aggregated features can emphasize more discriminative features than others by combining abundant spatial details and deep semantic information, thus better catering to the COD task. The calculation of the aggregated LF features f_{k-1}^l are similar to f_{k-1}^h, which can be seen in **Supp**.

Considering that bottom layers (at higher levels) focus more on HF details while top layers (at lower levels) care more about global information [38], we pass the aggregated HF/LF features into the bottom/top decoder layers along with the skip-connected encoded features $\{f_k^r\}_{k=1}^4$. To balance performance and efficiency, the integrated features $\{f_k^o\}_{k=1}^4$ passed to the decoder are defined as:

$$\begin{aligned} &f_1^o = conv1\left(con\left(f_1^r, up\left(f_1^h\right)\right)\right), f_2^o = f_2^r, \\ &f_3^o = f_3^r, f_4^o = conv1\left(con\left(f_4^r, f_3^l\right)\right), \end{aligned} \tag{9}$$

where $up(\cdot)$ and $con(\cdot)$ denote the up-sampling operation and the concatenation operation. $conv1$ represents 1×1 convolution, which is used for channel-level integration.

3.3. Segmentation-oriented Edge-assisted Decoder

In the segmentation-oriented edge-assisted decoder (SED) $\{D_k\}_{k=1}^4$, we propose to learn an auxiliary edge reconstruction task alongside the COD task to help generate precise segmentation maps with accurate boundaries. To be specific, as shown in Fig. 4, each decoder layer D_k consists of a reversible re-calibration segmentation (RRS) module and an ODE-inspired edge reconstruction (OER) module.

3.3.1 Reversible Re-calibration Segmentation Module

Due to complex camouflage, prediction maps inevitably have some ambiguous regions with low confidence, we adopt a reverse strategy to excavate cues from these low-confidence regions by reversing the attention, which erases detected regions and amplifies the response in low-confidence regions, thus re-calibrating the misclassified regions. Specifically, we repeat the coarse segmentation map $\{d_{k+1}^s\}_{k=0}^3$ as a 64-dimension tensor, normalize it to $[0, 1]$ with Sigmoid $S(\cdot)$, and reverse it by subtracting each element from 1. We then multiply the integrated feature f_k^o with the reversed map and concatenate it with the edge feature d_k^{fe} to obtain the segmentation result d_k^s:

$$d_k^s = conv3\left(con\left(d_k^{fe}, f_k^o \odot rv\left(S(rp(d_{k+1}^s))\right)\right)\right), \tag{10}$$

where $rp(\cdot)$ and $rv(\cdot)$ denote repeat and reverse.

3.3.2 ODE-inspired Edge Reconstruction Module

Existing methods tend to excavate edge information by incorporating certain priors within the residual network structure [13, 50]. However, either the intractable localization problem of the IS challenge or the ambiguous boundary problem of the ED challenge makes it difficult to design an appropriate edge prior for the COD task. In some cases, a biased prior can even reduce the segmentation performance.

Therefore, instead of exploiting prior knowledge, we focus on proposing an edge-friendly network architecture, i.e., the ODE-inspired edge reconstruction (OER) module. Compared with the traditional residual network structure that can be seen as the first-order Euler discretization approximation of ODE with nonnegligible truncation errors [46], the proposed OER module employs a higher-order ODE solver, specifically, a second-order Runge-Kutta (RK2), to provide more accurate numerical solutions in edge information processing. This better accommodates the fine-grained property of edges and facilitates the complete edge reconstruction, thus addressing the ED challenge. To ensure the flexibility of our OER module, we replace the fixed trade-off parameter in the RK2 solver with a weighted gate mechanism g_w with learnable coefficients. Given an input e_t, where $e_t = conv1\left(con\left(f_k^o, rp\left(d_{k+1}^s\right)\right)\right)$, the proposed OER module can be formulated as follows:

$$\begin{aligned} &e_{t+1} = e_t + g_w F_1 + (1 - g_w) F_2, \\ &g_w = S\left(\sigma_g con\left(F1, F2\right) + \mu_g\right), \\ &F_1 = F\left(e_t, \theta_t\right), F_2 = F\left(e_t + F_1, \theta_t\right), \end{aligned} \tag{11}$$

where $e_{t+1} = d_k^{fe}$, σ_g and μ_g are the learnable parameters in g_w. $\{F_i\}_{i=1}^2$ denotes the intermediate layers with shared parameters θ_t for efficiency. Following [1], we set F_i as a Conv-ReLU-Conv framework. To ensure the stability of OER, we apply the Hamiltonian system [9] to our OER module. By denoting Eq. (11) as $RK2(\cdot)$, the Hamiltonian-theory-guaranteed OER module is defined as follows:

$$\begin{aligned} &e_{t+1} = con\left(e_{t+1}^1, e_{t+1}^2\right), e_{t+1}^1 = e_t^1 + RK2_1\left(e_t^2\right), \\ &e_{t+1}^2 = e_t^2 - RK2_2\left(e_t^1 + RK2_1\left(e_t^2\right)\right), \end{aligned} \tag{12}$$

where e_t is split to e_t^1 and e_t^2 in channel-wise. Note that the OER module in Eq. (12) is a reversible and sta-

Methods	Publications	Backbones	*CHAMELEON* (76 images)				*CAMO* (250 images)				*COD10K* (2,026 images)				*NC4K* (4,121 images)			
			$M \downarrow$	$F_\beta \uparrow$	$E_\phi \uparrow$	$S_\alpha \uparrow$	$M \downarrow$	$F_\beta \uparrow$	$E_\phi \uparrow$	$S_\alpha \uparrow$	$M \downarrow$	$F_\beta \uparrow$	$E_\phi \uparrow$	$S_\alpha \uparrow$	$M \downarrow$	$F_\beta \uparrow$	$E_\phi \uparrow$	$S_\alpha \uparrow$
Common Setting: Single Input Scale and Single Stage																		
CPD [47]	CVPR19	ResNet50	0.048	0.775	0.874	0.857	0.113	0.675	0.723	0.716	0.053	0.578	0.776	0.750	0.072	0.719	0.808	0.787
SINet [6]	CVPR20	ResNet50	0.034	0.823	0.936	0.872	0.092	0.712	0.804	0.745	0.043	0.667	0.864	0.776	0.058	0.768	0.871	0.808
PFNet [28]	CVPR21	ResNet50	0.033	0.820	0.931	0.882	0.085	0.751	0.841	0.782	0.040	0.676	0.877	0.800	0.053	0.779	0.887	0.829
MGL-R [50]	CVPR21	ResNet50	0.031	0.825	0.917	0.891	0.088	0.738	0.812	0.775	0.035	0.680	0.851	0.814	0.053	0.778	0.867	0.833
LSR [25]	CVPR21	ResNet50	0.030	0.835	0.935	0.890	0.080	0.756	0.838	0.787	0.037	0.699	0.880	0.804	**0.048**	**0.802**	0.890	0.834
UGTR [48]	ICCV21	ResNet50	0.031	0.805	0.910	0.888	0.086	0.747	0.821	0.784	0.036	0.670	0.852	**0.817**	0.052	0.778	0.874	**0.839**
SLT-Net [2]	CVPR22	ResNet50	0.030	0.835	0.940	0.887	0.082	0.763	0.848	0.792	0.036	0.681	0.875	0.804	0.049	0.787	0.886	0.830
SegMaR-1 [14]	CVPR22	ResNet50	**0.028**	0.828	**0.944**	**0.892**	**0.072**	**0.772**	**0.861**	**0.805**	0.035	0.699	**0.890**	0.813	0.052	0.767	0.885	0.835
OSFormer [32]	ECCV22	ResNet50	**0.028**	**0.836**	0.939	0.891	0.073	0.767	0.858	0.799	**0.034**	**0.701**	0.881	0.811	0.049	0.790	**0.891**	0.832
FEDER-R50	—	ResNet50	**0.028**	**0.855**	**0.947**	**0.894**	**0.069**	**0.785**	**0.873**	**0.807**	**0.032**	**0.740**	**0.900**	**0.823**	**0.045**	**0.817**	**0.905**	**0.846**
SINet V2 [4]	TPAMI22	Res2Net50	0.030	0.816	0.942	0.888	0.070	0.779	0.882	0.822	0.037	0.682	0.887	0.815	0.048	0.792	0.903	0.847
BSA-Net [52]	AAAI22	Res2Net50	0.027	0.851	0.946	0.895	0.079	0.768	0.851	0.796	0.034	0.723	0.891	0.818	0.048	0.805	0.897	0.841
FEDER-R2N	—	Res2Net50	**0.026**	**0.856**	**0.947**	**0.903**	**0.066**	**0.807**	**0.897**	**0.836**	**0.029**	**0.748**	**0.911**	**0.844**	**0.042**	**0.824**	**0.913**	**0.862**
Other Setting: Multiple Input Scales (MIS)																		
ZoomNet [31]	CVPR22	ResNet50	0.024	0.858	0.943	0.902	0.066	0.792	0.877	0.820	0.029	0.740	0.888	**0.838**	0.043	0.814	0.896	0.853
FEDER-MIS	—	ResNet50	**0.023**	**0.869**	**0.959**	**0.906**	**0.064**	**0.801**	**0.893**	**0.827**	**0.028**	**0.756**	**0.913**	0.837	**0.041**	**0.832**	**0.915**	**0.859**
Other Setting: Multiple Stages (MS)																		
SegMaR-4 [14]	CVPR22	ResNet50	**0.025**	0.855	0.955	0.906	0.071	0.779	0.865	0.815	0.033	0.737	0.896	0.833	0.047	0.793	0.892	0.845
FEDER-MS-4	—	ResNet50	**0.025**	**0.874**	**0.964**	**0.907**	**0.067**	**0.809**	**0.886**	**0.822**	**0.028**	**0.752**	**0.917**	**0.851**	**0.042**	**0.827**	**0.917**	**0.863**

Table 1. Quantitative comparisons of the proposed FEDER and other state-of-the-art methods on four benchmarks. SegMaR-1 and SegMaR-4 denote SegMaR at one stage and four stages. R50 and R2N indicate ResNet50 and Res2Net50. The best results are marked in **bold**. For ResNet50 backbone in the common setting, the best two results are in **red** and **blue** fonts.

ble block [9], which further promotes the edge reconstruction performance. In this case, the final edge predictions $\{d_k^e\}_{k=1}^4$ can be acquired in the following manner:

$$d_k^e = conv3\left(e_{t+1}\right) = conv3\left(d_k^{fe}\right). \tag{13}$$

3.4. Loss Functions

The loss function of the proposed FEDER consists of two kinds of supervisions, namely the segmentation mask GT_s and edge GT_e of the camouflaged object, which correspond to the multi-scale segmentation maps $\{d_k^s\}_{k=1}^5$ and the multi-scale object edges $\{d_k^e\}_{k=1}^4$. Following [4], we employ the weighted binary cross-entropy loss L_{BCE}^w and weighted intersection-over-union loss L_{IoU}^w for segmentation supervision, which focuses more on those hard pixels. For edge supervision, we use the dice loss L_{dice} [29] to overcome the extreme imbalance between the positive and negative samples. Furthermore, to handle the multi-scale outputs, we up-sampling all the outputs to match the size of their corresponding ground truths during training. Therefore, the total loss of our FEDER is formulated as follows:

$$\begin{aligned} L_t = & \sum_{k=1}^{5} \frac{1}{2^{k-1}} \left(L_{BCE}^w \left(d_k^s, GT_s\right) + L_{IoU}^w \left(d_k^s, GT_S\right)\right) \\ & + \sum_{k=1}^{4} \frac{1}{2^{k-1}} L_{dice}\left(d_k^e, GT_e\right). \end{aligned} \tag{14}$$

4. Experiment

4.1. Experimental Setup

Datasets. We used four widely-used COD datasets for evaluation, including *CHAMELEON* [39], *CAMO* [18], *COD10K* [4], and *NC4K* [25]. *CHAMELEON* contains 76 hand-annotated images. *CAMO* has 1,250 camouflaged images with 1,000 training images and 250 testing images. Currently, *COD10K* is the largest COD benchmark, with 3,040 training images and 2,026 testing images. *NC4K* is a large-scale COD testing dataset, comprising 4,121 images. Following [4], we form the training set with 3,040 images from *COD10K* and 1,000 images from *CAMO*, while the remaining camouflaged images are used for testing.

Evaluation metrics. Four commonly-used metrics are employed for COD task, including mean absolute error M, adaptive F-measure F_β [27], mean E-measure E_ϕ [5], and structure measure S_α [3]. Larger F_β, $E\phi$, S_α, and smaller M indicate better segmentation performance.

Implementation details. The proposed FEDER is implemented in PyTorch on two RTX3090TI GPUs and is optimized by the Adam with momentum terms $(0.9, 0.999)$. Following the common setting [4, 6], our encoder (ResNet50 by default) is initialized with the model pre-trained on ImageNet [6]. In the training phase, the batch size is set to 36 and the learning rate is initialized to 0.0001, dividing by 10 every 80 epochs. For both training and inference phases, all images are resized as 384×384.

4.2. Comparison with the State-of-the-arts

Quantitative analysis. We compare the proposed FEDER with 12 cutting-edge techniques in three different settings, including the common setting (single input scale and single stage) and two other settings (multiple input scales (MIS) and multiple stages (MS)). In the MIS and MS settings, the proposed FEDER follows the practices of ZoomNet [31] and SegMaR [14], where FEDER-MS-4 means

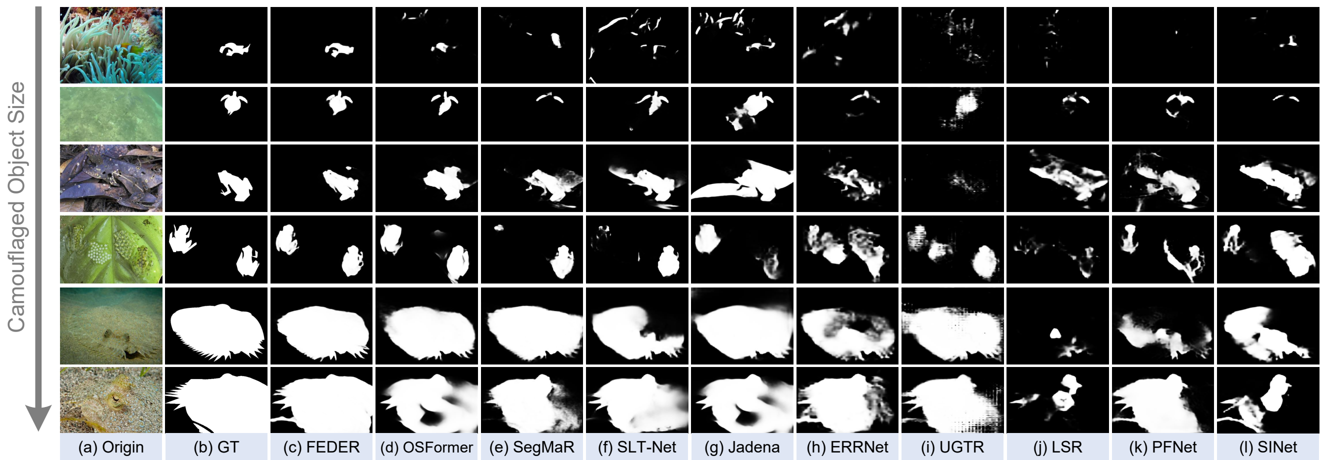

Figure 5. Visual comparisons of the proposed FEDER and other nine state-of-the-art methods. Our method can generate more accurate prediction maps with clearer boundaries than other methods. Please zoom in for more details.

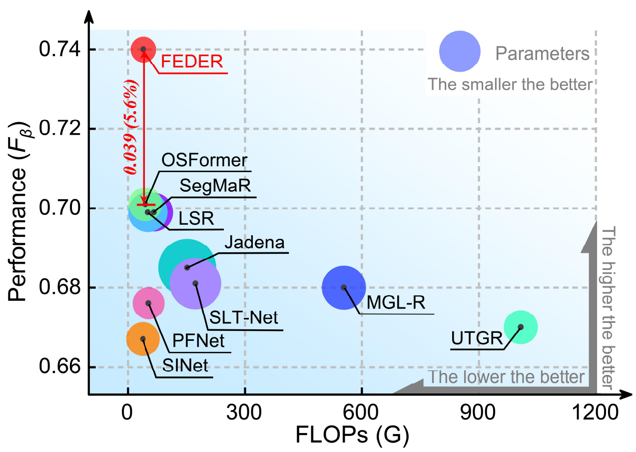

Figure 6. Performance-Params-FLOPs comparisons of some state-of-the-art deep learning-based COD methods on COD10K [4].

complete COD task with our FEDER at four stages following SegMaR-4. For a fair comparison, prediction maps of the above techniques are directly segmented by their provided models with no modifications. Besides, all prediction maps are evaluated with the same code. As shown in Tab. 1, our method achieves the best results in all settings and backbones, which comprehensively demonstrates the superiority of our FEDER. Notably, *COD10K* and *NC4K* are the two most challenging datasets in terms of the number of images and segmentation difficulty. In the common setting, the proposed FEDER surpasses the second best results 3.6% on average over all metrics on *COD10K* dataset and 2.6% on average on *NC4K* dataset. Such a big margin further confirms the effectiveness of the proposed method, both with the deep wavelet-like feature decomposition strategy and the ODE-inspired edge reconstruction module.

Qualitative analysis. Fig. 5 presents a visual comparison of our FEDER and other SOTAs. We select various typical and challenging camouflaged images and arrange them in order of the camouflaged object size, from smallest to largest. Note that most of these images suffer from the IS or ED challenge, which can confuse existing COD techniques, resulting in mislocalization, ambiguous boundaries, etc. In contrast to those methods, our FEDER can overcome such challenges and generate more competitive prediction maps in the following aspects: *(a) more accurate localization of small objects.* For those small objects under the IS challenge, precise localization is a significant problem due to subtle differences and can confuse most existing methods. Thanks to our HFA, LFA, and GFA modules, our FEDER can emphasize the inconspicuous discriminative features and thus ensure more accurate camouflaged object localization (in Rows 1 and 2). *(b) clearer edges on large objects.* For those large objects, our prediction maps can achieve much clearer boundaries than others (see Rows 7 and 8), which mainly attributes to our joint training strategy of edge and segmentation and our edge-friendly OER module. *(c) stronger suppression of redundant information.* In the IS challenge and degraded imaging scenarios, the detection performance can be inevitably influenced by redundant information, such as background noise. However, the proposed deep decomposition strategy can suppress the redundant information by filtering out those components with the most discriminative information, namely, the HF and LF components. Thus, as depicted in Rows 3 and 4 in Fig. 5, FEDER can generate more accurate prediction maps.

Efficiency analysis. We compare the performance, parameters, and FLOPs with other SOTAs on COD10K [4] in Fig. 6. As presented in Fig. 6, our proposed FEDER achieves the smallest FLOPs and parameters compared with other state-of-the-art deep learning-based COD techniques. Furthermore, our score in F_β is much higher than other methods and surpasses the second best one in 5.6%.

Methods		*COD10K* (2026 images)				*NC4K* (4121 images)			
		$M\downarrow$	$F_\beta\uparrow$	$E_\phi\uparrow$	$S_\alpha\uparrow$	$M\downarrow$	$F_\beta\uparrow$	$E_\phi\uparrow$	$S_\alpha\uparrow$
(a)	w/o DWD	0.035	0.697	0.861	0.794	0.049	0.776	0.873	0.817
(b)	FW–>LW	**0.032**	0.731	0.895	0.822	0.046	0.811	0.899	0.844
(c)	w/o HFA	0.033	0.728	0.895	0.817	0.047	0.805	0.891	0.840
(d)	w/o LFA	0.033	0.724	0.887	0.816	0.047	0.803	0.892	0.841
(e)	UFA–>GFA	0.033	0.720	0.894	0.810	0.047	0.804	0.890	0.838
(f)	FEDER	**0.032**	**0.740**	**0.900**	**0.823**	**0.045**	**0.817**	**0.905**	**0.846**

(a) Ablation study of DWD Component.

Methods		*COD10K* (2026 images)				*NC4K* (4121 images)			
		$M\downarrow$	$F_\beta\uparrow$	$E_\phi\uparrow$	$S_\alpha\uparrow$	$M\downarrow$	$F_\beta\uparrow$	$E_\phi\uparrow$	$S_\alpha\uparrow$
(a)	w/o OER	0.034	0.714	0.871	0.804	0.048	0.780	0.889	0.834
(b)	FC–>WGM	0.032	0.731	0.899	0.817	0.046	0.814	0.895	0.842
(c)	w/o HS	0.032	0.723	0.885	0.821	0.046	0.798	0.892	0.840
(d)	RB–>RK2	0.033	0.722	0.881	0.811	0.047	0.802	0.895	0.843
(e)	RK4–>RK2	**0.031**	**0.742**	**0.905**	**0.829**	**0.045**	0.816	**0.906**	**0.848**
(f)	FEDER	0.032	0.740	0.900	0.823	**0.045**	**0.817**	0.905	0.846

(b) Ablation study of OER module.

Table 2. Ablation study on *COD10k* [4] and *NC4K* [25], where w/o means without. (a) FW, LW, and UFA are short for fixed wavelet [40], learnable wavelet, and upsampling-based feature aggregation [51]. (b) FC, WGM, HS, RB, and RK4 are short for fixed coefficient, weighted gate mechanism, Hamiltonian system, residual block, and forth-order Runge-Kutta. The best results are marked in **bold**.

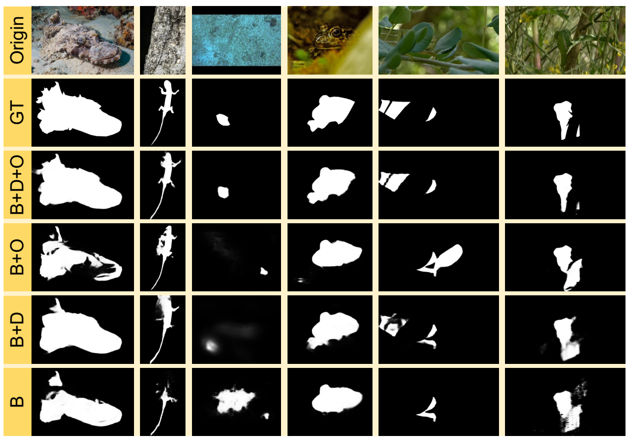

Figure 7. Visual verification of the effectiveness of the proposed components. B, D, and O denote baseline, DWD, and OER. Note that B+D+O is our proposed FEDER.

4.3. Ablation Study

We conduct the ablation study on the two largest datasets, namely *COD10K* and *NC4K*.

Effect of DWD component. We demonstrate the effectiveness of our DWD component both in visual verification (see Fig. 7) and quantitative analysis (see Tab. 2a). In Fig. 7, the networks with DWD component (Rows 3 and 5) exhibit more accurate localizations and stronger redundant information suppression capacities. In addition, we present more detailed information in Tab. 2a to verify the validity of each part in the DWD component. As presented in Tab. 2a, we prove the superiority of the DWD component (in (a)), learnable wavelet-like decomposition strategy (in (b)), HFA/LFA module (in (c) and (d)), and GFA module (in (e)).

Effect of OER module. The efficacy of our OER module is demonstrated visually by Figs. 7 and 8. In Fig. 7, methods with the OER module can generate the prediction maps with clearer edges. Besides, Fig. 8 illustrates the advancement of our OER module in generating accurate and clear boundary information. We provide detailed information about the superiority of our OER module in Tab. 2b. Specifically, (b) and (c) verify the effectiveness of our weighted gate mechanism (learnable coefficient) and Hamiltonian system. We further compare the performance of the OER module with

Figure 8. Visualizations of the reconstructed edge from the OER module, where the edges are marked in red for well-read.

different Runge-Kutta methods, i.e., RB (RK1), RK2, and RK4. Notably, the COD results with RK2 significantly outperform those with RB and are slightly lower than that with RK4. Therefore, we integrate RK2 into our OER module for a balance of performance and efficiency.

5. Conclusions

To address the IS and ED challenges, in this paper, we propose the FEDER model for COD. Specifically, we decompose the features into different frequency bands with learnable wavelets and filter out the most informative bands to excavate the subtle discriminative features with the HFA, LFA, and GFA modules, thereby solving the IS challenge. Besides, we propose to learn an auxiliary edge reconstruction task with our OER module to generate complete edges. Learning this auxiliary task along with the COD task thus facilitates the generation of precise segmentation results with accurate object boundaries, thus mitigating the ED challenge. Extensive experiments verify the superiority of our FEDER model in comparison with other SOTAs.

Acknowledgements: This research is partly supported by the National Key R&D Program of China (Grant No. 2020AAA0108303), and Shenzhen Science and Technology Project (Grant No. JCYJ20200109143041798) & Shenzhen Stable Supporting Program (WDZC20200820200655001) & Shenzhen Key Laboratory of next generation interactive media innovative technology (Grant No. ZDSYS20210623092001004).

References

[1] Bo Chang, Lili Meng, Eldad Haber, Lars Ruthotto, David Begert, and Elliot Holtham. Reversible architectures for arbitrarily deep residual neural networks. In *AAAI*, 2018. 3, 5

[2] Xuelian Cheng, Huan Xiong, Deng-Ping Fan, Yiran Zhong, Mehrtash Harandi, Tom Drummond, and Zongyuan Ge. Implicit motion handling for video camouflaged object detection. In *CVPR*, pages 13864–13873, 2022. 6

[3] Deng-Ping Fan, Ming-Ming Cheng, Yun Liu, Tao Li, and Ali Borji. Structure-measure: A new way to evaluate foreground maps. In *ICCV*, pages 4548–4557, 2017. 6

[4] Deng-Ping Fan, Ge-Peng Ji, Ming-Ming Cheng, and Ling Shao. Concealed object detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2021. 3, 6, 7, 8

[5] Deng-Ping Fan, Ge-Peng Ji, Xuebin Qin, and Ming-Ming Cheng. Cognitive vision inspired object segmentation metric and loss function. *Sci. Sin. Inf.*, 6:6, 2021. 6

[6] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *CVPR*, pages 2777–2787, 2020. 2, 3, 6

[7] Shang-Hua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip Torr. Res2net: A new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(2):652–662, 2019. 3

[8] Wooseok Ha, Chandan Singh, Francois Lanusse, Srigokul Upadhyayula, and Bin Yu. Adaptive wavelet distillation from neural networks through interpretations. *NIPS*, 34:20669–20682, 2021. 3, 4

[9] Eldad Haber and Lars Ruthotto. Stable architectures for deep neural networks. *Inverse Probl*, 34(1):014004, 2017. 5, 6

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016. 3

[11] Jianqin Yin Yanbin Han Wendi Hou and Jinping Li. Detection of the mobile object with camouflage color under dynamic background based on optical flow. *Procedia Eng.*, 15:2201–2205, 2011. 2

[12] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, pages 448–456. PMLR, 2015. 4

[13] Ge-Peng Ji, Lei Zhu, Mingchen Zhuge, and Keren Fu. Fast camouflaged object detection via edge-based reversible recalibration network. *Pattern Recognit.*, 123:108414, 2022. 5

[14] Qi Jia, Shuilian Yao, Yu Liu, Xin Fan, Risheng Liu, and Zhongxuan Luo. Segment, magnify and reiterate: Detecting camouflaged objects the hard way. In *CVPR*, pages 4713–4722, 2022. 1, 2, 6

[15] Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. Focal frequency loss for image reconstruction and synthesis. In *ICCV*, pages 13919–13929, 2021. 2

[16] Zhanghan Ke, Jiayu Sun, Kaican Li, Qiong Yan, and Rynson WH Lau. Modnet: Real-time trimap-free portrait matting via objective decomposition. In *AAAI*, pages 1140–1147, 2022. 3

[17] Ajay Kumar. Computer-vision-based fabric defect detection: A survey. *IEEE Trans. Ind. Electron.*, 55(1):348–363, 2008. 1

[18] Trung-Nghia Le, Tam V Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto. Anabranch network for camouflaged object segmentation. *Comput. Vis. Image Underst.*, 184:45–56, 2019. 6

[19] Bei Li, Quan Du, Tao Zhou, Yi Jing, Shuhan Zhou, Xin Zeng, Tong Xiao, Jingbo Zhu, Xuebo Liu, and Min Zhang. Ode transformer: An ordinary differential equation-inspired model for sequence generation. In *ACL*, pages 8335–8351, 2022. 3

[20] Boyi Li, Felix Wu, Kilian Q Weinberger, and Serge Belongie. Positional normalization. *NIPS*, 32, 2019. 4

[21] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. Multi-scale residual network for image super-resolution. In *ECCV*, pages 517–532, 2018. 4

[22] Kai Li, Yulun Zhang, Kunpeng Li, and Yun Fu. Adversarial feature hallucination networks for few-shot learning. In *CVPR*, pages 13470–13479, 2020. 4

[23] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *ICCV*, pages 10012–10022, 2021. 4

[24] Laetitia Loncan, Luis B De Almeida, José M Bioucas-Dias, Xavier Briottet, Jocelyn Chanussot, Nicolas Dobigeon, Sophie Fabre, Wenzhi Liao, Giorgio A Licciardi, Miguel Simoes, et al. Hyperspectral pansharpening: A review. *IEEE Geosci. Remote Sens. Mag.*, 3(3):27–46, 2015. 4

[25] Yunqiu Lv, Jing Zhang, Yuchao Dai, Aixuan Li, Bowen Liu, Nick Barnes, and Deng-Ping Fan. Simultaneously localize, segment and rank the camouflaged objects. In *CVPR*, pages 11591–11601, 2021. 6, 8

[26] Yue Ma, Yali Wang, Yue Wu, Ziyu Lyu, Siran Chen, Xiu Li, and Yu Qiao. Visual knowledge graph for human action reasoning in videos. In *ACM MM*, pages 4132–4141, 2022. 4

[27] Ran Margolin, Lihi Zelnik-Manor, and Ayellet Tal. How to evaluate foreground maps? In *CVPR*, pages 248–255, 2014. 6

[28] Haiyang Mei, Ge-Peng Ji, Ziqi Wei, Xin Yang, Xiaopeng Wei, and Deng-Ping Fan. Camouflaged object segmentation with distraction mining. In *CVPR*, pages 8772–8781, 2021. 2, 6

[29] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *3DV*, pages 565–571. IEEE, 2016. 6

[30] Yuxin Pan, Yiwang Chen, Qiang Fu, Ping Zhang, and Xin Xu. Study on the camouflaged target detection method based on 3d convexity. *Mod. Appl. Sci.*, 5(4):152, 2011. 2

[31] Youwei Pang, Xiaoqi Zhao, Tian-Zhu Xiang, Lihe Zhang, and Huchuan Lu. Zoom in and out: A mixed-scale triplet network for camouflaged object detection. In *CVPR*, pages 2160–2170, 2022. 6

[32] Jialun Pei, Tianyang Cheng, Deng-Ping Fan, He Tang, Chuanbo Chen, and Luc Van Gool. Osformer: One-stage camouflaged instance segmentation with transformers. In *ECCV*, pages 19–37. Springer, 2022. 6

[33] Natasha Price, Samuel Green, Jolyon Troscianko, Tom Tregenza, and Martin Stevens. Background matching and disruptive coloration as habitat-specific strategies for camouflage. *Sci. Rep.*, 9(1):1–10, 2019. 1

[34] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *AAAI*, pages 11908–11915, 2020. 4

[35] Dan Jeric Arcega Rustia, Chien Erh Lin, Jui-Yung Chung, Yi-Ji Zhuang, Ju-Chun Hsu, and Ta-Te Lin. Application of an image and environmental sensor network for automated greenhouse insect pest monitoring. *J. Asia Pac. Entomol.*, 23(1):17–28, 2020. 1

[36] Phil Sallee and Bruno Olshausen. Learning sparse multiscale image representations. *NIPS*, 15, 2002. 4

[37] Jiawei Shen, Zhuoyan Li, Lei Yu, Gui-Song Xia, and Wen Yang. Implicit euler ode networks for single-image dehazing. In *CVPRW*, pages 218–219, 2020. 3

[38] Chenyang Si, Weihao Yu, Pan Zhou, Yichen Zhou, Xinchao Wang, and Shuicheng Yan. Inception transformer. *arXiv preprint arXiv:2205.12956*, 2022. 5

[39] Przemysław Skurowski, Hassan Abdulameer, J Błaszczyk, Tomasz Depta, Adam Kornacki, and P Kozieł. Animal camouflage analysis: Chameleon database. *Unpublished manuscript*, 2(6):7, 2018. 6

[40] Radomir S Stanković and Bogdan J Falkowski. The haar wavelet transform: its status and achievements. *Comput. Electr. Eng.*, 29(1):25–44, 2003. 4, 8

[41] Martin Stevens and Sami Merilaita. Animal camouflage: current issues and new perspectives. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 364(1516):423–427, 2009. 1, 2, 3

[42] Cheng Tai and E Weinan. Multiscale adaptive representation of signals: I. the basic framework. *J. Mach. Learn. Res.*, 17(1):4875–4912, 2016. 4

[43] Christopher Torrence and Gilbert P Compo. A practical guide to wavelet analysis. *Bull. Am. Meteorol. Soc.*, 79(1):61–78, 1998. 4

[44] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. 4

[45] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *CVPR*, pages 13–19, 2020. 4

[46] E Weinan. A proposal on machine learning via dynamical systems. *Commun. Math. Stat.*, 1(5):1–11, 2017. 3, 5

[47] Zhe Wu, Li Su, and Qingming Huang. Cascaded partial decoder for fast and accurate salient object detection. In *CVPR*, pages 3907–3916, 2019. 6

[48] Fan Yang, Qiang Zhai, Xin Li, Rui Huang, Ao Luo, Hong Cheng, and Deng-Ping Fan. Uncertainty-guided transformer reasoning for camouflaged object detection. In *ICCV*, pages 4146–4155, 2021. 6

[49] Jaejun Yoo, Youngjung Uh, Sanghyuk Chun, Byeongkyu Kang, and Jung-Woo Ha. Photorealistic style transfer via wavelet transforms. In *ICCV*, pages 9036–9045, 2019. 3

[50] Qiang Zhai, Xin Li, Fan Yang, Chenglizhao Chen, Hong Cheng, and Deng-Ping Fan. Mutual graph learning for camouflaged object detection. In *CVPR*, pages 12997–13007, 2021. 5, 6

[51] Yulun Zhang, Zhifei Zhang, Stephen DiVerdi, Zhaowen Wang, Jose Echevarria, and Yun Fu. Texture hallucination for large-factor painting super-resolution. In *ECCV*, pages 209–225. Springer, 2020. 8

[52] Hongwei Zhu, Peng Li, Haoran Xie, Xuefeng Yan, Dong Liang, Dapeng Chen, Mingqiang Wei, and Jing Qin. I can find you! boundary-guided separated attention network for camouflaged object detection. In *AAAI*, 2022. 6