Camouflaged Object Detection with Feature Decomposition and Edge Reconstruction

Chunming He¹, Kai Li²*, Yachao Zhang¹, Longxiang Tang¹, Yulun Zhang³, Zhenhua Guo⁴, Xiu Li¹*
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 these 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.
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 FFeature 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 FFeature 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].
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 $\frac{W}{2^{k+1}} \times \frac{H}{2^{k+1}}$. 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 $\{f'_k\}_{k=1}^4$. 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 $d_5 = A_e(f_4)$, where $d_5$ 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-frequency bands, which are not easy to extract from the original image. To address this issue, we develop a learnable wavelet-like decomposition (LWD) module for deep adaptive feature decomposition. 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]. ODE-inspired networks are widely utilized in many fields, such as image dehazing [37] and machine translation [19].
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 \( \{ f^i_k \}_{k=1} \) and select the most informative HF and LF components for further refinement. We decompose \( f^i_k \) as

\[
(f^i_k)_{HF} = W_{HF} (f^i_k), \quad (f^i_k)_{LF} = W_{LF} (f^i_k),
\]

where \((f^i_k)_{HF}\) and \((f^i_k)_{LF}\) denote the HF and LF components of \( f^i_k \). \( W_{HF} \) 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 \( 3 \times 3 \) convolution layer, batch normalization (BN) [12], and ReLU. We then employ the joint attention module to accentuate those texture-rich regions for subtle structures of the two modules are illustrated in Fig. 3.

**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 \( (f^i_k)_{LF} \) as the input and outputs

\[
p^h_k = JA (PN (ResIN ((f^i_k)_{LF}))),
\]

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 \( \{ f^{h^i}_{k-1} \}_{k=1} \) that combines deep semantic information of the low-resolution feature \( (f^i_k)_{HF} \) (at a higher level) and the abundant spatial details of the high-resolution feature \( (f^h_{k-1})_{HF} \) (at a lower level) with the guidance of the attention map \( p^h_k \). Therefore, the aggregated feature \( f^{h^i}_{k-1} \) can better highlight the subtle discriminative features. To extract the attention-guided semantic information, we first generate the down-sampled aggregated feature \( f^{h^i}_{k-1} \) with the window-based linear model [23]:

\[
(f^{h^i}_{k-1})_{i} = \sigma_{down} ((f^{h}_{k-1})_{i} + \mu_{w}, \forall i \in s_{w}),
\]

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:

\[
\min_{\sigma_{w}, \mu_{w}} \sum_{i \in s_{w}} \left[ (p^h_{w})^2 ((f^{h^i}_{k-1})_{i} - ((f^{h^i}_{k-1})_{HF}))^2 + \epsilon \sigma_{w}^2 \right],
\]

where \( \epsilon \) is a constraint value for \( \sigma_{w} \). 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 \( \{ \sigma, \mu \} \), Eq. (4) can be rewritten as follows:

\[
f^{h^i}_{k-1} = \sigma_{i} \odot down((f^{h^i}_{k-1})_{HF}) + \mu_{i},
\]

where \( \odot \) is the Hadamard product. We then up-sample \( \{ \sigma_{i}, \mu_{i} \} \) as \( \{ \sigma_{h}, \mu_{h} \} \) and acquire the high-resolution aggregated feature \( f^{h^i}_{k-1} \) for enriching spatial details:

\[
f^{h^i}_{k-1} = GFA ((f^{h^i}_{k-1})_{HF}, (f^{h^i}_{k-1})_{HF}, p^h_k),
\]

\[
= \sigma_{h} \odot (f^{h^i}_{k-1})_{HF} + \mu_{h}.
\]
To iteratively acquire the aggregated features \( \{f_k^{h+1}\}_{k=2}^{4} \), we redefine GFA module by replacing \( (f_k^h)_{HF} \) with \( f_k^h \):
\[
f_k^h = GFA (f_k^h, (f_k^{h+1})_{HF}, p_k^{h}) ,
\]
where \( p_k^{h} = JA(ResB(f_k^{h})) \) and \( f_k^{h+1} = (f_k^{h+1})_{HF} \). 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 \( f_{k-1}^h \) is similar to \( f_k^{h+1} \), which can be seen in Supp.

Considering that bottom layers (at higher levels) focus more on HP 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 \( \{f_k^h\}_{k=1}^4 \). To balance performance and efficiency, the integrated features \( \{f_k^h\}_{k=1}^4 \) passed to the decoder are defined as:
\[
\begin{align*}
    f_1^1 &= \text{conv1} \left( \text{con} \left( f_1^1, \text{up} \left( f_1^h \right) \right) \right) , \\
    f_2^1 &= f_2^1 , \\
    f_3^1 &= f_3^1 , \\
    f_4^1 &= \text{conv1} \left( \text{con} \left( f_4^1, f_4^h \right) \right) ,
\end{align*}
\]
where \( \text{up} \cdot \) and \( \text{con} \cdot \) denote the up-sampling operation and the concatenation operation. \( \text{conv1} \) represents a \( 1 \times 1 \) convolution, which is used for channel-level integration.

### 3.3.3 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 thus re-calibrates the misclassified low-confidence regions. Specifically, we repeat the coarse segmentation map \( \{d_k \}_{k=0}^{2} \) 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 \( f_k^{h+1} \) with the reversed map and concatenate it with the edge feature \( d_k^{h+1} \) to obtain the segmentation result \( d_k^{h+2} \):
\[
d_k^{h+2} = \text{conv3} \left( \text{con} \left( d_k^{h+1}, f_k^h \circ \text{rv} \left( S(\text{rp}(d_k^{h+1})) \right) \right) \right) ,
\]
where \( \text{rp} \cdot \) and \( \text{rv} \cdot \) 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 \( e_t = \text{conv1} \left( \text{con} \left( f_1^{h+1}, \text{rp} \left( d_k^{h+1} \right) \right) \right) \), the proposed OER module can be formulated as follows:
\[
\begin{align*}
    e_{t+1} &= e_t + g_w F_1 + (1 - g_w) F_2 , \\
    g_w &= S (\sigma_g \text{con} \left( F_1, F_2 \right) + \mu_g) , \\
    F_1 &= F \left( e_t, \theta_1 \right) , \\
    F_2 &= F \left( e_t + F_1, \theta_1 \right) ,
\end{align*}
\]
where \( e_{t+1} = d_k^{h+1} \), \( \sigma_g \) and \( \mu_g \) are the learnable parameters in \( g_w \). \( \{F_1\}_{t=1}^2 \) denotes the intermediate layers with shared parameters \( \theta_1 \) for efficiency. Following [1], we set \( F_1 \) 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 \( \text{RK2} (\cdot) \), the Hamiltonian-theory-guaranteed OER module is defined as follows:
\[
\begin{align*}
    e_{t+1} &= \text{con} \left( e_{t+1}^{1}, e_{t+1}^{2} \right) , \\
    e_{t+1}^{1} &= e_t^{1} + \text{RK}^2 \left( e_t^{1} \right) , \\
    e_{t+1}^{2} &= e_t^{2} - \text{RK}^2 \left( e_t^{2} \right) ,
\end{align*}
\]
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-
table 1. Quantitative comparisons of the proposed FEDER and other state-of-the-art methods on four benchmarks. SegMaR-1 and SegMaR-2 denote SegMaR at one stage and four stages. R50 and R2N indicate ResNet50 and ResNet2N. The best results are marked in bold. For ResNet50 backbone in the common setting, the best two results are in red and blue fonts.

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 $\{d^s_k\}_{k=1}^{5}$ and the multi-scale object edges $\{d^e_k\}_{k=1}^{4}$. Following [4], we employ the weighted binary cross-entropy loss $L_{BCE}$ and weighted intersection-over-union loss $L_{IOU}$ 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:

$$L_e = \sum_{k=1}^{5} \frac{1}{2^{k-1}} (L_{BCE} (d^s_k, GT_s) + L_{IOU} (d^e_k, GT_e))$$

$$+ \sum_{k=1}^{4} \frac{1}{2^{k-1}} L_{dice} (d^e_k, GT_e).$$ (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 $F_\beta$ [27], mean E-measure $E_\phi$ [5], and structure measure $S_\alpha$ [3]. Larger $F_\beta, E_\phi, S_\alpha$, and smaller $M$ indicate better segmentation performance.

Implementation details. The proposed FEDER is implemented in PyTorch on two RTX3090 TI 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 x 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
Camouflaged Object Size

(a) Origin (b) GT (c) FEDER (e) SegMaR (f) SLT-Net (g) Jadena (h) ERRNet (i) UGTR (j) LSR (k) PFNet (l) SINet

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 $F_{\beta}$ is much higher than other methods and surpasses the second best one in 5.6%.
### 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 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


[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


[33] Natasha Price, Samuel Green, Jolyon Troscianko, Tom Tre-
genza, and Martin Stevens. Background matching and dis-
ruptive coloration as habitat-specific strategies for camou-

[34] Xu Qin, Zhihui Wang, Yuanchao Bai, Xiaodong Xie, and 
Huizhu Jia. Ffa-net: Feature fusion attention network for 

[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.*, 

[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 dehaz-

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

[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 

[41] Martin Stevens and Sami Merilaita. Animal camouflage: 

[42] Cheng Tai and E Weinan. Multiscale adaptive representation 
17(1):4875–4912, 2016. 4

[43] Christopher Torrence and Gilbert P Compo. A practical 
79(1):61–78, 1998. 4

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

[45] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wang-
 meng Zuo, and Qinghua Hu. Eca-net: Efficient channel at-
pages 13–19, 2020. 4

[46] E Weinan. A proposal on machine learning via dynamical 

[47] Zhe Wu, Li Su, and Qingming Huang. Cascaded partial de-
coder 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 

[50] Qiang Zhai, Xin Li, Fan Yang, Chenglizhao Chen, Hong 
Cheng, and Deng-Ping Fan. Mutual graph learning for cam-
ouflaged 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 

[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