Hu, X., Wang, S., Qin, X., Dai, H., Ren, W., Luo, D., Tai, Y. and Shao, L. (2023) High-resolution Iterative Feedback Network for Camouflaged Object Detection. In: 37th AAAI Conference on Artificial Intelligence (AAAI-23), Washington, DC, USA, 7-14 Feb 2023, pp. 881-889. ISBN 9781577358800 (doi: 10.1609/aaai.v37i1.25167)

The material cannot be used for any other purpose without further permission of the publisher and is for private use only.

There may be differences between this version and the published version. You are advised to consult the publisher’s version if you wish to cite from it.

https://eprints.gla.ac.uk/297303/

Deposited on 27 April 2023

Enlighten – Research publications by members of the University of Glasgow
http://eprints.gla.ac.uk
High-resolution Iterative Feedback Network for Camouflaged Object Detection

Xiaobin Hu¹, Shuo Wang², Xuebin Qin³, Hang Dai⁴, Wenqi Ren⁵, Donghao Luo¹, Ying Tai¹, Ling Shao⁶

¹Tencent Youtu Lab ²ETH Zurich ³Mohamed bin Zayed University of Artificial Intelligence ⁴University of Glasgow ⁵Sun Yat-sen University ⁶Terminus Group

{xiaobinhu, michaelluo, yingtai}@tencent.com, shawnwang.techn@gmail.com, xuebin@ualberta.ca, Hang.Dai@glasgow.ac.uk, renwq3@mail.sysu.edu.cn, ling.shao@ieee.org, *Corresponding Author

Abstract

Spotting camouflaged objects that are visually assimilated into the background is tricky for both object detection algorithms and humans who are usually confused or cheated by the perfectly intrinsic similarities between the foreground objects and the background surroundings. To tackle this challenge, we aim to extract the high-resolution texture details to avoid the detail degradation that causes blurred vision in edges and boundaries. We introduce a novel HitNet to refine the low-resolution representations by high-resolution features in an iterative feedback manner, essentially a global loop-based connection among the multi-scale resolutions. To design better feedback feature flow and avoid the feature corruption caused by recurrent path, an iterative feedback strategy is proposed to impose more constraints on each feedback connection. Extensive experiments on four challenging datasets demonstrate that our HitNet breaks the performance bottleneck and achieves significant improvements compared with 35 state-of-the-art methods. In addition, to address the data scarcity in camouflaged scenarios, we provide an application to convert the salient objects to camouflaged objects, thereby generating more camouflaged training samples from the diverse salient objects. Code will be made publicly available.

Introduction

Camouflaged object detection (COD) is a bio-inspired research area to detect hidden objects or animals that blend with their surroundings (Fan et al. 2021a). From biological and psychological studies (Cuthill 2019; Stevens and Merilaite 2009), the camouflage skill helps some animals prevent being the prey of their predators, and it also can cheat the human perception system that is sensitive to the coloration and the illumination around the edges. The camouflaged studies not only provide an effective way to deeply understand human perception system, but also benefit a wide range of downstream applications, such as medical image segmentation (Dong et al. 2021; Fan et al. 2020b,c), artistic creation (Chu et al. 2010), species discovery (Pérez-de la Fuente et al. 2012), and crack inspection (Fang et al. 2020).

In the last two decades, a growing interest is witnessed in developing algorithms capable of seeing targets through camouflage. Early methods aim to utilize the handcrafted low-level features (e.g., texture and contrast (Huerta et al. 2007), 3D convexity (Pan et al. 2011) and motion boundary (Yin et al. 2011)). These features still suffer from the limited capability of discriminating the foreground and the background in complex scenes. Recently, some CNNs-based frameworks have been proposed to analyze the visual similarities around boundaries between the camouflaged objects and their surroundings. The auxiliary information is extracted from the shared context as the boundary guidance for COD, such as features for identification (Fan et al. 2020a), classification (Le et al. 2019), boundary detection (Zhai et al. 2021) and uncertainties (Yang et al. 2021).

Although the approaches mentioned above have improved the performance, most methods discard the high-resolution details, including edges or textures, by down-sampling the high-resolution images. Fig. 2 shows an interesting phenomenon by evaluating the low-resolution (LR) and the high-resolution (HR) images on the same model well-trained on LR images, respectively. Although HR result suffers from a bit of over-segmentation with a lot of noise, it still has more high-frequency details like cat beards than that from LR. This implies that the high-resolution priors are crucial to the boundary and edge detection (Zhang et al. 2021; Wang et al. 2021a). The degradation of inputs from HR to LR leads
Based on the recursive operation, we design a novel recursive operation to refine the low-resolution features and computations of models in the inference stage. Our main contributions are summarized as:

- Adding additional training data to further improve the segmentation (RIR), and Iterative Feature Feedback (IFF). To reduce the computational cost for the HR feature maps in TFE via a global and cross-scale feedback strategy. To ensure the better aggregation of feedback feature, we use it-eration feature feedback (IFF) to impose constraints on feedback feature flow. In addition, we implement an application that converts the salient objects (Li et al. 2017; Zhao et al. 2021) as an image feature encoder. Then, we utilize the RIR module to recursively refine the LR feature extracted from TFE via a global and cross-scale feedback strategy.

- We propose a novel recursive operation to refine the low-resolution feature via a cross-scale feedback mechanism. The recursive operation is simple and can be easily extended to existing COD models.

- Based on the recursive operation, we design a novel framework, termed as High-resolution Iterative Feedback Network (HitNet) for COD task. To avoid the feature corruption caused by recurrent path, the corresponding iterative feedback loss with an iteration weight scheme is proposed for HitNet to penalize the output of each iteration.

- Our HitNet sets a new record, as shown in Fig. 1, breaking the performance bottleneck, compared with existing cutting-edge models on four benchmarks using four standard metrics. On COD10K, HitNet achieves \( F_{\text{m}} \) of 0.804, which is 7.5% higher than the second-best ZoomNet22 (Youwei et al. 2022).

Related Work

Camouflaged Object Detection. COD aims to spot the camouflaged object from its high-similarity surroundings (Fan et al. 2020a). It has wide applications (Fan et al. 2020b; Chu et al. 2010; Pérez-de la Fuente et al. 2012) and many COD methods (Youwei et al. 2022; Cheng et al. 2022) have been proposed. These methods can be categorized into two main classes: handcrafted-based and deep-learning-based. More specifically, most of the early works were developed based on the handcrafted features (e.g., colour and intensity features (Huerta et al. 2007), 3D convexity (Pan et al. 2011), and motion boundary (Yin et al. 2011)). But they are relatively less robust and prone to fail in complex scenarios. More studies resort to the powerful representation capacity of deep learning models to detect camouflaged objects in a data-driven way and have achieved impressive improvements against those handcrafted-based methods. On the one hand, deep models usually have many parameters, which ensures stronger representative capabilities for segmenting the camouflaged objects from their backgrounds. On the other hand, most of these deep models benefit by exploring the auxiliary knowledge, e.g., fixations, boundaries, and location. Nevertheless, most of models pay much attention to regional accuracy. At the same time, few of them explore the effectiveness of high-frequency information (in high-resolution), which plays a vital role in perceiving the clear boundaries or edges of camouflaged targets. Thus, it impedes the further improvements of COD models. To address this issue, we design a novel High-resolution Iterative Feedback Network, which sets a new record on all benchmarks.

Iterative Feedback Mechanism of Super-Resolution

allows the network to correct previous states (i.e., lower-resolution) with a higher-level output (i.e., higher-resolution) (Zamir et al. 2017; Hu et al. 2021). In image super-resolution, some studies proved certain improvements after using different feedback mechanisms, such as up- and down-projection units (Haris, Shakhnarovich, and Ukita 2018) and dual-state recurrent module (Han et al. 2018). However, most of these mechanisms are implemented by using recurrent structures (Li et al. 2019) while the information flows from the LR to HR images are still feed-forward. Recently, Li et al. (Li et al. 2019) proposed an image super-resolution feedback network to refine LR representation with HR information by outlining the edges and contours while suppressing smooth areas. Inspired by this work, we build our transformer-based high-resolution iterative feedback for COD. Different from Li et al. (Li et al. 2019), our feedback connection is designed as a global connection other than a local connection (Feng, Lu, and Ding 2019) and embedded into the multi-scale framework via a
feedback fusion block, which merges the information from multi-scale outputs. To avoid corruption of each iteration, we impose more constraints on each feedback connection by supervising each iteration with the corresponding loss.

**Vision Transformer.** The Transformer (Vaswani et al. 2017) was firstly proposed as a powerful tool in the domain of machine translation. Considering the superiority of transformer in modeling long-term dependencies, more recent studies have tried to exploit its potentials in different vision tasks, such as image classification (Dosovitskiy et al. 2020; Srinivas et al. 2021), object detection (Dai et al. 2021), and other low-level tasks (Yang et al. 2020). Thus, we adopt a Pyramid Vision Transformer (PVT) (Wang et al. 2021b) that uses a progressive shrinking pyramid structure to reduce the sequence length and a spatial-reduction attention layer to decrease the computation further when learning HR features.

**Proposed Method**

**Motivation.** Our motivation stems from the observation of degradation phenomenon, shown in Fig. 2. HR inputs generate more accurate predictions than LR inputs, especially for object boundaries. Thus, we aim to explore the feature interaction between high- and low-resolution for COD.

**Transformer-based Feature Extraction**

Currently, many of the vision transformers are GPU memory exhaustive and our HR features will further exaggerate the problem. To alleviate this issue, we choose the Pyramid Vision Transformer (PVT) (Wang et al. 2021b) as our feature extraction module, which can extract multi-scale features, and handle relatively higher resolution feature maps with less memory costs by its progressive shrinking strategy and spatial reduction attention mechanism.

![Figure 3: An overview of High-resolution Iterative Feedback Network (HitNet). Our HitNet consists of Transformed-based backbone for multi-scale feature extraction, multi-resolution iterative refinement to self-correct low-resolution features with high-resolution information via a cross-resolution iterative feedback mechanism, and iteration feature feedback to impose constraint on each iteration.](image)

**Multi-scale Feature Extraction.** PVT consists of four stages, and each stage includes a patch embedding and an encoder structure. The input features to each stage ($F_i$) are first divided into patches with size of $P_i$. After that, these features are fed into Transformer encoder structure to get the output features $X_i$ for the $i$-th. Then, we get the multi-scale features ($X_1, X_2, X_3, X_4$) with $(512, 320, 128, 64)$ number of channels and with $(\frac{1}{32}, \frac{1}{16}, \frac{1}{8}, \frac{1}{4})$ resolution of input images for further processing.

**Multi-Resolution Feedback Refinement**

The multi-scale feature $X$ extracted from the Transformer backbone are fed to a basic block $BA(\cdot)$ (Zhang et al. 2018) as shown in Fig. 3:

$$BA(X_i) = C_2(X_i) + C_1(C_2(X_i)) \cdot X_i,$$  (1)

where $X_i$ is the input feature of $i$-th scale produced by Transformer module, $C_2(\cdot)$ indicates two stacked convolutional layers with $3 \times 3$ filters. $C_1(\cdot)$ denotes the channel attention function (Zhang et al. 2018).

**Iterative Feedback Mechanism** is critical in this module to achieve high accuracy around the object boundary (see Fig. 6). The setting iterative number $in=1$ assumes the first iteration and no feedback feature transported from previous state. Thus, the $Y_1^{in}$ and $Y_3^{in}$ are the initial value (0) when
For iterative number \((in > 1)\), the feedback features are produced by previous iteration and then passed into feedback block \(FB(\cdot)\) as:

\[
\text{FB}(X_i + Y_i^{in}) = \text{Sq}(\text{Concat}(X_i \uparrow, Y_i^{in})),
\]

(2)

where \(Y_i^{in}\) is the feedback features of \(in\)-th iteration at \(i\)-th scale \((i \neq 2)\). Symbol \(\uparrow\) is the up-sampling operation from the size of \(X_i\) to \(Y_i^{in}\) to avoid degradation of the HR information. \(\text{Concat}(\cdot)\) indicates the channel-based concatenation operation between \(X_i\) and \(Y_i^{in}\), and \(\text{Sq}(\cdot)\) is feature size and channel compression using convolution layer with large kernel and stride\(^1\) to get identical size for \(i\)-th scale.

As shown in Fig. 3, with the prerequisite that the iterative number \((in > 1)\), the first scale structure receives \(X_1\) and \(Y_1^{in}\) and the output the feature can be defined as:

\[
S_1^{in} = \text{BA}(\text{FB}(X_1 + Y_1^{in})),
\]

(3)

Then, \(S_1^{in}\) is further fed into the next scale to generate next output feature as follows:

\[
S_2^{in} = \text{BA}(\text{Concat}(S_1^{in} \uparrow, X_2)),
\]

(4)

Finally, the features of the previous scale are transported to the next scale as:

\[
S_3^{in} = \text{BA}(\text{Concat}(S_2^{in} \uparrow, \text{FB}(X_3 + Y_3^{in}))),
\]

(5)

After ending at \(in\)-th iteration, \((in + 1)\)-th iteration starts from the first scale to the last scale in the same way. The feedback features \(Y_1^{in+1}\) and \(Y_3^{in+1}\) at \((in + 1)\)-th iteration are updated as follows:

\[
Y_1^{in+1} = Y_3^{in+1} = \text{Conv}(S_3^{in}),
\]

(6)

where \(\text{Conv}\) is a convolution layer with 3 kernel size and 1 padding. The segmentation prediction map of \(i\)-th iteration \((Y_i^{in})\) is obtained via two stacked convolution operation \(Y_i^{in} = \text{Conv}(\text{Conv}(S_3^{in}))\). The low and upper index of \(Y\) indicates the scale and the iteration information.

The design intuitions on different scales are mainly motivated to get a better cross-scale data flow. The feedback features are explicitly imported into the top and third top scales for the data flow. As the data flow works, the second-top scale can get the implicit feedback features from the top scale. From our experiments, this setting can decrease the computational cost but maintaining good performance. Our HitNet breaks the performance bottleneck due to the following three indispensable mechanisms:

- In each iteration, it outputs an intermediate HR segmentation map that is supervised with a segmentation loss, enabling the feedback features to learn HR cues.
- The HR feedback features merge with inputs in a feedback block, alleviating the degradation of HR information.
- It uses a feedback fusion mechanism to exploit the HR data flow in a multi-scale structure.

\(^1\)If \(i=1\), kernel = 8 with stride = 4 while \(i=3\), kernel = 1 with stride = 1.

**Iteration Feature Feedback**

To tailor satisfactory feedback feature flow and avoid the feature corruption caused by recurrent path, we present iteration feature feedback strategy to tie the each feedback feature with the segmentation ground-truth. Intuitively, the data flow of feedback features can be controlled by the loss function. Our basic loss function is defined as \(L = L_{\text{IoU}} + L_{\text{BCE}}\), where \(L_{\text{IoU}}\) is the weighted intersection-over-union (IoU) loss and \(L_{\text{BCE}}\) denotes the weighted binary cross entropy (BCE) loss. Unlike other recurrent structures (Wei, Wang, and Huang 2020), we compute the HR prediction loss in each iteration and use an iteration-weight scheme to penalize the output of each iteration when predicting a HR segmentation map:

\[
L_{\text{HIP}} = \sum_{in} (w \cdot in)L(Y^{in}) + L(Y'),
\]

(7)

where \(in\) is the current iteration number, \(N\) is the total iteration number, \(w\) is the weight parameter, \(Y^{in}\) is the output of \(in\)-th iteration, \(Y'\) is the output of graph-based resolution fusion. In this way, our iteration-weight scheme focuses on the features of deeper iterations by assigning higher weights.

In this session, to efficiently integrate the features from the previous module, we design an adaptive feature fusion module (shown in Fig. 4).

\[
Y' = \text{AFF}(T_1, T_2),
\]

(8)

where \(Y'\) is final prediction map, \(T_1\) is the \(Y^{in=4}\), \(\text{AFF}\) is the Adaptive Feature Fusion module. Specifically, when given the features \(T_1\) and \(T_2\), the global average pooling is used to get the shrunk features with the size of \(1 \times 1 \times C\) on the channel dimension. Afterwards, the operations \((f_c\) and \(f_a)\) are implemented on the shrunk features to get the channel-wise weights and adaptive feature-wise coefficients \((\alpha)\). \(f_c\) and \(f_a\) are the stacked combinations of operators \(\text{nn.Linear}+\text{ReLU}+\text{nn.Linear}+\text{Sigmoid}\). The features \(F_1\) and \(F_2\) are obtained after assigning the channel-wise weights on the channel features of \(T_1\) and \(T_2\). Finally, the final prediction map \(Y'\) can be achieved weighted by adaptive coefficients \(\alpha_1\) and \(\alpha_2\) as follows:

\[
Y' = F_1 \times \alpha_1 + F_2 \times \alpha_2,
\]

(9)

**Experiments**

**Experimental Settings**

**Datasets.** Our experiments are based on four widely-used COD datasets: (1) CHAMELEON (Skurowski et al. 2018) collects 76 high-resolution images from the Internet with the label of camouflage animals. (2) CAMO (Le et al. 2019) includes 2,500 images with eight categories. (3) COD10K (Fan et al. 2020a) is the largest collection containing 10,000 images that divided into 10 super-classes and 78 sub-classes from multiple photography websites. (4) NC4K (Lv et al. 2021) consists of 4,121 images and is commonly used to evaluate the generalization ability of models. Following previous studies and benchmarks (Zhai et al. 2021; Fan
Table 1: Quantitative results of our method and other 35 state-of-the-art methods on four benchmark datasets. The best results are highlighted in bold, and the second-best is marked in underline. Our HitNet outperforms the second-best model by a large margin. For a fair comparison, all results are either provided by the published paper or reproduced by an open-source model re-trained on the same training set with the recommended setting. ❍ means that the models are not evaluated in current dataset in their papers.

![Table 1](image)
Visual performance of the proposed HitNet. Our algorithm is capable of tackling challenging cases (e.g., complex edges with dense thorn, multiply camouflaged objects, partly occlusion, and global thin edges).

Figure 5

Figure 6: Visual performance of each iteration in our iterative feedback mechanism in our RIR module.

Figure 7: Configuration of MAE error vs. Inference time (ms) on Iteration Number (In).

Multi-scale connection fusion (denoted as ‘Multi-fusion’).

Evaluation of Different Backbones. To assess the contribution of the CNN-based and Transformer-based backbones, we substitute the Transformer backbone of HitNet with Res2Net-50 (Gao et al. 2021) used in 2022 DCO-FD as the version of ‘HitNet+Res2Net-50’. In addition, we also replace Transformer backbone with ResNet-50 (He et al. 2016) used in 2022 SegMaR as the version of ‘HitNet+ResNet-50’. For a fair comparison, the input resolutions of our different variants and the corresponding baselines are the same. As shown in Tab. 4, compared with 2022 DCO-FD and 2022 SegMaR models with the same backbones as ours, our HitNet achieves superior performance with high quantitative results. Our algorithm (‘HitNet+Res2Net-50’) without Transformer backbone still achieves the best performance compared with all 35 SOTA methods. But compared with the Res2Net-50 and ResNet-50 backbone, Transformer can achieve better performance due to its superiority of global receptive field. More experiments (i.e., inference time, computational complexity, and
Table 2: Ablation analyses of HitNet on COD10K dataset.

| Metric | w/o TFE | w/o RIR | w/o IFF-1 | w/o IFF-2 | HitNet |
|--------|--------|--------|-----------|-----------|--------|
| F_P ↑  | 0.745  | 0.712  | 0.793     | 0.801     | 0.804  |
| S_P ↑  | 0.851  | 0.833  | 0.860     | 0.862     | 0.869  |
| M ↓    | 0.026  | 0.031  | 0.023     | 0.024     | 0.023  |
| E_φ i  | 0.921  | 0.907  | 0.931     | 0.932     | 0.936  |

Table 3: Ablation study on indispensable factors of Iterative Feedback Mechanism on COD10K dataset.

| Configurations | Performance |
|----------------|-------------|
| Tc | FB | Multi-fusion | MAE ↓ |
| ✓ | × | × | 0.0245 |
| ✓ | ✓ | × | 0.0252 |
| ✓ | ✓ | ✓ | 0.0246 |
| ✓ | ✓ | ✓ | 0.0230 |

qualitative evaluation) are added to supplementary material.

Application

The camouflaged dataset is very scarce and rare only existing in camouflaged scenarios, and almost all public camouflaged datasets have been used in our paper. In contrast, there exists a large-scale salient dataset that is almost 100 times more than the camouflaged ones. It is an open question that how to well-utilize the abundant salient dataset to improve the camouflaged object accuracy without extra annotation labor. Thus, we adopt a cross-domain learning (CDL) technique that converts salient objects to camouflaged objects to achieve this goal. In addition, we propose a contrastive index to evaluate the camouflaged level. This index can be acted as the criterion to discard some hard cases with unchangeable intrinsic salient objects.

Cross-domain Learning. We employ the cycle-consistency structure (Zhu et al. 2017) to learn the camouflaged features and embed these features into the salient objects in an unsupervised cross-domain learning manner as shown in Fig. 8. The cycle-consistency loss can be formulated as:

$$L_{cyc}(G, F) = E_x[(F(G(x)) - x)||_1 + E_y[(G(F(y)) - y)||_1]$$

(10)

where G aims to construct fake images \( \{G(x)\} \) from salient samples \( \{x\} \) to get close to camouflaged domain \( Y \), while \( D(Y) \) tries to distinguish between the translated camouflaged samples \( \{G(x)\} \) and real camouflaged samples \( \{y\} \). \( F \) is another translator from camouflaged to salient objects.

The procedure is concluded as a min-max optimization task in the adversarial loss function used in CycleGAN.

To better select the converted camouflaged objects, we propose a contrastive index, considering the pixel-level similarity between object and its surroundings:

$$I_{sc} = \frac{1}{\text{Num}} \sum_i \| P_i - P_m \|_{2}^{|| \in (P_m - P_{std}, P_m + P_{std})}$$

(11)

where \( I_{sc} \) is the index of camouflaged level, \( P_i \) is \( i \)-th pixel intensity value, \( P_m \) is the mean value of images, \( P_{std} \) is the standard deviation, and \( i \) is the pixel index that belongs to

2Minimize the generator loss while maximizing the discriminator loss.

Figure 8: The overview of salient-to-camouflaged cross-domain learning pipeline. The \( S \) is the salient domain, and the \( C \) is the camouflaged domain. \( D_Y \) is the discriminator of the salient domain, and \( D_Y \) is the discriminator of the camouflaged domain.

Table 5: Quantitative results on different training strategies. ‘w/o’ means without any data strategy, ‘Salient data’ means adding salient data for training.

| Data Strategy | COD10K (Fan et al. 2020a) |
|---------------|---------------------------|
|               | S_P ↓ | E_φ ↓ | M ↓ |
| w/o           | 0.969 | 0.956 | 0.394 | 0.023 |
| Salient data  | 0.917 | 0.935 | 0.788 | 0.027 |
| CDL (Ours)    | 0.879 | 0.930 | 0.812 | 0.022 |

one \( \sigma \) rule to exclude the effect of extreme values. In Fig. 8, the car is an abandoned example detected as a high salient case by our contrastive index. Empirically, we set the threshold of \( I_{sc} \) as \( I_{sc} = 20 \).

Qualitative and Quantitative Evaluation. Our CDL strategy makes exciting and meaningful explorations on camouflaged domains from the following aspects. (1) As shown in Fig. 8, it converts salient objects to camouflage objects, which bridges the gap between salient and camouflage fields. (2) As shown in Tab. 5, it finds that the usage of salient object data cannot improve but severely deteriorate the performance of COD. Meanwhile, CDL strategy can make the distribution of salient objects closer to the camouflage object distribution. Thus, it can improve the accuracy of COD and reduce the relative MAE error by 4.3%.

Conclusion

We propose a novel high-resolution iterative feedback network (HitNet) to extract the informative and high-resolution representations for tackling the degradation issue of segmentation details on the COD task. HitNet can adaptively refine the low-resolution features with high-resolution information in an iterative feedback manner. More importantly, our approach achieves remarkable performance improvements and significantly outperforms 35 cutting-edge models on four challenging datasets. Finally, we introduce the cross-domain learning strategy to implement an application that converts the salient object to the camouflaged object, potentially enlarging the diversity of the COD dataset.
References

Chen, K.; Pang, J.; Wang, J.; Xiong, Y.; Li, X.; Sun, S.; Feng, W.; Liu, Z.; Shi, J.; Ouyang, W.; et al. 2019. Hybrid task cascade for instance segmentation. In CVPR.

Chen, Z.; Xu, Q.; Cong, R.; and Huang, Q. 2020. Global context-aware progressive aggregation network for salient object detection. In AAI.

Cheng, X.; Xiong, H.; Fan, D.-p.; Zhong, Y.; Harandi, M.; Drummond, T.; and Ge, Z. 2022. Implicit Motion Handling for Video Camouflaged Object Detection. In CVPR.

Chu, H.-K.; Hsu, W.-H.; Mitra, N. J.; Cohen-Or, D.; Wong, T.-T.; and Lee, T.-Y. 2010. Camouflage images. ACM TOG, 29(4): 51–1.

Cuthill, I. 2019. Camouflage. JOZ, 308(2): 75–92.

Dai, Z.; Cai, B.; Lin, Y.; and Chen, J. 2021. Up-detr: Unsupervised pre-training for object detection with transformers. In CVPR.

Dong, B.; Wang, W.; Fan, D.-P.; Li, J.; Fu, H.; and Shao, L. 2021. Polyp-pVT: Polyp Segmentation with PyramidVision Transformers. arXiv preprint arXiv:2108.06932.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. In ICLR.

Fan, D.-P.; Cheng, M.-M.; Liu, Y.; Li, T.; and Borji, A. 2017. Structure-measure: A New Way to Evaluate Foreground Maps. In ICCV.

Fan, D.-P.; Ji, G.-P.; Cheng, M.-M.; and Shao, L. 2021a. Concealed Object Detection. IEEE TPAMI.

Fan, D.-P.; Ji, G.-P.; Qin, X.; and Cheng, M.-M. 2021b. Cognitive vision inspired object segmentation metric and loss function. SSI, 6.

Fan, D.-P.; Ji, G.-P.; Sun, G.; Cheng, M.-M.; Shen, J.; and Shao, L. 2020a. Camouflaged object detection. In CVPR.

Fan, D.-P.; Ji, G.-P.; Zhou, T.; Chen, G.; Fu, H.; Shen, J.; and Shao, L. 2020b. Pranet: Parallel reverse attention network for polyp segmentation. In MICCAI.

Fan, D.-P.; Zhou, T.; Ji, G.-P.; Zhou, Y.; Chen, G.; Fu, H.; Shen, J.; and Shao, L. 2020c. Inf-net: Automatic covid-19 lung infection segmentation from ct images. IEEE TMI, 39(8): 2626–2637.

Fang, F.; Li, L.; Gu, Y.; Zhu, H.; and Lim, J.-H. 2020. A novel hybrid approach for crack detection. Pattern Recognition, 107: 107474.

Feng, M.; Lu, H.; and Ding, E. 2019. Attentive feedback network for boundary-aware salient object detection. In CVPR.

Gao, S.; Cheng, M.-M.; Zhao, K.; Zhang, X.-Y.; Yang, M.-H.; and Torr, P. H. 2021. Res2net: A new multi-scale backbone architecture. IEEE TPAMI, 43(02): 652–662.

Han, W.; Chang, S.; Liu, D.; Yu, M.; Witbrock, M.; and Huang, T. S. 2018. Image super-resolution via dual-state recurrent networks. In CVPR.

Haris, M.; Shakhnarovich, G.; and Ukita, N. 2018. Deep back-projection networks for super-resolution. In CVPR.

He, K.; Gkioxari, G.; Dollár, P.; and Girshick, R. 2017. Mask r-cnn. In ICCV.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In CVPR.

Hu, Y.; Yan, Y.; Ren, W.; Li, H.; Bayat, A.; Zhao, Y.; and Menze, B. 2021. Feedback Graph Attention Convolutional Network for MR Images Enhancement by Exploring Self-Similarity Features. In MIDL.

Huang, Z.; Huang, L.; Gong, Y.; Huang, C.; and Wang, X. 2019. Mask scoring r-cnn. In CVPR.

Huerta, I.; Rowe, D.; Mozero, M.; and González, J. 2007. Improving background subtraction based on a casuistry of colour-motion segmentation problems. In IbPRIA.

Jia, Q.; Yao, S.; Liu, Y.; Fan, X.; Liu, R.; and Luo, Z. 2022. Segment, Magnify and Reiterate: Detecting Camouflaged Objects the Hard Way. In CVPR, 4713–4722.

Le, T.-N.; Nguyen, T. V.; Nie, Z.; Tran, M.-T.; and Sugimoto, A. 2019. Anabranch network for camouflaged object segmentation. CVIU, 184: 45–56.

Li, A.; Zhang, J.; Lv, Y.; Liu, B.; Zhang, T.; and Dai, Y. 2021. Uncertainty-aware joint salient object and camouflaged object detection. In CVPR, 10071–10081.

Li, X.; Yang, F.; Cheng, H.; Chen, J.; Guo, Y.; and Chen, L. 2017. Multi-scale cascade network for salient object detection. In ACM MM, 439–447.

Li, Z.; Yang, J.; Liu, Z.; Yang, X.; Jeon, G.; and Wu, W. 2019. Feedback network for image super-resolution. In CVPR.

Lin, T.; Dollar, P.; Girshick, R.; He, K.; Hariharan, B.; and Belongie, S. 2017. Feature Pyramid Networks for Object Detection. In CVPR.

Liu, J.-J.; Hou, Q.; Cheng, M.-M.; Feng, J.; and Jiang, J. 2019. A Simple Pooling-Based Design for Real-Time Salient Object Detection. In CVPR.

Liu, N.; Han, J.; and Yang, M.-H. 2018. PiCANet: Learning pixel-wise contextual attention for saliency detection. In CVPR.

Loshchilov, I.; and Hutter, F. 2019. Decoupled weight decay regularization. In ICLR.

Luo, Z.; Mishra, A.; Achkar, A.; Eichel, J.; Li, S.; and JodoIn, P.-M. 2017. Non-local deep features for salient object detection. In CVPR.

Lv, Y.; Zhang, J.; Dai, Y.; Li, A.; Liu, B.; Barnes, N.; and Fan, D.-P. 2021. Simultaneously deep features for salient object detection. In CVPR.

Margolin, R.; Zelnik-Manor, L.; and Tal, A. 2014. How to evaluate foreground maps? In CVPR.

Mei, H.; Ji, G.-P.; Wei, Z.; Yang, X.; Wei, X.; and Fan, D.-P. 2021. Camouflaged object segmentation with distraction mining. In CVPR.
