# Supplementary Material for Feature Shrinkage Pyramid for Camouflaged Object Detection with Transformers

Zhou Huang[1,2†] Hang Dai[3†] Tian-Zhu Xiang[4*] Shuo Wang[5]
Huai-Xin Chen[2] Jie Qin[6] Huan Xiong[7]
[1]Sichuan Changhong Electric Co., Ltd. [2]UESTC [3]University of Glasgow
[4]G42 [5]ETH Zurich [6]CCST, NUAA [7]MBZUAI
chowhuang@std.uestc.edu.cn, hang.dai@glasgow.ac.uk, {tianzhu.xiang19, shawnwang.tech, qinjiebuaa, huan.xiong.math}@gmail.com, huaixinchen@uestc.edu.cn

## 1. Feature Shrinkage Decoder

The details of the feature shrinkage decoder (FSD) are introduced in the manuscript, and here we summarize the calculation process of FSD in Algorithm 1.

**Algorithm 1** Feature shrinkage decoder.

**Input:** $\{F_n^0 | n \in [12, 1]\}$, inputs for FSD layer "0".
**Output:** $\{P_i | i \in [0, 3]\}$, predictions for each layer of FSD.

The number of outputs in each layer of FSD is $num\_op = [6, 3, 2, 1]$
**for** $(i, m)$ **in** enumerate($num\_op$) :
  **for** $n = [m : 1]$ :
    **if** $n = m$ :
      $(dF_n^i, F_n^{i+1}) \leftarrow AIM(F_{2n}^i, F_{2n-1}^i)$ ;
    **else** :
      $(dF_n^i, F_n^{i+1}) \leftarrow AIM(\mathcal{C}\dashv\sqcup(dF_n^i, F_{2n}^i), F_{2n-1}^i)$ ;
  **end for**
    $P_i = \text{sigmoid}(dF_n^i)$
**end for**

## 2. More Experiments

### 2.1. Datasets

We conduct experiments on three COD datasets, including CAMO [8], COD10K [5], and NC4K [12].

**CAMO** is the first dataset for COD and contains 2.5K images (2K for training and 0.5K for testing) with manual annotations, of which 1.25K camouflaged images and 1.25K non-camouflaged images.

**COD10K** is currently the largest challenging dataset for COD. It contains 10K images with dense annotations (6K for training and 4K for testing) covering 78 categories.

**NC4K** is currently the largest test dataset for COD, containing 4,121 images. It not only contains binary ground truth maps, but also provides camouflaged object ranking annotations.

### 2.2. Evaluation Metrics

In our experiments, we adopt five kinds of evaluation metrics that are widely used in COD tasks for quantitative evaluation, including S-measure [2] ($S_m$), F-measure [1] ($F_\beta$), weighted F-measure [14] ($F_\beta^\omega$), E-measure [3] ($E_m$), and mean absolute error ($MAE$, $\mathcal{M}$).

**Mean absolute error.** $\mathcal{M}$ is to calculate the average absolute error of the prediction of camouflaged objects ($C$) and ground truth ($G$), which is defined as:

$$\mathcal{M} = \frac{1}{N}\sum_{i=1}^{N} |C(i) - G(i)|, \tag{1}$$

where $N$ is the total pixels of the image.

**S-measure.** Considering that camouflaged objects have complex shapes, we use $S_m$ that combines region-aware ($S_r$) and object-aware ($S_o$) to calculate structural similarity, which is defined as:

$$S_m = \alpha * S_o + (1 - \alpha) * S_r, \tag{2}$$

where $\alpha \in [0, 1]$ is the balance parameter and set to 0.5 in our experiments.

**F-measure.** $F_\beta$ is a comprehensive metric that takes into account both precision ($P$) and recall ($R$), and is defined as:

$$F_\beta = \frac{\left(1 + \beta^2\right) P \cdot R}{\beta^2 \cdot P + R}, \tag{3}$$

where $\beta$ is the balance parameter and $\beta^2$ is set to 0.3 in this paper. We report adaptive F-measure ($F_\beta^a$), mean F-measure ($F_\beta^m$) and maximum F-measure ($F_\beta^x$) in our experiments. The "adaptive" means that two times the average value of the prediction map pixels is used as the threshold for calculating precision and recall.

[†]Equal contributions. *Corresponding author: *Tian-Zhu Xiang*.

**Weighted F-measure.** $F_\beta^\omega$ is obtained based on the $F_\beta$ by combining the weighted precision defined by measure exactness and the weighted recall defined by measure completeness, which is calculated as:

$$F_\beta^\omega = \frac{\left(1+\beta^2\right) \times P^\omega \times R^\omega}{\beta^2 \times P^\omega + \mathrm{R}^\omega}, \tag{4}$$

**E-measure.** $E_m$ is an metric based on human visual perception, which can complete pixel-level matching and image-level statistics, which is denoted as:

$$E_m = \frac{1}{N} \sum_{i=1}^{N} \phi_{FM}\left(i\right), \tag{5}$$

where $\phi_{FM}$ denotes the enhanced-alignment matrix. We report adaptive E-measure ($E_\beta^a$), mean E-measure ($E_\beta^m$) and maximum E-measure ($E_\beta^x$) in our experiments.

## 2.3. Quantitative Experiments

We show more quantitative experimental results on three benchmark COD datasets. The methods used in the experiments for comparison include 10 SOD methods (BASNet [17], CPD [21], EGNet [27], SCRN [22], F$^3$Net [20], CSNet [6], SSAL [26], ITSD [28], UCNet [25], VST [11]) and 13 COD methods (SINet [5], SLSR [12], PFNet [15], MGL-R [24], UJSC [9], C$^2$FNet [18], UGTR [23], BSANet [29], OCE-Net [10], BGNet [19], SegMaR [7], ZoomNet [16], SINet-v2 [4]).

**Comprehensive Evaluation.** As shown in Tab. 2 and Tab. 3, we further list more comprehensive evaluation results on three COD datasets. It can be seen that our model achieves the best detection performance overall.

**Evaluation for Subclasses.** In addition to the overall quantitative comparison of the COD10K dataset, we also report quantitative results of some representative competitors on each subclass in Tab. 4. It can be seen that our model outperforms other competitors for most subclasses of the COD10K dataset. On the other hand, adapting and improving the model based on the results of each subclass is one of our future work.

## 2.4. Qualitative Comparison

Due to space limitations of the manuscript, we add more visual comparisons to this supplementary material for further demonstration of the performance of our model. Fig 1, 2, 3, and 4 show examples containing small, large, obscured, and boundary indistinguishable camouflaged objects, respectively. As can be seen from these visual comparisons, our model is more robust to a wide range of challenging scenarios, showing superior visual performance for more accurate and complete predictions.

Table 1. Ablation studies on COD10K and NC4K. ① is similar to [13], which adjusts our decoder to pairwise feature aggregation with overlap, and removes lateral supervision and feature interaction within the same layer. ② is a decoder that adds the lateral supervision and feature interaction within the same layer to ①. The difference between ours and ② is that our method is pairwise feature aggregation without overlapping.

| No. | COD10K | | | | | | NC4K | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $S_m\uparrow$ | $F_\beta^\omega\uparrow$ | $F_\beta^m\uparrow$ | $E_\phi^m\uparrow$ | $E_\phi^x\uparrow$ | $\mathcal{M}\downarrow$ | $S_m\uparrow$ | $F_\beta^\omega\uparrow$ | $F_\beta^m\uparrow$ | $E_\phi^m\uparrow$ | $E_\phi^x\uparrow$ | $\mathcal{M}\downarrow$ |
| ① | .849 | .732 | .761 | .887 | .922 | .029 | .872 | .804 | .832 | .901 | .927 | .038 |
| ② | .850 | **.736** | .768 | .893 | **.931** | **.026** | .878 | **.817** | .841 | .912 | .936 | .036 |
| **Ours** | **.851** | .735 | **.769** | **.895** | .930 | **.026** | **.879** | .816 | **.843** | **.915** | **.937** | **.035** |

## 2.5. More Ablation Experiments on FSD

Tab. 1 shows more ablation experiments of feature shrinkage decoder on COD10K and NC4K datasets. ① is similar to [13], which adjusts our decoder to pairwise feature aggregation with overlapping, and removes lateral supervision and feature interaction within the same layer. ② is a decoder that adds the lateral supervision and feature interaction within the same layer to ①. Our method differs from ② in that our method is pairwise feature aggregation without overlapping. Note that we retain other modules in these experiments.

By comparison of ① and ours, we can see that our decoder achieves superior performance over ①. In particular, our decoder significantly improves performance by 1.5%, 1.3%, 1.6% and 1.1% for $F_\beta^w$, $F_\beta^m$, $E_\phi^m$ and $E_\phi^x$, respectively, on the NC4K dataset. Although ① and ours both progressively aggregates adjacent features through a layer-by-layer shrinkage pyramid to to accumulate features for object prediction, our decoder introduces lateral supervision and feature flow within the same layer, which force the decoder to accumulate more critical camouflaged object cues for better object segmentation.

By comparison of ② and ours, we reduce the aggregation operations (*i.e.*, AIM) to alleviate the decoder structure by fusing adjacent features without overlapping. Specifically, ② contains 11 layers and 66 AIMs, while our decoder only contains 4 layers and 12 AIMs, which greatly reduces the computation. However, our decoder still achieves slightly better performance than ② with fewer aggregation operations. Experiments demonstrate the superior performance of the proposed FSD to other decoder structures.

## 2.6. Failure Cases

Although our proposed model achieves state-of-the-art performance, it does not detect camouflaged objects well in some very challenging scenes. As shown in Fig. 5, the results in the first and last three columns indicate the difficulty of detecting camouflaged objects under very low lighting conditions and a very similar appearance to the background,


Figure 1. Visual comparison with other competitors in detecting **small** camouflaged objects. Please zoom in for details.

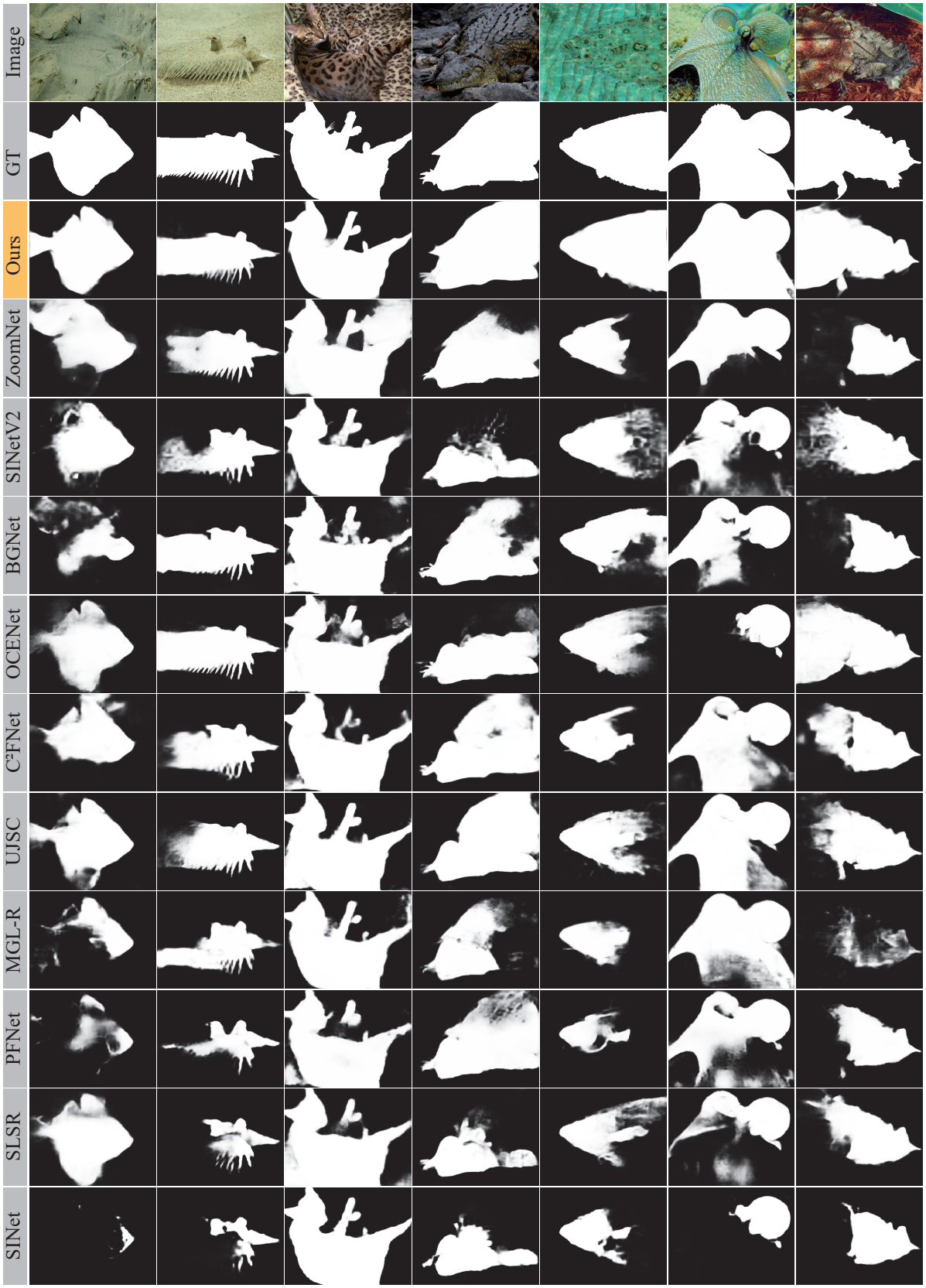


Figure 2. Visual comparison with other competitors in detecting **big** camouflaged objects. Please zoom in for details.


Figure 3. Visual comparison with other competitors in detecting **obscured** camouflaged objects. Please zoom in for details.


Figure 4. Visual comparison with other competitors in detecting camouflaged objects with **indistinguishable boundaries**. Please zoom in for details.

Table 2. Quantitative comparison with 23 SOTA methods on CAMO [8] dataset. Notes ↑ / ↓ denote the larger/smaller is better, respectively. The best and second best are **bolded** and underlined for highlighting, respectively.

| Methods | CAMO (250) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | $S_m \uparrow$ | $F_\beta^\omega \uparrow$ | $E_m^a \uparrow$ | $E_m^m \uparrow$ | $E_m^x \uparrow$ | $F_\beta^a \uparrow$ | $F_\beta^m \uparrow$ | $F_\beta^x \uparrow$ | $\mathcal{M} \downarrow$ |
| **Salient Object Detection** | | | | | | | | | |
| $BASNet_{19}$ | .618 | .413 | .719 | .661 | .708 | .525 | .475 | .519 | .159 |
| $CPD_{19}$ | .716 | .556 | .807 | .723 | .796 | .675 | .618 | .658 | .113 |
| $EGNet_{19}$ | .662 | .495 | .780 | .683 | .780 | .640 | .567 | .625 | .125 |
| $SCRN_{19}$ | .779 | .643 | .848 | .797 | .850 | .733 | .705 | .738 | .090 |
| $F^3Net_{20}$ | .711 | .564 | .802 | .741 | .780 | .661 | .616 | .630 | .109 |
| $CSNet_{20}$ | .771 | .642 | .847 | .795 | .849 | .730 | .705 | .740 | .092 |
| $SSAL_{20}$ | .644 | .493 | .765 | .721 | .780 | .605 | .579 | .601 | .126 |
| $ITSD_{20}$ | .750 | .610 | .830 | .780 | .830 | .692 | .663 | .694 | .102 |
| $UCNet_{20}$ | .739 | .640 | .811 | .787 | .820 | .716 | .700 | .708 | .094 |
| $VST_{21}$ | .787 | .691 | .866 | .838 | .866 | .746 | .738 | .756 | .076 |
| **Camouflaged Object Detection** | | | | | | | | | |
| $SINet_{20}$ | .751 | .606 | .834 | .771 | .831 | .709 | .675 | .706 | .100 |
| $SLSR_{21}$ | .787 | .696 | .855 | .838 | .854 | .756 | .744 | .753 | .080 |
| $PFNet_{21}$ | .782 | .695 | .852 | .842 | .855 | .751 | .746 | .758 | .085 |
| $MGL\text{-}R_{21}$ | .775 | .673 | .847 | .812 | .842 | .738 | .726 | .740 | .088 |
| $UJSC_{21}$ | .800 | .728 | .865 | .859 | .873 | .779 | .772 | .779 | .073 |
| $C^2FNet_{21}$ | .796 | .719 | .864 | .854 | .864 | .764 | .762 | .771 | .088 |
| $UGTR_{21}$ | .784 | .684 | .856 | .822 | .851 | .748 | .735 | .751 | .086 |
| $BSA\text{-}Net_{22}$ | .794 | .717 | .859 | .851 | .867 | .768 | .763 | .770 | .079 |
| $OCE\text{-}Net_{22}$ | .802 | .723 | .863 | .852 | .865 | .776 | .766 | .777 | .080 |
| $BGNet_{22}$ | .812 | .749 | .876 | .870 | .882 | .786 | .789 | .799 | .073 |
| $SegMaR_{22}$ | .815 | .753 | .872 | .874 | .884 | .795 | .795 | .803 | .071 |
| $ZoomNet_{22}$ | .820 | .752 | .878 | .878 | .892 | .792 | .794 | .805 | .066 |
| $SINet\text{-}v2_{22}$ | .820 | .743 | .875 | .882 | .895 | .779 | .782 | .801 | .070 |
| Ours | **.856** | **.799** | **.919** | **.899** | **.928** | **.829** | **.830** | **.846** | **.050** |

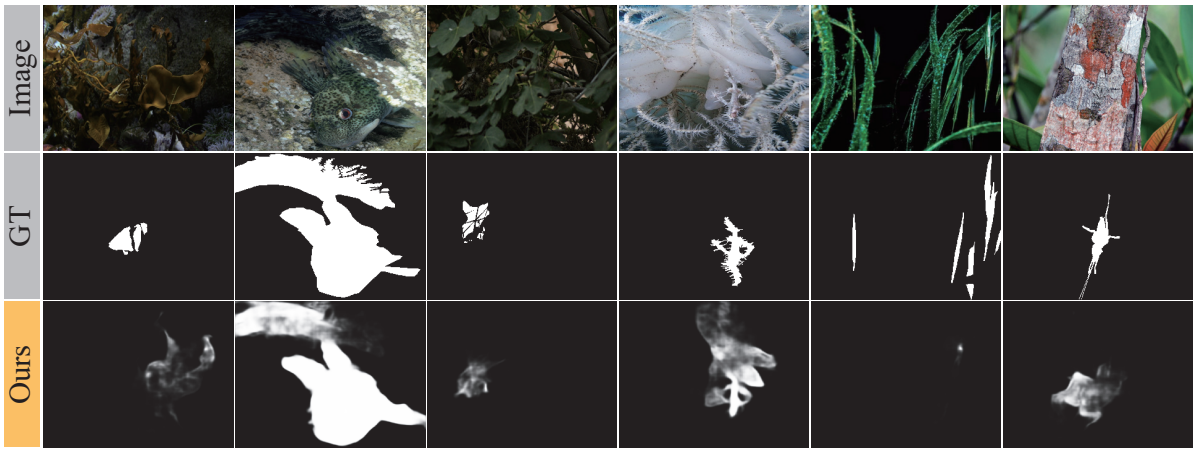


Figure 5. Failure cases in very challenging scenarios. Please zoom in for details.

respectively, which are potential directions for improvement in our future work.

## References

[1] Radhakrishna Achanta, Sheila Hemami, Francisco Estrada, and Sabine Susstrunk. Frequency-tuned salient region detec-

Table 3. Quantitative comparison with 23 SOTA methods on COD10K [5] and NC4K [12] datasets. Notes ↑ / ↓ denote the larger/smaller is better, respectively. "–" is not available. The best and second best are **bolded** and underlined for highlighting, respectively.

| Methods | COD10K (2,026) | | | | | | | | | NC4K (4,121) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $S_m\uparrow$ | $F_\beta^\omega\uparrow$ | $E_m^a\uparrow$ | $E_m^m\uparrow$ | $E_m^x\uparrow$ | $F_\beta^a\uparrow$ | $F_\beta^m\uparrow$ | $F_\beta^x\uparrow$ | $\mathcal{M}\downarrow$ | $S_m\uparrow$ | $F_\beta^\omega\uparrow$ | $E_m^a\uparrow$ | $E_m^m\uparrow$ | $E_m^x\uparrow$ | $F_\beta^a\uparrow$ | $F_\beta^m\uparrow$ | $F_\beta^x\uparrow$ | $\mathcal{M}\downarrow$ |
| **Salient Object Detection** | | | | | | | | | | | | | | | | | | |
| $BASNet_{19}$ | .634 | .365 | .676 | .678 | .735 | .421 | .417 | .451 | .105 | .695 | .546 | .784 | .762 | .786 | .618 | .610 | .617 | .095 |
| $CPD_{19}$ | .750 | .531 | .792 | .776 | .853 | .578 | .595 | .640 | .053 | .717 | .551 | .808 | .724 | .793 | .660 | .597 | .638 | .092 |
| $EGNet_{19}$ | .733 | .519 | .799 | .761 | .836 | .572 | .583 | .620 | .055 | .767 | .626 | .842 | .793 | .850 | .703 | .689 | .719 | .077 |
| $SCRN_{19}$ | .789 | .575 | .789 | .817 | .880 | .593 | .651 | .699 | .047 | .830 | .698 | .864 | .854 | .897 | .744 | .757 | .793 | .059 |
| $F^3Net_{20}$ | .739 | .544 | .818 | .795 | .819 | .588 | .593 | .609 | .051 | .780 | .656 | .853 | .824 | .848 | .710 | .705 | .719 | .070 |
| $CSNet_{20}$ | .778 | .569 | .791 | .810 | .871 | .589 | .635 | .679 | .047 | .750 | .603 | .812 | .773 | .793 | .672 | .655 | .669 | .088 |
| $SSAL_{20}$ | .668 | .454 | .782 | .768 | .789 | .529 | .527 | .535 | .066 | .699 | .561 | .805 | .780 | .812 | .653 | .644 | .654 | .093 |
| $ITSD_{20}$ | .767 | .557 | .787 | .808 | .861 | .573 | .615 | .658 | .051 | .811 | .680 | .855 | .845 | .883 | .717 | .729 | .762 | .064 |
| $UCNet_{20}$ | .776 | .633 | .867 | .857 | .867 | .673 | .681 | .691 | .042 | .811 | .729 | .883 | .871 | .886 | .776 | .775 | .782 | .055 |
| $VST_{21}$ | .781 | .604 | .837 | .837 | .877 | .620 | .653 | .682 | .042 | .831 | .732 | .887 | .877 | .901 | .758 | .771 | .792 | .050 |
| **Camouflaged Object Detection** | | | | | | | | | | | | | | | | | | |
| $SINet_{20}$ | .771 | .551 | .797 | .806 | .868 | .593 | .634 | .676 | .051 | .808 | .723 | .882 | .871 | .883 | .768 | .769 | .775 | .058 |
| $SLSR_{21}$ | .804 | .673 | .882 | .880 | .892 | .699 | .715 | .732 | .037 | .840 | .766 | .902 | .895 | .907 | .802 | .804 | .815 | .048 |
| $PFNet_{21}$ | .800 | .660 | .868 | .877 | .890 | .676 | .701 | .725 | .040 | .829 | .745 | .892 | .888 | .898 | .779 | .784 | .799 | .053 |
| $MGL\text{-}R_{21}$ | .814 | .666 | .865 | .852 | .890 | .681 | .711 | .738 | .035 | .833 | .740 | .889 | .867 | .893 | .778 | .782 | .800 | .052 |
| $UJSC_{21}$ | .809 | .684 | .882 | .884 | .891 | .705 | .721 | .738 | .035 | .842 | .771 | .903 | .898 | .907 | .803 | .806 | .816 | .047 |
| $C^2FNet_{21}$ | .813 | .686 | .886 | .890 | .900 | .703 | .723 | .743 | .036 | .838 | .762 | .898 | .897 | .904 | .788 | .795 | .810 | .049 |
| $UGTR_{21}$ | .817 | .666 | .850 | .853 | .890 | .671 | .712 | .741 | .036 | .839 | .747 | .886 | .875 | .899 | .778 | .787 | .807 | .052 |
| $BSA\text{-}Net_{22}$ | .818 | .699 | .894 | .891 | .901 | .723 | .738 | .753 | .034 | .841 | .771 | .903 | .897 | .907 | .805 | .808 | .817 | .048 |
| $OCE\text{-}Net_{22}$ | .827 | .707 | .883 | .894 | .905 | .718 | .741 | .764 | .033 | .853 | .785 | .904 | .903 | .913 | .812 | .818 | .831 | .045 |
| $BGNet_{22}$ | .831 | .722 | **.902** | **.901** | .911 | .739 | .753 | .774 | .033 | .851 | .788 | .911 | .907 | .916 | .813 | .820 | .833 | .044 |
| $SegMaR_{22}$ | .833 | .724 | .893 | .899 | .906 | .739 | .757 | .774 | .034 | – | – | – | – | – | – | – | – | – |
| $ZoomNet_{22}$ | .838 | .729 | .892 | .888 | .911 | **.741** | .766 | .780 | .029 | .853 | .784 | .904 | .896 | .912 | .814 | .818 | .828 | .043 |
| $SINet\text{-}v2_{22}$ | .815 | .680 | .863 | .887 | .906 | .682 | .718 | .752 | .037 | .847 | .770 | .898 | .903 | .914 | .792 | .805 | .823 | .048 |
| Ours | **.851** | **.735** | .900 | .895 | **.930** | .736 | **.769** | **.794** | **.026** | **.879** | **.816** | **.923** | **.915** | **.937** | **.826** | **.843** | **.859** | **.035** |


tion. In *CVPR*, pages 1597–1604. IEEE, 2009. 1

[2] 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. 1

[3] Deng-Ping Fan, Cheng Gong, Yang Cao, Bo Ren, Ming-Ming Cheng, and Ali Borji. Enhanced-alignment measure for binary foreground map evaluation. In *IJCAI*, 2018. 1

[4] Deng-Ping Fan, Ge-Peng Ji, Ming-Ming Cheng, and Ling Shao. Concealed object detection. *IEEE TPAMI*, 2022. 2

[5] Deng-Ping Fan, Ge-Peng Ji, Guolei Sun, Ming-Ming Cheng, Jianbing Shen, and Ling Shao. Camouflaged object detection. In *CVPR*, 2020. 1, 2, 8, 9

[6] Shang-Hua Gao, Yong-Qiang Tan, Ming-Ming Cheng, Chengze Lu, Yunpeng Chen, and Shuicheng Yan. Highly efficient salient object detection with 100k parameters. In *ECCV*, pages 702–721. Springer, 2020. 2

[7] 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. 2

[8] Trung-Nghia Le, Tam V Nguyen, Zhongliang Nie, Minh-Triet Tran, and Akihiro Sugimoto. Anabranch network for camouflaged object segmentation. *CVIU*, 184:45–56, 2019. 1, 7

[9] Aixuan Li, Jing Zhang, Yunqiu Lv, Bowen Liu, Tong Zhang, and Yuchao Dai. Uncertainty-aware joint salient object and camouflaged object detection. In *CVPR*, pages 10071–10081, 2021. 2

[10] Jiawei Liu, Jing Zhang, and Nick Barnes. Modeling aleatoric uncertainty for camouflaged object detection. In *WACV*, pages 1445–1454, 2022. 2

[11] Nian Liu, Ni Zhang, Kaiyuan Wan, Ling Shao, and Junwei Han. Visual saliency transformer. In *ICCV*, pages 4722–4732, 2021. 2

[12] 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. 1, 2, 8

[13] Mingcan Ma, Changqun Xia, Jia Li, et al. Pyramidal feature shrinking for salient object detection. In *AAAI*, volume 35, pages 2311–2318, 2021. 2

[14] Ran Margolin, Lihi Zelnik-Manor, and Ayellet Tal. How to evaluate foreground maps? In *CVPR*, pages 248–255, 2014. 1

[15] 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

Table 4. $S_m$ results for each sub-class on COD10K [5]. **Bolded** and underlined denote the best and second best scores.

| *Sub-class* | ITSD | UCNet | SINet | SLSR | PFNet | MGL-R | UJSC | C$^2$FNet | OCE-Net | ZoomNet | SINet-v2 | Ours |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Amphibian-Frog | 0.761 | 0.790 | 0.785 | 0.814 | 0.815 | 0.813 | 0.794 | 0.820 | 0.809 | **0.855** | 0.837 | 0.852 |
| Amphibian-Toad | 0.836 | 0.847 | 0.850 | 0.863 | 0.866 | 0.877 | 0.867 | 0.866 | 0.888 | 0.885 | 0.870 | **0.893** |
| Aquatic-BatFish | 0.813 | 0.820 | 0.836 | 0.809 | 0.867 | 0.786 | 0.817 | 0.834 | 0.906 | 0.890 | 0.873 | **0.907** |
| Aquatic-ClownFish | 0.692 | 0.707 | 0.693 | 0.784 | 0.805 | 0.608 | 0.709 | 0.737 | 0.771 | 0.813 | 0.787 | **0.851** |
| Aquatic-Crab | 0.791 | 0.788 | 0.804 | 0.817 | 0.805 | 0.839 | 0.811 | 0.828 | 0.839 | 0.836 | 0.815 | **0.864** |
| Aquatic-Crocodile | 0.752 | 0.815 | 0.761 | 0.783 | 0.753 | 0.785 | 0.808 | 0.817 | 0.843 | 0.829 | 0.825 | **0.857** |
| Aquatic-CrocodileFish | 0.689 | 0.785 | 0.734 | 0.791 | 0.780 | 0.815 | 0.751 | 0.764 | 0.738 | 0.805 | 0.746 | **0.846** |
| Aquatic-Fish | 0.780 | 0.791 | 0.767 | 0.816 | 0.799 | 0.821 | 0.821 | 0.818 | 0.835 | 0.841 | 0.834 | **0.854** |
| Aquatic-Flounder | 0.814 | 0.812 | 0.786 | 0.857 | 0.850 | 0.872 | 0.873 | 0.857 | 0.895 | 0.880 | 0.889 | **0.922** |
| Aquatic-FrogFish | 0.831 | 0.849 | 0.748 | 0.848 | 0.840 | 0.844 | 0.894 | 0.868 | 0.879 | **0.925** | 0.894 | **0.925** |
| Aquatic-GhostPipefish | 0.794 | 0.774 | 0.779 | 0.821 | 0.819 | 0.832 | 0.823 | 0.831 | 0.840 | 0.849 | 0.817 | **0.872** |
| Aquatic-LeafySeaDragon | 0.587 | 0.609 | 0.576 | 0.654 | 0.625 | 0.714 | 0.641 | 0.633 | 0.658 | 0.691 | 0.670 | **0.782** |
| Aquatic-Octopus | 0.846 | 0.827 | 0.843 | 0.873 | 0.863 | 0.895 | 0.865 | 0.869 | 0.897 | **0.889** | 0.887 | 0.885 |
| Aquatic-Pagurian | 0.654 | 0.624 | 0.643 | 0.682 | 0.640 | 0.683 | 0.679 | 0.650 | 0.646 | **0.724** | 0.698 | 0.710 |
| Aquatic-Pipefish | 0.718 | 0.755 | 0.726 | 0.782 | 0.774 | 0.805 | 0.780 | 0.792 | 0.810 | 0.807 | 0.781 | **0.828** |
| Aquatic-ScorpionFish | 0.779 | 0.731 | 0.740 | 0.799 | 0.773 | 0.753 | 0.806 | 0.778 | 0.815 | 0.834 | 0.808 | **0.851** |
| Aquatic-SeaHorse | 0.793 | 0.789 | 0.799 | 0.823 | 0.798 | 0.814 | 0.826 | 0.824 | 0.835 | 0.823 | 0.823 | **0.851** |
| Aquatic-Shrimp | 0.670 | 0.656 | 0.718 | 0.700 | 0.737 | 0.730 | 0.714 | 0.741 | 0.731 | 0.787 | 0.735 | **0.819** |
| Aquatic-Slug | 0.701 | 0.786 | 0.792 | 0.594 | **0.836** | 0.770 | 0.743 | 0.803 | 0.624 | 0.776 | 0.729 | 0.696 |
| Aquatic-StarFish | 0.797 | 0.876 | 0.799 | 0.856 | 0.852 | 0.846 | **0.903** | 0.892 | 0.868 | 0.892 | 0.890 | 0.889 |
| Aquatic-Stingaree | 0.779 | 0.669 | 0.724 | 0.789 | 0.785 | 0.747 | 0.781 | 0.791 | 0.815 | 0.818 | 0.815 | **0.881** |
| Aquatic-Turtle | 0.804 | 0.803 | 0.774 | 0.823 | 0.838 | 0.814 | 0.822 | 0.785 | 0.833 | **0.898** | 0.760 | 0.883 |
| Flying-Bat | 0.789 | 0.740 | 0.769 | 0.782 | 0.838 | 0.836 | 0.795 | 0.817 | 0.844 | 0.822 | 0.847 | **0.875** |
| Flying-Bee | 0.743 | 0.685 | 0.727 | 0.786 | 0.727 | 0.749 | 0.741 | 0.774 | 0.783 | 0.743 | **0.777** | 0.680 |
| Flying-Beetle | 0.916 | 0.917 | 0.911 | 0.931 | 0.903 | 0.821 | 0.923 | 0.922 | 0.926 | 0.931 | 0.903 | **0.932** |
| Flying-Bird | 0.796 | 0.815 | 0.807 | 0.830 | 0.826 | 0.841 | 0.836 | 0.846 | 0.851 | 0.867 | 0.835 | **0.873** |
| Flying-Bittern | 0.844 | 0.848 | 0.844 | 0.860 | 0.855 | 0.867 | 0.867 | 0.840 | 0.863 | **0.895** | 0.849 | 0.865 |
| Flying-Butterfly | 0.823 | 0.856 | 0.828 | 0.871 | 0.862 | 0.864 | 0.868 | 0.881 | 0.878 | 0.882 | 0.883 | **0.885** |
| Flying-Cicada | 0.834 | 0.843 | 0.851 | 0.875 | 0.887 | 0.891 | 0.886 | 0.875 | 0.900 | **0.916** | 0.883 | 0.909 |
| Flying-Dragonfly | 0.790 | 0.785 | 0.772 | 0.825 | 0.820 | 0.845 | 0.828 | 0.828 | 0.823 | 0.840 | 0.837 | **0.886** |
| Flying-Frogmouth | 0.929 | 0.932 | 0.896 | 0.941 | 0.945 | 0.936 | 0.932 | 0.936 | 0.942 | **0.961** | 0.941 | 0.947 |
| Flying-Grasshopper | 0.797 | 0.808 | 0.801 | 0.823 | 0.809 | 0.835 | 0.831 | 0.830 | 0.852 | 0.853 | 0.833 | **0.874** |
| Flying-Heron | 0.797 | 0.795 | 0.780 | 0.845 | 0.832 | 0.827 | 0.840 | 0.818 | 0.840 | **0.889** | 0.823 | 0.866 |
| Flying-Katydid | 0.760 | 0.768 | 0.771 | 0.797 | 0.798 | 0.833 | 0.811 | 0.824 | 0.840 | 0.847 | 0.809 | **0.853** |
| Flying-Mantis | 0.733 | 0.750 | 0.721 | 0.765 | 0.755 | 0.767 | 0.780 | 0.767 | 0.771 | 0.804 | 0.775 | **0.835** |
| Flying-Mockingbird | 0.767 | 0.795 | 0.735 | 0.834 | 0.800 | 0.796 | 0.832 | 0.843 | 0.821 | 0.863 | 0.838 | **0.875** |
| Flying-Moth | 0.838 | 0.890 | 0.863 | 0.894 | 0.893 | 0.877 | 0.908 | 0.901 | 0.929 | 0.916 | 0.917 | **0.941** |
| Flying-Owl | 0.805 | 0.834 | 0.832 | 0.861 | 0.855 | 0.865 | 0.851 | 0.868 | 0.890 | **0.896** | 0.868 | 0.876 |
| Flying-Owlfly | 0.837 | 0.782 | 0.772 | 0.851 | 0.752 | 0.845 | 0.866 | 0.861 | **0.886** | 0.879 | 0.863 | 0.872 |
| Other-Other | 0.753 | 0.755 | 0.737 | 0.807 | 0.764 | 0.812 | 0.809 | 0.790 | 0.812 | 0.846 | 0.779 | **0.899** |
| Terrestrial-Ant | 0.644 | 0.644 | 0.635 | 0.705 | 0.659 | 0.708 | 0.643 | 0.658 | 0.679 | 0.742 | 0.669 | **0.743** |
| Terrestrial-Bug | 0.845 | 0.820 | 0.844 | 0.828 | 0.840 | 0.872 | 0.843 | 0.848 | 0.897 | 0.865 | 0.856 | **0.874** |
| Terrestrial-Cat | 0.712 | 0.698 | 0.712 | 0.746 | 0.751 | 0.763 | 0.745 | 0.768 | 0.772 | 0.785 | 0.772 | **0.827** |
| Terrestrial-Caterpillar | 0.686 | 0.737 | 0.704 | 0.756 | 0.745 | 0.753 | 0.786 | 0.746 | 0.766 | 0.794 | 0.776 | **0.813** |
| Terrestrial-Centipede | 0.707 | 0.697 | 0.677 | 0.744 | 0.682 | 0.746 | 0.645 | 0.731 | 0.767 | 0.733 | 0.762 | **0.791** |
| Terrestrial-Chameleon | 0.765 | 0.746 | 0.759 | 0.786 | 0.776 | 0.834 | 0.814 | 0.810 | 0.803 | 0.833 | 0.804 | **0.845** |
| Terrestrial-Cheetah | 0.780 | 0.762 | 0.786 | 0.813 | 0.794 | 0.821 | 0.832 | 0.832 | 0.828 | 0.821 | 0.826 | **0.851** |
| Terrestrial-Deer | 0.691 | 0.711 | 0.701 | 0.738 | 0.737 | 0.751 | 0.761 | 0.760 | 0.757 | 0.791 | 0.757 | **0.798** |
| Terrestrial-Dog | 0.667 | 0.665 | 0.669 | 0.677 | 0.690 | 0.706 | 0.677 | 0.690 | 0.712 | 0.729 | 0.707 | **0.786** |
| Terrestrial-Duck | 0.694 | 0.694 | 0.693 | 0.709 | **0.800** | 0.732 | 0.710 | 0.747 | 0.750 | 0.726 | 0.746 | 0.784 |
| Terrestrial-Gecko | 0.789 | 0.831 | 0.825 | 0.861 | 0.830 | 0.815 | 0.868 | 0.867 | 0.879 | 0.856 | 0.848 | **0.908** |
| Terrestrial-Giraffe | 0.747 | 0.786 | 0.773 | 0.809 | 0.821 | 0.799 | 0.788 | 0.779 | 0.808 | 0.826 | 0.784 | **0.846** |
| Terrestrial-Grouse | 0.919 | 0.918 | 0.927 | 0.918 | 0.938 | 0.936 | 0.934 | 0.925 | 0.948 | 0.941 | 0.921 | **0.942** |
| Terrestrial-Human | 0.753 | 0.712 | 0.742 | 0.790 | 0.783 | 0.765 | 0.768 | 0.792 | **0.827** | 0.781 | 0.817 | 0.797 |
| Terrestrial-Kangaroo | 0.762 | 0.772 | 0.737 | 0.748 | 0.761 | 0.815 | 0.788 | 0.806 | 0.729 | 0.800 | **0.816** | 0.802 |
| Terrestrial-Leopard | 0.808 | 0.784 | 0.805 | 0.798 | 0.836 | 0.847 | 0.800 | 0.834 | 0.826 | 0.846 | 0.823 | **0.851** |
| Terrestrial-Lion | 0.773 | 0.807 | 0.761 | 0.832 | 0.818 | 0.814 | 0.815 | 0.833 | 0.843 | 0.814 | 0.813 | **0.859** |
| Terrestrial-Lizard | 0.786 | 0.804 | 0.800 | 0.820 | 0.819 | 0.829 | 0.830 | 0.823 | 0.838 | 0.852 | 0.830 | **0.853** |
| Terrestrial-Monkey | 0.829 | 0.644 | 0.675 | 0.808 | 0.720 | 0.855 | 0.877 | 0.835 | 0.797 | 0.898 | 0.888 | **0.913** |
| Terrestrial-Rabbit | 0.827 | 0.814 | 0.804 | 0.838 | 0.841 | 0.840 | 0.841 | 0.840 | 0.852 | 0.854 | 0.843 | **0.887** |
| Terrestrial-Reccoon | 0.756 | 0.790 | 0.738 | 0.748 | 0.774 | 0.619 | 0.780 | 0.788 | 0.801 | **0.837** | 0.766 | 0.791 |
| Terrestrial-Sciuridae | 0.804 | 0.811 | 0.821 | 0.798 | 0.831 | 0.841 | 0.838 | 0.852 | 0.844 | **0.897** | 0.842 | 0.856 |
| Terrestrial-Sheep | 0.487 | 0.754 | 0.490 | 0.750 | 0.582 | 0.565 | 0.561 | 0.686 | 0.540 | **0.761** | 0.500 | 0.493 |
| Terrestrial-Snake | 0.776 | 0.796 | 0.796 | 0.816 | 0.824 | 0.835 | 0.819 | 0.839 | 0.816 | **0.862** | 0.831 | 0.854 |
| Terrestrial-Spider | 0.716 | 0.724 | 0.736 | 0.757 | 0.759 | 0.782 | 0.766 | 0.769 | 0.793 | 0.802 | 0.771 | **0.808** |
| Terrestrial-StickInsect | 0.667 | 0.701 | 0.658 | 0.746 | 0.701 | 0.695 | 0.726 | 0.730 | 0.757 | 0.753 | 0.696 | **0.762** |
| Terrestrial-Tiger | 0.637 | 0.646 | 0.662 | 0.663 | 0.675 | 0.706 | 0.669 | 0.695 | 0.689 | 0.700 | 0.703 | **0.734** |
| Terrestrial-Wolf | 0.731 | 0.731 | 0.737 | 0.707 | 0.730 | 0.747 | 0.754 | 0.747 | 0.757 | **0.792** | 0.749 | 0.749 |
| Terrestrial-Worm | 0.659 | 0.710 | 0.682 | 0.739 | 0.792 | 0.766 | 0.804 | 0.691 | 0.782 | 0.808 | 0.806 | **0.812** |

[16] 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*, 2022. 2

[17] Xuebin Qin, Zichen Zhang, Chenyang Huang, Chao Gao, Masood Dehghan, and Martin Jagersand. Basnet: Boundary-aware salient object detection. In *CVPR*, pages 7479–7489, 2019. 2

[18] Yujia Sun, Geng Chen, Tao Zhou, Yi Zhang, and Nian Liu. Context-aware cross-level fusion network for camouflaged object detection. In *IJCAI*, pages 1025–1031, 2021. 2

[19] Yujia Sun, Shuo Wang, Chenglizhao Chen, and Tian-Zhu Xiang. Boundary-guided camouflaged object detection. In *IJCAI*, 2022. 2

[20] Jun Wei, Shuhui Wang, and Qingming Huang. $F^3$net: fusion, feedback and focus for salient object detection. In *AAAI*, pages 12321–12328, 2020. 2

[21] Zhe Wu, Li Su, and Qingming Huang. Cascaded partial decoder for fast and accurate salient object detection. In *CVPR*, pages 3907–3916, 2019. 2

[22] Zhe Wu, Li Su, and Qingming Huang. Stacked cross refinement network for edge-aware salient object detection. In *ICCV*, pages 7264–7273, 2019. 2

[23] 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. 2

[24] 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. 2

[25] Jing Zhang, Deng-Ping Fan, Yuchao Dai, Saeed Anwar, Fatemeh Sadat Saleh, Tong Zhang, and Nick Barnes. Uc-net: Uncertainty inspired rgb-d saliency detection via conditional variational autoencoders. In *CVPR*, pages 8582–8591, 2020. 2

[26] Jing Zhang, Xin Yu, Aixuan Li, Peipei Song, Bowen Liu, and Yuchao Dai. Weakly-supervised salient object detection via scribble annotations. In *CVPR*, pages 12546–12555, 2020. 2

[27] Jia-Xing Zhao, Jiang-Jiang Liu, Deng-Ping Fan, Yang Cao, Jufeng Yang, and Ming-Ming Cheng. Egnet: Edge guidance network for salient object detection. In *ICCV*, pages 8779–8788, 2019. 2

[28] Huajun Zhou, Xiaohua Xie, Jian-Huang Lai, Zixuan Chen, and Lingxiao Yang. Interactive two-stream decoder for accurate and fast saliency detection. In *CVPR*, pages 9141–9150, 2020. 2

[29] 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. 2