electronics

Article

Camouflaged Insect Segmentation Using a Progressive Refinement Network

Jing Wang [1,†], Minglin Hong [1,†], Xia Hu [2,†], Xiaolin Li [1,*], Shiguo Huang [1], Rong Wang [2] and Feiping Zhang [2]

[1] College of Computer and Information Sciences, Fujian Agriculture and Forestry University, Fuzhou 350002, China
[2] College of Forestry, Fujian Agriculture and Forestry University, Fuzhou 350002, China
* Correspondence: lixiaolin@fafu.edu.cn
† These authors contributed equally to this work.

Abstract: Accurately segmenting an insect from its original ecological image is the core technology restricting the accuracy and efficiency of automatic recognition. However, the performance of existing segmentation methods is unsatisfactory in insect images shot in wild backgrounds on account of challenges: various sizes, similar colors or textures to the surroundings, transparent body parts and vague outlines. These challenges of image segmentation are accentuated when dealing with camouflaged insects. Here, we developed an insect image segmentation method based on deep learning termed the progressive refinement network (PRNet), especially for camouflaged insects. Unlike existing insect segmentation methods, PRNet captures the possible scale and location of insects by extracting the contextual information of the image, and fuses comprehensive features to suppress distractors, thereby clearly segmenting insect outlines. Experimental results based on 1900 camouflaged insect images demonstrated that PRNet could effectively segment the camouflaged insects and achieved superior detection performance, with a mean absolute error of 3.2%, pixel-matching degree of 89.7%, structural similarity of 83.6%, and precision and recall error of 72%, which achieved improvements of 8.1%, 25.9%, 19.5%, and 35.8%, respectively, when compared to the recent salient object detection methods. As a foundational technology for insect detection, PRNet provides new opportunities for understanding insect camouflage, and also has the potential to lead to a step progress in the accuracy of the intelligent identification of general insects, and even being an ultimate insect detector.

Keywords: camouflaged insects; deep learning; insect detection; object segmentation; progressive refinement network

check for updates

Citation: Wang, J.; Hong, M.; Hu, X.; Li, X.; Huang, S.; Wang, R.; Zhang, F. Camouflaged Insect Segmentation Using a Progressive Refinement Network. *Electronics* **2023**, *12*, 804. https://doi.org/10.3390/electronics12040804

Academic Editors: Xiushan Nie, Guoqiang Zhong, Yongshun Gong, Bin Fan and Xin Li

Received: 5 January 2023
Revised: 27 January 2023
Accepted: 1 February 2023
Published: 6 February 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Segmenting an insect from its background is the necessary starting point for analyzing anything else about the insect. However, it is a challenging task to segment insects from complicated ecological images due to their various appearances, e.g., size, color and texture, even if they are of the same type [1]. Moreover, under natural selection, many insects have evolved an array of mechanisms that deceive the visual perceptual system of observers to avoid being detected [2]. In this regard, camouflaged insects perform best. However, their blurring of the boundary and lack of the intense contrast required for segmentation approaches further aggravate the difficulties of accurate insect detection [3].

Currently, a large number of image segmentations rely on extracted handcrafted features [4–10]. "Handcrafted" is a term in machine learning and computer vision referring to the application of some process, such as an algorithm or a manual procedure, to extract relevant features for identification from the raw data (image segmentation tasks) [1]. The handcrafted features are often low-level features, such as color, texture, shape, and appearance, which are sensitive to illumination changes, complicated backgrounds and different object positions [11,12]. Thus, these models tend to suffer from a high misdetection rate with regard to dealing with insect images shot in the wild.

Over the past few years, deep learning and convolutional neural networks (CNNs) have yielded a new generation of segmentation models [13–18]. Through the process of deep learning, CNNs extract the low-level features of the raw data via a series of nonlinear mappings and combine them into more abstract features (i.e., high-level features) [19]. These high-level features possess stronger representational ability than traditional handcrafted features and can express the comprehensive information of the original natural images. The full convolution neural network (FCN) [20] is the first method to employ fully convolutional networks for semantic segmentation. Subsequently, the improved methods based on FCN make great progress. Recent methods have achieved excellent results by using dilated convolution [21], deformable convolution [22], the attention mechanism [15,23], feature pyramid spatial pooling [24], and encoder–decoder structure [24,25]. Moreover, segmentation methods based on human hierarchy [26] or prototypes [27] have also received extensive attention. The results are significantly better than previous algorithms, and satisfactory results are achieved on object segmentation.

Methods for object segmentation based on deep learning show significant advantages; however, to distinguish the camouflaged insects from their backgrounds, three major challenges should be resolved: (1) Camouflaged insects have various shapes, scales, and positions, which are difficult to be accurately perceived from the whole images. (2) Due to similar texture or color, some noise from the multi-layer features extracted by CNNs is easily introduced to the fused features. (3) The boundaries of camouflaged insects are blurry, which interferes with segmentation refinement.

Specifically, camouflaged insects often have colors similar to their backgrounds, which makes them difficult to see. Bearing the colors and patterns similar to the background is most people's conception of camouflage, termed "background picturing" by [28], or "background matching" by [29]. According to [30], camouflage is the ability of prey (or predators) to prevent (or facilitate) predation by changing their features (surface luminance, body pattern, color, texture, or edges) as per that of the environment. Insects have evolved diverse camouflage strategies [31,32], such as background matching, disruptive coloration, transparency, and masquerade [33–35]. An insect with background-matching camouflage reproduces the same distribution of simple features as found in the background. Disruptive coloration works at a later stage in visual perception, and suppresses the grouping of simple features into the attributes that could potentially be recognized. Transparency allows light to pass through the whole body or body parts, such as the wings of dragonflies. Masquerade has its effect after perceptual segregation of an object, through mimicry of specific objects within the background.

Examples of camouflaged insects are shown in Figure 1a. The lichen katydid is good at disguising itself as a lichen to adapt to the environment. *Kallima inachus* tends to rest on dead vegetation and closes its wings to disguise itself as a withered leaf. The stick insect is shaped like a withered or bamboo stem and remains motionless to confuse predators. Phyllium is shaped like a leaf, with bite marks and veins that are similar to those of a leaf. *Biston betularia* often falls on lichens or stones that are similarly colored to itself. Its larvae can even hold its body in posture to perfectly resemble the shape of twigs and also toggle its body color. Obviously, compared with salient insects (Figure 1b) that are the most attention-seeking to observers, camouflaged insects can hardly be captured rapidly by the naked eye due to their high similarities to twigs, leaves, flowers and other complex backgrounds.

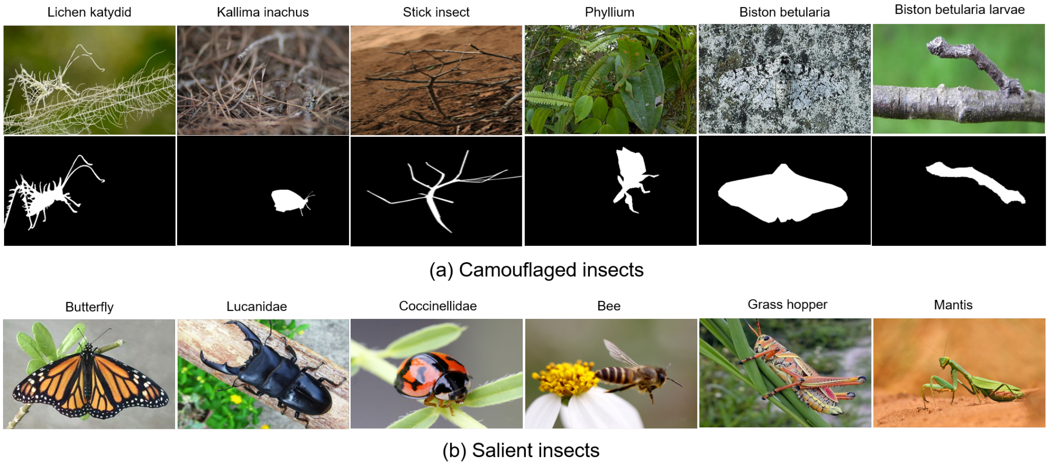

Figure 1. Examples of salient and camouflaged insects.

To cope with the above challenges, we developed three novel modules: the asymmetric receptive field (ARF) module, self-refinement module (SRM) and reverse guidance (RG) module. The ARF modules aim to alleviate the effect of various camouflaged insects and improve the ability to distinguish camouflaged insects from backgrounds, in which three parallel and asymmetric convolution layers with multiple receptive fields are adopted to extract anisotropic context information. SRM applies an initial attention strategy on the fused the multi-layer features to suppress distractors of the background to obtain an initial coarse segmentation map. For further outline refinement, RG modules are adopted to enhance the attention to boundaries by erasing the predicted foreground from the side output. The ARF, SRM and RG will further be integrated into the encoder–decoder framework, yielding the progressive refinement network (PRNet). PRNet is expected to improve recognition efficiency, promote camouflage research, and even improve the odds for the development of the ultimate insect detector.

2. Dataset

The COD10K [36] and CAMO [37] benchmarks are the two largest camouflage datasets, covering artificial camouflage, animal camouflage and insect camouflage. After discarding the duplicate insect images, we assembled a total of 1900 ecological images and ground truth pairs from the COD10K and CAMO for this study, which included 10 orders of typical camouflaged insects, such as Coleoptera, Hemiptera, Odonata, Neuroptera, Hymenoptera, Diptera, Lepidoptera, Phasmatodea, Mantodea, and Orthoptera (Figure 2). The number distributions of each order are 8, 130, 65, 26, 24, 4, 629, 127, 140, and 747, respectively.

Examples of some ecological image and ground truth pairs are shown in Figure 3. In each pair, the image is a color picture that contains foreground (objects) and background, and the ground truth is an objective and standard object-level annotation that indicates the location, scale, and shape of the objects in the image. In each image, there exists at least one camouflaged insect. Notably, there exist many challenging properties that were often encountered in real-world shooting, such as shape complexity, indefinable boundaries, occlusions, multiple insects, large insects, small insects, and being out of view. The property description is shown in Table 1.

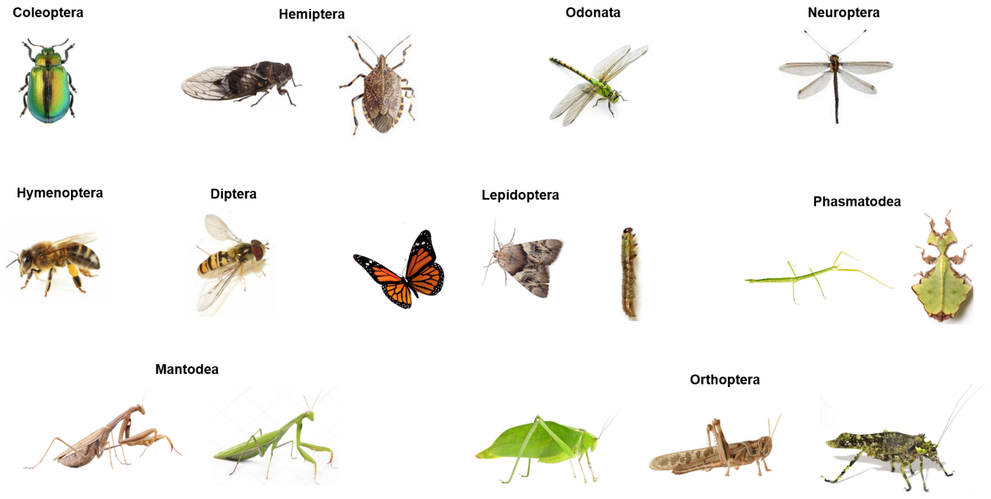

Figure 2. Specimens of typical camouflaged insects.

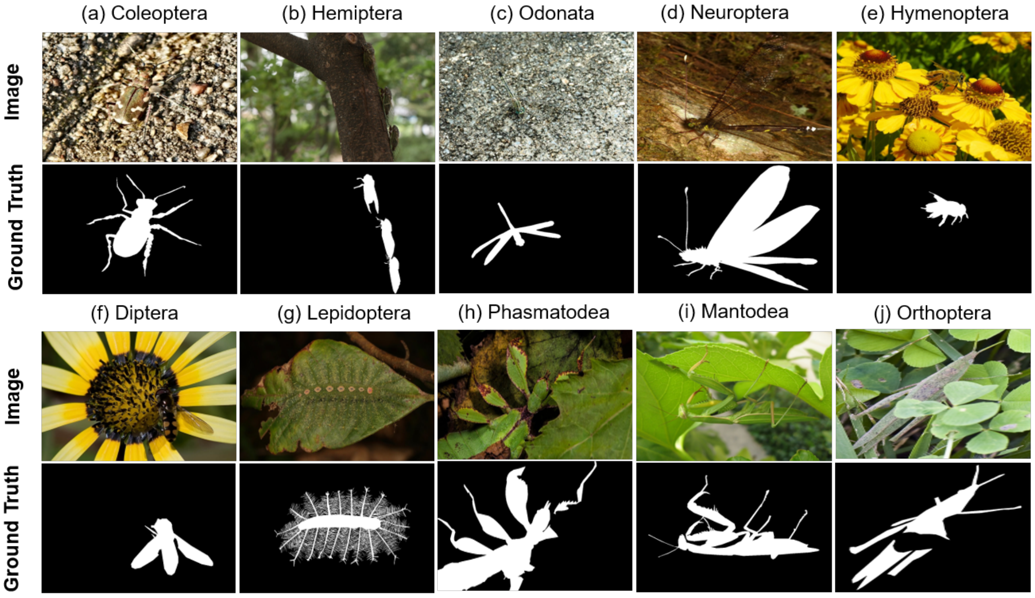

Figure 3. Examples of ecological image and ground truth pairs.

Table 1. Property descriptions.

Attribute	Description
Multiple Objects	Image contains at least two insects, e.g., Figure 3b
Small object	Ratio between insect area and whole image area is lower than 0.1., e.g., Figure 3e
Big Object	Ratio between insect area and whole image area is higher than 0.5., e.g., Figure 3d
Complex shape	Insect has thin parts and holes. e.g., Figure 3g
Indefinable boundaries	Insect has a similar color appearance, e.g., Figure 3c
Occlusion	Insect is partially occluded, e.g., Figure 3j
Out-of-View	Insect is clipped by image boundaries, e.g., Figure 3h

3. Method

To ‘see’ an object clearly, in general, visual stimuli received by our eyes are processed into neural signals by the visual nervous system and interpreted by visual centers in our brains. Researchers have proven that the information processing of the visual system is hierarchical in the cerebral cortex, and it is a constantly iterative and abstract process [38], as shown in Figure 4. According to this, the progressive refinement network (PRNet) was proposed in this paper. PRNet is composed of the following parts: (a) a backbone network for feature extraction; (b) three asymmetric receptive field (ARF) modules for deriving discriminative features, such as size and direction, (c) a self-refinement module (SRM), for abstracting outline and shape features, and (d) three reverse guidance (RG) modules for paying more attention to insect regions for making sharp outlines.

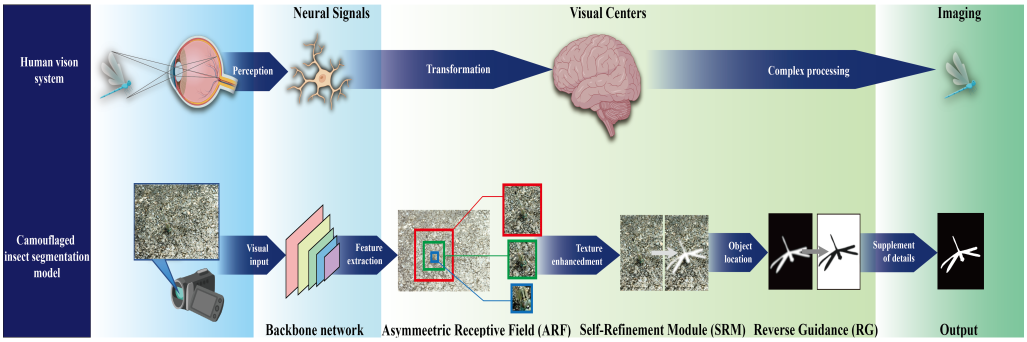

Figure 4. The process of PRNet for camouflaged insects, i.e., PRNet, consistent with the process of human vision.

Figure 5 shows the overall flowchart of our proposed model, which is built on an encoder–decoder architecture. Particularly, for the encoder, we use Res2Net [39] as the backbone, which is capable of capturing more multi-scale semantic features. We define the extracted multi-layer features as low-level features $\{X_0, X_1\}$ and high-level features X_2, X_3, X_4. The resolution of each feature X_k is $H/2^{k+1} \times W/2^{k+1}, k \in \{0, 1, 2, 3, 4\}$, covering diversified feature pyramids from high resolution, weak semantics to low resolution, strong semantics. Note that average pooling layers and the fully connected layers are removed to reduce the loss of details. For the decoder, we first introduce a novel ARF module, whose inputs are middle-level and high-level features, which is necessary to extract anisotropy contextual information in the horizontal, vertical and square kernel modes. Second, we propose a novel SRM with an initial attention strategy in order to aggregate the coarse information from the output of the proposed ARF module and to generate a refined camouflaged map based on the integrated features. Moreover, our SRM can be seen as a special self-attention module that makes full use of more fine-grained information. Finally, in the top-down decoding process, we use an RG module to supplement the missing regions and details of this refined map, which is essential to achieving a high-quality saliency map and clear boundaries. The details of the above are elaborated as follows.

3.1. Asymmetric Receptive Field

Since camouflaged insects often come from natural scenes, their sizes are varied and stochastic. Neuroscience experiments have verified that, in the human visual system, a set of various-sized receptive fields helps to highlight the area close to the retinal fovea, which is sensitive to small spatial information [40]. This motivated us to adopt receptive fields with various sizes to incorporate more discriminative camouflage cues after feature extraction. Additionally, the standard convolution operation of size $(2i-1) \times (2i-1)$ can be factorized as a sequence of two steps with $(2i-1) \times 1$ and $1 \times (2i-1)$ kernels, speeding up the inference efficiency without decreasing the representation capabilities [41]. As the $1 \times k$ and $k \times 1$ layers have non-square kernels, they are referred to as the asymmetric convolutional layers. Therefore, we proposed the asymmetric receptive field (ARF), which includes five branches $\{b_k, k = 1, \ldots, 5\}$. In each branch, the first convolutional (Bconv) layer had dimensions of 1×1 to reduce the channel size to 32. This was followed by two other layers: a $(2k-1) \times (2k-1)$ Bconv layer and a 3×3 Bconv layer with a specific dilation rate $(2k-1)$ when $k > 2$. The first four branches were concatenated, and then their channel size was reduced to 32 with a 1 × 1 Bconv operation. Finally, the 5th branch was added, and the whole module was fed to a ReLU [42] function to obtain the feature $\{ef_k, k = 2,3,4\}$. In brief, compared to the standard receptive field block structure [40], ARF added one more branch with a larger dilation rate to enlarge the receptive field and further replaced the standard convolution with two asymmetric convolutional layers. By using the ARF modules, comprehensive information with integrated anisotropy context from three levels was generated, and the approximate scales of insects in images could be acquired.

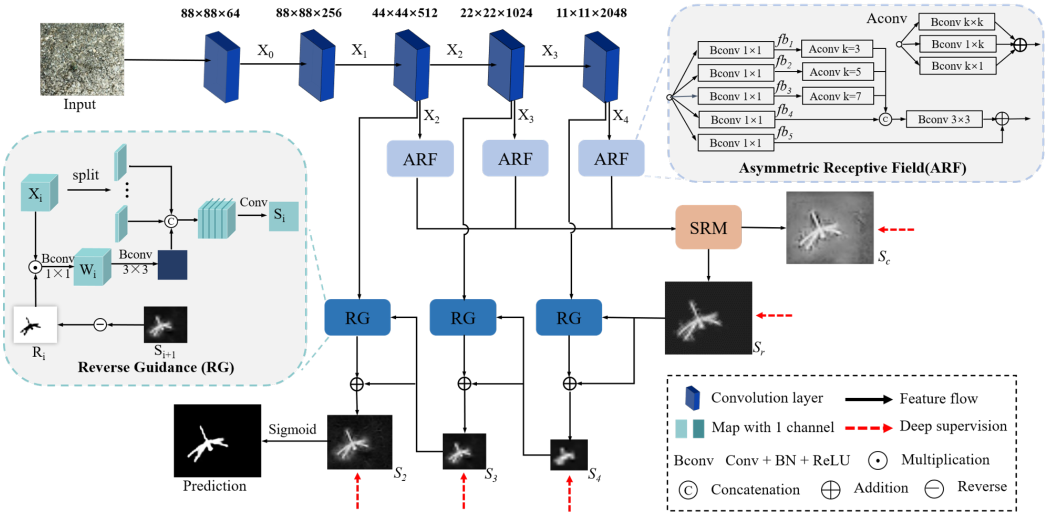

Figure 5. Overview of the proposed network, which consists of three main modules: asymmetric receptive field (ARF) module, self-refinement module (SRM) and reverse guidance (RG) module (Best viewed in color).

3.2. Self-Refinement Module

After obtaining the informative features from the previous ARF modules, in the SRM, we need to preliminarily segment the camouflaged object. Due to repeated down-sampling operations, such as pooling and convolution, the resolution of the prediction map is greatly reduced, which leads to blurred insect outlines. According to the recent evidence [43], high-level features have more global semantic information, which helps to differentiate which features are camouflaged objects in an image. However, due to the lack of details, the camouflage regions are blurred. In contrast, low-level features have detailed information, but it is difficult to determine the camouflage regions. To accurately locate camouflaged insects and obtain sharper outlines simultaneously, it is necessary to integrate multi-level features together. However, feature fusion across multiple levels easily introduces redundant information, resulting in the inaccurate location of targets. Therefore, it is necessary to reduce the differences between the three coarse features. To this end, we used the partial decoder component (PDC) [44] to extract the fusion features that contained high-level information. Such features could then be directly used to generate a coarse camouflaged map S_c, via a simple convolutional operation. However, due to the low resolution, the coarse map generated by the fusion features was far from the ground truth mask. Inspired by [45], we formulate the generated coarse camouflaged map as an attention mask, where such an attention mechanism can help denoise the features and generate a preliminary coarse segmentation map. In detail, the inputs of this SRM are the three coarse features $\{ef_i, i = 4,3,2\}$. We then use a PDC to integrate these features and use a 3 × 3 Bconv and a 1 × 1 Bconv to extract a coarse camouflaged map S_c as follows:

$$f_d = P_d(ef_4, ef_3, ef_2), \tag{1}$$

$$S_c = \text{iBconv}(f_d), \tag{2}$$

where P_d uses multiplication and concatenation to gradually decrease the gap between different features, and iBconv denotes multiple Bconv operators. To generate a more accurate camouflage map, we multiply this map S_c with the discriminative features to obtain a discriminative feature f_r, which can be described as follows:

$$f_r = f_d \odot S_c, \tag{3}$$

where $\odot$ denotes element-wise multiplying. Finally, we also use a 3 × 3 Bconv and a 1 × 1 Bconv to extract a fine-grained camouflaged map S_r.

3.3. Reverse Guidance Module

The human vision system pays great attention to the outline of the object, and often obtains the specific shape of the target object through the edge information and interprets the target object and so on. However, a key factor for insect camouflage is edge disruption [46–48]. As previously described above, the coarse camouflage map S_c was derived from the three highest layers, which could only capture a probable location of the camouflaged insect, ignoring boundary details. Moreover, direct up-sampling could further introduce more noise and make the boundary non-smooth. To this end, the RG module, which erases the predicted foreground from side-output, is proposed to refine such missing parts or details in the high-level prediction and apply residual architecture to further refine the predicted camouflaged map. As shown in Figure 5, the RG module aims to generate the corresponding edge attention map W_i by using a reverse attention map R_i. We further split the feature X_i with C channels into n groups (the number of channels in each group is c), and concatenate it with n reverse attention maps R_i so as to guide the features to focus on boundaries. To obtain a more complete camouflaged map, we iteratively add the predicted result of the latter layer S_{i+1} to the corresponding edge attention map W_i, which can be described as follows:

$$W_i = R_i \odot X_i, \tag{4}$$

$$x_i^1, ..., x_i^m, ..., x_i^n = \mathrm{split}(X_i), \tag{5}$$

$$F_i = \mathrm{concat}(x_i^1, W_i, ..., x_i^m, W_i, ..., x_i^n, W_i), \tag{6}$$

$$S_i = \mathrm{iBconv}(F_i) + S_{i+1} \tag{7}$$

Note that this reverse attention map R_i is obtained by erasing the foreground in the prediction, and it can be formulated as

$$R_i = 1 - \sigma(U(S_{i+1})), \tag{8}$$

where σ is the sigmoid function, and U is the up-sampling operation. Having access to high-level information, lower levels can learn more powerful features for refining the camouflage map in a coarse-to-fine manner, and output the final segmentation result, which is a binary bitmap.

In short, ARF captures contextual information from multi-layer features for coarse-grained refinement of fusion features. For fine-grained refinement, SRM and RG modules cover more useful information by applying an initial attention strategy on fusion features, and erasing the foreground to pay more attention to boundaries, respectively. These three modules progressively refine features from coarse to fine so as to achieve accurate segmentation. Finally, we integrate the ARF, SRM and RG into the encoder–decoder architecture, and the entire network can be trained end to end.

3.4. Implementation and Evaluation

To evaluate the performance of PRNet, 5-fold cross validation was adopted in this research. The camouflaged insect dataset was randomly split into 1520 images (80%) for training and 380 images (20%) for validation in each fold. The proposed method was implemented on the PyTorch platform. We used the Adam optimizer with a learning rate 1×10^{-4} for training, in which the epoch size was 40, and the batch size was 16. The encoder was initialized by the weights of Res2Net that is pre-trained on ImageNet. We resized all input images to 352 × 352 and used images with three scale rates $\{0.75, 1, 1.25\}$ during training.

PRNet is a supervised segmentation network to predict each pixel to be the insect or background, thus it is trained by minimizing the pixel position-aware (PPA) loss [49] of camouflaged maps $\{S_2, S_3, S_4, S_c, S_r\}$. PPA loss assigns different weights to different positions and pays more attention to hard pixels. PPA loss L_{ppa} is formulated as

$$L_{ppa} = L_{wbce} + L_{wiou}, \tag{9}$$

where L_{wbce} is a weighted binary cross entropy loss and L_{wiou} is a weighted intersection over Union (IoU) loss. The L_{wbce} loss function is formed as the following:

$$L_{wbce} = -\frac{1}{N}\frac{\sum_{i,j}(1+\gamma\alpha_{ij})\left[g_{ij}\log(p_{ij}) + (1-g_{ij})\log(1-p_{ij})\right]}{\sum_{i,j}\gamma\alpha_{ij}}, \tag{10}$$

where g_{ij} and p_{ij} represent the predicted values and ground truth of the pixel at location (i,j), respectively. N denotes the total number of pixels in an image, and γ is a hyperparameter. The weight α is calculated according to the difference between the center pixel and its surroundings, which can be defined as follows:

$$\alpha_{ij} = \left|\frac{\sum_{m,n\in A_{ij}} g_{mn}}{\sum_{m,n\in A_{ij}} 1} - g_{ij}\right|, \tag{11}$$

where A_{ij} is the area surrounding the pixel (i,j). If a_{ij} is large, the pixel at (i,j) is very different from its surroundings, which may represent an important pixel (e.g., edge) and deserves more attention. Similarly, α is assigned to L_{wiou} for emphasizing the importance of hard pixels, which can be defined as

$$L_{wiou} = 1 - \frac{1}{N}\frac{\sum_{i,j}(1+\gamma\alpha_{ij})g_{ij}p_{ij}}{\sum_{i,j}(1+\gamma\alpha_{ij})(g_{ij}+p_{ij}-g_{ij}p_{ij})} \tag{12}$$

In this paper, all of the output segmentation maps are upsampled to the same size as the ground truth G. Thus, the total loss can be defined as

$$L_{total} = L\left(\boldsymbol{G}, \sigma\left(\boldsymbol{R}_g^{\text{up}}\right)\right) + \sum_{t=1}^{5} L\left(\boldsymbol{G}, \sigma\left(\boldsymbol{R}_t^{\text{up}}\right)\right) \tag{13}$$

4. Results

4.1. Evaluation Metrics

Four popular metrics were utilized to evaluate its performance in camouflaged insect segmentation, i.e., mean absolute error (MAE) [50], enhanced-alignment measure (E_Φ) [51], structural similarity measure (S_α) [52], and weighted harmonic mean of precision and recall (F_β^w) [53]. For the predicted camouflage map C and its corresponding ground truth G, the image size is $W \times H$.

MAE calculates the pixel-wise absolute difference between the predicted camouflage map and ground truth, which is defined as

$$MAE = \frac{1}{W \times H}\sum_{i=1}^{W}\sum_{j=1}^{H}|C(i,j) - G(i,j)|, \tag{14}$$

where H and W are the height and the width of the map, respectively.

E_{Φ} is defined as

$$E_{\Phi} = \frac{1}{W \times H} \sum_{i=1}^{W} \sum_{j=1}^{H} \Phi_{FM}(i, j), \tag{15}$$

which can obtain image-level statistics and local pixel-matching information, based on the enhanced alignment matrix Φ_{FM}.

S_{α} uses the object-aware and region-aware structure similarities, namely S_o and S_r, to obtain structural similarity between the predicted camouflage map and ground truth, and is formulated as

$$S_{\alpha} = \alpha * S_o + (1 - \alpha) * S_r, \tag{16}$$

where α = 0.5.

F_{β}^{w} takes both the weighted precision ($Precision^{w}$) and weighted recall ($Recall^{w}$) into account, which is formulated as

$$F_{\beta}^{w} = (1 + \beta^2) \frac{Precision^{w} \cdot Recall^{w}}{\beta^2 \cdot Precision^{w} + Recall^{w}}, \tag{17}$$

where β^2 = 0.3.

4.2. Comparisons with State-of-the-Art Detection Methods

Our proposed PRNet was compared with seven recently published methods, including two detection methods designed for salient objects, i.e., BASNet [54] and F3Net [49], and five detection methods designed for camouflaged objects, i.e., SINet [44], PraNet [12], SINet-V2 [36], PFNet [55] and C2FNet [3]. For fair comparison, all results of these models were retrained with our camouflaged insect dataset, and the parameters were set as recommended in the corresponding papers. The open source codes were taken from http://dpfan.net/camouflage/ (accessed on 10 September 2022) and https://github.com/jiwei0921/SOD-CNNs-based-code-summary- (accessed on 10 September 2022). In addition, all the prediction maps are evaluated with the same code.

Quantitative Evaluation: The quantitative results of our PRNet against seven other object detection methods on the camouflaged insect dataset are presented in Table 2. Obviously, the performance of salient object detection methods, both BASNet and F3Net was not as good as that of all camouflaged object detection methods. Moreover, most models that employed Res2Net-50 as the backbone outperformed the models that employed ResNet-50. In terms of camouflaged object detection methods, PRNet was superior to PraNet and SINet which were also built on a customized UNet-based architecture. In particular, our method achieved the best performance on camouflaged insect detection with an *MAE* of 3.2%, E_{Φ} of 89.7%, S_{α} of 83.6% and F_{β}^{w} of 72%, which were improved by 0.1%, 0.4%, 0.3% and 0.1%, respectively, when compared to the second-best method.

Table 2. Quantitative comparison with state-of-the-art methods for camouflaged insect detection. ↑ / ↓ indicates that larger or smaller is better. The three best results are in red, green and blue colors, respectively.

Method	**Year**	**Field**	**Backbone**	*MAE* ↓	E_{Φ} ↑	S_{α} ↑	F_{β}^{w} ↑
BASNet [54]	2019	Camouflage	ResNet-50	0.068	0.770	0.692	0.483
F3Net [49]	2020	Salient	ResNet-50	0.113	0.638	0.641	0.362
SINet [44]	2020	Camouflage	ResNet-50	0.050	0.872	0.803	0.567
PraNet [12]	2020	Camouflage	Res2Net-50	0.040	0.857	0.801	0.662
SINet-V2 [36]	2022	Camouflage	Res2Net-50	0.033	0.889	0.836	0.719
PFNet [55]	2021	Camouflage	ResNet-50	0.033	0.893	0.83	0.715
C2FNet [3]	2021	Camouflage	Res2Net-50	0.035	0.887	0.833	0.712
PRNet	2021	Camouflage	Res2Net-50	0.032	0.897	0.836	0.720

Qualitative Evaluation: Furthermore, some visualization results are shown in Figure 6. Compared with other state-of-the-art models, our model achieved better visual effects by detecting more accurate and complete camouflaged objects with rich details. BASNet failed to detect some insects (see the 1st, 2nd and 3rd (last) rows). All boundaries of insects detected by F3Net were blurred. In particular, SINet and PraNet were weak at detecting the thin parts, such as the antennas (see 8th row) and feet (see 1st row). SINet-V2, PFNet and C2FNet interfused the noise from background in the last row and the last 4th row, in which the shapes of insects were complex.

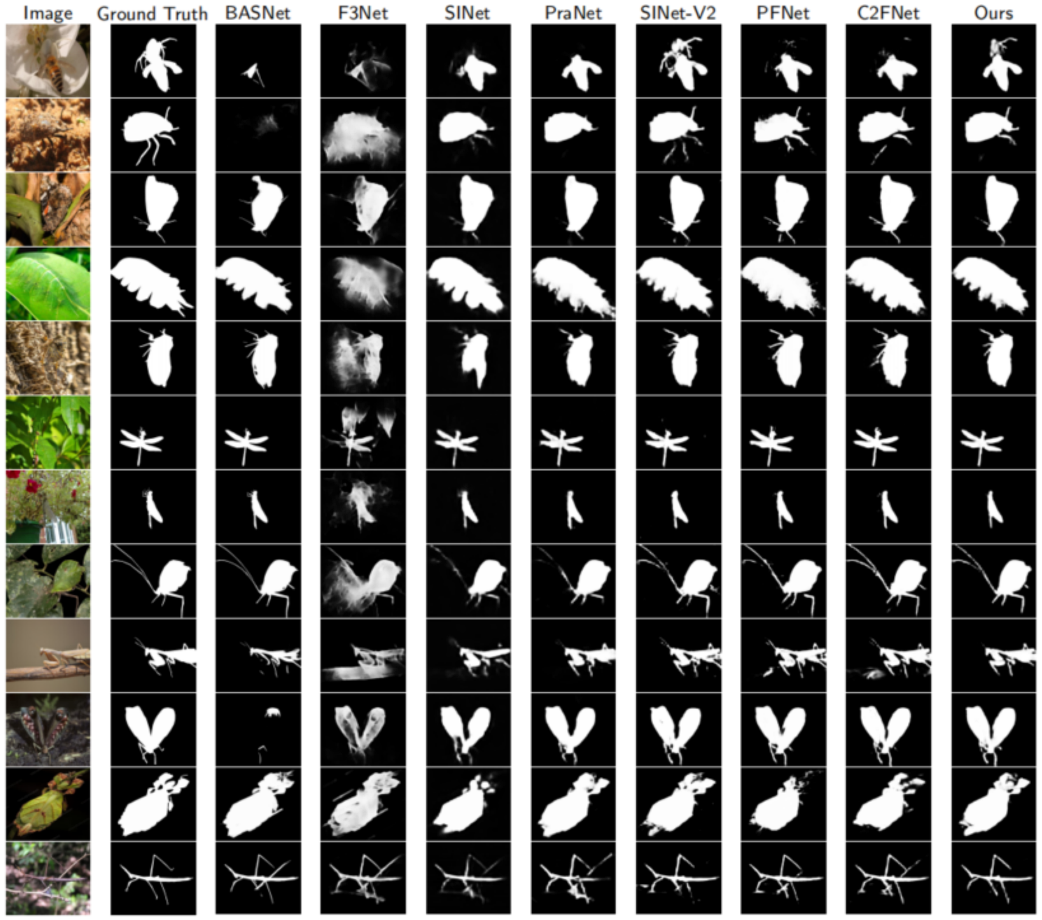

Figure 6. Visual comparison of different methods (Best viewed in color).

4.3. Effectiveness Verification of Each Module

To verify the effectiveness of each key module, they were removed from the complete model for contrast tests. Note that three tests were performed under the same training settings, and the results are summarized in Table 3. In the No.1 (SRM+RG) experiment, we replaced the ARF module with a convolution operation to adjust the features of the last three layers to a consistent number of channels. In the No.2 (ARF+RG) experiment, we removed the SRM module while keeping the ARF and RG modules, and then fused the features of the last three layers to generate an attention map via the combined operation of up-sampling and multiplication. In the No.3 (ARF+SRM) experiment, we replaced the RG with the combined operation of up-sampling and concatenation, and the ARF and SRM remained unchanged. No.4 (ARF+SRM+RG) was the complete model, consistent with the structure shown in Figure 5.

Table 3. Results of the effectiveness verification. The best results are highlighted in **bold** fonts.

No.	Module ARF	 SRM	 RG	*MAE* ↓	E_Φ ↑	S_α ↑	F_β^w ↑
1		√	√	0.033	0.89	0.834	0.715
2	√		√	0.040	0.897	0.831	0.657
3	√	√		0.161	0.830	0.672	0.326
4	√	√	√	**0.032**	**0.897**	**0.836**	**0.720**

Compared with the complete model (No.4), removing the ARF module (No.1) slightly degraded the detection performance. The reason for this phenomenon was the fact that the asymmetric receptive field brought the semantic strength and location accuracy of high-level features into full play but also inevitably introduced noise and fuzzy edges for the target object. When removing the SRM, i.e., No.2, the performance declined in several respects as well, which suggested that the lack of details went against extracting coarse camouflage maps. In particular, comparing No.3 with No.4, the use of RG significantly improved by 12.9%, 6.7%, 16.4%, and 39.4% in terms of *MAE*, E_Φ, S_α, and F_β^w, respectively. This result demonstrated that employing the RG module to mine edge information could help the model overcome the challenges in camouflaged object detection, such as complex shapes and indefinable boundaries. Thus, the contrast tests validated the rationality of integrating these modules into PRNet.

4.4. Generalization Verification

The generalization ability of the neural network model is of vital importance for its popularization and application. In other words, a good deep learning model should not only perform well on the existing images in academic datasets, but also perform well on fresh images. To further assess the generalization capability of the proposed PRNet, some camouflaged images from CHAMELEON. were directly fed into PRNet without retraining or any tune-up.

On the dataset CHAMELEON, our PRNet achieved the performance with an *MAE* of 3.8%, E_Φ of 84.1%, S_α of 83.1% and F_β^w of 65.7%. According to the visual results shown in Figure 7, despite the outlines of segmentation maps being vague, PRNet could segment most parts of camouflaged objects, indicating that PRNet might provide more chances for the discovery and detection of various camouflaged objects. For further accurate segmentation performance on various camouflaged objects, retraining the PRNet on a large number of images of relevant objects is the most practical and effective method.

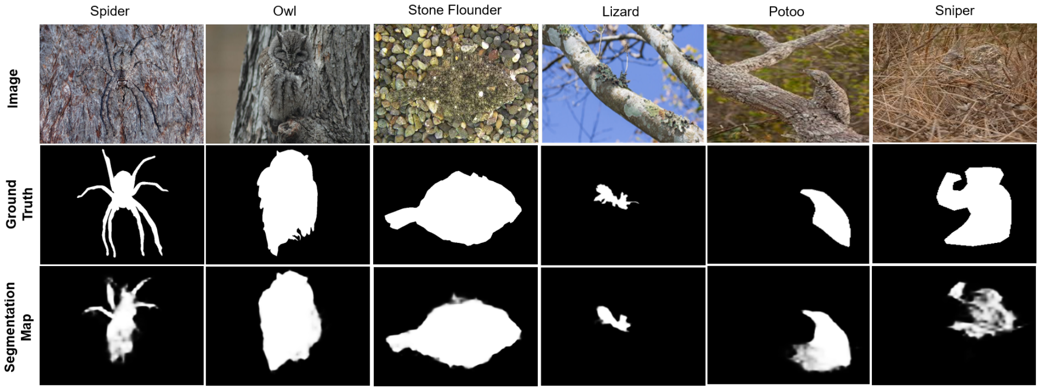

Figure 7. Generalization verification results of the proposed PRNet (Best viewed in color).

5. Discussion

This work proposed an end-to-end network PRNet that achieved the best performance on camouflaged insect detection compared with seven state-of-the-art detection methods designed for salient objects and camouflaged objects. PRNet integrated three key modules: ARF, SRM, and RG. The ARF module captured rich contextual information on the camouflaged images. SRM obtained a coarse segmentation map by suppressing distractors of the background. The RG module increased attention to insect edges by integrating reverse attention into multi-layer features.

As one of the toughest cases, camouflaged insect detection was often accompanied by weak boundaries, low contrast, and similar texture to the backgrounds, and thus, there were more challenges for camouflaged insect detection than salient insect or generic insect detection. Note that generic insects could be either salient or camouflaged, and camouflaged insects could be seen as difficult cases. Multi-level feature aggregation was explored for robust detection [56,57]. Recurrent and iterative learning strategies were also employed to progressively refine the prediction map [58,59]. Due to the effectiveness of feature enhancement, attention mechanisms [60,61] have also been applied to saliency detection [62,63]. In addition, edge cues were leveraged to refine the saliency map [54,64]. However, directly applying salient insect or generic insect detection methods to detect camouflaged insects may not yield the desired results, as the term "salient" was essentially the opposite of "camouflaged", i.e., standout versus immersion. This view was underpinned by quantitative and qualitative results. The saliency methods that achieved superior performance on salient object detection, e.g., BASNet [54] and F3Net [49], were inapplicable to camouflaged insect detection since they highlighted the most attention-seeking part of an image and discarded the seemingly unimportant pixels.

The recently proposed camouflaged object segmentation approaches achieved performance improvement to some extent; however, their performance degraded significantly for a number of challenging cases, such as complex shape, small size, being out-of-view, etc. As shown in Figure 6, PraNet [12] and PFNet [55] missed some parts of the insect body in the cases of insects with indefinable boundaries or thin parts. As a human visual system, a set of various-sized receptive fields helped to highlight the area close to the retinal fovea that is sensitive to details [40], which was overlooked by PraNet and PFNet. In addition, cross-level feature fusion also played a vital role in the success of camouflaged object detection. The ARF and SRM modules jointly considered both rich context information and effective cross-level feature fusion, yielding superior detection performance. Nevertheless, SINet [44] and C2FNet [3] utilized region and boundary information simultaneously, but the relationship between them was not fully captured, hence failing to correctly identify camouflage regions and interfusing the noise from the background when the camouflaged insect was oversize or out of view. The region and boundary were two key characteristics that distinguished camouflaged insects and backgrounds. After aggregating features in high levels and then predicting coarse regions, SINet-V2 [36] and our PRNet leveraged a set of recurrent reverse attention modules to establish the relationship between the region and boundary cues, which enabled the models to calibrate some misaligned predictions.

Our results went beyond recent studies in terms of both quantitative evaluation and visualization evaluation. The verification of module rationality, as well as the model generalization, were further done in the previous section. Extensive experiments showed that our PRNet outperforms the state-of-the-art methods on both quantitative and qualitative evaluations, which made it possible to become an ultimate universal insect detector for not only the camouflaged insects, but also the salient or generic insects.

Despite the superior performance of PRNet, due to the limited sample size and long-tailed distribution of challenging cases (see Figure 8), such as small insects and multiple insects, the dataset studied in this work is unable to take full advantage of deep learning models. The effect of deep learning depends on a large amount of data to some extent, and the emergence of new datasets will lead to rapid progress in computer vision [65–68]. With this in mind, we encourage researchers to construct a camouflaged insect dataset of large size and with high-quality annotation. Moreover, some measures should be taken in the future to address the challenging problem caused by the imbalance of samples of the camouflage dataset. The performance of small insect segmentation can also be improved by data augmentation in the training set, such as oversampling those images with small insects and augmenting each of those images by copy-pasting small objects many times [69]. In addition, generative adversarial networks (GANs) [70,71] can also alleviate the problem arising from a few discriminative features of the small objects. GANs can be used to generate a super-resolved representation of a small object that is very similar to a large object. The super-resolved feature is superimposed on the feature map of the original small target so as to enhance the feature expression to improve the segmentation performance of the small object. In addition to the small object problem, multiple insects and the congestion or occlusion between multiple insects will lead to the loss of feature information, causing false or missed segmentation. The short-term transformer block [72], lifted edges [73] and repulsion loss [74] might be effective in addressing these constraints. Additionally, our model shows highly correlated statistical dependencies between the predictions and inputs, but datasets may have different distributions in different domains. If the new domain were directly input to validate the model, the prediction results on the new dataset CHAMELEON might be inaccurate [75]. For example, the spider's feet cannot be completely distinguished from the background in Figure 7. Therefore, transfer learning is encouraged here as well to reduce the distribution difference between the new domain and the original domain.

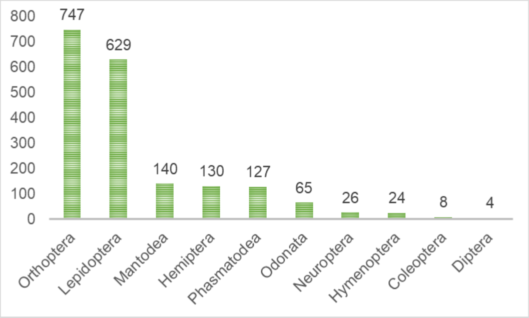

Figure 8. Number distribution of each order in our dataset.

6. Conclusions

In the present paper, we addressed the problem of segmenting insect objects. Camouflaged insects were the focus due to their extreme features in solving the detection problem. This was the first attempt and achieved success in applying deep learning techniques to camouflaged insect image detection. The main ingredients of our approach are the ARF, SRM and RG modules. ARF formulates a novel means of extracting contextual information that perceives the varied appearances of camouflaged insects. SRM fuses comprehensive features to suppress the distractors of the background exhibiting similar colors or textures to camouflaged insects. The refinement of insect outlines using the proposed RG module yields more robust segmentation results than recent approaches for the same task. Our method reaches state-of-the-art detection accuracy on 1900 images of camouflaged insects (MAE = 3.2%), approaching the performance of human experts. In future work, we will extend our camouflaged insect dataset as well as having high-quality annotations, and new techniques, such as weakly supervised learning, zero-shot learning, transfer learning, and multi-scale backbone, could also be explored.

Author Contributions: Conceptualization, X.L.; Data curation, X.L., R.W. and F.Z.; Methodology, M.H.; Project administration, S.H.; Supervision, S.H.; Validation, X.H.; Writing—original draft, J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Science Foundation of Fujian Province, China (2021J01125), Forestry Science and Technology Projects in Fujian Province, China (Memorandums [2021]35), and the Fund for Excellent Dissertations at Fujian Agriculture and Forestry University (1122YS01004).

Data Availability Statement: The COD10K dataset can be downloaded at https://mmcheng.net/cod/ (accessed on 10 September 2022). The CAMO dataset can be downloaded at https://sites.google.com/view/ltnghia/research/camo (accessed on 10 September 2022). The CHAMELEON dataset can be downloaded at https://github.com/lartpang/awesome-segmentation-saliency-dataset#chameleon (accessed on 10 September 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Valan, M.; Makonyi, K.; Maki, A.; Vondráček, D.; Ronquist, F. Automated Taxonomic Identification of Insects with Expert-Level Accuracy Using Effective Feature Transfer from Convolutional Networks. *Syst. Biol.* **2019**, *68*, 876–895. [CrossRef] [PubMed]
2. Stevens, M.; Merilaita, S. Animal camouflage: Current issues and new perspectives. *Philos. Trans. R. Soc. B Biol. Sci.* **2009**, *364*, 423–427. [CrossRef] [PubMed]
3. Sun, Y.; Chen, G.; Zhou, T.; Zhang, Y.; Liu, N. Context-aware Cross-level Fusion Network for Camouflaged Object Detection. In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 19–27 August 2021; pp. 1025–1031. [CrossRef]
4. Mele, K. Insect Soup Challenge: Segmentation, Counting, and Simple Classification. In Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops, Sydney, Australia, 1–8 December 2013; pp. 168–171. [CrossRef]
5. Deyi, M.; Yimin, C.; Qiming, L.; Chen, H.; Sheng, X. Region Growing by Exemplar-Based Hand Segmentation under Complex Backgrounds. *Int. J. Adv. Comput. Technol.* **2012**, *4*, 432–437. [CrossRef]
6. Wu, C.; Zhang, X. Total Bregman divergencebased fuzzy local information Cmeans clustering for robust image segmentation. *Appl. Soft Comput.* **2020**, *94*, 106468. [CrossRef]
7. Zhang, J.; Kong, F.; Wu, J.; Han, S.; Zhai, Z. Automatic image segmentation method for cotton leaves with disease under natural environment. *J. Integr. Agric.* **2018**, *17*, 1800–1814. [CrossRef]
8. Shajahan, S.; Sivarajan, S.; Maharlooei, M.; Bajwa, S.; Harmon, J.; Nowatzki, J.; Igathinathane, C. Identification and Counting of Soybean Aphids from Digital Images Using Shape Classification. *Trans. Am. Soc. Agric. Biol. Eng.* **2017**, *60*, 1467–1477. [CrossRef]
9. Zhang, P.; Li, C. Region-based color image segmentation of fishes with complex background in water. *IEEE Int. Conf. Comput. Sci. Autom. Eng.* **2011**, *1*, 596–600. [CrossRef]
10. Wang, Z.; Wang, K.; Liu, Z.; Wang, X.; Pan, S. A Cognitive Vision Method for Insect Pest Image Segmentation. *IFAC-PapersOnLine* **2018**, *51*, 85–89. [CrossRef]
11. Tang, H.; Wang, B.; Chen, X. Deep learning techniques for automatic butterfly segmentation in ecological images. *Comput. Electron. Agric.* **2020**, *178*, 105739. [CrossRef]
12. Fan, D.; Ji, G.; Zhou, T.; Chen, G.; Fu, H.; Shen, J.; Shao, L. PraNet: Parallel Reverse Attention Network for Polyp Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Lima, Peru, 4–8 October 2020; pp. 263–273. [CrossRef]
13. Weinstein, B. A computer vision for animal ecology. *J. Anim. Ecol.* **2017**, *87*, 533–545. [CrossRef]
14. Liu, Y.; Chen, K.; Liu, C.; Qin, Z.; Luo, Z.; Wang, J. Structured Knowledge Distillation for Semantic Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 2599–2608. [CrossRef]
15. Fu, J.; Liu, J.; Tian, H.; Li, Y.; Bao, Y.; Fang, Z.; Lu, H. Dual Attention Network for Scene Segmentation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3141–3149. [CrossRef]
16. Li, Y.; Chen, X.; Zhu, Z.; Xie, L.; Huang, G.; Du, D.; Wang, X. Attention-Guided Unified Network for Panoptic Segmentation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7019–7028. [CrossRef]
17. He, J.; Deng, Z.; Zhou, L.; Wang, Y.; Qiao, Y. Adaptive Pyramid Context Network for Semantic Segmentation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7511–7520. [CrossRef]
18. Zhou, T.; Li, J.; Wang, S.; Tao, R.; Shen, J. MATNet: Motion-Attentive Transition Network for Zero-Shot Video Object Segmentation. *IEEE Trans. Image Process.* **2020**, *29*, 8326–8338. [CrossRef]
19. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
20. Shelhamer, E.; Long, J.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 640–651. [CrossRef] [PubMed]
21. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 833–851.
22. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable Convolutional Networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 764–773.

23. Huang, Z.; Wang, X.; Huang, L.; Huang, C.; Wei, Y.; Liu, W. CCNet: Criss-Cross Attention for Semantic Segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019.
24. Liu, D.; Cui, Y.; Tan, W.; Chen, Y. SG-Net: Spatial Granularity Network for One-Stage Video Instance Segmentation. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual Conference, 19–25 June 2021; pp. 9811–9820. [CrossRef]
25. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015, Munich, Germany, 5–9 October 2015; pp. 234–241. [CrossRef]
26. Li, L.; Zhou, T.; Wang, W.; Li, J.; Yang, Y. Deep Hierarchical Semantic Segmentation. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 1236–1247. [CrossRef]
27. Zhou, T.; Wang, W.; Konukoglu, E.; Van Goo, L. Rethinking Semantic Segmentation: A Prototype View. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 2572–2583. [CrossRef]
28. Cuthill, I.C. Camouflage. *J. Zool.* **2019**, *308*, 75–92. [CrossRef]
29. Merilaita, S.; Stevens, M. Crypsis through background matching. In *Animal Camouflage: Mechanisms and Function*; Cambridge University Press: Cambridge, UK, 2011; pp. 17–33. [CrossRef]
30. Mondal, A. Camouflaged Object Detection and Tracking: A Survey. *Int. J. Image Graph.* **2020**, *20*, 2050028. [CrossRef]
31. Stevens, M.; Ruxton, G.D. The key role of behaviour in animal camouflage. *Biol. Rev. Camb. Philos. Soc.* **2019**, *94*, 116–134. [CrossRef]
32. Merilaita, S.; Scott-Samuel, N.; Cuthill, I. How camouflage works. *Philos. Trans. R. Soc. B Biol. Sci.* **2017**, *372*, 20160341. [CrossRef]
33. Théry, M.; Gomez, D. Chapter 7—Insect Colours and Visual Appearance in the Eyes of Their Predators. In *Advances in Insect Physiology: Insect Integument and Colour*; Academic Press: Cambridge, MA, USA, 2010; Volume 38, pp. 267–353. [CrossRef]
34. Cuthill, I.C.; Allen, W.L.; Arbuckle, K.; Caspers, B.; Chaplin, G.; Hauber, M.E.; Hill, G.E.; Jablonski, N.G.; Jiggins, C.D.; Kelber, A.; et al. The biology of color. *Science* **2017**, *357*, eaan0221. [CrossRef]
35. Cuthill, I.C.; Matchette, S.R.; Scott-Samuel, N.E. Camouflage in a dynamic world. *Curr. Opin. Behav. Sci.* **2019**, *30*, 109–115. [CrossRef]
36. Fan, D.P.; Ji, G.P.; Cheng, M.M.; Shao, L. Concealed Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 6024–6042. [CrossRef]
37. Le, T.; Nguyen, T.V.; Nie, Z.; Tran, M.T.; Sugimoto, A. Anabranch network for camouflaged object segmentation. *Comput. Vis. Image Underst.* **2019**, *184*, 45–56. [CrossRef]
38. Hubel, D.H.; Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.* **1962**, *160*, 106. [CrossRef] [PubMed]
39. Gao, S.; Cheng, M.; Zhao, K.; Zhang, X.; Yang, M.; Torr, P. Res2Net: A New Multi-Scale Backbone Architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 652–662. [CrossRef] [PubMed]
40. Liu, S.; Huang, D.; Wang, Y. Receptive Field Block Net for Accurate and Fast Object Detection. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 404–419. [CrossRef]
41. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826. [CrossRef]
42. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 807–814. [CrossRef]
43. Zhao, T.; Wu, X. Pyramid Feature Attention Network for Saliency Detection. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3080–3089. [CrossRef]
44. Fan, D.; Ji, G.; Sun, G.; Cheng, M.; Shen, J.; Shao, L. Camouflaged Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 2774–2784. [CrossRef]
45. Song, K.; Huang, Q.; Zhang, F.-e.; Lu, J. Coarse-to-fine: A dual-view attention network for click-through rate prediction. *Knowl.-Based Syst.* **2021**, *216*, 106767. [CrossRef]
46. Kang, C.; Stevens, M.; Moon, J.; Lee, S.; Jablonski, P.G. Camouflage through behavior in moths: The role of background matching and disruptive coloration. *Behav. Ecol.* **2014**, *26*, 45–54. [CrossRef]
47. Webster, R.J. Does disruptive camouflage conceal edges and features? *Curr. Zool.* **2015**, *61*, 708–717. [CrossRef]
48. Webster, R.J.; Hassall, C.; Herdman, C.M.; Godin, J.G.J.; Sherratt, T.N. Disruptive camouflage impairs object recognition. *Biol. Lett.* **2013**, *9*, 20130501. [CrossRef]
49. Wei, J.; Shuhui Wang, Q.H. F3Net: Fusion, Feedback and Focus for Salient Object Detection. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 12321–12328. [CrossRef]
50. Hornung, A.; Pritch, Y.; Krahenbuhl, P.; Perazzi, F. Saliency filters: Contrast-based filtering for salient region detection. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 733–740. [CrossRef]

51. Fan, D.; Gong, C.; Cao, Y.; Ren, B.; Cheng, M.M.; Borji, A. Enhanced-alignment Measure for Binary Foreground Map Evaluation. In Proceedings of the International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 13–19 July 2018; pp. 698–704. [CrossRef]
52. Fan, D.; Cheng, M.; Liu, Y.; Li, T.; Borji, A. Structure-Measure: A New Way to Evaluate Foreground Maps. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4548–4557. [CrossRef]
53. Margolin, R.; ZelnikManor, L.; Tal, A. How to Evaluate Foreground Maps? In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 248–255. [CrossRef]
54. Qin, X.; Zhang, Z.; Huang, C.; Gao, C.; Dehghan, M.; Jagersand, M. BASNet: Boundary-Aware Salient Object Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 7471–7481. [CrossRef]
55. Mei, H.; Ji, G.; Wei, Z.; Yang, X.; Wei, X.; Fan, D. Camouflaged Object Segmentation With Distraction Mining. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual Conference, 19–25 June 2021; pp. 8772–8781. [CrossRef]
56. Lee, G.; Tai, Y.; Kim, J. Deep Saliency with Encoded Low Level Distance Map and High Level Features. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 660–668. [CrossRef]
57. Zhao, J.; Liu, J.; Fan, D.; Cao, Y.; Yang, J.; Cheng, M. EGNet: Edge Guidance Network for Salient Object Detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 8778–8787. [CrossRef]
58. Zhang, X.; Wang, T.; Qi, J.; Lu, H.; Wang, G. Progressive Attention Guided Recurrent Network for Salient Object Detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 714–722. [CrossRef]
59. Wang, W.; Shen, J.; Cheng, M.M.; Shao, L. An Iterative and Cooperative Top-Down and Bottom-Up Inference Network for Salient Object Detection. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5961–5970. [CrossRef]
60. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.U.; Polosukhin, I. Attention is All you Need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30. . [CrossRef]
61. Woo, S.; Park, J.; Lee, J.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 3–19. [CrossRef]
62. Chen, S.; Tan, X.; Wang, B.; Hu, X. Reverse Attention for Salient Object Detection. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 236–252. [CrossRef]
63. Liu, N.; Han, J.; Yang, M. PiCANet: Learning Pixel-Wise Contextual Attention for Saliency Detection. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3089–3098. [CrossRef]
64. Su, J.; Li, J.; Zhang, Y.; Xia, C.; Tian, Y. Selectivity or Invariance: Boundary-Aware Salient Object Detection. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 3798–3807. [CrossRef]
65. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3213–3223. [CrossRef]
66. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
67. Johannes, A.; Picon, A.; Alvarez-Gila, A.; Echazarra, J.; Rodriguez-Vaamonde, S.; Navajas, A.D.; Ortiz-Barredo, A. Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case. *Comput. Electron. Agric.* **2017**, *138*, 200–209. [CrossRef]
68. Neuhold, G.; Ollmann, T.; Bulò, S.R.; Kontschieder, P. The Mapillary Vistas Dataset for Semantic Understanding of Street Scenes. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4990–4999. [CrossRef]
69. Kisantal, M.; Wojna, Z.; Murawski, J.; Naruniec, J.; Cho, K. Augmentation for small object detection. In Proceedings of the 9th International Conference on Advances in Computing and Information Technology, Sydney, Australia, 21–22 December 2019; pp. 119–133. [CrossRef]
70. Li, J.; Liang, X.; Wei, Y.; Xu, T.; Feng, J.; Yan, S. Perceptual Generative Adversarial Networks for Small Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1951–1959. [CrossRef]
71. Talas, L.; Fennell, J.; Kjernsmo, K.; Cuthill, I.; Scott-Samuel, N.; Baddeley, R. CamoGAN: Evolving optimum camouflage with Generative Adversarial Networks. *Methods Ecol. Evol.* **2020**, *11*, 240–247. [CrossRef]
72. Wang, Y.; Xu, Z.; Wang, X.; Shen, C.; Cheng, B.; Shen, H.; Xia, H. End-to-End Video Instance Segmentation with Transformers. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual Conference, 19–25 June 2021; pp. 8737–8746. [CrossRef]

73. Tang, S.; Andriluka, M.; Andres, B.; Schiele, B. Multiple People Tracking by Lifted Multicut and Person Re-identification. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3701–3710. [CrossRef]
74. Wang, X.; Xiao, T.; Jiang, Y.; Shao, S.; Sun, J.; Shen, C. Repulsion Loss: Detecting Pedestrians in a Crowd. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 7774–7783. [CrossRef]
75. Lv, F.; Liang, J.; Li, S.; Zang, B.; Liu, C.H.; Wang, Z.; Liu, D. Causality Inspired Representation Learning for Domain Generalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 8046–8056. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.