

融合注意力机制的弱监督迷彩伪装目标检测算法

杨 辉¹, 权冀川¹, 梁新宇¹, 郭安文¹, 王中伟²

(1. 陆军工程大学 指挥控制工程学院, 江苏 南京 210007; 2. 中国人民解放军 73658 部队)

摘要: 随着计算机硬件和人工智能技术的发展, 强监督目标检测算法已经取得了很大的成果。然而, 强监督目标检测算法需要在大规模、标注精度高的数据集上进行训练。但在某些特定领域, 上述条件要求过于苛刻。例如, 军事上常用的迷彩伪装目标的图像数据集就比公共数据集更难获得且标注难度更大。因此, 采用对数据集要求更低的弱监督目标检测算法来实现迷彩伪装目标的检测任务。由于图像中迷彩伪装目标与背景融合度较大, 导致原始浅层特征感知伪监督目标定位(Shallow feature-aware Pseudo supervised Object Localization, SPOL)算法的检测精度相对较低。本文的核心是在 SPOL 算法的基础上融合注意力机制, 通过加入注意力模块, 让模型更加关注迷彩伪装目标的区域, 以此来提高迷彩伪装目标的检测精度。

关键词: 目标检测; 弱监督算法; 注意力机制; 迷彩伪装目标

中图分类号: TP183

文献标识码: A

DOI: 10.20044/j.csdg.2097-1788.2022.03.013

引用格式: 杨辉, 权冀川, 梁新宇, 等. 融合注意力机制的弱监督迷彩伪装目标检测算法[J]. 网络安全与数据治理, 2022, 41(3): 81-91.

Weakly supervised camouflage object detection algorithm fused with attention mechanism

Yang Hui¹, Quan Jichuan¹, Liang Xinyu¹, Guo Anwen¹, Wang Zhongwei²

(1. Command & Control Engineering College, Army Engineering University of PLA, Nanjing 210007, China;

2. Unit 73658 of PLA, China)

Abstract: With the development of computer hardware and artificial intelligence technology, strongly supervised object detection algorithms have achieved great results. However, strongly supervised object detection algorithms need to be trained on large-scale datasets with high annotation accuracy. But in some specific fields, the above conditions are too demanding. For example, image datasets of camouflage objects commonly used in military are more difficult to be obtained and labeled than public datasets. Therefore, this paper adopted a weakly supervised algorithm with lower requirements for datasets to detect the camouflage objects in images. Due to the high degree of fusion between the camouflage objects and the image background, the detection accuracy of the original SPOL(Shallow feature-aware Pseudo supervised Object Localization) algorithm was relatively low. The core of this paper was to integrate the attention mechanism into the SPOL algorithm. After added an attention module, the model could pay more attention to the area with camouflage objects. So, the detection accuracy of the camouflage objects was improved.

Key words: object detection; weakly supervised algorithm; attention mechanism; camouflage object

0 引言

军事上采用迷彩伪装的目的是隐蔽自己、欺骗敌人、提高战场生存能力。相对于通用的目标检测任务, 图像中的迷彩伪装目标与背景环境融合度较大, 实现其目标检测任务更加困难。

目前, 对迷彩伪装目标检测研究的工作较少。

传统的检测方法主要把迷彩伪装目标看作是一种具有特殊纹理结构的目标, 并针对这一特性设计相应算法提取迷彩纹理, 从而实现迷彩伪装目标的检测。Bhajantri 等人^[1]将目标的迷彩伪装纹理作为一类物体, 然后对该类物体进行检测。Sengottuvelan 等人^[2]通过图像的结构信息, 确定图像中是否存在迷

彩伪装目标。Wu 等人^[3]根据目标在三维凸面上的灰度差异来检测迷彩伪装目标。尽管传统方法可以实现对迷彩目标的检测,但该类方法仅利用了图像的浅层特征信息,其检测效果相对较差。

近年来的研究工作主要是使用基于深度卷积神经网络的强监督目标检测算法完成迷彩伪装目标的检测任务。Deng 等人^[4]针对迷彩伪装目标的特性,在 RetinaNet^[5]算法的基础上嵌入了空间注意力和通道注意力模块。同时,基于定位置信得分构建了新的预测框过滤算法,有效实现了对迷彩伪装人员的检测。Wang 等人^[6]以 YOLO(You Only Look Once)v5 算法为基础,设计了一种针对迷彩伪装目标的检测算法,该算法在骨干网络中加入了注意力机制,同时加入非对称卷积模块增强了目标的语义信息,从而提升了迷彩伪装目标的检测精度。虽然强监督目标检测算法比传统方法的检测效果有了很大的提升,但该类算法模型需要在大规模标注精度高的数据集上进行训练,检测结果严重依赖于数据集标注的精度。目前的数据集标注工作主要是靠人工完成,而人工标注在很大程度上容易受人的主观因素影响,在军事应用领域很难获得大规模的且标注精度高的数据集。

在军事领域,受保密等特殊条件限制,很难构建包含迷彩伪装目标的大规模图片数据集。并且,图片中的迷彩伪装目标与图片背景的融合度较大,从本质上增加了目标检测的难度。同时在人工标注时也很容易造成误标或漏标,严重影响数据集的使用效果。若在小规模且标注精度低的数据集上训练强监督目标检测算法,则训练出来的模型对迷彩伪装目标的检测效果会很不理想。而弱监督目标检测算法可以很好地克服强监督目标检测算法的这一局限性。弱监督目标检测算法只需要带有图像级标签(不需要标注出目标在图像中的具体位置,只需要标明图像中包含物体的类别)的数据集就能实现目标检测,大幅降低了对数据集标注的要求。因此,弱监督目标检测算法比强监督目标检测算法具有更强的适应能力。

目前,基于类激活图^[7](Class Active Mapping, CAM)的模型是弱监督目标检测算法(Weakly Supervised Object Detection, WSOD)中最常用的模型之一。然而,基于 CAM 的模型最初是针对分类任务进行训练的,其设计目标与检测算法不一致。具体来说,分类

任务更加关注来卷积神经网络(Convolutional Neural Networks, CNN)深层的具有语义意义的特征。相比之下,源自 CNN 浅层的特征包含的语义信息较少,但语义更加丰富。在细节上,浅层特征边缘更清晰,失真更少。但由于以下两个缺陷,浅层和深层特征的直接融合对于 WSOD 是无效的:(1)由于监督信息不足,嵌入在浅层特征中的有意义的信息容易受到背景噪声干扰;(2)在原始区域中只有最具有辨别力的区域被激活。

为了克服上述缺陷,本文采用了一种简单但有效的浅层特征感知伪监督目标定位^[8](Shallow feature-aware Pseudo supervised Object Localization, SPOL)算法来实现迷彩伪装目标检测任务。该算法最大限度地嵌入浅层的低级特征,生成的 CAM 区域更加广泛和清晰,在公共数据集中取得了最好的结果,如表 1 所示。

表 1 SPOL 算法与其他主流算法
在公共数据集上的实验对比

算法模型	CUB-200 数据集		ImageNet-1K 数据集	
	Top-1	Loc	Top-1	Loc
ADL ^[9]	62.29		48.53	
I2C ^[10]	65.99		53.11	
PSOL ^[11]	77.44		58.00	
SPOL ^[8]	80.12		59.14	

然而迷彩伪装目标与背景相似度大,导致 SPOL 算法在该类目标上的检测精度较低。针对这一问题,本文分别融合了卷积块注意力模块(Convolutional Block Attention Module, CBAM)^[12]和 SE(Squeeze-and-Excitation)^[13]注意力模块,使模型更加关注迷彩伪装目标所在的区域,以此提高模型提取特征的能力。实验表明,加入 CBAM 模块后,算法的检测精度提高了 4.29%;加入 SE 模块后,算法的检测精度提高了 8.01%;同时加入 CBAM 模块和 SE 模块后,该指标提高了 7.23%,从而验证了本文算法在弱监督迷彩伪装目标检测任务中的有效性。

1 SPOL 算法结构

SPOL 算法模型主要包括两个阶段,CAM 生成和与类别无关的分割,其算法流程如图 1 所示(输入图像首先进入 CAM 生成模块(即 MFF-Net1)获得初始的 CAM。然后使用高斯分布先验获得高斯增强的 CAM。同时,将产生的伪标签作为与类别无关

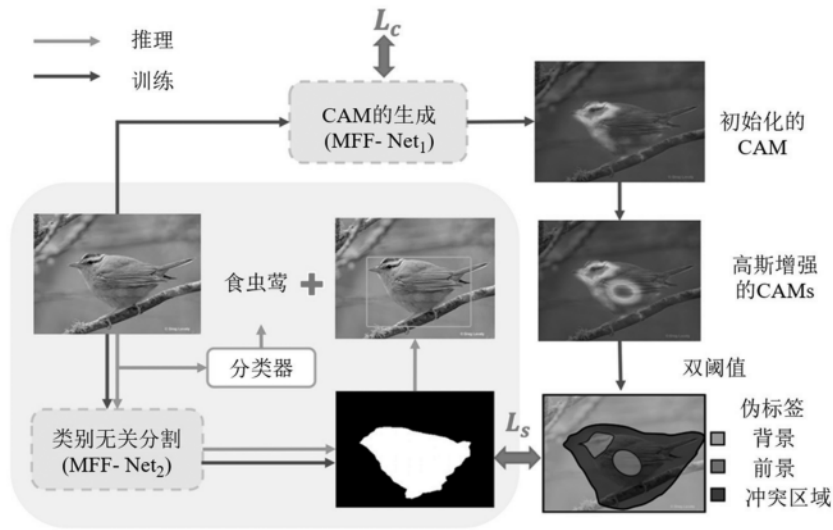


图 1 SPOL 算法流程

的分割模块(即 MFF-Net2)的监督)。在 CAM 生成阶段,乘法特征融合网络(Multiplicative Feature Fusion Net, MFF-Net)旨在聚合浅层和深层特征。与以往的融合方法不同,MFF-Net 网络中的特征以协同方式处理,即具有清晰背景的深层特征帮助抑制浅层的噪声,而浅层具有丰富局部结构的特征使对象边界更清晰。在与类别无关的分割阶段,使用高斯先验伪标签(Gaussian Prior Pseudo Label, GPPL)模块对初始的 CAM 进行细化,然后将其视为类不可知分割模块训练的伪标签。具体来说,充分利用初始的整个 CAM 作为加权系数,通过所有坐标的均值和方差计算得到物体的权重。然后,可以生成具有均值和方差的高斯分布,称为高斯先验伪标签。结合 GPPL 和原始 CAM,可以得到改善的 CAM。为了进一步细化这些区域,通过使用组合的 GPPL 和 CAM 作为伪标签设计了一个与类别无关的分割模型,在训练阶段,具有不同大小的 CAM 响应区域将分别被二值化为前景和背景。同时,使用两个预定义的阈值,其他部分将被忽略以避免训练期间标签的冲突。训练结束后,与初始 CAM 相比,获得的对象掩码将更加完整。最后,边界框提取器应用于对象掩码以获得最终的目标检测结果。

1.1 乘法特征融合

多尺度特征融合常用于强监督的目标检测^[14]、语义分割^[15]等任务。然而,这种策略对弱监督目标检测任务无效。因为来自浅层的特征包含太多的背景噪声。如果没有强有力的监督,浅层的特征就会

被背景噪声所掩盖,对最终的预测几乎没有效果。因此,该算法采用 MFF-Net 网络结构(见图 2)滤除浅层特征的背景噪声。首先以相同的分辨率(即 $H \times W$)对不同分支(即 X 、 Y 、 Z)的特征进行采样,然后通过逐元素乘法组合到后续的分类器中。

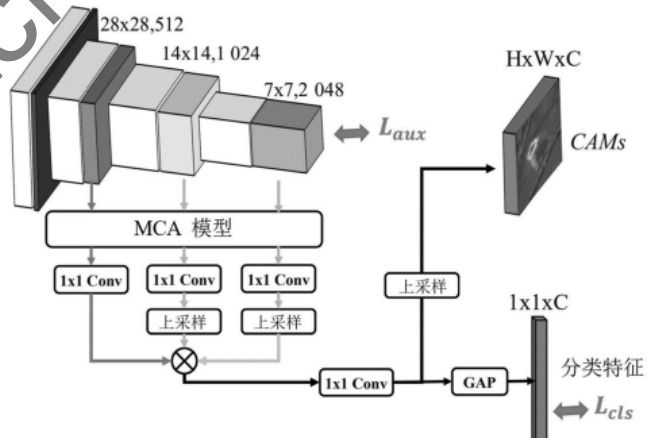


图 2 MFF-Net 网络结构

与以前的方法不同,MFF-Net 网络结构以协同方式处理浅层和深层特征,可以很好地利用浅层特征。为了详细说明这一点,图 3 介绍了四种具有不同监督风格的方法。具体来说,图 3(a)是原始分类模型(例如 VGG^[16]和 ResNet50^[17]),其中只有最后一层被监督。因为浅层特征远离监督并受到梯度消失问题的影响,图 3(b)显示了深度监督模型,其中深层和浅层特征都被直接监督以驱动网络学习更好的表示。但是由于感受野有限,浅层特征的语义较

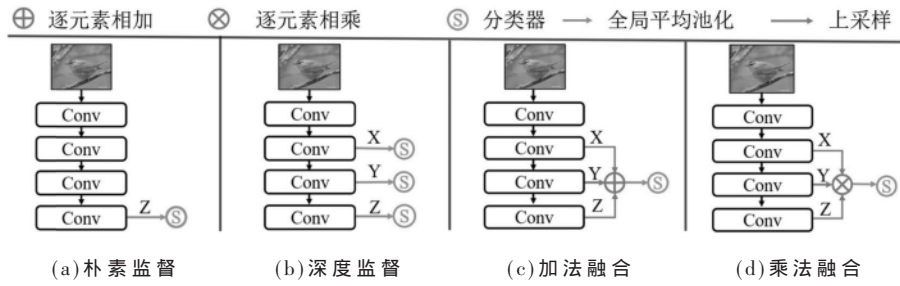


图3 不同监督方式

少并引入了更多的噪声。因此,这种直接监督对弱监督目标检测效果的帮助不大。与这些方法相比,特征融合提供了一种间接监督的形式,即在监督之前组合不同层的特征。

图3(c)和方程(3)展示常用的加法融合策略。但是,它没有考虑多尺度特征之间的相关性。如方程(4)所示,根据链式法则,在计算网络权重的梯度之前,特征相加对每个分支的梯度是相同的常数,与其他分支没有相关性。也就是说,当一个分支出错时,不会影响其他分支。在这种情况下,网络无法很好地学习每个分支,即使浅层特征是错误的,预测也是可以接受的。虽然它可以在测试阶段提高模型的稳定性,但在训练阶段却降低了模型容量并增加了算法的训练难度。与其他方法不同,在MFF-Net中,不同的分支通过乘法运算进行强耦合,如式(1)所示。式(2)说明了 X 分支的梯度不是恒定的,而是与 Y 和 Z 分支相关的。这三个分支在训练过程中会相互影响。当一个分支未能捕捉到优越的表示时,乘法机制会放大其错误,最终导致预测的错误。也就是说,MFF-Net为网络训练设置了强约束,其中每个分支都必须很好地学习表示。此外,在这种情况下, Y 和 Z 依赖于 X 。当 X 得到更好的表示时, Y 和 Z 将得到增强。因此,它们的融合可以产生更准确的预测。

$$F_{\text{mul}} = \frac{1}{H \times W} \sum_{i,j} (X_{ij} \times Y_{ij} \times Z_{ij}) \quad (1)$$

$$\frac{\partial F_{\text{mul}}}{\partial X_{ij}} = \frac{1}{H \times W} Y_{ij} \times Z_{ij} \quad (2)$$

$$F_{\text{add}} = \frac{1}{H \times W} \sum_{i,j} (X_{ij} + Y_{ij} + Z_{ij}) \quad (3)$$

$$\frac{\partial F_{\text{add}}}{\partial X_{ij}} = \frac{1}{H \times W} \quad (4)$$

1.2 基于乘法的通道注意力机制

卷积神经网络具有强大的特征提取能力,可以同时表示前景对象和非目标背景特征。浅层特征的某些部分是不必要的,可以视为噪声,会严重干扰最终预测。因此,如图4所示,在不同层的特征融合之前,采用通道注意力粗略地过滤掉噪声通道。区别于一次只关注一层的传统通道注意力方法,该算法设计了一个基于乘法的通道注意力(Multiplication based Channel Attention, MCA)模块同时处理各个层。

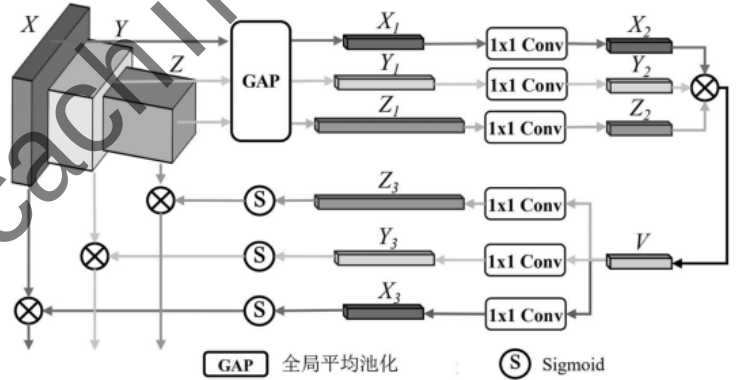


图4 MCA 模块

具体来说,对于输入的特征图 $X \in \mathbf{R}^{H_1 \times W_1 \times C_1}$ 、 $Y \in \mathbf{R}^{H_2 \times W_2 \times C_2}$ 和 $Z \in \mathbf{R}^{H_3 \times W_3 \times C_3}$,首先利用全局平均池化层分别实现 $X_1 \in \mathbf{R}^{1 \times C_1}$ 、 $Y_1 \in \mathbf{R}^{1 \times C_2}$ 和 $Z_1 \in \mathbf{R}^{1 \times C_3}$,然后利用3个平行的 1×1 Conv层将 X_1 、 Y_1 和 Z_1 分别转移到具有相同维度 $\mathbf{R}^{1 \times C'}$ 的 X_2 、 Y_2 和 Z_2 。最后,利用逐元素乘法融合得到潜在的特征向量 $V = X_2 \cdot Y_2 \cdot Z_2$, $V \in \mathbf{R}^{1 \times C'}$ 。这种潜在的特征向量表示不同层的特征的相互耦合。因此,可以将通道注意力同时应用于多个层。在相反的方向上,潜在的特征向量通过Sigmoid激活函数反馈到原始维度,即 $X_3 \in \mathbf{R}^{1 \times C_1}$ 、 $Y_3 \in \mathbf{R}^{1 \times C_2}$ 和 $Z_3 \in \mathbf{R}^{1 \times C_3}$ 。基于 X_3 、 Y_3 、 Z_3 ,MCA模块使

用乘法运算实现对每个对应层的通道注意。

1.3 类不可知分割引导的弱监督目标定位

由图 1 可见,虽然 MFF-Net1 网络结构已经产生了最初的 CAM,但它只关注最具辨别力的区域,还不足以提取准确的定位边界框。为了解决这个问题,进一步提出了伪监督类不可知分割模型,它利用了另一个 MFF-Net₂ 网络结构。这个与类别无关的分割模型,丢弃了类别信息,只关注定位信息,即输出仅代表前景或背景。该模型由伪标签生成模块和类别无关的分割模块两部分组成。

(1)分割伪标签生成。首先,通过高斯先验伪标签(GPPL)模块补充 CAM。CAM 上的每个点 (x, y) 都被视为一个样本。位置 (x, y) 处的响应对应于其权重,通过该设置,计算所有样本的 x 和 y 之间的均值 (μ_x, μ_y) 、方差 (σ_x^2, σ_y^2) 和相关系数 ρ 。然后,利用这些参数生成二维高斯分布,如式(5)和式(6),有助于定位物体重心并覆盖较宽的物体区域。

$$\Theta = \frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x \sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \quad (5)$$

$$f(x, y) = \frac{e^{-\frac{1}{2(1-\rho^2)}\Theta}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \quad (6)$$

然后将原始 CAM 与高斯增强的 CAM 集成在一起,采用元素最大值来获得完整的预测。下一步,增强的 CAM 被进一步转换为具有两个预定义阈值的 3 个部分,即高置信度区域对应于前景,低响应区域对应于背景,冲突区域对应于低置信度区域。

(2)类不可知分割和边界框提取。在获得前景和背景伪分割标签后,对类别无关的分割模型(即 MFF-Net2)进行训练。虽然只有一部分图像具有像素级标签,但分割模型可以捕获类似的上下文并自动覆盖前景。模型优化好后,可以从类不可知分割的预测掩码中提取边界框。最终的预测结合了提取的边界框和来自独立分类器的分类预测。

1.4 损失方程

CAM 的生成过程,除了应用分类损失外,还应用了附加损失。通过最后一个特征图和融合的分类特征计算两个损失,即 $L_{\max}$ 和 L_{cls} ,如图 2 所示。这两个损失都是使用交叉熵计算的。因此,联合损失 $L_c = L_{\text{cls}} + L_{\max}$ 用于优化 MFF-Net₁。对于类不可知分割,应用二元交叉熵损失来监督分割模型,如式(7)所示。然而,除了冲突区域外,只考虑伪前景和背景

区域。具体地,对于前景和背景, w_{ij} 等于 1,而对于那些冲突区域, w_{ij} 设置为零。最终损失冲突区域被忽略以避免误导网络。

$$L_s = - \frac{\sum_{(i,j)} w_{ij} [g_{ij} \log(p_{ij}) + (1-g_{ij}) \log(1-p_{ij})]}{H \times W} \quad (7)$$

其中 p_{ij} 和 g_{ij} 分别是位置 (i, j) 处的预测概率和真实标签。

2 改进的 SPOL 算法

由于迷彩伪装目标与图像背景融合度较大,SPOL 算法的检测精度相对较低。为了更进一步提高模型的检测精度,本文在 MFF-Net 网络模型中融入了注意力模块,使模型更加注重迷彩伪装目标所在的区域,以此增强模型提取特征的能力,得到包含伪装目标更全面的 CAM,进而提高模型的检测精度。

2.1 融合 CBAM 模块的 SPOL 算法

2.1.1 CBAM 模块

CBAM 将通道注意力和空间注意力串联起来,通过将通道注意力图和空间注意力图解耦以提高计算效率,并引入全局池化利用空间全局信息。模块结构如图 5 所示。

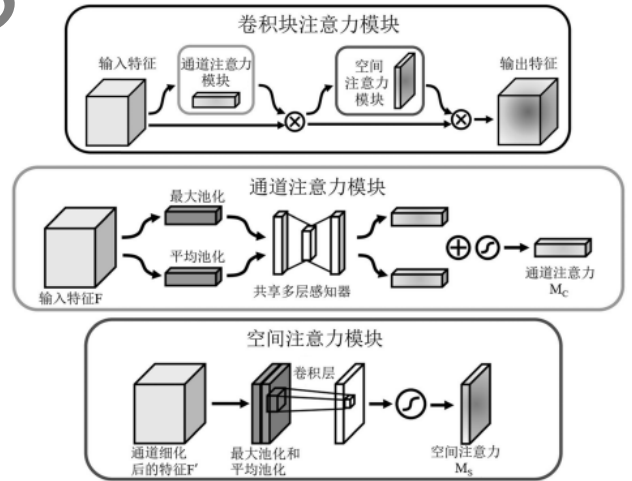


图 5 CBAM 结构

CBAM 模块有两个连续的子模块,通道注意力和空间注意力。给定输入的特征图 $X \in \mathbf{R}^{C \times H \times W}$,它依次推断一维通道注意力向量 $s_c \in \mathbf{R}^C$ 和二维空间注意力图 $s_s \in \mathbf{R}^{H \times W}$ 。通道注意力子模块采用两种不同的池化操作来聚合全局信息。它有两个并行分支,分别使用 MaxPool 和 AvgPool 操作:

$$F_{avg}^c = \text{GAP}^s(X) \quad (8)$$

$$F_{max}^c = \text{GMP}^s(X) \quad (9)$$

$$s_c = \sigma(W_2 \delta(W_1 F_{avg}^c) + W_2 \delta(W_1 F_{max}^c)) \quad (10)$$

$$M_c(X) = s_c X \quad (11)$$

式中 $\text{GAP}^s(\cdot)$ 和 $\text{GMP}^s(\cdot)$ 表示空间域中的全局平均池化和全局最大池化操作。空间注意力子模块对特征的空间关系进行建模,是对通道注意力子模块的补充。与通道注意力不同,它利用具有大内核的卷积层来生成注意力图。

$$F_{avg}^s = \text{GAP}^c(X) \quad (12)$$

$$F_{max}^s = \text{GMP}^c(X) \quad (12)$$

$$s_s = \sigma(\text{Conv}([F_{avg}^s; F_{max}^s])) \quad (14)$$

$$M_s(X) = s_s X \quad (15)$$

其中 $\text{Conv}(\cdot)$ 表示卷积运算,而 $\text{GAP}^c(\cdot)$ 和 $\text{GMP}^c(\cdot)$ 是通道域中的全局平均池化和全局最大池化操作。 $[\cdot]$ 表示通道上的串联。整个注意力过程可以概括为:

$$X' = M_c(X) \quad (16)$$

$$Y = M_s(X') \quad (17)$$

CBAM 模块依次结合通道注意力和空间注意力,可以利用特征的空间和跨通道关系强调网络关注什么以及关注哪里。即强调有用的通道以及增强

信息丰富的本区域。由于其轻量级设计,CBAM 模块可以无缝集成到任何 CNN 架构中,附加成本可以忽略不计。

2.1.2 融合 CBAM 模块的 MFF-Net 网络

原始的 MFF-Net 网络模型采用 ResNet50 网络提取特征。虽然该网络具有很强的特征提取能力,但是相比于通用的目标,迷彩伪装目标和背景融合度非常大。目标在图像中不易识别,从而增大了网络提取目标特征的难度。为了缓解这一问题带来的影响,本文在 ResNet50 网络中加入 CBAM 模块,让网络更加注意迷彩伪装目标所在的区域,以此增强网络的特征提取能力,增加的具体位置如图 6 所示。

2.2 融合 SE 模块的 SPOL 算法

2.2.1 SE 模块

SENet 的核心是一个 SE 模块,用于收集全局信息、捕获通道关系和提高表示能力。SE 模块分为两部分:挤压模块和激励模块,模块结构如图 7 所示。

挤压模块通过全局平均池化操作收集全局空间信息。激励模块使用全连接层和非线性层(ReLU 和 Sigmoid)捕获通道关系并输出注意力向量。然后,将输入的特征向量与注意力向量中的相应元素相乘来缩放输入特征的每个通道。一个以 X 为输入和 Y 为输出的挤压和激励块 F_{se} (带有参数 θ) 可以

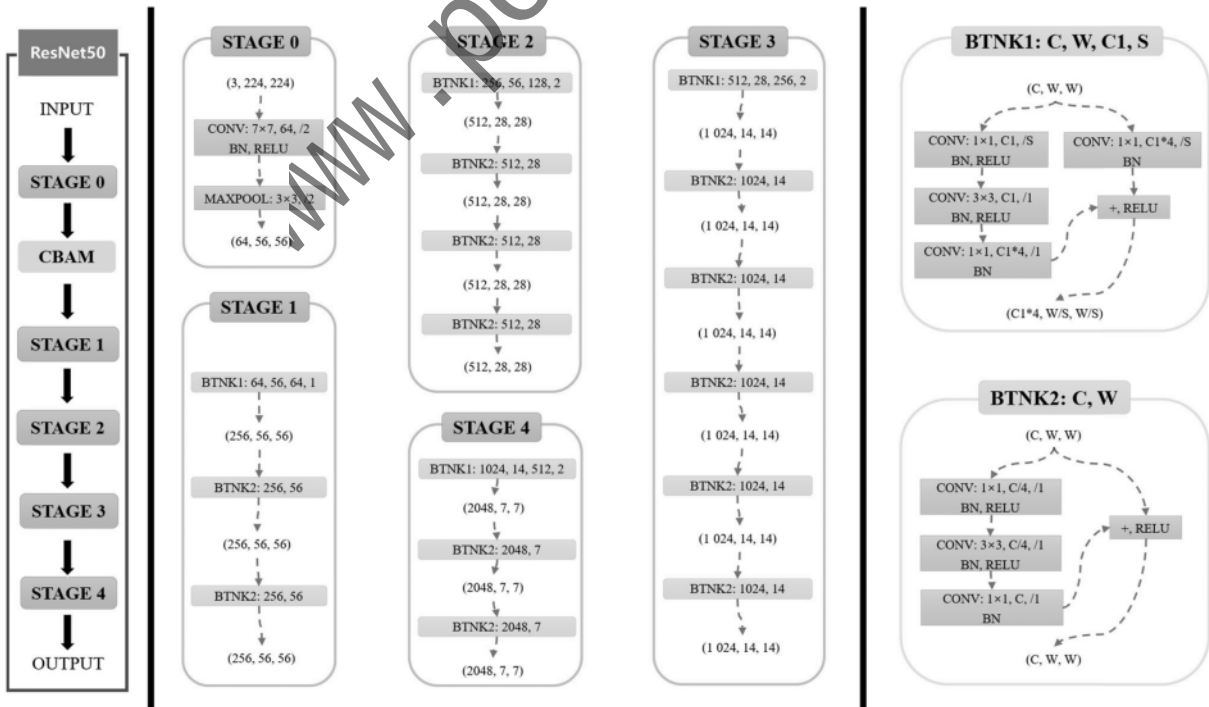


图 6 融合 CBAM 模块的 ResNet50 网络结构

表述为:

$$s = F_{se}(X, \theta) = \sigma(W_2 \delta(W_1 \text{GAP}(X))) \quad (18)$$

$$Y = sX \quad (19)$$

2.2.2 融合 SE 模块的 MFF-Net 网络

尽管在 ResNet50 网络中融合 CBAM 模块后,增强了网络提取特征的能力,从而提升了算法的检测效果,但仅在网络的 STAGE 0 模块后加入 CBAM 模块,这导致后续的几个模块在进行特征提取操作时无法着重关注迷彩伪装目标所在的区域。为了进一步利用注意力机制提高模型的检测精度,本文在 MFF-Net 网络模型的 ResNet50 中融入了 SE 模块,加入的具体位置如图 8 所示。

2.3 融合 CBAM 模块和 SE 模块的 SPOL 算法

为进一步验证融合注意力模块后目标检测算法的有效性,以及比较不同注意力模块对不同尺度目标检测效果的影响和作用,在上述研究的基础上尝试对两种注意力模块进行融合,改进形成的网络

结构如图 9 所示。

3 实验分析

3.1 数据集

实验采用的目标图像来源于课题组自建数据集 CSS^[18],该数据集包含雨林、丛林、雪地和山地 4 种野外环境下多类型的迷彩伪装目标。从实际战场侦察角度看,CSS 数据集涵盖了不同场景、不同天候、不同迷彩类型的多尺度伪装目标,可以满足多种迷彩伪装目标检测识别的实验需求。为了满足弱监督迷彩伪装目标检测算法训练和测试的要求,本文在 CSS 数据集的基础上进行了筛选与重新标注,构建了一个适合弱监督目标检测算法的数据集。该数据集的训练集由 6 100 张迷彩伪装目标图像组成,其中包括 1 000 张大目标图像、2 100 张中目标图像以及 3 000 张小目标图像。为了更好地验证本文设计的算法的有效性,构建了两个测试集用于算法的验证:第一个测试集由 350 张迷彩伪装目标图

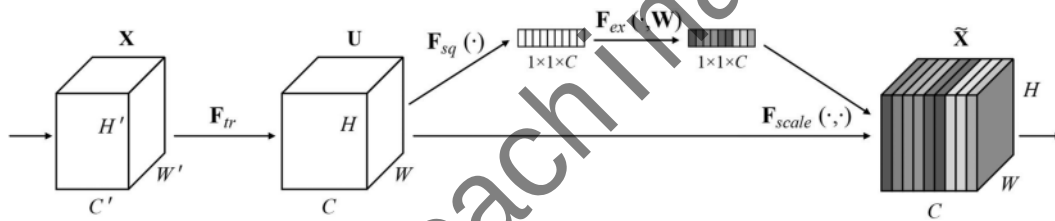


图 7 SE 模块结构

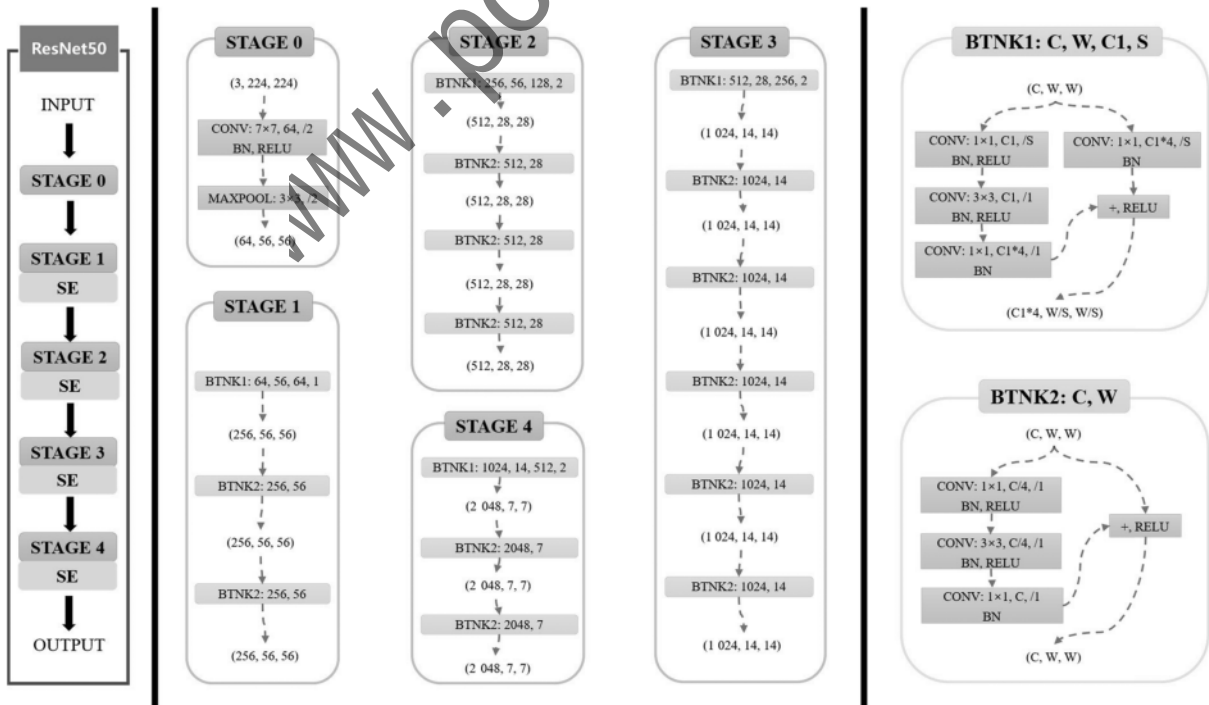


图 8 融合 SE 模块的 ResNet50 网络结构

像组成,只包含大目标和中目标的图像;第二个测试集包括大目标、中目标、小目标三个子集,三个子集都由 100 张迷彩伪装目标图像组成。其中,将目标所占像素点与全图像素点数量的百分比作为定义目标尺度的依据。小目标所占像素点与全图像素点数量的比值小于或等于 1%;中目标所占像素点与全图像素点数量的比值在 1%到 3%之间;大目标所占像素点与全图像素点数量的比值大于 3%,图 10 展示了不同尺度目标的图像样本。

3.2 实验环境

搭建实验环境,使用的服务器操作系统为 Ubuntu18.04,使用的应用软件环境及工具包包含 CUDA、Python3.8 等,使用 PyTorch 深度学习框架进行实验。实验所用的硬件设备配置如表 2 所示,训练算法模型时的重要参数设置如表 3 所示。

表 2 实验硬件环境配置

硬件名称	型号	数量
CPU	i9-9900K	1
内存	金士顿 16 GB	4
GPU	技嘉 GTX2080Ti 11 GB	1

表 3 训练参数列表

参数名	参数值
Batch Size (批量大小)	4
Learning Rate (学习率)	0.02
Moentum (动量)	0.9
Epochs (训练轮次)	20

3.3 评价指标

常用的弱监督目标检测评价指标有以下三种:

(1)Top-1 定位精度(Top-1 Loc):预测的物体类

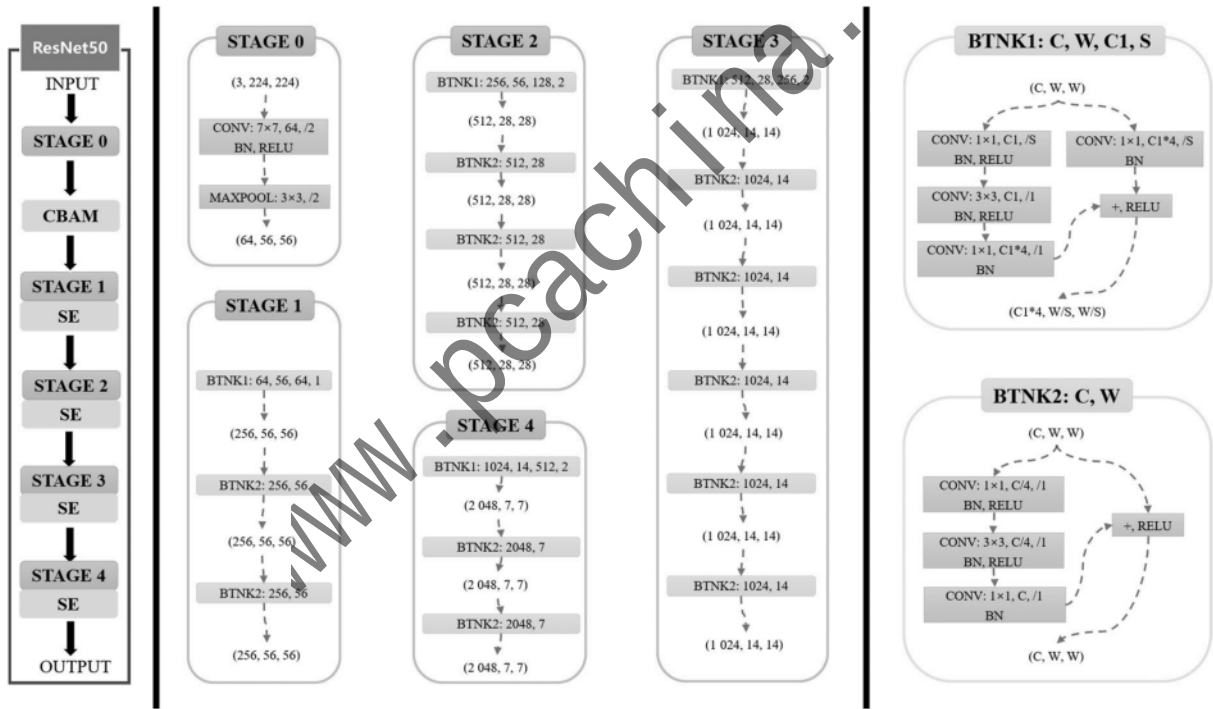


图 9 融合 CBAM 模块和 SE 模块的 ResNet50 网络结构



图 10 多尺度目标图像样本

别必须和物体的真实类别相同,且预测边界框与真实框的交并比的 (Intersection-over-Union, IoU) 数值超过 50%,即预测边界框与真实框的重合率超过 50%的比例。

(2)Top-5 定位精度(Top-5 Loc):对于预测的物体类别,取可能性最高的 5 类,只要这 5 类中包含物体的真实类别,就算分类正确。同时,预测的边界框与真实框的 IoU 数值超过 50%的比例。

(3)GT-known 定位精度(GT-known Loc):预测边界框与真实框的 IoU 数值超过 50%的比例。

本文是在迷彩伪装目标数据集上进行的对比实验,该数据集只有单类物体,故采用 Top-1 Loc 和 GT-known Loc 作为目标检测结果的评价标准。

3.4 改进算法的实验结果

表 4 比较了改进算法与原始算法在弱监督迷彩伪装目标图像数据集上的实验结果,采用的是大、中尺度目标混合的验证集。其中,CBAM-SPOL 表示融合 CBAM 模块的检测算法,SE-SPOL 表示融合 SE 模块的检测算法,CBAM-SE-SPOL 表示同时融合了 CBAM 模块和 SE 模块的检测算法。

表 4 不同算法的实验结果对比 (%)

算法	Top-1 Loc	GT-known Loc
SPOL	63.17	63.45
CBAM-SPOL	67.46	67.46
SE-SPOL	71.18	71.18
CBAM-SE-SPOL	70.40	70.74

由表 4 可见,融合注意力模块后,算法的检测精度比原始算法的检测精度有比较明显的提升。其中,加入 CBAM 模块后,Top-1 Loc 评价指标提高了 4.29%;加入 SE 模块后,该指标提高了 8.01%。然而,同时加入 CBAM 模块和 SE 模块后,该指标提高了 7.23%,仅比只加入 CBAM 模块提高 2.94%,而比只加入 SE 模块时降低 0.78%。主要原因是,同时融合 CBAM 模块和 SE 模块后,在模型训练过程中,当输入图像通过 CBAM 模块时,模型会选择性地丢弃一些不重要的信息,导致后续的 SE 模块局限于关注 CBAM 模块保留的特征信息,从而在一定程度上降低了模型提取特征的能力。但总体来说,加入注意力模块后的 MFF-Net 网络比原始的 MFF-Net 网络能够提取到更多的有效特征,可增强算法对隐蔽性高的迷彩伪装目标的检测能力。

为了验证注意力模块对各种尺度迷彩伪装目标检测效果的影响,分别采用含有的大目标、中目标、小目标图像的验证集对上述四种算法进行再次验证,并使用 Top-1 Loc 作为实验结果的评价指标,实验结果如表 5 所示。

表 5 不同算法对多尺度目标的检测结果对比 (%)

算法模型	Top-1 Loc		
	大目标	中目标	小目标
SPOL	67.68	32.58	3.41
CBAM-SPOL	72.45	39.76	6.32
SE-SPOL	85.86	42.53	7.29
CBAM-SE-SPOL	74.75	55.79	10.31

从表 5 中看出,对于三种不同尺度的迷彩伪装目标图像,融合注意力模块后的算法比原始 SPOL 算法的检测效果都有明显提升。综合比较得出,SE-SPOL 算法对大目标图像的检测效果最好,Top-1 Loc 评价指标提高了 18.18%;CBAM-SE-SPOL 算法对中目标图像的检测效果有显著的提升,评价指标比原始 SPOL 算法提高了 23.21%;同时,CBAM-SE-SPOL 算法对小目标图像检测效果的提升也很明显,评价指标提高了 6.90%。产生上述对比效果的主要原因是:(1)大目标在图像中所占的像素点数量多,SE 通道注意力本身对大目标的特征提取能力很强,而 CBAM 模块和 SE 模块的融合,反而会使模型丢失一些迷彩伪装目标的特征信息,所以当数据集主要包含大目标图像时,更适合用 SE-SPOL 算法进行检测。(2)中目标和小目标在图像中所占的像素点数量相对较少,先通过 CBAM 模块使模型关注这两类目标在图像中的大致区域,然后利用 SE 模块进一步加强模型对这些区域的特征提取能力,会显著改善检测效果;所以当数据集主要包含中目标或小目标时,更适合用 CBAM-SE-SPOL 算法进行检测。(3)小目标在图像中所占像素点数量太少,同时迷彩伪装目标与图像背景相似度较高,采用的图像级标签不含目标的位置信息,三点原因综合起来会导致模型训练时很难学习到小目标的特征。因此,虽然改进算法也能明显改善对小目标的检测效果,但检测精度还有很大的提升空间。

3.5 目标检测效果分析

图 11 展示了原始算法与加入不同注意力模块

的改进算法生成的像素级伪标签的比较结果。从图中可以看出,加入注意力模块后,改进算法生成的伪标签图像比 SPOL 算法更加清晰,覆盖的目标像素点更多。可见,改进算法中的 MFF-Net₂ 网络可以学习到更全面的目标信息,进而提升了模型的检测精度。

图 12 对比了改进算法和原始算法的检测效果。从第 1 行到第 3 行分别为大目标图像、中目标图像以及小目标图像。图中,绿色方框表示目标的真实框,红色方框表示算法的预测边界框,红色数值为目标检测的 IoU 比值。对比可见,本文提出的改进算法模型对目标的预测边界框更加准确,预测边界框与真实框的 IoU 比值更大,说明预测结果与真实情况更加接近。其中,SE-SPOL 算法对大目标的检测效果最好,IoU 比值最高;CBAM-SE-SPOL 算法对中、小目标的检测效果最好。

4 结论

本文以弱监督目标定位任务中的 SPOL 算法为基础,针对迷彩伪装目标图像这一类特殊场景,在算法中加入注意力模块,加强模型对迷彩伪装目标的特征提取能力。不同算法间的实验结果对比表明:本文的算法比原始算法在检测精度上得到了较大的提升。下一步的研究工作是针对实际场景中大量存在的模糊图像,研究采用去模糊算法对图像进行增强,以进一步提高模型的检测精度。

参考文献

- [1] BHAJANTRI N U, NAGABHUSHAN P. Camouflage defect identification: a novel approach[C]//9th International Conference on Information Technology(ICIT'06). IEEE, 2006: 145-148.
- [2] SENGOTTUVELAN P, WAHI A, SHANMUGAM A.

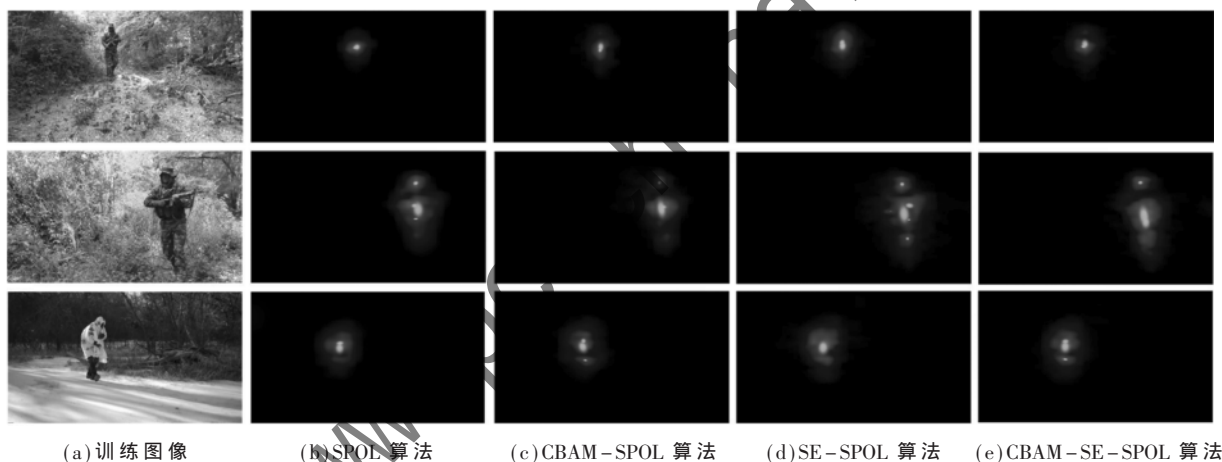


图 11 伪标签比较

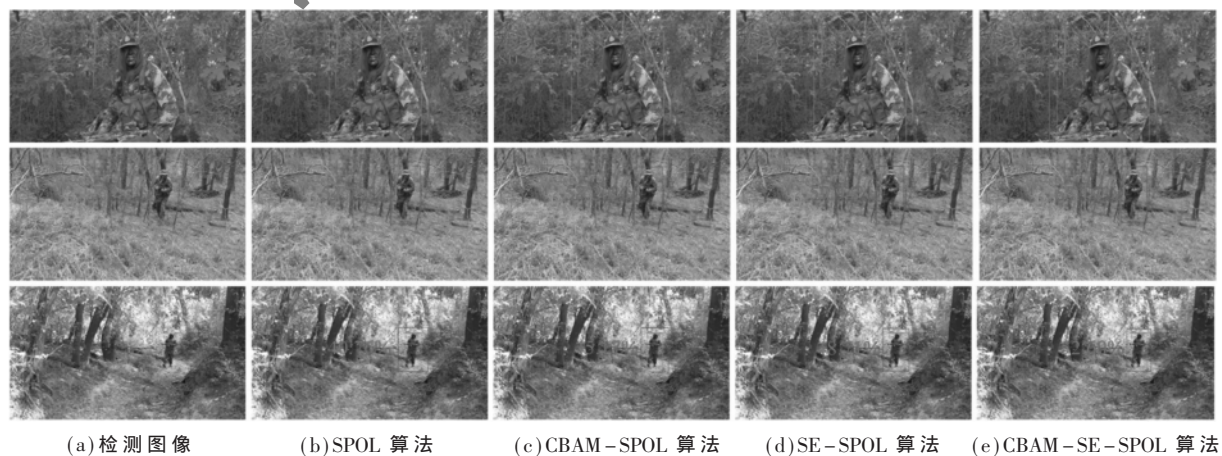


图 12 检测效果比较

- Performance of decamouflaging through exploratory image analysis[C]//2008 First International Conference on Emerging Trends in Engineering and Technology.IEEE, 2008:6-10.
- [3] 武国晶,吕绪良,邢海宁,等.三维凸面分析法在迷彩伪装检测中的应用[J].解放军理工大学学报(自然科学版),2015,16(6):582-586.
- [4] 邓小桐,曹铁勇,方正,等.改进 RetinaNet 的伪装人员检测方法研究[J].计算机工程与应用,2021,57(5):190-196.
- [5] LIN T Y,GOYAL P,GIRSHICK R,et al.Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision,2017:2980-2988.
- [6] 王杨,曹铁勇,杨吉斌,等.基于 YOLO v5 算法的迷彩伪装目标检测技术研究[J].计算机科学,2021,48(10):226-232.
- [7] ZHOU B,KHOSLA A,LAPEDRIZA A,et al.Learning deep features for discriminative localization[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016:2921-2929.
- [8] WEI J,WANG Q,LI Z,et al.Shallow feature matters for weakly supervised object localization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2021:5993-6001.
- [9] CHOE J,SHIM H.Attention-based dropout layer for weakly supervised object localization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019:2219-2228.
- [10] ZHANG X,WEI Y,YANG Y.Inter-image communication for weakly supervised localization[C]//European Conference on Computer Vision. Springer,Cham,2020:271-287.
- [11] ZHANG C L,CAO Y H,WU J.Rethinking the route towards weakly supervised object localization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020:13460-13469.
- [12] WOO S,PARK J,LEE J Y,et al.Cbam:convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision(ECCV),2018:3-19.
- [13] HU J,SHEN L,SUN G.Squeeze-and-excitation networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:7132-7141.
- [14] LIN T Y,DOLLÁR P,GIRSHICK R,et al.Feature pyramid networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2017:2117-2125.
- [15] LONG J,SHELHAMER E,DARRELL T.Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2015:3431-3440.
- [16] SIMONYAN K,ZISSERMAN A.Very deep convolutional networks for large-scale image recognition[J].arXiv preprint arXiv:1409.1556,2014.
- [17] HE K,ZHANG X,REN S,et al.Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016:770-778.
- [18] 梁新宇,林浩坤,杨辉,等.迷彩伪装目标图像语义分割数据集的构建[J].激光与光电子学进展,2021,58(4):214-220.

(收稿日期:2022-05-30)

作者简介:

杨辉(1997-),男,硕士研究生,主要研究方向:计算机视觉、目标检测。

权冀川(1974-),通信作者,男,博士,主要研究方向:系统工程。E-mail:574959336@qq.com。

梁新宇(1993-),男,硕士,主要研究方向:计算机视觉、图像分割。



版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部

www.pcachina.cn