

CAMOU-YOLO:一种迷彩伪装目标检测模型

王焯奎^{1,2}, 曹铁勇¹, 王 杨¹, 方 正¹, 刘亚九², 郑云飞^{1,3,4}, 付炳阳¹

(1. 陆军工程大学 指挥控制工程学院, 江苏 南京 210007;

2. 31401 部队, 吉林 长春 130000;

3. 陆军炮兵防空兵学院, 江苏 南京 211100;

4. 偏振光成像探测技术安徽省重点实验室, 安徽 合肥 230031)

摘 要:由于迷彩伪装目标与所处背景高度融合, 现有深度目标检测模型在此类目标上的检测效果并不出众。为提升对迷彩伪装目标的检测精度, 以 YOLOv5s 模型为基础, 提出了 CAMOU-YOLO——一种结合深度可分离卷积和动态注意力的迷彩伪装目标检测模型。针对迷彩伪装目标特征提取难的问题, 结合深度可分离卷积与残差结构设计了新的特征提取模块, 并对原有骨干网络进行改进, 在增强提取能力的同时, 减小了模型的参数量; 针对迷彩伪装目标定位难度大的问题, 在聚合网络中引入动态注意力机制, 强化了模型的空间感知能力, 使模型对迷彩伪装目标的定位更加精准。在一种公开的迷彩数据集上进行实验, CAMOU-YOLO 的 mAP@0.5、mAP@0.75 和 mAP@0.5:0.95 指标较原始模型提高了 3.2%、5.1%、2.3%, 在大、中、小目标上的召回率分别提高了 4.1%、2.7%、1.2%, 且参数量降低了 9.7%; 对比其他 7 种检测算法, CAMOU-YOLO 在检测精度上亦具有优势, 验证了所提模型对迷彩伪装目标检测任务的有效性。

关键词:迷彩伪装; 目标检测; YOLO; 深度可分离卷积; 动态注意力

中图分类号: TP751

文献标识码: A

文章编号: 1673-629X(2022)12-0029-08

doi:10.3969/j.issn.1673-629X.2022.12.005

CAMOU-YOLO: A Camouflaged Object Detection Model

WANG Ye-kui^{1,2}, CAO Tie-yong¹, WANG Yang¹, FANG Zheng¹,

LIU Ya-jiu², ZHENG Yun-fei^{1,3,4}, FU Bing-yang¹

(1. School of Command and Control Engineering, Army Engineering University of PLA, Nanjing 210007, China;

2. Unit 31401, Changchun 130000, China;

3. The Army Artillery and Defense Academy of PLA, Nanjing 211100, China;

4. Key Laboratory of Polarization Imaging Detection Technology, Hefei 230031, China)

Abstract: Since the camouflaged object being highly fused with the background, the existing object detection models based on deep learning have poor detection performance on such objects. Based on the YOLOv5s model, a camouflage object detection model combining depth separable convolution and dynamic attention is proposed to improve the detection accuracy. Aiming at the difficulty of feature extraction of camouflaged object, combined with the structure of depth separable convolution and residual, a new feature extraction module is designed, and the original backbone network is improved, which not only enhances the extraction ability but also reduces the number of parameters of the model. Aiming at the difficulty of positioning camouflaged object, the dynamic attention mechanism is introduced into the aggregation network to strengthen the spatial perception ability of the model and make the positioning of the camouflage object more accurate. Experimenting on a public camouflage dataset, CAMOU-YOLO's mAP@0.5, mAP@0.75 and mAP@0.5:0.95 are 3.2%, 5.1%, 2.3% higher than the original model, and recall rates on large, medium, and small objects are 4.1%, 2.7%, 1.2% higher, respectively. At the same time, the parameter quantity is reduced by 9.7%. Compared with the other seven detection algorithms, CAMOU-YOLO also has advantages in detection accuracy, which verifies the validity of the proposed model for camouflage target detection tasks.

Key words: camouflage; object detection; YOLO; depth separable convolution; dynamic attention mechanism

收稿日期: 2021-12-14

修回日期: 2022-04-14

基金项目: 国家自然科学基金青年科学基金(61801512); 国家自然科学基金(62071484); 江苏省优秀青年基金项目(BK20180080)

作者简介: 王焯奎(1992-), 男, 硕士, 研究方向为目标检测及对抗样本; 通信作者: 曹铁勇(1971-), 男, 博士, 教授, 博导, 研究方向为语音和图像处理。

0 引言

随着战争形态由信息化向智能化转变,无人作战等新兴领域对战场态势的感知能力提出了更高要求。在军事行动中,迷彩是较为常用的伪装手段,其能模拟自然环境的颜色特性,破坏伪装目标轮廓,使目标与周围背景相融合,从而实现伪装的效果。迷彩伪装对感知战场态势产生重大干扰,高效、精准地发现和检测迷彩伪装目标成为提升战场态势感知能力的关键。

在深度学习兴起前,一些传统方法利用统计特征检测迷彩伪装目标。辛浩然等人^[1]在 HSV 颜色空间处理迷彩图像,提取 H 分量进行小波变换生成模极大值,并通过模极大值发现图像突变点位置来提取目标边缘信息。武国晶等人^[2]改进三维凸面检测算子,通过检测灰度变化的峰值定位迷彩伪装目标。Bhajantri 等人^[3]将图像分割成若干个不相交的块,对每个块计算灰度共生矩阵(GLCM)特征,并使用分水岭算法检测伪装目标。Sengottuvelan 等人^[4]通过 GLCM 分析伪装目标的纹理特征,实施检测。然而这些传统算法局限于浅层特征提取,在检测精度和速度上均不够理想。

近年来,深度神经网络被大规模应用于目标检测任务中,并取得了良好的效果。根据候选区域(Region Proposals)生成方式的不同,可以将深度目标检测算法分为双阶段算法与单阶段算法。双阶段检测算法在第一阶段生成可能包含目标的候选区域,在第二阶段对候选区域的位置进行调整,并输出所包含目标的类别,典型算法包括 R-CNN^[5]、Fast RCNN^[6]、Faster RCNN^[7]、Cascade RCNN^[8]等。双阶段算法的检测精度较传统算法大幅提升,但存在检测速度较慢、训练优化困难等问题。单阶段检测算法直接在骨干网络提取出的特征图上生成候选区域,并完成后续检测,典型方法包括 YOLO 系列^[9-12]、SSD^[13]等。单阶段算法的检测速度较双阶段算法有明显的提升,能满足实时感知战场态势的要求,然而单阶段算法对迷彩伪装目标特征的提取能力较弱,导致其在迷彩目标检测任务上的效果并不出众。

因此,该研究以单阶段算法中性能较为先进的 YOLOv5 模型为基础,提出了面向迷彩伪装目标检测任务的 CAMOU-YOLO 模型。对骨干网络进行改进,结合深度可分离卷积^[14]与残差结构设计了新的提取模块,替换网络原有模块,既提升了模型对迷彩伪装目标的特征提取能力,又缩减了网络规模。此外,在聚合网络中引入动态注意力机制(Dynamic Attention Mechanism, DAM),强化模型对迷彩伪装目标语义特征和位置信息的感知能力。在检测精度较原始模型显著提升的同时,CAMOU-YOLO 的参数量下降明显,

满足了军事行动中实时、准确感知战场态势的要求。

1 相关工作

1.1 YOLOv5 网络结构

YOLOv5 模型由骨干网络、聚合网络、检测头组成,模型结构如图 1 所示。

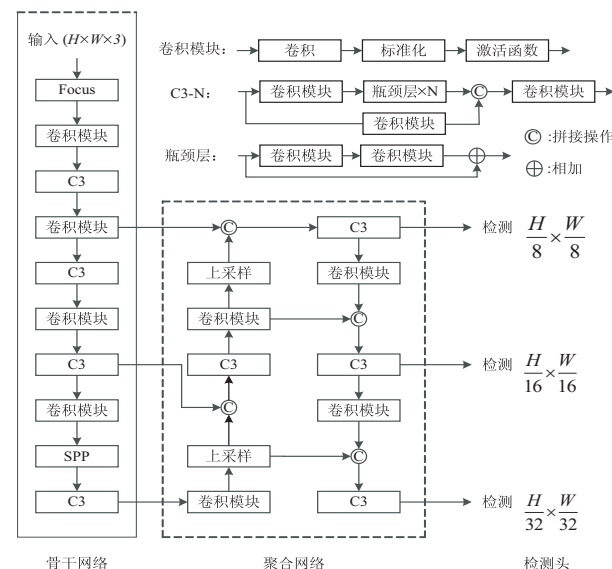


图 1 YOLOv5 网络结构

骨干网络主要由卷积与 C3 (CSP-Net^[15] with 3 Convolutions) 模块组成。卷积模块由卷积层、批量归一化层、激活函数组成;C3 模块将输入分成两路,一路经卷积模块处理,另一路通过卷积模块后,再由 N 个瓶颈层进行处理,最后在通道维度将两路输出拼接,送入卷积模块处理得到最后的输出。两路计算丰富了梯度传播的路径,增强了网络的学习能力。

经骨干网络提取的特征图,进入路径聚合网络(Path Aggregation Network, PANet^[16])进行双向融合。与传统特征金字塔不同,PANet 在经自底而上的融合后,特征图再进行一次自上而下的融合,同时双向融合间增加了横向连接。这种融合方式有利于低层细节信息的传导,使得顶层特征图得到底层特征图带来的丰富位置信息,提升了模型的定位精度。

检测时,YOLOv5 模型将图片分成 $S \times S$ 个网格,由目标中心所在区域的网格负责预测该目标。模型在网格上放置 B 个 Anchor,对每一 Anchor 输出 $(5+N)$ 个预测值,分别是置信度(confidence)、锚框中心距网格左上角坐标的偏移量(t_x, t_y)、锚框长宽的缩放比(t_w, t_h)和 N 个类别的概率。为了增加正样本的数量,YOLOv5 还采用了跨网格匹配规则,除了目标中心所在区域的网格之外,再找出离目标中心最近的两个网格,同时用这三个网格预测该目标,缓解了正负样本不平衡的问题。

模型损失函数由类别损失 l_{cls} 、置信度损失 l_{obj} 和

位置回归损失 l_{box} 三部分组成, 具体公式如下所示:

$$l_{\text{cls}} = - \sum_{i=1}^{S^2} \sum_{j=1}^B I_{ij}^{\text{obj}} \sum_{n=1}^N \text{BCE}(\widetilde{c}_n, c_n) \quad (1)$$

$$l_{\text{obj}} = - \sum_{i=1}^{S^2} \sum_{j=1}^B I_{ij}^{\text{obj}} \text{BCE}(\widetilde{C}_i, C_i) - \lambda_{\text{noobj}} \sum_{i=1}^{S^2} \sum_{j=1}^B I_{ij}^{\text{noobj}} \text{BCE}(\widetilde{C}_i, C_i) \quad (2)$$

$$l_{\text{box}} = 1 - \text{CIoU} \quad (3)$$

类别损失、置信度损失均采用了 BCE (交叉熵损失函数); 位置回归损失采用 CIoU 进行计算, CIoU 是对 DIoU 的改进, DIoU 考虑到了边界框与预测框重叠面积和中心点距离对目标位置回归的影响, 但忽略了两框的长宽比这个因素, CIoU 在其基础上添加了长宽比这一惩罚项, 使得目标框回归变得更加稳定。当第 i 个格子中的第 j 个预测框包含目标时, $I_{ij}^{\text{obj}} = 1$, $I_{ij}^{\text{noobj}} = 0$; 不包含目标时, $I_{ij}^{\text{obj}} = 0$, $I_{ij}^{\text{noobj}} = 1$ 。

\widetilde{C}_i 代表该类别预测概率, \widetilde{c}_n 为预测值, C_i 、 c_n 为类别标签值 (非 0 即 1)。

1.2 深度可分离卷积

标准卷积需要对输入的所有通道进行卷积, 产生了大量的计算, 如图 2(a) 所示。假设输入大小为 $M \times W \times H$, 输出大小为 $N \times W \times H$, 卷积核大小为 $M \times K \times K$, 由于输出通道为 N , 因此需要 N 个上述卷积核进行卷积操作, 由此可得标准卷积计算量为 $K^2 \times M \times N \times H \times W$ 。

如图 2(b)、(c) 所示, 深度可分离卷积将标准卷积分解为深度卷积和逐点卷积两部分。深度卷积对输入的每个通道分别应用一个 $K \times K$ 大小的卷积进行处理, 共有 M 个通道, 由此可得深度卷积的计算量为 $K^2 \times M \times H \times W$; 逐点卷积再将深度卷积的输出应用 N 个 $M \times 1 \times 1$ 大小的卷积进行组合, 由此可得逐点卷积计算量为 $M \times N \times H \times W$ 。综上所述, 深度可分离卷积相较于标准卷积缩减了模型大小, 计算量减少了:

$$\frac{K^2 \times M \times H \times W + M \times N \times H \times W}{K^2 \times M \times N \times H \times W} = \frac{1}{K^2} + \frac{1}{N} \quad (4)$$

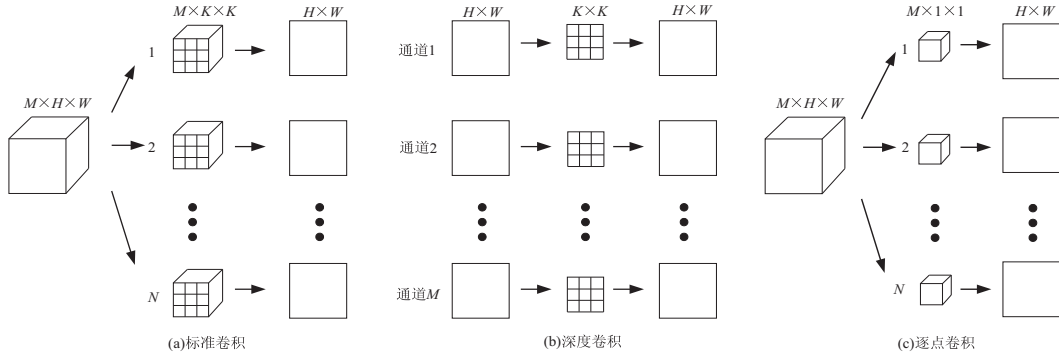


图 2 标准卷积和深度可分离卷积计算示意图

1.3 Dynamic ReLU

ReLU (修正线性单元, Rectified Linear Unit) 是深度神经网络中常用的激活函数, 计算简单高效, 有效解决了深度神经网络中存在的梯度爆炸和梯度消失问题, 提高了网络性能。但已有的 ReLU 及其变种 Leaky ReLU, PReLU 等激活函数的参数均固定不变, 对所有输入执行相同的处理。

Dynamic ReLU^[17] 是一种基于 ReLU 的激活函数, 其将全局上下文编码为一个超函数, 分段线性函数的斜率由超函数决定, 根据输入不同动态调整激活函数的参数。与参数固定的 ReLU 相比, Dynamic ReLU 在仅增加少量计算开销的情况下, 表征能力显著提升。

1.4 CBAM

注意力机制是人类视觉系统特有的处理机制, 在人类感知中起着重要作用。人类视觉系统的一个重要特性是人类不会试图一次性处理整个场景, 而是通过快速观察全局图像, 发现需要重点关注的目标区域, 而后聚焦重要区域, 获取更多目标的细节信息, 抑制其他

无用信息, 从而有效提高了人们处理信息的能力和效率。

与人类的视觉注意力机制类似, 深度学习中的注意力机制不仅告诉模型需要关注哪里, 且提高了模型的性能。具体来说就是从输入的大量数据中选择性地筛选出重要信息, 分配不同的关注权重, 使模型聚焦于重要的信息, 抑制不必要的信息。目前, 注意力机制在机器翻译、语音识别、图像分类、目标检测、目标分割等领域均得到了广泛应用。Woo 等人提出的 CBAM^[18] 从通道、空间两个维度依次应用注意力模块, 使网络关注输入的重要特征, 抑制无用特征, 更好地在通道和空间维度获取目标类别信息和位置信息, 从而引导模型正确地聚焦目标。CBAM 是一个简单而有效的前馈卷积神经网络注意力模块, 可以集成到任何 CNN 架构中, 额外增加的开销可以忽略不计。

2 CAMOU-YOLO 模型

迷彩伪装目标的外观与所处背景高度相似, 如何

高效提取迷彩伪装目标的特征信息,提升模型区分目标区域与背景区域的能力是研究的难点。

CAMOU-YOLO 以 YOLOv5 模型为基础,对其中的骨干网络和聚合网络进行改进,有效增强了原始模型对迷彩伪装目标特征的提取能力,检测精度明显提升,CAMOU-YOLO 结构如图 3 所示。

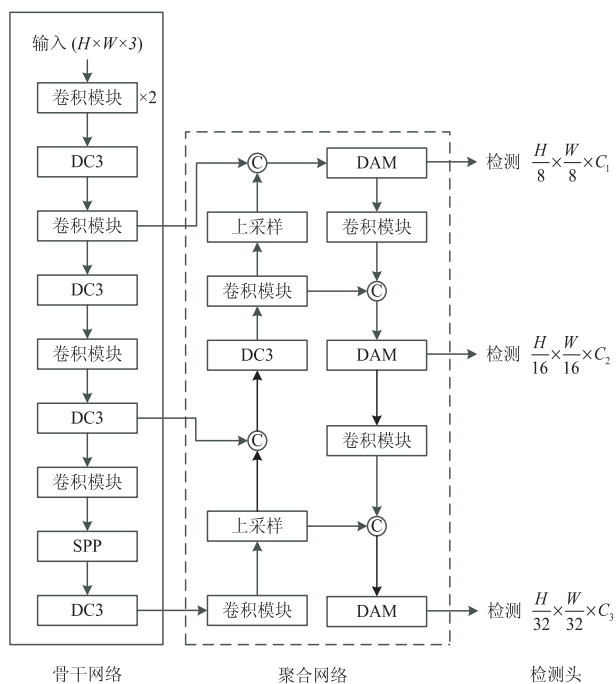


图 3 CAMOU-YOLO 网络结构

2.1 特征提取网络改进

原始特征提取网络中的 C3 结构将输入特征分为两路,通过卷积和残差瓶颈网络处理后,在通道维度进行拼接。C3 结构降低了计算量与内存消耗,然而该结构仍存在以下不足:(1)残差瓶颈网络为了减少计算量,首先使用一个 1×1 卷积对输入进行降维,其次使用 3×3 卷积进行处理,然而卷积在低维空间能够提取到的特征较少,这种先降维再进行卷积处理的网络结构将不可避免地导致信息丢失;(2)在具体应用部署中,需要将网络

的要求,减小模型的部署难度。

为解决上述问题,降低计算成本,提高网络效率,该研究改进了原有特征提取网络。使用 DC3 模块替换原网络结构中的 C3 模块,有效解决了原始特征提取网络存在的信息丢失问题,并减少了网络计算量。DC3 模块的具体结构如图 4 所示,首先在特征提取模块 C3 中引入深度可分离卷积代替标准卷积,深度可分离卷积参数量较标准卷积参数量大幅下降,有效减少了网络计算量;然而使用深度可分离卷积替代标准卷积后,网络性能会有略微下降,因此,借鉴 MobileNeXt^[19] 的类沙漏残差网络构建思想,对残差瓶颈结构中的深度可分离卷积结构进行重新设计,将深度卷积层放在高维空间(即残差路径两端),使其能够提取更丰富的特征,将两个 1×1 的逐点卷积层置于中间,用于通道数的缩减和扩展;最后,保留跨阶段层次结构网络设计,将其和类沙漏残差网络有效结合起来,这样的结构设计既能够提取更多的特征信息,保留梯度信息的跨多层传播,减少信息损失,又缩减了参数量,有效减小了网络规模。

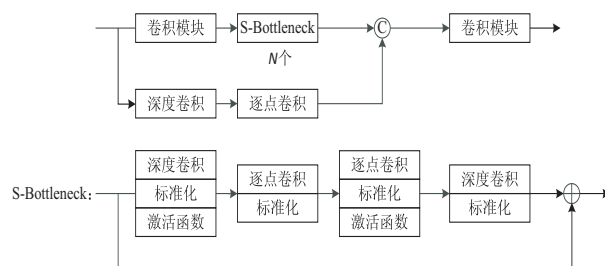


图 4 DC3 模块结构

2.2 聚合网络改进

图 5 中(a)、(b)、(c)分别为标注图, YOLOv5s 下采样 16 倍的特征图和检测结果,由图可得,经过高倍数下采样后, YOLOv5s 提取的中间特征图保留信息较少,最终导致目标定位不准确,出现了漏检、错检的情况。

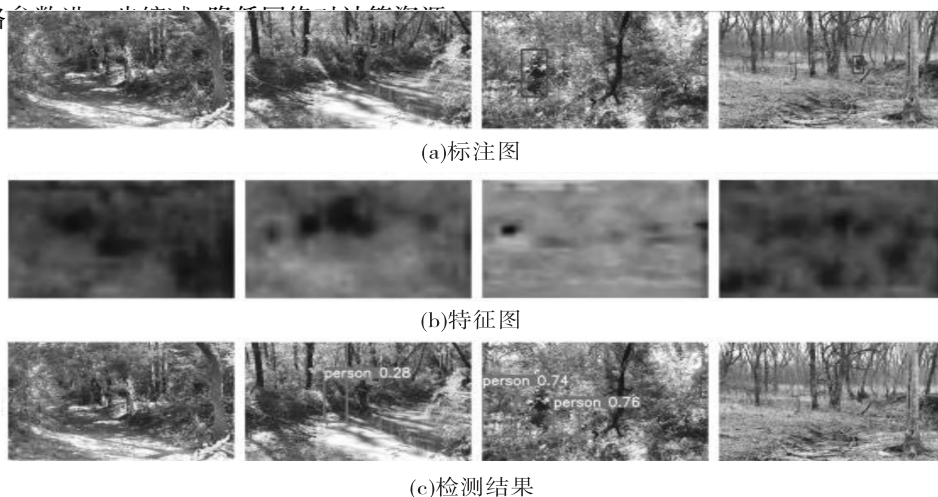


图 5 YOLOv5s 的检测结果和特征图

为解决特征图在高倍下采样后语义信息丢失严重的问题,该研究结合 Dynamic ReLU 激活函数和 CBAM 中的空间注意力机制,设计了 DAM 模块,将其嵌入聚合网络,提升模型的特征提取能力。

DAM 具体结构如图 6 所示,将 DC3 提取的特征分别送入两个分支进行处理:第一分支为 Dynamic ReLU 模块,对输入特征依次经过池化、全连接层、激活函数处理,通过学习输入特征图的全局信息自适应得到 ReLU 函数的 2 组系数 a_1, b_1, a_2, b_2 (a_1, a_2 为函数的斜率, b_1, b_2 为函数的截距);其次将输入特征图分别通过得到的两个 ReLU 函数进行处理并取最大值作为输出,使网络更好地学习输入特征图的全局信息,提升其特征表达能力。第二分支为空间注意力模块,首先对输入特征图的空间维度进行卷积、归一化、激活函数处理,得到所有空间区域的权重;其次,将上步所得权重与 Dynamic ReLU 模块处理后的输出相乘,使网络聚焦于重要区域的特征信息,进一步增强网络对重要特征信息的提取能力,使其能够更好地“捕捉”目标位置;最后将输出与 DC3 提取的特征进行融合,用于最后的检测。

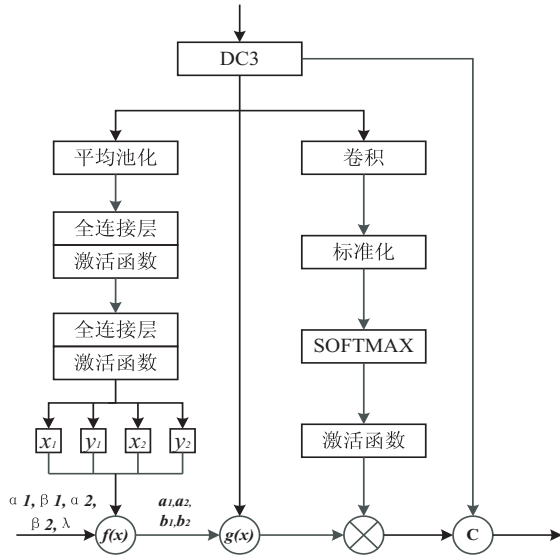


图6 DAM 模块结构

(图中, $\alpha_1, \beta_1, \alpha_2, \beta_2, \lambda$ 为超参数, $f(x) = [\lambda * x_1 + \alpha_1, y_1 + \beta_1, \lambda * x_2 + \alpha_2, y_2 + \beta_2]$, $g(x) = \max[a_1 * x + b_1, a_2 * x + b_2]$, C 表示 concat(拼接)操作)

图 7 中的(a)、(b)分别为 YOLOv5s 和添加 DAM 模块后的检测结果,(c)、(d)分别为 YOLOv5s 和添加 DAM 模块后的输出特征图。由图可得, YOLOv5s 对迷彩伪装目标的特征提取能力较弱,信息丢失严重,无法有效找到目标所在位置,从而出现了错检、漏检的情况;添加 DAM 模块后,提升了模型对迷彩伪装目标的特征提取能力,使模型能够更好地提取迷彩伪装目标的特征信息,对目标定位更加准确,检测精度得到一定

提高。

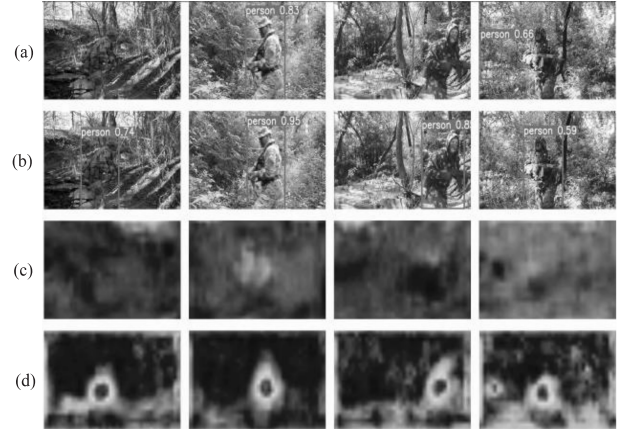


图7 YOLOv5s 和添加 DAM 模块后的检测结果和特征图

3 实验结果分析

3.1 数据集介绍

该研究在一个公开迷彩伪装数据集^[20]上进行实验。该数据集共有图片 3 278 张,包含丛林、草地、雪地、荒漠、开阔地 5 个场景,33 种迷彩样式,图像目标涵盖了大、中、小三种尺度和正面、侧身、背身、站立、半蹲、卧倒等多种姿态。数据集的典型样本如图 8 所示,(a)、(c)为原图,(b)、(d)为标注图。

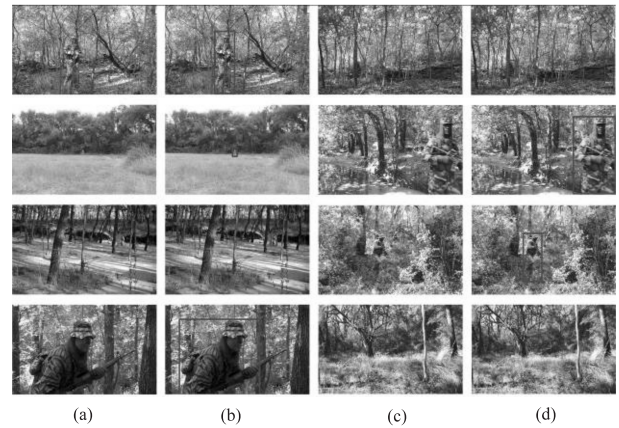


图8 数据集样本

3.2 评价指标

目标检测的预测结果共有 4 种,正样本被正确识别为正样本(True Positives, TP)、负样本被正确识别为负样本(True Negatives, TN)、负样本被错误识别为正样本(False Positives, FP)、正样本被错误识别为负样本(False Negatives, FN)。精确度(Precision)、召回率(Recall)计算分别如公式(5)、(6)所示:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

目标检测常用的性能评价指标为平均准确度均值

(Mean Average Precision, mAP), mAP 就是所有类别的平均准确度 (Average Precision, AP) 的均值, AP 的值即为 P-R 曲线下方的面积。

为了更好地反映算法检测性能,该研究没有采用 mAP@0.5 (交并比 (Intersection over Union, IoU) 为 0.5 时计算的 mAP) 作为评价指标,而是采用了与 coco^[21] 数据集一致的评价指标。该指标对算法检测性能评价更严格,使用 10 个 IoU 阈值 (0.5 至 0.95 间每隔 0.05 取一个值) 分别计算 mAP,然后取平均值得到最后的结果,并且该评价指标还将面积小于 32^2 , 大于 32^2 且小于 96^2 , 大于 96^2 像素的目标分别归为小目标、中目标、大目标,并计算这三类目标的 mAP 及平均召回率均值 (mean Average Recall, mAR), 从多个角度评价算法的检测性能。此外,实验还添加了每秒处理的图片数量 (FPS), 参数量 (Params) 和浮点运算次数

(Floating Point Of Operations, FLOPs) 三个指标来衡量算法的检测速度和网络结构大小。

3.3 实验环境及参数设置

文中算法在深度学习框架 pytorch1.7.1 上运行, 工作站配置为 Nvidia RTX 2080Ti 12G, Batch (批次) 设置为 32, 迭代次数为 300 轮, 图像大小为 512×288 像素, 初始学习率设置为 0.02, 使用 SGD (随机梯度下降) 优化方法, 动量值为 0.9, 采用了 mosaic 数据增强。对比算法包括 Faster RCNN、Cascade RCNN、SSD300、FCOS^[22]、ATSS^[23]、YOLOX^[24]、Deformable DETR^[25], 使用 MMDetection (一种目标检测算法工具箱) 复现的代码, 对比算法均使用默认参数。

3.4 实验结果与分析

CAMOU-YOLO 算法与原始算法及 7 种一阶段、二阶段检测算法的结果对比如表 1 所示。

表 1 不同算法的各类 mAP 比较

对比算法	mAP@0.5:0.95	mAP@0.5	mAP@0.75	mAP _S	mAP _M	mAP _L	FPS
Faster RCNN	0.568	0.896	0.636	0.397	0.567	0.709	8.3
Cascade RCNN	0.575	0.883	0.65	0.365	0.585	0.716	6.5
SSD300	0.527	0.886	0.545	0.382	0.512	0.686	53.3
FCOS	0.375	0.781	0.295	0.158	0.383	0.534	9.5
ATSS	0.533	0.887	0.561	0.324	0.553	0.663	12.5
YOLOX	0.597	0.902	0.69	0.442	0.584	0.729	28.6
Deformable DETR	0.589	0.902	0.693	0.427	0.589	0.716	4.5
YOLOv5s	0.595	0.905	0.672	0.418	0.601	0.715	131
CAMOU-YOLO	0.618	0.937	0.723	0.430	0.628	0.756	125

由表 1 结果可得, CAMOU-YOLO 算法较原始算法在 mAP@0.5:0.95、mAP@0.5、mAP@0.75、mAP_S、mAP_M、mAP_L 这 6 项指标上分别提高了 2.3%、3.2%、5.1%、1.2%、2.7%、4.1%, FPS 值有小幅降低, 但仍大幅领先于其他算法; 对比其他七种算法, CAMOU-YOLO 算法仅在 mAP_S 指标上稍低于 YOLOX 算法, 在其余评价指标上均处于领先地位, 多个指标的大幅提

升充分验证了 CAMOU-YOLO 算法较现有检测算法更能有效解决迷彩伪装目标检测任务。

为验证各改进模块对整体网络性能提升的有效性, 进一步开展了消融实验。逐步使用改进模块替换原始模块, 将实验结果分别跟原始算法性能进行比较, 结果如表 2 所示。

表 2 消融实验的结果比较

模型	map@0.5:0.95	mAP@0.5	mAP@0.75	mAP _S	mAP _M	mAP _L	mAR _S	mAR _M	mAR _L	Params	FLOPs
YOLOv5s	0.595	0.905	0.672	0.418	0.601	0.715	0.513	0.665	0.761	14.4 M	16.4 G
改进提取网络	0.607	0.935	0.697	0.41	0.615	0.742	0.524	0.681	0.783	10.6 M	11.8 G
改进聚合网络	0.614	0.928	0.72	0.425	0.627	0.749	0.524	0.685	0.786	15.3 M	16.8 G
CAMOU-YOLO	0.618	0.937	0.723	0.43	0.628	0.756	0.536	0.694	0.798	13 M	13.2 G

由表中结果分析可得, 改进的特征提取网络和聚合网络均能提升算法性能。单独改进特征提取网络, mAP@0.5:0.95、mAP@0.5、mAP@0.75、mAP_M、mAP_L、mAR_S、mAR_M、mAR_L 这 8 项指标上分别提高了

1.2%、3%、2.5%、1.4%、2.7%、1.1%、1.6%、2.2%, Params 减小了 26.4%, FLOPs 减少了 28%, 有效减小了网络模型, 在提升检测精度和召回率方面也有一定的贡献; 单独改进聚合网络, mAP@0.5:0.95、mAP@

0.5、 $mAP@0.75$ 、 mAP_s 、 mAP_M 、 mAP_L 、 mAR_s 、 mAR_M 、 mAR_L 这9项指标上分别提高了1.9%、2.3%、4.8%、0.7%、2.6%、3.4%、1.1%、2%、2.5%，在仅增加少量Params、FLOPs的情况下，显著提升了原始算法的检测精度和召回率；当两者结合使用时，既提升了算法的检测精度和召回率，又对网络结构进行了轻量化改进，较原始算法性能得到了较大的提升。

图9中(a)~(c)依次为标注图片、YOLOv5算法的检测结果和CAMOU-YOLO算法的检测结果，图中矩形框上方的白色部分为预测类别和置信度。由对比结果可得，CAMOU-YOLO算法对各种场景的目标及

不同尺寸的目标检测较原始算法的置信度更高；对比两者检测结果的矩形框可以发现，改进算法的检测矩形框更加接近标注图片，表明其定位更加精准，检测性能高于原始算法；此外，YOLOv5算法对一些目标还存在漏检的情况，而CAMOU-YOLO算法能够检测到目标，表明改进算法对迷彩伪装目标的特征提取更全面，减少了信息丢失，漏检问题得到了有效改善。综上所述，检测结果进一步验证了CAMOU-YOLO较原始模型性能提升明显，有效提升了迷彩目标特征提取能力，提高了检测精度和召回率，能较好地完成迷彩伪装目标检测任务。

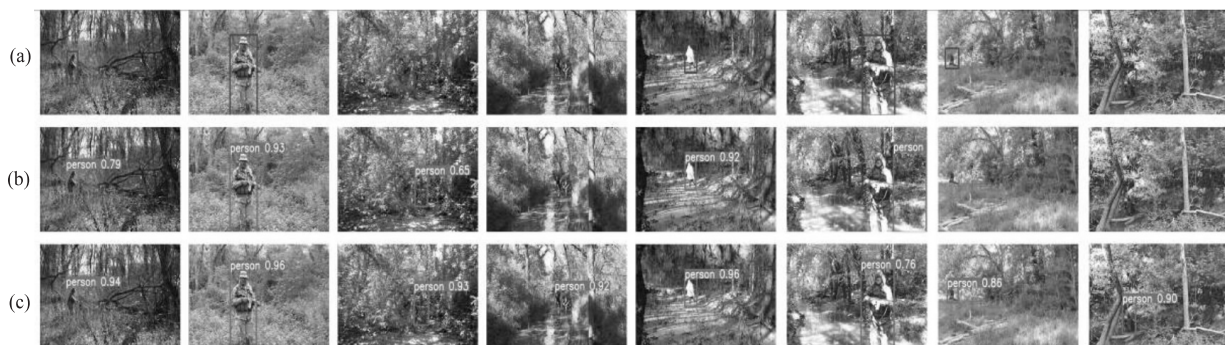


图9 YOLOv5和CAMOU-YOLO检测结果对比

4 结束语

针对现有目标检测算法无法高效、精准检测迷彩伪装目标的现状，基于YOLOv5模型，提出了一种结合深度可分离卷积和动态注意力机制的迷彩伪装目标检测模型CAMOU-YOLO。在原有模型中引入深度可分离卷积，借鉴类沙漏残差网络，设计了一种全新的特征提取模块——DC3模块，简化了特征提取网络，降低了网络参数数量和FLOPs值；结合Dynamic ReLU和注意力机制，设计了DAM模块，将其嵌入聚合网络中，从而有效增强了对迷彩伪装目标的特征提取能力，提升了检测定位精度。通过实验对比，验证了该模型的有效性，在较好达到网络轻量化要求的同时，检测性能也得到了显著提升，对迷彩伪装目标检测研究有一定的借鉴意义。

在未来的研究工作中，将继续优化模型结构，减少网络层数，进一步缩减网络规模，降低计算量，提升检测速度，推进迷彩伪装目标检测模型的发展应用。

参考文献:

- [1] 辛浩然, 张辉, 徐凡, 等. 基于模极大值算法的迷彩织物图像边缘检测[J]. 纺织科技进展, 2008(5): 15-16.
- [2] 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.
- [3] BHAJANTRI N U, NAGABHUSHAN P. Camouflage defect identification: a novel approach [C]//IEEE international conference on information technology. Bhubaneswar: IEEE, 2006: 145-148.
- [4] SENGOTTUVELAN P, WAHI A, SHANMUGAM A. Performance of decamouflaging through exploratory image analysis [C]//2008 first international conference on emerging trends in engineering and technology. Nagpur: IEEE, 2008: 6-10.
- [5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//IEEE conference on computer vision and pattern recognition. Columbus: IEEE, 2014: 580-587.
- [6] GIRSHICK R. Fast R-CNN [C]//IEEE international conference on computer vision. Santiago: IEEE, 2015: 1440-1448.
- [7] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [8] CAI Z, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection [C]//IEEE/CVF conference on computer vision and pattern recognition. Salt Lake City: IEEE, 2018: 6154-6162.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE, 2016: 779-788.
- [10] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//IEEE conference on computer vision and pattern

- recognition. Honolulu; IEEE, 2017; 6517–6525.
- [11] REDMON J, FARHADI A. YOLOv3: an incremental improvement [C]//IEEE conference on computer vision and pattern recognition. Washington; IEEE, 2018; 1–6.
- [12] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Scaled-YOLOv4: scaling cross stage partial network [C]//IEEE conference on computer vision and pattern recognition. Nashville; IEEE, 2021; 13024–13033.
- [13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C]//European conference on computer vision. Amsterdam; Springer, 2016; 21–37.
- [14] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2017-04-17]. <https://arxiv.org/pdf/1704.04861.pdf>.
- [15] WANG C Y, LIAO H, YE H I H, et al. CSPNet: a new backbone that can enhance learning capability of CNN [C]//IEEE/CVF conference on computer vision and pattern recognition workshops. Seattle; IEEE, 2020; 1571–1580.
- [16] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation [C]//IEEE/CVF conference on computer vision and pattern recognition. Salt Lake City; IEEE, 2018; 8759–8768.
- [17] CHEN Y, DAI X, LIU M, et al. Dynamic ReLU [C]//European conference on computer vision. Glasgow; Springer, 2020; 351–367.
- [18] WOO S, PARK J, LEE J Y, et al. Cbam: convolutional block attention module [C]//European conference on computer vision. Munich; Springer, 2018; 3–19.
- [19] DAQUAN Z, HOU Q, CHEN Y, et al. Rethinking bottleneck structure for efficient mobile network design [C]//European conference on computer vision. Glasgow; Springer, 2020; 680–697.
- [20] ZHENG Y F, ZHANG X W, WANG F, et al. Detection of people with camouflage pattern via dense deconvolution network [J]. IEEE Signal Processing Letters, 2019, 26(1): 29–33.
- [21] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context [C]//European conference on computer vision. Zurich; Springer, 2014; 740–755.
- [22] TIAN Z, SHEN C, CHEN H, et al. FCOS: fully convolutional one-stage object detection [C]//IEEE/CVF international conference on computer vision. Seoul; IEEE, 2019; 9626–9635.
- [23] ZHANG S, CHI C, YAO Y, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection [C]//IEEE/CVF conference on computer vision and pattern recognition. Seattle; IEEE, 2020; 9756–9765.
- [24] ZHENG G, SONGTAO L, FENG W, et al. YOLOX: exceeding YOLO series in 2021 [EB/OL]. [2021-08-06]. <https://arxiv.org/pdf/2107.08430.pdf>.
- [25] ZHU X, SU W, LU L, et al. Deformable detr: deformable transformers for end-to-end object detection [C]//International conference on learning representations. Vienna; ICLR, 2021; 1–16.
- +++++
- (上接第 28 页)
- 追踪的高光谱图像稀疏解混 [J]. 南京航空航天大学学报, 2019, 51(5): 577–585.
- [10] 袁 博. 空间与谱间相关性分析的 NMF 高光谱解混 [J]. 遥感学报, 2018, 22(2): 265–276.
- [11] BRUCKSTEIN A M, ELAD M, ZIBULEVSKY M. On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations [J]. IEEE Transactions on Information Theory, 2008, 54(11): 4813–4820.
- [12] HE W, ZHANG H, ZHANG L. Total variation regularized reweighted sparse nonnegative matrix factorization for hyperspectral unmixing [J]. IEEE Transactions on Geoscience & Remote Sensing, 2017, 55(7): 3909–3921.
- [13] QIAN Y, JIA S, ZHOU J, et al. Hyperspectral unmixing via $L_{1/2}$ sparsity-constrained nonnegative matrix factorization [J]. IEEE Transactions on Geoscience & Remote Sensing, 2011, 49(11): 4282–4297.
- [14] CANDÈS E J, WAKIN M B, BOYD S P. Enhancing sparsity by reweighted L_1 minimization [J]. Journal of Fourier Analysis & Applications, 2008, 14(5–6): 877–905.
- [15] 李 璠, 吴朝明, 张绍泉, 等. 变形 $L_{1/2}$ 正则化的高光谱图像稀疏解混 [J]. 激光与红外, 2021, 51(4): 515–522.
- [16] WANG Y, PAN C, XIANG S, et al. Robust hyperspectral unmixing with correntropy based metric [J]. IEEE Transactions Image Processing, 2015, 24(11): 4027–4039.
- [17] PENG J, SUN W, JIANG F, et al. A general loss-based non-negative matrix factorization for hyperspectral unmixing [J]. IEEE Geoscience and Remote Sensing Letters, 2020, 17(99): 1–5.
- [18] 汤 毅, 粘永健, 何 密, 等. $L_{1/2}$ 正则化的逐次高光谱图像光谱解混 [J]. 红外与激光工程, 2019, 48(7): 294–302.
- [19] QIAN Y, XIONG F, ZENG S, et al. Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery [J]. IEEE Transactions on Geoscience & Remote Sensing, 2017, 55(3): 1776–1792.
- [20] LI X, LU Q, DONG Y, et al. Robust subspace clustering by Cauchy loss function [J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(7): 2067–2078.
- [21] BECK A, TEOULLE M. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems [J]. IEEE Transactions on Image Processing, 2009, 18(11): 2419–2434.