

基于改进 EfficientDet 的雪豹红外相机图像检测方法

戴天虹, 刘超

(东北林业大学 机电工程学院, 哈尔滨 150040)

摘要:针对红外相机图像中雪豹存在难检测、难辨认等问题,提出了一个基于域迁移和新型注意力机制的 EfficientDet 雪豹检测算法。该算法首先采用图像增强来扩充训练样本提高模型的鲁棒性,并添加非雪豹图像优化数据集结构;其次,使用生成式对抗网络对夜间红外灰度图像进行域迁移,融合图像迁移前后的预测结果解决夜间红外图像目标识别能力弱的问题;最后通过在主干特征提取网络中加入注意力机制细化特征图来提高 RGB 和红外图像中雪豹的视觉显著性,并且改进特征融合网络结构整合更多有效信息。分析对比实验和消融实验的结果表明,该方法相比 Faster-RCNN、YOLOv3 和 SSD 的检测速度和精度更好,改进后的算法平均精确度为 97.4% 且检测速度可达 19 张/s,该检测模型更适合定位识别雪豹 RGB 与红外图像。

关键词:雪豹红外相机图像;目标检测;图像增强;域迁移;注意力机制

DOI:10.15938/j.jhust.2023.02.013

中图分类号: TP391.4 **文献标志码:** A **文章编号:** 1007-2683(2023)02-0108-09

Snow Leopard Infrared Camera Image Detection Method Based on Improved EfficientDet Model

DAI Tianhong, LIU Chao

(College of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150040, China)

Abstract: In view of the difficulty of snow leopard detection and recognition in infrared camera images, a snow leopard detection algorithm is proposed based on EfficientDet, which combines domain migration and new attention mechanism. Firstly, the algorithm adopts image enhancement to expand the training sample and adds non-snow leopard images to optimize the dataset structure to improve the robustness of the model. Secondly, the Generative Adversarial Networks is used to transfer the domain of night infrared gray image, and the image prediction results before and after migration are fused to solve the problem of weak target recognition ability in infrared image. Finally, the attention mechanism is added to the backbone feature extraction network to refine the feature map to improve the visual significance of snow leopard in RGB and infrared images, and the improved feature fusion network structure to integrate more effective information. The results of the contrast experiment and the ablation experiment show that the proposed method has better detection speed and accuracy than Faster-RCNN, YOLOv3 and SSD. The improved algorithm has an average accuracy of 97.4% and detection speed of 19 images/s. This detection model is more suitable for locating and recognizing snow leopard RGB and infrared images.

Keywords: infrared camera image of snow leopard; object detection; image enhancement; domain migration; attention mechanism

收稿日期: 2021-11-29

基金项目: 中央高校基本科研业务费专项资金(2572019CP17); 黑龙江省自然科学基金(C201414); 哈尔滨市科技创新人才项目(2014RFXXJ086)。

作者简介: 刘超(1997—),男,硕士研究生。

通信作者: 戴天虹(1963—),男,博士,教授,硕士研究生导师, E-mail: th_2000@sina.com。

0 引 言

雪豹被称为“雪山之王”,它们生活在寒冷、缺氧的高原地区,既是高山生态系统的旗舰物种和伞护物种,也是气候变化和水资源安全的指示物种。雪豹已被列入世界自然保护联盟濒危物种红色名录,随着全球气候变暖、猎物的减少、人兽冲突等因素对雪豹的生存构成日益严峻的威胁,对雪豹进行监测和保护目前刻不容缓,而科学高效的监测手段需建立在对基础数据的分析之上,因此精准、快速检测定位红外相机图像中的雪豹有助于生态学者进行种群评估和制定科学的保护应对方法。

传统红外相机监测方法^[1]目前仍面临诸多难题:监测影像大量冗余、人工检测识别负荷大、数据时效性差、预测分析不智能等。但近些年随着计算机视觉技术和红外成像技术的快速发展,基于深度学习的目标检测与识别算法在野生动物多样性调查、种群监测跟踪、种群密度评估等科研和保护工作中有着广泛的应用^[2-3]。刘文定等^[4]利用回归算法检测分割图像中野生动物区域并构建基于 VGG16 的全局-局部的双通道卷积神经网络模型自动识别马鹿、班羚等国家级陆生保护动物,该模型能有效减少复杂背景对动物识别的影响;Körschens 等^[5]使用 YOLO 自动定位识别大象的头部,来帮助生物学家识别野外遇到的大象;何嘉^[6]对在大型数据集上预训练的 SSD 模型进一步使用小型数据集迁移训练,得到适合云南地区的亚洲象检测模型;Schindler 等^[7]使用 Mask R-CNN 对夜间视频剪辑后的图像中的鹿等动物进行检测并识别其动作行为,该端到端模型可用于分析动物行为和统计种群数量;Yang 等^[8]在 RetinaNet 中添加时间和空间上下文信息模块检测视频中出现的大猩猩,此框架不仅可以检测受遮挡严重的目标还能同时处理多个视频帧。

近期对大型猫科动物如东北虎和雪豹的自动检测识别成为红外相机监测领域的研究热点:宫一男等^[9]使用微调(fine-tune)方法训练 YOLOv3 来检测识别昼(RGB)夜(灰度)红外相机图像包含的东北虎、东北豹等野生动物,但该算法对夜间红外图像的识别精度较低;史春姝等^[10]运用 SSD 检测并截取图像中虎的两侧和脸部,相比人工截取能节省大量时间;Kupyn 等^[11]基于改进的轻量级 FD-MobileNet 检测自然场景图像中的东北虎,该算法满足实时检测并且易于嵌入到边缘设备;Miguel 等^[12]应用鲁棒主成分分析方法、阈值化、二值形态学将红外相机图像

分为雪豹图像和其他类或空图像,但该算法较端到端模型处理流程多且复杂、计算开销大。上述研究主要聚焦于如何处理复杂背景对检测性能的影响,但目前大型猫科动物保护领域仍缺少对夜间红外灰度图像检测识别的针对性研究。

当前雪豹检测领域仍存在以下挑战:一方面,海量影像资料需要专业人员进行人工筛选和个体信息标注,传统人工标记方法耗费大量的时间和人力;另一方面,雪豹多在黄昏和夜间活动的特性导致红外相机记录的多是夜间由红外光源曝光的单色灰度图像,夜间红外灰度图像与日间 RGB 彩色图像相比具有分辨率差、对比度底、视觉效果模糊、含特征信息少等特点,现有模型对夜间红外灰度图像中的目标检测能力较弱。针对以上问题,本研究利用生成式对抗网络改善夜间红外图像的质量,通过数据增强和注意力机制提高模型对复杂背景下红外图像的检测能力。

1 本文算法

针对夜间红外图像存在的视觉效果模糊问题,本文使用生成式对抗网络 CycleGAN^[13]将夜间红外灰度域图像迁移到日间 RGB 彩色域;为增强雪豹在图像中的视觉显著性,本文提出了一种结合注意力机制的 EfficientDet^[14]检测方法。如图 1 所示,本文提出的模型主要包括域迁移、图像增强、坐标注意力机制模块、EfficientDet 检测模块 4 个部分。

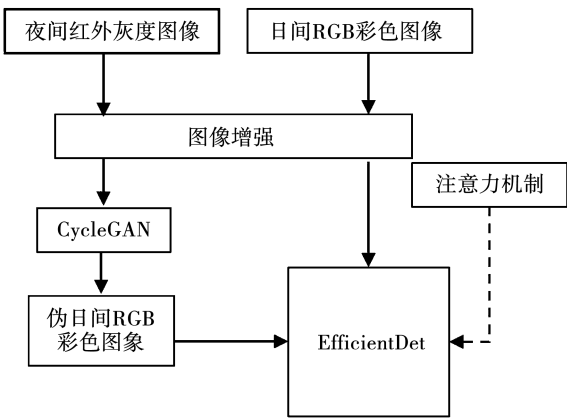


图 1 本文算法流程
Fig. 1 Algorithm flow of this paper

1.1 数据集及图像增强

本文雪豹的红外相机图像数据集来源于国际野生生物保护学会(wildlife conservation society, WCS),该数据集由 35 只雪豹的 790 张夜间红外灰度图像和 468 张日间 RGB 图像组成。

CNN 模型的预测能力受数据集的大小和丰富度影响,数据集越大越丰富训练得到的 CNN 模型提取图像特征的能力就越强,但在野生动物领域的计算机视觉应用过程当中通常没有充分的数据从头训练 CNN 模型。目前通常的做法是利用迁移学习^[15]把在大型数据集如 COCO 数据集上预训练的特征提取网络再使用自己的实验数据训练微调部分参数,由于 COCO 数据集和雪豹数据集的任务相似性较弱、域差异较大,单纯使用由雪豹组成的数据集直接进行迁移学习无法获得用于检测和识别的最佳网络模型,因此本文在纯雪豹数据集 S0 的基础上并入非雪豹数据集 D0 构建多元数据集 S1,用 S1 训练 EfficientDet 主干网络不仅能对雪豹进行准确的分类,还对图像中雪豹的高级语义特征具有很强的感知能力。由于东北虎与雪豹在外观和体型上具有一定的相似性且同为大型猫科动物,本文构建的非雪豹数据集 D0 全部是由来自动物和真实野外环境下的东北虎图像组成,各数据集的具体组成如表 1 所示。

表 1 检测数据集的组成

Tab. 1 Composition of detection dataset

数据集	训练集	测试集
S0	1 006	252
D0	1 500	375
S1	2 506	627

对纯雪豹数据集 S0 中的图像统计分析发现存在以下现象:背景环境高度杂乱、目标受其它物体遮挡、相机拖影现象产生高噪声、光照条件不受限制等,这些给雪豹检测带来了极大的困难,如图 2 所示。



图 2 难检测样本实例

Fig. 2 Samples difficult to detect

本文在 EfficientDet 训练阶段随机采用随机擦除、水平翻转、颜色噪声、放缩图片、HSV 色域扭曲等数据增强手段扩充训练样本,数据增强手段可以扩充复杂环境下的样本数量,增强模型的泛化能力和提高模型的鲁棒性,如图 3 所示。

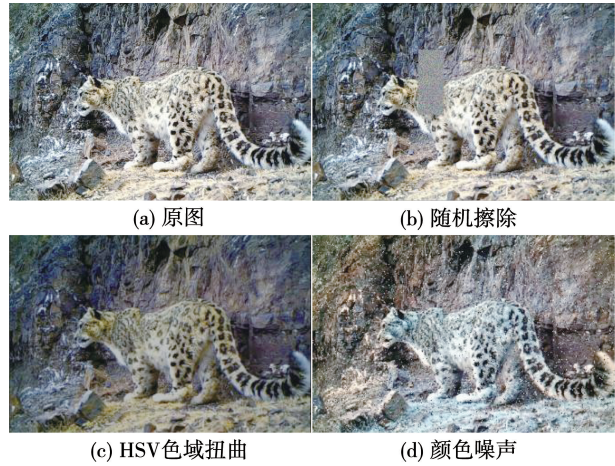


图 3 图像增强示意图

Fig. 3 Image enhancement examples

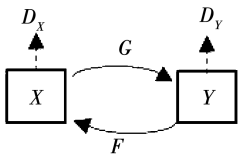


图 4 CycleGAN 网络原理图

Fig. 4 CycleGAN network diagram

1.2 夜间红外灰度图像域迁移

现在常用具备双光模式的红外相机进行非侵入式的野生动物监测:在日间拍摄 RGB 彩色图像,在夜间拍摄红外灰度图像。由于夜间红外灰度图像比日间 RGB 彩色图像的视觉效果差、辨识难度高、语义信息特征少,所以常规训练的检测模型从红外灰度图像中提取有用特征的难度更高。因此本文使用生成式对抗网络 CycleGAN 进行夜间红外灰度图像域迁移至日间 RGB 彩色域增加图像细节信息易于训练出更加健壮的检测模型。如图 4 所示, CycleGAN 由镜像对称的生成器 G 和 F 构成一个环形网络, G 将 X 域映射到 Y 域, F 则将 Y 域映射回 X 域。CycleGAN 与其他生成式对抗网络如相比无需成对图像训练即可进行源域到目标域图像的训练,因此适于野生动物监测领域的红外灰度图像域迁移。

D_X, D_Y 为判别器,其目标函数分别为:

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim p_{data}(y)} [\log D_Y(y)] + E_{x \sim p_{data}(x)} [\log(1 - D_Y(G(x)))] \quad (1)$$
$$L_{GAN}(F, D_X, Y, X) = E_{x \sim p_{data}(x)} [\log D_X(x)] +$$

$$E_{y \sim p_{\text{data}}(y)} [\log(1 - D_X(F(y)))] \quad (2)$$

CycleGAN 利用重构图和原始图计算 L1 loss,生成器 G 和 F 的循环一致损失函数为

$$L_{\text{cyc}}(G, F) = E_{x \sim p_{\text{data}}(x)} [\|F(G(x)) - x\|_1] + E_{y \sim p_{\text{data}}(y)} [\|G(F(y)) - y\|_1] \quad (3)$$

最终总体的目标函数为

$$L(G, F, D_X, D_Y) = L_{\text{GAN}}(G, D_Y, X, Y) + L_{\text{GAN}}(F, D_X, Y, X) + \lambda L_{\text{cyc}}(G, Y) \quad (4)$$

式中: λ 代表 2 个目标域的相对重要性权重。

CycleGAN 最终的优化目标为

$$G^*, F^* = \arg \min_{G, F} \max_{D_X, D_Y} L(G, F, D_X, D_Y) \quad (5)$$

本文使用源域为红外域和目标域为 RGB 域的图像训练 CycleGAN 后得到红外-RGB 图像生成器, 本文把夜间红外灰度图像经红外-RGB 生成器产生的新图像称为“伪日间 RGB 图像”。图 5 所示为部分图像的红外-RGB 生成器进行域迁移前后的示例, 很显然, 夜间红外图像中的目标与背景的区分度得到改善, 并且雪豹体征如斑点大小未受影响。

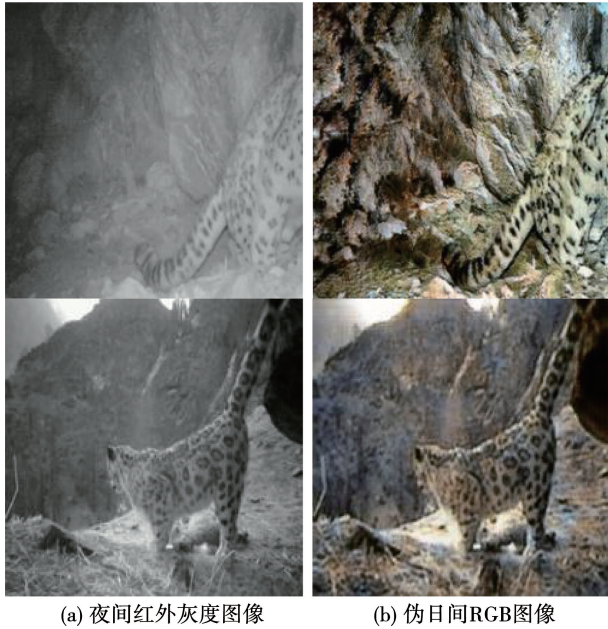


图 5 域迁移示例图

Fig. 5 Examples of domain migration

本文在训练过程中将同一图像的夜间红外图像和其伪日间 RGB 图像分别输入后续检测模型获得预测结果, 即夜间红外图像 X 经检测网络后输出目标检测框集合 T_i^X 、类别 S_i^X 和置信度 TS_i^X , 伪日间 RGB 图像 Y 输入检测网络后输出目标检测框合集 T_i^Y 、类别 S_i^Y 和置信度 TS_i^Y , 为进一步提高预测精度, 本文对两个预测结果进行融合, 计算公式如下:

$$m = \arg \max_i (TS_i^X), n = \arg \max_j (TS_j^Y) \quad (6)$$

$$R = \begin{cases} TS_m^X, TS_m^X > TS_n^Y \\ TS_n^Y, \text{otherwise} \end{cases} \quad (7)$$

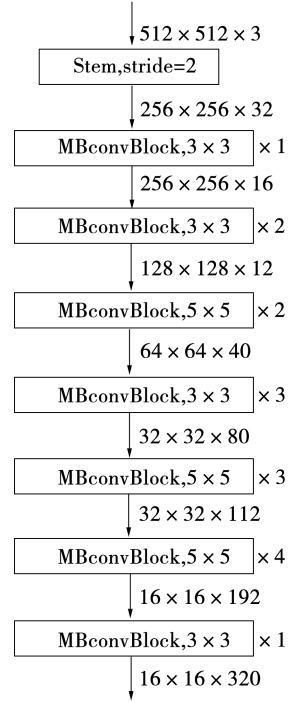


图 6 EfficientDet 主干特征提取网络结构图

Fig. 6 EfficientDet feature extraction backbone

1.3 EFFICIENTDET-D0 模型及改进

目前基于深度学习的目标检测算法主要分为以 R-CNN 系列为代表的两阶段算法和以 YOLO 系列、SSD 为代表的单阶段算法, 单阶段算法检测速度快但精度上逊色于两阶段算法。一个检测算法的效率是衡量其能否应用到实际中重要指标之一, 对此最近谷歌大脑团队专门提出一种基于回归的实时目标检测模型 EfficientDet, 每秒浮点运算 (FLOPS) 约为同类单级网络的 1/28, 轻量级的网络结构设计使其能准确高效地应用到边缘设备。

EfficientDet-D0 以 EfficientNet-B0^[16] 为主干特征提取网络。EfficientNet-B0 是由一个用于调整通道数的 Stem 卷积层和 16 个 MBConvBlock 堆叠而成, MBConvBlock 首先使用 1×1 的卷积对输入进行升维 ($\times$ 代表卷积运算), 然后使用深度可分离卷积 (depthwise separable convolution)^[17] 替代标准卷积操作来提取高维的特征语义信息, 深度可分离卷积能极大减少特征提取过程的参数数量进而降低计算开销, 最后将 MBConvBlock 的输入与上一步的输出逐元素相乘作为该运算块的最终输出, 如图 6 所示。

引入特征金字塔是提升的多尺度图像识别精度的有效途径之一^[18], 特征金字塔网络目前也在目标检测、语义分割、行为识别任务中对提升模型性能有

突出的表现。主干特征提取网络的底层特征图分辨率高包含较多的位置和细节信息,高层的特征图分辨率较低但具有更强的特征语义信息,因此融合不同层次的特征所得的特征图具有更多的细节信息和语义信息。如图 7 所示,经典的特征金字塔结构 (feature pyramid networks, FPN)^[19] 自顶向下进行多尺度特征融合, PANet^[20] (path aggregation network) 在 FPN 的基础上加入自底向上的融合路径,由于不同输入特征层具有不同的语义信息且对特征融合的最终输出贡献不同,所以 FPN 和 PANet 对处理输入特征平等处理的方式仍存在改进空间,因此 EfficientDet-D0 在其双向特征金字塔模块 Bi-FPN 中引入可学习的权值表示不同输入特征层的重要程度,本文为融合更多的特征且在保持很小计算开销的情况下添加了一个类似残差结构的跨层级连接,主干特征提取网络 EfficientDet-B0 产生的特征图经过 3 次自顶向下和自底向上的多尺度特征融合,最终得到含多尺度信息的聚合特征图送入下个阶段的分类框预测网络和边界框回归网络获得预测结果。本文考虑到雪豹检测任务的特殊性在 Bi-FPN 结构上作出进一步地改进,如图 8 所示,提出了一种新型的多跨级连接双向金字塔结构 (muti-cross BiFPN),此结构在仅增加少量计算的情况下引入额外的跨级连接,该结构不仅能够融合更多不同尺度的信息,还能改善 Bi-FPN 经过多次级联后信息流失的现象。

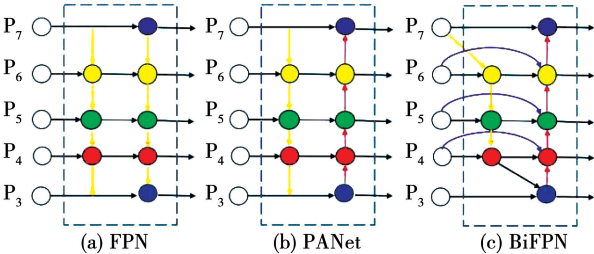


图 7 不同特征融合结构对比示意图
Fig.7 Comparison of different feature fusion structures

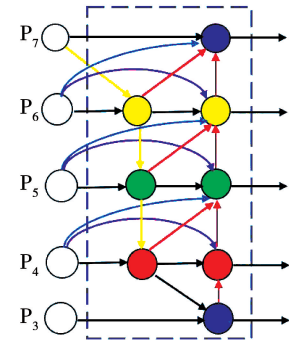


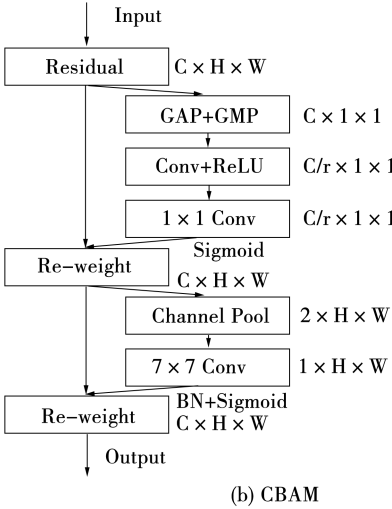
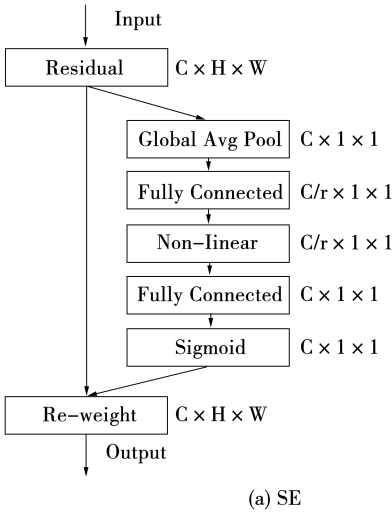
图 8 多跨级双向金字塔结构
Fig.8 Muti-cross Bi-FPN

1.4 坐标注意力机制

雪豹是一种具有很强的伪装能力的岩栖性动物,其体色与周围的环境极为相似且雪豹的姿势多变,在实际应用中使人工检测并识别红外相机图像中的雪豹变得尤为困难,如图 9 所示。有时单凭人的肉眼很难发现图像中确实存在的雪豹;此外,卷积神经网络提取的特征显著性易受红外相机拍摄角度、光线强弱、背景噪声等因素影响。针对上述问题,本文在卷积神经网络中添加注意力机制快速滤除图像中无用的信息并聚焦于感兴趣区域,提高网络对图像中关键信息的响应能力。



图 9 难发现样本示例
Fig.9 Hard to find sample instances



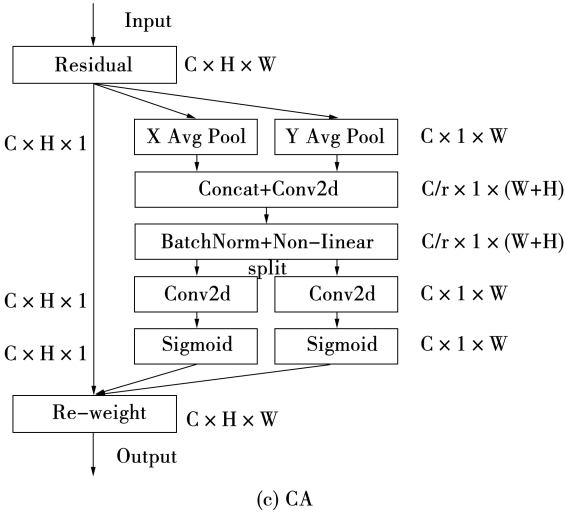


图 10 不同的注意力机制对比图

Fig. 10 Comparison of different attention mechanisms

目前轻量级神经网络主要采用 Squeeze-and-Excitation (SE)^[21] 和 Convolutional Block Attention Module (CBAM)^[22] 来提高模型的性能。如图 10(a) 所示, SE 仅对所有特征图生成通道注意力, 且每个输入特征图中的空间信息经全局平均池化后完全损失, 而空间信息对检测和分割任务尤为重要。如图 10(b) 所示, CBAM 首先对所有特征图做全局平均池化和全局最大池化生成通道注意力加权到原特征图, 然后再在通道上进行最大池化和平均池化得到两个包含空间信息的特征图, 最后尝试使用较大的卷积核融合空间信息, 虽然使用 7×7 的卷积能够获得更大的感受野, 但仍然无法覆盖全局获得任意相对位置的信息, 因此这种加权也仅考虑了局部范围的空间信息。针对上述问题, 本文引入了一种新颖且高效的坐标注意力机制^[23], 通过嵌入位置信息到通道注意力, 使移动网络获取更大区域的信息同时避免引入大的开销。

常用的通道注意力机制使用全局平均池化 (global average pooling, GAP) 来对空间信息进行全局编码, GAP 公式为

$$z_c = \frac{1}{HW} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad (8)$$

坐标注意力则对输入特征分别沿水平方向 X 和垂直方向 Y 聚合编码为一对包含位置信息的一维特征向量, 水平方向 X 得到的一维特征为

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (9)$$

垂直方向 Y 得到的一维特征为:

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (10)$$

将式(9)和(10)两个方向维度的池化结果拼接, 然后使用 1×1 的卷积压缩通道获得注意力特征图 $f \in \mathbb{R}^{(C/r) \times (H \times W)}$:

$$f = \delta(F_1([z^h, z^w])) \quad (11)$$

其中: F_1 为水平和池化结果的连接操作, r 为控制模块大小的一个超参数; δ 为非线性激活函数。

经过标准化处理和非线性激活函数后, 重新按照水平和垂直方向分为两组特征图, 再经 1×1 卷积融合后归一化获得加权系数, 最后加权到原始输入特征图获得坐标注意力输出特征图 $y_c(i, j)$:

$$g^h = \sigma(F_h(f^h)) \quad (12)$$

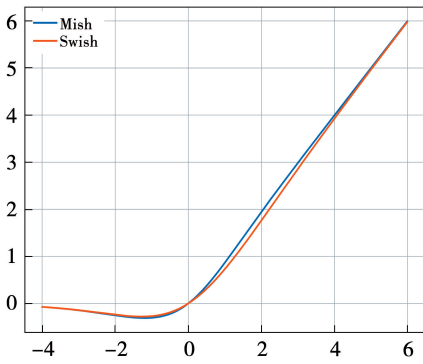
$$g^w = \sigma(F_w(f^w)) \quad (13)$$

$$y_c(i, j) = x_c(i, j) \times g^h(i) \times g^w(j) \quad (14)$$

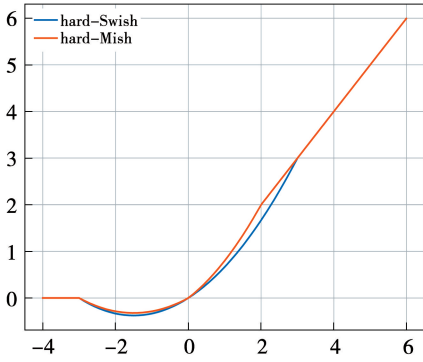
此外, 本文还对注意力机制中使用的非线性激活函数做了相关研究, 为解决 ReLU 的负值截断和梯度饱和现象, 如图 11(a) 所示, 激活函数 swish 在零点附近允许较小的负梯度流动, 而近期新提出的 Mish 激活函数被证明比 ReLU 和 Swish 的效果更好^[24], Mish 激活函数的光滑特性能进一步提高梯度下降效果。如图 11(b) 所示, 本文因受到 Swish 函数在移动网络计算开销大的而提出的 hard-Swish 函数这一过程的启发, 为了降低 Mish 激活函数中 tanh 在轻量级网络中的计算开销, 本文创新性地设计了 hard-mish 作为非线性激活函数应用到本文的检测网络模型之中:

$$f = \begin{cases} 0, & x \leq -3 \\ \frac{x(x+3)}{7}, & -3 < x \leq 0 \\ \frac{x(x+3)}{5}, & 0 < x \leq 2 \\ x, & x > 2 \end{cases} \quad (15)$$

综上, 坐标注意力机制分别沿两个空间方向聚合特征生成一对方向感知的特征映射, 这种注意力生成方式利用通道之间的关系和位置信息来精确定位感兴趣区域。如图 12 所示, EfficientDet 仅使用主干特征提取网络的 $P3$ 、 $P4$ 、 $P5$ 、 $P6$ 、 $P7$ 共 5 个不同尺度的特征图作为 Multi-cross BiFPN 结构的输入, 本文在以上 5 个有效特征层后分别嵌入 5 个坐标注意力 (coordinate attention, CA) 模块促使模型进一步自适应细化特征, BiFPN 继而能够融合更准确的特征信息, 最终提高 EfficientDet 对目标的定位能力。



(a) Swish、Mish激活函数曲线图



(b) hard-Swish、hard-Mish激活函数曲线图

图 11 激活函数 hard-Swish 和 hard-mish

Fig. 11 Activation function hard-Swish and hard-mish

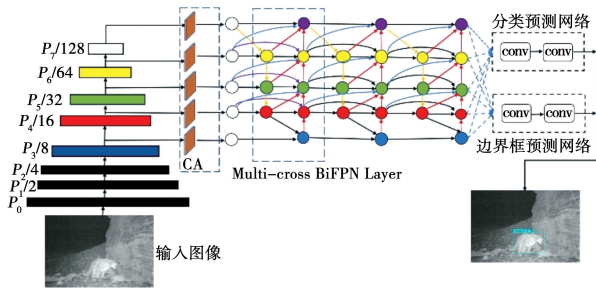


图 12 改进后的 EfficientDet-D0

Fig. 12 The improved structure of EfficientDet-D0

2 实验及结果分析

2.1 实验参数设置及评价指标

本文实验在 Intel (R) Core (TM) i3 - 9100F CPU、NVIDIA GTX2080Ti GPU (11G) 硬件环境下和 pytorch1.6.0、CUDA10.1 的 Ubuntu16.04 软件环境下进行。将训练时的 batch size 设为 8,采用 Adam 优化器,学习率和权重衰减系数分别为 0.001 和 0.0005。

在目标检测中一般采用平均精确度 (average precision, AP) 作为衡量模型检测能力的主要评估指标,AP 由精准率和召回率计算得出。查准率 (Preci-

sion) 和召回率 (Recall) 的计算公式为:

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

式中:TP 为预测为真阳性样本数量;FP 为真阴性样本数量;FN 为假阳性样本的数量。

AP 为精度-召回 (Precision-Recall) 曲线下方面积大小:

$$AP = \int_0^1 Precision \cdot Recall \cdot dr \quad (18)$$

2.2 实验与分析

为了验证 EfficientDet-D0 原始算法性能比其他常用基础检测算法的性能更优异,本文分别设置 Faster-RCNN、YOLOv3 和 SSD 作为对比实验,使用同一数据集 S1 进行训练和测试,对比实验的结果如表 2 所示。EfficientDet-D0 的雪豹检测 AP 值为 89.1%,分别比 Faster-RCNN、YOLOv3 和 SSD 的 AP 值高 4.5%、5.1%、3.0%,EfficientDet-D0 训练后的模型大小是 YOLOv3 的 1/16 且检测速度为 SSD 的 2.4 倍,EfficientDet-D0 的模型大小、检测速度和平均精度均优于其他 3 种算法。进一步通过消融实验论证本文图像数据增强、域迁移以及坐标注意力 (CA) 对 EfficientDet 检测精度的影响,以 EfficientDet-D0 为基准模型分别进行 4 组实验分析本文各模块的性能:第 1 组实验仅采用数据增强;第 2 组实验在第 1 组实验的基础上增加域迁移;第 3 组实验加在第 1 组实验基础上加入 CA,为了对比 CA 与其他模块对夜间红外图像的改进效果,值得注意的与第 2 组实验类似本文仅在训练阶段使用域迁移,在验证阶段使用 RGB 图像和红外图像验证训练后模型效果;第 4 组实验综合添加数据增强、域迁移和 CA。由表 3 可见,基准模型单纯使用数据增强后 AP 提高了 2.2%,证明扩充训练样本后模型的鲁棒性得到提高;使用数据增强和域迁移后的模型比基准模型 AP 高 4.5%,AP 比单纯使用数据增强提高 2.3%,证明域迁移能有效提升对夜间红外图像的检测精度;加入数据增强和 CA 后的 AP 比基准模型提高了 6.7%,比单纯使用数据增强 AP 提高 4.5%,证明 CA 能提高目标在日间 RGB 图像的夜间红外灰度图像中的视觉显著性;将三种改进手段同时加入后比基准模型的 AP 高 8.3%。实验表明数据增强对模型性能的贡献最小,分别加入域迁移和 CA 能明显提高模型精度,尤其是在使用数据增强的基准模型后加入 CA 模块比使用域迁移后 AP 提升幅度还要高 2.2%。

表 2 不同模型检测性能对比实验

Tab. 2 Parameters of simulation model			
算法	模型大小/MB	检测速度帧/s	AP/%
Faster R-CNN	108. 2	9	84. 6
YOLOv3	235. 1	17	84. 0
SSD	90. 7	8	86. 1
EfficientDet-D0	15. 1	19	89. 1

表 3 消融实验

Table 3 Ablation Experiment			
数据增强	域迁移	CA	AP/%
			89. 1
√			91. 3
√	√		93. 6
√		√	95. 8
√	√	√	97. 4

注:“√”代表对应实验采用该方法/模块。

由于在测试过程中使用日间 RGB 和夜间红外灰度图像验证 Efficient 的改进效果,但未确定 CA 模块对两种图像的影响程度是否存在差异,为更直观地比较 EfficientDet-D0 在加入 CA 前后模型特征提取能力的差异,本文通过在测试阶段的模型特征提取过程中添加类激活图^[25](class activation mapping, CAM)进行可视化实验来分析探究 2 类图像的哪些部分对识别过程起到贡献作用以及比较加入 CA 前后的差异。由于卷积神经网络最后的卷积层具有最抽象的目标语义信息,CAM 对最后的特征图进行通道级加权调整后将分类的输出映射回原图上生成具有类别信息的热力图,其表达式如下:

$$L_{CAM}^c = \text{ReLU}(\sum_{k=1}^n w_k^c A_k^L)$$

(19)

其中: A_k^L 为第 k 个通道的最高层 L 层特征图; w_k^c 为经 softmax 分类第 c 个神经元的权重。

类激活图中不同颜色揭示了网络对不同区域的关注程度,红色区域对网络的影响最大,颜色从红渐变至黄色至蓝色的渐变意味着区域对网络的影响逐渐变小。从图 13(b)上方两图可以看出,加入坐标注意力前,日间 RGB 图像与雪豹体色相近的部分岩石区域呈现红色,夜间红外灰度图像除雪豹外还有几部分光源和曝光比较严重的部分区域呈现红色或浅黄色,说明模型在特征提取过程中受背景信息的不程度地干扰;从图 13(b)下方两图可以看出,加入坐标注意力后,只有雪豹的躯干部分显示红色,表明添加坐标注意力后模型在精准响应感兴趣区域的同时能够有效抑制冗余背景信息的干扰,进而提取到更多关键信息。

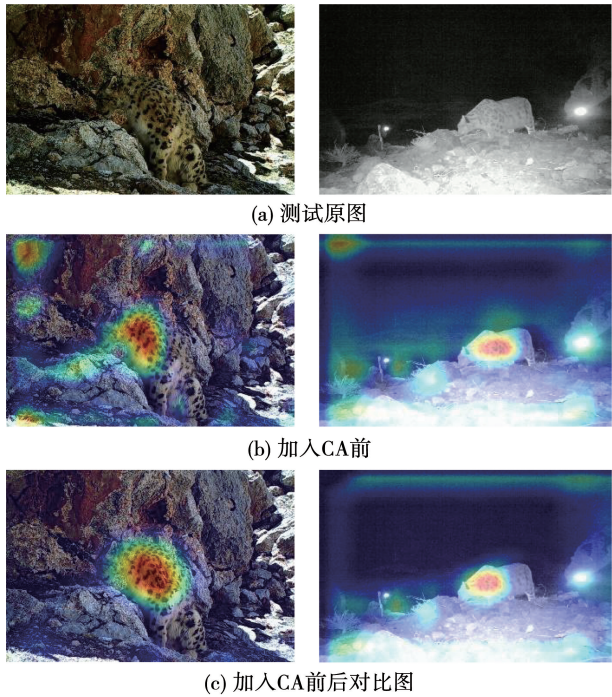


图 13 EfficientDet-D0 改进前后的类激活图对比
Fig. 13 CAM of CA improvement comparison

3 结 论

本文提出一种结合域迁移和注意力机制的 EfficientDet 雪豹检测方法,首先通过添加非雪豹图像改善迁移学习野生动物检测识别领域中存在的域差异问题,并且采用图像增强扩充样本增加模型的鲁棒性,利用生成式对抗网络对夜间红外域图像进行迁移并融合图像迁移前后的预测结果来进一步提高模型的检测精度。其次对于 EfficientDet 不仅改进其特征金字塔结构,还添加坐标注意力机制来进一步提高雪豹在日间 RGB 图像和夜间红外图像视觉显著性,本研究为自动检测具有杂乱背景的野生动物红外相机图像提供了新的解决方案。

参 考 文 献:

[1] 李治霖,康霭黎,郎建民,等. 探讨基于红外相机技术对大型猫科动物及其猎物的种群评估方法[J]. 生物多样性,2014,22(6):725.
LI Zhilin, KANG Aili, LANG Jianmin, et al. On the Assessment of Big Cats and Their Prey Populations Based on Camera Trap Data[J]. Biodiversity Science, 2014, 22(6): 725.
[2] SCHNEIDER S, TAYLOR G W, LINQUIST S, et al. Past, Present and Future Approaches Using Computer Vision for Animal Reidentification from Camera Trap Data[J]. Methods in Ecology and Evolution, 2019, 10

- (4): 461.
- [3] SCHNEIDER S, TAYLOR G W, KREMER S. Deep Learning Object Detection Methods for Ecological Camera Trap Data[C]//2018 15th Conference on Computer and Robot Vision (CRV). IEEE, 2018: 321.
 - [4] 刘文定, 李安琪, 张军国, 等. 基 ROI-CNN 的赛罕乌拉国家级自然保护区陆生野生动物自动识别[J]. 北京林业大学学报, 2018, 40(8): 123.
LIU Wending, LI Anqi, ZHANG Junguo, et al. Automatic Identification Method for Terrestrial Wildlife in Saihanwula National Nature Reserve in Inner Mongolia of Northern China Based on ROI-CNN[J]. Journal of Beijing Forestry University, 2018, 40(8): 123.
 - [5] KÖRSCHENS M, BARZ B, DENZLER J. Towards Automatic Identification of Elephants in the Wild[J]. arXiv Preprint arXiv:1812.04418, 2018.
 - [6] 何嘉. 基于深度学习的野生动物智能检测与识别[D]. 深圳: 深圳大学, 2019.
 - [7] SCHINDLER F, STEINHAGE V. Identification of Animals and Recognition of Their Actions in Wildlife Videos Using Deep Learning Techniques[J]. Ecological Informatics, 2021, 61: 101215.
 - [8] YANG X, MIRMEHDI M, BURGHARDT T. Great Ape Detection in Challenging Jungle Camera Trap Footage Via attention-based Spatial and Temporal Feature Blending [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019: 255.
 - [9] 宫一男, 谭孟雨, 王震, 等. 基于深度学习的红外相机动物影像人工智能识别: 以东北虎豹国家公园为例[J]. 兽类学报, 2019, 39(4): 458.
GONG Yinan, TAN Mengyu, WANG Zhen, et al. AI Recognition of Infrared Camera Image of Wild Animals Based on Deep Learning: Northeast Tiger and Leopard National Park for Example[J]. Acta Theriologica Sinica, 2019, 39(4): 458.
 - [10] 史春妹, 谢佳君, 顾佳音, 刘丹, 姜广顺. 基于目标检测的东北虎个体自动识别[J]. 生态学报, 2021, 41(12): 4685.
SHI Chunmei, XIE Jiajun, GU Jiayin, et al. Amur Tiger Individual Automatic Identification Based on Object Detection[J]. Acta Ecologica Sinica, 2021, 41(12): 4685.
 - [11] KUPYN O, PRANCHUK D. Fast and Efficient Model for Realtime Tiger Detection in the Wild[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019: 310.
 - [12] MIGUEL A, BEARD J S, Bales-Heisterkamp C, et al. Sorting Camera Trap Images [C]//2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP). IEEE, 2017: 249.
 - [13] ZHU J Y, PARK T, ISOLA P, et al. Unpaired Image-to-image Translation Using Cycle-consistent Adversarial Networks[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 2223.
 - [14] TAN M, PANG R, LE Q V. Efficientdet: Scalable and Efficient Object Detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781.
 - [15] 王莉莉, 冯其帅, 陈德运, 等. 一种基于正则化判别分析的迁移学习算法[J]. 哈尔滨理工大学学报, 2019, 24(2): 89.
WANG Lili, FENG Qishuai, CHEN Deyun, et al. A Transfer Learning Algorithm Based on Regularized Discriminant Analysis[J]. Journal of Harbin University of Science and Technology, 2019, 24(2): 89.
 - [16] TAN M, LE Q. Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks [C]//International Conference on Machine Learning. PMLR, 2019: 6105.
 - [17] CHOLLET F. Xception: Deep Learning with Depthwise Separable Convolutions [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251.
 - [18] 赵升, 赵黎. 基于双向特征金字塔和深度学习的图像识别方法[J]. 哈尔滨理工大学学报, 2021, 26(2): 44.
ZHAO Sheng, ZHAO Li. On Image Recognition Using Bidirectional Feature Pyramid and Deep Neural Network [J]. Journal of Harbin University of Science and Technology, 2021, 26(2): 44.
 - [19] 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.
 - [20] WANG K, LIEW J H, ZOU Y, et al. Panet: Fewshot Image Semantic Segmentation with Prototype Alignment [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 9197.
 - [21] HU J, SHEN L, SUN G. Squeeze-and-excitation Networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132.
 - [22] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional Block Attention Modul[C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3.
 - [23] HOU Q, ZHOU D, FENG J. Coordinate Attention for Efficient Mobile Network Design [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713.
 - [24] MISRA D. MISH: A Self Regularized Nonmonotonic Neural Activation Function [J]. arXiv Preprint arXiv: 1908.08681, 2019.
 - [25] 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.