

研究论文 ORIGINAL PAPERS

基于改进孪生网络的 野生动物单目标跟踪方法

殷子璇¹, 赵亚琴^{1*}, 肖治术², 肖文宏², 虞秋萍¹, 许智皓¹

(1. 南京林业大学机械电子工程学院, 南京 210037;
2. 中国科学院动物研究所农业虫害鼠害综合治理研究国家重点实验室, 北京 100101)

稿件运行过程

收稿日期: 2024-10-11

修回日期: 2024-11-15



关键词: 单目标跟踪;
野生动物;
孪生网络;
注意力机制

Keywords: Single target tracking;
Wildlife;
Siamese network;
Attention mechanism

中图分类号: Q958.1; TP391.4

文献标志码: A

文章编号:

2310-1490(2025)-03-0533-11

DOI: 10.12375/ysdwxb.20250307

摘要

野生动物作为生态系统的重要组成部分,其动态监测对于维系生态平衡、理解物种间相互作用及评估生态系统健康状况具有至关重要的意义。野生动物监测主要通过无人机机载相机和固定的红外相机来捕捉动物的自然行为。然而,由于野生动物行为的不可预测性,在实际跟踪过程中,常会出现目标较小、多尺度变化以及动物身体被遮挡等问题。为了应对这些挑战,提出一种基于改进孪生网络的动物目标跟踪方法,将跟踪问题转化为相似性学习问题。在孪生关系网络(SiamRN)的特征提取阶段引入多头注意力机制,包括串联窗口自注意力运算和滑动窗口自注意力运算,增强模型对小目标的精准跟踪能力。同时,多头注意力机制的引入降低了网络的参数量和复杂度,提高了运算效率。在公开数据集和自制数据集上进行实验,结果表明本研究采用的野生动物跟踪方法的成功率和准确率分别为0.698和0.928,优于主流的孪生网络跟踪方法,该方法能够准确跟踪和定位野生动物目标,实现野生动物监测。

Single Target Tracking Method for Wildlife Based on Improved Siamese Network

YIN Zixuan¹, ZHAO Yaqin^{1*}, XIAO Zhishu²,
XIAO Wenhong², YU Qiuping¹, XU Zhihao¹

基金项目: 国家自然科学基金面上项目(32371583)

第一作者简介: 殷子璇(2000—),女,博士研究生;主要从事计算机视觉相关研究。E-mail: yinzixuan0929@163.com

*通信作者: 赵亚琴, E-mail: yaqinzhao@163.com

- (1. College of Mechanical and Electronic Engineering, Nanjing Forestry University, Nanjing 210037, China;
2. State Key Laboratory of Integrated Management of Pest Insects and Rodents in Agriculture, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China)

Abstract: As an important component of the ecosystem, the dynamic monitoring of wildlife is of great significance for maintaining ecological balance, understanding species interactions, and assessing the health status of the ecosystem. Wildlife monitoring mainly relies on unmanned aerial vehicle (UAV) onboard cameras and fixed infrared cameras to capture the natural behavior of animals. However, due to the unpredictability of wildlife behavior, there are issues with small targets, multi-scale variations, and animal body occlusion in the actual tracking process. To address these challenges, this paper proposed an animal target tracking method based on an improved Siamese network, which transforms the tracking problem into a similarity learning problem. Introducing multi head attention mechanism in the feature extraction stage of Siamese relation network (SiamRN), including concatenated window self-attention operation and sliding window self-attention operation, the precise tracking ability of small targets is enhanced. At the same time, the introduction of multi head attention mechanism reduces the number and complexity of network parameters and improves computational efficiency. The experiment was conducted on both public and self-made datasets, and the results showed that the success rate and accuracy of the wild animal tracking method used in this paper are 0.698 and 0.928, respectively, which was superior to mainstream Siamese network tracking methods. The method proposed in this paper can accurately track and locate wildlife targets, achieving wildlife monitoring.

野生动物作为自然生态的重要组成部分,其保护对于维持生态平衡至关重要^[1]。如何有效监测动物行为、健康状态以及种群活动等是野生动物保护的关键。传统的动物监测方法虽然能为相关工作提供一定的数据支持,但受限于人力资源和复杂自然环境,难以持续、准确地监测动物种群动态^[2]。随着科技的发展,现代野生动物监测技术结合自动化和智能化技术,能够全天候地实现野生动物的自动、快捷和准确监测。

早期野生动物追踪采用射频识别(radio frequency identification, RFID)技术,RFID技术通过给动物植入耳标、佩戴项圈或腿带等装置,实现了对动物个体的识别和追踪^[3]。GPS定位装置也被用于捕获、传输动物的位置数据^[4]。Ullmann *et al.*^[5]还利用加速度计记录了动物运动数据,进而分析动物的行为模式。此外,结合北斗与GPS定位技术的卫星追踪器也被应用于鸟类的救助与放飞过程^[6-7]。这些方法丰富了野生动物监测方式,为野生动物保护提供了重要的数据。但值得注意的是,这些方法需要动物携带RFID或GPS装置,这些携带的装置可能会对动物造成一定的压力,进而影响其自然行为。

计算机视觉和深度学习技术在人工智能领域发展迅速,能够在数据收集的同时,最大限度地减少对野生动物的干扰,因此,其展现出了巨大的潜力^[8]。张雪莹等^[9]展示了深度学习技术在野生动物监测中的显著进展。在野生动物的检测和识别方面,杨帆^[10]通过引入注意力机制改进YOLOv3算法,有效提升了对野生动物的检测和识别能力。Kim *et al.*^[11]利用Faster R-CNN模型对26种亚马逊鸚鵡进行了分类。黄志静等^[12]提出了一种基于深度残差收缩网络的野生动物识别模型,该模型通过减少图像中的噪声干扰来提高野生动物图像识别的准确性。这些研究成果不仅推动了野生动物保护工作的发展,也为计算机视觉技术在生态保护领域的应用提供了新的思路和方法。

近年来,深度学习也被应用于野生动物的单目标跟踪任务中^[13],例如,阎婧宇等^[14]以YOLOv7-Tiny和多目标跟踪器Bot-SORT分别作为检测和跟踪的基础框架,应用于无人机的动物跟踪,以应对野外环境中的跟踪挑战。马光凯等^[15]针对圈养动物,提出了GhostNet-DeepSORT算法,用于自动跟踪东北虎(*Panthera tigris altaica*)幼崽群体活动。在跟踪技术

的研究中,孪生网络被广泛采用。许群等^[16]提出了基于注意力特征融合的孪生网络跟踪框架,用于精准跟踪东北虎个体。

众多研究聚焦于基于孪生网络的跟踪方法,例如,SiamFC (fully-convolutional Siamese networks) 算法将跟踪任务定义为相似性匹配任务,以较快的速度实现了目标跟踪任务^[17];SiamRPN (Siamese region proposal network) 把孪生网络与区域推荐网络融合,将目标跟踪问题抽象成为样本检测问题,提高了准确率与运算速度^[18]。SiamBAN (Siamese box attention network) 去掉了图像中可能包含目标对象的初始猜测,降低了模型参数,提升了速度,但是在复杂情况下跟踪的辨别能力不足^[19]。孪生关系网络 (Siamese relation network, SiamRN)^[20]引入了2个关键模块:关系检测器和精化模块,以抑制假阳性检测并过滤背景干扰,从而显著提升了目标跟踪的准确性和鲁棒性。但是,这些研究是在特定场景下的跟踪。野生动物跟踪会面对动物快速移动、目标较小和尺度变化等情况。此外,复杂多变的野外环境也会增加跟踪的难度,因此,需对现有孪生网络的跟踪方法进行改进,以便应用于野生动物跟踪。

本研究将野生动物监测与孪生网络跟踪技术

相结合,开发了一种针对野生动物个体的单目标跟踪框架。在此框架中,选择SiamRN网络为基础网络,并在特征提取的主干网络部分引入Swin-Transformer^[21],减少网络模型参数,同时保持高效的特征提取能力。此外,还在特征提取过程中融入多头注意力机制,进一步增强图像的分类能力,使模型在复杂背景下的目标跟踪更加准确和稳定,进而适应多样化的生态环境。

1 研究方法

1.1 数据集构建

本研究建立了包括麋鹿(*Elaphurus davidianus*)、虎(*Panthera tigris*)、狮(*Panthera leo*)、马、企鹅及其他鸟类共6类动物的数据集。其中,其他鸟类数据集选取了鸟类公开数据集中相机拍摄的部分鸟类可见视频^[22]。麋鹿数据集在中国盐城大丰拍摄,其余动物视频从纪录片中获取。所有数据集共计60个视频,产生19 302张图像帧。将整个数据集按照7:3的比例划分为训练集和验证集。其中,训练集包含13 512张图像帧,验证集包括5 790张图像帧。数据集样例如图1所示。



图1 数据集样本

Figure 1 Sample dataset

在构建数据集时,采用 Labelimg 对每一帧图像中的动物目标区域的边界框中心点坐标和宽、高进行标注^[23]。此外,采用数据增强技术对图像进行旋转、缩放和平移等操作,增加训练集的多样性,以便使模型应用于复杂多变的场景中。

1.2 SiamRN 结构分析

孪生神经网络以 2 个样本为输入,输出其嵌入高维度空间的表征,以比较 2 个样本的相似程度。在目标跟踪任务中,孪生网络的 2 个分支共享相同的权重,其中一个分支用于提取目标的特征,另一分支用于提取背景的特征^[24]。SiamRN^[20]是一种基于此原理构建的孪生神经网络模型,专门针对目标跟踪任务进行了优化设计。它使用深度残差网络 ResNet50 作为骨干网络,从图像中提取多尺度特征,使用基于区域的卷积神经网络的级联结构对位置框进行分类。SiamRN 引入关系检测器(relation detector),关系检测器由 3 个不同的非线性可学习比较器组成,即全局检测器、局部检测器和 patch 检测器。这些检测器用于比较目标生成的特征与多尺度回归特征之间的关系。全局检测器旨在通过深度嵌入来

比较它们的全局特征,局部检测器在像素和通道学习更详细的比较,patch 检测器用来学习不同 patch 之间的关系。精细化模块(refinement module)将关系检测器的输出转换为匹配分数,随后在匹配分数和分类分支的互相关特征图之间进行复制,最后使用卷积运算生成精细的分类置信得分。精细化模块可以将关系检测器有效集成到 Siamese 框架中,以抑制假阳性检测并过滤背景干扰。

1.3 模型改进

SiamRN 网络结构参数量较大,导致其在实际应用中难以实现高效跟踪。因此,本研究在特征提取阶段引入 Swin-Transformer block,利用 Swin-Transformer block 的多头自注意力机制增强模型对输入的关注程度,使模型能够更有针对性地处理不同区域的信息,并基于窗口级别进行自注意力计算。改进后的 SiamRN 网络不仅能够有效减少模型参数量,降低计算复杂度,还保持了强大的特征提取能力,有助于在保证跟踪精度的同时,提升模型的运行效率和适应性。模型改进的整体结构如图 2 所示。

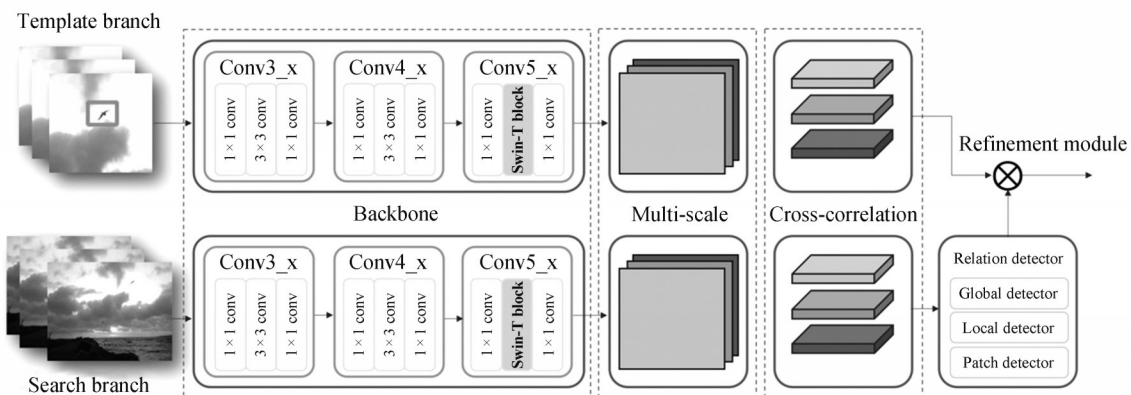


图 2 跟踪模型网络结构

Figure 2 Tracking model network structure

输入一对野生动物图像,初始帧被裁剪为 127×127 的图像用做模板图像,搜索分支的每帧图像被裁剪为 255×255 ,使用改进后的 ResNet50 进行特征提取,抽取 Conv3、Conv4 和 Conv5 输出的特征图,获得多尺度特征图,然后将模板的特征与搜索图像的特征进行互相关运算,得到分类和回归结果。将回归结果送入关系检测器中进行运算,度量回归结果得到目标动物与模板之间的关系并与分类结果在精细化模块融合,最终得到当前帧目标动物的

位置。

1.3.1 多尺度特征提取

SiamRN 主干网络通常使用 ResNet50 来提取图像特征,但是对小目标的识别能力较弱,对视频图像中连续帧的动物位置变化不够敏感且计算复杂。此外,无人机机载处理器资源需要平衡跟踪精度和运算速度。为了适应上述场景,本研究引入 Swin-Transformer 中的多头自注意力机制。与传统的多头自注意力模块不同,Swin-Transformer 是基于移位窗

口构建的。通过串联窗口自注意力运算(W-MSA)以及滑动窗口自注意力运算(SW-MSA)来提高模型对图像中不同范围信息的获取能力,通过基于窗口级别进行的自注意力计算,降低计算复杂度。

图3是主干网络的特征提取部分,在ResNet50的基础上,改进了原网络的残差结构,将Bottleneck-block的Conv5的 3×3 卷积替换为3个Swin-Transformer block的级联结构。每个Swin-Transformer block是由LayerNorm(LN)层、多头自注意力模块、残差连接和具有GELU非线性的2层感知器(MLP)组成。Swin-Transformer block可以公式化为

$$\hat{z}^l = W\text{-MSA}(LN(z^{l-1})) + z^{l-1}, \quad (1)$$

$$z^l = MLP(LN(\hat{z}^l)) + \hat{z}^l, \quad (2)$$

$$\hat{z}^{l+1} = SW\text{-MSA}(LN(\hat{z}^l)) + \hat{z}^l, \quad (3)$$

$$z^{l+1} = MLP(LN(\hat{z}^{l+1})) + \hat{z}^{l+1}. \quad (4)$$

式中: \hat{z} 和 z 分别为W-MSA模块和MLP模块的输出特征, l 为第 l 个block。

在Swin-Transformer block中,自注意力机制通过查询、键和值矩阵来计算注意力权重。自注意力机制的计算公式为

$$attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}} + B\right)V. \quad (5)$$

式中: $Q, K, V \in \mathbb{R}^{M^2 \times d}$ 分别为查询矩阵、键矩阵和值矩阵, $M \times M = M^2$ 为窗口的patch数量; d 为查询/键矩阵维度; $B \in \mathbb{R}^{M^2 \times M^2}$,是相对位置偏置矩阵,通过参数化一个较小的矩阵 $\hat{B} \in \mathbb{R}^{(2M+1) \times (2M+1)}$ 生成。

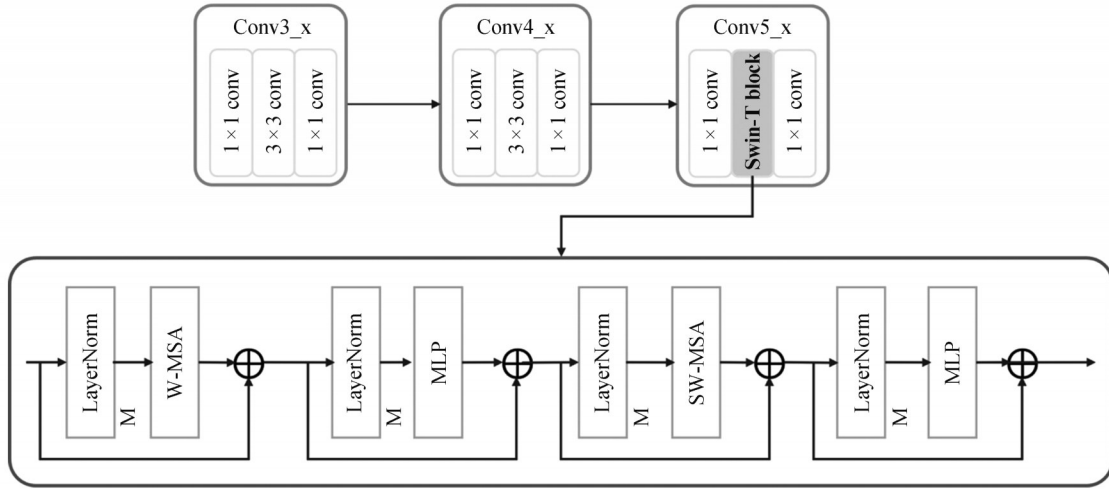


图3 特征提取网络

Figure 3 Feature extraction network

特征图采用线性层降采样,通道变为原始特征图通道数的2倍,使用 2×2 卷积核,步长为2。W-MSA在每个窗口内部实现多头自注意力机制,分块处理大规模输入,在减少计算量的同时,可捕捉动物图像的局部特征关系,并在softmax之前加上可学习的相对位置编码(位置编码将绝对位置编码成一个表示位置的向量),将注意力计算限制在每个窗口内来减少计算量。SW-MSA通过滑动窗口的方式在一定程度上建模了长距离的全局关系,使得改进后的ResNet50模型在处理远距离像素之间的关系时保持一定的上下文信息。

在特征提取过程中,不同层级的卷积层负责捕捉动物图像不同尺度的特征。利用Conv3、Conv4和

修改后的Conv5的特征获得动物图像多尺度特征。Conv3提取的是图像局部形状和纹理组合特征,Conv4提取的是高级局部特征及目标的部分结构,如动物的眼睛等面部特征,Conv5位于ResNet50的第3个残差块,它属于网络中的深层卷积层,负责捕捉高层次的语义特征。如图3所示,在特征提取网络中,输入的动物视频帧先经过Conv3、Conv4两个卷积层,然后在Conv5阶段首先通过 1×1 卷积核降维来减小计算负担,之后再输入到Swin-Transformer block中。在Swin-Transformer block中,特征被划分为多个局部窗口,每个局部窗口内的特征通过自注意力机制进行信息交互。通过移位窗口机制,相邻窗口之间的特征也进行信息交互,从而在全局范围

内捕获动物特征之间的依赖关系,这对于处理图像中大范围区域内动物的位置变化更有优势。此外,通过多层级的注意力机制,自动聚焦动物身体的关

键特征,忽略图像中的背景噪声和其他干扰因素,最后使用一个 1×1 卷积层升维,将特征维度提升回到原来的维度。主干网络参数见表1。

表1 主干网络参数

Table 1 Backbone network parameters

阶段 Stage	输入尺寸 Input size	卷积核尺寸 Kernel size	步长 Stride	填充 Padding	输出尺寸 Output size
Conv1	255×255	7×7	2	3	128×128
	127×127	3×3	2	1	64×64
Conv2	64×64	1×1	1	0	64×64
	64×64	3×3	1	1	64×64
	64×64	1×1	1	0	64×64
Conv3	64×64	1×1	1	0	64×64
	64×64	3×3	2	1	32×32
	32×32	1×1	1	0	32×32
Conv4	32×32	1×1	1	0	32×32
	32×32	3×3	1	1	32×32
	32×32	1×1	1	0	32×32
Conv5	$32 \times 32^*$	1×1	1	0	32×32
	16×16	1×1	1	0	16×16

注: *. 该层为Swin-Transformer block,由LayerNorm(LN)层、多头自注意力模块、残差连接和具有GELU非线性的2层MLP组成。

Note: *. This layer is a Swin-Transformer block, consisting of a LayerNorm(LN) layer, a multi head self attention module, residual connections, and a 2-layer MLP with GELU nonlinearity.

1.3.2 相似度度量

从多个阶段输出的特征图提取不同层级的特征后,基于区域卷积神经网络的级联结构对特征进行分类。由于分类和回归分支是独立更新的,会导致分类置信度最高的位置对应的方框并不是跟踪目标最准确的方框,因此,通过关系检测器比较目标生成的特征与多尺度回归特征之间的关系,解决区域和目标不匹配的问题。

首先基于互相关生成特征图上每个点对应的回归参数。这些参数用于在原始图像上截取出目标区域。接着,利用ROI Pooling技术对这些区域进行池化,随后,将池化后的区域特征分别通过3个不同的检测器(全局检测器、局部检测器和patch检测器)进行运算,每个检测器从不同的角度和尺度评估目标特征,各产生3个评分,这3个分数分别反映各检测

器认为该区域与目标动物相似的程度。将3个分数加权平均输出,并与分类分支生成的分类结果相乘,以此实现对非目标区域的抑制作用。最后通过卷积运算生成一个精细的分类置信度,根据分类置信度获得目标的所在位置。

1.4 损失函数

IoU损失为回归损失函数,回归损失定义为:

$$L_{\text{reg}} = 1 - \text{IoU}. \quad (6)$$

对于关系检测器的损失,采用MSE损失进行匹配分数回归,其形式如下

$$L_{\text{matching}} = (r_{i,j} - y_{i,j})^2. \quad (7)$$

式中: (i,j) 为特征图中的点, $y_{i,j}$ 为该点的分类结果, $r_{i,j}(\in (0,1))$ 为该点的回归关系分数。

本研究采用交叉熵损失(L_{cls})计算分类分支的损失函数,则模型总损失函数被定义为

$$\mathcal{L} = \lambda_1 * \mathcal{L}_{cls} + \lambda_2 * \mathcal{L}_{reg} + \lambda_3 * \mathcal{L}_{matching} \quad (8)$$

式中: 权重 λ_1 、 λ_2 和 λ_3 是用来平衡不同损失项重要性的超参数, 且满足 $\lambda_1 + \lambda_2 + \lambda_3 = 1$ 。

2 结果与分析

2.1 实验平台参数设置

所有实验均在配备 2080ti GPU 的个人电脑上进行。该程序在 PyTorch(v1.6.0) 框架下编写。此外, CUDA v10.1 被应用于训练模型, 加快了处理速度; Python 3.7 版本也被使用。其他使用的库是 NumPy (v1.21.2)、pycocotools、Cython、Matplotlib、tqdm、yacs 和 OpenCV。在训练中的初始学习率设置为 0.1, 学习率将每 30 个 epoch 衰减为原来的 1/10。训练采用随机梯度下降法更新各层的参数。

由于训练关系检测器需要同一序列中的目标样本, 而野生动物视频样本往往数量有限。本研究使用 Few-shot^[25] 来训练网络, 以便解决数据不平衡问题。此外, 关系检测器可以通过这个方法学习一种可以跨任务迁移的深度度量方法, 用以衡量区域与目标之间的关系, 帮助过滤掉背景中的干扰物。不同于其他方法可能依赖于已知的物体类别来检测目标, 本研究不需要类别信息, 而是专注于判断两个物体是否为同一个, 这对于跟踪任务尤为重要。通过这种方式, 本研究的模型可以在样本量极小的情况下, 有效地区分目标与背景, 从而提高跟踪的准确性。

2.2 模型性能评估指标

2.2.1 精确率 (precision)

精确率定义为模型输出的预测框中心位置与 ground truth 框中心位置之间的误差。

首先计算预测框中心点与 ground truth 框中心点的欧式距离得到中心位置像素差, 计算公式为

$$\rho = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (9)$$

式中: (x_2, y_2) 为目标预测框的中心位置, (x_1, y_1) 为 ground truth 框的中心位置。

精确率的计算公式为

$$P = \frac{F_p}{F} \quad (10)$$

式中: F_p 为像素误差低于给定阈值的帧数, F 为视频序列总帧数, 阈值设置为 1 ~ 50。

2.2.2 成功率 (success rate)

成功率定义为模型输出预测框的准确度, 即这个视频在指定阈值下跟踪模型的成功跟踪率。计算模

型输出的整个序列所有帧的预测框和 ground truth 真值框区域内像素的交并比 (IoU), 其计算公式为

$$IoU = \frac{|a \cap b|}{|a \cup b|} \quad (11)$$

式中: a 为预测目标框面积, b 为 ground truth 目标框面积。

如果某一帧的 IoU 高于阈值, 认为该帧跟踪成功。成功率计算公式为

$$S = \frac{F_s}{F} \quad (12)$$

式中: F_s 为跟踪成功的帧数, F 为视频序列总帧数, 阈值范围设置为 0 ~ 1。成功率越高, 表明预测框与 ground truth 真值框之间的重叠程度越高。

2.3 跟踪算法比较分析

分别将本研究与主流的 5 种基于孪生网络的目标跟踪方法 SiamFC、SiamRPN、SiamRPN++、SiamBAN 和 SiamRN 在数据集上进行跟踪实验, 采用上述评价指标对跟踪结果进行比较分析。

图 4(a) 和 (b) 分别展示了所有方法跟踪结果的成功率和精确率, 本研究方法的成功率和精确率最高, 分别达到 0.698 和 0.928, 尤其是本研究方法的精确率比最优方法 SiamRN 高出 0.031, 表明本研究预测输出的定位框与 ground truth 之间的重叠度更高, 定位误差更小。此外, 当考虑不同的定位阈值时, 本研究方法实现了最高的定位精度。在不同的重叠阈值下, 本研究方法输出的预测框与 ground truth 之间的重叠程度也最高。

2.4 参数量与计算量比较分析

在野生动物跟踪的实际应用中, 模型参数量和计算量的要求与精度同样重要。参数表示模型的复杂性, 模型越复杂, 其参数越大。GFLOPs 是考量一个网络结构计算量的标准, 计算量对应的是运算速度。从表 2 可以看出, 本研究使用的模型无论是参数量还是计算量, 都远远小于基准算法 SiamRN。

表 2 改进模型与基准模型参数量与计算量

Table 2 Improved model and benchmark model parameters and computational results

模型 Model	主干网络 Backbone	参数量 Params	计算量 GFLOPs
基准网络 Benchmark network	ResNet50	45.52×10^6	7.57
本研究 Ours	本研究	23.51×10^6	3.12

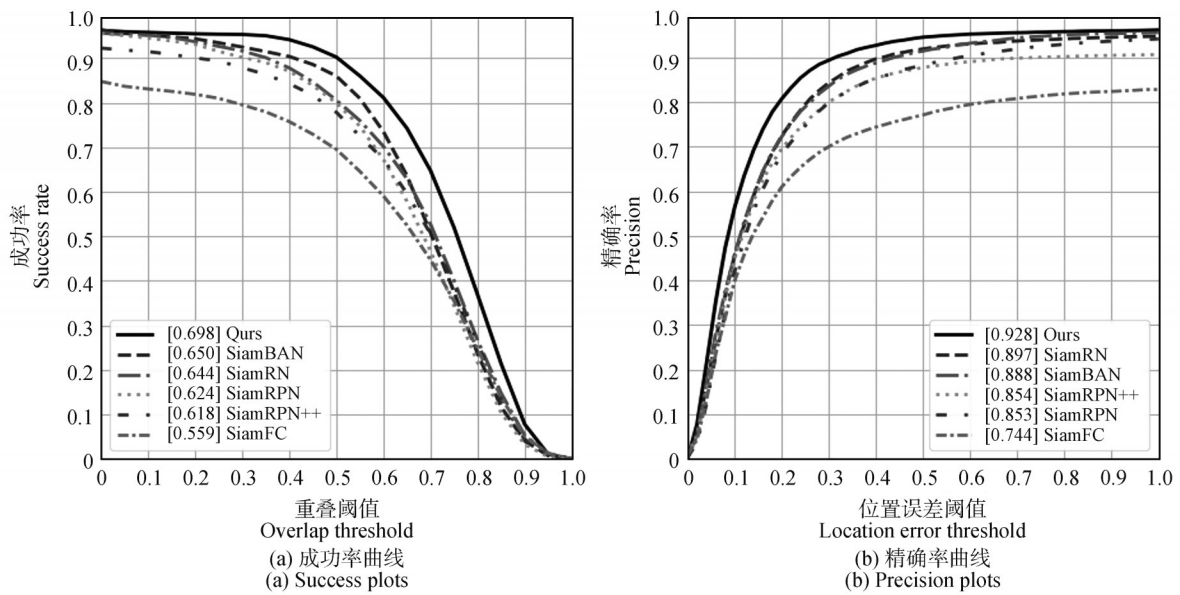


图4 跟踪算法的成功率(a)与精确率(b)

Figure 4 Success rate(a) and precision(b) of tracking algorithm

2.5 单目标跟踪实验结果分析

小目标在图像中占据的像素较少,这导致其特征信息不丰富,特别是在目标快速移动时,连续帧之间的变化可能导致目标特征的丢失或混淆,从而增加跟踪的难度。如图5(a)所示,本研究算法成功在第8帧、第38帧中跟踪目标,而SiamRN算法则出现漏检情况。

当目标出现尺度变化时,如果算法不能适应尺度变化,便会逐渐积累误差,导致跟踪结果逐渐偏离目标的真实位置。如图5(b)所示,在SiamRN中从第51帧开始出现误差,到第96帧时误差逐渐累积,最终在第198帧目标丢失,而本研究方法可以有效应对尺度变化,实现对动物目标的稳定跟踪。此外,在图5(b)中,目标与背景颜色相似,容易出现混淆,这也是在跟踪过程中SiamRN方法丢失目标的原因之一。

在处理密集或紧靠的动物群体时,由于动物个体的外观特征非常相似,容易导致跟踪算法出现误检。如图5(c)所示,从第246帧开始,SiamRN方法开始出现误检情况,持续至第314帧,此阶段的跟踪结果与初始设定的目标存在偏差。在第190帧中的2只企鹅相互靠近时,SiamRN方法还出现了目标混淆的情况,而本研究方法未发现此类误检。

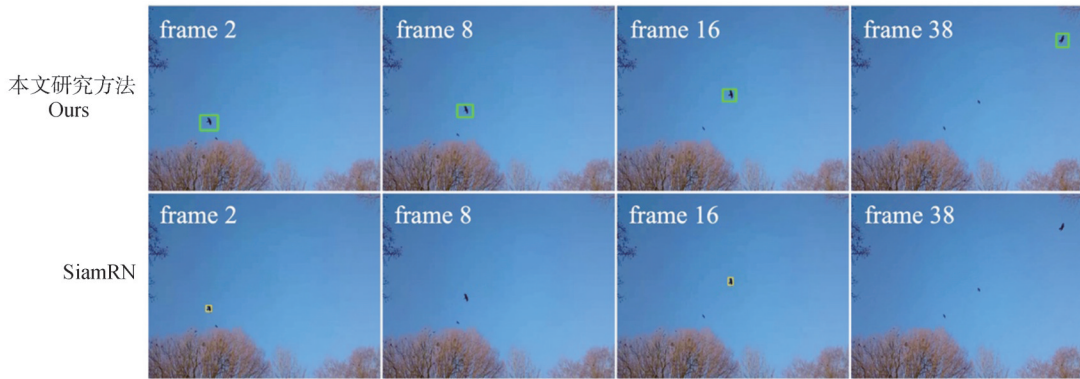
3 讨论

在监测野生动物时,研究人员通常使用无人

机^[26]、相机^[27]捕捉画面,通过分析这些图像来确定野生动物的行为活动状态。然而,由于野生动物在自然环境中的行为不可预测,相机拍摄的画面常常会出现目标快速移动或运动模糊的情况。此外,在野生动物实际跟踪过程中,会出现目标较小、尺度变化、相似性干扰、野生动物相互遮挡或被环境完全遮挡等问题,这些复杂情形都会导致算法难以正确估计目标的位置和运动^[28]。

本研究选取麋鹿、虎、狮、马、企鹅及其他鸟类为研究对象,构建了6类野生动物数据集,旨在探索深度学习网络实现野生动物实时跟踪的有效方法。在特征阶段,采用多头注意力机制^[21]能够准确捕捉目标对象的关键特征,有效增强对小目标的跟踪能力。滑动窗口自注意力运算可以在不同尺度上提取目标特征,提高对目标尺度变化的适应性。串联窗口自注意力运算和滑动窗口自注意力运算的结合使本研究算法能够在目标被遮挡或快速移动的情况下保持稳定的跟踪性能。此外,利用模型输出的跟踪结果还能生成动物的运动轨迹,也可应用于动物的姿态和行为识别,如图6所示的站立、行走和飞行等行为。与其他方法相比,本研究提出的算法在面对小目标快速移动、目标尺度变化以及密集动物群体等挑战性场景时,能够取得更为优异的跟踪效果。

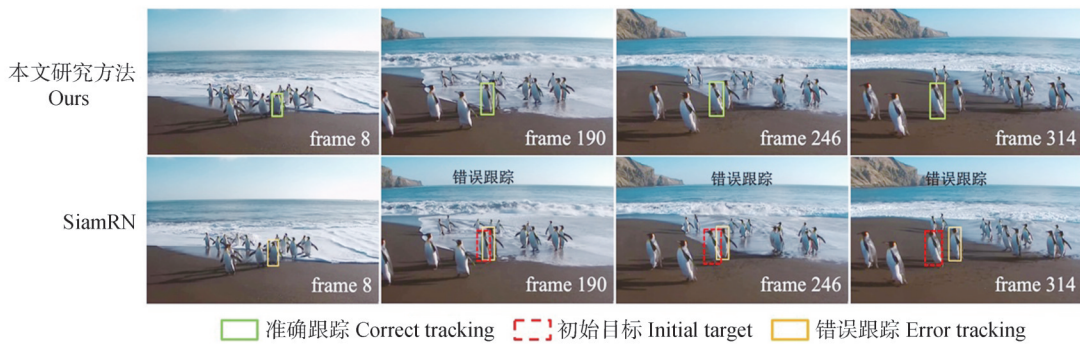
目前基于视频的野生动物单目标跟踪研究相对较少,尽管本研究取得了一些成果,但仍存在如下局限性:本研究集中在有限的物种上,在一定程度上限



(a) 快速移动的小目标跟踪结果对比
(a) Comparison of tracking results for fast moving small targets



(b) 尺度变化的目标跟踪结果对比
(b) Comparison of target tracking results with scale changes



(c) 外观特征相似的目标跟踪结果对比
(c) Comparison of target tracking results with similar appearance features

图5 跟踪结果对比

Figure 5 Comparison of target tracking results

制了研究的广度和算法的适用性。未来的研究将扩充数据集,纳入更多种类的野生动物,以增加物种多样性,更加全面地评估算法的通用性和鲁棒性。此外,本研究的网络设计主要针对单个动物目标的跟踪,当需要同时跟踪多个动物目标时,可能会遇到目标间的相互作用和潜在冲突问题。因此,未来的研究还需要探索新的策略,以有效管理和解决多目标跟踪中的复杂情况。

4 结论

野生动物跟踪对于野生动物保护至关重要。本研究设计了一种基于改进的孪生网络框架的跟踪方法,以便解决跟踪野生动物时常见的目标较小、尺度变化和模型参数大等问题。首先,在跟踪框架的主干网络 ResNet50 中引入了 Swin-Transformer,以减少模型参数量并保持强大的特征提取能力。其次,通

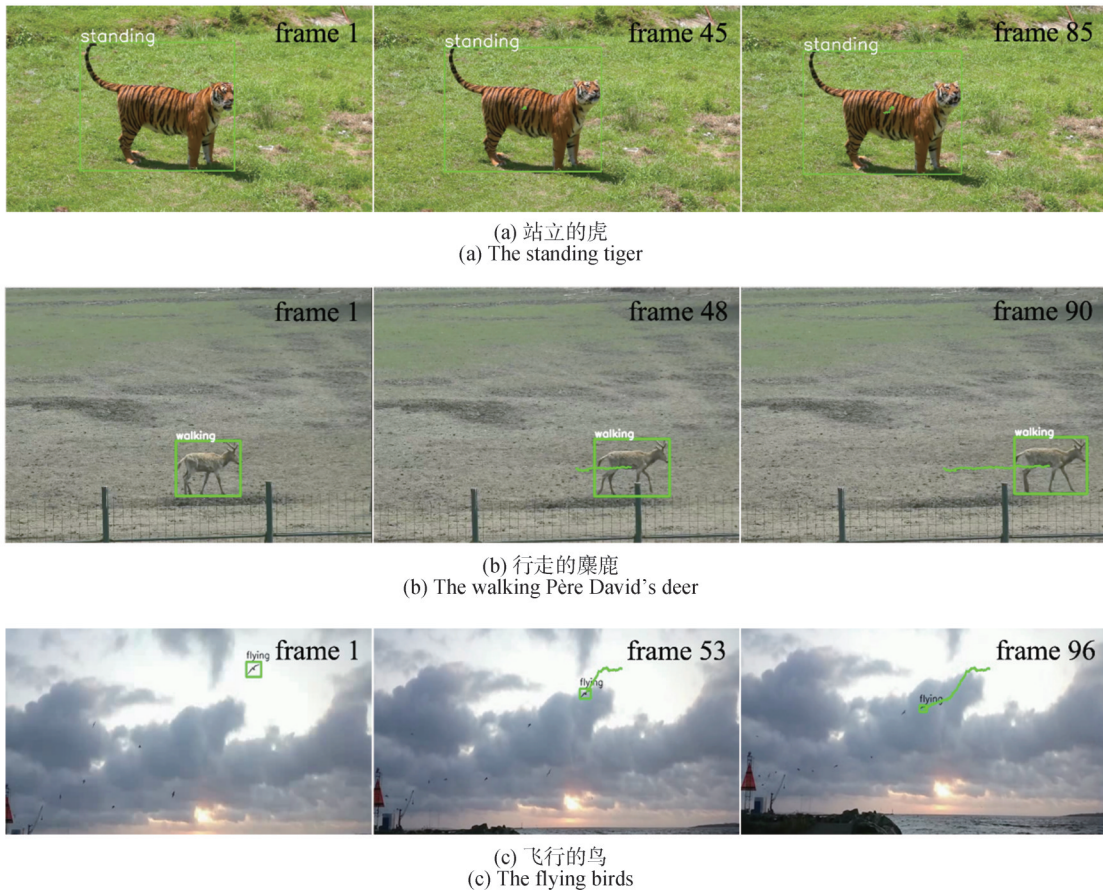


图6 跟踪轨迹及行为识别

Figure 6 Tracking trajectory and behavior recognition

过引入串联窗口自注意力运算和滑动窗口自注意力运算,在特征提取阶段增强模型对图像中不同范围信息的获取能力,通过窗口级别的自注意力计算提高模型的适应性和灵活性。此外,本研究构建了一个野生动物跟踪数据集并进行测试实验。结果显示,本研究提出的跟踪模型在性能上优于5种主流的孪生网络跟踪算法,提高了跟踪的准确性和鲁棒性,更适合实际应用场景,为野生动物监测和保护提供了一种可行且有效的解决方案,有助于推动野生动物保护工作的科学化和智能化发展。

参考文献:

- [1] 李晓艳. 野生动物监测现状与技术方法的应用研究[J]. 农业灾害研究, 2023, 13(10): 16-18.
LI X Y. Research on the current situation and application of technical methods of wildlife monitoring[J]. Journal of Agricultural Catastrophology, 2023, 13(10): 16-18.
- [2] 张媛媛, 穆浩文, 孙哲, 等. 智能科技助力野生动物保护[J]. 科学, 2024, 76(3): 42-46.
ZHANG Y Y, MU H W, SUN Z, *et al.* Intelligent technology pro-

otes wildlife conservation[J]. Science, 2024, 76(3): 42-46.

- [3] SIMANUNGKALIT G, BARWICK J, COWLEY F, *et al.* Use of an ear-tag accelerometer and a radio-frequency identification (RFID) system for monitoring the licking behaviour in grazing cattle [J]. Applied Animal Behaviour Science, 2021, 244: 105491.
- [4] ARAVIND D, ANUPRIYA S, AARTHI A T, *et al.* An automatic wildlife tracking system using GPS and wireless sensor networks [R/OL]. Pune: National Conference on Electronics, Communication and Computing (NCECC-2017), 2017 [2024-10-28]. <https://www.researchgate.net/publication/333878429>.
- [5] ULLMANN W, FISCHER C, KRAMER-SCHADT S, *et al.* The secret life of wild animals revealed by accelerometer data: how landscape diversity and seasonality influence the behavioural types of European hares[J]. Landscape Ecology, 2023, 38(12): 3081-3095.
- [6] 李秀明, 李晓民, 宋玉波, 等. 基于卫星跟踪技术的三江平原大白鹭幼鸟活动区分析[J]. 野生动物学报, 2015, 36(4): 395-401.
LI X M, LI X M, SONG Y B, *et al.* Analysis of movements of a juvenal great egret on the Sanjiang Plain by satellite tracking[J]. Chinese Journal of Wildlife, 2015, 36(4): 395-401.

- [7] 郝焱, 陈江文, 晏迪, 等. 长岭龙凤湖湿地野生鸟类救治放飞与卫星跟踪[J]. 野生动物学报, 2019, 40(4): 964-968.
HAO Y, CHEN J W, YAN D, *et al.* Wild bird rescue-release and satellite tracking in Changling Longfeng Lake wetland [J]. Chinese Journal of Wildlife, 2019, 40(4): 964-968.
- [8] KAYS R, CROFOOT M C, JETZ W, *et al.* Terrestrial animal tracking as an eye on life and planet [J]. Science, 2015, 348(6240): aaa2478.
- [9] 张雪莹, 张浩林, 韩莹莹, 等. 基于深度学习的野生动物监测与识别研究进展[J]. 野生动物学报, 2022, 43(1): 251-258.
ZHANG X Y, ZHANG H L, HAN Y Y, *et al.* Research progress of the wildlife monitoring and identification based on deep learning [J]. Chinese Journal of Wildlife, 2022, 43(1): 251-258.
- [10] 杨帆. 基于人工智能下的野生动物识别研究与应用[J]. 中国高科技, 2024(16): 72-74.
YANG F. Research and application of wild animal recognition under artificial intelligence [J]. China High and New Technology, 2024(16): 72-74.
- [11] KIM J I, BAEK J W, KIM C B. Image classification of Amazon parrots by deep learning: a potentially useful tool for wildlife conservation[J]. Biology, 2022, 11(9): 1303.
- [12] 黄志静, 邵慕义, 张庭瑞, 等. 基于深度学习的野生动物识别[J]. 电子测试, 2022(22): 69-71; 10.
HUANG Z J, SHAO M Y, ZHANG T R, *et al.* Wild life recognition based on deep learning[J]. Electronic Test, 2022(22): 69-71; 10.
- [13] MANOHAR N, SHARATH KUMAR Y H, KUMAR G H. An approach for the development of animal tracking system[J]. International Journal of Computer Vision and Image Processing, 2018, 8(1): 15-31.
- [14] 阎婧宇, 谢永华. 基于YOLOv7-Tiny算法的无人机实时跟踪野生动物方法[J]. 野生动物学报, 2024, 45(2): 251-261.
YAN J Y, XIE Y H. Real time tracking of wild animals using unmanned aerial vehicles based on YOLOv7-Tiny algorithm [J]. Chinese Journal of Wildlife, 2024, 45(2): 251-261.
- [15] 马光凯, 吴伟, 刘丹, 等. 圈养东北虎幼崽群体活动的自动跟踪方法研究[J]. 野生动物学报, 2022, 43(3): 605-613.
MA G K, WU W, LIU D, *et al.* Automatic tracking method of group activities of captive *Panthera tigris altaica* cubs [J]. Chinese Journal of Wildlife, 2022, 43(3): 605-613.
- [16] 许群, 谢永华. 基于注意力机制融合多特征的东北虎个体自动跟踪方法[J]. 生物多样性, 2024, 32(3): 141-154.
XU Q, XIE Y H. Automatic individual tracking method of Amur tiger based on attention mechanism fusion of multiple features [J]. Biodiversity Science, 2024, 32(3): 141-154.
- [17] BERTINETTO L, VALMADRE J, HENRIQUES J F, *et al.* Fully-convolutional Siamese networks for object tracking [C]//HUA G, JÉGOU H. Computer vision: ECCV 2016 workshops. Cham: Springer International Publishing, 2016: 850-865.
- [18] LI B, YAN J J, WU W, *et al.* High performance visual tracking with Siamese region proposal network [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018. Salt Lake City: IEEE, 2018: 8971-8980.
- [19] CHEN Z D, ZHONG B N, LI G R, *et al.* Siamese box adaptive network for visual tracking [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020. Seattle: IEEE, 2020: 6668-6677.
- [20] CHENG S Y, ZHONG B N, LI G R, *et al.* Learning to filter: Siamese relation network for robust tracking [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20-25, 2021. Nashville: IEEE, 2021: 4421-4431.
- [21] LIU Z, LIN Y T, CAO Y, *et al.* Swin Transformer: hierarchical vision transformer using shifted windows [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021. Montreal: IEEE, 2021: 10012-10022.
- [22] SVANSTRÖM F, ALONSO-FERNANDEZ F, ENGLUND C. A dataset for multi-sensor drone detection [J]. Data in Brief, 2021, 39: 107521.
- [23] 王景鑫, 潘欣. 一种基于Labeling的辅助标注方法[J]. 科技创新与应用, 2023, 13(29): 145-148.
WANG J X, PAN X. An auxiliary labeling method based on labeling [J]. Technology Innovation and Application, 2023, 13(29): 145-148.
- [24] CHICCO D. Siamese neural networks: an overview [M]//CARTWRIGHT H. Methods in molecular biology: Vol. 2190: artificial neural networks. 3rd ed. New York: Humana Press, 2020: 73-94.
- [25] THRUN S. Is learning the n -th thing any easier than learning the first? [C]//TOURETZKY D S, MOZER M C, HASSELMO M E. Proceedings of the 9th international conference on neural information processing systems. Cambridge: MIT Press, 1996: 640-646.
- [26] 胡健波, 张健. 无人机遥感在生态学中的应用进展[J]. 生态学报, 2018, 38(1): 20-30.
HU J B, ZHANG J. Unmanned aerial vehicle remote sensing in ecology: advances and prospects [J]. Acta Ecologica Sinica, 2018, 38(1): 20-30.
- [27] 杨铭伦, 张旭, 郭颖, 等. 基于YOLOv5的红外相机野生动物图像识别 [J]. 激光与光电子学进展, 2022, 59(12): 1215015.
YANG M L, ZHANG X, GUO Y, *et al.* Recognition of wild animals using infrared camera images based on YOLOv5 [J]. Laser & Optoelectronics Progress, 2022, 59(12): 1215015.
- [28] FLEMING C H, DRESCHER-LEHMAN J, NOONAN M J, *et al.* A comprehensive framework for handling location error in animal tracking data [EB/OL]. BioRxiv (2021-07-24) [2024-10-08]. <https://doi.org/10.1101/2020.06.12.130195>.