基于深度学习的无人机单目标跟踪

谢志丰,周　诺,梁　军*
(华南师范大学 软件学院,广东 佛山 528225)

摘　要:无人机单目标跟踪,是指对无人机运动过程中拍摄的视频进行实时处理,进而准确、稳定地跟踪一个移动目标。无人机单目标跟踪受环境影响较大,存在光照变化、背景干扰、目标遮挡、相似目标干扰等问题,使得追踪准确性尚有待提高。针对上述问题,以 SiamRPN++为基础,对其模型和损失函数进行创新性优化。主要研究贡献:在网络骨架(Backbone)方面,通过引入注意力机制网络结构 SENet,与原有模型的 ResNet50 组成 Se_ResNet50,提升对单目标跟踪的准确性和有效性;在损失函数方面,使用 Balanced L1 Loss 提升关键的回归梯度,在分类、整体定位以及精确定位中实现更加平衡的训练;在 SiamRPN++的结构基础上,对 Backbone 和 Loss 函数进行优化。实验使用 ILSVRC2013 和 ILSVRC2014 的 DET 数据集进行训练,以 VOT2018 和 OTB100 为测试数据集检验训练精度。最终追踪准确性在原基础上得到了一定的提高。

关键词:无人机;深度学习;目标跟踪;注意力机制;平衡 L1 损失;SENet
中图分类号:TP301　　**文献标识码**:A　　**文章编号**:1673-629X(2024)01-0185-08
doi:10.3969/j.issn.1673-629X.2024.01.027

Single Target Tracking for UAV Based on Deep Learning

XIE Zhi-feng,ZHOU Nuo,LIANG Jun*
(School of Software,South China Normal University,Foshan 528225,China)

Abstract:Unmanned aerial vehicle (UAV) single target tracking refers to real-time processing of videos captured during UAV movement,accurately and stably tracking a moving target. UAV single target tracking is greatly affected by the environment,with issues such as changes in lighting,background interference,target occlusion,and interference from similar targets,resulting in the need for further improvement in tracking accuracy. We focus on these issues and creatively optimize the SiamRPN++ model and loss function. The main research contributions are as follows:in terms of network backbone,we introduce the attention mechanism network structure SENet and combine it with the original ResNet50 model to form Se_ResNet50,improving the accuracy and effectiveness of singles target tracking. In terms of loss function,we use Balanced L1 Loss to enhance the key regression gradients and achieve a more balanced training in classification,overall localization,and precise localization. Based on the structure of SiamRPN++,we optimize the Backbone and Loss functions. Experiments were conducted using the ILSVRC2013 and ILSVRC2014 DET datasets for training and VOT2018 and OTB100 as test datasets to verify training accuracy. Ultimately,tracking accuracy was improved to a certain extent compared to the original.

Key words:unmanned aerial vehicle (UAV);deep learning;object tracking;attention mechanism;balanced L1 loss;SENet

0　引　言

近年来随着科技的快速发展,无人机逐步应用在各个领域,且发挥了巨大的作用。如,无人机可应用在军事领域内的通讯、侦察任务,执行自然灾害或事故发生地的搜寻、救助任务,实现物资派送、地质勘查、电力巡线等,无人机还具有成本低、安全性高和机动性强的特点。随着无人机技术的逐步推进,其研究和应用前景将得到进一步提升。

目标跟踪技术属于无人机应用的关键技术之一,同时也是无人机执行任务时信息获取源的主要技术。随着人工智能、深度学习领域的不断推广,借助计算机视觉或者深度学习技术可以让无人机更加智能,不再依赖地面端的信息处理,有效减少了信息传递中受到的环境干扰,极大提高了系统实时性和在复杂环境中的稳定性。

尽管基于深度学习的单目标跟踪算法在大多数检测和跟踪场景中表现出优秀的性能,但在困难环境中,例如存在小尺寸或相似物体、目标形变和目标遮挡等

收稿日期:2023-06-29　　**修回日期**:2023-10-30
基金项目:广东省基础与应用基础研究基金(2022A1515140110);广东省基础与应用基础研究基金(重点项目)(2020B1515120089)
作者简介:谢志丰(2001-),男,研究方向为深度学习;通信作者:梁　军(1982-),男,博士,CCF 会员(E200012341M),研究方向为图论、人工智能。

因素,模型的检测和跟踪性能往往显著下降。解决此类问题最简单的办法是提高精确度,但会带来计算资源消耗大、算法复杂度高等问题。因此,在保持一定精确度的前提下,控制算法复杂度或采用轻量级模型变得至关重要。

当前研究中,针对困难环境下的单目标跟踪,许多工作致力于在轻量级模型和可控算法复杂度的基础上提高性能。这些方法旨在平衡模型的准确性和计算效率,以应对目标的变化和遮挡等复杂情况。通过优化网络架构、引入注意力机制、设计精细的损失函数等手段,这些方法在保持较低计算资源需求的同时,提升了目标检测和跟踪的准确率。

在未来的研究中,仍需要进一步探索和发展更多适应困难环境的单目标跟踪算法。这些算法应具备高精确度、可控算法复杂度和轻量级模型的特点,以应对小型或相似物体、目标形变和目标遮挡等挑战。通过更深入的研究和创新去实现更加鲁棒和高效的单目标跟踪算法。

下面是该文的主要贡献:

(1)为了提高单目标跟踪的准确性和有效性,采用了自注意力机制网络结构,并将其与 ResNet50[1] 模型相结合,设计了 Se_ResNet50 的网络骨架。这个改进的网络骨架能够更好地关注重要的特征,并且在保留原有特征信息的基础上,提升对目标的识别和跟踪能力。

(2)在损失函数方面,使用了 Balanced L1 Loss。这种损失函数能够平衡分类、整体定位以及精确定位之间的训练权重,使得回归梯度变得更加平滑和稳定,从而提升了跟踪的精度。它能够让训练过程更加平衡,并在不同的训练阶段中得到更优秀的结果,从而提升单目标跟踪的表现。

1 背景知识

单目标跟踪的主要方法分为两种,一种是基于相关滤波的跟踪算法。该算法通过当前已知目标训练出一个滤波器。滤波器和当前目标做相关运算后,可以得到高斯响应图。以此为模板寻找下一帧中响应最高的点,该点即为预测的目标位置。它是基于回归判别模型的典型方法,能够快速运转计算是因为其利用循环矩阵,使用快速傅里叶变换实现时域到频域的转换[2]。目前经典的基于滤波算法有:CSK[3],KCF[4],DCF[5],SRDCF[6] 等。

深度学习跟踪算法是一种基于深度学习技术的目标跟踪方法,具有较强的学习和表征能力。与传统的跟踪算法相比,深度学习跟踪算法能够自动学习目标的特征表征,并在跟踪过程中不断更新和优化这些表征,从而大大提高了跟踪的准确性和鲁棒性。同时,由于深度学习模型可以学习到更加抽象和高层次的特征表示,因此该算法对目标的光照、尺度变化、遮挡等复杂情况具有较强的适应能力,能够实现在复杂环境下的实时跟踪[7]。神经网络也分为主流的三大类,分别是卷积神经网络(CNN)、循环神经网络(RNN)和生成式对抗网络(GAN)。

当前,目标追踪任务中往往基于卷积神经网络进行设计,通过其强大的特征提取能力,得到的不错的追踪效果。如,2012 由 Krizhevsky 等人提出的 AlexNet[8] 和 2016 年 He 等人提出的 ResNet 在目前的主流追踪器 SiamRPN[9] 中具有良好效果。

SiamRPN 是一种深度学习目标跟踪方法,通过将目标跟踪问题转化为目标匹配问题,并结合分类和回归技术来实现目标跟踪。在该方法中,使用卷积神经网络提取目标的特征,同时采用区域生成网络(RPN)辅助目标的定位。通过将分类和回归相结合,SiamRPN 能够实现更加准确和鲁棒的目标跟踪。值得注意的是,SiamRPN 能够自适应地学习目标的特征表示,从而具有更强的适应性和鲁棒性,能够在各种场景下实现高效的目标跟踪[2]。

SiamRPN 系列除了两者外,还有各种改进系列,如 SiamRPN + +[10], DaSiamRPN[11], SiamMask[12], Deeper and Wider SiamRPN[13] 等。SiamRPN 系列展现出来强大的生命力,不仅在改进上各个版本有不同的改进思路,并且在 VOT 和 OTB 等测试评估中取得了显著成绩。

在近几年来,基于深度学习的单目标跟踪算法愈加精进,SiamRPN++,STARK[14] 和 MixFormer[15] 在相关挑战上都取得了很好的成绩。尤其在 OTB-100 数据集上,AiATrack[16] 和 DiMP-NCE+[17] 获得了最优结果。在 VOT2018 数据集上 TREG[18] 和 Ocean[19] 表现十分突出。然而其他系列模型较为庞大,在轻量级无人机上无法部署。因此,SiamRPN 系列更为适合应用在无人机上,但其追踪准确性还有提高的空间。该文将基于 SiamRPN++进行改进,通过结合自注意力机制和对 Loss 函数进行改进,有效提高追踪的精度。

各时间节点的代表性目标跟踪算法如图 1 所示。

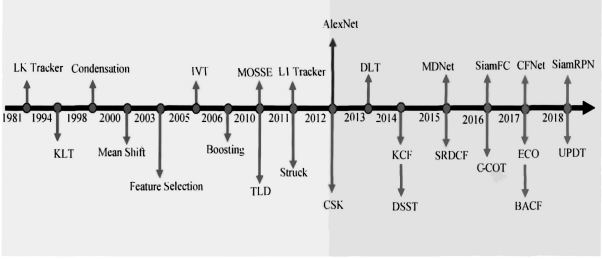

图 1 各时间节点的代表性目标跟踪算法

2　无人机单目标跟踪模型 SB_SiamRPN++

该文主要研究的是基于 SiamRPN++模型的 Backbone 和 Loss 改进,先对 SiamRPN++基础理论内容和整体网络进行阐述。SiamRPN++是在 SiamRPN 的基础上进行改进,作者引入了 Faster RCNN 中的 RPN,使用分支相关特征图提取特征,用于预测目标位置和置信度。SiamRPN++打破了严格平移不变性限制和目标相似限制,可以使用 ResNet 深层次网络进行训练。SiamRPN++使用多层特征融合的方式,合理使用浅层次中获取的图像特征与深层次中获取的语义信息。SiamRPN++还引进了 Depthwise Cross Correlation 模块,在减少计算量的同时使得分支更加平衡。SiamRPN++在 SiamRPN 上的优化使其感受野更广,卷积层更多,细粒度更高。

SiamRPN++的整体网络结构图如图 2 所示。

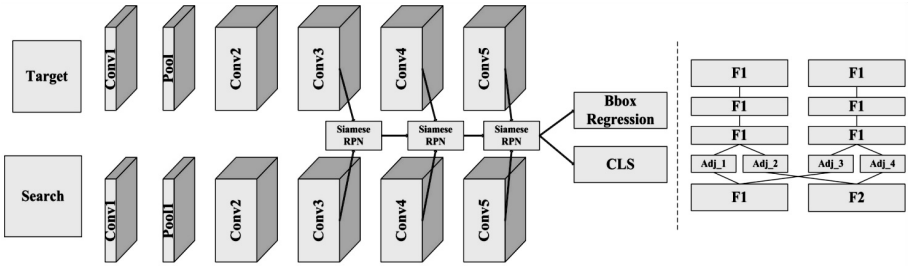

图 2　SiamRPN++结构图(左)与 Siamese RPN 模块图(右)

在 SiamRPN++的基础上,该文主要从注意力机制和损失函数出发。注意力机制有 SE,CBAM,ECA 和 CA 等,CBAM 和 CA 计算复杂度高,ECA 通道间的相关性建模能力相对较弱,综合后决定选取 SE 注意力机制模块,其简单高效,引入参数量少,能够显著提升性能。Smooth L1 Loss 在样本均衡方面存在一定问题,在小目标、遮挡目标等具有优化空间, Balanced L1 Loss 能在一定程度上缓解此问题。

综上所述,该文在原有的 SiamRPN++基础上对 Backbone 和 Loss 方面进行优化,在 Backbone 方面,将原有模型的 ResNet50 与 SE Block 融合成 SE_ResNet50,Loss 方面将原有模型中的 L1 Loss 更改为 Balanced L1 Loss。整体网络结构与 SiamRPN++基本一致,在每次计算中加入了 SE Block 模块。经过优化后的网络结构图取名为 SB_SiamRPN++,如图 3 所示。

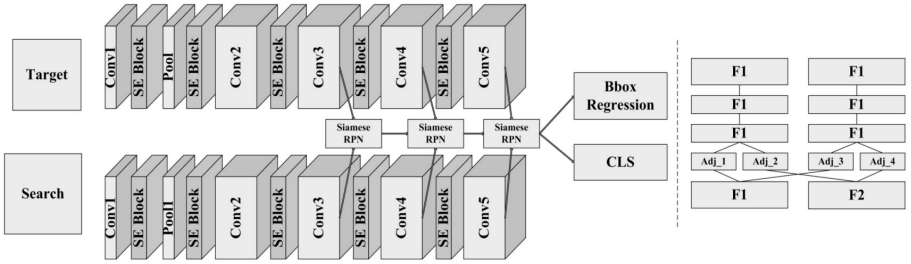

图 3　SB_SiamRPN++结构图(左)与 Siamese RPN 模块图(右)

在 Loss 方面,官方 SiamRPN++模型 pysot 中使用的是 L1 Loss。该文采用 Balanced L1 Loss 对其进行优化,引入 α,β,γ 参数,调整 loss 计算策略,从压缩维度求和转为 Balanced L1 Loss 策略,loss_weight 均衡策略不变。

在实验中,观察学习梯度逐步调整、优化参数,最终设置参数为:$\alpha = 0.7,\beta = 1.0,\gamma = 1.5$,具体实验过程在实验分析部分会指出。

Balanced L1 Loss 在 VOT2018 数据集测试中,EAO 有显著提升。其在迭代过程中,迭代效率大大提高。在使用 L1 Loss 的 pysot 模型训练中,需要 10 ~ 20 次迭代得到最好结果,修改为 Balanced L1 Loss 函数后在 1 ~ 10 个迭代中就能得到最好结果,说明在梯度收敛中,Balanced L1 Loss 的收敛速度比 L1 Loss 更快。

2.1　SENet 模块

SENet[20]称 Squeeze-and-Excitation Networks,是由 Momenta 胡杰团队提出的新网络结构,并夺下 ImageNet 2017 Image Classification 任务的桂冠。SENet 在 ImageNet[21]数据集中的 Top-5 error 降低至 2.251%,性能提高了 0.74%[22]。

SENet 采用一种全新的特征重标定策略,通过深度学习自动获取每个特征通道的重要程度,以此为依据提升重要特征并抑制无用特征[5]。和以往特征通道融合方法不同,SENet 并没有引入新的空间维度,而是通过新策略使特征通道之间的相互依赖关系直接呈现出来。SENet 的核心思想从网络的 loss 入手,根据 loss 学习特征权重,使有效的特征图权重增大,无效或者效果小的特征图权重减小,以此为训练模型达到更好的

效果。Sequeeze-and-Excitation block 是一个子结构，可以嵌入到其他分类和检测模型。

图 4 为 SENet 结构图。SE Block 属于自注意力机制函数，自注意力是一种特殊注意力机制，可以通过对所有位置的特征向量取加权平均得到，能更好地提高并行运算效率，提高了模型的可解释性。在传统的卷积神经网络中，对局部特征编码会导致在捕捉长距离方面无法进行建模，自注意力机制能计算不同位置间的相对关系，从而更好捕捉全局信息，所以在捕捉长距离的依赖关系中更具优势。

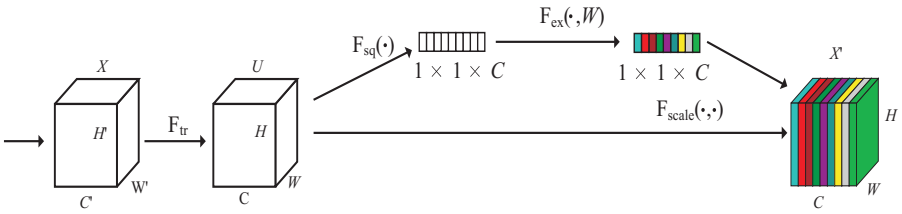

图 4 SENet 模块

SE Block 作为一个子网络结构，其结构非常简单，容易部署，不需要引入新的函数或者层，可以嵌入到任何复杂模型中开发新型 SENet，广泛用于计算机视觉等领域。但是它的灵活性也有一定的限制，无法直接应用于标准卷积转换。它减低了计算复杂度，在 SE_ResNet50 得到精度甚至与 SE_ResNet101 媲美，在训练过程中，加入了 SENet 模块一定程度上降低了收敛错误率。但在一定的网络深度下进行 BP 优化时，靠近输入层的部分网络可能出现梯度消散问题，导致模型难以优化。因此，除了结合该方法外，该文进一步对 Loss 进行修改，一定程度上解决了该问题。

2.2 Balanced L1 损失函数

目标检测的损失函数分为分类损失和回归损失，是一个多任务损失函数。Balanced L1 Loss 损失函数是在 Fast R-CNN 中使用的损失函数。通常，在回归损失前引入参数 λ 进行调整，当分类效果较好时，损失值较为精确，但是会忽略回归重要性。为了均衡不同任务，需要对参数进行调整。在参数调整中要注意的是，回归时没有边界限制的，直接增加回归损失权重是不可行的，容易产生巨大梯度不利于训练[23]。

$$L_{p,u,t_u,v} = L_{\text{cls}}(p,u) + \lambda[u \geqslant 1]L_{\text{loc}}(t^u,v) \tag{1}$$

基于上述问题，作者提出了 Balanced L1 Loss，Balanced L1 Loss 是在 Smooth L1 Loss 基础衍生出来的。在 Smooth L1 Loss 中，设置拐点区分 inliers 和 outliers，并对 outliers 设置 1.0 进行梯度截断。

Balanced L1 Loss 主要目的是要显著提升 inliers 的梯度，使准确的训练点在训练中能发挥更重要的作用。作者通过参数 λ 来调整回归损失上界，通过调整参数 α，γ 可以得到更加平衡的训练，作者并没有使用超参数 β，该参数用于控制难易样本的权重。总的来说，Balanced L1 Loss 的核心思想是提升关键的回归梯度，平衡样本及任务，以便能在分类、识别、定位和回归中实现更加平衡的训练。

Balanced L1 Loss 的公式如下[23]：

$$L_{\text{loc}} = \sum_{i \in x,y,w,h} L_b(t_i^u - v_i) \tag{2}$$

$$\frac{\partial L}{\partial x} = \begin{cases} \alpha \ln(b|x|+1), & \text{if } |x| < 1 \\ \gamma, & \text{otherwise} \end{cases} \tag{3}$$

$$L_b(x) = \begin{cases} \ln(b|x|+1)\dfrac{\alpha}{b|x|+1} - \alpha|x|, & \text{if } |x| < 1 \\ \gamma|x| + C, & \text{otherwise} \end{cases} \tag{4}$$

其中，α，γ 满足以下公式[23]：

$$\alpha \ln(b+1) = \gamma \tag{5}$$

3 实验分析

3.1 相关环境与数据集说明

为模拟无人机中较为低端的计算环境，该文在 Windows 下运行，使用 CPU，在原有官方 pysot 模型上修改相应配置，在对应代码将 WORLD_SIZE 设置为 1，MASTER_ADDR 设置为 localhost，MASTER_PORT 可以设置成任意未被占用的端口，如 6789 等。在 torch.device 函数中出现无法使用 GPU 的情况，则将 GPU 参数改为 CPU 参数。将代码中的 RANK 设置为 0。该文使用的是 python-3.10.7，pytorch 为 1.11.0+cu113。

在数据集上，训练采用的是 ILSVRC2013 和 ILSVRC2014 DET dataset，而测试实际追踪效果则是在 VOT2018[24] 和 OTB100[25] 数据集上。ILSVRC 全称为 IMAGENET Large Scale Visual Recognition Challenge（示例见图 5）。自 2010 始，每一年都会举办 ILSVRC 图像分类和目标检测大赛。ImageNet 数据集在目前深度学习图像领域应用中占据主流，关于图像识别、分类、定位、检测等都基于此来开展。ImageNet 数据集有 1 400 多万幅图片，涵盖多达 2 万多个类别，并在许多图片中有明确的类别标注和物体标注[26]。VOT2018 包含了 60 个高质量的真实场景视频序列，涵盖了各种目标类别和运动模式。这些视频序列在不同的环境中捕捉到了目标物体的运动，如户外、室内和

复杂背景等。OTB100 数据集包含了 100 个具有挑战性的视频序列。这些视频序列涵盖了不同的目标类别(如人、车、动物等)、运动模式和背景条件。这两个数据集具有很强的挑战性,十分适合无人机应用。

图 5　ILSVRC 数据集示例

3.2　部分实验指标说明

在目标追踪中,常用的评估标准有 Robustness, Accuracy 和 Expected Average Overlap(EAO),它们用于衡量追踪算法在准确性和鲁棒性方面的表现。

Robustness(鲁棒性):鲁棒性用于评估目标追踪算法在面对各种挑战性条件时的稳定性和可靠性。一种常用的鲁棒性指标是在多个测试视频序列上计算系统失效的频率。对于每个测试序列,当算法丢失跟踪目标且不再恢复时,认为系统失效。通过计算失效频率来得到鲁棒性指标,失效频率越低表示算法越鲁棒。

Accuracy(准确率):是用来评估目标位置估计精确度的指标。它通过计算预测框与目标真值框之间的重叠程度(IoU)来确定预测框的准确性。如果预测框与目标真值框的 IoU 大于等于设定的阈值(通常为 0.5),则该预测框被视为准确的。最后,准确率被计算为准确的预测框数除以总的预测框数。准确率越高,表示算法在目标定位方面表现越精确。

Expected Average Overlap(EAO):EAO 用于综合衡量目标追踪算法在准确性和鲁棒性之间的平衡。它基于预测框与目标真值框之间的重叠程度,通过衡量平均重叠面积(Average Overlap,AO)和失效率(failure rate)来计算。

具体来说,首先计算每个测试序列上的平均重叠面积,然后通过对平均重叠面积进行加权平均,权重为各个序列的失效率。最后得到的加权平均值即为 EAO,值越高表示算法在准确性和鲁棒性方面表现越好。

3.3　Balanced L1 损失函数参数设置

对 Balanced L1 Loss 中的 $\alpha = 0.5, \gamma = 1.5$ 进行调整。随机选取部分数据集,并以此作为标准训练数据集进行训练,对 α, γ 参数进行微调,迭代次数为 10 ~ 20,β 均为 1.0,具体微调参数后 VOT 最优评估结果如表 1 所示,OTB 最优评估结果如表 2 所示。

表 1　Balanced L1 Loss 调整 α, γ 的 VOT 最优评估结果

Backbone & Loss	Accuracy	Robustness	Lost Number	EAO
① $\alpha = 0.5, \gamma = 1.5$	0.326	1.241	265.0	0.089
② $\alpha = 2.0, \gamma = 1.5$	0.339	1.185	253.0	0.099
③ $\alpha = 1.0, \gamma = 1.0$	0.320	1.133	242.0	0.094
④ $\alpha = 0.7, \gamma = 1.5$	0.309	0.983	210.0	0.106

表 2　Balanced L1 Loss 调整 α, γ 的 OTB 最优评估结果

Backbone & Loss	Success	Precision
① $\alpha = 0.5, \gamma = 1.5$	0.238	0.348
② $\alpha = 2.0, \gamma = 1.5$	0.263	0.385
③ $\alpha = 1.0, \gamma = 1.0$	0.261	0.409
④ $\alpha = 0.7, \gamma = 1.5$	0.255	0.368

在表 1 的 VOT 最优评估表中,Accuracy 最高分数的参数为② $\alpha = 2.0, \gamma = 1.5$,这个参数函数与 Smooth L1 Loss 类似,在梯度和 Loss 方面都十分接近;但是在 EAO 指标中,④ $\alpha = 0.7, \gamma = 1.5$ 的效果最好,该参数梯度更偏向平缓,Loss 数值比 Smooth L1 Loss 收敛得更快,EAO 更高表示训练更有效。在 Robustness 指标中,对比原参数① $\alpha = 0.5, \gamma = 1.5$,② $\alpha = 2.0, \gamma = 1.5$ 和③ $\alpha = 1.0, \gamma = 1.0$ 逐层降低,④ $\alpha = 0.7, \gamma = 1.5$ 下降幅度最高,下降了 20.79%,说明了在平缓梯度后提高了有效样本在训练过程中的重要性,减少了误差。在 Lost Number 指标中,其数值按顺序逐步递减,在④ $\alpha = 0.7, \gamma = 1.5$ 中递减了 20.75%。

在表 2 OTB 最优评估表中,② $\alpha = 2.0, \gamma = 1.5$ 在 Success 取 0.263 的高分;③ $\alpha = 1.0, \gamma = 1.0$ 取得 0.261 分数,与②几乎没有差距,说明提高梯度只在一定程度上有提升,依然具有一定的限制性;④ $\alpha = 0.7, \gamma = 1.5$ 取得 0.255,比②低 3.04%,可能在训练过程中梯度放缓,精细化程度更高,需要一定的迭代次数。在 Precision 指标中,③ $\alpha = 1.0, \gamma = 1.0$ 获得分数最高,说明在该梯度与 Loss 数值下训练模型能更好地提高模型的正确率和预测率。

综合表 1 和表 2 的数据指标,可以选取③ $\alpha = 1.0, \gamma = 1.0$ 参数完整训练,会使模型的精确度显著提高,但误差方面优化程度不高,且考虑到数据集训练样本不够多的情况,易出现过拟合现象。在整体模型鲁棒性方面,选择④ $\alpha = 0.7, \gamma = 1.5$ 是最优解,其 VOT 评估分数高达 0.106,用该参数完整训练数据集后得到的模型鲁棒性是最好的。综合上述,在数据集训练样本较少的情况下,选取鲁棒性更高的模型更具有优势,因此选择参数④ $\alpha = 0.7, \gamma = 1.5$ 进行最终的完整训练。

3.4 实验结果分析

在消融实验部分,选择全部数据集进行测试,基本参数和通用模型参数不变,控制训练周期,实验取 1 ~ 10 次迭代的最优结果。VOT 评估结果如表 3 所示,OTB 评估结果如表 4 所示。

表 3 消融实验下不同条件下的各模型 VOT 最优评估结果

Backbone & Loss	Accuracy	Robustness	Lost Number	EAO
①ResNet50 & Weight_L1_Loss	0.225	1.789	382.0	0.053
②SE_ResNet50 & Weight_L1_Loss	0.420	1.241	245.0	0.119
③ResNet50 & Balanced_L1_Loss	0.310	1.161	248.0	0.094
④SE_ResNet50 & Balanced_L1_Loss	0.399	0.918	196.0	0.131

表 4 消融实验下不同条件下的各模型 OTB 评估结果

Backbone & Loss	Success	Precision
①ResNet50 & Weight_L1_Loss	0.321	0.434
②SE_ResNet50 & Weight_L1_Loss	0.354	0.510
③ResNet50 & Balanced_L1_Loss	0.184	0.316
④SE_ResNet50 & Balanced_L1_Loss	0.345	0.485

综合 VOT 和 OTB 的评估结果,对比原有模型 ResNet50 & Weight L1 Loss,在使用 SENet 和 ResNet50 融合后的 Se_ResNet50 与 Balanced L1 Loss 的 SiamRPN++,其在 VOT 和 OTB 表现优异。在 VOT 评估中,与原有模型相比,Accuracy 略微下降,在 Robustness、Loss Nmber 和 EAO 评估指标方面,都有显著提升。Robustness 比原有模型提升了 40.6%,Loss Nmber 提升了 40.6%,EAO 提升了 33.6%。在 OTB 评估中,Success 基本不变,Precision 提升了 11.7%。上述数据表明优化后的 SiamRPN++,准确度相对于原模型略微提高,稳定性和鲁棒性显著提高。

上述实验结果是优化后的 Backbone & Loss 和原模型中的 Backbone & Loss 进行比较,除此之外,实验还对官方模型 MobileNetV2, SE_ResNet101, SE_ResNet152,ResNet101 和 ResNet152 进行了实验,迭代次数和训练数据集均与上述一致,并进行 VOT 和 OTB 评估,具体评估结果如表 5、表 6 所示。

表 5 SE_ResNet40,SE_ResNet101,SE_ResNet152,ResNet50,MobileNetV2,ResNet101 和 ResNet152 的 VOT 最优评估结果

Backbone & Loss	Accuracy	Robustness	Lost Number	EAO
①SE_ResNet50 & Balanced_L1_Loss	0.399	0.918	196.0	0.131
②SE_ResNet101 & Balanced_L1_Loss	0.337	1.110	237.0	0.101
③SE_ResNet152 & Balanced_L1_Loss	0.337	1.086	232.0	0.101
④ResNet50 & Weight_L1_Loss	0.415	1.545	330.0	0.098
⑤MobileNetV2 & Weight_L1_Loss	0.303	1.124	240.0	0.093
⑥ResNet101 & Weight_L1_Loss	0.346	1.559	333.0	0.079
⑦ResNet152 & Weight_L1_Loss	0.332	1.803	385.0	0.074

表 6 SE_ResNet40,SE_ResNet101,SE_ResNet152,ResNet50,MobileNetV2,ResNet101 和 ResNet152 的 OTB 最优评估结果

Backbone & Loss	Success	Precision
①SE_ResNet50 & Balanced_L1_Loss	0.345	0.485
②SE_ResNet101 & Balanced_L1_Loss	0.236	0.371
③SE_ResNet152 & Balanced_L1_Loss	0.235	0.369
④ResNet50 & Weight_L1_Loss	0.331	0.470
⑤MobileNetV2 & Weight_L1_Loss	0.205	0.318
⑥ResNet101 & Weight_L1_Loss	0.190	0.257
⑦ResNet152 & Weight_L1_Loss	0.185	0.283

为了更好地显示出网络深度对模型性能的影响,从表 5 中的④,⑥和⑦中可得知加深原有模型的网络深度对模型的影响。在不优化 Loss 函数的前提下,通过加深 ResNet 网络深度,并不能提升模型性能。在各项指标方面,均有明显下降。对比 ResNet50,ResNet101 的 EAO 指标下降了 19.39%,ResNet152 的 EAO 指标下降了 24.49%。说明了加深网络深度无法提升模型性能。从表 5 的①,②和③中,Loss 优化为 Balanced L1 Loss,也可以看出不同网络深度对整体模型的影响。在使用 SE_ResNet101 和 SE_ResNet152 增加网络深度后,与 SE_ResNet50 相比,两者在 Accuracy 指标都下降了 15.54%;在 Robustness 两者略微上升;在 Lost Number 方面两者上升了至少 18.37%;在 EAO 方面,SE_ResNet101 和 SE_ResNet152 指标相同,都下降了 22.90%。上述数据表明在增加网络深度后,

VOT 各项评估指标均有降低,在 Accuracy 和 EAO 方面降低尤甚。对比表 6 数据,OTB 评估数据也表示在增加网络深度后 Success 和 Precision 指标均有较大程度降低。总体来说,SiamRPN++基础模型中缓解平移不变性的问题上具有一定限度。

对比原模型中的 ResNet50 & Weight_L1_Loss,使用 Balanced L1 Loss 的 SE _ ResNet101 和 SE _ ResNet152 在 Accuracy 方面显著下降,但是在 Robustness 有明显提升,两者 EAO 与原模型相比差距较小,而在 SE_ResNet50 中除了 Accuracy 方面下降,其他均有明显提升。横向对比表 5 中的②和⑤,③和⑦,也能验证在优化 Backbone 和 Loss 后有明显提升,其中最显著提升的是 SE_ResNet50 & Balanced L1 Loss。在表 6 中,SE_ResNet50 & Balanced L1 Loss 的表现最佳,对比原模型 ResNet50 & Weight_L1_Loss,Success 和 Precision 分别提高了 4.23% 和 3.19% 。综上所述,优化 SE_ResNet50 & Balanced L1 Loss 后的 SB_SiamRPN++模型鲁棒性和精确度方面都有不同程度的提高。

在上述相同的实验条件下,经过 10 ~ 20 次迭代训练后,SB_SiamRPN++与 SiamCar,SiamFC 和 SiamRPN++所得到的 VOT,OTB 数据结果如表 7、表 8 所示。

表 7　SB_SiamRPN++,SiamCar,SiamFC 和 SiamRPN++的 VOT 最优评估结果

Network	Accuracy	Robustness	Lost Number	EAO
①SB_SiamRPN++	0.399	0.918	196.0	0.131
②SiamCar	0.471	1.793	383.0	0.095
③SiamFC	0.451	0.965	206.0	0.133
④SiamRPN++	0.415	1.545	330.0	0.098

表 8　SB_SiamRPN++,SiamCar,SiamFC 和 SiamRPN++的 OTB 最优评估结果

Network	Success	Precision
①SB_SiamRPN++	0.345	0.485
②SiamCar	0.338	0.482
③SiamFC	0.455	0.614
④SiamRPN++	0.331	0.470

在表 7 中,Accuracy 方面 SiamCar 最高,SB _ SiamRPN++最低,相差 0.072;在 Robustness 方面,SB_ SiamRPN++最低,SiamCar 最高,相较于 SiamRPN++,SB_SiamRPN++降低了 40.58% ,数据与 SiamFC 相近;在 Loss Number 方面,与 Robustness 类似,SB _ SiamRPN++相较于 SiamRPN++提高了 40.61% ;在 EAO 方面,SiamFC 与 SB_SiamRPN++实验数据相近,两者都在较大程度上领先其他两个模型。

在表 8 中,Success 和 Precision 方面 SiamFC 的数据结果最优,SB_SiamRPN++在 OTB 中并不占据优势,但相对于 SiamRPN++模型仍有较大程度上的提升,其与 SiamCar 的实验数据相近。

从整体实验数据表明,预测 SB_SiamRPN++最终训练效果相对于 SiamRPN++有较大提升,与 SiamCar 所呈现的效果相似。SiamFC 的准确性和预测性会比 SB_SiamRPN++更高,SB_SiamRPN++的 Robustness 指标上,对于 SiamFC 上并不占据优势。整体来看 SiamFC 会比 SB_SiamRPN++效果更好,SB_SiamRPN ++对比其他两者占据优势。

4　结束语

目前,无人机在各个领域都有广泛的应用,并且单目标跟踪技术备受关注。在任务协助和军事等领域,无人机表现出色,而单目标跟踪算法的不断改进可以更好地帮助人类完成困难任务或自主完成高难度任务。该文对 SiamRPN++算法进行了修改和优化,主要集中在 Backbone 和 Loss 函数方面,取得了显著的改进效果。相对于 SiamRPN++算法,在保持准确度基本不变的情况下,降低了误差率,并提高了模型的鲁棒性。

尽管与一些最新的目标追踪算法相比,该算法的准确率略有不足,但这些追踪算法模型通常庞大且复杂,无法在无人机上得到很好的应用。而 SiamRPN 系列算法具有较强的实时性和轻量级模型架构,因此在实际应用中,对该算法进行改进具有较强的实用性。

该研究中的创新点:

(1)Backbone:使用 SENet 与 ResNet 融合,简化了数据复杂性,增强了算法精确度,使预测结果更加精准。SENet 的 Sigmoid 和 Scale 操作极大地减少了参数量。SE_ResNet50 网络具有更多非线性,在拟合通道间的复杂相关性上有明显提升,还融合了 ResNet50 的网络优点,网络深度更深,且不存在梯度消失问题。

(2)Loss:使用 Balanced L1 Loss,能更好地提升回归梯度,使学习更加平衡,提高了回归任务中的精准度。

在未来可以继续改进的地方:

(1)在数据集方面,可以选用更大更多的数据集,如:ILSVRC2018 和 ILSVRC2019,MS COCO 数据集等。但碍于机器设备原因,无法进行巨大数据集的训练。

(2)在 Backbone 方面,还能将 ResNet 更换成 ResNeXt,能更好地提升计算能力,需要修改的工程量以及匹配通道数等各项参数上要花费更长的时间去优化、提升。

(3)在 Loss 方面,该文是修改回归损失函数,还能

将分类损失函数作适当调整,也可以使用其他损失函数,如 Smooth L1 Loss,KL Loss 等。

回顾第一章中所提到的发展图,基于深度学习算法的单目标跟踪模型,经历了从简单模型到复杂模型,从复杂模型中提炼关键和结构优化的过程。该文实现了无人机视频的单目标跟踪算法 SiamRPN++的优化,除了单目标跟踪任务,多目标跟踪任务也十分重要。多目标跟踪相比于单目标跟踪,其目标识别、定位和预测等问题更加复杂,还需要克服跟踪目标数量繁多和目标种类不一等难点。未来,无人机领域可能会扩展到民用,无人机的智能化和人性化使操作更加快捷方便,在娱乐、拍摄和运输等方面能带来极大的便利。

参考文献:

[1] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]//2016 IEEE conference on computer vision and pattern recognition (CVPR). Las Vegas:IEEE,2016:770-778.

[2] 李 玺,查宇飞,张天柱,等. 深度学习的目标跟踪算法综述[J]. 中国图象图形学报,2019,24(12):2057-2080.

[3] HENRIQUES J F,CASEIRO R,MARTINS P,et al. Exploiting the circulant structure of tracking-by-detection with kernels[C]//Computer Vision - ECCV 2012. Florence:Springer,2012:702-715.

[4] HENRIQUES J F,CASEIRO R,MARTINS P,et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2014,37(3):583-596.

[5] WANG Q,GAO J,XING J,et al. Dcfnet:discriminant correlation filters network for visual tracking[J]. arXiv:1704.04057,2017.

[6] DANELLJAN M,HÄGER G,KHAN F S,et al. Learning spatially regularized correlation filters for visual tracking[C]//2015 IEEE international conference on computer vision (ICCV). Santiago:IEEE,2015:4310-4318.

[7] 王红涛,邓淼磊,赵文君,等. 基于深度学习的单目标跟踪算法综述[J]. 计算机系统应用,2022,31(5):40-51.

[8] KRIZHEVSKY A,SUTSKEVER I,HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM,2017,60(6):84-90.

[9] LI B,YAN 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. Salt Lake City:IEEE,2018:8971-8980.

[10] LI B,WU W,WANG Q,et al. SiamRPN++:evolution of siamese visual tracking with very deep networks[C]//2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Long Beach:IEEE,2019:4277-4286.

[11] ZHU Z,WANG Q,LI B,et al. Distractor-aware siamese networks for visual object tracking[C]//Proceedings of the European conference on computer vision (ECCV). [s. l.]:[s. n.],2018:101-117.

[12] WANG Q,ZHANG L,BERTINETTO L,et al. Fast online object tracking and segmentation:a unifying approach[C]//2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Long Beach:IEEE,2019:1328-1338.

[13] ZHANG Z,PENG H. Deeper and wider siamese networks for real-time visual tracking[C]//2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Long Beach:IEEE,2019:4586-4595.

[14] YAN B,PENG H,FU J,et al. Learning spatio-temporal transformer for visual tracking[C]//2021 IEEE/CVF international conference on computer vision (ICCV). Montreal:IEEE,2021:10428-10437.

[15] CUI Y,JIANG C,WANG L,et al. MixFormer:end-to-end tracking with iterative mixed attention[C]//2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR). New Orleans:IEEE,2022:13598-13608.

[16] GAO S,ZHOU C,MA C,et al. Aiatrack:attention in attention for transformer visual tracking[C]//European conference on computer vision. [s. l.]:Springer,2022:146-164.

[17] GUSTAFSSON F K,DANELLJAN M,TIMOFTE R,et al. How to train your energy-based model for regression[J]. arXiv:2005.01698,2020.

[18] CUI Y,JIANG C,WANG L,et al. Target transformed regression for accurate tracking[J]. arXiv:2104.00403,2021.

[19] ZHANG Z,PENG H,FU J,et al. Ocean:object-aware anchor-free tracking[C]//Computer vision - ECCV 2020. Glasgow:Springer,2020:771-787.

[20] HU J,SHEN L,SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF conference on computer vision and pattern recognition. Salt Lake City:IEEE,2018:7132-7141.

[21] DENG J,DONG W,SOCHER R,et al. ImageNet:a large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. Miami:IEEE,2009:248-255.

[22] 朱命昊. 基于深度注意网络的图像分类[D]. 西安:西安电子科技大学,2021.

[23] PANG J,CHEN K,SHI J,et al. Libra R-CNN:towards balanced learning for object detection[C]//2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Long Beach:IEEE,2019:821-830.

[24] KRISTAN M. The sixth visual object tracking VOT2018 challenge results[C]//Computer vision - ECCV 2018. [s. l.]:Springer,2018.

[25] WU Y,LIM J,YANG M H. Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2015,37(9):1834-1848.

[26] 卢宁宁. 基于深度学习的目标检测与识别的应用研究[D]. 南京:东南大学,2021.