控制与决策
Control and Decision

基于注意力特征融合的无人机多目标跟踪算法

刘芳, 浦昭辉, 张帅超

引用本文:
刘芳,浦昭辉,张帅超. 基于注意力特征融合的无人机多目标跟踪算法[J]. 控制与决策, 2023, 38(2): 345−353.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2021.1098

基于注意力特征融合的无人机多目标跟踪算法

刘 芳†, 浦昭辉, 张帅超

(北京工业大学 信息学部，北京 100124)

摘 要: 随着无人机技术的不断发展,无人机多目标跟踪已成为无人机应用的关键技术之一. 针对无人机视频中的复杂背景干扰、遮挡、视点高度和角度多变等问题,提出一种基于注意力特征融合的无人机多目标跟踪算法. 首先,将改进的卷积注意力模块引入残差网络,建立三元组注意力特征提取网络;其次,在特征金字塔网络的结构上加入新的特征融合通道,设计多尺度特征融合模块,增强模型对多尺度目标的特征表达能力;最后,根据目标的重识别特征匹配与检测框匹配得到目标轨迹. 仿真实验结果表明,该算法可有效提升无人机多目标跟踪的精度,具有较好的鲁棒性.

关键词: 无人机；计算机视觉；多目标跟踪；卷积神经网络；注意力机制；特征融合

中图分类号: TP391 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2021.1098

引用格式: 刘芳,浦昭辉,张帅超. 基于注意力特征融合的无人机多目标跟踪算法[J]. 控制与决策, 2023, 38(2): 345-353.

UAV multi-target tracking algorithm based on attention feature fusion

LIU Fang†, PU Zhao-hui, ZHANG Shuai-chao

(Faculty of Information Technology，Beijing University of Technology，Beijing 100124，China)

Abstract: With the development of UAV technology, multi-target tracking of UAV video has become one of the key technologies in the application of UAVs. Aiming at the problems of complex background interference, occlusion, variable viewpoint of height and angle in UAV multi-target tracking video, a multi-target tracking algorithm based on attention feature fusion is proposed. Firstly, the improved convolution attention module is introduced into the residual network, and a three-tuple attention feature extraction network is established. Then, a new feature fusion channel is added to the structure of the feature pyramid network, and a multi-scale feature fusion module is designed to enhance the model's ability to express the features of multi-scale targets. Finally, the target trajectory is obtained by target re-identification feature matching and bounding box matching. The simulation results show that the algorithm effectively improves the accuracy and robustness of the UAV multi-target tracking.

Keywords: UAV；computer vision；multi-target tracking；convolutional neural network；attention mechanism；feature fusion

0 引 言

随着无人机在军用及民用领域的广泛应用,以无人机为平台的图像获取和处理技术在军事、交通、物流和摄影等诸多领域得到快速发展. 基于无人机视觉的多目标跟踪技术已成为一项重要的研究课题. 而无人机采集的视频中往往存在背景干扰、遮挡、视点高度和角度多变等情况,因此在复杂环境下实现鲁棒、高精度的目标跟踪成为无人机多目标跟踪的主要研究方向.

近年来,人工智能技术的快速发展促进了深度学习在目标跟踪领域的应用[1]. 其中,卷积神经网络因具有强大的目标特征提取能力,在多目标跟踪任务上被广泛使用. 在线实时多目标跟踪(sort)算法[2]利用卷积神经网络提供的检测结果,结合卡尔曼滤波预测和匈牙利匹配算法实现了对于多目标的检测与跟踪,但无法应对目标被遮挡的情况,一旦出现遮挡便会丢失目标;深度关联度量在线实时多目标跟踪(deepsort)算法[3]在sort算法基础上进行改进,在数据的关联跟踪部分引入重识别特征,使得部分被遮挡的物体能够被重新识别. 在无人机视觉领域中,文献[4]提

收稿日期: 2021-06-24；**录用日期:** 2021-10-27.
基金项目: 国家自然科学基金项目(61171119).
责任编委: 巩敦卫.
†通讯作者. E-mail: liufang@bjut.edu.cn.

出了一种基于薄板样条函数的无人机多目标跟踪方法,优化了位置漂移等问题;文献[5]使用Darknet与ResNet提取检测结果与重识别特征,搭建了无人机平台的检测与跟踪算法,但由于使用两个神经网络导致其结构较为冗余;文献[6]提出了一种联合提取检测特征和重识别特征的多目标跟踪模型,并使用无人机视频进行了测试.

由于在无人机采集的视频中,存在复杂背景和追踪目标所占像素较小等因素,跟踪目标受到复杂环境的遮挡、目标与目标之间的相互遮挡情况频繁发生. 一般的多目标跟踪算法在这种场景下对跟踪目标精确检测与重识别的能力有限,导致无人机视频跟踪结果中出现多个目标轨迹相互跳变、单个目标跟踪轨迹碎片化等问题,使得多目标跟踪算法在无人机场景下的多目标跟踪精度大幅降低. 由此可见,鲁棒的无人机多目标跟踪算法的设计与应用仍面临着挑战.

使用注意力机制对检测与重识别进行优化是目前机器视觉领域的研究前沿,文献[7]在无人机目标检测任务上使用了双重注意力机制,文献[8]使用注意力机制实现了重识别网络感受野的自适应. 以上注意力机制在处理跟踪任务时往往由于其过多的参数而导致模型较大,不能满足处理跟踪任务的实时性需要,因此需要对注意力模块的性能与结构进行优化,进一步平衡其速度与性能指标.

综上所述,针对无人机多目标跟踪中目标检测与重识别精度较低、跟踪轨迹碎片化等问题,提出一种基于注意力特征融合的无人机多目标跟踪算法. 首先,为了使多目标跟踪算法拥有更为精确的检测与重识别能力,设定添加卷积注意力模块[9]的骨干网络作为基线算法. 本模型在此基础上设计了改进的三元组注意力特征网络(triple attention-residual network, TA-ResNet). 通过提取特征在不同维度上的联合注意力信息,减少模型参数并有效融合卷积核不同位置的注意力信息. 优化特征提取以提升检测精度,并增强对目标的重识别能力. 其次,在多目标跟踪算法的特征提取部分设计一种改进的特征金字塔模块,通过加入新的特征融合通道,将不同尺度的特征在上采样层进行多尺度加权融合. 同时结合可变形卷积[10]在上采样层进行插值采样,构建针对多目标跟踪的特征融合模块(layers aggregation block, LA-Block). 仿真实验结果表明,该算法在满足实时性要求的前提下,可有效提升无人机多目标跟踪的精度,降低重识别错误次数.

1 算法模型

为了保证无人机在复杂环境中完成精确有效的多目标跟踪任务,提出一种基于注意力特征融合的无人机多目标跟踪算法. 该算法模型的总体网络结构如图1所示.

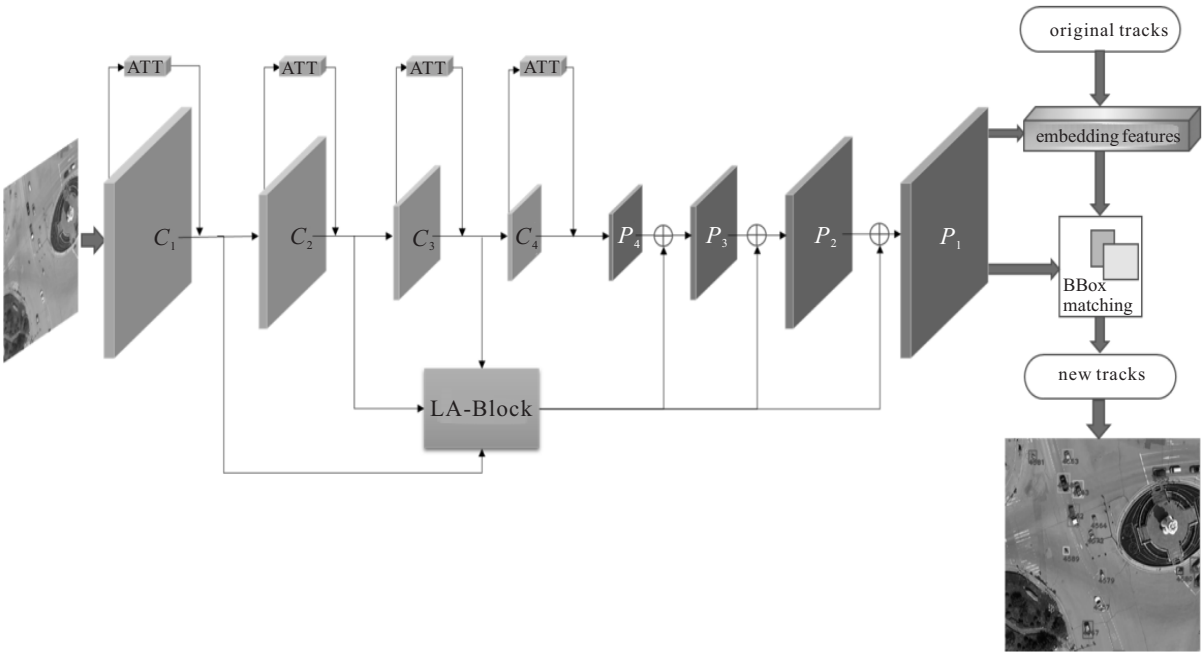

图1 算法总体网络结构

该算法主要分为两部分:第1部分为基于改进三元组注意力机制的特征提取网络(TA-ResNet),通过引入注意力机制,帮助残差网络[11]更好地学习无人机视频中目标的位置和语义信息;第2部分为基于特

征金字塔的特征融合多目标跟踪模块,在特征金字塔网络[12](FPN)的结构上设计新的特征融合通道,构建多尺度特征融合模块(LA-Block),增强模型对多尺度目标的特征表达能力. 在训练过程中,网络对输出特征的每个位置均会生成是否含有目标中心点的概率值,计算概率值与真实值的差得到模型损失并优化损失函数. 在测试推理过程中,对LA-Block的输出进行采样回归,得到多目标检测框与重识别特征. 经过检测框与重识别特征的级联数据匹配得到多目标的跟踪结果.

1.1　三元组注意力特征提取网络

注意力机制作为深度学习模型的通用辅助模块,已经在目标检测、目标跟踪及重识别等任务上得到了广泛的应用. 注意力机制通过卷积池化提取特征通道间的关系,能更好地优化卷积神经网络的权重分布,对无人机多目标跟踪模型使用注意力进行优化,可以有效提高跟踪精度. 但是,由于常见的注意力模块的级联池化操作会使特征在降维过程中损失大量的空间信息,不利于多目标跟踪中的特征匹配任务,导致频繁的目标轨迹序号切换问题. 针对一般注意力模块中空间信息损失过多的问题,提出改进的三元组注意力结构. 通过分别提取各个维度上的注意力特征,组成跨维度注意力特征三元组,再将三元组特征加权融合,得到对卷积神经网络进行权重优化的结果. 改进的注意力机制在进行池化的过程中可保留大量空间位置信息,将其引入残差网络模型,搭建三元组注意力特征提取网络TA-ResNet.

改进三元组注意力通过并联的形式表示注意力权重,其具体结构如图2所示. 分别对输入特征按照(C、H、W)三个维度进行降维,得到通道维与特征高度维的维度注意力(width attention)、通道维与特征宽度维的维度注意力(height attention)以及特征高度与宽度维的空间注意力(spatial attention).

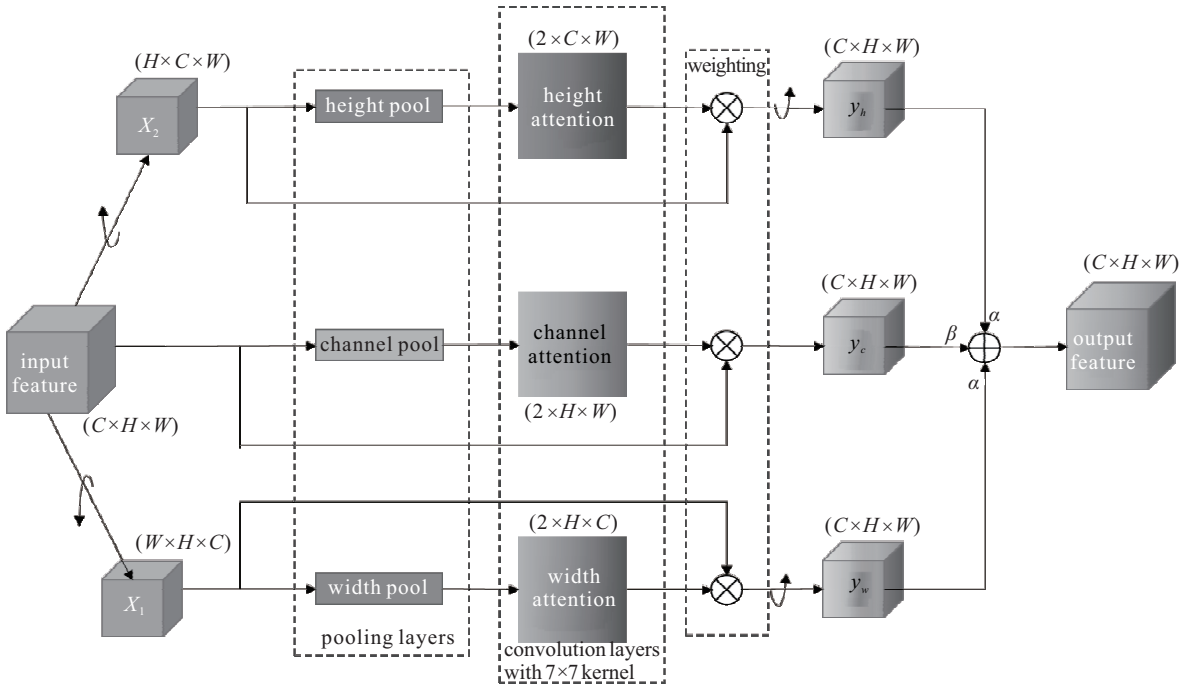

图 2　三元组注意力结构

width attention的提取方式是输入特征按照W维度在$H \times C$平面上进行降维. 首先对输入特征$x_1(C \times H \times W)$转置,得到维度顺序为$(W \times H \times C)$的$\hat{x}_1$. 其次对$\hat{x}_1$进行降维,得到降维后的结果$\hat{x}_1^*(2 \times H \times C)$. 将$\hat{x}_1^*$通过一个大小为$7 \times 7$的卷积核并经过Sigmod激活函数进行注意力提取,再进行维度还原便得到channel维与height维的关联注意力权重张量,定义其输出y_w如下:

$$y_w = \overline{\hat{x}_1 \sigma(\psi_1(\hat{x}_1^*))}. \tag{1}$$

其中:$\psi_1(\,)$为卷积操作,$\sigma(\,)$为Sigmod激活函数.

同理,height attention将特征按照height维度在$C \times W$平面上进行降维,提取channel维与width维的关联注意力并加权,其输出y_h如下:

$$y_h = \overline{\hat{x}_2 \sigma(\psi_2(\hat{x}_2^*))}, \tag{2}$$

其中$\hat{x}_2^*(2 \times C \times W)$由输入特征改变维度顺序并降维得到. 将$\hat{x}_2^*$经过$7 \times 7$卷积核$\psi_2(\,)$与Sigmod激活函数$\sigma(\,)$后与输入特征相乘加权,再经过维度顺序还原得到三元组注意力中的特征横向位置与通道的联合

注意力y_h.

三元组中的空间注意力的提取则是将特征按照channel维度在$H \times W$平面进行降维,提取空间注意力y_c,其公式如下:

$$y_h = x_3\sigma(\psi_3(\hat{x}_3)). \tag{3}$$

其中$\hat{x}_3$为输入特征经过降维得到的特征,其维度为$(2 \times H \times W)$. 按照相同的卷积激活操作得到权重后进行加权,得到特征纵向位置与横向位置的联合注意力,即空间注意力y_c.

将得到通道跨维度关联的注意力特征与空间注意力特征进行加权融合,能够得到最终输出的三元组特征注意力作为注意力模块的输出. 将跨维度注意力模块作用于输入特征,能够使其在通道上包含更多空间维度的特征响应. 因此,为了最大优化跟踪任务中的重识别特征,需要将模型在通道维度上的注意力进行进一步的改进. 通过设定空间注意力以及跨维度注意力特征权重,使输入特征在经过注意力模块后得到的输出特征包含更多的跨维度信息,更精确地反应目标的具体语义信息,其加权过程为

$$Y = \alpha(y_w + y_h) + \beta y_c. \tag{4}$$

其中:Y为注意力模块的输出;α、β为三元组注意力权值,为避免信息冗余,分别设定为0.4和0.2.

传统注意力机制的级联池化降维操作,其注意力权重往往在一个$(1 \times 1 \times C)$的一维向量上进行学习,再对每个位置进行注意力加权. 而三元组注意力则是通过并联的三次降维,使得注意力权重的学习在3个二维张量$(2 \times C \times W$、$2 \times H \times C$、$2 \times H \times W)$上进行,使得学习需要的参数量仅取决于卷积核的大小而与通道个数无关. 对于多目标跟踪中的重识别任务而言,通常需要提取到很高的特征维度才能获得很好的表达能力(通常为64～256),这使得传统注意力模型的参数较多. 传统注意力结构SE模块、CBAM模块与本算法三元组注意力模块的运算参数量分别如下:

$$P_{\text{SeNet}} = 2C^2/s, \tag{5}$$

$$P_{\text{CBAM}} = 2C^2/s + 2f^2, \tag{6}$$

$$P_{\text{T-Att}} = 6f^2. \tag{7}$$

其中:C为特征的通道维度(64～256),s为下采样倍数(通常为4),f为卷积核的大小. 将SE模块及CBAM模块与三元组注意力的运算参数量进行对比可以发现,三元组注意力在提取的特征通道数较高时,由于其结构只涉及了一次,其参数量明显低于传统注意力结构.

三元组注意力机制的关键便是通过改变输入特征张量的维度顺序,按照不同方向对于输入特征张量进行单次降维,再将获得的注意力权重进行加权融合,得到更为注重维度之间相互关系的注意力机制. 基于骨干网络ResNet34,引入三元组注意力机制构建的TA-ResNet特征提取网络,其结构参数如表1所示.

表1 三元组注意力提取特征网络模型

layer	type	kernel	output size
X	input	/	$1\,088 \times 608$
conv-1	conv	$7 \times 7, 64$, stride 2	272×152
conv-2	conv	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3 +$ T-ATT	136×76
conv-3	conv	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4 +$ T-ATT	68×38
conv-4	conv	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6 +$ T-ATT	34×19
conv-4	conv	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3 +$ T-ATT	17×9

1.2 特征融合多目标跟踪模块

在无人机多目标跟踪任务中,由于跟踪目标较多,单个目标较小,提取特征后需要经过上采样网络对高维特征进行降维来获得位置信息与语义信息平衡的特征,用于Re-ID特征及目标框的级联匹配. 特征金字塔网络在多尺度上采样的同时融合了同尺度的下采样特征,构成了简单的特征融合模块. 而在无人机多目标跟踪任务中,无人机飞行拍摄时存在视野较大、视距较远等特点. 在现有无人机多目标跟踪算法中,语义信息大多被固定在低分辨率的高维特征上,对目标识别能力有限,导致后续跟踪匹配任务输出的跟踪轨迹的精度较低.

为了解决多尺度特征语义信息分配不均而导致的跟踪精度下降等问题,本文算法在注意力特征提

取网络后加入LA-Block,在特征金字塔组成的上采样模块中加入了使用可变型卷积进行上采样的特征融合通道,能够使特征金字塔中每层融合特征的语义信息更为丰富,位置信息更为精确,从而提升网络输出特征的质量. LA-Block使用基于可变形卷积(deformable convolution,简称deform-conv)的网络层进行上采样. deform-conv卷积核在每一个元素上额外增加一个方向参数,在训练过程中通过对生成方向参数的卷积核进行额外的权重优化,得到能生成更有利于反映目标特征的卷积核. 使用可变形卷积进行上采样可以使特征包含更为丰富的空间位置信息,进一步增强多尺度特征对目标的表达能力.

LA-Block结构如图3所示.

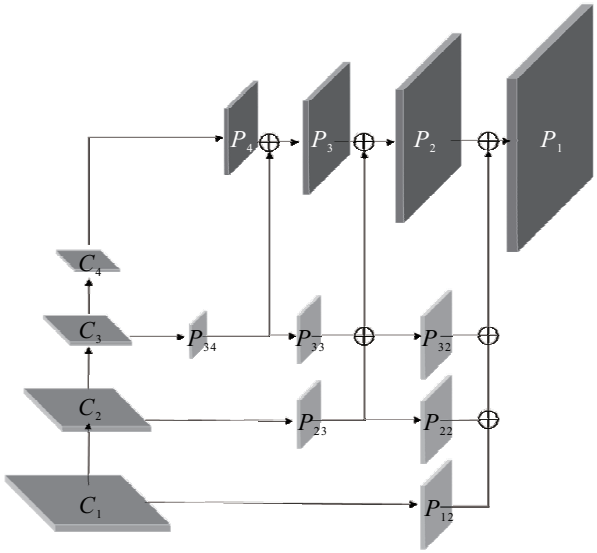

图 3　LA模块结构

在图3中: $\{C_1, C_2, C_3, C_4\}$为三元组注意力特征提取网络各个下采样阶段输出的特征,$\{P_1, P_2, P_3, P_4\}$为LA-Block在各个阶段上采样中的输出结果. 各阶段具体步骤如下.

step 1: 将下采样层得到的输出特征C_4经过保持特征尺度的可变形卷积层得到P_4,即

$$P_4 = \psi(C_4), \tag{8}$$

其中$\psi(\,)$为保持尺度的可变形卷积.

step 2: 将P_4与C_3经过上采样deform-conv层得到的输出P_{34}进行加权融合,并经过转置卷积上采样得到P_3,即

$$P_3 = T(\delta_3 P_4 + \varepsilon_3 \cdot \zeta(C_3)). \tag{9}$$

其中: $T(\,)$为转置卷积上采样,$\zeta(\,)$为deform-conv上采样,δ、ε为权重参数.

step 3: 将P_3与经过deform-conv上采样得到的P_{33}、经过deform-conv进行同尺度采样得到的P_{23}加权融合,再经转置卷积上采样得到P_2,即

$$P_2 = T(\delta_2 \cdot P_3 + \varepsilon_2 \cdot (\zeta(P_{34}) + \psi(C_2))). \tag{10}$$

step 4: 同理,将P_2与经过deform-conv上采样得到的P_{32}、P_{22}与经过deform-conv同尺度采样得到的P_{12}加权融合,再经转置卷积上采样得到P_1,即

$$P_1 = T(\delta_1 \cdot P_2 + \varepsilon_1 \cdot (\zeta(P_{33}) + \zeta(P_{23}) + \psi(C_1))). \tag{11}$$

经过4个阶段的特征融合采样操作,可以得到LA特征融合模块的具体公式,即

$$\begin{cases} P_1 = T(\delta_1 \cdot P + \varepsilon_1 \cdot (\zeta(P_{33}) + \zeta(P_{23}) + \psi(C_1))), \\ P_2 = T(\delta_2 \cdot P_3 + \varepsilon_2 \cdot (\zeta(P_{34}) + \psi(C_2))), \\ P_3 = T(\delta_3 \cdot P_4 + \varepsilon_3 \cdot \zeta(C_3)), \\ P_4 = \psi(C_4). \end{cases} \tag{12}$$

为了避免特征信息冗余,同时为了满足后续分组关联任务所需要的特征尺度的要求,权重组设定为$\delta_{1,2,3} = \{0.7, 0.6, 0.5\}, \varepsilon_{1,2,3} = \{0.1, 0.2, 0.5\}$.

1.3　分组预测与特征关联

输入图片经过基于注意力机制的下采样网络及特征融合模块后,生成采样倍率为4的输出特征,再将网络输出特征进行分组预测以得到两帧间数据关联任务所需要的Re-ID特征及检测框.

首先,将网络输出特征并行通过3个3×3卷积和1×1卷积得到3个针对输出特征的降维采样结果,即3个特征头(feature head). 对3个特征头分别进行中心点响应热图、目标框大小回归和中心点偏移量回归. 中心点热图(center-point heatmap)特征头的形状为(n, H, W). 其中: n为所检测的目标种类数量,H和W为高度和宽度,反应了多目标预测的中心点位置. 目标框大小的形状(B-box size)与中心点偏移(center offset)特征头的形状均为$(2, H, W)$,框大小回归给出了热图中每个位置上目标框的宽高(W, H)预测值,而偏移量回归则为了弥补中心相应热图中由于下采样产生的中心点位移,给出了热图中每个位置上的中心点偏移量(x, y). 如果热图在某位置没有中心点响应,则其B-box size与center offset均为0. 其次,算法根据中心点热图中存在响应的点的坐标,在未降维的输出特征的相应坐标位置上直接提取高维特征组,作为当前帧中全部检测目标的Re-ID特征.

使用特征关联算法对多目标特征进行跟踪匹配时,采用级联匹配器对特征进行关联匹配. 具体而言,首先初始化跟踪序列,根据第1帧的检测框生成原始的多目标轨迹集,保存重识别特征组,并建立长度为60帧的搜索区间,以找到再次出现的被遮挡的目标

并链接正确的轨迹. 使用卡尔曼滤波器[13]预测当前帧的Re-ID特征组所表示的多目标的位置,并通过轨迹集中的多目标位置计算马氏距离,将马氏距离过远的匹配附加上惩罚项,组成代价矩阵(cost matrix). 而后利用匈牙利算法[14]结合cost matrix对重识别特征组与已有轨迹集中的多目标进行二元匹配,将匹配命中的目标加入已经生成的轨迹中. 在当前帧的检测结果中找到未被匹配的目标,将这些目标的边界框与上一帧的目标框进行交并比(IOU)计算,并在匈牙利二元匹配的过程中,使用交并比结果作为匹配的权重,得到目标框的二次匹配组合. 最后,将轨迹集中超过搜索区间长度且仍未被匹配目标的轨迹保存并移出待匹配集,对当前帧未被匹配的目标进行新轨迹创建并加入匹配集,至此,将当前帧中已完成匹配的目标作为现有轨迹的一个新的节点,通过将新的轨迹统一整合,构建新的轨迹集合,完成对于当前帧所检测目标的轨迹跟踪.

2 仿真实验

为验证本算法在无人机多目标跟踪任务上的有效性,选用VisdroneMOT[15]多目标跟踪数据集进行仿真测试. Visdrone2018-MOT多目标跟踪数据集中包括63段有完整标注的无人机多目标跟踪视频,主要目标为行人及车辆. 该数据集中包括运动场地、步行街区、城市道路、高速公路、郊野公园等不同场景. 数据集中存在大量由无人机飞行时复杂环境产生的问题,按照不同的情况对数据集进行统计分析,其结果如表2所示.

表 2 数据集场景分析

视频序列总数	单帧最少目标数大于20	背景遮挡	目标相互遮挡	双向运动
63	43	34	37	46

可见表2中将情况分为无人机与目标的双向运动、密集目标、背景遮挡、目标相互遮挡4个问题. 以上问题在数据集中大量存在,并且均有可能造成跟踪结果中的目标重识别错误、目标检测错误、目标丢失等,导致无人机目标跟踪失败.

实验阶段将数据集随机分为由55段跟踪数据组成的训练样本集和由8段跟踪数据组成的评估测试样本集. 其中: 训练集共计22 970帧,测试集共计4 095帧. 仿真实验使用MOT-challenge[16]上定义的多项指标来评估本算法模型的跟踪轨迹结果,如多目标跟踪准确度(multiple object tracking accuracy, MOTA)、识别F值(identification F-score, IDF1)以及表示目标丢失次数的ID序号切换次数(ID-switch)等.

实验平台采用AMD 2 600x处理器,Nvidia RTX 2080显卡,内存为16 G、Ubuntu 18.04操作系统. 本算法与实验中使用的对比算法的骨干网络均选择在ImageNet[17]上预训练的开源残差网络模型进行特征提取. 本算法在骨干网络的基础上实现了使用注意力机制对特征提取优化,并添加了特征多层融合模块,使算法模型具有更好的重识别能力. 对算法模型在Visdrone2018-MOT训练集上进行30个epoch训练,训练batch size设置为4,初始学习率为2e-4,在第15、25个epoch结束后分别进行倍率为10的学习率衰减,最终学习率为2e-6.

3 实验结果分析

3.1 多目标跟踪结果定性分析

通过对测试集中无人机视频的目标进行跟踪来验证本算法对无人机多目标跟踪的优化效果,并将测试序列的多目标跟踪结果进行可视化,其结果如图4和图5所示.

(a) uav0000126_00001[180,221,250]

(b) uav0000126_00001[265,280,320]

图 4 算法在遮挡条件下的跟踪结果展示

(a) uav0000305_00000[1,16,33]

(b) uav0000071_03240[40,55,70]

(c) uav0000150_02310[6,40,72]

(d) uav0000084_00000[1,15,40]

图 5　算法在不同环境下的可视化跟踪结果展示

图4(a)和图4(b)中为无人机视角下目标被遮挡情况的跟踪结果. 图4展示了标有ID序号的车辆在行驶过程中依次掉头并经过树木遮挡后被重新跟踪并被正确赋予先前ID的结果,可以看到本算法在目标被部分遮挡时仍可以正确识别跟踪目标,在目标被完全遮挡后仍可以找回目标并赋予正确的ID序号,没有出现轨迹跳变的情况,保留了目标的完整轨迹.

图5展示了几种不同环境下的无人机多目标跟踪结果,视频帧序列从左到右为时间顺序. 从图5(a)可以看到,本算法可使无人机在俯视视角下针对十字路口处的密集车辆目标完成正确跟踪;从图5(b)中的可视化结果看出,算法可以正确在夜晚环境下完成对多个交通目标的跟踪;图5(c)中小目标较多,而算法通过注意力特征融合模块,在高分辨率的特征图上更好地提取到了小目标的特征响应,从而达到了准确的小目标检测跟踪效果;从图5(d)中可以看到,针对密度较大的人群目标,在无人机与目标产生双向运动的情况下,算法仍可以对密集行人目标进行准确跟踪定位识别.

3.2　算法模块可行性验证分析

为了验证算法注意力特征提取模块对于性能提升的有效性,将算法与使用ResNet34骨干网络的跟踪算法进行单一变量的对比实验. 通过该实验,对引入注意力特征在跟踪特征提取上的优化效果进行检验. 实验选择了MOTA, IDF1指标与ID-switch次数来进行对比,同时也列出了不同注意力机制在网络上的额外参数量.

实验结果如表3所示. 从表3可以看出,TA-ResNet所需要的额外参数为0.048 M, 远低于SENet[18]和CBAM额外参数量. 但在相同条件下与CBAM在VisDrone多目标跟踪数据集上的多目标跟踪准确度MOTA仅相差了0.9 %, 比SENet的MOTA值领先了2.4 %. 通过对比本算法与其他特征提取网络的识别F值IDF1可以看出,添加了改进三元组注意力机制的模型达到了最高的目标识别跟踪率. 在反应目标ID序号丢失次数的指标ID-switch中,特征

提取网络由于添加了改进的三元组注意力机制,得到了更好的重识别特征,目标丢失次数比其他算法有了明显的降低,有效减少了目标轨迹切换次数,提高了跟踪完整性. 以上结果均表明,TA-ResNet在仅附加极少额外注意力参数的情况下,有效优化了网络的特征提取性能.

表 3 注意力提取特征网络比较

method	MOTA/%	IDF1/%	ID-switch	overhead parameters/M
ResNet	38.78	55.42	1 963	/
SENet	41.02	58.69	1 737	2.514
CBAM	44.35	57.33	1 643	2.532
ours	43.43	58.81	1 509	0.048

为了验证所提算法中特征融合金字塔模块(LA-Block)和FPN特征金字塔对无人机视觉下的多目标跟踪器优化的有效性,将本算法与使用普通上采样级联模型的骨干网络的跟踪算法进行单一变量的对比实验. 通过对比算法在测试样本集上的跟踪精度结果,能够反应本算法特征融合模块对于跟踪结果的优化. 其结果如表4所示.

表 4 注意力提取特征网络比较

method/%	MOTA/%	MOTP/%	recall/%	precision/%
①ResNet+FPN	48.76	36.45	50.16	81.38
②TA-ResNet	43.43	35.19	45.02	79.84
③TA-ResNet+FPN	51.50	37.14	52.17	82.45
④ours	54.82	40.19	56.70	83.31

对比表4中的方法②和方法③可以看出,特征金字塔结构可以明显提升多目标跟踪任务中的各项指标,引入特征金字塔结构可以使特征进行对应尺度的融合,使底层特征的语义信息具有一定的提升. 而对比方法①与方法③可以看出,在同样使用特征金字塔结构进行语义信息优化时,含有注意力机制的底层特征可以进一步地提升包括跟踪精度MOTA、MOTP在内的多项跟踪指标. 同时优化的特征对于目标召回率和召回精度等反应检测的指标也有显著的提升,帮助模型减少目标丢失,提高检测精度. 在此基础上引入融合了多层特征融合LA模块的对比实验③,将高层的语义信息逐层融合到底层特征中,进一步优化模型的检测跟踪性能. 本算法通过增强语义信息的特征,使其对于无人机多目标跟踪有明显的改进效果,显著提升了跟踪结果的各项指标.

3.3 多目标跟踪算法对比实验

为了对比实验阶段的不同主流的多目标跟踪网络通过相同的数据训练得到的模型,比较不同多目标跟踪算法在仿真实验的无人机数据集上的各项指标,并验证本文算法的多目标跟踪性能,对比算法包括deep-sort、JDE、FairMOT[19]. 在实验中,分别采用多目标跟踪准确度(MOTA)、识别F值(IDF1)来评判算法的跟踪精确度,使用帧率(FPS)评价指标对算法在测试集上的平均运行速度进行定量分析.

表5中结果表明,在无人机航拍数据上与主流多目标跟踪算法对比,从目标跟踪精度指标来看,本文算法的IDF1值有明显提高,达到了63.47 %,比FairMOT算法高出了1.5 %. 本文算法在MOTA指标上优于deep-sort和单阶段检测跟踪器JDE,达到了54.82 %,这也表明本文所提算法具有相对更优的特征提取能力和更好检测跟踪精度. 本文算法通过引入改进的三元组注意力特征,虽然在MOTA指标上仍落后FairMOT算法不到 0.5 %,但是在速度上相较其有着明显的提升,达到32 fps,该速度同时也优于其他对比算法. 综上所述,本文算法基本满足无人机的多目标跟踪任务要求,并且取得了速度与精度的更好平衡.

表 5 注意力提取特征网络比较

method	MOTA/%	IDF1/%	FPS
deep-sort	49.23	57.53	27
JDE	53.96	58.31	21
FairMOT	55.31	61.93	24
ours	54.82	63.47	32

4 结 论

本文提出的基于注意力特征融合的无人机多目标跟踪算法能够很好地改善目前无人机多目标跟踪任务中由于复杂背景因素干扰、遮挡、视点高度和角度多变而产生的问题. 利用三元组注意力机制建立了TA-ResNet,在此基础上构建多尺度特征融合模块

LA-Block,将空间尺寸不一的复数特征图通过可变形卷积进行级联上采样并加权融合,增强了特征对目标的表达能力,并利用级联匹配将特征关联为完整多目标跟踪轨迹. 仿真实验结果表明,与主流多目标跟踪算法相比,该跟踪算法具有较高的跟踪精度及运行速度,并且取得了较低的ID切换次数与较高的目标识别率. 本文所提算法在增加最小计算量的情况下有效提升了网络性能,但是特征融合模块的计算代价依然较高. 因此,下一步将继续优化网络结构,以期进一步提高无人机多目标跟踪的精度和速度.

参考文献(References)

[1] 李月峰, 周书仁. 在线多目标视频跟踪算法综述[J]. 计算技术与自动化, 2018, 37(1): 73-82.
(Li Y F, Zhou S R. Survey of online multi-object video tracking algorithms[J]. Computing Technology and Automation, 2018, 37(1): 73-82.)

[2] Bewley A, Ge Z Y, Ott L, et al. Simple online and realtime tracking[C]. 2016 IEEE International Conference on Image Processing. Phoenix, 2016: 3464-3468.

[3] Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric[J/OL]. 2017: arXiv: 1703.07402.

[4] 余仁伟, 朱浩, 蔡昌恺. 基于薄板样条函数的无人机多目标跟踪算法[J]. 仪器仪表学报, 2021, 42(3): 168-176.
(Yu R W, Zhu H, Cai C K. Multi-object tracking algorithm for UAV based on the thin plate spline function[J]. Chinese Journal of Scientific Instrument, 2021, 42(3): 168-176.)

[5] 王旭辰, 韩煜祺, 唐林波, 等. 基于深度学习的无人机载平台多目标检测和跟踪算法研究[J]. 信号处理, 2022, 38(1): 157-163.
(Wang X C, Han Y Q, Tang L B, et al. Multi target detection and tracking algorithm for UAV platform based on deep learning[J]. Journal of Signal Processing, 2022, 38(1): 157-163.)

[6] Wang Z D, Zheng L, Liu Y X, et al. Towards real-time multi-object tracking[C]. The 16th European Conference on Computer Vision (ECCV). Glasgow, 2020: 107-122.

[7] 王胜科, 任鹏飞, 吕昕, 等. 基于中心点和双重注意力机制的无人机高分辨率图像小目标检测算法[J]. 应用科学学报, 2021, 39(4): 650-659.
(Wang S K, Ren P F, Lü X, et al. Small target detection algorithm of UAV high resolution image based on center point and dual attention mechanism[J]. Journal of Applied Sciences, 2021, 39(4): 650-659.)

[8] 王松, 纪鹏, 张云洲, 等. 自适应感受野网络的行人重识别[J]. 控制与决策, 2022, 37(1): 119-126.
(Wang S, Ji P, Zhang Y Z, et al. Adaptive receptive network for person re-identification[J]. Control and Decision, 2022, 37(1): 119-126.)

[9] Woo S, Park J, Lee J Y, et al. CBAM: Convolutional block attention module[C]. The 15th European Conference on Computer Vision (ECCV). Munich, 2018: 3-19.

[10] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[C]. IEEE International Conference on Computer Vision (ICCV). Venice, 2017: 764-773.

[11] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, 2016: 770-778.

[12] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 936-944.

[13] Kalman R E. A new approach to linear filtering and prediction problems[J]. Journal of Basic Engineering, 1960, 82(1): 35-45.

[14] Zhu H. Group role assignment via a Kuhn-Munkres algorithm-based solution[J]. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2012, 42(3): 739-750.

[15] Fan H, Du D, Wen L, et al. VisDrone-MOT2020: The vision meets drone multiple object tracking challenge results[C]. The 16th European Conference on Computer Vision (ECCV). Glasgow, 2020: 713-727.

[16] Milan A, Leal-Taixé L, Reid I, et al. MOT16: A benchmark for multi-object tracking[J/OL]. 2016, arXiv:1603.00831.

[17] Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.

[18] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 7132-7141.

[19] Zhang Y, Wang C, Wang X, et al. Fairmot: On the fairness of detection and re-identifification in multiple object tracking[J/OL]. 2020, arXiv: 2004.01888.

作者简介

刘芳(1971−), 女, 副教授, 从事人工智能、机器视觉、无人机视觉导航、深度学习等研究, E-mail: liufang@bjut.edu.cn;

浦昭辉(1997−), 男, 硕士生, 从事机器视觉、多目标跟踪的研究, E-mail: pzh18432017@gmail.com;

张帅超(1997−), 男, 硕士生, 从事机器视觉、目标跟踪的研究, E-mail: Zhangshuaichao@emails.bjut.edu.cn.

(责任编辑：闫　妍)