基于改进 YOLOv8 与 ByteTrack 的航拍多目标跟踪算法

郑明宇[1],邵慧超[2],邵延华[1],楚红雨[1]

(1.西南科技大学信息工程学院,四川 绵阳 621010;2.立得空间信息技术股份有限公司,湖北 武汉 430070)

摘 要: 目标检测与多目标跟踪技术日益成熟,但在复杂场景下执行航拍多目标跟踪任务时,目标尺寸小、尺寸变化大、遮挡等问题仍会导致检测与跟踪效果不理想。为此,提出一种基于改进 YOLOv8 与 ByteTrack 的航拍多目标跟踪算法(YBTrack)。构建 MSA-YOLO(Multi-Scale Attention YOLO)检测器,设计空间-深度卷积替换 YOLOv8 原有卷积,将空间信息转换为通道维度,有效保留目标细节,减少多尺度特征图融合过程中信息丢失导致的漏检、误检。同时设计轻量加速空间-通道注意力模块,用于颈部卷积,降低计算复杂度,并作为检测头前的特征细化模块,进一步增强算法对目标特征信息提取的能力。为提高跟踪效果,对 ByteTrack 跟踪模型进行优化,设计空间-外观相似度矩阵(ASM),提升模型区分相似目标的性能,并提出目标校正函数,减少卡尔曼滤波器产生的误差积累,降低目标偏移、丢失率。将 MSA-YOLO 与优化后的 ByteTrack 结合,开展多目标跟踪实验。其中,MSA-YOLO 在 VisDrone2019-DET 数据集上平均精度均值(mAP@0.5)提高了 9.4 百分点;多目标跟踪算法在 VisDrone2019-MOT、MOT17 数据集上多目标跟踪准确率(MOTA)分别提升了 11.2 和 8.3 百分点,ID 调和均值(IDF1)分别提升了 8.9 和 7.4 百分点,实验结果表明本文所提方法跟踪效果显著。此外,与其他多目标跟踪算法的对比实验也证明了本文算法的优越性。

关键词: YOLOv8;ByteTrack;多目标跟踪;特征信息;卡尔曼滤波器

中图分类号: TP391 **文献标志码:** A **DOI:** 10.19678/j.issn.1000-3428.0252029

Multi-Target Tracking Algorithm for Aerial Photography Based on Improved YOLOv8 and ByteTrack

ZHENG Mingyu[1], SHAO Huichao[2], SHAO Yanhua[1], CHU Hongyu[1]

(1.School of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, Sichuan, China;
2.Leador Spatial Information Technology Co., Ltd., Wuhan 430070, Hubei, China;)

【Abstract】 Object detection and multi-target tracking technologies are becoming increasingly mature. However, when performing aerial multi-target tracking tasks in complex scenarios, issues such as small target size, large size variation, and occlusion lead to unsatisfactory detection and tracking performances. Therefore, this paper proposes an aerial multi-target tracking algorithm based on an improved YOLOv8 and ByteTrack method called YBTrack. First, a detector called Multi-Scale Attention YOLO (MSA-YOLO) is constructed. The original convolution in YOLOv8 is replaced with a space-depth convolution that transforms spatial information into channel dimensions, effectively preserves target details, and reduces missed and false detections caused by information loss during multiscale feature map fusion. Simultaneously, a lightweight accelerated space-channel attention module is designed for neck convolution to reduce computational complexity. This module also acts as a feature refinement module before the detection head, further enhancing its ability to extract target feature information. Next, the ByteTrack tracking model is optimized to improve the tracking performance. A Spatial-Appearance Similarity Matrix (ASM) is designed to enhance the model's ability to distinguish similar targets. In addition, a target correction function is proposed to reduce the error accumulation of the Kalman filter and decrease the target offset and loss rates. Finally, MSA-YOLO and the optimized ByteTrack are combined for multi-target tracking experiments. MSA-YOLO achieves a 9.4 percentage points improvement in mAP@0.5 on the VisDrone2019-DET dataset. The multi-target tracking algorithm improves the Multiple Object Tracking Accuracy (MOTA) by 11.2 and 8.3 percentage points and the Identification F1 Score (IDF1) by 8.9 and 7.4 percentage points on the VisDrone2019-MOT and MOT17 datasets, respectively. Experimental results demonstrate the excellent tracking performance of the proposed method. Furthermore, comparison experiments with other multitarget tracking algorithms confirm the superiority of the proposed algorithm.

基金项目: 国家自然科学基金(61601382, 62106209)。

作者简介: 郑明宇(CCF 会员),女,硕士研究生,主研方向为目标检测、目标跟踪;邵慧超,高级工程师;邵延华(CCF 会员),副教授、博士;楚红雨(通信作者),教授、博士。

收稿日期: 2025-01-09 **修回日期:** 2025-03-11 **E-mail:** 49456338@qq.com

计算机工程 www.ecice06.com

【Key words】YOLOv8; ByteTrack; multiple target tracking; feature information; Kalman filter

0 引言

随着无人机技术的迅猛发展与“智慧监测”理念的提出,先进的无人机技术与智能“检测+跟踪”算法的结合,将大幅提升监测系统的智能化水平,可广泛应用在人群聚集监测和交通流动监管等场景。目前,航拍复杂场景下基于深度学习的目标检测与多目标跟踪正是研究热点之一[1]。

基于深度学习的目标检测主要有以 YOLO[2] 为代表的单阶段检测方法和以 Faster R-CNN[3] 为代表的双阶段检测方法。单阶段方法将目标检测简化为一个端到端的回归问题,较二阶段方法,通常检测速度更快;YOLO 系列算法在速度与部署方面表现优异,处理常规场景图像效果良好;但面对小目标、目标遮挡或存在复杂背景干扰的情况时,对目标特征信息提取不足,易出现漏检、误检和精度低等问题[4]。为提升目标检测性能,LI 等[5] 提出动态注意力尺度序列融合算法,使用动态检测头提高了对不同类型和尺寸目标的表达能力。SU 等[6] 针对复杂场景中小目标的稠密与遮挡问题,优化特征提取和目标定位,提高小目标检测精度。何植仟等[7] 使用可变形卷积获取丰富的语义信息,提升模型对小目标的检测效果。这些研究为提升目标检测性能提供了重要思路,但仍需进一步优化检测精度与速度的均衡。

基于深度学习的多目标跟踪算法主要分为两类:一类是基于检测的跟踪方法(TBD),其依赖检测器在每一帧中检测目标,再通过跟踪算法将同一目标在不同帧中进行关联,实现连续跟踪;另一类是基于特征关联的联合检测跟踪方法(JDT),其通过提取每个目标的外观、运动轨迹、位置等特征来关联相邻帧中的检测结果[8]。WOJKE 等[9] 基于简单在线实时跟踪(SORT)[10] 提出了 DeepSORT,通过引入外观特征提取模块,增强对复杂背景和遮挡的鲁棒性,但面向复杂场景易出现目标切换、丢失的问题。CAO 等[11] 提出 OC-SORT,通过结合目标检测器的测量信息进行虚拟轨迹修正,该方法在处理遮挡和复杂运动时表现优异,但严重依赖检测器性能。FairMOT[12] 引入双分支结构结合并行检测与重识别(Re-ID)提升检测精度。随着 Transformer[13] 架构的提出,SUN 等[14] 设计了 TransTrack 算法,通过自注意力机制增强目标关联,该算法在遮挡严重的场景中表现出色,但对航拍小目标检测和跟踪的精度较低。

针对现有多目标跟踪算法在遮挡等复杂场景中跟踪精度低、目标丢失等问题,研究者提出了多种改进方法。ZHANG 等[15] 提出 ByteTrack 多目标跟踪算法,在保持轻量化的同时,通过区分高、低置信度目标,并对低置信度目标重新关联,显著提高了小目标和遮挡目标的跟踪精度,但该算法对低置信度的依赖增加了其对检测器性能的要求。卡尔曼滤波器易受外界干扰产生误差积累,从而降低对下一帧运动轨迹的预测精度,增加目标偏移与丢失的次数。AHARON 等[16] 在 ByteTrack 基础上,构建成本矩阵,提高目标在动态变化场景中的匹配精度。WANG 等[17] 通过计算目标之间的外观相似性,提升检测与特征提取的统一性,提高算法在复杂场景中的跟踪效果。YOU 等[18] 为提高卡尔曼滤波器对复杂运动模式的预测能力,改进卡尔曼滤波器的状态向量与协方差矩阵,提高轨迹连续性,减少目标的丢失与偏移。这些改进均推动了多目标跟踪算法的发展。

尽管现有算法在对航拍小目标的检测与跟踪上取得了一定的进展,但对小目标的漏检、误检率依然较高;此外,当前跟踪算法对目标的跟踪丢失和偏移次数仍较多,这些问题均会对系统造成干扰,影响多目标跟踪效果。针对上述不足,为进一步提高航拍多目标跟踪的精度,本文提出了基于改进 YOLOv8[19] 与 ByteTrack 的航拍多目标跟踪算法(YBTrack)。该算法利用 YOLOv8 在速度与精度平衡方面的优势,且与 ByteTrack 的高效关联策略相结合,实现了高效的多目标跟踪。具体而言,本文首先通过改进 YOLOv8 提供高质量的检测结果,然后与 ByteTrack 的鲁棒性关联策略形成互补,关联低置信度检测框,弥补检测器在航拍复杂场景中的不足,最后优化 ByteTrack 算法,提升跟踪精度并且减少目标丢失和 ID 切换,进而加强整体跟踪性能。本文主要工作如下:

1)设计空间-深度卷积模块(SPD-DPConv),减少多尺度特征图融合过程中信息提取不足导致的漏检、误检,保留小目标细节,提高模型检测精度。

2)设计轻量加速空间-通道注意力模块(LASA),减少检测模型计算复杂度,并作为检测头前的特征细化模块,增强目标特征信息提取能力。

3)引入空间-外观相似度矩阵(ASM),加强 ByteTrack 跟踪模型区分相似目标的性能,减少跟踪目标的偏移与丢失。

计算机工程 www.ecice.com

4)提出目标校正函数,改善目标跟踪过程中卡尔曼滤波器产生的误差积累,进一步减少跟踪目标偏移与丢失,提升多目标跟踪精度。

1 本文方法

1.1 YBTrack

YBTrack 包含检测分支和跟踪分支。对于检测分支 MSA-YOLO(Multi-Scale Attention YOLO),在 YOLOv8 的基础上结合设计的 SPD-DPConv 模块与 LASA 模块,增强检测模型对目标特征信息的提取能力,提升检测器的检测精度,输出更优的检测参数到跟踪环节;对于跟踪分支 ByteTrack,设计 ASM,增强算法对相似目标的区分能力,引入目标校正函数降低卡尔曼滤波器产生的误差积累,提升卡尔曼滤波器的目标预测-校正性能。随后利用匈牙利匹配算法匹配检测框与现有目标轨迹,并利用不同帧的轨迹关联信息,对跟踪轨迹进行更新,得到当前帧跟踪结果。整体模型能够增强算法对检测框的处理能力,同时也能提高跟踪器的跟踪精度。本文 YBTrack 算法结构与流程如图 1 所示(彩色效果见《计算机工程》官网 HTML 版,下同)。

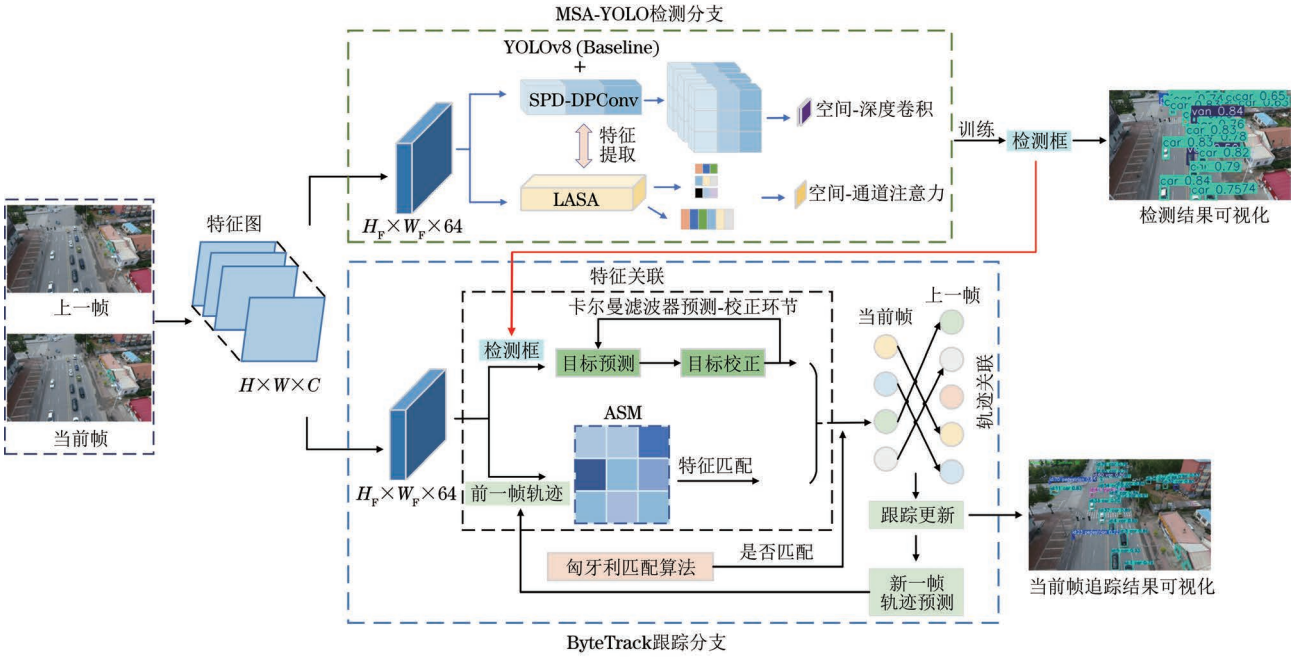


图 1 YBTrack 算法结构

Fig. 1 YBTrack algorithm structure

1.2 MSA-YOLO

改进 YOLOv8 后的 MSA-YOLO 模型结构如图 2 所示。引入 SPD-DPConv 和 LASA 模块,提升模型对目标特征信息的提取能力,克服其在处理复杂场景、多尺度航拍目标及遮挡目标时的局限性,为后续跟踪输入更精准的检测信息,提升系统在动态环境中的跟踪性能。

1.2.1 空间-深度卷积模块

在卷积神经网络中,空间特征通常反映目标在二维空间中的位置信息,而通道特征则更多与目标的类别、形状、颜色等高层语义信息相关。然而传统卷积网络对这两类特征的处理较为独立,难以充分融合这两类信息。为加强特征图的空间-通道融合与信息提取能力,引入空间-深度模块(SPD)[20],设计深度密集卷积层(DPConv),从而构建新的 SPD-DPConv 模块。SPD-DPConv 可将空间特征与通道特征分别进行提取,再进行融合,减少因尺度差异导致的漏检、误检,提高模型在多尺度航拍目标上的检测性能。SPD-DPConv 模块如图 3 所示。

对于大小为 $S\times S\times C_1$ 的中间特征图,设下采样率为 b,得到 $b\times b$ 个形状为 $\frac{S}{b}\times\frac{S}{b}\times C_1$ 的子特征图。此模块主要由两个部分组成:

1)SPD 层,主要作用是将空间维度的信息转换到深度维度。如图 3 所示,设下采样率为 2,便得到 4 个形状为 $\frac{S}{2}\times\frac{S}{2}\times C_1$ 的子特征图,通道数由原来的 C_1 增加为 $4C_1$,从而保留更多细节信息,提升模型捕捉目标局部特征和位置信息的能力,增强模型对多尺度变化目标的识别能力。

2)DPConv 层,用于保留空间细节,减少目标检测中的信息丢失,降低因尺寸变化带来的影响,包含

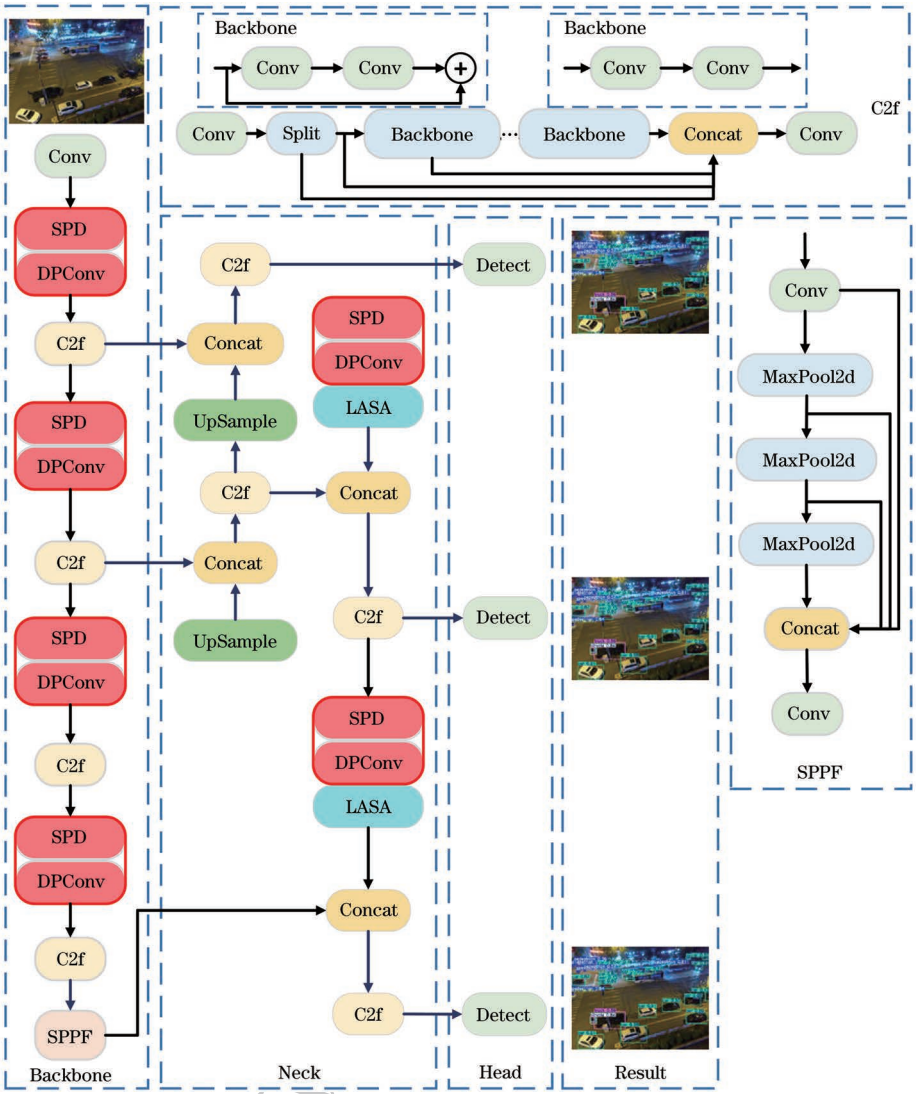


图 2 MSA-YOLO 网络结构

Fig. 2 MSA-YOLO network structure

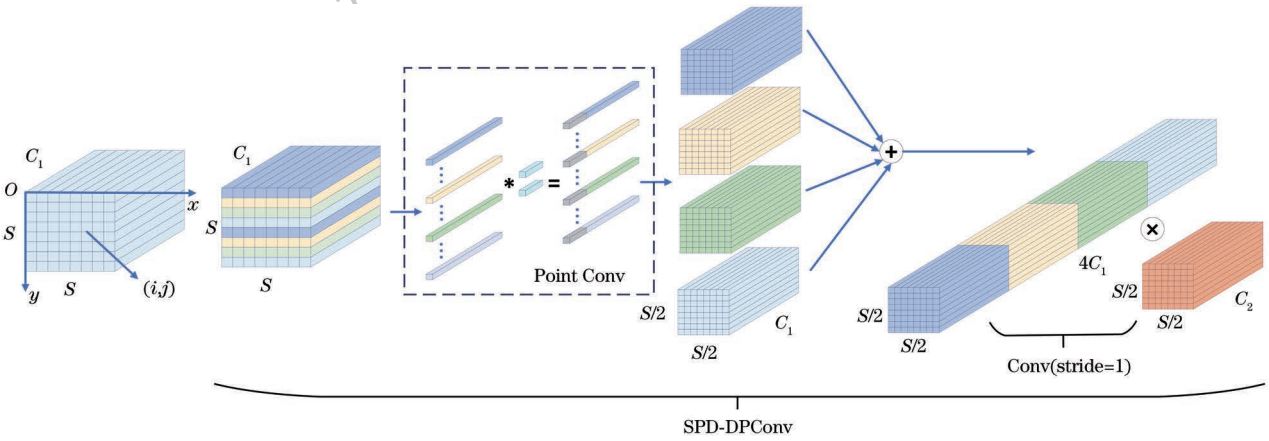


图 3 空间-深度卷积模块

Fig. 3 Space-depth convolution module

一个点卷积和一个非步长卷积(步长为 1),通过 1×1 的点卷积在通道维度上实现信息交互与整合,以在不改变特征图尺寸的情况下高效融合通道信息,点卷积计算公式如式(1)所示:

$$Y_{i,j,c'}=\sum_{c=1}^{4C}X_{i,j,c}\cdot W_{c,c'} \tag{1}$$

然后使用非步长卷积对输出特征进一步处理,将不同通道的信息整合在一起,保持特征图的尺寸

不变，计算公式如式(2)所示：

$$Y_{i,j,c'}=\sum_{m=1}^{k}\sum_{n=1}^{k}\sum_{c=1}^{4C}X_{(i+m-1),(j+n-1),c}\cdot \boldsymbol{W}_{m,n,c,c'} \tag{2}$$

式中：X 和 Y 分别为输入、输出特征图；C 和 C' 为其对应的通道数；(i,j) 表示特征图的空间坐标；k 为卷积核的大小；$\boldsymbol{W}_{c,c'}$ 为点卷积权重矩阵；$\boldsymbol{W}_{m,n,c,c'}$ 为卷积核权重。

1.2.2 轻量加速空间-通道注意力模块

针对卷积计算带来的大规模参数问题，为使模型保持较低的计算复杂度，并进一步加强目标特征信息的提取能力，设计了 LASA 模块。如图 4 所示，此模块结合空间信息增强模块(SIEM)与通道细化模块(CRAM)实现对特征的聚焦。

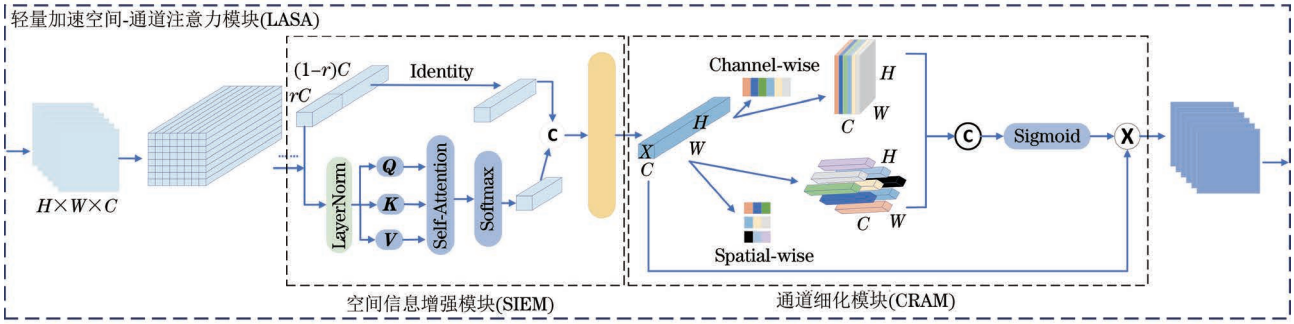


图 4 轻量加速空间-通道注意力模块

Fig.4 Lightweight acceleration space-channel attention module

1)SIEM：基于单头自注意力机制(SHSA)[21]，通过只生成一个注意力头 Attention($\boldsymbol{Q}$,$\boldsymbol{K}$,$\boldsymbol{V}$)，应用在输入通道的一部分($C_P=rC$)，剩余通道[$C_P=(1-r)C$]保持不变，且采用输入特征图的初始通道作为代表，避免对所有通道重复计算，减少内存开销和计算量；同时保留长距离依赖的表达能力，识别更大感受野下的空间上下文，提升输入特征图的空间信息表达能力。SHSA 层描述如式(3)～式(6)所示：

$$\mathrm{SHSA}(X)=\mathrm{Concat}(\widetilde{X}_{\mathrm{att}},X_{\mathrm{res}})W^{\mathrm{O}} \tag{3}$$

$$\widetilde{X}_{\mathrm{att}}=\mathrm{Attention}(\widetilde{X}_{\mathrm{att}},X_{\mathrm{res}})W^{\mathrm{O}} \tag{4}$$

$$\mathrm{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{Softmax}(\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}/\sqrt{d_{\mathrm{qk}}})\boldsymbol{V} \tag{5}$$

$$\widetilde{X}_{\mathrm{att}},X_{\mathrm{res}}=\mathrm{Split}(X,[C_P,C-C_P]) \tag{6}$$

式中：W^{O} 是投影权重；查询(Query，$\boldsymbol{Q}$)、键(Key，$\boldsymbol{K}$)、值(Value，$\boldsymbol{V}$)都来自输入数据 X 进行不同的特征变换；d_{qk} 是查询 $\boldsymbol{Q}$ 和键 $\boldsymbol{K}$ 的维度(默认为 16)，而 Concat(•)是连接操作。为保持内存访问的一致性，采用初始的 C_P 通道作为整个特征图的代表。

2)CRAM：采用无参数注意力模型(SimAM)[22]，无需额外的学习参数，仅利用通道内已有的信息进行特征强化，消除通道间的冗余特征，突出重要通道，使目标特征更明确，保持低计算开销，精准捕捉细微的特征差异，增强目标与背景的区分能力。

1.3 ByteTrack 算法优化

为减少跟踪目标偏移、丢失的情况，进一步提升跟踪准确性，对多目标跟踪算法 ByteTrack[15] 进行优化，具体实现流程如图 5 所示。

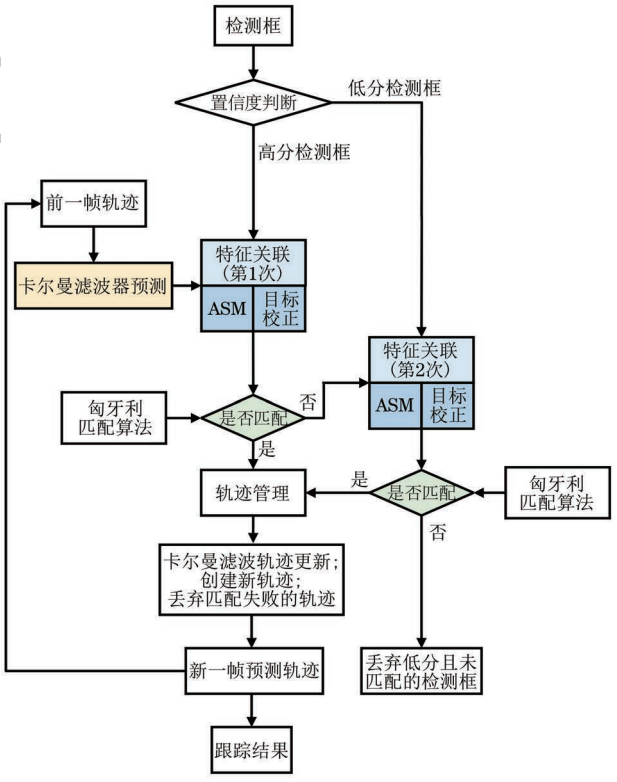


图 5 改进的 ByteTrack 算法流程

Fig.5 The improved ByteTrack algorithm flow

得到候选检测框后，将检测信息输入跟踪环节，对高分检测框和低分检测框筛选后进入特征关联阶段。首先，卡尔曼滤波器利用前一帧的信息对高分检测框进行轨迹预测，生成预测边界框。然后，使用匈牙利算法对预测边界框和检测框进行匹配，确定

目标的实际位置。如果匹配失败,则进行第 2 次特征关联,并再次进行轨迹匹配;若仍然无法匹配,则选择丢弃该目标。最后,卡尔曼滤波器根据匹配结果更新轨迹,并丢弃匹配失败的轨迹,得到目标的实际位置信息。

1.3.1 空间-外观相似度矩阵

为增强模型对相似目标的处理能力,减少目标的偏移、丢失,根据当前帧与上一帧的信息,以外观和空间特征的关联矩阵为基础进行相似性度量[17],构建 ASM,精确计算目标之间的相似性,提升目标匹配的准确性,算法实现流程如下:

1)输入当前帧与上一帧的检测框集合 $D_t=\{d_1^t,d_2^t,\cdots,d_m^t\}$、外观特征向量$\{\boldsymbol{f}_1^t,\boldsymbol{f}_2^t,\cdots,\boldsymbol{f}_m^t\}$、空间特征信息$\{s_1^t,s_2^t,\cdots,s_m^t\}$。

2)使用边界框位置、速度、加速度、目标空间信息,计算空间相似度矩阵 $\boldsymbol{A}_{\text{space}}$。

3)由跟踪器提取的目标视觉特征向量 $\boldsymbol{f}_i=[f_{i,1},f_{i,2},\cdots,f_{i,n}]$(其中 n 为特征向量维度),计算外观相似度矩阵 $\boldsymbol{A}_{\text{appearance}}$。

4)将空间相似度矩阵和外观相似度矩阵按权重 α 融合,得到 ASM,通过最大化 ASM,找到最优匹配,提升目标 ID 连续性。

5)利用 ASM 匹配结果,结合目标校正函数,校正卡尔曼滤波器的预测状态,减少目标偏移、丢失。

输入当前帧检测框集合 $D_t=\{d_1^t,d_2^t,\cdots,d_m^t\}$、外观特征向量$\{f_1^t,f_2^t,\cdots,f_m^t\}$、空间特征信息$\{s_1^t,s_2^t,\cdots,s_m^t\}$,然后与跟踪器提取的视觉特征构建关联矩阵,最后将新检测的目标与已有轨迹匹配,确保目标 ID 的连续性。

对目标 i 和目标 j,空间相似度矩阵如式(7)所示:

$$\boldsymbol{A}_{\text{space}}(i,j)=\exp\left(-\frac{d(X_i^t,X_j^{t-1})}{\sigma_{\text{s}}}\right) \tag{7}$$

式中:σ_{s} 为控制距离影响的超参数,用于平滑空间距离的影响;$d(X_i^t,X_j^{t-1})$表示当前帧中目标 i 和上一帧中目标 j 的空间距离。$d(X_i^t,X_j^{t-1})$公式如式(8)所示:

$$d(X_i^t,X_j^{t-1})=\sqrt{(x_i-x_j)^2+(y_i-y_j)^2}+\beta|v_i-v_j|+\alpha|a_i-a_j| \tag{8}$$

式中:α 和 β 分别为控制速度和加速度的权重,用于调节空间距离度量对目标动态变化的适应能力。

对目标 i 和目标 j,外观相似度矩阵如式(9)所示:

$$\boldsymbol{A}_{\text{appearance}}(i,j)=\frac{\boldsymbol{f}_i^t\cdot\boldsymbol{f}_j^{t-1}}{\|\boldsymbol{f}_i^t\|\cdot\|\boldsymbol{f}_j^{t-1}\|} \tag{9}$$

式中:$\boldsymbol{f}_i^t$ 和 $\boldsymbol{f}_j^{t-1}$ 分别表示当前帧和上一帧中目标的外观特征向量。假设当前帧中有 M 个目标,上一帧中有 N 个目标,则外观相似度矩阵如式(10)所示:

$$\boldsymbol{A}_{\text{appearance}}=\begin{bmatrix}s_{1,1} & s_{1,2} & \cdots & s_{1,N}\\ s_{2,1} & s_{2,2} & \cdots & s_{2,N}\\ \vdots & \vdots & & \vdots\\ s_{M,1} & s_{M,2} & \cdots & s_{M,N}\end{bmatrix} \tag{10}$$

将空间相似度矩阵和外观相似度矩阵加权融合,得到最终的 ASM,如式(11)所示:

$$\boldsymbol{A}_{\text{ASM}}(i,j)=\varepsilon\boldsymbol{A}_{\text{space}}(i,j)+(1-\varepsilon)\boldsymbol{A}_{\text{appearance}}(i,j) \tag{11}$$

式中:ε 为用于平衡空间特征和外观特征在相似度融合中权重比例的参数,取值范围为[0,1]。

1.3.2 目标校正函数

利用 ASM 特征匹配相似度高的目标对当前帧的目标状态进行校正,保证卡尔曼滤波器的预测不偏离真实值。对于目标 i,通过其与上一帧中目标 j 之间的关联相似度对预测状态进行校正,目标校正函数如式(12)所示:

$$X_i^t=(1-\lambda)X_i^t+\lambda X_j^{t-1} \tag{12}$$

式中:λ 为校正系数,由 ASM 特征匹配相似度动态调整。匹配值越高,ASM 的值越大,校正强度随之增强。

ASM 通过匹配当前帧与上一帧中的目标,避免目标在遮挡或快速移动时被错误关联。完成基于 ASM 的初步匹配后,算法进入传统的卡尔曼滤波器观测更新阶段。此时,卡尔曼滤波器利用传感器的观测数据(目标检测框)与预测状态进行融合更新,确保目标的精确跟踪,减少跟踪丢失与偏移。更新公式如式(13)~式(15)所示:

$$K_t=\boldsymbol{P}_{t|t-1}\boldsymbol{H}^{\text{T}}(\boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^{\text{T}}+\boldsymbol{R})^{-1} \tag{13}$$

$$X_t=X_{t|t-1}+K_t(Z_t-\boldsymbol{H}X_{t|t-1}) \tag{14}$$

$$\boldsymbol{P}_t=(\boldsymbol{I}-K_t\boldsymbol{H})\boldsymbol{P}_{t|t-1} \tag{15}$$

式中:$\boldsymbol{R}$ 为观测噪声协方差矩阵;$\boldsymbol{P}_{t|t-1}$ 为预测状态协方差矩阵;$\boldsymbol{H}$ 为观测模型矩阵;Z_t 为当前帧的观测值;K_t 为卡尔曼增益;$\boldsymbol{I}$ 为单位矩阵。

2 实验与分析

2.1 数据集和参数配置

1)数据集。

本文首先在 VisDrone2019-DET 数据集上对目标检测算法进行训练与验证,其中训练集、验证集、测试

计算机工程 www.ecice.com

集分别包含 6 741、548、1 610 张图像；标注信息分布如图 6 所示，共包含 10 个类别的标签；为减少训练时间和资源消耗，提升实验效率，分别选取 1 725、548、1 610 张图像作为训练集、验证集和测试集。

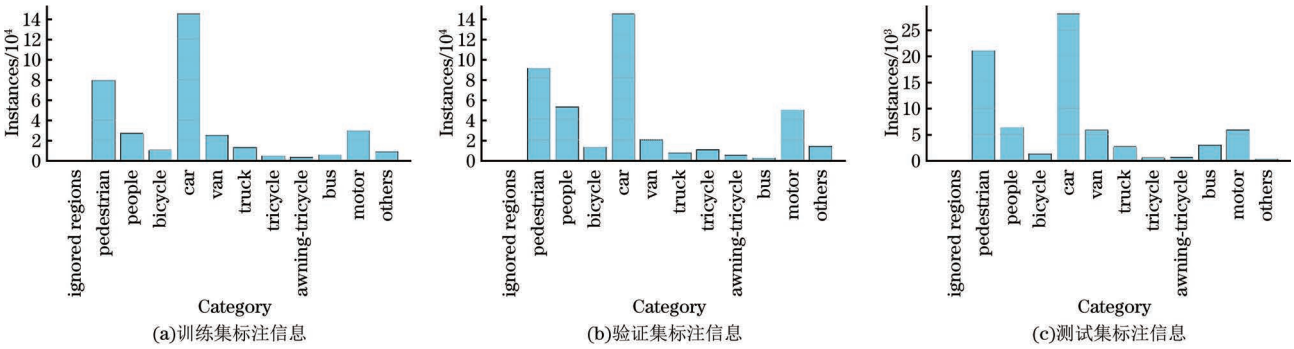


图 6 VisDrone2019-DET 标注信息的分布

Fig. 6 Distribution of annotation information in VisDrone2019-DET

然后在 VisDrone2019-MOT 和 MOT17 数据集上对跟踪性能进行验证。VisDrone2019-MOT 数据集由无人机在实际环境中拍摄而成，涵盖了晴天、阴天、夜晚等多种场景。实验部分选取了验证集中的 10 个视频序列（包含 2 620 张图像）以及测试集中的 15 个视频序列（包含 3 658 张图像）进行评估，VisDrone2019-MOT 测试集图片示例如图 7 所示。



图 7 VisDrone2019-MOT 视频序列示例

Fig. 7 VisDrone2019-MOT video sequence example

MOT17 数据集为多目标跟踪的权威数据集，广泛用于学术研究和算法对比测试，MOT17 数据集视频序列示例如图 8 所示。



图 8 MOT17 视频序列示例

Fig. 8 MOT17 video sequence example

2）实验环境配置。

在 Intel® i5-12600KF @ 3.70 GHz 处理器、16 GB 内存、NVDIA RTX4060 8 GB 显存的 GPU，Windows 10 系统、PyTorch 2.0 及 CUDA 11.3 的环境下进行训练、验证和测试。实验具体参数如表 1 所示。

表 1 实验参数

Table 1 Experimental parameters

参数名称	参数信息
训练轮次	300
批次大小	8
图像大小/像素	640×640
初始学习率	0.001
动量	0.05

2.2 评价指标

为有效地评估模型性能，采用常见评价指标对检测模型和多目标跟踪模型分别进行评估，具体评价指标如表 2 所示，其中，（↑）表示指标越高效果越好，（↓）表示指标数值越高效果越差；N_{TP}、N_{FP} 和 N_{FN} 分别表示模型预测正例、假正例、假负例数量；P_{AP} 中的 P 和 R 分别表示目标检测精度和召回率；N_{GT} 为图像中标注的目标个数；c 表示所有匹配的目标对数量；d 表示目标预测框与真值框之间的距离；N_{IDTP}、N_{IDFP} 和 N_{IDFN} 分别代表正确分配的目标 ID 数量、错误分配的目标 ID 数量和错过的目标 ID 数量[1]。

2.3 实验结果与分析

2.3.1 模块消融实验结果与分析

1）SPD-DPConv 模块有效性验证。

为验证 SPD-DPConv 模块的有效性，开展与现有优秀卷积模块SCConv[23]、ODConv[24]、PCConv[25]

表 2 评价指标

Table 2 Evaluation index

评价指标	定义	计算公式
$P(\uparrow)$	精确率	$P=\frac{N_{TP}}{N_{TP}+N_{FP}}$
$R(\uparrow)$	召回率	$R=\frac{N_{TP}}{N_{TP}+N_{FN}}$
$P_{AP}(\uparrow)$	关键点相似度为[0.50, 0.55, …, 0.95]时 10 个位置的平均精度(AP)	$P_{AP}\int_0^1 P(R)\mathrm{d}R$
$P_{mAP}(\uparrow)$	平均精度均值(mAP),下文中 mAP@0.5 指在置信度阈值为 0.5 时的 mAP 值	$P_{mAP}=\frac{1}{n}\sum_{i=1}^{n}P_{APi}$
$N_{Params}(\downarrow)$	参数量	—
$R_{MOTA}(\uparrow)$	多目标跟踪准确率(MOTA)	$R_{MOTA}=1-\frac{\sum_i N_{FNi}+N_{FPi}+N_{IDswi}}{\sum_i N_{GTi}}$
$R_{MOTP}(\downarrow)$	多目标跟踪精确率(MOTP)	$R_{MOTP}=\frac{\sum_t^i d_{t,i}}{\sum_t c_t}$
$V_{IDF1}(\uparrow)$	ID 调和均值(IDF1)	$V_{IDF1}=\frac{2N_{IDTP}}{2N_{IDTP}+N_{IDFP}+N_{IDFN}}$
$N_{IDsw}(\downarrow)$	目标 ID 切换次数	—

的消融实验,实验结果如表 3 所示,其中,√表示使用该方法,最优数据加粗标示,下同。加入 SPD-DPConv 卷积后,检测精度明显提升,证明此卷积模块的有效性。此外,与其他现有优秀卷积的实验对比中,SPD-DPConv 卷积检测精度更高,进一步说明本文卷积更加适用于处理航拍复杂场景。

表 3 SPD-DPConv 卷积有效性验证

Table 3 SPD-DPConv convolution validity verification %

Model	SPD-DPConv	SCConv[23]	ODConv[24]	PCConv[25]	精确率	召回率	mAP@0.5
YOLOv8[19]					45.4	35.1	34.2
YOLOv8	√				**52.4**	**40.5**	**41.3**
YOLOv8		√			50.6	37.6	38.4
YOLOv8			√		43.6	33.2	32.4
YOLOv8				√	44.2	33.9	33.6

2)LASA 模块有效性验证。

为验证 LASA 模块及其子模块的有效性与鲁棒性,在 VisDrone2019-DET 数据集上对此模块进行了消融实验,且分别测试子模块和主模块对网络精确度提升的影响。实验结果如表 4 所示。

表 4 LASA 子模块对 YOLOv8 的影响

Table 4 Impact of LASA submodule on YOLOv8 %

Model	精确率	召回率	mAP@0.5
YOLOv8[19]	45.4	35.1	34.2
+SIEM	45.9	35.6	34.5
+CRAM	47.3	36.2	36.5
+LASA	**49.2**	**36.9**	**37.3**

从表 4 的实验结果可知,SIEM,CRAM,LASA 子模块分别加入 YOLOv8 模型中,对检测精度有不同的提升效果,在同时加入 SIEM 与 CRAM 模块后,检测精度出现大幅度的提升,证明加入 LASA 模块有助于处理复杂场景下的航拍目标。

2.3.2 检测器消融实验结果与分析

为验证本文检测模型 MSA-YOLO 的性能,在 VisDrone2019-DET 数据集上进行了消融实验。实验选择 YOLOv5、YOLOv8 和最新的 YOLO11 作为基线模型(Baseline),旨在全面评估本文模块的通用性和有效性,并且与其他系列 YOLO 模型进行实验对比。实验结果如表 5 所示。

计算机工程 www.ccice06.com

表 5 不同组件对 YOLO 系列模型的影响

Table 5 The impact of different components on YOLO series models

Baseline	SPD-DPConv	LASA	Detector	精确率/%	召回率/%	mAP@0.5/%	参数量/10^6
YOLOv5[19]			YOLOv5	39.5	28.7	27.7	30.7
YOLOv5	√		YOLOv5A	41.5	32.1	31.1	65.7
YOLOv5	√	√	YOLOv5B	45.1	34.4	33.3	56.5
YOLOv6			YOLOv6	36.0	28.6	25.9	42.3
YOLOX[19]			YOLOX	44.6	34.9	34.1	32.1
YOLOv9[26]			YOLOv9	43.4	34.3	33.4	39.4
YOLOv10[27]			YOLOv10	44.5	33.9	33.9	37.9
YOLO11[28]			YOLO11	35.7	30.7	28.4	28.4
YOLO11	√		YOLO11A	48.7	38.3	38.7	58.7
YOLO11	√	√	YOLO11B	53.7	41.9	41.5	41.5
YOLOv8[19]			YOLOv8	45.4	35.1	34.2	34.2
YOLOv8	√		YOLOv8A	52.8	40.5	41.3	55.4
YOLOv8	√	√	MSA-YOLO(Ours)	**55.5**	**42.3**	**43.6**	**36.6**

从表 5 可知，不同的 YOLO 模型，在替换 SPD-DPConv 卷积后精确率、召回率、mAP@0.5 相较于原始 YOLO 模型均有明显提升。具体而言，这是因为 YOLOv5 的 Focus 层与快速空间金字塔池化(SPPF)结构对高分辨率特征保留不足，限制了 SPD-DPConv 在小目标检测中的潜力，同时 YOLO11 的极简结构使其对新增加模块的敏感性较高，性能增益受限；相比之下，YOLOv8 的骨干网络采用的跨阶段连接和自适应特征融合机制，可更好地利用 SPD-DPConv 的多尺度特征提取能力，因此在替换卷积后，精确率、召回率、mAP@0.5 提升效果更优，分别提升了 7.4、5.4、7.1 百分点。

其次，对比所有系列的 YOLO 模型，替换卷积且加入 LASA 模块的 YOLOv8 模型呈现更好的检测效果，精确率、召回率、mAP@0.5 分别提升了 10.1、7.2、9.4 百分点，这主要得益于 LASA 模块的自适应参数分配优化特征选择与 YOLOv8 的轻量化设计高度适配，通过减少冗余通道计算，显著降低了计算开销，与只添加 SPD-DPConv 模块相比参数量减少了 1.88×10^7，表明 LASA 模块的引入不仅大幅减少了模型的整体计算参数量，还进一步提升了模型的检测精度，在 YOLOv5 与 YOLO11 中引入 SPD-DPConv 与 LASA 模块，MSA-YOLO 的精确率、召回率、mAP@0.5 分别至少提升了 1.8、0.4、2.1 百分点。

通过对比上文各模块的消融实验可知：这两个模块相辅相成，证明了本文所提出方法的有效性与优越性。

2.3.3 检测结果可视化

为了展示算法的有效性，进行了检测算法结果可视化，全方位展示改进后的检测模型 MSA-YOLO 的有效性。从图 9 中可以看出，MSA-YOLO 可以得到更高置信度的检测框，有效地防止了误检和漏检，提升了模型抗遮挡的能力，并且大幅度提升了检测精度。

2.3.4 目标跟踪算法消融实验

为验证 ASM 与目标校正函数在多目标跟踪中对跟踪性能的影响，在 VisDrone2019-MOT 视频数据集的 15 个测试序列上开展消融实验，测试序列涵盖了晴天、阴天、夜晚等不同的航拍复杂场景，确保了实验结果的代表性和多样性，在 15 个测试序列上的实验结果平均值如表 6 所示。

从表 6 可知，针对不同的跟踪算法，引入 ASM 和目标校正函数后，MOTA 的平均提升 0.9～2.6 百分点，IDF1 平均提升 0.9～2.3 百分点，同时 ID 切换次数最大减少了 25 次，显著降低了目标丢失和切换的频率。在 MSA-YOLO 检测器下，与 YOLOX[19]+ByteTrack[15]的组合相比，结合 ASM 和目标校正函数后，MOTA 和 IDF1 分别提高了 11.2 和 8.9 百分点，ID 切换次数减少了 49 次。这表明，本文提出的跟踪方法在复杂场景下，即使存在目标遮挡和光照变化等挑战，依然能有效提升跟踪精度，并显著降低 ID 切换的频率。

此外，当使用 MSA-YOLO 作为检测器时，在不同的跟踪算法中引入 ASM 与目标校正函数均能提升跟踪性能；分别使用 DeepSORT[9]、OC-SORT[11]、FairMOT[12]与 ByteTrack[15]分别作为跟踪器时，ByteTrack 跟踪效果最为显著，MOTA 提升了 2.6 百分点，IDF1 提升了 1.8 百分点，ID 切换次数减少了 21 次。这进一步表明，在航拍复杂场景

计算机工程 www.ecicao6.com

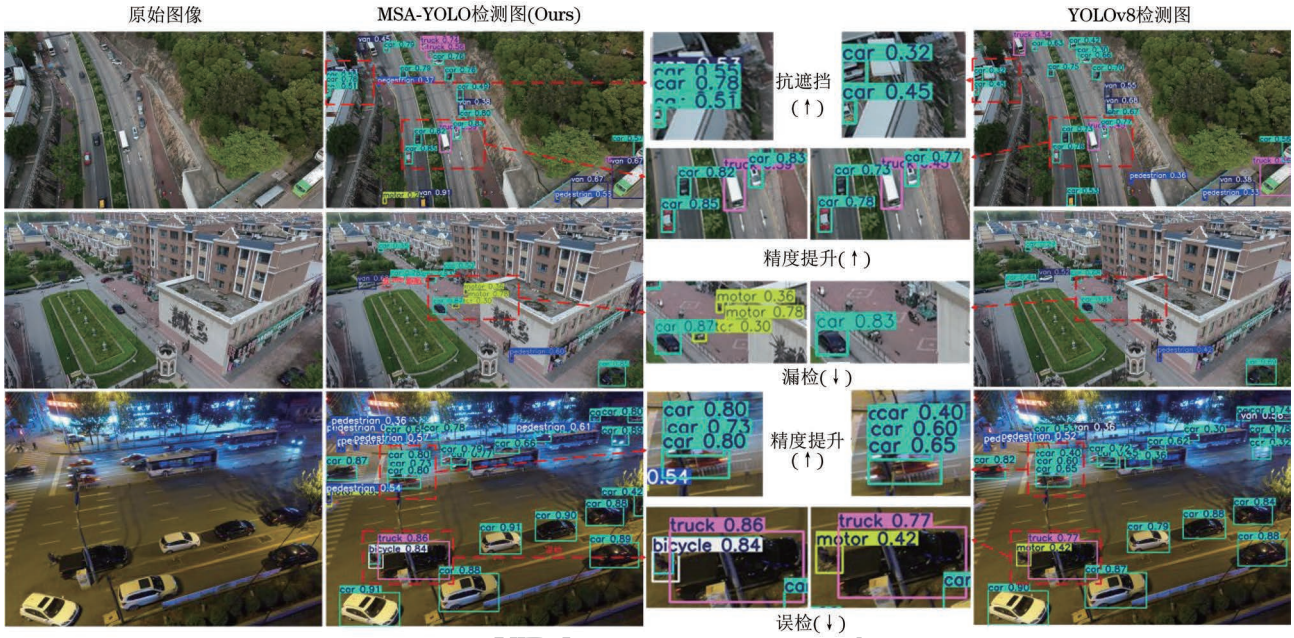


图 9 检测结果可视化

Fig. 9 Visualization of detection results

表 6 组件对多目标跟踪模型的影响

Table 6 Influence of components on multi-target tracking model

Detector+Tracker	ASM	目标校正函数	MOTA/%	MOTP/%	IDF1/%	ID 切换次数/次
YOLOX[19]+ByteTrack[15]			53.6	68.2	65.5	196
YOLOX+ByteTrack	√	√	55.2	67.4	67.8	172
YOLOv8[19]+ByteTrack			54.3	68.5	67.6	190
YOLOv8+ByteTrack	√	√	55.8	66.2	68.5	176
YOLO11[28]+ByteTrack			54.5	68.1	67.1	228
YOLO11+ByteTrack	√	√	55.4	67.5	68.3	203
MSA-YOLO+DeepSORT[9]			53.2	69.9	65.2	211
MSA-YOLO+DeepSORT	√	√	54.1	69.1	66.6	186
MSA-YOLO+OC-SORT[11]			61.7	57.2	72.7	177
MSA-YOLO+OC-SORT	√	√	62.9	55.9	73.5	152
MSA-YOLO+FairMOT[12]			53.7	67.4	66.9	214
MSA-YOLO+FairMOT	√	√	54.6	65.8	67.3	194
MSA-YOLO+ByteTrack			62.2	56.2	72.6	168
MSA-YOLO+ByteTrack (Ours)	√	√	**64.8**	**53.3**	**74.4**	**147**

下，ASM 与目标校正函数更适合应用于 ByteTrack 算法，能够取得更佳的跟踪效果，不仅提升了跟踪精度，还有效降低了目标丢失和切换的频率。

2.3.5 基于新算法的跟踪可视化

为展示 YBTrack 在航拍复杂场景下的跟踪性能，在 VisDrone2019-MOT 视频序列中进行了可视化对比实验，与 ByteTrack[15] 与 OC-SORT[11] 算法进行对比，可视化结果如图 10 所示。

在图 10 中，选择了 3 个具有代表性的场景：阴天，晴天，夜间，在每一个视频场景中分别选取相同的两帧，在 YBTrack、ByteTrack[15] 和 OC-SORT[11] 上开展实验。可视化结果展示了算法在不同的 3 个复杂场景中的表现，可以看出，本文 YBTrack 几乎跟踪到了所有的目标，并保持跟踪 ID 不变，且在有强光干扰的晴天场景和黑暗的夜间场景，YBTrack 均可以跟踪更多的目标；而 ByteTrack 与 OC-SORT 却出现了较多的跟踪丢失、偏移的情况，且跟踪精度低。这进一步表明本文多目标跟踪算法在航拍复杂场景下的鲁棒性。

2.3.6 与其他跟踪方法的对比

为验证 YBTrack 航拍多目标跟踪性能，与主流的多目标跟踪算法开展了对比实验。为了确保实验

计算机工程 www.ecice06.com



图 10 VisDrone2019-MOT 视频序列跟踪结果可视化

Fig. 10 Visualization of VisDrone2019-MOT video sequence tracking results

公平性，所有实验都使用 YOLOv8 作为检测器，且分别在 VisDrone2019-MOT 与 MOT17 视频数据集的 15 个序列上进行实验，得到各项指标的平均值，实验结果如表 7、表 8 所示。

表 7 VisDrone2019-MOT 数据集上不同跟踪方法对比

Table 7 Comparison of different tracking methods on VisDrone2019-MOT dataset %

Model	MOTA	MOTP	IDF1
SORT[10]	39.3	79.6	45.2
DeepSORT[9]	52.7	69.5	62.8
OC-SORT[11]	57.3	60.2	67.2
FairMOT[12]	52.5	71.8	62.7
TransTrack[14]	53.9	70.2	64.4
AIPT[29]	58.1	57.3	70.5
MOTRv2[30]	57.9	59.4	69.8
ByteTrack[15]	54.3	68.5	67.6
YBTrack (Ours)	**64.8**	**53.3**	**74.4**

表 8 MOT17 数据集上不同跟踪方法对比

Table 8 Comparison of different tracking methods on MOT17 dataset %

Model	MOTA	MOTP	IDF1
SOTR[10]	67.6	71.6	65.2
DeepSORT[9]	74.6	67.7	75.8
OC-SORT[11]	79.8	67.1	77.4
FairMOT[12]	75.6	69.5	76.8
TransTrack[14]	78.5	64.2	75.3
AIPT[29]	85.6	64.2	81.6
MOTRv2[30]	83.7	65.3	79.2
ByteTrack[15]	80.3	66.8	76.9
YBTrack (Ours)	**88.6**	**59.6**	**84.3**

从表 7 与表 8 的实验结果中可知，YBTrack 在所有跟踪指标上均表现更优。在 VisDrone2019-DET 数据集上，MOTA、IDF1 分别比其他跟踪模型至少高 6.7、3.9 百分点；在 MOT17 数据集上，MOTA、IDF1 分别比其他跟踪模型至少高 3.0、2.7 百分点，说明本文跟踪方法在处理航拍复杂动态目标时精度与鲁棒性均更优。

3 结束语

本文提出了一种基于改进 YOLOv8 与 ByteTrack 的航拍多目标跟踪方法，以提升在航拍复杂场景下的检测与多目标跟踪效果。该方法通过优化 YOLOv8 网络，用空间-深度卷积模块替换原有卷积，捕捉更细致的目标特征图细节；设计轻量化空间-通道注意力模块，增强特征图的空间依赖捕捉能力，降低计算复杂度，得到了高精度的检测器 MSA-YOLO；优化了 ByteTrack 算法，设计空间-外观相似度矩阵，加强对相似目标的处理，提出目标校正函数，抑制卡尔曼滤波器的误差积累，提升了跟踪效果，减少了跟踪目标偏移、丢失现象。最后，在 VisDrone2019-DET 数据集上进行了检测算法的消融与对比实验，在 VisDrone2019-MOT、MOT17 视频数据集上验证了跟踪算法，并与多种先进多目标跟踪算法进行对比。实验结果表明，本文 YBTrack 显著提升了检测性能和多目标跟踪能力，在航拍复杂场景下对目标的捕捉效果尤为出色。后续研究将致力于提升小目标实时追踪的

效果，进一步降低参数计算量，以便更好地服务低空经济。

参考文献

[1] 苑玉彬，吴一全，赵朗月，等. 基于深度学习的无人机航拍视频多目标检测与跟踪研究进展[J]. 航空学报，2023，44(18)：028334.
YUAN Y B，WU Y Q，ZHAO L Y，et al. Research progress of UAV aerial video multi-object detection and tracking based on deep learning[J]. Acta Aeronautica et Astronautica Sinica，2023，44(18)：028334.(in Chinese)

[2] JIANG P Y，ERGU D，LIU F Y，et al. A review of Yolo algorithm developments[J]. Procedia Computer Science，2022，199：1066-1073.

[3] REN S Q，HE K M，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.

[4] 王国明，贾代旺. 基于 YOLOv8 的小目标检测模型的优化[J]. 计算机工程，2025，51(12)：294-303.
WANG G M，JIA D W. Optimization of small object detection model based on YOLOv8[J]. Computer Engineering，2025，51(12)：294-303.(in Chinese)

[5] LI H D，QU H C. DASSF：dynamic-attention scale-sequence fusion for aerial object detection[EB/OL]. [2024-12-01]. https://arxiv.org/abs/2406.12285.

[6] SU J，QIN Y C，JIA Z，et al. MPE-YOLO：enhanced small target detection in aerial imaging[J]. Scientific Reports，2024，14(1)：17799.

[7] 何植仟，曹立杰. UAVAI-YOLO：无人机航拍图像的小目标检测模型[J]. 智能科学与技术学报，2024，6(2)：262-271.
HE Z Q，CAO L J. UAVAI-YOLO：dense small target detection algorithm based on UAV aerial images[J]. Chinese Journal of Intelligent Science and Technology，2024，6(2)：262-271.(in Chinese)

[8] 蒋凌云，杨金龙. 检测优化的标签多伯努利视频多目标跟踪算法[J]. 计算机科学与探索，2023，17(6)：1343-1358.
JIANG L Y，YANG J L. Detection optimized labeled multi-bernoulli algorithm for visual multi-target tracking[J]. Journal of Frontiers of Computer Science and Technology，2023，17(6)：1343-1358.(in Chinese)

[9] WOJKE N，BEWLEY A，PAULUS D. Simple online and realtime tracking with a deep association metric[C]//Proceedings of the IEEE International Conference on Image Processing (ICIP). Beijing，China：IEEE Press，2018：3645-3649.

[10] BEWLEY A，GE Z Y，OTT L，et al. Simple online and realtime tracking[C]//Proceedings of the IEEE International Conference on Image Processing (ICIP). Phoenix，USA：IEEE Press，2016：3464-3468.

[11] CAO J K，PANG J M，WENG X S，et al. Observation-centric SORT：rethinking SORT for robust multi-object tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver，Canada：IEEE Press，2023：9686-9696.

[12] ZHANG Y F，WANG C Y，WANG X G，et al. FairMOT：on the fairness of detection and re-identification in multiple object tracking[J]. International Journal of Computer Vision，2021，129(11)：3069-3087.

[13] HAN K，WANG Y H，CHEN H T，et al. A survey on vision transformer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2023，45(1)：87-110.

[14] SUN P Z，CAO J K，JIANG Y，et al. TransTrack：multiple object tracking with transformer[EB/OL]. [2024-12-01]. https://arxiv.org/abs/2012.15460.

[15] ZHANG Y F，SUN P Z，JIANG Y，et al. ByteTrack：multi-object tracking by associating every detection box[C]// Proceedings of European Conference on Computer Vision. Berlin，Germany：Springer，2022：1-21.

[16] AHARON N，ORFAIG R，BOBROVSKY B Z. BoT-SORT：robust associations multi-pedestrian tracking[EB/OL]. [2024-12-01]. https://arxiv.org/abs/2206.14651.

[17] WANG Y H，HSIEH J W，CHEN P Y，et al. SMILEtrack：SiMIlarity LEarning for occlusion-aware multiple object tracking[C]//Proceedings of the AAAI Conference on Artificial Intelligence. [S.l.]：AAAI Press，2024：5740-5748.

[18] YOU L X，CHEN Y J，XIAO C，et al. Multi-object vehicle detection and tracking algorithm based on improved YOLOv8 and ByteTrack[J]. Electronics，2024，13(15)：3033.

[19] TERVEN J，CÓRDOVA-ESPARZA D M，ROMERO-GONZÁLEZ J A. A comprehensive review of YOLO architectures in computer vision：from YOLOv1 to YOLOv8 and YOLO-NAS[J]. Machine Learning and Knowledge Extraction，2023，5(4)：1680-1716.

[20] SUNKARA R，LUO T. No more strided convolutions or pooling：a new CNN building block for low-resolution images and small objects[C]// Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Berlin，Germany：Springer，2023：443-459.

[21] YUN S，RO Y. SHViT：single-head vision transformer with memory efficient macro design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle，USA：IEEE Press，2024：5756-5767.

[22] YANG L X，ZHANG R Y，LI L D，et al. SimAM：a simple，parameter-free attention module for convolutional neural networks[C]//Proceedings of International Conference on Machine Learning. [S.l.]：PMLR，2021：11863-11874.

[23] LI J F，WEN Y，HE L H. SCConv：spatial and channel reconstruction convolution for feature redundancy[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver，Canada：IEEE Press，2023：6153-6162.

[24] LI C，ZHOU A J，YAO A B. Omni-dimensional dynamic convolution[EB/OL]. [2024-12-01]. https://arxiv.org/abs/2209.07947.

[25] CHEN J R，KAO S H，HE H，et al. Run，don't walk：chasing higher FLOPS for faster neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver，Canada：IEEE Press，2023：12021-12031.

[26] WANG C Y，YEH I H，MARK LIAO H Y. YOLOv9：learning what you want to learn using programmable gradient information[C]// Proceedings of European Conference on Computer Vision. Berlin，Germany：Springer，2024：1-21.

[27] WANG A，CHEN H，LIU L H，et al. YOLOv10：real-time end-to-end object detection[EB/OL]. [2024-12-01]. https://arxiv.org/abs/2405.14458.

[28] HUANG J W，WANG K B，HOU Y，et al. LW-YOLO11：a lightweight arbitrary-oriented ship detection method based on improved YOLO11[J]. Sensors，2024，25(1)：65.

[29] ZHANG Y K，XIE H S，JIA Y H，et al. AIPT：Adaptive information perception for online multi-object tracking[J]. Knowledge-Based Systems，2024，285：111369.

[30] ZHANG Y A，WANG T C，ZHANG X Y. MOTRv2：bootstrapping end-to-end multi-object tracking by pretrained object detectors[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver，Canada：IEEE Press，2023：22056-22065.

文字编辑 金胡考
栏目编辑 宋 圆

计算机工程 www.ciceonline.com