

# 光电工程

## Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊  
Scopus CSCD

### 面向事件相机探测无人机的双视图融合检测方法

李森, 陈诺, 安玮, 李博扬, 凌强, 李卫星

#### 引用本文:

李森, 陈诺, 安玮, 等. 面向事件相机探测无人机的双视图融合检测方法[J]. 光电工程, 2024, 51(11): 240208.

Li M, Chen N, An W, et al. Dual view fusion detection method for event camera detection of unmanned aerial vehicles[J]. *Opto-Electron Eng*, 2024, 51(11): 240208.

<https://doi.org/10.12086/oe.2024.240208>

收稿日期: 2024-09-02; 修改日期: 2024-10-12; 录用日期: 2024-10-12

### 相关论文

#### 无人机视角下的道路损伤检测算法MAS-YOLOv8n

王晓燕, 王禧钰, 李杰, 梁文辉, 牟建宏, 毕楚然

光电工程 2024, 51(10): 240170 doi: 10.12086/oe.2024.240170

#### 自适应前景聚焦无人机航拍图像目标检测

肖振久, 吴正伟, 张杰浩, 曲海成

光电工程 2024, 51(9): 240149 doi: 10.12086/oe.2024.240149

#### 改进YOLOv7的无人机视角下复杂环境目标检测算法

张润梅, 肖钰霏, 贾振楠, 陈中, 陈梓华, 袁彬, 曹炜威, 宋妮妮

光电工程 2024, 51(5): 240051 doi: 10.12086/oe.2024.240051

#### 基于改进YOLOv5s的无人机图像实时目标检测

陈旭, 彭冬亮, 谷雨

光电工程 2022, 49(3): 210372 doi: 10.12086/oe.2022.210372

更多相关论文见光电期刊集群网站 



光电工程  
Opto-Electronic Engineering

<http://cn.ojournal.org/oe>



 OE\_Journal



Website

DOI: 10.12086/oe.2024.240208

CSTR: 32245.14.oe.2024.240208

# 面向事件相机探测无人机的双视图融合检测方法

李 淼, 陈 诺, 安 玮\*, 李博扬,  
凌 强, 李卫星

国防科技大学电子科学学院, 湖南 长沙 410073

**摘要:** 随着低空无人机的广泛应用, 实时检测此类低慢小目标对维护公共安全至关重要。传统相机以固定曝光时间成像获得图像帧, 在光照变化时难以自适应, 导致在强光照等场景下存在探测盲区。事件相机作为一种新型的神经形态传感器, 逐像素感知外部亮度变化差异, 在复杂光照条件下依然可以生成高频稀疏的事件数据。针对基于图像的检测方法难以适应事件相机稀疏不规则数据的难题, 本文将二维目标检测任务建模为三维时空点云中的语义分割任务, 提出了一种基于双视图融合的无人机目标分割模型。基于事件相机采集真实的无人机探测数据集, 实验结果表明, 所提方法在保证实时性的前提下具有最优的检测性能, 实现了无人机目标的稳定检测。

**关键词:** 事件相机; 低空无人机; 低慢小目标检测; 多视图融合

**中图分类号:** TP391

**文献标志码:** A

李淼, 陈诺, 安玮, 等. 面向事件相机探测无人机的双视图融合检测方法 [J]. 光电工程, 2024, 51(11): 240208

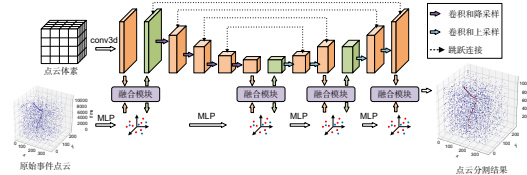
Li M, Chen N, An W, et al. Dual view fusion detection method for event camera detection of unmanned aerial vehicles[J]. *Opto-Electron Eng*, 2024, 51(11): 240208

## Dual view fusion detection method for event camera detection of unmanned aerial vehicles

Li Miao, Chen Nuo, An Wei\*, Li Boyang, Ling Qiang, Li Weixing

College of Electronic Science and Technology, National University of Defense Technology, Changsha, Hunan 410073, China

**Abstract:** With the widespread application of low-altitude drones, real-time detection of such slow and small targets is crucial for maintaining public safety. Traditional cameras capture image frames with a fixed exposure time, which makes it challenging to adapt to changes in lighting conditions, resulting in the detection of blind spots in intense light and other scenes. Event cameras, as a new type of neuromorphic sensor, sense differences in external brightness changes pixel by pixel. They can still generate high-frequency sparse event data under complex lighting conditions. In response to the difficulty of adapting image-based detection methods to sparse and irregular data from event cameras, this paper models the two-dimensional object detection task as a semantic segmentation task in a three-dimensional spatiotemporal point cloud and proposes a drone object segmentation model based on dual-view fusion. Based on the event camera collecting accurate drone detection datasets, the experimental results



收稿日期: 2024-09-02; 修回日期: 2024-10-12; 录用日期: 2024-10-12

基金项目: 国家自然科学基金资助项目 (62401591, 62401589); 国家自然科学基金创新研究群体项目 (61921001)

\*通信作者: 安玮, anwei@nudt.edu.cn。

版权所有©2024 中国科学院光电技术研究所

show that the proposed method has the optimal detection performance while ensuring real-time performance, achieving stable detection of drone targets.

**Keywords:** event camera; low altitude unmanned aerial vehicles; small target detection; multi view fusion

## 1 引言

低空空域的开放为无人机技术的广泛应用提供了广阔空间, 无人机逐渐被应用于各种行业<sup>[1]</sup>, 如远程监控<sup>[2]</sup>、精准农业<sup>[3]</sup>、地质勘探<sup>[4]</sup>、气象观测<sup>[5]</sup>、电力巡检<sup>[6]</sup>、应急救援<sup>[7]</sup>、快递物流<sup>[8]</sup>及影视拍摄<sup>[9]</sup>等。然而, 无人机数量的激增及其使用行为的难以规范, 不仅加剧了空域管理的难度, 还引发了诸多安全隐患和社会问题。无人机的不当使用, 如非法侵入禁飞区、侵犯个人隐私和干扰航空交通等, 已成为制约无人机产业健康发展的关键因素。因此, 探索并发展一种能够实时、准确地检测与跟踪低空空域内无人机活动的技术体系, 对于维护公共安全、保障个人隐私和优化空域资源配置具有重要意义<sup>[10-13]</sup>。

传统的检测技术, 如声波探测<sup>[14]</sup>、无线电探测<sup>[15]</sup>和雷达探测<sup>[16]</sup>等, 虽然具备较高的检测精度和范围, 但往往伴随着高昂的设备成本和复杂的配置要求, 限制了其在某些场景下的应用。相比之下, 基于机器视觉的检测技术以其成本低廉、易于配置的优势, 逐渐成为低空无人机检测领域的研究热点。

传统基于帧的相机在极端光照条件下(如过曝或黑夜场景)性能受限<sup>[17]</sup>。这种限制直接影响了无人机目标的有效捕捉与识别。为了克服这一缺点, 事件相机<sup>[18]</sup>作为一种新型的神经形态传感器, 凭借其独特的优势, 为低空无人机检测技术的发展提供了新思路。动态视觉传感器(DVS, dynamic vision sensor)及其地址事件表示(AER, address-event representation)机制, 为处理高速、高动态范围视觉信息提供了一种高效的方法。在AER中, 每个事件点都包含了关键信息: 事件发生的像素位置、时间戳以及事件极性, 这些信息对于捕捉和分析快速移动或变化的目标非常有效。

根据事件数据的不同表征方式, 基于事件相机目标检测可以分为基于稀疏表示和基于稠密表示的检测方法。基于稀疏表示的方法直接处理事件数据, 这些方法将事件数据视为非结构化数据, 包括通过脉冲神经网络<sup>[19-20]</sup>(SNN)直接处理原始事件和通过图神经网络<sup>[21-22]</sup>(GNN)对时空动态进行建模。然而, 这些方法

受到对专用硬件的依赖或性能不足的限制。为了解决这些问题, 基于稠密表示的方法将稀疏事件转换为类似图像的格式, 如基于极性对事件积累生成事件帧, 然后采用通用的目标检测器进行检测<sup>[23]</sup>。时间表面(TS)将每个像素值存储为该像素位置最近事件的时间戳<sup>[24]</sup>。然而, 由于时间维度的压缩, 这两种方法在目标检测任务中的性能有限。为了保留事件流的时间信息, 体素网格通过插值将事件堆叠到多个时间窗口中, 将事件流转换为4D张量<sup>[25]</sup>。EventPillars方法<sup>[26]</sup>和ASTMNet方法<sup>[27]</sup>中设计了可学习的编码器, 它们通过网络端到端的学习事件表示, 在基于事件的对象检测任务中具有较好的性能。最近, GET方法<sup>[28]</sup>采用了一种基于组的视觉变换器骨干网络, 并基于Transformer设计了目标检测网络, 在事件对象检测和分类任务中展现了最先进的性能。

目前, 基于事件相机的目标检测性能最优方法是将事件数据映射为类似图像帧的表示, 然后进行后续处理。然而这种方法并不适合于远距离无人机等小目标的检测。由于远距离的无人机目标尺寸很小, 在像平面中目标没有任何形态特征, 和噪声十分相似, 网络无法有效地学习到目标特征, 所以很难直接将目标与噪声进行区分。为了实现运动小目标的高效检测, 本文将事件数据视为三维空间中的点云。具体来说, 事件相机输出的事件流可以看做是一种由二维空间位置信息和时间信息组成的点云, 与激光雷达点云相比, 这种由事件点组成的特殊点云不含目标的深度信息, 但是时间信息组成了其第三维。本文将这种特殊的点云称为事件点云, 而基于事件的小目标检测可以视为从事件点云中分割具有线状拓扑结构的无人机轨迹。

## 2 事件相机介绍

### 2.1 事件相机与传统相机的区别

事件相机作为一种新型的动态视觉传感器, 异步捕获像素级亮度变化。事件相机的高时间分辨率和异步性显示出捕捉具有快速运动和高动态范围场景的巨大潜力, 其应用非常广泛<sup>[18, 29-30]</sup>, 如自动驾驶汽车、



机器人技术和安全监控。与以固定帧率捕获整个场景详细纹理的传统静态相机(如 RGB 和红外相机)不同, 事件相机以异步的方式对辐射强度变化的像素进行二元响应, 专注于捕捉移动对象的运动和结构。

原始事件数据在时空维度上表示为异步二进制离散点  $E(x, y, t, p)$ , 其中:  $x \in [1, W]$ ,  $y \in [1, H]$  表示空间维度上的位置,  $W, H$  表示事件数据的宽度和高度范围,  $t$  表示时间戳,  $p \in [0, 1]$  表示亮度变化的极性值(即  $p = 0$  表示像素亮度下降,  $p = 1$  表示增加)。事件相机与传统相机的输出差异如图 1 所示, 传统相机等时间间隔的输出场景图像, 每一张图像记录了某一时刻的场景状态; 而事件相机并无帧的概念, 它异步的输出稀疏事件点, 当场景有运动时, 它输出事件点, 当场景静止时, 事件相机无任何输出。相比于传统相机, 事件相机具有以下优势: 1) 低冗余: 只关注于运动的目标, 它只输出场景中运动部分的信息, 因此其数据量远低于传统相机; 2) 低功耗: 由于只对感知变化区域, 事件相机功耗极低, 通常为毫瓦级, 仅为传统相机的十分之一甚至是百分之一; 3) 高动态范围: 可达 120 dB, 这为事件相机在强光和低光场景下捕获目标信号奠定了基础。

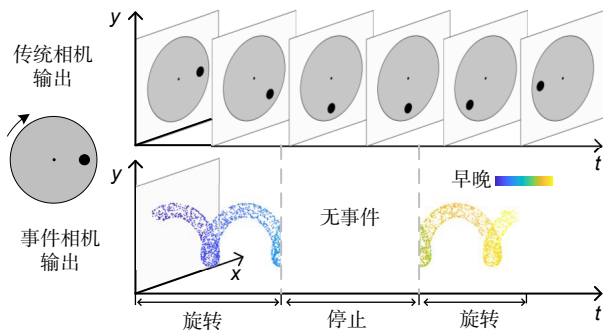


图 1 事件相机与传统相机输出比较

Fig. 1 Comparison of output between event camera and traditional camera

## 2.2 事件相机无人机检测数据集

事件相机由于具有信息冗余低、功耗低和动态范围高的优点, 特别适合用于低空安全监测, 采用事件相机可以实现长时间、全天候的低空无人机检测。

由于至今尚无公开的基于事件相机的无人机目标检测数据集, 因此本实验组自行采集与构建数据集进行模型的训练和评估。构建的低空无人机数据集 Ev-UAV 共包含 60 段事件相机记录的无人机飞行数据, 每段数据持续时间 10~20 s。该数据集包含各种飞行

姿态(如盘旋、倾斜和高速起降等)以及各种场景(湖边、城市和树林等)。数据集逐事件点的对目标和背景进行了标注。部分数据可视化见图 2。数据集按照 4 : 1 : 1 的比例划分为训练集、验证集和测试集。

## 3 基于事件点云分割的无人机小目标检测算法

由于远距离的无人机目标尺寸很小, 在像平面中目标没有任何形态特征, 和噪声十分相似, 所以很难直接将目标与噪声进行区分。本文将每个事件视为  $xyt$  空间中的一个三维点, 这样小目标在空间中的运动轨迹表现为一条连续的曲线, 而噪声由于是随机出现, 不具有运动连续性, 所以在三维空间中表现为一些杂乱无规则的点。所提出的方法基于二者在三维空间中的这种区别, 通过分割三维空间中的连续的曲线, 实现对目标的检测。所提方法的输入并不是转化为帧的事件图像, 而是针对一段时间内的所有事件, 基于目标的运动轨迹对目标和噪声进行区分。

图 3 展示了所提出的无人机小目标检测方法。它是一个双分支网络, 上半部分为体素特征提取分支, 下半部分为点特征提取分支, 两个分支之间在不同的阶段进行交互融合。体素分支采用类似 UNet<sup>[31]</sup> 的结构, 首先使用一个输入卷积从原始输入中提取上下文信息, 然后执行四个下采样阶段, 最后执行四个上采样阶段以恢复原始输入状态。点云分支包含四层 MLP, 来提取不同维度的逐点特征。体素特征与点特征的融合依次发生在输入卷积、第四次下采样、第二次上采样和最后一次上采样阶段之后。

体素特征提取分支能够捕获到点云数据的全局上下文信息和空间结构, 而点特征提取分支则通过多层感知机 (MLP) 提取每个点的局部和细节特征。两者的结合实现了全局与局部、结构与细节特征的互补, 有助于更全面地提取目标特征。此外, 多源特征的融合使得模型在面对复杂环境和小目标检测时具有更强的鲁棒性。即使部分特征受到噪声的影响, 其他特征仍能提供有效的信息支持。

### 3.1 点特征提取分支

为了从事件流中提取有效的点特征, 本文设计了一个由多层 MLP 组成的特征提取分支。该分支首先将事件流划分为具有固定时间间隔的时间窗口, 每个窗口内的事件点集合作为处理单元。给定事件流  $E$ , 首先将其划分为具有固定时间间隔的时间窗口



$E = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$ , 作为基本的处理单元。点特征提取分支的输入可以描述为:  $\varepsilon_i = \{e_k = (x_k, y_k, t_k, p_k) | k = 1, 2, \dots, N\}$ , 其中, 每个事件点表示为空间坐标  $(x, y)$  和时间坐标  $t$ , 以及事件的极性  $p$ ,  $N$  为点的数量。

点特征提取分支由四层 MLP 组成, 每个 MLP 层由一个线性拟合和一个非线性激活函数组成, 单层

的 MLP 可以描述为:  $X_{\text{out}} = f(WX_{\text{in}} + b)$ , 其中  $X_{\text{in}}$  为该层的输入,  $W$  和  $b$  为神经元之间的权值和偏置, 函数  $f$  为激活函数, 通常为 sigmoid 函数或 tanh 函数。

分支中的四层 MLP 的维度分别为: 32、256、128 和 32, 即可以得到四个不同维度的点特征  $X_i \in \mathbb{R}^{N \times d}$ ,  $d = \{32, 256, 128, 32\}$ 。

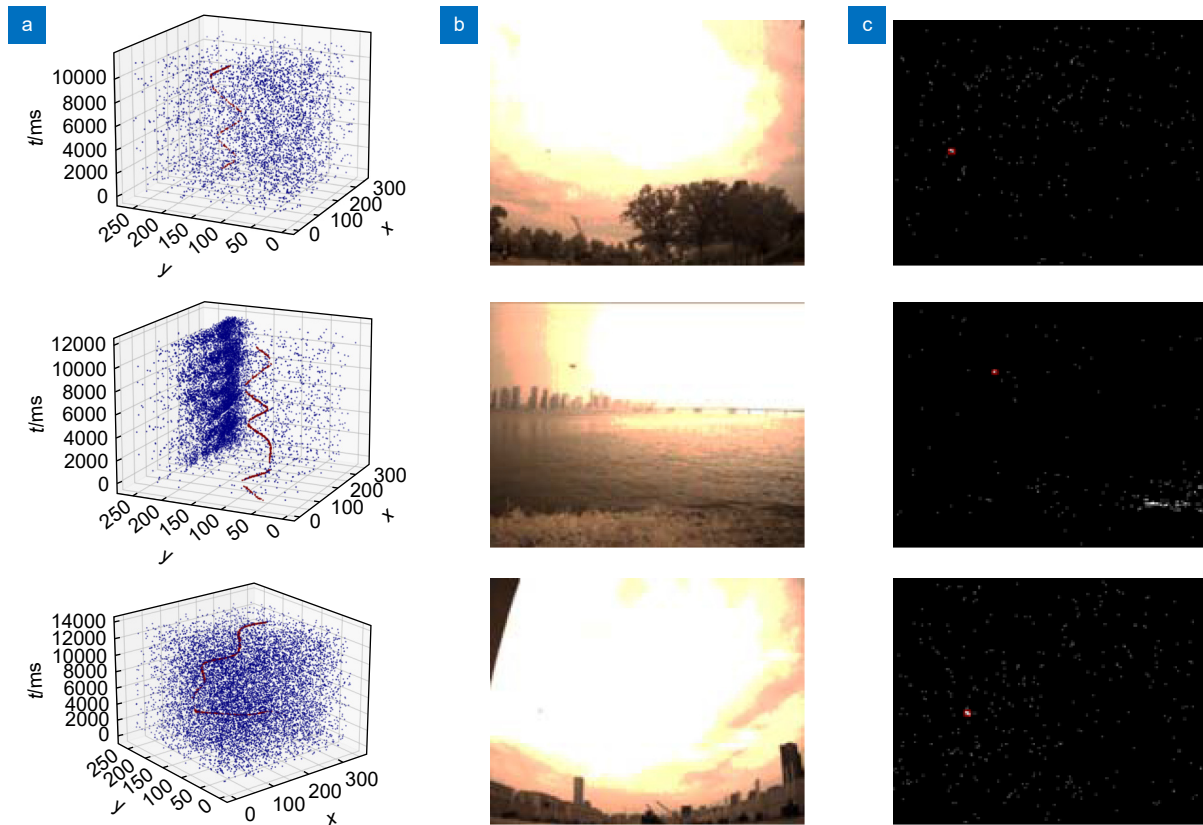


图 2 事件相机无人机检测数据集。(a) 事件点云, 蓝色为背景, 红色为无人机目标; (b) 可见光图像帧; (c) 事件帧, 红色框内为目标

Fig. 2 Event camera drone detection dataset. (a) Event point cloud, with blue background and red drone target; (b) Visible light image frame; (c) Event frame, with the target within the red box

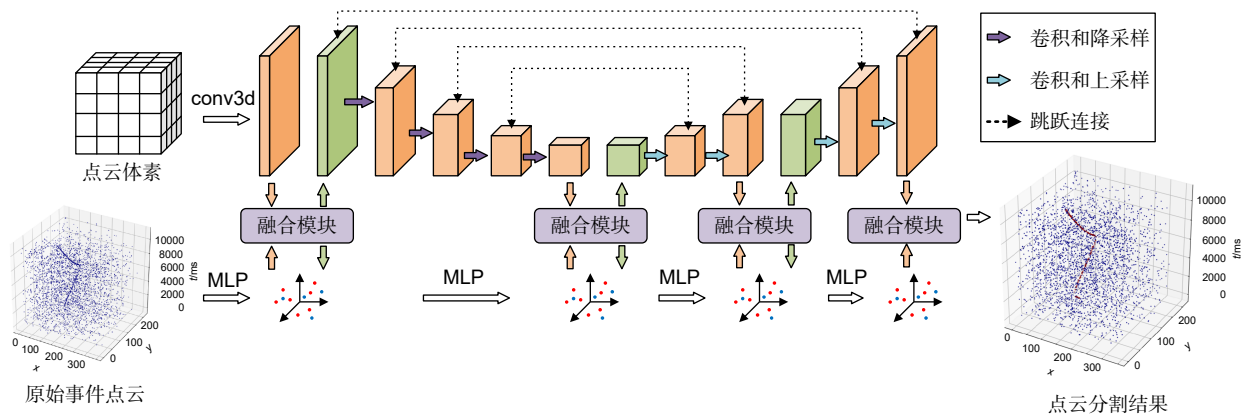


图 3 点-体素融合分割网络

Fig. 3 Point-voxel fusion segmentation network

### 3.2 体素特征提取分支

体素特征提取分支以事件体素为输入, 由原始事件点云进行离散化生成, 每个体素单元的分辨率为  $2 \text{ pixel} \times 2 \text{ pixel} \times 10 \text{ ms}$ , 这种表示方式不仅减少了数据冗余, 还便于捕捉局部时空特征, 如图 4 所示。体素特征提取分支的输入可以描述为:  $V \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times \frac{T}{10}}$ 。该分支是一个类似于 UNet 的结构, 通过逐步下采样以捕获高层次的语义信息, 并通过逐步上采样以恢复空间分辨率并融合多尺度特征。具体而言, 该分支包含一个输入层, 四个下采样和四个上采样阶段, 这 9 个阶段的维度分别为 32、64、128、256、256、128、128、64 和 32。下采样每个阶段的特征都通过跳跃连接与对应的上采样阶段特征连接。

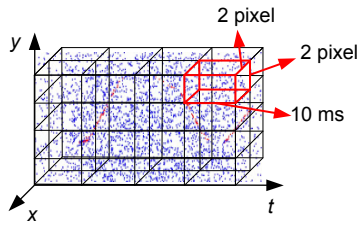


图 4 事件点云体素化  
Fig. 4 Event point cloud voxelization

### 3.3 双分支特征融合模块

在点分支网络中, 对于每一点, 都会学习到一个“点特征”对应每个点的细节特征。在体素分支网络中, 体素化将整个事件点云平均的划分为不同的子空间, 因此一个子空间内包含多个事件点, 这样提取的方式具有更大的感受野, 可以更好地捕获全局特征。双分支特征融合模块将包含每个点细节特征的“点特征”和包含更多全局特征的“体素特征”在网络的不同阶段进行融合, 使得网络可以有效地学习到由粗到细的特征, 实现更好的分割效果。

对于某一事件点  $e_i$ , 来自点云特征提取分支的特征可表示为  $X_i^p \in \mathbb{R}^d$ , 来自体素的特征可表示为  $X_i^v \in \mathbb{R}^d$ 。  $X_i^p$  通过不同的 MLP 层计算可得, 不同的层具有不同的维度  $d$ 。  $X_i^v$  根据该点在体素的位置确定, 即  $X_i^v = V(x, y, t)$ 。

双分支特征融合的核心任务是在大量无用信息的干扰下将有用信息聚合在一起。加法和级联是聚合多个特征的常用操作, 但它们都会受到许多非信息特征的影响。它们可以表述为

$$X_{\text{con}} = \text{concat}(X_1, \dots, X_L), \quad (1)$$

$$X_{\text{add}} = \sum_{i=1}^L X_i, \quad (2)$$

其中:  $X_{\text{con}}$  和  $X_{\text{add}}$  分别为级联和加法融合特征向量, 由于点特征的多样性, 不同视图分支的特征在重要性上并不一致, 但是加法和级联特征融合策略忽略了每个特征向量的有用性, 在融合过程中将大量无用特征与有用特征结合。门控机制可以通过测量每个特征的重要性来自适应地聚合信息, 受其启发, 本文设计了一种基于门控机制的多视图分支特征融合模块, 该模块可以描述为

$$\psi = \text{softmax}(\text{conv}(\text{concat}(X_p, X_v))), \quad (3)$$

$$X_{\text{door}} = \psi \cdot X_p + (1 - \psi) \cdot X_v. \quad (4)$$

该模块首先将两个视图的特征拼接, 经过  $3 \times 3$  卷积和 softmax 后转化为概率权重  $\psi$ 。概率权重表示两种在不同空间位置的特征的重要性。然后对相应通道的概率权重  $\psi$  分离, 并将分离后的概率权重与原始的点特征和体素特征分别进行加权, 得到具有不同重要性的点特征和体素特征。最后, 经过门控加权的两个视图特征向量相加融合得到最终的特征融合结果  $X_{\text{door}}$ 。在特征融合结果中, 门控机制凸出了更重要的特征, 抑制了不重要的特征。融合过程如图 5 所示。

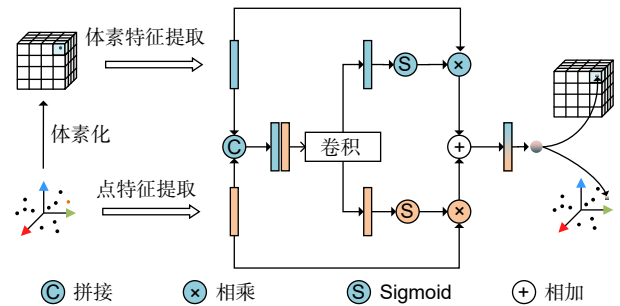


图 5 点-体素特征交互融合模块  
Fig. 5 Point-voxel feature interaction fusion module

## 4 结果与分析

### 4.1 评价指标

本文将事件目标检测任务视为事件点云分割任务, 将点云 IoU 作为评价指标。为了方便与基于帧的目标检测方法进行对比, 本文同时将 AP 作为评价指标。本文直接检测目标运动轨迹, 因此将轨迹检测率和虚警率也作为评价指标。点云 IoU 表示预测的分割结果与真实标签的重叠程度。IoU 的计算方法是将预测结果和真实标签的区域进行交集和并集运算, 然后计算

它们的比值。IoU 值越大预测点云与真值点云之间的重叠度越高, 分割效果越好, 可表示为

$$IoU = \frac{TP}{TP + FP + FN}, \quad (5)$$

其中:  $TP$  表示正确预测的目标,  $FP$  表示错误预测的目标,  $FN$  表示为错误预测的背景。

AP 是根据 Precision-Recall 曲线计算的面积。它对不同召回率下的 Precision 进行插值, 并计算插值曲线下的面积。AP 是评估目标检测算法在不同召回率下的综合性能指标。

当分割的事件点和真值轨迹的交并比为 80% 以上时, 认为对目标轨迹检测成功, 标记为正确预测的轨迹  $TP$ 。若将噪声错误检测为轨迹实例, 认为对目标轨迹检测错误, 标记为错误预测的轨迹  $FP$ 。若将真实轨迹预测为背景噪声, 则将该轨迹标记为错误预测的背景  $FN$ 。轨迹检测率  $P_d$  指的是正确检测的轨迹实例占全部真值轨迹实例的比例, 定义为

$$P_d = \frac{TP}{TP + FN}. \quad (6)$$

轨迹虚警率  $P_f$  指的是错误检测的轨迹实例占全部真值轨迹实例的比例, 定义为

$$P_f = \frac{FP}{TP + FN}. \quad (7)$$

## 4.2 实验训练设置

对提出的网络进行训练, 随机初始化模型参数, 采用 Adam 优化器进行优化, 每次输入 0.5 s 的事件数据, 训练批大小为 16, 动量为 0.9, 衰减系数设置为 0.0002, 最大迭代次数为 80000。初始学习率设置为 0.0002, 迭代次数到达 60000 时, 学习率线性衰减到 0.000002, 以促进损失函数进一步收敛。实验在单

张 NVIDIA 3090 GPU 上进行, 完成一次完整的训练大约需要 10 h。

## 4.3 实验结果与分析

### 4.3.1 对比实验

为了验证所提算法对低空无人机目标的检测能力, 分别对通用目标检测器 (YOLOv7<sup>[32]</sup>、SSD<sup>[33]</sup>、FastRCNN<sup>[34]</sup>)、点云目标分割网络 (Pointnet<sup>[35]</sup>、Pointnet++<sup>[36]</sup>)、基于事件相机的目标检测网络 (RVT<sup>[37]</sup>、SAST<sup>[38]</sup>) 以及小目标检测网络 (DNA-Net<sup>[39]</sup>、OSCAR<sup>[40]</sup>) 四类网络在 Ev-UAV 数据集上进行了对比实验。点云目标分割网络的输入为时间  $\Delta t$  内的事件点云。为了适应网络的输入, 通用目标检测器和小目标检测网络的输入为事件帧。事件帧  $L$  将时间  $\Delta t$  内的同一像素的事件进行积累, 像素  $(i, j)$  处的值为该时间段内产生的事件数量, 可以表示为

$$L(i, j) = \sum_t^{t+\Delta t} e(i, j, t, p). \quad (8)$$

为了比较的公平性, 对于所有的对比方法,  $\Delta t$  均与本文所提网络保持一致, 设置为 0.5 s。

对比实验结果如表 1 所示, 从中可以看到: 1) 相比于基于帧的方法, 基于点的方法性能普遍更优。这主要是因为远距离无人机目标尺寸较小, 没有明显的形状轮廓等特征, 模型难以有效地检测到目标; 而基于点云的方法可以直接分割目标运动轨迹特征, 因此具有更好的性能。即使是专为小目标检测设计的基于帧的网络也具有较低的检测率和较高的虚警率。2) 多视图融合方法优于单一视图的方法。尽管使用了多视图融合, 但由于紧凑的网络设计, 本文的 PVNet 仍然显示出非常有竞争力的运行延迟。相比于次优的方

表 1 不同算法在 Ev-UAV 上的实验结果

Table 1 Experimental results of different algorithms on Ev-UAV dataset

| 方法         | 框架 | $P_d/\%$   | $P_f/\%$    | AP         | IoU        | 推理时间/ms |
|------------|----|------------|-------------|------------|------------|---------|
| SSD        | 帧  | 88.3(+8.2) | 13.3(-10.1) | 62.7(+6.8) | —          | 18.7    |
| FastRCNN   | 帧  | 88.7(+7.8) | 13.5(-10.3) | 63.2(+6.3) | —          | 19.5    |
| YOLOV7     | 帧  | 89.8(+6.7) | 10.3(-7.1)  | 65.3(+4.2) | —          | 16.5    |
| RVT        | 帧  | 90.3(+6.2) | 10.1(-6.9)  | 65.7(+3.8) | —          | 10.5    |
| SAST       | 帧  | 90.1(+6.4) | 10.5(-7.3)  | 65.5(+4.0) | —          | 13.2    |
| DNA-Net    | 帧  | 90.5(+6.0) | 9.91(-6.7)  | 65.9(+3.6) | —          | 20.3    |
| OSCAR      | 帧  | 90.3(+6.2) | 10.2(-7.0)  | 66.3(+3.2) | —          | 25.3    |
| Pointnet   | 点云 | 91.3(+5.2) | 8.35(-5.1)  | 66.1(+3.4) | 66.7(+3.6) | 19.8    |
| Pointnet++ | 点云 | 93.5(+3.0) | 7.33(-4.1)  | 67.3(+2.2) | 67.5(+2.8) | 25.3    |
| Ours       | 点云 | 96.5       | 3.21        | 69.5       | 70.3       | 15.2    |



法 Pointnet++, 所提方法的检测率 $P_d$ 提高了 3%, 虚警率 $P_f$ 下降了 4.1%, mAP 提高了 2.2%, IoU 提高了 2.8%, 而运行时间缩短了 10.1 ms。

选取三组具有代表性的场景, 利用不同的检测算法对同一场景进行检测, 可视化结果如图 6 所示。从图中可以看出, 基于帧的算法 YOLOv7 出现了漏检, 且出现了较多虚警, 基于点云的算法 PointNet++ 无漏检, 但是存在少量的虚警, 而所提算法可以在无虚警率情况下有效地检测出所有小目标无人机。

4.3.2 消融实验

1) 视图的有效性

为了验证多视图融合的有效性, 对单个视图和双

视图融合进行了消融实验, 结果如表 2 所示。从表中可以看出, 相比于仅使用单视图, 所提方法的检测率 $P_d$ 最高可以提高 6.9%, 虚警率 $P_f$ 可以降低 6.6%, IoU 最高可以提高 7.1%, 而推理时间依旧满足实时检测要求。以上结果说明视图融合的方法在可接受的复杂性开销下可以有效地提高性能。

2) 融合方法的有效性

为了验证门控机制对多视图融合的有效性, 采用不同的融合方法进行了消融实验, 结果如表 3 所示。可以看到, 门控融合方法与加法、乘法以及拼接相比, 检测率 $P_d$ 分别提高了 6.3%, 4.7%, 2.7%, 虚警率 $P_f$ 分别降低了 5.4%, 4.4%, 3.1%, IoU 分别提升了

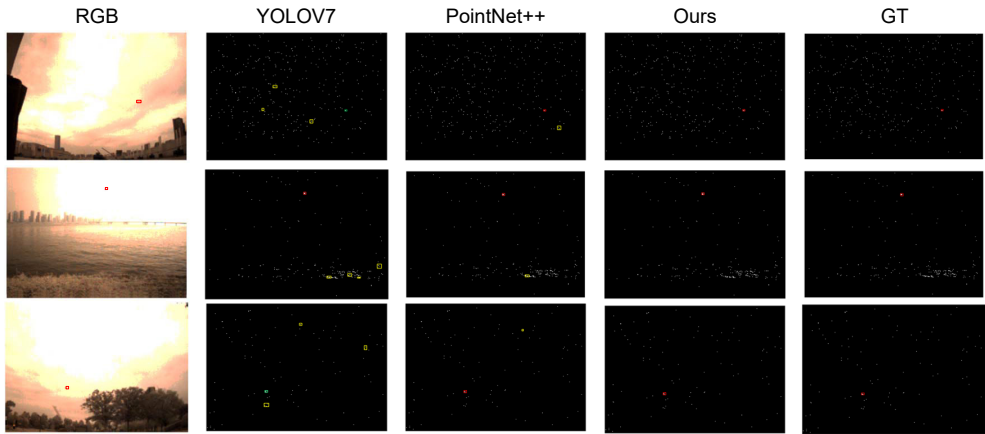


图 6 可视化检测结果对比。红色框为正确检测结果, 黄色框为虚警, 绿色框为漏检

Fig. 6 Visual comparison of detection results. The red box represents the correct detection result, the yellow box represents false alarm, and the green box represents missed detection

表 2 不同视图的影响。V 和 V+表示不同的体素分辨率: V 表示 4 pixel×4 pixel×20 ms, V+表示 2 pixel×2 pixel×10 ms

Table 2 The impact of different views. V and V+ represent different voxel resolutions: V represents 4 pixel × 4 pixel × 20 ms, V+ represents 2 pixel × 2 pixel × 10 ms

| 方法  | $P_d/\%$    | $P_f/\%$     | AP          | IoU/ $\%$   | 推理时间/ms |
|-----|-------------|--------------|-------------|-------------|---------|
| P   | 89.6 (+6.9) | 9.86(−6.65)  | 66.3 (+3.2) | 63.2 (+7.1) | 7.2     |
| V   | 90.6 (+5.9) | 8.98 (−5.77) | 65.7 (+3.8) | 63.8 (+6.5) | 10.2    |
| V+  | 92.5 (+4.0) | 5.33 (−2.12) | 66.8 (+2.7) | 64.7 (+5.6) | 12.3    |
| PV  | 94.3 (+2.2) | 4.32 (−1.11) | 67.2 (+2.3) | 69.2 (+1.1) | 19.2    |
| PV+ | 96.5        | 3.21         | 69.5        | 70.3        | 20.3    |

表 3 不同融合方法的有效性

Table 3 The effectiveness of different fusion methods

| 方法     | $P_d/\%$    | $P_f/\%$     | AP          | IoU         |
|--------|-------------|--------------|-------------|-------------|
| Add    | 90.2 (+6.3) | 8.63 (−5.42) | 67.3 (+2.2) | 66.8 (+3.7) |
| Multi  | 91.8 (+4.7) | 7.68 (−4.47) | 66.9 (+2.6) | 64.2 (+6.1) |
| Concat | 93.8 (+2.7) | 6.38 (3.17)  | 67.6 (+1.9) | 65.3 (+5.0) |
| Ours   | 96.5        | 3.21         | 69.5        | 70.3        |

3.7%, 6.1% 和 5%, 这说明了门控融合机制对多视图融合的有效性。

## 5 结 论

本研究针对神经形态传感器(事件相机)在低空无人机探测中面临的目标尺寸小、特征少、事件稀疏不规则等问题,突破传统的图像帧处理思路,将无人机目标检测任务建模为三维点云语义分割任务,设计了一种基于双视图融合的事件无人机目标分割模型和一种用于融合双视图特征的门控融合机制。实验中对自建的事件相机无人机目标检测数据集的定量和定性评估证明,与当前最先进的目标检测和点云分割网络相比,所提方法在保持低虚警的情况下具有最高的检测性能。

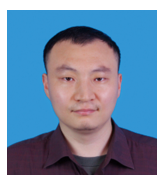
**利益冲突:** 所有作者声明无利益冲突

## 参考文献

- [1] Bouguettaya A, Zarzour H, Kechida A, et al. Vehicle detection from UAV imagery with deep learning: a review[J]. *IEEE Trans Neural Netw Learn Syst*, 2022, **33**(11): 6047–6067.
- [2] Chen H Y, Liu D B, Yan X W. Infrared image UAV target detection algorithm based on IDOU-YOLO[J]. *Journal of Applied Optics*, 2024, **45**(4): 723–731.  
陈海水, 刘登斌, 晏行伟. 基于 IDOU-YOLO 的红外图像无人机目标检测算法[J]. *应用光学*, 2024, **45**(4): 723–731.
- [3] Han J T, Tan K, Zhang W G, et al. Identification of salt marsh vegetation “fairy circles” using random forest method and spatial-spectral data of unmanned aerial vehicle LiDAR[J]. *Opto-Electron Eng*, 2024, **51**(3): 230188.  
韩江涛, 谭凯, 张卫国, 等. 协同随机森林方法和无人机 LiDAR 空谱数据的盐沼植被“精灵圈”识别[J]. *光电工程*, 2024, **51**(3): 230188.
- [4] Park S, Choi Y. Applications of unmanned aerial vehicles in mining from exploration to reclamation: a review[J]. *Minerals*, 2020, **10**(8): 663.
- [5] Sziroczak D, Rohacs D, Rohacs J. Review of using small UAV based meteorological measurements for road weather management[J]. *Prog Aerosp Sci*, 2022, **134**: 100859.
- [6] Wang Z Y, Gao Q, Xu J B, et al. A review of UAV power line inspection[C]//*Proceedings of 2020 International Conference on Guidance, Navigation and Control*, Tianjin, 2022: 3147–3159. [https://doi.org/10.1007/978-981-15-8155-7\\_263](https://doi.org/10.1007/978-981-15-8155-7_263).
- [7] Khan A, Gupta S, Gupta S K. Emerging UAV technology for disaster detection, mitigation, response, and preparedness[J]. *J Field Robot*, 2022, **39**(6): 905–955.
- [8] Li Y, Liu M, Jiang D D. Application of unmanned aerial vehicles in logistics: a literature review[J]. *Sustainability*, 2022, **14**(21): 14473.
- [9] Mademlis I, Mygdalis V, Nikolaidis N, et al. High-level multiple-UAV cinematography tools for covering outdoor events[J]. *IEEE Trans Broadcast*, 2019, **65**(3): 627–635.
- [10] Xi Y D, Yu Y, Ding Y Y, et al. An optoelectronic system for fast search of low slow small target in the air[J]. *Opto-Electron Eng*, 2018, **45**(4): 170654.  
奚玉鼎, 于涌, 丁媛媛, 等. 一种快速搜索空中低慢小目标的光电系统[J]. *光电工程*, 2018, **45**(4): 170654.
- [11] Zhang R M, Xiao Y F, Jia Z N, et al. Improved YOLOv7 algorithm for target detection in complex environments from UAV perspective[J]. *Opto-Electron Eng*, 2024, **51**(5): 240051.  
张润梅, 肖钰菲, 贾振楠, 等. 改进 YOLOv7 的无人机视角下复杂环境目标检测算法[J]. *光电工程*, 2024, **51**(5): 240051.
- [12] Chen X, Peng D L, Gu Y. Real-time object detection for UAV images based on improved YOLOv5s[J]. *Opto-Electron Eng*, 2022, **49**(3): 210372.  
陈旭, 彭冬亮, 谷雨. 基于改进 YOLOv5s 的无人机图像实时目标检测[J]. *光电工程*, 2022, **49**(3): 210372.
- [13] Zhang M C, Niu C H, Liu L S, et al. Infrared small target detection algorithm for UAV detection system[J]. *Laser Technol*, 2024, **48**(1): 114–120.  
张明淳, 牛春晖, 刘力双, 等. 用于无人机探测系统的红外小目标检测算法[J]. *激光技术*, 2024, **48**(1): 114–120.
- [14] Sedunov A, Haddad D, Salloum H, et al. Stevens drone detection acoustic system and experiments in acoustics UAV tracking[C]//*Proceedings of 2019 IEEE International Symposium on Technologies for Homeland Security*, Woburn, 2019: 1–7. <https://doi.org/10.1109/HST47167.2019.9032916>.
- [15] Chiper F L, Martian A, Vlădeanu C, et al. Drone detection and defense systems: Survey and a software-defined radio-based solution[J]. *Sensors*, 2022, **22**(4): 1453.
- [16] de Quevedo Á D, Urzaiz F I, Menoyo J G, et al. Drone detection and radar-cross-section measurements by RAD-DAR[J]. *IET Radar Sonar Navig*, 2019, **13**(9): 1437–1447.
- [17] Gallego G, Delbrück T, Orchard G, et al. Event-based vision: a survey[J]. *IEEE Trans Pattern Anal Mach Intell*, 2020, **44**(1): 154–180.
- [18] Shariff W, Dilmaghani M S, Kieley P, et al. Event cameras in automotive sensing: a review[J]. *IEEE Access*, 2024, **12**: 51275–51306.
- [19] Paredes-Vallés F, Scheper K Y W, De Croon G C H E. Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: from events to global motion perception[J]. *IEEE Trans Pattern Anal Mach Intell*, 2020, **42**(8): 2051–2064.
- [20] Cordone L, Miramond B, Thierion P. Object detection with spiking neural networks on automotive event data[C]//*Proceedings of 2022 International Joint Conference on Neural Networks*, Padua, 2022: 1–8. <https://doi.org/10.1109/IJCNN55064.2022.9892618>.
- [21] Li Y J, Zhou H, Yang B B, et al. Graph-based asynchronous event processing for rapid object recognition[C]//*Proceedings of 2021 IEEE/CVF International Conference on Computer Vision*, Montreal, 2021: 914–923. <https://doi.org/10.1109/ICCV48922.2021.00097>.
- [22] Schaefer S, Gehrig D, Scaramuzza D. AEGNN: asynchronous event-based graph neural networks[C]//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, New Orleans, 2022: 12361–12371. <https://doi.org/10.1109/CVPR52688.2022.01205>.
- [23] Jiang Z Y, Xia P F, Huang K, et al. Mixed frame-/event-driven fast pedestrian detection[C]//*Proceedings of 2019 International Conference on Robotics and Automation*, Montreal, 2019: 8332–8338. <https://doi.org/10.1109/ICRA.2019.8793924>.

- [24] Lagorce X, Orchard G, Galluppi F, et al. HOTS: a hierarchy of event-based time-surfaces for pattern recognition[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, **39**(7): 1346–1359.
- [25] Zhu A, Yuan L Z, Chaney K, et al. Unsupervised event-based learning of optical flow, depth, and egomotion[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 2019: 989–997. <https://doi.org/10.1109/CVPR.2019.00108>.
- [26] Wang D S, Jia X, Zhang Y, et al. Dual memory aggregation network for event-based object detection with learnable representation[C]//*Proceedings of the 37th AAAI Conference on Artificial Intelligence*, Washington, 2023: 2492–2500. <https://doi.org/10.1609/aaai.v37i2.25346>.
- [27] Li J N, Li J, Zhu L, et al. Asynchronous spatio-temporal memory network for continuous event-based object detection[J]. *IEEE Trans Image Process*, 2022, **31**: 2975–2987.
- [28] Peng Y S, Zhang Y Y, Xiong Z W, et al. GET: group event transformer for event-based vision[C]//*Proceedings of 2023 IEEE/CVF International Conference on Computer Vision*, Paris, 2023: 6015–6025. <https://doi.org/10.1109/ICCV51070.2023.00555>.
- [29] Chen N F Y. Pseudo-labels for supervised learning on dynamic vision sensor data, applied to object detection under ego-motion[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Salt Lake City, 2018: 644–653. <https://doi.org/10.1109/CVPRW.2018.00107>.
- [30] Afshar S, Nicholson A P, Van S A, et al. Event-based object detection and tracking for space situational awareness[J]. *IEEE Sensors Journal*, 2020, **20**(24): 15117–15132.
- [31] Huang H M, Lin L F, Tong R F, et al. UNet 3+: a full-scale connected UNet for medical image segmentation[C]//*Proceedings of ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing*, Barcelona, 2020: 1055–1059. <https://doi.org/10.1109/ICASSP40776.2020.9053405>.
- [32] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, 2023: 7464–7475. <https://doi.org/10.1109/CVPR52729.2023.00721>.
- [33] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//*Proceedings of the 14th European Conference on Computer Vision*, Amsterdam, 2016: 21–37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2).
- [34] Girshick R. Fast R-CNN[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*, Santiago, 2015: 1440–1448. <https://doi.org/10.1109/ICCV.2015.169>.
- [35] Charles R Q, Su H, Kaichun M. PointNet: deep learning on point sets for 3D classification and segmentation[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 2017: 77–85. <https://doi.org/10.1109/CVPR.2017.16>.
- [36] Charles R Q, Yi L, Su H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 2017: 5105–5114. <https://dl.acm.org/doi/abs/10.5555/3295222.3295263>.
- [37] Gehrig M, Scaramuzza D. Recurrent vision transformers for object detection with event cameras[C]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, 2023: 13884–13893. <https://doi.org/10.1109/CVPR52729.2023.01334>.
- [38] Peng Y S, Li H B, Zhang Y Y, et al. Scene adaptive sparse transformer for event-based object detection[C]//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, 2024: 16794–16804. <https://doi.org/10.1109/CVPR52733.2024.01589>.
- [39] Li B Y, Xiao C, Wang L G, et al. Dense nested attention network for infrared small target detection[J]. *IEEE Trans Image Process*, 2023, **32**: 1745–1758.
- [40] Dai Y M, Li X, Zhou F, et al. One-stage cascade refinement networks for infrared small target detection[J]. *IEEE Trans Geosci Remote Sens*, 2023, **61**: 5000917.

## 作者简介



李森(1988-), 男, 博士, 优聘副研究员, 分别于2012年和2017年获得国防科技大学硕士和博士学位。主要从事红外弱小目标探测研究。

E-mail: [lm8866@nudt.edu.cn](mailto:lm8866@nudt.edu.cn)



陈诺(1997-), 男, 博士研究生, 分别于2019年2022年获得中南大学学士和硕士学位, 目前正在攻读国防科技大学信息与通信工程博士学位。主要从事事件相机目标检测、弱小目标探测研究。

E-mail: [chennuo97@nudt.edu.cn](mailto:chennuo97@nudt.edu.cn)



【通信作者】安玮(1969-), 女, 博士, 教授, 分别于1995年和1999年获得国防科技大学硕士和博士学位。主要从事智能光电感知系统研究。

E-mail: [anwei@nudt.edu.cn](mailto:anwei@nudt.edu.cn)

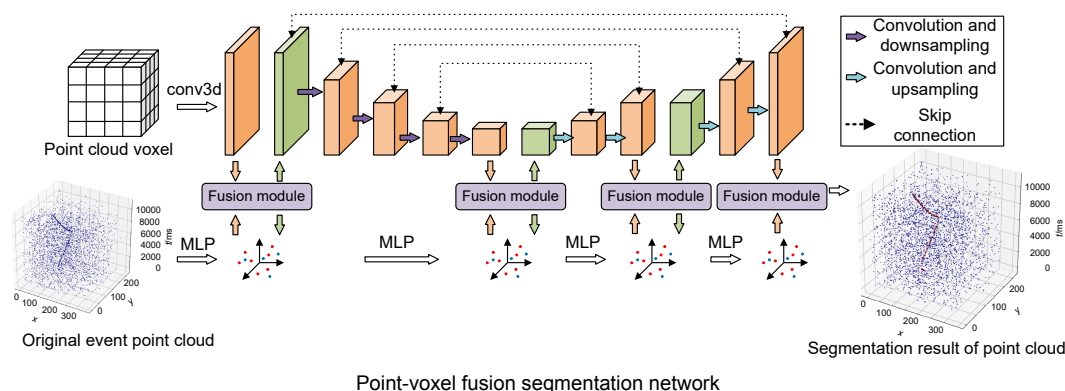


扫描二维码, 获取PDF全文



# Dual view fusion detection method for event camera detection of unmanned aerial vehicles

Li Miao, Chen Nuo, An Wei\*, Li Boyang, Ling Qiang, Li Weixing



**Overview:** With the proliferation of UAVs in various industries, concerns over airspace management, public safety, and privacy have escalated. Traditional detection methods, such as acoustic, radio, and radar detection, are often costly and complex, limiting their applicability. In contrast, machine vision-based techniques, particularly event cameras, offer a cost-effective and adaptable solution. Event cameras, also known as dynamic vision sensors (DVS), capture visual information in the form of asynchronous events, each containing the pixel location, timestamp, and polarity of an intensity change. This address-event representation (AER) mechanism enables efficient high-speed and high-dynamic-range visual data processing. Traditional cameras capture image frames with a fixed exposure time, which makes it difficult to adapt to changes in lighting conditions, resulting in the detection of blind spots in strong light and other scenes. Event cameras perform well in this situation. The existing object detection algorithms are suitable for synchronous data such as image frames, but cannot handle asynchronous data such as data streams. The proposed method treats UAV detection as a 3D point cloud segmentation task, leveraging the unique properties of event streams. This paper introduces a dual-branch network, PVNet, which integrates voxel and point feature extraction branches. The voxel branch, resembling a U-Net architecture, extracts contextual information through downsampling and upsampling stages. The point branch utilizes multi-layer perceptrons (MLPs) to extract point-wise features. These two branches interact and fuse at multiple stages, enhancing feature representation. A key innovation lies in the gated fusion mechanism, which selectively aggregates features from both branches, mitigating the impact of non-informative features. This approach outperforms simple addition or concatenation fusion methods, as demonstrated through ablation studies. The method is evaluated on a custom dataset, Ev-UAV, containing 60 event camera recordings of UAV flights under various conditions. Evaluation metrics include intersection over union (IoU) for segmentation accuracy and mean average precision (mAP) for detection performance. Compared to frame-based methods like YOLOV7, SSD, and FastRCNN, event-based methods like RVT and SAST, and point cloud-based methods like PointNet and PointNet++, the proposed PVNet achieves superior performance, with an mAP of 69.5% and IoU of 70.3%, while maintaining low latency. By modelling UAV detection as a 3D point cloud segmentation task and leveraging the strengths of event cameras, the PVNet model effectively addresses the challenges of detecting small, feature-scarce UAVs in event data. The results demonstrate the potential of event cameras and the proposed fusion mechanism for real-time and accurate UAV detection in complex environments.

Li M, Chen N, An W, et al. Dual view fusion detection method for event camera detection of unmanned aerial vehicles[J]. *Opto-Electron Eng*, 2024, 51(11): 240208; DOI: [10.12086/oe.2024.240208](https://doi.org/10.12086/oe.2024.240208)

Foundation item: Project supported by National Natural Science Foundation of China (62401591, 61921001), and the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (61921001)

College of Electronic Science and Technology, National University of Defense Technology, Changsha, Hunan 410073, China

\* E-mail: [anwei@nudt.edu.cn](mailto:anwei@nudt.edu.cn)