DOI:10.20189/j.cnki.CN/61-1527/E.202604007

事件语义分割的可靠区域引导学习研究综述

姚俊萍,孔鑫琰,李晓军*,程　鹏
(火箭军工程大学,西安 710025)

摘　要:事件相机具有高时间分辨率、高动态范围和低延迟等特性,可为复杂光照与高速运动场景下的语义分割提供有效的边缘与时序信息。围绕事件语义分割中的可靠区域引导学习问题,在界定相关概念与任务特点的基础上,从事件数据时空编码表示和可靠性证据来源出发,系统梳理了基于事件活跃区域、边缘显著区域和高置信度预测区域的典型方法。然后,结合常用数据集和评价指标,对相关模型的性能表现与技术特点进行了比较分析,总结了该方向在稀疏事件表征、可靠区域筛选、边界保持和预测一致性建模等方面的研究进展。本文进一步指出,现有研究仍存在可靠性证据来源单一、区域互补关系利用不足和动态场景适应能力有限等问题。最后,从稀疏时空建模、边界度量与先验构建、多区域协同引导和动态可靠区域建模等方面展望未来研究方向,以期为事件语义分割中的可靠区域引导学习研究提供参考。

关键词:事件语义分割;可靠区域引导;事件活跃区域;边缘显著区域;高置信度预测区域
中图分类号:TP391.4　　**文献标志码**:A　　**开放科学标识码(OSID)**:
文章编号:2097-2741(2026)04-0095-16

听语音　聊科研　与作者互动

稳健的语义分割是自动驾驶、机器人等智能设备自主感知的基础[1-4]。然而,基于帧的相机在帧率和动态范围方面存在固有局限,导致图像质量低下,阻碍了在高速运动、低光照或过曝等复杂条件下对具有判别性特征的提取[5-6]。在这些场景中,受生物视觉启发而产生的事件相机因其独特优势而成为更优选择,其通过异步检测光强变化,提供高时间分辨率和高动态范围,在挑战性条件下表现出稳健的性能[7-8]。2024 年 6 月,Gehrig 等[9]提出将事件相机与每秒 20 帧的帧式摄像头结合使用,可实现与 5 000 帧摄像头相当的低延迟性能,且带宽需求仅相当于 45 帧摄像头。这主要归因于事件相机异步测量光强变化的特性,具备高时间分辨率和稀疏性,能显著降低带宽和延迟需求[10]。相较于基于帧的相机,事件相机的这些特性使其在高动态、低光照或高速运动等复杂环境下拥有更优的感知能力与能效比。

然而,事件数据本身呈现稀疏且非结构化的特征,使得语义分割任务面临诸多挑战。一是事件数据具有与传统帧图像显著不同的数据分布特性,使得直接迁移基于帧的分割方法往往难以获得理想效果[11];二是在保持分割精度的同时,冗余计算开销难以被有效降低[12]。这些问题成为当前事件语义分割研究面临的主要瓶颈[13-14]。

针对上述问题,研究人员逐渐意识到事件数据中存在显著差异化的区域特征:一些区域在语义上往往更加关键或在表征上更为可靠,而部分区域则可能受到噪声或背景中非关键运动的干扰。传统基于全局统一处理的方式未能有效区分这些差异,导致模型在训练与推理过程中对背景或噪声区域投入过多计算资源,影响整体分割性能。Gallego 等[15]提出边缘与高对比度像素是事件数据最活跃

收稿日期:2026-01-26　　**修回日期**:2026-04-25　　**录用日期**:2026-05-06
基金项目:陕西省自然科学基础研究计划项目(2025JC-YBMS-783)
第一作者:姚俊萍(1978—),女,教授,主要从事信息系统与数据工程研究。
*通信作者:李晓军(1981—),男,副教授,主要从事智能信息处理与分布式计算、数据工程研究。

引用格式:姚俊萍,孔鑫琰,李晓军,等. 事件语义分割的可靠区域引导学习研究综述[J]. 火箭军工程大学学报,2026,40(4):95-110. *YAO Junping, KONG Xinyan, LI Xiaojun, et al. Research on reliable region-guided learning for event semantic segmentation[J]. Journal of Rocket Force University of Engineering, 2026, 40(4): 95-110.*

的区域，也是最具判别力的视觉信息源，分割模型理应优先关注这些区域。因此，引入可靠区域的引导机制具有重要意义。一方面，能够有效缓解事件数据固有的稀疏性与时空分布不规则性限制。通过在特征学习过程中突出事件密集且语义显著的区域，模型能够聚焦于关键结构，从而显著提升分割精度。另一方面，能够避免在冗余或低置信度区域上进行无效计算，进而降低整体处理能耗。同时，该策略在减少冗余特征传递与计算开销的基础上，可进一步提高模型的推理速度，实现了精度与效率的平衡。这种设计为事件语义理解提供了一种兼具高效性与鲁棒性的解决思路。基于上述思路，研究者提出事件活跃、边缘显著和高置信度预测区域等多种可靠区域引导的语义分割方法[5,8]，为事件语义分割的性能提升奠定了重要基础。此外，不同可靠区域反映的可靠性依据并不完全相同，如何理解其适用特点、互补关系与潜在局限，仍有必要进行系统梳理和分析。

近年来，虽存在若干关于事件相机视觉的综述，但对于“事件语义分割”这一子研究多是简略介绍，常与检测、光流等研究并列讨论，缺乏以“问题—方法—数据—分析—展望”为主线的系统化梳理。已有综述通常更侧重于事件相机视觉任务的整体发展或代表性方法归纳，对事件语义分割中不同可靠区域建模思想之间的内在联系、共性瓶颈和潜在研究空白讨论相对不足。因此，本文围绕事件语义分割中的可靠区域引导学习展开综述，系统梳理基于事件活跃区域、边缘显著区域和高置信度预测区域的代表性方法，并结合方法比较分析不同可靠区域引导策略的适用特点与局限，在此基础上，讨论可靠区域协同建模、动态建模等潜在研究方向，以期为后续研究提供更具系统性和启发性的参考。

1 问题描述

事件相机以异步事件流记录场景亮度变化，在高速运动和复杂光照条件下能够提供高时间分辨率、高动态范围的视觉信息，但其稀疏、异步和非结构化特征也使语义分割面临语义表达不足、边界结构模糊以及冗余事件干扰等问题。为在统一框架下分析事件语义分割中的关键区域建模问题，首先，给出事件语义分割任务及事件数据时空编码表示；在此基础上，从可靠性证据来源出发，将可靠区域归纳为事件活跃区域、边缘显著区域和高置信度预测区域3类，分别对应事件响应强度、结构边界显著性和模型预测置信度，为后续方法综述、性能比较与发展趋势分析提供问题基础和分类依据。

1.1 事件语义分割

语义分割是计算机视觉中的一项基础任务[16-19]，目标是对图像中每个像素赋予一个语义类别标签。形式上，设输入的图像为

$$I\in\boldsymbol{R}^{H\times W\times C} \tag{1}$$

式中：H 和 W 分别表示图像的高度和宽度；C 表示通道数。模型通过学习式(2)所示的参数化函数，将每个像素映射到 K 个语义类别之一。

$$f_\theta:\boldsymbol{R}^{H\times W\times C}\rightarrow\{0,1,\cdots,K-1\}^{H\times W} \tag{2}$$

式中：θ 为模型参数。

近年来，随着深度神经网络的发展，针对语义分割已有许多研究，但大多数方法依赖基于帧的相机所获取的图像，难以应对高速运动或低照度等极端场景，其图像常伴随模糊与细节丢失，限制了语义信息的表达能力。相比之下，事件相机具备高动态范围、高时间分辨率及低功耗等优势，能在复杂环境下保持稳定感知，为语义分割提供全新的数据形式[11,20]。

与基于帧的相机工作原理不同，事件相机仅检测每个像素光强的变化，并异步生成事件流[21]。每个事件 e_i 可表示为四元组，即

$$e_i=(x_i,y_i,t_i,p_i) \tag{3}$$

式中：x_i 和 y_i 分别代表事件发生的坐标；t_i 为事件发生的时间戳；$p_i\in\{-1,+1\}$ 代表事件亮度变化的正负极性。事件触发的条件是局部亮度变化达到设定的阈值 C，即

$$\log I(x_i,y_i,t_i)-\log I(x_i,y_i,t_i-\Delta t)\geqslant p_iC \tag{4}$$

整个事件序列形成的事件数据可表示为

$$e=\{e_1,e_2,\cdots,e_N\} \tag{5}$$

事件相机具有无运动模糊、超高动态范围(大于120 dB)、低延迟及低功耗等特性[22]，使得其具有显著的应用潜力和技术优势。

1.2 事件数据时空编码表示

原始事件以异步脉冲序列形式输出，若直接送入深度模型，容易因空间极度稀疏与时间异步导致特征利用低效。更重要的是，不同的时空编码方式会直接影响“事件活跃区域”“边缘显著区域”等可靠区域在特征空间中的显现形式，从而对后续可靠区域引导策略产生根本性影响。因此，

本节系统梳理事件数据的主流时空编码表示方法，并分析其对后续可靠区域建模与语义分割框架设计的影响。

1.2.1　帧化累积表示

帧化累积表示主要是按照传统相机的处理方式，将事件数据映射到二维平面上，使得稀疏事件数据转换为稠密表示形式，可以直接使用深度学习的方式对数据进行处理。即首先在固定的时间窗 $[t_0, t_0+\Delta t]$（或固定事件数 N）内对事件进行时间或极性上的投影，获得与图像尺寸一致的 2-D 张量 $I \in R^{H\times W\times C}$，从而实现对神经网络的复用。这类时空编码方法的主要优势是可以对主流的语义分割框架进行“无缝迁移”，具有较高的处理效率，但存在转换时抛弃事件的精细时间顺序信息的不足。在事件累积策略上，主要有固定时间窗[23-24]和固定事件数[25-26]两种设计。

对事件数据进行累积操作后，需要对其内部时间细节进行详细编码，现有研究多采用多通道变体的方式进行编码，目的是将累积操作时丢失的“窗口内顺序”等信息重新找回。现有研究采用 2-通道[27]、4-通道[28]和 6-通道[11,29]的设计，充分利用了事件的运动敏感特性。不同于多通道变体编码的方式，部分研究[30-32]采用 B-bin 编码的方式，强调把时间轴离散化，让网络捕获窗口内的先后顺序，类似于“帧内再切帧”的操作，保留时间顺序的同时统计增强细节信息。

帧化累积表示凭借其实现简洁、迁移难度低、实践友好等特点，仍是事件语义分割任务的主流起点，且其保留了事件活跃区域的计数梯度信息。但在对极端动态或超低延迟场景的适应上，体素化和纯脉冲流正逐步取代其地位。

1.2.2　时空体素/张量表示

时空体素/张量表示的核心思想是将原始事件序列映射到规则的栅格 $V \in R^{B\times H\times W}$ 中，在空间平面上保持像素对齐，并在时间维度上将时间窗离散成 B 个区间（bin）。与帧化累积表示不同，这一设计既保留了窗口内的相对时序，又构造出稠密张量，可直接送入卷积神经网络（convolutional neural network，CNN）或 Transformer 网络等网络模型中。

在具体实现上，文献[33-34]对每个 bin 进行计数，将正负极性合并后输入模型；文献[35]采用 5-bin 体素化序列的方式输入 3-D UNet 网络，两者的共性是用纯计数即可满足 3-D 卷积提取时序。进一步地，文献[36]采用相同的 6-bin 计数结合时间统计，并在网络内部引入时空位置传播以强化 bin 间依赖。这类做法在计数后补充时间均值与方差，不仅实现了保留时间顺序的同时，还增加了对事件“新旧程度”的感知能力。

时空体素/张量表示法在时间信息和运动细节上明显优于帧化累积表示法，既保留了事件活跃区域，又对显著边缘区域有较好响应，更适用于高速运动或极端光照等场景。然而，三维稀疏张量充分保留事件关键信息，导致显存和计算开销较大，且网络还需要对“时空稀疏”进行特定优化，以避免大量空体素造成的效率浪费[37]。针对时空体素/张量的不同应用场景，近期研究将事件数据切片后再进行混合编码，或引入自适应时间窗口，以在分辨率与资源消耗之间取得更优平衡[38]。

1.2.3　时间表面表示

时间表面表示是一种将事件数据映射到二维像素平面的紧凑表示法[39-42]。给定一个事件 (x_i, y_i, t_i, p_i)，记录最近一次事件的时间戳 t_i，并将其当作灰度值，即可得到

$$S_p(x,y;t)=\exp\left(-\frac{t-t_{\text{last}}^{(p)}(x,y)}{\tau}\right) \tag{6}$$

式中：$t_{\text{last}}^{(p)}(x,y)$ 表示截止时刻 t 像素上最后一个极性为 p 的事件时间戳；τ 为时间衰减常数，使较近事件的权重较大，较远事件的权重较小。式（6）在局部邻域内形成一幅“衰减的时间记忆图”，记录了近期事件的时间痕迹，形成一个局部时空上下文，既保留了亚毫秒级顺序，又能直接映射到二维网格。

与其他事件编码表示法相比，时间表面表示法高度压缩信息，每个像素只保留一个时间戳，导致其在纹理丰富的场景中表示效果会下降。但其天然包含运动边缘几何信息的特性，有利于边缘可靠区域的显现。为了降低时间表面对噪声的敏感度，可以在时空窗口内对事件进行滤波操作来计算每个像素值[43]。

1.2.4　纯脉冲流表示

纯脉冲流表示是指不设定任何先验时间窗或空间累积的前提下，把事件相机输出的原始四元组 $e=(x,y,t,p)$ 逐事件送入网络，并在网络内部使用异步记忆、图结构或脉冲神经元的方式进行增量更新。该表示法可以完全保持事件时间顺序，不引入 Δt 累积延迟。比如，文献[44]采用多尺度层级记忆平面的方式，每产生一条事件就立即以稀疏注意

力写入层级记忆，使网络始终保持与物理事件流同步，避免任何帧化或体素化带来的信息丢失和等待延迟；文献[9]采用异步稀疏有向图节点的方式进行处理，把每一条事件当作一个新节点实时插入一张随时间增长的有向时空图，然后，仅对与之相关的局部子图做增量消息传递，从而完全摆脱任何帧、体素或时间窗累积。

纯脉冲流表示法完全保持事件信息，在微小边缘和高置信区域表现更佳。此外，该方法在处理时能够实现零累积和超低时延效果，但也存在一定的局限，如难以预测实时帧率、触发处理器“抖动”现象。

1.3 事件语义分割的可靠区域引导学习

在明确事件数据的多种时空编码形式之后，下一步需要关注的问题是：如何在不同事件表示中识别并利用具有较高语义判别可靠性的区域。因此，本节从可靠区域的概念与形式化定义出发，给出事件活跃区域、边缘显著区域和高置信度预测区域3类典型可靠区域划分，为后续的技术综述奠定理论基础。

在事件语义分割任务中，由于事件数据存在天然的稀疏性和不规则性特征[15]，语义分割模型在处理此类数据时可能会出现伪目标预测、边缘模糊或目标边界不清晰等问题。因此，可靠区域引导学习的核心在于从事件表示中筛选具有较高判别价值的区域，并利用其引导特征学习、损失约束或推理计算，从而提升事件语义分割的精度、效率与鲁棒性。

在事件语义分割任务中，可靠区域一般是指时空维度或语义特征层面，具备足够信息密度、运动响应显著或能够提供稳定语义判别的关键区域，这些区域可以为语义分割任务提供稳定可靠的特征支撑[45]。其可形式化定义如下。

设时间窗口 Δt 内采样得到的事件集合为

$$\varepsilon_{t,\Delta t}=\{e_i=(x_i,y_i,t_i,p_i)\mid t_i\in\Delta t\} \tag{7}$$

式中：(x_i,y_i)表示事件触发位置；t_i 表示事件时间戳；p_i 表示事件极性。经过事件帧、体素网格或时间表面等方式编码后，可得到事件表示图 X。设 Ω 表示事件表示图或特征图上的空间位置集合，$u\in\Omega$ 表示其中一个像素位置或局部区域。

从可计算角度看，区域可靠性可以从事件响应、结构边界和语义预测3个方面进行度量。首先，事件活跃度 $A(u)$用于描述位置 u 邻域内的事件触发强度，可由事件计数、事件密度或局部运动响应计算得到；其次，结构显著性 $B(u)$用于描述位置 u 是否位于目标边缘、轮廓或空间梯度较大的区域，由事件表示图的梯度强度、边缘响应或轮廓连续性进行度量。最后，语义置信度 $C(u)$用于描述模型对位置 u 的语义预测是否可靠，可由最大类别概率、预测熵或伪标签置信度进行度量。

基于上述3类可靠性证据，可靠区域可进一步划分为事件活跃区域、边缘显著区域和高置信度预测区域。

1)事件活跃区域(event-active region，EAR)是指在局部时间窗口内事件触发频率较高、事件密度较大或运动响应较强的区域。该类区域通常反映物体运动、亮度变化或高对比度结构所引起的有效事件响应。结合前文定义，则事件活跃区域可定义为

$$\Omega_A=\{u\in\Omega\mid A(u)\geqslant\tau_A\} \tag{8}$$

式中：τ_A 为事件活跃度阈值。该类区域通常对应事件触发密集、运动响应明显或亮度变化显著的位置，能够为模型提供较强的动态信息。

2)边缘显著区域(salient edge region，SER)是指在事件表示图或特征图中具有明显边缘响应、轮廓结构或较大空间梯度的区域。该类区域通常对应目标边界、纹理边缘或结构突变位置，对目标轮廓建模和边界细化具有重要作用，有助于缓解事件语义分割中的边界模糊和目标轮廓不清晰问题。设 $B(u)$表示位置 u 的结构显著性，则边缘显著区域可定义为

$$\Omega_B=\{u\in\Omega\mid B(u)\geqslant\tau_B\} \tag{9}$$

式中：τ_B 为结构显著性阈值。

3)高置信度预测区域(high-confidence region，HCR)是指模型在语义预测过程中具有较高类别置信度或较低不确定性的区域。该类区域通常由预测概率、预测熵、伪标签置信度或类别一致性等指标确定，可用于伪标签筛选、自训练、损失重加权和预测校准等过程。设 $C(u)$ 表示位置 u 的语义置信度，则高置信度预测区域定义为

$$\Omega_C=\{u\in\Omega\mid C(u)\geqslant\tau_C\} \tag{10}$$

式中：τ_C 为语义置信度阈值。若模型在位置 u 的类别预测概率为 $P(c\mid u)$，则语义置信度可表示为

$$C(u)=\max_{c\in C}P(c\mid u) \tag{11}$$

式中：C 表示类别集合，$c\in C$ 表示类别索引。除最大类别预测概率外，语义置信度也可以通过预测熵或不确定性估计进行计算。

因此，所讨论的可靠区域可表示为由3类区域

共同构成的集合，即

$$\Omega_r = \Omega_A \cup \Omega_B \cup \Omega_C \tag{12}$$

在研究中，Ω_r 的 3 类子区并非互斥，可相交形成多重权重掩码。需要指出的是，3 类可靠区域并非彼此独立，而是从运动响应、结构边界和语义置信等不同层面描述区域可靠性，其间具有一定互补性，但也可能受到背景运动、伪边缘或模型置信偏差等因素影响。因此，可靠区域引导学习不仅需要关注单一区域线索的提取，还需进一步考虑不同可靠性证据之间的协同关系，以形成更加稳健的区域判断机制。

基于以上定义，以可靠性证据的来源为划分依据，可将现有方法归纳为基于事件活跃区域的方法、基于边缘显著区域的方法和基于高置信度预测区域的方法，分类细节如图 1 所示。

1）基于事件活跃区域的方法。利用输入数据本身提供的可靠性线索，事件相机在物体运动、光照变化或边缘处触发密集。将事件数量密集的位置视为可靠区域，数量稀疏区域则视为背景或低价值区域。

2）基于边缘显著区域的方法。依据数据表征中空间梯度强度与轮廓连续性，利用在边缘处事件多、梯度大的特性确定可靠区域。

3）基于高置信度预测区域的方法。以模型输出的置信度/不确定性作为“可靠性”度量，将高置信像素归为可靠区域，用于自训练、伪标签与损失重加权等决策校准过程。

综上，从可靠性证据来源维度出发，开展事件语义分割的可靠区域引导学习研究综述，从核心思想、优势局限等维度对各类方法进行对比归纳。

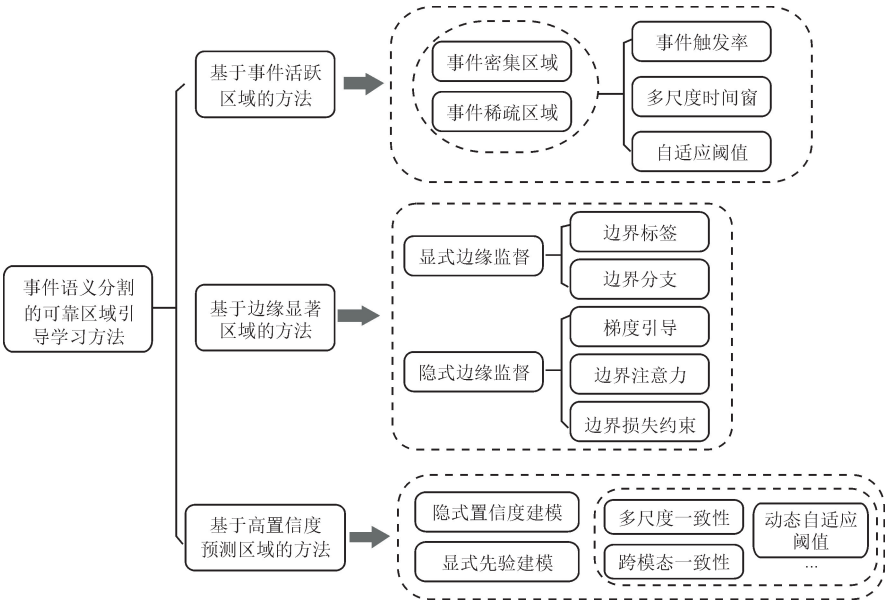


图 1　事件语义分割的可靠区域引导学习方法分类

Fig. 1　Classification of reliable region-guided learning approaches for event semantic segmentation

1.4　常用数据集与评价指标

在事件语义分割任务中，数据集与评价指标共同构成方法开发与性能评估的基础，不仅影响模型的训练过程，也直接关系到不同方法之间的公平比较和结果分析。为此，本节对当前常用数据集及评价指标进行简要介绍，为后续方法综述与性能对比建立实验协议提供参考。

1.4.1　数据集

目前，事件语义分割研究中常用数据集主要为 DDD17[46] 和 DSEC-Semantic[12]。这两类数据集包含车辆驾驶的多种真实场景，但在数据质量和场景多样性上有着不同的侧重表现。

DDD17 数据集是首个公开的事件相机与基于帧的相机结合的端到端自动驾驶数据集。该数据集涵盖了高速公路和城市道路等多种真实驾驶环境，并包含不同时间段的环境场景采集。同时，该数据集提供了语义分割标签，为每一个像素分配所属标签。

DSEC-Semantic 数据集是基于原始 DSEC 数据集[47]提出的首个大规模、高质量的事件语义分割数据集，同时提供了配对的 RGB 图像和精细粒度的语义分割标签。该数据集专注于在复杂环境下

记录真实的驾驶场景，侧重在极具挑战的光照环境，如快速光暗变化、强逆光、极低光照环境，尤其在进出隧道等典型高速场景中展现出其优越性。

总体来看，DDD17侧重中低分辨率、类别较少的自动驾驶场景，更适合评估方法在基础语义分割精度、事件稀疏场景下的鲁棒性；DSEC-Semantic则在高分辨率、复杂光照和高速运动场景下提供了更丰富的语义类别和更细粒度的像素标注，更适合考察方法在极端光照变化、高速运动和细粒度结构下的表现。

1.4.2 评估指标

基于事件的语义分割任务中，主要包含像素准确率、平均交并比、吞吐量/帧率和浮点运算次数等评估指标。在多数研究中，平均准确率和交并比为通用评价指标。

1)像素准确率

像素准确率 P_{Acc} 定义为所有类别上被正确分类的像素数占总像素数的比例，其表达式为

$$P_{\mathrm{Acc}}=\frac{\sum_{i=1}^{K}TP_i}{\sum_{i=1}^{K}N_i} \tag{13}$$

式中：K 为类别总数；TP_i 为第 i 类中被正确预测的像素数；N_i 为第 i 类的总像素数，即该类的真实像素总数。

2)平均交并比

平均交并比 mIoU 是语义分割任务中常用的评价指标，用来衡量预测区域与真实标注区域的重叠程度。

对于 K 类像素，有以下公式：

$$\begin{aligned} IoU_c&=\frac{TP_c}{TP_c+FP_c+FN_c}\\ I_{\mathrm{mIoU}}&=\frac{1}{K}\sum_{c=1}^{K}IoU_c \end{aligned} \tag{14}$$

式中：K 为类别数量；IoU_c 为第 c 类预测区域与真实区域的重合程度；TP_c 为预测正确的像素数量；$TP_c+FP_c+FN_c$ 表示类别 c 的预测区域与真实区域的并集大小。

3)吞吐量/帧率

吞吐量 FPS 表示单位时间内可处理的帧数，即平均帧时的倒数，可表示为

$$\mathrm{FPS}=\frac{N_{\mathrm{frames}}}{\sum_{i=1}^{N_{\mathrm{frames}}}t_i}=\frac{1}{\frac{1}{N_{\mathrm{frames}}}\sum_{i=1}^{N_{\mathrm{frames}}}t_i} \tag{15}$$

式中：t_i 为第 i 帧端到端推理耗时；N_{frames} 为测试图像帧数。

4)浮点运算次数

浮点运算次数 FLOPs 包括加法累加运算(accumulate operations，Acc)和乘加运算(multiply-accumulate operations，MAC)，有以下计算公式：

$$\begin{aligned} &\mathrm{MAC}=T\times C_{\mathrm{out}}\times H_{\mathrm{out}}\times W_{\mathrm{out}}\\ &\mathrm{ACC}=\theta_{t-1}S_{\mathrm{H}}S_{\mathrm{W}}C_{\mathrm{out}}+T\times C_{\mathrm{out}}\times H_{\mathrm{out}}\times W_{\mathrm{out}}+\theta_t\\ &\mathrm{FLOPs}=\mathrm{MAC}+\mathrm{ACC} \end{aligned} \tag{16}$$

式中：T 为时间步数；θ_{t-1}、θ_t 分别为前/当前时间步脉冲数；S_{H}、S_{W} 分别为卷积核高和宽；C_{out} 为输出通道数，H_{out}、W_{out} 分别为输出特征图尺寸。

5)模型参数量

模型参数量 Params 反映深度学习模型的存储开销与部署难度，参数量越小，模型越适合在边缘设备等算力受限平台上运行。设网络共有 L 层，第 l 层的权重和偏置参数分别为 $W^{(l)}$、$b^{(l)}$，$|\cdot|$ 表示张量中标量元素的个数，则总参数量可定义为

$$\mathrm{Params}=\sum_{l=1}^{L}(|W^{(l)}|+|b^{(l)}|) \tag{17}$$

在事件语义分割任务中，可靠区域引导策略若能在不显著增加参数量的前提下提升精度，将更具实际应用价值。

6)时间步数

对于基于时间展开的网络，如脉冲神经网络或时序编码模型，设每个样本在时间维度上离散为 T 个时间步(time steps)，时间步越多，理论上可以获得越丰富的时间动态信息，但也会带来成倍的计算与能耗开销。

综上，事件语义分割的评价体系不应局限于传统的 Acc 和 mIoU，还需综合考虑吞吐量、计算复杂度、参数规模以及时间步数等因素，才能全面反映不同可靠区域引导策略在精度、实时性与能耗之间的折中。

2 基于事件活跃区域的方法

事件相机检测每个像素的光强变化，在物体运动、光照变化或边缘处会触发大量事件。这些触发密集的时空位置被称为“事件活跃区域”，包含关键的运动和结构信息，而其他区域往往为背景或无用信息。此类方法通常以事件计数或活动度作为可靠性度量，将高事件密度区域视为语义上更重要或

观测更可靠的区域，通过设计加权损失、注意力掩码或稀疏计算策略，引导语义分割网络在这些区域投入更多表达能力与计算资源。该类方法常见模型架构如图 2 所示，通过与真实标签预测得到的损失进行权重调整，使特征提取阶段聚焦于事件活跃区域。

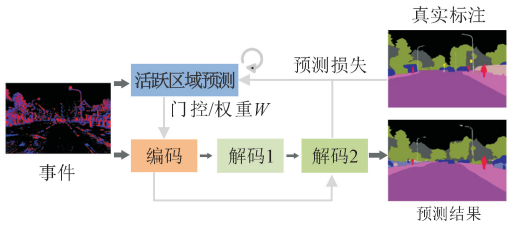


图 2　基于事件活跃区域的方法架构

Fig. 2　Architecture of methods based on event active regions

在基于事件活跃区域的建模方面，越来越多的工作以 CNN 与 Transformer 等人工神经网络（artificial neural network，ANN）为基础展开[48-50]。受限于事件数据的稀疏性，传统卷积方法在局部感受野内容易丢失上下文信息。相比之下，Transformer 技术的自注意力机制可以从稀疏输入中找到长距离的依赖关系[51]，使得稀疏分散的事件之间建立联系。针对这一思路，文献[52]提出以事件计数图为先验信息构造注意力权重，抑制噪声并增强事件的显著区域。该方法对给定时间窗内每个像素的事件频次做统计，频次高的位置即为事件更“活跃”的区域。类似地，文献[53]把相邻关键帧之间在固定时间窗内发生的事件数量作为强度变化的度量，事件数量多的区域归类为“活跃区域”。该方法直接继承了 EvSegSNN[54] 的网络结构与训练流程，并引入事件活跃区域选择机制，在几乎不损失精度情况下，显著降低了延迟与能耗。

需要指出的是，Transformer 中标准自注意力机制的二次计算复杂度限制了模型的扩展性[55-56]。在自注意力计算过程中，没有任何事件的空白区域也会被编码为标记并参与计算，造成不必要的计算与内存开销。针对这一瓶颈，文献[36]提出了后验注意力机制与 Transformer 架构相结合的方法，将活跃区域作为注意力掩码，引导模型关注事件丰富的区域。值得注意的是，该模型在 ANN 结构的基础上引入无监督优化策略，显著提升了语义分割性能。

基于事件稀疏性的设计路径，核心在于将计算与注意力聚焦于真正“活跃”的区域，抑制噪声对判别边界的影响，在其他区域仅施加轻量约束或直接跳过计算，从而在几乎不损失精度的前提下，降低延迟与能耗。除侧重单模态事件稀疏特征建模的研究外，也有工作显式挖掘图像帧与事件数据之间的互补性。例如，文献[57]采用双分支整合帧数据和事件数据，提取事件数据中所有非零体素的位置作为参考点集，这组点集则为“事件活跃区域”。相较于以往的密集参考点集，该方法能显著降低大特征图的计算量。

总体而言，现有基于事件活跃区域的策略虽然能够通过聚焦事件密集区域在一定程度上提升事件语义分割的效率与精度，但其关键局限在于方法仍以事件密度或触发强度作为主要依据进行筛选与计算分配，并未在建模机制上真正实现对事件活跃区域、非事件区域与噪声区域的细致区分，从而导致注意力/计算资源仍可能被低价值区域占用，时空信息的动态权衡也缺乏更强的结构化约束[58]。针对此问题，未来可结合稀疏 Transformer[59] 或 Mamba[60] 等高效序列建模框架，设计面向事件数据的自适应稀疏化与选择性更新机制，在空间上通过稀疏注意力/Token 选择抑制冗余计算，在时间上通过状态选择性传播强化关键时刻的表达，从而实现更稳定的时空协同建模、更高的鲁棒性与泛化能力。

3　基于边缘显著区域的方法

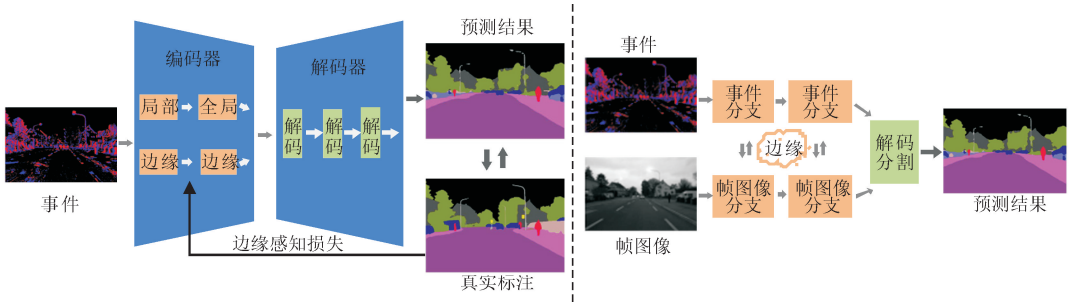
相较于仅依赖事件计数的活跃区域，物体轮廓、运动边界等“边缘显著区域”在事件数据中往往具有更强的判别性和结构约束能力。本节围绕这一更细粒度的可靠区域线索，综述显式边缘监督与隐式边缘监督两类方法，并分析其在边界精度与能耗之间的平衡。

事件相机天然在显著梯度变化区域，如物体边缘、轮廓、运动边界等产生大量事件，因此，边缘显著性在事件数据中非常突出。在语义分割研究中，此类区域对准确划定各物体的边界具有极大价值。对于边缘显著区域的研究，通常可分为显式边缘监督与隐式边缘监督两类，常见模型架构如图 3 所示。主要分为仅事件边缘分割和跨模态边缘融合分割，利用边缘显著特征进行显式或隐式监督，保证模型在边缘处准确划定各物体的边界，提高语义分割任务的准确性。

3.1　显式边缘监督方法

显式边缘监督是指通过人工标注或半自动标

注等方式得到明确边缘标记，并将这些标记直接作为监督信号参与训练。考虑到事件数据先天稀疏、信息主要集中于有限的显著区域，一个直观思路是明确告知模型应将注意力分配到何处。文献[30]指出事件与运动边缘具有天然强相关性，进而提出 SELSAM 方法。SELSAM 方法借助 Segment Anything Model 生成“干净”的目标轮廓，再结合边缘算子过滤物体内部纹理边缘，仅保留与分类相关的外轮廓，避免“伪边缘”误导分割。文献[61]提出 SpikeBRGNet 方法，将边缘显著区域和脉冲发放相结合，在结构上显式设置 Boundary Aware 分支来有效挖掘边缘区域信息；在损失函数设计上，提出 Boundary Region Guided Loss 设计。以上方法直接对显著的“边缘区域”施加强监督，从而使模型的学习重心放到“边缘显著且可靠”的区域上，优点是可控性强、目标明确，不足之处是依赖高质量标注信号。



(a)仅事件边缘分割 (b)跨模态融合分割

图 3 基于边缘显著区域的方法架构

Fig. 3 Architecture of methods based on edge-salient regions

除结构层面的引导外，显式边缘监督亦可用于提升伪标签生成的质量与鲁棒性。由于先前的伪标签生成方法所产生的标签质量较差，影响语义分割的效果，文献[62]提出事件辅助的低光照增强模块，利用事件边缘信息进行显式边缘监督。该模块利用事件相机提供的边缘信息，结合 Canny 边缘检测，设计边缘一致性(相似度)损失，使增强后的图像在低光照条件下仍可生成准确的标签。

总体而言，上述研究虽然处于模型训练的不同环节，但共同体现了“以显式边缘信息作为监督信号”的思路。一方面，保证了训练过程中特征提取的可靠性，另一方面，使得生成的数据标签更接近真实边界，从而形成互补的整体优化框架。

3.2 隐式边缘监督方法

与显式边缘监督不同，隐式边缘监督不直接提供边缘标注，而是通过网络结构设计(如多尺度、注意力机制、门控等)、特征提取方式(如全脉冲模式[63-64])等方式，让模型在训练过程中自然学习到边缘显著性特征。如文献[65]提出采用多尺度和全脉冲编码的方法 MFSSeg。该方法专门设计的 EFS-Res 可利用深度可分离卷积改善脉冲特征图的分布，从而实现更精确的分割边缘。

常规的体素/网格累积设计会丢失事件数据的时序连续性并造成高频或边界信息损失，导致边界区域细节变差。针对这一不足，文献[66]提出 METHOD 方法，把事件的时序特征与光流统一成稠密、时序一致的运动张量。光流提供粗粒度动态位移，事件时序特征提供细粒度边界线索，两者结合能够选择丰富语义区域并保持更清晰的边界。

与文献[53]方法类似，部分研究通过解决事件与图像的跨模态融合问题提升分割精度。同时，考虑到事件体素编码方法可以保留事件动态信息和边缘结构，文献[67]提出跨模态注意力软对齐 ASA 模块，让事件体素编码表示中的动态信息和边缘结构与 RGB 模态丰富的纹理和语义进行互补，有效提升了事件语义分割性能。文献[68]提出边缘引导注意力模块 EGA，从事件体素中识别并利用边缘特征，并借助事件光流与帧光流的局部一致性突出边缘运动信息，从而有效抑制噪声事件的干扰。具体而言，当某一区域的事件响应较为密集，且与图像帧在局部运动信息上具有较高一致性时，该区域将被赋予更高权重，从而隐式增强真实边缘区域的特征响应。

现有事件语义分割多依赖高质量像素级标注，但事件数据非结构化与稀疏的特性会导致精细标注成本居高不下。针对上述问题，部分研究者考虑如何在不依赖昂贵标注的情况下，提高事件相机的语义分割精度，同时降低噪声和偏差[18]。文献[69]提出弱监督条件下的点级标注方法，并在不同类别相邻的边界区域引入正则化进行约束。实验结果

表明，结合标签传播与事件边缘特性，在弱监督条件下仍能实现有效的语义分割。

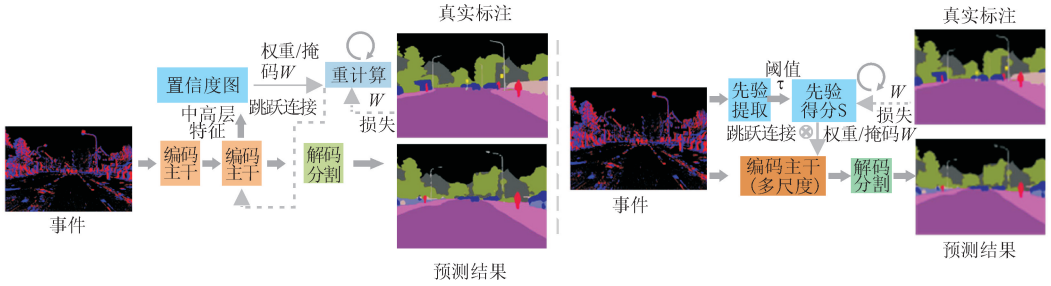
综合而言，基于边缘显著区域的方法充分利用了事件数据在物体轮廓、运动突变处存在高密度触发的特性，并通过显式或隐式监督的方式强调这些区域的特征表达。其优势在于能够精准定位显著的结构边界，从而提升分割的细节还原能力与目标轮廓的完整性。与基于事件活跃区域的方法不同，边缘显著区域强调捕捉空间结构边界的显著特征，适合优化边界精度和视觉细节。但现有方法仍存在明显不足，多数工作依赖外部信息提供监督信号，例如，通过事件重建得到伪强度图、借助额外模态或标签生成策略构造边界/区域伪标签，再将其用于训练边界分支或显著性分支。这类“外源监督”在一定程度上提升了可学习性，却也引入了额外误差与偏置，使模型性能与泛化能力对外部流程高度敏感。面对上述问题，未来可通过从数据内部挖掘边界信息，避免对外部重建或伪标签辅助的依赖。

4　基于高置信度预测区域的方法

与基于原始事件统计（活跃度）或几何梯度（边缘显著性）的可靠区域不同，本节关注的是从模型输出中反推得到高置信度预测区域，并围绕隐式置信建模与显式先验建模两种路线展开综述。

在可靠区域引导的事件语义分割研究中，高置信度预测区域作为一种重要的可靠性度量，被广泛应用于不同的模型设计中。其基本思想是基于模型输出的置信度分布，将预测概率较高的区域视为可靠区域，并在后续的训练优化与特征融合过程中给予更大权重。

现有研究主要可以分为显式先验建模和隐式置信建模两种思路，常见模型架构如图 4 所示。两者区别在于可靠度来源不同，前者通过置信/不确定性预测获取特征回归值，后者则通过先验提取获取可靠区域。



（a）隐式置信建模　　（b）显式先验建模

图 4　基于高置信度预测区域的方法架构

Fig. 4　Architecture of methods based on high confidence prediction regions

4.1　隐式置信建模方法

学习式的隐式置信建模方法强调由神经网络模型自动学习置信度分布，摆脱对手工先验特征或规则阈值的依赖。其核心思想是在生成语义分割结果的同时，对模型预测不确定性进行隐式建模，从而为高置信度区域的筛选提供依据。常见的做法包括利用多任务学习框架将置信度估计作为辅助任务，或通过不确定性建模直接输出隐式置信度。

面向低延迟事件分割，文献[44]在层级记忆机制中显式利用置信度，训练阶段以高置信区域更新短时/长时记忆单元，减少错误写入导致的记忆污染，实现与物理事件流的同步、稀疏而稳定的状态维护。文献[70]则将传统点积注意力替换为似然自注意力，在事件点云内部直接估计区域的预测可靠性，以此调节注意力分配与消息聚合，天然贴合事件数据的非均匀触发分布。此类方法的共同优势在于置信学习与分割表征统一优化，可自适应不同事件率与场景动态，有效降低端到端时延与错误放大。

此外，文献[26]提出的 HPL-ESS 方法采用基于 DaFormer[71]改造的骨干网络对混合伪标签进行迭代优化。尽管其未显式以“高置信区域”作为门控信号，但动机与机制均是为了提升伪标签质量、避免错误强化，与高置信思想一致。

总体而言，学习式隐式置信建模凭借其端到端可学习性与结构灵活性，能够更好地适配事件数据的稀疏与异步特征，并为“可靠样本优先”的记忆更新与推理调度提供依据。

4.2　显式先验建模方法

部分研究者依赖显式先验建模来生成高置信度区域，例如，文献[31]提出事件活动建模模块 AEIM，通过多尺度汇聚获得场景活动图，并据此学

习像素级注意掩码，选择性地保留高置信事件细节；文献[36]提出基于时空传播的事件语义分割方法KWYAF，其中，顺序运动编码与ER2SM模块通过局部窗口掩蔽与可靠区域筛选策略，显式识别高置信度区域，并将这些区域作为时空传播源点，使可靠预测扩散至邻近稀疏区域，从而提升整体分割性能。此类方法的局限在于对活动度阈值、窗口大小等人为先验信息过于依赖，对静态或低活动场景的适应性相对不足。

显式先验建模和学习式隐式置信建模两种方法在置信度的度量方式上有所不同，前者更具有可解释性，而后者则具有更强的灵活性与端到端优化潜力。总体而言，基于高置信度预测区域的方法能够有效降低噪声干扰、提高分割精度并提升计算效率，但其性能仍然依赖于模型在初始阶段的预测质量，若高置信度区域本身存在系统性偏差，则可能导致误差的放大。因此，如何进一步提高高置信度区域的判定可靠性，仍是未来研究需要关注的关键问题。

5 方法比较与性能评估

在统一评估协议下，对典型方法在DDD17与DSEC-Semantic数据集上的性能表现进行综合比较与分析。表1汇总了各方法在主流数据集上的Acc、mIoU、Time Steps、Params、FLOPs与FPS等指标。

需要说明的是，前述章节主要依据可靠区域的来源对方法进行划分，而本节则从网络实现架构的角度，将现有方法归纳为3类。①ANN方法：以CNN、Transformer等人工神经网络架构为主体，通常基于实值激活、密集张量计算和梯度反向传播优化进行事件特征建模与语义预测，是当前事件语义分割中应用较为广泛的实现范式；②SNN方法：以脉冲神经网络为主体，通常基于脉冲编码、事件驱动计算和神经元时序动态进行事件特征建模与语义预测，在信息表达机制上更契合事件相机异步、稀疏的输出特性；③Hybrid方法：以多模态融合或异构网络协同为主要特征，通常通过融合事件流与帧图像信息，或在同一系统中联合采用ANN与SNN模块，实现“事件-帧”互补感知与“密集-稀疏”协同计算，是兼顾语义表达能力和事件稀疏特性的混合实现范式。两种划分维度相互独立又在一定程度上相互补充，有助于从“可靠区域建模思路-网络架构实现方式”两个层面，对不同方法的性能差异进行更为立体和系统的分析。

由表1可以看出，在ANN架构方法中，EV-SegNet在DSEC-Semantic上的Acc与mIoU均明显落后于KWYAF、ESEG等后续工作，一方面反映出早期方法受到事件标注稀缺与监督不足的限制，另一方面也说明在缺乏显式“可靠区域”建模时，模型更容易受到噪声事件与背景冗余触发的干扰。相比之下，EV-Segformer在DDD17上的mIoU与EV-SegNet性能相近，但在DSEC-Semantic上有显著提升，说明其引入的“后验注意力”机制更加契合高动态、强对比场景下事件数据的统计特性。与基于事件活跃区域的思路不同，HMNet、KWYAF与ESEG通过显式建模高置信度预测或边缘显著区域，在保持网络规模可控的前提下显著增强了边界判别力，其Acc与mIoU已处于ANN范式的第一梯队，但在复杂场景、跨模态融合能力方面仍与混合架构存在差距。

在Hybrid架构方法中，引入帧模态以补充语义细节与纹理信息，使整体性能较纯事件ANN有大幅提升。HALSIE重点优化跨模态时空特征提取与低能耗实时推理，相比其他混合方法，其在保证较高FPS和较低能耗的同时，更侧重于工程可部署性，Acc与mIoU仍略有不足。Hybrid-Seg、EIS-Net与BRENet等方法则在多模态融合框架中显式融入“可靠区域引导”，如基于帧域边缘或高置信区域构建事件域先验，从而在ANN与Hybrid范式中取得较优的综合表现。其中，BRENet在DDD17上的mIoU达到78.56%，在保证较高精度的同时兼顾了多模态融合的稳定性，体现出可靠区域策略与跨模态结构结合的潜力。

对于SNN架构方法，整体来看，其精度水平尚普遍低于最优的ANN与Hybrid方法，但借助事件驱动与稀疏计算机制，在时延与能耗方面具有天然优势，更适合低功耗、强实时的边缘场景。EvSeg-SNN、Spiking-DeepLab与Spiking-FCN等早期方法在mIoU上仍有明显提升空间，而SegSNNnet与MFS-EFS通过引入可靠区域或多尺度事件特征后，在较小参数量和有限时间频数下实现了精度与效率更好的平衡。值得注意的是，SpikingEDN与Spike-BRGNet的精度已接近甚至超过部分ANN方法，其中，Spike-BRGNet在DSEC-Semantic上的mIoU比SpikingEDN提升约了0.019 1，同时保持典型SNN能耗与时延优势。结合其采用的可靠区域引导策略，说明在SNN框架内显式挖掘和利用

可靠区域，可以在不显著增加模型复杂度的前提下有效提升事件语义分割性能。

表 1　事件语义分割方法对比分析

Table 1　Comparative analysis of event semantic segmentation methods

方法	所属策略类型	网络架构	DDD17 数据集						DSEC-Semantic 数据集					
			Acc/%↑	mIoU/%↑	Time steps	Params/M↓	FLOPs/G↓	FPS↑	Acc/%↑	mIoU/%↑	Time steps	Params/M↓	FLOPs/G↓	FPS↑
EV-SegNet[11]	Base	ANN	89.76	54.81	—	—	—	—	88.61	51.76	—	—	—	—
E2VID[17]	Base	ANN	83.24	44.77	—	—	—	—	76.67	40.70	—	—	—	—
ESS[12]	Base	ANN	87.86	52.46	—	—	—	—	84.04	44.87	—	—	—	—
EV-Segformer[29]	EAR	ANN	94.72	54.29	—	—	—	—	90.38	56.04	—	—	—	—
HMNet[44]	HCR	ANN	—	—	—	—	—	—	89.80	55.00	—	—	—	—
KWYAF[36]	EAR/HCR	ANN	90.04	57.69	—	—	—	—	90.87	57.75	—	—	—	—
ESEG[30]	SER	ANN	90.68	59.97	—	—	—	—	91.47	57.55	—	—	—	—
HALSIE[14]	Base	Hybrid	92.50	60.66	—	—	—	—	89.01	52.43	—	—	—	—
Hybrid-Seg[57]	EAR	Hybrid	95.07	67.31	—	—	—	—	94.27	66.57	—	—	—	—
EISNet[31]	HCR	Hybrid	96.04	75.03	—	—	—	—	95.12	73.07	—	—	—	—
BRENet[66]	SER	Hybrid	96.61	78.56	—	—	—	—	95.85	74.94	—	—	—	—
EvSegSNN[54]	Base	SNN	—	45.54	20	8.55	—	—	—	—	—	—	—	—
Spiking-DeepLab[72]	Base	SNN	—	44.63	5	15.89	4.05	100	—	44.58	5	15.89	14.02	100
Spiking-FCN[72]	Base	SNN	—	42.87	5	18.64	35.23	72.73	—	38.52	5	18.64	17.18	126.46
SegSNNnet[53]	EAR	SNN	—	51.93	1	0.41	—	114.29	—	47.91	1	0.33	1.17	759.85
MFS-EFS[65]	SER	SNN	—	62.80	2	0.92	8.11	—	—	46.10	2	1.11	28.88	—
SpikingEDN[23]	EAR	SNN	—	53.15	1	8.60	—	—	—	53.04	1	8.50	—	—
Spike-BRGNet[32]	SER	SNN	—	54.72	5	7.68	—	—	—	54.95	5	7.68	—	—

注：“—”表示原始文献未公开该项数值。

由表 1 可以看出，在同一网络架构范式内，显式建模事件活跃区域、边缘显著区域或高置信度预测区域的可靠区域引导方法，往往能在 Acc 与 mIoU 上稳定优于未采用该策略的基线模型；在 ANN、Hybrid 与 SNN 3 类架构之间，则存在精度、实时性与能耗之间的典型折中。进一步来看，当前方法的性能提升主要来自对局部可靠线索的有效利用，但仍存在可靠性证据来源相对单一、不同可靠区域协同不足以及动态场景适应能力有限等问题。未来研究可在现有可靠区域选择策略基础上，进一步关注多区域协同引导与动态可靠区域建模，使事件活跃、边缘显著和语义高置信等信息形成更加统一的可靠性判断机制。

6　结论与展望

本文从“问题描述—关键技术—数据与评估—分析与展望”出发，对事件语义分割领域近年面向可靠区域引导学习的代表性方法进行了系统梳理与对比，重点围绕事件活跃区域、边缘显著区域与高置信度预测区域 3 类可靠性证据，总结其核心思想、典型架构与适用条件，并结合公开数据集与通用指标对方法性能进行了对照分析与归纳。尽管相关研究已在提升分割精度、降低冗余计算和改善

边界质量等方面取得进展，但现有方法仍普遍存在可靠性证据来源较单一、区域互补关系利用不足以及动态可靠性建模不充分等问题。基于上述总结与分析，该领域仍存在如下值得关注的研究机会。

1）面向选择性计算的稀疏时空建模研究

在事件语义分割中，现有方法往往仍以事件密度或触发强度作为主要依据进行筛选与计算分配，并未在建模机制上真正实现对事件可靠区域的细致区分。未来可引入稀疏 Transformer 与 Mamba 等高效序列建模框架，在空间维度，通过可学习的稀疏注意力、Token/区域选择与多尺度窗口机制，将计算预算自适应地分配给语义贡献更高的区域；在时间维度，借助选择性状态更新与长程依赖建模，避免对全时序进行均匀处理，从而兼顾效率与时空一致性。同时，未来可开展与不确定性评估、动态阈值以及时序窗口自适应策略相应的联合设计研究，使稀疏化决策由“固定规则”转向“数据驱动”。事件语义分割有望从“基于事件密度的启发式聚焦”进一步走向“机制驱动的选择性计算”。

2）面向时空统计驱动的边界度量与先验构建研究

针对现有方法依赖外部信息提供监督信号，导致存在额外误差与偏置的问题，从数据内部挖掘边界信息，避免对外部重建或伪标签的依赖，是颇具潜力的研究方向。其核心思路是将边界视为事件流自身的可分离结构信号，通过时空统计与一致性原则形成可学习的“边界先验”。例如，可利用事件在时间维度上的聚集/稀疏变化或局部时空梯度不连续性等，构造与语义边界高度相关的边界性度量，从而为边界学习提供更稳定、可迁移且与事件数据特性更一致的监督信号。

3）高置信可靠区域的判定可靠性估计方法研究

现有方法通常将 Softmax 置信度阈值作为高置信预测区域的判定依据，并据此进行伪标签筛选、记忆更新或注意力调节。然而，在训练早期、跨场景或噪声干扰下，容易出现“高置信但预测错误”的区域信息，进而在迭代过程中被持续强化，导致误差扩散。因此，将高置信预测区域判定从单一阈值规则提升为更稳健的可靠性估计机制，是一项极具研究价值的探索方向，可在隐式置信建模/显式先验建模框架下，引入不确定性与校准约束，使置信度与真实正确率更一致，避免错误在迭代过程中被持续放大，从而实现高置信可靠区域的更稳健判定与误差抑制。

4）多区域协同引导与动态可靠区域建模研究

现有可靠区域引导方法多围绕事件活跃区域、边缘显著区域或高置信度预测区域中的某一类线索展开建模，虽然能够在特定场景下提升分割性能，但对不同可靠性证据之间互补关系的利用仍不充分。未来，可进一步探索多区域协同引导机制，通过注意力融合、门控调节或图结构传播等方式，使不同可靠区域之间形成信息互补，从而降低对单一线索的依赖。此外，事件流具有显著的时变性，可靠区域的分布会随运动速度、光照条件和场景复杂度等因素的变化而发生变化。相较于固定阈值、固定时间窗或静态掩码策略，动态可靠区域建模能够根据事件率变化，其局部运动强度和预测不确定性自适应地调整区域权重，有助于提升模型在复杂光照、高速运动和跨场景条件下的鲁棒性。因此，多区域协同引导与动态可靠区域建模有望成为后续事件语义分割研究的重要方向。

参考文献

[1] 金立生，韩广德，谢宪毅，等. 基于强化学习的自动驾驶决策研究综述[J]. 汽车工程，2023，45(4)：527-540.
Jin Lisheng，Han Guangde，Xie Xianyi，et al. Review of autonomous driving decision-making research based on reinforcement learning[J]. Automotive Engineering，2023，45(4)：527-540.

[2] 王晨，杜晨曦，刘瑞军，等. 基于 RGB-D 图像的语义分割方法综述[J]. 计算机辅助设计与图形学学报，2025，37(1)：100-119.
Wang Chen，Du Chenxi，Liu Ruijun，et al. A review of semantic segmentation methods based on RGB-D images [J]. Journal of Computer-Aided Design & Computer Graphics，2025，37(1)：100-119.

[3] 蔡子悦，袁振岳，庞明勇. 深度学习的点云语义分割方法综述[J]. 计算机工程与应用. 2025，61(11)：22-30.
Cai Ziyue，Yuan Zhenyue，Pang Mingyong. Survey ondeep-learning-based point cloud semantic segmentation[J]. Computer Engineering and Applications，2025，61(11)：22-30.

[4] Zhao Jingyuan，Zhao Wenyi，Deng Bo，et al. Autonomous driving system：a comprehensive survey[J]. Expert Systems with Applications，2024，242：122836.

[5] Liu Zhanwen，Li Yuhang，Wang Yang，et al. Boosting visual recognition in real-world degradations via unsupervised feature enhancement module with deep channel prior[J]. IEEE Transactions on Intelligent Vehicles，2024，9(11)：7208-7221.

[6] Sayed M, Brostow G. Improved handling of motion blur in online object detection[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2021: 1706-1716.

[7] Lichtsteiner P, Posch C, Delbruck T. A 128×128 120 dB 15 μs latency asynchronous temporal contrast vision sensor[J]. IEEE Journal of Solid-State Circuits, 2008, 43(2): 566-576.

[8] Peng Yansong, Li Hebei, Zhang Yueyi, et al. Scene adaptive sparse transformer for event-based object detection[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2024: 16794-16804.

[9] Gehrig D, Scaramuzza D. Low-latency automotive vision with event cameras[J]. Nature, 2024, 629(8014): 1034-1040.

[10] Chen Li, Wu Penghao, Chitta K, et al. End-to-end autonomous driving: challenges and frontiers[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 10164-10183.

[11] Alonso I, Murillo A C. EV-SegNet: semantic segmentation for event-based cameras[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2020: 1624-1633.

[12] Sun Zhaoning, Messikommer N, Gehrig D, et al. ESS: learning event-based semantic segmentation from still images[C]//Computer Vision-ECCV 2022. Cham: Springer, 2022: 341-357.

[13] Messikommer N, Gehrig D, Loquercio A, et al. Event-based asynchronous sparse convolutional networks[C]//Computer Vision-ECCV 2020. Cham: Springer, 2020: 415-431.

[14] Biswas S D, Kosta A, Liyanagedera C, et al. HALSIE: hybrid approach to learning segmentation by simultaneously exploiting image and event modalities[C]//2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ, USA: IEEE, 2024: 5952-5962.

[15] Gallego G, Delbrück T, Orchard G, et al. Event-based vision: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(1): 154-180.

[16] 董成荣,姚俊萍,李晓军,等. 面向分布式复杂数据样本的联邦语义分割方法综述[J]. 计算机应用研究, 2024, 41(6): 1610-1617.
Dong Chengrong, Yao Junping, Li Xiaojun, et al. Survey on federated semantic segmentation methods for distributed complex data samples[J]. Application Research of Computers, 2024, 41(6): 1610-1617.

[17] Rebecq H, Ranftl R, Koltun V, et al. High speed and high dynamic range video with an event camera[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(6): 1964-1980.

[18] Wang Lin, Kim T K, Yoon K J. Joint framework for single image reconstruction and super-resolution with an event camera[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(11): 7657-7673.

[19] 孙先涛,闻勇,陈文杰,等. 基于语义分割与旋转目标检测的机器人抓取位姿估计[J]. 控制与决策, 2024, 39(9): 2913-2922.
Sun Xiantao, Wen Yong, Chen Wenjie, et al. Robot-grasping pose estimation based on semantic segmentation and rotating target detection[J]. Control and Decision, 2024, 39(9): 2913-2922.

[20] Yao Zhen, Chuah M C. Event-guided low-light video semantic segmentation[C]//2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ, USA: IEEE, 2025: 3330-3341.

[21] Chen Guang, Cao Hu, Conradt J, et al. Event-based neuromorphic vision for autonomous driving: a paradigm shift for bio-inspired visual sensing and perception[J]. IEEE Signal Processing Magazine, 2020, 37(4): 34-49.

[22] Zheng Xu, Liu Yexin, Lu Yunfan, et al. Deep learning for event-based vision: a comprehensive survey and benchmarks[EB/OL]. (2024-04-11)[2026-03-26]. https://doi.org/10.48550/arXiv.2302.08890.

[23] Zhang Rui, Leng Luziwei, Che Kaiwei, et al. Accurate and efficient event-based semantic segmentation using adaptive spiking encoder-decoder network[J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(5): 9326-9340.

[24] 于男男,王超毅,乔羽,等. 基于事件相机的图像语义分割方法[J]. 计算机辅助设计与图形学学报, 2025, 37(9): 1560-1572.
Yu Nannan, Wang Chaoyi, Qiao Yu, et al. Event-based image semantic segmentation[J]. Journal of Computer-Aided Design & Computer Graphics, 2025, 37(9): 1560-1572.

[25] Shedligeri P, Mitra K. High frame rate optical flow estimation from event sensors via intensity estimation[J]. Computer Vision and Image Understanding, 2021: 103208.

[26] Jing Linglin, Ding Yiming, Gao Yunpeng, et al. HPL-

ESS: hybrid pseudo-labeling for unsupervised event-based semantic segmentation[C]//2024 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Piscataway, NJ, USA: IEEE, 2024: 23128-23137.

[27] Maqueda A I, Loquercio A, Gallego G, et al. Event-based vision meets deep learning on steering prediction for self-driving cars[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 5419-5427.

[28] Zhu Zihao, Yuan Liangzhe, Chaney K, et al. EV-FlowNet: self-supervised optical flow estimation for event-based cameras[C]//Robotics: Science and Systems XIV. Pittsburgh, Pennsylvania: Robotics: Science and Systems Foundation, 2018: 1-9.

[29] Jia Zexi, You Kaichao, He Weihua, et al. Event-based semantic segmentation with posterior attention[J]. IEEE Transactions on Image Processing, 2023, 32: 1829-1842.

[30] Zhao Yucheng, Lyu Gengyu, Li Ke, et al. ESEG: event-based segmentation boosted by explicit edge-semantic guidance[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(10): 10510-10518.

[31] Xie Bochen, Deng Yongjian, Shao Zhanpeng, et al. EISNet: a multi-modal fusion network for semantic segmentation with events and images[J]. IEEE Transactions on Multimedia, 2024, 26: 8639-8650.

[32] Long Xianlei, Zhu Xiaxin, Guo Fangming, et al. SLT-Net: efficient event-based semantic segmentation with spike-driven lightweight transformer-based networks [C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, NJ, USA: IEEE, 2025: 4331-4338.

[33] Gu Xusheng, Qiu Changjie, Lin Xiuhong, et al. DSTF: dual-stream spatio-temporal fusion network for event-based data[C]//Pattern Recognition and Computer Vision. Singapore: Springer, 2025: 111-125.

[34] Gu Fuqiang, Li Yuanke, Long Xianlei, et al. MambaSeg: harnessing mamba for accurate and efficient image-event semantic segmentation[C]//Proceedings of the Fortieth AAAI Conference on Artificial Intelligence. Washington, DC: AAAI Press, 2026: 4302-4310.

[35] Wang Lin, Chae Y, Yoon S H, et al. Evdistill: asynchronous events to end-task learning via bidirectional reconstruction-guided cross-modal knowledge distillation[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2021: 608-619.

[36] Li Ke, Lyu Gengyu, Chen Hao, et al. Know where you are from: event-based segmentation via spatio-temporal propagation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(5): 4806-4814.

[37] Deng Yongjian, Chen Hao, Liu Hai, et al. A voxel graph CNN for object classification with event cameras[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 1162-1171.

[38] Silva D A, Smagulova K, Elsheikh A, et al. A recurrent YOLOv8-based framework for event-based object detection [J]. Frontiers in Neuroscience, 2025, 18: 1477979.

[39] Lagorce X, Orchard G, Galluppi F, et al. HOTS: a hierarchy of event-based time-surfaces for pattern recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(7): 1346-1359.

[40] Sironi A, Brambilla M, Bourdis N, et al. HATS: histograms of averaged time surfaces for robust event-based object classification[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 1731-1740.

[41] Low W F, Sonthalia A, Gao Zhi, et al. Superevents: towards native semantic segmentation for event-based cameras[C]//International Conference on Neuromorphic Systems 2021. New York, USA: ACM, 2021: 1-8.

[42] Bolten T, Pohle-Fröhlich R, Tönnies K. Instance segmentation of event camera streams in outdoor monitoring scenarios[C]//Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. Rome, Italy: SCITEPRESS - Science and Technology Publications, 2024: 452-463.

[43] Fang Huachen, Wu Jinjian, Hou Qibin, et al. Fast window-based event denoising with spatiotemporal correlation enhancement[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47 (3): 1381-1394.

[44] Hamaguchi R, Furukawa Y, Onishi M, et al. Hierarchical neural memory network for low latency event processing[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2023: 22867-22876.

[45] Al-Huda Z, Peng Bo, Yang Yan, et al. Object scale selection of hierarchical image segmentation using reliable regions[C]//2019 IEEE 14th International Conference on Intelligent Systems and Knowledge Engineer-

ing (ISKE). Piscataway, NJ, USA: IEEE, 2019: 1081-1088.

[46] Binas J, Neil D, Liu S C, et al. DDD17: end-to-end DAVIS driving dataset[EB/OL]. (2017-11-04)[2026-03-26]. https://doi.org/10.48550/arXiv.1711.01458.

[47] Gehrig M, Aarents W, Gehrig D, et al. DSEC: a stereo event camera dataset for driving scenarios[J]. IEEE Robotics and Automation Letters, 2021, 6(3): 4947-4954.

[48] Liu Xin, Tian Su, Tao Fei, et al. A review of artificial neural networks in the constitutive modeling of composite materials[J]. Composites Part B: Engineering, 2021, 224: 109152.

[49] Peng Yansong, Zhang Yueyi, Xiong Zhiwei, et al. Get: group event transformer for event-based vision[C]// 2023 IEEE/CVF International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2023: 6015-6025.

[50] Weng Wenming, Zhang Yueyi, Xiong Zhiwei. Event-based video reconstruction using transformer[C]// 2021 IEEE/CVF International Conference on Computer Vision. Piscataway, NJ, USA: IEEE, 2021: 2543-2552.

[51] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.

[52] Zhang Yitong, Wei Yingmei, Guo Yanming, et al. Exploiting event temporal dynamics and sparsity characteristics for RGB-event fusion semantic segmentation [C]//Proceedings of the 2025 International Conference on Multimedia Retrieval. NewYork, USA: ACM, 2025: 1831-1839.

[53] Hareb D, Martinet J, Miramond B. Enhanced neuromorphic semantic segmentation latency through stream event[C]//Neural Information Processing. Singapore: Springer, 2025: 46-60.

[54] Hareb D, Martinet J. EvSegSNN: neuromorphic semantic segmentation for event data[C]//2024 International Joint Conference on Neural Networks (IJCNN). Piscataway, NJ, USA: IEEE, 2024: 1-8.

[55] Papa L, Russo P, Amerini I, et al. A survey on efficient vision transformers: algorithms, techniques, and performance benchmarking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 7682-7700.

[56] Tay Y, Dehghani M, Bahri D, et al. Efficient transformers: a survey[J]. ACM Computing Surveys, 2022, 55(6): 1-28.

[57] Li Hebei, Peng Yansong, Yuan Jiabei, et al. Efficient event-based semantic segmentation via exploiting frame-event fusion: a hybrid neural network approach [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2025, 39(17): 18296-18304.

[58] Wang Zuowen, Hu Yuhuang, Liu S C. Exploiting spatial sparsity for event cameras with visual transformers [C]//2022 IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE, 2022: 411-415.

[59] Yuan Jingyang, Gao Huazuo, Dai Damai, et al. Native sparse attention: hardware-aligned and natively trainable sparse attention[C]//Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Vienna, Austria: Assaciation for Computationa Linguistics, 2025: 23078-23097.

[60] Dao T, Gu A. Transformers are SSMs: generalized models and efficient algorithms through structured state space duality[C]//ICML'24: Proceedings of the 41st International Conference on Machine Learning. Vienna: PMLR, 2024: 10041-10071.

[61] Long Xianlei, Zhu Xiaxin, Guo Fangming, et al. SpikeBRGNet: efficient and accurate event-based semantic segmentation with boundary region-guided spiking neural networks[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(3): 2712-2724.

[62] Wu Zehao, Xia Yuanqing, Hu Rui. A modified automatic label generation method under low-light environments for event-based semantic segmentation[J]. Neurocomputing, 2025, 638: 130136.

[63] Lobo J L, Del S J, Bifet A, et al. Spiking neural networks and online learning: an overview and perspectives[J]. Neural Networks, 2020, 121: 88-100.

[64] Yamazaki K, Vo-Ho V K, Bulsara D, et al. Spiking neural networks and theirapplications: a review[J]. Brain Sciences, 2022, 12(7): 863.

[65] Su Qiaoyi, He Weihua, Wei Xiaobao, et al. Multi-scale full spike pattern for semantic segmentation[J]. Neural Networks, 2024, 176: 106330.

[66] Yao Zhen, Ying Xiaowen, Chuah M C. Rethinking RGB-event semantic segmentation with a novel bidirectional motion-enhanced event representation[EB/OL]. (2025-05-02)[2026-03-26]. https://arxiv.org/html/2505.01548v1.

[67] Xie Chuyun, Gao Wei, Guo Ren. Cross-modal learning for event-based semantic segmentation via attention

soft alignment[J]. IEEE Robotics and Automation Letters,2024,9(3):2359-2366.

[68] Liu Yuhan, Deng Yongjian, Chen Hao, et al. Event-based video frame interpolation with edge guided motion refinement[EB/OL]. (2024-04-28)[2026-03-26]. https://doi.org/10.48550/arXiv.2404.18156.

[69] Cho H, Yoon S H, Kweon H, et al. Finding meaning in points: weakly supervised semantic segmentation for event cameras[C]//Computer Vision-ECCV 2024. Cham: Springer, 2025: 266-286.

[70] Annamalai L, Thakur C S. EventASEG: an event-based asynchronous segmentation of road with likelihood attention[J]. IEEE Robotics and Automation Letters, 2024, 9(8): 6951-6958.

[71] Hoyer L, Dai Dengxin, Van G L. Daformer: improving network architectures and training strategies for domain-adaptive semantic segmentation[C]//2022 IEEE/CVF conference on computer vision and pattern recognition. Piscataway, NJ, USA: IEEE, 2022: 9914-9925.

[72] Kim Y, Chough J, Panda P. Beyond classification: directly training spiking neural networks for semantic segmentation[J]. Neuromorphic Computing and Engineering, 2022, 2(4): 044015.

Research on Reliable Region-Guided Learning for Event Semantic Segmentation

KONG Xinyan, YAO Junping*, LI Xiaojun, CHENG Peng

(Rocket Force University of Engineering, Xi'an 710025, China)

ABSTRACT: Event cameras feature a high temporal resolution, wide dynamic range, and low latency, providing effective edge and temporal information for semantic segmentation in complex illumination and high-speed motion scenarios. Focusing on reliable region-guided learning in event semantic segmentation, this study defined relevant concepts and task characteristics, and systematically reviewed typical methods based on event-active, edge-salient, and high-confidence prediction regions from the perspectives of spatiotemporal encoding representations of event data and sources of reliability evidence. Consequently, the performances and technical characteristics of related models were comparatively analyzed using common datasets and evaluation metrics, and the progress of the research on sparse event representation, reliable region selection, boundary preservation, and prediction consistency modeling was summarized. This review further identified several issues in existing studies, such as single-source reliability evidence, the inadequate exploitation of regional complementary relationships, and the limited adaptability to dynamic scenarios. Finally, future research directions were discussed in terms of sparse spatiotemporal modeling, boundary measurement and prior construction, multi-region collaborative guidance, and dynamic reliable region modeling, with the goal of providing a reference for research on reliable region-guided learning in event semantic segmentation.

KEYWORDS: event semantic segmentation; reliable region-guided; event active region; edge salient region; high confidence prediction region

(编辑:于福兴)