arXiv:2304.04220v1 [cs.CV] 9 Apr 2023

面向足球视频动作定位的主动学习

Silvio Giancola*, 1 Anthony Cioppa*, 1, 2 JuliaGeorgieva[3] Johsan Billingham[4]Andreas Serner[5] Kerry Peek[6] Bernard Ghanem[1] MarcVan Droogenbroeck[2]IVUL、KAUST [2]TELIM、Li`ège 大学 [3]
科廷大学科廷联合健康学院[4]国际足联足球技术与创新部 [5]国际足联医疗部 [6]悉尼大学医学与健康学院

摘要

足球是一项复杂且动态的运动，每场比赛中都有许多动作同时发生。分析足球视频颇具挑战，需要识别细微且多样的时空模式。尽管计算机视觉近年来取得了进展，但当前算法在利用有限的标注数据进行学习时仍面临重大挑战，导致其检测这些模式的性能下降。本文提出一种主动学习框架，用于选择下一步最值得标注的视频样本，从而大幅减少标注工作量，并加快动作定位模型的训练，使其更快达到最高准确率。我们的方法利用不确定性采样的理念，选择最具挑战性的视频片段用于下一轮训练，从而加快算法的学习过程。我们证明，所提出的主动学习框架能够有效减少在足球视频中准确进行动作定位所需的训练数据。在 *SoccerNet-v2* 上使用 *NetVLAD++*进行动作定位时，仅使用三分之一的数据集便取得了相近的性能，这表明该方法在减少标注时间和提高数据效率方面具有显著潜力。我们还在两个新的数据集上进一步验证了该方法，这些数据集侧重于对头球和传球动作进行时间定位，证明了该方法对足球中不同动作语义的有效性。我们相信，用于动作定位的主动学习框架将助力动作定位算法的进一步应用，并加快体育领域的标注活动。

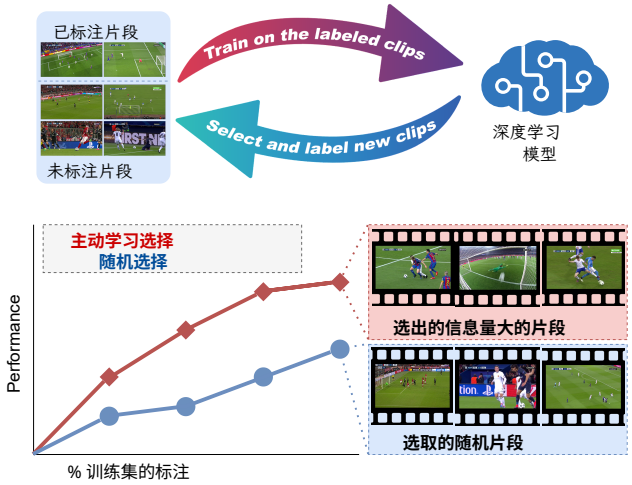


图1。**用于动作定位的主动学习**。给定一个视频片段数据集，我们的主动学习框架会迭代地（**i**）在已标注片段上训练深度学习模型，并（**ii**）选择接下来由预言机进行标注的视频片段。通过主动选择信息量最大的下一个视频样本进行标注，我们加速了用于训练动作定位模型的未标注数据集标注过程。

1. 引言

视频分析是一个快速发展的领域，在监控 [50],体育 [53], 和自动驾驶 [39]等各个领域有着众多应用。视频分析中的一项基本任务是动作定位，其目标是识别视频序列中以单个时间戳为锚点的特定动作，并对其进行精确定位。近年来，这项任务因其在视频搜索 [15, 48], 、视频摘要 [7, 13], 和活动识别 [3, 25]等多种应用中的重要性而备受关注。

传统上，动作定位采用监督学习技术来实现，即使用标注数据集训练分类器，使其能够识别视频中的动作并在时间上对其进行定位。然而，为大规模视频数据集添加标注既耗时又昂贵，限制了监督学习方法的可扩展性和适用性。主动学习是一种有前景的方法，它可以通过选择进行标注所需的最具信息量的样本，减少对大型数据集进行完整标注的需求。

本文提出了一种主动学习框架

(*) 表示贡献相同 联系方式:
silvio.giancola@kaust.edu.sa、anthony.cioppa@uliege.be。

针对足球视频中的动作定位，旨在减少标注工作量，并根据标注数量提升系统的整体性能。我们的框架如图 1所示，将主动学习与最先进的动作定位方法相结合，通过从未标注数据集中选择具有信息量的样本，迭代扩充标注集。我们在多个基准数据集上评估了该方法，并将其与未利用主动学习的朴素随机选择方法进行了比较。我们还分析了不同查询策略对系统性能的影响。

贡献。 我们的贡献概括如下： **(i)** 我们提出了首个用于动作定位任务的主动学习框架，该框架会迭代地选择相关片段以进行下一轮标注。 **(ii)** 我们在多个基准数据集上，将基于不确定性采样的几种主动学习选择策略与最先进的动作定位方法进行了比较。 **(iii)**我们提供了全面的分析，展示了该框架能够显著减少达到预期性能所需的标注量。

2. 相关工作

2.1. 体育视频理解

体育视频分析具有挑战性且内涵丰富，近年来已逐渐成为热门研究课题 [36, 53]。大规模数据集的出现对推动这些任务的发展起到了至关重要的作用。此类大规模数据集的例子包括 Pappalardo等人开发的 [38], Yu 等人开发的 [64], Soc-cerDB [29], SoccerTrack [42], 和 DeepSportRadar [57]。Giancola等人提出的 [17], SoccerNet 数据集已成为足球视频理解领域最全面的标注数据资源。它包含十项不同任务的基准，例如动作定位 [11], 、摄像机校准 [6],和球员跟踪 [8]。

近年来，基于深度学习的方法凭借出色的性能以及从原始数据中提取高层特征的能力，已成为许多体育视频分析任务的首选方法。例如，基于深度学习的自动化方法在球员跟踪 [35] 、存在遮挡时的重识别 [49], 3、羽毛球中的二维羽毛球轨迹重建 [34], 、橄榄球中的医疗风险评估 [37],、战术分析 [52], 、传球可行性评估 [1], 以及人才选拔 [10]等任务中均取得了令人瞩目的成果。此外，基于深度学习的方法还使研究人员能够有效利用大规模数据集。尽管取得了这些成功，基于深度学习的方法仍面临若干挑战，例如处理含噪和不完整的数据、考虑复杂的比赛场景，以及泛化到新的领域。因此，该领域仍有很大的改进和研究空间。

体育视频理解。此外，为体育视频分析任务标注大规模数据集是一个耗时且昂贵的过程，需要大量人力资源和专业知识。作为一种解决方案，Vandeghen 等. [58] 提出了一种用于球员检测的半监督方法，利用了一个大型未标注数据集。Vats 等. [61]提出了一种使用 Transformer 进行球员识别的弱监督方法。为解决这一数据问题，我们提出了一种用于足球动作定位的主动学习方法，旨在减少所需的标注数据量，同时保持较高的任务性能。

2.2. 动作定位

动作定位是足球视频理解中的一项重要任务，它需要在未经剪辑的足球转播视频中定位特定事件，例如点球、进球或角球。不同于时间活动定位， [3], 动作定位使用单个时间戳描述事件，这遵循了足球规则中对动作的定义 [27]。近期研究探索了使用大规模数据集的方式，例如 SoccerNet [17], , 该数据集已从 3 个类别扩展至 17 个类别，以涵盖比赛中可能发生的所有动作 [11]。这一数据集引起了研究界的广泛关注，公开挑战赛 [18], 便是明证；这表明，当前动作定位在研究界和工业界都受到高度关注，相关研究也十分活跃。

Giancola 等人提出了首个动作定位方法 [17], , 该方法基于时间池化。之后，他们通过聚合时间上下文 [19]对该方法进行了改进。Rongved 等人 [41]提出了一种方法：以 5 秒为滑动窗口，直接将 3D 残差网络应用于视频帧。Vanderplaetse 等人 [59] 和 Xarles 等人 [18] 将视觉特征与音频特征相结合，采用了多模态方法。Cioppa等人 [7] 引入了上下文感知损失函数，以建模动作周围的时间上下文。Vats 等人 [60]使用多塔卷积神经网络来考虑动作位置的不确定性；Tomei等人 [55] 则对特征提取器进行微调，并采用掩码策略，将注意力集中在动作发生后的帧上。

目前，SoccerNet-v2 上的最先进成果由 Soares 等人取得。 [46, 47], 他们通过提出一种基于锚点的方法赢得了 SoccerNet 2022 挑战赛。该方法将锚点定义为由一个时间瞬间和一个动作类别组成的对，其中时间瞬间经过密集采样。对于每个锚点，模型都会推断检测置信度和细粒度的时间位移，其中时间位移表示预测动作确切发生的时间。他们的方法显著提升了时间精度。Hong 等人 [24], 作为 2022 年挑战赛的亚军，提出了首个精确时间定位（PTS）方法，在该方法中，特征提取器和定位

头部均以端到端方式进行训练。他们采用轻量级的 RegNet 架构，包括一个 GSM [51]模块和一个 GRU [5] 模块，并在其之上将每一帧分类为某个动作类别或背景。其他方法也专注于时空编码器，采用 [9], 基于图的架构 [4], 或基于 Transformer 的架构 [65]。

尽管表现令人印象深刻，所有最先进的方法都依赖监督学习，而监督学习需要大规模的标注数据集。然而，在体育视频分析中，需要识别的动作可能会随时间发生变化，这就需要重新标注数据集。在这项工作中，我们研究了如何利用主动学习技术高效地重新标注此类数据集。主动学习旨在选择最具信息量的样本进行标注，从而在保持较高任务性能的同时，最大限度地减少标注工作量。

2.3. 主动学习

主动学习已成功应用于广泛的领域，包括图像理解 [16,26], 视频理解 [22], 自然语言处理 [54], 语音识别 [21], 和化学 [12]。主动学习的主要目标是选择信息量最大的未标注样本进行标注，并使用尽可能少的标注数据来达到特定性能。主动学习的主要策略包括不确定性采样 [30, 33, 56], 多样性最大化 [43,62], 委员会查询 [14, 20, 28, 45], 和 期望误差 [23,31,32,63]。关于主动学习的全面且更通用的文献综述，请参阅 [40,44] 。

不确定性采样。 这些方法对迄今训练的大多数动作定位模型最容易产生混淆的未标注数据进行采样。Tong等人提出使用支持向量机算法，为图像检索执行有效的相关性反馈。Joshi等人提出的主动学习方法会选择模型认为最难分类的未标注数据。该选择基于分类器输出的熵，或基于“最佳与次佳”（BvSB）范式。

多样性最大化。 这些主动学习方法选择能够最佳代表整个可用未标注数据集空间的样本。Yang 等. [62] 提出了一种最大化样本多样性的方法。他们在多种视觉识别任务上研究了这一方法，包括动作识别、物体分类、场景识别和事件检测。同样，Sener 等. [43]将选择过程建模为核心集问题。他们通过最小化图像与数据集中剩余样本之间的 L2 距离，抽取具有代表性的图像子集。

委员会式查询。在这种范式中，根据在同一标注数据集上训练的一组学生算法之间分歧最大的原则，选择下一批待标注的数据。Seung 等人 [45] 引入了这种方法，并在贝叶斯框架中进一步分析了它——Freund的研究等人. [14]。Houlsby 等人. [26] 进一步研究了委员会查询与信息增益理论之间的关系。

预期错误。 这些方法试图学习一种与特定样本分类错误相关的度量。学习主动学习（LAL） [32] 旨在回归候选样本所带来的错误减少量。主动学习得分通过监督方式基于训练数据集中的错误损失进行学习。同样，Yoo 等人 [63] 也学习一个与任何任务无关的“损失预测模块”。通过这样做，他们会主动选择预测损失较高的样本，期望这些样本能提供显著的新信息，以便在下一训练步骤中将损失降至最低。

面向时间视频分析的主动学习。虽然主动学习已在通用设置上得到广泛分析，但只有少数工作将这些方法应用于时间视频分析。Brandla 等。 [2]提出了一种用于时间活动定位（TAL）算法的主动学习方法，该方法基于不确定性采样 [33]。Heilbron等。 [22]通过对不同主动学习范式进行实证研究，进一步探讨了TAL中的主动学习，其中LAL [32]表现最佳。

遵循先前的文献，我们正式提出了首个用于动作定位的主动学习工作流。我们分析了几种不确定性采样 方法，为更多用于动作定位的主动学习方法奠定基础。

3. 用于动作定位的主动学习

我们提出了首个用于动作定位任务的主动学习框架。我们的框架旨在使用最少量的标注数据训练出准确的动作定位模型。根据不确定性采样主动学习领域的文献，我们确定了实现这一目标的三个关键步骤：**(1)** 在一个标注数据集上训练动作定位模型，该数据集在每个主动学习步骤中不断扩充。**(2)** 使用主动学习算法从未标注数据池中选择信息量最大的 数据。**(3)** 由预言机标注选定片段，并将新数据和标注纳入标注集。我们的完整框架概览如图2所示。

形式上，给定一个视频 v，动作定位任务是识别该视频中的所有动作定位结果 $\mathbf{S} = \{s_1, ..., s_M\}$。一个定位结果 s_m 由动作类别（例如，点球、进球等）和一个时间锚点组成。在每个主动学习步骤τ中，一个动作定位模型（其推理函数记为 f_τ）使用一组带标注样本 $\mathcal{L}_\tau$进行训练。模型细节和训练过程将在第 3.1节中介绍。随后，一个主动学习算法 g 从样本池 $\mathbf{C}^*$ 中选择一组最优的未标注样本 $\mathcal{U}_\tau$。第 3.2节介绍了几种主动学习算法。随后，第3.3节所述的预言机提供动作定位结果的真实标注（即，动作

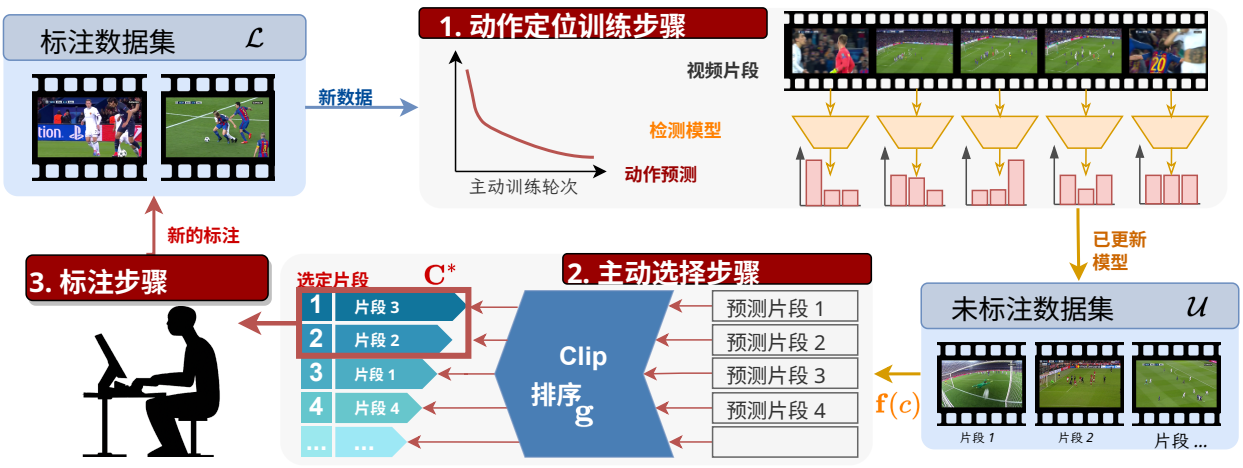


图 2. **用于动作定位的主动学习流程**。我们从一个小型标注数据集 $\mathcal{L}$ 开始，并在其上训练一个动作定位模型，其推理函数记为 f。使用训练好的模型，我们从未标注数据集 $\mathcal{U}$ 中选择接下来要标注的样本。为此，我们首先收集模型对每个片段的预测结果 $f(\mathcal{U})$，并将预测结果传入我们的选择函数 g，该函数会对片段进行排序，以选出 $\mathbf{C}^*$。随后，将所有选定片段交给预言机（人工标注员），由其标注该片段中所有动作的类别和位置。然后，将这些新的标注数据添加到标注数据集中，用于下一轮训练。重复这一过程，直到达到所需的性能，或未标注数据集为空。

类别 k 和时间锚点 t)用于所选样本$\mathbf{C}^*$。随后，标注集 $\mathcal{L}_\tau$ 会根据 $\mathcal{L}_{\tau+1} = \mathcal{L}_\tau \cup \mathbf{C}^*$，加入这些新标注的片段实例 $\mathbf{C}^*$。这一过程会重复进行，直到模型达到期望的性能或集合 $\mathcal{U}_\tau$ 耗尽。由于最耗时的步骤是对样本进行标注，因此我们框架的目标是通过提出一种高效的主动选择算法，尽量减少查询预言机的次数。

3.1. 模型训练步骤

数据集。 动作定位的数据集通常由一组 L 未剪辑视频组成 $\mathbf{V} = \{v_1, ..., v_L\}$，每个视频都标注有一组 $\mathbf{S}$ 动作定位结果 $s = \{k, t\}$，其类别为 k 个类别中的 K 一类，并以单个时间戳 t作为锚点。由于硬件限制（例如 GPU 内存或计算时间），目前尚无法在较长的未剪辑视频上训练动作定位模型，因此通常使用从视频中提取的片段进行训练。在本文中，我们认为每个视频 v_l 都可以视为一组 N固定长度且不重叠的片段$\mathbf{C_l} = \{c_l^1, ..., c_l^N\}$。每个片段 c_l^n 都可以标注一个按时间锚定的动作定位结果列表 $\mathbf{S}^{\mathbf{n}}_{\mathbf{l}} = \{s^n_{l,1}, ..., s^n_{l,M}\}$。

视频编码器。 典型的动作定位模型由视频编码器 $\mathbf{H}$ 和动作定位头 $\mathbf{A}$组成。给定由 J 帧组成的裁剪后的视频片段 c_l^n，视频编码器会为每帧提取紧凑的特征表示 $\mathbf{H}(\mathbf{c}^{\mathbf{n}}_{\mathbf{l}}) = \{h^n_{l,1}, ..., h^n_{l,J}\}$。这种帧特征编码器通常在外部数据集上预训练，然后在动作定位数据集上冻结或微调。由于特征维度各不相同，通常会使用主成分分析统一其维度，从而得到更加紧凑且标准化的帧表示。此外，这些特征编码器既可以应用于整个视频片段，以利用时间信息；也可以独立应用于片段中的每一帧，后者通常所需的计算能力和内存更少。动作定位基线模型通常会使用可学习的、基于卷积神经网络的编码器提取帧特征，例如基于帧的 ResNet 编码器，或基于视频的 I3D/C3D 编码器 [17]。

动作定位头。 给定一组紧凑的帧特征表示 $\mathbf{H}(\mathbf{c}^{\mathbf{n}}_{\mathbf{l}}) = \{h^n_{l,1}, ..., h^n_{l,J}\}$，动作定位头 $\mathbf{A}$ 会在时间维度上组合描述符，并为当前片段 c_l^n输出预测动作定位结果列表 $\hat{}\mathbf{S}^{\mathbf{n}}_{\mathbf{l}} = \mathbf{A}(\mathbf{H}(\mathbf{c}^{\mathbf{n}}_{\mathbf{l}})) = \{\hat{}s^n_{l,1}, ..., \hat{}s^n_{l,M'}\}$。这些预测结果可以通过两种方式获得。第一类动作定位模型 [7] 直接回归预测位置和类别。在这项工作中，我们聚焦于第二类模型，它首先输出每帧或每个片段的类别分数（包括背景）。随后，使用非极大值抑制算法对随时间变化的预测类别分数进行处理，提取动作的精确定位 t。因此，动作定位方法 f_τ的完整数学函数可表示为 $\mathbf{A} \circ \mathbf{H}$。

训练。 我们将主动学习步骤 n 中大小为 $|\mathcal{L}_\tau|$ 的标注数据集 $\mathcal{L}_\tau$ 定义为 $\mathcal{L}_\tau = \{(c_1^{train}, \mathbf{S}_1)..... (c_{|\mathcal{L}_\tau|}^{train}, \mathbf{S}|\mathcal{L}_\tau|)\}$。在每个主动学习步骤中，动作定位基线模型都在 $\mathcal{L}_\tau$ 上进行训练。在我们的框架中，我们考虑了几种训练范式。第一种是在每个主动学习步骤中从头开始训练动作定位模块

。这种训练可以持续到收敛，也可以训练特定数量的轮次。其优点是，深度学习模型通常能因此得到更好的训练。然而，缺点是每轮训练可能都要耗费大量时间。一种更快的训练范式是对上一个主动学习步骤中得到的模型进行微调，直到收敛或训练固定数量的轮次。这样可以缩短训练时间，但无法确保模型收敛。例如，如果网络在最初几个步骤中由于训练数据量不足而发散，之后可能就无法恢复。我们将在实验部分研究这些训练范式。

推理。在测试时，模型会对完整视频生成预测结果，而它仅使用片段进行训练。解决这种不匹配的一种常见方法是将视频拆分为有重叠或无重叠的片段。每个片段都会被独立处理，结果则沿视频进行聚合。动作定位性能通过平均精度均值（mAP）进行评估，指标依据在给定时间容差内成功定位动作 δ。主要的相关指标是 Avg-mAP，即对不同取值下的 mAP 进行平均，δ表示真实动作定位结果与预测动作定位结果之间的时间容差。我们采用典型的指标 [18]紧Avg-mAP（其中 δ 范围从1到5 秒）和宽松Avg-mAP（其中 δ 范围从5 到60 秒）。

3.2. 主动选择步骤

下一步包括从未标注数据集（定义为一组未标注视频片段）中选择片段$\mathcal{U}_\tau = \{c_1^u, ..., c_{|\mathcal{U}_\tau|}^u\}$，其大小为 $|\mathcal{U}_\tau|$。主动选择步骤的目标是创建一个函数 g，该函数从 $\mathbf{C}^*$ 中选择一组新的片段 $\mathcal{U}_\tau$。主要挑战在于确保该函数选择的样本能够对改进动作定位模型产生最大的影响。如第 2 节所述，存在许多主动学习工作流。在本文中，我们聚焦于 不确定性采样这一特定情况。具体而言，我们分析在前一个主动学习步骤中训练的动作定位模型的预测结果。预测结果是每个类别对应的一组 $(K+1)$概率值，可以按帧或按片段给出。对于黑盒模型，类别置信度分数是预测结果返回的唯一不确定性信息。遵循图像分类主动学习领域的相关研究，我们利用不确定性度量（UM）和熵度量（EM）构建了两个选择器。下面我们将展示如何针对动作定位预测结果实现这些方法。

不确定性度量。不确定性度量（UM）仅考虑与每个片段或帧相关联的置信度分数。给定置信度分数 p_k，主动学习分数与其到混淆置信度0。5 的距离成反比。熵度量在形式上定义如下：

$$\mathrm{UM} = 1 - 2 \times |p_k - 0.5| \,. \tag{1}$$

在动作定位这一特定情况下，该分数按帧计算，然后在所有帧上取平均值或进行最大池化。

熵度量。熵度量（EM）考虑所有类别的置信度分布。这种估计需要获取所有类别的置信度分数，而对于黑盒算法来说可能无法获取，因为这类算法只返回所选类别的最高置信度。基于置信度分数列表p_1，..... p_K，我们提取出一个主动学习分数，该分数与预测结果之间分布的均匀性成反比。熵度量正式定义如下：

$$\mathrm{EM} = -\sum_{i=1}^{K} p_i \log(p_i) \,. \tag{2}$$

选择样本。我们利用函数 g 来选出主动评分最高的 top-k 个信息量最大的片段$\mathbf{C}^*$。在这项工作中，我们研究了几种在每个主动学习步骤中选择片段数量$\mathbf{C}^*|$ 的方法。第一种方法是在每个主动学习步骤中选择固定数量的片段。第二种方法则是逐渐增加片段数量。这样做的优点是，在开始阶段模型仍需要信息量很大的片段时，只选择相关片段。

3.3. 标注步骤

一旦主动学习步骤选出片段，就需要由预言机（在实际场景中即人工标注员）进行标注，并提供真实动作定位结果。$\mathbf{C}^*$ 中的片段由人工标注，随后将片段及其对应的标注添加到 $\mathcal{L}_\tau$ 中。在被动学习设置中，预言机通常会在 $\mathcal{U}_\tau$ 中随机选择几个片段，这可能会标注出冗余信息。本文表明，我们的主动学习框架能够选出相关片段，相较于随机选择提升性能，从而节省时间和金钱。

4. 实验

4.1. 实验设置

我们的主动学习框架不依赖于数据集、动作定位训练参数和主动学习选择算法。本节将介绍我们在各种设置下开展实验的技术细节。

数据集。 在本研究中，我们利用三个数据集来评估用于足球视频动作定位的主动学习框架：SoccerNet-v2、SoccerNet-ball（公开数据集）以及

数据集	比赛	标注	类别	密度
SoccerNet-v2	550	110,458	17	2.23/分钟
SoccerNet-ball	9	11,041	2	13.62/分钟
FWWC19-header	52	6,527	1	1.39 次/分钟

表 1. **数据集**。 我们在三个用于足球视频动作定位的数据集上研究主动学习框架。

FWWC19-header（私有）。表格 1概述了每个数据集的主要特征。

SoccerNet-v2 包含 550 场比赛，标注了来自 17 个通用动作类别的 110，458 个动作定位结果，例如进球、点球、红黄牌和任意球。这些带有时间戳的标注全面呈现了足球视频中发生的各类动作。

SoccerNet-ball 包含 9 场公开比赛，标注了 11，041 个与球相关的事件，例如传球和带球推进。该数据集提供了有关球相关动作的宝贵信息，而这正是比赛中的关键方面。此外，比赛中事件的高密度要求具备精准的时间定位能力。

FWWC19-header 是一个私有数据集，包含 52 场 2019 年国际足联女子世界杯（FWWC19）比赛，针对 5 类头部冲击事件进行了标注，包括有意头球、无意头球、头球争顶、尝试头球以及其他头部冲击事件。该数据集有助于深入了解头部冲击相关事件；这类事件是足球及其他接触性运动中一个重要的医疗问题。

动作定位方法。 在本研究中，我们考察了两种动作定位基线方法，以支持我们针对足球视频中动作定位的主动学习框架：NetVLAD++ [19] 和精确时间定位 [24]。表 2概述了每种基线方法的主要特征。

基线	编码器	动作定位头	训练
NETVLAD++	ResNet152	NetVLAD	Head
PTS	ResNet18	GRU	编码器+头部

表 2. **足球视频动作定位基线**。 我们在足球视频动作定位的两个基线上研究了我们的主动学习框架，即 NetVLAD++ 和精确时间定位。

NetVLAD++ [19] 通过对时间上连续的帧特征进行池化，学习识别片段中出现的动作类别。由于特征编码器被冻结，且定位头轻量，因此采用主动学习方式训练的速度非常快。其一项主要特点是采用弱监督方式进行训练，不考虑片段中动作的精确定位，这显著加快了预言机的标注过程。然而，其缺点是动作定位的精确度较低。

精确时间定位（PTS） [24] 通过结合密集帧特征与 GRU，学习识别特定帧上是否发生了特定动作。该紧凑型编码器与 GRU 一起进行端到端训练，在 SoccerNet-v2 上取得了最先进的性能。在本研究中，我们选择了可在单块 GPU 上运行的 ResNet 18 特征编码器，因为它比配备 GRU 的 RegNet 编码器快得多。

指标。 对于动作定位，除非另有说明，我们采用宽松的 Avg-mAP [17]，。对于主动学习，我们分析动作定位性能随用于训练模型的数据比例（即标注数据集的大小）变化的学习曲线。遵循 [22]，我们估计学习曲线下面积（AULC）。优秀的主动学习器应当比随机采样器具有更高的 AULC。此外，我们还提出了两个指标：（i）$\mathcal{M}_{data}^{10\%}$：仅使用 10% 的数据时的 Avg-mAP 性能；（ii）$\mathcal{M}_{perf}^{90\%}$：达到最终 Avg-mAP 性能 90% 所需的数据比例。

技术细节。 除非另有说明，我们的实验均采用动作定位模型 NetVLAD++ [19]，使用带有 ResNET PCA512 特征的配置，并按照原始实现中定义的相同训练参数，将模型训练至收敛。在每次主动学习迭代中，我们选择数量为 $|\mathbf{C}^*|$ 的样本，相当于数据集的 1%。在每个动作定位训练步骤中，我们都从头开始重新训练。

4.2. 初始结果

我们首先将我们的框架与两种主动学习选择算法——不确定性度量（UM）和熵度量（EM）——进行比较，并以随机采样器（RS）作为对照。图3 展示了动作定位性能（宽松平均平均精度）随标注数据集大小的变化。表 3 报告了主要指标曲线下面积、 $\mathcal{M}_{data}^{\%}$ 和 $\mathcal{M}_{perf}^{\%}$。初始结果表明，我们的学习框架显著加快了训练速度。采用熵度量（EM）时，模型性能以更快的速度收敛，因此只需较少的标注数据即可达到更高的性能。具体而言，我们采用EM的设置取得了47.96%的曲线下面积指标，而随机采样（RS）为45.11%。此外， $\mathcal{M}_{perf}^{90\%}$ EM为16%，而随机采样（*RS*）为45%，这表明我们只需三分之一的数据即可达到动作定位性能的90%。最后，$\mathcal{M}_{data}^{5\%}$ 表明仅使用5%的数据时，我们即可达到37.24%的动作定位指标平均平均精度，而随机采样时为32.06%。有趣的是，与随机采样（RS）相比，不确定性度量（UM）仅带来了有限的改进。

4.3. 加速主动学习框架

本节分享一些能够加快主动学习框架运行速度的发现。具体而言，**(i)**我们为 NetVLAD++引入了一个更快的调度器，且性能相近；**(ii)** 我们引入了自适应主动学习调度器（AdapAL）；**(iii)** 我们研究了持续训练，即在每个主动学习步骤中微调模型，而非从头开始训练。

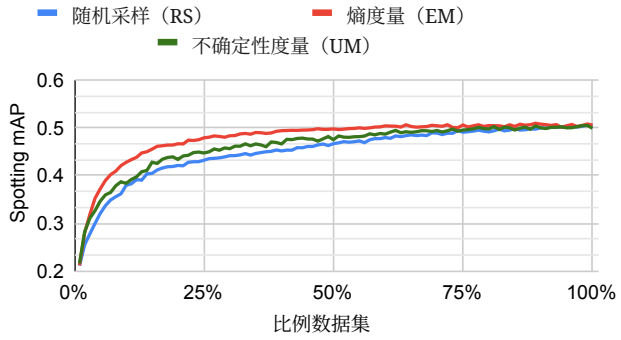


图 3. **主动学习** 对比**随机采样**。我们使用熵度量（EM）进行的不确定性采样能够以更快的速度收敛到最优解，并使用更少的数据。实际上，主动学习只需要随机采样器所需数据量的36%，即可达到相近的性能（$\mathcal{M}_{perf}^{90\%}$）；而相近数量的数据最多可带来 18% 的性能提升（$\mathcal{M}_{data}^{4\%}$）。

方法	RS	UM	EM
曲线下面积 (↑)	45.11	46.24	**47.96**
$\mathcal{M}_{perf}^{90\%}$ (↓)	45.00	31.00	**16.00**
$\mathcal{M}_{perf}^{99\%}$ (↓)	99.00	-	**64.00**
$\mathcal{M}_{data}^{5\%}$ (↑)	32.06	34.64	**37.24**
$\mathcal{M}_{data}^{10\%}$ (↑)	37.98	38.38	**42.79**

表 3. **主动学习** 对比。**随机采样**。 我们提出的基于熵度量（EM）的主动学习框架优于随机采样（RS）和不确定性度量（UM）。

更快的模型训练。 首先，我们采用了更快的学习率调度器。我们不再从 10^{-3}开始，并在每次验证损失进入平台期时按 10 的比例降低学习率，直到达到 10^{-8}，而是从 10^{-2} 开始，并将其降至 10^{-4}。此外，我们将判定平台期的耐心值从 10 个5 个训练轮次。这样一来，训练时间实际上缩短了一半，同时仍能取得与原始训练调度器相当的性能，如图4所示。

自适应主动学习调度器。 其次，我们调整主动学习（AL）步骤，逐步增加每个 AL 步骤中所选取并标注的样本数量。在该机制下，我们可以增加数据集 $\mathcal{L}_t$ 的比例，而不仅仅是数据集的 1%。在实践中，我们选择在数据集达到 15% 后增加 2%，达到 25% 后增加 5%，达到 40% 后增加 10%。这样一来，我们将 AL 步骤从 100 减少到 30，节省了 70% 的时间。图4 说明，自适应 AL 步长不会影响 NetVLAD++在 SoccerNet-v2 上的训练性能。

持续训练。第三，我们研究从上一个主动学习步骤恢复训练是否有助于缩短训练时间。停止条款-

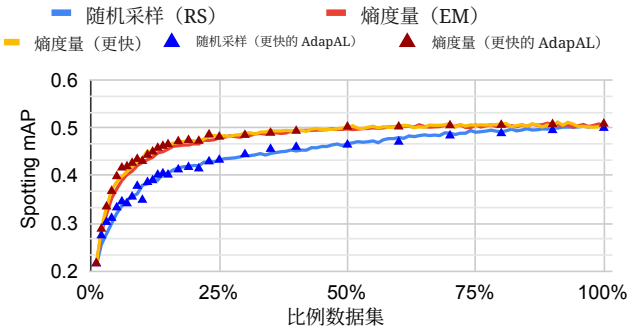


图 4。**更快的训练和自适应主动学习（Ada-pAL）范式**。我们在此展示，对于我们的实验，可以显著减少主动学习所需的时间，同时不会以任何方式降低主动学习训练的性能。

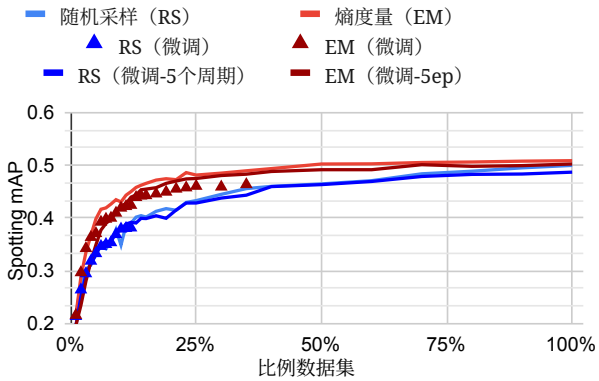


图5。**有限训练轮数的微调效果**。从训练轮数有限的模型进行微调，可以使下一步主动学习中的训练更加稳定。

主动学习设置	训练	曲线下面积 (↑)	$\mathcal{M}_{data}^{10\%}$ (↑)	$\mathcal{M}_{perf}^{90\%}$ (↓)
$RS_{1\%}$	原始。	45.11	37.98	45.00
$EM_{1\%}$	原始。	47.96	42.79	16.00
$EM_{1\%}$	fast	**48.28**	**43.68**	**14.00**
RS_{AdapAL}	fast	44.64	34.91	40.00
EM_{AdapAL}	fast	**48.01**	**43.06**	**13.00**
RS_{AdapAL}	5 集	44.13	37.78	40.00
EM_{AdapAL}	5 集。	**46.94**	**42.30**	**19.00**

表 4。**消融实验**。我们提出的基于熵度量（EM）的主动学习框架，在所有主动学习设置下均优于随机采样（RS）。

图 5 中的三角形序列表明，使用原始参数进行持续训练的朴素实现会导致损失发散，从而妨碍后续微调。相反，我们提出先进行 20 个 epoch 的引导训练，并在每个主动学习步骤中对模型微调 5 个 epoch，同时将学习率固定为 10^{-2}。我们可以看到

Data	指标	RS	EM	UM
SoccerNet-ball	曲线下面积 (↑)	40.47	42.18	**42.88**
SoccerNet-ball	$\mathcal{M}_{data}^{10\%}$ (↑)	36.55	**42.41**	41.88
SoccerNet-ball	$\mathcal{M}_{perf}^{90\%}$ (↓)	23.00	**8.00**	9.00
FWWC19-header	曲线下面积 (↑)	42.67	44.28	**44.59**
FWWC19-header	$\mathcal{M}_{data}^{10\%}$ (↑)	35.18	42.32	**42.39**
FWWC19-header	$\mathcal{M}_{perf}^{90\%}$ (↓)	35.00	**12.00**	**12.00**

表5. **数据集泛化**。 EM 和 UM 在另外两个数据集上优于 RS。随着类别多样性降低，EM 与 UM 之间的差距变小。

在每个主动学习步骤中继续训练 5 个 epoch，仍然保持了 EM 优于 RS 的相同趋势。

表格4 总结了（i）加快训练、（ii）调整主动学习调度器，以及（iii）继续训练在指标方面的实际表现。尽管性能差异最小，但 RS 与 EM 之间的差距仍然存在。最重要的是，这些技巧使实验运行速度提升了一个数量级。

4.4. 泛化分析

数据集泛化。 我们依次在另外两个数据集上对框架进行了实验，即 SoccerNet-ball 和 FWWC19-header。由于 SoccerNet-ball 的动作更加密集，NetVLAD 的超参数++经过调整，时间窗口设为 1秒，NMS设为1 秒。我们选择了加速的主动学习设置，包括更快的训练调度器、自适应主动学习步骤，以及在完成 20 个训练周期的引导训练后，每个主动学习步骤进行 5 个训练周期的持续微调。表 5 详细说明了结果，特别显示，与随机采样相比，UM 和 EM 显著提升了两个数据集上的训练效率。有趣的是，在这两个数据集中，UM与 EM之间的差距小于其在 SoccerNet-v2 上的差距。我们推测，这种行为源于 SoccerNet-头球和FWWC19-header 中类别数量较少，分别为 2 和 1（见表 1）。事实上，在二元分类器的情况下，UM 和 EM 对样本的排序实际上是相似的（见公式（1）和（2））。

架构泛化。我们分析了主动学习框架对其他动作定位方法的泛化能力，尤其是精确时间定位 [24]。不同于 NetVLAD++，精确时间定位为每帧而非每个片段生成类别预测分数。为了估计每个片段的主动学习分数，我们通过平均池化或最大池化聚合每帧的主动学习分数。前者会考虑片段所有帧的平均不确定性，后者则会采样包含单个不确定帧的片段，以便接下来进行训练。同样地，我们选择了相同的加速主动学习设置，因为精确时间定位的速度要慢得多

数据集	AL	曲线下面积 (↑)	$\mathcal{M}_{data}^{10\%}$ (↑)	$\mathcal{M}_{perf}^{90\%}$ (↓)
SoccerNet-v2	RS	27.76	13.59	60.00
SoccerNet-v2	平均EM	28.53	16.14	**50.00**
SoccerNet-v2	最大EM	**28.83**	**17.62**	**50.00**
SoccerNet-v2	平均UM	28.26	13.58	60.00
SoccerNet-v2	最大UM	28.80	15.64	**50.00**
SoccerNet-ball	RS	56.37	52.28	19.00
SoccerNet-ball	平均EM	54.26	46.57	40.00
SoccerNet-ball	最大EM	57.64	51.79	25.00
SoccerNet-ball	平均UM	54.92	49.53	35.00
SoccerNet-ball	最大UM	**58.48**	**53.80**	**12.00**
FWWC19-header	RS	29.77	21.77	-
FWWC19-header	平均EM	34.81	18.45	**40.00**
FWWC19-header	最大EM	33.92	23.43	70.00
FWWC19-header	平均UM	**35.93**	19.14	50.00
FWWC19-header	最大UM	35.26	**24.86**	**40.00**

表6. **架构泛化**。 与 PTS [24]结合时，EM 和 UM 主动选择函数的表现也优于 RS 选择。

比 NetVLAD 训练++。与 NetVLAD 的研究结果一致++，我们在表4 中展示了UM 和EM在全部三个数据集上均优于RS。此外，我们发现了一个类似的趋势：类别数量较少时，UM 往往优于EM。最后，最大池化似乎效果更好，这意味着包含单个不确定帧的片段通常能为训练模型提供更多信息。

5. 结论

总而言之，我们提出的主动学习框架会选择信息量最大的待标注视频样本，从而减少标注工作量，并加快动作定位模型的训练。我们利用不确定性采样，选择下一个最具挑战性的视频片段进行训练，从而加快模型的学习过程。我们表明，该框架能够有效减少在足球视频中准确进行动作定位所需的训练数据：在 SoccerNet-v2 上使用 NetVLAD++ 时，仅用三分之一的数据集便可达到相近的性能。这表明该框架在减少标注工作量方面具有显著潜力。此外，我们还在两个新的数据集上验证了这一方法，这两个数据集侧重于对头球和传球动作进行时间定位。未来，我们将探索其他适用于动作定位任务的主动学习范式，例如多样性最大化、委员会查询和期望误差。

致谢。 本研究部分得到了阿卜杜拉国王科技大学（KAUST）赞助研究办公室通过视觉计算中心（VCC）资助项目以及SDAIA-KAUST数据科学与人工智能卓越中心（SDAIA-KAUST AI）的支持。A. Cioppa获得F.R.S.-FNRS资助。感谢Eloise Arnold协助设计FWWC19-header数据集的可靠性协议。

参考文献

[1] Adri`a Arbu´es Sang¨uesa、Adri`an Mart´ın、Javier Fern´andez、Coloma Ballester和Gloria Haro。使用球员身体朝向对足球比赛中的传球可行性进行建模。载于IEEE/CVF计算机视觉与模式识别会议研讨会（CVPRW），第3875–3884页，美国华盛顿州西雅图，2020年6月。电气与电子工程师协会（IEEE）。2[2] Sunil Bandla和Kristen Grauman。从未剪辑视频中主动学习动作检测器。载于IEEE国际计算机视觉会议（ICCV），第1833–1840页，澳大利亚新南威尔士州悉尼，2013年12月。电气与电子工程师协会（IEEE）。3[3] Fabian Caba Heilbron、Victor Escorcia、Bernard Ghanem和Juan Carlos Niebles。ActivityNet：用于人类活动理解的大规模视频基准。载于IEEE/CVF计算机视觉与模式识别会议（CVPR），第961–970页，美国马萨诸塞州波士顿，2015年6月。电气与电子工程师协会（IEEE）。1、2[4] Alejandro Cartas、Coloma Ballester和Gloria Haro。使用无监督球员分类的基于图的足球动作定位方法。载于国际ACM多媒体体育内容分析研讨会（MMSports），第93–102页，葡萄牙里斯本，2022年10月。ACM。3[5] Kyunghyun Cho、Bart van Merrienboer、Dzmitry Bahdanau和Yoshua Bengio。关于神经机器翻译的性质：编码器—解码器方法。载于SSST-8，即第八届句法、语义、结构与统计翻译研讨会论文集，第103–111页，卡塔尔多哈，2014年。计算语言学协会。3[6] Anthony Cioppa、Adrien Deli`ege、Silvio Giancola、Bernard Ghanem和Marc Van Droogenbroeck。通过多视角空间定位与重识别扩展Soccer-Net。Scientific Data，9(1)：1–9，2022年6月。2[7] Anthony Cioppa、Adrien Deli`ege、Silvio Giancola、Bernard Ghanem、Marc Van Droogenbroeck、Rikke Gade和Thomas B. Moeslund。用于足球视频动作定位的上下文感知损失函数。载于IEEE/CVF计算机视觉与模式识别会议（CVPR），第13123–13133页，美国华盛顿州西雅图，2020年6月。电气与电子工程师协会（IEEE）。1、2、4[8] Anthony Cioppa、Silvio Giancola、Adrien Deliege、Le Kang、Xin Zhou、Zhiyu Cheng、Bernard Ghanem和Marc Van Droogenbroeck。SoccerNet-tracking：足球视频中的多目标跟踪数据集与基准。载于IEEE国际计算机视觉与模式识别会议研讨会（CVPRW），CVsports，第3490–3501页，美国路易斯安那州新奥尔良，2022年6月。电气与电子工程师协会（IEEE）。2[9] Abdulrahman Darwish和Tallal El-Shabrway。STE：用于足球视频动作定位的时空编码器。载于国际ACM多媒体体育内容分析研讨会（MMSports），第87–92页，葡萄牙里斯本，2022年10月。ACM。3[10] Tom Decroos、Lotte Bransen、Jan Van Haaren和Jesse Davis。行动胜于进球。载于ACM SIGKDD国际知识发现与数据挖掘会议，第1851–1861页。ACM，2019年7月。2[11] Adrien Deli`ege、Anthony Cioppa、Silvio Giancola、Meisam J. Seikavandi、Jacob V. Dueholm、Kamal Nasrollahi、Bernard Ghanem、Thomas B. Moeslund和Marc Van Droogenbroeck。SoccerNet-v2：用于全面理解足球转播视频的数据集与基准。在 IEEE 国际计算机视觉与模式识别会议研讨会（CVPRW），CVsports，第 4508–4519 页，美国田纳西州纳什维尔，2021 年 6 月。最佳 CVSports 论文奖。 2[12] Vasilios Duros、Jonathan Grizou、Weimin Xuan、Zied Hosni、De-Liang Long、Haralampos N Miras 和 Leroy Cronin。人类与机器人在巨型多金属氧酸盐的发现和结晶过程中的较量。《应用化学》，129(36):10955–10960，2017。 3[13] A. Ekin、A.M. Tekalp 和 R. Mehrotra。自动足球视频分析与摘要。IEEE 图像处理汇刊，12(7):796–807，2003 年 7 月。 1[14] Yoav Freund、Sebastian H. Seung、Eli Shamir 和 Naftali Tishby。使用委员会查询算法进行选择性采样。《机器学习》，28(2-3):133–168，1997 年 8 月。 3[15] Gregory G. Castanon，Casta˜n´on。通过大型视频语料库进行探索性搜索。博士论文，美国波士顿大学，2016。 1[16] Yarin Gal、Riashat Islam 和 Zoubin Ghahramani。基于图像数据的深度贝叶斯主动学习。在国际机器学习会议（ICML），第 1183–1192 页。PMLR，2017。 3[17] Silvio Giancola、Mohieddine Amine、Tarek Dghaily 和 Bernard Ghanem。SoccerNet：用于足球视频动作定位的可扩展数据集。在 IEEE/CVF 计算机视觉与模式识别会议研讨会（CVPRW），第 1792–179210 页，美国犹他州盐湖城，2018 年 6 月。电气和电子工程师协会（IEEE）。 2, 4, 6[18] Silvio Giancola、Anthony Cioppa、Adrien Deli`ege、Flori- ane Magera、Vladimir Somers、Le Kang、Xin Zhou、Olivier Barnich、Christophe De Vleeschouwer、Alexandre Alahi、Bernard Ghanem、Marc Van Droogenbroeck、Abdulrahman Darwish、Adrien Maglo、Albert Clap´es、Andreas Luyts、An- drei Boiarov、Artur Xarles、Astrid Orcesi、Avijit Shah、Baoyu Fan、Bharath Comandur、Chen Chen、Chen Zhang、Chen Zhao、Chengzhi Lin、Cheuk-Yiu Chan、Chun Chuen Hui、Dengjie Li、Fan Yang、Fan Liang、Fang Da、Feng Yan、Fufu Yu、Guanshuo Wang、H. Anthony Chan、He Zhu、Hongwei Kan、Jiaming Chu、Jianming Hu、Jianyang Gu、Jin Chen、Jo˜ao V. B. Soares、Jonas Theiner、Jorge De Corte、Jos´e Hen- rique Brito、Jun Zhang、Junjie Li、Junwei Liang、Leqi Shen、Lin Ma、Lingchi Chen、Miguel Santos Marques、Mike Aza- tov、Nikita Kasatkin、Ning Wang、Qiong Jia、Quoc Cuong Pham、Ralph Ewerth、Ran Song、Rengang Li、Rikke Gade、Ruben Debien、Runze Zhang、Sangrok Lee、Sergio Escalera、Shan Jiang、Shigeyuki Odashima、Shimin Chen、Shoichi Ma- sui、Shouhong Ding、Sin-wai Chan、Siyu Chen、Tallal El- Shabrawy、Tao He、Thomas B. Moeslund、Wan-Chi Siu、Wei Zhang、Wei Li、Xiangwei Wang、Xiao Tan、Xiaochuan Li、Xiaolin Wei、Xiaoqing Ye、Xing Liu、Xinying Wang、Yan- dong Guo、Yaqian Zhao、Yi Yu、Yingying Li、Yue He、Yujie Zhong、Zhenhua Guo 和 Zhiheng Li。SoccerNet 2022 挑战赛结果。在 ACM 多媒体体育内容分析研讨会（MMSports），第 75–86 页，葡萄牙里斯本，2022 年 10 月。ACM。 2, 5[19] Silvio Giancola 和 Bernard Ghanem。面向足球转播动作定位的时间感知特征池化。在

IEEE 国际计算机视觉与模式识别会议（CVPR），第 4490–4499 页，美国田纳西州纳什维尔，2021 年 6 月。2, 6[20] Ran Gilad-Bachrach、Amir Navot 和 Naftali Tishby。让委员会查询成为现实。载于 NeurIPS，第 18 卷，2005 年。3[21] Dilek Hakkani-T¨ur、Giuseppe Riccardi 和 Allen Gorin。用于自动语音识别的主动学习。载于 IEEE 国际声学、语音与信号处理会议（ICASSP），第 4 卷。IEEE，2002 年。3[22] Fabian Caba Heilbron、Joon-Young Lee、Hailin Jin 和 Bernard Ghanem。接下来该标注什么？一项关于动作定位主动学习的实证研究。载于 ECCV，《计算机科学讲义》第 11215 卷，第 212–229 页。Springer 国际出版，2018 年。3, 6[23] Steven C. Hoi、Rong Jin、Jianke Zhu 和 Michael R. Lyu。批量模式主动学习及其在医学图像分类中的应用。载于国际机器学习会议（ICML），第 417–424 页，2006 年。3[24] James Hong、Haotian Zhang、Micha¨el Gharbi、Matthew Fisher 和 Kayvon Fatahalian。在视频中检测时间精确、粒度细致的事件。CoRR，abs/2207.10213，2022 年。2, 6, 8[25] Kristina Host 和 Marina Iva˘si´c-Kos。基于计算机视觉的体育运动中人体动作识别概述。Heliyon，8(6)，2022 年 6 月。1[26] Neil Houlsby、Ferenc Husz´ar、Zoubin Ghahramani 和 M´at´e Lengyel。用于分类和偏好学习的贝叶斯主动学习。CoRR，abs/1112.5745，2011 年。3[27] IFAB。竞赛规则。技术报告，国际足球协会理事会，瑞士苏黎世，2022 年。2[28] Juan Eugenio Iglesias、Ender Konukoglu、Albert Montillo、Zhuowen Tu 和 Antonio Criminisi。通过主动学习结合生成式和判别式模型，对 CT 扫描进行语义分割。载于《医学成像信息处理》，《计算机科学讲义》第 6801 卷，第 25–36 页。Springer Berlin Heidelberg，2011 年。3[29] Yudong Jiang、Kaixu Cui、Leilei Chen、Canjin Wang 和 Changliang Xu。SoccerDB：用于全面视频理解的大规模数据库。载于 ACM 多媒体体育内容分析国际研讨会（MMSports），第 1–8 页。ACM，2020 年 10 月。2[30] Ajay J. Joshi、Fatih Porikli 和 Nikolaos Papanikolopou-los。用于图像分类的多类别主动学习。载于 IEEE/CVF 计算机视觉与模式识别会议（CVPR），第 2372–2379 页，美国佛罗里达州迈阿密，2009 年 6 月。电气与电子工程师协会（IEEE）。3[31] Ajay J. Joshi、Fatih Porikli 和 Nikolaos P. Papanikolopou-los。用于多类别图像分类的可扩展主动学习。IEEE 模式分析与机器智能汇刊，34(11):2259–2273，2012 年 11 月。3[32] Ksenia Konyushkova、Raphael Sznitman 和 Pascal Fua。从数据中学习主动学习。载于 NeurIPS，第 30 卷，2017 年 3。[33] David D. Lewis 和 Jason Catlett。用于监督学习的异质不确定性采样。《机器学习论文集》，第 148–156 页，1994 年。3[34] Paul Liu 和 Jui-Hsien Wang。MonoTrack：从单目羽毛球视频重建羽毛球轨迹。

在IEEE/CVF计算机视觉与模式识别会议研讨会（CVPRW）中，第3512–3521页，美国路易斯安那州新奥尔良，2022 2 年6月。电气与电子工[35]程师协会（IEEE）。Adrien Maglo、Astrid Orcesi和Quoc-Cuong Pham。利用少量比赛特定标注实现团队运动球员的高效跟踪。在IEEE/CVF计算机视觉与模式识别会议研讨会（CVPRW）中，第3460–3470页，美国路易斯安那州新奥尔良，2022年6月。电气与电子工程师协会（IEEE）。2[36] Thomas B. Moeslund、Graham Thomas和Adrian Hilton。体育运动中的计算机视觉。Springer，2014年。2[37] Naoki Nonaka、Ryo Fujihira、Monami Nishio、Hidetaka Mu-rakami、Takuya Tajima、Mutsuo Yamada、Akira Maeda和Jun Seita。橄榄球端到端高风险擒抱检测系统。在IEEE/CVF计算机视觉与模式识别会议论文集，第3550 3559页，2022年 2 – 。[38] Luca Pappalardo、Paolo Cintia、Alessio Rossi、Emanuele Massucco、Paolo Ferragina、Dino Pedreschi和Fosca Gi-annotti。足球比赛中的时空事件公共数据集。《科学数据》，6(1):1–15，2019年10月。2[39] Monirul Islam Pavel、Siok Yee Tan和Azizi Abdul-lah。基于深度学习的视觉自主车辆系统：系统性文献综述。《应用科学》，12(14):1–51，2022年7月。1[40] Pengzhen Ren、Yun Xiao、Xiaojun Chang、Po-Yao Huang、Zhihui Li、Brij B. Gupta、Xiaojiang Chen和Xin Wang。深度主动学习综述。ACM计算调查，54(9):1–40，2021年10月。3[41] Olav Rongved、Markus Stige、Steven Hicks、Vajira Tham-bawita、Cise Midoglu、Evi Zouganeli、Dag Johansen、Michael Riegler和P˚al Halvorsen。足球中的自动事件检测与分类：使用多种模态的潜力。《机器学习与知识提取》，3(4):1–25，2021年12月。2[42]Atom Scott、Ikuma Uchida、Masaki Onishi、Yoshinari Kameda、Kazuhiro Fukui和Keisuke Fujii。SoccerTrack：一个利用鱼眼和无人机视频进行足球数据集构建与跟踪的算法。在IEEE/CVF计算机视觉与模式识别会议研讨会（CVPRW）中，第3568–3578页，美国路易斯安那州新奥尔良，2022 2 年6月。[43]电气与电子工程师协会（IEEE）。Ozan Sener和Silvio Savarese。卷积神经网络的主动学习：核心集方法。CoRR，abs/1708.00489，2017年。3[44] Burr Settles。主动学习文献综述。计算机科学技术报告1648，2009年。3[45] H. Sebastian Seung、Manfred Opper和Haim Sompolin-sky。委员会查询。在第五届年度计算学习理论研讨会论文集，第287–294页。ACM，1992年7月。3[46] J o˜ao V. B. Soares和Avijit Shah。使用密集检测锚点进行动作定位的再探讨：Soccer-Net挑战赛投稿 2022 2206 07846, 2022 2 。CoRR，[47]abs/. . Joao V. B. Soares、Avijit Shah和Topojoy Biswas。使用密集检测锚点在足球视频中实现时间精确的动作定位。在IEEE国际图像处理会议（ICIP）中，第2796–2800页，法国波尔多，2022年10月。电气与电子工程师协会（IEEE）。2[48] Mattia Soldan、A. Pardo、Juan Le’on Alc’azar、Fabian Caba Heilbron、Chen Zhao、Silvio Giancola和Bernard

Ghanem。《Mad：用于电影音频描述视频语言grounding 的可扩展数据集》。2022 IEEE/CVF 计算机视觉与模式识别会议（CVPR），第 5016–5025 页，2021。1[49] Vladimir Somers、Christophe De Vleeschouwer 和 Alexandre Alahi。《用于遮挡人物再识别的基于身体部位的表征学习》。载于 IEEE/CVF 冬季计算机视觉应用会议（WACV），第 1613–1623 页，美国夏威夷州威可洛亚，2023 年 1 月。电气与电子工程师协会（IEEE）。2[50] G. Sreenu 和 M. A. Saleem Durai。《智能视频监控：通过深度学习技术进行人群分析的综述》。大数据杂志，6(1)，2019 年 6 月。1[51] Swathikiran Sudhakaran、Sergio Escalera 和 Oswald Lanz。《用于视频动作识别的门控移位网络》。载于 IEEE/CVF 计算机视觉与模式识别会议（CVPR），第 1099–1108 页，美国华盛顿州西雅图，2020 年 6 月。电气与电子工程师协会（IEEE）。3[52] Genki Suzuki、Sho Takahashi、Takahiro Ogawa 和 Miki Haseyama。《基于深度极限学习机及战术特征的足球视频团队战术估计》。IEEE Access，7:153238–153248，2019。2[53] Graham Thomas、Rikke Gade、Thomas B. Moeslund、Peter Carr 和 Adrian Hilton。《体育领域的计算机视觉：当前应用与研究主题》。计算机视觉与图像理解，159:3–18，2017 年 6 月。1, 2[54] Cynthia A. Thompson、Mary Elaine Califf 和 Raymond J. Mooney。《用于自然语言句法分析与信息提取的主动学习》。载于国际机器学习会议（ICML），第 406–414 页。Citeseer，1999。3[55] Matteo Tomei、Lorenzo Baraldi、Simone Calderara、Simone Bronzin 和 Rita Cucchiara。《RMS-net：用于足球事件定位的回归与掩码》。载于 IEEE 国际模式识别会议（ICPR），第 7699–7706 页，意大利米兰，2021 年 1 月。电气与电子工程师协会（IEEE）。2[56] Simon Tong 和 Edward Chang。《用于图像检索的支持向量机主动学习》。载于第九届 ACM 国际多媒体会议论文集，第 107–118 页。ACM，2001 年 10 月。3[57] Gabriel Van Zandycke、Vladimir Somers、Maxime Istasse、Carlo Del Don 和 Davide Zambrano。《DeepSportradar-v1：用于体育理解的高质量标注计算机视觉数据集》。载于 ACM 多媒体内容分析体育研讨会（MMSports），第 1–8 页，葡萄牙里斯本，2022 年 10 月。ACM。2[58] Renaud Vandeghen、Anthony Cioppa 和 Marc Van Droogenbroeck。《通过半监督训练改进足球运动员和足球检测》。载于 IEEE 计算机视觉与模式识别国际会议研讨会（CVPRW），CVsports，第 3480–3489 页，美国路易斯安那州新奥尔良，2022 年 6 月。电气与电子工程师协会（IEEE）。2[59] Bastien Vanderplaetse 和 Stephane Dupont。《结合音频与视频流改进足球动作定位》。载于 IEEE 计算机视觉与模式识别国际会议研讨会（CVPRW），CVsports，第 3921–3931 页，美国华盛顿州西雅图，2020 年 6 月。电气与电子工程师协会（IEEE）。2[60]Kanav Vats、Mehrnaz Fani、Pascale Walters、David A. Clausi 和 John Zelek。《粗略标注的……中的事件检测》通过并行多感受野一维卷积处理的体育视频。 *CoRR*，abs/2004.06172，2020年。 2[61] Kanav Vats、William McNally、Pascale Walters、David A. Clausi 和 John S. Zelek。通过 Transformer 和弱监督学习识别冰球运动员。载于 *IEEE/CVF*计算机视觉与模式识别会议研讨会（*CVPRW*），第 3450–3459 页，美国路易斯安那州新奥尔良，2022 年 6 月。电气与电子工程师协会（IEEE）。 2[62] Yi Yang、Zhigang Ma、Feiping Nie、Xiaojun Chang 和 Alexander G. Hauptmann。通过不确定性采样和多样性最大化进行多类别主动学习。 *IJCV*，113(2):113–127，2014 年 11 月。 3[63] Donggeun Yoo 和 In So Kweon。主动学习中的损失学习。载于 *IEEE/CVF*计算机视觉与模式识别会议（*CVPR*），第 93–102 页，美国加利福尼亚州长滩，2019 年 6 月。电气与电子工程师协会（IEEE）。 3[64] Junqing Yu、Aiping Lei、Zikai Song、Tingting Wang、Hengyou Cai 和 Na Feng。足球转播视频综合数据集。载于 *IEEE*多媒体信息处理与检索会议（*MIPR*），第 418–423 页，美国佛罗里达州迈阿密，2018 年 4 月。电气与电子工程师协会（IEEE）。 2[65] He Zhu、Junwei Liang、Chengzhi Lin、Jun Zhang 和 Jianming Hu。基于 Transformer 的足球视频动作定位系统。载于 *ACM*多媒体内容分析体育研讨会（*MMSports*），第 103–109 页，葡萄牙里斯本，2022 年 10 月。ACM。 3