arXiv:2301.13576v1 [cs.AI] 31 Jan 2023

体育任务：MediaEval 2022 视频中的乒乓球击球动作细粒度动作检测与分类

皮埃尔-埃蒂安马丁[1], 乔丹卡朗德尔[2], 鲍里斯 曼桑卡尔[3], 珍妮 贝努瓦-皮诺[3],雷诺 佩泰里[2], 洛朗 马斯卡里利亚[2] 和 朱利安莫利耶[4]

[1]马克斯·普朗克演化人类学研究所 *CCP*系，德国莱比锡 *D-04103*[2]拉罗谢尔大学 *MIA*，法国拉罗谢尔[3]波尔多大学、法国国家科学研究中心、波尔多国立综合理工学院、波尔多计算机科学研究所，法国塔朗斯[4]波尔多大学 *IMS*，法国塔朗斯

摘要

体育视频分析是一个广泛的研究领域，其应用非常多样，例如比赛事件检测、视频摘要或运动员的细粒度动作分析。作为 MediaEval 2022 基准测试计划的一部分，本任务旨在检测和分类体育视频中的细微动作。我们聚焦于乒乓球比赛录像。自 2019 年开展以来，本任务一直提供分类挑战，使用在自然条件下录制、且每次击球均标注了已知时间边界的未修剪视频。自 2021 年起，本任务还提供基于未标注未修剪视频的击球检测挑战。今年，我们扩充了训练集、验证集和测试集，以确保每个数据集中都包含所有击球动作。现在，该数据集与 [1, 2] 中使用的数据集相似。本研究旨在为希望进一步评估自身体育表现的教练和运动员开发工具。

1. 引言

动作检测和分类是计算机视觉领域的主要挑战之一 [3]。在过去几年中，专注于动作分类的数据集数量大幅增加，其复杂性也不断提升 [2]。大量研究致力于使用动作捕捉系统分析体育动作。然而，附着在身体上的传感器和标记点往往会干扰运动员的自然行为。因此，这一问题推动了利用摄像机进行视频录制的非侵入式方法的发展。

这项体育视频分类项目由波尔多大学体育学院（STAPS）、计算机科学实验室波尔多计算机科学研究所（LaBRI）以及拉罗谢尔大学的MIA实验室发起 [1]。该项目旨在开发用于乒乓球击球识别的人工智能和多媒体索引方法。最终目标是评估

MediaEval'22：多媒体评测研讨会，*2023*年*1*月*13*日至*15*日，挪威卑尔根及线上
mediaeval.sport.task@diff.u-bordeaux.fr (P. Martin)
www.eva.mpg.de/comparative-cultural-psychology/staff/pierre-etienne-martin (P. Martin) 0000-0002-9593-4580 (P. Martin)
© 2022年本文作者版权所有。根据知识共享署名4.0国际许可协议（CCBY4.0），允许使用。CEUR研讨会论文集（CEUR-WS.org）
得到新阿基坦大区的CRISP项目——体育表现计算机视觉，以及MIRES联合会的支持。

个人运动员，尤其是学生的表现，从而制定最佳训练策略。为此，我们与志愿运动员合作录制了一组视频 TTStroke-21 。

UCF-101 [4],HMDB [5, 6], AVA [7]和 Kinetics [8, 9, 10]等数据集应用于动作识别领域，多年来涵盖的视频样本和类别数量不断增加。专注于体育细粒度分类的数据集寥寥无几，例如 FineGym [11]和 TTStroke21 [12]。

为应对数据集日益增加的复杂性，一些分类方法会尽可能充分地利用时间信息。例如，[13]仅使用 RGB数据从视频中学习时空依赖关系。另一种做法是整合从视频中提取的其他模态，例如光流 [14, 15, 16]。此外，在 TTStroke-21 数据集中，击球分类颇具挑战性，因为两次击球之间的动作在视觉上高度相似。

以下各节将介绍今年的体育任务及其子任务，并说明数据集以及下载数据集时的具体使用条款。有关该任务的补充信息，请参阅 MediaEval 网站上的专门页面[2]。

2. 任务描述

体育任务基于 TTStroke-21 数据库 [17, 12]。该数据库包含在自然条件下进行乒乓球比赛的录像。通过本任务提供的数据集重点收录了使用 GoPro摄像机以每秒 120 帧的速度拍摄、并由专业选手标注的视频。本任务为研究人员提供了一个机会，可利用高质量的视频和标注解决体育领域的细粒度分类问题。与 MediaEval 2021 届的体育任务相比 [18], ，数据集得到了扩充，数据的组织和分发方式也有所不同。本任务包含两个子任务：从修剪后视频中进行击球分类，以及从未修剪视频中进行击球检测。每个子任务都有各自的数据集。

研究人员可以参加其中一个或两个子任务，并为每个子任务最多提交五个运行结果。参与者必须为每个运行结果填写该子任务测试集专用的 XML文件。XML文件的内容因子任务而异。运行结果必须以压缩包（zip 文件）的形式提交，每个子任务中的每次运行都应放在不同的目录中。参与者还应提交一篇工作说明论文，介绍其方法，并说明是否使用了任何外部数据（例如其他数据集或预训练网络）来计算运行结果。然而，禁止使用往年体育任务数据集 TTStroke-21 上训练的预训练模型。本任务被视为全自动任务：系统获得视频后，应在没有任何人工干预的情况下生成结果。鼓励参与者在提交作品时公开发布代码。今年与 2021 届类似，两个子任务的基线均已公开分享[3] [19]。

2.1. 子任务 1- 击球分类

在此子任务中，参与者需要对一组修剪后视频进行分类，每段视频中仅包含一次乒乓球击球，也可能完全没有击球。共有 20 种可能的击球类别，另加一种非击球类别。为此，提供了两个标注集：一个训练集和一个

[2]https://multimediaeval.github.io/editions/2022/tasks/sportsvideo/ [3] https://github.com/ccp-eva/SportTaskME22

验证集，分别包含 807 段和 230 段修剪后视频。还需对一个包含 118 段修剪后视频的未标注测试集进行分类。不同集合中的修剪后视频可能来自相同的未修剪视频，但截取时间点不同且彼此不重叠。

具体而言，邀请参与者填写 XML 文件，并将默认标签“Unknown”替换为参与者所用方法分配的击球类别。所有提交结果都将根据总体准确率进行排名，并提供各类别准确率的详细信息。

在去年的赛事中，使用 SWIN-Transformers 的方法取得了最高的总体准确率，达到 74.2% [20]；ResNet-50 模型紧随其后，准确率为 68.8%， [21]大幅领先于去年的基线（20.4%）。方法的效果似乎与模型架构有关，同时也取决于训练时是否考虑击球类别之间的相似性以及类别不平衡。今年的任务采用与 [1, 2] 相同的 TTStroke-21 数据库划分，以便更好地与 MediaEval 基准范围之外的既有研究进行比较。

2.2. 子任务 2——击球检测

在此子任务中，参与者需要对一组未修剪视频进行分段，以检测其中的击球，不论击球类别为何。为此，任务提供了两个标注集：训练集和验证集，分别包含 16 段和 6 段未修剪视频。参与者还需对一个由 6 段未修剪视频组成的未标注测试集进行时间分段。训练集、验证集和测试集中的视频互不重复；不过，同一位球员可能出现在不同的数据集中。

具体来说，参与者需要在提供的测试集XML文件中填写击球动作的时间边界（视频的帧索引）。所有提交结果都将根据平均精度均值（mAP）和时间交并比（IoU）进行评估。这两项指标通常用于图像分割，但经过调整后也适用于本任务：

- **mAP：**每个击球动作都视为一个需要进行时间检测的对象。当预测结果与真实标注之间的时间IoU高于某个IoU阈值时，即视为检测成功。与COCO挑战赛 [22]类似，采用从0.5到0.95、步长为0.05的20个阈值。该指标将用于对参与者进行最终排名。
- **IoU：**所有视频中真实标注与预测击球动作之间逐帧的重叠度。

在去年的评测中，只有两位参与者为这一困难的子任务提交了运行结果 [23, 24]。他们并未提升基线的mAP结果， [24] 但使用YOLOv5模型将IoU提高到了0.247，而基线为0.144。

3. 数据集描述

该数据集在波尔多大学体育学院录制。它由在自然条件下使用 GoPro摄像机拍摄的、以球员为中心的视频构成，不含标记点或传感器（见图1）。专业乒乓球教练设计了一套专门的分类体系，用于描述所有可能的击球。该数据集包含 20 种乒乓球击球类别：8 种发球、6 种进攻性击球和 6 种防守性击球。击球也可进一步划分为两个超类：Forehand 和 Backhand。该数据集由专业球员通过众包标注平台进行标注。非击球样本根据击球标注推断得出。



图1： 来自 TTStroke-21 的同一击球的关键帧

为了能够分享该数据集，我们使用基于单次检测器（SSD）框架和 ResNet 主干网络的 OpenCV 深度学习人脸检测器，对每段原始视频的每一帧中球员的面部进行模糊处理。我们实现了一种跟踪方法，以降低误报率。检测到的人脸会被模糊处理，视频则重新编码为 MPEG-4 格式。

与去年的版本相比，今年的分类数据集新增了更加多样化的视频样本，内容更加丰富。源视频经过修剪后，按类别归档并分配到训练集、验证集和测试集中。本子任务共采用 1,155 个修剪后视频，包含超过 210,000 帧。对于检测子任务，采用了来自 28 个视频、总时长 100 分钟的乒乓球比赛视频；这些视频以每秒 120 帧的速度录制，并分配到训练集、验证集和测试集中。其帧数超过 718,000 帧。两个子任务的视频分辨率均为 1920×1080，总计占用 46.1 GB 的磁盘空间。每个子任务均提供验证集，以便参与者进行更好的比较。在提交测试集结果时，也可以使用该验证集进行训练。

4. 特殊条件

尽管系统会自动对人脸进行模糊处理以保护匿名性，但部分人脸可能被误检测，因此仍有一些运动员可以被识别。为尊重运动员的个人数据，本数据集受一项使用协议约束，该协议称为特殊条件。

这些特殊条件适用于 MediaEval 研讨会任务“体育任务：从视频中对乒乓球击球进行细粒度动作检测与分类”的视频使用。这些条件对应于用户与代尔夫特大学签署的《*MediaEval 2022*研究集使用协议》中提及的具体使用协议。完全接受这些特殊条件 是提供视频以供参与 MediaEval 2022评测活动的必要前提。用户必须完整阅读这些条件，例如，在任何出版物中使用视频之前，遮挡视频中的人脸（模糊处理或添加黑色横幅），并于2023年1月30日前销毁数据。

5. 讨论

与去年一样，MediaEval 体育任务设有两个子任务：i) 分类和 ii) 击球检测。自该任务启动以来，分类方法的性能已有提升。我们希望有更多人参与检测子任务，从而进一步提升体育领域中兴趣时刻的检测效果。我们鼓励参与者分享遇到的困难和取得的结果，即使结果看起来不够理想。所有研究，即使未能成功，

可能会启发未来的方法。

致谢

作者谨向为TTStroke-21做出贡献的运动员、教练和标注人员表示感谢。

参考文献

[1] P. Martin, J. Benois-Pineau, R. Péteri, J. Morlier, 《孪生时空卷积神经网络中的三维注意力机制：在乒乓球比赛视频动作分类中的应用》, 载于: ICPR, IEEE计算机学会, 2021。[2] P. Martin, 《利用时空卷积神经网络对视频中的动作进行细粒度检测与分类：乒乓球领域的应用》（Détection et classification fines d'actions à partir de vidéos par réseaux de neurones à convolutions spatio-temporelles. Application au tennis de table）, 博士论文, 拉罗谢尔大学, 法国, 2020。网址: https://tel.archives-ouvertes.fr/tel-03128769。[3] P. Martin, J. Benois-Pineau, R. Péteri, A. Zemmari, J. Morlier, 《用于动作识别的三维卷积网络：体育手势识别的应用》, 施普林格国际出版公司, 2021。[4] K. Soomro, A. R. Zamir, M. Shah, 《UCF101：包含野外视频中101个人体动作类别的数据集》, CoRR abs/1212.0402（2012）。[5] H. Kuehne, H. Jhuang, E. Garrote, T. A. Poggio, T. Serre, 《HMDB：用于人体运动识别的大型视频数据库》, 载于: ICCV, IEEE计算机学会, 2011, 第2556–2563页。[6] H. Jhuang, J. Gall, S. Zuffi, C. Schmid, M. J. Black, 《迈向动作识别的理解》, 载于: ICCV, IEEE计算机学会, 2013, 第3192–3199页。[7] C. Gu, C. Sun, D. A. Ross, C. Vondrick, C. Pantofaru, Y. Li, S. Vijayanarasimhan, G. Toderici, S. Ricco, R. Sukthankar, C. Schmid, J. Malik, 《AVA：时空定位原子视觉动作视频数据集》（2018）, 第6047–6056页。[8] W. Kay, J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev, M. Suleyman, A. Zisserman, 《Kinetics人体动作视频数据集》, CoRR abs/1705.06950（2017）。[9] L. Smaira, J. Carreira, E. Noland, E. Clancy, A. Wu, A. Zisserman, 《关于Kinetics-700-2020人体动作数据集的简短说明》, CoRR abs/2010.10864（2020）。[10] A. Li, M. Thotakuri, D. A. Ross, J. Carreira, A. Vostrikov, A. Zisserman, 《AVA-Kinetics定位人体动作视频数据集》, CoRR abs/2005.00214（2020）。[11] D. Shao, Y. Zhao, B. Dai, D. Lin, 《FineGym：用于细粒度动作理解的层次化视频数据集》, 载于: CVPR, IEEE, 2020, 第2613–2622页。[12] P. Martin, J. Benois-Pineau, R. Péteri, J. Morlier, 《利用双流时空卷积神经网络进行细粒度体育动作识别》, Multimedia Tools and Applications 79（2020）20429–20447。[13] Z. Liu, H. Hu, 《用于视频动作识别的时空关系网络》, IEEE Access 7（2019）14969–14976。[14] J. Carreira, A. Zisserman, 《动作识别何去何从？一种新模型与Kinetics数据集》, 载于: CVPR, IEEE计算机学会, 2017, 第4724–4733页。[15] G. Varol, I. Laptev, C. Schmid, 《用于动作识别的长期时序卷积》, IEEE模式分析与机器智能汇刊 40（2018）1510–1517。[16] P. Martin, J. Benois-Pineau, R. Péteri, J. Morlier, 《针对基于三维卷积网络的细粒度动作分类, 运动估计方法的最优选择》, 载于: ICIP, IEEE, 2019, 第554–558页。[17] P. Martin, J. Benois-Pineau, R. Péteri, J. Morlier, 《基于孪生时空卷积神经网络的体育动作识别：乒乓球领域的应用》, 载于: CBMI, IEEE, 2018, 第1–6页。[18] P. Martin, J. Calandre, B. Mansencal, J. Benois-Pineau, R. Péteri, L. Mascarilla, J. Morlier, 《体育视频：从视频中对乒乓球击球动作进行细粒度检测与分类——MediaEval 2021》, 载于: [25], 2021。网址: http://ceur-ws.org/Vol-3181/paper3.pdf。[19] P. Martin, 《采用注意力机制的三维卷积神经网络MediaEval 2022基准体育任务基线方法》, 载于: MediaEval, CEUR研讨会论文集, CEUR-WS.org, 2022。[20] Y. Qian, L. Yu, W. Liu, A. Hauptmann, 《学习无偏Transformer以进行长尾体育动作分类》, 载于: [25], 2021。网址: http://ceur-ws.org/Vol-3181/paper52.pdf。[21] P. Martin, 《MediaEval 2021 [25], 体育视频 2021 任务的时空卷积神经网络基线方法》, 载于: 3181。网址: http://ceur-ws.org/Vol- /paper .pdf。13

[22] T. Lin, M. Maire, S. J. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, 《Microsoft COCO：上下文中的常见物体》，载于：D. J. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars（编），《计算机视觉——ECCV 2014——第13届欧洲会议》，瑞士苏黎世，2014年9月6–12日，论文集，第V部分，第8693卷，计算机科学讲义，Springer，2014，第740–755页。[23] A. Zahra, P. Martin, 《用于乒乓球击球检测的双流网络》，载于：[25], 2021。网址：http://ceur-ws. org/Vol-3181/paper55.pdf。[24] B. J, M. T. T, B. B, J. S, L. N. N, 《YOLOV5在乒乓球击球检测与分类中的应用》，载于：[25], 2021。网址：http://ceur-ws.org/Vol-3181/paper38.pdf。[25] S. Hicks, K. Pogorelov, A. Lommatzsch, A. G. S. de Herrera, P. Martin, S. Z. Hassan, A. Porter, A. Kasem, S. Andreadis, M. Lux, M. G. Ocaña, A. Liu, M. Larson（编），MediaEval 2021研讨会工作笔记论文集，线上会议，2021年12月13–15日，第3181卷，*CEUR*研讨会论文集，CEUR-WS.org，2022。网址：http: //ceur-ws.org/Vol-3181。