主观与客观质量评估数据集数据表

Nabajeet Barman*‡、Yuriy Reznik†和Maria Martini‡
*Brightcove UK Ltd，伦敦，英国，nbarman@brightcove.com
†Brightcove Inc，西雅图，美国，yreznik@brightcove.com
‡金斯顿大学，伦敦，英国，m.martini@kingston.ac.uk

arXiv:2305.02142v1 [cs.MM] 3 May 2023

摘要—多年来，人们创建了许多主观和客观质量评估数据集，并将其提供给研究界。然而，目前还没有标准流程来记录数据集的各个方面，例如源序列详情、测试对象人数、测试方法、编码设置等。此类信息对于数据集的用户往往至关重要，因为它能帮助他们快速了解数据集的动机和范围。如果没有这样的模板，每位读者只能自行从相关出版物或网站中汇总信息，这是一个繁琐且耗时的过程。在某些情况下，缺少用于指导文档编制流程的模板，可能会导致某些重要信息被无意遗漏。

本文通过提出一种数据表模板，填补这一简单但重要的空白，用于记录多媒体数据主观和客观质量评估数据集的各个方面。本研究的贡献旨在简化现有数据集和新数据集的文档编制流程，并提高其可复现性。提出的数据表模板可在 GitHub[1]，上获取，此外还提供了几个开源视听主观和客观数据集的数据表示例。

索引词—**体验质量，主观评估，客观评估，数据集，数据库，多媒体，开源**

I. 引言

在过去二十年里，视频流媒体已无处不在，目前约占互联网总流量的 82% [1]。这在很大程度上得益于多媒体流媒体各方面的进步，从编解码器的改进 [2–4] ，到更完善的 CDN、改进的传输与分发机制 [5, 6]，再到性能更强、质量更高的终端用户设备，例如智能手机、智能电视和笔记本电脑。然而，为了让这类通过互联网传输多媒体内容的视频流媒体服务持续发展，确保终端用户对服务的体验质量（QoE）感到满意至关重要。

QoE 在 ITU-T 建议书 P.10/G.100 [7] 中被定义为“应用或服务用户的愉悦或烦恼程度”。多年来，人们开展了大量研究，致力于开发各种质量指标和模型，以帮助预测最终用户感知到的多媒体应用体验质量 [8]。此类质量指标既可以是简单的图像质量指标，例如峰值信噪比和结构相似性指数 [9] ，也可以是更复杂的视频质量指标，例如视频多方法评估融合 [10] 和 ITU-T 建议书 P.1204 [11]。

体验质量领域取得进展的原因之一，例如体验质量模型和指标的改进，以及基于体验质量的视频流工作流优化，是开源数据集的创建与可用性不断提高。从 21 世纪初的早期数据集（如 VQEG-HD3 [12] 和 Live VQA [13],），到较新的数据集（如 AVT-VQDB-UHD1 [14], 、GamingVideoSET [15], 、Live YouTube-Gaming [16] 和 BC-KU MultiScreen Dataset [17]），均体现了这一点。

A. 动机

多年来，已经创建了许多主观和客观质量评估数据集，并向社区开放 [13–24]。为了简化并规范主观测试和客观质量（模型）评估的流程，人们提出了 ITU-T P. 808 [25], ITU-T P.809 [26], ITU-T P.910 [27], ITU-T P. 913 [28],ITU-R BT.500 [29],以及 ITU-T P.1401 [30] 等各种标准。这些标准针对多个方面提供了详细建议，例如视频序列的选择、主观测试流程（测试环境、参与者选择、测试方法等）以及模型性能评估。

然而，目前没有用于说明数据集各个方面文档编制流程的“标准”或模板。在缺乏此类模板的情况下，数据集的各个方面只能由数据集创建者自行报告。缺少模板可能会无意中导致数据集的重要信息被遗漏。这也意味着，收集并记录相关出版物中的信息这一负担被转移给了最终用户；这一过程耗时、繁琐，而且通常无法复现。

B. 先前的工作／努力

记录数据集的需求并非 QoE 数据集所独有。例如，数据来源已在其他领域得到广泛研究，如数据库领域 [31, 32]。同样，近年来，许多研究聚焦于记录机器学习数据集的创建和使用过程。这类研究的例子包括模型卡 [33] 和数据表 [34], ，它们使数据集创建者能够记录机器学习模型和数据集的各个方面。由于它们……，这类研究已在机器学习社区中获得广泛认可。

[1]https://github.com/NabajeetBarman/datasheet-for-qoe-datasets

在提升数据集创建者与用户之间的沟通和透明度方面具有很高的实用价值。

本文受到 Gebru 等人 [34],的研究启发，该研究为基于人工智能/机器学习的数据集提出了一份数据表。然而，所提出的数据表模板（以及其他类似工作）并不适用于体验质量数据集，因为这些模板是为典型的基于人工智能/机器学习的数据集设计的；这类数据集通常规模庞大（数百万张图像、数十亿条文本、十万个视频），且其属性侧重于模型开发流程。另一方面，体验质量数据集通常规模小得多，侧重于对视听内容进行主观和客观评估。

C. 贡献

本文提出了一个“数据表”模板，用于记录体验质量（主观和客观评估）数据集的各个方面。所提出的数据表由归入六个不同部分的各类问题/方面组成。每个字段/问题都附有详细说明。数据表既可由数据集创建者填写，也可由最终用户填写，之后可进行共享，以便他人更轻松地理解其工作并复现研究结果。

该数据表模板可在 GitHub [35] 上公开获取，提供多种格式（*Google* 表格、*.xlsx*、*.odt*、*.pdf* 和 *.html*），可用于记录新建和现有数据集的各个方面。除了提供所提出的数据表模板外，为了便于理解，GitHub 仓库 [35]中还提供了现有开源数据集的多个数据表示例。

D. 模板开发流程

本节详细介绍数据表模板的创建过程。作者首先根据其在体验质量评估领域的丰富经验，创建了数据表模板初稿。具体而言，作者在创建并记录十多个开源数据集时积累的经验，以及使用类似第三方开源数据集开展各种工作的经验，都用于设计模板初稿。随后，作者使用该模板为三个开源数据集创建了示例数据表：GamingVideoSET [15], AVT-VQDB-UHD-1 [14]和 BC-KU 多屏数据集 [36]。根据填写示例数据表的经验，我们发现了缺失或顺序不当的字段、拼写错误，以及标题和描述不够清晰的问题，并据此改进了数据表模板初稿。此后，

随后，我们收集了来自 Sony（德国）、伊尔梅瑙工业大学（德国）和金斯顿大学（英国）的体验质量研究人员的反馈，并根据这些反馈进一步改进了数据表，最终形成了当前提议的数据表模板版本。

II. 拟议的数据表模板

图1展示了所提议的数据表模板的概要。该数据表涵盖了任何传统二维视听体验质量数据集的各个方面，从数据集概览到主观和客观质量评估的详细信息。所提议的字段均为可选项，提供了灵活性，以便于

- 数据集概览：此处提供关于数据集的一般信息，以便读者快速了解该数据集的内容，以及它是否符合自己的兴趣。
  - 例如：数据集名称、许可证、联系信息等。
- 数据集描述：介绍数据集的各个方面，例如：
  - 源视频：帧率、位深度、分辨率等。
  - 编码设置：编解码器、编码器类型、考虑的伪影等。
  - 处理后的视频序列：PVS 数量、容器格式等。
- 主观质量评估：提供有关主观测试的详细信息，例如：
  - 主观测试设置：测试环境、显示尺寸、分辨率和观看距离等。
  - 测试方法：测试软件、测试受试者人数和测试序列等。
  - 测试参与者：有关参与者招募方式及其人口统计信息。
  - 主观评分分析：异常值检测、变换等。
- 客观质量评估：提供有关模型评估的详细信息。
  - 质量指标：所考虑的指标及其类型、所使用的池化方法等。
  - 模型评估：所考虑的性能度量、模型复杂度等。
- 伦理考量：记录任何相关的伦理/GDPR相关问题，例如伦理审查流程、知情同意书以及数据集的预期用途。
- 补充信息：提供数据集创建者、资助等额外详情。以及此前未涵盖的附加信息。

图1：所提出的数据表模板各个部分的概要。

数据集创建者可以根据需要添加任何其他信息。为便于理解，拟议的数据表分为六个不同的部分，下面将逐一介绍。

A. 数据集概览

顾名思义，本节首先向最终用户概述数据集，旨在帮助读者快速了解数据集的内容，以及它是否符合自己的兴趣。本节介绍数据集的名称、创建日期/年份、数据集存储库下载链接、许可证、必需引用以及联系信息等详细信息。

B. 数据集描述

第二部分概述了数据集中源视频和编码视频表示的各种特征，并进一步分为三个子部分。第一部分“源视频”记录了数据集中提供或使用的源视频各方面的信息（数量和类型、位深度、动态范围、分辨率等）。这相当重要，因为读者可能对特定类型内容的数据集感兴趣（例如，10 位高动态范围游戏内容）。第二部分“编码设置”收集编码参数方面的信息，例如编码器类型、码率控制、编解码器、分辨率、位深度等。最后一部分“处理后的视频序列”则记录有关编码视频序列的详细信息，例如使用的序列数量，以及用于媒体播放的容器格式。

C. 主观质量评估

对必须遵循的测试设置、方法论和流程进行详细说明，包括数据处理指南（如异常值检测等），这些内容在各种

ITU 建议书 [27–29] 有助于确认所报告的主观测试结果的可靠性、可重复性和有效性。因此，本节汇总了涵盖主观质量评估各个方面的相关信息：主观测试设置（测试环境、显示器、观看距离、评分量表等）、测试方法（播放软件、测试参与者和测试序列的数量等）、测试参与者信息（测试参与者的人口统计信息等）以及主观数据分析（离群值分析、分数转换等）。本节所呈现的信息有助于读者更好地理解主观测试评估结果。

D. 客观质量评估

客观质量评估包含利用信号保真度等客观测量值来预测人类观察者所感知视觉质量的方法和模型。大多数 QoE 数据集都包含对各种图像和视频质量指标的性能评估。本节收集的数据旨在涵盖模型性能评估的不同方面，例如所考虑的质量指标、所使用的实现及其版本、模型的训练和测试方式，以及用于量化所评估模型和指标性能的各种测量方法。

E. 伦理考量

QoE 数据集通常包含多媒体数据，其中可能包含个人数据，例如人物图像或视频；在某些情况下，也可能包含暴力或令人不安的场景。例如，对于包含游戏视频的数据集，其中可能有一些观众会觉得令人不安的暴力场景。此外，QoE 数据集通常还包含涉及人类测试对象的主观质量评估。因此，对于数据集使用者而言，了解数据集设计过程中考虑的相关伦理审查批准和伦理考量通常非常重要。本节收集了各个方面的信息，例如数据集的预期用途、伦理批准、主观测试期间使用的样本知情同意书，以及读者在使用该数据集之前必须了解的其他相关伦理或《通用数据保护条例》（GDPR）问题。

F. 补充信息

本节包含一些补充问题，旨在收集不属于其他各节核心内容的信息，包括创建者/作者的信息、数据集创建的资金来源，以及读者需要了解的任何数据集机密信息。此外，本节还为数据集创建者提供机会，补充模板中的各项问题未涵盖的其他相关信息。

注： 需要注意的是，所提出的模板包含的功能远多于上述讨论的内容。有关所提出的数据表的完整概述，请参阅 GitHub[35] 中的实际数据表模板，或本文末尾的附录。

G. 示例数据表

为了帮助读者更好地理解所提出的模板，我们为以下数据集提供了示例模板：

1) GamingVideoSET（2018年） [15]：该数据集包含源视频，以及用于游戏视频质量评价的主观和客观评估结果。2) AVT-VQDB-UHD1（2020年） [14]：该数据集包含源视频，以及使用三种不同编解码器编码的视频的主观和客观评分；该数据集部分用于制定 ITU-T 建议书 P.1204 [11]。3) BC-KU 多屏数据集（2023年） [17]：这是一个较新的数据集，包含在由三种不同设备——移动设备、平板电脑和电视——组成的多屏设置下获得的主观和客观评估结果。

III. 讨论、结论与未来工作

本文首次尝试创建数据表模板，以便更好地记录主观和客观质量评估数据集。这一模板并非定稿，也不尽完善；我们预计将根据与其他专家的讨论和反馈，持续改进。本文撰写之际，我们正在征求多家多媒体流式传输公司和组织的资深体验质量研究人员的进一步反馈，例如视频质量专家组 (VQEG)[2] 和 Qualinet[3]。随后，我们将把讨论意见和反馈纳入改进版本的开发，并在 GitHub 仓库中更新。我们相信，这一数据表模板可以帮助数据集创建者、数据集使用者，以及在适用情况下，基于这些数据集发表论文或作品的审稿人。

虽然填写这份数据表会增加数据集创建者的工作负担，但我们相信其益处远大于成本，人工智能/机器学习领域也有类似情况。此外，尽管所提出的模板更适用于传统的二维视频/视听数据集，但也很容易适配其他数据集，例如仅音频、沉浸式视频（虚拟现实/增强现实、360度视频、光场、点云和三维网格）以及计算机视觉数据集。未来，我们计划与其他体验质量专家合作，为此类体验质量数据集创建更具个性化的数据表。

致谢

Nabajeet Barman 谨向 Rakesh Rao、Saman Zadtootaghaj 和 Steven Schmidt 致谢，感谢他们对模板草稿提出的反馈。

参考文献

[1] *Cisco Visual Networking IndexCompleteForecastHighlights*, https :/ / www. cisco . com / c / dam / m / en us/ _ solutions/service- provider/vni- forecast- highlights/pdf/ Global Device Growth Traffic Profiles.pdf, 2018年.[2] *ITU-T 建议 H.264:* 高级视频编码 *for* 通用视听服务, https://www.itu.int/rec/T-REC-H.264, 2021年8月.

[2] https://vqeg.org/vqeg-home/ [3] http://www.qualinet.eu/

[3] ITU-T 建议书 H.265：高效视频编码，https://www.itu.int/rec/T-REC-H.265，2021年8月。[4] ITU-T 建议书 H.266：多功能视频编码，https://www.itu.int/rec/T-REC-H.266，2022年4月。[5] HTTP 实时流媒体，RFC 8216，https://tools.ietf.org/html/rfc8216，2019年。[6] ISO/IEC 23009-1:2019 信息技术 — HTTP 动态自适应流媒体（动态自适应流媒体） — 第 1 部分：媒体呈现描述和分段格式，https://www.iso.org/standard/79329.html，2019年。[7] ITU-T 建议书：性能、服务质量和体验质量词汇，https://www.itu.int/rec/T-REC-P.10-201711-I，2017年11月。[8] N. Barman 和 M. G. Martini，“HTTP 自适应视频流媒体的体验质量建模——综述与开放挑战”，IEEE Access，第 7 卷，第 30 831–30 859 页，2019年。DOI：10.1109/ACCESS.2019.2901778。[9] Z. Wang、A. C. Bovik、H. R. Sheikh 和 E. P. Simon- celli，“图像质量评价：从误差可见性到结构相似性”，IEEE 图像处理汇刊，第 13 卷，第 4 期，第 600–612 页，2004年。[10] Netflix，视频多方法评估融合，https://github.com/Netflix/vmaf。[11] ITU-T 建议书 P.1204：通过可靠传输提供的流媒体服务的视频质量评价，适用于最高 4K 分辨率，https://www.itu.int/rec/T-REC- P.1204，2020年1月。[12] 消费者数字视频库，https://www.cdvl. org/。[13] LIVE 视频质量评价数据库，http://live.ece.utexas.edu/research/Quality/live video.html 。[14] R. R. R. Rao、S. G¨oring、W. Robitza、B. Feiten 和 A. Raake，“AVT-VQDB-UHD-1：大型 UHD-1 视频质量数据库”，载于《2019 IEEE 多媒体研讨会》，2019年12月，第 1–8 页。[15] N. Barman、S. Zadtootaghaj、S. Schmidt、M. G. Mar- tini 和 S. M¨oller，“GamingVideoSET：游戏视频流媒体应用数据集”，载于《2018年第16届网络与系统游戏支持年度研讨会（NetGames）》，2018年，第 1–6 页。DOI：10 . 1109 / NetGames.2018.8463362。[16] LIVE-YouTube 游戏视频质量数据库，https: // live.ece.utexas.edu/research/LIVE-YT-Gaming/index. html。[17] N. Barman、Y. Reznik 和 M. G. Martini，“多屏视频流媒体应用的主观数据集”，载于《2023年第十五届多媒体体验质量国际研讨会》，2023年，第 1–6 页。[18] LIVE Netflix 视频体验质量数据库，http://live.ece.utexas.edu/research/LIVE NFLXStudy/ nflx index.html 。[19] LIVE-NFLX-II 主观视频体验质量数据库，http : / / live . ece. utexas . edu / research / LIVE NFLX II / live nflx plus.html 。[20] LFOVIA 视频体验质量数据库，https://www.iith.ac.in/~lfovia/downloads.html 。[21] EPFL-PoliMI 视频质量评价数据库，http: // vqa.como.polimi.it/，2003年。[22] LIVE HTTP 视频流媒体体验质量数据库，http : / / live . ece . utexas . edu / research / Quality / TVSQ VQA database.html 。[23] N. Barman、S. Schmidt、S. Zadtootaghaj 和 M. G. Martini，编解码器压缩效率评估之使用客观和主观质量评价的 *MPEG-5* 第 *2* 部分（*LCEVC*），2022年。arXiv：2204 . 05580[cs.MM]。[24] *Live* 移动设备卡顿视频数据库 *II*，http:// live .ece . utexas.edu/research/LIVEStallStudy/liveMobile.html 。[25] *ITU-T* 建议书 *P.808*：采用众包方法对语音质量进行主观评价，https://www.itu.int/rec/T-REC-P.808/en，2021年6月。[26] *ITU-T* 建议书 *P.809*：游戏质量的主观评价方法，https://www.itu.int/rec/T- REC-P.809/en，2018年6月。[27] *ITU-T* 建议书 *P.910*：多媒体应用的主观视频质量评价方法，https://www.itu.int/rec/T-REC-P.910-202111-I，2021年11月。[28] *ITU-T* 建议书 *P.913*：在任何环境下对互联网视频以及分发质量电视的视觉质量、音频质量和视听质量进行主观评价的方法，https://www.itu.int/rec/T-REC-P.913-202106-I，2021年6月。[29] *ITU-T* 建议书 *BT.500*：电视图像质量主观评价方法，https://www.itu.int/rec/R-REC-BT.500-14- 201910-I/en，2019年10月。[30] *ITU-T* 建议书 *P.1401*：客观质量预测模型的统计评估、鉴定和比较方法、指标及程序，https://www.itu.int/rec/T-REC-P.1401-202001- I，2020年1月。[31] A. P. Bhardwaj、S. Bhattacherjee、A. Chavan、A. Deshpande、A. J. Elmore、S. Madden 和 A. G. Parameswaran，“DataHub：大规模协作式数据科学与数据集版本管理”，*CoRR*，第 abs/1409.0798 卷，2014年。arXiv：1409.0798。 [在线]。可通过以下网址获取：http://arxiv.org/abs/1409.0798。[32] J. Cheney、L. Chiticariu 和 W.-C. Tan，“数据库中的数据溯源：原因、方法与位置”，数据库基础与趋势，第 1 卷，第 4 期，第 379–474 页，2009年4月。[33] M. Mitchell、S. Wu、A. Zaldivar、P. Barnes、L. Vasserman、B. Hutchinson、E. Spitzer、I. D. Raji 和 T. Gebru，“用于模型报告的模型卡”，载于公平性、问责制与透明度会议论文集，FAT* ’19 系列，美国佐治亚州亚特兰大：ACM，2019年，第 220–229 页。[34] T. Gebru、J. Morgenstern、B. Vecchione、J. W. Vaughan、H. Wallach、H. D. III 和 K. Crawford，“数据集数据表”，*ACM* 通讯，第 64 卷，第 12 期，第 86–92 页，2021年11月，ISSN：0001-0782。[35] N. Barman 和 Y. Reznik，*QoE* 数据集数据表，https://github.com/NabajeetBarman/datasheet-for-qoe-datasets，2023年。[36] N. Barman、Y. Reznik 和 M. G. Martini，*BC-KU* 多屏幕数据集，https : / / github . com / NabajeetBarman/Multiscreen-Dataset，2023年。

附录

为便于理解该模板，此处提供了一个开源数据集 GamingVideoSET 的示例数据表。数据表模板及所有示例数据表均可在 GitHub 仓库 [35] 中找到。

QoE 数据集数据表。本模板适用于数据集创建者以及任何使用已发布数据集的人员。

注意：如果一项工作中使用或提出了多个数据集，请为每个数据集分别填写一份表格。

模板详情 （not （待修改）	创建日期:	2023年4月28日	此模板首次公开发布的日期
	修改日期	NA	模板修改日期（如适用）
	版本	v 1.0	当前版本号

第一部分：数据集概览		**值**	**其他 备注/网址 （如适用）**	**字段描述**
提供通用信息 关于您的数据集。它 将使读者能够 快速理解这是什么 数据集是关于什么的，以及它是否是 与他们的兴趣相关。	数据集名称	GamingVideoSET		数据集名称
	数据集缩写	NA		如果与数据集名称不同，则填写数据集名称的简称
	版本	v1.0		版本（默认：v1.0）
	创建/发布日期/年份	2018		数据集首次发布的日期。如果无法提供确切日期，请填写创建年份
	修改/更新日期 （dd-mm-yyyy）	NA		数据集最后更新日期（如不适用或自创建以来未更新，请填写不适用）
	代码仓库	https://kingston.box.com/v/游戏VideoSET		代码仓库链接
	引用（BibTex/PlainTex/URL）	N. Barman、S. Zadtootaghaj、S. Schmidt、M. G. Martini 以及 S. Möller，《GamingVideoSET：用于游戏的视频数据集 视频流应用，2018年第16届年度 网络与系统游戏支持研讨会 （NetGames），阿姆斯特丹，荷兰，2018年，第1-6页。		请添加必需引用，或提供包含该引用的网址
	许可证	NA		MIT许可证、Apache 2.0许可证、BSD许可证等。
	开放获取?	Yes	密码保护 数据集是	完全开放还是需要密码保护?
	联系信息	Nabajeet Barman (nabajeetbarman4@gmail.com)		作者/创建者的联系信息，用于发送有关数据集的问题或意见
	数据集大小	30 GB		完整数据集的大致大小（以 GB 为单位）
	附加信息?	NA		如有其他附加信息，请填写

第二部分：数据集描述		**值**	**附加 备注/网址 （如适用）**	**字段描述**
源视频 有关各个方面的信息 有关源视频各方面的信息 研究中使用的那些	源序列数量	24		所考虑的源序列数量（音频、视频等）
	内容类型	游戏		源序列的类型（游戏、自然、动画、计算机生成等）
	源序列是否可用?	Yes		是否将源序列作为数据集的一部分提供?
	源数据存储库	NA		如果数据集包含作为贡献内容的源视频，请填写 NA。
	位深度	8		8 位、10 位、12 位、16 位等。
	动态范围	SDR		标准动态范围/高动态范围（请尽可能补充详细信息，例如传递特性、伽马等。）
	帧率	30		源视频的帧率，例如 24、25、30、60、120、144 等。
	分辨率	1920x1080		源视频分辨率
	未经处理/用户生成	未经处理	无损采集	源序列是未经处理的，还是由用户生成／已压缩的?
	视频格式	原始视频		源视频的格式，例如 RawVideo、ProRes、HEVC 编码等。
	视频容器	YUV		MP4、MKV、WEBM、Y4M
	音频格式（如适用）	NA		音频格式（如有）
	视频/音频/视听	仅视频		源序列类型（例如，仅视频、仅音频、视听）
	SI/TI 信息	Yes		数据集中的 SI/TI [ITU-T 建议书 913] 是否可用?
	附加信息?	None		如有，请添加任何附加信息。
编码设置 有关编码的信息 数据集所使用的设置（也 简称 Hypothetical 参考电路（HRC））中 一些作品）。	编码器实现	FFmpeg		例如：FFmpeg、VTM、HM 等。
	编码器类型	软件		软件/硬件/两者
	软件实现类型	实际		参考/实际/两者
	码率控制	CBR		视频缓冲验证器、恒定码率、平均码率等
	编码速度/模式	极快		例如，FFmpeg：极快、中等、慢速等
	视频编解码器	H.264	Main 4.0 配置文件	H.264、高效视频编码、VP9、AV1、通用视频编码、AVS2、AVS3
	分辨率	1080p、720、480p		考虑的编码分辨率
	帧率	30		考虑的编码帧率
	位深度	8		编码视频序列比特率
	考虑的伪影	压缩和缩放		编码、缩放、网络（卡顿、质量变化、丢包等）。请补充详细信息。
	考虑的场景	HAS 直播流媒体		所考虑的流媒体场景类型
	附加信息?	NA		如有，请添加任何附加信息
处理后的视频序列	PVS 数量	576		PVS 总共有多少个?
	容器格式	MP4		MP4、MKV、WEBM、解码原始视频（YUV）等。
	附加信息	None		如有任何其他信息，请补充。

第三节：主观质量评估		**值**	**其他 评论/网址 （如适用）**	**字段描述**
主观测试设置	主观测试环境	基于实验室		受控（实验室）、公共环境、基于家庭、众包、混合
	ITU建议书	BT.500-14		请注明测试是否遵循了任何特定的 ITU建议书，例如 BT.500-14、P.910 或 P.913。
	主观测试方法和评分 量表	绝对类别评分 (1-5)		绝对类别评分（1–5分）、绝对类别评分（0–10分）、双刺激损伤量表、连续质量评分? 如果采用的不是标准化方法，请详细说明。
	评分量表	总体		总体 / 连续（使用滑块）/ 两者皆可
	显示器类型	台式显示器		电视、移动设备、平板电脑、台式显示器
	显示尺寸	24"	ViewSonic 显示器 显示器	所用显示器尺寸
	显示器分辨率	1920x1080		设备型号链接（如有）
	观看距离	3H		与屏幕之间的观看距离
	观看角度	Yes	论文中	
	主观测试示意图或照片	NA		是否有展示实验室测试布置的示意图或照片？如果有，请注明其位置。 有吗?
	测试室中的光强	NA		是否测量了设备、室内照度、伽马值等？如果是，请补充详细信息
	附加信息	NA		如有，请添加有关测试环境的任何附加信息。

测试方法	视频播放软件	VLC		用于播放视频的播放器（例如 VLC、MPV、Matlab、FFplay 等）
	测试软件	专有软件		您使用了哪种软件来播放视频并记录主观意见评分？例如： Matlab、MSU VQMT、TUIL AVRateNg、Subjectify.us、专有软件、纸质版。请添加适用的参考
	培训	Yes	使用其他游戏序列	在正式测试之前是否进行过培训？
	播放列表随机化	Yes		是否针对每场测试/每位参与者对播放列表进行了随机化？请解释。
	播放	参与者自主控制		刺激材料播放是如何进行的？例如，先播放固定时长的视频，然后显示灰屏固定时长，并由测试参与者自主控制。
	升频（如使用）	双三次插值		所用升频滤波器（双三次插值、Lanczos-3、Lanczos-5、超分辨率）
	测试类型	仅视频		仅视频、视听、仅音频
	测试受试者人数	24		参与主观测试的人数
	每场测试的受试者人数	1		单场测试中有多少名测试对象参与？
	测试序列数量	90		主观测试中纳入的测试序列数量
	附加信息	NA		如有，请添加有关测试方法的任何附加信息
测试参与者	测试对象招募	大学生和教职员工		研究中的测试对象是如何招募的？大学生和教职员工、众包平台、亲友等
	是否进行了预筛选？	两项测试均在实验室进行		在主观测试中，是否进行了视力测试或其他任何预筛查测试？ 参与者进行了视力测试或其他预筛选测试？如果有，请说明具体测试项目（例如 Sellen 视力测试，或使用 Ishira 图表进行的色盲测试、使用立体蝴蝶测试进行的 3D 视觉检查等）
	专家还是非专家	非专家		请说明测试对象是专家（从事视频流媒体工作的人）还是非专业人士，还是两者兼有？
	性别分布	男性占32%，女性占68%。		测试对象的性别分布是怎样的？
	年龄分布	（中位数）29		测试参与者的年龄分布情况如何？
	其他人口统计信息 信息	Yes		是否有其他人口统计信息数据，例如观看设备偏好、数量的... 每小时可观看的视频、游戏体验等是否可用（如适用）？如果可以，请提供。 详情
	附加信息	NA		如有关于测试参与者的任何附加信息，请添加
主观评分分析	筛选后	Yes	请采用基于 IQR 的方式。 详情请参阅论文	您是否进行了任何离群值分析？如果是，您采用了哪种方法论？例如，相关性分析。基于 Z 分数
	是否移除主观评分？	Yes		离群值分析后是否移除了任何主观评分？
	主观评分	MOS		主观评分如何报告？仅报告主观平均分/DMOS？
	主观评分变换	Yes	请参阅 Ref-Pezulli	您是否对原始主观评分进行了任何映射或变换？如果是，采用了哪种方法？（例如，用于估计总体平均体验质量的广义线性模型（GLM）[Ref-Pezulli]。请补充详细信息？
	单项主观评分	是，作为附加数据集的一部分	参见下方注释 1	是否提供单独的主观评分？
	统计分析？	No		是否进行了任何统计分析，例如计算置信区间等？
	附加信息	NA		如有关于主观评分后处理和分析的任何附加信息，请补充。 可用
其他意见		None		补充上述内容未涵盖的任何有关主观测试的信息。 与读者相关
第四节：客观质量评估		**值**	**补充评论/网址（如适用）**	**字段描述**
质量指标	考虑的质量指标	峰值信噪比、结构相似性指数、视频多方法评估融合		您在这项工作中评估了哪些质量指标？例如：峰值信噪比、结构相似性指数、视频多方法评估融合、LPIPS,
	指标类型	传统方法、机器学习		您使用的是哪种类型的指标？机器学习/深度学习/传统方法、混合
	池化	VQMT 工具	参见注释 2	使用了哪种池化方法（例如均值、闵可夫斯基求和、调和平均数等）以获得最终的视频质量评分（如适用）。
	实现及版本	数据集仅提供了指标分数 在数据集中可用。		如有，请添加所用指标的开源实现链接网址列表（FFmpeg、GitHub、商业软件、专有软件等）。此外，请注明该指标的版本和/或您在本研究中使用的实现方式
	附加信息?	NA		添加任何可能与用户相关的质量指标附加信息
模型评估	性能度量 已考虑	皮尔逊线性相关系数和斯皮尔曼等级相关系数		采用哪些指标来评估质量指标的性能（例如，皮尔逊线性相关系数、斯皮尔曼等级相关系数、均方根误差、肯德尔相关系数、决定系数等）
	已执行映射	None		是否执行了任何变换/拟合（线性/三阶多项式/逻辑回归）再计算性能分数?
	统计显著性检验	Yes	参见注释 2	是否进行了任何统计显著性检验，以比较所比较/评估的指标的性能? 比较/评估?
	模型训练	NA		是否对模型进行了训练/重新训练？如果是，如何进行?
	模型复杂度	NA		如果可以，请补充说明模型复杂度是如何评估的。
	附加信息?	NA		添加任何可能与用户相关的模型评估附加信息。
其他意见		None		请添加上述未涵盖的有关客观评估的任何附加信息 且可能与读者相关
第五节：伦理考量		**值**	**补充评论/网址（如适用）**	**字段描述**
请在此部分 说明任何与伦理/《通用数据保护条例》相关的信息	伦理审查	Yes	柏林工业大学	在创建数据集之前，是否进行了任何伦理审查流程?
	预期用途	仅限非商业研究用		请描述此数据集的预期用途?
	知情同意书（用于主观评估）测试）	NA		链接至在测试对象参与前提供给他们的样本知情同意书，如果可用
	伦理/GDPR相关问题?	某些序列可能包含暴力场景	请参阅该文档。 自述文件中的数据集	是否存在任何可能与该数据集的读者/用户相关的伦理问题?
	还有其他意见吗?	NA		是否还有与用户相关的其他伦理考量或信息? 这个数据集?
第六节：补充信息		**值**	**补充评论/网址（如适用）**	**字段描述**
	是预览版/网页版的有源视频吗?	No		是否有可供查看/收听的预览/样本序列，以便了解该彩票计划? 数据集？当数据集规模巨大，而读者希望确认时，这很有用。 在下载之前，请确认其符合您的要求。

回答其他问题。 此处提到的问题，例如 适用。如果有任何 您#+#+#+#+other information that you 如有需要补充的内容，请添加。 在这些内容的末尾 问题	该数据集由谁创建?	学术合作	金斯顿大学， 伦敦和 技术 大学，柏林， 德国	公司、大学、合作
	是谁资助创建了该 数据集?	欧盟和德国研究基金会	欧盟地平线2020 资助协议编号 643072 和德国研究基金会 项目 MO 1038/21-1。	项目、公司资助、自筹资金、无资助
	数据集中是否有任何内容 属于机密或可能会 限制其使用?	适用公平使用政策	请参阅 中的自述文件 数据集	例如，该数据集是否仅用于研究用途和非商业用途
	谁进行了主观 测试?	作者		主观测试是由数据集的作者/创建者进行的吗？还是外包给了 外部第三方？请提供详细信息。
	数据集所在的现有作品 正在使用	请参阅参考文献 [3-10]	其他内容可在此处找到 在 Google Scholar 上	请添加您或其他人使用过此数据集的现有研究成果的参考文献 （这适用于可能已发布一段时间的现有数据集）
请在此处补充 任何附加信息 您认为可能 对读者有帮助的内容 在本文中未作讨论 数据表问卷	注1	对 VQA 性能的更详细分析是 在后续出版物中进行。请参阅参考文献 [1]。		附加评论/信息
	注释 2	使用 [2] 中提供的 VQMT 工具进行指标 计算		其他评论/信息
	注释 3			其他评论/信息
	注4			其他评论/信息
	注5			其他评论/信息
	注6			其他评论/信息

参考资料

[1]	Nabajeet Barman、Steven Schmidt、Saman Zadtootaghaj、Maria G. Martini 和 Sebastian Möller。2018年。《被动游戏视频的视频质量评价指标评估》流媒体。在第 23 届分组视频研讨会（PV '18）论文集。美国纽约州纽约市计算机协会，7–12。https://doi.org/10.1145/3210424.3210434
[2]	VQMT - 视频质量测量工具，https://github.com/rolinh/VQMT
[3]	N. Barman、E. Jammeh、S. A. Ghorashi 和 M. G. Martini，《基于机器学习的被动游戏视频流媒体应用无参考视频质量估计》，发表于 IEEE Access，第7卷，第74511–74527页，2019年，doi：10.1109/ACCESS.2019.2920477】【。
[4]	Utke、M.、Zadtootaghaj、S.、Schmidt、S. 等。NDNetGaming——开发一种用于游戏视频质量预测的无参考深度 CNN。《Multimed Tools Appl》81，3181–3203（2022年）。https://doi.org/10.1007/s11042-020-09144-6
[5]	S. Zadtootaghaj、N. Barman、S. Schmidt、M. G. Martini 和 S. Möller，《NR-GVQM：一种无参考游戏视频质量度量》，2018年 IEEE 国际多媒体研讨会（ISM），中国台湾台中，2018年，第131-134页，doi：10.1109/ISM.2018.00031。
[6]	S. Göring、R. R. R. Rao 和 A. Raake，《nofu——一种用于游戏内容的轻量级无参考像素视频质量模型》，2019年质量第十一届国际会议上的..... 多媒体体验（QoMEX），柏林，德国，2019年，第 1-6 页，doi：10.1109/QoMEX.2019.8743262。
[7]	R. R. R. Rao 等人，“基于比特流的视频质量模型 ITU-T P.1204.3 在游戏内容上的大规模评估”，2020 年 IEEE 第 22 届多媒体信号国际研讨会 处理（MMSP），坦佩雷，芬兰，2020 年，第 1-6 页，doi：10.1109/MMSP48831.2020.9287055。
[8]	J. Le Tanou 和 N. Francisco，《视频游戏直播的编码优化》，IBC 2022，阿姆斯特丹，
[9]	Sam Van Damme，用于视频流游戏的客观质量评估， https://libstore.ugent.be/fulltxt/RUG01/002/785/859/RUG01-002785859_2019_0001_AC.pdf
[10]	Nasim Jamshidi Avanaki、Steven Schmidt、Thilo Michael、Saman Zadtootaghaj 和 Sebastian Möller。2022年。Deep-BVQM：一种基于深度学习和比特流的视频质量模型。载于第30届 ACM 国际多媒体会议（MM '22）论文集。计算机协会，美国纽约州纽约市，915–923。 https://doi.org/10.1145/3503161.3548374

参考文献（提供模板描述中使用的参考资料）

Ref-ITUP913	ITU-T 建议书《互联网视频和分发质量电视在任何环境下的视频质量、音频质量和视听质量主观评估方法》。 https://www.itu.int/rec/T-REC-P.913-202106-I. 2021年6月
Ref-LPIPS	Zhang, R., Isola, P., Efros, A. A., Shechtman, E. 和 Wang, O.，“学习型感知图像块相似度 （LPIPS）。”https://torchmetrics.readthedocs.io/en/stable/image/learned_perceptual_image_patch_similarity.html（2023）
Ref-Pezulli	S. Pezzulli、M. G. Martini 和 N. Barman，“基于主观测试的质量评分估计——超越受试者的主观平均分”，载于 IEEE 多媒体汇刊，第 23 卷，第 2505–2519 页，2021 年， doi: 10.1109/TMM.2020.3013349.
Ref-SSIM	Wang, Z., Bovik, A. C., Sheikh, H. R. 和 Simoncelli, E. P.,《图像质量评价：从误差 可见性到结构相似性》，IEEE 图像处理汇刊 13(4), 600–612 (2004)
Ref-VMAF	Netflix，《VMAF - 视频多方法评估融合》，https://github.com/Netflix/vmaf

除非另有明确说明，本建议模板采用的术语均定义于 ITU-T P.10/G.100 以及其他相关 ITU-T 建议书（P.910、P.913、P.1401）和 ITU-R BT.500 中。

数据表结束