平均意见分（MOS）与恰可察觉差异（JND）之间是否存在关系?

Jingwen Zhu[1], Hadi Amirpour[2],Wei Zhou[3],Patrick Le Callet[11] 南特大学´，南特中央理工学院、法国国家科学研究中心、LS2N、UMR 6004，南特，法国[2] 克拉根福大学信息技术系（ITEC），奥地利[3] 卡迪夫大学计算机科学与信息学院，卡迪夫，英国

arXiv:2602.17010v1 [eess.IV] 19 Feb 2026

摘要—**评估感知视频质量对于确保现代流媒体应用具备较高的体验质量（QoE）至关重要。现有主观数据集和视频质量指标（VQMs）覆盖了广泛的质量范围，但许多实际应用场景——尤其是面向高端用户的场景——关注高质量情形，需要更精细的粒度。恰可察觉差异（JND）已成为描述这些高端质量区域感知阈值的关键概念，并在感知码率阶梯构建中发挥重要作用。然而，JND 与更广泛使用的平均意见分（MOS）之间的关系仍不明确。本文基于一个现有的 JND 数据集开展了一项劣化类别评分（DCR）主观测试，考察 MOS 与 1st 和 2ndJND 的 75% 满意用户比例（SUR）点之间的对应关系。我们发现，JND 点处的 MOS 值通常符合理论预期（例如， 1st JND 的 75% SUR 对应 4.75），但反向映射——从 MOS 推断 JND——则存在歧义，因为不同 PVS 索引之间的置信区间相互重叠。统计显著性分析进一步表明，参与者人数有限的 DCR 测试可能无法检测出参考视频与 JND 视频之间有意义的差异。**

索引词—**恰可察觉差异（JND）、满意用户比例（SUR）、平均意见分（MOS）、视频质量**

I. 引言

评估人类感知的视频质量对于流媒体行业至关重要，有助于确保终端用户获得良好的体验质量（QoE） [1]。研究人员开展了大量研究，收集用于视频质量评估的主观数据集 [2], [3], [4], [5], ，并利用基于学习的方法，在这些数据集上训练和开发了先进的客观视频质量指标（VQM） [6], [7], [8], [9], [10], [11] 。这些指标旨在以计算高效且可扩展的方式估计人类对质量的感知，并广泛用于编码优化和自适应流媒体系统。

在受控条件下由人类观众对视频质量进行评分的主观测试，为评估客观指标提供了基准真值。视频质量指标（VQM）的有效性通常通过其预测结果与主观分数的相关程度来衡量，采用的统计指标包括皮尔逊相关系数或斯皮尔曼相关系数。高性能方法旨在与人类主观评分实现较强的相关性，从而确保其预测用户感知质量的可靠性。因此，客观视频质量指标的设计与基准测试仍与主观数据集的可用性和质量密切相关，这些数据集能够捕捉难以直接建模的细微感知效应。

现有的大多数数据集旨在涵盖**全范围**的质量等级。如表 I 所示，典型的主观视频质量研究会要求参与者按照国际电信联盟（ITU） [12]的建议，针对单个刺激采用 ACR-HR（绝对类别评分—隐藏参考）方法，使用从“优秀”到“差”的量表对视频质量进行评分；或者针对成对刺激，按照 DCR（退化类别评分）方法，使用从“无法察觉”到“非常恼人”的量表进行评分。由此得到的 MOS 或 DMOS（差分平均意见分）值可作为真实值，用于训练和评估客观视频质量指标、优化编码流程等。

表 I：ACR 用于单个刺激、DCR 用于成对刺激的评分量表

分数	ACR	DCR
5	优秀	无法察觉
4	Good	可察觉但不令人厌烦
3	Fair	略微令人烦恼
2	Poor	令人烦恼
1	Bad	非常令人烦恼

对于终端用户，尤其是高级订阅用户而言，他们期望在自己的设备上享受尽可能出色的视觉质量。这些用户最关心的是所交付的视频在感知上是否与原始视频无法区分，而非质量谱中端或低端的画质下降。传统的主观质量评估方法（如 ACR 和 DCR）会覆盖整个质量范围（表 I），从“差”到“优秀”，或从“非常令人烦恼”到“无法察觉”。这些方法适用于通用基准测试，但往往缺乏足够的精细度，难以区分高质量表示之间的细微差异。因此，多个高质量版本可能获得相近的高平均主观评分（例如，介于 4.5 和 5.0 之间），这使得在质量阶梯的顶端优化感知透明性或提高比特率使用效率时，可供参考的信息有限。

为解决这一局限，近年来的研究开始关注恰可察觉

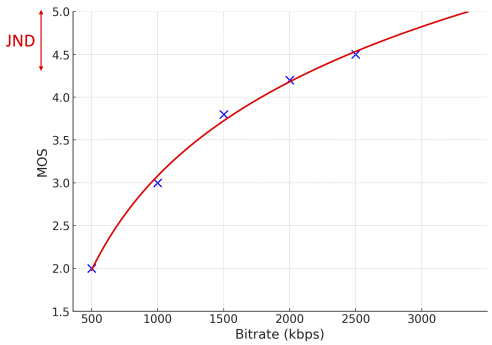


图 1：平均意见分与恰可察觉差异之间的关系。

差异（JND）这一概念 [13],[14], [15], [16]。JND 定义了观众相对于锚点——通常是无瑕疵视频或参考视频——刚开始察觉到差异时的感知阈值 1st JND [17], [18]。这对于自适应流媒体系统中的码率阶梯构建等应用尤为重要，因为确定比特率提升到何种程度后不再带来感知上的收益至关重要 [19], [20]。通过识别感知无损与有损表示之间的界限，JND 使服务提供商能够降低最高质量表示的比特率，同时不牺牲视觉质量。这不仅提高了压缩效率，还减少了带宽使用量和 CDN 成本。在此背景下，JND 可与平均意见分互为补充，在高质量区域提供更精细的分辨率，从而为编码、质量监测和用户体验优化提供更充分的决策依据。

本文旨在探讨的一个研究问题是平均意见分与恰可察觉差异之间的关系。具体而言，对于具有 1st 或 2nd 恰可察觉差异值的视频，其典型的平均意见分是多少？应该是 5、4.5，还是其他数值？（见图 1）据我们所知，此前尚无研究明确考察过这种关系。为填补这一空白，我们基于现有的恰可觉差视频数据集 [21], 设计了一项主观测试，以便更好地理解平均意见分与恰可察觉差异之间的关系。定义两者之间的关系仍是一项尚待解决的挑战，因为目前尚不清楚，对于恰好处于 1st 恰可察觉差异所定义的感知阈值上的视频，其平均意见分应为多少。

II. 主观测试设计概述

本节概述了本工作所采用的主观测试设计。我们首先介绍构成分析基础的HD-VJND数据集。随后，我们引入满意用户比例（SUR）的概念，该概念汇总不同观察者的个体恰可察觉差异点，从而在群体层面理解感知阈值。最后，我们展示基于DCR的主观数据集，该数据集使用从HD-VJND数据集中选取的视频样本构建，通过与参考内容的成对比较来研究恰可察觉差异与平均意见分之间的关系，因为原始数据集缺少平均主观评分。

*A. HD-VJND*数据集

我们基于现有的恰可察觉差异数据集，即HD-VJND数据集 [21]，设计了主观测试。HD-VJND包含180个源参考内容（SRC）视频，这些视频采用内容选择技术从600多个候选中选出，旨在确保视觉内容的多样性 [22]。每个源参考内容时长约10秒，分辨率为1080p。所有源内容均以1080p和720p两种分辨率进行编码，恒定速率因子取值范围为17至31，步长为0.25。在恰可察觉差异搜索方法方面，采用了稳健二分搜索（RBS） [14]，并辅以预处理步骤以缩短搜索时间。

测试环境为参与者的家庭环境，使用下列所列的若干款 55 英寸显示器：

- 索尼 KD-55XH8094 x4
- 三星 QE55Q74TATXXC x2
- 三星 UE55TU8075U x2
- LG NanoCell 55NAN091 x2

共有 20 名参与者参加了这项研究，来自 10 个家庭，每个家庭有 2 名参与者。按照 ITU建议 [12]，参与者被要求坐在距离屏幕三倍屏幕高度的位置。他们还被要求拉上窗帘，并关闭任何直接照向屏幕的灯光，以避免反光。更多详情请参阅 [21]。

B. 使用 *SUR* 对恰可察觉差异分布进行建模

恰可察觉差异的取值因不同观察者而异。为了汇总一组参与者的恰可察觉差异数据， [14]中提出了满意用户比率（SUR）。如图2所示，个体恰可察觉差异的分布可以用概率密度函数（pdf）及其对应的累积分布函数（cdf）来建模。SUR曲线定义为 $SUR = 1 - cdf$，表示感到满意（即未察觉到任何差异）的用户所占比例。先前的研究 [14], [21] 通常将 1st 恰可察觉差异处75%的SUR阈值用作识别 2nd 恰可察觉差异的锚点；类似地， 2nd 恰可察觉差异的75% SUR点也常被用作 3rd 恰可察觉差异的锚点。

在HD-VJND数据集中，为了实现流畅播放，使用源视频在原始分辨率下的近无损压缩版本（恒定速率因子=1）作为1080p下 1st 恰可察觉差异搜索的锚点，而非原始视频。随后，将 1st 恰可察觉差异的75% SUR点用作 2nd恰可察觉差异搜索的锚点。 1st 恰可察觉差异搜索仅在1080p的编码分辨率下进行，而 2nd 恰可察觉差异搜索则进行两次——一次在1080p，一次在720p。这一设计基于凸包假设： 1st 恰可察觉差异应出现在与源相同的分辨率上，而 2nd 恰可察觉差异可能出现在1080p或720p。在两种编码分辨率下进行 2nd 恰可察觉差异搜索，有助于更好地理解在构建码率阶梯时应选择哪种分辨率。

C. 基于 *DCR* 的平均意见分收集

为了探究恰可察觉差异与平均意见分之间的关系，我们采用 DCR 作为主观测试方案，因为在

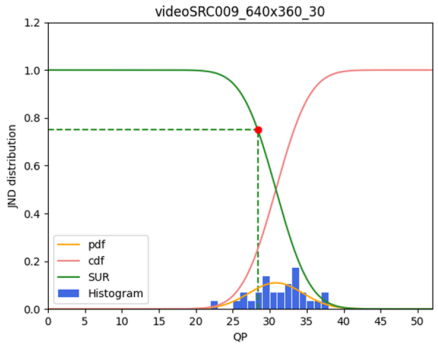


图 2：恰可察觉差异的满意用户比（SUR）示意图。

恰可察觉差异的主观测试中，观察者可以获得参考视频，并需要回答他们是否能察觉失真版本与参考视频之间的差异。因此，我们决定采用 DCR 而非 ACR，因为我们希望相对于参考视频评估失真程度。

DCR数据集基于HD-VJND数据集构建。我们保留了HD-VJND数据集中的相同源内容，并为每个源内容选取了 9 个处理后视频序列（PVS），如表II所示。PVS 的选取原则是，首先纳入 1st JND 的 75% SUR 点（PVS 索引 1），以及 1080p 和 720p 下的 2nd JND 点（PVS 索引 2 和 3）。随后，我们在 PVS 索引 1、2 和 3 附近额外抽取视频点。这些 PVS 将帮助我们了解 1st 和 2nd JND 在 75% SUR 处的平均意见分。

表 II：DCR数据集的 PVS 选取

PVS 索引	描述
0	参考视频
1	1st 恰可察觉差异的75%SUR 1080p
2	1080p下 2nd 恰可察觉差异的75%SUR
3	720p下 2nd 恰可察觉差异的75%SUR
4～7	PVS 1与PVS 2之间的4个点
8	PVS 1的恒定速率因子-1
9	PVS 3的恒定速率因子 +1

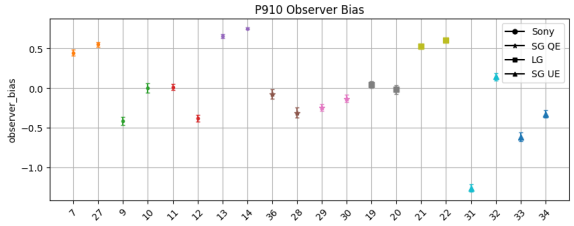
主观测试环境与第二节A部分所述的 HD-VJND数据集保持一致。为避免疲劳，对观察者的测试时长限制为最长 1 小时。

III. 结果与分析

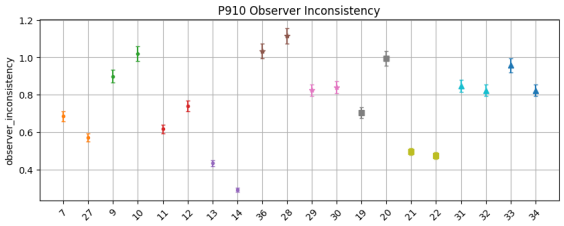
A. 数据筛选

收集 DCR数据集后，我们首先按照 ITU-T 建议书 P. 910 [23] 进行数据筛选。20 名参与者的偏差和一致性如图 3 所示。

观察者按家庭环境分组，同一种颜色表示来自同一家庭的参与者。例如，观察者 7 和观察者 27 来自相同的家庭环境，并使用同一台索尼电视。一个有趣的观察是，来自相同家庭环境的参与者通常表现出相似的偏差和不一致性。然而，对于使用



(a) 20 名参与者的偏差



(b) 20 名参与者的不一致性

图 3：DCR数据集中20名参与者的偏差与不一致性。共享相同家庭环境的参与者以相同颜色表示，而相同的标记形状表示相同的电视参考视频。

不同家庭环境中同一型号电视的参与者，并未观察到这种规律，这表明观看条件的影响可能大于显示类型的影响。这种现象也可能受到这样一个事实的影响：共同生活的参与者通常是伴侣，他们拥有相似的教育背景和习惯，因此在主观测试中的行为可能也较为相似。

B. 恰可察觉差异锚点处的平均意见分

我们计算了整个收集到的DCR数据集中PVS索引0、1、2、3对应的平均主观评分；结果如表 III 所示。原始分数 S_{pvs} 使用公式 (1) 映射为 S'_{pvs}，使参考视频的平均意见分为5。

$$S'_{pvs} = S_{pvs} - S_{ref} + 5 \tag{1}$$

表 III：整个数据集中PVS索引0–3处的平均主观评分和95%置信区间（CI）

PVS 索引	描述	平均主观评分	95%置信区间
0	参考视频	5	[4.90，5.10]
1	75%SUR of 1st 恰可察觉差异 1080p	4.79	[4.64，4.93]
2	1080p 条件下 2nd 恰可察觉差异的 75%SUR	4.51	[4.31，4.72]
3	720p 条件下 2nd 恰可察觉差异的 75% SUR	4.27	[4.01, 4.52]

理想情况下，对于对应于 1st 恰可察觉差异的 75% SUR 的视频（即 PVS 索引 1），SUR 的定义意味着 75% 的观察者无法察觉该视频与参考视频之间的任何差异。因此，在 DCR 研究中，这些观察者会投票选择“无法察觉”，分数为 5（见表 I）。其余 25% 的观察者能够察觉差异，会投票选择“可察觉但不令人厌烦”，分数为 4。因此，恰可察觉差异的 75% SUR 点的理想平均意见分应使用公式（2）计算，结果为 4.75。

$$MOS_{ideal}(75\%\ \text{SUR}_{\textbf{1st JND}}) = 5 \times 0.75 + 4 \times 0.25 = 4.75 \tag{2}$$

对于 2nd 恰可察觉差异的 75% SUR，在恰可察觉差异搜索过程中使用了 PVS 索引 1 的视频作为锚点。理想情况下，如果在 DCR 主观研究中使用同一视频（PVS 索引 1）而非 PVS 索引 0 作为参考视频，那么我们应当得到 2nd 恰可察觉差异的 75% SUR 对应的平均意见分 4.75。然而，在实际的 DCR 研究中，我们使用 PVS 索引 0 作为参考视频（它也是 1st 恰可察觉差异搜索的锚点）。由于该参考视频质量更高，感知差异也随之增大，因此我们预期得到的平均意见分会低于 4.75。恰可察觉差异的 75% SUR 点（即 PVS 索引 2）的理想平均意见分使用公式（3）和公式（4）计算。

$$\Delta = 5 - MOS_{ideal}(75\%\ \mathrm{SUR}_{1\mathrm{st\ JND}}) = 0.25 \tag{3}$$

$$\begin{aligned} MOS_{\mathrm{ideal}}(75\%\ \mathrm{SUR}_{\mathbf{2nd\ JND}}) &= MOS_{\mathrm{ideal}}(75\%\ \mathrm{SUR}_{\mathbf{1st\ JND}}) - \Delta \\ &= 4.75 - 0.25 = 4.5 \end{aligned} \tag{4}$$

从表 III 可以看出，1st 恰可察觉差异的 75% SUR 为 4.79，接近该视频的理想平均意见分。720p 中 2nd 恰可察觉差异的 75% SUR 点的平均意见分略低于 1080p。1080p 版本更接近根据公式（3）和公式（4）计算出的理想平均意见分。

图 4 展示了整个数据集中所有处理后视频序列索引的平均 DCR 主观意见分数及其 95%置信区间。正如预期，随着压缩强度增加，平均意见分通常会降低。1st 恰可察觉差异的 75% SUR 点处的平均意见分的 95%置信区间为 $[4.64, 4.93]$，这可解释为该恰可察觉差异等级的真实平均意见分有 95% 的概率落在此范围内。

尽管不同处理后视频序列索引的平均意见分略有差异，但它们的 95%置信区间往往存在重叠。因此，恰可察觉差异与平均意见分之间的关系本质上是单向的：

- **从恰可察觉差异推断平均意见分是可行的：**例如，我们可以合理推断，与第 1个 恰可察觉差异的 75% SUR 对应的平均意见分位于 95%置信区间 $[4.64, 4.93]$ 内。
- **从平均意见分推断恰可察觉差异存在歧义：**例如，如果某段视频在主观测试中的平均意见分为 4.7，我们就无法有把握地判断它对应的是第 1个 恰可察觉差异的 75% SUR，还是第 2个 恰可察觉差异的 75% SUR。

先前的研究 [24], [25]考察了视频质量指标（VQMs）（例如 VMAF [6], ）与 1st 恰可察觉差异之间的映射关系。由于大多数视频质量指标（包括 VMAF）都是基于主观测试中获得的平均意见分或差分平均意见分进行训练的，因此从视频质量指标到恰可察觉差异的映射也具有单向性。换言之，尽管恰可察觉差异等级可以与特定的视频质量指标分数相关联，但反向映射并非唯一确定—— [26]中也指出了这一点。

C. 统计显著性分析

在上一节中，我们获得了不同恰可察觉差异阶数对应的 DCR 平均意见分。本节针对每个源参考内容，在不同处理后视频序列索引之间进行单因素方差分析显著性检验，并计算存在统计显著差异的源内容占源内容总数的百分比。结果汇总于表 IV。可以观察到，只有 10.71% 的源内容表现出统计显著

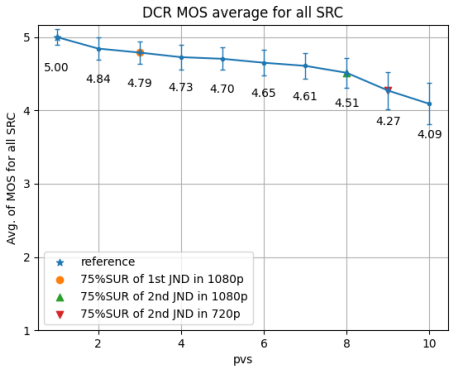


图 4：1080p 中 1st 恰可察觉差异和 2nd 恰可察觉差异，以及 720p 中 2nd 恰可察觉差异时，75% SUR 对应的 DCR 平均意见分值。（注：原始平均意见分已映射，以使参考视频的平均意见分为 = 5）

差异（p 值 < 0.05），即参考视频与 1st 恰可察觉差异阶数的 75% SUR 点之间存在差异。这表明，DCR 主观研究通常无法得出 PVS 0（参考视频）与 PVS 1（ 1st 恰可察觉差异阶数的 75% SUR）之间存在显著差异的结论，这可能是因为两段视频的质量过于相似，且参与者人数有限。然而，如表 III 和图 4 所示，平均意见分仍反映出这两段视频之间存在轻微差异。

表 IV：采用单因素方差分析得出的整个数据集中存在显著差异的视频百分比

PVS 索引对	显著差异百分比
(0, 1)	10.71%
(0, 2)	38.40%
(0, 3)	64.29%
(2, 3)	6.25%
(1, 2)	7.14%
(1, 3)	30.36%

IV. 结论

本文探讨了恰可察觉差异与通过 DCR 主观研究获得的平均意见分之间的关系，旨在回答以下问题：与 1^{st}和 2nd 恰可察觉差异相对应的平均意见分是多少？我们首先基于现有的 HD-VJND 数据集设计并开展了一项 DCR 研究。通过分析不同环境和显示设备下参与者的行为，我们发现，在相同设置下的参与者往往表现出相似的偏差和不一致性。

我们的结果揭示了恰可察觉差异与 DCR 主观意见分之间存在单向映射关系。具体而言，我们可以合理估算给定恰可察觉差异水平对应的平均意见分（例如，1st 恰可察觉差异的 75% SUR），但反向关系并不成立：知道某个视频的平均意见分，并不能确定其恰可察觉差异水平。

我们还进行了统计显著性分析，结果表明，参与者人数较少的 DCR 研究可能无法检测出恰可察觉差异视频与参考视频之间有意义的差异，因为它们的质量往往非常接近。未来，我们计划增加参与者人数，以提升双刺激连续质量评估协议的区分能力。

参考文献

[1] Wei Zhou、Hadi Amirpour、Christian Timmerer、Guangtao Zhai、Patrick Le Callet 和 Alan C. Bovik，《感知视觉质量评估：原理、方法与未来方向》，2025年，版本号：1。[2] Yilin Wang、Sasi Inguva 和 Balu Adsumilli，《用于视频压缩研究的 YouTube UGC 数据集》，载于*2019* 年 *IEEE* 第 *21* 届多媒体信号处理国际研讨会（*MMSP*）。IEEE，2019年，第 1–5 页。[3] Zeina Sinno 和 Alan Conrad Bovik，《感知视频质量的大规模研究》，*IEEE* 图像处理汇刊，第 28 卷，第 2 期，第 612–627 页，2018年。[4] Vlad Hosu、Franz Hahn、Mohsen Jenadeleh、Hanhe Lin、Hui Men、Tam´as Szir´anyi、Shujun Li 和 Dietmar Saupe，《康斯坦茨自然视频数据库》，2017年。[5] Z Ying、M Mandal、D Ghadiyaram 和 AC Bovik，《大规模社交视频质量（LSVQ）数据库》，在线：*https://github. com/baidut/PatchVQ*，第 8 卷，2020年。[6] Zhi Li、Anne Aaron、Ioannis Katsavounidis、Anush Moorthy 和 Megha Manohara，《迈向实用的感知视频质量指标》，*Netflix* 技术博客，第 6 卷，第 2 页，2016年。[7] Rakesh Rao Ramachandra Rao、Steve G¨oring、Werner Robitza、Alexan- der Raake、Bernhard Feiten、Peter List 和 Ulf W¨ustenhagen，《4K/UHD 基于比特流的模型标准：ITU-T P.1204.3——模型细节、评估、分析与开源实现》，载于*2020*年第十二届国际多媒体体验质量会议（*QoMEX*），阿斯隆，爱尔兰，2020年5月。[8] Haoning Wu、Chaofeng Chen、Jingwen Hou、Liang Liao、Annan Wang、Wenxiu Sun、Qiong Yan 和 Weisi Lin，《FAST-VQA：采用片段采样的高效端到端视频质量评估》，欧洲计算机视觉会议（*ECCV*）论文集，2022年。[9] Chenlong He、Qi Zheng、Ruoxi Zhu、Xiaoyang Zeng、Yibo Fan 和 Zhengzhong Tu，《COVER：全面的视频质量评估器》，载于*IEEE/CVF* 计算机视觉与模式识别会议（*CVPR*）研讨会论文集，2024年6月，第 5799–5809 页。[10] Hadi Amirpour、Jingwen Zhu、Patrick Le Callet 和 Christian Tim- merer，《用于 HTTP 自适应流媒体的实时视频质量指标》，载于*ICASSP 2024*，2024年4月，第 3810–3814 页。[11] Hadi Amirpour、Jingwen Zhu、Wei Z hou、Patrick Le Callet 和 Christian Timmerer，《VQM4HAS：用于 HTTP 自适应流媒体中 HEVC 视频的实时质量指标》，*IEEE* 多媒体汇刊，第 1–12 页，2025年。[12] 国际电信联盟，《P.910：多媒体应用的主观视频质量评估方法》，2021年11月。[13] Jingwen Zhu，真实环境中的高端视频流媒体质量：满意用户比的测量与预测，博士论文，南特大学，2024年。[14] Haiqiang Wang、Ioannis Katsavounidis、Jiantong Zhou、Jeonghoon Park、Shawmin Lei、Xin Zhou、Man-On Pun、Xin Jin、Ronggang Wang、Xu Wang、Yun Zhang、Jiwu Huang、Sam Kwong 和 C. C. Jay Kuo，“VideoSet：基于JND测量的大规模压缩视频质量数据集”，视觉通信与图像表示期刊，第46卷，第292–302页，2017年7月。[15] Yun Zhang、Huanhua Liu、You Yang、Xiaoping Fan、Sam Kwong和C. C. Jay Kuo，“基于深度学习的压缩视频恰可察觉差异与感知质量预测模型”，*IEEE*视频技术电路与系统汇刊，第32卷，第3期，第1197–1212页，2022年3月，会议名称：IEEE视频技术电路与系统汇刊。[16] Hadi A mirpour、Jingwen Zhu、Raimund Schatz、Patrick Le Callet和Christian Timmerer，“探索提升用户满意度的比特率成本：恰可察觉差异（JND）视角”，载于*2024*年数据压缩会议（*DCC*）。IEEE，2024年，第432–441页。[17] Joe Yuchieh Lin、Lina Jin、Sudeng Hu、Ioannis Katsavounidis、Zhi Li、Anne Aaron和C.-C. Jay Kuo，“编码图像/视频JND测试的实验设计与分析”，载于数字图像处理应用 *XXXVIII*。2015年9月，第9599卷，第324–334页，SPIE。[18] Di Yuan、Tiesong Zhao、Yiwen Xu、Hong Xue和Liqun Lin，“视觉恰可察觉差异：视频编码中的一种感知测量”，*IEEE Access*，第7卷，第29014–29022页，2019年，出版方：电气与电子工程师学会（IEEE）。[19] Vignesh V Menon、Hadi Amirpour、Mohammad Ghanbari和Chris- tian Timmerer，“面向感知的逐标题编码，用于自适应视频流”，载于*ICME2022*。2022年7月，第1–6页，IEEE。[20] Masaru Takeuchi、Shintaro Saika、Yusuke Sakamoto、Tatsuya Na- gashima、Zhengxue Cheng、Kenji Kanai、Jiro Katto、Kaijin Wei、Ju Zengwei和Xu Wei，“利用恰可察觉差异估计实现感知质量驱动的自适应视频编码”，载于*2018*年图像编码研讨会（*PCS*），美国加利福尼亚州旧金山，2018年6月，第179–183页，IEEE。[21] Jingwen Zhu、Anne-Flore Perrin和Patrick Le Callet，“压缩高清视频的恰可察觉差异与满意用户比的主观测试方法优化和预测框架”，载于*2022*年图像编码研讨会（*PCS*）。IEEE，2022年，第313–317页。[22] Suiyi Ling、Yoann Baveye、Deepthi Nandakumar、Sriram Sethuraman和Patrick Le Callet，“迈向更优质的高质量视频评估”，载于第一届视觉多媒体应用体验质量（*QoE*）研讨会论文集，2020年，第3–9页。[23] ITU-R，“多媒体应用的主观视频质量评估方法”，ITU-R 建议书P.910，2022年。[24] Hadi Amirpour、Raimund Schatz和Christian Timmerer，“介于二与六之间？迈向VMAF码率阶梯中JND步长的准确估计”，载于*QoMEX2022*，2022年9月，第1–4页。[25] Jingwen Zhu、Hadi Amirpour、Raimund Schatz、Christian Timmerer和Patrick Le Callet，“通过视频质量指标增强VMAF代理的满意用户比（SUR）预测”，载于*2023*年*IEEE*国际视觉通信与图像处理会议（*VCIP*）。IEEE，2023年，第1–5页。[26] Jingwen Zhu、Hadi Amirpour、Raimund Schatz、Patrick Le Callet和Christian Timmerer，“超越曲线和阈值——为压缩视频的满意用户比引入不确定性估计”，载于*2024*年图像编码研讨会（*PCS*）。IEEE，2024年，第1–5页。