arXiv:2601.10448v1 [cs.MM] 15 Jan 2026

采用 VVC、LCEVC 和 ML-VVC 编码的超高清视频主观评价

Naeem Ramzan，SMIEEE，Muhammad Tufail Khan 计算、工程与物理科学学院 苏格兰西部大学
naeem.ramzan@uws.ac.uk

摘要—**本文介绍了一种多层视频编码配置的主观质量评估结果，该配置将低复杂度增强视频编码（LCEVC）作为增强层应用于通用视频编码（VVC）基础层之上。评估采用此前为 MPEG 多层视频编码评估制定的相同测试方法和条件，LCEVC 增强层使用 LCEVC 测试模型（LTM）8.1 版进行编码。**

测试比较了由高清VVC基层与 LCEVC 增强层生成的重建超高清输出，以及两种参考情况：上采样VVC基层解码和多层 VVC（ML-VVC）。考虑了两个工作点，其增强层分别约占总比特率的 10% 和 50%。主观评估采用降级类别评分（DCR）方法，由 25 名参与者完成，数据集包含 15 个标准动态范围和高动态范围序列。

报告结果包括平均意见分（MOS）及其95%置信区间，从而能够比较既定测试范围内不同编码方法和工作点的感知质量。

索引词—**低复杂度增强视频编码（LCEVC）、通用视频编码（VVC）、多层视频编码、空间可扩展性、主观质量评估、降级类别评分（DCR）。**

I. 引言

在 [1] 中，报告了根据多层视频编码测试计划 [5] 进行主观测试的结果。该测试考察了三种MPEG多层视频编码技术，这些技术在功能、性能和复杂度方面具有不同特征，并且能够进行双层编码，包括一个高清基础层和一个超高清增强层（2倍空间可扩展性）。这些技术包括 SHVC（HEVC可扩展扩展）、VVC多层配置文件和 LCEVC。在该测试中，LCEVC使用2025年1月发布的参考软件版本进行测试，即LTM 7.0 [2], ，并采用了一些受限配置。该报告发布后，LTM发布了新版本，即LTM 8.1 [7]。本文提供了一项测试的结果，该测试采用了 [1]和 [5]中使用的完全相同的方法，但使用了LTM 8.1，并采用参考软件手册中规定的配置。

II. 方法

本节介绍本文所报告测试中采用的方法。

A. 数据集

表 I 列出了本文使用的序列。有关这些序列的更多详情，请参阅 [5]。

表 I 序列列表

序列	分辨率	帧数和帧率	色彩空间
BodeMuseum	3840×2160	600（60 fps）	标准动态范围，BT.2020
地铁	3840×2160	600（60 fps）	标准动态范围，BT.2020
OberbaumSpree	3840×2160	600（60 帧/秒）	标准动态范围，BT.2020
地铁树	3840×2160	600（60 帧/秒）	标准动态范围，BT.2020
滨水区	3840×2160	240（24 帧/秒）	标准动态范围，BT.2020
大型足球广告	3840×2160	600（60 帧/秒）	混合对数伽马，BT.2020
AMS01	3840×2160	600（60 fps）	混合对数伽马，BT.2020
AMS02	3840×2160	600（60 fps）	混合对数伽马，BT.2020
AMS05	3840×2160	600（60 帧/秒）	混合对数伽马，BT.2020
女子足球	3840×2160	600（60 帧/秒）	感知量化，BT.2020
绿色山脉1	3840×2160	250（25 fps）	感知量化，BT.2020
KitchenPressin	3840×2160	600（60 fps）	感知量化，BT.2020
RiverPlate1	3840×2160	600（60 帧/秒）	感知量化，BT.2020
蒂尔加滕大道	3840×2160	600（60 帧/秒）	感知量化，BT.2020
WalkInPark	3840×2160	600（60 fps）	感知量化，BT.2020

B. 测试简介

本次测试旨在评估结合VVC基础层的LCEVC，与经上采样的VVC基础层之间的性能。测试内容的分辨率为高清（VVC基础层）+ 超高清（LCEVC增强层），其中高清版本为VVC基础层的UHS上采样版本。每个增强层都叠加在基础层之上，以实现两个目标工作点之一：（a）E10：增强层占总比特率的10%；（b）E50：增强层占总比特率的50%。为便于与 [1], 中报告的测试进行校准对比，本次测试还评估了上采样基础层和ML-VVC。

C. 配置

本次测试中的 VVC基础层和 ML-VVC增强层使用 VTM 23.8 [6]进行编码，采用 CTC 随机访问配置以及相应的标准动态范围 [8] 和高动态范围 [9] CTC 测试条件。本次测试采用了平坦量化矩阵设置，并禁用了局部量化自适应。测试所用的配置与以下配置相同：

表 I 量化参数

序列名称	增强层百分比	BL QP	ML-VVC QP	LCEVC	
				SW 2	SW 1
BodeMuseum	10%	29	41	4100	1000
	50%		34	1250	3000
地铁	10%	29	39	2750	32767
	50%		33	1000	32767
OberbaumSpree	10%	29	37	2500	32767
	50%		33	900	1000
地铁树	10%	31	41	3000	32767
	50%		35	1500	32767
滨水区	10%	27	43	3500	32767
	50%		35	1500	32767
大型足球广告	10%	31	40	2750	32767
	50%		34	1100	32767
H3_AMS01	10%	29	42	3000	32767
	50%		35	1400	32767
H3_AMS02	10%	29	41	3000	32767
	50%		35	1300	32767
H3_AMS05	10%	29	41	2500	32767
	50%		34	1250	1250
女子足球	10%	31	38	1750	32767
	50%		34	700	32767
GreenMountain1	10%	27	35	3000	32767
	50%		31	1000	32767
厨房调料	10%	27	39	2500	3000
	50%		30	600	32767
拉普拉塔河 1_	10%	27	38	2250	32767
	50%		31	900	32767
蒂尔加滕大道	10%	29	38	3000	32767
	50%		33	1000	32767
公园漫步	10%	27	34	2500	3000
	50%		30	750	32767

在 [5] 中使用的配置。本次测试中的 LCEVC增强层使用 LTM 8.1 编码，配置如 [7] 所述。用于生成基础层并对其进行上采样的下采样器和上采样器与 [5] 中所述的一致。

D. 量化参数

本次测试使用的量化参数列于表 II。具体而言，基础层和 ML-VVC 使用的量化参数（QPs）与 [5]中使用的相同。对于 LCEVC，则采用一种基于凸包法的方法，根据特定的码率和失真确定最佳量化参数对，从而设定两个标准量化参数 SW1 和 SW2。

E. 测试设计与执行

测试在苏格兰西部大学的观看实验室进行。评估采用劣化类别评分（DCR）方法。测试以盲测方式进行，编码均经过匿名处理并以随机顺序呈现，共有 25 名受试者参与。评估方法与 [5] 中所述的方法密切一致。

参与者使用 11点评分量表对视觉质量进行评分，范围从 0（质量最低）到 10（质量最高），如表 III 所示。

表 III 11级损伤量表

评分	损伤项目	
10	不可感知	
9	略微可感知	某处
8		各处
7	可感知	某处
6		各处
5	明显可感知	某处
4		各处
3	令人烦恼	某处
2		各处
1	极其令人烦恼	某处
0		各处

III. 实验结果

图1展示了实验结果的图表。此外，表 IV 和表 V 报告了总体平均意见分数，并标出了正向和负向的 95%置信区间（置信区间）。结果按增强层（EL）占比为 10% 和 50% 的情况（E10 / E50）分组展示。

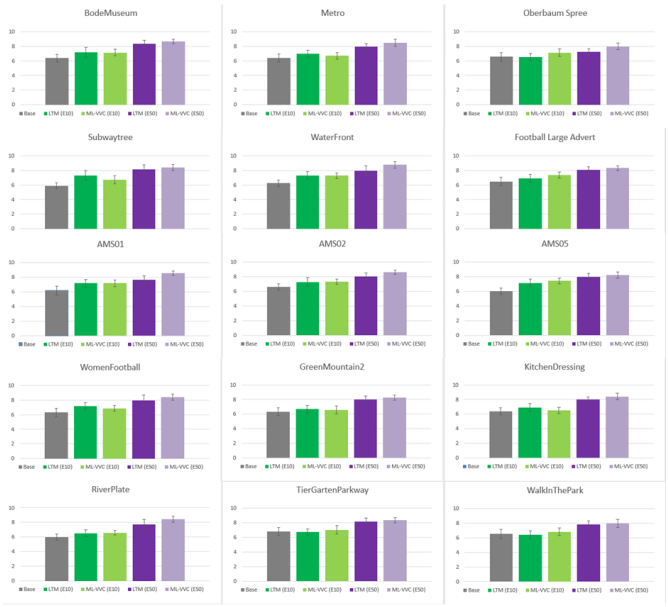

图1. 各序列的平均意见分数和置信区间图，涵盖三种方案：基础层（灰色）、LCEVC E10（深绿色）、ML-VVC E10（浅绿色）、LCEVC E50（深紫色）和 ML-VVC E50（浅紫色）

IV. 结论

本研究重新评估了在 VVC 基层上应用 LCEVC 增强层的主观表现，采用 LTM 8.1 以及此前用于 MPEG 多层视频编码评估的相同测试计划和实验室条件

表 IV 多层方法对比（E10）

序列名称	Base		LCEVC E10		ML-VVC E10	
	MOS	CI	MOS	CI	MOS	CI
博德博物馆	6.36	0.51	7.16	0.65	7.12	0.46
地铁	6.40	0.57	6.96	0.50	6.68	0.44
OberbaumSpree	6.56	0.59	6.52	0.42	7.12	0.52
地铁树	5.88	0.39	7.28	0.69	6.72	0.53
滨水区	6.24	0.46	7.32	0.50	7.28	0.40
大型足球广告	6.48	0.57	6.88	0.55	7.36	0.45
AMS01	6.16	0.61	7.16	0.50	7.16	0.44
AMS02	6.64	0.41	7.28	0.58	7.32	0.37
AMS05	6.04	0.40	7.16	0.53	7.44	0.38
女子足球	6.32	0.58	7.20	0.45	6.88	0.39
绿色山脉1	6.32	0.53	6.68	0.53	6.56	0.57
厨房调料	6.40	0.48	6.88	0.55	6.52	0.42
拉普拉塔河	6.00	0.42	6.52	0.46	6.60	0.34
蒂尔加滕大道	6.80	0.54	6.76	0.43	7.04	0.60
公园漫步	6.52	0.59	6.40	0.57	6.80	0.51

表 V 多层方法对比（E50）

序列名称	Base		LCEVC E50		ML-VVC E50	
	MOS	CI	MOS	CI	MOS	CI
博德博物馆	6.36	0.51	8.32	0.47	8.64	0.29
地铁	6.40	0.57	7.92	0.45	8.48	0.49
OberbaumSpree	6.56	0.59	7.20	0.48	7.96	0.43
地铁树	5.88	0.39	8.16	0.58	8.40	0.42
滨水区	6.24	0.46	7.96	0.67	8.76	0.46
大型足球广告	6.48	0.57	8.04	0.47	8.32	0.33
AMS01	6.16	0.61	7.64	0.54	8.52	0.30
AMS02	6.64	0.41	8.04	0.47	8.64	0.29
AMS05	6.04	0.40	8.00	0.48	8.24	0.43
女子足球	6.32	0.58	7.96	0.73	8.44	0.42
绿色山脉1	6.32	0.53	8.00	0.48	8.24	0.35
厨房调料	6.40	0.48	8.00	0.38	8.40	0.45
拉普拉塔河	6.00	0.42	7.76	0.67	8.44	0.42
蒂尔加滕大道	6.80	0.54	8.20	0.45	8.36	0.38
公园漫步	6.52	0.59	7.80	0.54	7.96	0.53

评估涵盖 15 个超高清标准动态范围和高动态范围序列，并考察了相对于总比特率的两个增强层比特率工作点：E10 和 E50。

在所测试的内容中，基于 LCEVC 的增强和 ML-VVC 均始终取得了高于对 VVC基础层进行简单上采样的主观质量评分。在 E10 工作点，增强方法与上采样基层之间观测到的平均意见分数差异通常为正值，但也观察到因序列而异的变化，并且在某些情况下，差异仍处于报告的 95%置信区间内。在 E50 工作点，所有序列中两种多层方法的平均意见分数均更高，这反映了分配给增强层的比特率有所增加。

直接比较 LCEVC 和 ML-VVC 时，结果表明，在测试设计的限制范围内，两种方法在等效工作点下均能提供相当的主观质量。两种增强方法之间的差异取决于序列，并且在大多数情况下处于相应的置信区间内，表明在所测试条件下，两者之间不存在统计上稳健的差异。

需要指出的是，这些发现仅适用于测试计划所定义的所选配置、编码器版本、量化设置以及下采样/上采样过程。结果未涉及编码器或解码器复杂度、实现成本、延迟，也不适用于评估工作点之外的性能。后续研究可将此分析扩展至更多比特率点、其他重采样方法及不同的编码器配置，从而更全面地刻画多层编码方法之间的权衡。

V. 致谢

我们感谢 V-Nova 团队提供本次评估所用的编码器配置、量化参数设置和编码比特流。所有材料均以匿名形式提供，并附有相应脚本，使实验室能够在事先不了解各比特流所采用技术的情况下自动解码。

参考文献

[1]《MPEG 多层视频编码技术主观评估最终报告》，ISO/IEC JTC 1/SC 29/AG 5 N178，2025年10月。

[2] 7 0 – –

[3] 910 2021

https://git.mpeg.expert/MPEG/Video/LCEVC/LTM/-/tags/LTM ITU-T P.（）建议书，多媒体应用的主观视频质量评估方法。[4] ITU-R BT.500-14建议书（2019年），电视图像质量主观评估方法。[5] ISO/IEC JTC 1/SC 29/AG 5，《MPEG 多层视频编码技术评估测试计划勘误》，文件 AG5N167，第20次会议，韩国大田，2025年7月。[6] https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware VTM/-/tree/VTM-23.8?ref type=tags [7] https://git.mpeg.expert/MPEG/Video/LCEVC/LTM [8] F. Bossen、X. Li、V. Seregin、K. Sharman、K. S¨uhring，《标准动态范围 4:2:0 10 位视频的 VTM 和 HM 通用测试条件及软件参考配置》，JVET-AK2010，2025年3月。[9] A. Segall、E. Franc¸ois、W. Husak、S. Iwamura、D. Rusanovskyy，《HDR/WCG 视频的 VTM 和 HM 通用测试条件及评估流程》，JVET-AC2011，2023年2月。