# 数字音视频编解码技术标准工作组®

AVS M8987: 2025年06月11日

来源: 李卓元 廖峻骐 唐传波 张昊田 卞逸凡 李瑶 盛锡华 梁雄壮 李宇琪

冯新民 左之睿 李佳林 汪家振 蒋易恒 向辉 李礼 刘东 吴枫

(中国科学技术大学)

标题: 一种用于 EEM 的测试集 USTC-TD 及其测试基准

状态: 输入提案

# 摘要

本提案提出一种用于 EEM 测试的新视频编码测试数据集 USTC Test Dataset (USTC-TD)。在本提案中,首先,基于该数据集,将其与 EEM 中现采用的 AVS 测试集及现端到端视频编码相关研究中常采用的数据集(HEVC CTC, VVC CTC, MCL-JCV, UVG)进行了定量的评估分析(空域丰富度,时间丰富度),验证了其能够良好地补偿现有数据集的测试缺陷。进一步地,基于该数据集,分析了 EEM 及其他视频编码标准(AVS, VVC)和基于学习的视频编码方案(DCVC-DC,DCVC-FM,DCVC-RT)在客观质量测试指标(PSNR,VMAF)下的性能测试结果,提供了该数据集相关的测试基准。

# 正文

# 1. USTC-TD 视频测试数据集构建

基于以往视频编码测试数据集(HEVC CTC [1], VVC CTC [2], MCL-JCV [3], UVG [4]), 本提案所提出的新视频测试数据集旨在涵盖更多典型的视频运动内容特征。相比于图像测试数据,视频测试数据具有时域特性,且视频帧中通常包含多个任意形状的运动物体和不同类型的运动,导致复杂的运动场,这往往对视频编码方案提出了挑战。因此,在所提出的新测试数据集中捕捉了具有多种时域相关类型的视频数据,包括不同类型的运动场景以及镜头运动。图1中展示了 USTC-TD 数据集中所有采集视频序列的部分帧画面,每个视频的具体配置信息列于表1,其中不同的运动类型在表1中以不同颜色进行显著区分。具体地,在测试序列捕捉过程中,测试序列从每个完整拍摄的测试视频序列

© 本文档和所包含技术信息的知识产权归提出者所有,仅供"数字音视频编解码技术标准工作组"工作范围内应用,并非公开发表,提出者保留申请专利或在公开出版物发表等相关权利。对于侵犯上述权利的行为,文档提出者和"数字音视频编解码技术标准工作组"保留公开谴责或提出诉讼的权利。



图 1 USTC-TD 视频数据集中各视频序列的示意图

(帧率 30fps,总帧数 9000 帧)中筛选,最终选出具有代表性的 300 帧子集作为最终测试序列,其序列总长度为 10 秒,与 HEVC 和 VVC 通用测试条件 (Common Test Conditions, CTC [1-2])中 10 秒的测试序列长度保持一致。

测试序列	颜色空间	运动类型	场景类型	分辨率	帧率
USTC-Badminton	YUV420	Medium	动态场景	1080P	30 fps
USTC-BasketballDrill	YUV420	Medium	动态场景	1080P	30 fps
USTC-BasketballPass	YUV420	Medium	动态场景	1080P	30 fps
USTC-BicycleDriving	YUV420	Complex	动态场景	1080P	30 fps
USTC-Dancing	YUV420	Complex	静态背景	1080P	30 fps
USTC-ParkWalking	YUV420	Complex	动态场景	1080P	30 fps
USTC-Running	YUV420	Complex	动态场景	1080P	30 fps
USTC-ShakingHands	YUV420	Complex	动态场景	1080P	30 fps
USTC-Snooker	YUV420	Tiny	动态场景	1080P	30 fps
USTC-FourPeople	YUV420	Tiny	静态背景	1080P	30 fps

# 2. USTC-TD 视频测试数据集及相关数据集评估分析

为了全面验证所提出的新视频测试数据集在内容因素上的多样性,并对所提

出 USTC-TD 视频测试数据集的优势进行定性分析,在不同的图像/视频特征上对 USTC-TD 数据集进行了评估,并与以往的视频通用测试数据集进行了比较。其中,视频数据集包括 HEVC CTC [1]、VVC CTC [2]、MCL-JCV [3] 和 UVG [4]。在图像/视频特征分析方面,我们选取了空间丰富度(Spatial Information, SI)[5]和时间丰富度(Temporal Information, TI)[6],从空间、时间相关性这两个维度对各数据集的特性进行表征。具体地,所采用指标定义如下:

**空间丰富度(Spatial Information, SI)**: SI 被用作表示边缘能量。根据以往数据集质量评价指标定义 [7], SI 被定义为图像或视频帧亮度分量中边缘幅值的均方根值,具体计算方式如下,

$$Score_{SI} = \sqrt{\frac{L}{1080}} \sqrt{\sum \frac{s_r^2}{P}}$$

其中, $s_r = \sqrt{s_v^2 + s_h^2}$ 指代每个像素的边缘强度。 $s_v$ 与 $s_h$ 指代图像或视频帧分别经过垂直和水平方向的 Sobel 算子进行滤波处理。其中,P表示滤波后视频帧中的总像素数,L 表示视频帧的垂直分辨率。归一化因子  $\sqrt{\frac{L}{1080}}$ 用于降低 SI 对于视频帧和分辨率的依赖性。对于视频数据集,参考 UVG 数据集中的质量评价方式,SI 的取值为所有视频帧中 SI 结果的最大值。

**时间丰富度(Temporal Information, TI)**: TI 被用作表示时域变化。根据数据集质量评价指标定义 [7],TI 被定义为连续帧 $F_{n-1}$ 与 $F_n$ 之间时域变化量标准差最大值,该指标反映了视频中相邻帧之间像素变化的强度,其计算公式如下:

$$Score_{TI} = \max_{1 \leq n \leq N-1} \left\{ std_{0 \leq i \leq W-1, 0 \leq j \leq H-1} [F_n(i,j) - F_{n-1}(i,j)] \right\}$$

其中,W、H、N分别表示视频帧的宽度、高度以及总帧数。

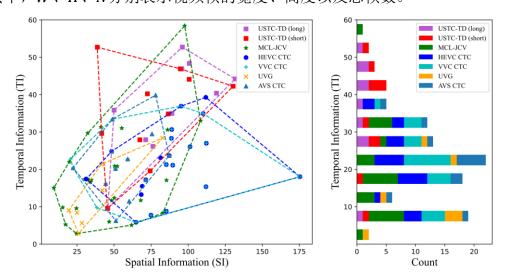


图 2 不同视频测试数据集在时域信息(TI)和空间信息(SI)特征上的可视化评估结果;其中,散点图展示了 TI 与 SI 的对应关系,相应的凸包表示各数据集在特征空间中的覆盖范围;柱状图则表示在不同 TI 得分区间下,各数据集中视频的数量分布情况;Short 与 Long 分别代表 USTC-TD 前 96 帧与全长序列

综合上述测试指标,我们对所提出数据集及相关的先前测试视频数据集进行了测试,TI 和 SI 的得分共同用于评估数据集在空间和时间维度上的多样性,对比结果如图 2 所示。通过与其他数据集的对比,我们发现 USTC-TD 能够与其他数据集协同覆盖多种图像/视频特征,验证了所提出数据集在内容多样性方面的优势。具体地,在 USTC-TD 的测试协助下,测试数据的时间复杂度涵盖范围能够从 5 覆盖至 55,特别是在其他视频数据集中缺乏的 40 至 55(运动高变化)区间内进行了补充,这使得该数据集能够与其他数据集协同实现对时间多样性的更优覆盖,使得对视频压缩方案的评估更加全面和稳健。具体 USTC-TD 中各视频的分见下表,展示了 USTC-TD 视频数据集的定量结果。其中,较高的分数以红色表示,较低的分数以蓝色表示,Short Videos 对应于前 96 帧,Long Videos 对应于序列全长 300 帧。

Video Sequence	Short	Videos	Long Videos	
, in a sequence	SI	TI	SI	TI
USTC-Badminton	67.37	27.91	71.35	27.94
USTC-BasketballDrill	130.31	42.25	131.66	44.19
USTC-BasketballPass	94.63	46.92	95.74	46.89
USTC-BicycleDriving	38.66	52.70	96.09	52.73
USTC-Dancing	74.34	19.61	7.22	26.16
USTC-FourPeople	45.49	9.54	45.77	9.59
USTC-ParkWalking	72.48	40.22	119.07	40.40
USTC-Running	86.52	34.84	89.02	34.97
USTC-ShakingHands	100.67	44.11	100.74	48.37
USTC-Snooker	41.62	29.63	50.01	35.85

## 3. 测试基准

## 3.1 测试条件

基于该数据集,测试了 EEM 及视频编码标准(AVS, VVC)和基于学习的视频编码方案(DCVC-DC, DCVC-FM, DCVC-RT)在客观质量测试指标(PSNR, VMAF)下的性能测试结果,提供了该数据集相关的测试基准。

具体的各编码器配置如下:

EEM: v6.1-release model, rtx3090, encode 模式

AVS3: HPM-15.1, Low-delay P 配置

VVC: VTM-23.0, Low-delay B 配置

DCVC-DC (CVPR 2023) [8]: YUV420 release model, rtx3090, encode 模式 DCVC-FM (CVPR 2024) [9]: YUV420 release model, rtx3090, encode 模式

DCVC-RT (CVPR 2025) [10]: YUV420 release model, rtx3090, encode 模式 SEVC (CVPR 2025) [11]: RGB release model, rtx3090, encode 模式 DCMVC (CVPR 2025) [12]: RGB release model, rtx3090, encode 模式

#### 3. 2 RD-Curve

这里我们展示所有方法相对比下的 RD-Curve 图, 其中图中纵坐标为 YUV-PSNR (Y:U:V=6:1:1), 横坐标为 bpp (bits per pixel)。

其中,图中各编码器具体测试的各码率点下 Avg YUV PSNR 及 bpp:

#### **EEM-6.1:**

Rate0: bpp=405.0848, PSNR=34.601416 Rate1: bpp=920.03744, PSNR=37.629253 Rate2: bpp=1393.6544, PSNR=38.959979 Rate3: bpp=3362.7162, PSNR=41.505575

#### **HPM-15.1:**

Rate0: bpp=469.886, PSNR=36.0752892 Rate1: bpp=721.71264, PSNR=37.389481 Rate2: bpp=1130.428, PSNR=38.731476 Rate3: bpp=1815.2268, PSNR=39.952665 Rate4: bpp=3280.1614, PSNR=41.350416 Rate5: bhp=6282.764, PSNR=42.96621

# VTM-23.0:

Rate0: bpp=256.99552, PSNR=34.635072 Rate1: bpp=386.67816, PSNR=36.02614 Rate2: bpp=583.61192, PSNR=37.271515 Rate3: bpp=896.09672, PSNR=38.539345 Rate4: bpp=1520.1638, PSNR=39.972706 Rate5: bpp=2522.5823, PSNR=41.144961

## DCVC-DC (CVPR 2023) [8]:

Rate0: bpp=1048.2648, PSNR=38.196702 Rate1: bpp=1579.7266, PSNR=39.517531 Rate2: bpp=2612.3458, PSNR=40.93548 Rate3: bpp=4385.3029, PSNR=42.2178

## DCVC-FM (CVPR 2024) [9]:

Rate0: bpp=953.21864, PSNR=38.269877 Rate1: bpp=1625.1318, PSNR=39.822654 Rate2: bpp=2801.5844, PSNR=41.216474 Rate3: bpp=4834.1496, PSNR=42.490997

## DCVC-RT (CVPR 2025) [10]:

Rate0: bpp=126.9, PSNR=31.51249 Rate1: bpp=224.2, PSNR=33.512264 Rate2: bpp=389.6, PSNR=35.309546

Rate3: bpp=669.6, PSNR=36.998242

Rate4: bpp=1118.4, PSNR=38.46706

Rate5: bpp=1887.7, PSNR=39.857576

Rate6: bpp=3190.2, PSNR=41.126878

Rate7: bpp=5452.6, PSNR=42.331684

# **SEVC (CVPR 2025) [11]:**

Rate0: bpp=1133, PSNR=39.572526

Rate1: bpp=1642.9, PSNR=40.486876

Rate2: bpp=2601.6, PSNR=41.479344

Rate3: bpp=4112.2, PSNR=42.390654

# DCMVC (CVPR 2025) [12]:

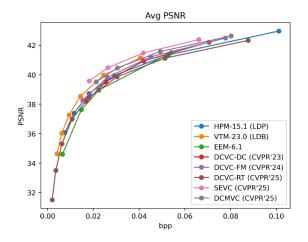
Rate0: bpp=1320.8765, PSNR=39.517028

Rate1: bpp=1898.3072, PSNR=40.46566

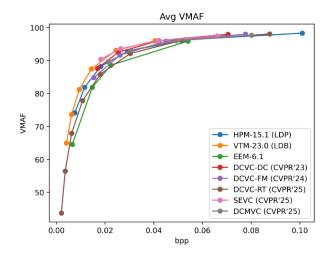
Rate2: bpp=3062.4566, PSNR=41.574128

Rate3: bpp=4985.8237, PSNR=42.635264

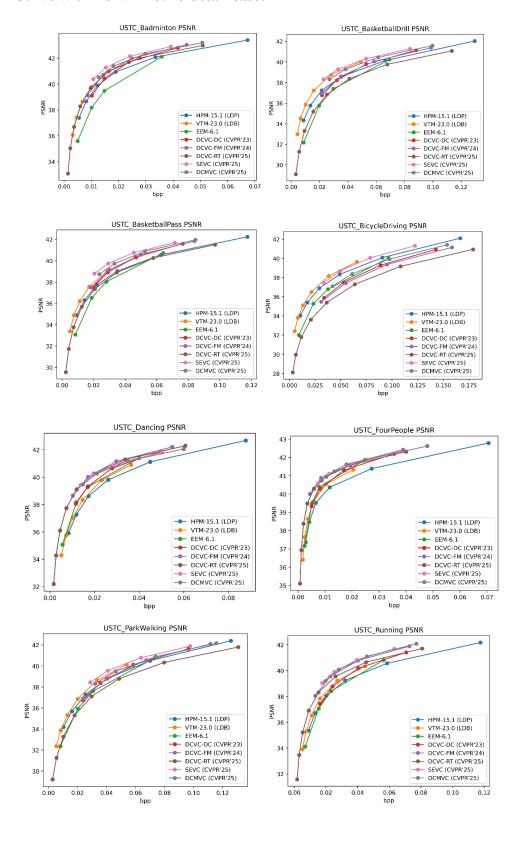
# **RD-Curve with PSNR metric:**

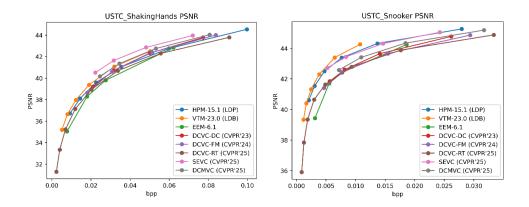


#### **RD-Curve with VMAF metric:**

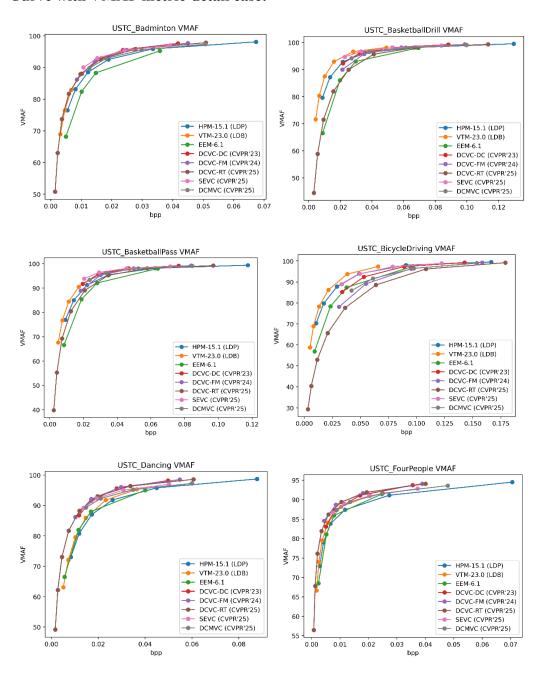


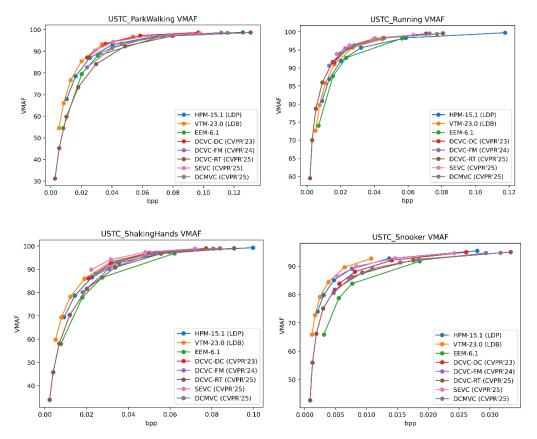
## **RD-Curve with PSNR metric-detail case:**





# **RD-Curve with VMAF metric-detail case:**





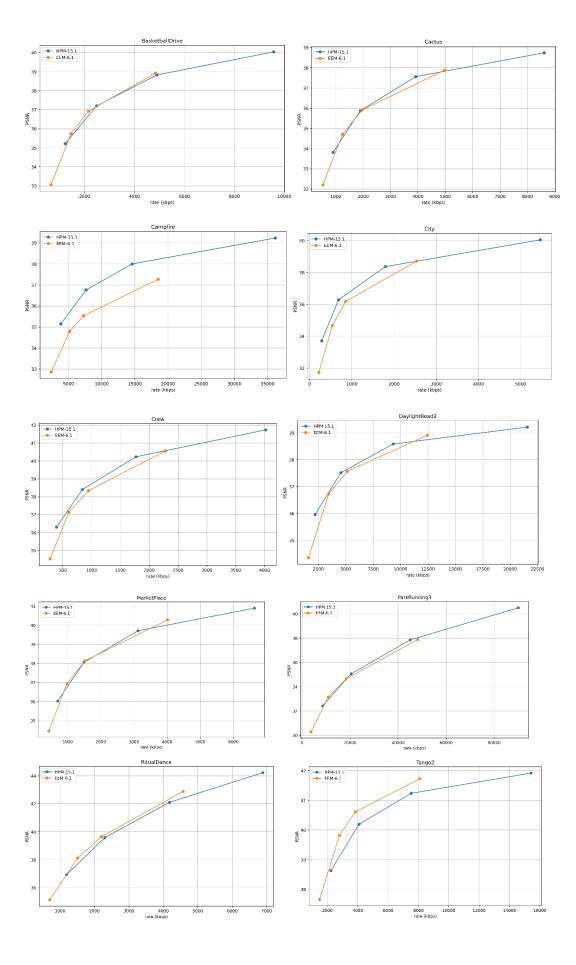
#### 3.3 BD-Rate

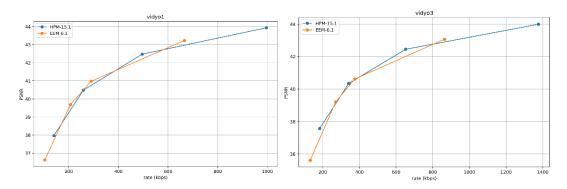
这里我们展示所有方法相对于 HPM (anchor)下的 BD-Rate, 其中表中最上侧为 HPM-15.1 基准, 向下依次为各对比编码器下的相对 BD-Rate 结果。

	BD-PSNR	BD-PSNR Y	BD-PSNR U	BD-PSNR V	BD-VMAF
HPM-15.1	0.0	0.0	0.0	0.0	0.0
VTM-23.0	-14.88%	-16.77%	-6.70%	-6.19%	-15.59%
EEM-6.1	15.61%	33.54%	-27.96%	-26.63%	25.25%
DCVC-DC	2.92%	1.52%	19.47%	0.59%	-13.54%
DCVC-FM	-4.17%	11.61%	-49.18%	-47.60%	-1.99%
DCVC-RT	8.86%	20.62%	-32.20%	-27.66%	15.23%
SEVC	-26.54%	-9.09%	-4.30%	-23.89%	-22.50%
DCMVC	-13.49%	5.85%	-8.98%	-17.68%	-3.58%

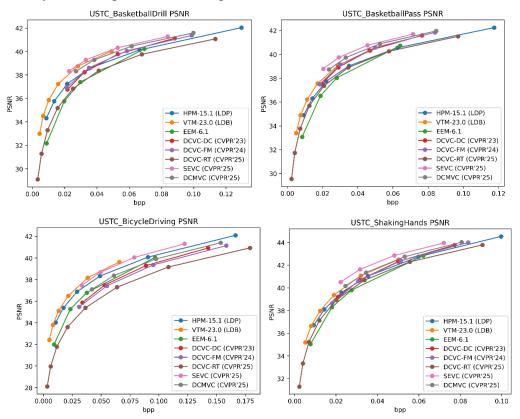
## 3.4 对比分析

这里我们基于上述在新测试集 USTC-TD 上的测试结果,结合 EEM-6.1 在 AVS 通用测试序列上开源的通用测试结果展开对比。这里展示 EEM-6.1 下相比于 HPM-15.1 的 RD-Curve 图,其中,图中纵坐标为 YUV-PSNR(Y:U:V=6:1:1),横坐标为 bit rate(kbps)。





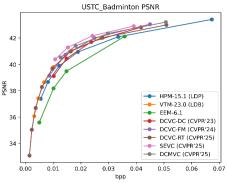
目前,EEM-6.1 在 AVS 的通用测试序列上,其在 Tango2, BasketballDrive, MarketPlace, RitualDance 等多个运动稍复杂的场景下取得了优于 HPM-15.1 (LDP) 的性能。为了进一步使得测试集的评估更加全面,我们结合 AVS 通用测试序列特性 (图 2),重点对比一下 4 个目前为 AVC-CTC 所未覆盖时域信息分数范围下的测试序列,其中包括 USTC-BasketballDrill, USTC-BasketballPass, USTC-BicycleDriving, USTC-ShakingHands。



首先,在有场景切换特点且时间丰富度分数最高的 BicycleDriving 序列上,HPM-15.1, VTM-23.0, EEM-6.1 编码器的性能优于绝大多数现有的端到端视频编码方法。在时间丰富度分数次高的三个序列上,EEM-6.1 的性能比较展示出了相比于现有的传统编码器及端到端编码器的一定差距。同时,在有前景物体运动及镜头运动特点的 Badminton 测试集上,EEM-6.1 也展示出了同样的趋势。

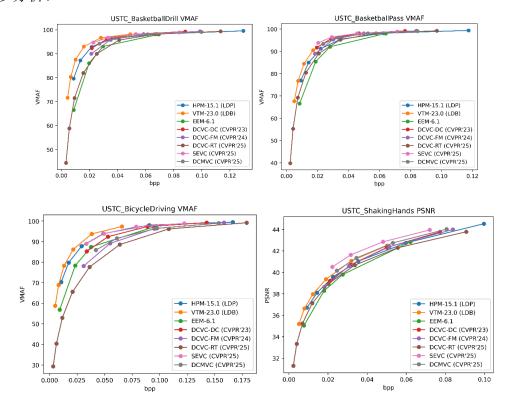
其次,在 USTC-TD 中其他测试序列的结果上,EEM-6.1 与绝大多数编码器

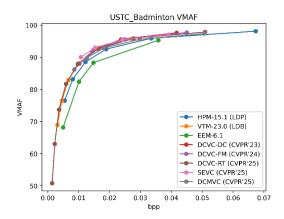
相比性能相当,这里我们同样展示上述 5 个序列的 BD-rate 结果(每个序列上性能最低的编码器用红色标出)。



	BasketballDrill	BasketballPass	BicycleDriving	ShakingHands	Badminton
HPM-15.1	0.0	0.0	0.0	0.0	0.0
VTM-23.0	-25.54%	-17.64%	-15.52%	-10.22%	-15.59%
EEM-6.1	30.83%	16.15%	29.91%	13.78%	33.45%
DCVC-DC	3.17%	-12.97%	42.31%	-1.04%	-4.22%
DCVC-FM	-0.55%	-16.14%	51.47%	0.05%	-16.03%
DCVC-RT	31.71%	5.96%	95.93%	8.92%	-15.08%
SEVC	-28.56%	-35.34%	-9.93%	-24.84	-30.59%
DCMVC	-17.31%	-26.08%	23.00%	-10.59%	-22.32%

进一步地,我们结合 VMAF 进一步对上述 5 个具有测试特点的序列进行进一步分析。





通过 VMAF 指标下的性能对比,我们可以进一步验证上述在 PSNR 指标下结果对比的现象,同时表明所提供的 USTC-TD 数据集具备丰富的时间内容设置,能够有效区分不同编码器在各种运动场景下的性能差异。

# 4. 总结

本提案针对 EEM 的通用测试提供一套新的测试集,建议标准组采纳所提供的 测试集中全部或部分具有代表性运动特性的序列(BasketballDrill, BasketballPass, BicycleDriving, Badminton, ShakingHands),以弥补现有测试数据集的时间丰富性不足,以用于后续在 EEM 上技术改进提案的更全面测试。

# 5. 参考资料

**Related Paper:** Zhuoyuan Li, Junqi Liao, Chuanbo Tang, Haotian Zhang, Yuqi Li, Yifan Bian, Xihua Sheng, Xinmin Feng, Yao Li, Changsheng Gao, Li Li, Dong Liu, and Feng Wu. USTC-TD: A test dataset and benchmark for image and video coding in 2020s. IEEE Transactions on Multimedia, 2025.

下载地址: https://zhuoyuanli1997.github.io/files/USTC-TD/USTC-TD.pdf

Related Homepage: https://esakak.github.io/USTC-TD/

本提案中所有的测试基于对齐 EEM 编码器所处理的 YUV 格式,基于 RGB 的基准在网站及对应论文中已开源。

## 参考文献

- [1] F. Bossen et al., "Common test conditions and software reference configurations," JCTVC-L1100, vol. 12, no. 7, p. 1, 2013.
- [2] J. Boyce et al., "JVET Common Test Conditions and Software Reference Configurations," JVET document, JVET-J1010, 2018.
- [3] H. Wang, W. Gan, S. Hu, J. Y. Lin, L. Jin, L. Song, P. Wang, I. Katsavounidis, A. Aaron, and C.-C. J. Kuo, "MCL-JCV: a JND-based H. 264/AVC video quality assessment dataset," in 2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016, pp. 1509–1513.

- [4] A. Mercat, M. Viitanen, and J. Vanne, "UVG dataset: 50/120fps 4k sequences for video codec analysis and development," in Proceedings of the 11th ACM Multimedia Systems Conference, 2020, pp. 297–302.
- [5] S. Wang, J. Zheng, H.-M. Hu, and B. Li, "Naturalness preserved enhancement algorithm for non-uniform illumination images," IEEE Transactions on Image Processing, vol. 22, no. 9, pp. 3538–3548, 2013.
- [6] A. Mercat, M. Viitanen, and J. Vanne, "UVG dataset: 50/120fps 4k sequences for video codec analysis and development," in Proceedings of the 11th ACM Multimedia Systems Conference, 2020, pp. 297–302.
- [7] S. Winkler, "Analysis of public image and video databases for quality assessment," IEEE Journal of Selected Topics in Signal Processing, vol. 6, no. 6, pp. 616–625, 2012
- [8] Jiahao Li, Bin Li, and Yan Lu. Neural video compression with diverse contexts. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 22616–22626, 2023.
- [9] Jiahao Li, Bin Li, and Yan Lu. Neural video compression with feature modulation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 26099–26108, 2024.
- [10] Jia, Z., Li, B., Li, J., Xie, W., Qi, L., Li, H., Lu, Y. Towards Practical Real-Time Neural Video Compression, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2025.
- [11] Yifan Bian, Chuanbo Tang, Li Li, and Dong Liu. Augmented Deep Contexts for Spatially Embedded Video Coding. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 26099–26108, 2025.
- [12] Chuanbo Tang, Zhuoyuan Li, Yifan Bian, Li Li, and Dong Liu. Neural video compression with context modulation. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 26099–26108, 2025.

# 中国数字音视频编解码技术标准工作组 会员提案专利披露与许可承诺表

根据《中国数字音视频编解码技术标准工作组知识产权政策》第十四、十五、十六条等相关规定,AVS会员在向工作组各专题组提交技术提案时应填写本《会员提案专利披露与许可承诺表》,作为该提案的必要组成部分同时提交。

专题组名称: 音	<b>顷</b> 🗆	视	頻☑	系统□	DRM 🗆				
提案 AVS 文档编号:	1	M8526	提案日期:	2025年	5月30日				
提案标题:									
提案会员名称:									
提案代表姓名 (印刷体): 李卓元 Email: <u>zhuoyuanli@mail.ustc.edu.cn</u>									
提案代表通讯地均				•					
230026		<i>/</i>	7,2 /	3 420112 ( 3 100 %)					
<u>====================================</u>	13609289	323		:					
愿填写。下列表格均表 A: 提案会员在与申请和专利计划。 [2]	其实际知晓			不涉及提案会员和	中他人的专利、专利				
		H- 《디 Icot Lov · 나		<sup>八</sup> 교육구 하나 '눅					
表 B-1: 在中华人民	<b>共和国</b> L》	<b>犬骨授权</b> 日	的专利和/蚁巳	公廾的专利申请					
如果本提案中包含	提案会员或	其关联者在	中华人民共和国	己获得授权的专利和	1/或已公开的专利申请,				
提案会员应当填写下表:									
专利名称	申请号	申请日	由提案会员自愿	经出的专利主要技术	许可承诺(请选择)				
	或	或	内容介绍及其与	5标准内容相关性说明	目 (a) RAND 免费许可或				
	专利号	授权日	(可)	另加附件)	(b) 加入 AVS 专利池				

# 表 B-2: 在中华人民共和国未公开的专利申请 □

2

如果提案会员的缺省许可义务不是 RAND-RF 或者 POOL, 当提案会员或其关联者有与此提案相关的未公 开的中华人民共和国专利申请时,提案会员必须选中此表。

提案会员可在下表中自愿披露上述未公开的专利申请的具体情况及选择许可承诺:

序	专利名称	申请号	申请日	专利主要技术内容介绍及其与标准内	许可承诺(请选择)
号		或		容相关性说明(可另加附件)	(a) RAND 免费许可或
		受理号			(b) 加入 AVS 专利池
1					(b) 加入 AVS 专利池
2					

≢n o.	大山北人	日中中国子	外已获得授	141年上41日	/武士 八 正統	4.土利山)	
衣匠ご	住甲华人	、比共州国人	少103X待按	似的安利和	/ 以口公开日	7) 安利甲屑	ш

如果本提案中包含提案会员或其关联者在中华人民共和国之外已获得授权的专利和/或已公开的专利申请,提案会员应当填写下表:

序	专利名称	申请号	申请日	申请或者	由提案会员自愿给出的专利主要技术	许可承诺(请选择)
号		或	或	获得授权	内容介绍及其与标准内容相关性说明	(a) RAND 免费许可或
		专利号	授权日	的国家	(可另加附件)	(b) 加入 AVS 专利池或
				/地区		(c) RAND
1						
2						

# 表 B-4: 在中华人民共和国之外未公开的专利申请 □

如果提案会员的缺省许可义务不是 RAND-RF 或者 POOL, 当提案会员或其关联者有与此提案相关的在中华人民共和国之外的未公开的专利申请时,提案会员必须选中此表。

提案会员可在下表中自愿披露上述未公开的专利申请的具体情况及选择许可承诺:

序	专利名称	申请号	申请日	申请的	专利主要技术内容介绍及其与标准内	许可承诺(请选择)
号		或		国家	容相关性说明(可另加附件)	(a) RAND 免费许可或
		受理号		/地区		(b) 加入 AVS 专利池或
						(c) RAND
1						
2						

# 表 C: 自愿填写

提案会员自愿在其实际知晓的范围内披露的本提案可能涉及到的他人专利和专利申请:

序	专利名称	申请号	申请日	申请或者	专利的主要技术内容介绍及其与标	申请人或
号		或	或授权	获得授权	准内容相关性说明(可另加附件)	专利权人
		专利号	日	的国家/地		
				X		
1.						
2.						