

ICS: 33.160
CCS: M60



世界超高清视频产业联盟标准

T/UWA 010-2022

智能视频处理系统通用技术规范

General technical specifications for intelligent video processing systems

2022-02-10 发布

2022-02-10 实施

世界超高清视频产业联盟

目 次

前言.....	III
1 范围.....	1
2 规范性引用文件.....	1
3 术语和定义.....	1
4 缩略语.....	2
5 概述.....	2
6 技术要求.....	3
6.1 智能视频画质提升.....	3
6.1.1 智能超分.....	3
6.1.2 智能增强.....	3
6.1.3 智能插帧.....	3
6.1.4 智能 HDR 转换.....	4
6.1.5 综合画质提升.....	4
6.2 智能老片修复.....	4
6.2.1 划痕去除.....	4
6.2.2 噪点去除.....	4
6.2.3 智能上色.....	4
6.2.4 综合修复提升.....	4
6.3 智能视频编辑.....	4
6.3.1 特定子图像去除.....	4
6.3.2 黑边裁剪.....	4
6.3.3 智能横竖屏转换.....	5
6.3.4 智能字幕.....	5
6.4 智能视频编码.....	5
6.4.1 内容自适应编码.....	5
6.4.2 ROI 编码.....	5
6.5 视频格式.....	5
6.5.1 输入编码格式.....	5
6.5.2 输出编码格式.....	5
6.5.3 输入封装格式.....	5
6.5.4 输出封装格式.....	6
7 测试方法.....	6
7.1 智能视频画质提升.....	6
7.1.1 智能超分.....	6
7.1.2 智能增强.....	6
7.1.3 智能插帧.....	6
7.1.4 智能 HDR 转换.....	6
7.1.5 综合画质提升.....	7

7.2	智能老片修复	7
7.2.1	划痕去除	7
7.2.2	噪点去除	7
7.2.3	智能上色	7
7.2.4	综合修复提升	7
7.3	智能视频编辑	8
7.3.1	特定子图像去除	8
7.3.2	黑边裁剪	8
7.3.3	智能横竖屏转换	8
7.3.4	智能字幕	8
7.4	智能视频编码	8
7.4.1	内容自适应编码	8
7.4.2	ROI 编码	8
7.5	视频格式	9
7.5.1	输入编码格式	9
7.5.2	输出编码格式	9
7.5.3	输入封装格式	9
7.5.4	输出封装格式	9
附录 A	(资料性) 测试项目和技术要求对照	10
附录 B	(规范性) 测试系统配置	11
附录 C	(规范性) 源视频配置	12
C.1	总则	12
C.2	4K	12
C.3	4K 转 25 fps 高清	12
C.4	4K 转 15 fps 高清	13
C.5	4K 转标清	13
C.6	内容自适应编码	13
C.7	隔行	13
C.8	老片	13
C.9	特定子图像	13
C.10	黑边	13
C.11	横竖屏转换	14
C.12	字幕	14
附录 D	(规范性) 技术指标计算公式	15
D.1	图像质量提升率	15
D.2	召回率	15
D.3	字错误率	15
D.4	VMAF 标准差	15
D.5	CRF 模式下 ROI 编码的平均率变化	15
D.6	交并比	16

前 言

本文件按照 GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由世界超高清视频产业联盟提出并归口。

本文件起草单位：北京百度网讯科技有限公司、中国电子技术标准化研究院、中央宣传部电影卫星频道节目制作中心、中国移动通信集团有限公司、中视奇云(北京)科技有限公司、上海交通大学、华为技术有限公司、工业和信息化部电子第五研究所、中兴通讯股份有限公司、京东方科技集团股份有限公司、中国信息通信研究院、杭州当虹科技股份有限公司、北京集创北方科技股份有限公司、广东博华超高清创新中心有限公司、杭州趣看科技有限公司、百视通网络电视技术发展有限责任公司、深圳市洲明科技股份有限公司、TCL实业控股股份有限公司、海信视像科技股份有限公司、阿里巴巴(中国)有限公司、北京小米移动软件有限公司、北京锐马视讯科技有限公司、北京中科大洋科技发展股份有限公司、北京市博汇科技股份有限公司。

本文件主要起草人：邢怀飞、孙齐锋、尤莉、李冉、查丽、邢刚、刘立军、宋利、曹菲菲、王冰、连林江、韩松、宋飞、孙强、柳春旭、陈仁伟、李婧欣、赵晓莺、王建、黄林轶、尹芹、顿胜堡、刘明玺、王晓明、王加良、蔡佳音、王亚军、陈勇、张晋芳、肖铁军、钱代友、郭佩佩、方晖、王金东、赵鹏、樊磊、韦胜钰、何栋梁、李鑫、丁文鹏、汤毅、张尚荣、王烨东、张晓波、黄永衡、刘国栋、冯亚楠、陈梅丽、张萌萌、陈冠男、翟梦冉、黄成、彭旭辉、白莹杰、张家斌、彭海。

智能视频处理系统通用技术规范

1 范围

本文件规定了智能视频处理系统在智能视频画质提升、智能老片修复、智能视频编辑、智能视频编码和视频格式等方面的通用技术要求，描述了对应的测试方法。

本文件适用于智能视频处理系统的设计、开发、测试和使用。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 9813.1 计算机通用规范 第1部分：台式微型计算机

GB/T 9813.2 计算机通用规范 第2部分：便携式微型计算机

GB/T 9813.3 计算机通用规范 第3部分：服务器

GY/T 155 高清晰度电视节目制作及交换用视频参数值

GY/T 307 超高清晰度电视系统节目制作和交换参数值

GY/T 315 高动态范围电视制作和交换图像参数值

GY/T 340—2020 超高清晰度电视图像质量主观评价方法 双刺激连续质量标度法

3 术语和定义

下列术语和定义适用于本文件。

3.1

智能视频处理系统 **intelligent video processing systems**

应用人工智能技术进行视频优化的系统。

注：视频优化包括但不限于视频画质提升、老片修复、视频编辑效率提升、视频压缩效率提升等，该系统可以基于云平台或者终端进行实现。

3.2

内容自适应编码 **content adaptive encoding**

根据对视频内容特性的分析生成最优码率和相应分辨率的编码配置的编码方法。

注：采用人工智能模型或机器学习模型，预测最优编码配置，实现在同等视频质量条件下的编码码率最低。

3.3

智能上色 **intelligent coloring**

应用人工智能技术将视频进行色彩渲染的方法。

3.4

划痕 **scratch**

老的胶片上，由于胶片介质划伤造成的图像条状缺失；或者老的磁带上，由于磁头非正常接触导致介质划伤带来的图像条状损伤；或者其他因素导致老片的条状损失。

3.5

噪点 **noise**

视频上的颗粒状图像损伤。

注：噪点包括雪花噪声、椒盐噪声、高斯噪声、伪影、块效应、模糊和斑点等。

3.6

特定子图像 specific subimage

视频图像中具有特定含义的图像区域。

4 缩略语

下列缩略语适用于本文件。

CRF: 恒定码率系数 (Constant Rate Factor)

CMAF: 通用媒体应用格式 (Common Media Application Format)

DASH: 基于 HTTP 传输的动态自适应流 (Dynamic Adaptive Streaming over HTTP)

HDR: 高动态范围 (High Dynamic Range)

HLS: HTTP直播流媒体 (HTTP Live Streaming)

ROI: 感兴趣区域 (Region Of Interest)

VMAF: 视频质量多方法融合评价 (Video Multimethod Assessment Fusion)

5 概述

智能视频处理系统, 在传统的视频处理基础上集成了基于人工智能的视频处理算法, 主要用于视频内容再生产场景和老片修复场景。从视频处理 workflow 角度, 智能视频处理系统的功能架构主要包括智能视频画质提升、智能老片修复、智能视频编辑、智能视频编码和视频格式。从系统逻辑实现的角度, 技术构成主要包括基础层、处理层、平台层和交互层。本文件按照视频处理 workflow, 结合逻辑实现, 规范了相关的技术要求和测试方法。测试项目和技术要求的对照应按附录 A, 各测试项目的测试系统配置应按附录 B, 各测试项目的源视频配置应按附录 C, 各测试项目中技术指标的计算公式应按附录 D。

智能视频处理系统功能架构如图 1 所示, 主要包括智能视频画质提升、智能老片修复、智能视频编辑、智能视频编码和视频格式五个功能模块。智能视频画质提升主要完成智能超分、智能增强、智能插帧、智能 HDR 转换, 实现综合画质提升。智能老片修复主要完成划痕去除、噪点去除、智能上色, 实现老片综合修复提升。智能视频编辑主要完成特定子图像去除、黑边裁剪、智能横竖屏转换和智能字幕。智能视频编码主要完成内容自适应编码和 ROI 编码。视频格式主要规定输入编码格式、输出编码格式、输入封装格式和输出封装格式。

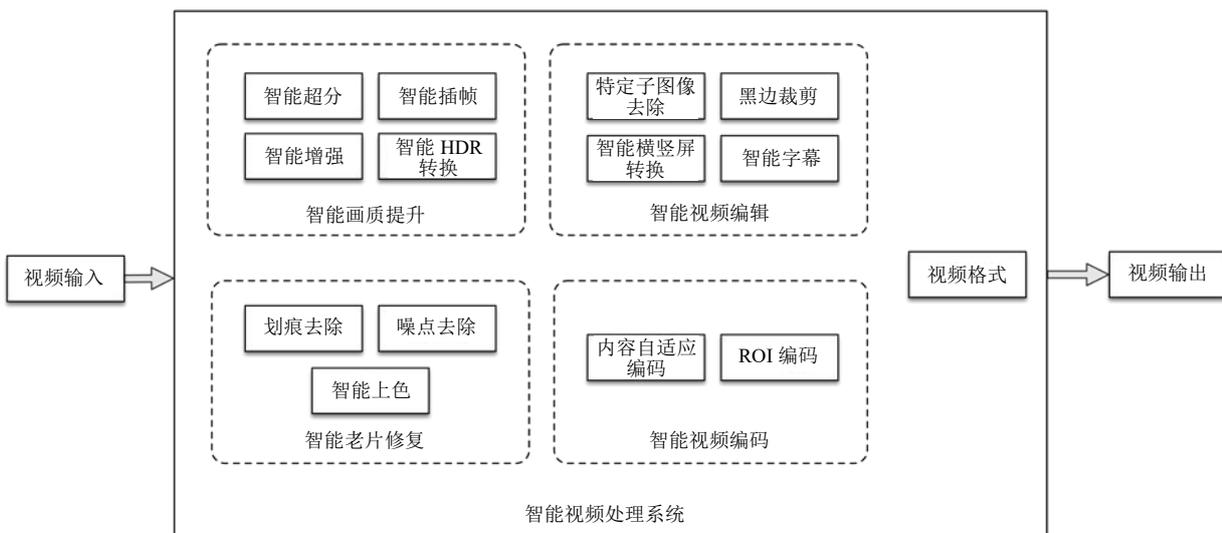


图 1 智能视频处理系统功能架构

智能视频处理系统技术架构如图 2 所示, 主要由基础层、处理层、平台层和交互层四个技术集成层构成。基础层主要包括计算、存储、网络和容器等基础设施。处理层主要包括视频编辑与合成、视频分

析与分片、视频处理与转码和视频合并与封装等处理模块。平台层主要包括平台接入、媒资管理、任务队列和任务调度等平台组件。交互层主要包括交互界面和 API 等交互方式。

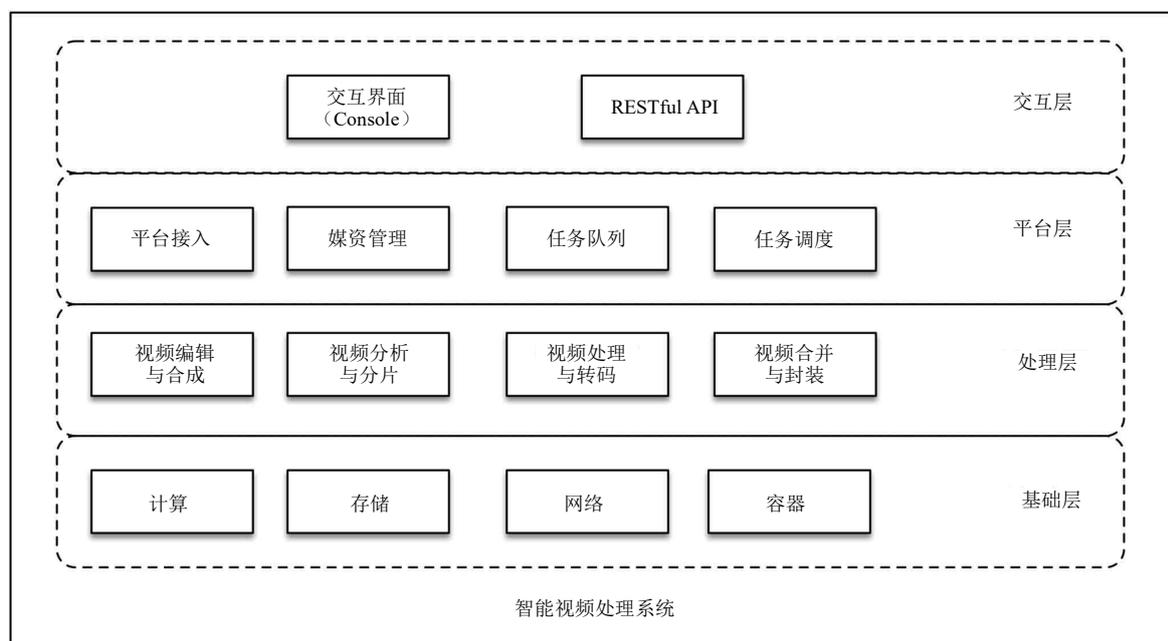


图 2 智能视频处理系统技术架构

6 技术要求

6.1 智能视频画质提升

6.1.1 智能超分

智能视频处理系统满足以下智能超分要求。

- 应支持标清 SD (720×576) 视频转高清 HD (1920×1080) 视频。
- 应支持高清 HD (1920×1080) 视频转 UHD 超高清 4K (3840×2160) 视频。
- 宜支持从高清 HD (1920×1080) 视频转 UHD 超高清 8K (7680×4320) 视频。
- 至少应支持 2 倍的分辨率提升。
- 宜支持 4 倍的分辨率提升。

6.1.2 智能增强

智能视频处理系统满足以下智能增强要求。

- 应支持视频中图像边缘纹理的智能增强。
- 应支持视频中文字边缘纹理的智能增强。
- 宜支持特定 ROI 区域的自适应增强，如人脸。
- 应支持对隔行视频的去隔行处理。

6.1.3 智能插帧

智能视频处理系统满足以下智能插帧要求。

- 最小应支持 2 倍插帧。
- 宜支持 4 倍插帧。

6.1.4 智能 HDR 转换

智能视频处理系统满足以下智能 HDR 转换要求。

- a) 应支持 PQ、HLG 两种 HDR 转换曲线。
- b) 应支持从 GY/T 155 到 GY/T 315 的动态范围转换。
- c) 应支持从 GY/T 155 到 GY/T 307 的色彩空间转换。
- d) 应支持 8 bit 位深到 10 bit 位深的转换。
- e) 应支持 H.265、AVS2 或 AVS3 编码格式中的一种。
- f) 应支持在编码中写入 HDR 元数据。
- g) 宜支持不同 HDR 标准之间的相互转换, 如 HLG 转 HDR10 或 HDR Vivid 等。

6.1.5 综合画质提升

应用智能超分、智能增强、智能插帧和智能 HDR 转换中一种或多种处理后, 输出的目标视频比源视频的图像质量提升率不低于 10%, 图像质量提升率按附录 D 中公式 (D.1) 计算。

6.2 智能老片修复

6.2.1 划痕去除

智能视频处理系统满足以下划痕去除要求。

- a) 应支持划痕智能检测, 划痕智能检测的召回率不低于 70%。召回率按附录 D 中公式 (D.2) 计算。
- b) 应支持划痕去除, 划痕去除强度能按照阈值调整。
- c) 宜提供可视化环境, 支持人工辅助调整划痕检测结果。

6.2.2 噪点去除

智能视频处理系统满足以下噪点去除要求。

- a) 应支持噪点智能检测。
- b) 应支持噪点去除, 噪点去除强度能按照阈值调整。
- c) 宜提供可视化环境, 支持人工辅助调整噪点检测结果。

6.2.3 智能上色

智能视频处理系统应支持智能上色。

6.2.4 综合修复提升

应用划痕去除、噪点去除、智能上色中一种或多种处理后, 输出的目标视频比源视频的图像质量提升率不低于 10%, 图像质量提升率按附录 D 中公式 (D.1) 计算。

6.3 智能视频编辑

6.3.1 特定子图像去除

智能视频处理系统满足以下特定子图像去除要求。

- a) 应支持特定子图像智能检测, 特定子图像智能检测召回率不低于 95%。召回率按附录 D 中公式 (D.2) 计算。
- b) 应支持指定区域的特定子图像智能去除。

6.3.2 黑边裁剪

智能视频处理系统满足以下黑边裁剪要求。

- a) 应支持在视频图像左、右区域的黑边智能检测, 黑边智能检测召回率不低于 95%。召回率按附录 D 中公式 (D.2) 计算。
- b) 应支持指定区域的黑边智能裁剪。

6.3.3 智能横竖屏转换

智能视频处理系统满足以下智能横竖屏转换要求。

- a) 应支持基于画面主体将横屏视频转换成 9:16 竖屏视频。
- b) 存在多个画面主体时，智能选择画面主体进行裁剪。
- c) 视频场景频繁切换时或存在多画面主体裁剪时，输出画面平稳不抖动。

6.3.4 智能字幕

智能视频处理系统满足以下智能字幕要求。

- a) 应支持视频中的语音自动生成字幕，语音自动生成字幕识别字错误率不高于 10%。字错误率按附录 D 中公式 (D.3) 计算。
- b) 应支持 ASS 或 SRT 字幕格式文件。

6.4 智能视频编码

6.4.1 内容自适应编码

智能视频处理系统满足以下内容自适应编码要求。

- a) 应支持按视频场景进行智能内容自适应编码。
- b) 按 7.4.1 的测试方法，指定分辨率为高清 HD (1920x1080) 时，各目标视频片段的 VMAF 平均分不低于 91 分且不高于 94 分，不同目标视频片段之间的 VAMF 平均分之差不超过 2 分。
- c) 按 7.4.1 的测试方法，各目标视频片段中，单帧 VMAF 分数比 VAMF 平均分低 5% 以上的比例不超过 8%。
- d) 按 7.4.1 的测试方法，内容自适应编码的各目标视频片段，对比同等码率下 CBR 转码的目标视频片段，VMAF 平均分至少提高 3 分，VMAF 标准差降低至少 40%。VMAF 标准差按附录 D 中公式 (D.4) 计算。

6.4.2 ROI 编码

智能视频处理系统满足以下 ROI 编码要求。

- a) 应支持智能视频 ROI 编码。
- b) 在 CRF 编码模式下 ROI 编码的平均码率变化应低于 5%。CRF 模式下 ROI 编码的平均码率变化按附录 D 中公式 (D.5) 计算。

注：常见 ROI 区域内容包括人脸、人体、文字等。

6.5 视频格式

6.5.1 输入编码格式

智能视频处理系统满足以下输入编码格式要求。

- a) 应支持不少于 2 种分发域输入编码格式，如 H.264、H.265、AVS2 等。
- b) 应支持不少于 2 种制作域输入编码格式，如 Apple ProRes 422、DNxHD—NxHR、XAVC—I Class 300/480 等。

6.5.2 输出编码格式

智能视频处理系统应支持不少于 2 种输出编码格式，如 H.264、H.265、AVS2 等。

6.5.3 输入封装格式

智能视频处理系统满足以下输入封装格式要求。

- a) 应支持不少于 5 种分发域输入封装格式，如 MP4、MOV、TS、FLV、AVI、MKV、3GP 等。
- b) 应支持不少于 2 种制作域输入封装格式，如 Mov、MXF、AVI 等。

6.5.4 输出封装格式

智能视频处理系统满足以下输出封装格式要求。

- a) 应支持 MP4、TS 输出封装格式。
- b) 宜支持 FLV、MOV 输出封装格式。
- c) 宜支持 DASH 协议、HLS 协议或 CMAF 协议。

7 测试方法

7.1 智能视频画质提升

7.1.1 智能超分

智能超分测试步骤如下。

- a) 将 C.5 的 4K 转标清源视频进行超分，输出高清 HD (1920×1080) 目标视频，目标视频能保持图像细节清晰。
- b) 将 C.3 的 4K 转 25 fps 高清源视频进行超分，输出 UHD 超高清 4K (3840×2160) 目标视频，目标视频能保持图像细节清晰。
- c) 将 C.3 的 4K 转 25 fps 高清源视频进行超分，输出 UHD 超高清 8K (7680×4320) 目标视频，目标视频能保持图像细节清晰。
- d) 使用附录 B 中的媒体分析工具检测上述目标视频，目标视频分别符合 6.1.1 a)、b)、c) 的要求。
- e) 步骤 b) 输出的目标视频符合预期，则判定满足 2X 超分要求。
- f) 步骤 c) 输出的目标视频符合预期，则判定满足 4X 超分要求。

7.1.2 智能增强

智能增强测试步骤如下。

- a) 将 C.3 的 4K 转 25 fps 高清源视频进行智能增强，输出的目标视频能保持边缘纹理锐化、符合人眼主观感受。
- b) 将 C.7 的隔行源视频进行去隔行处理，输出的目标视频无明显隔行效应。

7.1.3 智能插帧

智能插帧测试步骤如下。

- a) 将 C.3 的 4K 转 25 fps 高清源视频进行智能插帧，输出 50 fps 目标视频，目标视频能保持画面流畅、无明显抖动、无明显闪烁。
- b) 将 C.4 的 4K 转 15 fps 高清源视频进行智能插帧，输出 60 fps 目标视频，目标视频能保持画面流畅、无明显抖动、无明显闪烁。
- c) 使用附录 B 中的媒体分析工具检测上述目标视频，目标视频分别符合 6.1.3 a)、b) 的要求。

7.1.4 智能 HDR 转换

智能 HDR 转换测试步骤如下。

- a) 将源视频进行智能 HDR 转换，输出分辨率为 UHD 超高清 4K (3840×2160)、动态范围为 GY/T 315、色彩空间为 GY/T 307、位深为 10 bit、HDR 格式为 HDR10、编码标准为 H.265、AVS2 或 AVS3 的目标视频。目标视频能突出亮部细节和暗部细节，色彩饱和度更高。
- b) 使用附录 B 中的媒体分析工具检测步骤 a) 输出的目标视频，在分辨率、动态范围、色彩空间、位深、HDR 格式和编码标准方面分别符合 6.1.4 b)、c)、d)、e) 的要求。
- c) 将步骤 a) 的目标视频，进行 HLG 转 HDR10 或 HDR Vivid，输出目标视频。目标视频能突出亮部细节和暗部细节，色彩饱和度更高。
- d) 使用附录 B 中的媒体分析工具检测步骤 c) 输出的目标视频，目标视频的元数据为 HDR 10 或 HDR Vivid。

7.1.5 综合画质提升

对源视频进行智能超分、智能增强、智能插帧和智能 HDR 转换中一种或多种处理后，输出目标视频。

比对目标视频和源视频，图像质量画质提升主观评价的测试阶段、测试图像的显示和评分标度遵守 GY/T 340—2020 中 5.5、5.6、5.7 的规定。结果分析、结果说明遵守 GY/T 340—2020 中 5.8、5.9 的规定。

用双刺激连续质量标度法得到源视频图像评分与目标视频图像评分，按附录 D 中公式 (D.1) 计算画质提升率。

注：评分过程中应考虑以下因素。

- a) 清晰度，包括但不限于：画面清晰、无明显模糊、无呼吸效应。
- b) 干净度，包括但不限于：无明显噪点、马赛克和块效应，实现艺术效果的块效应除外。
- c) 保真度，包括但不限于：无明显掉色，亮度和对比度明显提升。
- d) 流畅度，包括但不限于：无明显卡顿、抖动、重影和拖尾，流畅性明显提升。
- e) 画面锐度，画面锐化程度明显提升。
- f) 字幕质量，包括但不限于：字幕无模糊，字幕和画面同步。
- g) 整体质量，视频整体质量明显提升。

7.2 智能老片修复

7.2.1 划痕去除

划痕去除测试步骤如下。

- a) 设置划痕强度，将源视频进行划痕检测和去除，输出目标视频。目标视频划痕消除效果明显。
- b) 使用附录 B 中的媒体分析工具检测目标视频。
- c) 按附录 D 中公式 (D.6) 计算划痕检测交并比，大于 0.1 为检测正确。
- d) 结合步骤 c) 的检测结果，按附录 D 中公式 (D.2) 计算划痕检测召回率。

7.2.2 噪点去除

噪点去除测试步骤如下。

- a) 设置噪点强度，将源视频进行噪点检测和去除，输出目标视频。目标视频能保留画面细节，并提升主观清晰度。
- b) 使用附录 B 中的媒体分析工具检测目标视频。

7.2.3 智能上色

智能上色测试步骤是将源视频进行智能上色，输出的目标视频没有闪烁、没有抖动、色彩稳定。

7.2.4 综合修复提升

对源视频进行划痕去除、噪点去除和智能上色中一种或多种处理后，输出目标视频。

比对目标视频和源视频，图像质量主观评价的测试阶段、测试图像的显示和评分标度遵守 GY/T 340—2020 中 5.5、5.6、5.7 的规定。结果分析、结果说明遵守 GY/T 340—2020 中 5.8、5.9 的规定。

用双刺激连续质量标度法得到源视频图像得分与目标视频图像评分，按附录 D 中公式 (D.1) 计算画质提升率。

注：评分过程中应考虑以下因素。

- a) 清晰度，包括但不限于：画面清晰，无明显模糊，无呼吸效应。
- b) 干净度，包括但不限于：划痕、噪点明显减少，不影响主观体验。
- c) 保真度，包括但不限于：无明显掉色，亮度和对比度有提升，画面无扭曲。
- d) 流畅度，包括但不限于：无明显卡顿、抖动、重影和拖尾，流畅性提升。
- e) 颜色，颜色接近自然色。
- f) 整体质量，视频整体质量提升。

7.3 智能视频编辑

7.3.1 特定子图像去除

特定子图像去除测试步骤如下。

- a) 将源视频进行特定子图像去除，输出目标视频。
- b) 比对源视频和目标视频，记录水印检测和去除数据。
- c) 按附录 D 中公式 (D.6) 计算特定子图像检测交并比，大于 0.5 为检测正确。
- d) 结合步骤 c) 的检测结果，按附录 D 中公式 (D.2) 计算特定子图像检测召回率。

7.3.2 黑边裁剪

黑边裁剪测试步骤如下。

- a) 将源视频进行黑边去除，输出目标视频。
- b) 比对源视频和目标视频，记录黑边检测和裁剪数据。
- c) 按附录 D 中公式 (D.6) 计算黑边检测交并比，大于 0.95 为检测正确。
- d) 结合步骤 c) 的检测结果，按附录 D 中公式 (D.2) 计算黑边检测召回率。

7.3.3 智能横竖屏转换

智能横竖屏转换测试步骤如下。

- a) 将源视频进行智能横转竖操作，输出目标视频。
- b) 使用附录 B 中的媒体分析工具检测视频显示宽高比 9:16。
- c) 视频场景切换时，应自动切换到新场景的主体，无抖动。

7.3.4 智能字幕

智能字幕测试步骤如下。

- a) 将源视频进行语音自动生成字幕，并生成 ASS 或 SRT 格式的字幕。
- b) 对比语音自动生成字幕信息与源视频字幕信息，按附录 D 中公式 (D.3) 计算语音自动生成字幕字错误率。

7.4 智能视频编码

7.4.1 内容自适应编码

内容自适应编码测试步骤如下。

- a) 指定目标分辨率为高清 HD (1920×1080)，并将源视频进行内容自适应编码，输出目标视频。
- b) 使用附录 B 中的媒体分析工具检测目标视频，计算各目标视频片段 VMAF 平均分、目标视频片段间 VMAF 平均分之差、标准差、单帧 VMAF 分数低于 VMAF 平均分 5% 以上的比例，计算结果满足 6.4.1 b) 和 c) 的技术要求。
- c) 指定目标分辨率为高清 HD (1920×1080)，指定目标码率与步骤 a) 一致，将源视频进行 CBR 编码，并输出目标视频。
- d) 使用附录 B 中的媒体分析工具检测目标视频，计算步骤 c) 各目标视频片段 VMAF 平均分、VMAF 标准差。
- e) 计算步骤 a) 输出的各目标视频片段相对步骤 c) 输出的各目标视频片段的 VAMF 平均分提高和 VMAF 标准差降低，计算结果满足 6.4.1 d) 的技术要求。

7.4.2 ROI 编码

ROI 编码测试步骤如下。

- a) 在 CRF 模式下，指定 CRF 值为 23，将源视频进行 ROI 编码，输出目标视频。
- b) 将同一源视频进行非 ROI 编码，输出目标视频。
- c) 使用附录 B 中的媒体分析工具检测两个目标视频的平均码率。
- d) 按附录 D 中公式 (D.5) 计算 CRF 模式下 ROI 编码的平均码率变化。

7.5 视频格式

7.5.1 输入编码格式

7.5.1.1 分发域输入编码格式测试步骤如下。

- a) 将 H.264、H.265、AVS2 编码格式的源视频分别上传系统，进行 MP4 编码格式转码，输出目标视频。
- b) 使用附录 B 中的媒体分析工具检测目标视频的转码时长与格式，目标视频与源视频的时长保持一致，格式符合预期。

7.5.1.2 制作域输入编码格式测试步骤如下。

- a) 将 Apple ProRes 422、DNxHD—NxHR、XAVC—I Class 300/480 编码格式的源视频分别上传系统，进行 MP4 格式转码，输出目标视频。
- b) 使用附录 B 中的媒体分析工具检测目标视频的转码时长与格式，目标视频与源视频的时长保持一致，格式符合预期。

7.5.2 输出编码格式

输出编码格式测试步骤如下。

- a) 将 MP4 格式的源视频上传系统，进行 H.264、H.265、AVS2 格式转码，输出目标视频。
- b) 使用附录 B 中的媒体分析工具检测目标视频的转码时长与格式，目标视频与源视频的时长保持一致，格式符合预期。

7.5.3 输入封装格式

7.5.3.1 分发域输入封装格式测试步骤如下。

- a) 将 MP4、MOV、TS、FLV、AVI、MKV、3GP 格式的源视频分别上传系统，进行 MP4 格式转码，输出目标视频。
- b) 使用附录 B 中的媒体分析工具检测目标视频的转码时长与格式，目标视频与源视频的时长保持一致，格式符合预期。

7.5.3.2 制作域输入封装格式测试步骤如下。

- a) 将 Mov、MXF、AVI 格式的源视频分别上传系统，进行 MP4 格式转码，输出目标视频。
- b) 使用附录 B 中的媒体分析工具检测目标视频的转码时长与格式，目标视频与源视频的时长保持一致，格式符合预期。

7.5.4 输出封装格式

输出封装格式测试步骤如下。

- a) 将 MP4 格式的源视频上传系统，进行 MP4、TS、FLV、MOV、DASH、HLS、CMAF 格式转码，输出目标视频。
- b) 使用附录 B 中的媒体分析工具检测目标视频的转码时长与格式，目标视频与源视频的时长保持一致，格式符合预期。

附录 A
(资料性)
测试项目和技术要求对照

智能视频处理系统的测试包括主观评价和客观测试，主观评价和客观测试关联互补。主观评价利用人对视频的主观感受给出视频质量评价，评价处理前后的视频主观质量提升。客观测试验证系统功能和技术指标。各测试项目和技术要求对照见表 A.1。

表 A.1 测试项目和技术要求对照表

序号	测试项目	技术要求章条号
1	智能超分	6.1.1
2	智能增强	6.1.2
3	智能插帧	6.1.3
4	智能 HDR 转换	6.1.4
5	综合画质提升	6.1
6	划痕去除	6.2.1
7	噪点去除	6.2.2
8	智能上色	6.2.3
9	综合修复提升	6.2
10	特定子图像去除	6.3.1
11	黑边裁剪	6.3.2
12	智能横竖屏转换	6.3.3
13	智能字幕	6.3.4
14	内容自适应编码	6.4.1
15	ROI 编码	6.4.2
16	输入编码格式	6.5.1
17	输出编码格式	6.5.2
18	输入封装格式	6.5.3
19	输出封装格式	6.5.4

附录 B
(规范性)
测试系统配置

测试系统主要包括显示器、计算机等硬件，媒体分析工具、VMAF 工具等软件，各测试项目的测试系统配置明细见表 B.1。显示器遵守 GY/T 340—2020 表 2 的规定。计算机遵守 GB/T 9813.1、GB/T 9813.2 或 GB/T 9813.3 的规定。

表 B.1 测试系统配置明细表

序号	测试项目	显示器	计算机	媒体分析工具	VMAF工具	主观评价人员
1	智能超分	☑	☑	☑		
2	智能增强	☑	☑			
3	智能插帧		☑	☑		
4	智能 HDR 变换		☑	☑		
5	综合画质提升	☑	☑			☑
6	划痕去除		☑	☑		
7	噪点去除		☑	☑		
8	智能上色		☑			
9	综合修复提升	☑	☑			☑
10	特定子图像去除	☑	☑			
11	黑边裁剪	☑	☑			
12	智能横竖屏转换	☑	☑	☑		
13	智能字幕	☑	☑			
14	内容自适应编码	☑	☑	☑	☑	
15	ROI 编码	☑	☑	☑		
16	输入编码格式	☑	☑	☑		
17	输出编码格式	☑	☑	☑		
18	输入封装格式	☑	☑	☑		
19	输出封装格式	☑	☑	☑		
示例 1：如 MediaInfo 等媒体分析软件。 示例 2：如 FFMPEG 等 VMAF 工具。						

附录 C (规范性) 源视频配置

C.1 总则

智能视频处理系统测试需要 11 个源视频。源视频包括 4K、4K 转 25 fps 高清、4K 转 15 fps 高清、4K 转标清、内容自适应编码、隔行、老片、特定子图像、黑边、横竖屏转换和字幕 11 个测试数据集，其中 4K 源视频用于其他源视频制作。各测试项目的源视频配置见表 C.1。

表 C.1 源视频配置表

序号	测试项目	源视频										
		4K	4K 转 25 fps 高清	4K 转 15 fps 高清	4K 转标清	内容自适应编码	隔行	老片	特定子图像	黑边	横竖屏转换	字幕
1	智能超分		☑		☑							
2	智能增强		☑				☑					
3	智能插帧		☑	☑								
4	智能 HDR 变换		☑									
5	综合画质提升		☑									
6	划痕去除						☑					
7	噪点去除						☑					
8	智能上色						☑					
9	综合修复提升						☑					
10	特定子图像去除							☑				
11	黑边裁剪								☑			
12	智能横竖屏转换									☑		
13	智能字幕											☑
14	内容自适应编码					☑						
15	ROI 编码		☑									
16	输入编码格式		☑									
17	输出编码格式		☑									
18	输入封装格式		☑									
19	输出封装格式		☑									

C.2 4K

4K 源视频满足以下要求。

- a) 为 UHD 超高清 4K (3840×2160) 视频。
- b) 总数不少于 20 个片段。
- c) 每个视频片段的内容应是动态的，并持续 10 s~15 s。
- d) 应包括细节丰富的图像、纹理丰富的图像和包含文字图像。
- e) 应包括人体、人脸的图像，人脸或人体部分包括不同肤色。
- f) 应包括运动图像，运动图像帧率为 50 帧/秒。

C.3 4K 转 25 fps 高清

4K 转高清源视频满足以下要求。

- a) 由 C.2 中 UHD 超高清 4K (3840×2160) 视频转换的高清 HD (1920×1080) 视频。
- b) 高清 HD (1920×1080) 视频的颜色转换空间为 BT.709，位深为 8 bit，帧率为 25 fps，动态范围为 SDR。
- c) 用双刺激连续质量标度法主观评价源视频图像得分宜为 3 分左右。

C.4 4K 转 15 fps 高清

4K 转 15 fps 高清源视频满足以下要求。

- a) 由 C.2 中 4K 源视频转换的高清 HD (1920×1080) 视频。
- b) 高清 HD (1920×1080) 视频的颜色转换空间为 BT.709, 位深为 8 bit, 帧率为 15 fps, 动态范围为 SDR。

C.5 4K 转标清

4K 转标清源视频满足以下要求。

- a) 由 C.2 中 UHD 超高清 4K (3840×2160) 视频转换的标清 SD (720×576) 视频。
- b) 标清 SD (720×576) 视频的颜色转换空间为 BT.709, 位深为 8 bit, 帧率为 25 fps, 动态范围为 SDR。

C.6 内容自适应编码

内容自适应编码源视频满足以下要求。

- a) 源视频总数不少于 3 个视频片段。
- b) 每个视频片段时长不低于 5 s。
- c) 每个视频片段至少包含 5 个场景, 场景包括但是不限于: 动画片、运动、游戏、动物世界、户外、人脸、电影、广告、音乐。

C.7 隔行

隔行源视频满足以下要求。

- a) 源视频总数不少于 5 个片段。
- b) 每个源视频的内容应是动态的, 并持续 10 s~15 s。
- c) 源视频分辨率应不低于标清 SD (720×576)。
- d) 源视频包含隔行视频。

C.8 老片

老片源视频满足以下要求。

- a) 源视频总数不少于 20 个片段。
- b) 每个源视频片段的内容应是动态的, 并持续 10 s~15 s。
- c) 应包括带划痕的图像。
- d) 应包括带噪声的图像, 图像噪声包括但不限于: 雪花噪声、椒盐噪声、高斯噪声、伪影、块效应、模糊和斑点。
- e) 应包含黑白视频。

C.9 特定子图像

特定子图像源视频由厂商提供, 并满足以下要求。

- a) 源视频总数不少于 50 个片段。
- b) 每个源视频的内容应是动态的, 并持续 10 s~15 s。
- c) 源视频分辨率应不低于标清 SD (720×576)。
- d) 应包括系统已覆盖的 20 个不同特定子图像。
- e) 特定子图像应当在视频中持续存在。

C.10 黑边

黑边源视频满足以下要求。

- a) 源视频总数不少于 20 个片段。
- b) 每个源视频的内容应是动态的, 并持续 10 s~15 s。
- c) 源视频分辨率应不低于标清 SD (720×576)。

- d) 应包含 10%~20% 无黑边视频样本, 其余为有黑边样本。
- e) 应包含黑边面积占比小于 5% 的样本。
- f) 应包括左黑边样本和右黑边样本样本。
- g) 应包括非对称黑边样本。

C.11 横竖屏转换

横竖转换源视频满足以下要求。

- a) 源视频总数不少于 20 个片段。
- b) 每个源视频的内容应是动态的, 并持续 10 s~15 s。
- c) 源视频分辨率应不低于标清 SD (720×576)。
- d) 图像应为横屏, 宽高比为 4:3 或 16:9。
- e) 视频画面应以人物作为主体。
- f) 视频画面应包含多画面主体。
- g) 数据集中应包含有场景切换的视频。

C.12 字幕

字幕源视频满足以下要求。

- a) 源视频总数不少于 20 个片段。
- b) 每个源视频的内容应是动态的, 并持续 10 s~15 s。
- c) 源视频分辨率应不低于标清 SD (720×576)。
- d) 源视频图像画面包含中文字幕, 字幕位于画面下方五分之一区域内。
- e) 字幕清晰, 在显示器上可人工辨识。
- f) 每个片段的中文语音信息包含不少于 100 个汉字。

附录 D
(规范性)
技术指标计算公式

D.1 图像质量提升率

图像质量提升率按公式 (D.1) 计算。

$$E=(b-a)/a \times 100\% \quad \text{(D.1)}$$

式中:

E ——图像质量提升率;

a ——源视频图像主观评价得分;

b ——目标视频图像主观评价得分。

D.2 召回率

召回率按公式 (D.2) 计算。

$$\text{Recall} = \frac{S_d}{S_l} \quad \text{(D.2)}$$

式中:

Recall——召回率;

S_d ——检测正确的对象数目;

S_l ——视频中真实的对象数目。

D.3 字错误率

字错误率按公式 (D.3) 计算。

$$\text{WER} = \frac{S+D+I}{S+D+C} \quad \text{(D.3)}$$

式中:

WER——字错误率;

S ——被错误替换的字数;

D ——未识别到的字数;

I ——识别出的验证数据中不存在的字数;

C ——识别正确的字数。

D.4 VMAF 标准差

VMAF 标准差按公式 (D.4) 计算。

$$v_{std} = \sqrt{\frac{\sum_{i=1}^n (v_i - \bar{v})^2}{n-1}} \quad \text{(D.4)}$$

式中:

i ——视频片段中各帧图像的顺序号;

v_{std} ——视频片段的 VMAF 标准差;

v_i ——视频片段中各帧图像的 VMAF;

\bar{v} ——视频片段的 VMAF 平均值。

D.5 CRF 模式下 ROI 编码的平均率变化

CRF 模式下 ROI 编码的平均码率变化按公式 (D.5) 计算。

$$\text{Change} = \frac{\text{ABS}(B_{ROI} - B_{normal})}{B_{normal}} \quad \text{(D.5)}$$

式中:

Change——码率变化;

B_{ROI} ——ROI 编码的平均码率;

B_{normal} ——普通编码的平均码率。

D.6 交并比

交并比按公式 (D.6) 计算。

$$IoU = \frac{P_d \cap P_l}{P_d \cup P_l} \dots\dots\dots (D.6)$$

式中：

IoU ——交并比；

P_d ——检测出来的对象像素集合；

P_l ——真实的对象像素集合。



世界超高清视频产业联盟标准

智能视频处理系统通用技术规范
General technical specifications for intelligent
video processing systems
T/UWA 010-2022

*

世界超高清视频产业联盟 发布

*

开本：880×1230 1/16 印张：1 $\frac{1}{2}$ 字数：36千字

2021年7月第一版 2021年7月第一次印刷
印数：200册