UHD World Association
世界超高清视频产业联盟

UHD World Association
世界超高清视频产业联盟

UHD Word Association
www.theuwa.com

超高清视频智能处理系统应用白皮书

(V1.0)

世界超高清视频产业联盟

前 言

本文件由 UWA 联盟视频体验工作组组织制订，并负责解释。

本文件发布日期：2023 年 11 月 15 日。

本文件由世界超高清视频产业联盟提出并归口。

本文件主要起草单位：

北京百度网讯科技有限公司、中国电信集团有限公司、中国联合网络通信有限公司、上海交通大学 、北京邮电大学、中国信息通信研究院、华为技术有限公司、德科仕通信（上海）有限公司、优酷网络技术（北京）有限公司、工业和信息化部电子第五研究所、中移（杭州）信息技术有限公司、杭州当虹科技股份有限公司、北京数码视讯科技股份有限公司、百视通网络电视技术发展有限责任公司、苏州涌现智能科技有限公司、北京爱奇艺科技有限公司、上海数字电视国家工程研究中心、深圳市中兴微电子技术有限公司、上海兆言网络科技有限公司、世界超高清视频产业联盟、中关村现代信息消费应用产业技术联盟。

本文件主要起草人：

邢怀飞、尤莉、曹菲菲、查丽、罗传飞、贾武、宋利、闫石、王亚军、程宝平、周奎翰、韩善禄、黄伟、吴雪波、蔡佳音、李静、尚璟、刘灏、韦胜钰、蔡佳、张黎敏、王冰、韩松、詹颖、容敏华、黄挺、李日、任博豪、张翰、周骋、杨蕊菡、张媵英、朱蕤 、汤毅、张丽莉、王志航、殷惠清、孔德辉、钱雷、闫伟、陈红、刘璇。

免责说明：

1， 本文件免费使用，仅供参考，不对使用本文件的产品负责。

2， 本文件刷新后上传联盟官网，不另行通知。

目录

1. 视频处理产业发展历程

随着视频技术的发展，视频应用经历了从标清、高清到超高清（4K/8K）的产业化升级过程，超高清视频技术已经在广播电视领域全面应用，相关其他产业的落地也正在逐步推进。当前阶段，通过超高清采集和制作的视频内容还比较少，存量的经典视频可以通过基于 AI 的超高清重制，补足 4K/8K 超高清视频内容缺口。传统的视频超高清重建是通过多张低分辨率的图像获取高分辨率图像的处理过程。从 2006 年开始，基于人工智能的超高清视频智能技术逐渐兴起，在超高清视频的采、制、传、呈、用产业链中，解决了历史视频的低成本、高效率超高清重制问题，有效增加了超高清视频供给。

图 1 超高清视频产业链

视频是人类获得信息的最重要途径之一，承载了海量非结构化数据，应用也最为广泛与人们的生活密不可分。从视频技术视角，视频处理产业主要经历了以下发展阶段。

20 世纪 40 年代到 50 年代是视频技术发展的关键期，信息论与视频编码的理论体系逐步成型。

20 世纪 70 年代，电荷耦合器件（Charge Coupled Device，CCD）技术的发明使得视频的采集成本逐渐降低，使得视频处理具备了数字化处理的基础。

20 世纪 90 年代，随着计算机通信技术的不断发展，视频数字化浪潮逐渐兴起，各种数字视频压缩技术和处理技术得到了快速发展，包括 ITU/MPEG 等组织构建了一系列的编解码标准，并且逐渐地应用到了广播电视、通信、互联网等领域。

从 2012 年开始，以深度学习(Deep Learning，DL)为代表的人工智能技术在视频领域再次兴起。人工智能技术越来越多融入到了整个视频从诞生到显示的端到端的流程，涵盖视频的拍摄、后期的制作、视频的压缩转码、视频的处理、视频的理解分析、视频码率的自适应分发和传输，以及视频的播放和显示的每一个环节。尤其是在视频处理和视频转码方面，AI 技术既能够提升视频的画质，提升用户的主观体验，同时也能够通过 AI 辅助编码的方式降低视频码率，降低运营商的带宽的费用。超高清视频智能逐渐成为主流视频处理技术。

从 2017 年开始，5G 移动通信技术、互联网技术、超高清视频技术和人工智能技术的进一步发展，视频技术呈现高清化、实时化和沉浸式的发展趋势。尤其是近几年移动互联网技术蓬勃发展，也催生了移动短视频、互联网直播等多种新型的视频传播应用，根据 Cisco 的互联网流量的报告，视频流量占据了整个互联网的 90%以上。人们对于视频体验的需求也越来越高，超高清 4K/8K 技术也正在不断发展，呈现产业爆发趋势。

从 2022 年开始，随着 ChatGPT、文心一言等预训练大语言模型的发布，文生视频等技术在产业逐步应用，将极大加速超高清视频智能处理技术发展。

20 世纪 40~50 年代	**信息论与视频编码的理论体系逐步成型**
20 世纪 70 年代	CCD 技术降低视频采集成本，视频处理具备了数字化基础
20 世纪 90 年代	视频技术数字化，数字视频压缩技术和数字视频处理技术快速发展
2012 年	人工智能技术兴起，与视频技术融合发展，超高清视频智能技术逐步主流
2017 年	5G、超高清、人工智能技术融合发展，对超高清视频智能诉求变强
2022 年	随着 ChatGPT、文心一言等预训练大语言模型的发布，将极大加速超高清视频智能处理技术发展

图 2 视频处理产业发展历程

从产业升级的时间窗口来看，以分辨率为重要用户体验指标的产业升级在加速，从模拟标清到数字高清用了 30 年，高清到 4K 差不多 10 年时间，4K 到 8K 不到 5 年，对视频传输的带宽需求也提到了两三百兆，8K 时间窗已然开启。

2. 超高清视频智能处理技术发展现状

2.1 超高清视频智能处理系统架构现状分析

超高清视频智能处理是应用人工智能技术将视频优化为超高清视频的过程。超高清视频智能处理是视频全生命周期过程的重要组成部分，人工智能技术能够对视频处理的每一个环节进行赋能。从视频处理工作流的视角来看，超高清视频智能处理系统功能架构的功能架构如图 3 所示，主要包括：智能画质提升、智能老片修复、智能视频编码、智能视频编辑等功能模块。

通过人工智能技术对输入的视频进行分辨率提升、帧率提升、纹理细节增强、高动态范围图像(High-Dynamic Range, HDR)转化等，提升视频的画质，使得视频画质更好地显示和呈现。通过人工智能技术，对

老片进行智能划痕去除、智能噪点去除、智能上色等修复，再进行综合画质提升（包括智能超分、智能插帧、智能增强、智能 HDR 转换等），以符合超高清播放需求。通过智能技术分析视频，进行智能化的编辑（包括特定子图像去除、智能横竖屏转换、黑边裁剪等），使得视频符合更广范围的分发终端需求，提升视频生产的效率。通过智能视频编码（包括内容自适应编码（Content Adaptive Encoding）、感兴趣区域（Region of Interest, ROI）编码等），与视频分发相结合，保证用户体验的同时节省分发的带宽。人工智能技术基于模型的学习和表达能力，实现对视频的综合智能处理。

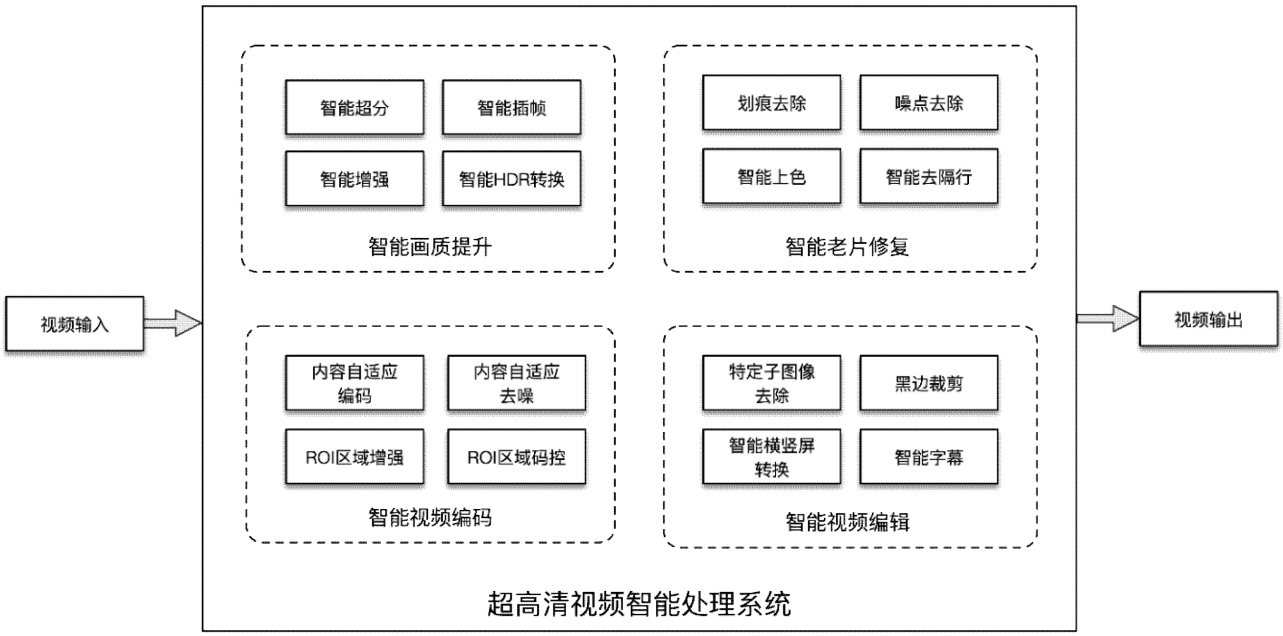

图 3 超高清视频智能处理系统功能架构

超高清视频智能处理系统技术架构如图 4 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组件。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（例如 RESTful API）等交互方式。REST (REpresentation State Transfer) 描述了一个架构样式的网络系统，指的是一组架构约束条件和原则。RESTful 指的是满足这些约束条件和原则的应用程序或设计。依托超高清视频智能处理系统，可以综合实现智能画质提升、智能老片修复、智能视频编辑、智能视频编码等。

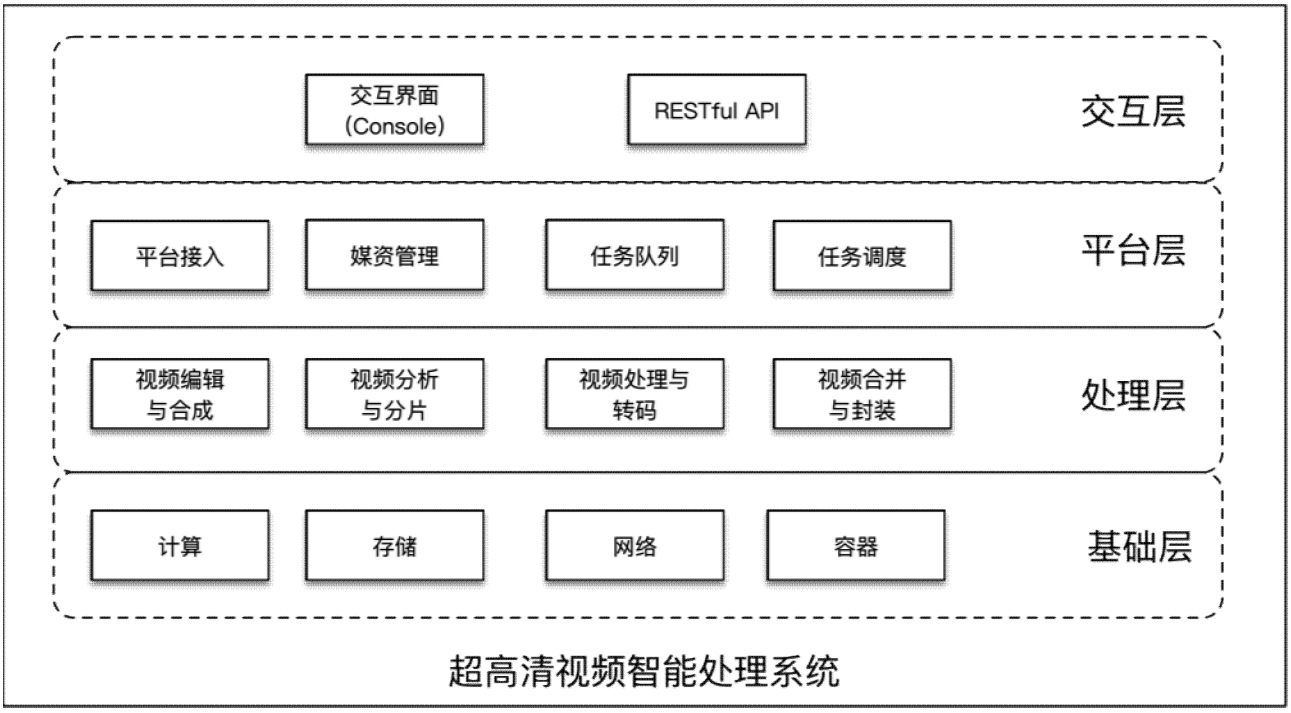

图 4 超高清视频智能处理系统技术架构

2.2 超高清趋势下智能画质提升越来越重要

智能画质提升涉及到视频质量评价指标相关的多个维度的提升，常见的影响视频质量的技术指标有视频的分辨率、帧率、高动态范围（亮度）、色域（色彩丰富度）、色深，还包括视频编码码率、编码标准（格式）、视频码率控制方式、视频内容本身等多个因素。人工智能技术在以上维度，均可助力画质提升。通过基于深度学习的智能视频超分增强技术，可以将原本帧率的低清晰度的视频，上采样到高帧率、高清晰度的视频，与传统的差值算法相比，借助于模型强大的学习能力，可以从海量数据中恢复出更多的细节信息，从而提升了视频画质。智能插帧技术使得运动视频看起来更流畅。

下面从技术角度出发，分析当前超高清视频智能处理系统的发展现状，为进一步技术应用提供参考。

2.2.1. 智能超分技术趋于成熟

智能超分是应用人工智能技术将低清图像增强到超高清图像的过程，是画质提升中的核心算法之一。由于任意一幅低清图像，都可能存在多幅合理的高清图像与之对应，因此智能超分本质上是一个病态问题，具有非常大的挑战性。

传统的智能超分算法，有基于预测、梯度、统计、切片、稀疏表达等的各种算法。近 10 年以来，从 AlexNet 在 ImageNet 图像分类任务上的成功开始，深度学习成为显学，也被广泛应用到智能超分任务中，从早期的基于卷积神经网络（CNN）的算法（如 SRCNN），到最近的基于生成对抗网络（GAN）的算法

（例如 SRGAN）。在两大超分比赛 NTIRE 和 PRIM 以及蓬勃发展的行业需求推动下，众多基于深度学习的智能超分算法不断涌现，出现了不同网络结构、不同损失函数、不同模型训练策略的算法。

基于深度学习的超分算法发展主要集中在模型框架、主干网络设计、训练策略这几个方向，各个方向的技术演进情况以及未来发展方向如下。

（1）智能超分模型框架持续升级，迭代上采样模型框架潜力巨大

智能超分模型框架有两种分类方式：一是，根据超分模型上采样步骤在超分模型中的位置，将智能超分模型框架分为前置上采样、后置上采样、逐步上采样和迭代上采样四种；二是，根据是否使用生成对抗网络（GAN）结构，将智能超分模型框架分为 GAN 和非 GAN 两种。

早期模型采用前置上采样模型框架，先将低清图像放大到目标分辨率，模型只需学习相同分辨率下的变换，降低了学习难度。前置上采样模型框架容易放大原图噪声、画面容易模糊，并且模型在大分辨率高维空间上运算的计算量远大于其他框架。后置上采样模型框架降低了计算量，开始逐渐流行。它使用神经网络对原图进行特征提取，然后在网络末端加入可学习的上采样层，输出高分辨率。后置上采样运算都在低维空间，学习难度有所增加，此外高倍超分（4 倍、8 倍）任务效果不好。逐步上采样模型框架应运而生，通过将高倍超分拆解成多个低倍超分，降低了学习难度，例如 LapSRN、ProSR 等。逐步上采样能处理高倍超分，支持多种倍率的超分，但通常网络设计复杂、训练过程不够稳定。迭代上采样模型框架，将上采样层和降采样层交替连接（DBPN，SRFBN），实现对低清和高清图像联系的深度挖掘，具有非常大的潜力和前景。

使用生成对抗网络结构(GAN)的模型框架，可以产出更好的主观效果，但在原始图像质量较差的情况下，也可能会生成奇怪的噪声。未使用生成对抗网络结构（非 GAN）的模型框架，通常在产出的客观指标（例如 PSNR、SSIM）上更优，但主观上可能存在画面过度平滑的问题。

（2）智能超分网络设计结构丰富，呈现全局和局部结合发展趋势

深度神经网络设计在计算机视觉领域的研究成果很好地应用到了智能超分任务上，目前已经应用到智能超分任务上的网络设计常用组件包括：残差块（全局残差和局部残差）、多路径学习（全局多路径（LapSRN）、局部多路径（MSRN）、尺度多路径（MDSR））、全链接块（Dense Block）、Transformer 注意力机制（通道注意力、非局部注意力）、高级卷积（膨胀卷积、分组卷积、深度可分离卷积、视频超分中用到的 3D 卷积）、区域递归学习、金字塔池化、小波变换、反次像素、xUnit 等。智能超分网络设呈现全局和局部结合的发展趋势。

（3）智能超分模型训练策略多元，需要逐步丰富供给和优化效果

智能超分模型训练策略包括：损失函数的选择、批归一化（Batch Normalization）、课程学习(Curriculum learning, CL)、多元监督等。

损失函数的选择策略，可选 L1 和 L2 损失，感知损失（对应主观质量，使用与训练的 VGG 或 ResNet 网络提取）、风格损失（保证颜色、纹理、对比度的一致性），对于采取 GAN 框架的智能超分模型还有 GAN 网络常见的对抗损失，实际使用中通常根据需求组合多个损失函数作为训练目标。

批归一化策略可以有效加速训练过程，并使训练更加稳定，在智能超分模型的训练中经常被采用。但是，批归一化处理会丢失图像的缩放信息、降低神经网络的范围灵活性，去掉批归一化可以减少大量计算，很多新的智能超分模型开始通过增大模型来替换批归一化。

课程学习策略是指模型训练时从一个简单任务开始，逐步增加难度直到满足需求，在高倍智能超分任务中应用较多。

多元监督策略通常在损失函数中引入更多的项，通过反向传播作用到模型的学习中。例如 LapSRN 采用的逐步上采样的框架，在不同的超分倍率上都有对应的损失函数。

2.2.2. 智能插帧算法持续迭代

早期电影电视的帧率普遍在 25fps 至 30fps 之间，已经无法满足刷新率不断提升的新款显示器需求，为了消除低帧率视频的卡顿感，需要用插帧技术来提升视频帧率，给用户带来丝滑流畅的观看体验。传统的帧融合或重复帧算法，直接将前后两帧的融合结果或重复帧作为新增的中间帧插入到原视频中，该算法计算简单，但帧融合会产生拖影、重复帧会带来画面卡顿，通常可作为特殊场景、低端设备的兜底方案。目前广泛使用的基于运动估计和运动补偿（Motion Estimate and Motion Compensation，MEMC）的运动补偿算法和基于光流的时间插值算法，根据计算得到的相邻帧运动估计或光流信息，生成符合原视频平滑运动关系的中间帧，可以有效提升视频流畅度，但对于运动复杂的场景，在有限算力条件下很难计算出精确的结果。常见的智能插帧技术包括：基于 CNN 的视频插帧算法、基于流式的视频插帧算法、基于 GAN 的视频插帧算法。

（1）基于 CNN 的视频插帧算法持续发展

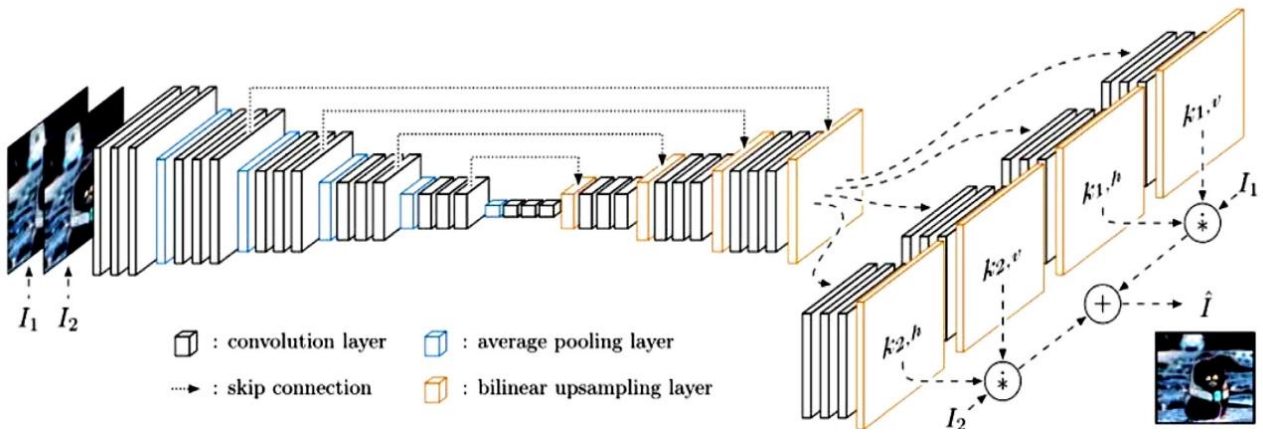

图 5 卷积神经网络框架

近年来，卷积神经网络（CNN）被广泛应用到插帧任务中。Niklaus 等于 2017 年提出了基于自适应卷积核的视频插帧算法，他们通过 CNN 神经网络，为每一个像素预测了一个 2 维卷积核，相比常见的将运动估计和像素预测分成两步的视频插帧算法，这一算法的自适应卷积核同时包含了这两个步骤，可以更好地处理运动估计无法处理的遮挡、亮度突变等问题。然而，预测每个像素都需要一次 2 维卷积运算导致计算量过大，Niklaus 等又提出了基于分离自适应卷积核（SepConv）的视频插帧算法，使用两个 1 维卷积来近似前面的 2 维卷积，并使用一个编码-解码神经网络一次性预测所有像素的卷积核，大大减少了计算量。基于卷积核的算法无法处理剧烈运动的情况（像素运动的范围超过了卷积核尺寸），Cheng 等于 2020 年提出了基于可变分离卷积核（DSepConv）的视频插帧算法，相比之前的算法，缩减了卷积核的尺寸，采用类似的编码-解码神经网络，在预测卷积核的同时，预测表征运动信息的偏移量和掩模，通过可变形卷积，获得中间帧的结果。上述算法通常适用于中间帧的预测，对于两帧之间任意时间点的预测，基于流式的视频插帧算法。

（2）基于光流的视频插帧算法占据主流

基于流式的视频插帧算法是指根据对连续帧之间对应实体的自然运动，实现对中间帧的仿真，对运动估计的计算越精准，插帧的效果越好。深度学习技术的应用，发展和优化了传统的光流插值算法和运动补偿算法。

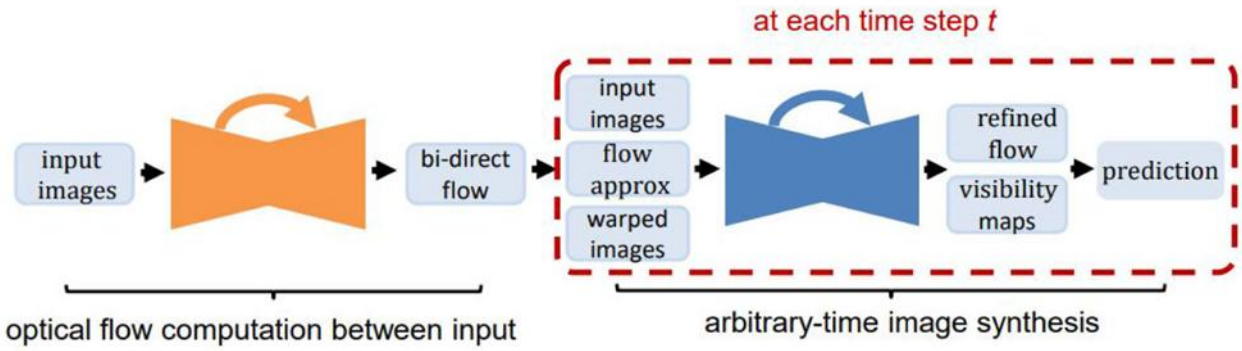

图 6 MEMC-Net

2018 年，SuperSloMo 给出了基于深度学习和光流法的基本框架，使用一个扩展的 U-Net 结构来计算双向光流，可以同时插出多帧来，在一致性上超越了传统算法和基于 CNN 的算法。

基于光流的时间插值算法主要面临两个问题，一是如何准确地获取大范围的运动信息，二是如何处理光流遮挡（前帧多个像素对应后帧的一个像素）/空洞（后帧像素找不到对应的前帧像素）导致的主观质量问题。MEMC-Net 算法使用深度学习算法实现了传统的 MEMC 框架，运动向量和补偿都通过可学习的卷积神经网络来获取，全新设计了自适应翘曲（warp）层，整合了光流信息与学习得到的补偿卷积核，以解决像素遮挡和空洞问题，该自适应翘曲层还能提升画质，在视频增强、超分等任务中应用。DAIN（深度感知视频插帧）算法将深度信息引入视频插帧算法，分别通过光流估计网络、深度估计网络、上下文特征估计网络、卷积

核估计网络来计算不同的信息，通过自适应翘曲层整合前面得到的所有结果，最后通过一个综合网络输出插帧结果，深度信息的引入可以帮助解决像素遮挡的问题和大范围移动的问题，基于上下文信息可以更好地合成中间帧，DAIN 算法在标准数据集上获得了比 MEMC-Net 算法更好的结果。BIN 算法将视频插帧和视频增强问题联合起来解决，可以同时降低运动模糊和上变频帧速率，开发了一个金字塔模块来周期性地合成清晰的中间帧，金字塔模块具有可调的空间接收场和时间范围，从而有助于可控的计算复杂性和恢复能力，金字塔模块集成了一个递归模块，可以迭代合成出时间平滑的结果，得到比 DAIN 更好的效果。

（3）基于 GAN 的视频插帧算法应用探索

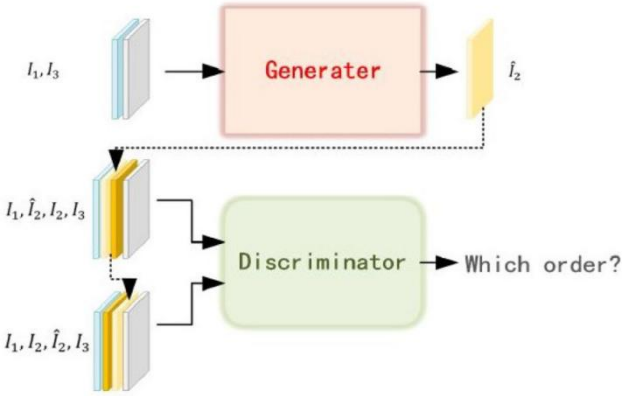

图 7 基于 GAN 的视频插帧算法基本原理

GAN 被引入到视频插帧算法中，催生了一系列基于 GAN 的视频插帧算法。FINNiGAN 是第一个使用基于 GAN 的视频插帧网络，使用了一个 SIN（结构插值网络）结构来保留画面细节，显著地减少传统算法处理高速运动画面时由于光流计算错误产生的画面鬼影和画面撕裂问题。基于 GAN 的算法，还能减少参数量，提高算法实时性。Li et al. 等于 2018 年 通过一个多尺度的 CNN 结构，实现了对大范围运动的支持，同时使用一个 WGAN-GP 网络，使插帧结果更加自然。同样在多尺度的网络结构的基础上，FIGAN 利用空间变换网络表征的光流信息，在不同的网络层级融合了感知损失，在保持相当效果的前提下，比其他算法快了 47 倍。FI- MSAGAN 引入了注意力机制到生成网络中，融合了局部信息和全局信息，以处理移动物体，不论是在性能上还是效果上，都表现得极具竞争力。

2.2.3. HDR 产业生态逐渐完善

高动态范围（high dynamic range, HDR）是一种提升视频动态范围的技术，目的是使视频更真实地反映现实场景。相比标准动态范围（standard dynamic range, SDR）的视频，HDR 视频的亮度和对比度更高，由于暗部更暗，亮部更亮，画面细节更加丰富。华为、苹果等公司，在手机上就可以实现 HDR 视频内容的拍摄、录制和编辑，极大地降低了 HDR 视频内容创作门槛。

（1）HDR 现状分析

现阶段 HDR 技术落地应用的产业挑战较大，主要体现在三个方面：一是部分技术方案的专利费用高导致产业链成本居高不下，支持的设备未形成规模，生态呈现碎片化；二是 HDR 多种技术标准共存，标准间的兼容性较差，不能覆盖主流终端的适配、认证及测试过程，导致终端呈现效果差异明显，用户难以获得一致的视觉体验；三是传统 HDR 制作流程复杂，运用 HDR 技术的超高清片源匮乏，高质量片源供给不足，超高清频道专区少、时长短，用户侧的超高清需求被抑制。

HDR 有很大的潜在市场，经过多年的发展，目前有 HDR10、HLG、HDR Vivid、HDR10+、Dolby Vision 等多个标准。HDR10 具有开放和免费的优势，但是静态元数据兼容性不足，无法保证相同内容在不同终端设备上一致化、最佳的显示效果；HDR10+是三星牵头开发的动态 HDR 标准，但是片源比较少，在国内基本没有使用；Dolby Vision 是一个完全封闭的生态，商业授权费用很高。下图对各个标准做出了一些对比的总结。

标准	**HDR10**	**HLG**	**HDR10+**	**Dolby Vision**	**HDR Vivid**
厂家	CTA	NHK/BBC	Sumsung	Dolby	UWA
峰值亮度（nits）	1k-4k	1k	10k	10k	10k
位深	10bit	10bit	10bit/12bit	12bit	10bit/12bit
传输曲线	PQ	HLG	PQ	PQ	PQ/HLG
色彩空间	Rec.2020	Rec.2020	Rec.2020	Rec.2020	Rec.2020
元数据	静态（SMPTE ST 2086，MaxFALL，MaxCLL）	动态	动态（SMPTE ST 2094 10）	动态	动态，现有标准增加动态元数据
成本	免费	免费	免费	收费	免费
兼容性	不兼容	兼容 SDR 好	不兼容	取决于 Profile 兼容性比较强	兼容性高

表 1 HDR 生态标准对比总结

HDR Vivid 是 UWA 联盟推进的国产 HDR 标准，主要有有三大核心技术：动态元数据（Dynamic Metadata）、色调映射(Tone Mapping)和饱和度调节（Saturation Adjustment）。HDR Vivid 兼容现有 HDR10 标准，在 HDR10 标准的基础上，增加了动态元数据视频中每一帧或每一个场景中的关键特性都将被提取，根据每帧或每个场景动态调整显示效果，同时显示终端基于动态元数据并结合自身的显示能力对源视频进行映射处理（Tone Mapping），为显示终端提供了更加准确的动态范围映射方式，实现不同厂家设备的显示一致性，最大限度还原 HDR 内容原有的艺术效果、还原创作者的创作意图。与业界私有 HDR 技术相比，

HDR Vivid 是一种更开放更普世的技术标准及开放方案。对产业各方来说，技术开放性以及安全性都更适于产业部署。

通过 UWA 标准的形式，定义了超高清视频呈现处理 HDR 过程，截止到 2022 年的 12 月份，UWA 联盟总共发表了 3 个部分的标准：涵盖从元数据和适配、应用指南和技术要求到测试方法标准的规定；其中第 1 部分，元数据及适配；第 2 部分，也包括应用指南 后期制作两个子部分；第 3 部分的技术要求和测试方法，包括显示设备、便携式显示设备、播放设备和播放软件这 4 个子部分。

制作端可以不修改已有的 HDR10 和 HLG HDR 制作流程，只增加动态元数据产生环节。采用符合调色习惯的图形界面，协助制作人员人工调试 HDR Vivid 动态元数据。提供自动化动态元数据生成工具，支撑大量内容的批量生产，降低开发成本。在元数据的 Codec 传输上，支持 HEVC/AVS2/AVS3/VVC 元数据的适配。HDR Vivid 的端到端的解决方案如下图所示。

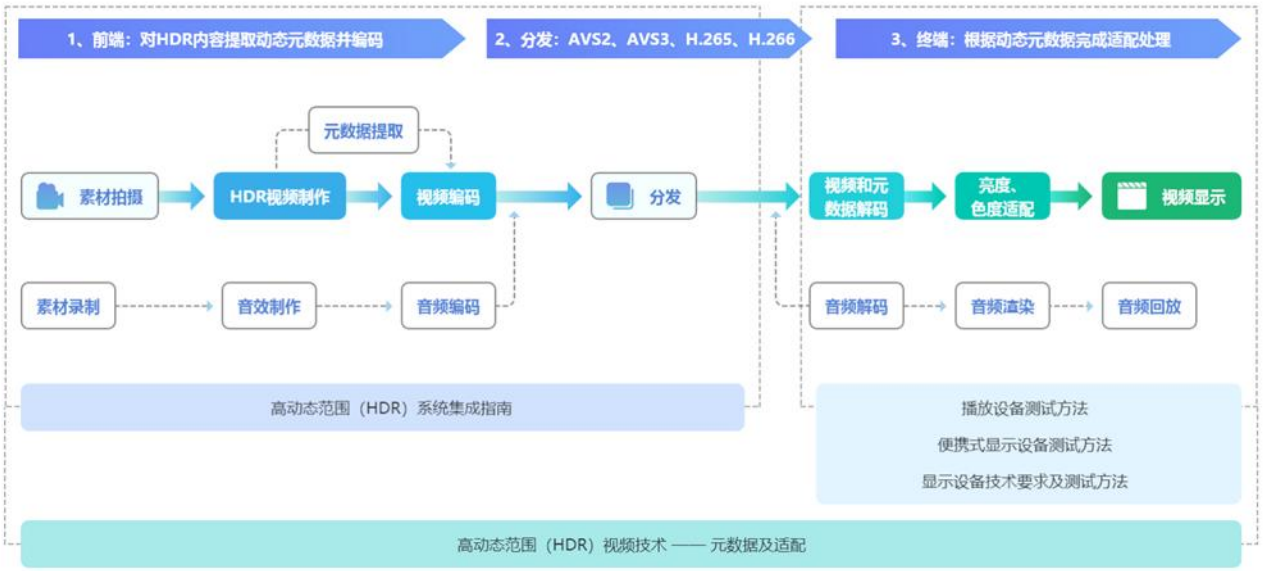

图 8 HDR Vivid 端到端解决方案

HDRvivid 涵盖了生态链端到端，方案上主要体现在视频制作、视频编码、视频传送和视频播放等环节。

（2）HDR 应用挑战

HDR 的内容是非常匮乏的，原因是传统的 HDR 制作工艺非常复杂，制作门槛高，运用 HDR 技术制作的片源少，超高清高质量的片源供给不足，不能满足用户对于超高清日渐增长的需求。

在当前 UGC 视频的时代背景下，尤其是短视频行业的快速发展极大地降低了视频的制作门槛，但是现在的 HDR 内容主要还是 PGC 为主。主要的编辑软件包括 Final Cur Pro X、Adobe Premiere Pro CC、Davinci Resolve 等等。

所以，如何快速产出或者补足 HDR 超高清内容的素材，就成为了一个巨大的市场机会。基于 AI 的智能视频处理技术有着广阔的发展空间，可以通过智能处理，进行存量老电影、电视剧素材的修复，通过超分技术进行标清到 4K 的转换，通过 HDR 技术进一步提升效果，达到接近真 4K 的标准节目效果。支持真的 HDR 播放和显示的播放器和屏幕还比较少，OTT 大屏、不同移动端的终端屏幕参数也参差不齐，终端呈现的效果差异也很明显，用户难以获得一致的视觉体验。所以使用一定的策略，进行色调映射和各种设备的适配，各种 HDR 格式和标准之间的转换，是一个长期需要解决和优化的问题。

在智能手机的生态上，HDR 的解码、显示等基础技术的支持也还没有那么普遍。比较基础的是对 HDR 标准的支持，目前 HDR 标准和格式共存，标准之间的兼容性也不一样。目前 HDR Vivid 的生态已构建完成，在国内处于快速发展的态势，制作工具随着 BMD 公司的 Davinci 产品、Filmlight 和索贝工具的全面支持，使生产内容的速度和成本大为降低；HDR Vivid 的内容规模已上线 2 万多个小时，国内 HDR Vivid 后续正在大批量生产中；编解码系统，当虹、数码视讯已提供支持能力；MTK、高通、海思等等芯片已全面支持；终端方面，已经有华为、荣耀、小米、夏普、康佳等国内企业、三星等国际企业的手机、电视支持。

AI 在视频领域的技术发展，为视频内容的重建提供了新的技术手段。基于 AI 的超分辨率技术可以实现标清到高清（SD 转 HD）、或者高清到 4K 甚至 8K 的分辨率的提升，可以弥补大量的图像细节；通过基于 AI 的逆色调映射（Inverse Tone Mapping）技术和色彩增强技术，可以实现对比度、色彩饱和度等多个层面的提升。这些提升的细节，需要用 HDR 视频的高动态范围和宽色域来进行表达。NTIRE 2021 首次举办了 HDR 视频图像生成技术的大赛。根据典型的应用场景，可以将智能视频重制划分为智能画质提升和智能老片修复两个分类。其中智能老片修复可以极大地提升传统的人工修复效率，而超分和 HDR 则进一步提升弥补细节，调节亮度和饱和度，尽量提升到接近真 4K 的水平。

（3）基于 AI 的 SDR 转 HDR 上变换算法

目前网络上分发的视频资源仍以 SDR 为主流，HDR 视频资源相对稀缺。由此，一些研究人员开始提出 HDR 重建技术，通过算法将现有的 SDR 图像/视频转换为 HDR 图像/视频，以弥补 HDR 资源的不足。最早的 HDR 重建方法通过融合一组不同曝光度的图像实现，由于需要提供不同曝光度的输入图像，其比较适合静态场景，而无法处理动态场景的视频。单帧 HDR 重建（single image HDR reconstruction）算法尝试从单幅 SDR 图像生成 HDR 图像，在应用层面更为泛用。这类方法一般提取一些图像特征，然后构造逆向色调算子（inverse tone mapping），将 SDR 的像素值映射回 HDR 范围。然而，由于曝光不足和过曝区域中信息已经丢失，恢复细节较为困难。

自 2017 年以来，一些基于深度学习的 HDR 重建方法被提出，其中又以单帧重建为主流。早期的深度学习方法尝试让网络学习从单幅 SDR 图像生成多张不同曝光程度的图像，再用传统方法将这些生成的图像融

合，也有方法直接让网络端到端地学习从 SDR 图像到 HDR 图像的映射。随后一些研究（SingleHDR 和 HDRTVNet）从原理出发，对现实中 HDR 到 SDR 的转换过程进行了分析建模，然后用多个网络分别学习其子过程，使方法更具备理论基础。

在 NTIRE2021 的单帧 HDR 重建比赛中，第二名的 HDRUNet 在 HDR 重建时额外考虑了降噪过程，令生成图像具有更低的噪声水平和更丰富的细节。同时，同时使用超分模型和逆色调映射模型融合的方法，取得更好的细节效果，同时提升动态范围，成为一个技术趋势。

2.3 智能老片修复提效显著已逐步走向应用

拍摄年代较早的视频内容，受限于当时的成像设备和存储介质，通常质量较差，为了适应超高清时代对画面质量的要求，需要对老片进行修复。以常见的胶片老电影数字化修复为例，大致的步骤包括：

a)胶片数字化，将胶片内容通过图像采集卡转换为数字内容；

b)画面修复，修复画面中的划痕、污点、噪点，将黑白画面转成彩色画面等；

c)画面质量提升，通过调色、调整对比度等多种手段，提升画面主观质量；

d)画质审核及其分发。

传统对老的视频素材画面的修复和质量提升，主要依赖人工进行手工修复，人力成本高，修复时间长。基于人工智能的智能老片修复算法可以在人工修复前提供粗修，大大节省人工成本，加快修复效率。

2.3.1. 智能划痕去除分割填补分步法落地

智能划痕去除分为两个步骤：智能划痕分割和智能划痕填补。

智能划痕分割需要给出画面上划痕区域的精确像素点集合，是一个典型的视频图像分割问题。常见的图像分割算法，有从 FR-CNN 检测算法演进而来的 Mask RCNN 算法，在 FRCNN 预测检测框的基础上，新增了一个预测分割结果的分支，Mask RCNN 沿用 FRCNN 的两阶段检测框架，使用了带有特征金字塔的 ResNet-FPN 作为主干，在 FRCNN 既有的 ROIAlign 之后添加了卷积层，进行分割结果的预测。常用的 Pixel2Pixel 的网络，如自编码解码网络，也可以用来预测分割结果，UNet 在此基础上，将降维前的特征和升维后的特征拼接起来，在小目标分割领域获得了更好的效果。在视频图像分割领域，端到端的模型设计已经成为主流。例如，有基于循环神经网络来对时间维度和空间维度对视觉特征联合建模的方法 PDB-ConvLSTM，这一方法设计了一个金字塔膨胀卷积（PDC）模块在不同尺度上同时提取空间特征，然后将提取到的空间特征输入到一个深度双向卷积 LSTM 网络（DB-ConvLSTM）中学习时空信息。另一大类方法是通过两个平行的网

络（如 MATNet），从原始图像和光流信息中分别提取特征，然后用来做分割的预测，这种方法可以充分利用运动信息，面对运动目标可以得到更精确的分割结果。

智能划痕填补是一个视频修复（Video Inpainting）问题，是指在视频图像中的损坏区域中，填充符合常理和时空一致性的内容。不同于图像修复（Image Inpainting）只需要考虑单张图像内的空间信息，视频修复还要考虑视频前后帧的时间信息，使修复结果能保持瞬时一致性，不损失视频播放的流畅度。基于深度学习的视频修复算法大致分为三类：

a)基于 3D 卷积的算法。以 LGTSM 为例，整体采用 GAN 架构，提出了可学习的门控时移模块，并基于这个模块设计了 3D 门卷积的生成网络，这个网络可以结合相邻帧的信息，更加精确地修复不规则的损坏区域，并且不影响正常的区域。

b)基于 Transformer 注意力机制的算法。相比基于 3D 卷积的算法，Transformer 模块可以更好地融合时间维度和空间维度的全局信息。基于 Transformer 注意力机制的视频修复算法也不断涌现， 例如 STTN，设计了一个融合前后帧信息和帧内信息的时空 Transformer 网络，将自注意力填充到损坏区域，在标准数据集上获得了比 LGTSM 更好的效果。FuseFormer 则在 transformer 模块的前向传播过程中做了改进，将需要处理的视频图像切片（patch）进行重叠，聚合了更多的信息，增加了感受野，使修复的结果在边界上的过度也更加自然。

c)基于光流的算法。引入光流信息来辅助修复，解决了基于 3D 卷积算法和基于 Transformer 注意力机制算法的修复结果存在瞬时不一致的问题，获得了更好的瞬时一致性。典型的基于光流的视频修复方法将视频修复任务看作像素传播（pixel propagation）问题，如 DFVI（深度光流引导的视频修复），天然可以更好地保持瞬时一致性。常见的光流方法大致可分为三个阶段：a）光流完成（flow completion），完成损坏区域的光流信息补充;b)像素传播（pixel propagation），在光流信息引导下通过双向转移像素到损坏区域;c)内容幻想（Content hallucination），使用一个预训练的图像修复网络生成损坏区域的最终内容。上述步骤通常是分开的、并且有很多额外的处理（例如图像融合、求解线性方程等），计算量大，并且额外处理无法用 GPU 加速。E2EFGVI，将上述三个步骤全部转化为可训练的网络模块，提出了一种端到端的光流引导视频修复框架，在光流完成阶段，在 mask 视频中一步完成操作，不包含了多个复杂步骤；对于像素传播，E2EFGVI 升级为特征传播，在特征空间中使用可变卷积模块实现，引入了特征级别的操作还能弥补光流估计的误差，得到更好的效果；对于内容幻想模块，E2EFGVI 提出了一种时间焦点 Transformer 结构来建模空间和时间维度上的长程依赖关系。E2EFGVI 在标准数据集的性能超过了 LGTSM，STTN 和 FuseFormer。

2.3.2. 智能噪点去除端到端自适应效果显

视频图像在拍摄、数字化、处理、传输等不同阶段都有可能产生各种各样的噪声，根据噪声和视频信号的关系，可以分为三种形式：

a)加性噪声，噪声与视频信号无关，是叠加到视频信号上的，例如信道噪声及光导摄像管的摄像机扫描图像时产生的噪声；

b)乘性噪声，噪声与视频信号相关，例如飞点扫描器具产生的噪声、电视图像中的相干噪声、电影胶片中的颗粒噪声；

c)量差噪声，视频信号在量化过程中产生量化误差，反映到接收端的噪声。

智能噪点去除算法如下：

a)基于滤波器的算法。经典的滤波器算法，例如中值滤波器、均值滤波器、高斯滤波器等，利用人工设计的低通滤波器去除视频图像中的噪声。

b)基于模型的算法。基于模型的算法试图对自然图像或噪声的分布进行建模，然后使用模型分布作为先验，试图获得清晰的图像与优化算法。基于模型的算法通常将去噪任务定义为基于最大后验的优化问题，其性能主要依赖于图像的先验。常见的基于模型算法包括：非局部自相似（NSS）模型、稀疏模型、梯度模型和马尔可夫随机场（MRF）模型。基于模型的方法有很强的数学推导性，但在使用过程中涉及复杂的优化问题，去噪过程非常耗时，并且在大量噪声的情况下恢复纹理细节的性能显著下降。

c)基于学习的算法。基于学习的方法侧重于学习有噪声图像到干净图像的潜在映射，基于深度网络的算法已成为主流方法，因为其比基于滤波、基于模型的算法获得了更好的去噪效果。DnCNN 使用了残差块和批归一化处理，主要针对高斯噪声进行去噪；FFDNet 侧重去除更复杂的不通噪声，将噪声水平图作为输入，使卷积神经网络能适用不同噪声水平的图像；CBDNet 优化了 FFDNet 提出的噪声水平图，通过一个 5 层的全连接神经网络来自动估算噪声水平，实现了端到端的自适应去噪神经网络。DeamNet 尝试提升神经网络的可解释性，参考了传统的基于模型的去噪方法，提出一种自适应一致性先验，实现了一种端到端的可训练和可解释的深度去噪神经网络。

2.3.3. 智能上色帧间稳定性待进一步提升

给黑白影片上色是智能老片修复的一个重要课题，深度学习的发展，使智能上色成为可能。上色问题是一个经典的 Pixel2Pixel 问题，根据实际使用到的信息的不同，上色算法大致可以分成以下几种：

a)基于输入黑白图像本身的上色算法。上色算法可以视为一个通用的图像到图像的转换算法，基本的思路是通过多个编码层降维、然后再通过解码层升维重建出结果，整体的网络框架选用 GAN 结构，通过生成对抗，训练得到较好的上色器。常见的选择有用编码解码网络或类 U-Net 作为生成网络的主干（ICGAN，

cGAN、PixColor 等），在训练中引入残差块、批处理化等常用深度神经网络技术来解决梯度消失问题，实现了黑白图像到彩色图像的变换。

b)基于参考彩色图像的上色算法。基于参考图像的上色算法，可以使上色结果更加符合用户的意图，如基于实例的深度上色（Deep Exemplar-based colorization）算法。该算法包含两个子网络：相似度网络和上色网络。相似度网络通常使用一个标准的 VGG19 结构来提取参考图像和上色结果的特征，得到参考图像和目标的相似度，上色网络则将输入的黑白图像和参考图像的色彩通道以及相似度结果拼接起来，使用一个 U-Net 结构来生成最终的上色结果。

c)有参考文字描述的上色算法。用户使用文字描述，也可以向智能上色算法表达上色的意图。以 Text2Colors 为例，这一算法使用了两个 GAN 网络：文字调色板生成网络（TPN）和基于调色板的上色网络（PCN），TPN 使用一个颜色-文字的数据库训练，负责根据文字描述构建调色板信息，PCN 负责将调色板信息和输入的黑白图像变换为彩色图像。

d)基于深度学习的上色算法。基于深度学习的上色算法的上色效果主要由数据驱动，不同的训练集得到模型的上色效果差异巨大，并且缺乏较好的评价手段。

有参考彩色图像的上色算法，更方便引入用户的创作意图，在未来智能老片修复的实际需求中，可以得到更好地落地应用。

2.4 智能视频编辑具备广阔的市场发展空间

视频编辑是传统视频生产的重要组成部分，随着人工智能技术和网络技术的不断发展，除了传统的广播电视、互联网长视频之外，短视频技术取得了迅速的发展。智能手机的普及，极大的降低了视频拍摄的门槛，生产了越来越多的 UGC 视频内容，智能视频编辑需求越来越大。

Kamua 采用 GPU 算力进行视频内容的智能编辑，比如横屏转竖屏、使用 AI 进行目标识别，在 BMD Davinci Resolve 软件里边进行使用。或者仅将人工智能技术应用于对预先存在的视频片段执行特定任务，例如专业的后期制作的软件包括 Davinci，Baselight 等都融入了更多到了 AI 的插件。

云端的视频编辑，可以应用更多的 AI 能力，比如智能视频拆条、字幕生成、横转竖等的能力，极大地提升生产效率。

2.5 智能视频编码驱动的感知编码技术落地

人工智能是视频压缩的巨大希望，像 4K/8K 超高清视频编码、AR、VR 等需要大量数据才能达到更好的效果，一方面需要更好的视频压缩技术，另一方面需借助人工智能技术来进一步降级更多的带宽，以提升用户的体验。现有编解码器（如 AVC、HEVC、AV1 和 AVS3），可以与 AI 技术相结合，研发 AI 驱动的编码技术，其中基于 AI 的感知编码技术是一个重要研发方向。

内容自适应转码是云转码的一个热点技术，其基本思想是根据被转码视频内容本身去设置一个合适的码率阶梯（Ladder）配置，而不是传统的根据视频的分辨率去设置码率。内容自适应有不同的基本类型：基于视频类别、基于视频、基于视频分段、基于帧。内容自适编码的目标是通过 AI 模型强大的表征能力，在特定的编码器基础上，找到视频的编码参数（编码码率、CRF 等）、视频分辨率、视频质量之间的关系。但是从编码器原理的角度来看，每一个视频序列，甚至每一个视频帧都有唯一的码率-失真（Rate-Distortion）曲线，能够合理的根据视频内容的复杂度，给出最优的码率分配，是可以通过 AI 能力进行优化的问题。

从 2015 年开始，国际上以 Netflix，YouTube 等为代表，很多 AI 技术被应用到视频编码的过程中。Netflix 首先提出了“Per Title”编码的技术，针对每一个视频，采用不同的编码参数（码率、分辨率）进行多次编码，得到视频的码率、分辨率以及视频质量之间的关系，这里采用 VMAF 作为视频质量指标。该方法计算量巨大，无法满足直播等低延迟条件下的应用需求。YouTube 采用神经网络估计编码的 CRF 值，得到预期视频的恒定质量的目标。BeamR 构建了一个帧级别的内容自适应编码方法，采用一个具有自研专利技术的质量判别指标，对输入到编码器的每一帧进行编码，如果达不到预期的编码质量，那么就对视频帧进行重新编码，该技术通过恒定质量的编码，达到尽量节省码率的目的。Brightcove 提供云转码服务，支持端到端的内容自适应的编码到分发能力，通过对视频内容以及网络传输条件的分析，使用机器学习模型产出最适合当前网络分发的编码配置，达到端到端提升用户体验的目标。Fraunhofer 提出 FOKUS 方法，基于机器学习驱动的编码解决方案，通过多种模型，估计视频场景的时间、空间复杂度、分辨率、帧率、色彩、亮度等视频内容复杂度与视频码率、分辨率和质量之间的关系，通过 CDN 进行进一步的分发。Hivision、Microsoft、ATEME 等公司也都推出了类似的方案。

随着短视频、超高清应用的快速发展，国内的视频应用也异军突起，基于云的转码与处理也快速发展，在 toB 的方向上各个云厂商也提出了类似窄带高清、高清低码、感知编码等转码的能力。百度智能云基于”智感超清”的产品能力，在构建了短视频场景上的数百万场景的训练数据集，通过 AI 模型学习了视频编码质量（VMAF），与视频的编码 CRF（码率、分辨率）之间的关系，在预期的目标质量输出的条件下，达到最优码率输出目的。

随着 AI 技术的进一步发展，可以进一步构建端到端的视频内容、网络负载、视频编码配置之间的关系，构建用户体验驱动的智能视频编码的方案。在基于传统编码框架设计的编码工具性能已趋于极限的背景下，探究智能化的视频编码技术是当前行业的重点发力方向。未来视频数据的消费场景不再单纯局限于人眼视觉，服务于机器视觉的视频编码也将迎来巨大应用市场。

3. 超高清视频智能处理系统应用现状

3.1 超高清视频智能处理系统已多样化供给和广泛应用

超高清视频智能处理系统主要有公有云服务、私有化软件、软硬件一体机三种产品形态。

在公有云服务形态下，内容拥有者或内容运营方直接使用公有云服务，不需要算力等硬件投入和算法模型、工程界面等软件投入，但数据出厂需要数据传输和安全加密等成本投入。

在私有化软件形态下，内容拥有者或内容运营方投入自建算力集群、集成算法模型等软件，数据不出厂不需要数据传输和安全加密等成本投入。在软硬一体机形态下，内容拥有者或内容运营方直接投入软硬一体机，快速部署使用，缺点是算力不够弹性，只适用于小规模算力需求场景。

按照实时性要求不同，超高清视频智能处理可分为实时处理和离线处理两种方式。实时性要求不高时，使用小算力集群进行离线处理即可；在时间紧、任务重时，离线处理则需使用大算力集群。在线处理时效性高，但成本高。离线处理的视频，内容方可再增加主观效果审核、内容安全审核等环节，保证内容质量和新增内容合规。直播场景需实时在线处理，非直播场景可离线处理。老片智能修复主要为离线处理方式。

超高清视频智能处理标准产业已有初步积累。在系统层面，世界超高清视频产业联盟（UWA）发布了 T/UWA 010-2022《智能视频处理系统通用技术规范》，电子行业标准 2022-0592T-SJ《超高清视频智能处理系统通用技术规范》已完成公开征求意见。在技术方面，围绕视音频的超高清 4K 和 8K 标准、AVS2/3 编解码标准、画质和音质 HDR Vivid 和 Audio Vivid 标准等已经发布并经产业实际商用环境验证。超高清视频智能处理解决方案可支持相关技术标准。

超高清视频智能处理的解决方案，产业已经有丰富供给，百度智能云、华为、中国移动、当虹、数码视讯、涌现科技的解决方案介绍详见附录 A，此外阿里云、腾讯云、火山引擎等也提供相关解决方案。

当前超高清视频智能处理系统在动画、高清内容转制 4K/8K 内容等方面应用效果良好，高清内容转制的 8K 内容在 8K 终端上看主观效果不俗。超高清视频智能处理系统生成 4K/8K 内容的成本比原生拍摄 4K/8K

内容的成本低两个数量级，促进了超高清视频供给增加。超高清视频智能处理系统处理也存在不足，如处理的视频画面质量不如原生 4K/8K 内容、画面复杂内容的提升感知有限等。造成超高清视频智能处理系统不足的主要原因是数据集不够丰富。内容方拥有丰富数据集，中长视频大屏端受短视频挑战，经营压力较大，对人工智能等新技术投入谨慎。超高清视频智能处理解决方案提供者缺乏数据集，算力和算法投入大，商业端变现周期长。

超高清视频智能处理技术的发展是一个螺旋上升的过程。国家超高清战略，极大推进了超高清产业的发展，超高清视频智能处理技术广泛应用于广播电视、文教娱乐、安防监控、实时通信等行业，在老片修复、超高清重制、视频增强、智能编解码等场景极大的提高了超高清视频生产效率、降低了内容生产成本。在广播电视、文教娱乐、安防监控、实时通信行业的超高清视频智能处理应用案例详见附录 B。

3.2 广播电视：老旧内容高效修复、视频超高清化重制

广播电视领域拥有非常丰富且有价值的历史影像资料，包括：纪录片、电影电视、新闻资料等，其中的很多内容都具备历史文化价值。当前随着 5G、大屏等基础设施的不断发展，用户对于视频体验的要求也越来越高，超高清 4K/8K 的显示技术逐渐成为了主流，高清化/超高清化成为必不可少的消费体验。随着互联网技术的快速发展，在广播电视的分发渠道也从传统的有线电视，向 IPTV、OTT 进行扩展，同时屏幕终端方面，除了机顶盒+电视机屏幕的方式，OTT TV 也可以面向移动屏幕的多个端进行分发，这样就对内容的多样的智能化的处理提出了需求。因此老片修复、超高清重制等需求在广播电视领域迅速增加。当前有很多厂商构建了超高清视频智能处理解决方案，以公有云、私有化、一体机等产品形态交付落地。通过将 AI 粗修+人工精修相结合，大大提升了视频处理的效率和成本。但是针对不同场景不同类型的视频内容，智能处理的效果还需进一步细化加强，智能处理效果的规范标准还需进一步统一。目前，超高清视频智能处理系统主要解决以下问题：

（1）提升修复效率

标清视频每秒 25 帧，10 分钟的视频共计 15000 帧，人工修复人均每小时修复 300 到 600 帧视频，且需要多人协同，耗费巨大的时间成本和人力成本。划痕、脏斑、噪波等问题是很多老视频存在的问题，非常影响观看体验，但是又在视频中持续时间长，单帧修复费时，问题不容易辨认等特点，是人工修复的痛点和难点。

（2）提升视频画质

老视频出品年代早，受当时视频、存储等技术的限制，存在较大量的黑白视频，加之存储条件差异，视觉效果普遍较差，只能通过超分或者图像增强手段来提升画面质量。

（3）攻克工具短板

当前人工修复的软硬件工具，均是国外厂商提供，比较老旧落后，且售后维护服务较差。超高清视频智能处理工具国内解决方案供给丰富，典型解决方案介绍详见附录 A。

（4）提供弹性算力

海量的视频内容资源，包括老片翻新后的视频，或者年代稍近，画质稍好的视频，都需要进行超高清的重制，以满足超高清视频指标的要求，可以借助于云端海量 GPU/NPU 算力资源，或者借助专业化硬件处理产品进行处理。

在广播电视领域的电影行业，很多存量的老电影，来自于不同的年代。在介质方面，有胶片、磁带等不同介质，由于存储介质随着时间变化或者其他历史原因，会产生划痕、脏点、霉变等，经过数字化之后就成为各种噪声。老电影的修复是一个专业领域，多个国有电影厂有专业修复科室，以及中国电影资料馆等机构。电影修复是一个非常耗费人力的过程，一个电影修复师每天能修复 6000 帧左右的画面，而修复的 6000 至 8000 帧，片长大约 4 至 5 分钟，如果遇到修复难度大的片子，修复师每人每天只能修复几百帧，片长只有大约 20 至 30 秒。通过 AI 技术可以自动检测划痕，可以采用智能填补方式明显减少/去除划痕，可以采用超高清重制方式生产超高清版本，将全新画质的片子呈现在观众面前，极大地提升了内容生产效率。

3.3 文教娱乐：提升画质保障体验、码率优化降低带宽

在超高清发展的大趋势下，各大互联网视频平台也在不断朝着更清晰的播放体验升级，提升画质体验是提升用户粘性的关键因素。此外，由于互联网视频平台 IT 支出里，视频分发加速的带宽成本占比非常大，因此降低带宽成本也成为了主要诉求。随着互联网红利逐渐见顶，降本增效成为了各大平台发展的主旋律。

（1）提升画质

在长视频点播场景，各大平台均在 1080P 之后推出 4K+HDR 的画质选项，爱奇艺推出了帧绮映画，腾讯视频推出了臻彩视界、优酷视频推出了帧享影音、B 站推出了杜比视界。这类画质的视频均采用了 HDR 标准，互联网场景里，HDR10、HDR10+、DolbyVision 用的比较多，同时部分平台还引入了全景声音质，从画质和音质全方位带来更加沉浸感的体验。从内容供给上，部分是真 HDR 视频进行二次编码，有部分内容则通过 AI 能力将 SDR 转换成 HDR 视频，并进行增强超分处理。

在短视频领域，大部分内容是 UGC 内容，这类视频画质参差不齐，格式多种多样，且有部分画质较差的视频需要尽可能提升观看体验。这里大部分厂商用到的能力包括：去黑边、去黑帧、去水印 logo、增强超分、感知编码等，通过多种技术手段进行智能处理，提升画面清晰度和内容质量。在增强超分提升清晰度方

面，有两种技术方案：一种是在服务端的编解码预处理环节，通过 AI 模型进行处理后再分发播放；一种在播放端进行实时处理，即服务端下发低质的视频，在终端解码后通过 AI 模型处理后渲染播放。两种方案各有优劣，服务端处理算力充足，模型可以更大效果更佳，但带宽无法节省；终端处理算力有限，模型效果相对较差，但可以节省下行带宽成本。

在直播领域，同样可以采用短视频处理方案，达成降本增效的目标。与点播处理场景不同，直播场景对于处理延迟的要求更高，需要更加轻量级的模型对视频进行实时的智能化的处理。

（2）降低分发成本

降低带宽和存储成本的主要方式是在信源端提升视频的压缩效率、降低视频码率大小，与此同时还需要保证视频画质，这里大部分厂商用的手段有两方面：

升级编码器或编码标准：各家厂商和视频平台，均在投入研发资源进行编码器的优化，相比开源的编码器，例如：X264、X265，经过优化的编码器一般能达到同等画质下降低 20-50%的码率大小。此外，各大视频平台通过升级更新的编码标准，例如：从当前主流的 H.264 升级到 H.265、AV1 等，可以大幅降低码率。不过新的编码标准在处理成本和播放终端适配上是一个新的挑战。

AI 驱动的智能感知编码：包括基于内容自适应的码率控制（CAE）、人眼感兴趣区域（ROI）的处理的与码率控制优化等。内容自适应编码 CAE 方向在互联网场景里也应用广泛，因为不同视频甚至同一个视频的不同画面的复杂度是不太一样的，所需要的编码码率也是不一样的，因此可以通过场景分片，利用 AI 模型去预测最优的码率，简单画面可以降低码率，复杂画面可以提升码率，做到更细粒度的码率控制，从而降低码率，提升画质。ROI 编码是说，人眼通常对于所关注的画面区域更加敏感，而人眼关注的区域一般是前景主体区域，包括：人脸、人体、物品、文字等。通过 AI 模型对这些主体区域进行识别和跟踪，并对主体区域进行重点的码率分配，和去噪、增强等处理，提升主体区域的画质，从而可以大幅提升主观清晰度。而对于非 ROI 区域可以降低码率降低视频大小。主观画质提升了，也可以一定程度的降低整体文件的码率，以达到节省大小的同时，还能保证画质不会太差。

在互联网，尤其是 UGC 领域很多视频是由于拍摄环境不好，或者多次压缩、编辑而产生了很多的伪影、噪声、马赛克、抖动、模糊等等质量的问题。当前也是一个海量视频产生的时代，那么如何提升互联网视频的用户体验，同时也降低视频的分发带宽，是各个视频运营商亟待解决和优化的问题。通过基于 AI 的画质重建技术，可以极大程度地解决上述 UGC 场景下所固有的视频质量问题。

3.4 安防监控：视频处理还原现场、自适应编码控成本

安防监控也是超高清视频智能处理技术的发力点，从看得见到看得清、看得懂，核心诉求主要两方面：降低码率大小，从而降低上行带宽成本和存储成本；提升画质以提升后续 AI 分析的准召率和可观看性，更好的还原监控细节场景。

（1）降低存储成本

监控场景的视频量非常大，需要存储 30 天至半年不等，用于内容查看、监管等等，因此存储的成本压力也是较大的。另外，随着视联网的兴起，很多摄像头视频数据需要上传至云端进行存储、处理和分析，导致出口上行带宽成本压力大，因此压缩视频码率，降低存储和带宽成本也成为了核心诉求。

当前的主流压缩方案有两类：一是引入更优质的编码器，例如：将 264 升级为 265 编码，二是针对监控场景做特定的压缩算法优化，包括：剪裁掉重复精致画面片段等。交付形态包含了硬编和软编两种方案，在算力上需要更高的要求，这里需要衡量算力成本的增加和存储带宽成本的节省的收益差，才能更好的将智能处理技术方案落地推广。

（2）提升画质

监控场景有很多摄像头比较老旧，输出的视频画质清晰度较差，会影响后续内容回看，以及后续 AI 分析的准召率，因此提升视频清晰度在超高清视频发展的大趋势下的需求逐渐增多。监控场景类的视频问题有其特殊性，例如：光线弱导致画面亮度低细节不清晰、大雾雨雪天气导致画面噪点多质量差、过度压缩导致马赛克等问题。这次可以采用 AI 模型进行处理，包括：调节亮度、去噪、细节增强等，以提升画面清晰度。此外，对于监控类的内容，变化的画面以及人物是主要关注对象，因此可以通过感知编码针对人物进行重点优化，通过内容自适应编码对运动画面进行重点优化，从而提升视频清晰度。在安防监控领域，由于摄像机的型号、压缩标准、光照、天气等一系列影响因素，都会影响采集到的视频画质。同样也可以使用 AI 技术对画质进行重建，一方面可以使得视频让人眼看起来更清晰，另外也可以让基于智能化的分析，带来更好的效果。

3.5 实时通信：轻量运算降低能耗、联动编码自适网络

实时通信场景下，超高清视频智能处理同样发挥着重要的作用。清晰度、流畅度以及延时是衡量实时通信的三个重要指标，需根据应用场景寻找权衡三者的最佳编码方案。其中画质的提升与其他行业的述求和方案大致相同，由于泛终端设备性能，网络条件的各种变化，实时通信场景下的视频编码方法不仅要满足高效、低延时需求，同时还需要考虑到通信带宽紧缺问题。目前实时通信的主要挑战如下：

（1）绿色算力

为了实现实时通信在更多类设备上的运行，编解码算法需要满足两点，1)编码压缩率高；2)算法算力开销小。由于一些家庭设备采用电池，并且固定安装，更换不方便，所以对于算力的要求很高。为了解决此问题，首先需要有合理的触发机制，来激活对应的视频前处理，进行业务处理，同时在后续保证前处理与编码算法选择，也要保证算力和能耗的最佳比。这样才能保证设备长时间运行，满足用户的日常需求。

（2）提升流畅度

实时通信的设备很多情况下处于家宽网络，由于路由器等设备的网络情况复杂，因此保持实时通话流畅性具有很大的挑战。除了引入通常的网络抗丢包、抖动技术，如 ARQ、FEC、不对等保护等手段，还需要有拥塞控制，其中拥塞控制的核心目的就是让“发送速率”尽可能去逼近“可用速率“。

（3）提升编码效率

目前，大多应用追求高清视频、超高清视频和 4K 视频等高分辨率的实时视频体验。在未来全息、XR 等新型应用不断涌现的背景下，给当前通信资源带来了巨大压力，势必需要大幅提升当前视频编码效率。

在这一背景下，语义视频编码方法以保留视频语义信息并保证传输视频质量为目标，有望解决上述挑战。相较于传统视频编码技术，语义编码利用深度神经网络模型提取视频更高维度的语义特征，进一步挖掘视频数据帧内和帧间的语义相关性，对不同部分的视频内容进行分别处理从而提高压缩率。目前典型的语义编码方法主要选取关键帧作为共享知识库并将帧的内容根据变化情况进行区分，多帧之间共同且变化较少的元素作为帧间的静态元素，变化较多的元素作为动态元素。对于静态元素，根据具体任务使得网络在关键帧之间少传输甚至不传输静态背景的变化信息。对于动态元素，通过挖掘关键帧之间例如人脸关键点位置关系等动态元素的时域或者空间相关性，从而实现根据动态元素前后帧的特征变化和共享知识库来重构两个关键帧之间的视频片段。语义编码也能处理点云等数据提取据局部结构特征，利用 3D 数据的稀疏性来进行分层特征学习挖掘深层的语义信息，从而使得仅传输部分表示局部区域的数据并根据数据之间的相关性来重构 3D 数据，从而大大减少视频传输的数据量。

4. 超高清视频智能处理系统测评方法

超高清视频智能处理系统主要包括：智能画质提升、智能老片修复、智能视频编辑、智能视频编码等功能。近年来，涌现了海量的智能处理算法，在实际应用中，由于影响视频质量的诸多因素，如压缩算法、传输带宽、设备性能等，超高清视频质量往往无法达到预期的效果。因此，如何对其处理能力进行评测成为制约

其实际应用的关键。T/UWA 010-2022《智能视频处理系统通用技术规范》（原 CUVA 010-2021）是行业内系统的测评规范，其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。参考其第 7 章测试方法对超高清视频智能处理系统进行评测，包括测试视频序列确定、测试视频处理、主观质量评测和客观质量评测几个阶段。

根据测量方法的不同，视频质量评测技术可分为主观质量评测方法、客观质量评测方法，其中客观性评测方法又可以分为全参考模（主动式）、全参考方法（被动式）两大类。

主动式视频质量测试采用测试仪表向被测设备或系统发送一个标准的参考视频流文件，然后接收经过被测设备或系统处理后的视频流文件，通过一定的算法比较这两个视频流文件的差异，可计算出视频质量分值。这种全参考模型的测量方法一般比较复杂，计算量较大，测试结果也相对比较准确，它适合于在实验室环境对超高清视频智能处理系统做功能性验证测试。被动式视频质量测试则采用测试仪表对网络中的实际视频流进行监测和捕获，然后对其数据流特征进行分析，并计算出视频质量分值。这种无参考模型的测量方法，计算量小、测试速度快，适用于对超高清视频智能处理系统输出的视频流进行实时监测和故障告警。

与传统的编解码的质量评测方式类似，智能老片修复类的任务，建议采用主观与客观相结合的方法进行测试。在这个测试过程中，视频测评序列本身非常重要。可以采用通过超高清原始视频加入噪声、损伤或者降低分辨率的方式，制作测试视频。可以参考的流程如下：

a)将 4K （或 1080P）原始测试序列 A，经过处理，比如降分辨率，降帧率等，并加入各类视频损伤（如加上划痕、雪花、彩虹、失焦、黑边等），作为测试序列 B；

b)用智能修复系统对测试序列 B 进行修复，以达到较好的修复效果。得到修复后序列 C；

c)对原始测试序列 A 和 修复后序列 C 进行比较，得到测试结果值 VQ。

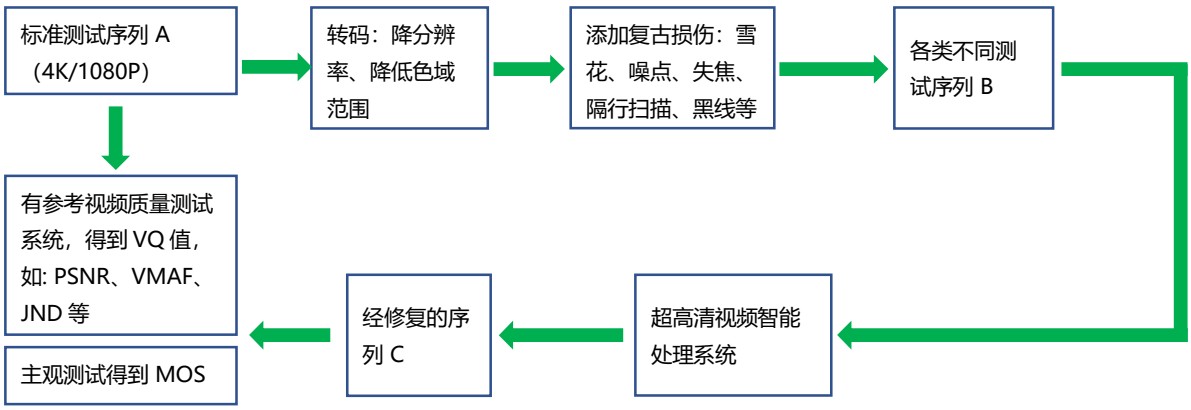

图 9 测试视频制作流程

4.1 测评视频序列确定

超高清重置测评视频序列：

a) 源测试序列可以从既有高质量 4K 视频下采样得到，亦可以采用现有低清视频。

b) 总数应不少于 20 个片段，每个视频内容应是动态的，持续 10～15s。

c) 视频画面应细节纹理丰富，包含文字、人体、人脸（不同肤色、不同种族），包含运动画面。

d) 视频规格：高清 HD（1920x1080），颜色空间 BT.709，位深 8bit，25FPS，标准动态范围。

老片修复场景测评视频序列：

a)源视频总数不少于 20 个片段，内容应是动态的，持续 10～15s。

b)应包含带划痕的画面。

c)应包含有噪声的画面，噪声包括但不限于雪花噪声、椒盐噪声、高斯噪声、伪影、块效应等。

d)应包含黑白视频。

4.2 测评视频处理

在确定了测试视频序列后，需要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。

a) 智能超分：目标视频能够保持图像细节清晰，且判定满足超分要求。

b) 智能增强：输出的目标视频能够保持边缘纹理锐化、符合人眼主观感受，输出的目标视频无明显的隔行效应。

c) 智能插帧：目标视频能够保持画面流畅、无明显抖动、无明显闪烁。

d) 智能 HDR：目标视频输出分辨率为 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。目标视频能突出亮部细节和暗部细节，色彩饱和度更高。

对于老片修复场景，还应进行划痕去除、噪点去除、智能上色等处理。

a)划痕去除：支持划痕去除，划痕去除强度能按照阈值调整。

b)噪点去除：支持噪点去除，噪点去除强度能按照阈值调整。

c)智能上色：支持智能上色。

4.3 主观质量测评

使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，受试者观看基准视频图像和目标视频图像，两种图像以随机方式先后展示，要求受试者对每一对图像对的质量进行评分，得到源视频图像主观质量分数 a 和目标视频主观质量评分 b 。

可按以下公式计算画质提升率：

$E=(b-a)/a \times 100\%$

其中，E 表示视频图像的质量提升率，a 表示源视频图像主观质量评分，b 表示目标视频主观质量评分。

超高清重置主观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。

老片修复主观打分则应考虑以下因素：

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。

4.4 客观质量测评

超高清视频体验质量，取决于视频的清晰度、流畅度等因素，涵盖了视频的分辨率、帧率、码率、编码和终端多个维度的指标。超高清视频的客观质量评测基于图像层，通过对视频画面关键质量指标的衡量，刻

画视频不同维度的质量，如根据视频图像的模糊度、块效应、对比度、噪点度、色彩丰富度和曝光度等指标从多个维度评价视频质量，使评价结果更加符合用户的主观体验。

从对参考视频的依赖上，客观评价指标评价方法分为：全参考方法（Full Reference, FR）、部分参考方法（Reduced Reference，RR）和无参考（No Reference，NR）方法。

4.4.1. 全参考（主动式）测评

全/半参考方法主要包括峰值信噪比（Peak signal to noise ratio，PSNR）、结构相似性指数（Structural Similarity Index，SSIM）和视频多方法评估融合（Video Multimethod Assessment Fusion，VMAF）等。其他还有基于统计特性：NIQE，VIF；有基于深度特征的：如 LPIPS，DISTS 等。

a) PSNR，即峰值信噪比，是峰值信号的能量与噪声的平均能量之比。PSNR 是最普遍、最广泛使用的评鉴画质的客观量测法，虽然和人眼看到的视觉品质不完全一致，但目前仍作为对照其他指标的基线。人们一般使用该指标来衡量被压缩后的视频的失真程度，值越大越好，一般取值范围 20-40。PSNR 的优点是计算复杂度低，但其局限性是和主观评价有一定差距，并且计算需要原片源片作为参考，无源片无法计算

b) SSIM，即结构相似性指数，是从亮度、对比度与结构来对两幅图像的相似性进行评估。SSIM 应用于局部可抵抗失真程度突变，效果更好。实际是对各种局部窗口的 SSIM 做平均，并用高斯加权函数对每个局部的统计值进行加权防止出现块效应。但是该算法的局限性也是需要原片源片作为参考，否则无法计算

c) VMAF，即视频多方法评估融合，本质上是模拟人眼评价的主观结果。VMAF 算法相对于前两个算法更贴近于人眼视觉的视频评价标准，可以提供更接近于用户的主观评价。它将人类视觉建模和机器学习结合，模拟人眼给出的客观评分（5 分制），优点是采用大量的主观评价数据集作为训练集，可自定义训练算法和模型，从而构建符合自主业务需求的质量评价标准，且可无限接近人眼主观感受。缺点是不同分辨率、不同视距和类型的视频评测得分不能直接比较，需换算处理。

4.4.2. 无参考（被动式）测评

由于近些年互联网视频的爆发增长以及其参考源难以获取的特性，无参考的质量评价方法逐渐成为近些年来的研究热点，无参考方法无需原始视频信息，直接根据待评价视频的信息评价视频质量，具有更好的灵活性和通用性，以及更广泛的应用价值。目前，无参考方法主要包括深度学习、无参考图像空间质量评估器（Blind/Referenceless Image Spatial Quality Evaluator，BRISQUE)等，内核是依赖人工智能技术。

a) 基于深度学习的方法：主要通过构建深度神经网络，学习图像的视觉特征以构建图像质量评价模型，或直接通过端到端来学习失真图像到图像视觉质量的函数表达。

b) 无参考图像空间质量评估器（BRISQUE）方法：是一个经典的利用 NSS 进行 NR-IQA 的模型。它不需要对图像进行频域分解，仅仅使用简单的归一化过程，就使得数据呈现有规律的分布。模型简单且高效，可扩展性强，计算的复杂度较低。

参考 623 YD/T 3776/3777/3778/3779-2020 宽带视频服务用户体验评估系列测试标准，该标准所提出的 uVES 视频质量评测方法也可适用于对超高清视频智能处理系统所输出的视频流进行实时监测和质量评价。

uVES 视频质量评价标准按照其处理信号层级分为 Mode0, Mode1, Mode2 三层模型，三层模型所需的输入信息按获取难度递增，且评价精细化程度递增。采用分层形式有利于根据实际应用条件进行灵活调整，如不具备完全解码能力时可仅采用基于码流层的模型。

a)uVES Mode0：为基于码流层的轻量级模型，它主要衡量视频经过不同的编码方式和使用不同的码率在不同尺寸大小的显示终端呈现，用户所获得的源体验质量每种视频分辨率都存在显示质量的极限，如果要提升用户体验，需要提升视频分辨率；终端设备的屏幕越大，对视频分辨率要求越高，不同分辨率的体验差异越明显。

b)uVES Mode1：基于码流层的复杂模型，与 Mode0 相比，Mode1 综合考虑显示质量和视频源压缩质量。Mode1 需要从编码数据分组及比特流中采集视频帧关键编码信息，衡量视频压缩对于视频源质量的损伤情况，主要衡量的指标包含终端屏幕尺寸、帧类型及帧大小、编码量化参数、运动矢量信息以及帧内编码单元跳过比例。因此，Mode1 的计算复杂度相对变大，模型准确度更高，适合计算精度要求较高的场景。

c)uVES Mode2：基于图像层模型，Mode2 通过对视频画面关键质量指标的衡量，刻画视频不同纬度的质量。Mode2 需要从播放器连续采集视频帧的图像层关键质量信息，根据图像的模糊度、块效应、对比度、噪点度、颜色度等指标评价视频源质量。

针对智能处理的超高清视频的质量评价， 应综合考虑主客观质量评测, 通过客观参数及主观画质等方面多维度评测，计算处理方法的有效性。转换处理后的评测，采用人工主观测评的方式，对修复后的序列帧的画质进行打分。只有各维度客观评测和总体体验观感的主观评测都达标才可通过评测。

4.5 主客观融合质量评价

基于标准比对源的主客观融合质量评价方法，采用有损视频序列作为智能处理系统的输入源生成待评价视频，在优化结果评价阶段，引入符合商业播放要求的视频作为评分标准，结合双刺激主观评价和有参考客观评价方法，对经智能处理后的优化视频开展主客观融合质量评价。

根据有损视频序列及标准比对源获取方式的不同分为两种评价路径，如下图所示。

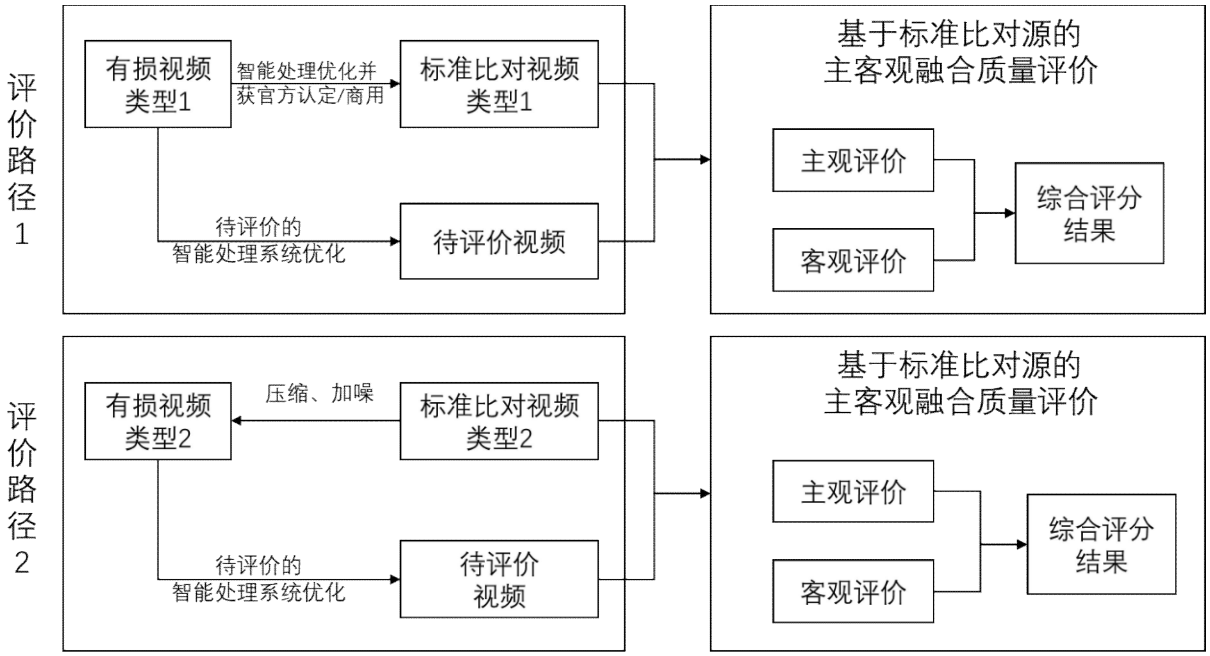

图 10 基于标准比对源的主客观融合质量评价方法

其中，

a)有损视频-类型 1：可采用现有画面抖动、模糊、分辨率较低或带噪点、划痕等的有损视频。

b)有损视频-类型 2：对既有超高清高质量视频进行下采样或加噪，得到高清/标清有损视频。

c)标准比对视频-类型 1：采用智能处理优化且达到商用标准的优质超高清视频，作为标准比对源。

d)标准比对视频-类型 2：已获官方认定的或商用的超高清高质量视频，作为标准比对源。

e)待评价视频：针对有损视频，经智能处理优化后，输出待评价视频。

f)智能处理优化包含但不限于：智能超分、智能增强、智能插帧、智能上色和智能 HDR 转换中的一种或多种方式。

g)测试视频总数应不少于 20 个片段，每个视频内容应是动态的，持续 10～15s。

h)视频画面应细节纹理丰富，包含文字、人体、人脸（不同肤色、不同种族），包含多种颜色信息，包含运动画面。

i)视频规格：

高清 HD（1920x1080），颜色空间 BT.709，位深 8bit，25FPS，标准动态范围；4K 超高清 HDR（3840×2160），颜色空间 BT.709 或 DCI-P3 或 BT.2020，位深 10bit，50FPS，高动态范围；8K 超高清 HDR（7680×4320），颜色空间 BT.709 或 DCI-P3 或 BT.2020，位深 10 或 12bit，50FPS，高动态范围。

主客观融合的计算法方法，综合主、客观评价结果，按照如下比例进行计算，得出最终评分。

评价项目	评分占比

有参考主观评价	70%
有参考客观评价	30%
合计	100%

表 2 主客观融合得分计算表

5. 超高清视频智能处理产业问题和建议

5.1 超高清视频智能处理模型不够通用，按应用场景训练专用模型

超高清视频智能处理涉及到从视频的生产、视频编解码、视频图像的画质提升以及视频的播放和显示等整个视频工作流的多个环节，AI 已经越来越多的融入到了整个流程，其中超高清视频画质提升处理和智能视频编码是两个非常重要的技术手段。超高清视频智能画质提升是指利用人工智能（AI）模型对视频/图像进行画质重建，提升了输入到视频编码器的视频画质，这样可以提升终端的视频体验。智能视频编码是利用 AI 技术进行辅助的编码优化的技术方向。

智能画质提升涉及到视频质量评价指标相关的多个维度的提升，常见的影响视频质量的技术指标有视频的分辨率、帧率、高动态范围（亮度）、色域（色彩丰富度）、色深，还包括视频编码码率、编码标准（格式）、视频的码率控制方式、视频内容本身等多个因素。人工智能技术在以上维度，均可助力画质提升。通过基于深度学习的智能视频超分增强技术，可以将原本低帧率、低清晰度的视频，上采样到高帧率、高清晰度的视频，与传统的差值算法相比，借助于模型强大的学习能力，可以从海量数据中恢复出更多的细节信息，从而提升了视频画质。智能插帧技术使得运动视频看起来更流畅。

超高清视频画质提升处理技术，包括以 AI 超分辨率为代表的低层级的图像/视频处理技术、AI 智能插帧、基于 AI 的 SDR 转 HDR 的技术等等，都取得了非常快速的技术发展。以 AI 超分为例，从最早的比较简单的 3 层网络 SRCNN 开始，得益于 ResNet 残差学习方法的使用，超分网络的层数也不断的增加，比如 VDSR 的模型，网络层数达到了 20 层以上。随着演进了更复杂的网络结构，比如基于生成对抗网络（GAN）的超分模型。网络结构中，一个是生成网络，用来从一个低分辨率的图像生成一个高分辨的图像；另外一个是判别网络，用来判定生成网络生成的图像的真假，在训练网络的时候，这两个网络相互博弈最终达到平衡。在推理阶段，可以通过 GAN 网络生成细节纹理非常逼真的图像，取得更好的主观的效果。从实现方案上来看，也从单帧图像的超分方案，逐渐演进到了视频级别（多帧）的方案，将低分辨率作为输入，然后进行帧间对齐，特征

提取，特征融合，最后重建生成高分辨率视频。随着深度网络的使用，模型的训练和预测的复杂度逐渐升高，主观效果有很大的提升，但是随之而来的对算力的要求也越来越高。这样在实际应用场景上的落地，也提出了很高的算法优化和工程优化的要求。

尽管超高清视频智能处理的技术发展很快，但是人们也需要认识到现实中的视频类型、噪声等复杂很多，在模型的泛化能力和使用能力方面需要进一步进行优化。首先学术界基本都是在特定数据集上进行模型的训练和测试，比如超分或者 HDR 模型的训练，低分辨率的图像是采用下采样或者人工降质的图像进行模拟的，与实际场景的图像存在一定的差距。在模型训练过程中，一些模型采用了比如 MSE 这样的损失函数，也不一定能够反映真实的人眼感知的质量。所以，人工智能处理视频画质相关技术的成熟度需要进一步提升，部分场景下的处理后效果还达不到业务使用的要求。对于不同年代不同类型的视频，画质问题点是不一样，需要 AI 模型具备更广的适配性，同时需要深耕垂类场景做精细化的调优。AI 模型需要大量的数据输入进行训练，这里的数据集的多样性、丰富性、代表性非常关键，也是当前模型效果提升的一个瓶颈。此外，如何评价视频重建后的画质，也是视频处理和编码领域一个持续以来的技术难题。例如对于老旧电影的艺术性，应该尽量的去遵循原始导演的创作意图，如何做到『修旧如旧』， 对于 AI 模型自动处理也是非常大的挑战。

除了画质效果外，视频处理速度当前也需进一步提升。智能处理技术需要依赖大量的 GPU 算力，而且当前的处理速度也相对较慢，部分场景里无法达到业务的速度需求，同时也增加了成本压力。因此不断优化技术，提升性能，从而提升产品的性价比，也是业内急需解决的问题。

5.2 超高清视频智能处理测评码流缺失，共建产业视频测评码流池

超高清视频智能处理解决方案的应用还处于市场探索期，当前业内上下游还没有达成共识。T/UWA 010-2022《智能视频处理系统通用技术规范》（原 CUVA 010-2021）于 2021 年 7 月发布，尚处于推广期。

对于业务方，如何评测厂商的产品方案是一个大难题。一是测试数据不一定完整，不能很好的覆盖各类处理场景；二是评测工具、评测人员不一定专业，主观评价的成本较高；三是评测标准行业共识有待提升，不统一，无法量化，无法更好的与厂商协同。

对于技术服务厂商，当前是各自为战构建自己的产品方案，行业内统一标准没有共识，依靠自己的资源进行摸索，产品方案提升速度有限。

建议在 UWA 联盟内组织打造一套客观性测试的码流，并且组织实际客户场景的视频测试。

5.3 超高清视频内容供需量严重不平衡，智能处理加速供给补缺口

当前超高清视频智能处理解决方案的商业化落地还处于早期。部分能力在互娱、广播电视、安防监控等场景里得到一定应用，但是离大规模商用还存在一定距离。在互娱场景，需要进一步降低成本，提升性价比，扩大使用范围。在广播电视领域，需要进一步提升处理效果，并挖掘更多历史存量视频内容处理后的商业化价值，带来超高清视频智能处理方案的正向循环促进。在安防监控领域，需进一步探索应用创新，深入到各类垂类场景里优化落地。户外大屏应用、家庭应用，超高清视频交易平台。

5.4 大屏巨幕时代大屏用户体验待提升，加强中长视频超高清布局

人机界面决定了人和数字世界的连接，人机界面可见两个趋势。一是平面显示的巨幕化和超高清化，包括家庭电视和专业影院；二是 VR/AR 等个人设备的发展。这两个趋势带来的巨幕化、超清化体验将大大促进消费升级，对 4K 和 8K 超高清内容的需求强烈，尤其是 8K 内容，逼近人眼的分辨率，可近距离观看，具有相当的沉浸感，能带给人身临其境的感受。

大屏仍具有独特的价值。内容为王，创意爆发，在数字世界或者元宇宙时代，视频内容传递什么样的文化和内涵，给人带来何种冲击，具有无穷想象空间。相较于短视频给人密集的信息冲击、即时刺激和趣味，大屏适合主打沉浸式的、能够给人带来深度人生感悟的中长视频，配合立体声效打造家庭影院效果，为快节奏生活时代带来一些不一样的慢节奏体验。

大屏产业需要破局。抖音、快手为代表的短视频企业，打造了极致体验，获得用户偏爱，商业表现极其亮眼。大屏则存在套餐套娃，体验较差，内容触达率低等问题。大屏产业需要一场自我革命和破局，在视频体验、易用性、明确简单的收费模式等方面进行革新，打造用户购买电视即可得内容的服务模式，打造家庭影院级的沉浸体验。电视台、互联网视频和电影等在大屏端的发展则是冷暖自知。原生的 4K 和 8K 内容越来越多，同时 AI 等革命性技术大放异彩，极为有力的缓解了内容痛点问题。

综上，借助人机界面新变化和超高清体验提升，呼吁中长视频内容方，应该在大屏巨幕时代，加强布局 8K/4K 超高清内容高地，牵引消费者认知，端到端打造用户体验，提供给消费者新的不一样的人文体验。

6. 超高清视频智能处理产业未来展望

6.1 超高清视频智能处理模型向大模型演进

近些年来，随着软硬件设施的进步，一些学术机构和企业开始推出大模型(Foundation Model)，利用海量规模的预训练模型和训练数据，在许多任务上打败以往的中小模型。以往 AI 模型的常用训练范式为监督学习，其需要大量标注数据的支撑，训练数据的质和量对于模型性能有着关键性影响。然而，对不同的任务需要分别收集和标注大量数据，其成本往往较高。借助迁移学习思想，可以先在上游任务上使用大量数据预训练模型，然后再微调模型，使其能够胜任下游任务。这样对于下游任务而言，需要的训练数据较少，降低了数据收集成本。由于下游模型继承了上游模型的部分知识，上游模型性能对于迁移后的下游模型性能有着重要影响。得益于大模型强大的泛化能力，将大模型用作上游预训练模型，可以让下游的小模型取得很好的效果。如今，“大规模预训练+微调”成为一种新的 AI 发展趋势。

随着 transformer 结构的提出，大模型首先在 NLP 领域诞生，比如 GPT、BERT、RoBERTa 等。这些语言大模型采用自监督训练范式，虽然使用了海量训练数据，但并不需要人工标注。语言大模型表现出惊人的语义理解能力，在文本翻译、聊天对话等任务上几乎达到人类水平。在视觉领域，基于视觉 transformer (Vision Transformer, ViT)、去噪扩散概率模型 (Denoising Diffusion Probabilistic Model, DDPM)等结构和方法，也诞生了一批大模型，如 DALL·E、V-MoE、CoAtNet、Stable Diffusion 等。这些大模型目前主要应用于图像分类、图像生成、图像理解等任务上，也取得了出乎意料的表现。在图像生成任务上大模型尤为成功，近期的 Stable Diffusion 能够在短时间内生成大量高质量画作，并可以通过微调学习不同作画风格，在大众中引发广泛讨论。

在智能视频处理领域，常见的任务如超分辨率、色彩增强、去噪、智能上色等为 low-level 视觉任务，目前相关的大模型研究并不多。一些研究基于 transformer 结构提出了相对较大的模型，如 IPT、SwinIR 等，并在图像超分、图像复原等任务上进行了验证，相比以前模型有一定指标提升。还有如 VRT、RVRT 等使用 transformer 的视频级模型，其一次输入多个视频帧，相比图像级模型能够更充分利用帧间信息互补。基于 DDPM 方法，谷歌针对先后提出了 SR3 和 SR3+模型。SR3 通过迭代细化的方式实现图像超分，在主观测试中，测试者有接近 50%的概率误认为超分后的图像为真实照片。SR3+在 SR3 结构上小幅改进，并引入盲超分退化模型，能够生成更真实合理的细节。SR3+使用多达 61M 幅图像训练，模型参数量达 400M，验证了增大模型和数据规模能够显著提升超分效果。同样基于 DDPM，谷歌还提出了一个通用的 image-to-image 模型：

Pallette，其被应用于图像上色、图像修补、图像扩展、JPEG 复原等图像复原任务上，在实验中也取得了比 GAN 更真实合理的复原效果。

虽然大模型在各个领域都涌现出超强的能力，但是其推理效率低和算力需求大是产业落地中不可忽视的问题。low-level 视觉任务本身耗费算力就较大，一般规模的模型在目前的视频处理应用中已经较慢。前面提到的一些基于 transformer 和 DDPM 的图像/视频处理模型，其相比于如今百亿、千亿级的语言大模型，规模并不算大，但其推理所需算力依然很高。直接部署的话，在一般硬件条件下推理效率极低，甚至可能显存不足无法使用。因此如果要落地应用，未来也许还是需要通过微调将大模型能力迁移训练到更小的模型上，或者是在某些算力充足的特殊应用场景下用来获得最佳处理效果。

6.2 AIGC 加速超高清视频智能处理技术发展

AIGC（AI Generated Content）是使用 AI 技术自动生成文字、图片、音频、视频和代码等内容，其发展主要得益于生成算法、预训练模式和多模态等 AI 技术的融合，以及大数据的积累和预训练模型的研究等。AIGC 的发展对视频处理和创作是一场新的变革，通过 AI 生成图像和视频大大降低了创作者的门槛，而且相关研究者也致力于对 AIGC 生成高质量长视频进行推进。在超高清视频处理领域，与 AIGC 相关度比较高的是 text2image 和 text2video 两大技术，即用户输入文本后自然语言处理模型理解，图像或者视频生成模型则生成与文本描述符合的图片或者视频。

早期的 text2image 主要基于 GAN 方式，最早可以追溯到 2016 年 Reed 等人使用无条件 GAN 实现，之后 GAN 的变种用来进一步提升生成能力，如文本条件 GAN(text-conditional GAN)中的生成器根据提取的文本特征试图生成真实的图片，这种方式生成的图片只有 64x64 分辨率大小。之后通过增加 what-where 信息和 StackGAN 的设计可以生成 128x128 和 256x256 的分辨率。随着 transformer 的发展，相关技术被广泛用于 text2image 上。基于 transformer 的 DALL-E(GPT-3 的变种)由于其强大的学习能力和大规模的训练数据可以根据模糊的语言概念生成高质量的 256x256 图像。之后，基于扩散模型的 DALL-E 2 可以运用 CLIP(contrastive language-image pre-training)模型直接学习图像和文本描述之间的相互关联，生成 1024x1024 的图像而且具有更高的真实性和艺术性。Google 提出的 Imagen 可以根据场景描述生成高质量、高分辨率的图像，其框架中也是使用扩散模型达到生成高分辨率图像的目标。

Text2video 合成是目前相对比较新的研究方向，早期的文本转视频工作主要是在简单领域或者场景上的视频生成，如数字移动和特定的人体动作等。首个工作是 2017 年的基于循环注意力 VAE 的 Sync-DRAW 模型实现视频生成，之后 GAN 的生成方式也被引入。由于数据和模型的发展，文字转视频在近两年大放异彩。

微软亚洲研究院在 2021 年开发的多模态模型 NUWA 可以实现文本、图像、视频之间的生成、转换和编辑，主要采用的技术是 3D transformer 编码器和解码器。升级版的 NUWA-Infinity 采用全局自回归嵌套局部自回归的生成机制，从而生成全局一致且细节丰富的高质量图像和视频，而且生成的视频可以是任意大小的高分辨率或者时间长度。Meta 于 2022 年 9 月 29 日推出 Make-A-Video 是基于 text2image 模型和无监督方式的视频数据训练得到视频生成模型。Google 也在 2022 年针对视频品质设计了 Imagen Video，在 Imagen 的基础上构建级联的视频扩散模型，用户在输入文本之后经过自然语言处理预训练模型和基本视频扩散模型以每秒 3 帧的速度生成 24x48 的 16 帧图像，之后使用空间和时间超分模型生成每秒 24 帧、总长 128 帧且分辨率为 1280x768 的 5.3 秒视频。Google 还发布的另一款文本转视频工具是 Phenaki 模型，为了使得生成视频有任意的长度采用双向掩码 transformer，而且生成的视频贴近文本描述，因此可以用一系列的文本产生有连贯性的多个视频，该模型目前能够生成 2 分钟以上的长视频。清华大学提出的 CogVideo 可以生成时长 4 秒左右的分辨率为 480p 的视频，内容一致性和运动真实性方面都比较好。

虽然 AIGC 的 text2img 和 text2video 得到了爆发式的发展，但是也面临着多方面的挑战。对于抽象文本的理解目前的模型尚存在不足，甚至会导致生成模型输出不相关的结果。目前 AIGC 的是实现基本都是基于大规模的模型和海量数据训练得到，在生成阶段甚至需要几秒钟直至几分钟得到图像，而生成视频时间耗费更长，大模型在不同端侧的部署和对算力需求的提升也是投入工业化使用的重要瓶颈。

7. 参考文献

[1]贾川民等，视频处理与压缩技术，中国图像与图形学报

[2]Ultra HD Forum Guidelines, https://ultrahdforum.org/wp-content/uploads/UHD-Guidelines-V2.5-Fall2021.pdf

[3]HDR Vivid 系列标准， http://theuwa.com/download/

[4]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，中华人民共和国广播电视和网络视听行业标准

[5]Rishi Bommasani etc，On the Opportunities and Risks of Foundation Models，https://arxiv.org/pdf/2108.07258.pdf

[6]Netflix，Per-Title Encode Optimization，https://netflixtechblog.com/per-title-encode-optimization-7e99442b62a2

[7]Zhihao Wang, Jian Chen, Steven C.H. Hoi，

[8]Deep Learning for Image Super-resolution: A Survey，https://arxiv.org/abs/1902.06068

附录 A: 超高清视频智能处理解决方案提供商

A.1 百度智感超清，让内容焕发新生

A.1.1. 厂商介绍

百度是拥有强大互联网基础的领先 AI 公司，以“用科技让复杂的世界更简单”为使命，坚持技术创新，致力于“成为最懂用户，并能帮助人们成长的全球顶级高科技公司”。 人工智能时代的 IT 技术栈包含“芯片-框架-模型-应用”四层，百度是全球为数不多在这四层进行全栈布局的人工智能公司。从高端芯片昆仑芯，到飞桨深度学习框架，再到文心预训练大模型，到搜索、智能云、自动驾驶、小度等应用，各个层面都有领先业界的自研技术。

百度智能视频云团队专注于提供音视频相关的端到端产品能力和解决方案。其中智感超清智能视频处理是核心产品之一，主要致力于通过 AI 能力进行自动的老片修复和画质提升，从而大幅提升视频画质体验。经过三年多的持续优化，已具备完善的智感超清产品体系，同时，在文教娱乐、广电传媒、安防监控等场景落地了多个案例，帮助客户大大提升视频处理的效率，减轻了人工修复的压力和成本，一定程度快速弥补了超高清内容的缺失，让经典内容焕发新生，给客户带来更加高清沉浸感的视频体验。

A.1.2. 方案介绍

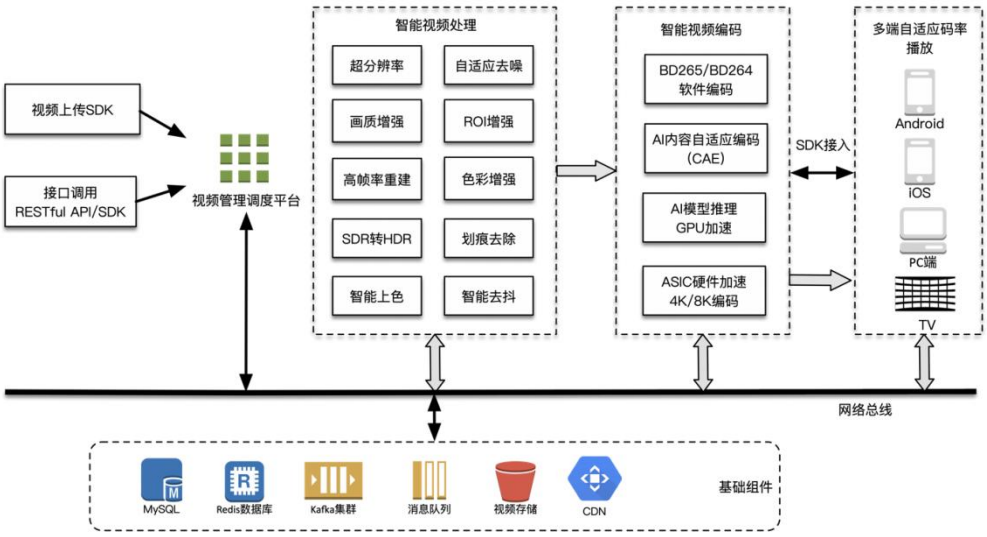

图 11 百度智能云智感超清产品架构图

(1)主要功能

智能编码：采用客观性质量评价指标（PSNR，SSIM，VMAF 等），标注了海量的视频数据，建立视频特征数据库，通过 AI 深度学习模型进行训练，通过模型预测可以得到待编码的视频在一定视频质量下的最优编码参数。简单的场景分配较低的码率就能获得较好的画面质量，同时能降低文件大小，而较复杂的场景则分配较高的码率去获得较好的画面质量，通过智能编码，在人眼主观上实现一个恒定质量的用户体验，同时在码率上能平均节省 20%-30%。

感观增强：智感超清产品基于 AI 技术实现了创新性的视频增强解决方案，借助于深度学习技术，模型可以对画面的颜色、亮度、对比度，边缘纹理信息进行增强，让画面看上去更漂亮。从视频压缩的角度出发，为了突破基于传统信号压缩编码极限，必须充分利用人眼视觉系统（HVS）的特性。在帧内码率分配环节，可以将重点考虑人眼感兴趣的区域，比如运动、人脸、肤色、文章、纹理等区域等，采用合适的量化参数，确保将有限的码率分配到人眼更关注的重点区域，取得更优的主观质量。

超分辨率：百度智能云智感超清采用 AI 的方法上采样重建到高分辨率的视频（如 SD->HD 或 HD->4K），提升视频画面的细节。包括视频空间分辨率提升（即画面低清升分至高清）、视频时间分辨率提升（即低帧率插帧至高帧率）、色深分辨率提升（由 RGB 通道 8bit 色深提升至 10bit 及以上）以及色域分辨率的提升（由 BT.709 颜色空间提升到 BT.2020），产出满足行业标准的真 4K 视频。

智能 SDR to HDR：在百度智能云智感超清产品里，借助 AI 模型能力，将 SDR 内容重制（Remastering）到 HDR，拓展亮度和色彩空间，适配 HDR 显示能力，保留创作者意图，包括：将 BT.709 转换到 BT.2020，将 8bit 转换到 10bit，同时将暗部细节增强、过曝细节修复。支持的 HDR 格式包括：HDR10/PQ、HLG、DolbyVision、Vivid 等。

老片修复：针对老视频存在的划痕脏斑，画面噪声，低分辨率及黑白颜色等问题，通过专有大数据驱动的方式训练 AI 模型，得到修复各个问题的划痕脏点分割、区域填补、自适应去噪、多维度视频升分辨率及基于参考帧智能上色等 AI 算法。另外，基于数十万量级的专业修复图像帧显著提升了修复算法的鲁棒性与泛化能力。

（2）核心优势

去划痕算法创新：基于注意力机制的循环迭代式划痕分割神经网络，对视频帧的划痕、噪斑进行有效分割与定位。该网络模型能有效利用视频前后帧内容的相关性，建模视频内容与划痕内容在时序上的特性从而有效预测视频帧各像素属于划痕噪斑的概率值，通过设置不同的划痕噪斑概率阈值，从而可调节地检测出划痕区域。

可配置的自适应 AI 去噪算法：该模型引入全局与局部图像信号进行建模，利用图像空间相关性重构图像信号并进行噪声信号分离达到去噪目的。同时，该模型将噪声水平作为先验条件进行输入，可以根据输入噪声的强弱自适应进行去噪，因此，其具有较强可配置能力。

全景声音质：全景声是基于声道、对象、场景的三维声技术，提供采集制作系统、高效率音频编码、灵活的播放渲染方案，支持多声道变换、音效增强等。支持全景声 wanos、杜比 Atmos、DTS5.1 等格式，带来沉浸感的声音体验。

关键技术国产化：系统结合百度自研的飞桨（PaddlePaddle）人工智能推理框架， 配合百度自研的昆仑 AI 芯片，采用国产 HDR vivid 标准，将相关 AI 算法运行在有自主知识产权的软件及硬件设备上，推动相关国产化技术框架、标准的应用和发展，有利于实现国产化技术标准与视频修复行业生态的良性互动。

A.1.3. 应用效果

（1）广播电视领域

用于历史老旧电影电视、纪录片等影像的修复和高清化转换，修复效率相比人工提升 20 倍+。

将低清转换成高清甚至 4K/8K，支持 HDR 和全景声，适应大屏播放，快速弥补超高清内容不足。

节目内容播出/发布前的花屏、黑边、抖动、静音等音画质量问题智能检测，保证播出内容质量安全。

（2）文教娱乐领域

BD265 编码器+CAE 内容自适应算法，帮助客户提升视频画质，降低视频大小，从而降低带宽和存储成本 30%+。

可将长短视频进行 4K/HDR/全景声的转换，大幅提升播放体验，呈现更丰富的细节，帮助提升付费会员转化，或者提升购买转化（例如：电商类短视频）。

（3）安防监控领域

将摄像头等终端设备的图片流/视频流进行增强超分等处理，提升画质清晰度，从而提升观看体验，以及后续 AI 感知分析的准确性。

通过画面遮挡、场景变换、过亮过暗等质量检测能力，帮助实时检测摄像头等终端设备是否正常运行，以及图片/视频流是否稳定正常传输。

A.2 昇腾视频增强，助力 8K/4K 超高清内容供给

A.2.1. 厂商介绍

华为昇腾视频增强平台由昇腾平台和视频增强行业解决方案构成。昇腾视频增强解决方案依托于昇腾硬件的高性能算力和 CANN 异构计算架构，致力于为广电媒资、电影电视、影像库、xUGC 内容创作者等超高清行业伙伴提供一种人工智能的手段，既可以使老的经典内容焕发青春，也可以对新拍的内容进行之前增强，从而加持 4K/8K 超高清内容制作和生产，极大缓解 8K/4K 超高清内容不足的痛点。

本方案已在广电媒资、互联网视频、超高清内容制作、安防监控场景等多个行业领域内形成落地。昇腾伙伴可使用多地昇腾计算中心进行线上视频增强，也可使用昇腾线下服务器，算法和应用层开放，希望能聚集产业伙伴，烘托超高清产业加速前进。

A.2.2. 方案介绍

图 12 具有自主知识产权的 AI 视频修复增强训练推理平台

（1）主要功能

本方案提出端到端全国产化 4K/8K 视频修复增强解决方案，涵盖噪声修复、斑块划痕修复、人脸增强、空间超分、时间插帧和 HDR 色彩还原完整流程。

噪声修复：昇腾提出一种自适应噪声提取和迁移学习方法。提取不同强度的常见噪声以及模拟的电影胶片上的污点、划痕和斑块，训练出修复常见噪声的基础模型。对于带有不常见的噪声视频，采用自适应噪声提取方法从视频中提取噪声模式，为每个视频建立专有噪声集。基于迁移学习方法在专有噪声集上微调基础模型，从而支持对不同噪声视频的自适应去噪。随着去噪模型学习的噪声视频越多，不断自我进化，可以提升去噪效果和片源适应性。

板块划痕修复：提供斑块划痕检测模型，用于检测画面中可能存在的污点和划痕，并交给填充模型结合前后多帧内容信息进行内容填补。自动化处理和修复视频中 60%~70%的斑块划痕，助力老片修复过程。

空间超分辨率：自适应地对上下文信息进行对比，根据相似度选择最合适的信息辅助当前帧的超分，过滤无关信息，使系统无需额外进行场景转换检测就能处理画面突变的情况。本系统基于场景丰富多样的高质量数据集，使用随机的“高质到低质”视频退化过程来模拟不同视频的劣化，最大程度覆盖视频类型、场景、退化的多样性，使得系统能处理多种不同片源。本系统也支持高位深视频和 HDR 视频进一步提升分辨率。

人脸特写区域增强：本系统集成人脸检测、人脸三维姿态、姿态纠正、人脸像素分割和人脸增强网络等多种行业先进人工智能算法。人脸检测算法从视频中检测人脸位置，人脸关键点检测算法检测眼睛鼻子和嘴巴位置，仿射变换纠正人脸姿态，采用空间域生成对抗网络的人脸增强算法生成细粒度纹理，人脸分割算法选取人脸像素区域并逆变换融合到原画面中，保证增强的人脸与背景图像没有割裂感。

时间插帧：本系统采用光流估计的方案来估计像素点的运动趋势，进而预测物体的运动。光流估计的准确性是算法的核心。考虑了不同尺度上物体运动的情况，也考虑了运动向量的高阶信息，极大地提升了光流的准确性，在大位移和重复纹理的场景下也表现出理想的效果。

色彩还原：在色彩方面，本系统将原视频像素颜色转换到色度均匀的 CIELAB 色彩空间，将颜色从 BT.709 拓展到 BT.2020，还原的同时避免了颜色的跳变。

（2）核心优势

图 13 CANN 通过多层次算法优化，高效推动 AI 视频增强能力落地

自动化噪声提取和算法演进：提出了一种自动化的噪声提取方法，能够从时间和空间两个维度在片源中合适的部分提取出噪声，随后可对去噪模型进行调整，让去噪模型在不需要人工干预的情况下自我学习演进。

低质视频退化模拟方法：通过各项同性高斯核和各向异性高斯核模糊真实视频，随机选用双线性、双三次插值下采样到随机大小，并且随机加入不同强度的高斯噪声和压缩噪声。通过对多种退化方法的随机采样

组合，可以得到比双三次插值退化更加接近于真实低质量视频的模拟结果，有助于提升视频超分网络的泛化性能，使超分算法适用于多种不同类型和场景的片源。

基于时序一致性的人脸增强方法：基于时序一致性损失函数的生成对抗网络模型。对于单帧人脸图像，对其进行仿射变换来模拟视频中人脸姿态变换，人脸姿态变换前后的增强结果间的范数即为两帧间的一致性损失。一致性损失在保留 GAN 网络增强能力的同时，降低人脸姿态对增强结果的影响，进而提高人脸增强网络的稳定性。

A.2.3. 应用效果

（1）广播电视

应用于广东博华超高清中心、深广电、总台上海总站、上海交大和上海数字电视国家工程中心、成都/武汉/西安 AICC 和本地广电媒资客户，批量和自动化离线处理老旧视频影像素材，对素材进行噪声去除、划痕去除，老片重新焕发生机，2021 年建党 100 周年献礼节目《伟大征程》平台支撑和技术赋能。

（2）文教娱乐

应用于华为视频、星视达等伙伴，将多种不同规格的 8 比特 SDR、高位深 SDR、HDR 的 4K 视频进一步离线处理到 8K 分辨率，广告和宣传片生产域，替代成本高昂的直接 4K/8K 视频渲染过程，对渲染后的低分辨率（1080P）视频进行离线增强，处理到 4K/8K，减少制作域内容生产成本。

（3）安防监控

应用于武汉人工智能中心，高速道路监控场景，在线提升监控视频清晰度、质量和观感，进而提升工作人员进行视频分析的准确性。

A.3 中国移动 AIoTel，为视频物联网注智赋能

A.3.1. 厂商介绍

中国移动智慧家庭运营中心于 2014 年 3 月成立，以“深耕数智家庭创新能力，让网络更智能，让生活更多彩” 为愿景目标，秉承“技术为根、人才为本、创新为魂”的发展理念，不断满足人民群众对美好数字生活的向往。

公司以和家亲 IoT、移动看家、和家智话等业务为牵引，引领家庭物联网的发展，已累计服务和家亲用户超 3 亿，和家智话用户 2000 万，移动看家用户 9000 万，平安乡村服务覆盖全国 30 万个行政村。面向产业链赋能 300 余家生态合作伙伴，覆盖智能家居行业 Top30 品牌、1500 余款智能终端，平台接入 5000 多万台终端，同时携手 10 余家国产化多媒体芯片合作伙伴，发布了 20 余款视频物联网芯片，实现了能力、业务、产业链上下游高效协同。

中国移动将协同终端产业链合作伙伴，共同构建“超清化、智能化、多态化、泛在化”视频物联网能力体系，持续推进视频物联网产业规模化升级。

A.3.2. 方案介绍

AIoTel（AIoT+Telphony，多媒体物联通信能力）是中国移动围绕“AI+IoT+Video+Comm”的设计理念，面向视频物联网自主研发的新型基础设施，牵头制定 ITU、CCSA 等国际标准、行业标准 10 余项。聚焦包括智能家居在内的物联网领域，构建承载泛家庭视频业务的高清视频交换网络，挖掘智能物联网多媒体通信的增量市场，提供了泛终端、泛网络、全场景、电信级的多媒体物联通信服务，填补了软件定义电信级物联网多媒体通信解决方案的行业空白，创新地支持电信级的视频通话、视频对讲、视频监控、智能喇叭、视频会议、家居控制、智能识别等功能。

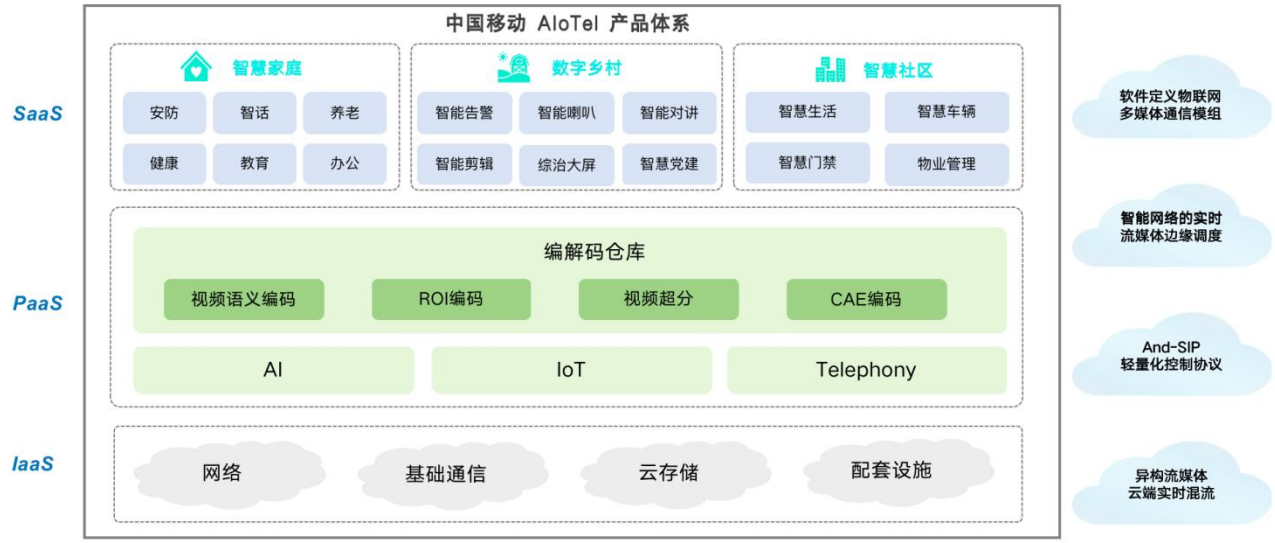

图 14 AIoTel 解决方案总视图

（1）主要功能

分级超分：物联网视频面向消费者传递信息的场景和媒介的界限也在被不断突破和延展，终端的形态更加丰富。智能电视、智能冰箱、智能音箱、早教机、智能手环等大、中、小、微屏形态各异。屏幕的多样性需要视频面向各个分辨率的视频，都能够匹配，并以高质量的画面呈现；而设备的多样性，又需要在做超分的时候，适配不同设备的算力。为解决这类问题，中国移动采用了分级超分的方案。在模型训练时，根据特定场景建立场景化数据集，进行迁移学习；训练好的网络，利用重参数化的思想对模型进行轻量化操作，使得其算法复杂度满足移动端算力；在模型部署时，会根据不同设备的硬件情况选择相应算法。

内容编码：针对应用群体的特征，对用户(老人/儿童)、不同显示要素，为编码到应用各个层面，提供不同的能力。一方面，不同区域或者对象的视觉重要性不同，关注程度不同，各种失真也具有不同的敏感和容忍程度，据此可以分配不同的编码资源，有选择地控制质量，提升视频压缩率。另一方面，可以通过识别和跟踪视频中人物或物体，及时准确地识别和捕捉所关注的事件，自动地进行标记和提醒，用于日常生活辅助、家居看护等方面。

语义编码：类脑感知的多媒体语义编码借鉴人脑的视觉感知机理，从宏观的角度理解图像、视频的内在含义，并对其进行结构化建模就可以实现 “从像素到语义”的革新。大量研究表明大脑传输带宽仅为 8Mbps，却可瞬间完成复杂的视觉认知，人眼视网膜分布十亿个视感细胞，而人脑可将数据压缩近 18 万倍。语义编码针对特定任务展开，关注目标符号含义的准确性表征，提取与人脑认知相关的语义特征进行编码，解码端利用语义信息及生成对抗网络（GAN）等机器视觉技术进行计算重构，并将用户主观体验实时反馈至网络，对感知特征进行动态的更新、优化和自我演进，从而较大提升用户对视频的直观感知。

（2）核心优势

场景化模型构建：在 AI 视频应用领域，如视频超分/ROI 编码/视频语义编码等，模型的优劣与训练数据集质量息息相关。为此，中国移动对如实时通信/安防监控/音视频会议/人脸面部等应用场景进行分类，采集实际的应用数据，构建分类数据集，进行模型分类训练，提升模型的场景应用效果。

智适应编码能力：在算法上，结合传统编码压缩技术与 AI 技术，提升编码效率，降低编码复杂度；积极探索语义视频编码，突破传统编码框架束缚，将视频模型的压缩性能提升一个量级。在应用上，针对不同终端、不同用户特征、不同网络环境等场景，提供对分辨率、帧率、抗丢包等的自适应编解码，实现超高清、高清、实时交互体验。

语义通信系统：基于语义编码的语义通信系统应用于人脸视频对话和场景化监控视频传输与存储，在技术层面上，改变像素级的多媒体通信传统范式，为移动通信发展提供新途径，在产业发展方面，可将多媒体

通信编码码率降低 1 个数量级以上，对比商用 x265，可降低 80%码率，较大提升安防监控、5G 多媒体通信、视频点直播、VR/AR、元宇宙等产品的市场竞争力。

A.3.3. 应用效果

（1）安防监控领域

解决方案应用于移动看家业务，通过提供场景化家庭安防产品包来保障家庭安全，以便捷方式满足用户的安防需求，应用效果如下：

a）**品类拓展**：从自有到生态，引入 15 大品类、300 家品牌，TOP30 全覆盖。

b）**场景延展**：打造 8 大场景，看家、看门、看院、看娃、看店、看车、看鱼塘果园、看鸡场猪圈等，覆盖从家庭到泛家庭的安全防护服务。

（2）实时通信领域

解决方案应用于和家智话业务，用于智能音箱、手机 VoLTE、机顶盒等交互通信，应用效果如下：

a）**终端赋能**：为家庭智能设备如智能音箱、手机应用、机顶盒、一体机等提供音视频实时通信能力，打通设备通信壁垒，赋能 1000 余款终端，激活设备 4500 余万。

b）**应用生态**：基于实时音视频应用，扩展了智能对讲、远程看家、乡村喇叭、门铃对讲、智能提醒等多个应用类别，以家庭设备为基础，建立了完善的通信生态圈，累计订购用户超 4000 万。

A.4 当虹超高清制播，纵享极致视听

A.4.1. 厂商介绍

当虹科技是一家定位于大视频领域，主要面向传媒文化、泛安全、智能网联汽车等方向，提供智能视频解决方案与视频云服务的高科技企业。

在传媒文化方向，当虹科技作为中国数字音视频编解码技术标准工作组会员，参与每一代视频编码标准的制定，是国内 4K/8K 超高清实时编码器的核心供应商。公司深度支持中央广播电视总台 CCTV-8K 超高清频道、奥林匹克频道（CCTV-16）开播，8K 编解码产品在 2022 年央视春晚、北京冬奥会、陕西全运会等重大活动中成功落地应用。同时，公司与咪咕、腾讯视频等平台合作，共同构建 5G+8K 直播、4K IP 化平台等。此外，还参与花果山、马栏山等超高清视频产业园标杆项目。

在泛安全方向，当虹科技是国内极少数同时具备“视频编码”与“智能 AI 识别” 双引擎技术的企业，可在视频质量基本不变前提下，最大减少 90%的传输成本和存储成本。以边缘智能计算、视频联网、AI 解析、大数据分析为基础，致力于海量监控视频的“视频+AI+大数据”的场景化深度应用，有效助力金融、能源、交通、应急、教育、公安、政府等泛安全行业客户。

未来，当虹科技将在大视频领域继续深耕，努力通过智能视频技术让世界更清晰、更安全、更美好。

A.4.2. 方案介绍

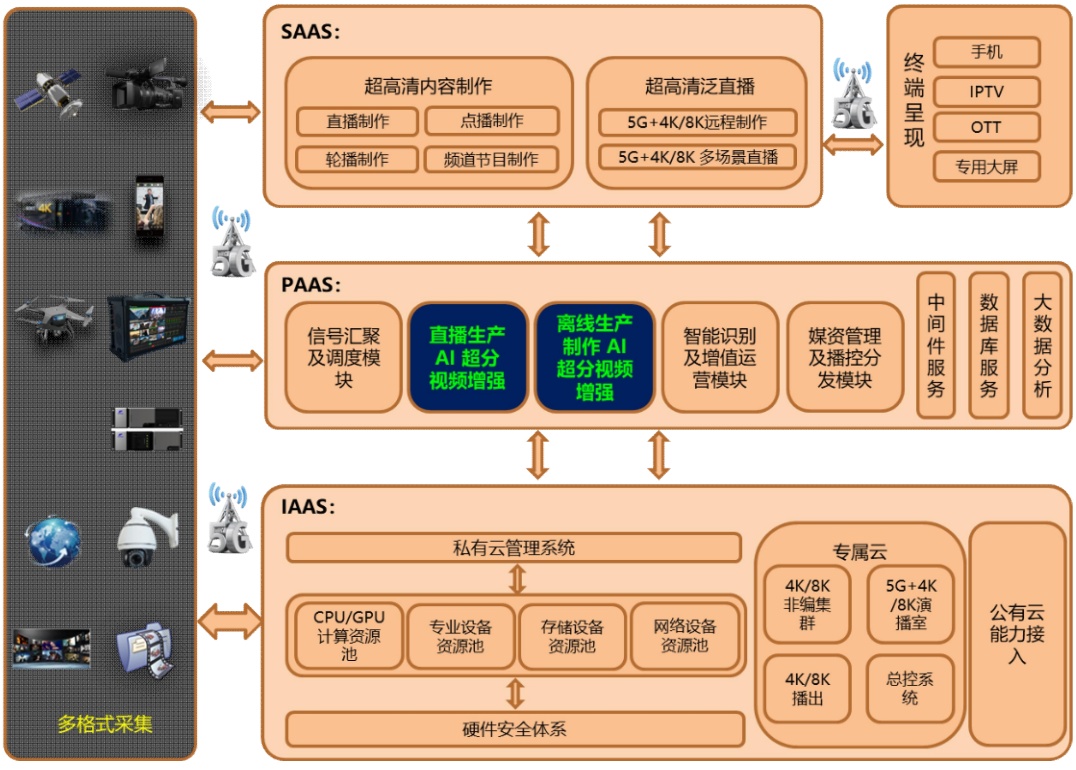

图 15 当虹超高清视频智能处理解决方案

（1）主要功能

视频超分。基于 AI 深度学习模型，通过海量高分辨率视频以及图像素材的机器学习，训练高频细节来获得图像高精度复原，同时集成了针对文字、人脸、物件的多种去噪优化算法，在分辨率提升的同时，智能消除 ROI 区域图像的瑕疵，为用户超高清内容生产以及超高清专区建设提供高画质，高可靠的文件转码解决方案。

视频插帧。支持整倍数帧率转换(25p 到 50p)，非整数倍帧率转换(24p 到 30p/50p)，以及帧率倍频(50i 到 50p)等多场景帧率转换技术，该技术基于制作域的深度学习框架，通过海量运动、平移素材训练，准确估算运动轨迹，并选择最佳图像效果融入视频，在满足任意帧率的转换的同时，使用户获得更清晰，更流畅画质体验。

画质增强。经过长期对视频图像的分析与计算，当虹科技研究并提供十多项画质增强算法技术，致力于解决图像经过多次压缩导致的画质下降或者非专业拍摄导致图像画质不佳等问题，通过 VMAF 主观评价的智能学习框架进行每一项算法的研究与优化，其中包括对图像的综合色调调节、动态亮度调整、边沿锐化、帧间/帧内去噪、防抖动、抗锯齿等，为用户全面的画质提升需求提供多样化的选择。

视频修复。对于历史沉淀的经典老片，大多数布满划痕和灰雾，整体的色调也显得格外的陈旧，但无论从美学价值还是历史传承，老片修复是具有深远的意义的。当虹科技支持老旧视频的修复能力，可针对老片的划痕进行消除、色彩偏差进行纠正、拍摄环境造成雾化以及模糊进行去除，多方面原因造成的块效应、噪声进行处理等，使得修复后的老片焕然一新，还原真正应有的美学价值。

HDR 转换。支持 SDR 与 HLG、PQ 以及色域 BT.709 与 BT.2020 的相互转换，助力台内、新媒体以及运营商等用户 SDR/HDR 的内容同播制作以及真 4K HDR 内容生产，同时开放二十余项基于制作域的参数供用户进行自定义配置，一方面能够确保上下变换后色彩保持一致，另一方面充分还原导演的创作意图。

（2）核心优势

极致视听呈现。在视频上实现“5G+4K/8K+AI”，支持 4K/8K 60fps 高帧率视频转码，10bit 高位深、4:2:2 色度采样、BT.2021 宽色域以及 HLG/PQ/SDR 等多种亮度曲线，并完美适配 AVS3 标准与 HDR Vivid 标准，还原真实的人眼场景。在音频上除支持杜比音频外，还对 Audio Vivid 完成适配，打造身临其境的音频环境。整体可实现“8K 超高清+三维声”，带来良好的视听效果。

多格式兼容。支持多种主流摄像机格式如 XAVC、MPEG2、XDCAM、AVC-I、DVCPRO、ProRes、DNxHD 等，支持多种台内制播格式如 ProRes、DNxHD/DNxHR、XAVC、MPEG2、DV 等，支持多种互联网及新媒体格式如 H264、H.265、AVS2、AVS3 等。适用于多种业务场景。

AI 编码压缩。具备自研的编码压缩算法。帧间划分场景片段，简单场景与复杂场景重新决策编码参数。帧内划分平滑区域与台标、人物、文字等区域，将更多码率分配到人眼敏感区域上，提升画质。保证画质的同时降低码率，平均可节省约 30%带宽并达到同等质量效果，为用户降低运营成本。

软硬件设备国产化。当虹科技“磐为”系列产品融合华为“泰山”系列鲲鹏高性能处理器，实现编码器从硬件平台、CPU、操作系统、数据库到应用软件全面国产化。

A.4.3. 应用效果

（1）广播电视领域

解决方案应用于 IPTV、OTT、城市大屏，主要应用效果如下：

a）自有 AI 编码压缩算法可节约 15%~30%带宽， 为用户降低运营成本。

b）AI 超分视频增强技术可扩大超高清内容生产，为用户提供高质量高性能的转码能力。

c）覆盖总台 8K 超高清频道、奥林匹克频道，各省级 IPTV、各大卫视，以及央视频、芒果 TV 等新媒体平台，为千万家庭共享视听盛宴保驾护航。

（2）文教娱乐领域

解决方案应用于咪咕视频，高校影音资料智能修护实验室等，主要应用效果如下：

a） 基于超高清视频、AI 画质增强等技术积累，为各大体育赛事直播提供超高清视频技术支持。并在传输分发等关键环节护航，为用户在手机上就能享受不一样的超高清、互动观赛体验。

b）应用于传媒院校影音资料数字化采集、存储与网络化传输，通过智能修复对历史资料焕新，并赋能教研实践，为特色优势学科建设、人才培养、视听产业融合发展、国家主旋律电视文艺创作注入内生动力。

A.5 数码视讯，全面引领超高清大视听数字产业

A.5.1. 厂商介绍

数码视讯科技集团由清华科技园及公司核心团队于 2000 年共同发起成立，2010 年在深交所挂牌上市。公司致力于视频技术服务、加密技术、5G 技术、AI 技术等领域研发，为全球 110 多个国家和地区的运营商、企业、政府、金融等客户提供优质、精准的服务，让全球用户畅享更智能、更安全、更美好的数字生活。

数码视讯深度参与国产视频编码（AVS 标准）、ChinaDRM、DCAS、TVOS、C-DOCSIS、应急广播、广电 5G 广播标准制定及应用探索，在理论与实践层面不断推进视频技术与解决方案的国产化落地进程。

集团现已成立北京、深圳、武汉、西安等多个研发基地，在成立 20 年之际荣获“国家科学技术进步”二等奖。随着规模和业务范围的拓展，现已服务 20 余家国家级客户、34 家省级客户、200 余家市级客户。

在智能视频处理领域，数码视讯依托自有算法不断寻求视频画面效果与运算资源之间的平衡，尤其是对于 4K/8K 超高清视频，具备完善的超清产品智能处理产品体系。作为唯一 100%参与国内全部 4K 上星及 8K 开播频道播出编码系统的支持方，数码视讯拥有丰富的超高清视频处理经验，可以为用户提供不同场景下清晰、美丽、流畅的视觉体验。

A.5.2. 方案介绍

（1）主要功能

数码视讯具有完善的编转码产品体系及技术体系，可以根据用户不同需求提供不同形态，不同技术标准的产品方案，尤其是超高清编码产品体系已经全面应用智能视频处理技术，实现画面质量的显著提升。

10K118 超高清编码器系列

数码视讯 10K118 超高清编码系列支持 4K/8K 超高清 AVS2/AVS3/HEVC 节目编码，作为 AVS 工作组最早一批会员单位，数码视讯深度参与了历代 AVS 标准的制定，也是国内最早实现 AVS 标准产业化的企业之一。因此在 10K118 系列中使用了自有 AVS2/AVS3 算法库，完整保留了 AVS 系列的编码特性，充分展现了 AVS 系列算法出色的压缩能力，产品获得了国家科学技术进步奖二等奖。

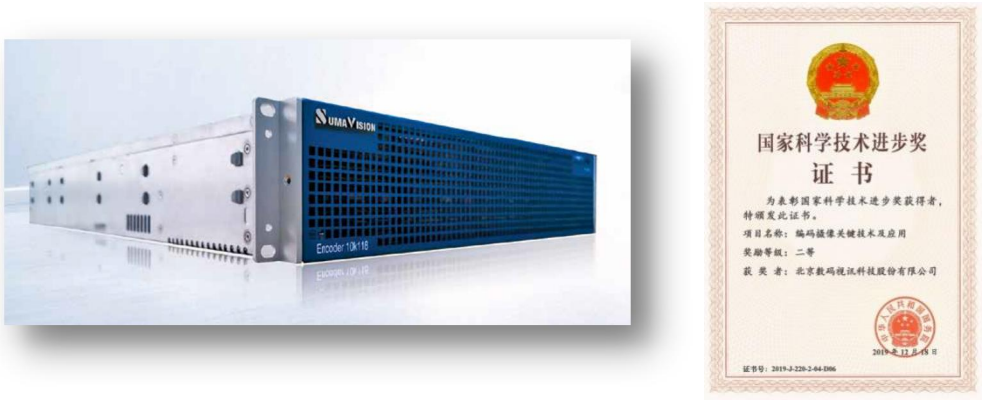

图 16 8K AVS3 编码器及所获奖项

在智能视频处理方面，数码视讯支持通过自有图像感知编码方案来提高超高清视频的感知视觉质量，可以在不增加图像带宽情况下，显著提升人脸等重要区域质量。

图 17 图像感知技术效果对比图

（2）核心优势

支持中国标准。数码视讯 10K118 超高清编码器系列全面支持中国标准，不但在编码算法上采用了自研 AVS2/AVS3 算法库，还支持中国 HDR 标准-HDR Vivid 以及中国三维声标准-Audio Vivid，实现了中国超高清标准的全面落地。

智能图像处理。数码视讯 10K118 超高清编码器系列支持完善的智能图像处理技术，可以分析和识别视频场景和区域，并根据人眼感知特性，对不同区域进行码率加权，可重点加权人脸、台标字幕、平缓背景等人眼敏感的区域，结合传统的码率控制模式，建立各部分复杂度分布图、确定适合的帧类型、宏块运动的重复参考迭代关系，使码率控制的结果更为合理，从而提升视频编码的总体质量和细节质量。

A.5.3. 应用效果

基于数码视讯 10K118 超高清编码器系列先进的技术以及优异的产品质量，在市场上获得了用户的一致好评，先后用于央视总台 CCTV-8K 频道、CCTV-4K 超高清频道、奥林匹克 4K 频道，北京台冬奥纪实 8K 频道、冬奥纪实 4K 频道以及上海台欢笑剧场 4K 频道播出编码系统的建设，为中国 4K/8K 超高清内容生产贡献了自己的一份力量。

央视总台CCTV-8K频道

CCTV 4K超高清频道

CCTV16 奥林匹克4K频道

北京台冬奥纪实8K频道

北京台 冬奥纪实4K频道

上海台 欢笑剧场 4K频道

图 18 参与 4k/8k 频道建设

A.6 涌现科技硬件视频编码，大视频时代智能加速引擎

A.6.1. 厂商介绍

涌现科技（EMERGETECH）是一家聚焦基于人工智能的人眼视觉与机器视觉融合编解码技术的国家高新技术企业，将人工智能与视频编解码算法深度融合，通过平衡算法和芯片设计优化，为行业提供高性能、高密度、低功耗、低成本的专用芯片、模组、软件工具及针对特定场景的全栈应用解决方案。公司在智能视频编码研究及基于异构计算的高性能多媒体处理器设计方面产出多项创新成果，累计授权及在申请自主知识产权超百项，打造了以“芯片—视频加速卡—视频加速服务器”为核心的产品矩阵，全面打通“算法-芯片-应用场景”。

A.6.2. 方案介绍

（1）主要功能

涌现科技基于自研 ASIC 的高性能视频编解码芯片推出了一系列高通量高性能的 Seirios 视频加速卡，相比目前通用的软编方案，Seirios 视频加速卡可完美解决软编编解码速度慢、计算资源消耗大、难以支持多任务处理、以及功耗高能耗大和不利于减碳节能等问题，硬件编码综合性能国际领先，并且可提供全套国产化方案，兼顾高性能及高性价比。

目前涌现 Seirios 系列产品已广泛应用于互联网视频平台、云桌面/云手机/云游戏、广电超高清、智慧城市等领域的数据中心和边缘计算节点，能够帮助客户快速、有效处理视频流，提升效率的同时，降低硬件、服务器成本投入，释放通用算力，降低设备功耗，节省运营成本。

（2）核心优势

①高密度、高并发：支持 H.264/H.265/VP9 多种格式编解码，Seirios P4 单卡支持 8 路 4K@60fps 或 64 路 1080P@30fps 编解码；支持一机多卡，提升编码密度。

②超强兼容：Seirios 兼容主流 x86 和 ARM 架构，原生支持 FFmpeg 和 GStreamer 框架，采用可插拔模块化设计，提供 M.2、U.2、PCIe 等标准接口，使用灵活、便于集成，实现软件到硬件方案的无缝升级。

③智能视频云转码：多种格式相互转码、一进多出转码，最高输出一路原始分辨率+ 三路小分辨率码流 ，全面匹配不同设备，满足直播、点播、视频监控等不同场景需求。

④AI 辅助视频加速：Seirios 支持动态帧级设置 ROI、动态分配编解码资源、自适应帧率变化编码等超高清视频智能及编解码算法，提升视频的整体质量，优化观看体验。

A.6.3. 应用效果

Seirios 高效视频转码云桌面解决方案：将 Seirios 安装在后端物理机服务器上，充分发挥其高压缩比、高并发的特性，进行画面云端渲染及高速转码，在保障画质的同时释放服务器算力。

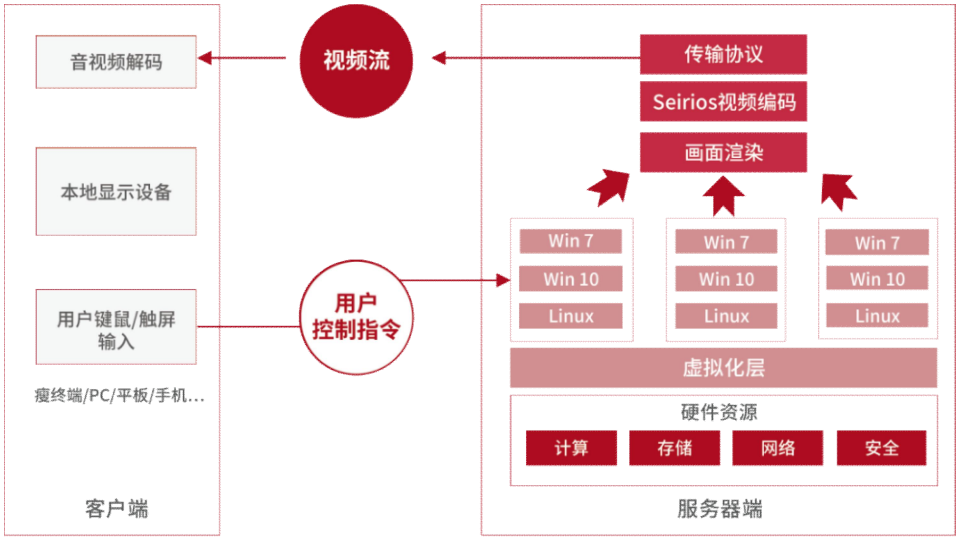

图 19 Seirios 高效视频转码云桌面解决方案

①高密度编码，降低 CPU 负载：在同等画质基线下，Seirios P4 单卡可同时处理 64 路 1080P@30fps 的虚拟桌面编码。实际案例中，相较于传统编码方案，Seirios 对 CPU 的占用率平均降低 36%以上，能够有效提升服务器荷载云终端数量，提供更高品质的画面呈现。

②快速编码，降低带宽：在云桌面场景下文字办公、网页浏览、视频播放等超过 26 个测试用例中，使用 Seirios 进行视频编码，在固定画质的基准前提下，相较于传统 X264 编码方案，最大码流节省 70%，平均有效降低带宽 40%以上，单帧编码耗时降低约 30%。

③高度适配性、高成熟度和高稳定性：目前 Seirios 已经完成与中科曙光、宝德、飞腾、麒麟软件、统信等 40 多家芯片、操作系统、整机厂商的适配认证工作。

A.7 博华超高清创新中心，打造 AI 超高清内容生成平台

A.7.1. 厂商介绍

国家超高清视频创新中心（深圳）于 2022 年 10 月由工业和信息化部批准成为国家超高清视频创新中心共建单位，其依托主体为广东博华超高清创新中心有限公司。公司成立于 2019 年 5 月，注册资金 1 亿元，目前股东有智能视听研究院、创维-RGB、创维数字、康佳集团、南方新媒体、当虹科技、视源股份、新视创伟、索贝数码和博冠股份十家单位，涵盖了采集、制作、传输、显示等产业链关键环节的龙头企业和科研单位。

按照国创中心的主要目标和任务方向，中心建设了“四个平台”和“七个实验室”，包括共性技术研发平台（超高清视频前端采集及内容制作实验室、编解码技术关键技术实验室、超高清+AI 创新应用技术实验室）、测试验证平台（超高清视频测试认证实验室、5G+4K/8K 传输验证平台）、中试孵化平台、行业支撑服务平台（超高清全媒体开放实验室、超高清内容制作公共服务平台）。

国创中心致力于在超高清视频领域开展共性关键技术研发和成果转移扩散，重点围绕视音频编解码、超高速接口、超高清视频测试验证、超高清+AI 创新应用等方向，开展技术研发和平台建设，推进关键共性技术的产业化，构建超高清视频产业生态圈。

图 20 智能超分系统完成商用落地，斩获中国软件行业协会大奖

A.7.2. 方案介绍

（1）主要功能

博华智能超高清超分内容生成平台主要包括如下流程：视频上传、AI 去噪、视频插帧、超分辨率、HDR 转换、编码封装和转换后视频回传。

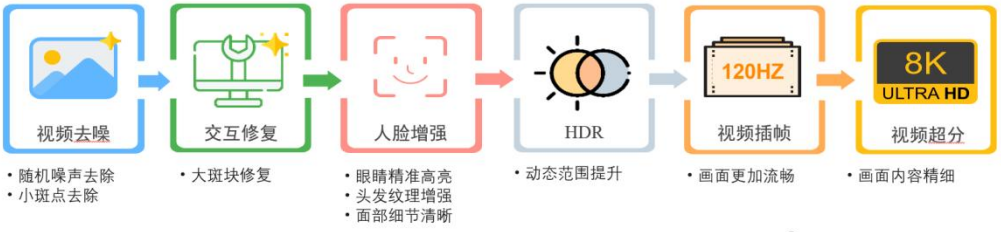

图 21 博华智能超高清超分内容生成平流程

博华智能超分采用昇腾 Atlas 计算平台，MindSpore 神经网络框架，让 Ascend NPU 加速神经网络计算，以此满足超分算法对超大规模算力的需求。系统采用前后端分离，支持云上和端侧部署，用户可灵活选择。系统可满足高效率、低成本、高可靠超高清内容生产要求。

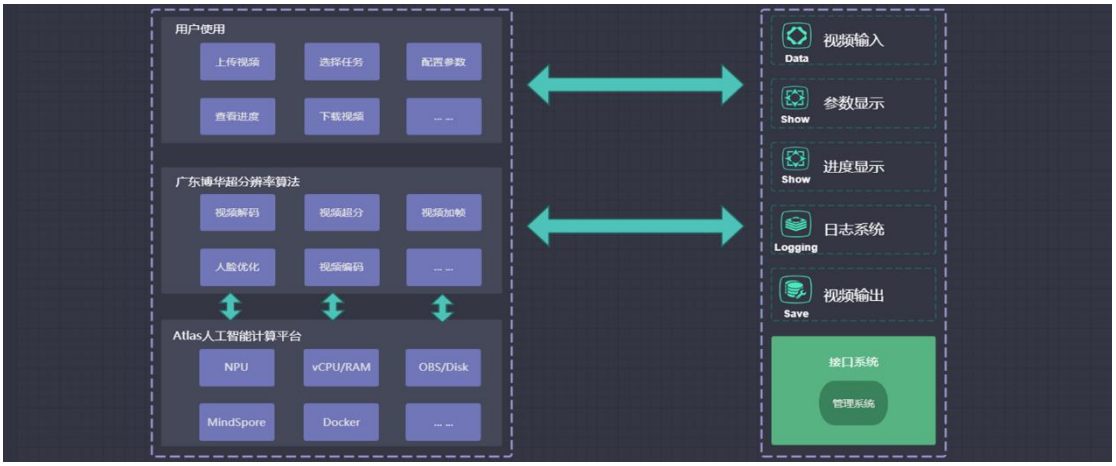

图 22 平台系统框架图

（2）核心优势

智能超分系统从计算平台、算法模型和评价系统等方面做了深度优化，满足用户生产系统要求。

1)轻量级超分模型。视频超分算法模型具有量级轻、性能高特点，并基于昇腾计算平台做了深度优化，在单机昇腾计算平台可以达到 6 秒/帧的解码、图像增强、超分、倍帧和编码。在一定算力配合下，支持实时在线超分。

2)场景变换算法。系统支持自动和手动处理多场景模式，提高了系统场景支持的泛化能力。

3)帧间信息算法提升。视频超分的一个重要特征是利用帧间信息，算法在三维卷积和非局部模运算量大，光流估计的精度方面有效利用帧间信息提高了超分算法性能。

4)视频质量评价方法。系统支持客观和主管相结合更符合人类感知的视频评价方法。

5)系统支持线上和线下部署，符合广播电视台和内容运营商生产系统要求。

A.7.3. 应用效果

智能超分平台已部署深圳昇腾创新中心算力平台，对外可提供视频超分服务。智能超分平台利用场景自适应的 AI 超高清技术攻克 4K/8K 产业发展难题、丰富 4K/8K 内容供给, 通过利用“超高清+AI”技术降低

整个产业在超高清内容制作的成本，通过内容为牵引打造丰富的 4K/8K 产业集聚发展、补齐 4K/8K 产业发展短板，形成高水平有特色的“超高清+AI”创新制作平台；在超高清视频制作技术研究领域，基于目前的“超高清+AI”成果，扩展 AI 超高清内容方面的成果输出，如图像智能去噪和增强，图像智能填补，图像智能上色，内容智能分析等，为南方新媒体股份有限公司，中广电传媒有限公司等主要客户提供优质的 4K/8K 超高清内容。

通过设立“超高清+AI”应用实验室，拓展产业应用，通 4K/8K 超高清视频产业与 AI 人工智能相互促进的正向循环，一方面，通过 AI 技术创新，在超分辨率和画质修复，个性化推荐内容制作等 4K/8K 超高清视频领域上发挥独特的作用。另一方面，利用丰富的数据资源和鹏城云脑算力，训练高效的超高清智能网络模型，加快更多的 4K/8K 超高清视频在智能交通、智慧医疗等应用场景落地，促进了 4K/8K 超高清视频的发展，开阔市场的空间，从而，随着 4K/8K 超高清内容增多，市场对 4K/8K 的设备需求增长，更多的设备进入市场必然会产出高质量的 4K/8K 数据，进一步促进 AI 人工智能的发展，打造一个正向的产业循环。

附录 B: 超高清视频智能的行业应用案例

B.1 广播电视：电影频道智感超清修复

B.1.1. 案例内容

（1）用户痛点

电影频道作为唯一的国家级电影专业传媒机构，拥有海量的电影片库，但是现存档的电影版本主要为高清甚至标清版本，不能适应超高清技术的持续发展。市场上对电影电视媒体提出更清晰、更智能、更融合的要求，观众对超高清视频内容需求也是呈急速上升的发展趋势，但是目前存在超高清视频产能不足的问题。

- 电影频道之前的修复手段主要依赖于人工手动方式，据统计：一部 90 分钟的影片约为 13 万帧，逐帧修复，平均每部电影，需要 25 人/日（将近一个月），最难的画面 1 人 1 天只能修复 20 秒，因此电影修复工作需耗费大量的人工成本和时间成本。
- 修复使用的设备均为进口产品，从商务采购到设备到货安装部署完成，已是比较落后的版本了；问题解决、服务升级等均不到位，没有更好的支撑。此外，国产化是行业发展趋势，进口产品无法满足此要求，无法做到自主可控。

（2）解决方案

百度智能云交付了智感超清一体机，该一体机里内置了多个老片修复和画质提升的模型，包括：去噪、去划痕、去抖动、智能上色、色彩增强、细节增强、超分、SDRtoHDR 等，其中同一类算子，根据不同的影片类型，进一步细分了场景模型，例如：超分又细分为 速度优先模型、质量优先模型等。同时每类算子，均开放了强度调节参数，修复人员可以根据不同片子情况配置最适合的模版。

整体业务流程如下图所示：首先通过扫描仪将胶片转成数字化格式，然后通过智感超清一体机进行自动化处理一遍，该环节可完成去噪声、去划痕、增强超分处理，然后人工进行精修，包括：部分 AI 没有去除的斑点、噪点、划痕等，并进行调色处理，最终达到可播出的质量要求。

本案例中，电影频道与百度智能云达成了深度合作，电影频道将人工修复的近百万帧数据输出给百度智能云，智感超清通过数据输入进行模型训练，并反馈给电影频道优化效果，人工二次评测提出问题，模型再次训练调优，如此反复，模型效果得到了较大的提升，并具备了更好的泛化能力，能适应不同类型的影片。

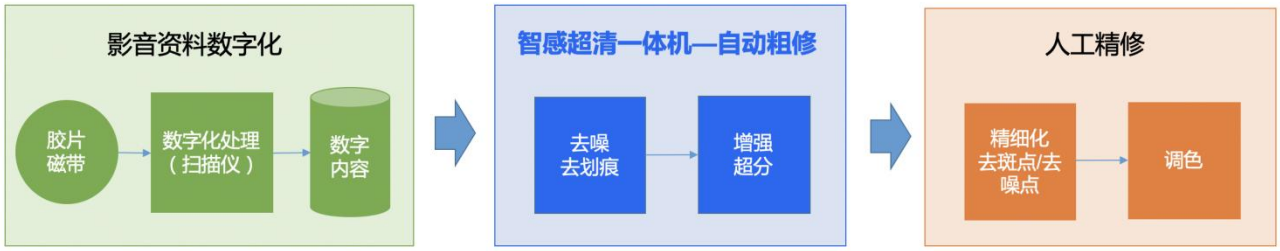

图 23 电影频道老片修复流程

B.1.2. 案例成效

通过智感超清一体机的应用，大大节约了时间成本和人力成本，提高了修复生产效率。据统计：人工 8 小时平均修复 3000 到 5000 帧。AI 智能修复，24 小时不间断工作，每天修复 28.5 万帧。修复效率是成几何倍增长。通过大量不同类型电影的数据训练，项目成果将更适用于电影节目的超分处理，将为电影频道 4K 超高清内容建设，4K 电影频道开播，4K 电影网络播映等多种业务需求提供技术支撑，更好地利用高新技术引领新时代电影高质量发展。以 AI 人工智能为核心技术的智感超清修复软硬件一体机，也将填补国内修复设备的空白，推进修复设备国产化的进程。

老电影经过 AI 智能修复，将继续得以保护与传承，让更多观众享受优质老电影的魅力，并继续在年轻观众中广泛传播，让珍贵的老电影重新焕发产业生机，社会效益巨大。同时也将彻底改变视频修复行业的工作模式，通过大量的修复数据训练，提升检测修复脏点、划痕等视频问题的准确率，将修复工作者从简单重复的劳动中解脱出来。

如下是修复前后的效果视频的图片截选：

图 24 修复前后的对比图

B.1.3. 创新与亮点

本案例创新性地提出基于 AI 的视频修复解决方案。通过人工神经网络，对视频帧空间及时间维度内容相关性进行建模，并结合 AI 领域领先的注意力机制，全局与局部信号建模与生成对抗式网络等技术，研发了划痕躁斑分割、区域填补、自适应去噪、基于参考帧上色及多维度分辨率提升等 AI 算法模型。

具体技术创新有以下几点：

- 无监督可交互的去噪神经网络算法:使用带噪声图片的神经网络无监督训练方案，从而摆脱现在有监督神经网络对带噪及无噪成对图片的训练数据依赖，解决真实场景下带噪图像对应的干净图像获取难的问题。
- 级联 U 型超分辨率重建深度卷积神经网络：基于 U 型网络进行画面的超分辨率重建，同时，引入级联提升框架，通过多级级联 U 型网络提升超分变率重建的效果。
- 端到端逆色调映射卷积网络：结合画面的全局与局部图像特征，通过神经网络映射，能同时完成 8bit 至 10bit 或更高位深转换以及 BT.709 到 BT.2020 颜色空间映射算法，将业内已有算法模型速度提升 6 倍，效果持平。
- 采用异构计算 4K 编码技术：基于 GPU+CPU+ASIC 的异构加速计算的架构，支持国产化昆仑的深度学习推理芯片，通过模型量化减少每个权重系数所需的比特数来压缩网络，实现预测的加速。基于 ASIC 编码芯片实现对 4K 编码的硬件加速，实现 4K 视频处理和编码的高效率生产。

B.2 广播电视：上海交大和总台智能影像修复

B.2.1. 案例内容

由上海交通大学联合中央广播电视总台、华为研发的人工智能视频增强技术修复的影像被《伟大征程》及中共上海一大会址纪念馆选中使用，曾在 2020 年中国国际服务贸易交易会展出。2022 年 4 月 6 日，中央电视台科教频道 CCTV10《时尚科技秀》栏目播出节目——《智能影像修复》介绍了该项新技术，它能够让《红楼梦》、《西游记》等脍炙人口的经典影视作品，以及《开国大典》这样珍贵的历史影像重新焕发出生机。

B.2.2. 案例成效

图 25 CCTV10 科教频道 焕然一新，神奇的影像修复技术

历史上有众多的经典影视作品，在今天依然受到大众的喜爱和关注，但是由于拍摄、制作时的设备限制，以及部分胶片、磁带长期保存遭到的损伤，在高清、超高清内容日益普及并逐渐成为主流的今天，这些作品的影像质量已经无法满足观众的需要。以往个别老旧影像的修复主要依靠人工来完成，但是视频的每一秒都至少由 24 帧画面组成，逐帧修复工作量极大，成本高、周期长，也无法满足批量、快速的修复要求，成为行业面临的一个难题。

2020 年 12 月，上海交通大学与中央广播电视总台签署了关于深化落实《超高清视音频制播呈现国家重点实验室》协作的合作协议，依托国重实验室在上海交大共同建设“智能媒体技术研究实验室”。用人工智能技术修复老旧影像，让总台音像资料馆中的大量珍贵历史影像重新投入节目的制作，为建党百年盛大庆典《伟大征程》献礼，是王延峰教授团队在实验室共建后接到的首个重大紧急任务。为了解决这一问题，张娅和张小云两位牵头教授积极调研总台的需求与影像资料情况，基于团队在视频处理和人工智能领域的积累综合研判当前国际主流技术方向，同时联合上海云视、华为等多家上下游企业进行技术攻关，经过持续的算法创新和系统优化，AI 视频增强平台应运而生。

B.2.3. 创新与亮点

AI 视频增强平台汇聚了当前人工智能与视频处理领域的先进算法，融合了面向真实场景的视频超分辨率技术、AI 指导 AI 修复、人机耦合、AI 人脸增强、AI 智能插帧、AI 智能去噪和细节增强等多个维度的技术，

能够实现从空间分辨率、时间帧率、色彩和对比度等多个方向上的视频质量增强，一举解决原来影像存在的划痕、噪点闪烁、细节模糊、运动拖尾等问题，并且使原本低分辨率、隔行扫描的影像达到 3840x2160 的 4K 超高清分辨率，帧率达到 50 帧，支持宽色域和 HDR 动态范围。尤其是针对图像增强任务的不适定性导致的生成画面模糊、细节不够等问题，研究团队提出了高频细节局部自回归采样、基于任务解耦框架的有参超分等创新算法，显著提升了图像分辨率和细节质量；针对视频插帧面临的块效应、运动伪影等挑战，提出了 MEMC（运动估计和运动补偿）模型驱动和场景深度感知的插帧算法，充分发挥了传统插值框架和深度学习两者的优势。

B.3 广播电视：央视总台/北京台 8K 频道智能视频编码案例

B.3.1. 案例内容

数码视讯基于 10K118 超高清编码产品，先后服务于央视总台 CCTV-8K 频道，以及北京台冬奥纪实 8K 频道播出编码系统的建设，为频道提供高质量的 AVS3 播出节目流。

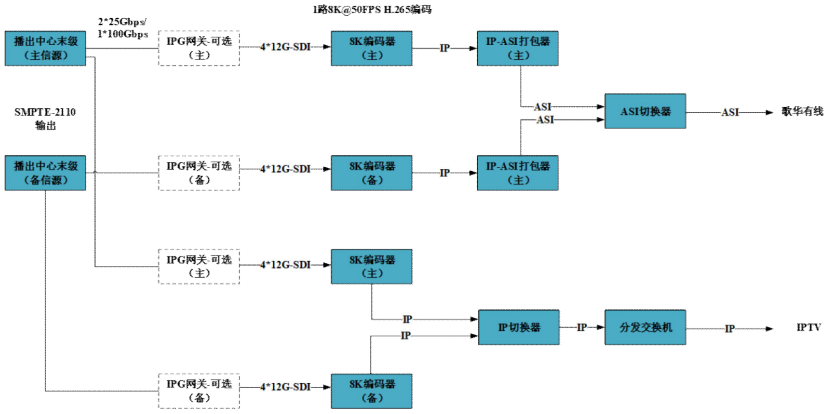

图 26 北京台 8K 频道组网图

如上图所示，以北京台 8K 频道组网为例，数码视讯 10K118 超高清编码产品在系统中不但负责将基带 IP 信号编码为 AVS3 格式压缩信号，还会基于自有智能视频处理方案，对视频内容进行智能处理，进一步提升视频画面质量。

B.3.2. 案例成效

数码视讯 10K118 超高清编码产品应用了一种基于人脸感知的编码方案来提高超高清视频中人脸区域的视觉质量。首先使用专门定制的人脸感知模型来精确和快速地定位人脸区域。然后，基于分层映射算法生成人脸感知图。最后，以人脸感知图为指导，优化编码过程，包括模式决策、块划分和比特分配。此方法在央视总台以及北京台 8K 频道建设中均得到了实现，证明了其有效性，尤其是基于 AVS3 标准下，进一步提升了中国标准的应用效果。通过数码视讯感知编码方法显著提高了人脸区域的客观质量和主观质量，而非人脸区域引入的失真相对不可见。此外，算法还节省了模式决策和块划分所需的计算量，从而进一步降低了计算复杂度。

图 27 数码视讯编码优化对比图

B.3.3. 创新与亮点

数码视讯 10K118 超高清编码产品所应用的智能视频处理方法主要有以下亮点：

a)提出了一种改进的人脸检测模型来快速准确地定位人脸，比现有的常规模型更加准确高效。

b)定位人脸后，生成分层感知图，从而在后续编码过程中合理应用人脸信息。对于不同场景下不同大小的人脸，进行了更加细致的划分，从而提升编码策略优化的合理性。

c)利用感知图对人脸区域赋予更高的重要性，执行优化编码策略，包括模式决策、块划分和 QP 分配。并且在新的人脸辅助模式决策和块划分结构下，图像编码复杂度也有所降低，缓解了超高清场景下的耗时问题。

d)数码视讯方案可以灵活地与其他特定的检测模型相结合，例如文本检测或电视标志检测。而且，此方法不改变视频编码标准的语法，因此可以很好地适用于其他具有类似架构的编码产品。

B.4 文教娱乐：浙江传媒学院历史影音资料智能修护实验室

B.4.1. 案例内容

图 28 当虹老片智能修复流程图

据浙江传媒学院官网显示，该校仅图书馆的馆藏音视频时长就高达 97101 小时。未来，还将对中国电视艺术发展史上形成的历届奖杯、证书、文史、音视频资料等档案和数据进行收集、整理、展示。而当虹科技助力浙江传媒学院建成实验室。通过数字化采集、存储以及网络化传输，可以更好地达到数据可溯化、可量化、可共享，从而延长资料寿命。视音频资料在完成数字化后，再由智能修复提质平台通过人工智能技术对历史资料进行焕新。

B.4.2. 案例成效

图 29 实验室建成效果图（背景显示“AI+人工”修复调色成效）

以 86 版《西游记》为例，按照以往人工修复的方式，一集《西游记》完整的修复工作，所需时间可以长达 2-3 个月。而如果使用 AI 修复的方式，仅需 10-20 个小时就可以完成一集的修复，效率提升上百倍。

另外作为中国广播电视艺术资料研究中心的主要研究平台，实验室主要服务于国内主旋律电视艺术研究，进一步挖掘中国广播电视大奖的深厚资源和品牌价值。通过历史影音的智能修护，对具有标杆示范作用、体现行业较高艺术水准的优秀广播电视和网络视听作品（节目）等相关作品资料，进行收集整理、数字化保存、展览展示、学术研究、教学实践等。

B.4.3. 创新与亮点

a) AI 助力人工修复。项目专家组指出，艺术修复的关键是“修旧如旧”。通过“AI+人工”的高效协同，会让历史影像修复更严谨、更精细，从而真正还原艺术创作的时代性和美学性。

b) 配套可以储存海量资源的媒资服务器。通过视频上载、编目、加工、检索、预览、分发、存储管理等智能一体化的融合生产功能，实现影音资料数字化、移动碎片化、搜索便捷化、编目智能化等多维度全流程管理。

c) 在教研实践中，提升中国电视艺术理论研究和艺术创作能力，加强人工智能技术实训，培养应用型、复合型、创新型人才。为特色优势学科建设、人才培养、视听产业融合发展、国家主旋律电视文艺创作注入内生动力。

B.5 安防监控：移动看家

B.5.1. 案例内容

移动看家是中国移动面向泛家庭场景打造的智能安防及亲情看护服务，利用多媒体通信技术、人工智能技术，满足用户从看家护院、防火防盗、亲情看护的多方位安全诉求，打造泛安全领域智能安防看护基础设施，通过视频技术致力于构建数字生活的基础视联网服务。

在安防监控场景下，平坦静态画面占比超 40%，为满足聚类业务场景下多屏播放的特性，移动看家引入智能编码、云/终端超分等技术，达到视频码率压降、大屏画质提升等效果，实现云存和带宽资源的降本增效。

典型应用场景如下：

（1）家庭看护

家庭看护基于 AIoTel 技术，赋予猫眼门铃、摄像头、智能台灯等家居设备通信能力，家中的老人或小孩可通过一键呼叫功能，呼出 VoLTE 视频电话，实现通信必达，随时随地看护家庭成员。

（2）数字乡村

对接各地政府天网/雪亮、综治平台等，面向农村用户提供联防联治监控解决方案，打造了农村群防群治综合超大规模社会治理平台，有效支撑地方政府开展乡村治理、安全防护工作。

（3）智慧店铺

智慧店铺实现了店铺夜间无人值守、客流分析、积分管理等能力，有效支撑了数十万店铺的日常管理和精准营销工作。

B.5.2. 案例成效

移动看家订购用户已突破 4000 万，注册用户超 9000 万，服务超 4500 万个家庭，自研多中心分布式智能调度架构支撑超 4000 万台设备接入与存储，云存储规模达到 700PB，在消费类安防领域云存储规模居行业前列。

（1）智能编码

智能编码可以根据输入源的运动情况、单帧内图像复杂度、目标码流范围、感兴趣区域（如人脸区域），动态调整编码的 GOP、帧率、输出码率，在同等编码主观质量前提下得到比常规编码更小的码流。还支持跳帧参考、自定义量化矩阵、主观因子等丰富编码能力，进一步改善编码质量。在静态场景下，平均码率降低 50%。

片源名称	分辨率	运动属性	帧率	协议	目标码率(kbps)		实际码率(kbps)		Smart相对CBR节省百分比
					最小码率	最大码率	CBR	Smart	
StaticSample	1080P	静止/几乎不动	25	H.264	300	2500	2000	362	-81.90%
				H.265	200	1800	1500	237	-84.20%
SmallMotionSample		小运动/运动量不大		H.264	300	2500	2000	803	-59.85%
				H.265	200	1800	1500	536	-64.27%
MiddleMotionSample		中等运动		H.264	300	2500	2000	909	-54.55%
				H.265	200	1800	1500	721	-51.93%
MiddleMotionSample				H.264	300	2500	2000	1120	-44.00%
				H.265	200	1800	1500	934	-37.73%
LargeMotionSample1		剧烈运动		H.264	300	2500	2000	1933	-3.35%
				H.265	200	1800	1500	1426	-4.93%
LargeMotionSample2				H.264	300	2500	2000	2096	4.80%
				H.265	200	1800	1500	1581	5.40%

表 3 智能编码测试结果

（2）智能编码+ROI

增加基于人脸识别的 ROI 技术后，单帧码率波动不大，系列平均值波动更小。按 1080p 25fps 折算，静止或小运动场景下，HEVC/H.265 格式可稳定输出 500Kbps 以下码率，AVC/H.264 格式可稳定输出 750Kbps 以下码率。低于 CBR 对 HEVC/H.265(1.5Mbps)、AVC/H.264(2Mbps)的码率要求。可在平均码率基本不变的前提下，提高人眼感兴趣区域的清晰度。

（3）云端/终端超分

视频云端/终端超分技术分为两个部分，既有发送端也有接收端。在云端利用视频处理提升编码压缩率，通过利用场景分类、内容理解、质量分析以及 JND 来决策编码参数；通过利用视频编码视频前处理提升视频质量达到更好效果。在终端利用视频编码提升视频处理效果，通过编码决策优化对视频进行后处理；通过分层编码发现对质量更有价值的信息。在监控的直播和回放场景，最大可支持 4 倍视频超分。

增强前　　增强后

图 30 视频超分处理效果

B.5.3. 创新与亮点

（1）强大的研发能力和技术应用

移动看家以视频算力为中心，开放多元泛在连接、异构网络接入、敏捷算力输出、多维立体呈现等能力，以 AIoTel 为技术底座，创新推出智能提醒、时光轨迹、智能对讲、视频巡检等视联算力服务及一键呼叫摄像头、支持 VoNR 视频对讲的智能可视门锁等硬件产品，满足泛家庭场景下安防服务、亲情看护、安全监测等不同需求。

（2）强大的智慧家庭生态

中国移动积极联合产业链，从前期的方向共商、研发协同、技术创新，到后期的场景孵化、品牌营销、渠道拓展，和家亲一站式赋能平台为生态伙伴提供技术、业务与服务多方位的赋能服务，实现生态价值的闭环。

B.6 实时通信：和家智话

B.6.1. 案例内容

和家智话是一套面向多形态终端的智能通信系统，其立足 AIoTel 解决方案，依托超高清音视频、交互式通信、智能物联网等核心技术，以家庭固话码号为基础，赋能多形态家庭智能终端，打通全国 31 省 BOSS、接入全国 IMS 核心网，实现大中小屏等多形态终端及 VoLTE 手机的互联互通。可满足老人、小孩等不同用户群体在家庭场景下的通信需求，为用户提供高清、必达、互通、可互动的多媒体通信增值服务。

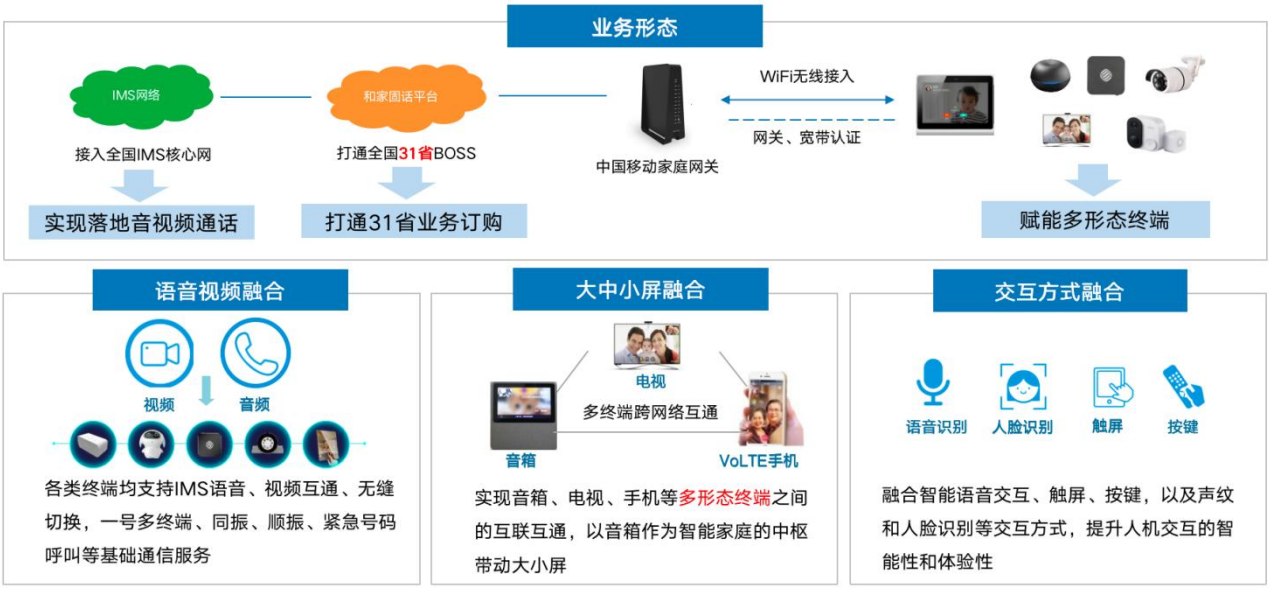

图 31 和家智话业务总视图

典型应用场景如下：

（1）亲情沟通

提供多终端设备无缝切换能力，充分满足家庭场景化需求。当用户拨打智能终端虚拟号时，家中的大屏电视、中屏音箱、智能手表等视联网终端均能收到音视频来电。

（2）远程医疗

借助 5G 全千兆高带宽、低延时特性，提供流畅的高清远程健康服务，可为用户提供多方视频通话及互动体验，用户通过智能语音交互发起“专家咨询”，可实现挂号、预约指定专家。

（3）在线教育

作为 5G 智慧教育最后一公里的关键环节，在线教育服务利用机顶盒、智慧大屏等智能终端，为用户提供多媒体远程互动教学，在家中体验课堂式教学。

（4）智能提醒

创新打造轻量级电话提醒增值服务。用户可通过和家亲 APP 或 H5 页面，对自己、家人发起电话提醒，支持视频、语音、文字图片等播报，支持 VoLTE 视频，满足个人提醒及家人关爱等需求。

B.6.2. 案例成效

和家智话累计服务用户 2000 万，累计覆盖 15 个品类 700 余款设备，致力于构建全行业领先的视联多媒体通信平台。

（1）非对称视频编码

在通信双方算力不均衡的场景下，结合动态编码、云端视频处理等技术实现非对称视频通信，通过自适配发送端与接收端的编码类型、编码参数、分辨率、帧率，满足 H.264/H.65 下分辨率 360P~4K 的异构终端通信，实现大中小屏与 IMS 间的视频无缝衔接。

（2）信源信道联动编码

视频物联网通信网络质量波动较大，依托终端和平台对通信视频质量的监测，向编码器实时反馈上下行速率、丢包率、抖动率、使用等信道信息，结合自适应长期参考帧编码、可变分辨率编码等技术，动态调整视频码率和分辨率，提升信源对信道的适应性，增强通话体验。算法优化后，高清视频占比提升 37%，通话卡顿率下降 22%，平均 MoS 分提升 20%。

（3）语义视频通信

探索语义编码技术在会议、1v1 等人脸视频通话场景下的应用，结合视频结构化、特征编码、GAN 等技术，实现超低码率的窄带视频通信，对比商用 x265，可降低 80%码率，突破传统视频通信上限。

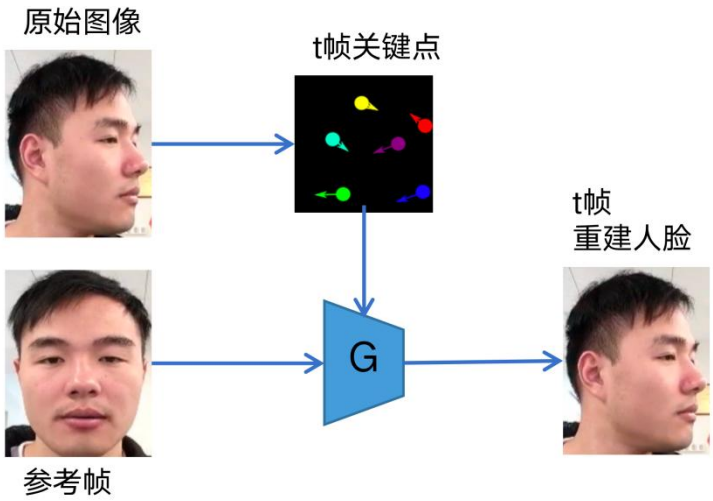

图 32 语音通信效果图

B.6.3. 创新与亮点

和家智话业务面向移动互联网、视频物联网，打造了一套高清、低碳、异构接入的实时通信系统。依托 AIoTel 能力底座，创新提出了双流异显 4K 视频通话、交互式多媒体响应、软件定义 VoLTE 通信模组等技术。

（1）双流异显 4K 超高清

创新提出基于机顶盒的双流异显 4K 超高清视频通话技术，通过超高清编解码弹性伸缩技术，提高运行覆盖率，基于网络传输控制算法优化增加 4K 媒体传输鲁棒性，同时解决了机顶盒同时解码两路 4K 超高清视频流的性能瓶颈问题。

（2）交互式多媒体响应（IMR）

创新提出 And-SIP 轻量化通信协议，在视频传输的基础上提供多媒体数据交互服务，使硬件在通话、直播等多媒体交互过程中支持开锁等智能家居设备控制功能，较大拓展了视频实时通信的服务场景。

（3）软件定义 VoLTE 通信模组（SD-VCM）

创新提出软件定义 VoLTE 通信模组技术，通过融合 SIM 认证、多网接入、慢心跳等机制，让智能物联网终端无需通信模组或芯片即可获得电信级通信服务。并通过多媒体引擎软硬件自适应技术，研发轻载化编解码算法，解决了跨品牌、跨平台、跨系统的智能物联网终端设备适配难题。


UWA
UHD World Association
世界超高清视频产业联盟