



# 构建智能实时网络， 使能 5G 视频业务繁荣

## Building Smart Real-Time Networks to Enable Prosperity of 5G Video Services

吕达 /LYU Da

郑清芳 /ZHENG Qingfang

(中兴通讯股份有限公司, 中国 深圳 518057)  
(ZTE Corporation, Shenzhen 518057, China)

**摘要:** 5G 将促进视频业务的大繁荣, 包括极大地改善现有的视频业务体验和催生新型的视  
频服务形态。为应对 5G 视频业务所面临的超低时延、高可靠及高体验质量等方面的挑战,  
中兴通讯提出构建智能实时视频网络 (SmartRTN) 的理念, 并围绕这一理念, 创新性地研  
发出一系列技术和方案, 包括基于内容智能分析的低码高清视频编码技术、超低时延的网络  
传输、基于深度学习的内容处理与增强、结合边缘计算以及网络切片的组网方案和智能调度  
策略等。这些技术和方案被应用于视频业务端到端各个环节, 有效地解决了困扰 5G 视频业  
务发展的技术瓶颈问题。

**关键词:** 低码高清; 实时通信; 超分辨率; 智能调度; 体验质量

**Abstract:** 5G is expected to bring prosperity of video applications, including significantly  
improving existing applications and bringing forth new exciting applications. To meet  
the challenges of 5G video applications, such as ultra low latency, high reliability and  
high quality of experience, ZTE proposes the concept of constructing smart real time  
video network (SmartRTN). Based on this concept, ZTE innovatively develops a series of  
technologies and solutions, including low bitrate high quality video compression based on  
content intelligent analysis, ultra low latency video transportation, smart video processing  
and enhancement based on deep learning, networking solutions and intelligent scheduling  
strategies combining edge computing and network slicing. These technologies and  
solutions have been applied in all end-to-end video service processes, effectively solving  
the technical bottlenecks that beset 5G video service development.

**Keywords:** low bitrate high quality; real time communication; super resolution; smart  
scheduling; quality of experience

DOI: 10.12142/ZTETJ.202101013

网络出版地址: [https://kns.cnki.net/kcms/  
detail/34.1228.TN.20210126.1653.002.html](https://kns.cnki.net/kcms/detail/34.1228.TN.20210126.1653.002.html)

网络出版日期: 2021-01-27

收稿日期: 2020-12-08

## 1 5G 视频业务概述

### 1.1 5G 促进视频业务持续发展

5G 具有大宽带、低时延的特点, 它解决了视频业务发展的关键瓶颈问题, 大大促进了视频业务的发展。5G 不仅使传统视频业务, 如安防、视频会议、点播、直播等, 获得了迅速发展, 还使由视频业务衍生的远程教育、远程医疗等远程交互业务也获得了巨大发展。更进一步地, 面向家庭

和娱乐场景的超高清视频、沉浸式视频、全景视频、3D 视频也获得了高速发展的机会。

根据 Cisco 可视化网络指数 (VNI) 预测, 到 2022 年, 全球互联网协议 (IP) 视频流量将占总流量的 82%, 如图 1 (a) 所示。所有形式的 IP 视频 (包括互联网视频、IP 视频点播、视频流游戏、视频会议和基于文件共享的视频文件) 的总和将继续保持在总 IP 流量的 80%~90%。2017—2022 年, 全球

视频流量的复合年增长率为 26%。随着网络的广泛部署以及市场竞争的发展, 移动视频业务发展迅猛, 2022 年移动视频流量将占据总移动数据业务的 79%, 并保持 46% 的年复合增量率, 如图 1 (b) 所示。

### 1.2 新型视频业务下的端到端技术指标

视频业务的形态不断增加, 对端到端的技术指标提出差异化要求: 准实时直播时延的可接受范围为 1~3 s; 实

时互动直播的时延要控制在 500 ms 以内；视频会议要求端到端时延需要在 200 ms 以内、编码时延需在 100 ms 以内、操作指令时延在 30 ms 以内；而对于实时性要求较强的增强现实（AR）/虚拟现实（VR）业务及云游戏业务，端到端时延一般需要控制在 100 ms 以内、编码时延需要在控制在 10 ms 以内。

### 1.3 5G 视频业务端到端的质量仍需提升

5G 给网络状况带来的提升只是视频业务繁荣的必要非充分条件，我们还必须从视频业务端到端全流程的角

度来设计完整的技术体系。5G 只是为视频的高效传输提供底层网络支撑。如何协同利用人工智能（AI）、云计算、边缘计算等新技术，来构建端到端的视频技术体系，以及如何从智能实时网络、智能化处理、端云边协同高性能计算及存储、智能部署等多个角度、业务全流程，来提升视频业务采集、预处理、编码、传输、解码、后处理、渲染各个环节的处理效率和业务质量，以给用户提供更清晰、高流畅度、低时延、强交互的极致用户体验，是中兴通讯正在努力的方向。

## 2 中兴通讯打造下一代智能实时网络（SmartRTN）

中兴通讯基于多年技术积累和产品研发工作，从信源、信道、用户体验、业务部署及运维等多方面综合考虑，提出通过构建智能实时视频网络来使能视频业务繁荣的理念。围绕这一理念，中兴通讯创新性地研发出一系列技术，并将这些技术成功应用于视频业务端到端各环节，例如：

（1）在信源方面，中兴通讯结合业界最新视频编码标准的进展，通过对视频内容的智能分析，合理地分配码率，尽可能在保证较高画质体验的前提下提升数据压缩比；进一步地引入基于 AI 的图像生成技术，使特定场景内容（如人脸等）取得了极致的压缩比。

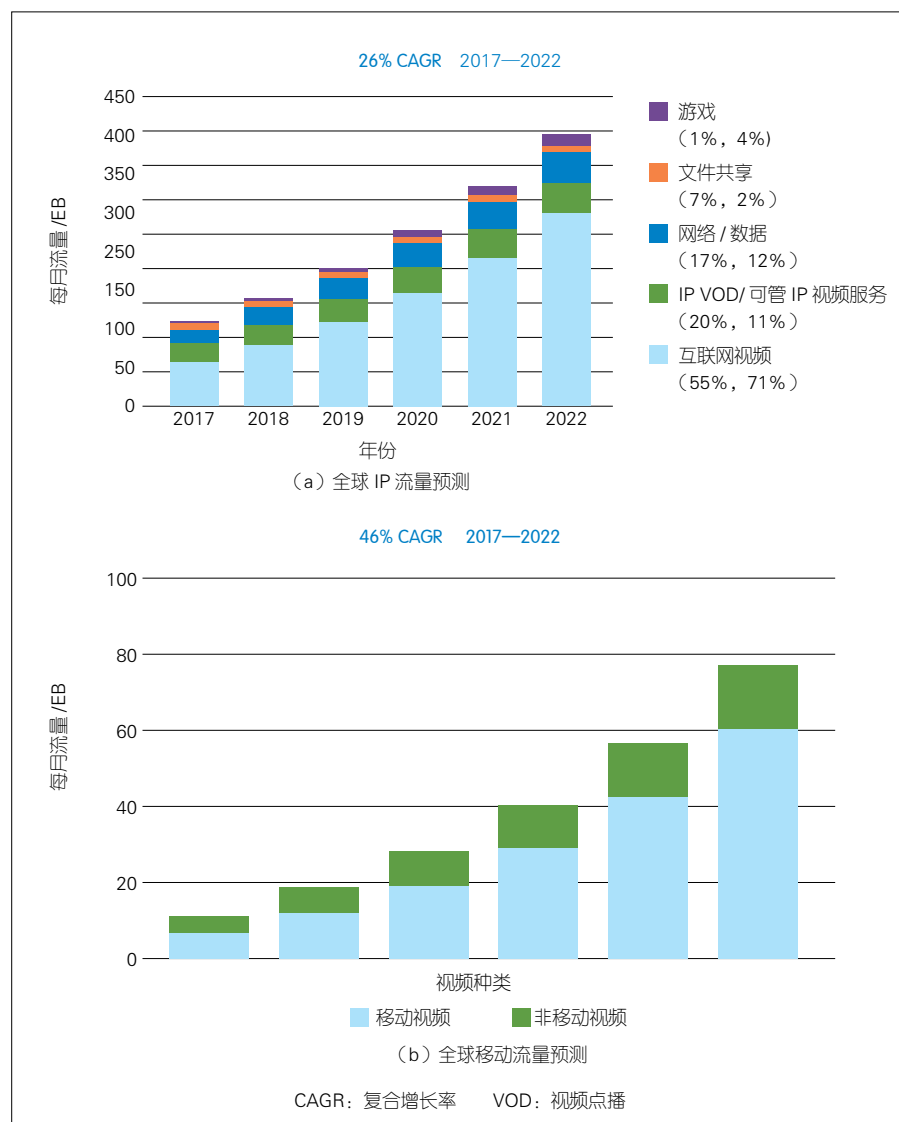
（2）在信道方面，中兴通讯自主研发传输协议，通过控制视频编码与传输之间的协同机制，有效降低了传输时延。精巧设计的抗丢包策略，实现了弱网环境下的可靠传输。

（3）在用户体验方面，中兴通讯研发了一系列技术，以对不同环节予以改善。例如，在成像环节，增强在不同光照条件下的画面清晰度；在显示环节，通过虚拟背景技术保护用户隐私；在会议场景中，通过对人脸以及人物动作的识别，使会场管理更加便利。

（4）在业务部署和运维方面，中兴通讯借助 5G 的网络切片，实现了用户服务质量（QoS）差异化保障；使用边缘计算，实现业务的就近接入和媒体的下沉处理；通过智能路由，实现最优路径的选择；通过智能用户体验质量（QoE）检测，及时发现故障并无感修复。

### 2.1 低码高清

视频低码高清是指在保证视频画



▲图 1 Cisco 可视化网络指数全球 IP 流量预测和全球移动流量预测（2017—2022 年）<sup>[1-2]</sup>

面质量的前提下, 尽可能提升压缩比、降低视频码率, 它可以从视频编码、视频前后处理等多个维度进行提升。视频编码主要分为基于现有成熟编码的优化和新一代编码技术的引入。

### 2.1.1 挖掘现有视频编解码标准的最大潜力

基于目前产品广泛使用的 H.264/H.265 编码, 我们实现针对不同场景的内容感知编码 (CAE) 优化:

#### (1) 基于感兴趣区域 (RoI) 编码优化

在典型的视频通信场景中, 人们的主要关注点在于人脸及周边区域, 而非背景区域。如图 2 所示, 在视频通信发送端引入实时人脸检测和基于 RoI 的编码算法, 并对不同区域设置不同的码率, 可使最终实现的 RoI 编码在保持画面主观质量不下降的前提下, 实现 20% 的码率节省。

#### (2) 基于屏幕内容特性的压缩编码优化

无论是视频通信还是云电脑的应用, 视频的内容来源主要包括两类: 屏幕内容分享和摄像头视频。屏幕内容和摄像头采集生成的视频内容有本质差别, H.265 已有专门针对屏幕内容的高效视频压缩编码 (HEVC) - 屏幕图像编码 (SCC) [3] 压缩标准。如图 3 所示, 考虑到现有大规模部署的 H.264 系统, 针对视频会议的辅流文档共享、云电脑的屏幕内容分享场景, 中兴通讯对屏幕内容进行分类压缩, 采用调色板、文字特征提取等压缩方式, 在确保文字区域无损清晰的前提下, 使图像传输带宽降低 10% 以上。

#### (3) 基于动态帧率的编码优化

在视频通信或云电脑的实际使用场景中, 经常会出现阶段性画面无变化的情形, 比如, 在视频交互通信中播放幻灯片 (PPT) 文档内容、云电脑

用户操作不太频繁。动态帧率的编码优化能够根据场景的运动剧烈程度来动态实时调整帧率, 比如在 PPT 分享或屏幕应用静止时, 可以通过自动降低帧率实现至少 10% 的综合带宽降低效果。

### 2.1.2 研发新一代编码技术

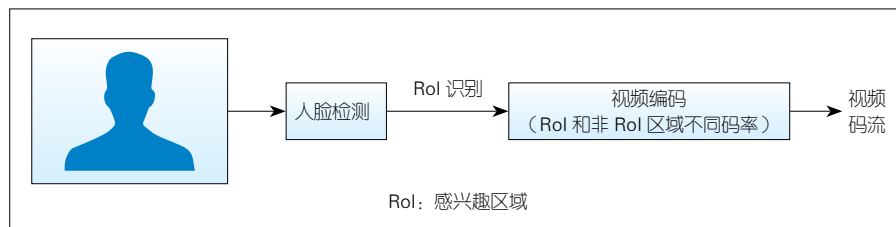
除了前述基于现有 H.264/H.265 进行码率、帧率等方面的编码优化外, 中兴通讯还积极参与研发新一代视频编解码技术。目前, 全球最新视频 Codec 标准主要以多功能视频编码 (VVC, 也称 H.266) [4]、开放媒体联盟视频标准 (AV1) [5] 和第 3 代数字音视频编解码技术标准 (AVS3) [6] 为主流,

同时基本视频编码 (EVC) [7] 和低复杂度增强视频编码 (LCEVC) [8] 针对特定场景 (如降低编码复杂度、充分利用现有硬件等) 也有一定的应用空间。部分最新视频编码码率降低效果对比结果具体如图 4 所示。

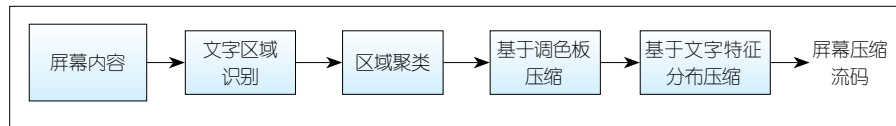
中兴通讯持续参与 HEVC、VVC 标准的制定工作, 并在当前动态图像专家组 (MPEG) 的两个特别工作组 (AHG) 中担任领导职位。

### 2.1.3 AI 进一步提升压缩比

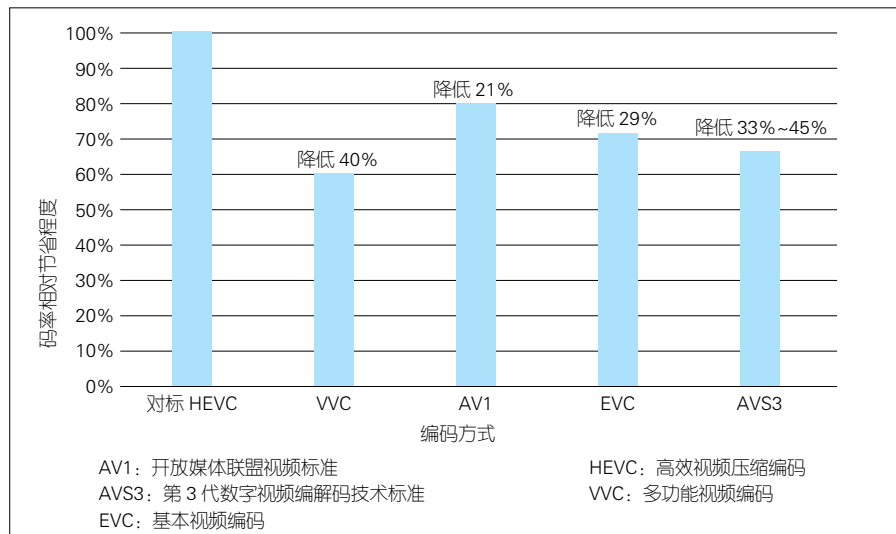
AI 在各个领域中的应用得到了迅猛发展, 并在特定的业务场景中, 带来了新的解决方法。关注用户真正的场景需求有可能颠覆传统的视频编



▲ 图 2 感兴趣区域编码示例



▲ 图 3 中兴通讯针对屏幕内容的文字压缩技术



▲ 图 4 最新 Codec 码率降低对比图 (相对 HEVC)

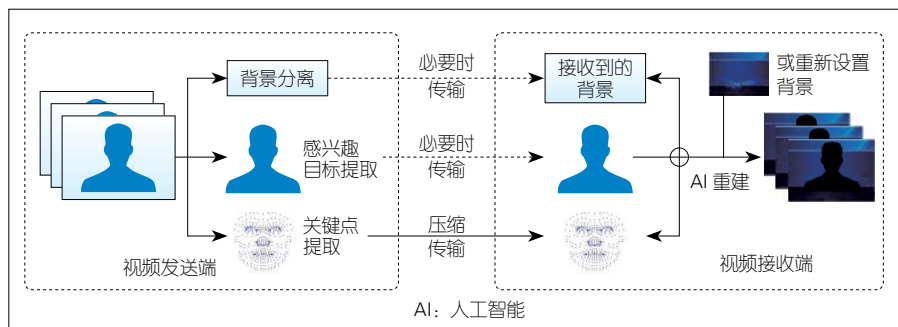
解码技术，并带来极致的压缩比。例如，在 SmartRTN 网络中，针对个人视频通信这种场景，传输的视频帧主要由变化很小的背景图片和运动的人脸构成，用户的关注点主要是表情的交流。基于生成式对抗网络（GAN）的人脸生成技术可以对摄像头获取的内容的关键信息进行重构，以形成新的解决方案，具体如图 5 所示。与传统的基于像素域的信号处理技术相比，中兴通讯提出基于感兴趣目标和稀疏关键点检测的方法，对视频信息进行编码。基于运动驱动感兴趣目标，并结合该场景下背景信息重构压缩后的视频帧，该方法使码率得到了更加有效的压缩——可以达到传统算法码率的 1/10。同时，重构后的视频可以任意切换光照模式和视角，在虚拟会场中可以实现统一的光照模式及物体的任意视角，为个人视频通信业务提供更具沉浸式的临场感和更加真实的眼神交流体验。

## 2.2 超低时延传输

### 2.2.1 不同视频传输协议

针对不同的视频业务场景，常见的视频传输协议有实时消息传输协议（RTMP）、通用媒体应用格式（CMAF）、低时延 HTTP 实时流媒体（LHLS）、Web 实时通信（WebRTC）等，其技术特性对比如表 1 所示。

为了满足实时音视频通信对低时延传输的需求，中兴通讯研发了超低时延传输协议，在传输层参考快速用户数据报网络连接（QUIC）<sup>[11]</sup> 协议的基础上做了大量的重新设计，例如加密机制、多路径、前向纠错（FEC）支持、优先级管理、可配置的拥塞控制算法等，以满足实时流传输场景的需求。在媒体传输的应用层上，中兴通讯研发的协议与 WebRTC、RTMP 等协议兼



▲图 5 中兴通讯基于人工智能的编码解决方案

▼表 1 不同传输协议对比

协议	理想延迟区间/s	降低时延机制	应用场景
RTMP/FLV	1~2	流式传输 + 协议栈优化	泛娱乐直播
CMAF <sup>[9]</sup>	2~3	索引预生成及 HTTP 编码块传输	OTT、体育直播
LHLS <sup>[9]</sup>	2~3	索引预生成及 HTTP 编码块传输	低时延直播
RTP/WebRTC <sup>[10]</sup>	0.2~1	链接及传输协议优化，尽可能低的抖动缓冲区（jitterbuffer）	超低时延直播、视频通话

CMAF：通用媒体应用格式

FLV：Flash 视频格式

HTTP：超文本传输协议

LHLS：低延迟 Http 实时流媒体

OTT：指越过运营商的互联网（视频）业务

RTMP：实时消息传输协议

RTP：实时传输协议

WebRTC：Web 实时通信

容，可适配各种不同的实时音视频应用场景的需求。

### 2.2.2 融合编码和传输技术

实时通信系统需要考虑视频编码器和传输协议的协同控制。传输协议和编解码器不同步或网络条件不稳定，容易引发延迟现象或故障。

（1）有两份视频编码时，选择最合适的一份以避免拥塞。

斯坦福的 Salsify 项目<sup>[10]</sup>创新性地体现了新的组合方式——编解码器速率控制和传输拥塞控制。Salsify 的编解码器可保证发送者不会在网络拥塞时发送帧（必要时会丢弃已经编码的帧），且不固定帧的发送速率。同时，编解码器还可被允许生成更接近可用网络容量的帧，且生成每个帧的两个版本：一个质量略高于先前的成功案例，另一个则质量略低。应用程序在查看每个选项的实际压缩大小后，从这些选项中进行选择（或不选）。官方的测试结果表明<sup>[11]</sup>，Salsify 比现有

的商用系统（如 Skype、FaceTime 和 WebRTC）在时延控制和视频质量上更为优秀。

（2）采用编码与传输的管道机制，边编码边传输

音视频采集、编码、传输、解码、渲染等流程是相互联动和影响的。采集、编码与传输形成管道，可以有效降低时延。例如，视频编码编完一个切片后，在编下一个切片的同时，可传输刚编完的切片数据；若采用 SVC 或 LCEVC 编码，则可以编完一个层，且在编下一个层的同时，立刻传输已编完的层数据。

### 2.2.3 拥塞控制技术

实际的网络状态是复杂多变的，丢包、延时和网络带宽都在时刻变化，这就对网络拥塞控制算法提出了很高的要求。网络拥塞是指发送的数据超过了网络所能承载的传输能力。尽管基础通信设施在不断地完善，网络拥塞的情况在 5G 时代还是会有可能出现。

针对实时音视频传输的拥塞控制, 中兴通讯提出适应多场景的拥塞控制模块, 包括传统的基于传输控制协议 (TCP) 的瓶颈带宽和往返时延 (BBR)<sup>[12]</sup>、基于用户数据报协议 (UDP) 的谷歌拥塞控制 (GCC)<sup>[13]</sup> 和基于机器学习的拥塞控制功能。这些拥塞控制模块可以被选择部署在云端或者集成在发送端。

(1) 针对视频专网等高可靠环境, 通信双方可以采用 TCP 方式传输实时音视频数据。此时发送端自动采用基于 TCP 的控制模块。目前主要采用的拥塞控制算法是 BBR 系列。

(2) 对于弱网不可靠环境, 通信双方采用 UDP 方式传输实时音视频数据, 发送端则自动采用 UDP 系列的控制算法, 如 GCC。

(3) 另外, 中兴通讯提出的拥塞控制模块还包括支持基于大数据驱动的智能拥塞控制决策模块。该模块通过收集发送端、传输网络、接收端等多方的信息, 形成对网络拥塞程度的预测, 从而推动发送端选择不同的编码参数、不同的传输协议、拥塞控制参数 (详细技术原理可参考本文 2.4.3 节)。

#### 2.2.4 FEC、自动重传请求 (ARQ) 等弱网对抗技术

FEC 也叫前向纠错码, 是视频业务系统网络保证可靠传输质量的重要方法。FEC 可以对  $n$  份原始数据增加  $m$  份数据, 并能通过  $n+m$  份中的任意  $n$  份数据, 还原原始数据, 即如果有任意小于等于  $m$  份的数据失效, 仍然能通过剩下的数据还原出来。当前的 FEC 算法使用范特蒙矩阵或者柯西矩阵, 来实现纠错码的功能。通过在传统 FEC 算法上做自适应改进, 中兴通讯的视频 FEC 方案可根据网络条件, 实现延时自调整、网络自适应、冗余自增减等功能。

ARQ 也是抵抗网络丢包的一种重要手段。中兴通讯视频系统使用的是基于否定确认包 (NACK) 的丢包重传技术。NACK 是一种通知技术, 其触发通知的条件刚好与确认包 (ACK) 相反。在未收到消息时, NACK 通知发送方“我未收到消息”, 即通知未达。NACK 在接收端检测到数据丢包后, 发送 NACK 报文到发送端。发送端根据 NACK 报文中的序列号, 在发送缓冲区找到对应的数据包, 并将其重新发送到接收端。ARQ 和 FEC 配合使用, 可以在不大幅增加网络冗余的条件下, 实现较好的抗丢包效果。在实际应用中, 中兴通讯视频系统能抵抗 80% 的网络丢包, 满足 95% 以上的使用场景。

### 2.3 视频智能分析

#### 2.3.1 暗景增强实现低光照下的视频画质提升

在视频通信场景中, 由于场地变换、光照摄像头角度变化等因素, 通常会出现由关键人脸部分光照不均匀导致的暗影现象, 这影响了用户体验。中兴通讯通过对大量 3D 人脸在不同光照模式下的数据进行模拟训练, 实现了基于 2D 图像对光照条件的预测, 并通过光照条件的映射实现了自然光照场景下人脸图像的非线性变换模拟, 使之达到了光照均匀的效果, 提升了暗光场景下的人脸画质。

#### 2.3.2 人像分割及背景虚化

视频通信可以随时随地通过移动终端接入。虽然这极大地方便了客户使用, 但同时也导致客户个人私密信息出现在视频中。因此, 基于语义分割的背景和背景虚化功能就成为了视频通信产品不可或缺的功能。

中兴通讯基于神经网络架构搜索技术构建了轻量级模型, 在自收集的

Portrait 数据集上进行训练, 实现了端侧的语义分割算法, 并通过网络模块轻量化设计、模型剪枝及模型蒸馏等提速方案, 得到了 300 kB 大小的轻量级语义分割模型。通过端侧的部署加速和前后端处理的项目流程优化, 我们在骁龙 845 手机芯片上实现了高达 33 帧/秒的实时推理过程。

背景替换和虚化技术是基于实时人像分割技术的应用。在使用轻量化深度神经网络对输入图完成人像分割任务之后, 所得的人像分割网络输出背景为 0、人像为 1 的图像, 并与输入图进行相乘可保留人像信息。背景替换的图片可以首先将网络输出的图像取反, 然后进行相乘生成替换的背景图像, 最后将人像信息和背景图像合成一张图片, 即可得到所需的背景替换, 具体如图 6 所示。

#### 2.3.3 人脸识别

中兴通讯基于大规模私有人脸数据集、深度卷积神经网络的人脸特征编码模型以及度量学习方法, 在人脸识别领域有着长期的技术积累。特别地, 在视频人脸识别处理中, 中兴通讯提出综合视频空域信息的代表帧融合和特征增强方法, 相应的处理流程如图 7 所示。表 2 给出了中兴通讯视频人脸识别方案在标准测试集 YouTube Faces 上的准确率比较。该方法大大提高了人脸特征的泛化性, 同时提高了对运动/失焦模糊、低分辨、视频编解码噪声的耐受力, 并在多个开源测试集上达到了较高的准确率。

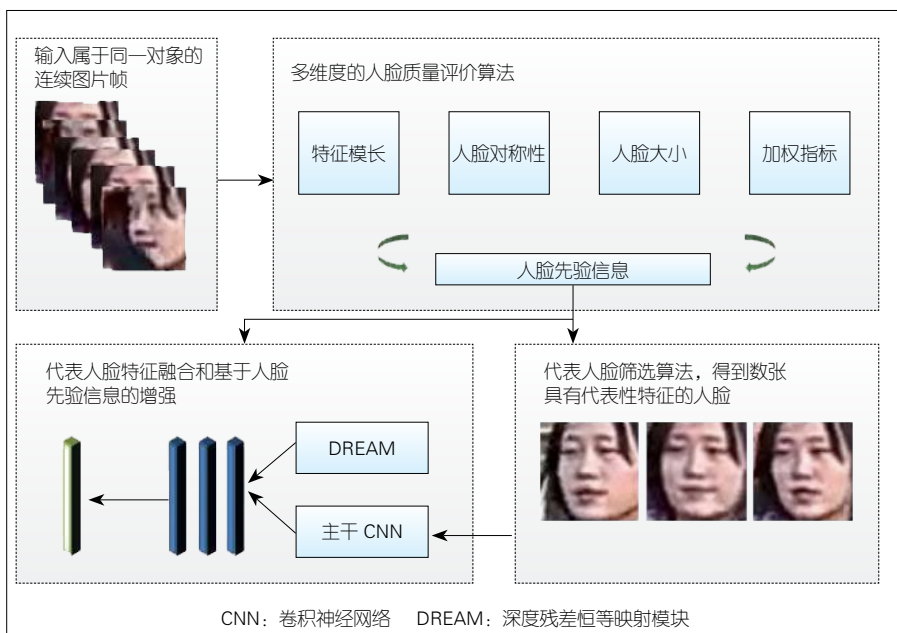
### 2.4 智能部署

#### 2.4.1 融合移动边缘计算 (MEC) 和网络切片

基于 5G 端到端网络切片技术, 对专用网络进行优化, 可实现视频服



▲图 6 暗景增强、背景虚化、背景替换效果对比



▲图 7 中兴通讯视频人脸识别处理流程<sup>[4]</sup>

▼表 2 中兴通讯视频人脸识别方案的准确率比较

测试方法	准确率 / %	所用帧数	融合方法
代表帧 + 特征增强	95.98	3	平均池化
特征融合	95.40	1	/
代表帧	95.80	3	平均池化
单帧	95.04	1	/
FaceNet <sup>[15]</sup>	95.12	100	平均相似度
DAN <sup>[16]</sup>	94.28	所有帧	平均池化
DeepFace <sup>[17]</sup>	91.40	100	平均池化
NAN <sup>[18]</sup>	95.72	所有帧	加权池化
C-FAN <sup>[19]</sup>	96.50	所有帧	加权池化

C-FAN: C-特征聚集网络 DAN: 判别聚集网络 NAN: 神经聚集网络

务加速、视频服务网络与其他网络业务隔离服务, 解决网络拥塞和时延问题。支持 5G 接入侧的 MEC 视频服务下沉, 不仅可实现媒体就近接入、就

近处理, 为用户带来更低时延的视频体验, 还可同时降低对骨干网带宽占用。更进一步地, 融合 5G 网络切片和 MEC 可对基站、频率专享等组成 5G

虚拟专网, 可以满足高端客户的高安全、高可控、高性能要求, 如图 8 所示。

#### 2.4.2 智能路由调度

由于 RTN 网络服务用户的网络条件和质量各异, 基于强大的大数据分析和 AI 预判能力的支持, 中兴通讯实现了实时的智能路由调度, 具体如图 9 所示。针对统一接入调度模块, 用户侧接入调度除了选择就近边缘接入外, 核心的网络路由可以选择进行如下操作:

(1) 基于大数据提取多维度网络路由质量评价指标, 生成当前路由优劣评分;

(2) 基于现有的评价模型, 实现了未来 5~10 min 内网络质量的预判;

(3) 实时统计网络各个节点、不同粒度的质量参数 (如带宽、往返时延等), 并综合前两者的评分结果, 实现当前路由表的实时动态调整。

另外, 为了保证低时延, 网络架构设计与传统的内容分发网络 (CDN) 分层设计稍有差异。其中, 核心中继服务器采用扁平 Mesh 组网架构, 内部链路更短、更灵活, 可支持采用动态选路的方式来调整构建的网状结构。中继服务器之间采用优化过的 QUIC 协议实现数据传输, 使内部链路延迟达到 30 ms 左右。

#### 2.4.3 智能 QoE 监测

实时视频服务的 QoE 受到实时音视频采集、前处理、编码、传输、解码、后处理、渲染各个环节的影响。一旦某一个环节出现问题, 如传输过程中的网络丢包、采集环节中的系统不兼容, 都会直接导致实时音视频服务出现质量问题, 影响用户体验。因此, 我们需要建立端到端的实时音视频服务智能 QoE 监测和优化系统。

如图 10 所示, 实时音视频服务

智能 QoE 监测和优化系统分为数据收集、健康度评估和智能优化 3 个部分。

(1) 数据收集。该部分主要收集端到端的全链路实时音视频通信数据，包括终端设备数据、网络环境数据等。

- 终端设备数据：设备机型、用户 IP、视频流的分辨率、帧率，在前处理、编码、解码、后处理、渲染等过程中的 CPU 使用率，图形处理器 (GPU) 使用率以及内存使用率等；
- 网络环境数据：上下行网络丢包、抖动、时延等。

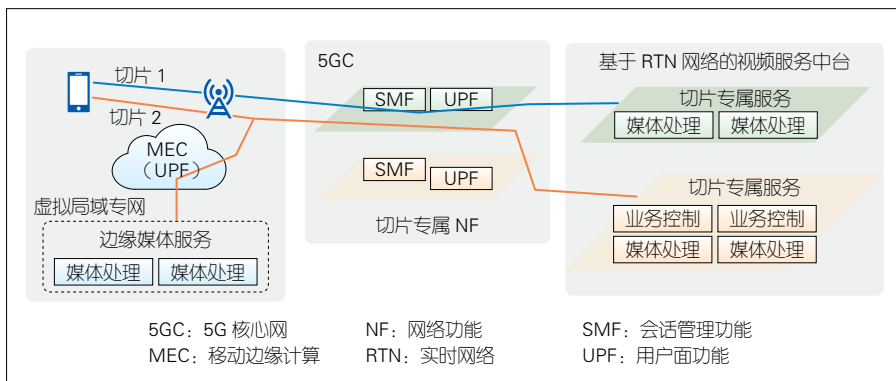
包、抖动、时延等。

(2) 健康度评估。该部分的核心思想是对收集到的监控数据进行过滤、汇聚、实时计算，并进行实时音视频通信质量评估，快速识别和感知实时音视频通信中的问题。中兴通讯通过构建 / 更新一组机器学习模型，判断当前全链路服务状态的健康程度，并将其作为后续智能优化阶段的触发条件。具体来说，该部分包括：首先，基于“异常状态监控指标与正常状态监控指标处于不同分布”的假设，选用 QoS 指标（时延、码率、CPU 等）<sup>[20]</sup> 和无参考视频质量评估得分，构造样本特征空间；然后，在此基础上构造多个异构自动编码器<sup>[21]</sup>，并利用它们在训练集上的预测残差值进行正则化模型筛选；最后，通过模型筛选的多个自动编码器的投票结果，将被作为当前状态健康度的评估值。

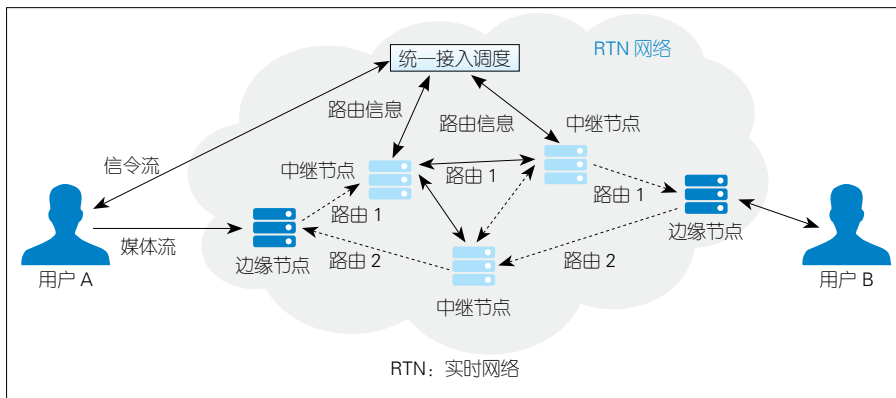
(3) 智能优化。当前状态的健康度低于阈值时，就需要进行智能优化。这里我们将智能优化过程视为马尔科夫决策过程，利用强化学习求解当前状态下的最优策略。具体来说，我们将健康度的前后提升比率定义为奖励，将网络状态的可观测信息（时延、丢包、阻塞情况）定义为状态空间，将网络参数组合的可调选项（纠错策略、重传策略、缓冲器的缓冲值和缓冲区大小）定义为动作空间，利用动作探索的奖励反馈实时更新深度策略网络<sup>[22]</sup>，并逐步实现当前状态下的最佳网络配置组合。

### 3 结束语

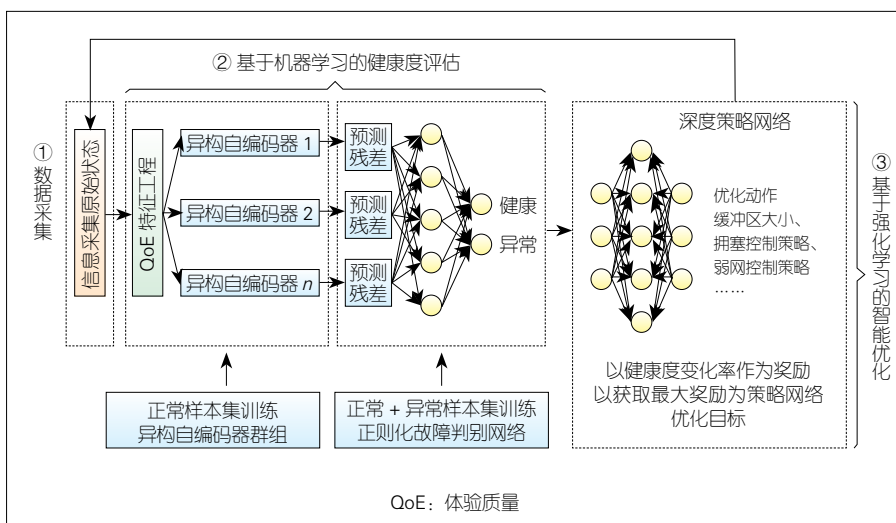
随着 5G 商用落地以及相关设施的完善，视频的使用体验将不断升级，视频的业务形态将不断创新，视频的应用场景也将不断延展。“万物视频化”的趋势对底层技术支撑体系提出了新的、更高的要求。为了使能视频



▲图 8 融合 5G 网络切片、移动边缘计算的实时音视频网络及服务示例



▲图 9 智能路由调度



▲图 10 实时音视频服务智能用户体验质量监测

业务繁荣, 中兴通讯提出了构建智能实时视频网络的理念, 基于自身长期在网络通信、视频多媒体、AI 等领域的持续耕耘和积累沉淀, 创新性地研发了一系列技术和产品, 并使之应用于视频业务端到端流程的各环节。中兴通讯构筑 SmartRTN 综合技术体系, 着眼于改善最终用户的体验, 有效解决了内容增强、高效压缩、可靠传输以及智能运维等问题, 为 5G 视频业务的不断演化和纵深拓展提供了牢固的基础。

#### 参考文献

- [1] Cisco. Cisco visual networking index (VNI) complete forecast update, 2017—2022 [EB/OL]. (2018-12)[2020-12-05]. [https://www.cisco.com/c/dam/m/en\\_us/network-intelligence/service-provider/digital-transformation/knowledge-network-webinars/pdfs/1213-business-services-ckn.pdf](https://www.cisco.com/c/dam/m/en_us/network-intelligence/service-provider/digital-transformation/knowledge-network-webinars/pdfs/1213-business-services-ckn.pdf)
- [2] Cisco. Cisco visual networking index (VNI) global and americas/EMEAR mobile data traffic forecast, 2017—2022 [EB/OL]. (2019-03)[2020-12-05]. [https://www.cisco.com/c/dam/m/en\\_us/network-intelligence/service-provider/digital-transformation/knowledge-network-webinars/pdfs/190320-mobility-ckn.pdf](https://www.cisco.com/c/dam/m/en_us/network-intelligence/service-provider/digital-transformation/knowledge-network-webinars/pdfs/190320-mobility-ckn.pdf)
- [3] ITU. HEVC-SCC [EB/OL]. [2020-12-05]. <http://www.itu.int/rec/T-REC-H.265>
- [4] MPEG. VVC [EB/OL]. [2020-12-05]. <https://mpeg.chiariglione.org/standards/mpeg-i/satellite-video-coding>
- [5] Alliance for Open Media. AV1 [EB/OL]. [2020-12-05]. <http://aomedia.org/>
- [6] AVS Work Group. AVS3 [EB/OL]. [2020-12-05]. <http://www.avs.org.cn/>
- [7] MPEG. MPEG5-EVC [EB/OL]. [2020-12-05]. <https://mpeg.chiariglione.org/standards/mpeg-5/essential-video-coding>
- [8] MPEG. MPEG5-LCEVC [EB/OL]. [2020-12-05]. <https://mpeg.chiariglione.org/standards/mpeg-5/low-complexity-enhancement-video-coding>
- [9] 视频传输延迟分析及解决方案: CMAF, LHLS [EB/OL]. (2018-09-21)[2020-12-05]. <https://cloud.tencent.com/developer/article/1346159>
- [10] WebRTC 1.0: Real-Time Communication Between Browsers [EB/OL]. <https://www.w3.org/TR/webrtc/>
- [11] QUIC, a multiplexed stream transport over UDP [EB/OL]. <https://www.chromium.org/quic>
- [12] Salsify. Video is better when the codec and transport work together [EB/OL]. [2020-12-05]. <https://snr.stanford.edu/salsify/>
- [13] Salsify. Salsify 测试结果 [EB/OL]. [2020-12-05]. <http://web.mit.edu/6.829/www/currentsemester/materials/slides-salsify-lecture.pdf>
- [14] DING Z Z, ZHENG Q F, HOU C H, et al. Improving face recognition in surveillance video with judicious selection and fusion of representative frames [C]//ACM Multimedia Asia(MMA Asia' 20). NY, USA: ACM, 2021. DOI:10.1145/3444685.3446259
- [15] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: a unified embedding for face recognition and clustering [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA: IEEE, 2015: 815-823. DOI:10.1109/cvpr.2015.7298682
- [16] RAO Y M, LU J W, ZHOU J. Learning discriminative aggregation network for video-based face recognition [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 3781-3790
- [17] TAIGMAN Y, YANG M, RANZATO M, et al. DeepFace: closing the gap to human-level performance in face verification [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 1701-1708. DOI:10.1109/cvpr.2014.220
- [18] YANG J L, REN P R, ZHANG D Q, et al. Neural aggregation network for video face recognition [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 4362-4371. DOI:10.1109/cvpr.2017.554
- [19] GONG S X, SHI Y C, KALKA N D, et al. Video face recognition: component-wise feature aggregation network (C-FAN) [C]//2019 International Conference on Biometrics (ICB). Crete, Greece: IEEE, 2019: 1-8. DOI:10.1109/icb45273.2019.8987385
- [20] InfoQ. 音视频质量评估白皮书 [EB/OL]. (2019-08-26)[2021-12-05]. <https://www.infoq.cn/article/xt9vNLcC6dIkSvu9l6M2>
- [21] HINTON G, OSINDERO S, YW T. A fast learning algorithm for deep belief nets [J]. Neural computation, 2006, 18(7): 1527-1554. DOI: 10.1162/neco.2006.18.7.1527
- [22] 桑顿, 巴图. 强化学习: 第 2 版 [M]. 俞凯, 译. 北京: 电子工业出版社, 2019

#### 作者简介



吕达, 中兴通讯股份有限公司云视频与能源研究院院长、高级工程师; 研究方向为通信技术和协议、互联网技术、云计算技术、视频技术、数字家庭网络及业务等; 先后从事数字程控交换机、固网软交换、IPTV、视频会议、通信网络供电等产品架构设计与研发管理工作, 曾主持完成数字程控交换机、多媒体视讯、视频会议等重大产品项目; 发表论文多篇, 申请专利 8 项。



郑清芳, 中兴通讯股份有限公司云视频首席科学家; 研究方向为人工智能、计算机视觉、视频编解码、视频通信、人机交互、多媒体芯片与系统等; 先后从事视频智能编目系统、视频搜索系统、手机 3D 成像系统及应用、车载成像与识别系统、人脸识别、3D 立体视觉芯片、视频会议等系统及产品的架构设计与核心技术研发; 发表论文多篇, 申请专利 2 项。