研究笔记 · 2026 年 11 月

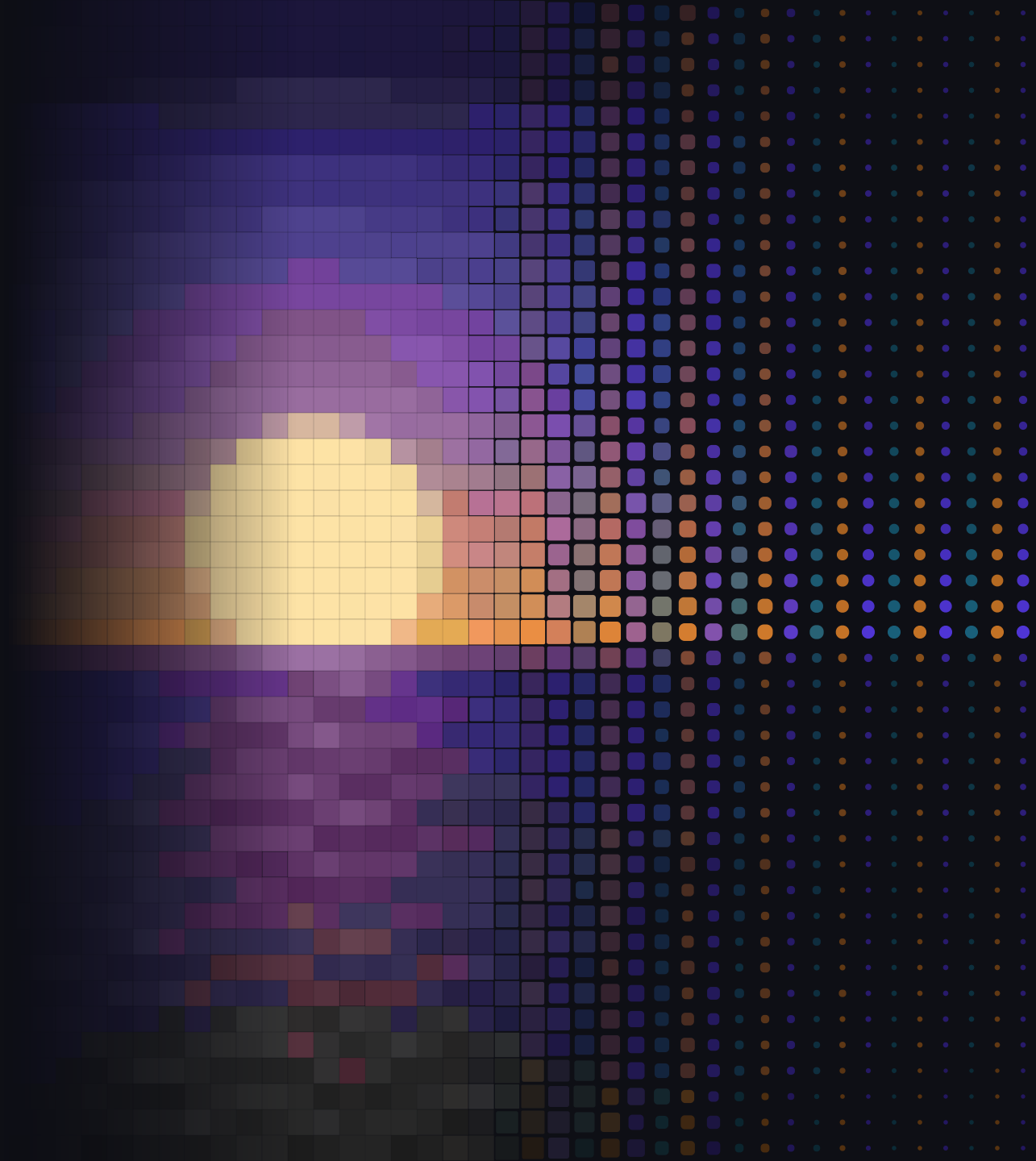
2026 年的神经编码

围绕编解码器，而非取而代之

播放 · 拍照 · 通话 · 生成式 AI · 游戏 · 芯片

谭帅

tanshuai.com · i@tanshuai.com



已是主流，但在编解码器外围

01 神经网络已在标准编解码器两端成为主流

摄像头 ISP、编码端下采样和播放端超分。中间的码流仍是 AV1、HEVC、AVC 或 VVC，由固定功能硬件解码。

02 端到端神经编解码器只在闭环中上线

通话、生成式 AI 的 tokenizer 和一款 3A 游戏：一家公司掌控两端，并且可以退回标准编解码器。

03 主流屏幕上的神经码流要等 H.267

第一版定于 2029 年底；首批芯片约在 2031 年，2030 年代中期普及 估算 。

≈75%

移动数据流量中视频所占的比例（2025 年底）

12.6 亿

2025 年智能手机出货量，全部用固定功能硬件解码视频

≈12 倍

唯一公开的神经解码芯片设计的每像素能耗，对比硬件 HEVC 解码器

神经网络进入编解码的三条路径



编解码器外围：不需要任何人许可。标准内部：需要一个标准。取代编解码器：需要掌控两端。

2026 年神经编码落地在哪里

	标准编解码器外围 预处理、增强、超分	端到端神经编解码器 编码器和解码器都由学习得到	标准内部 标准码流或接口中的神经工具
01 播放与流媒体	● 大规模部署 Netflix 神经下采样；YouTube、Meta 超分；电视 AI 画质芯片	● 未部署 没有流媒体服务向电视或手机发送神经视频格式；终端都用硬件解码标准编解码器	● 标准化中 H.267 第一版定于 2029 年底，神经工具在候选之列
02 拍照与图像	● 大规模部署 旗舰手机的 AI ISP；CES 2026 发布全神经网络 ISP 的 IP	● 演示与试点 JPEG AI 已是 ISO 标准：有手机演示，没有产品	● 未部署 正在普及的是传统格式：Chrome 155 默认支持 JPEG XL
03 通话与会议	● 大规模部署 接收端超分、神经降噪	● 大规模部署 Teams 两方通话自 2021 年用神经音频；MLVC 视频正在 Teams 中推出	● 有限部署 Opus 1.5 加入可选、向后兼容的机器学习工具
04 生成式 AI	—	● 大规模部署 每个视频和语音模型都运行一个学习得到的 tokenizer，例如压缩比 1:192 的视频 VAE	—
05 游戏与图形	● 大规模部署 神经超分与插帧：仅 DLSS 4 就覆盖 250 多款游戏和应用	● 有限部署 一款 3A 游戏已上线神经纹理；英伟达 SDK 仍是 beta	● 标准化中 DirectX 正在为着色器加入矩阵运算（预览）
06 专业链路：卫星、机器人	● 演示与试点 在服务器端修复低码率画面	● 演示与试点 抗丢包神经编解码器处于研究与试点	—

各行按覆盖人数排序。

两端是神经网络，中间是标准

编码端

Netflix 的神经下采样自 2022 年起运行，任何设备都无需改动；在 Netflix 的测试中，约 77% 的观众更喜欢它的结果 厂商说法。

屏幕端

YouTube 的超分自 2025 年 10 月起登上电视；Meta 在服务器和手机上都运行超分。

电视里

三星 NQ8 AI Gen3 处理器宣称有 768 个用于超分的神经网络。

码流仍是标准格式

30%

Netflix 观看量中 AV1 的占比（2025 年 12 月）

90%+

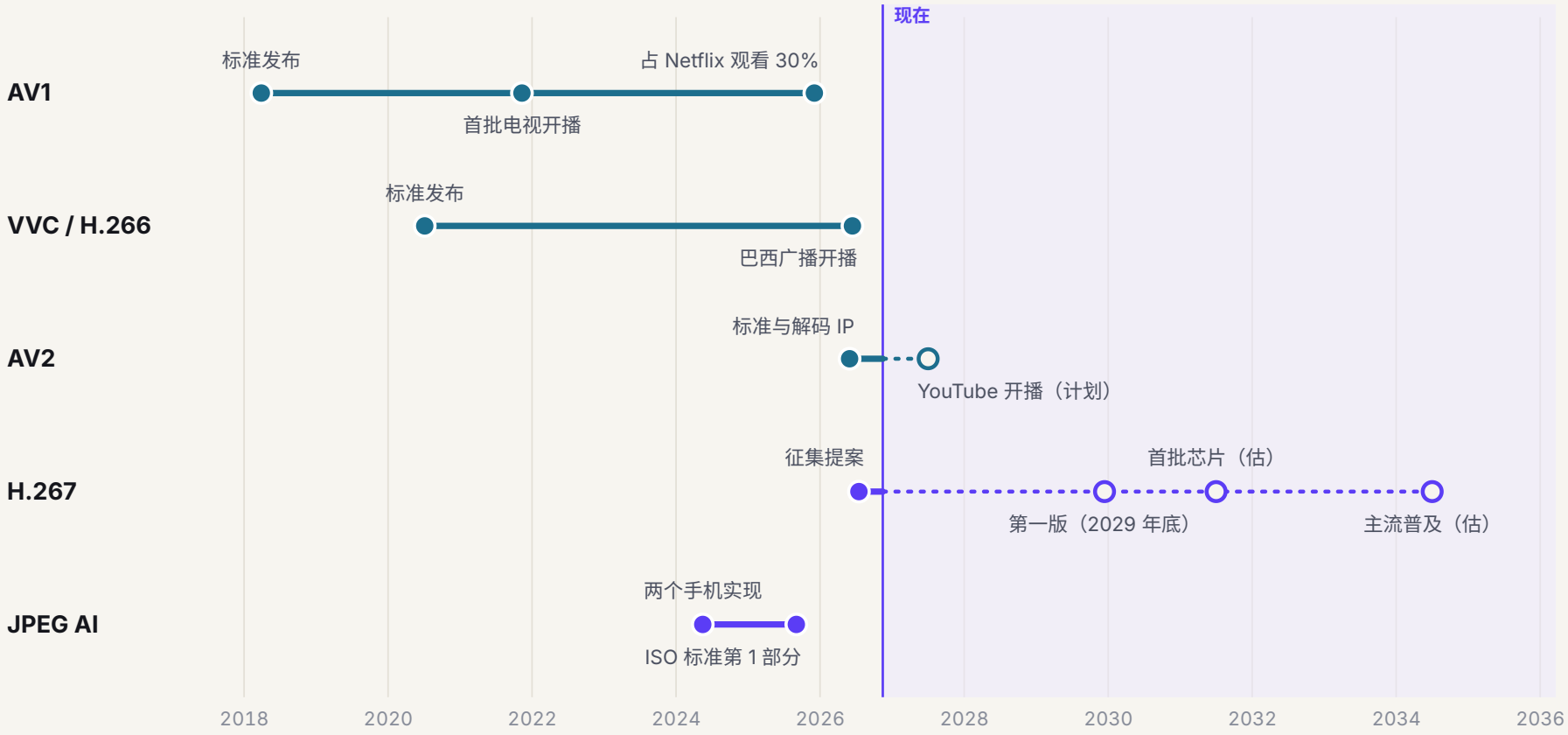
YouTube 内容库中已有 AV1 版本的比例，按观看时长加权（2026 年 3 月）

3 周

从 AV2 1.0.0 版（2026 年 5 月 28 日）到三家供应商宣布解码 IP 客户

时间表

从标准定稿到主流屏幕，要花掉大半个十年



AV1 从标准定稿（2018 年）到占 Netflix 观看量的 30%（2025 年 12 月），用了将近八年。空心标记为目标或我的估算。

H.267 决定神经工具能否进入每一块屏幕

2025 年 10 月
征集证据：来自 16 家公司和实验室的五份响应；神经方法在 8K HDR 上增益最明显

2026 年 7 月
在日内瓦发布征集提案

2027 年 1 月
在布里斯班评估提案

2027 年下半年
第一个测试模型

2029 年底
标准第一版

≈28%
传统 ECM 路线相对 VVC 节省的码率

6–14%
神经环路滤波节省的码率，每像素 1 万至 2.2 万次以上乘加

SEI
VVC 码流已经可以标示一个神经后处理滤波器，供解码端执行

我的判断：传统内核，加上少数节省的码率足以抵消解码代价的神经工具 分析。

相机里是神经网络，网页上是传统格式

- **相机**

  高通的 Cognitive ISP 自 2023 年起在手机上做实时分割；一款完全运行在 NPU 上的全 AI ISP 在 CES 2026 上亮相。

- **JPEG AI**

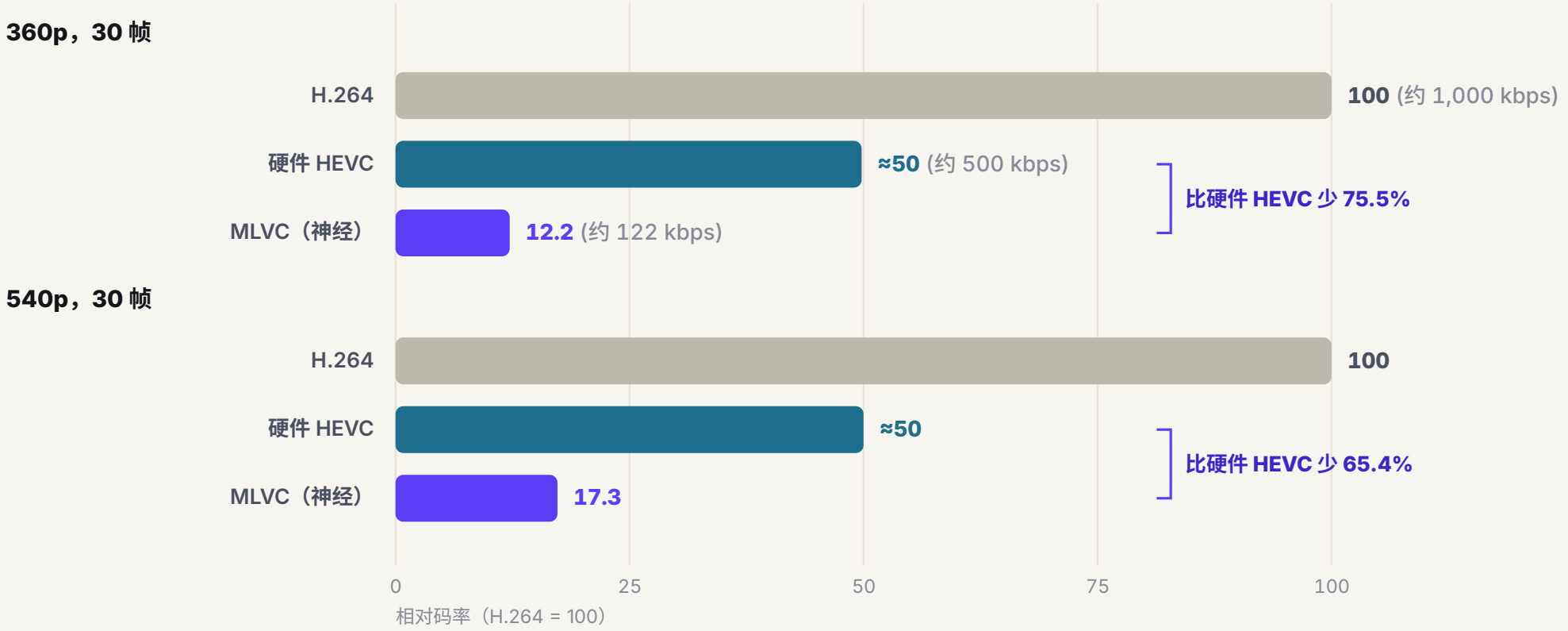
  2025 年发布：比 VVC 帧内节省 12.5–27.9%，每像素 7 千到 21.6 万次乘加。我没有找到任何默认使用它的产品。

- **网页**

  Chrome 155 于 2026 年 10 月 6 日默认启用 JPEG XL，解码器用 Rust 编写，内存安全。

> **必须到处都能打开的格式仍是传统的；神经编码在一家公司自己的流程里胜出。**

通话，是端到端神经编解码器落地的地方



Satin 自 2021 年起用于所有 Teams 和 Skype 双人通话。MLVC 自 2026 年 7 月 24 日起以 MIT 许可开源，正在 Teams 点对点通话中推广，最高 540p30，最多占用一半 NPU 厂商说法。

并非每个平台都转向神经编解码，原因很有启发

标准内部，可选

Opus 1.5 和 1.6

神经增强、丢包隐藏、深度冗余和带宽扩展，全部默认关闭，并与 2012 年的标准完全兼容。

主动选择传统方案

Meta MLow（2024）

Meta 超过 20% 的通话跑在 ARMv7 手机上，每天还有数千万次 WhatsApp 通话跑在用了十年以上的手机上。MLow 在 6 kbps 下音质约为 Opus 的两倍，计算量少 10%。

起决定作用的是存量设备的复杂度预算。神经编解码器最先落地在设备新、两端都由一家厂商掌控的地方。

规模最大、却没人称之为编解码器的部署

- **学习型编解码器无处不在**

  每个语音和视频模型都在读写一个学习型编解码器的 token 或潜变量。

- **不是分发格式**

  每个模型都有自己的潜空间：没有码率控制，没有一致性测试，没有第二个解码器。

- **一项成本**

  对提供生成式视频服务的公司来说，压缩率更高的潜变量能让生成成本超比例下降 分析。

两个例子

12.5 Hz · 1.1 kbps

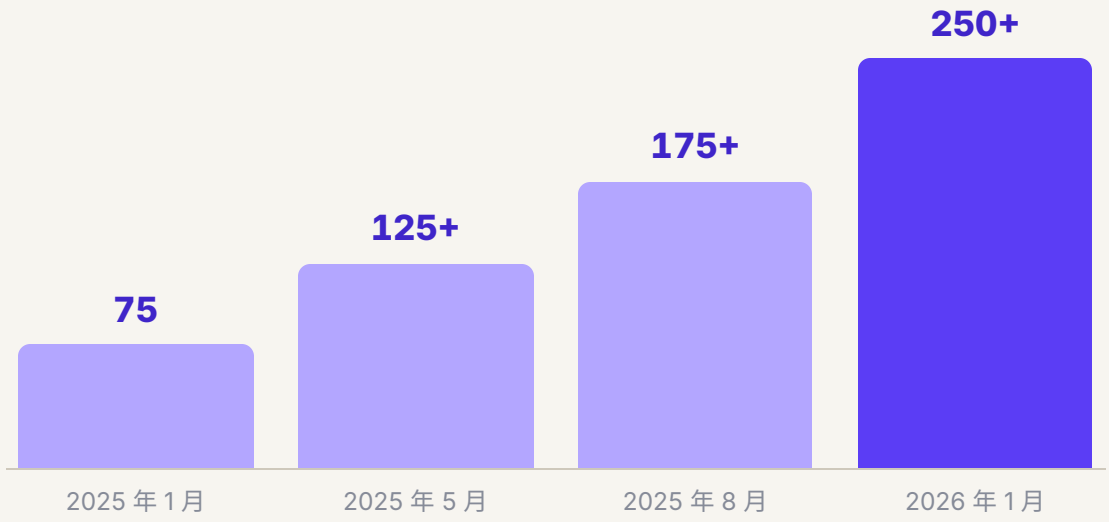
Kyutai Mimi 的语音 token，码率低于任何电话编解码器

1:192

LTX-Video 的视频 VAE：一个潜变量 token 代表 8 帧上 32×32 像素

神经重建已是主流，神经资源刚刚起步

DLSS 4 多帧生成：游戏和应用



神经纹理压缩

英伟达演示场景，GTC 2026

纹理显存 6.5 GB → 0.97 GB



《刺客信条：幻景》，已上市

只用于部分资源；对比块压缩格式



口径不同，不能直接比较

DirectX Linear Algebra 于 2026 年 4 月进入公开预览：在着色器内做神经解码，任何厂商的 GPU 都走同一条代码路径。

卫星、无人机和机器人：最契合，市场却小

- **为什么契合**

  一个所有者掌控两端，带宽稀缺，内容范围窄到可以专门训练。

- **中国**

  2026 年第 49 号公告自 11 月 18 日起施行，修订无人机相关出口管制；其中一个条目以语义图像压缩作为指标。

- **美国**

  FCC 于 2025 年 12 月 22 日把外国制造的无人机及其关键部件列入涵盖清单（Covered List）。

> **对有出口合规能力的专业公司来说，这是可靠的细分市场，而不是主流市场。**

每像素能耗高出一个数量级

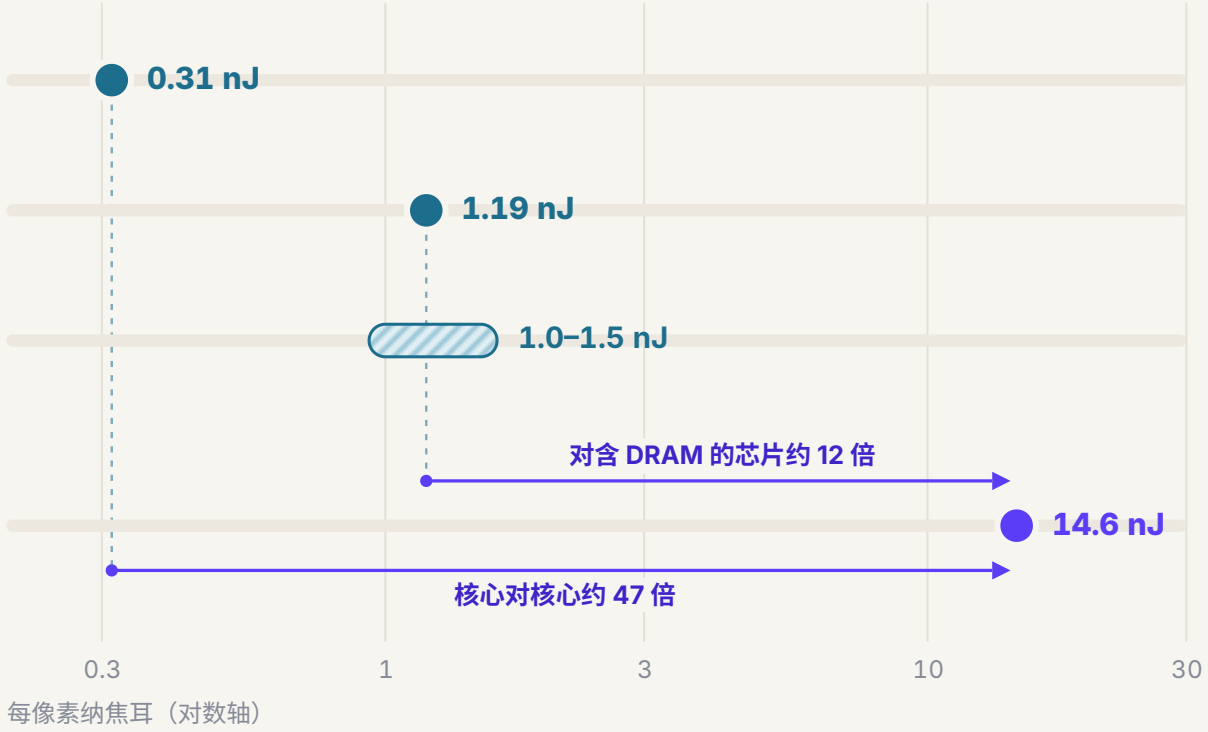
HEVC 解码芯片（核心）
40 nm，4K30，实测

HEVC 解码芯片（含 DRAM）
40 nm，4K30，实测核心加仿真 DRAM

VVC 解码器
28 nm，4K60，推算

神经视频解码器（NVCA）
28 nm，1080p25，仅综合；未计权重与 DRAM

按 1080p30 计，约为 74 mW 对 908 mW。



按 1080p30 计：含 DRAM 的 HEVC 芯片约 74 毫瓦，NVCA 约 900 毫瓦。HEVC 芯片用的是更老的工艺节点，NVCA 没有计入权重和 DRAM：这个比较对神经设计有利。

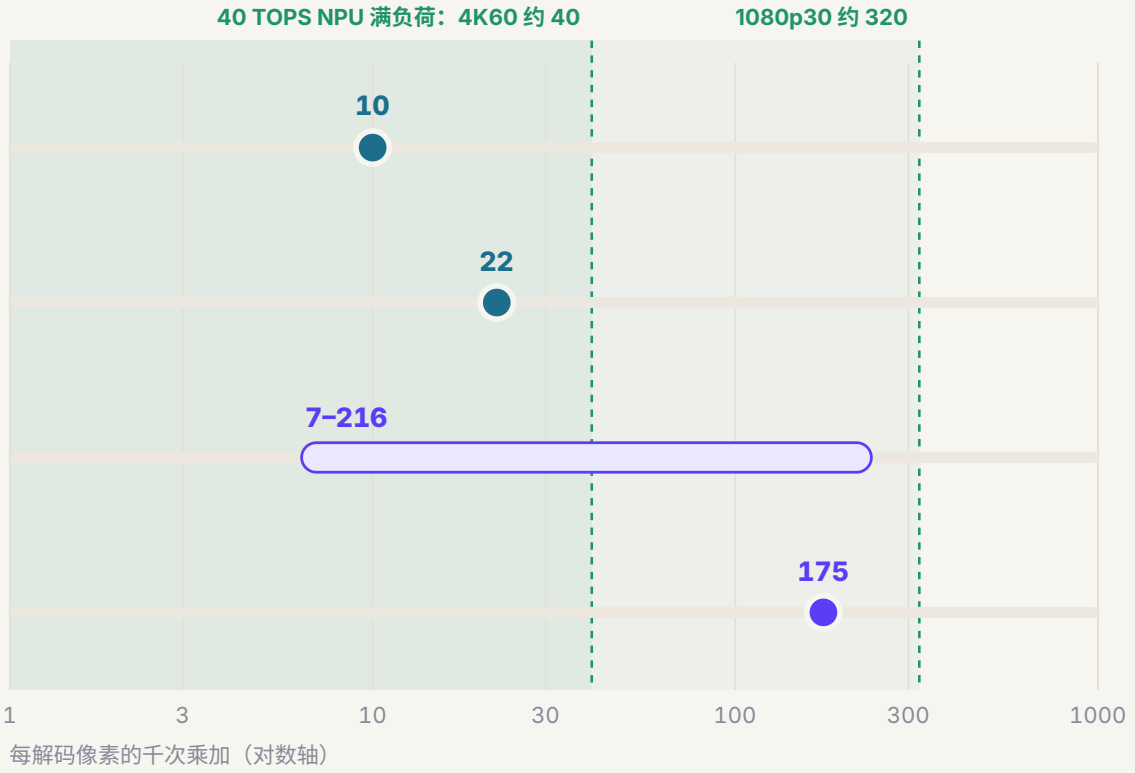
神经解码器对 NPU 的要求

神经环路滤波，极低复杂度
JVET 探索 · 约省 6% 码率
1080p30 需 1.2 TOPS · 4K60 需 10

神经环路滤波，低复杂度
JVET 探索 · 约 8%
1080p30 需 2.7 TOPS · 4K60 需 22

JPEG AI 解码器，基础到高配
比 VVC 帧内省 12.5–27.9%

DCVC-RT，完整的神经视频解码器
比 VVC 参考编码器省 21%（论文设置）
1080p30 需 22 TOPS · 4K60 需 174



DCVC-RT 每像素约需 17.5 万次乘加：按满利用率计，1080p30 约需 22 TOPS，4K60 约需 174 TOPS 估算。一颗 40 TOPS 的 NPU 满负荷运行，在 4K60 下每像素约能负担 4 万次。

访存流量和逐比特一致才是硬约束

DRAM 占大头

先看访存，再看乘加

在 MIT 的 HEVC 芯片上，DRAM 约占系统能耗的四分之三：每像素 1.19 纳焦中占 0.88。神经解码器要在帧与帧之间保留庞大的特征图；DCVC-RT 的作者发现，限制速度的是内存 I/O，而不是运算。

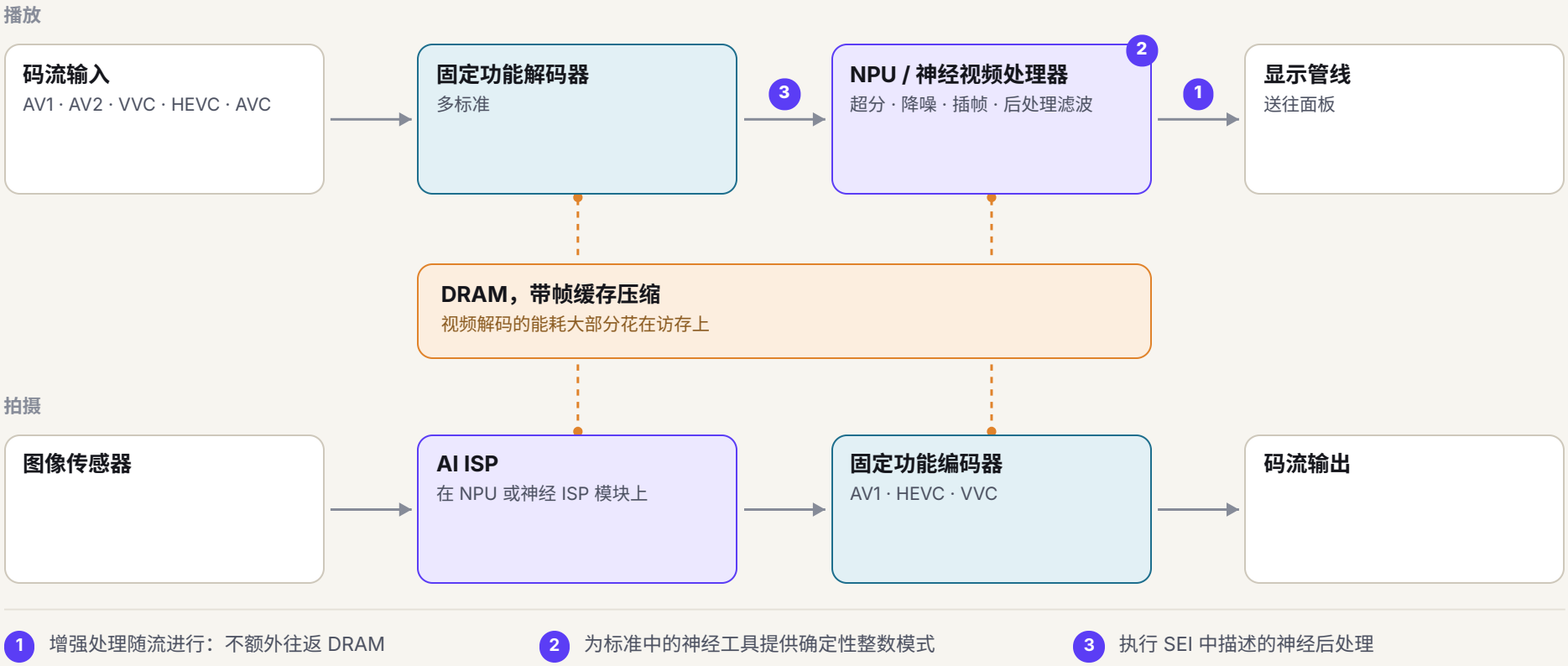
精确性

每个解码器都必须算出相同的比特

只要两台设备算出的概率不同，熵解码就会失败。解决办法：整数模型（DCVC-RT）、共享尺度的超先验（MLVC）、16 位整数推理（JVET 的 SADL）、一致性测试点（JPEG AI）。

对芯片团队来说，这又回到了 HEVC 整数变换那套规矩：标准必须精确定义运算。

已经落地的分工：固定功能编解码器加 NPU



腾讯沧海：第一代芯片已部署超过 10 万片；第二代加入 H.266 和片上超分。Chips&Media 和 Allegro DVT 同时销售编解码 IP 和神经视频 IP。

按角色看，我会做什么

SoC 和电视芯片团队

- 为每个必须支持的标准配固定功能解码器，现在就要有 AV2
- NPU 按显示分辨率下的增强处理定规模，不额外往返 DRAM
- 支持神经后处理滤波 SEI 和确定性整数模式
- H.267 在 2029 年或更晚才开始设计

IP 供应商

- 现在就做 AV2 和 VVC 解码器
- 后处理滤波加速器和帧缓存压缩
- 面向神经处理的一致性工具

视频平台

- 编码端的神经前处理最先回本
- 设备条件允许时，在客户端做超分
- 端到端神经编解码器只用在自己掌控的闭环里

神经工具在哪里变得可授权

标准	神经内容	状态（2026 年 11 月）
H.267（工作名）	未定：传统工具（ECM）和神经工具都是候选	2027 年 1 月评估提案；2029 年底第一版
VVC 和 HEVC，配合 VSEI	在 SEI 中描述神经后处理滤波；解码端可选	已生效；2025 年 10 月扩展
AV2	传统工具，免版税	2026 年 5 月 28 日发布 1.0.0 版；YouTube 计划 2027 年初推出网页端流
JPEG AI	端到端学习型图像编解码器	第 1 部分于 2025 年发布；未发现产品默认支持
IEEE 1857.11	为 NPU 设计的神经图像编码	2024 年 12 月发布
面向机器的视频编码	为机器分析而非人眼观看而编码	2026 年初进入国际标准草案阶段
Opus 机器学习扩展	可选，向后兼容	已进入 libopus 1.5 和 1.6

645

字节跳动持有的超越 VVC 专利族数（GreyB，2025）

0.50 · 1.00–1.50 美元

VVC 专利池费率：每部手机 · 每台 4K 电视，一类地区（约 3.4 · 6.7–10.1 元）

0 美元

AOMedia 成员就 AV1 和 AV2 收取的专利费

钱流向了谁

钱跟着标准和平台走

现在就赚钱

平台

留住节省下来的钱

Netflix 神经下采样；YouTube、Meta 超分；微软以 MIT 许可开源 MLVC

现在就赚钱

终端和芯片厂商

卖差异化

摄像头 ISP；电视超分画质芯片；腾讯自研沧海编解码芯片

现在就赚钱

IP 供应商

每个新标准都授权一次

AV2 定稿三周内，Chips&Media、Allegro DVT、芯原都宣布了 AV2 解码客户

现在就赚钱

GPU 厂商

构建平台锁定

DLSS 4 覆盖 250 多款游戏和应用；神经纹理 SDK 处于 beta

按台收费

专利权人和专利池

按台收费

VVC：每部手机 0.50 美元，每台 4K 电视 1.00–1.50 美元；H.267 工具 2029 年后

以被收购退出

神经编解码创业公司

被收购

WaveOne → 苹果（2023）；iSIZE → 索尼（2023）；Deep Render → InterDigital（2025）

> 最知名的三家神经编解码创业公司 WaveOne、iSIZE 和 Deep Render，分别被苹果、索尼和 InterDigital 收购。

公司可以在哪里切入

六个切入点，以及各自的风险

切入点	谁来买	为什么是现在	主要风险
编码端神经前处理	流媒体和用户视频平台、广播机构	设备无需改动；已在 Netflix 得到验证	大平台自己做
显示端增强 IP 和调校	电视、机顶盒和手机芯片厂商	设备里已经有 NPU	同质化；厂商自研团队
闭环神经编解码器	企业通信、工业和远程呈现厂商	MLVC 证明了可行，而且代码已开源	必须掌控两端；NPU 碎片化
编解码器加 NPU 的芯片或 IP	SoC 厂商	AV2 周期正在开始	在位者实力强
评测、一致性和确定性推理	所有交付神经工具的公司	逐比特一致是硬性要求	不捆绑销售则市场很小
向 H.267 贡献标准提案	专利许可方、大平台	提案将于 2027 年 1 月评估	周期长、成本高、结果不确定

你会听到的说法

一个说法讲清四件事之前，都要打折扣

说法	证据怎么说
“神经编解码器已经胜过 VVC。”	在实验室里是。DCVC-RT 比 VVC 参考编码器节省 21%；它的解码器每像素约需 17.5 万次乘加。
“AI 编解码器几年内就会取代 H.26x 和 AV1。”	在播放领域不会：H.267 第一版在 2029 年底，芯片约在 2031 年 估算。眼下到来的编解码器 AV2 是传统的。
“MLVC 把带宽削减 88%。”	这是在 360p 下相对 H.264。相对硬件 HEVC：360p 下 75.5%，540p 下 65.4%，基于微软自己的测试集。
“神经纹理把显存减少 85–96%。”	那是演示场景。已上市的游戏提升约 30%，而且只用在部分资源上。
“JPEG AI 是 ISO 标准，所以相机和浏览器都会采用。”	未发现任何产品默认支持。网页采用的是传统格式 JPEG XL。

问清比较基准、指标、分辨率和解码代价。四项讲不全的说法，是研究成果，不是产品。

值得关注的节点与开放问题

关键节点

2027 年 1 月	评估 H.267 提案：多少提案依赖神经工具，每像素代价多大
2027 年初	YouTube 首批 AV2 网页流；已发布的 SoC 中出现 AV2 解码器
2027 年下半年	H.267 第一个测试模型
2027 年	MLVC 走出 Teams 点对点通话
2027–2028 年	第一个默认使用 JPEG AI 的产品
2029 年底	H.267 第一版开启芯片周期

开放问题

01 大规模下的精确性：在不同厂商的 NPU 上做逐比特一致的整数解码，要损失多少压缩率？

02 先看访存，再看乘加：神经后处理每像素的 DRAM 流量

03 在 H.267 里，神经工具的每像素代价到多少就不再划算

04 除了通话，哪些闭环最先采用端到端神经编解码器

05 解码器生成细节时，保真度该怎么衡量

在这些领域，我们可以一起探讨，包括技术交流和技术合作。

i@tanshuai.com · tanshuai.com