研究筆記 · 2026 年 11 月

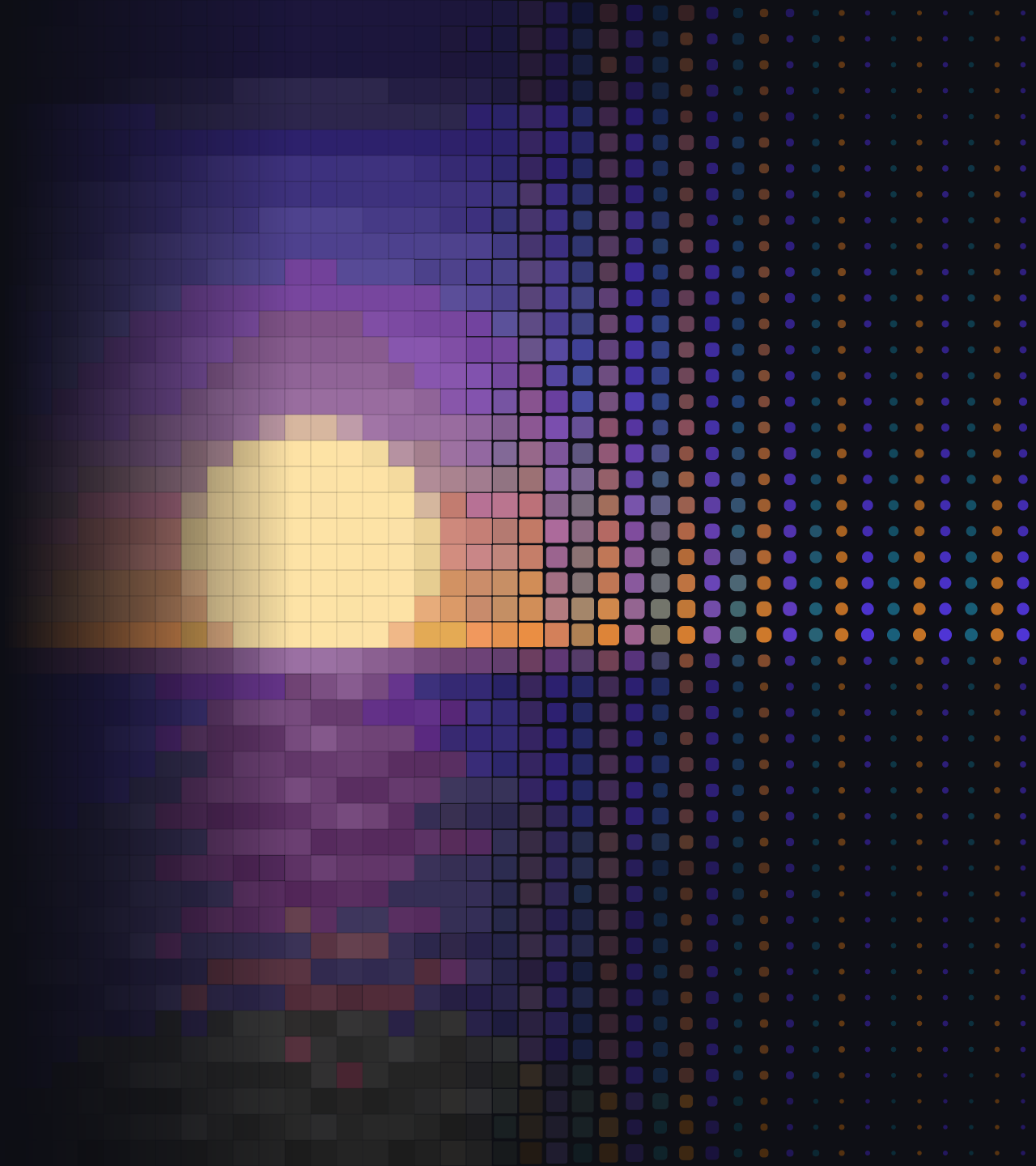
2026 年的神經編碼

在編解碼器周邊，而非取而代之

播放 · 攝影 · 通話 · 生成式 AI · 遊戲 · 晶片

譚帥

tanshuai.com · i@tanshuai.com



已是主流，但在編解碼器周邊

01 神經網路已是標準編解碼器兩端的主流

相機 ISP、編碼端降取樣與播放端超解析度。中間的位元串流仍是 AV1、HEVC、AVC 或 VVC，由固定功能硬體解碼。

02 端對端神經編解碼器只在封閉迴路中落地

通話、生成式 AI 的 tokenizer，以及一款 AAA 遊戲：由一家公司掌控兩端，並可退回標準編解碼器。

03 主流螢幕上的神經位元串流要等 H.267

第一版預定 2029 年底完成；首批晶片約在 2031 年，2030 年代中期才會成為主流 估算 。

≈75%

行動數據流量中影片的占比（2025 年底）

12.6 億

2025 年智慧型手機出貨量，全都以固定功能硬體解碼影片

≈12×

唯一公開的神經解碼晶片設計的每像素能耗，相對於硬體 HEVC 解碼器

神經網路進入編解碼器的三種方式



在編解碼器周邊，不需要任何人許可；在標準內部，需要一項標準；取代編解碼器，需要掌控兩端。

落地情況

2026 年神經編碼在哪裡落地

	標準編解碼器周邊 前處理、畫質增強、升頻	端對端神經編解碼器 學習式編碼器與解碼器	標準內部 標準位元串流或 API 中的神經工具
01 播放與串流	● 大規模部署 Netflix 降取樣器；YouTube 與 Meta 超解析度；電視 AI 升頻晶片	● 未部署 沒有串流服務傳送神經視訊格式；電視與手機以硬體解碼標準編解碼器	● 標準化中 H.267 第一版預定 2029 年底完成；神經工具名列候選
02 攝影與影像	● 大規模部署 旗艦手機的 AI ISP；CES 2026 展出以 IP 形式提供的全神經網路 ISP	● 展示與試行 JPEG AI 已是 ISO 標準：有手機展示，沒有產品	● 未部署 正在普及的是傳統格式：Chrome 155 預設啟用 JPEG XL
03 通話與會議	● 大規模部署 接收端超解析度與神經網路噪音抑制	● 大規模部署 Teams 雙人通話自 2021 年起使用神經音訊；MLVC 視訊正在 Teams 推出	● 有限部署 Opus 1.5 加入可選用、向下相容的機器學習工具
04 生成式 AI	—	● 大規模部署 每個影片與語音模型都執行學習式 tokenizer，例如壓縮比 1:192 的影片 VAE	—
05 遊戲與繪圖	● 大規模部署 神經網路升頻與畫格生成：光是 DLSS 4 就支援 250 多款遊戲與應用程式	● 有限部署 一款 AAA 遊戲已採用神經紋理；輝達的 SDK 仍為測試版	● 標準化中 DirectX 正為著色器加入矩陣運算（預覽版）
06 專用鏈路：衛星、機器人	● 展示與試行 在伺服器端修復低位元率畫面	● 展示與試行 抗遺失神經編解碼器處於研究與試行階段	—

各列依各領域觸及的人數排序。

兩端是神經網路，中間是標準

- **在編碼端**

  Netflix 的神經網路降取樣器自 2022 年起運作，任何裝置都不需改動；在 Netflix 的測試中，約 77% 的觀眾偏好它的結果 廠商說法 。

- **在螢幕端**

  YouTube 自 2025 年 10 月起在電視上提供超解析度；Meta 在伺服器與手機上都執行超解析度。

- **在電視內**

  三星 NQ8 AI Gen3 處理器標榜以 768 個神經網路進行升頻。

位元串流仍是標準格式

30%

Netflix 觀看量中 AV1 的占比（2025 年 12 月）

90%+

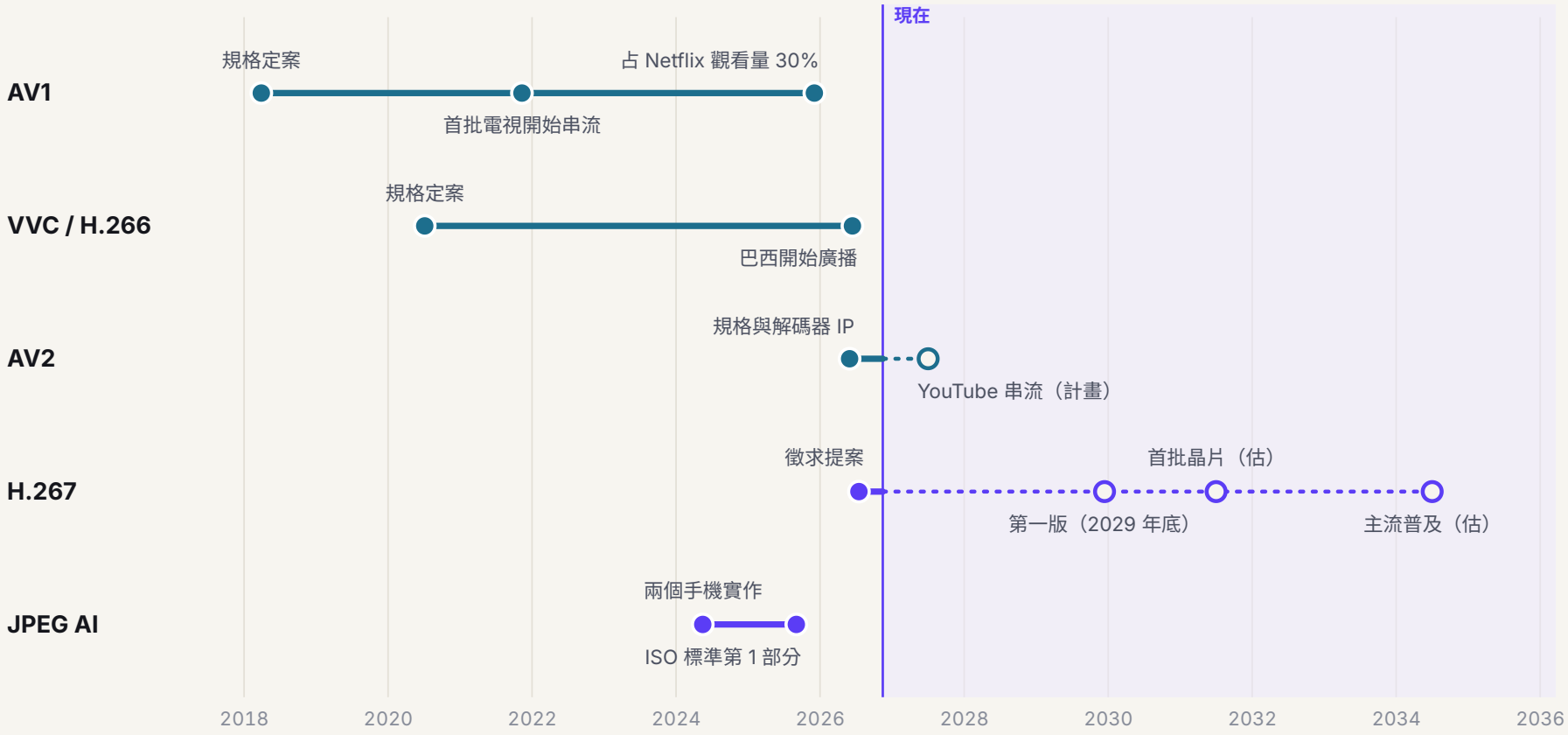
YouTube 內容庫中已提供 AV1 的比例，按觀看時數計（2026 年 3 月）

3 週

從 AV2 1.0.0 版（2026 年 5 月 28 日）到三家廠商都有了解碼器 IP 客戶

時程

從標準定案到主流螢幕，要花上大半個十年



AV1 從規格定案（2018）到占 Netflix 觀看量的 30%（2025 年 12 月），花了將近八年。空心標記為目標或我的估算。

H.267 決定神經工具能否走進每一塊螢幕

2025 年 10 月
徵求證據：來自 16 家公司與實驗室的五份回應；神經網路的增益在 8K HDR 上最明顯

2026 年 7 月
在日內瓦發布徵求提案

2027 年 1 月
在布里斯本評估提案回應

2027 年下半年
第一個測試模型

2029 年底
標準第一版

≈28%
傳統 ECM 路線相對 VVC 的節省幅度

6–14%
神經網路迴路濾波器在每像素 1 萬到 2.2 萬次以上 MAC 下的節省幅度

SEI
VVC 串流已可標示神經後處理濾波器，供解碼器套用

我的解讀：以傳統工具為核心，只在節省的位元值得付出解碼成本之處加入少數神經工具 分析 。

相機內是神經網路，網頁上是傳統格式

- **相機**

  高通的 Cognitive ISP 自 2023 年起在手機上執行即時分割；CES 2026 展出了一套在 NPU 上執行的完整 AI ISP。

- **JPEG AI**

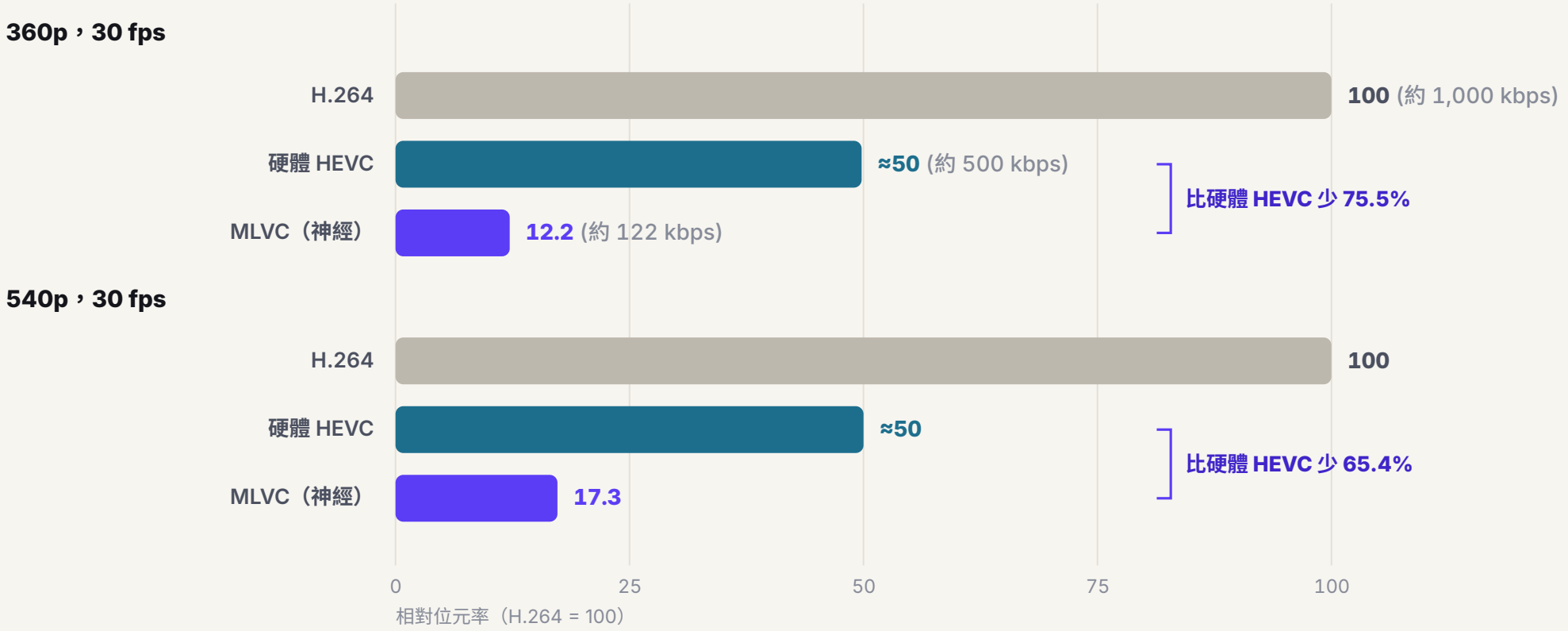
  2025 年發布：在每像素 7 千到 21.6 萬次 MAC 下，比 VVC 畫面內編碼省 12.5–27.9%。我沒有找到任何預設採用它的產品。

- **網頁**

  Chrome 155 於 2026 年 10 月 6 日預設啟用 JPEG XL，採用以 Rust 撰寫、記憶體安全的解碼器。

> **必須在任何地方都能開啟的格式仍是傳統格式；神經編碼在單一公司的處理流程內勝出。**

端對端神經編解碼器在通話中落地



Satin 自 2021 年起用於所有 Teams 與 Skype 雙人通話。MLVC 自 2026 年 7 月 24 日起以 MIT 授權開源，正在 Teams 點對點通話中推出，最高 540p30，最多占用一半的 NPU 廠商說法。

並非每個平台都改用神經網路，原因值得借鏡

標準內部，可選用

Opus 1.5 與 1.6

神經網路增強、封包遺失隱藏、深度冗餘與頻寬擴展，全都預設關閉，並與 2012 年的標準完全相容。

刻意選擇傳統做法

Meta MLow（2024）

Meta 超過 20% 的通話在 ARMv7 手機上進行，每天還有數千萬通 WhatsApp 通話在用了十年以上的手機上進行。MLow 在 6 kbps 下的品質約為 Opus 的兩倍，運算量少 10%。

決定因素是既有裝置能承受的複雜度預算。神經編解碼器會先在裝置較新、且由單一廠商掌控兩端的地方落地。

規模最大、卻沒人稱之為編解碼器的部署

- **學習式編解碼器無所不在**

  每個語音與影片模型都在讀寫學習式編解碼器的 token 或潛在表示。

- **不是發行格式**

  每個模型都有自己的潛在空間：沒有位元率控制、沒有一致性測試，也沒有第二個解碼器。

- **一項成本**

  對提供生成式影片服務的業者而言，壓縮率更高的潛在表示，能讓生成成本以超過等比例的幅度下降 分析。

兩個例子

12.5 Hz · 1.1 kbps

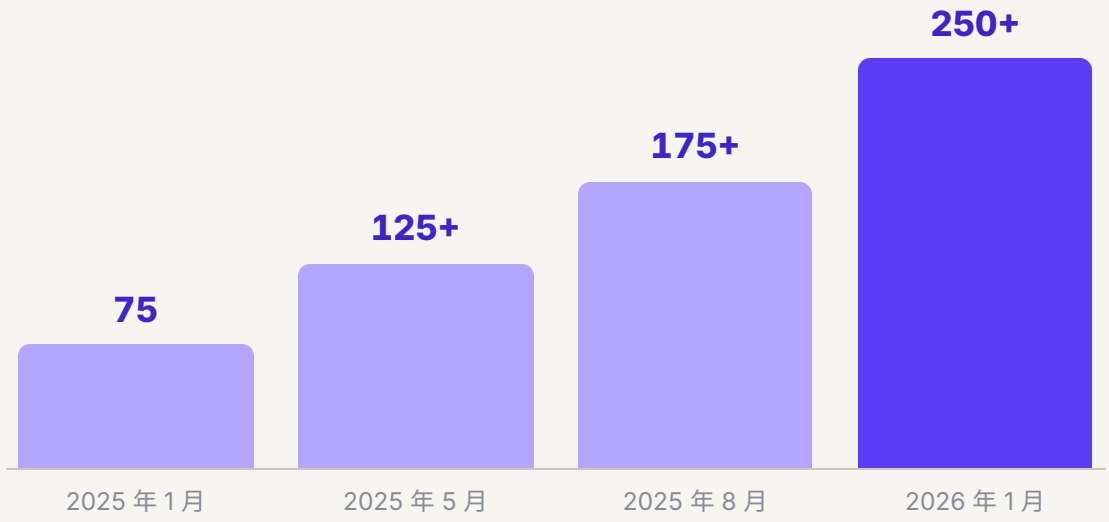
Kyutai Mimi 的語音 token，低於任何電話語音編解碼器

1:192

LTX-Video 的影片 VAE：一個潛在 token 代表 8 個影格中的 32×32 像素

神經重建已是主流，神經素材剛起步

DLSS 4 多畫格生成：遊戲與應用程式



神經紋理壓縮

輝達展示場景，GTC 2026

紋理記憶體 6.5 GB → 0.97 GB

記憶體少 85%

Assassin's Creed Mirage，已上市

僅用於部分素材；相對區塊壓縮格式

壓縮率約高 30%

衡量方式不同，無法直接比較

DirectX Linear Algebra 於 2026 年 4 月進入公開預覽：在任何廠商的 GPU 上，以同一套程式路徑在著色器內做神經解碼。

衛星、無人機與機器人：最契合，但市場小

- **為何契合**

  一個擁有者掌控兩端，頻寬稀缺，內容範圍又窄到足以針對性訓練。

- **中國**

  2026 年第 49 號公告於 11 月 18 日生效，修訂無人機出口管制；其中一項以語意影像壓縮作為門檻指標。

- **美國**

  FCC 於 2025 年 12 月 22 日將外國製無人機及其關鍵元件列入其「涵蓋清單」。

> **對具備出口合規能力的專業廠商是穩健的利基市場，但不是主流市場。**

每像素能耗高出一個數量級

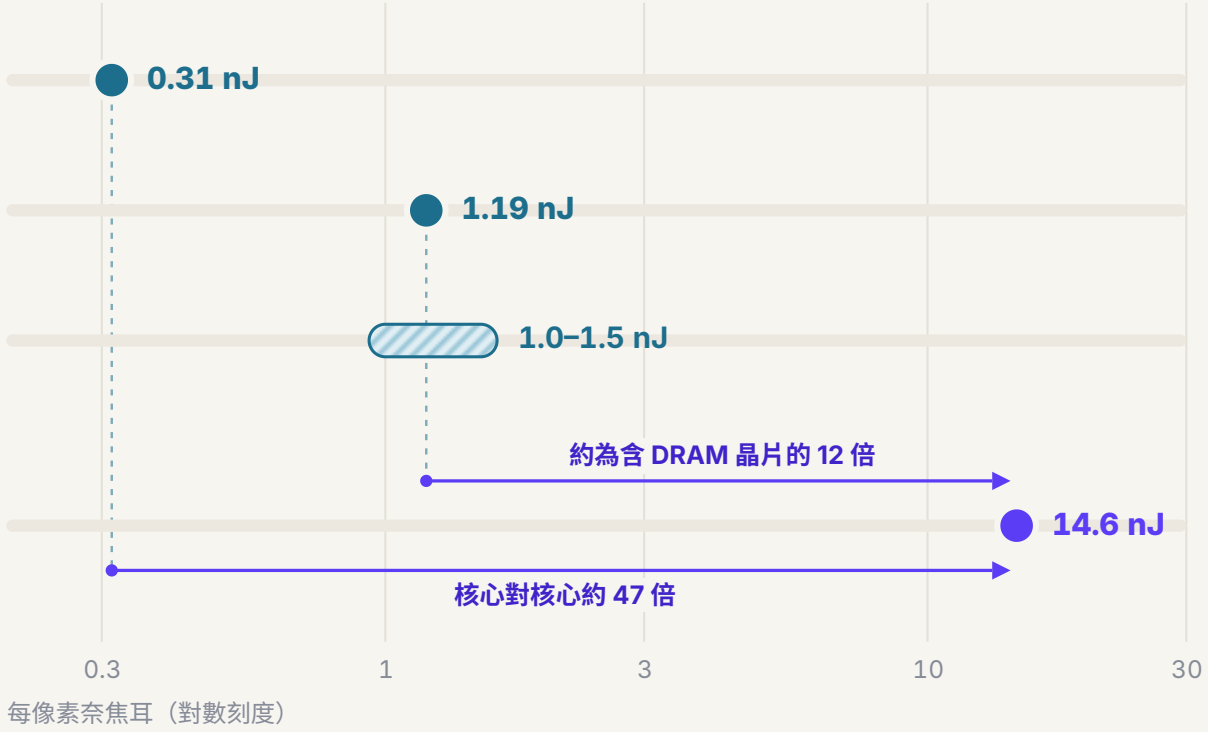
HEVC 解碼晶片（核心）
40 nm，4K30，實測

HEVC 解碼晶片（含 DRAM）
40 nm，4K30，實測核心加模擬 DRAM

VVC 解碼器
28 nm，4K60，推估

神經視訊解碼器（NVCA）
28 nm，1080p25，僅合成；未計權重與 DRAM

以 1080p30 計，約為 74 mW 對 908 mW。



以 1080p30 計：含 DRAM 的 HEVC 晶片約 74 mW，NVCA 約 900 mW。HEVC 晶片製程較舊，NVCA 又未計權重與 DRAM：這個比較對神經設計有利。

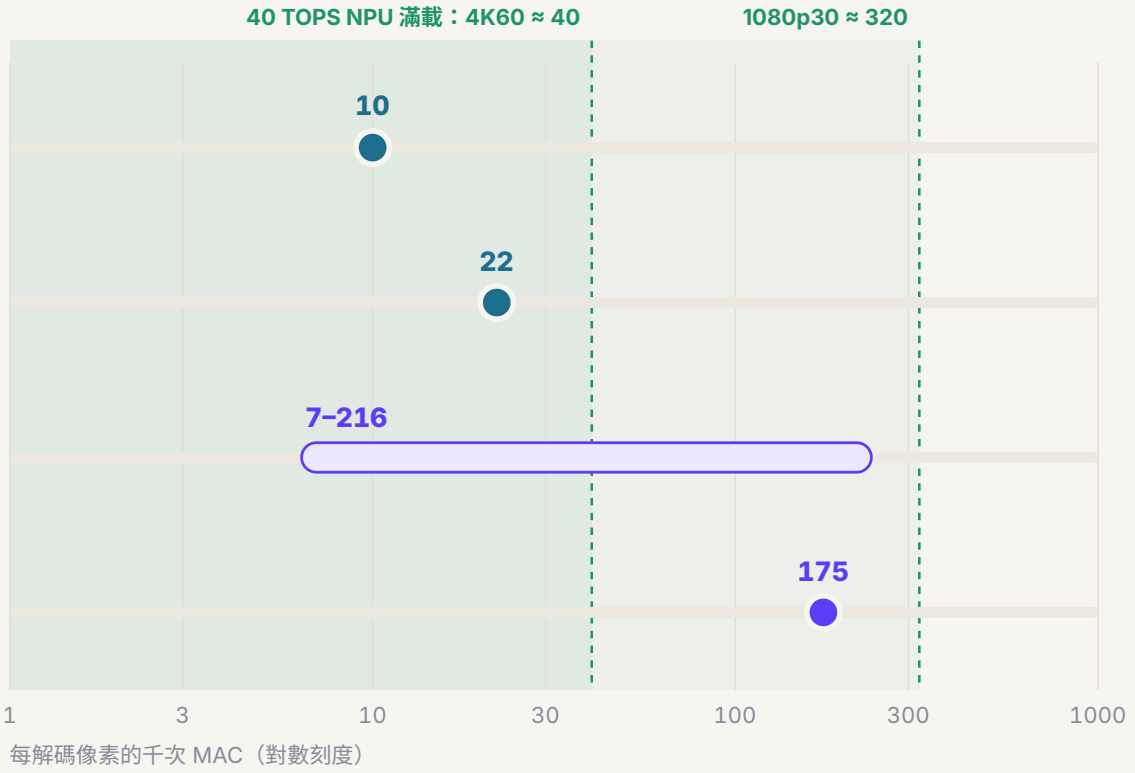
神經解碼器對 NPU 的要求

神經網路迴路濾波器，極低複雜度
JVET 探索 · 約省 6% 位元
1080p30 需 1.2 TOPS · 4K60 需 10

神經網路迴路濾波器，低複雜度
JVET 探索 · 約 8%
1080p30 需 2.7 TOPS · 4K60 需 22

JPEG AI 解碼器，基本到高階
比 VVC 畫面內編碼省 12.5–27.9%

DCVC-RT，完整的神經視訊解碼器
比 VVC 參考編碼器省 21%（論文設定）
1080p30 需 22 TOPS · 4K60 需 174



DCVC-RT 每像素約需 17.5 萬次 MAC：在完全利用的前提下，1080p30 約需 22 TOPS，4K60 約需 174 TOPS 估算。一顆滿載的 40 TOPS NPU 在 4K60 下每像素約可負擔 4 萬次。

記憶體存取流量與逐位元一致，才是硬限制

DRAM 占大宗

記憶體先於 MAC

在 MIT 的 HEVC 晶片中，DRAM 約占系統能耗的四分之三：每像素 1.19 nJ 中的 0.88 nJ。神經解碼器要在影格之間保留龐大的特徵圖；DCVC-RT 的作者發現，限制速度的是記憶體 I/O，而不是算術運算。

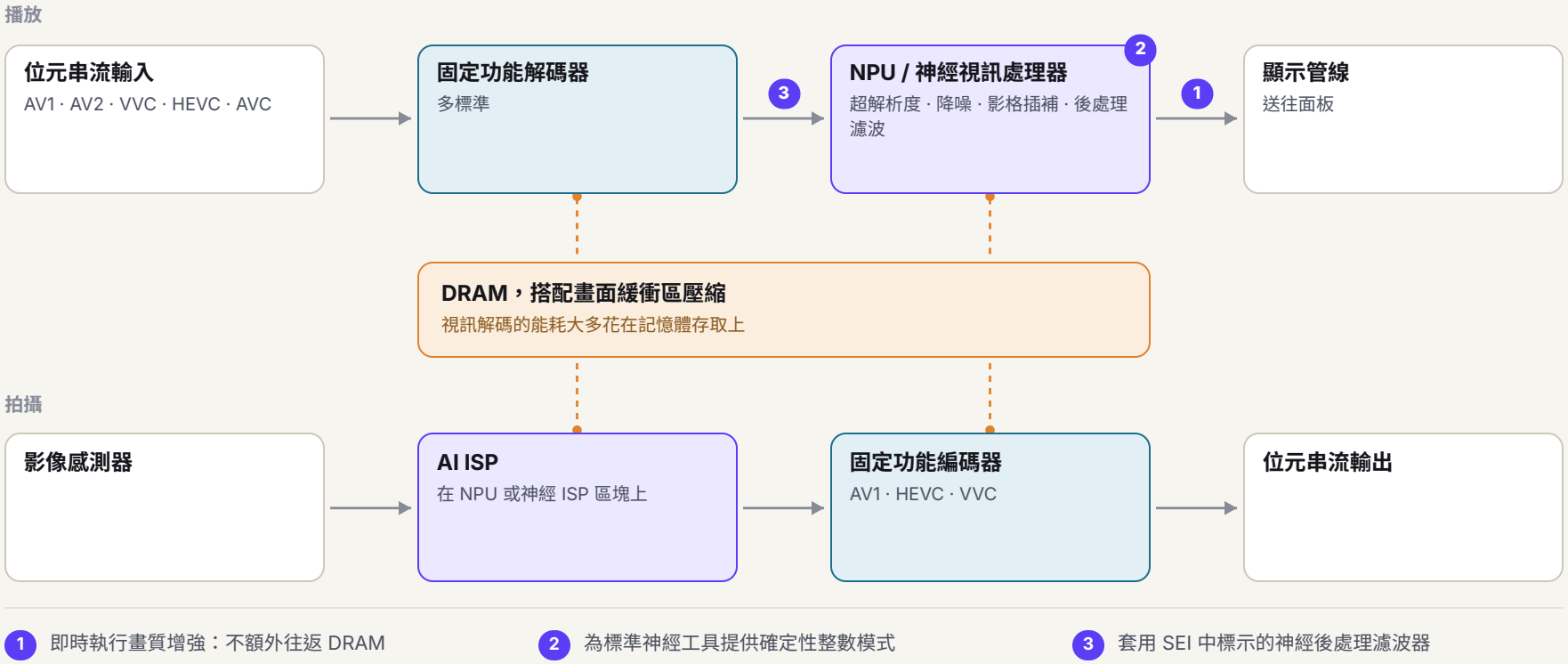
精確性

每個解碼器都必須算出相同的位元

兩台裝置只要算出的機率不同，熵解碼就會失敗。解法：整數模型（DCVC-RT）、共享尺度的超先驗（MLVC）、16 位元整數推論（JVET 的 SADL）、一致性檢查點（JPEG AI）。

對晶片團隊來說，這又是 HEVC 整數轉換的那套紀律：標準必須精確定義算術運算。

已落地的分工：固定功能編解碼器加上 NPU



> 騰訊的 Canghai：第一代晶片已部署超過 10 萬顆；第二代加入 H.266 與晶片內超解析度。Chips&Media 與 Allegro DVT 同時銷售編解碼器 IP 與神經視訊 IP。

依角色來看，我會做什麼

SoC 與電視晶片團隊

- 為每個必須支援播放的標準配置固定功能解碼器，現在就加入 AV2
- NPU 依顯示解析度下的畫質增強來定規模，且不額外往返 DRAM
- 支援神經後處理濾波器 SEI 與確定性整數模式
- 把 H.267 當作 2029 年或更晚才啟動的設計

IP 廠商

- 現在就做 AV2 與 VVC 解碼器
- 後處理濾波加速器與畫面緩衝區壓縮
- 神經處理的一致性工具

影音平台

- 編碼端的神經前處理最先回本
- 裝置條件允許時，採用用戶端超解析度
- 端對端神經編解碼器只用在你掌控的封閉迴路中

神經工具在哪裡成為可授權的技術

標準	神經內容	2026 年 11 月狀態
H.267（暫定名稱）	未定：傳統（ECM）與神經工具都是候選	2027 年 1 月評估提案回應；2029 年底推出第一版
VVC 與 HEVC，搭配 VSEI	以 SEI 描述神經後處理濾波器；解碼器可選擇是否套用	現行有效；2025 年 10 月擴充
AV2	傳統工具，免權利金	2026 年 5 月 28 日推出 1.0.0 版；YouTube 網頁串流預計 2027 年初推出
JPEG AI	端對端學習式影像編解碼器	第 1 部分於 2025 年發布；未發現產品預設支援
IEEE 1857.11	為 NPU 設計的神經影像編碼	2024 年 12 月發布
Video Coding for Machines	為機器分析而非人眼觀看而編碼	2026 年初進入國際標準草案階段
Opus 機器學習擴充	可選用且向下相容	已納入 libopus 1.5 與 1.6

645

字節跳動持有的 VVC 後繼標準專利家族（GreyB，2025）

0.50 · 1.00–1.50 美元

VVC 專利池每支手機與每台 4K 電視的費率，Region 1

0 美元

AOMedia 成員對 AV1 與 AV2 收取的權利金

錢跟著標準與平台走

已有收入

平台

留住省下的錢

Netflix 神經網路降取樣器；YouTube 與 Meta 超解析度；微軟以 MIT 授權開源 MLVC

已有收入

裝置與晶片廠商

銷售差異化

相機 ISP；電視升頻處理器；騰訊自研的 Canghai 編解碼晶片

已有收入

IP 廠商

每個新標準都授權一次

規格定案三週內，Chips&Media、Allegro DVT 與 VeriSilicon 都有了 AV2 解碼器客戶

已有收入

GPU 廠商

建立平台綁定

DLSS 4 支援 250 多款遊戲與應用程式；神經紋理 SDK 仍為測試版

按裝置收費

專利權人與專利池

按裝置收費

VVC：每支手機 0.50 美元，每台 4K 電視 1.00–1.50 美元；H.267 工具在 2029 年後

被併購出場

神經編解碼新創公司

被收購

WaveOne → 蘋果（2023）；iSIZE → Sony（2023）；Deep Render → InterDigital（2025）

最知名的三家神經編解碼新創公司 WaveOne、iSIZE 與 Deep Render，分別被蘋果、Sony 與 InterDigital 收購。

六種切入方式，以及各自的風險

切入點	買方	為何是現在	主要風險
編碼端的神經前處理	串流與使用者影音平台、廣播業者	裝置不需改動；已在 Netflix 驗證	大平台會自己做
顯示端畫質增強 IP 與調校	電視、機上盒與手機晶片廠商	裝置內已有 NPU	商品化；廠商自建團隊
封閉迴路神經編解碼器	企業通訊、工業與遠端臨場廠商	MLVC 證明可行，且程式碼開源	必須掌控兩端；NPU 碎片化
編解碼器加 NPU 的晶片或 IP	SoC 廠商	AV2 週期正要開始	既有大廠實力強
評測、一致性與確定性推論	所有出貨神經工具的業者	逐位元一致是硬性要求	除非搭售，否則市場小
向 H.267 提交標準提案	授權方、大平台	2027 年 1 月評估提案回應	週期長、成本高、結果不確定

你會聽到的說法

說法交代清楚四件事之前，都要打折扣

說法	證據怎麼說
「神經編解碼器已經勝過 VVC。」	在實驗室裡是。DCVC-RT 比 VVC 參考編碼器省 21%；其解碼器每像素約需 17.5 萬次 MAC。
「AI 編解碼器幾年內就會取代 H.26x 與 AV1。」	播放領域不會：H.267 第一版在 2029 年底，晶片約在 2031 年 估算。現在到來的編解碼器 AV2 是傳統設計。
「MLVC 讓頻寬減少 88%。」	那是在 360p 下相對 H.264。相對硬體 HEVC：360p 為 75.5%，540p 為 65.4%，用的是微軟自己的測試集。
「神經紋理讓顯示記憶體減少 85–96%。」	那是展示場景。已上市的遊戲約提升 30%，而且只用在部分素材上。
「JPEG AI 是 ISO 標準，所以相機和瀏覽器都會採用。」	未發現產品預設支援。網頁採用的是傳統格式 JPEG XL。

問清楚比較基準、指標、解析度與解碼成本。交代不出這四項的說法，是研究成果，不是產品。

接下來看什麼

觀察重點與未解問題

觀察指標

2027 年 1 月	評估 H.267 提案回應：有多少仰賴神經工具，每像素成本多少
2027 年初	YouTube 首批 AV2 網頁串流；已發表的 SoC 內建 AV2 解碼器
2027 年下半年	H.267 的第一個測試模型
2027 年	MLVC 走出 Teams 點對點通話
2027–2028 年	第一個預設採用 JPEG AI 的產品
2029 年底	H.267 第一版啟動晶片週期

未解問題

01 大規模下的精確性：在各廠商的 NPU 上做到逐位元一致的整數解碼，要犧牲多少壓縮率？

02 記憶體先於 MAC：神經後處理每像素的 DRAM 存取流量

03 在 H.267 中，神經工具的每像素成本到多少就不再划算

04 通話之外，哪些封閉迴路會最先採用端對端神經編解碼器

05 解碼器會自行生成細節時，該如何衡量保真度

這些領域，我們可以透過技術交流與技術合作一起探索。

i@tanshuai.com · tanshuai.com