智能图像视频编码的未来发展之路

李　礼[1]　　贾川民[2]
[1] 中国科学技术大学　　[2] 北京大学

图像视频编码是信息产业的基础性和前瞻性技术，它不仅支撑着数字媒体产业的蓬勃发展，更是推动新一代信息技术革命的关键使能技术。最新的视频编码标准 H.266/VVC 能够实现 400~600 倍的压缩效率，这不仅使 8K 超高清视频的实时传输成为可能，更能为全球视频传输行业节约巨大的带宽和存储成本。此外，在深空、深海等特殊应用场景中，高效图像视频编码技术也颇为重要。例如，在深空探测领域，火星探测器与地球之间的通信带宽极其有限，通常只有几十比特每秒。通过先进的图像视频编码技术，科学家们能够在有限的带宽条件下传输火星表面图像和视频数据，为人类探索火星提供关键数据支持。同样，在深海探测中，受限于水下通信的特殊环境，高效的图像视频编码技术成为实现深海科考数据传输的核心保障。

经典图像视频编码技术基于混合编码框架，这一框架自 20 世纪 80 年代提出以来，经过数十年的发展与完善，成为视频压缩标准的基石。其核心模块包括预测、变换、量化和熵编码，每个模块都在提升压缩效率过程中发挥着独特的作用。预测模块通过时域或空域相关性减少冗余信息；变换模块，如离散余弦变换（discrete cosine transform, DCT），将空域信号转换到频域以集中能量；量化模块在保证视觉质量的前提下降低数据精度；而熵编码模块则利用统计特性进一步压缩数据。基于这一框架，国际标准组推出了一系列影响深远的视频编码标准，如图 1 所示。从早期的 H.261 起步，到 MPEG-2 推动数字电视革命，再到 H.264/AVC[1] 成为互联网视频的基石，以及 H.265/HEVC[2] 支持 4K 超高清视频，直至最新的 H.266/VVC[3] 实现 8K 超高清视频的高效压缩，每一代标准都在前代基础上实现了相同主观质量下约 50% 的码率节省。我国数字音视频编解码技术标准（audio video coding standard, AVS）工作组相关工作的起步比国际标准晚了 10 年左右，但经过 20 余年的发展，目前已经达到国际先进水平[4]。

然而，随着经典编码框架的日益成熟，其发展正面临瓶颈：单个编码工具的性能提升已进入“边际效应递减”阶段。自 H.265/HEVC 标准发布以来，单个工具的优化往往只能带来不足 1% 的性能提升，而付出的计算复杂度代价却呈指数级增长。与此同时，深度神经网络（deep neural networks, DNN）作为一种革命性的方法论，为多媒体领域带来了变革性发展，其不仅能够自动学习图像视频中的复杂统计特性，还可以建立从像素空间到特征空间的高效映射。这也给图像视频编码领域带来了新的机遇。

基于深度学习的智能图像视频编码技术的高速发展始于 2015 年，在近十年的发展中诞生了“混合智能编码”以及“端到端智能编码”两种范式。混合智能编码范式强调深度学习技术与经典混合编码框架的深度融合，通过在经典编码框架中引入神经网络模块，在保持混合编码基本框架的同时提升编码效率。典型方法包括基于深度学习的环路滤波[5]、基于深度学习的帧内预测[6] 和帧间预测[7] 等。这种范式的优势在于能够充分利用经典编码框架的成熟性和神经网络的强大学习

DOI：10.11991/cccf.202511012
基金项目：国家自然科学基金委面上项目（62171429）
通信作者：李礼，E-mail: lil1@ustc.edu.cn

©《计算》编辑部版权所有

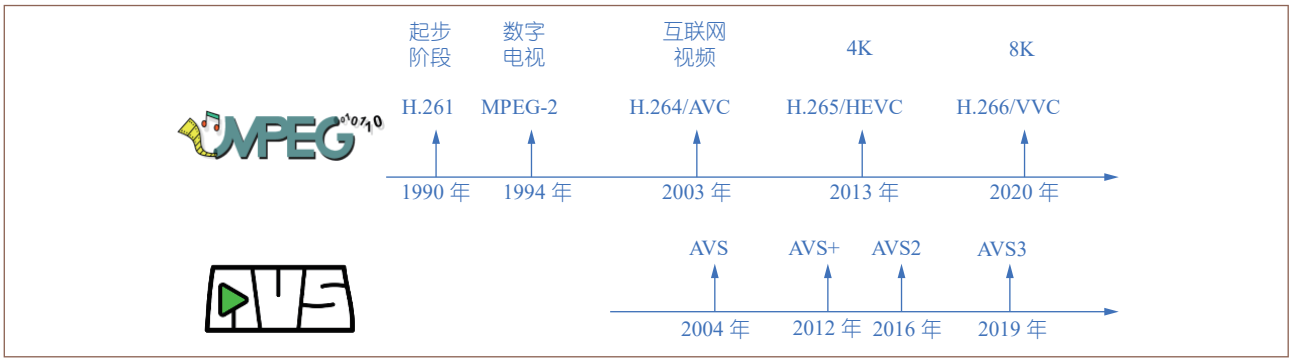


图 1　基于混合编码框架的国际国家经典视频编码标准发展历程

能力，有望基于现有硬件框架实现广泛应用。端到端智能编码范式则完全采用基于神经网络的全新架构，实现了从像素到比特流的端到端优化。这种方法突破了混合编码框架局部优化的局限，能够全局优化率失真性能。代表性工作如 Ballé等[8] 提出的变分自编码器框架以及 Li 等[9] 提出的条件视频编码框架等。这种范式的优势除了全局优化外，还在于其对新兴场景的适配性，如面向机器视觉的图像视频压缩[10] 等。

总体来说，基于深度学习的智能图像视频编码技术在过去十年取得了突破性进展，压缩效率显著提升，应用场景不断拓展。然而，学术界对智能编码的技术本质仍存在诸多争议。智能编码与经典编码的本质区别究竟在于何处？是仅仅体现在更高的压缩效率上，还是从根本上改变了编码的范式？智能技术为编码赋予了怎样的新内涵——是更强的语义理解能力、更灵活的场景适应性，还是更全局的率失真优化？反过来，编码技术又为智能技术带来了哪些新的可能性——是更智能的系统，更准确的评价指标，还是更紧致的训练数据？这些根本性问题不仅关系到智能编码的理论发展，也直接影响着其实际应用的方向和路径。本文旨在深入探讨智能图像视频编码技术的本质和未来发展路径，为智能图像视频编码技术的大规模应用奠定理论基础和路线指导，推动该技术从实验室走向产业化，从专业应用走向普及化，最终实现技术经济价值与社会效益的最大化。

智能图像视频编码的技术本质

为了探讨智能图像视频编码的技术本质，本研究从智能与编码的相互关系出发，首先分析智能为编码技术带来的变革性影响，其次探讨编码对智能技术的反哺作用，如图 2所示。

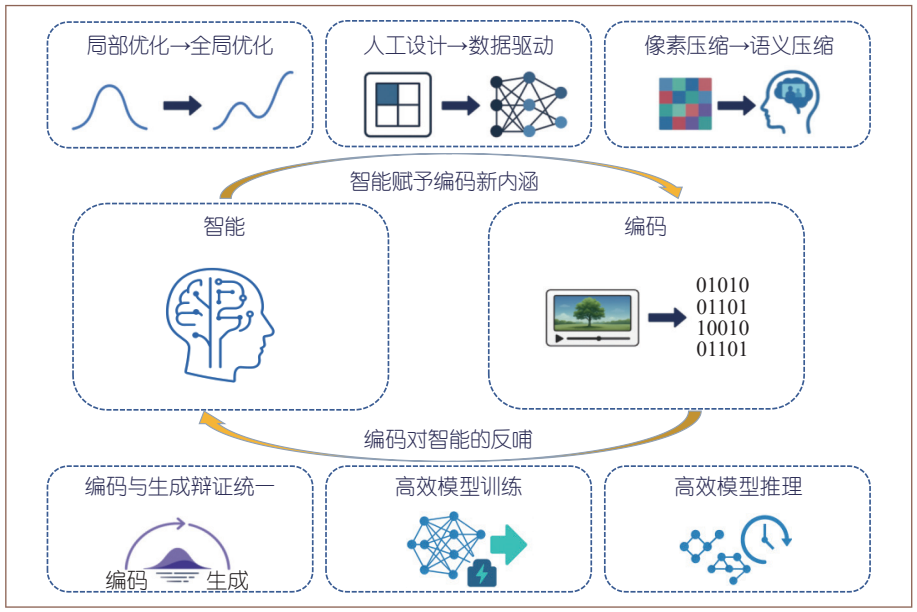


图2　智能与编码的相互关系示意

智能赋予编码的新内涵

智能技术的引入为图像视频编码带来了全新的优化目标、计算模型和压缩对象，推动了编码技术的范式转变。

新的优化目标：从局部优化到全局优化

传统编码技术以像素级保真度为优化目标，采用模块化设计，各环节（如预

测、变换、量化)独立优化,难以实现全局最优。而智能编码通过端到端学习,实现了以下优化目标的突破:

端到端全局优化 基于神经网络的编码框架能够联合优化编码器、量化模块和熵编码器,实现全局最优,这对于复杂的视频编码框架而言尤其重要。

面向机器视觉的语义优化 将优化目标从像素保真度转向语义保真度,提出将多尺度特征损失、面向机器视觉的任务损失等作为优化目标。

面向人类视觉的主观优化 利用生成对抗网络(generative adversarial networks, GAN)和扩散模型(diffusion model)等学习原始图像的分布,在相同码率(尤其是低码率)下实现更优的主观质量。

新的计算模型:从人工设计到数据驱动

传统编码依赖人工设计的变换基(如 DCT)和线性预测模式,而智能编码通过神经网络自动学习最优信号表示:

卷积神经网络(convolutional neural network, CNN)通过层次化特征提取建立空间相关性模型,在刻画信号统计特性的同时能更准确地捕捉图像信号的细微特征。

Transformer 架构 利用自注意力机制建立长程依赖模型,在刻画信号局部特征的基础上更好地捕捉信号的全局相关性。

新的压缩对象:从像素到语义的跃迁

智能技术拓展了编码的物理内涵,将压缩对象从图像视频扩展到其衍生物,例如深度神经网络从图像视频中提取的特征、表征图像视频的神经网络权重等。

特征压缩 直接压缩高层语义特征而非原始像素,能够在提升语义信息压缩效率的同时有效降低神经网络在边缘计算端的运行复杂度;压缩特征而非图像视频像素也可以在一定程度上起到隐私保护的作用。

模型权重压缩 利用隐式神经表达、3D 高斯溅射等,通过压缩神经网络权重间接实现高效图像视频压缩,该方法的主要优点是极低复杂度以及通过连续函数实现任意时空分辨率等。

编码对智能的反哺

编码不仅受益于智能的发展,也为智能技术的发展提供了新的框架以及更高效的训练和推理方法。

统一框架:编码与生成的辩证统一

编码与生成的本质均是尽可能准确地估计概率分布,二者殊途同归地收敛到了统一的智能框架。

统一框架 大语言模型本质上是以压缩作为优化目标训练出的生成模型。而在视觉领域,越来越多的生成模型也逐渐开始使用自回归(同压缩优化目标相同)进行训练,最新生成模型的多尺度自回归就和压缩模型的多层超先验殊途同归。

“压缩即智能” 智能的本质在于从数据中提取有效信息并将其压缩为简洁的表示。智能模型的高压缩率通常意味着智能模型能够从大量数据中提取关键信息,并生成更简洁的表示形式。这种压缩能力反映了智能系统对数据的理解和抽象能力。此外,压缩后的表示应能推广到新任务或新数据,体现智能模型的泛化能力。

模型训练:高效模型训练

编码技术能够显著减少输入或传输数据量,有助于提升模型训练效率。

低维隐变量 将输入图像或视频压缩至隐空间进行降维,使智能模型可以在更低维的数据上进行训练,有助于加速模型收敛并提升模型泛化性。

分布式训练 在分布式训练中,压缩模型参数和梯度更新可以显著减少通信带宽需求,从而加快深度模型训练速度。

模型推理:高效模型推理

压缩技术可以在剪枝、量化、蒸馏等方面显著提升模型推理效率。

参数剪枝 压缩技术可以通过去掉不重要的权重,甚至 Transformer 模型中的词元(token)等,减少模型运行时的实际参数数量,从而降低模型计算复杂度。

模型量化 量化技术将浮点数权重转换为低精度的整数表示,减少模型推理的计算和存储需求。进一步,模型可以使用混合精度推理,利用硬件加速来加快运行速度。

模型蒸馏 蒸馏基于压缩的思想,通过让学生模型模仿教师模型的输出和中间层信息,使其在减少参数量和计算成本的同时仍保持较高的性能。

智能图像编码技术发展现状和前行之路

智能图像编码技术主要包括混合智能编码架构、

端到端智能编码架构以及面向机器视觉的编码，下面将针对上述 3 个方面进行详述。

混合智能编码架构

近年来，深度学习技术已在经典混合视频编码框架上得到广泛研究，这些技术的应用场景主要可以分为提升编码效率和降低编码复杂度两大类。前者主要提出用神经网络模块替换其中的手工设计模块，或使用神经网络对编码过程中的数据进行增强。常见的方法包括将卷积神经网络用于帧内预测、帧间预测质量增强和环路滤波等，如图 3 所示。同时，也有一些方法尝试使用训练的神经网络替代变换中的 DCT 模块，以及改进熵编码模块中的概率估计过程。此外，在基于重采样的视频编码中，上下采样神经网络也被用来替换传统的线性滤波器。在另一类用于降低编码复杂度的方法中，深度学习工具常被用于帧内预测、帧间预测模式的快速抉择以及块划分的加速。这些方法使用离线训练的网络模型来辅助或者替换混合编码框架中的穷举在线搜索优化方法。

这些深度学习工具通过针对单个模块的训练数据进行拟合，通常在编码通路的某个环节上有着不错的性能。但考虑到传统编码框架复杂性，这些工具往往不能面向编码全局率失真性能进行端到端训练，同时各个模块之间也无法联合训练，这在一定程度上限制了神经网络在编码上的性能极限。未来的研究可以从编码框架的在线优化与网络离线训练相结合的角度入手，以联合优化的方式探索更高效编码性能的实现。

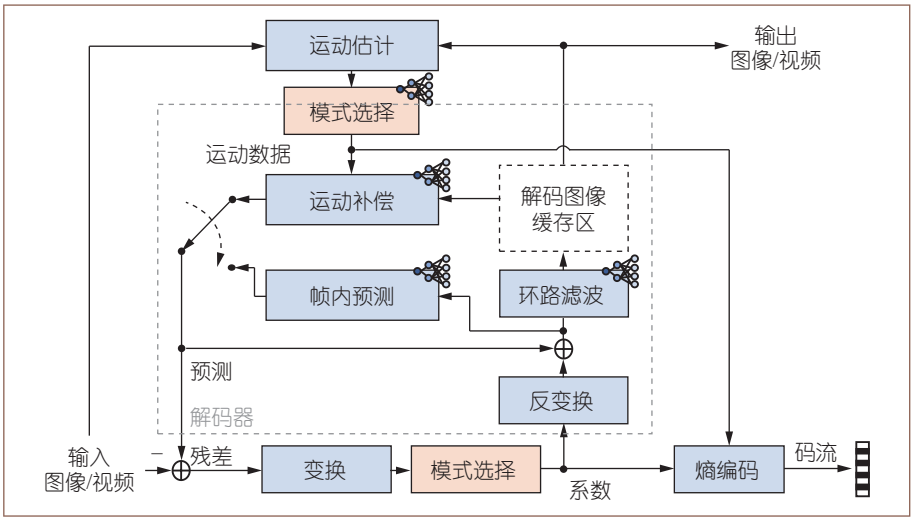


图3　混合智能编码架构示意

端到端智能编码架构

端到端智能编码的基本架构如图 4 所示，经过近几年的迅速发展已经取得了卓越的成果。最先进的图像编码器相较于传统编码标准 H.266/VVC 在率失真性能上已经有了超过 15% 的性能提升；端到端视频编码器也在迅速发展的过程中逐步解决了可变码率、误差传播等关键问题，相较于 H.266/VVC 有接近 30% 的性能提升。此外，一些工作也聚焦于实际应用场景，开发出了可用于低功耗设备的实时编码或解码的端到端图像/视频编解器。在特殊场景下，端到端图像/视频编码也展现出更易扩展的功能性，如可伸缩编码、多视角编码等。这些成果说明端到端编码的巨大潜力及其研究价值。

未来，端到端图像视频编码主要有两条发展路径：一是追求更高的性能上限和多样化的功能性。一方面，端到端编码尚未达到性能瓶颈，仍有空间进一步探索编码的性能上限；另一方面，端到端编码的高度可扩展性可以支持更多模态、更多场景的协同智能。因此，后续的工作需要更多考虑编码自身的特性，面对更多样的应用场景同时结合最新的视觉模型（如扩散模型等）来实现真正的“智能”和“编码”统一框架。二是追求低延时、可落地和硬件扩展。编码最终追求实际应用中的实现，尽管已经有工作聚焦于端到端实时编解码，但目前端到端编码仍不具备实际落地部署的能力。因此在探索性能上限的同时，端到端编码还需要沿着更低延时、更低功耗以及硬件适配等落地方向前进，从而将卓越的性能成就带到生活的实际应用中。

面向机器视觉的编码

随着人工智能（artificial intelligence, AI）技术的快速发展，面向机器视觉的图像视频压缩技术已成为多媒体处理领域的重要研究方向，其基本架构如图 5所示。当前研究大致有如下几条技术路径：一是基于可伸缩编码架构构建层次化

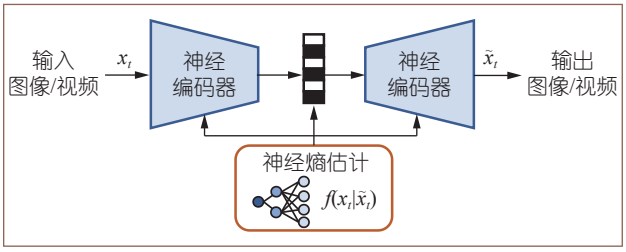


图4　端到端智能编码架构示意

特征表示，使单码流具备多任务适应性。这类方法采用码率自适应机制，在率失真优化理论框架下实现分类、检测、分割和重建等下游任务性能的渐进式提升。二是使用多目标联合优化方法突破传统重建失真这一单一评价体系。构建了融合像素级保真度与任务级语义保持度的复合型损失函数，使得一段码流可以同时适配多种不同的任务。三是摒弃像素重建环节，直接对视觉任务特征进行压缩，建立端到端的特征压缩通道。这类方法通过借鉴现有的一些压缩技术，分析不同视觉任务的特征分布特性，设计针对性的量化、熵编码策略，在保证下游任务性能的同时实现高效压缩。

尽管面向机器视觉的压缩技术取得了显著进展，现有技术体系仍面临严峻挑战，尤其是跨任务泛化能力较弱的问题。不少方法仅针对单一或少数几个视觉任务进行了优化，当迁移至新领域时性能显著下降，使之无法适应多样化的下游任务需求。如何建立更加普适的机器视觉压缩方法是未来需要解决的一个重要科学问题。

智能图像视频编码标准发展现状和前行之路

智能图像视频编码标准的发展历程如图 6 所示，主要包括运动图像专家组（Moving Picture Experts Group, MPEG）开展的基于混合编码架构的神经网络视频编码（neural network-based video compression, NNVC）标准化活动；电气与电子工程师协会（Institute of Electrical and Electronics Engineers, IEEE）和联合图像专家组（Joint Photographic Experts Group, JPEG）同期制定的端到端图像压缩标准 IEEE 1857.11 和 JPEG AI，以及 AVS 标准组正在开展的端到端视频编码标准端到端探索模型（end-to-end exploration model, EEM）活动；MPEG 和面向机器的数据压缩（data compression for machines, DCM）标准组正在制定的面向机器视觉的图像视频压缩标准机器视频编码（video coding for machines, VCM）

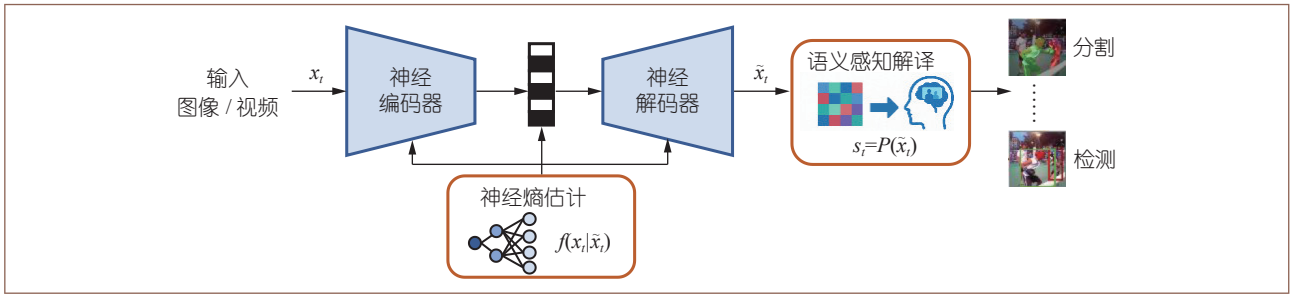


图 5　面向机器视觉的端到端智能编码架构示意

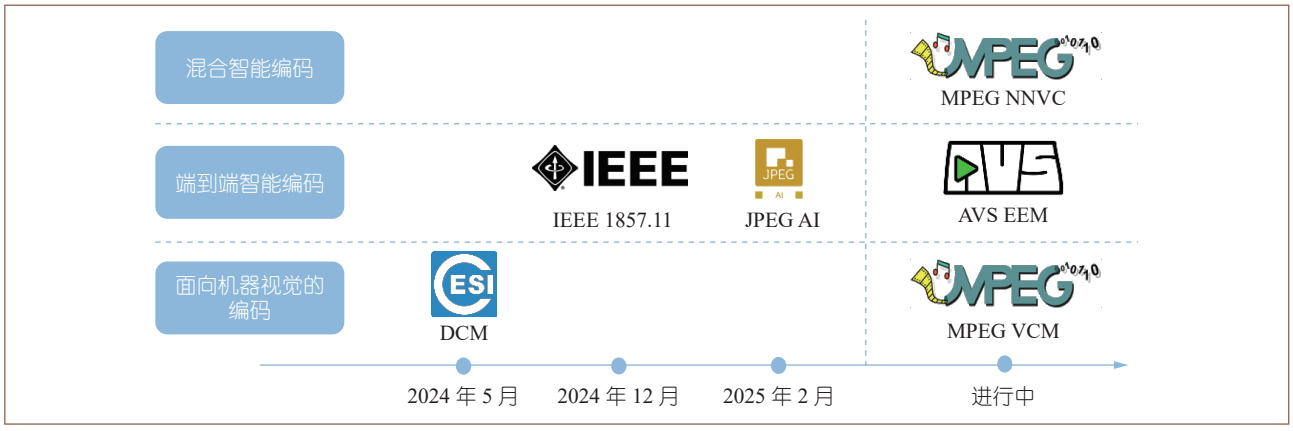


图 6　智能图像视频编码标准工作发展进程

和 DCM。接下来将针对上述 3 个方面进行详述。

NNVC 发展现状和前行之路

NNVC 主要基于 H.266/VVC，通过神经网络优化传统编码流程中的关键环节，如帧内预测、环路滤波和运动补偿等，从而提高压缩效率。近年来，MPEG 在工作组内积极探讨基于深度学习的编码工具的实用化方案，并有诸多企业（如 Fraunhofer HHI、腾讯、华为、字节跳动等）投入研究。特别地，基于深度学习的环路滤波在低复杂度限制下（低于 7 kMAC/pixel），仍然可以带来显著的性能提升（超过 5%）。然而，在进一步迈向实用化的过程中，NNVC 仍面临计算复杂度高、硬件资源需求大和标准化进程缓慢等挑战，仍处于实验性研究和初步实现阶段。

NNVC 的未来前行之路将围绕轻量化和易部署两个方向展开，以克服当前计算复杂度高和硬件资源需求大的挑战。在轻量化方面，未来研究将致力于进一步设计轻量化网络结构，引入知识蒸馏、剪枝和低比特量化等，以减少模型参数量，同时保持甚至提升压缩性能。在易部署方面，未来的方案需要兼容现有编解码框架，如 H.266/VVC，并面向图形处理器（graphics processing unit, GPU）、神经网络处理器（neural processing unit, NPU）和现场可编程门阵列（field programmable gate array, FPGA）进行优化，以提高推理效率并降低硬件要求。同时，NNVC 的标准化进程也将进一步加快，使神经网络模型在不同软硬件平台上具备良好的可复现性和跨平台兼容性。

IEEE 1857.11/JPEG AI 和 EEM 发展现状和前行之路

在端到端图像压缩方面，IEEE 1857.11 与 JPEG AI 是 IEEE 以及 JPEG 两大标准化组织率先进行端到端图像编码技术标准化的尝试成果。IEEE 1857.11 标准由 IEEE 1857.11 子工作组牵头制定，并已于 2024 年底正式发布。该标准充分利用了端到端图像编码的最新技术进展，同时兼顾实际部署需求，标准化了适用不同计算资源的多种智能图像编码方法，在压缩性能上超越了传统编码框架如 H.266/VVC。JPEG AI 是由 JPEG 标准制定的基于学习的图像编码标准，已于 2025 年中正式发布。JPEG AI 标准不仅能够生成紧凑的压缩域表示，在相同的主观视觉质量下大幅提升压缩效率，同时也为下游图像处理和计算机视觉任务提供了一种高效率的处理方法。在端到端视频压缩方面，国内音视频标准组 AVS 牵头形成了首个基于纯神经网络的参考软件 AVS-EEM，在多家单位的共同努力下，目前已成功迭代更新 4 个版本。该方案基于端到端条件编码框架进行搭建，在复杂度约束下加入了多尺度时域参考机制，显著提升了编码效率。目前，AVS-EEM 参考软件的编码端复杂度低于 300 kMAC/pixel，解码端复杂度低于 200 kMAC/pixel。在压缩性能方面，以 AVS3 参考软件 HPM15.1 为基准，AVS-EEM 参考软件在亮度−色度（luminance-chrominance, YUV）分量上取得了和 HPM15.1 相似的压缩性能。

尽管智能图像视频编码标准已经取得了一些阶段性进展，但相关技术和标准的进一步推广应用，仍须对标准化进程进行进一步深入思考和完善。首先，基于实现互联互通这一基本目标，解码器的核心网络结构依然需要严格标准化；然而，对于网络模型参数是否也需要标准化，则应根据具体应用场景进行细致考虑，国标标准工作组 MPEG 也正为这一问题进行深入研讨。其次，编码标准中测试条件的选择与定义直接关系到图像视频质量评价标准的有效性，因此，编码标准的制定应当与图像视频质量评价标准紧密结合、相辅相成。再次，随着端到端图像编码技术的不断进步，更加轻量化的编码方法层出不穷，新的编码范式（如基于隐式表达的方法）也给智能图像视频编码的标准化带来了新的挑战和机遇。最后，智能视频编码标准的一大优势在于其可以针对特定模态和应用场景进行深度优化，例如在监控视频中应注重低延迟和高可靠性，在沉浸式视频和 3D 视频中应追求更高的压缩效率和视觉体验。通过聚焦具体场景的需求，智能视频编码标准将能够更好地服务于行业应用，推动智能视频编码技术的持续创新和落地应用。

VCM/DCM 发展现状和前行之路

VCM 旨在制定面向机器视觉的视频编码标准。基于混合编码框架的经典视频编码标准主要针对人眼感知优化，而 VCM 通过结合高效特征提取和任务感知压缩，使压缩后的视频在尽可能降低码率的同时满足机器视觉任务（如目标检测等）需求。目前，VCM 主要包括两条技术路线：其中一条路线是复用现有基于混合编码框架的视频编码标准，通过结合特征预处理、特

征重排等技术使特征更好地适配视频编码器；另一条路线是端到端特征编码，通过自编码器面向机器视觉任务对特征进行端到端优化。然而，VCM 仍面临诸多挑战，包括如何在不同任务间找到通用的优化策略，以及如何兼顾传统视频编解码标准的兼容性和新技术的创新性。DCM 的主要目标是探索面向机器的数据压缩标准，在最大化数据压缩的同时保留关键信息，使其适用于 AI 计算任务。与 VCM 有多个技术路线不同，DCM 更加专注于端到端数据编码方法，如变分自编码器和自注意力机制等。DCM 近年来在面向任务的端到端点云压缩、面向任务的视频压缩等方面取得了突破，使得压缩后的数据不仅码率更低，同时仍能提供足够的信息供机器进行高效推理。然而，DCM 仍需解决模型计算开销大以及模型可解释性弱等问题。

VCM 和 DCM 标准工作组未来将在多个方向上持续演进。VCM 工作组未来的发展将更多聚焦于与现有视频编码标准的融合，使其既能服务于人类视觉，又能在 AI 任务中提供更优的特征表达；DCM 工作组未来将聚焦神经网络压缩的稳定性和泛化能力问题，确保在不同 AI 任务下都能保持高效的压缩率和良好的视觉任务性能。此外，随着人工智能物联网（artificial intelligence of things, AIoT）、自动驾驶、智能监控等领域对高效视觉数据处理需求的增长，VCM 和 DCM 预计将进一步拓展应用边界，并可能形成更加标准化的解决方案，推动机器视觉编码技术走向更广泛的产业应用。

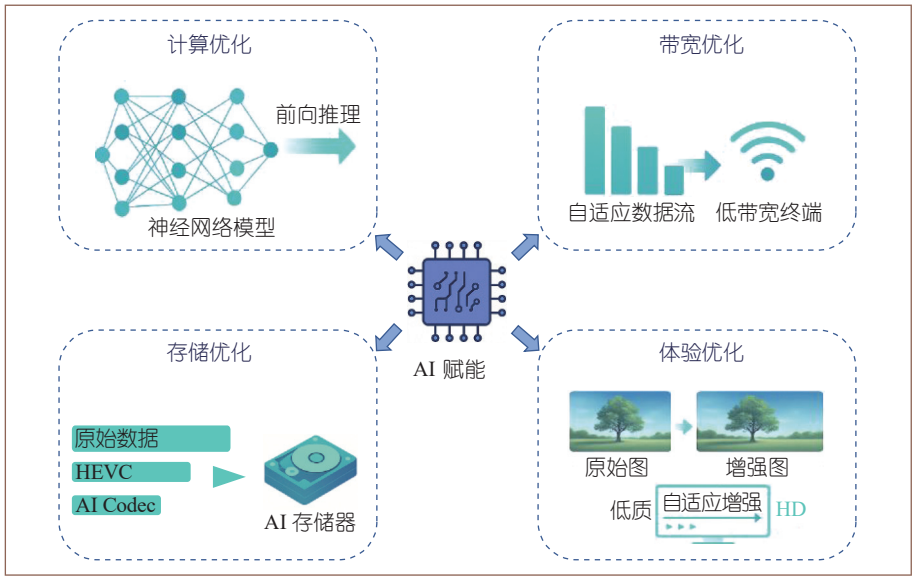


图7　面向优化计算、带宽、存储和体验的智能图像视频编码发展之路示意

智能图像视频编码产业化现状和发展之路

智能图像视频编码产业化现状

智能图像视频编码技术近年来在学术界取得了显著突破，但其产业化进程仍处于早期阶段。目前国内的应用仍主要体现在智能增强、超分辨率等图像视频编码的非内核技术中。涉及图像视频的编码内核技术，目前主要在安防领域有部分公司尝试基于神经网络的图像压缩应用。国外相比国内的应用规模更大，例如 YouTube 已使用深度学习技术优化 AV1 编码，节约一定的带宽；微软 Teams 也使用智能码率控制提升视频实时编码质量。但是，智能编码想要和经典编码一样规模化应用于数字经济的方方面面仍面临诸多限制。首先，智能编码算法通常需要大量的计算资源，这对硬件设备提出了较高的要求；其次，现有的智能编码模型在泛化性方面表现仍存在不足，难以适应多样化的应用场景；此外，行业内缺乏统一的技术标准，导致不同厂商的解决方案难以兼容，也是限制技术推广的重要原因之一。

智能图像视频编码产业发展之路

优化计算、带宽、存储和体验，解决产品应用瓶颈

智能图像视频编码的广泛应用面临算力、泛化性、标准化等问题。如图 7 所示，为了推动技术的产业化，需要从计算、带宽、存储和体验 4 个方面进行优化，以满足用户需求。

计算优化　智能编码算法通常依赖于深度学习模型，这些模型在推理过程需要大量计算资源。为了降低计算成本，可以通过模型压缩、量化和剪枝等技术来减少模型的复杂度。此外，还可以利用专用硬件（如 AI 加速芯片）提升计算效率。

带宽优化　在低带宽环境下，智能编码技术可以通过内容感知编码、内容自

适应码率动态调整等方法优化视频传输效率。

存储优化 智能编码技术可以通过基于神经网络的熵模型技术使用高效编码或转码等方式减少图像视频文件的存储空间。这对于需要长期存储大量图像视频数据的行业(如安防监控)具有重要意义。

体验优化 智能编码技术不仅可以提升压缩效率,还可以通过超分辨率、去噪、色彩增强等技术提升视频的视觉质量。

从降本增效到新增长点:智能编码的商业模式创新

如图 8 所示,智能图像视频编码技术不应仅被视为降本增效的工具,而应像“网约车”一样作为新的增长点来实现利润增长。通过提供个性化服务、开发新型应用场景,企业可以更好地将智能编码技术转化为其核心竞争力。

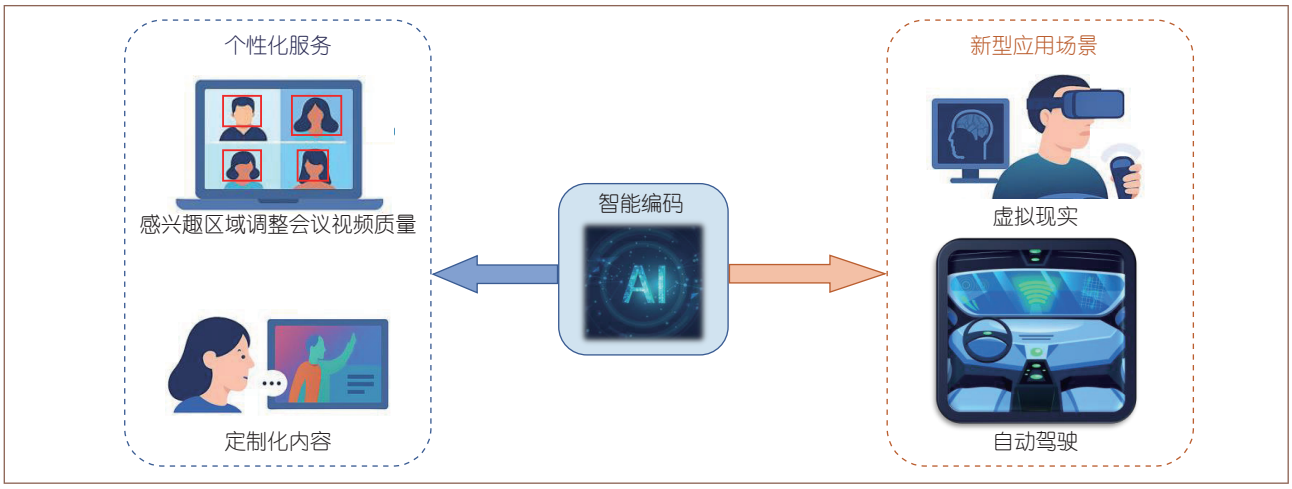


图8 面向智能编码商业模式创新的智能图像视频编码发展之路示意

个性化服务 智能编码技术可以根据用户需求提供个性化的视频服务。例如,在视频会议中,智能编码算法可以根据参会者的注意力区域动态调整视频质量,从而提升会议体验。在流媒体平台中,智能编码技术可以根据用户观看习惯和网络状况,提供定制化的视频流,从而增强用户黏性。

新型应用场景 智能编码技术可以应用于虚拟现实(virtual reality, VR)、增强现实(augmented reality, AR)等新兴领域。例如,在 VR 视频中,智能编码技术可以通过视点预测和区域优化编码来减少数据传输量,从而提升 VR 体验的流畅性。此外,智能编码技术还可应用于生物图像、医疗影像和自动驾驶等领域,为这些行业提供高效的图像视频处理解决方案。

结束语

本文讨论了智能图像视频编码的技术内涵,并在此基础上,对智能图像视频编码技术、标准和产业化发展现状进行了总结,同时提出了相应的发展建议。希望智能图像视频编码领域学术界和工业界的从业者深入思考该领域的技术和标准发展之路,为未来大规模的产业化推广做好技术和标准准备。 ■

致谢: 本文中部分观点来自 CCF YOCSEF 合肥于 2024 年 11 月 16 日举办的闭门技术论坛“智能图像视频编码路在何方?”。感谢与会嘉宾的建议和思辨。



李 礼
CCF 高级会员、YOCSEF 合肥 2025—2026 届主席、多媒体技术专委会执行委员。中国科学技术大学特任教授。主要研究方向为图像视频智能编码。
lil1@ustc.edu.cn



贾川民
CCF 专业会员、多媒体技术专委会执行委员。北京大学助理教授。主要研究方向为多模态数据智能编码。
cmjia@pku.edu.cn

参考文献

[1] Thomas Wiegand, Gary J. Sullivan, Gisle Bjøntegaard, et al. Overview of the H. 264/AVC Video Coding Standard[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2003, 13(7): 560−576.

[2] Gary J. Sullivan, Jens-Rainer Ohm, Woo-Jin Han, et al. Overview of the High Efficiency Video Coding (HEVC) Standard[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2012, 22(12): 1649−1668.

[3] Benjamin Bross, Ye-Kui Wang, Yan Ye, et al. Overview of the Versatile Video Coding (VVC) Standard and its Applications[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(10): 3736−3764.

[4] Jiaqi Zhang, Chuanmin Jia, Meng Lei, et al. Recent Development of AVS Video Coding Standard: AVS3[C]//2019 Picture Coding Symposium (PCS). Ningbo: IEEE, 2019: 1−5.

[5] Tianyi Li, Mai Xu, Ce Zhu, et al. A Deep Learning Approach for Multi-Frame In-Loop Filter of HEVC[J]. *IEEE Transactions on Image Processing*, 2019, 28(11): 5663−5678.

[6] Jiaohao Li, Bin Li, Jizheng Xu, et al. Fully Connected Network-Based Intra Prediction for Image Coding[J]. *IEEE Transactions on Image Processing*, 2018, 27(7): 3236−3247.

[7] Hyomin Choi, Ivan V. Bajić. Deep Frame Prediction for Video Coding[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019, 30(7): 1843−1855.

[8] Johannes Ballé, Valero Laparra, Eero P. Simoncelli. End-to-end Optimized Image Compression [EB/OL]. (2016−11−05)[2025−07−01]. https://doi.org/10.48550/arXiv. 1611.01704.

[9] Jiaohao Li, Bin Li, Yan Lu. Deep Contextual Video Compression[J]. *Advances in Neural Information Processing Systems*, 2021, 34: 18114−18125.

[10] Lingyu Duan, Jiaying Liu, Wenhan Yang, et al. Video Coding for Machines: a Paradigm of Collaborative Compression and Intelligent Analytics[J]. *IEEE Transactions on Image Processing*, 2020, 29: 8680−8695.

Future Development Path of Intelligent Image and Video Coding

Li Li[1], Chuanmin Jia[2]

1. University of Science and Technology of China

2. Peking University

Abstract: In recent years, intelligent image and video coding technologies have made significant progress. In terms of coding performance, they have approached or even surpassed classical image and video coding methods based on signal statistics, while enabling coding schemes optimized for both subjective quality and machine vision. However, their widespread adoption remains challenging due to high encoding/decoding complexity and the lack of dedicated hardware support. As the development of classical image and video coding methods has been stuck in a bottleneck, this article first examines the technical essence of intelligent image and video coding, then conducts an in-depth discussion focusing on the technological, standardization, and industrialization pathways of these coding technologies, aiming to promote further advancement in this field.

Keywords: image and video coding; intelligent coding; feature coding; video coding for machine; autoencoder

摘　要: 智能图像视频编码技术近年来取得了长足发展，从编码性能角度已经接近甚至超越了基于信号统计的经典图像视频编码方法，并可以使能面向主观质量、机器视觉优化的编码方法。但是，目前因为编解码复杂度高、缺乏专用硬件等原因仍难以得到广泛应用。在经典图像视频编码方法发展举步维艰之际，本研究首先讨论智能图像视频编码的技术本质，并进一步聚焦智能图像视频编码技术、标准、产业化之路进行深入分析与探讨，以期推动该领域的进一步发展。

关键词: 图像视频编码；智能编码；特征编码；面向机器的视频编码；自编码器

中图分类号: TN919.81; TP391

中文引用格式: 李礼, 贾川民. 智能图像视频编码的未来发展之路 [J]. 计算, 2025, 1(7): 69−77.

英文引用格式: Li Li, Chuanmin Jia. Future Development Path of Intelligent Image and Video Coding[J]. *Computing Magazine of the CCF*, 2025, 1(7): 69−77.

（本文责任编委：杨　易）