

# 机器视觉编码关键技术与未来展望

袁 冯<sup>1,2</sup>, 张立谦<sup>1</sup>, 潘兆庆<sup>1</sup>, 王瀚漓<sup>3</sup>, 李 礼<sup>4</sup>, 赵铁松<sup>5</sup>, 张 云<sup>6</sup>

(1. 天津大学电气自动化与信息工程学院, 天津 300072; 2. 徐州医科大学医学人工智能学院, 徐州 221004; 3. 同济大学电子与信息工程学院, 上海 201804; 4. 中国科学技术大学信息科学技术学院, 合肥 230026; 5. 福州大学物理与信息工程学院, 福州 350108; 6. 中山大学电子与通信工程学院, 深圳 518107)

**摘 要:** 随着人工智能与智能视觉技术的快速发展, 图像和视频已由传统的人眼观看媒介逐步演变为支撑机器感知、理解与决策的重要信息载体。传统视觉编码方法主要面向像素重建质量和人眼感知质量进行优化, 难以充分满足目标检测、语义分割、目标跟踪和行为识别等机器视觉任务对语义信息与判别特征高效保持的需求。近年来, 面向机器视觉的编码技术受到学术界和工业界的广泛关注, 并逐渐成为视觉信息压缩领域的重要研究方向。本文围绕机器视觉编码的关键技术与发展趋势进行系统综述。首先, 介绍传统图像视频编码标准面向机器视觉任务的优化方法, 重点分析任务感知码率分配、机器视觉率失真优化以及编码工具优化等关键技术。随后, 系统梳理基于深度学习的智能编码方法, 并按照编码对象将其归纳为智能图像编码、智能视频编码和智能特征编码3个主要方向, 分析相关技术由像素重建向语义保持、任务性能保持及边缘-云协同智能处理演进的发展脉络。最后, 总结当前机器视觉编码在评价体系、跨任务与跨模型泛化、计算复杂度、标准兼容性以及人机协同与隐私安全等方面面临的主要挑战, 并对统一率失真建模、通用视觉表征编码、轻量化智能编码、人机协同编码及标准化等未来发展方向进行展望。

**关键词:** 机器视觉编码; 视频编码标准; 智能图像编码; 智能视频编码; 特征编码; 人机协同

**中图分类号:** TN919.81

**文献标志码:** A

**引用格式:** 袁冯, 张立谦, 潘兆庆, 等. 机器视觉编码关键技术与未来展望[J]. 数据采集与处理, 2026, 41(5): 1240-1267. YUAN Feng, ZHANG Liqian, PAN Zhaoqing, et al. Coding for machine vision: Key technologies and future perspectives[J]. Journal of Data Acquisition and Processing, 2026, 41(5): 1240-1267.

## 引 言

随着人工智能技术的快速发展, 视频数据的功能属性正由传统的人眼感知媒介<sup>[1-4]</sup>逐步拓展为支撑机器感知、智能分析、决策生成与自动化控制的重要信息载体<sup>[5-7]</sup>。在智慧城市中, 海量监控视频广泛服务于行人检测、车辆识别、交通流量统计及异常事件预警<sup>[8]</sup>; 在智能制造中, 生产线视频可支撑表面缺陷检测、装配状态识别与机器人协同控制<sup>[9]</sup>; 在智慧医疗中, 内窥镜、超声、手术导航及医学影像视频则为病灶识别、器官分割和辅助诊断等任务提供重要依据<sup>[10]</sup>。由此可见, 视频数据的应用价值已不再局限于面向人眼的观看与事后回放, 而是日益体现于实时感知、自动分析和智能决策等机器视觉应用之中。与此同时, 视频数据规模也呈现持续快速增长趋势。一方面, 摄像头、无人机、车载传感器、工业相机和医疗成像设备等视觉采集终端的大规模部署, 使视频数据源不断增加; 另一方面, 分辨率、帧率、动态范

**基金项目:** 国家自然科学基金(62322116)。

**收稿日期:** 2026-08-13; **修订日期:** 2026-09-20

围以及多视角采集能力的持续提升,也使单路视频产生的数据量显著增长。尤其在智能监控、自动驾驶和工业检测等典型场景中,系统通常需要对多路高清视频进行长时间、连续采集、传输与分析,由此对网络带宽、边缘计算能力及云端存储资源提出了更高要求。因此,如何在有限带宽与计算资源约束下实现视频数据的高效压缩、传输与智能利用,已成为视频信息处理和机器视觉领域亟待解决的关键问题。

传统视频编码技术在降低视频传输带宽与存储开销的需求驱动下不断发展。早期视频编码主要面向电视广播、视频通信和多媒体存储等各类观看场景,其核心目标是在有限码率条件下尽可能获得较高的重建质量和良好的主观视觉体验。围绕这一目标,高级视频编码(Advanced video coding, AVC)<sup>[11]</sup>、高效视频编码(High efficiency video coding, HEVC)<sup>[12-15]</sup>和通用视频编码(Versatile video coding, VVC)<sup>[16-17]</sup>等国际视频编码标准相继制定,并逐步形成了以块划分<sup>[18]</sup>、帧内预测<sup>[19-20]</sup>、帧间预测<sup>[15]</sup>、运动估计与补偿<sup>[21]</sup>、变换量化<sup>[22]</sup>、熵编码<sup>[23]</sup>及环路滤波<sup>[24-25]</sup>为核心的混合视频编码框架。其中,帧内预测利用同一帧内像素或区域之间的空间相关性降低空间冗余,帧间预测通过运动估计与补偿挖掘相邻帧之间的时间相关性,变换量化与熵编码进一步减少信号表示和统计冗余,环路滤波则用于降低编码过程中产生的块效应及其他重建失真。经过多代标准持续演进,传统视频编码技术在压缩效率和重建质量方面取得了显著提升。然而,其编码优化范式本质上仍以人眼感知为中心。无论是峰值信噪比(Peak signal-to-noise ratio, PSNR)、结构相似性(Structural similarity, SSIM)等客观质量指标<sup>[26-28]</sup>,还是基于人类主观感知构建的视频质量评价方法<sup>[29-31]</sup>,主要关注压缩视频相对于原始视频的像素保真度、结构相似性或视觉自然度。

人类视觉感知与机器视觉分析对视频信息的关注重点存在明显差异。对于人类观看而言,纹理连续性、边缘自然度、亮度一致性以及整体视觉舒适度通常具有较高的重要性<sup>[32-36]</sup>;对于目标检测、语义分割、目标跟踪和行为识别等机器视觉任务而言,目标区域、轮廓边界、局部判别纹理、运动信息以及高层语义关系往往具有更直接的任务贡献<sup>[8-10]</sup>。例如,在交通监控场景中,天空、道路和建筑等大面积背景区域虽然影响整体视觉质量,但其对车辆检测和车牌识别等任务的贡献通常较为有限;相较之下,小目标车辆、行人轮廓、车牌字符以及跨帧运动轨迹所占像素比例较低,却可能直接影响机器视觉任务的分析结果。在场景理解和行为识别等任务中,背景所包含的环境上下文及其与目标之间的关系同样是重要的判别依据。机器视觉对信息的关注具有任务依赖性,其编码需求不能简单等同于人眼视觉感知需求。这种感知目标上的差异使传统视频编码方法直接应用于机器视觉场景时面临编码优化目标与机器任务需求不匹配的问题。特别是在低码率条件下,编码器可能为维持整体重建质量而分配较多比特用于背景纹理,却对小目标、目标边界及细粒度判别区域进行较强量化,从而引发检测框定位偏差、分割边界模糊、目标跟踪中断乃至行为识别错误。此外,部分面向主观视觉质量设计的环路滤波、去块及质量增强工具虽然有助于改善人眼观看体验,却可能在一定程度上削弱机器视觉模型所依赖的高频结构与局部判别特征。因此,传统面向人眼感知建立的视频编码优化机制难以直接满足机器视觉任务对任务相关信息高效保持的需求,亟须根据机器感知特性重新审视视频编码中的信息重要性度量、比特资源分配及率失真优化机制。

在上述背景下,视频编码研究逐步由面向人眼观看的传统压缩范式<sup>[37-41]</sup>,拓展至面向机器理解的智能编码范式<sup>[42-43]</sup>。不同于传统视频编码以像素重建质量和主观视觉质量为主要优化目标,面向机器视觉的视频编码更加关注压缩码流对下游视觉任务的支撑能力,其核心目标是在给定码率约束下尽可能保持目标检测、语义分割、目标跟踪、行为识别及场景理解等任务的分析性能。围绕这一目标,研究者逐步将目标区域、语义分割掩膜、视觉显著性、任务损失、深层视觉特征以及任务感知率失真模型等信息引入编码过程,并通过编码前预处理、感兴趣区域比特分配、量化参数自适应调控、编码工具优化、压缩域语义分析以及端到端任务驱动学习等技术,使有限的编码资源能够更加集中地服务于任务相关信

息的表示与保持。伴随机器视觉需求不断深化<sup>[44-45]</sup>,视频编码的优化目标也逐步由像素保真和感知保真向语义保真与任务性能保持演进,其关注重点由压缩后视觉内容的重建质量进一步延伸至压缩表示所承载的机器可理解信息及其任务效用。这一范式转变突破了传统视频编码以人类视觉感知为核心的优化边界,也推动视频编码与计算机视觉、深度学习及边缘智能等技术进一步融合,为智慧城市、工业视觉、智慧医疗、自动驾驶及边缘-云协同智能分析等场景中的大规模视觉数据高效处理提供了新的技术路径。

本文旨在系统梳理机器视觉编码的关键技术、代表性方法及其演进脉络,相关技术路线及其相互关系如图1所示。首先,以传统混合视频编码标准为基础,介绍视频编码的基本框架与率失真优化机制,并重点综述面向机器视觉任务的标准编码器优化方法,包括基于感兴趣区域的码率分配、任务感知量化控制、压缩域视觉分析以及机器视觉率失真建模等关键技术。其次,进一步梳理基于深度学习的智能编码方法,并根据编码对象将相关研究划分为智能图像编码、智能视频编码和智能特征编码3个主要方向,系统分析其由像素重建优化向语义信息保持、任务性能保持以及边缘-云协同推理演进的技术路径。最后,总结现有研究在评价体系、跨任务与跨模型泛化、标准兼容性、编码复杂度以及人机协同等方面面临的主要问题,并对机器视觉编码未来的发展方向进行展望。

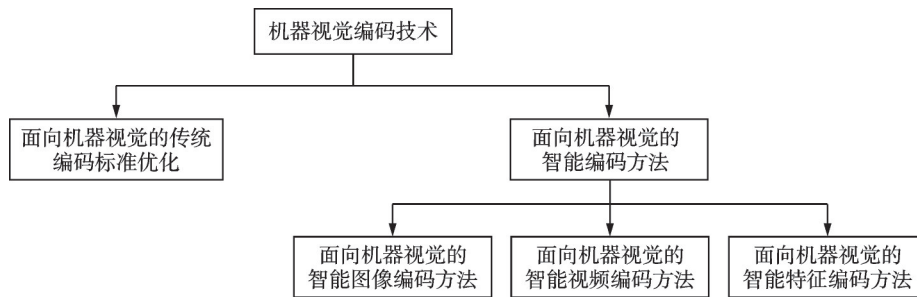


图1 机器视觉编码技术脉络结构图

Fig.1 Structure diagram of machine vision coding technologies

## 1 面向机器视觉的传统编码标准优化

传统图像与视频编码标准长期以人眼感知为主要优化目标,其性能评价通常围绕PSNR、SSIM及主观视觉质量展开。然而,随着越来越多的视频数据被直接用于目标检测、语义分割、目标跟踪和图像识别等机器视觉任务,编码过程中产生的像素失真与下游任务性能退化之间并不存在严格的一致关系。Reich等<sup>[5]</sup>针对图像分类、目标检测、语义分割和光流估计等任务开展研究,结果表明JPEG(Joint photographic experts group)和H.264等传统编码方法在高压缩率条件下会显著影响深度视觉模型的输出,其中语义分割任务的平均交并比(mean Intersection over union, mIoU)在低码率条件下可下降超过80%。Fischer等<sup>[8]</sup>面向自动驾驶场景分析压缩失真对机器视觉任务的影响,发现小目标、道路边界及局部判别纹理等信息在压缩后更容易引起检测与分割性能下降。Pham-Dinh等<sup>[46]</sup>基于运动图像专家组(Moving picture experts group, MPEG)的面向机器的视频编码(Video coding for machines, VCM)评价框架,系统分析不同编码配置对机器视觉任务性能的影响;Partanen等<sup>[47]</sup>则构建了UVG-VCM数据集,为机器视觉编码方法提供了相对统一的数据与评测基础。上述研究分别从任务性能退化分析和评价基准构建等角度揭示了传统编码失真与机器视觉任务性能之间的差异,也表明仅依赖像素级重建质量难以充分反映压缩数据对机器分析的有效性。因此,有必要面向机器视觉任务重新审视传统编码标准中的优化目标、编码策略及性能评价机制。

为提升传统图像与视频编码器面向机器视觉任务的压缩效率,相关研究开始通过调整码率分配策略,使有限的编码资源优先用于任务关键区域,其基本思路如图2所示。Stankiewicz等<sup>[48]</sup>提出基于感兴趣区域重定向的机器视觉视频编码方法,通过对目标区域进行裁剪、放大并结合填充操作保持编码输入尺寸不变,使有限码率优先用于目标相关内容,从而在不改变编码器输入格式的情况下提升目标区域的有效分辨率。实验结果表明,该方法在腾讯视频数据集(Tencent video dataset, TVD)目标跟踪任务的特定编码配置下最高可节省约57%的码率。Róžek等<sup>[49]</sup>在此基础上进一步引入基于感兴趣区域的视频重定向策略,使对检测和跟踪任务更为重要的区域获得更高的有效空间分辨率,在不同任务与编码配置下节省了3%~57%的码率。与基于感兴趣区域裁剪或重定向的方法不同,Fischer等<sup>[50]</sup>进一步将神经网络生成的目标检测显著性信息引入VVC编码控制过程,通过自适应调节非显著编码树单元(Coding tree unit, CTU)的量化参数,在基本保持检测准确率的条件下最高节省约29%的码率。Xie等<sup>[51]</sup>则利用脉冲神经网络预测机器视觉任务可容忍的失真区域,引导编码器将更多比特资源分配至任务相关内容。总体而言,此类方法无需改变现有标准编码框架的基本结构,能够通过感兴趣区域、显著性或任务敏感性信息对码率分配过程进行调控,因此具有较好的标准兼容性与工程部署潜力。然而,其编码性能通常依赖前端目标检测、显著性预测或感兴趣区域估计的准确性与稳定性;一旦任务关键区域出现错检、漏检或定位偏差,可能造成有限比特资源的不合理分配,并进一步影响机器视觉任务性能及整体编码效率。

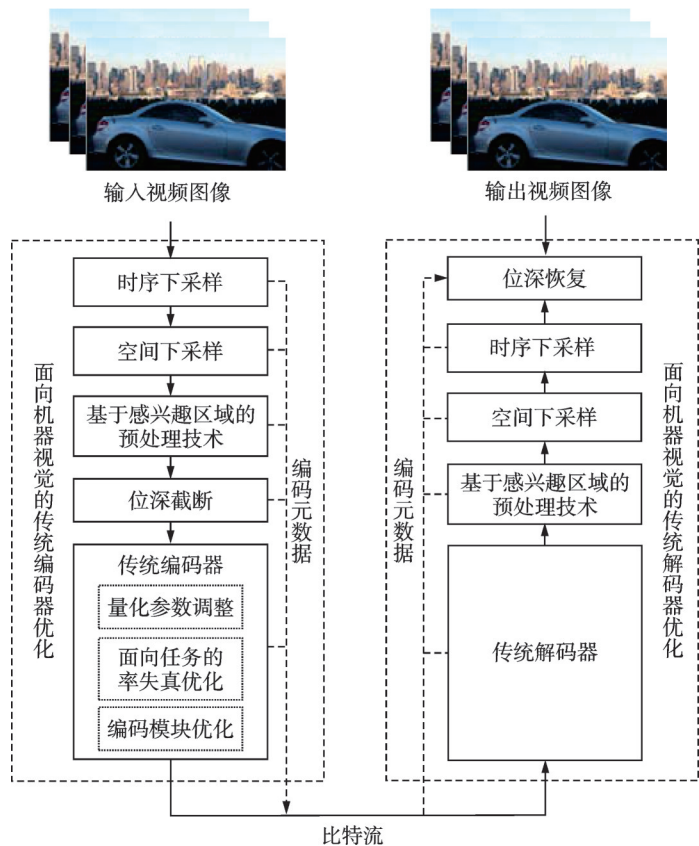


图2 面向机器视觉的传统编码标准优化技术路线

Fig.2 Technical routes of optimization for traditional coding standard for machine vision

为进一步突破传统视频编码标准在机器视觉任务中的压缩效率瓶颈,研究者开始从率失真优化准则入手,将任务性能、语义信息及机器感知特征纳入编码决策过程。Ai等<sup>[52]</sup>以车牌识别任务为例,将拍摄距离、编码码率与识别性能进行联合建模,构建面向识别任务的率失真优化方法。实验结果表明,在相近平均精度均值(mean Average precision, mAP)条件下,该方法相较VVC可节省43.99%的码率,同时将编码时间降低73.76%。Liu等<sup>[53]</sup>提出基于深度Transformer的恰可识别差异预测模型(Deep Transformer-based just recognizable difference prediction model, DT-JRD),用于刻画机器视觉模型对压缩失真的感知阈值,并利用Transformer对图像中的机器识别失真进行建模。基于该失真模型,在基本保持目标检测准确率的同时,相较VVC标准编码器可节省29.58%的码率。Sun等<sup>[54]</sup>则将编码失真信息显式引入下游视觉模型,使任务网络能够学习压缩伪影与任务输出之间的关联,从而提高视觉模型对压缩失真的鲁棒性。Huang等<sup>[55]</sup>和Fernández-Menduina等<sup>[56]</sup>从图像编码角度分别提出视觉分析驱动和特征保持导向的率失真建模方法,进一步表明任务网络的中间特征能够作为传统像素失真度量的有效补充。总体而言,此类方法的核心在于将任务性能、任务敏感特征或机器感知失真显式融入率失真优化过程,使编码决策由传统的像素失真最小化逐步转向面向下游视觉任务的任务失真最小化,从而提升有限码率条件下机器视觉信息的保持效率。

除自适应码率分配和率失真优化外,传统视频编码标准中的具体编码工具也需要结合机器视觉任务需求进行重新评估与优化。Gou等<sup>[57]</sup>面向VCM场景提出VVC帧内编码快速决策方法,利用视觉特征信息和编码单元划分信息提前终止冗余搜索,在尽可能保持机器视觉任务性能的同时有效降低编码复杂度。Fischer等<sup>[58]</sup>系统分析了VVC帧内编码工具及环路滤波工具对神经目标检测性能的影响,研究表明,VVC相较HEVC在人眼视觉质量方面获得的编码增益并不能完全转化为机器视觉任务性能收益。在部分场景下,关闭特定环路滤波工具反而能够在保持相近mAP的条件下进一步降低编码码率。Lorkiewicz等<sup>[59]</sup>则利用神经网络在解码端生成色度信息,从而减少色度分量的传输开销,并在目标跟踪任务中平均节省约12%的码率,部分测试序列的码率节省可达60%。总体而言,此类研究表明,传统编码工具在人眼感知优化与机器任务优化之间并不存在完全一致的收益关系。因此,面向机器视觉的标准编码优化不仅需要调整码率分配策略和率失真准则,还应从编码工具层面对预测、划分、滤波、颜色分量表示及复杂度控制等机制进行任务导向的重新设计,以减少对任务关键结构、纹理及语义信息的不利影响。

总体而言,上述方法主要在保持现有标准编码框架兼容性的基础上,通过任务感知码率分配、机器视觉率失真优化以及编码工具调整与重构等方式提升压缩效率,相关代表性方法及其技术特点比较如表1所示。其主要优势在于能够充分复用JPEG、AVC、HEVC和VVC等成熟编码标准及其软硬件生态,具有较好的工程部署能力与系统兼容性,适用于智慧监控、自动驾驶和边缘视觉分析等既有视频系统的渐进式升级。然而,上述方法在本质上仍建立于以像素重建为核心的传统编码框架之上,机器视觉任务通常仅通过外部先验、任务损失或编码参数调控等方式参与编码决策,尚未从表示层面对任务相关语义信息与任务无关冗余进行显式建模与联合优化。因此,当数据分布、下游任务模型或目标尺度发生变化时,其性能容易受到感兴趣区域估计误差、任务失真度量泛化能力不足以及局部编码工具优化受限等因素影响。如何在保持标准兼容性与工程可部署性的同时,引入更加稳健的语义先验、构建更具泛化能力的任务失真度量,并实现多类编码工具之间的协同优化,仍是传统标准框架下机器视觉编码进一步提升性能的重要研究方向。

表1 面向机器视觉的传统编码标准优化技术比较

Table 1 Comparison of traditional coding standard optimization techniques for machine vision

| 文献          | 核心思想                              | 关键技术                                              | 主要优势                                                    | 性能表现                                                         |
|-------------|-----------------------------------|---------------------------------------------------|---------------------------------------------------------|--------------------------------------------------------------|
| [5,8,46,47] | 从机器任务退化角度重新审视标准编码失真,并建立VCM评价基础。   | 分类、检测、语义分割、光流估计与自动驾驶场景评测;MPEG VCM测试条件和UVG-VCM数据集。 | 揭示PSNR、SSIM等眼质量指标与mAP、mIoU等机器任务指标之间的不一致性,为后续任务导向优化提供依据。 | 低码率压缩下JPEG、H.264等可使语义分割mIoU下降超过80%;相关基准统一了不同编码工具的任务评价。       |
| [48-51]     | 依据检测框、目标边界、视觉显著性或机器可容忍失真区域重新分配码率。 | 感兴趣区域比特分配、视频重定向、检测显著性CTU量化控制、脉冲神经网络失真区域预测。        | 能够复用现有标准编码器,使有限比特优先服务检测、跟踪和分割等任务关键区域。                   | 感兴趣区域方法节省约57%的码率;视频重定向可节省3%~57%码率;显著性VVC最高节省约29%码率。          |
| [52-56]     | 将机器视觉任务失真或任务敏感特征显式纳入率失真优化准则。      | 识别导向率失真优化、DT-JRD、压缩失真嵌入、视觉分析驱动率失真优化、特征保持率失真模型。    | 使编码决策由像素失真最小化进一步转向任务性能保持,能够更准确描述机器可察觉失真。                | 识别导向率失真优化相较VVC取得43.99%编码效率增益,并减少73.76%编码时间;DT-JRD节省29.58%码率。 |
| [57-59]     | 面向机器任务重新评估并优化传统编码标准中的具体编码工具。      | VVC帧内快速决策、环路滤波工具选择、解码端色度合成。                       | 在保持标准兼容性的同时降低编码复杂度和非关键分量传输开销,提高任务相关视觉信息保留能力。            | 色度合成在目标跟踪中平均节省约12%码率,部分序列最高可达60%。                            |

2 面向机器视觉的智能编码方法

随着深度学习技术的快速发展,机器视觉编码研究逐步由依赖传统标准编码器的任务导向优化,转向以深度神经网络为核心的智能编码范式。经过多代标准演进,传统混合编码框架在预测、变换、量化、熵编码及率失真优化等方面已形成较为成熟的技术体系,仅依赖局部参数调整、区域级码率分配或编码工具选择,难以从表示机制层面进一步挖掘机器视觉任务所需信息的压缩潜力。深度神经网络具备较强的特征提取、非线性映射和端到端联合优化能力,能够根据下游任务需求学习更加紧凑且具有任务判别性的视觉表示,因此逐渐成为机器视觉编码的重要技术基础。

总体来看,现有智能编码方法主要沿3条技术路径发展,其整体框架如图3所示。第1类是面向机器的智能图像编码,以单帧图像为编码对象,通过引入任务损失、语义特征约束或人机联合优化目标,引导压缩网络优先保持图像分类、目标检测和语义分割等任务所需的空

间结构与语义信息;第2类是面向机器的智能视频编码,以连续视频序列为编码对象,在端到端框架中联合建模运动信息、时序上下文、潜变量表示及机器视觉任务目标,以充分挖掘视频在空间域和时间域中的任务相关冗余;第3类是面向机器的智能特征编码,其不再以完整像素重建为必要目标,而是直接压缩视觉网络产生的中间特征或高层语义表示,以支持边缘-云协同推理和机器间视觉信息传输。上述3类方法推动机器视觉编码由像素域压缩逐步拓展至特征域和语义域,并使编码优化目标由重建质量保持进一步转向语义信息保持与任务性能保持,为构建更加高效的

任务导向视觉表示与传输机制提供了新的技术路径。

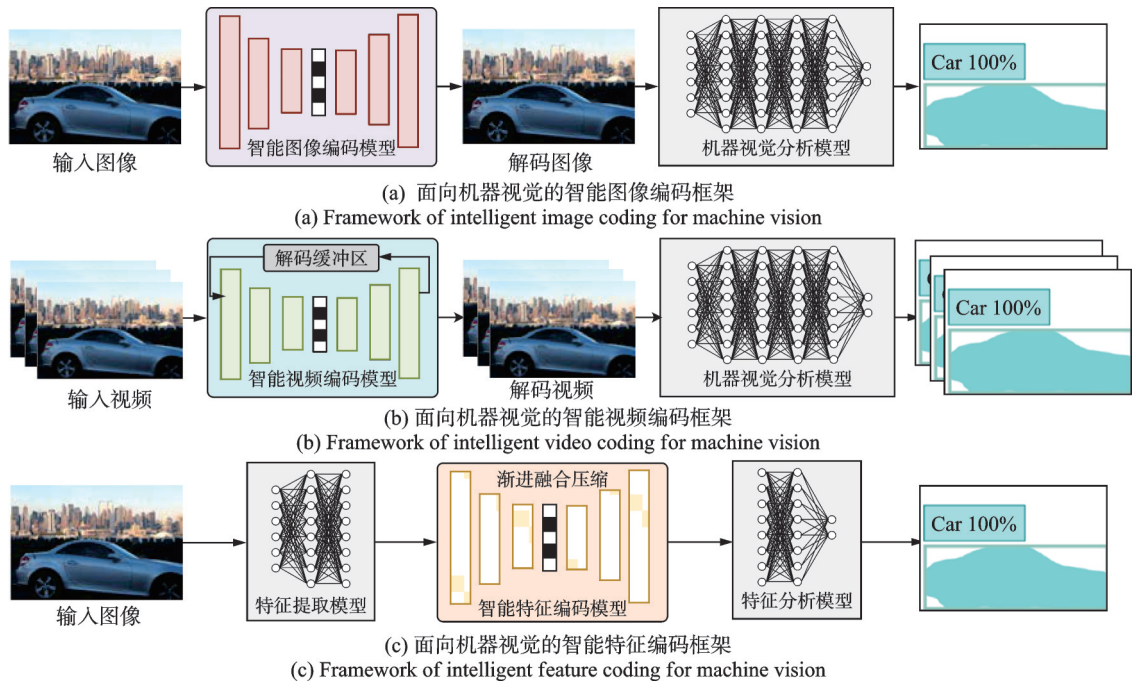


图3 面向机器视觉的智能编码技术路线

Fig.3 Technical routes of intelligent coding for machine vision

## 2.1 面向机器视觉的智能图像编码方法

面向机器视觉的智能图像编码研究,早期主要关注压缩失真对下游视觉任务性能的影响。传统图像编码方法虽然能够在较低码率下获得较好的像素重建质量和主观视觉质量,但其优化目标主要围绕人眼感知建立,难以保证压缩图像中任务相关结构与语义信息的有效保持。Fischer等<sup>[60]</sup>分析了神经图像压缩方法在机器到机器通信场景中的性能,研究表明,即使学习式压缩能够改善感知质量,若其训练目标仍以像素重建或人眼感知为主,所学习的压缩表示也未必有利于下游机器视觉任务。针对这一问题,Chamain等<sup>[61]</sup>将图像编解码器与目标检测网络进行联合微调,使压缩模型能够显式感知下游任务需求,在约0.1每像素比特数(Bits per pixel, bpp)的低码率条件下,与仅对任务网络进行微调的基线方法相比,检测mAP相对提升约4.6%。Le等<sup>[62]</sup>提出面向机器视觉任务的端到端图像编码框架,将目标检测和实例分割损失直接引入压缩模型训练过程,相较VVC在目标检测和实例分割任务上分别取得37.87%和32.90%的Bjontegaard平均码率差(Bjontegaard delta rate, BD-rate)节省。在此基础上,相关研究开始由单一机器任务优化向人机协同与多码率自适应方向拓展。Liu等<sup>[63]</sup>面向机器视觉与人眼视觉共同消费场景提出ICMH-Net,通过潜变量分区、传输、重建与聚合机制,在机器任务性能与图像重建质量之间实现权衡。Le等<sup>[64]</sup>引入内容自适应推理优化策略,使编码器能够根据输入图像内容动态调整压缩表示。实验结果表明,相较原始预训练编解码器,该方法在实例分割任务中平均节省3.66%的BD-rate,在低码率条件下最高可节省9.85%的BD-rate。此外,Liu等<sup>[65]</sup>和Tatsumi等<sup>[66]</sup>分别从可变码率学习和无需再训练的自适应量化角度提升压缩模型对不同码率条件的适应能力,从而增强智能图像编码方法在实际多码率部署场景中的灵活性。总体而言,上述研究主要通过任务驱动的联合优化、人机协同建模以及码率自适应机制,对图像编码模型内部的表示学习与压缩过程进行调整,使压缩表示逐步由面向像素重建转向面向机器视觉任务的信息保持。

在编码前端优化方面,相关方法通过信息筛选与语义增强,在图像进入压缩器之前突出任务相关内容,其基本框架如图4所示,以提升低码率条件下机器视觉任务相关信息的保持能力。Suzuki等<sup>[67]</sup>提出识别感知的图像预变换方法,在编码前增强与目标识别相关的结构信息,从而提高压缩图像对后续识别任务的适应性。Lu等<sup>[68]</sup>提出预处理增强的机器视觉图像压缩方法,通过抑制背景冗余并突出目标结构,在保持下游任务准确率的同时,相较BPG等传统编码方法可节省约20%的码率。Yang等<sup>[69]</sup>设计了语义预处理器,利用高层语义先验突出机器视觉任务关注区域;Patwa等<sup>[70]</sup>则围绕语义信息保持开展图像压缩研究,以减少编码过程对任务关键语义的破坏。Qi等<sup>[71]</sup>通过学习非均匀重要性质量图,引导编码器根据不同空间区域的任务重要程度实施差异化比特分配。Yang等<sup>[72]</sup>进一步提出任务可切换预处理器,使统一的前端预处理框架能够适应图像分类、目标检测和语义分割等不同机器视觉任务。此外,Chen等<sup>[43]</sup>提出色度感知的面向机器视觉学习式图像压缩方法,将色度信息的处理引入图像压缩模型设计,从颜色分量的表示与压缩角度拓展了面向机器视觉的图像编码思路。总体而言,上述方法通过在正式编码前强化任务相关区域、结构和语义信息,同时抑制背景及非关键内容,使有限的编码资源能够更加集中地用于机器视觉任务敏感信息的表示与传输。在高层语义建模方面,相关方法进一步引入语义概念、结构先验和生成模型,以减少对传统逐像素重建范式的依赖,并提升极低码率条件下的压缩效率。Chang等<sup>[73-74]</sup>分别从概念编码和语义先验建模角度出发,探索利用更高层次的语义表示替代部分像素级信息传输。对于人脸等具有稳定几何结构和较强语义先验的视觉对象,生成模型与结构先验能够在极低码率条件下有效重建任务相关内容,从而显著提升智能图像编码效率。Zhang等<sup>[75]</sup>提出机器感知驱动的层级生成式人脸图像压缩方法,在多任务视觉分析场景中,相较直接传输原始RGB图像最高可节省99.6%的码率,而分割mIoU仅下降约0.8%。Mao等<sup>[76]</sup>利用基于风格的生成对抗网络(Style-based generative adversarial network, StyleGAN)生成先验构建可伸缩人脸图像编码方法,在约0.002 4 bpp条件下,相较VVC可使人脸解析任务的频率加权交并比(Frequency-weighted intersection over union, FWIoU)提升52.44%,并显著改善极低码率条件下的身份识别性能。Fang等<sup>[77]</sup>则以人脸草图作为结构先验,提出兼顾人类视觉与机器视觉的人脸图像编码方法,在机器视觉任务中可节省40.9%~86.6%的码率,同时保持较好的主观重建质量。与传统图像压缩侧重完整像素细节重建不同,上述研究表明,在极低码率条件下,优先编码目标结构、高层语义及任务判别线索,并利用生成先验恢复部分非关键视觉细节,可以在显著降低传输码率的同时维持较高的机器视觉任务性能。

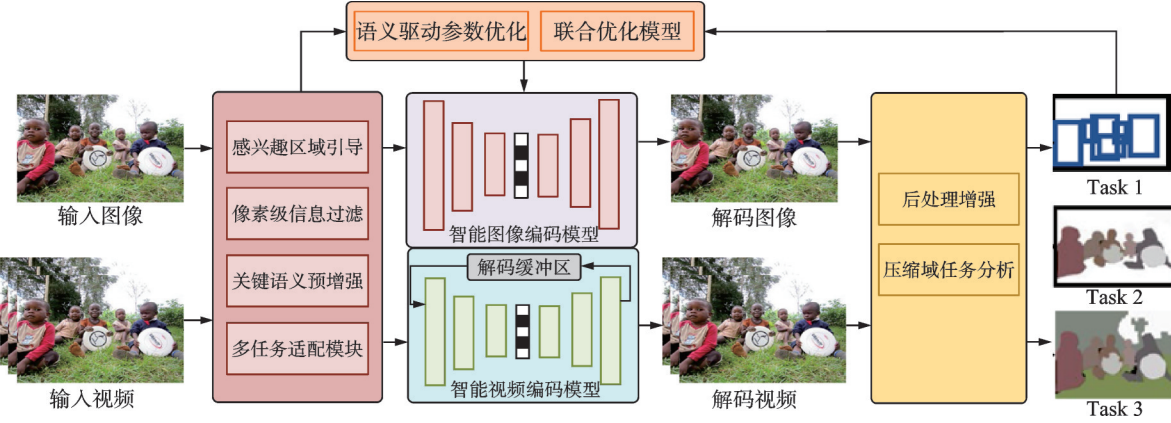


图4 面向机器视觉的智能图像/视频编码优化方法框架  
Fig.4 Framework of optimization methods for intelligent image/video coding for machine vision

为减少传统“压缩-解码-像素域分析”流程对完整图像重建的依赖,相关方法开始从压缩域表示学习与任务网络适配等角度重新设计智能图像编码框架,其基本思路如图4所示。Liu等<sup>[78]</sup>提出学习式压缩域语义分割方法,通过动态通道选择与静态通道选择机制抑制压缩表示中的冗余信息,在相同 mIoU 条件下,相较已有压缩域方法最高可节省 15.8% 的码率;与基于像素域重建的分析方法相比,该方法还可分别降低约 83.6% 的码率开销和 44.8% 的推理时间。Akbari等<sup>[79]</sup>提出基于深度语义分割的分层图像压缩方法(Deep semantic segmentation-based layered image compression, DSSLIC),将深度语义分割结果作为分层图像压缩的重要先验,以增强压缩表示对高层语义信息的保持能力。Hoang等<sup>[80]</sup>通过编码器与任务解码器之间的匹配机制,使解码端能够恢复语义分割结果,提高压缩表示对语义分割任务的适配能力。Duan等<sup>[81]</sup>则从压缩域语义推理出发,将已有像素域视觉模型适配到量化潜变量上,使视觉任务能够直接基于压缩表示执行,减少完整图像重建的需要。总体而言,上述方法通过联合优化压缩表示与下游任务网络,使机器视觉分析逐步由“先重建、后分析”的串行处理模式转向面向压缩表示的直接推理或协同推理,在降低像素重建依赖的同时,为进一步减少传输码率与推理开销提供了新的实现路径。

在人机共同消费视觉内容的应用场景中,智能图像编码不仅需要保持机器视觉任务性能,还需兼顾必要的人眼重建质量。因此,可伸缩编码逐渐成为人机协同图像压缩的重要技术路线,其基本框架如图5所示。图5中的语义层、结构层和纹理层为概念性划分,旨在表征机器分析与人眼重建对信息的差异化需求,并不代表各类方法均采用统一的三级编码结构。具体方法可采用基础层与增强层、潜变量分组或残差补偿等组织方式,不同层次的信息亦可能存在交叠。Liu等<sup>[82]</sup>提出语义至信号的可伸缩图像压缩框架,将机器视觉所需的语义信息与人眼重建所需的信号信息进行分层编码,使基础码流能够支持机器分析,增强码流则用于改善人眼观看质量。Choi等<sup>[83]</sup>提出面向人机视觉的可伸缩图像编码器,将潜变量划分为基础层和增强层,其中基础层主要承载机器视觉任务所需信息,增强层用于补充图像重建细节。实验结果表明,该方法在机器视觉任务上相较最佳替代方案可节省 37%~80% 的码率,同时保持与先进图像编解码器相当的重建质量。Hu等<sup>[84]</sup>、Yang等<sup>[85]</sup>和 Tatsumi等<sup>[86]</sup>分别从可伸缩图像编码、可伸缩人脸编码以及显式残差增强层设计等角度出发,探索机器语义信息与人眼视觉信息的分层表示与解耦。Cui等<sup>[87]</sup>将语义信息与视觉显著性信息共同引入可伸缩图像编码过程,使压缩表示能够同时服务机器任务分析与人类视觉感知。Shen等<sup>[88]</sup>提出渐进式条件编码框架,在 COCO2017 数据集上实现对目标检测、实例分割和图像重建等任务的协同支持。总体而言,上述方法通过对机器任务相关语义信息与人眼重建所需视觉信息进行分层表示和渐进传输,使同一码流能够根据不同终端需求灵活支持机器分析与人类观看,为构建统一的人机协同视觉编码体系提供了重要技术基础。

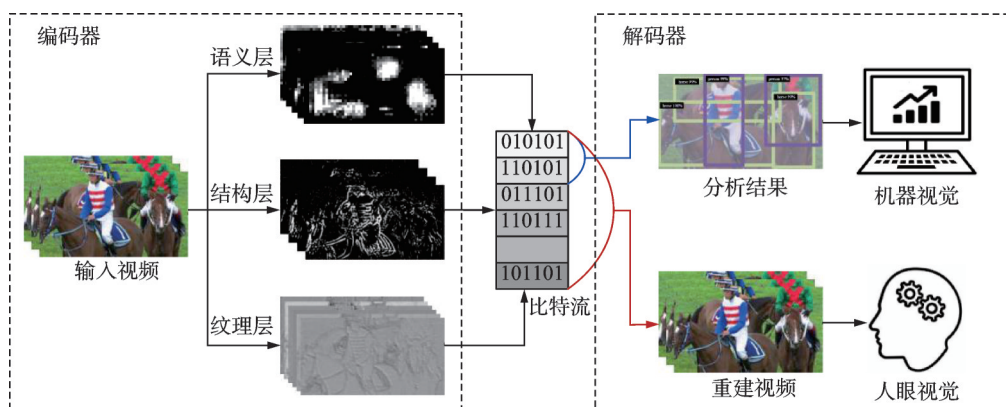


图5 面向人机协同视觉的智能图像与视频编码概念框架

Fig.5 Conceptual framework of intelligent image and video coding for human-machine collaborative vision

面向实际应用的智能图像编码不仅需要关注压缩效率,还需进一步考虑前端感知、隐私保护、在线适配以及端云协同等系统级问题。Gao等<sup>[89]</sup>提出数字视网膜概念,通过视频流、特征流和模型流在端-边-云之间的协同处理,分别支持人眼观看、机器视觉分析和模型更新,提高城市大规模视觉数据的传输与分析效率。Lou等<sup>[90]</sup>进一步提出模型中心的前端视觉感知范式,强调摄像端应根据后端智能分析模型的需求进行自适应采样与编码,从而增强前端感知与后端任务之间的协同性。针对协同视觉分析中的隐私安全问题,Azizian等<sup>[91]</sup>引入隐私保护自编码器,使云端能够完成目标检测等视觉任务,同时降低敏感视觉内容被恢复的风险。Li等<sup>[92]</sup>提出面向在线计算机视觉服务的自适应压缩方法(AdaCompress),根据在线计算机视觉服务中的网络状态与任务需求动态调节压缩强度,以提高实际部署环境下的传输适应能力。Li等<sup>[93]</sup>则将隐式神经表示引入人机协同图像压缩,为神经场表示、连续视觉信号建模及在线自适应编码提供了新的研究思路。上述研究表明,智能图像编码正由单一压缩模型优化逐步拓展至前端感知、网络传输与下游任务协同设计,以更好地适应复杂动态环境下的视觉信息处理需求。与此同时,模型效率、跨任务泛化能力以及统一评价体系也逐渐成为智能图像编码研究的重要方向。Ladune等<sup>[94]</sup>、Zhang等<sup>[95]</sup>和Yu等<sup>[96]</sup>分别从熵模型优化、预训练视觉骨干的低秩适配以及参数高效多任务适配等角度出发,在降低模型开销的同时提升压缩模型对不同任务和视觉模型的适应能力。Choi等<sup>[97]</sup>发布 CompressAI-Vision 开源框架,为图像与视频压缩方法在不同计算机视觉任务中的统一测试与性能比较提供了工具基础。Harell等<sup>[98]</sup>从率失真理论角度分析面向机器视觉的编码问题,并在图像分类、目标检测、实例分割和语义分割等任务中验证理论指导下编码优化的有效性。上述研究表明,轻量化模型设计、跨任务与跨模型泛化,以及统一的评价工具和理论分析框架,正在成为提升智能图像编码可部署性、可复用性与可比较性的重要基础。

总体而言,上述智能图像编码方法主要通过端到端任务驱动优化、压缩域语义分析、编码前端语义增强、人机协同可伸缩编码以及系统级自适应等方式,提高压缩表示对机器视觉任务的支撑能力,相关代表性方法及其技术特点比较如表2所示。相较传统以像素重建为核心的编码方法,上述方法能够直接围绕目标检测、语义分割、实例分割和人脸识别等下游任务学习更加紧凑且具有任务判别性的视觉表示,并在低码率条件下优先保持目标轮廓、语义结构及其他任务关键信息,从而提升机器视觉分析的编码效率。然而,智能图像编码通常依赖深度网络联合训练及任务损失约束,其性能容易受到训练数据分布、下游任务类型、视觉模型结构以及人机视觉需求差异等因素影响,跨任务与跨模型泛化能力仍有待提升,与现有编码标准及软硬件体系的兼容性也相对有限。此外,生成先验、隐私保护、在线自适应和多任务适配等技术虽然进一步拓展了智能图像编码的应用范围,但也带来了模型复杂度增加、推理开销上升以及评价标准不统一等问题。未来,如何在保证机器视觉任务性能的基础上,进一步增强跨任务与跨模型泛化能力,降低端侧部署与推理复杂度,构建统一的机器视觉率失真评价体系,并兼顾人眼重建质量与隐私安全需求,仍是面向机器视觉的智能图像编码进一步发展的重要研究方向。

## 2.2 面向机器视觉的智能视频编码方法

早期面向机器视觉的智能视频编码研究主要围绕低码率视频理解、评价基准构建以及任务自适应编码机制展开。与静态图像相比,视频不仅包含单帧中的空间结构与语义信息,还蕴含运动线索、跨帧目标关联以及长时行为演化等时序信息,因此压缩失真对视频目标检测、目标跟踪、动作识别和视频理解等任务的影响更加复杂。Duan等<sup>[99]</sup>较早提出视频编码应由传统“压缩-重建”范式向协同压缩与智能分析范式转变,强调像素编码、特征表示与下游视觉分析之间应进行协同设计。Yang等<sup>[6]</sup>进一步从紧凑视觉表示压缩角度系统梳理面向机器视觉的视频编码方法,并将像素重建、特征表示和智能分析纳入统一技术框架。Tian等<sup>[100]</sup>提出融合传统视频编码与神经语义压缩的编码框架,并构建涵盖动作识别、多目标跟踪和视频目标分割的低码率视频理解基准。在UCF101、HMDB51和Kinetics400数据集

表 2 面向机器视觉的智能图像编码技术比较

Table 2 Comparison of intelligent image coding techniques for machine vision

| 文献      | 核心理念                                | 关键技术                                                                             | 主要优势                                           | 性能表现                                                                                                                |
|---------|-------------------------------------|----------------------------------------------------------------------------------|------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|
| [60-66] | 将检测、实例分割或人机联合目标直接融入神经网络压缩训练。        | 编解码器与任务网络联合微调、任务损失约束、潜变量分区聚合、内容自适应推理、可变码率与自适应量化。                                 | 突破单纯像素重建目标,使压缩表示主动保留目标轮廓、语义结构和任务判别线索,并适应多码率部署。 | 约 0.1 bpp 下检测精度提升约 4.6%; 相较 VVC 在检测和实例分割上分别节省 37.87% 和 32.90% BD-rate; 内容自适应推理平均节省 3.66% BD-rate。                   |
| [67-72] | 在编码前端对任务相关内容进行筛选、增强和差异化质量分配。        | 识别感知预变换、语义预处理器、语义保持压缩、非均匀重要性质量图、任务可切换预处理。                                        | 在图像进入压缩器之前抑制背景冗余并突出目标结构,使有限比特更集中于机器任务关注区域。     | 预处理增强方法在保持下游任务准确率时,相较 BPG 等传统编码器可节省约 20% 码率。                                                                        |
| [73-77] | 利用高层语义概念、生成模型和结构先验替代对全部像素细节的保真传输。   | 概念编码、语义先验建模、层级生成式人脸压缩、StyleGAN 先验、草图结构先验。                                        | 适合人脸等结构化对象,在极低码率下优先保存身份、轮廓、关键点和语义约束。           | 层级生成式人脸压缩最高节省 99.6% 码率且仅牺牲约 0.8% 分割 mIoU; StyleGAN 先验在 0.002 4 bpp 下使 FWIoU 较 VVC 提升 52.44%; 草图先验节省 40.9%~86.6% 码率。 |
| [78-81] | 减少对完整像素域重建的依赖,使压缩表示与任务网络共同适配。       | 学习式压缩域语义分割、动态/静态通道选择、DSSLIC、编码器-解码器匹配、压缩域语义推理架构。                                 | 能够直接利用压缩域或语义域信息完成分析,降低解码成完整图像后再推理的冗余。          | 学习式压缩域语义分割在相同 mIoU 下较已有压缩域方法最高节省 15.8% 码率,较像素域方法节省约 83.6% 码率和 44.8% 推理时间。                                           |
| [82-88] | 面向人机共同消费,将机器语义层与人眼重建层进行分层组织。        | 语义至信号可伸缩压缩、基础层/增强层潜变量划分、显式残差增强层、语义显著性建模、渐进条件编码。                                  | 同一码流可按需支持机器分析和人类复核,兼顾任务准确率与主观重建质量。             | 可伸缩图像编码相较最佳替代方案在机器视觉任务上节省 37%~80% 码率,并保持与先进图像编解码器相当的重建质量。                                                           |
| [89-98] | 面向实际系统补充前端感知、隐私保护、在线适配、模型效率和统一评价能力。 | 数字视网膜、模型中心前端感知、隐私保护自编码器、AdaCompress、隐式神经表示、熵模型优化、低秩适配、CompressAI-Vision、率失真理论建模。 | 推动智能图像编码从单模型压缩走向感知、传输、任务、隐私和评价体系的协同优化。         | 二值概率模型在学习式图像压缩挑战赛 (Challenge on learned image compression, CLIC) 条件下较高斯或拉普拉斯潜变量模型节省约 18% 码率;统一评估框架提升了不同方法的可比较性。     |

上,该方法在保持相同识别性能的条件下,相较 VVC 可节省超过 50% 的码率。在编码控制方面, An 等<sup>[101]</sup>则从时间维度引入任务信息,利用目标跟踪结果进行自适应时间重采样,在目标检测任务上 BD-mAP 分别提升了 1.29%、0.47% 和 2.44%,并在目标跟踪任务上取得最高 6.86% 的 Bjøntegaard 平均多目标跟踪准确率差 (Bjøntegaard delta multiple object tracking accuracy, BD-MOTA) 提升。上述研究初步建立了面向低码率视频理解的评价与优化框架,并推动任务信息由空间域编码控制进一步拓展至时间维度,为后续从运动建模、时序语义保持和端到端联合优化等方向提升智能视频编码效率奠定了基础。

在低码率视频理解与任务自适应编码研究的基础上,智能视频编码开始向人机协同码流、语义保持表示以及多任务联合优化等方向拓展,以增强压缩表示对不同视觉消费需求的适应能力。Lin等<sup>[102]</sup>提出深度可伸缩视频编码方法(DeepSVC),通过可伸缩视频编码结构使同一码流能够同时支持机器视觉分析与人眼观看。Cui等<sup>[103]</sup>提出双上下文可伸缩视频编码方法,分别对机器分析上下文和人眼重建上下文进行分层建模,以实现不同视觉消费需求之间的协同。Benjak等<sup>[104]</sup>设计学习式混合视频编码器,使视频重建、目标检测和实例分割等任务共享统一的压缩表示。实验结果表明,与仅采用帧内编码结构的模型相比,引入帧间预测机制后可获得约58%的BD-rate改善。Tian等<sup>[105]</sup>提出SMC++,利用无监督掩膜学习机制筛选并保持视频中的任务相关语义信息;在0.04 bpp条件下,相较面向人眼的智能视频编码器和传统视频编码标准,其动作识别Top-1准确率分别提升22%和10%。Zhang等<sup>[106]</sup>提出时空解耦潜变量组合模型(Spatial-temporal decoupled latent composition, STDLC),通过时空解耦的潜变量组合对空间语义与时间动态信息进行紧凑表示,从而提升动作识别任务的率失真性能,并在K400数据集上相较传统视频编码标准节省约22.58%的BD-rate。Sheng等<sup>[107]</sup>提出通用神经视频编码框架(Versatile neural video coding, VNVC),通过共享中间特征支持视频重建、视频增强与机器视觉分析。实验结果表明,相较采用IPP帧结构的VVC参考软件,该方法在视频重建方面分别基于PSNR和MS-SSIM节省约7.0%和49.9%的BD-rate,同时保持较好的机器视觉任务性能。总体而言,上述研究表明,智能视频编码正由面向单一重建或单一任务的优化模式,逐步转向对空间语义、时间动态及人机视觉需求进行统一建模,为构建兼顾机器分析效率与人眼重建质量的通用视频压缩表示提供了新的技术路径。

随着研究不断深入,面向机器视觉的智能视频编码逐渐由基础的语义保持与任务优化,拓展至复杂失真补偿、任务自适应控制、隐私保护以及多路视频资源调度等系统级方向。Dao等<sup>[108]</sup>针对感兴趣区域保持较高质量、背景区域被强压缩的人机协同编码场景,提出梯度引导的扩散修复方法,利用内容感知监督与自适应细化机制恢复退化背景,改善重建画面的视觉质量与整体一致性,以满足机器分析之外的人眼观看需求。Zhu等<sup>[109]</sup>面向多路视频智能分析场景提出自适应神经编解码与资源调度系统,在带宽和时延约束下联合优化视频码率与计算资源分配策略,最高可提升约15%的分析准确率。针对视觉数据传输过程中的隐私安全问题,Gou等<sup>[110]</sup>将灵活自编码器引入VCM隐私保护,在抑制敏感视觉细节泄露的同时兼顾目标检测mAP与码率性能。Sun等<sup>[111]</sup>提出对称熵约束视频编码框架(SEC-VCM),通过双向熵约束建立视频解码过程与视觉骨干特征提取过程之间的语义对应关系,保留任务相关信息并抑制冗余内容。Ge等<sup>[112]</sup>在深度视频压缩框架中引入任务感知编码控制机制,使编码器能够根据不同下游任务需求动态调整压缩表示,相较已有端到端视频编码方法可节省约25%的码率。上述研究表明,智能视频编码的优化目标正由单一的语义信息保持逐步扩展至任务适应性、复杂失真鲁棒性、系统资源约束与隐私安全等多维需求的协同优化,进一步推动其从面向单一压缩模型的设计走向面向实际智能视频系统的联合优化。

总体而言,上述智能视频编码方法并非智能图像编码在时间维度上的简单延伸,而是需要围绕空间语义保持、时间冗余建模、运动信息利用、跨帧任务一致性以及系统部署约束进行协同设计,相关代表性方法及其技术特点比较如表3所示。现有方法主要通过低码率视频理解、人机协同可伸缩码流、语义保持表示、时空潜变量建模、任务感知编码控制、压缩失真补偿以及多路视频编解码与资源调度等方式,提高压缩表示对动作识别、目标跟踪、视频目标分割及多路视频分析等下游任务的支撑能力。相较传统以像素重建为核心的视频编码方法,上述方法能够更加充分地利用视频中的运动线索、跨帧关联与时序语义,使有限码率优先用于任务相关信息的表示与传输,从而提升低码率条件下的机器视觉分

析性能。然而,与智能图像编码相比,智能视频编码通常涉及更复杂的时空联合建模与多帧依赖关系,因而具有更高的训练复杂度、推理开销与系统部署成本,其性能也更容易受到视频内容变化、运动模式差异、时序跨度、下游任务模型结构以及长时依赖建模能力等因素影响。如何在保证机器视觉任务性能的基础上,进一步降低端到端视频编码与推理复杂度,增强跨场景、跨任务与跨模型泛化能力,并在人眼重建质量、机器分析性能、隐私安全和实时传输等多重需求之间实现有效权衡,仍是面向机器视觉的智能视频编码进一步发展的重要研究方向。

表 3 面向机器视觉的智能视频编码技术比较

Table 3 Comparison of intelligent video coding techniques for machine vision

| 文献         | 核心思想                                      | 关键技术                                                    | 主要优势                                            | 性能表现                                                                                                                           |
|------------|-------------------------------------------|---------------------------------------------------------|-------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------|
| [6,99-101] | 围绕低码率视频理解建立评价基准,并利用任务信息指导空间和时间编码决策。       | VCM 范式、紧凑视觉表示、低码率视频理解基准、可学习 Codec Control、目标跟踪驱动的时间重采样。 | 从视频检测、跟踪、动作识别和理解任务出发,显式考虑运动线索、跨帧关联和长时语义保持。      | 低码率视频理解方法相较于传统视频编码标准节省超过 50% 比特;时间重采样在检测任务上获得 1.29%、0.47% 和 2.44% 的 BD-mAP 改善,在跟踪任务上最高提升 6.86% BD-MOTA。                        |
| [102,103]  | 构建可同时服务机器分析和人眼观看的人机协同视频码流。                | DeepSVC、双上下文可伸缩视频编码、机器分析上下文和人眼重建上下文分层建模。                | 通过可伸缩表示满足不同视觉消费需求,使基础码流优先服务机器任务,增强信息以进一步提升观看质量。 | 在同一码流中兼顾机器视觉任务性能和人眼重建质量,提升视频码流在监控复核等场景中的复用能力。                                                                                  |
| [104-106]  | 在端到端框架中学习语义保持视频表示,并利用帧间结构和时空潜变量提升任务率失真性能。 | 学习式混合视频编码、帧间结构建模、无监督掩膜语义学习、时空解耦潜变量组合、动作识别率任务优化。         | 相较于逐帧图像压缩,更充分利用视频时序冗余和语义连续性,提升动作识别等动态任务的低码率鲁棒性。 | 学习式混合视频编码引入帧间结构后获得约 58% 的 BD-rate 改善;SMC++ 在 0.04 bpp 下较智能视频编码器和传统标准分别提升 22% 和 10% Top-1 准确率;STDLC 在 K400 上节省约 22.58% BD-rate。 |
| [107]      | 将视频重建、视频增强和机器分析统一到可变神经视频编码结构中。            | VNVC、统一潜变量表示、重建增强联合优化、任务感知训练。                           | 同一压缩表示同时兼顾人眼视觉质量和机器分析性能,体现人机协同优化优势。             | 在视频重建质量上取得明显提升,PSNR 和 MS-SSIM 指标下分别节省约 7.0% 和 49.9% BD-rate,同时保持较好的机器视觉任务性能。                                                   |
| [108-112]  | 面向复杂真实场景进一步考虑低码率失真补偿、隐私保护、任务适应和多路资源调度。    | 梯度引导扩散修复、自适应神经编解码与调度、隐私保护自编码器、对称熵约束、任务感知编码器控制。          | 将智能视频编码从单一语义保持扩展到带宽、时延、隐私、安全和任务模型变化等系统级约束。      | 多路视频分析调度最高提升约 15% 分析准确率;任务感知编码器控制相较于有端到端视频编码方法节省约 25% 码率。                                                                      |

### 2.3 面向机器视觉的智能特征编码方法

面向机器视觉的智能特征编码主要源于协同智能和边缘-云分割推理等应用需求<sup>[114]</sup>。在此类场景中,前端设备无需将完整图像或视频传输至云端,而是首先执行视觉网络的部分计算,再将产生的中间特征进行压缩与传输,由后端继续完成后续推理。图6展示了Alvar等<sup>[113]</sup>利用ResNet-34提取的深层特征经量化后的平铺结果,可以发现图中不同通道呈现不同的空间响应,体现了中间特征与原始像素表示的差异。Bajic等<sup>[115]</sup>较早系统地分析了协同智能面临的关键挑战,指出模型切分能够在端侧计算开销、网络传输带宽和云端推理能力之间实现灵活权衡,但深度网络产生的中间特征通常具有较大的数据规模,其传输开销可能成为新的系统瓶颈。Shao等<sup>[116]</sup>提出BottleNet++,通过轻量化神经网络联合优化中间特征压缩与信道传输,并在训练中考虑信道噪声的影响。相较直接传输未经压缩的中间特征,该方法在加性高斯白噪声信道下最高可将带宽需求降至原来的1/64,在二元擦除信道下最高可实现256倍的比特压缩比,同时将分类准确率下降控制在2%以内。Yuan等<sup>[117]</sup>则从边缘视觉分析系统角度提出可伸缩特征压缩方法,强调特征码流设计需要综合考虑端侧推理时延、传输带宽以及云端任务性能,从而实现计算与通信资源的协同优化。不同于具有明确空间结构和像素分布规律的图像或视频信号,深度特征通常具有高维、稀疏、通道相关性强以及任务依赖性显著等特点,因此早期研究主要围绕特征张量的统计特性及其紧凑表示展开。Choi等<sup>[118]</sup>面向协同目标检测提出深度特征压缩方法,并通过压缩增强训练减弱特征失真对检测性能的影响,在基本保持检测准确率的条件下最高可减少约70%的通信开销。Singh等<sup>[119]</sup>从训练阶段引导网络学习具有更高可压缩性的中间特征,使特征压缩性能提高约一个数量级,同时获得一定的正则化收益。Choi等<sup>[120]</sup>进一步利用前向与后向预测机制降低深度特征张量中的残差冗余。Shah等<sup>[121]</sup>则利用退火收缩机制学习更加紧凑的特征表示。总体而言,上述研究表明,深度特征在数据结构、统计分布和任务敏感性等方面均与像素信号存在明显差异,因此难以直接沿用传统像素编码机制,需要围绕通道相关性、尺度结构、激活分布及任务重要性等特征属性重新设计表示与压缩方法。



图6 量化深层特征张量的平铺可视化示例(引自文献[113]图2)

Fig.6 Example of tiled visualization of a quantized deep feature tensor (from Fig.2 in Ref. [113])

随着MPEG面向机器的特征编码(Feature coding for machines,FCM)标准化工作的持续推进,面向机器视觉的特征编码逐步由单一算法设计拓展至统一编码流程、测试模型构建、复杂度优化以及压缩后特征恢复等方向。Eimon等<sup>[122]</sup>围绕面向机器视觉的特征编码标准及其消费级应用开展研究,指出在边缘设备与云端之间直接传输中间特征,可有效降低完整图像或视频远程传输带来的带宽开销。该研究进一步分析了FCM在智能监控、智能交通和消费电子等场景中的应用模式,实验结果表明,相较远程推理方案,FCM在保持任务准确率的同时可平均降低75.90%的码率需求。随后,Eimon等<sup>[123]</sup>提出面向可伸缩机器视觉的特征编码测试模型,通过统一特征提取、特征打包、编码传输和任务推理等处理流程,为不同视觉任务下特征压缩方法的性能比较提供标准化测试基础。实验结果显示,该测试模

型在多种机器视觉任务上可实现平均 85.14% 的码率降低,进一步验证了标准化特征编码流程在提升协同推理传输效率方面的有效性。面向实时视觉分析需求,Perera 等<sup>[124]</sup>对 FCM 处理流程进行复杂度优化,重点降低特征编码与解码过程中的计算时延,以增强特征压缩在实时应用场景中的部署能力。总体而言,上述研究推动特征编码由早期面向单一任务和特定网络的算法优化,逐步走向统一测试流程、低复杂度实现和压缩后特征恢复,为 FCM 的标准化与实际部署奠定了重要技术基础。

多尺度特征压缩是智能特征编码领域的重要研究方向之一。目标检测和实例分割等视觉网络通常同时依赖浅层高分辨率特征与深层高语义特征,不同尺度特征在目标定位、类别判别和边界恢复等任务环节中承担不同作用,因此需要针对尺度间冗余关系与任务重要性设计更加精细的联合建模和比特分配机制,其整体框架如图 7 所示。图中  $p_2 \sim p_5$  表示端侧视觉模型提取的不同尺度特征,  $P_2' \sim P_5'$  表示经过压缩、传输和解码后的特征,云端通过对  $P_5'$  进行池化生成  $P_6'$ 。Zhang 等<sup>[125]</sup>提出多尺度特征压缩(Multi-scale feature compression, MSFC)方法,针对掩膜区域卷积神经网络(Mask region-based convolutional neural network, Mask R-CNN)等多任务网络中多尺度特征传输开销较大的问题,设计多尺度特征融合、单流特征编解码和多尺度特征重建模块,将原本分散的多层特征整合为更加紧凑的传输表示。该方法可将 32 bit 浮点特征平均压缩至 0.012 bit,同时将目标检测和实例分割准确率下降控制在 2% 以内,体现了多尺度特征联合编码在协同推理中的压缩潜力。Chen 等<sup>[126]</sup>提出层级残差特征压缩方法,利用基础层特征保留主要语义信息,并通过残差增强层补偿低码率压缩造成的细节损失,从而提高多任务视觉分析中的特征重建质量。该方法在 COCO 和 OpenImages 数据集上,针对目标检测和实例分割任务 BD-rate 分别节省了 84.06%/88.46% 和 91.58%/88.72%。Ding 等<sup>[127]</sup>从特征稀疏性出发构建层级压缩框架,通过识别并优先保留对任务输出贡献更大的稀疏激活区域,减少任务无关特征的传输开销。随着研究深入,多尺度特征编码逐步由简单的特征融合与层级压缩转向跨尺度相关性建模和任务重要性驱动的联合优化。Liu 等<sup>[128]</sup>提出多尺度互压缩方法,针对不同尺度特征之间存在显著冗余相关性的特点,通过小尺度基础特征预测大尺度特征,减少跨尺度冗余信息,相较 VVC 特征锚点和 VVC 图像锚点 BD-rate 分别节省了 90.10% 和 74.97%。Pan 等<sup>[129]</sup>采用多感受野卷积网络对多尺度特征进行压缩,利用不同尺度的感受野联合建模局部结构与全局语义信息,以增强压缩特征对目标检测和实例分割任务的适应能力。Chen 等<sup>[42]</sup>提出语义一致的整体感知网络,通过尺度间去相关与自适应特征融合挖掘全局和局部相关性,减少多尺度特征的冗余传输。同时,结合上下文引导的细节重建与敏感性感知的信号和语义保真约束,在压缩过程中兼顾局部细节与整体语义,为目标检测和实例分割提供紧凑的特征表示。Zhang 等<sup>[130]</sup>进一步建模层级语义相关性与空间语义相关性,使不同尺度之间的语义依赖能够直接参与压缩表示学习。Liu 等<sup>[131]</sup>则依据多尺度特征的重要程度进行端到端比特分配,使有限编码资源优先分配至对下游任务贡献更大的尺度与通道。总体而言,多尺度特征编码正由各尺度相对独立的压缩方式逐步演进为跨尺度相关性、空间语义依赖与任务重要性协同建模的联合压缩范式,其核心在于充分抑制尺度间冗余的同时,优先保持对下游视觉任务具有较高判别价值的结构与语义信息。

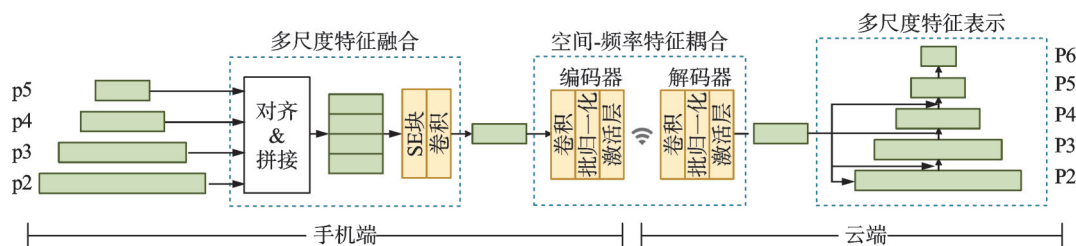


图7 多尺度特征编码框架

Fig.7 Framework of multi-scale feature coding

此外,智能特征编码的应用范围也在持续拓展,由单一图像特征压缩逐步延伸至双目视觉、视频分析、人机协同以及结构化视觉任务等场景。Jin等<sup>[132]</sup>将联合视觉特征压缩应用于双目图像机器编码,针对左右视图之间存在显著相关性的特点,通过联合压缩视觉特征,使立体图像表示与下游机器分析共享更加紧凑的中间表示。Chen等<sup>[133]</sup>提出面向智能感知的中间深度特征压缩方法,强调前端设备无需传输完整视觉信号,而应优先传输能够支撑后端分析的紧凑中间表征,从而降低智能感知系统中的通信开销。Sun等<sup>[134]</sup>提出可变换视频特征压缩框架(Transformable video feature compression, Trans-VFC),针对视频特征在不同机器任务和码率条件下复用能力有限的问题,引入可变换视频特征压缩机制,使压缩特征能够根据下游任务需求进行适配与转换。Xia等<sup>[135]</sup>从VCM角度提出超越特征与信号二元划分的可伸缩编码范式,尝试在机器语义表示与人眼信号表示之间建立统一的层级编码结构,使同一码流能够按需支持机器分析与人类观看。Huang等<sup>[136]</sup>研究兼顾人眼与机器视觉需求的通用特征提取与压缩方法,探索构建跨任务、跨视觉消费对象的统一特征表示。Yan等<sup>[137]</sup>提出基于压缩特征图的语义可伸缩图像编码方法,使不同语义粒度的下游任务能够通过渐进解码获得相应层次的特征表示。Wang等<sup>[138]</sup>面向可伸缩人脸压缩引入深度特征重建机制,使人脸图像在低码率条件下仍能够较好地保持身份与结构信息。Wang等<sup>[139]</sup>则通过多粒度约束的师生学习获得紧凑人脸特征表示,以降低人脸分析任务中的特征传输与存储开销。上述研究表明,智能特征编码正由面向单一任务的中间特征压缩,逐步向跨场景、跨任务和人机协同条件下的可变换、可伸缩与统一视觉表示演进,其研究重点也由单纯降低特征码率进一步转向提升压缩表示的通用性、可复用性与任务适配能力。

总体而言,上述智能特征编码方法主要面向协同智能和边缘-云分割推理场景,通过对深度网络中间特征进行压缩与传输,减少完整图像或视频信号远程传输所带来的带宽开销,相关代表性方法及其技术特点比较如表4所示。与智能图像编码和智能视频编码相比,特征编码直接作用于机器视觉网络的中间表示,能够围绕通道重要性、尺度相关性、特征稀疏性、任务敏感性以及多任务共享关系等特征属性进行针对性压缩,因此在低码率条件下具有较强的任务信息保持能力和通信效率。同时,随着FCM标准化、多尺度联合压缩和可伸缩特征表示等研究不断推进,特征编码正由面向特定网络和单一任务的中间特征压缩,逐步向统一、可复用和可扩展的机器视觉表示演进。然而,现有方法仍面临特征格式与接口缺乏统一规范、模型切分位置敏感、跨网络结构兼容性不足、跨任务与跨场景泛化能力有限以及标准化与软硬件生态尚不完善等问题。如何构建统一且可复用的特征表示与编码接口,在保证下游任务性能的同时降低特征编解码和端侧计算复杂度,并进一步增强压缩表示在不同模型、任务和应用场景之间的迁移与泛化能力,仍是面向机器视觉的智能特征编码进一步发展的重要研究方向。

表 4 面向机器视觉的智能特征编码技术比较

Table 4 Comparison of intelligent feature coding techniques for machine vision

| 文献        | 核心思想                                | 关键技术                                                     | 主要优势                                     | 性能表现                                                                     |
|-----------|-------------------------------------|----------------------------------------------------------|------------------------------------------|--------------------------------------------------------------------------|
| [115-117] | 面向协同智能和边缘-云分割推理,压缩中间特征以替代完整图像或视频传输。 | 模型切分、可学习瓶颈层、边缘视觉分析、可伸缩特征压缩、端侧时延与云端精度联合考虑。                | 在端侧计算、传输带宽和云端推理之间建立折中,使特征码流直接服务后端视觉任务。   | BottleNet++最高实现64倍甚至256倍带宽压缩,分类准确率下降小于2%;可伸缩特征压缩进一步兼顾延迟和任务精度。            |
| [118-121] | 针对深度特征不同于自然图像的统计特性,学习更适合压缩的特征表示。    | 压缩增强训练、深度特征时空重排、可压缩特征学习、前后向预测、3D CNN二值特征、特征图重排和预量化、退火收缩。 | 从通道、尺度、激活分布和残差冗余出发设计特征压缩方式,避免直接套用像素编码假设。 | 协同目标检测中特征压缩最高减少约70%通信开销且不损失检测准确率;8 bit量化重排后基本保持任务准确率;可压缩特征学习使压缩性提高一个数量级。 |

续表

| 文献        | 核心思想                                       | 关键技术                                                             | 主要优势                                          | 性能表现                                                                                                                                                           |
|-----------|--------------------------------------------|------------------------------------------------------------------|-----------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------|
| [122-124] | 随着 MPEG FCM 推进,特征编码走向统一流程、标准化测试模型和压缩后特征恢复。 | FCM 标准、Feature Coding Test Model、实时 FCM 流程。                      | 为不同视觉任务提供可复用的特征提取、打包、编码、传输和推理流程,增强方法之间的可比性。   | FCM 相较远程推理平均减少 75.90% 码率;测试模型在多任务上平均降低 85.14% 比特率。                                                                                                             |
| [125-131] | 面向检测和实例分割中的多尺度特征,进行跨尺度融合、层级残差补偿和任务重要性比特分配。 | MSFC、多尺度特征融合、层级残差特征压缩、特征稀疏学习、多尺度互压缩、多感受野卷积、层级/空间语义相关性建模。         | 能够同时保留浅层定位细节和深层语义信息,减少尺度间冗余并优先保护对任务贡献更大的特征。   | MSFC 将 32 bit 浮点特征平均压缩到 0.012 bit 且检测/分割准确率下降小于 2%;层级残差方法在 COCO 和 OpenImage 上 BD-rate 节省 84.06%/88.46%和 91.58%/88.72%;多尺度互压缩较 VVC 特征锚点和图像锚点节省 90.10% 和 74.97%。 |
| [132-137] | 将特征编码从单一图像特征扩展到双目视觉、视频特征和人机协同可伸缩表示。        | 双目联合视觉特征压缩、中间深度特征传输、TransVFC、超越特征与信号的可伸缩编码、通用特征提取与压缩、语义可伸缩特征图解码。 | 使压缩特征能够在不同视图、不同机器任务和不同视觉消费对象之间复用,推动语义层与信号层统一。 | 相关方法主要提升跨任务和跨场景复用能力,为同一码流按需支持机器分析与人眼观看提供了扩展路径。                                                                                                                 |
| [138,139] | 面向人脸等结构化视觉对象,学习紧凑且可重建的身份相关特征表示。            | 可伸缩人脸压缩、深度特征重建、多粒度约束师生学习、紧凑人脸特征表示。                               | 利用人脸身份、结构和语义先验降低低码率下的特征传输与存储开销,并保持识别与分析能力。    | 在低码率人脸压缩和人脸分析场景中增强身份与结构信息保持能力,体现特征编码与结构化任务结合的优势。                                                                                                               |

3 总结与展望

随着智能视觉应用的快速发展,视觉数据的主要消费对象已由传统的人眼观看逐步拓展至机器分析与智能决策。面向机器视觉的编码技术正是在这一背景下形成的重要研究方向,其核心目标不再局限于提升压缩图像或视频的像素重建质量与主观视觉质量,而是在有限码率、计算资源和传输带宽约束下,尽可能保持目标检测、语义分割、实例分割、目标跟踪、动作识别和人脸识别等下游机器视觉任务的分析性能。本文围绕机器视觉编码的关键技术与发展脉络进行了系统梳理。首先,分析传统图像与视频编码标准面向机器视觉任务时存在的优化目标不匹配问题,并总结任务感知码率分配、机器视觉率失真优化以及编码工具调整与重构等代表性技术;随后,从智能图像编码、智能视频编码和智能特征编码 3 个方面,系统梳理基于深度学习的智能编码方法,并分析其由像素重建优化向语义信息保持、任务性能保持以及边缘-云协同推理演进的技术路径。

总体而言,传统编码标准面向机器视觉任务的优化方法具有较好的标准兼容性和工程可部署性,能够依托 JPEG、AVC、HEVC 和 VVC 等成熟编码框架及其软硬件生态进行渐进式改进,适用于视频监控、自动驾驶和边缘视觉分析等既有系统的升级。然而,这类方法本质上仍建立于以像素重建为核心的传统编码体系之上,机器视觉任务通常仅通过外部先验、任务损失或编码参数调控等方式参与编码决策,对任务相关语义信息与任务无关冗余之间的关系缺乏显式建模。相比之下,基于深度学习的智

能编码方法能够利用神经网络的表征学习与端到端联合优化能力,直接学习更具任务相关性和判别性的紧凑视觉表示,在低码率条件下表现出更强的语义与任务信息保持能力。其中,智能图像编码主要关注单帧空间结构、语义信息及人机协同重建;智能视频编码进一步引入运动信息、跨帧关联与时序语义建模;智能特征编码则面向协同智能与边缘-云分割推理,通过直接压缩和传输中间特征减少视觉数据通信开销,并逐步向统一、可复用和可伸缩的机器视觉表示发展。

尽管现有研究已在编码效率、任务性能保持和协同推理等方面取得显著进展,机器视觉编码仍处于持续演进阶段,在评价体系、跨任务与跨模型泛化、标准兼容性、模型复杂度、人机协同以及隐私安全等方面仍存在一系列亟待解决的问题。尤其随着视觉基础模型、多模态智能、边缘智能和生成式模型等技术快速发展,机器视觉编码的研究对象、任务需求与系统形态也在不断拓展。未来研究需要进一步突破面向单一任务、单一模型和固定场景的优化范式,推动编码表示、任务模型与实际视觉系统之间更加统一和高效的协同设计。基于上述分析,未来可重点围绕以下几个方面开展研究。

### (1) 构建统一的机器视觉编码评价体系

当前,面向机器视觉的编码研究涵盖图像、视频和特征等不同编码对象,下游应用涉及图像分类、目标检测、语义分割、实例分割、目标跟踪、动作识别等多类视觉任务。由于不同研究采用的数据集、任务网络、编码配置及评价指标存在较大差异,不同方法之间往往难以进行统一、客观和公平的性能比较。传统 PSNR、SSIM 和视频多方法评估融合 (Video multimethod assessment fusion, VMAF) 等质量指标主要衡量像素重建或人眼感知质量,难以准确反映压缩失真对机器视觉任务的实际影响;而 mAP、mIoU、Top-1 Accuracy 和 MOTA 等任务指标又与具体数据集、视觉模型及任务设置高度相关,尚缺乏统一的机器视觉率失真解释框架。面向未来,应建立覆盖图像、视频和特征编码的统一评价体系,在规范数据集、任务模型、码率计算方式和测试条件的基础上,综合考虑码率、任务性能、人眼重建质量、计算复杂度、能耗和传输时延等因素,并探索能够刻画压缩表示任务效用的通用失真度量与率失真模型,从而为不同机器视觉编码方法的公平比较和标准化发展提供统一评价基础。

### (2) 提升跨任务、跨模型和跨场景泛化能力

现有智能编码方法大多围绕特定训练数据、下游任务和视觉模型进行联合优化,在固定测试条件下能够获得较高的压缩效率,但当任务类型、模型结构或场景分布发生变化时,其性能往往出现不同程度的下降。例如,针对目标检测优化的编码模型未必能够有效适配语义分割、目标跟踪或动作识别任务,在特定数据集上学习得到的任务先验也可能难以迁移至开放环境和未知场景。对于智能视频编码与特征编码,运动模式、时序跨度、模型切分位置以及中间特征结构的变化还会进一步增加表示迁移与任务适配的难度。实际部署中,下游视觉模型升级还可能影响原有压缩表示的可用性。对于输出重建图像或视频的系统,原有码流在解码接口兼容时仍可使用,但需重新评估新模型的任务性能;对于直接传输中间特征的系统,特征维度、分布或语义含义的变化则可能使新模型无法直接利用原有表示。因此,可结合稳定的特征接口与模型版本管理,通过轻量化适配模块、知识蒸馏和参数高效微调降低重新训练与部署的成本;若原有表示已丢失新任务所需的信息,则仍需更新编码模型。未来应由面向“特定任务-特定模型”的联合优化逐步转向更加通用的视觉表示压缩,探索多任务联合学习、视觉基础模型适配、任务可切换编码、参数高效适配、自监督与无监督表示学习以及可伸缩特征接口等技术,增强压缩表示在不同任务、视觉模型和应用场景之间的迁移能力与鲁棒性,为机器视觉编码在开放环境中的实际应用奠定基础。

### (3) 降低智能编码模型复杂度并增强实时部署能力

端到端智能编码虽然能够突破传统像素重建目标的限制,但通常依赖复杂的深度神经网络及联合

优化过程,带来较高的训练成本、推理时延和存储开销。尤其在智能视频编码和智能特征编码中,时空相关性建模、多尺度特征处理、概率熵模型以及任务网络联合推理进一步增加了系统复杂度,与摄像端和边缘设备对低功耗、低时延及实时处理的需求存在明显矛盾。面向实际部署,应更加重视压缩效率、任务性能与计算资源之间的协同优化,通过轻量化网络结构、参数高效适配、低复杂度熵模型、动态计算、早退机制、网络剪枝、低比特量化以及硬件友好算子等方式降低编解码与推理开销。同时,还需要结合端-边-云协同架构,根据设备算力、网络带宽和任务时延要求动态划分编码、特征提取与任务推理过程,从系统层面实现计算、通信和存储资源的联合优化。

#### (4) 加强标准兼容与智能编码标准化研究

传统编码标准优化具有成熟的软硬件生态和较好的工程兼容性,但对高层语义与任务需求的建模能力相对有限;基于深度学习的智能编码具有更强的表示学习和任务优化能力,但在比特流规范、模型互操作性、硬件支持及标准化生态等方面仍有不足。未来机器视觉编码需要在标准兼容性与智能表示能力之间建立更加合理的技术衔接。一方面,可继续在 HEVC 和 VVC 等成熟编码框架中引入任务感知码率控制、机器视觉率失真优化、语义辅助信息及智能编码工具,以降低产业升级成本;另一方面,应持续推进 VCM、FCM、JPEG AI 等相关标准化工作,逐步建立面向机器视觉的统一编码架构、测试条件和互操作机制。对于特征编码,还需重点解决特征格式、模型切分接口、网络输入输出规范、测试模型及比特流语法等问题。通过构建开放、稳定且具有互操作性的标准体系,可进一步推动机器视觉编码由算法验证走向规模化工程部署和产业应用。

#### (5) 兼顾人机协同、隐私保护与安全可信需求

在智慧城市、自动驾驶、工业检测和智慧医疗等实际场景中,视觉数据往往需要同时服务机器实时分析与人类复核,单纯围绕机器任务性能进行编码优化难以满足复杂应用需求。面向人机共同消费的编码系统需要在机器分析效率与人眼重建质量之间建立更加灵活的协同机制。可伸缩人机编码、语义与视觉细节分层表示、生成式重建以及特征级传输等技术,为同一码流按需支持机器分析与人类观看提供了可行路径。与此同时,视觉数据通常包含身份、位置、行为及场景环境等敏感信息,压缩与传输过程中的隐私泄露和安全风险也不容忽视。机器视觉编码需要由传统的“码率-失真”优化进一步拓展至“码率-任务-感知-隐私-安全”等多目标协同优化。未来可结合隐私保护表征学习、可控视觉重建、特征加密、安全协同推理和访问权限控制等技术,使编码表示在保持下游任务性能的同时减少敏感信息泄露风险,逐步构建兼顾压缩效率、机器分析、人类复核与数据安全的可信机器视觉编码体系。

## 4 结束语

本文围绕机器视觉编码的关键技术与发展趋势,系统梳理了视觉编码由面向人眼重建向面向机器理解演进的研究脉络。随着图像与视频数据越来越多地直接服务于智能监控、自动驾驶、工业检测和边缘视觉分析等机器视觉应用,编码系统的优化目标正由传统的像素与感知质量保持,逐步转向对任务相关语义、运动信息和深层特征的高效表示与传输。围绕这一转变,本文总结了传统编码标准优化、智能图像编码、智能视频编码和智能特征编码等主要技术路线,并分析了任务感知码率分配、机器视觉率失真优化、端到端任务驱动压缩、人机协同可伸缩编码以及协同智能特征压缩等关键技术。总体而言,机器视觉编码正在由单一压缩环节优化向感知、编码、传输与推理协同设计演进。随着视觉基础模型、边缘智能和多模态感知等技术不断发展,机器视觉编码有望进一步突破传统信号压缩范式,为构建高效、通用、可信且易于部署的智能视觉系统提供重要技术支撑。

## 参考文献:

- [1] 何小海, 李鑫磊, 魏海涛, 等. 深度学习驱动的视频编码: 方法、进展与展望[J]. 数据采集与处理, 2026, 41(2): 515-542.  
HE Xiaohai, LI Xinlei, WEI Haitao, et al. Deep learning-driven video coding: Methods, progress, and perspectives[J]. Journal of Data Acquisition & Processing, 2026, 41(2): 515-542.
- [2] WANG S, WANG H. Discretized Gaussian mixture hyperprior for learned image compression with mask module[C]// Proceedings of 2022 IEEE International Conference on Multimedia and Expo (ICME). Taipei, China: IEEE, 2022: 1-6.
- [3] YUAN F, PAN Z, LEI J, et al. Multi-domain spatial-temporal redundancy mining for efficient learned video compression[J]. IEEE Transactions on Broadcasting, 2025, 71(3): 808-820.
- [4] YUAN F, PAN Z, LEI J, et al. Mining temporal redundancy using long short-term motion aggregation and global-local decorrelation for learned video compression[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2026, 36(4): 5510-5524.
- [5] REICH C, HAHN O, CREMERS D, et al. A perspective on deep vision performance with standard image and video codecs [C]//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA: IEEE, 2024: 5712-5721.
- [6] YANG W, HUANG H, HU Y, et al. Video coding for machines: Compact visual representation compression for intelligent collaborative analytics[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(7): 5174-5191.
- [7] TANG Z, YI X, WANG H. Toward learned image compression for multiple semantic analysis tasks[J]. IEEE Transactions on Broadcasting, 2025, 71(4): 1022-1033.
- [8] FISCHER K, BLUM C, HERGLOTZ C, et al. Robust deep neural object detection and segmentation for automotive driving scenario with compressed image data[C]//Proceedings of 2021 IEEE International Symposium on Circuits and Systems (ISCAS). Daegu, Korea: IEEE, 2021: 1-5.
- [9] CHAO C, MU X, GUO Z, et al. IAMF-YOLO: Metal surface defect detection based on improved YOLOv8[J]. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 5016817.
- [10] ZHOU Z, SIDDIQUEE M M R, TAJBAKHS N, et al. UNet++: Redesigning skip connections to exploit multiscale features in image segmentation[J]. IEEE Transactions on Medical Imaging, 2020, 39(6): 1856-1867.
- [11] WIEGAND T, SULLIVAN G J, BJONTEGAARD G, 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.
- [12] PAN Z, KWONG S, SUN M T, et al. Early MERGE mode decision based on motion estimation and hierarchical depth correlation for HEVC[J]. IEEE Transactions on Broadcasting, 2014, 60(2): 405-412.
- [13] SULLIVAN G J, OHM J R, HAN W J, 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.
- [14] PAN Z, ZHANG Y, KWONG S. Efficient motion and disparity estimation optimization for low complexity multiview video coding[J]. IEEE Transactions on Broadcasting, 2015, 61(2): 166-176.
- [15] PAN Z, LEI J, ZHANG Y, et al. Fast motion estimation based on content property for low-complexity H.265/HEVC encoder [J]. IEEE Transactions on Broadcasting, 2016, 62(3): 675-684.
- [16] BROSS B, WANG Y K, YE Y, 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.
- [17] YUAN F, LEI J, PAN Z, et al.  $\lambda$ -domain rate control via wavelet-based residual neural network for VVC HDR intra coding [J]. IEEE Transactions on Image Processing, 2024, 33: 6189-6203.
- [18] YU-WEN H, AN J, HUANG H, et al. Block partitioning structure in the VVC standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3818-3833.
- [19] PFAFF J, FILIPPOV A, LIU S, et al. Intra prediction and mode coding in VVC[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3834-3847.
- [20] PAN Z, CHEN J, PENG B, et al. Efficient chroma intra prediction via exemplar colorization network for versatile video coding

- [J]. *IEEE Transactions on Multimedia*, 2025, 27: 4713-4724.
- [21] JIN D, LEI J, PENG B, et al. Deep affine motion compensation network for inter prediction in VVC[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(6): 3923-3933.
- [22] ZHAO X, KIM S H, ZHAO Y, et al. Transform coding in the VVC standard[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(10): 3878-3890.
- [23] SCHWARZ H, COBAN M, KARCZEWICZ M, et al. Quantization and entropy coding in the versatile video coding (VVC) standard[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(10): 3891-3906.
- [24] KARCZEWICZ M, HU N, TAQUET J, et al. VVC in-loop filters[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(10): 3907-3925.
- [25] PAN Z, YI X, ZHANG Y, et al. Efficient in-loop filtering based on enhanced deep convolutional neural networks for HEVC [J]. *IEEE Transactions on Image Processing*, 2020, 29: 5352-5366.
- [26] WANG Z, SIMONCELLI E P, BOVIK A C. Multiscale structural similarity for image quality assessment[C]//*Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003. Pacific Grove, CA, USA: IEEE, 2003: 1398-1402.
- [27] WANG Z, BOVIK A C. Mean squared error: Love it or leave it? A new look at signal fidelity measures[J]. *IEEE Signal Processing Magazine*, 2009, 26(1): 98-117.
- [28] HORÉ A, ZIOU D. Image quality metrics: PSNR vs. SSIM[C]//*Proceedings of 2010 20th International Conference on Pattern Recognition*. Istanbul, Turkey: IEEE, 2010: 2366-2369.
- [29] PAN Z, ZHANG H, LEI J, et al. DACNN: Blind image quality assessment via a distortion-aware convolutional neural network[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(11): 7518-7531.
- [30] PAN Z, YUAN F, WANG X, et al. No-reference image quality assessment via multibranch convolutional neural networks[J]. *IEEE Transactions on Artificial Intelligence*, 2023, 4(1): 148-160.
- [31] PAN Z, YUAN F, LEI J, et al. VCRNet: Visual compensation restoration network for no-reference image quality assessment [J]. *IEEE Transactions on Image Processing*, 2022, 31: 1613-1627.
- [32] WANG J, XU M, DENG X, et al. MW-GAN for perceptual quality enhancement on compressed video[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(7): 4224-4237.
- [33] ZHANG M, BAI H, SHANG W, et al. MDEformer: Mixed difference equation inspired transformer for compressed video quality enhancement[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2025, 36(2): 2410-2422.
- [34] YUE J, YE M, JI L, et al. A survey of deep-learning-based compressed video quality enhancement[J]. *IEEE Transactions on Broadcasting*, 2025, 71(4): 977-992.
- [35] PAN Z, YU W, LEI J, et al. TSAN: Synthesized view quality enhancement via two-stream attention network for 3D-HEVC [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(1): 345-358.
- [36] PAN Z, YUAN F, YU W, et al. RDEN: Residual distillation enhanced network-guided lightweight synthesized view quality enhancement for 3D-HEVC[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, 32(9): 6347-6359.
- [37] LEE J S, EBRAHIMI T. Perceptual video compression: A survey[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2012, 6(6): 684-697.
- [38] LUO Z, ZHU C, HUANG Y, et al. VMAF oriented perceptual coding based on piecewise metric coupling[J]. *IEEE Transactions on Image Processing*, 2021, 30: 5109-5121.
- [39] LIU Z, WANG H, SU T. Learned image compression with multi-scale spatial and contextual information fusion[C]//*Proceedings of 2022 IEEE International Conference on Image Processing (ICIP)*. Bordeaux, France: IEEE, 2022: 706-710.
- [40] TANG Z, WANG H, YI X, et al. Joint graph attention and asymmetric convolutional neural network for deep image compression[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 33(1): 421-433.
- [41] PAN Z, ZHANG G, PENG B, et al. JND-LIC: Learned image compression via just noticeable difference for human visual perception[J]. *IEEE Transactions on Broadcasting*, 2025, 71(1): 217-228.

- [42] CHEN R, YUAN F, ZHAO T, et al. SHPNet: Semantic-consistent holistic perception network for efficient multi-scale feature compression[J].IEEE Transactions on Broadcasting, 2026. DOI: 10.1109/TBC.2026.3720881.
- [43] CHEN R, ZHANG L, ZHANG Y, et al. CA-LIC: Chroma-aware learned image compression for machines[J]. IEEE Transactions on Multimedia, 2026. DOI:10.1109/TMM.2026.3722423.
- [44] TANG Z, YI X, WANG H. Collaborative human-machine perception for cross-modal image compression[C]//Proceedings of 2025 International Symposium on Machine Learning and Media Computing (MLMC). Harbin, China: IEEE, 2025: 1-6.
- [45] YI X, WANG H, KWONG S, et al. Task-driven video compression for humans and machines: Framework design and optimization[J]. IEEE Transactions on Multimedia, 2023, 25: 8091-8102.
- [46] PHAM-DINH K, PAKDAMAN F, ZHANG H, et al. Evaluating the emerging MPEG video coding for machines in semantic segmentation[C]//Proceedings of 2025 International Symposium on Multimedia (ISM). Naples, Italy: IEEE, 2025: 141-148.
- [47] PARTANEN T, ANTTILA M, KORTELAHTI R, et al. UVG-VCM: Benchmarking dataset for machine-oriented visual data compression[C]//Proceedings of 2026 18th International Conference on Quality of Multimedia Experience (QoMEX). Cardiff, UK: IEEE, 2026: 1-4.
- [48] STANKIEWICZ O, GRAJEK T, MAĆKOWIAK S, et al. Region-of-interest-based video coding for machines[C]// Proceedings of 2024 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). Niagara Falls, ON, Canada: IEEE, 2024: 1-6.
- [49] RÓŻEK S, STANKIEWICZ O, MAĆKOWIAK S, et al. Video coding for machines using region-of-interest-based retargeting [J]. EURASIP Journal on Image and Video Processing, 2025, 2025(1): 18.
- [50] FISCHER K, FLECKENSTEIN F, HERGLOTZ C, et al. Saliency-driven versatile video coding for neural object detection [C]//Proceedings of ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Toronto, ON, Canada: IEEE, 2021: 1505-1509.
- [51] XIE W, LI Z, LU Y, et al. Firing bits where it matters: Spiking-guided just recognizable distortion modeling for machine-centric video coding[C]//Proceedings of the Fortieth AAAI Conference on Artificial Intelligence. [S.l.]: AAAI, 2026, 40(22): 18674-18682.
- [52] AI D, LIU T, HAN C, et al. Rate-distortion optimization in video coding for machines: A license plate recognition perspective [J]. IEEE Signal Processing Letters, 2026, 33: 1856-1860.
- [53] LIU J, ZHANG Y, WANG X, et al. DT-JRD: Deep transformer-based just recognizable difference prediction model for video coding for machines[J]. IEEE Transactions on Multimedia, 2026, 28: 114-127.
- [54] SUN Y, ZHAO Y, LIU M, et al. Embedding compression distortion in video coding for machines[C]//Proceedings of 2025 IEEE International Conference on Multimedia and Expo (ICME). Nantes, France: IEEE, 2025: 1-6.
- [55] HUANG Z, JIA C, WANG S, et al. Visual analysis motivated rate-distortion model for image coding[C]//Proceedings of 2021 IEEE International Conference on Multimedia and Expo (ICME). Shenzhen, China: IEEE, 2021: 1-6.
- [56] FERNÁNDEZ-MENDUIÑA S, PAVEZ E, ORTEGA A. Image coding for machines *via* feature-preserving rate-distortion optimization[J]. IEEE Transactions on Multimedia, 2026, 28: 4970-4984.
- [57] GOU A, SUN H, ZENG X, et al. Fast VVC intra encoding for video coding for machines[C]//Proceedings of 2023 IEEE International Symposium on Circuits and Systems (ISCAS). Monterey, CA, USA: IEEE, 2023: 1-5.
- [58] FISCHER K, HERGLOTZ C, KAUP A. On intra video coding and in-loop filtering for neural object detection networks[C]// Proceedings of 2020 IEEE International Conference on Image Processing (ICIP). Abu Dhabi, United Arab Emirates: IEEE, 2020: 1147-1151.
- [59] LORKIEWICZ M, RÓŻEK S, STANKIEWICZ O, et al. Video coding for machines with neural-network-based chroma synthesis[J]. IEEE Access, 2025, 13: 112777-112784.
- [60] FISCHER K, FORSCH C, HERGLOTZ C, et al. Analysis of neural image compression networks for machine-to-machine communication[C]//Proceedings of 2021 IEEE International Conference on Image Processing (ICIP). Anchorage, AK, USA:

IEEE, 2021: 2079-2083.

- [61] CHAMAIN L D, RACAPE F, BEGAINT J, et al. End-to-end optimized image compression for machines, a study[C]//Proceedings of 2021 Data Compression Conference (DCC). Snowbird, UT, USA: IEEE, 2021: 163-172.
- [62] LE N, ZHANG H, CRICRI F, et al. Image coding for machines: An end-to-end learned approach[C]//Proceedings of ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Toronto, ON, Canada: IEEE, 2021: 1590-1594.
- [63] LIU L, HU Z, CHEN Z, et al. ICMH-net: Neural image compression towards both machine vision and human vision[C]//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, ON, Canada: ACM, 2023: 8047-8056.
- [64] LE N, ZHANG H, CRICRI F, et al. Learned image coding for machines: A content-adaptive approach[C]//Proceedings of 2021 IEEE International Conference on Multimedia and Expo (ICME). Shenzhen, China: IEEE, 2021: 1-6.
- [65] LIU A, YU H, LIU J, et al. Learning-based image coding for machine intelligence with variable-rate[C]//Proceedings of 2025 IEEE International Symposium on Circuits and Systems (ISCAS). London, UK: IEEE, 2025: 1-5.
- [66] TATSUMI Y, ZENG Z, WATANABE H. Training-free adaptive quantization for variable rate image coding for machines [C]//Proceedings of 2026 IEEE International Conference on Consumer Electronics (ICCE). Dubai, United Arab Emirates: IEEE, 2026: 1-6.
- [67] SUZUKI S, TAKAGI M, HAYASE K, et al. Image pre-transformation for recognition-aware image compression[C]//Proceedings of 2019 IEEE International Conference on Image Processing (ICIP). Taipei, China: IEEE, 2019: 2686-2690.
- [68] LU G, GE X, ZHONG T, et al. Preprocessing enhanced image compression for machine vision[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(12): 13556-13568.
- [69] YANG M, HERRANZ L, YANG F, et al. Semantic preprocessor for image compression for machines[C]//Proceedings of ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Rhodes Island, Greece: IEEE, 2023: 1-5.
- [70] PATWA N, AHUJA N, SOMAYAZULU S, et al. Semantic-preserving image compression[C]//Proceedings of 2020 IEEE International Conference on Image Processing (ICIP). Abu Dhabi, United Arab Emirates: IEEE, 2020: 1281-1285.
- [71] QI Y, FENG R, ZHANG Z, et al. Image coding for machines based on non-uniform importance allocation[C]//Proceedings of 2023 IEEE International Conference on Visual Communications and Image Processing (VCIP). Jeju, Korea: IEEE, 2023: 1-5.
- [72] YANG M, YANG F, MURN L, et al. Task-switchable pre-processor for image compression for multiple machine vision tasks [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(7): 6416-6429.
- [73] CHANG J, MAO Q, ZHAO Z, et al. Layered conceptual image compression via deep semantic synthesis[C]//Proceedings of 2019 IEEE International Conference on Image Processing (ICIP). Taipei, China: IEEE, 2019: 694-698.
- [74] CHANG J, ZHAO Z, YANG L, et al. Thousand to one: Semantic prior modeling for conceptual coding[C]//Proceedings of 2021 IEEE International Conference on Multimedia and Expo (ICME). Shenzhen, China: IEEE, 2021: 1-6.
- [75] ZHANG Y, JIA C, CHANG J, et al. Machine perception-driven facial image compression: A layered generative approach[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(4): 3825-3836.
- [76] MAO Q, WANG C, WANG M, et al. Scalable face image coding via StyleGAN prior: Toward compression for human-machine collaborative vision[J]. IEEE Transactions on Image Processing, 2024, 33: 408-422.
- [77] FANG X, DUAN Y, DU Q, et al. Sketch assisted face image coding for human and machine vision: A joint training approach [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(10): 6086-6100.
- [78] LIU J, SUN H, KATTO J. Semantic segmentation in learned compressed domain[C]//Proceedings of 2022 Picture Coding Symposium (PCS). San Jose, CA, USA: IEEE, 2022: 181-185.
- [79] AKBARI M, LIANG J, HAN J. DSSLIC: Deep semantic segmentation-based layered image compression[C]//Proceedings of ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brighton, UK: IEEE, 2019: 2042-2046.

- [80] HOANG T M, ZHOU J, FAN Y. Image compression with encoder-decoder matched semantic segmentation[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA: IEEE, 2020: 619-623.
- [81] DUAN Z, MA Z, ZHU F. Unified architecture adaptation for compressed domain semantic inference[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(8): 4108-4121.
- [82] LIU K, LIU D, LI L, et al. Semantics-to-signal scalable image compression with learned revertible representations[J]. International Journal of Computer Vision, 2021, 129(9): 2605-2621.
- [83] CHOI H, BAJIĆ I V. Scalable image coding for humans and machines[J]. IEEE Transactions on Image Processing, 2022, 31: 2739-2754.
- [84] HU Y, YANG S, YANG W, et al. Towards coding for human and machine vision: A scalable image coding approach[C]//Proceedings of 2020 IEEE International Conference on Multimedia and Expo (ICME). London, UK: IEEE, 2020: 1-6.
- [85] YANG S, HU Y, YANG W, et al. Towards coding for human and machine vision: Scalable face image coding[J]. IEEE Transactions on Multimedia, 2021, 23: 2957-2971.
- [86] TATSUMI Y, ZENG Z, WATANABE H. Explicit residual-based scalable image coding for humans and machines[C]//Proceedings of 2025 IEEE International Workshop on Multimedia Signal Processing (MMSP). Beijing, China: IEEE, 2025: 90-95.
- [87] CUI T, WANG Y, WANG Y, et al. Semantic and saliency-aware scalable image coding toward human-machine collaboration [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(11): 11527-11539.
- [88] SHEN T, LIU Y. Progressive conditional coding for integrated human-machine vision[J]. IEEE Access, 2026, 14: 99322-99335.
- [89] GAO W, MA S, DUAN L, et al. Digital retina: A way to make the city brain more efficient by visual coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(11): 4147-4161.
- [90] LOU Y, DUAN L Y, LUO Y, et al. Towards efficient front-end visual sensing for digital retina: A model-centric paradigm[J]. IEEE Transactions on Multimedia, 2020, 22(11): 3002-3013.
- [91] AZIZIAN B, BAJIĆ I V. Privacy-preserving autoencoder for collaborative object detection[J]. IEEE Transactions on Image Processing, 2024, 33: 4937-4951.
- [92] LI H, GUO Y, WANG Z, et al. AdaCompress: Adaptive compression for online computer vision services[C]//Proceedings of the 27th ACM International Conference on Multimedia. Nice, France: ACM, 2019: 2440-2448.
- [93] LI H, ZHANG X. Human-machine collaborative image compression method based on implicit neural representations[J]. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 2024, 14(2): 198-208.
- [94] LADUNE T, PHILIPPE P, HAMIDOUCHE W, et al. Binary probability model for learning based image compression[C]//Proceedings of ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain: IEEE, 2020: 2168-2172.
- [95] ZHANG Y, DUAN Z, HUANG Y, et al. Low-rank adaptation of pre-trained vision backbones for energy-efficient image coding for machines[C]//Proceedings of 2025 IEEE International Conference on Image Processing (ICIP). Anchorage, AK, USA: IEEE, 2025: 1163-1168.
- [96] YU Y, XING Y, HUANG Z, et al. MT-MoA: A unified parameter-efficient multi-task adaptation framework toward image coding for machines[J]. IEEE Transactions on Multimedia, 2026. DOI:10.1109/TMM.2026.3705198.
- [97] CHOI H, HAN H, ROSEWARNE C, et al. Compressai-vision: Open-source software to evaluate compression methods for computer vision tasks[C]//Proceedings of 2025 IEEE International Conference on Image Processing Workshops (ICIPW). Anchorage, AK, USA: IEEE, 2025: 434-439.
- [98] HARELL A, FOROUTAN Y, AHUJA N, et al. Rate-distortion theory in coding for machines and its applications[J]. IEEE

- Transactions on Pattern Analysis and Machine Intelligence, 2025, 47(7): 5501-5519.
- [99] DUAN L, LIU J, YANG W, et al. Video coding for machines: A paradigm of collaborative compression and intelligent analytics[J]. IEEE Transactions on Image Processing, 2020, 29: 8680-8695.
- [100] TIAN Y, LU G, YAN Y, et al. A coding framework and benchmark towards low-bitrate video understanding[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(8): 5852-5872.
- [101] AN E, KIM M, SEO K D, et al. Tracking-based adaptive temporal resampling for video coding for machines[C]//Proceedings of 2025 IEEE International Conference on Advanced Visual and Signal-Based Systems (AVSS). Tainan, Taiwan, China: IEEE, 2025: 1-6.
- [102] LIN H, CHEN B, ZHANG Z, et al. DeepSVC: Deep scalable video coding for both machine and human vision[C]//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Ontario, Canada: ACM, 2023: 9205-9214.
- [103] CUI T, WANG Y, FANG Z, et al. Dual-contextual scalable video coding for human-machine collaborative intelligence[J]. Displays, 2026, 95: 103638.
- [104] BENJAK M, KHAN S, CHEN Y H, et al. Learned hybrid video coding for human perception and multiple machine vision tasks[C]//Proceedings of 2025 IEEE International Conference on Image Processing (ICIP). Anchorage, AK, USA: IEEE, 2025: 1996-2001.
- [105] TIAN Y, LING X, GENG C, et al. SMC: Masked learning of unsupervised video semantic compression[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2026, 48(2): 1992-2011.
- [106] ZHANG C, SUN H, KATTO J. STDLC: Video coding for machines WITH spatial-temporal decoupled latent composition [C]//Proceedings of ICASSP 2026—2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Barcelona, Spain: IEEE, 2026: 8367-8371.
- [107] SHENG X, LI L, LIU D, et al. VNVC: A versatile neural video coding framework for efficient human-machine vision[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(7): 4579-4596.
- [108] DAO L T H, GIA VIEN A, LEE J, et al. Gradient-guided diffusion-based restoration of extremely compressed backgrounds for video coding for machines[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2026, 36(1): 685-701.
- [109] ZHU A, QI J, ZHANG S, et al. Machine-centric high-accuracy multi-video analytics with adaptive neural codecs[J]. IEEE Transactions on Networking, 2025, 33(5): 2537-2552.
- [110] GOU A, SUN H, ZENG X, et al. Privacy-preserving with flexible autoencoder for video coding for machines[C]//Proceedings of 2024 IEEE International Symposium on Circuits and Systems (ISCAS). [S.l.]: IEEE, 2024: 1-5.
- [111] SUN Y, LIU M, YAO C, et al. Symmetric entropy-constrained video coding for machines[J]. IEEE Transactions on Image Processing, 2026, 35: 6915-6930.
- [112] GE X, LUO J, ZHANG X, et al. Task-aware encoder control for deep video compression[C]//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2024: 26036-26045.
- [113] ALVAR S R, BAJIĆ I V. Multi-task learning with compressible features for collaborative intelligence[C]//Proceedings of 2019 IEEE International Conference on Image Processing (ICIP). Taipei, China: IEEE, 2019: 1705-1709.
- [114] LIU W, WANG H. DFTC: Differentiable feature topology compression for efficient 3D-LLMs[C]//Proceedings of 2026 International Symposium on Machine Learning and Media Computing (MLMC). Nanping, Fujian, China: IEEE, 2026: 1-6.
- [115] BAJIĆ I V, LIN W, TIAN Y. Collaborative intelligence: Challenges and opportunities[C]//Proceedings of ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Toronto, ON, Canada: IEEE, 2021: 8493-8497.
- [116] SHAO J, ZHANG J. BottleNet++: An end-to-end approach for feature compression in device-edge co-inference systems[C]//Proceedings of 2020 IEEE International Conference on Communications Workshops (ICC Workshops). Dublin, Ireland: IEEE, 2020: 1-6.

- [117] YUAN Z, RAWLEKAR S, GARG S, et al. Split computing with scalable feature compression for visual analytics on the edge[J]. *IEEE Transactions on Multimedia*, 2024, 26: 10121-10133.
- [118] CHOI H, BAJIĆ I V. Deep feature compression for collaborative object detection[C]//*Proceedings of 2018 25th IEEE International Conference on Image Processing (ICIP)*. Athens, Greece: IEEE, 2018: 3743-3747.
- [119] SINGH S, ABU-EL-HAJJA S, JOHNSTON N, et al. End-to-end learning of compressible features[C]//*Proceedings of 2020 IEEE International Conference on Image Processing (ICIP)*. Abu Dhabi, United Arab Emirates: IEEE, 2020: 3349-3353.
- [120] CHOI H, COHEN R A, BAJIĆ I V. Back-and-forth prediction for deep tensor compression[C]//*Proceedings of ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona, Spain: IEEE, 2020: 4467-4471.
- [121] SHAH M A, RAJ B. Deriving compact feature representations via annealed contraction[C]//*Proceedings of ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Barcelona, Spain: IEEE, 2020: 2068-2072.
- [122] EIMON M E H, MERLOS J, PERERA A, et al. Enabling next-generation consumer experience with feature coding for machines[C]//*Proceedings of 2025 IEEE International Conference on Consumer Electronics (ICCE)*. Las Vegas, NV, USA: IEEE, 2025: 1-4.
- [123] EIMON M E H, MERLOS J, PERERA A, et al. Feature coding for scalable machine vision[J]. *IEEE Consumer Electronics Magazine*, 2026, 15(3): 48-59.
- [124] PERERA A, EIMON M E H, MERLOS J, et al. FCM-RT: Real-time feature coding for machines[C]//*Proceedings of the 33rd ACM International Conference on Multimedia*. Dublin, Ireland: ACM, 2025: 13525-13527.
- [125] ZHANG Z, WANG M, MA M, et al. MSFC: Deep feature compression in multi-task network[C]//*Proceedings of 2021 IEEE International Conference on Multimedia and Expo (ICME)*. Shenzhen, China: IEEE, 2021: 1-6.
- [126] CHEN C, XU M, LI S, et al. Residual based hierarchical feature compression for multi-task machine vision[C]//*Proceedings of 2023 IEEE International Conference on Multimedia and Expo (ICME)*. Brisbane, Australia: IEEE, 2023: 1463-1468.
- [127] DING D, CHEN Z, LIU Z, et al. Hierarchical image feature compression for machines *via* feature sparsity learning[J]. *IEEE Signal Processing Letters*, 2024, 31: 1159-1163.
- [128] LIU T, XU M, LI S, et al. Learnt mutual feature compression for machine vision[C]//*Proceedings of ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Rhodes Island, Greece: IEEE, 2023: 1-5.
- [129] PAN Z, WANG H, ZHAO T, et al. Multi-scale feature compression via multi-receptive-field convolutional neural network for machine vision[J]. *IEEE Transactions on Multimedia*, 2026, 28: 6944-6956.
- [130] ZHANG L, PAN Z, LI L, et al. Learning hierarchical and spatial semantic correlations for enhanced multi-scale feature compression[J]. *IEEE Transactions on Broadcasting*, 2026, 72(1): 193-206.
- [131] LIU J, ZHANG Y, GUO Z, et al. Multiscale feature importance-based bit allocation for end-to-end feature coding for machines[J]. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2025, 21(9): 1-19.
- [132] JIN D, LEI J, PENG B, et al. Stereo image coding for machines with joint visual feature compression[J]. *IEEE Transactions on Artificial Intelligence*, 2026, 7(9): 4880-4892.
- [133] CHEN Z, FAN K, WANG S, et al. Toward intelligent sensing: Intermediate deep feature compression[J]. *IEEE Transactions on Image Processing*, 2020, 29: 2230-2243.
- [134] SUN Y, ZHAO Y, LIU M, et al. TransVFC: A transformable video feature compression framework for machines[J]. *Pattern Recognition*, 2026, 171: 112091.
- [135] XIA S, LIANG K, YANG W, et al. An emerging coding paradigm VCM: A scalable coding approach beyond feature and signal[C]//*2020 IEEE International Conference on Multimedia and Expo (ICME)*. London, UK: IEEE, 2020: 1-6.

- [136] HUANG K, AN P, YANG C, et al. Generic feature extraction and compression for human and machine-oriented vision[J]. *Signal Processing*, 2026, 240: 110369.
- [137] YAN N, LIU D, LI H, et al. Semantically scalable image coding with compression of feature maps[C]//*Proceedings of 2020 IEEE International Conference on Image Processing (ICIP)*. Abu Dhabi, United Arab Emirates: IEEE, 2020: 3114-3118.
- [138] WANG S, WANG S, ZHANG X, et al. Scalable facial image compression with deep feature reconstruction[C]//*Proceedings of 2019 IEEE International Conference on Image Processing (ICIP)*. Taipei, China: IEEE, 2019: 2691-2695.
- [139] WANG S, WANG S, YANG W, et al. Teacher-student learning with multi-granularity constraint towards compact facial feature representation[C]//*Proceedings of ICASSP 2021—2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Toronto, ON, Canada: IEEE, 2021: 8503-8507.

#### 作者简介:



袁冯(1996-),男,讲师,研究方向:图像/视频编码、人工智能, E-mail: yuanfeng\_1996@tju.edu.cn。



张立谦(1999-),女,博士研究生,研究方向:机器视觉编码、计算机视觉。



潘兆庆(1986-),通信作者,男,教授,博士生导师,研究方向:多媒体通信、视频编码, E-mail: pan\_zhaoqing@tju.edu.cn。



王瀚涛(1978-),男,教授,博士生导师,研究方向:计算机视觉、多媒体信息处理、机器学习、视频编码和具身智能。



李礼(1990-),男,教授,博士生导师,研究方向:图像压缩、视频压缩、光场压缩和点云压缩。



赵铁松(1984-),男,教授,博士生导师,研究方向:多媒体信号处理、视频编码、机器视觉编码和人机协同视觉。



张云(1982-),男,教授,博士生导师,研究方向:图像/视频编码与通信、视觉感知计算、图像质量评价和三维视频处理。

(编辑:张黄群)

## Coding for Machine Vision: Key Technologies and Future Perspectives

YUAN Feng<sup>1,2</sup>, ZHANG Liqian<sup>1</sup>, PAN Zhaoqing<sup>1\*</sup>, WANG Hanli<sup>3</sup>, LI Li<sup>4</sup>, ZHAO Tiesong<sup>5</sup>, ZHANG Yun<sup>6</sup>

(1. School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China; 2. School of Medical Artificial Intelligence, Xuzhou Medical University, Xuzhou 221004, China; 3. College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China; 4. School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China; 5. College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China; 6. School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen 518107, China)

**Abstract:** With the rapid development of artificial intelligence and intelligent vision technologies, images and videos have evolved from media primarily intended for human viewing into essential sources of information for machine perception, understanding, and decision-making. Conventional visual coding methods primarily optimize pixel reconstruction fidelity and perceptual quality for human vision. However, these objectives do not necessarily ensure the efficient preservation of semantic information and discriminative features required by machine vision tasks, such as object detection, semantic segmentation, object tracking, and action recognition. Coding for machine vision has therefore attracted increasing attention from both academia and industry as an important research direction in visual information compression. This paper systematically reviews its key technologies and development trends, covering the optimization of conventional coding standards and deep-learning-based coding methods. First, machine-oriented optimization methods for conventional image and video coding standards are examined, with particular attention to task-aware bit allocation, rate-distortion optimization for machine vision, and coding tool optimization. Second, deep-learning-based methods are organized into three categories according to the coding object: Learned image coding, learned video coding, and learned feature coding. The evolution of these technologies from pixel reconstruction toward semantic preservation, task performance preservation, and edge-cloud collaborative intelligent processing is analyzed. Finally, the paper discusses major challenges concerning evaluation protocols, generalization across tasks and models, computational complexity, compatibility with existing standards, and the joint requirements of human-machine collaboration and privacy protection. Future research directions are also outlined, including unified rate-distortion modeling, general-purpose visual representation coding, lightweight intelligent coding, human-machine collaborative coding, and standardization.

### Highlights:

1. A systematic review of coding for machine vision is presented, covering conventional coding standard optimization and learned image, video, and feature coding.
2. Representative methods are compared in terms of core ideas, key techniques, and coding performance, tracing the shift from pixel reconstruction to semantic preservation and task-oriented compression.
3. Major challenges and future directions are discussed, including unified rate-distortion modeling, general-purpose visual representation coding, lightweight deployment, human-machine collaborative coding, and standardization.

**Key words:** coding for machine vision; video coding standard; learned image coding; learned video coding; feature coding; human-machine collaboration

---

**Foundation item:** National Natural Science Foundation of China (No.62322116).

**Received:** 2026-08-13; **Revised:** 2026-09-20

**\*Corresponding author, E-mail:** pan\_zhaoqing@tju.edu.cn.