新一代通用视频编码标准 H.266/VVC:现状与发展

万帅[1,2]，霍俊彦[3]，马彦卓[3]，杨付正[3]
（1. 西北工业大学电子信息学院，710129，西安；2. 皇家墨尔本理工大学工程学院，VIC3001，
澳大利亚墨尔本；3. 西安电子科技大学通信工程学院，710071，西安）

摘要：相比于上一代标准，新一代通用视频编码标准（H.266/VVC）在同等质量下能够节省大约50%的码率，且适用于多种多样的视频应用场景。论文从H.266/VVC的关键技术出发，对标准的现状、实现和应用发展进行深入探讨。H.266/VVC沿用既往标准中的双层码流体系和混合编码框架，针对帧内预测、帧间预测、变换、量化、环路滤波等所有主要编码模块进行了技术革新，并为屏幕内容视频等应用提供了高效的专用编码工具。H.266/VVC标准目前已处于实用化阶段，官方参考软件VTM和开源编解码器VVenC/VVdeC是目前最具代表性的软件编解码实现。对H.266/VVC的性能分析可以看出：H.266/VVC针对高分辨率视频取得的编码增益更为突出；主要编码工具对性能的贡献通常以复杂度为代价，但也有部分编码工具在提升编码性能的同时可降低整体编码复杂度。H.266/VVC的硬件实现面临诸多挑战，发展明显滞后于软件实现，现有研究主要集中在对具体编码模块的硬件加速方面。H.266/VVC标准发布之后，下一代视频编码标准的发展目前仍围绕混合编码框架进行探索，聚焦在两大方向：超越VVC的增强压缩关注更为先进的、非神经网络的编码工具，基于神经网络的视频编码则探索采用神经网络的编码工具。除此之外，部分或完全跳出现有混合编码框架的端到端视频编码也在飞速发展，未来视频编码标准与神经网络结合成为趋势，但面临着计算资源依赖和稳定结构两方面的考验。

关键词：H.266/VVC标准；视频编码标准；编码模块；编解码器；神经网络
中图分类号：TP391.4 **文献标志码**：A
DOI：10.7652/xjtuxb202404001 **文章编号**：0253-987X(2024)04-0001-17

The New-Generation Versatile Video Coding Standard H.266/VVC: State-of-the-Art and Development

WAN Shuai[1,2], HUO Junyan[3], MA Yanzhuo[3], YANG Fuzheng[3]
(1. School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129, China;
2. School of Engineering, Royal Melbourne Institute of Technology, Melbourne VIC3001, Australia;
3. School of Electronic Engineering, Xidian University, Xi'an 710071, China)

Abstract: Compared with the previous-generation standard, the new-generation versatile video coding standard H.266/VVC saves about 50% of the bit rate given the same quality and applies to a wide range of video application scenarios. The status quo, implementation, and application development of H.266/VVC are discussed in this paper from the perspective of its key technologies. H.266/VVC retains the dual-layer bitstream structure and hybrid coding framework of the

收稿日期：2023-07-04。 作者简介：万帅（1979—），女，教授，博士生导师。 基金项目：国家自然科学基金资助项目（62171353，62101409）；陕西省自然科学基础研究计划资助项目（2024JC-YBMS-463）。
网络出版时间：2023-11-17 网络出版地址：https://link.cnki.net/urlid/61.1069.T.20231116.1522.002

previous standard while introducing technological innovations to all the major coding modules such as intra-frame prediction, inter-frame prediction, transformation, quantization, and loop filtering. Moreover, it provides efficient specialized coding tools for applications such as screen content videos. Currently, the H. 266/VVC standard is in a practical stage, with the official reference software VTM and the open-source codecs VVenC/VVdeC serving as the most prominent software codec implementations. An analysis of the performance of H. 266/VVC reveals that it achieves more notable coding gains for high-resolution videos. While some of the main coding tools contribute to improved performance, they may also increase complexity. Nevertheless, certain coding tools manage to enhance coding performance while reducing overall coding complexity. The hardware implementation of H. 266/VVC faces many challenges, and its development lags behind software implementation. Following the release of H. 266/VVC, the development of a next-generation video coding standard still focuses on the hybrid coding framework. There are two main directions: Enhanced compression in Beyond VVC concentrates on more advanced, non-neural-network-based coding tools, while neural-network-based video coding explores the use of neural-network-based coding tools. Furthermore, there is rapid progress in the development of end-to-end video coding that partially or completely deviates from the existing hybrid coding framework. In the future, the combination of video coding standards with neural networks will become a trend while facing the challenges of computational resource dependence and keeping a stable structure.

Keywords: H. 266/VVC standard; video coding standard; coding modules; encoder/decoder; neural network

作为一种主要通过视觉来感知世界的生物，人类对于视觉媒体的需求是天然的，并且随着信息技术的发展与日俱增。近年来，视频应用逐渐呈现出两大趋势：高清化与多样化。一方面，高清、超高清视频的获取成本大大降低，随之而来的则是爆发式的数据增量。另一方面，互联网生态的不断丰富也促生了各式各样的视频应用。人们不再满足于被动观看，而是热衷于主动进行记录和分享（如短视频、运动摄影）、通过视频进行交互（如视频会议、屏幕内容分享、游戏视频）、参与沉浸式体验（如 360°全景视频、多视角立体视频）等。相应地，各类视频应用涉及的视频格式多样、特色各异，为视频压缩编码的统一架构和普遍有效性带来挑战。

针对大幅提高编码效率和应对多样化视频类型的迫切需求，国际电信联盟 ITU-T 的视频编码专家组与国际标准化组织 ISO/IEC 的运动图像专家组于 2015 年组建联合视频探索小组（Joint Video Exploration Team，JVET），共同开展新一代视频编码国际标准的制定工作。值得一提的是，这两个组织曾经有过 3 次非常成功的合作：DVD 的核心技术 H. 262/MPEG-2 标准[1]，获得广泛应用的先进视频编码 H. 264/AVC 标准[2]，压缩性能突出的高效视频编码 H. 265/HEVC 标准[3]。

经过两年的探索，JVET 在 Joint Exploration Model（JEM）参考软件平台[4]上取得了较好的编码增益，为新一代视频编码标准的研发做好了技术储备。2018 年 4 月，JVET 将新一代标准命名为通用视频编码（versatile video coding，VVC）[5]。VVC 标准的目标是以统一架构编码不同类别的视频。在同一次会议上，JVET 建立了第一版 VVC 测试模型（VVC test model，VTM）[6]，正式开启了 VVC 标准的制定。2018 年 4 月—2020 年 7 月是 VVC 标准形成的关键时期，期间 JVET 共召开了 10 次会议，对 6 000 多份技术提案进行了深入讨论，将性能优异的工具采纳至标准。在这个过程中，VTM 编码性能迅速提升，标志着 VVC 标准的快速发展。2020 年 7 月，随着 JVET 会议落下帷幕，通用视频编码标准 VVC 正式形成[7]。随后，ITU-T 批准 VVC 标准并正式定名为 ITU-T H. 266。因此，在正式场合通常将该标准写为 H. 266/VVC[8]。

H. 266/VVC 标准的编码性能卓越。相比 H. 265/HEVC，H. 266/VVC 在同等质量的条件下能够节省大约 50%的码率[9]。同时，其解码复杂度不超过 H. 265/HEVC 的两倍，编码复杂度增加与压缩性能增益基本保持正比。此外，H. 266/VVC 标准能够应对更多样的视频格式和内容，为已有和

新兴的视频应用提供高效、灵活、统一的编码压缩框架。H.266/VVC 出色的编码性能来源于其在标准化过程中引入的新型编码工具和语法架构[7-11]，本文将对 H.266/VVC 关键技术进行梳理和剖析。

H.266/VVC 标准面向广泛的应用场景，除了电视广播、视频会议、视频点播等传统业务，还包括自适应视频流、屏幕内容视频、多视点视频、可分层视频、全景视频等新兴业务。目前，H.266/VVC 标准已经处于实用化阶段。在标准参考软件 VTM 的基础上，工业界开发了更为高效的软硬件编解码实现，如开源编解码器[12]、商用编解码器、播放器、比特流分析软件等[13]，随之涌现出各式各样相关应用。H.266/VVC 标准也受到应用类标准的认可，现已被欧洲电信标准协会数字视频广播标准以及巴西下一代广播电视标准采纳[13]。与此同时，视频编解码标准并未停止演进的步伐。在标准制定完成之后，JVET 围绕提高视频编解码性能进行着持续的探索，形成了两大探索方向：超越 VVC 的增强压缩（enhanced compression beyond VVC capability，beyond VVC）[14] 和基于神经网络的视频编码（neural network-based video coding，NNVC）[15]。本文针对 H.266/VVC 标准的编解码器实现和未来技术发展走向进行了探讨与展望。

1　H.266/VVC 关键技术

视频编码的关键技术包含高层语法设计和编码工具两大方面。H.266/VVC 沿用了既往标准中的双层码流体系，包含视频编码层（video coding layer，VCL）和网络适配层（network abstract layer，NAL）。原始视频数据划分为编码单元后送入混合编码框架进行编码，遵循标准语法描述生成 VCL 比特流，再与对应高层语法一起组合、封装，构成 NAL 单元（NAL unit，NALU）。NALU 可以作为载荷直接在网络上进行传输。

1.1　码流结构

H.266/VVC 的编码比特流可包含一个或多个编码视频序列（coded video sequence，CVS）。CVS 是时域独立可解码的基本单元，每个 CVS 以帧内随机接入点图像或逐渐解码刷新图像作为起始。CVS 码流结构如图 1 所示。每个 CVS 包含一个或多个按解码顺序排列的访问单元（access unit，AU）。每个 AU 包含一个或多个同一时刻的图像单元（picture unit，PU），每个 PU 包含且仅包含一幅完整图像的编码数据。当一个 AU 包含多个 PU 时，每个 PU 可以是特定质量或分辨率（可分级视频流）图像，也可以是多视点视频的某一视点，以及深度、反射率等属性信息。AU 中的不同 PU 被归属为不同的层，一个 CVS 中所有同层的 PU 组成了编码视频序列层（coded layer video sequence，CLVS）。

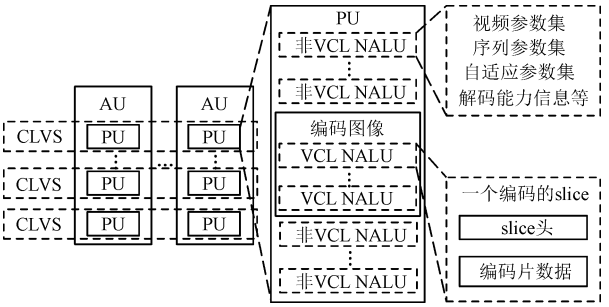


图 1　CVS 结构

Fig. 1　CVS structure

每个 PU 为一幅图像的码流，包含一个或多个片（slice）。片与片之间进行独立编解码，该设计有利于在数据丢失时进行重新同步。slice 编码数据打包形成的 NALU 称为 VCL NALU。除此之外，PU 还包含非 VCL NALU，如参数集、访问单元分割符等。参数集包含视频中不同层级编码单元的共用信息，是独立编码的数据类型。其中，视频参数集承载视频分级的信息，表达 PU 间的依赖关系，配合参考帧管理支持可分级视频编码、多视点视频编码等需求。序列参数集包含 CVS 的共用编码参数，如图像格式、编码块尺寸限制、档次与层级等。H.266/VVC标准新引入了自适应参数集、解码能力信息等参数集。

参数集作为非 VCL NALU 进行传输，为传递关键数据提供了高鲁棒机制。参数集的独立性使得其可以提前发送，也可以在需要增加新参数集的时候再发送。参数集可以被多次重发或者采用特殊技术加以保护，甚至采用带外发送的方式。

slice 头及其之上的语法结构通常称为高层语法。高层语法设计的目的是为了网络传输和存储的需要，对视频编码数据进行有效组织和封装，保证码流接口的友好性、随机接入性、误码恢复能力、互动性、向后兼容性等。网络适配层是高层语法中最关键的组成之一。视频压缩数据根据其内容特性被分成具有不同特性的 NALU，并对 NALU 的内容特性进行标识。网络可以根据 NALU 及其标识优化视频传输性能，不再需要具体分析视频数据的内容特性。H.266/VVC 高层语法的详细信息可参考文献[16]。

1.2 编码框架

类似于 H.265/HEVC，H.266/VVC 仍采用基于编码树单元(coding tree unit，CTU)的划分结构。待编码图像首先被分割成 slice，每个 slice 由相同大小的 CTU 组成。为匹配 4K、8K 等视频的编码需求，H.266/VVC 中 CTU 亮度块的最大允许尺寸为 128×128 像素。每个 CTU 按照二叉树、三叉树、四叉树递归划分为不同尺寸的编码单元(coding unit，CU)[17] 作为大多数编码工具的基本单位。slice 到 CU 的划分结构如图 2 所示。

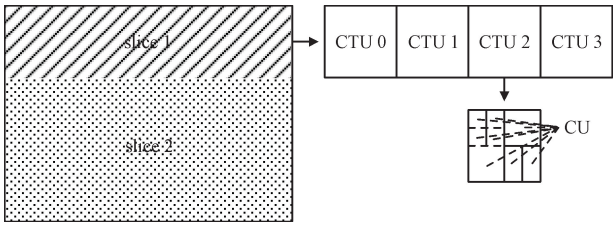


图 2 slice 到 CU 之间的划分示例

Fig. 2 Example of partitioning slice into CU

H.266/VVC 整体编码框架如图 3 所示。对每个 CU 进行编码时，通常流程如下：首先，通过帧内或帧间预测去除图像的空、时间相关性；再次，将预测残差送入变换模块生成能量较为集中的变换系数；之后，将变换系数送到量化模块实现多对一的映射；最后，再送入熵编码模块以输出码流。为了得到与解码器一致的重建信号，H.266/VVC 编码器包含完整的解码器，如图 3 中黄色底色部分所示。编码控制模块往往通过拉格朗日率失真优化[18-19] 选择最优的编码参数[20]。H.266/VVC 在图 3 所示的各个模块都引入了新工具，下文将分模块进行介绍。

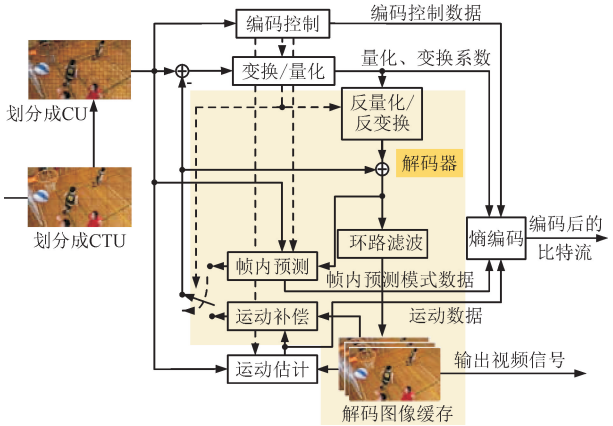


图 3 H.266/VVC 编码框架

Fig. 3 Framework of H.266/VVC encoding

1.3 帧内预测编码

帧内预测编码使用当前图像内已编码像素值预测待编码像素值，从而有效去除视频空域相关性。H.266/VVC 的帧内预测包含参考像素获取、预测值计算和预测值修正 3 个步骤[21]，如图 4 所示。图中{}里的内容为 H.266/VVC 采用的代表性技术。

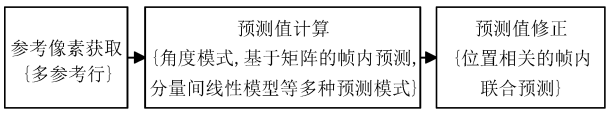


图 4 H.266/VVC 帧内预测编码

Fig. 4 H.266/VVC intra prediction coding

在获取参考像素时，H.266/VVC 允许使用邻近额外 2 行/列参考像素[22]，扩展了参考像素范围。为提高角度预测的准确性，H.266/VVC 引入了高效插值滤波器。传统的预测模式包括平面、直流和 65 种角度模式，以适配具有不同纹理特性的编码块。对于宽高不等的矩形块，H.266/VVC 引入宽角度模式[23]。此外，帧内子区域划分[24] 使用重建子区域作为后续子区域的参考。H.266/VVC 还引入位置相关的帧内联合预测(position dependent intra prediction combination，PDPC)[25]，利用空间相关性强的参考像素对预测值进行修正。

基于矩阵的帧内预测(matrix-based intra prediction，MIP)模式是 H.266/VVC 中采用神经网络思想的新技术。MIP 源于多层神经网络[26-27]，为权衡计算复杂度，最终使用矩阵乘法近似实现一层全连接网络。MIP 预测过程如图 5 所示，参考像素作为输入向量与 MIP 预测矩阵相乘得到部分预测值，再通过上采样得到最终预测值。其中，对参考像素下采样和对输出向量上采样有利于降低矩阵乘法次数，同时减少内存。

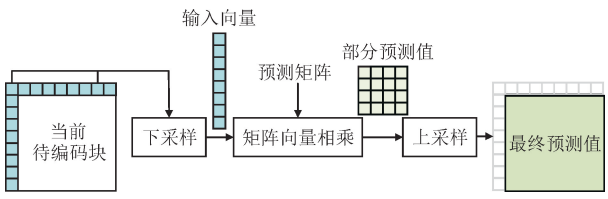


图 5 基于矩阵的帧内预测原理框图

Fig. 5 Diagram of matrix-based intra prediction

采用亮度信号对色度信号进行分量间预测是 H.266/VVC 的特色之一。如图 6 所示，分量间线性模型(cross-component linear mode，CCLM)预测

模式[28]基于亮度色度局部相关性建立分量间线性模型，根据该模型和亮度重建值计算色度预测值。CCLM 的关键是利用参考像素的亮度色度值确定线性模型的系数。H.266/VVC 采纳了本文作者提出的参考像素子集方案[29]，该方法在不降低性能的前提下具有更低复杂度，且对不同块尺寸采用统一方案，利于硬件实现。

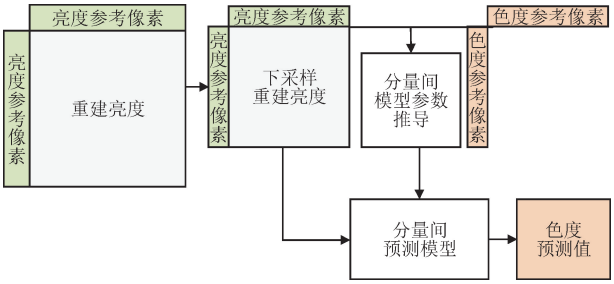


图 6　分量间线性模型原理框图

Fig. 6　Diagram of cross-component linear mode

1.4　帧间预测编码

帧间预测编码利用视频的时域相关性，只编码图像间的运动信息及预测残差，可大幅度提高编码效率。根据运动矢量(motion vector，MV)和预测残差的编码方式，H.266/VVC 帧间编码模式可以分为跳过、合并和高级运动矢量预测 3 大类，如表 1 所示。具体地，跳过模式采用预测运动矢量表示 MV 信息，且不编码运动矢量差值和预测残差。合并模式仅编码预测运动矢量和预测残差，不编码运动矢量差值。高级运动矢量预测模式则对预测运动矢量、运动矢量差值和残差均进行编码。

表 1　H.266/VVC帧间预测模式分类及特点

Table 1　Classification and characteristics of H.266/VVC inter prediction modes

帧间预测数据	跳过	合并	高级运动矢量预测
预测运动矢量	有	有	有
运动矢量差值	无	无	有
预测残差	无	有	有

H.266/VVC 帧间预测过程可分为运动矢量的预测、运动矢量的确定、运动补偿 3 个步骤，每个步骤都引入了多项新技术[30-31]，如图 7 所示。图中，{}里的内容为 H.266/VVC 采用的代表性技术。

1.4.1　运动矢量预测

运动矢量预测列表中按一定规则放置与当前 CU 空域、时域相关性强的已编码块 MV，作为当前 CU 的预测运动矢量的候选值。当选择列表中的某个候选运动矢量预测作为当前 CU 的预测运动矢量时，只需编码选中运动矢量预测在列表中的索引值。H.266/VVC 引入了基于历史的候选运动矢量预测，利用先前已编码块的运动信息存储为历史信息并用于构造运动矢量预测列表。

图 7　H.266/VVC 帧间预测编码

Fig. 7　H.266/VVC inter prediction coding

传统的帧间预测中，同一个 CU 内所有像素采用相同的运动矢量。H.266/VVC 引入了基于子块的时域 MV 预测，使用单一模式标识 CU 内各子块的不同 MV 信息，提升了 MV 的表示效率。

仿射运动补偿是 H.266/VVC 的特色帧间编码技术。对于存在旋转、缩放、拉伸等非平移运动的编码块，块中各像素的运动矢量虽然不同，但具有一定的规律性，可以通过高阶变形模型以极少的模型参数来描述[32-34]。

1.4.2　运动矢量确定

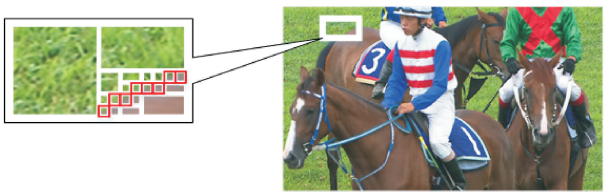
针对合并模式，H.266/VVC 引入了解码端运动矢量修正和带有运动矢量差值索引的合并模式。解码端运动矢量修正是解码端基于前后向运动矢量的对称偏移，利用前后向参考块的匹配程度确定调整偏移量，对运动矢量进行修正。带有运动矢量差值索引的合并模式并未编码实际的运动矢量差值，而是根据运动矢量差值出现的概率预先定义一个固定的高概率偏移值集合，用集合中的索引确定 MV 的偏移量。

针对高级运动矢量预测模式，H.266/VVC 引入了对称运动矢量差值和运动矢量差值的自适应精度表示。对称运动矢量差值针对双向预测的 CU，只编码其前向运动矢量差值，后向运动矢量差值则根据对称一致性推导得到。运动矢量差值的自适应精度表示允许针对不同运动剧烈程度的视频内容，以 CU 为单位自适应选择不同运动矢量差值精度，以兼顾运动矢量表示范围和精度。

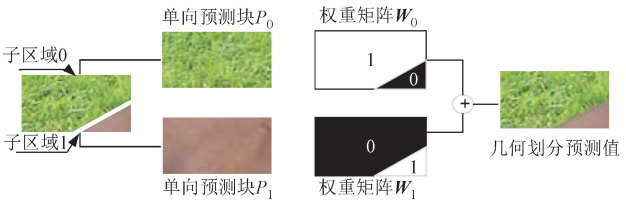
1.4.3　运动补偿

H.266/VVC 引入了联合帧内帧间预测，其运动补偿通过融合帧内和帧间预测值实现。

几何划分帧间预测具有一定分割的理念[35]。当运动物体具有非水平或垂直边缘时，采用矩形划分将在边缘处产生大量小块，需要编码大量的块划分及 MV 信息，如图 8(a)所示。几何划分帧间预测使用斜线将矩形 CU 划分成两个不规则子区域以匹配实际的运动，如图 8(b)所示。划分线以极坐标形式用角度和偏移量来高效表示。各子区域分别利用不同运动信息获得补偿，并对划分线附近区域以软混合的方式进行加权融合，以模拟自然场景中柔和的边缘过渡。



(a)矩形划分



(b)几何划分

图 8 非规则运动物体的划分

Fig. 8 Partition of irregular moving objects

H. 266/VVC 引入了基于光流的预测值修正。对于普通运动，双向光流[36]可以利用前向预测参考块和后向预测参考块的一致性，估计前向预测参考块和后向预测参考块间的光流，修正运动矢量及相应的预测值。对仿射运动补偿，光流预测细化[37]为 4×4 像素子块的每个像素计算光流补偿值。

为处理亮度渐变的场景，除了 slice 级加权运动补偿，H. 266/VVC 还引入了 CU 级的双向加权运动补偿预测。该技术针对局部亮度渐变的场景，在 CU 层传输线性加权预测的参数。

1.5 变换编码

预测残差空间分布通常较分散，采用变换编码可将其映射到分布集中的变换域，进一步去除空间冗余。H. 266/VVC 引入多项变换新技术[38]，其过程如图 9 所示。预测残差通常首先经过主变换得到一次变换系数；对于采用 DCT-Ⅱ作为主变换核的一次变换系数，选择性使用二次变换，得到最终变换系数。H. 266/VVC 支持变换跳过模式，直接对残差进行量化。针对帧间预测残差，子块变换[39]仅对残差能量大的部分区域进行变换，其余区域残差强制归零。当编码块宽或高等于最大变换尺寸时，变换系数仅保留低频部分，高频系数置零，同样达到减少变换系数能量的目的。图中，{} 里的内容为 H. 266/VVC采用的代表性技术。

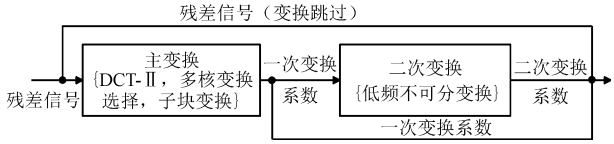


图 9 H. 266/VVC 编码端变换编码

Fig. 9 H. 266/VVC transform coding in encoder

不同预测模式下的残差信号往往具有不同特性[40]，H. 266/VVC 引入多核变换选择以匹配多种预测残差特性。

H. 266/VVC 引入低频不可分变换[41]，对以 DCT-Ⅱ为主变换核的变换系数进行二次变换，使得残差能量更集中。首先，将左上角的 N 个低频变换系数转化为一维向量，并将其余位置的变换系数置零；然后，对该一维向量进行低频不可分变换变换得到长度为 $R(R\leqslant N)$的向量。由此，二次变换可以达到减少变换系数的目的。与 MIP 类似，低频不可分变换的变换核通过大量数据训练得到。

1.6 量化

除传统的标量量化外，H. 266/VVC 引入了高效的依赖量化[42]。与传统的标量量化不同，依赖量化中当前变换系数的量化值依赖于前一个变换系数的量化值。依赖量化利用了变换系数间的相关性，使得变换系数经量化后在 M 维向量空间更紧密(M 代表变换块中变换系数的个数)。从解码器的角度来看，H. 266/VVC 中的依赖量化定义了两种不同的标量量化器 Q_0、Q_1，对应设计了依赖量化的 4 种状态。在反量化时，变换系数按照编解码顺序重建，每处理一个系数，依赖量化相应更新一次状态，第 $k+1$ 个系数所使用的标量量化器(Q_0 或 Q_1)，由第 k 个量化索引值和当前的依赖量化的状态决定。

编码端可采用率失真优化量化[43]的思想，利用状态间随时间(对应于系数组中的序号)转换形成的栅格，对整个系数组最优的量化路径进行搜索，选择最佳量化索引[44-45]。依赖量化以系数组为单位实现最优量化，隐含了矢量量化的思想。

1.7 熵编码

熵编码用于进一步去除数据间的统计冗余。对于高层语法元素，H. 266/VVC 采用简单的熵编码

方法，例如定长码、零阶指数哥伦布码等，有利于快速解析语法元素。对于片级以下的语法元素，H.266/VVC采用基于上下文的自适应算术编码[46]获取较高编码性能。其中，上下文建模是利用以已编码的语法元素为条件进行编码的思想进行概率模型预测。H.266/VVC 引入双概率更新模型以及低复杂度的概率迭代算法，可获得准确的概率模型。

1.8　环路滤波

H.266/VVC 仍采用基于块的混合编码框架，因此方块效应、振铃效应、颜色偏差以及图像模糊等常见编码失真效应仍然存在。为了降低各类失真对视频质量的影响，H.266/VVC 采用环路滤波[47]，包括亮度映射与色度缩放、去方块滤波、样点自适应补偿和自适应环路滤波。其中，去方块滤波和样点自适应补偿延续了 H.265/HEVC 的算法。

亮度映射与色度缩放[48]的核心思想是为亮度平坦区域分配更多码字，为纹理复杂区域分配较少的码字。其中，基于动态分段线性模型的亮度映射技术根据概率分布将亮色度原始值域范围扩展到指定位深的像素值域范围。

自适应环路滤波[49]运用维纳滤波的思想，以原始帧和重建帧之间的最小均方误差为优化目标，根据维纳-霍夫方程求解得到自适应环路滤波滤波系数。自适应环路滤波包括亮度自适应环路滤波、色度自适应环路滤波和分量间自适应环路滤波。分量间自适应环路滤波提出利用亮度对色度进行补偿，补充色度的纹理细节，提升色度质量。分量间自适应环路滤波使用未经过自适应环路滤波滤波的亮度重建值进行色度修正，便于并行执行不同的自适应环路滤波滤波。

1.9　面向屏幕内容的编码算法

屏幕内容，如计算机桌面分享、文档演示、游戏动画等，是一种特殊视频类型，通常由计算机生成。相比自然视频，屏幕内容视频不受摄像机镜头的物理限制，不存在传感器噪声，常含有更少的颜色类型、更多的重复图形、更锐利的物体边缘。针对上述特点，H.266/VVC 标准引入多种屏幕内容编码工具[50-51]。

帧内块复制的预测过程与帧间预测类似，在当前帧已经完成重建的区域内搜索与当前块匹配的参考块，进而将参考块进行复制得到预测块。参考块与当前块之间的位移用块矢量来描述。

在局部区域，计算机生成的内容通常只使用少量的颜色。调色板模式直接对这些数量较少的颜色集进行编码，以提升编码效率。调色板可以是分量间联合调色板，也可以是单分量的调色板。

为了削弱颜色失真效应，屏幕视频经常使用 4∶4∶4颜色格式。为有效利用颜色分量间的相关性，H.266/VVC 标准采用了自适应颜色变换，允许使用颜色转换模块将视频残差信息转换到 YCgCo 颜色空间进行变换、量化、熵编码等操作，提高编码性能，降低计算复杂度。

2　H.266/VVC 编解码器

随着 H.266/VVC 标准的正式发布，涌现出大量相关软硬件，包括开源编解码器、商用编解码器、播放器等[13]。其中，H.266/VVC 标准的官方参考软件 VTM 和基于 VTM 开发的开源编解码器 VVenC/VVdeC 是目前最具代表性的 H.266/VVC 编解码实现，对于学术研究和产品开发都具有重要的价值。

2.1　H.266/VVC 软件实现

2.1.1　VVC 测试模型 VTM

视频编码标准只规定码流的语法语义，并不对编码器进行限制。然而，为了对语法元素进行合理设计，标准应明确可能的编码方式，从而形成系统化的标准编解码器测试模型。VVC 测试模型 VTM 是 H.266/VVC 标准的官方参考软件，由 JVET 开发和维护[6]。作为标准实现的基本参考，VTM 可用于验证和评估 H.266/VVC 编解码器的性能，帮助理解标准语法的内涵和解码过程，并可作为开发实际产品的基础。除了常规的视频编码，VTM 还支持多视角编码、全景视频编码、深度图编码等，能够满足沉浸式场景、三维视频等不同应用的需求。

VTM 功能齐全，使用方便，来源权威，针对 H.266/VVC标准的编解码器性能评估通常以 VTM 为基准，相应的结果需在通用测试条件下进行对比[52]。通用测试条件由 JVET 制定，规定了标准测试序列[53]和不同应用场景下的编解码参数设置，包括标准动态范围、高动态范围、360°全景、非 4∶2∶0 色度格式等测试场景。相应地，标准测试序列涵盖了不同分辨率下的自然场景视频(class A～E)、屏幕内容视频(class F)、高动态视频(class H)以及 360°全景视频(class S)等。

针对不同的测试场景，通用测试条件规定了相应的测试条件，包括全帧内、随机访问和低延迟设置。随机访问设置通常提供 1 s 左右的随机接入间隔，适用于娱乐类应用，如广播、流媒体等；低延迟设置适用于对时延敏感的交互式应用，如视频会议、直

播等;全帧内设置则阻断了帧间误差传播,为信道环境较差的场景,如丢包严重场景,提供更高的鲁棒性。

BD-rate(Bjøntegaard delta bit rate,用符号 ΔR 表示)[54-55]是视频编码中使用的客观度量指标。在一定比特率或质量范围内,该指标可以比较两种不同的视频编解码器,或同一视频编解码器不同配置下的率失真性能。该值为负数时,表示压缩效率提高。通用测试条件规定了各测试条件下使用的量化参数,通常为 22、27、32 和 37,得到相应的编码码率和质量(通常以峰值信噪比衡量)后,即可求解出 BD-rate。

2.1.2 VVenC 开源编码器

VTM 的开发以获得最大编码增益为目标,并未针对编解码速度进行全面优化,也不支持多线程实现,因此并不能满足实际应用的需求。VVenC[12]是 VTM 之外受到广泛关注的 H.266/VVC 开源编码器,由德国 Fraunhofer HHI 研究所基于 VTM 开发。VVenC 最大的特点是快速高效,可以用更低的复杂度获得近似 VTM 的性能,并针对实用性进行了大量优化。VVenC 编码器对 VTM 编码框架的各大模块均设计了快速算法[56]。此外,VVenC 支持真实应用场景下的实用功能,如多线程加速、可变码率控制和感知质量优化等,还支持编码预处理、高动态范围、变分辨率编码以及屏幕内容编码等功能。

通过对配置集进行帕累托优化,VVenC 设置了 5 个可选的质量/速度预设档位[57],分别为极慢、慢、中等、快和极快,可根据应用需求在编码时间和复杂度之间进行权衡和选择。以 H.265/HEVC 的参考软件 HM 为基准,VTM 以及 VVenC 在不同预设档位和线程数设置下的 BD-rate 增益 ΔR 和编码器运行时间如图 10 所示。

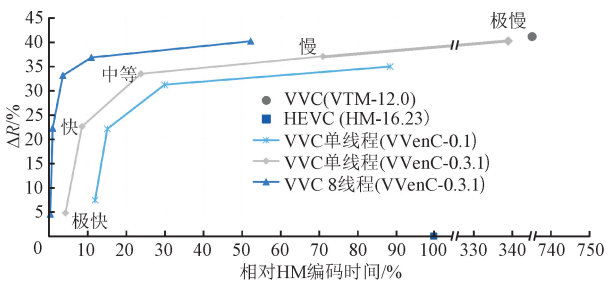


图 10 VTM 及 VVenc 相对于 HM 的编码性能[12]

Fig. 10 VTM and VVenc coding performance compared to HM[12]

2.2 H.266/VVC 算法性能分析

为深入分析 H.266/VVC 的算法性能,本节将其与 H.265/HEVC 的性能进行比较,并进一步在 VTM 中关闭各编码工具以分析具体工具对性能的贡献。此外,讨论了采用各主要工具的像素占比情况和限制条件。

2.2.1 与 H.265/HEVC 性能对比及分析

表 2 给出了随机访问配置下,H.266/VVC 与 H.265/HEVC 的性能对比。可以看出,在相同的峰值信噪比下,H.266/VVC 可节省 38.42%的编码码率。若采用主观质量作为质量测度,可节省约 50% 的编码码率。与此同时,编码复杂度为 H.265/HEVC的 7 倍,解码复杂度为 163%。在序列测试集中,class A1 与 class A2 为 4K 高分辨率视频。可以看出,H.266/VVC 针对高分辨率视频取得的编码增益更为突出。

表 2 随机访问配置下 H.266/VVC 与 H.265/HEVC 性能对比

Table 2 Coding performance comparison between H.266/VVC and H.265/HEVC under random access configuration

序列集	ΔR/%			相对复杂度/%	
	Y 通道	U 通道	V 通道	编码	解码
class A1	−40.60	−40.71	−47.16	586	158
class A2	−44.02	−41.73	−40.78	679	170
class B	−37.41	−50.12	−48.51	672	161
class C	−33.86	−36.28	−38.21	923	164
平均值	−38.42	−42.87	−43.95	713	163

2.2.2 H.266/VVC 编码算法性能分析

对 H.266/VVC 各编码工具的性能评价需综合考虑率失真性能和算法复杂度两个因素。率失真性能可通过与不含该工具的 VTM 基准进行性能比较,得到相同重建质量下的码率变化量(以 BD-rate 衡量)。针对特定的编码工具,文献[58]采用在 VTM 中关闭该工具导致的 BD-rate 上升量来衡量该工具对编码性能的贡献。编码和解码复杂度通常通过计算与 VTM 基准编码时间和解码时间的比值进行衡量。

H.266/VVC 主要编码工具全称及缩写如表 3 所示。

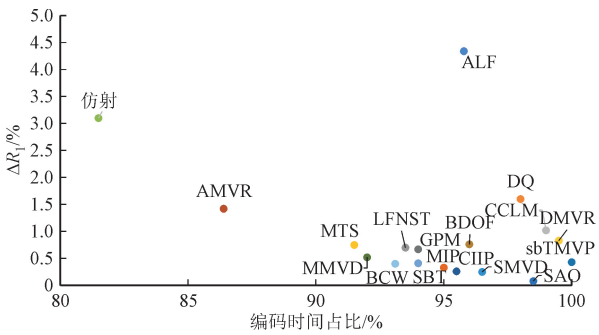
表 3　H.266/VVC 主要编码工具全称及缩写

Table 3　Full names and abbreviations of the main coding tools in H.266/VVC

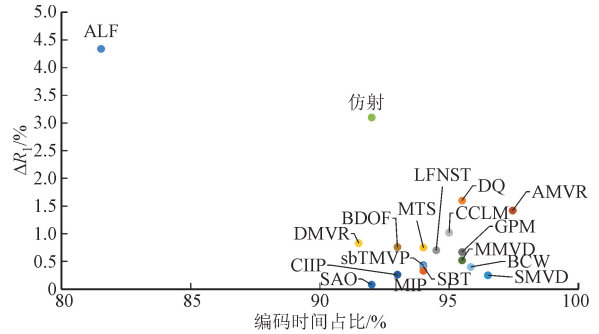
序号	中文全称	英文全称	英文缩写
1	运动矢量差值的自适应精度表示	adaptive motion vector resolution	AMVR
2	自适应环路滤波	adaptive loop filter	ALF
3	多核变换选择	multiple transform selection	MTS
4	带有运动矢量差值索引的合并模式	merge-mode with motion vector difference	MMVD
5	低频不可分变换	low-frequency non-separable transform	LFNST
6	双向加权运动补偿预测	bi-prediction with CU level weight	BCW
7	双向光流	bi-directional optical flow	BDOF
8	几何划分	geometric partitioning mode	GPM
9	基于矩阵的帧内预测	matrix-based intra prediction	MIP
10	联合帧内帧间预测	combined inter and intra prediction	CIIP
11	子块变换	sub-block transform	SBT
12	依赖量化	dependent quantization	DQ
13	分量间线性模型	cross-component linear mode	CCLM
14	对称运动矢量差值	symmetric motion vector difference coding	SMVD
15	样点自适应补偿	sample adaptive offset	SAO
16	解码端运动矢量修正	decoder-side motion vector refinement	DMVR
17	基于子块的时域运动矢量预测	subblock-based temporal motion vector prediction	sbTMVP
18	仿射	affine	

图 11 给出了 H.266/VVC 主要编码工具的性能和编解码复杂度的关系。图中：ΔR_1 表示各编码工具关闭后相比关闭前 VTM 基准的 BD-rate，反映了关闭该编码工具后压缩效率的下降（对应于码率的上升）情况；编/解码时间占比分别表示编码工具关闭后相比 VTM 基准的编/解码时间比。从图中可以看出，性能提升最明显的是自适应环路滤波，关闭该工具使得 BD-rate 增加 4.34%。与此同时，关闭自适应环路滤波后，编码时间为 VTM 基准编码时间的 96%，解码时间为 VTM 基准解码时间的 87%。由此可见，该工具在解码端具有较高的复杂度。值得注意的是，有相当一部分编码工具，在关闭后其编解码时间比反而上升，其主要原因是不采用该类编码工具通常需要将图像划分为更小的 CU 进行编解码，而更小的 CU 单元会引入更多的处理环节，从而引起编解码时间增加。

此外，编码工具的使用像素占比也可作为评估新工具有效性的测度。图 12 给出了 H.266/VVC 主要新增工具在全帧内、随机访问和双向低延迟这 3 种典型设置下的使用像素占比。综合图 11 和图 12 可以看出，使用像素占比与工具的性能贡献大致呈线性关系。例如，在参与评估的工具中，自适应环路滤波影响的像素数量最多，而其性能贡献也最大。但是也有例外。例如，仿射技术，影响像素数相对较少，但性能贡献也较大。这是因为符合非平移运动假设的像素数量未必很多，但采用仿射技术能够开展高效预测。更为详细的数据请参考文献[58]。



(a)各工具的率失真性能-编码时间



(b)各工具的率失真性能-解码时间

图 11 主要编码工具的率失真性能和编解码复杂度

Fig. 11　Rate-distortion performance and encoding/decoding complexity of key coding tools

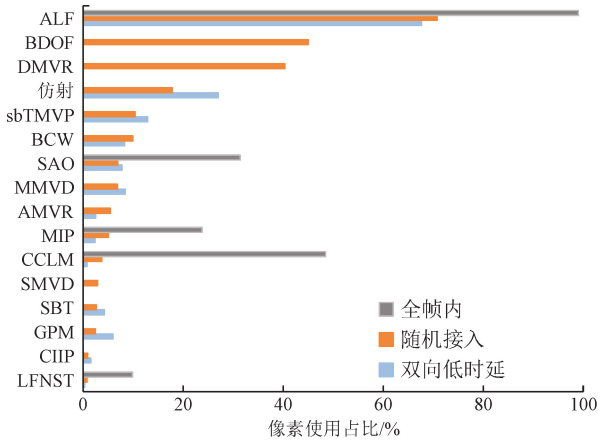


图 12 主要新增编码工具像素使用占比

Fig. 12 Pixel occupancy of key new coding tools

在 H. 266/VVC 标准发展和制定的过程中，每个新编码工具的加入都经过了详细的评估，以确保在尽可能在不增加或少增加编解码复杂度的前提下，新编码工具的加入可与其他编码工具协同作用，获取更佳的编码效率。考虑到各编码工具适用场景可能不同，标准中设置了一些编码工具使用的限制条件。例如：最可能模式列表、MIP 和多核变换选择等工具仅针对亮度分量开启使用；多核变换选择仅针对帧内编码残差使用；当使用 CCLM 模式时，禁用多参考行、PDPC 工具等等。

2.3 H. 266/VVC 硬件实现

H. 266/VVC 为方便软硬件实现和优化提供了必要的支持。与 H. 265/HEVC 一样，H. 266/VVC 支持波前技术，方便实现多线程并行编码。通过多线程并行运算，设计合理的数据结构和程序流程，以及充分利用单指令多数据操作的优化方法，目前已经可在现有主流的 x86 硬件平台上实现 4K 分辨率视频的 H. 266/VVC 码流实时解码，但同样分辨率的实时编码在目前主流硬件平台尚不可行。虽然 H. 266/VVC 最高支持 8K 分辨率视频，相应的硬件编解码实现仍任重道远。

尽管 H. 266/VVC 标准制定过程中充分考虑了算法的实现复杂度，但其在硬件实现方面仍面临巨大的挑战[59]。这些挑战主要来源于为有效实现超高分辨率视频编码引入的特性。首先，高分辨率视频本身数据量大，对其进行实时编解码需要高吞吐率和高处理速度。其次，为提高编码效率，H. 266/VVC 发展了已有编码技术，或引入了更为复杂的新技术[60]，导致对硬件运算能力的要求有进一步的提高。例如，预测模式的增多需要更多的硬件资源，计算预测值所需的逻辑也更加复杂。H. 266/VVC 新引入的 MIP 和多核变换选择需要存储一定数量的预训练矩阵，尽管在标准制定时已对矩阵系数的数量和精度进行了限制，但其对存储空间仍有较高的要求。帧间预测的精细运动估计和补偿算法则对计算能力和带宽提出了更高的要求。

从整体架构来说，H. 266/VVC 沿袭了混合式编码框架，具有并行处理能力强的特点。混合式编码框架可复用性较好，因此现有针对硬件实现的研究工作主要集中在对具体编码模块的硬件加速方面，尤其是最为耗时的预测模块、变换模块和滤波模块等。例如，文献[61]针对 VVC 中的帧内预测模式进行分析，遴选出 18 种对性能贡献较大的子模式，针对专用集成电路设计了优化算法。文献[62]针对 VVC 中的快速运动估计模块，利用整像素位置的率失真代价推断最佳预测的分像素位置，从而提高硬件友好性。文献[63]提出了统一的多核变换选择的硬件架构，便于 VVC 变换模块的硬件实现。针对自适应环路滤波，文献[64]针对亮度和色度提出了优化的扫描机制，从而针对自适应环路滤波实现内存的高效管理。

虽然面临诸多挑战，H. 266/VVC 的硬件实现已处于实用化阶段。目前，市场已出现支持 H. 266/VVC 的芯片产品。例如：MediaTek 的 Pentonic 2000 智能电视系统级芯片(system on chip，SoC)是世界首批支持 H. 266/VVC 解码的芯片[65]，RealTek[66] 和 LG[67] 也相继发布了支持 H. 266/VVC 的 SoC 产品。

3 发 展

JVET 在 H. 266/VVC 发布之后持续对视频编码的发展进行研究，以探索标准的未来走向和关键技术。现阶段的探索仍然以 H. 266/VVC 架构为基础。根据编码工具的不同类别，继 H. 266/VVC 之后，视频编码标准的发展聚焦在两大方向：Beyond VVC 探索更为先进的、非神经网络的编码工具；NNVC 则探索基于神经网络的编码工具。

3.1 基于 H. 266/VVC 框架的编码技术演进：Beyond VVC

Beyond VVC 探索的方式属于技术演进[14]，在 H. 266/VVC 编解码模块的基础上，通过改进编码工具增强编码性能。Beyond VVC 目前已经取得了令人关注的进展，相比于 H. 266/VVC 参考软件 VTM11. 0，Beyond VVC 参考软件平台 ECM (enhanced compression model) 8. 0 在全帧内配置

下可以达到 9.86%的编码增益，在随机访问配置下可以达到 19.86%的编码增益[68]。Beyond VVC 的性能提升来源于预测、变换、环路滤波等模块的综合提升[69]。

3.1.1　预测模块

帧内/帧间预测是 Beyond VVC 提升编码效率的主要模块，其增益很大程度上得益于模板的使用[70]。Beyond VVC 将模板定义为当前待编码块上方和左方的区域，在编码当前块时该区域已完成重建，如图 13 所示。模板与待编码块具有强相关性，模板与参考区域的关系可以作为待编码块的先验信息。

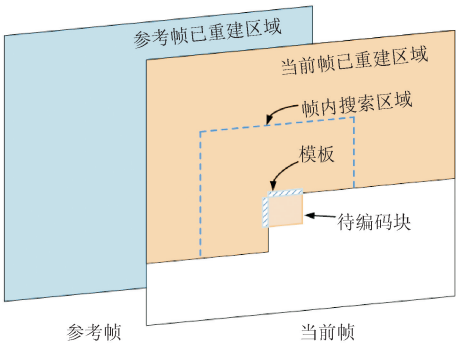


图 13　Beyond VVC 基于模板算法

Fig. 13　Template based algorithm in Beyond VVC

在帧内预测中，可利用模板估计优先预测模式或模型参数，从而高效表示预测模式等信息。基于模板的帧内模式推导[71]技术针对模板开展帧内预测，根据帧内预测的结果和模板重建信息的预测精度确定 1～2 种优先的帧内预测模式。解码端模式推导[72]技术分析模板的梯度直方图，根据梯度方向确定 1～2 种优先的帧内预测模式。帧内模板匹配模式可在当前帧已重建区域中通过模板开展搜索寻找到匹配模板，为待编码块确定预测块，从而提高编码效率。

在帧间预测中，模板广泛地应用于运动矢量预测、运动矢量确定和预测值获取过程中。例如，针对参考帧已重建区域，可利用模板对运动矢量列表的多个候选进行排序，也可利用模板进行搜索提升运动矢量的精度。考虑到模板与待编码块的特性可能存在差异，通常利用模板确定多个候选以及候选顺序，最佳预测模式需根据其用于待编码块的率失真性能确定。

在色度帧内预测中，建立精细分量间预测模型是提升编码效率的有效途径。卷积分量间预测模型[73]建立邻域内多位置亮度与色度的预测模型，并依据模板区域计算模型参数，自适应建立亮度与色度卷积预测模型。多模型线性模型为亮度与色度建立分段线性预测模型。为了获取更准确的预测模型，Beyond VVC 通过调整 CCLM 的模型参数[74]，或者通过在预测模型中添加梯度因子与位置信息等方法[75]提升预测模型的准确性。另外，由作者提出的直接块矢量模式，充分利用亮度已有块矢量信息推导色度块矢量[76]。进一步地，Beyond VVC 采用了色度融合技术，设计了传统色度预测模式与分量间预测模式的自适应加权方案。

3.1.2　变换模块

Beyond VVC 进一步拓展了多核变换选择的变换核候选集，同时利用新提出的可分离 KL 变换变换核将多核变换选择应用于帧间编码。Beyond VVC 还提出不可分离一次变换[77]，对于符合相应特性的预测残差，经一次变换后可实现能量的高度集中。另外，基于与已重建模板像素的连续性，对部分变换系数符号进行预测，可进一步提升变换系数的编码效率[78]。

3.1.3　环路滤波

Beyond VVC 中的自适应环路滤波引入了基于边带分类和基于残差样本的分类方法。此外，分量间联合样点自适应补偿[79]通过对亮度值和色度值联合分类，确定补偿值，可得到更为精细的补偿值。双边滤波器[80]对去方块滤波的输出图像进行处理，综合考虑样本的空间相邻度和像素相似度生成补偿值，可达到去除噪声同时保留边缘的目的。

3.2　基于神经网络的编码工具探索：NNVC

2020 年 10 月的 JVET 会议上，正式确定建立 NNVC 的探索实验，研究利用神经网络强大的非线性表达能力提升视频编码性能。目前，NNVC 探索实验[15]主要包括基于神经网络的环路滤波技术、基于神经网络的帧内/帧间预测技术、基于神经网络的超分辨率技术以及基于神经网络的后处理技术等。

基于神经网络的环路滤波采用神经网络滤波器对重建视频进行处理[81-83]，研究重点为有限计算复杂度下的网络架构设计、网络模型输入的编码参数、与传统滤波处理算法的融合等。基于神经网络的帧间预测重点研究采用神经网络生成高质量参考帧、替换传统双向加权预测的多帧预测等[84]。基于神经网络的帧内预测[85]使用神经网络从当前图像已重建的像素产生更加准确的预测块，或是根据已重建的亮度信息使用神经网络产生准确的色度信

息[86]，从而更好地利用视频的空间信息。基于神经网络的超分辨率研究通过编码低分辨率视频节省码率，解码端将低分辨率视频超分恢复到高分辨率[87]。基于神经网络的后处理对重建视频进行滤波，但不同于环内滤波，滤波后图像不会作为后续图像的参考，网络可以利用编码端提供的一些辅助信息，提升滤波性能[88]。

基于神经网络的视频编码工具在 H.266/VVC 的基准上能够取得较好的性能增益[89]，但该类编码工具也具有高计算复杂度的特点。文献[90]提出使用 CP 分解以及融合相邻 1×1 卷积等操作有效降低计算复杂度。如何在神经网络的性能增益和计算复杂度之间进行合理权衡是目前聚焦的研究趋势。2023 年 4 月的 JVET 会议上，与会专家达成共识，下一阶段的主要任务是建立不同复杂度下的滤波方案[91]，研究如何统一设计滤波器的输入、如何开展高效的训练以及统一滤波器如何与现有编解码器高效融合。上述任务的确立标志着 NNVC 的发展进入了新的阶段。

4 总结与展望

新一代视频编码标准 H.266/VVC 面向多样化的应用，从高层语法和编码工具两个层面进行了大量技术革新，能够适应不同的网络环境和应用需求。H.266/VVC 仍然采用成熟的混合编码框架，对于软硬件实现十分友好，具有良好的实用化前景。H.266/VVC的技术革新体现在所有主要编码模块，包括帧内预测、帧间预测、变换、量化、环路滤波等。标准参考模型 VTM 具有出色的压缩性能，在此基础上也出现了更为高效的开源编解码器和相关应用。

由于各行各业对视频应用的需求迫切且持续增长，H.266/VVC 具有广阔的应用前景。然而，从标准发布到大规模商用往往存在较长的周期。例如，H.265/HEVC 于 2013 年标准制定完成，直至近年来才具有较大的市场占比。基于此，预估 H.266/VVC 的市场接纳期也会持续较长时间。H.266/VVC 最先进入的领域预计是以高效软件编码器为主的互联网媒体领域，如短视频应用等。此外，H.266/VVC的实用有望推动可分级视频、高动态视频、屏幕视频等应用的发展。H.266/VVC 硬件实现的应用率先出现在允许较大尺寸的电视芯片领域。至于其他消费类应用、军工应用等领域，则更多地需要成熟的芯片方案和可接受的算力来支撑。另外，H.266/VVC 的相关应用也可能拓展到图像压缩领域。例如，高效图像文件压缩格式源于 H.265/HEVC的帧内编码，在图像压缩获得了有效应用，H.266/VVC 中的帧内编码技术也可能产生类似的发展。

虽然 H.266/VVC 标准已经进入实用化阶段，JVET 对 H.266/VVC 后续标准的发展仍保持着不懈探索。在现有标准架构下，JVET 分别从采用非神经网络工具 Beyond VVC 和神经网络工具 NNVC 两个不同的方向进行持续研究，都取得了显著进展。然而，哪一个方向才是下一代视频编码标准的走向，目前尚未形成明确结论。

值得关注的是，部分跳出或完全跳出现有混合编码框架的端到端视频编码也在飞速发展。基于神经网络的端到端视频编码大致可分为 3 类：基于残差编码的方案，基于条件编码的方案，基于 3D 自编码器的方案。基于残差编码的方案借鉴传统混合视频编解码器的思路，通过神经网络编码工具进行运动补偿生成预测帧，并将其与当前帧的残差进行编码，如 DVC[92]。基于条件编码的方案将时域相关的帧或特征作为当前帧编码的条件，如 DCVC 及其后续一系列改进[93]。基于 3D 自编码器方案则是相关图像编解码器的自然延伸，扩大了网络输入的维度[94-95]。除上述方案之外，隐式神经网络表达采用过拟合的神经网络进行信源表示，将视频压缩问题转化为神经网络压缩问题，从全新的角度实现了视频编码[96-97]。

目前，针对静止图像，国际标准化组织正在致力于研究使用端到端的神经网络进行高效压缩，即正在进行中的 JPEG AI 标准[98]。预计使用神经网络进行视频编码的国际标准也将会出现在或远或近的未来。无论是与传统编码框架融合，还是采用全新的端到端实现，现阶段基于神经网络的视频编码面临两方面的考验。一是如何能够克服其对计算资源的依赖。二是如何能够定义一个通用的稳定结构，使得日益变化的网络能够以极低的代价在终端上进行迭代更新。未来视频编码标准将如何与神经网络结合，又在何时能够获得实用化，将在对相关技术的探索中逐渐明朗。

参考文献：

[1] ISO. Information technology: generic coding of moving pictures and associated audio information: part 2 video: ISO/IEC 13818-2: 2013 [S]. Vernier, Geneva, Switzerland: ISO, 2013.

[2] ISO. Information technology: coding of audio-visual objects: part 10 advanced video coding: ISO/IEC 14496-10: 2022 [S]. Vernier, Geneva, Switzerland: ISO, 2022.

[3] ISO. Information technology: high efficiency coding and media delivery in heterogeneous environments: part 2 high efficiency video coding: ISO/IEC 23008-2: 2023 [S]. Vernier, Geneva, Switzerland: ISO, 2023.

[4] CHEN Jianle, KARCZEWICZ M, HUANG Yuwen, et al. The joint exploration model (JEM) for video compression with capability beyond HEVC [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(5): 1208-1225.

[5] BROSS B. Versatile video coding (draft 1): JVET-J1001 [EB/OL]. (2018-04-25) [2023-06-01]. https://jvet-experts.org/JVET-J1001-v1.doc.

[6] CHEN Jianle, ALSHINA E. Algorithm description for versatile video coding and test model 1 (VTM 1) [EB/OL]. (2018-04-25) [2023-06-01]. https://jvet-experts.org/JVET-J1002-v2.doc.

[7] ISO. Information technology: coded representation of immersive media: part 3 versatile video coding: ISO/IEC 23090-3: 2022 [S]. Vernier, Geneva, Switzerland: ISO, 2022.

[8] BROSS B, WANG Yekui, YE Yan, 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.

[9] BROSS B, CHEN Jianle, OHM J R, et al. Developments in international video coding standardization after AVC, with an overview of versatile video coding (VVC) [J]. Proceedings of the IEEE, 2021, 109(9): 1463-1493.

[10] 万帅，霍俊彦，马彦卓，等. 新一代通用视频编码 H.266/VVC：原理、标准与实现 [M]. 北京：电子工业出版社，2022.

[11] 朱秀昌，唐贵进. H.266/VVC：新一代通用视频编码国际标准 [J]. 南京邮电大学学报(自然科学版)，2021，41(2)：1-11.
ZHU Xiuchang, TANG Guijin. H.266/VVC: versatile video coding international standard [J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2021, 41(2): 1-11.

[12] WIECKOWSKI A, BRANDENBURG J, BARTNIKET C, et al. Open optimized VVC encoder (VVenC) and decoder (VVdeC) implementations [EB/OL]. (2020-10-06) [2023-06-01]. https://jvet-experts.org/JVET-T0099-v3.doc.

[13] SULLIVAN G J. Deployment status of the VVC standard [EB/OL]. (2023-01-11) [2023-06-01]. https://jvet-experts.org/JVET-AC0021-v4.doc.

[14] SEREGIN V, CHEN Jie, LI Guichun, et al. EE2: summary report of exploration experiments on enhanced compression beyond VVC capability [EB/OL]. (2023-01-13) [2023-06-01]. https://jvet-experts.org/JVET-AC0024-v4.doc.

[15] ALSHINA A, GALPIN F, LI Yue, et al. EE1: summary report of exploration experiments on neural network-based video coding [EB/OL]. (2023-01-14) [2023-06-29]. https://jvet-experts.org/JVET-AC0023-v3.doc.

[16] WANG Yekui, SKUPIN R, HANNUKSELA M M, et al. The high-level syntax of the versatile video coding (VVC) standard [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3779-3800.

[17] HUANG Yuwen, AN Jicheng, HUANG Han, et al. Block partitioning structure in the VVC standard [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3818-3833.

[18] SULLIVAN G J, WIEGAND T. Rate-distortion optimization for video compression [J]. IEEE Signal Processing Magazine, 1998, 15(6): 74-90.

[19] ORTEGA A, RAMCHANDRAN K. Rate-distortion methods for image and video compression [J]. IEEE Signal Processing Magazine, 1998, 15(6): 23-50.

[20] 公衍超，张森森，何书婷，等. 面向监控视频内容自适应的通用视频编码量化参数级联 [J]. 西安交通大学学报，2023，57(12)：30-40.
GONG Yanchao, ZHANG Sensen, HE Shuting, et al. Content adaptive quantization parameter cascading for surveillance video coding using versatile video coding [J]. Journal of Xi'an Jiaotong University, 2023, 57(12): 30-40.

[21] PFAFF J, FILIPPOV A, LIU Shan, et al. Intra prediction and mode coding in VVC [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3834-3847.

[22] CHANG Y J, JHU H J, JIAN Huiyu, et al. Intra prediction using multiple reference lines for the versatile video coding standard [C]//Proceedings Volume 11137, Applications of Digital Image Processing XLII. Bellingham, WA, USA: SPIE, 2019: 1113716.

[23] ZHAO Liang, ZHAO Xin, LIU Shan, et al. Wide angular intra prediction for versatile video coding [C]//

2019 Data Compression Conference (DCC). Piscataway, NJ, USA: IEEE, 2019: 53-62.

[24] DE-LUXÁN-HERNÁNDEZ S, GEORGE V, MA J, et al. An intra subpartition coding mode for VVC [C]//2019 IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE, 2019: 1203-1207.

[25] SAID A, ZHAO Xin, KARCZEWICZ M, et al. Position dependent prediction combination for intra-frame video coding [C]//2016 IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE, 2016: 534-538.

[26] PFAFF J, HELLE P, MANIRY D, et al. Neural network based intra prediction for video coding [C]//Proceedings Volume 10752, Applications of Digital Image Processing XLI. Bellingham, WA, USA: SPIE, 2018: 1075213.

[27] HUO Junyan, SUN Yu, WANG Haixin, et al. Unified matrix coding for NN originated MIP in H.266/VVC [C]//ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ, USA: IEEE, 2022: 1635-1639.

[28] LI Junru, WANG Meng, ZHANG Li, et al. Sub-sampled cross-component prediction for emerging video coding standards [J]. IEEE Transactions on Image Processing, 2021, 30: 7305-7316.

[29] HUO Junyan, DU Hongqing, LI Xinwei, et al. Unified cross-component linear model in VVC based on a subset of neighboring samples [J]. IEEE Transactions on Industrial Informatics, 2022, 18(12): 8654-8663.

[30] CHIEN W J, ZHANG Li, WINKEN M, et al. Motion vector coding and block merging in the versatile video coding standard [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3848-3861.

[31] YANG Haitao, CHEN Huanbang, CHEN Jianle, et al. Subblock-based motion derivation and inter prediction refinement in the versatile video coding standard [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3862-3877.

[32] HEITHAUSEN C, VORWERK J H. Motion compensation with higher order motion models for HEVC [C]//2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ, USA: IEEE, 2015: 1438-1442.

[33] HUANG Han, WOODS J W, ZHAO Yao, et al. Control-point representation and differential coding affine-motion compensation [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(10): 1651-1660.

[34] LI Li, LI Houqiang, LIU Dong, et al. An efficient four-parameter affine motion model for video coding [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(8): 1934-1948.

[35] GAO Han, LIAO Ruling, REUZÉ K, et al. Advanced geometric-based inter prediction for versatile video coding [C]//2020 Data Compression Conference (DCC). Piscataway, NJ, USA: IEEE, 2020: 93-102.

[36] LIU Hongbin, ZHANG Li, ZHANG Kai, et al. Two-pass bi-directional optical flow via motion vector refinement [C]//2019 IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE, 2019: 1208-1211.

[37] LUO Jiancong, HE Yuwen, CHEN Wei. Prediction refinement with optical flow for affine motion compensation [C]//2019 IEEE Visual Communications and Image Processing (VCIP). Piscataway, NJ, USA: IEEE, 2019: 1-4.

[38] ZHAO Xin, KIM S H, ZHAO Yin, et al. Transform coding in the VVC standard [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3878-3890.

[39] ZHAO Yin, GAO Han, YANG Haitao, et al. CE6: sub-block transform for inter blocks(test 6.4.1) [EB/OL]. (2019-01-16) [2023-06-01]. https://jvet-experts.org/JVET-M0140-v3.doc.

[40] ZHAO Xin, CHEN Jianle, KARCZEWICZ M, et al. Joint separable and non-separable transforms for next-generation video coding [J]. IEEE Transactions on Image Processing, 2018, 27(5): 2514-2525.

[41] KOO M, SALEHIFAR M, LIM J, et al. Low frequency non-separable transform (LFNST) [C]//2019 Picture Coding Symposium (PCS). Piscataway, NJ, USA: IEEE, 2019: 1-5.

[42] 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.

[43] KARCZEWICZ M, YE Y, CHONG I. Rate distortion optimized quantization [EB/OL]. (2008-09-01) [2023-06-29]. https://www.itu.int/wftp3/av-arch/video-site/0801_Ant/VCEG-AH21.doc.

[44] MARCELLIN M W, FISCHER T R. Trellis coded quantization of memoryless and Gauss-Markov sources

[J]. IEEE Transactions on Communications, 1990, 38(1): 82-93.

[45] JOSHI R L, CRUMP V J, FISCHER T R. Image subband coding using arithmetic coded trellis coded quantization [J]. IEEE Transactions on Circuits and Systems for Video Technology, 1995, 5(6): 515-523.

[46] MARPE D, SCHWARZ H, WIEGAND T. Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2003, 13(7): 620-636.

[47] KARCZEWICZ M, HU Nan, TAQUET J, et al. VVC in-loop filters [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3907-3925.

[48] LU Taoran, PU Fangjun, YIN Peng, et al. Luma mapping with chroma scaling in versatile video coding [C]//2020 Data Compression Conference (DCC). Piscataway, NJ, USA: IEEE, 2020: 193-202.

[49] TSAI C Y, CHEN C Y, YAMAKAGE T, et al. Adaptive loop filtering for video coding [J]. IEEE Journal of Selected Topics in Signal Processing, 2013, 7(6): 934-945.

[50] NGUYEN T, XU Xiaozhong, HENRY F, et al. Overview of the screen content support in VVC: applications, coding tools, and performance [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3801-3817.

[51] XU Xiaozhong, LIU Shan. Overview of screen content coding in recently developed video coding standards [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(2): 839-852.

[52] BOSSEN F, BOYCE J, SUEHRING K, et al. VTM common test conditions and software reference configurations for SDR video [EB/OL]. (2020-11-19) [2023-06-01]. https://jvet-experts.org/JVET-T2010-v2.doc.

[53] BOSSEN F, BOYCE J, SUEHRING K, et al. Test sequences [EB/OL]. (2018-04-20) [2023-07-03]. ftp://jvet@ftp.hhi.fraunhofer.de.

[54] BJØNTEGAARD G. Calculation of average PSNR differences between *RD*-curves [EB/OL]. (2021-03-28) [2023-06-01]. https://www.itu.int/wftp3/av-arch/video-site/0104_Aus/VCEG-M33.doc.

[55] STRÖM J. Summary information on BD-rate experiment evaluation practices [EB/OL]. (2023-05-17) [2023-06-01]. https://jvet-experts.org/JVET-R2016-v1.doc.

[56] WIECKOWSKI A, BRANDENBURG J, HINZ T, et al. VVenc: an open and optimized VVC encoder implementation [C]//2021 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). Piscataway, NJ, USA: IEEE, 2021: 1-2.

[57] BRANDENBURG J, WIECKOWSKI A, HENKEL A, et al. Pareto-optimized coding configurations for VVenC, a fast and efficient VVC encoder [C]//2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP). Piscataway, NJ, USA: IEEE, 2021: 1-6.

[58] CHIEN W, BOYCE J, CHEN Y, et al. JVET AHG report: tool reporting procedure and testing(AHG13) [EB/OL]. [2020-12-14](2023-06-01). https://jvet-experts.org/JVET-T0013-v2.doc.

[59] 范益波. 视频编解码芯片设计原理 [M]. 北京：科学出版社，2022.

[60] 陈俊煜，孙斌，黄晓峰，等. H.266/VVC 二维变换的统一硬件结构 [J]. 浙江大学学报(工学版)，2023，57(9)：1894-1902.
CHEN Junyu, SUN Bin, HUANG Xiaofeng, et al. Unified hardware architecture for 2D transform in H.266/VVC [J]. Journal of Zhejiang University(Engineering Science), 2023, 57(9): 1894-1902.

[61] BORGES V, PERLEBERG M, PORTO M, et al. Efficient architecture for VVC angular intra prediction based on a hardware-friendly heuristic [C]//2023 IEEE 14th Latin America Symposium on Circuits and Systems (LASCAS). Piscataway, NJ, USA: IEEE, 2023: 1-4.

[62] CHEN Shushi, HUANG Leilei, LIU Jiahao, et al. An error-surface-based fractional motion estimation algorithm and hardware implementation for VVC [EB/OL]. (2023-02-13) [2023-06-01]. https://arxiv.org/abs/2302.06167.

[63] SILVEIRA B, NETO L, PALOMINO D, et al. Multiple transform selection hardware design for 4K @ 60fps real-time versatile video coding [C]//2022 IEEE International Symposium on Circuits and Systems (ISCAS). Piscataway, NJ, USA: IEEE, 2022: 1-5.

[64] FARHAT I, HAMIDOUCHE W, GRILL A, et al. Adaptive loop filter hardware design for 4K ASIC VVC decoders [J]. IEEE Transactions on Consumer Electronics, 2022, 68(2): 107-118.

[65] DE LOOPER C. MediaTek wants to power next-generation TVs and Chromebooks [EB/OL]. [2023-11-10]. https://ustoday.news/mediatek-wants-to-power-next-generation-tvs-and-chromebooks/.

[66] RealTek. RealTek launches world's first 4K UHD set-top box SoC (RTD1319D) supports VVC/H. 266 video decoding, GPU with 10-bit graphics, multiple CAS, and HDMI 2. 1a [EB/OL]. (2022-08-29) [2023-06-01]. https://www. realtek. com/zh-tw/component/zoo/item/realtek-launches-world-s-first-4k-uhd-set-top-box-soc-rtd1319d.

[67] LG. LG QNED99 75 inch 8K smart QNED TV and LG QNED99 86 inch 8K smart QNED TV [EB/OL]. [2023-04-17]. https://www. lg. com/hk_en/tv/lg-75qned99cpb.

[68] SEREGIN V, CHEN Jie, LÉANNEC F L, et al. JVET AHG report: ECM software development (AHG6) [EB/OL]. (2023-01-03) [2023-06-01]. https://jvet-experts. org/JVET-AC0006-v1. doc.

[69] COBAN M, LÉANNEC F L, LIAO Ruling, et al. Algorithm description of enhanced compression model 8 (ECM 8) [EB/OL]. (2023-04-06) [2023-06-01]. https://jvet-experts. org/JVET-AC2025-v1. doc.

[70] LI Xiang, LIEN-FEI C, DENG Zhipin, et al. JVET AHG report: ECM tool assessment (AHG7) (2023-04-18) [2023-06-01]. https://jvet-experts. org/JVET-AD0007-v1. doc.

[71] CAO Keming, HU Nan, SEREGIN V, et al. EE2-related: fusion for template-based intra mode derivation [EB/OL]. [2021-07-01]. https://jvet-experts. org/JVET-W0123-v1. doc.

[72] ABDOLI M, GUIONNET T, MORA E, et al. Non-CE3: decoder-side intra mode derivation with prediction fusion using planar [EB/OL]. (2019-07-04) [2023-06-01]. https://jvet-experts. org/JVET-O0449-v2. doc.

[73] ASTOLA P, LAINEMAY, YOUVALARI R, et al. AHG12: convolutional cross-component model (CCCM) for intra prediction [EB/OL]. (2022-04-13) [2023-06-01]. https://jvet-experts. org/JVET-Z0064-v1. doc.

[74] LAINEMA J, AMINLOU A, ASTOLA P, et al. AHG12: slope adjustment for CCLM [EB/OL]. (2022-01-12) [2023-06-01]. https://jvet-experts. org/JVET-Y0055-v2. doc.

[75] KUO Chewei, HONG-JENG J, XIU Xiaoyu, et al. EE2-1. 1b and 1. 2: filter-based linear model and gradient linear model [EB/OL]. (2022-07-21) [2023-06-01]. https://jvet-experts. org/JVET-AA0125-v2. doc.

[76] HUO Junyan, HAO Xue, MA Yanzhuo, et al. EE2-3. 1: direct block vector mode for chroma prediction [EB/OL]. (2023-01-01) [2023-6-01]. https://jvet-experts. org/JVET-AC0071-v1. doc.

[77] YAN Ning, XIU Xiaoyu, CHEN Wei, et al. Non-EE2: on non-separable primary transform for intra blocks [EB/OL]. (2023-01-01) [2023-06-01]. https://jvet-experts. org/JVET-AC0163-v1. doc.

[78] XIU Xiaoyu, CHEN Yiwen, YAN Ning, et al. EE2-4. 2: enhanced sign prediction [EB/OL]. (2023-01-01) [2023-06-01]. https://jvet-experts. org/JVET-AC0137-v1. doc.

[79] KUO Chewei, XIU Xiaoyu, CHEN Yiwen, et al. EE2-5. 1: cross-component sample adaptive offset [EB/OL]. (2021-07-01) [2023-06-01]. https://jvet-experts. org/JVET-W0066-v1. doc.

[80] STRÖM J, WENNERSTEN P, ANDERSSON K, et al. EE2: bilateral filter in VTM, EE2 and VVenC [EB/OL]. (2021-04-22) [2023-06-01]. https://jvet-experts. org/JVET-V0094-v4. doc.

[81] LI Junru, LI Yue, LIN Chaoyi, et al. A neural-network enhanced video coding framework beyond VVC [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2022: 1780-1784.

[82] LI Yue, ZHANG Kai, ZHANG Li. EE1-1. 7: deep in-loop filter with additional input information [EB/OL]. (2023-01-05) [2023-06-01]. https://jvet-experts. org/JVET-AC0177-v1. doc.

[83] CHANG Renjie, WANG Liqiang, XU Xiaozhong, et al. EE1-1. 1: more refinements on NN based in-loop filter with a single model [EB/OL]. (2023-01-05) [2023-06-01]. https://jvet-experts. org/JVET-AC0194-v1. doc.

[84] ZHAO Lei, WANG Shiqi, ZHANG Xinfeng, et al. Enhanced motion-compensated video coding with deep virtual reference frame generation [J]. IEEE Transactions on Image Processing, 2019, 28(10): 4832-4844.

[85] DUMAS T, GALPIN F, BORDES F. EE1-3. 2: neural network-based intra prediction with learned mapping to VVC intra prediction modes [EB/OL]. (2023-01-04) [2023-06-01]. https://jvet-experts. org/JVET-AC0116-v1. doc.

[86] 霍俊彦, 王丹妮, 马彦卓, 等. 基于轻量级全连接网络的 H. 266/VVC 分量间预测 [J]. 通信学报, 2022, 43(2): 143-155.
HUO Junyan, WANG Danni, MA Yanzhuo, et al. Efficient cross-component prediction for H. 266/VVC based on lightweight fully connected networks [J]. Journal on Communications, 2022, 43(2): 143-155.

[87] CHANG Renjie, WANG Liqiang, XU Xiaozhong, et al. EE1-2.2: GOP level adaptive resampling with CNN-based super resolution [EB/OL]. (2023-01-05) [2023-06-01]. https://jvet-experts.org/JVET-AC0196-v1.doc.

[88] SANTAMARIA M, YANG Ruiying, CRICRI F, et al. EE1-1.11: content-adaptive post-filter [EB/OL]. (2023-01-03) [2023-06-01]. https://jvet-experts.org/JVET-AC0055-v1.doc.

[89] EADIE S, GALPIN F, LI Yue, et al. JVET AHG report: NNVC software development (AHG14) [EB/OL]. (2023-04-20) [2023-06-01]. https://jvet-experts.org/JVET-AD0014-v3.doc.

[90] SHINGALA J N, KADARAMANDALGI S, SHYAM A, et al. AHG11: complexity reduction on neural-network loop filter [EB/OL]. (2022-07-18) [2023-06-01]. https://jvet-experts.org/JVET-AA0080-v2.doc.

[91] ALSHINA E, GALPIN F. BoG report on NN-filter design unification [EB/OL]. (2023-04-27) [2023-06-01]. https://jvet-experts.org/JVET-AD0380-v7.doc.

[92] LU Guo, OUYANG Wanli, XU Dong, et al. DVC: an end-to-end deep video compression framework [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 10998-11007.

[93] SHENG Xihua, LI Jiahao, LI Bin, et al. Temporal context mining for learned video compression [J]. IEEE Transactions on Multimedia, 2023, 25: 7311-7322.

[94] PESSOA J, AIDOS H, TOMÁS P, et al. End-to-End learning of video compression using Spatio-Temporal autoencoders [C]//2020 IEEE Workshop on Signal Processing Systems (SiPS). Piscataway, NJ, USA: IEEE, 2020: 1-6.

[95] SUN Wenyu, TANG Chen, LI Weigui, et al. High-quality single-model deep video compression with frame-Conv3D and multi-frame differential modulation [C]//Computer Vision-ECCV 2020. Cham, Germany: Springer International Publishing, 2020: 239-254.

[96] ZHANG Yunfan, VAN ROZENDAAL T, BREHMER J, et al. Implicit neural video compression [EB/OL]. (2021-12-21) [2023-07-01]. https://arxiv.org/abs/2112.11312.

[97] CHEN Hao, HE Bo, WANG Hanyu, et al. NeRV: neural representations for videos [C]//Advances in Neural Information Processing Systems. San Francisco, CA, USA: Curran Associates Inc., 2021: 21557-21568.

[98] JPEG. Overview of JPEG AI [EB/OL]. [2023-07-02]. https://jpeg.org/jpegai/.

（编辑　陶晴）