视频编码的技术基础及发展方向

周建同[1]，杨海涛[1]，刘东[2]，马祥[1]，王田[1]
（1. 华为技术有限公司，广东 深圳 518129；2. 中国科学技术大学，安徽 合肥 230026）

摘　要：现有视频编码采用基于块的混合编码架构，利用预测、变换、量化和熵编码技术实现对视频信号的高效压缩。在现有架构基础上进一步优化，提供针对视频图像信号局部特性的更加灵活的处理和编码。基于机器学习的视频编码技术有望部分或全面地改变现有的混合编码框架，给视频编码带来新的研究思路。未来视频除了现有的二维平面视频，还需要编码面向 AR/VR 应用的球面视频数据和体视频数据，这些新的视频源数据格式也给视频编码技术研究带来新的机会和挑战。

关键词：视频编码；机器学习；虚拟现实；球面视频；体视频
中图分类号：TP393　　**文献标识码**：A
doi: 10.11959/j.issn.1000−0801.2017248

Trends and technologies of video coding

ZHOU Jiantong[1], YANG Haitao[1], LIU Dong[2], MA Xiang[1], WANG Tian[1]
1. Huawei Technologies Co., Ltd., Shenzhen 518129, China
2. University of Science and Technology of China, Heifei 230026, China

Abstract: The current video coding uses block based hybrid architecture, which uses predictive, transform, quantization and entropy coding techniques to efficiently compress video signals. Further optimizations on current architectures provide more flexible processing and coding for local characteristics of video image signals. Video coding based on machine learning was expected to change the existing hybrid coding framework partially or comprehensively, and bring new research ideas to video coding. In addition to existing 2D video signal, the future of video also needs to spherical video coding and volumetric video coding for AR/VR applications, the new video source data format of the video encoding technology has brought new opportunities and challenges.

Key words: video coding, machine learning, virtual reality, spherical video, volumetric video

1　引言

视频编码技术是数字媒体应用的关键基础性技术，从 ITU 在 1984 年发布业界第一个数字视频编码标准 H.120 至今，经过近 30 年的发展，产生一代代视频编码技术，支撑了 DVD、数字电视和 IPTV[1]、互联网视频服务[2,3]、视频监控、AR/VR 等视频相关产业和应用的发展。

视频编码技术的主要目的是解决大数据量视频信息的高效存储和传输，在有限资源的情况下，保证尽可能地提高用户视觉体验。而随着视频采集和显示设备的更新换代，视频体验在过去

收稿日期：2017−07−15；修回日期：2017−08−08

30 年持续提升和变革，视频产业也空前繁荣。目前空间分辨率达到 4K（4 096×2 160 像素）、时间分辨率超过 50 帧/s 的超高清视频已经开始商用。而新兴的虚拟现实和增强现实等浸入式视频应用更是牵引视频从过去 30 年平面视频提供的零自由度体验向三自由度全景视频（球面数据）和六自由度视频（体数据）发展，如图 1 所示。对于三自由度全景视频空间分辨率 4K 只是“起步价”，未来空间分辨率将达到 8K、16K 乃至更高，时间分辨率要求也随之提升，90 帧/s 乃至更高的帧率能提供更好的用户体验。而六自由度视频更是颠覆了人们的传统视频体验，提供视频信息呈现空间内的漫游体验，具备多视点和多视角的特点。这些新的浸入式视频体验和应用给视频编码技术提出了更高的要求，也对现有的存储和传输系统提出了很大的挑战。即便使用目前最新最高效的通用视频编码国际标准高性能视频编码 HEVC 技术，由于压缩效率不够，码流的速率仍远远超出现有网络的承载能力，极大地限制了相关业务的发展。因此进一步提高视频编码的压缩效率、提供面向三自由度视频和六自由度视频的高效编码方案有迫切的需求。本文将从以下 3 个方面，对视频编码的技术发展进行初步探讨：现有混合编码架构的持续演进；基于机器学习理论的视频编码技术突破；六自由度视频体数据的技术探索。

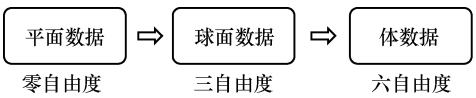


图 1　浸入式视频发展趋势

2　现有混合编码架构的持续演进

基于混合编码框架的现有视频编码的技术仍在不断演进。现有视频编码技术的基本框架即所谓混合编码框架在 30 年来没有发生大的变化，仍然沿袭了图像分块结构，配合预测—变换—量化—熵编码的流程，只是每项具体方法，如预测方法、变换方法等，不断演进，提供更为精细的局部信号自适应能力，通过几百种处理算法的灵活组合模式，提升视频的压缩效率，然后在摩尔定律的帮助下，解决复杂度提升带来的问题。其理论思路主要是从信号处理理论出发，推导和设计不同的预测方法、变换方法等，再利用率失真优化理论来进行编码模式选择，从一组编码方法或参数中根据率失真代价最小的准则选择最优的方法或参数。

混合编码架构作为业界主流的视频编码技术架构体系，其技术研究和标准化主要由两个国际标准组织——ISO/IEC WG11 运动图像专家组 MPEG 和 ITU-T SG16 的视频编码专家组 VCEG 联合推动。最新一代的视频编码标准 HEVC/H.265 在 2013 年发布一年后，MPEG 组织于 2014 年 10 月在法国斯特拉斯堡召开了第一届未来视频编码（FVC）技术论坛[4]。谷歌、华为、高通和 Netflix 等公司受邀做专题报告，结合各自产品与服务，从不同角度阐述工业界对视频压缩编码技术发展的观点，并提出更高压缩效率的诉求。2015 年 2 月会议中，MPEG 启动面向未来视频压缩编码的相关标准工作，包括制定工作计划[5]、进一步汇总和梳理工业界需求[6]、征集潜在应用于未来视频压缩编码的技术方案[7,8]。2015 年 10 月，MPEG 与 VCEG 两个组织宣布建立联合视频探索组（Joint Video Exploration Team，JVET）开展面向 FVC 的技术研究探索[6]。2017 年 4 月，JVET 确定未来视频编码标准时间表，于 2017 年 10 月公开征集标准候选技术方案，2018 年 4 月启动正式标准化工作流程，2020 年完成标准化并发布。FVC 针对业界热点应用，将更好地满足 4K/8K 超高清视频、高动态范围 HDR 视频和 VR 全景视频的应用需求。其中，HDR 视频编码和 VR 全景视频编码会基于 4K/8K 超高清视频编码算法架构，叠加个性化编码工具以及系统层适配参数，满

足各自特殊需求。从 2015 年至今，FVC 通过建立联合探索模型（joint exploration test model，JEM），吸引了业界诸多公司和研究机构的力量，高通[10]、三星、华为、联发科、英特尔、微软等公司以及德国的 HHI 研究所等都在积极投入。截至目前，FVC 的编码性能已经比 HEVC/H.265 在超高清视频[11]上提升 32%以上[12]，其性能结果如图 2 所示，标志着面向下一代视频压缩标准的技术储备已趋成熟。

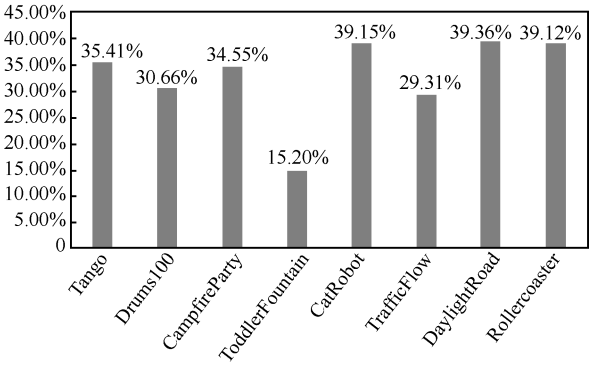


图 2　FVC 在超高清视频序列上的编码性能

当前 FVC 编码模型引入多项编码技术，涵盖超高清视频图像块划分、帧内预测、帧间预测、空间变换、环路滤波和熵编码等视频编/解码器（如 codec）关键模块。和之前的视频编码标准技术相比，如下 4 项关键技术对现有编码架构改变较大，并带来显著的性能提升。

2.1　灵活的图像块划分技术

HEVC/H.265 使用基于四叉树的块划分方法[13]，将一个图像区域 CTU 作为四叉树的根节点，按照节点一分为四的方式（如图 3（a）所示），将 CTU 递归划分成若干个叶节点，每个叶节点为基本的编码单元。

FVC 在 QT 划分树的基础上加入了 4 种基于二叉树[14]和三叉树[15]的划分方式，如图 3（b）、图 3（c）、图 3（d）、图 3（e）所示；相应地，划分树也扩展为两级：首先，CTU 作为第一级划分树的根节点，第一级划分树上的节点仅可使用 QT 划分方式继续划分，此与 HEVC 相同；然后，第一级划分树叶节点上“生长出”第二级划分树，即第一级划分树叶节点为第二级划分树的根节点，第二级划分树上的节点可使用两种 BT 划分方式和两种 TT 划分方式之一继续划分。

上述划分方式可称为 QT-BT/TT 方式，即一种 QT 级联 BT/TT 的二级划分树结构；与单一 QT 相比，它不但使得划分产生的 CU 形状多样，能够更加灵活地适配图像内容，而且解决了四叉树一分为四导致划分过于细碎、CU 数目较多的弊端，从而提高编码效率约 10%。

2.2　解码端运动矢量推导

目前编码过程中，运动矢量（motion vector，MV）信息耗费较多的编码比特。如果能够减少运动矢量传输的比特消耗，则可以有效提升视频编码效率。因此在传统视频编码技术中引入解码端运动矢量推动（decoder side motion vector derivation，DMVD）技术，可以有效降低 MV 信息的编码开销。在 2010 年 DMVD 概念首次被引入视频编码标准研究中[16]，解码端推导的两个思路原理分别如下。

基于空间相关性的模板匹配 MV 导出方案。在当前块的周围已重建的相邻区域中确定一个 L 形模板（图 4 中灰色区域），并在参考图像中找到与模板相匹配的位置，将其 MV 作为当前块的 MV 或者 MV 预测值[17,18]。图 4 中 W 为 L 形模板的尺寸。

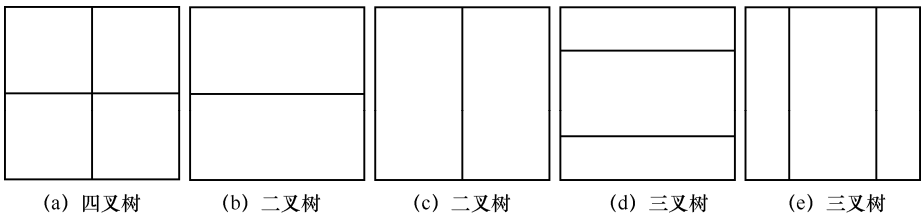


图 3　3 种图像块划分示意

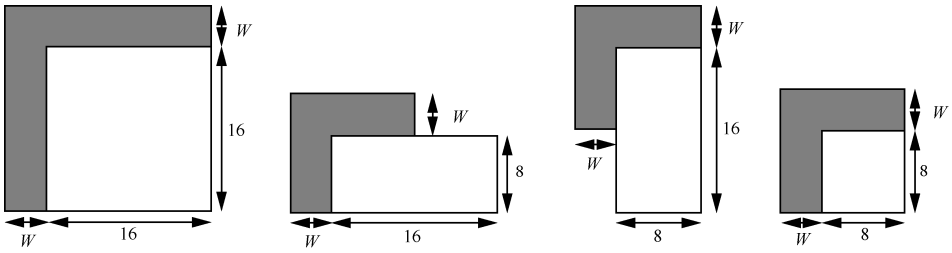


图 4　当前块的模板示意

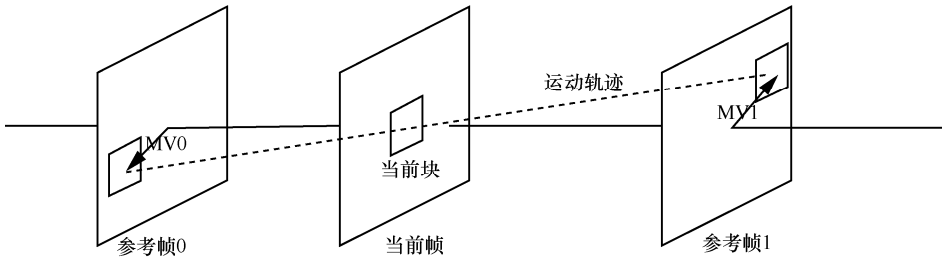
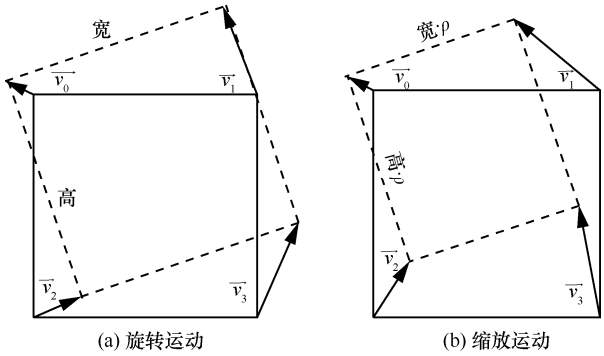


图 5　基于帧率上采样的 MV 导出方法

基于时间相关性的帧率上采样的 MV 导出方案[19]，对当前块基于镜像搜索的方法，在相邻两帧中找到两个匹配误差最小的块的位置（如图 5 所示），从而导出当前块的 MV。通过将这两种理论的有效结合，FVC 的解码端运动矢量导出技术其性能达到 6%[20]。

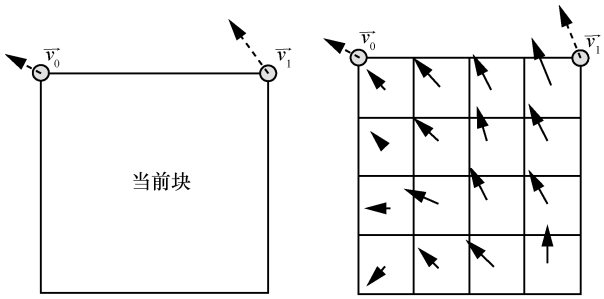
2.3　基于仿射变换模型帧间预测技术

现有视频编码标准中的运动补偿模型均基于传统的平动模型理论，但其并不能有效编码包含复杂运动内容的序列，如旋转（如图 6（a）所示）和缩放（如图 6（b）所示）等，其中$\vec{v}_0$、$\vec{v}_1$、$\vec{v}_2$、$\vec{v}_3$为每个角点位置的运动矢量。仿射运动模型可以有效描述此类复杂运动，从而提高预测的准确度，进而提高编码效率。传统的仿射运动模型使用较为复杂，2015 年基于放射变换运动模型的帧间预测技术被提出并成为 FVC 参考中的基础算法[21-23]。通过使用当前块上方两个角点位置的运动矢量$\vec{v}_0$、$\vec{v}_1$（如图 7（a）所示)，基于 4 点仿射模型推导出当前块中每个像素或者每个子块的运动矢量$\vec{v}_0$、$\vec{v}_1$（如图 7（b）所示）。使用这种方法可以有效地提高含有复杂运动内容的视频序列的编码效率，平均可以提高编码性能约 3%以上，对于某些复杂运动序列，编码效率提升甚至可以达到 30%。



(a) 旋转运动　(b) 缩放运动

图 6　旋转运动和缩放运动示意



(a) 当前块的两个角点的MV　(b) 基于仿射模型导出的MV

图 7　利用仿射运动模型导出 MV

2.4　基于维纳滤波器的自适应环路滤波

维纳滤波器是一种基于最小均方误差的线性滤波器，可以有效地应用于去噪等应用场景。而

视频编码引入的模糊等失真可看作一种叠加在原始视频信号上的噪声信号。因此，维纳滤波也可以用于最小化编码失真视频信号与原始视频信号之间的均方误差，使得滤波后的视频信号尽量接近原始信号。在 2007 年，维纳滤波器作为一种后处理滤波器引入编码标准领域[24]，对解码之后的视频信号进行恢复。2008 年，参考文献[25,26]提出一种块级的自适应环路滤波器，将维纳滤波引入编码环内，并且进入 H.265 的 KTA 模型。在 H.265 标准化过程中得到了持续的改进[27]，但是由于解码端复杂度较高的原因，最终并未进入 H.265 的标准。目前，ALF 由于其优秀的编码性能，又进入 FVC 的编码参考软件 JEM 中，并成为 FVC 的关键技术之一。其编码性能平均约为 6%。

除了以上主要技术，FVC 在帧内预测、变换等相关模块均有进一步优化改进，使得整体性能较前一代视频编码技术得到大幅度提升。但从近几年的研究和标准化工作来看，在传统编码框架基础上的优化难度也越来越大，整个编码系统的复杂度也越来越高，给实际应用带来了一定的难度。面对不断增长的视频数据量和编码技术发展需求，寻找和研究更好的编码理论，寻求算法和性能突破就变得尤为重要。

3 基于机器学习理论的视频编码技术突破

回顾现有视频编码方法的两个问题：（1）其理论假设常常是理想的平稳信号，而自然视频的特性十分复杂，一般不满足理想信号的假设。能否基于自然视频的自身特性设计优化的编码方法？如何“自动”地设计这些编码方法而不是依赖人工调整参数？（2）在现有的编码算法设计中，衡量失真所用的指标是均方误差，等价于优化重建视频的信号保真度（如峰值信噪比（peak signal to noise ratio，PSNR）），但是信号保真与重建视频的视觉质量并非一一对应，有时甚至倒挂。能否在率失真优化时使用视觉失真度量？如何面向视觉质量优化来设计编码方法？

近年来蓬勃发展的深度学习技术为解决上述问题提供了可能的方案。深度学习脱胎于仿生的人工神经网络，借力大数据和高性能计算，在各种机器学习问题特别是涉及语音、自然语言、图像、视频等类型数据的问题中表现出色，在一大批任务中取得了目前最好的结果。将深度学习引入视频编码，针对上述问题有两个潜在的优势。

使用神经网络进行图像视频编码早在20世纪八九十年代就有研究[28]，但其性能一直未能在大量测试数据上稳定地超越 JPEG 和视频编码标准，因此这类方法沉寂了相当长一段时间。从深度学习兴起后，将深度学习用于图像视频编码的研究又逐渐兴起。

Toderici 等人[29,30]提出了基于深度网络特别是递归神经网络（recurrent neural network，RNN）的图像压缩方法。该方法的 3 个主要特点：采用类似自编码器（auto-encoder）的网络结构，并在其中插入二值量化层，直接得到二进制码流；采用多层次残差学习网络结构，第一层对原始块进行逼近，第二层对原始块和第一层输出结果之间的残差进行逼近，第三层对原始块和第一、第二层输出结果之和之间的残差进行逼近；对二进制码流又设计了基于RNN的熵编码方法，进一步提高压缩效率。实验结果显示，该方法在 SSIM 等视觉质量评价准则下显著优于 JPEG。该方法可以视作一种可伸缩编码方案（因其具有多层次结构），也可以视作一种变换量化+熵编码的方案。Dumas 等人[31]对自编码器进行改进，提出了一种随机“赢者全得”自编码器，用于图像压缩获得与 JPEG 相当的性能。Prakash 等人[32]使用 CNN 检测图像的感兴趣区域，然后基于检测结果调整不同区域的量化参数，但是量化参数是手动设定的，没有率失真优化过程。Ballé 等人[33,34]在网络中使用近似熵实现了图像编码率失真优化，其失真度量函数是 MSE 或经过简单变

换的 MSE。具体来说，以参考文献[34]为例，其思路如图 8 所示。

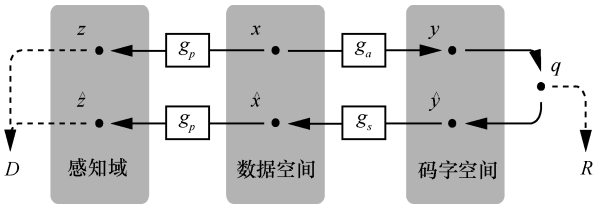


图 8　Ballé 等提出的基于非线性变换的编码框架[34]

图 8 中，x 是输入图像，g_a 是编码所用的正变换，也可以称为分析模块，通过这个分析模块把原始图像 x 映射到码字空间 y，即：

$$y=g_a(x;\Phi) \tag{1}$$

其中，Φ 表示这个分析模块里面的参数，这些参数需要通过训练获得。

之后，码字空间通过量化得到离散的符号，即图 8 中所示的 q。为了重建出输入图像，q 首先经过反量化得到 $\hat{y}$，接着通过一个反变换模块（也可以称为合成模块）g_s，最终得到输入图像的重建图像 $\hat{x}$。值得一提的是，参考文献[34]还实现了笔者前文所提的面向视觉质量优化的编码，即整个编码系统的失真度量并不是在原始的数据空间，而是利用了基于感知的变换 g_p 将图像信号变换到感知域之后再计算失真。这样，整个编码系统可以看成一个面向视觉质量优化的深度学习图像编码系统，在上述系统中发挥重要作用的是分析和合成这两个模块，其网络结构如图 9 所示。

从图 9 不难看出，分析与合成网络均可分为 3 个阶段，每个阶段均包含卷积、下/上采样和 GDN/IGDN 激活 3 种类型操作，执行每种类型操作的网络参数的数量在各操作模块下标出。图 9 中各卷积操作网络模块中列出了卷积滤波器大小（$N\times N$）和输入输出特征数量（输出特征数×输入特征数）。这个分析与合成网络的特色之处在于使用了 GDN 和 IGDN 这两个新型的激活函数。关于更多技术细节，感兴趣读者可以参考文献[33]。

由于 CNN 已被用于去除 JPEG 压缩带来的人工痕迹[35]，也可以类似地用于视频编码中的后处理。韩国 KAIST 的 Park 等[36]提出了一种基于 CNN 的视频编码后处理方案，并报告了显著的码率节省，但是训练 CNN 和使用 CNN 进行编码的是同一组序列，这可能由于过拟合而无法证实 CNN 方案的普适性。中国科学技术大学的 Dai 等人[37]研究了基于 CNN 的帧内编码后处理方案，提出了一种新的网络结构，并通过合理的实验验证了该方案的普适性。该方案取代去块效应（deblocking）滤波和像素自适应偏移（sample adaptive offset, SAO），用于 HEVC 帧内编码获得了平均 4.6%的码率节省。图 10 是 Dai 等人提

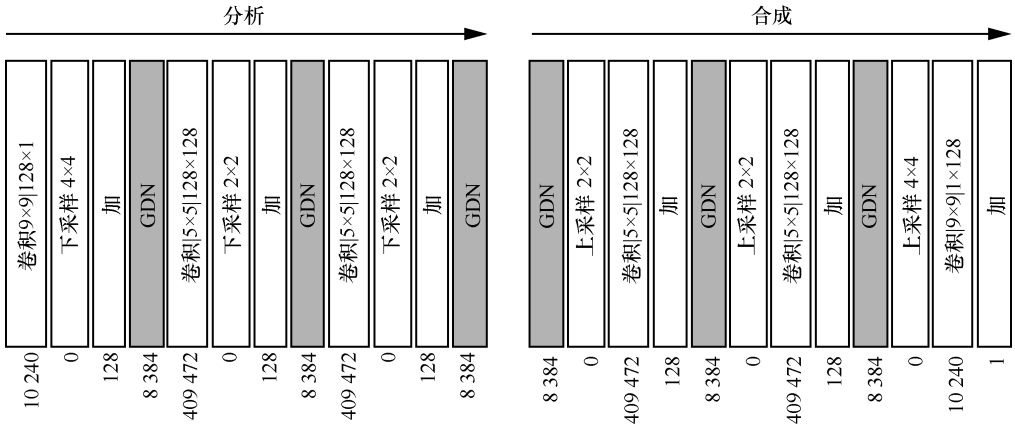


图 9　编码框架中的分析和合成模块结构

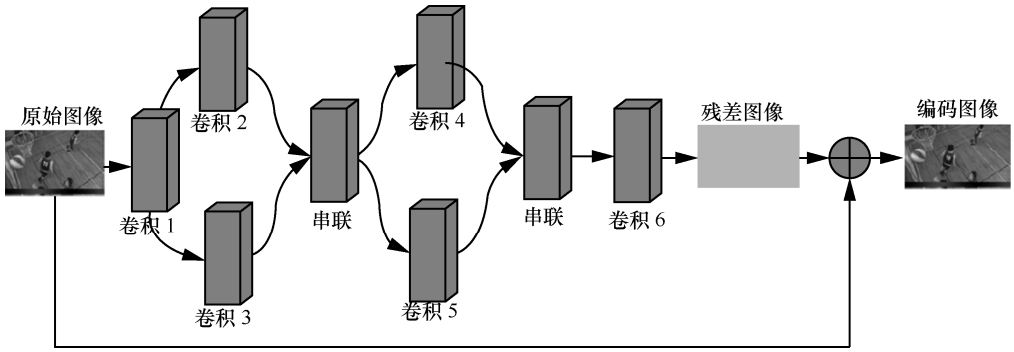


图 10　Dai 等人提出的用于视频帧内后处理的网络结构[37]

出的用于视频帧内后处理的网络结构。

该网络把经过帧内压缩之后的图（未进行任何后处理）作为输入，对应的原始图像作为标签进行训练。网络共有 4 层，第二、三层使用了多尺度的卷积核，并且网络采用了残差训练。注意到原来 HEVC 的后处理操作包含了去块效应滤波和像素自适应偏移这两个步骤，且两个步骤均基于信号处理相关理论人工设计了一些滤波参数，而 Dai 等人把帧内后处理操作直接形式化成了一个端到端的映射问题进行优化，并且取得了更好的性能。

目前国内外都有一些研究者正在进行基于深度学习的其他视频编码方法的研究，例如帧内预测方法[38]、亚像素插值方法、码率控制方法、基于深度学习超分辨率的编码方法、熵编码方法等。总体来说，基于深度学习的图像视频编码研究正处在起步阶段，几乎所有的现有工作都是以信号保真度作为优化目标，即网络训练时使用 MSE 作为损失函数。面向视觉质量优化的深度学习视频编码目前仍是空白，在未来，通过深度学习联合发展视觉质量评价方法和面向视觉质量优化的视频编码方法，可能是比较有前景的研究课题。

4　六自由度视频体数据的编码技术探索

全浸入式视觉体验能够在视场内漫游，要求视频数据提供 6 个自由度的数据信息，即数据需要支持在三维坐标空间中，沿 x、y、z 3 个坐标轴方向移动和围绕这 3 个坐标轴转动的 6 个运动情况下的视觉信息呈现。由于视觉信号采集、处理和呈现的难度，今天比较普及的是三自由度视频，即围绕 x、y、z 3 个坐标轴转动情况下的视觉信息呈现。VR 360°视频就是这种视频数据的典型应用。如上文所述，其原始数据呈现为球面数据形态，相比于平面视频数据，球面视频数据在实际应用中要求更高的分辨率，比如 8K、16K。由于球面数据可以通过简单的映射运算转化为传统的平面视频数据，因此可以基于平面视频编码技术进行处理，取得不错的编码效果，其压缩性能也随着平面视频编码技术的提升而提升。但六自由度视频则具有完全不同的数据形态，一般称之为体视频数据[39]。

体视频数据一般有 3 种方式表达：三维网格表达、点云表达和超多视点表达。三维网格和点云都是计算机图形学的经典技术原理，随着视频体验向六自由度发展，视频信号处理技术和计算机图形技术正逐渐融合，成为未来体视频数据发展的技术理论基础。

利用三维网络（3D mesh）来描述三维空间内的对象，对于一个视觉空间内的三维对象，使用记录对象表面的几何信息、拓扑信息和纹理信息来表达。几何信息用来记录对象表面每个网格顶点的坐标位置，拓扑信息用来记录这些顶点之间

的链接关系。通过几何信息和拓扑信息，可以有效表达三维对象的空间形状，再结合对每个网格内的纹理数据信息，即可以表达一个三维空间内的体视频数据。因此对于三维网格数据的编码，需要全面考虑几何数据、拓扑数据和纹理数据的联合高效编码。三维网格的编码在计算机图形学中已经多有研究，而如何利用三维网格高效表达体视频数据以及其高效编码，还需要进一步的研究。

点云是三维空间中一组离散点数据的组合，每个点数据包含该点的空间位置信息、颜色信息、光学反射信息等[40]。因此一个用点云描述的体视频数据就是一组时间上连续采样的空间点云数据序列。点云数据的编码还是一个全新的问题，其数据表达方式、编码架构等还需要结合应用场景的需求进行深入研究[41]。

超多视点表达是基于传统二维视频数据表达方式最直接的扩展到体视频数据的表达方式，即用二维视频数据记录体视频数据各个视角方向的二维视频信息。这样得到的视频数据就是一组二维视频数据。每一个二维视频数据可以采用现有的视频编码技术进行编码，再借助相邻视角方向视频数据之间的相关性，采用多视角编码技术进一步编码，从而得到更高的编码效率[42]。

由于六自由度体视频技术目前还处在相当早期的阶段，因此关于体视频数据的采集、处理、编码、传输和呈现还需要大量的研究工作。国际运动图像专家组 MPEG 在 2016 年已经开始了面向浸入式六自由度视频的长期标准研究工作，该项目称为 MPEG-I（immersive）。MPEG-I 也提出了从三自由度到六自由度的技术演进路标，希望在 2022 年前后，能够完成业界第一个面向六自由度体视频数据的视频编码技术标准。

5　结束语

尽管视频编码理论、技术和标准已经经过几十年的发展，但随着新视频采集和呈现技术以及新应用的不断涌现，仍然面临强烈的技术发展需求。一方面，针对传统的二维视频数据，基于现有混合编码架构的视频编码技术仍呈现出不断改进的空间，通过对视频图像更加灵活的块划分技术、更加精准的运动模型表达和运动信息编码技术等，视频的编码效率可以得到显著的提升。而基于机器学习的全新编码思路和架构正得到越来越多的探索和研究，有望在不远的将来，部分或全部改变现有的视频编码架构和设计原理，从而给视频编码领域带来更大的突破。另一方面，在视频数据源方面，三自由度球面视频数据和六自由度体视频数据给视频编码技术的研究带来更多的挑战和机会，也给人们带来全新的视频体验和应用。视频编码技术的发展需要新的应用牵引，而新视频应用的繁荣也需要视频编码技术不断的突破和创新。

参考文献：

[1] 施唯佳, 蒋力, 贾立鼎. OTT TV 和 IPTV 的技术比较分析[J]. 电信科学, 2014, 30(5): 15-19,26.
SHI W J, JIANG L, JIA L D. Technique comparative analysis of OTT TV and IPTV[J]. Telecommunications Science, 2014, 30(5): 15-19,26.

[2] 魏峥, 施唯佳, 祝谷乔. 互联网视频中多屏互动技术的应用[J]. 电信科学, 2014, 30(5): 27-32, 39.
WEI Z, SHI W J, ZHU G Q. Multi-screen interaction technologies on internet streaming video[J]. Telecommunications Science, 2014, 30(5): 27-32, 39.

[3] 张敏, 宋杰, 刘晓峰. 电信运营商面对 OTT 的战略选择[J]. 电信科学, 2014, 30(2): 142-146,151.
ZHANG M, SONG J, LIU X F. Strategic selection of telecom operators to counter OTT[J]. Telecommunications Science, 2014, 30(2): 142-146, 151.

[4] MPEG. Presentations of the brainstorming session of the future of video coding standardization: MPEG-w15050[S]. 2014.

[5] MPEG. Steps towards a future video compression standard:

MPEG-w15272[S]. 2015.

[6] MPEG. Requirements for a future video coding standard: MPEG-w15090[S]. 2015.

[7] MPEG. Request for contributions on future video compression technology: MPEG-w15273[S]. 2015.

[8] JVET. Joint call for evidence on video compression with capability beyond HEVC: JVET-F1002[S]. 2017.

[9] MPEG. Joint group on future video coding technology exploration (JVET): MPEG-w15897[S]. 2015.

[10] ITU. Coding tools investigation for next generation video coding: ITU-T SG16-C806[S]. 2015.

[11] JVET. JVET common test conditions and software reference configurations: JVET-B1010[S]. 2016.

[12] JVET. Algorithm description of joint exploration test model 6: JVET-F1001[S]. 2017.

[13] YUAN Y, KIM I K, ZHENG X, et al. Quadtree based nonsquare block structure for inter frame coding in high efficiency video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(12): 1707-1719.

[14] AN J, CHEN Y W, ZHANG K, et al. Block partitioning structure for next generation video coding: COM 16– C966[S]. 2015.

[15] JVET. Multi-type-tree: JVET-D0117[S]. 2016.

[16] YANG H, FU J, LIN S, et al. Description of video coding technology proposal by Huawei Technologies & Hisilicon Technologies[C]//ISO/IEC JTC1/SC29/WG11, JCTVC-A111, April 15-23, 2010, Dresden, Germany.[S.1.:s.n.], 2010.

[17] KAMP S, WIEN M. Description of video coding technology proposal by RWTH Aachen University[C]//JVT on Video Coding of ITU-T VCEG and ISO/IEC MPEG 1st Meeting, JCTVC, JCTVC-A112, April 15-23, 2010, Dresden, Germany.[S.1.:s.n.], 2010.

[18] KAMP S, WIEN M. Decoder-side motion vector derivation for block-based video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(12): 1732-1745.

[19] CHIU Y, XU L, ZHANG W, et al. Description of video coding technology proposal: self derivation of motion estimation and adaptive (Wiener) loop filtering[C]//JCT-VC 1st Meeting, JCTVC-A106, April 15-23, 2010, Dresden, Germany.[S.1.:s.n.], 2010.

[20] CHEN J, CHIEN W J, KARCZEWICZ M, et al. Further improvements to HMKTA-1.0[J]. Doc VECG-AZO7, 2015.

[21] LIN S, CHEN H, ZHANG H, et al. Affine transform prediction for next generation video coding[J]. ITU-T SG16 Doc COM16-C1016, 2015.

[22] CHEN H, LIANG F, LIN S. Affine SKIP and MERGE modes for video coding[C]//2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP), Oct 19-21, 2015, Xiamen, China. New Jersey: IEEE Press, 2015: 1-5.

[23] LI L, LI H, LIU D, et al. An efficient four-parameter affine motion model for video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017.

[24] WITTMANN S, WEDI T. Transmission of post-filter hints for video coding schemes[C]//2007 IEEE International Conference on Image Processing, Sept 16-Oct 19, San Antonio, TX, USA. New Jersey: IEEE Press, 2007: 81-84.

[25] ITU. Adaptive (Wiener) filter for video compression:ITU-T SG16 Contribution C, VCEG-C437[S]. 2008.

[26] ITU. Adaptive loop filter for improving coding efficiency: ITU-T SG16 Contribution C, VCEG-C402[S]. 2008.

[27] 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.

[28] JIANG J. Image compression with neural networks–a survey[J]. Signal Processing: Image Communication, 1999, 14(9): 737-760.

[29] TODERICI G, O'MALLEY S M, HWANG S J, et al. Variable rate image compression with recurrent neural networks[J]. arXiv preprint arXiv:1511.06085, 2015.

[30] TODERICI G, VINCENT D, JOHNSTON N, et al. Full resolution image compression with recurrent neural networks[J]. arXiv preprint arXiv:1608.05148, 2016.

[31] DUMAS T, ROUMY A, GUILLEMOT C. Image compression with stochastic winner-take-all auto-encoder[C]//2017 IEEE International Conference on Acoustics (ICASSP 2017), March 5-9, 2017, New Orleans, USA. New Jersey: IEEE Press, 2017: 1512-1516.

[32] PRAKASH A, MORAN N, GARBER S, et al. Semantic perceptual image compression using deep convolution networks[J]. arXiv preprint arXiv:1612.08712, 2016.

[33] BALLÉ J, LAPARRA V, SIMONCELLI E P. End-to-end optimization of nonlinear transform codes for perceptual quality[J]. arXiv preprint arXiv:1607.05006, 2016.

[34] BALLÉ J, LAPARRA V, SIMONCELLI E P. End-to-end optimized image compression[J]. arXiv preprint arXiv:1611.01704, 2016.

[35] DONG C, DENG Y, CHANGE Loy C, et al. Compression artifacts reduction by a deep convolutional network[C]//2017 IEEE International Conference on Computer Vision (ICCV 2015), Dec 7-13, 2015, Santiago, Chile. New Jersey: IEEE Press, 2017: 576-584.

[36] PARK W S, KIM M. CNN-based in-loop filtering for coding efficiency improvement[C]//2016 IEEE Image, Video, and Multi-

dimensional Signal Processing Workshop (IVMSP), July 11-12, 2016, Bordeaux, France. New Jersey: IEEE Press, 2016: 1-5.

[37] DAI Y, LIU D, WU F. A convolutional neural network approach for post-processing in HEVC intra coding[C]//2017 International Conference on Multimedia Modeling (MMM 2017), January 4-6, 2017, Reykjavik, Iceland. Heidelberg: Springer, 2017: 28-39.

[38] LIU Z, YU X, CHEN S, et al. CNN oriented fast HEVC intra CU mode decision[C]//2016 IEEE International Symposium on Circuits and Systems (ISCAS 2016), May 22-25, 2016, Montreal, Canada. New Jersey: IEEE Press, 2016: 2270-2273.

[39] LAFRUIT G, QUACKENBUSH S, FOESSEL S, et al. Technical report of the joint ad hoc group for digital representations of light/sound fields for immersive media applications[R]. 2016.

[40] TULVAN C, MEKURIA R, LI Z, et al. Use cases for point cloud compression[R]. 2016.

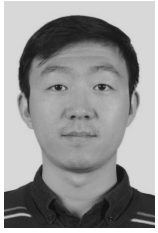
[41] MEKURIA R, LI Z, TULVAN C. Call for proposals for point cloud compression[R]. 2017.

[42] PALOMO C M. Interactive image-based rendering for virtual view synthesis from depth image[D]. Rio de Janeiro: Pontifícia Universidade Católica do Rio de Janeiro, 2009.

[作者简介]

周建同（1980−），男，华为技术有限公司主任工程师，主要研究方向为多媒体应用系统和视频通信。

杨海涛（1983−），男，华为技术有限公司主任工程师，主要研究方向为图像视频处理、压缩和通信。



刘东（1983−），男，中国科学技术大学副教授，主要研究方向为图像视频压缩和多媒体数据挖掘。

马祥（1987−），男，华为技术有限公司工程师，主要研究方向为视频压缩。

王田（1967−），男，华为技术有限公司媒体技术实验室主任，主要研究方向为多媒体通信系统、虚拟/增强现实和计算机视觉。