

基于学习的视频编码技术进展

朱秀昌,唐贵进

(南京邮电大学 江苏省图像处理与图像通信重点实验室,江苏 南京 210003)

摘要: 近来,深度神经网络为视频编码提供了一种新的有前途的解决方案,即基于学习的视频编码技术。文中着重回顾了两类基于学习的视频编码技术,一类是用神经网络替代或协助传统编码框架中某些“功能模块”的压缩,另一类是全神经网络实现的“端到端系统”压缩。分别给出了两类技术中一些近来有代表性的研究成果,展示了它们的优越之处、长足进展及发展潜力。在给出这类技术目前存在问题的同时,也简要展望了今后的研究方向。

关键词: 视频编码;神经网络;深度学习;编码工具;端到端系统

中图分类号: TN919.8 **文献标志码:** A **文章编号:** 1673-5439(2022)02-0001-12

Advances in learning-based video coding technologies

ZHU Xiuchang, TANG Guijin

(Jiangsu Province Key Lab on Image Processing & Image Communication,
Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Deep neural networks have provided a new and promising solution for video coding, namely learning-based video coding techniques. This paper reviews the two types of learning-based video coding technologies: (1) methods that use neural networks to replace or assist the compression of certain functional modules in the traditional coding framework, and (2) the end-to-end system compression implemented by full neural networks. Some recent representative studies on the two types of technologies are summarized, and their advantages, great progresses and development potentials are elaborated. Finally, we present problems of these technologies, and identify the future research directions.

Keywords: video coding; neural network; deep learning; encoding tools; end-to-end system

近年来,在媒体的传播和存储中,视频应用占了大部分^[1],如网络视频、数字电视、视频直播、安全监控、视频通信、云存储等业务。由于视频的分辨率(如4K、8K视频)和保真度(如高动态范围和高比特精度)快速增加,更有效地压缩视频数据的体量而不影响视频的内容就尤为重要。

为此,视频编码(或压缩)技术需要在给定的比特率负担下,采用适当的算法来最小化解码重建失真,或最大化感知质量(Quality of Experience, QoE)。几十年来,广大研究人员专注于视频编码框架和工具的研究和改进,形成一系列非常成熟的视频编码

技术以及相关的视频编码标准,并已为市场提供了广泛的、高性能的视频压缩服务。

1 视频编码

1.1 视频压缩的需求

视频压缩的需求来自两方面的压力,一方面是视频本身的数据量激增,另一方面是视频的应用范围扩大。让我们感受一下高清视频数据的“海量”,一帧未压缩4K(3 840×2 160)视频的数据量约23.7 MB,以帧频60 Hz实时传送时要求速率至少为1.4 Gb/s,这对一般百兆、千兆网络是难以胜任的。

由此可见,随着视频质量的提升,如分辨率从 20 世纪 80 年代的 352×288 的标准视频到如今高分辨率 2K、4K、8K 视频,像素精度从 8 bit 提高到 10 bit、12 bit,帧频从 30 帧/s 提高到 60 帧/s、120 帧/s 甚至更高,更不用说 3D 视频、多视点视频、虚拟现实了。与此同时,视频的新业态、新用法层出不穷,如近年来的短视频、社交视频、直播视频的普及,在自动驾驶、远程监控、云存储、太空航行中的视频传输等。这两方面的原因使得生成的视频数据一直在呈指数方式快速增长。这样,虽然数十年来视频压缩能力和传输带宽也在不断增长,但还是远远赶不上视频数据量的剧增,这必然会对视频编码技术提出越来越高的要求。

1.2 传统视频编码框架

自 20 世纪 70 年代视频信号数字化之后,人们开始用统计信号处理的方法对数字视频进行压缩处理:在不丢失或很少丢失信源信息的情况下,找到“最紧凑表示”,达到对信源数据压缩的目的。

视频是多帧图像的时间序列,其中存在大量的冗余信息。所谓“最紧凑表示”,实际上就是削减视频信息中冗余后的表示。这里的冗余信息大致包含画面的空间冗余,序列的时间冗余,人眼关注的视觉冗余,编码符号的统计冗余等。针对不同的冗余信息,传统编码采用了多项压缩措施,如帧内预测、帧间预测、运动估计、变换、量化、熵编码、环路滤波、预处理和后处理等,或称“编码工具”。

传统的视频编码就是用这些编码工具组成的混合编解码系统进行的。其中主要是指预测编码和变换编码这两类工具的混合。这样的混合编码系统最早出现在 1988 年国际电联 (ITU-T) 颁布的第一个视频编码标准 H.261 中。此后,ITU-T 和国际标准化组织 (ISO/IEC) 合作推出的一系列国际标准,如 H.261/MPEG-1^[2]、H.262/MPEG-2、H.264/AVC、H.265/HEVC^[3] 和最新的 H.266/VVC^[4],中国的视频编码标准 AVS1、AVS2 和 AVS3,国际开放媒体联盟 (Alliance for Open Media, AOM) 的企业标准 AV1、AV2 等也都延用了这种“混合编码”框架。

近 40 年来,视频编码技术尽管取得了非常突出的成绩,但是要想进一步提升压缩性能,传统的方法遇到了不小的困难。为此,研究人员将目光转向人工神经网络 (Artificial Neural Network, ANN) 技术。

1.3 深度神经网络

ANN 技术从 20 世纪 50 年代末问世以来,其发展至今虽几经波折,但总的趋势是性能不断提高,应

用范围不断扩大。将 ANN 技术应用于视频压缩起始于 20 世纪 80 年代,在简单的多层感知机 (Multi-Layer Perceptron, MLP) 网络上进行的,由于效果不佳而此后长时间内进展甚微。自从深度神经网络 (Deep NN, DNN)^[5],尤其是卷积神经网络 (Convolutional NN, CNN)、循环神经网络 (Recurrent NN, RNN) 以及生成对抗网络 (Generative Adversarial Network, GAN) 的陆续出现,对基于学习的视频编码 (Learning-based Video Coding, LVC) 发展起到了很大的推动作用。现在可以看出,LVC 方式具有传统压缩方式不具备的优点,尽管现在尚处于起步阶段,还有许多技术难点需要克服,但是,其优越的压缩性已获得了研究人员的广泛认可。

在 LVC 中最为常用的网络为 CNN 和 RNN。CNN 是在 MLP 基础上发展起来的深度网络。随着网络的层数增加,相邻层之间的密集连接使得网络参数的数量呈平方增加,阻碍了神经网络的计算。为了解决这个问题,CNN 采用了参数共享、稀疏连接和池化等技术,将复杂的神经网络计算简化为类似卷积的运算,使大规模神经网络的训练成为可能。RNN 是一种在时间上传递的神经网络,网络的深度就是时间的长度。一般前向网络的每一层神经元信号只能向下一层传播,样本的处理在时刻上是独立的。而 RNN 的神经元在这个时刻的输出可以直接影响下一时刻的输入,因此 RNN 能够提取时间序列的信息,很适合用来处理视频序列问题。

ANN 理论中的“通用近似定理” (Universal Approximation Theory) 告诉我们,神经网络,尤其是深度神经网络理论上可以实现任意函数的逼近功能,能够处理复杂的非线性问题,因而在视频压缩方面具有超越传统方法的能力。但是,由数据驱动的神经网络必须通过训练优化才能获得特定的函数功能,这就涉及到网络训练中常用的误差反传 (Backward Propagation, BP) 机制和梯度下降 (Gradient Descent, GD) 优化算法。在训练的过程中,BP 技术将多层神经网络中目标函数 (或损失函数) 的误差由输出层向输入层反向传播;由梯度下降法来引导目标函数逐渐趋于极小值。

1.4 基于学习的视频编码

我们知道,视频编码本质上是一个消除非线性信息冗余的过程,只不过在传统编码中将它简化成一系列线性过程来处理。如上所述,“深度网络”可以实现复杂的函数功能,提供更多的非线性建模能力。因此,在目前基于学习的视频编码中,几乎都采

用深度网络,并取得了很好的压缩性能。

基于学习的视频编码方法和传统编码方法的区别在于深度网络是数据驱动的,可以通过无监督学习建模的方法来自动完成数据压缩的任务。而传统方法要靠手工提取信号特征,采用相对固定的方法对视频信号进行处理。2015年至今,LVC已成为一个非常活跃的研究领域。

下文在对传统视频编码方法简单介绍的基础上,着重对基于学习的视频编码的技术分类、实现方式、存在问题和未来展望进行一些回顾,重点分析“模块化编码工具”和“端到端全学习系统”,并介绍这一研究领域中出现的一些新进展。

2 基于学习的视频编码的实现

在基于学习的视频编码技术中,目前大致有两种主要的实现方式。第一种是模块化神经网络视频编码(MODularized Neural Video Coding, MOD-NVC),简称“模块化工具”,它使用基于学习的方法,在传统的混合编码框架中改进编码模块。第二种是端到端神经网络视频编码(End-to-End NVC, E2E-NVC),简称“端到端系统”,它充分利用深度神经网络,以端到端全局优化的学习方式紧凑地表示输入视频。

2.1 模块化工具

MOD-NVC是将基于学习的模块化编码工具集成到传统的视频编码框架中。它既可以作为一个独立的模块替代原模块,例如基于学习的环内滤波;也可以作为原模块的增强工具,或者编码策略的一个部分,例如帧内预测的模式判别部分。这样,传统基于规则的编码工具性能可以通过神经网络数据驱动的学习方式得到进一步的改进,实现更紧凑的内容表示。模块化工具几乎覆盖了传统视频编码框架中的主要功能模块:预处理、帧内/帧间预测、变换和量化、熵编码、后处理、环内滤波、编码控制等。

2.2 端到端系统

E2E-NVC方式可以不拘泥于传统的混合编码框架,常用深度神经网络,通过端到端的基于学习的方法来完成压缩处理。自从2015年Google提出一种使用循环网络的图像压缩通用框架,人们对这一类方法的研究兴趣迅速增加。E2E-NVC方式的帧内编码效果已经达到、甚至超过传统的HEVC的水平,但帧间编码尚有欠缺,正在努力改进之中。3篇国内学者的综述文章较全面回顾了这一方面的相关研究^[6-8]。

3 模块化编码工具

3.1 预处理模块

预处理的内容没有严格的定义,包含的方法较多。除了和常规的去噪、去模糊等对应的基于学习的预处理外,值得关注的还有上/下采样、显著性区域划分和编码模式选择等预处理方法。

(1) 上/下采样

上/下采样是降低视频空间分辨率(像素密度)或时间分辨率(帧频)的一种预处理方法,就是在编码前对原始视频进行下采样,直接降低它的数据量,减轻后续的编码负担;在解码后再进行相应的上采样,还原视频的分辨率。

上/下采样主要有两种实现方式,一种是传统的下采样滤波不动,用深度网络做上采样滤波。如Afonso等^[9]在编码端用传统的低通滤波器进行空间下采样,用SVM决定是否对输入帧做下采样;在解码器端,则用CNN将解码后的视频上采样到原始分辨率。另一种是上采样和下采样滤波都用深度网络来完成,这样具有更多的灵活性。如Jiang等^[10]提出的用两个CNN分别作编码端的下采样和解码端的上采样,中间可以用现有的编解码器。

(2) 显著性区域划分

人类视觉系统(Human Visual System, HVS)有一定的偏好,对视域中某些部分或对象特别关注,或特别感兴趣。根据这一特性,可将图像划分为重点关注的显著性区域和一般的区域。编码时对不同区域的区域进行不同程度的压缩,对显著性区域进行轻度压缩,对一般区域进行重度压缩,从而获得总体较高的压缩率,而不影响QoE。例如,在不同速率的LVC中,Kirmemis等^[11]发现在不同的比特率可能存在不同的最佳感知-失真权衡点,并提出了一种确定最佳权衡点的实际方法。实验结果表明,该方法在显著性检测和感知压缩质量方面均优于现有的感知编码算法。

(3) 编码模式选择

深度学习技术还可应用于视频编码的块划分、编码模式的选择等。例如Kuang等在文献[12]中给出了屏幕内容编码中各个深度的二叉树划分方法,通过卷积的各层对编码单元候选模式的预测来得到分等级的深度特征提取。

3.2 帧内预测模块

传统的帧内预测利用当前编码块左边和上边的两条已编码像素来预测当前块内的像素值。由于存

在局部结构的不同,编码块中像素可以从多达几十、近百个方向中选取最好的预测。例如 VVC 中共设置了 93 个角度划分的预测方向,沿着不同方向的直线对块内像素进行预测。这种方式增加了编码模式数据,且沿直线预测缺乏对编码块内的纹理结构的适应性,限制了编码性能的进一步提升。深度学习网络可以利用周围更多的解码行和列作参考,来实现结构自适应的帧内预测。

自 2017 年以来,出现了多种基于深度网络的帧内预测方法。例如, Wang 等^[13]先用一种多尺度 CNN 生成 HEVC 的初步预测块,再用更多的上下文参考像素来改进,生成一个更准确的预测块。与 HEVC 参考软件 HM16.9 相比,该方法平均可节省 3.4% 的比特率。

最近, Brand 等^[14]使用单一的条件自动编码器(Conditional Auto Encoder, CAE)网络来对亮度和色度分量进行帧内预测。这种方法将 16×16 预测块邻近的参考行/列扩大为 4 像素宽度,以参考更多的内容,能够预测传统模式无法准确预测的结构。这种方法学习到的潜在空间变量本身就是预测函数索引,可取代经典帧内编码中使用的模式索引。该方法还可在亮度和色度编码之间提供跨通道预测,以避免分别发送色度通道的潜在变量。与 VVC 的参考模型 VTM 相比,亮度分量和色度分量的 BD-Rate (Bjontegaard Delta) 增益分别为 1.13% 和 1.21%。

3.3 帧间预测模块

和帧内预测利用空间相关性类似,帧间预测利用的是视频的时间相关性,即用先前重建的帧作参考,进行帧间预测,另外加上运动矢量进行补偿,提高预测精度。此外,分数精度的运动矢量、灵活块划分等都可以提高帧间预测的效率。

在运动矢量估计方面已有较大的进展。如 Yan 等^[15]考虑到精确的运动矢量在帧间预测中的重要性,提出了一种分数像素参考 CNN(Fractional Pixels Reference CNN, FRCNN)来预测分数位像素值。FRCNN 方法与以前的插值或超分辨率方法不同,它不是在高分辨率图像中预测像素值,而是由接近当前编码帧的参考帧生成分数位像素。例如,对于整像素周围三个半像素各训练一个单独的 CNN。FRCNN 在低延迟 P 帧和 B 帧,随机访问(RA)情况下,与 HM16.7 相比,分别实现了平均 3.9%、2.7% 和 1.3% 的比特率节省。使用联合空时 CNN(Spatial and Temporal CNN, STCNN)作帧间预测的 Mao 等^[16]用当前块和两个参考块的空间相邻像素,以及

参考帧和当前帧之间的时间距离作为 CNN 的辅助信息,以此来提高双向帧间预测的精度,在不同的情况下,可获得 2%~5% 的比特率节省。

Yang 等^[17]提出了一种基于循环 AE(Recurrent AE, RAE)网络的视频编码框架。RAE 充分利用了视频帧之间的时间相关性来压缩运动和残差信息,其结构如图 1 所示。图 1 中 $\uparrow 2$ 和 $\downarrow 2$ 分别表示步幅为 2 的上下采样。每个卷积层有 128 个滤波器,当压缩运动信息时,所有卷积核为 3×3 ,当压缩残差信息时为 5×5 。和传统的编码标准相比,这种循环自动编码器扩展了参考帧的范围。在低时延 P 帧 x265 的实验中,该方法的 PSNR 和 SSIM 均优于之前所有基于学习方法。

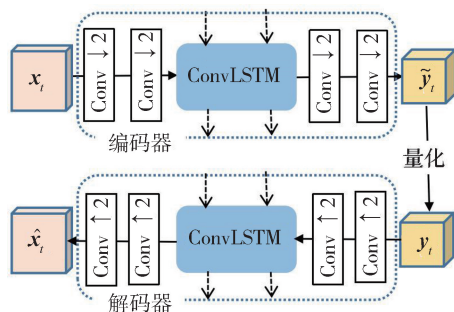


图 1 RAE 网络的结构

3.4 变换模块

变换模块将视频数据或预测残差转换为变换系数,然后进行量化和熵编码。传统视频编码多采用线性变换,不利于消除图像的非线性冗余。由此,人们试图用非线性性能更好的 ANN 来实现变换。在自动编码器的基础上, Liu 等^[18]提出了一种基于 CNN 的方法来实现一种类似于 DCT 的变换。该变换由一个 CNN 和一个全连接层组成,其中由 CNN 来对输入块进行预处理,由全连接层来完成变换。在其实现过程中,全连接层由 DCT 的变换矩阵进行初始化,然后与 CNN 一起进行训练,用联合率失真作为代价函数,其中速率是由量化系数的 L1 范数来估计。这种变换训练后优于固定的 DCT。

Yang 等^[19]提出了一种基于学习的帧内编码中的非线性变换。这种方法将预测的方向性信息合并到残差域中。然后,设计了一个 CNN 模型及相应的反映变换效率的损失函数,利用帧内预测信号来减少残差的方向性,实现了比传统 DCT 更好的去相关性和能量压缩。和 HEVC 参考软件相比,该方法对自然场景视频的 BD-Rate 增益为 1.79%。

3.5 量化模块

量化是在不损失感知质量的情况下删除不敏信息的一种有损压缩工具。传统的量化通常采用计算和存储成本较低的均匀标量量化。基于学习的量化工具也常采用这种量化方式,其中软量化判决性能较硬判决好,但实现复杂。为此,Wang等^[20]在帧内视频编码中构建了一种基于深度学习的系数自适应偏移的硬判决量化模型,可以自动调节硬判决输出,使之趋近软判决的水平,平均实现了2.17%的比特率节约,并有利于硬件的实现。

由于普通量化器的梯度几乎处处为零,不利于神经网络的优化训练,因此现已提出若干近似量化的方法来克服这个困难。其中Tsubota等^[21]在综合比较了现有的均匀量化近似方法后,评估了解码器和熵模型的不同近似组合,并且获得了最佳近似组合,性能优于现有的量化近似方法。实验表明,通过添加噪声的近似比舍入近似更好。

3.6 熵编码模块

熵编码早先使用的是哈夫曼(Huffman)编码,发展到现在,普遍使用算术编码,尤其是高效的上下文自适应二进制算术编码(Context Adaptive Binary Arithmetic Coding, CABAC)。

CABAC包括二值化、上下文建模和二进制算术编码3个步骤。其中前两步都是手工设计的,可能无法准确地估计语法元素的概率,从而限制了CABAC效率的提高。为了解决这一问题, Ma等^[22]提出了一种基于CNN的算术编码(CNN-based AC, CNNAC)方法来对HEVC帧内预测残差的语法元素进行编码。CNNAC使用CNN直接估计语法元素的概率分布,将语法元素的值及其估计的概率分布输入到一个多层算术编码器中来执行熵编码。与HEVC相比,该方法实现了平均4.7%的BD-Rate增益。

3.7 环内滤波模块

为了减轻重建图像的可视感知损伤,在视频编解码中引入了“后处理”模块。后处理主要依靠特别设计的自适应滤波器来增强重建视频质量,提升QoE。后处理通常有两种实现方式,一种是“环外滤波”,即滤波器设置在解码端重建图像的后面,和编码器无关。另一种是“环内滤波”,如HEVC中的去方块滤波(Deblocking Filter, DF)和样本自适应偏移(Sample Adaptive Offset, SAO),除了解码器设置滤波器外,编码器也需要同样的滤波器,并且都位于编码环路内。这样编解码联合处理,改进重建图像

质量。

考虑到内容的纹理特性对环内滤波器的影响, Jia等^[23]提出了像素分类和滤波联合的CNN自适应环内滤波器,其结构如图2所示。对每个LCU(Largest Coding Unit)用AlexNet网络对其中的像素进行分类,然后利用对应类别的CNN对像素进行滤波。训练过程中,对每个像素用所有 N 个滤波网络进行滤波,将该像素归类为滤波性能最好的滤波器所属类别,然后根据分类的像素重新训练 N 个卷积滤波器,并根据滤波性能重新分类,该过程迭代进行,从而同时得到 N 个滤波器和近似最优的像素分类。该方法有效地提高了环路滤波器的内容自适应性,可以平均节省约6.0%的码率。

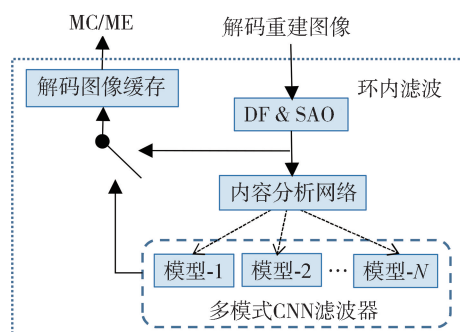


图2 联合分类与滤波的环内滤波器

Zhang等^[24]提出一种基于回归残差CNN(Recursive Residual CNN, RRCNN)的环内滤波器,在减少比特率的同时改进重建帧质量。RRCNN不需要为高、低比特率而训练不同的模型,只用单一模型处理不同的比特率,但需为滤波亮度和色度分别设计不同的网络。RRCNN采用了可切换的机制,允许在通常的滤波器(如DF和SAO)和RRCNN之间以3种方式切换,切换可以在帧或者CTU层面上根据率失真代价来决定。对比HEVC, RRCNN可获得平均8.7%的比特率节约。另一种CNN环内滤波器MFRNet由Ma等^[25]提出,在传统视频编码框架内利用CNN进行环内滤波和后处理滤波,允许在网络的其余层中重用早期层表示。结果表明,环内处理和后处理均有显著的PSNR增益。

3.8 率失真优化模块

视频编码的基本问题就是编码图像 x 的模式 m 在编码速率 $R(x, m)$ 和解码图像失真 $D(x, m)$ 之间通过拉格朗日常数 λ 来进行权衡,可归结为率失真优化(Rate Distortion Optimization, RDO)问题,即最小化目标函数 $J(x, m)$:

$$\min_m J(x, m) = D(x, m) + \lambda R(x, m)$$

上面提到的编码工具的作用都是在保持一定重建图像质量的前提下,尽可能降低比特率。下面列举两个针对不同问题的优化工具。

在 HEVC 编码分区判别方面,针对传统视频编码中率失真优化模型,Xu 等^[26]通过 CNN 和 LSTM 预测整个 CTU 分区结构,替换了传统的模式决策,以确定模式划分是否应该提前终止。在基于感知的视频编码方面,Zhu 等^[27]在 HEVC 中采用率失真优化,按照时空关注图选择恰当的量化参数。时空关注图是根据编码块的运动矢量,由基于 CNN 的空间关注和时域关注预测计算得到。

4 端到端全学习系统

传统的混合编码框架中,编码工具都基本是在特定的编解码结构下独自进行 RD 优化操作。目前,很多基于学习的编码工具的优化方式也基本如此,虽然可以提升传统视频编码算法的性能,但没有建立一个端到端的整体优化的编码框架。2016 年,Toderici 等^[28]提出了首个 E2E 图像编码方法,2019 年 Lu 等^[29]提出了首个 E2E 视频编码方法,向人们展示了 E2E 图像/视频编码技术的可能性。随后出现的 E2E 方案大多数仍然遵循传统的混合编码定义,采用有监督学习方法,用不同的算法来有效地表示空域纹理、时域运动和预测残差^[30-31]。

4.1 几项关键技术

(1) 单模型变速率压缩

近年来,卷积 AE 常用于视频编码,通过设置不同的 λ 参数来调整不同的比特率和 RD 折中。这样,不同的比特率可能需要不同的网络模型,这使得硬件实现颇具挑战性。为此,Choi 等^[32]提出了条件卷积 AE,使用单个网络模型就可实现可变速率压缩,并且没有明显的编码效率损失。

(2) 非线性激活和量化

为了生成更紧凑的特征表示,Ballé 等^[33]建议取代传统的 ReLU 等非线性激活,使用广义分裂归一化(Generalized Divisive Normalization, GDN)的方法,理论上证明它与人类视觉感知更加一致。其随后在文献[34]中的一项研究显示,GDN 在压缩任务中优于其他非线性激活函数,如 ReLU、LeakReLU 和 tanh 等。

为了误差的反向传播,量化操作必须在 E2E 学习框架中可导。现已开发了多种方法来近似连续分布的微分,如添加均匀噪声^[33]、随机舍入^[28]和软到硬的向量量化^[35]等。

(3) 运动表示

在基于学习的视频编码中,Lu 等^[29]引入了运动表示的光流,取得了与 HEVC 相当的性能,但是编码效率在低比特率情况下受了较大的损失。Chen 等^[36]扩展了他们的非局部注意力优化图像压缩(NLAIC)方法,用于帧内编码和残差编码,并将二阶流-to-流预测用于更紧凑的运动表示,在不同内容和比特率上显示出一致的 RD 增益。Rippel 等在文献[30]中进行了另一项探索,使用复合特征对运动流和残差进行联合编码,嵌入到聚合的多帧信息中,高效地生成运动流和残差的编码。

针对时间信息在视频编码中的有效表示,Liu 等^[31]提出一种利用一阶光流和二阶光流来预测时间相关性的方法。采用一种单级无监督学习方法将光流封装为量化的相继帧特征,然后进行上下文自适应熵编码,以去除二阶相关性。在低时延场景下,该方法压缩性能优于 HEVC 和当前基于学习的方法。

(4) 注意力机制编码

Li 等^[37]开发了一种基于 CNN 的内容加权图像压缩系统,利用一个分离的三层 CNN 生成了一个基于空间复杂度的重要性图,以此进行自适应比特分配,显著改进了重建图像的主观质量。在非局部注意优化方面,Chen 等^[36]提出了一种基于变分 AE 的端到端编码结构,用非局部操作来捕获潜在特征和超先验相关性,用隐式注意机制为显著图像区域分配更多的比特。以 PSNR 和 MS-SSIM 衡量,该模型优于现有的同类编码方法。

(5) 概率模型

概率分布估计在视频数据压缩中起着至关重要的作用。假设特征元素为高斯分布,Ballé 等^[38]利用超先验(hyperprior)估计了高斯尺度模型(Gaussian Scale Mixture, GSM)的潜在特征的参数。后来,Hu 等^[39]使用分层超先验(从粗到细)来改进多尺度表示中的熵模型。

(6) 损失函数

基于 SSIM 的损失函数与视觉感知比较接近,其应用可以提高重建质量,特别是在低比特率下。此外,对于感知优化编码,将对抗损失^[40]或 VGG 损失^[41]度量的感知损失嵌入到学习中,可以产生较好的视觉效果。

4.2 基于 AE 的编码系统

自编码器(AE)由编码器和解码器两部分组成,不需要人工干预图像特征,具有数据压缩、重建的功

能,非常适合用于图像/视频编码。信号输入到 AE 后,在编码端通过减少隐层神经元数目实现数据的压缩,在解码端则增加神经元数目来重构输入信号。但是,基本 AE 没有对编码进行优化,其压缩能力有限,重建图像质量不佳,它常常要和其他网络技术共同应用,才有可能获得良好的编码效果。

一种改进的方法就是在编码之前做数据变换,如图 3 所示^[42]。原始图像 x 由转换函数 g_a 转换成 $y = g_a(x)$, 然后对 y 做量化和编码后为 $\hat{y}$; 解码后 $\hat{y}$ 由非线性变换 g_s 转换成 $\hat{x} = g_s(\hat{y})$ 。为了兼顾速度和质量,可以最小化联合率失真函数得到。重建图像失真 D 可由感知变换 $h_p(\cdot)$ 的结果 z 和 $\hat{z}$ 由计算得到,码率 R 则可根据量化后得到的代码计算得到。

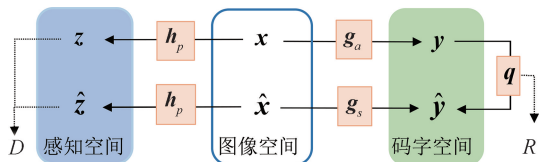


图 3 编码前的数据变换

另一种改进的基于 AE 的端到端方案由 Golinski 等^[43]提出,在解码端使用反馈回归模块将提取的历史潜在变量信息反馈回编码端,且显式地使用运动估计模块辅助运动信息的提取和压缩,该方案在 MS-SSIM 上的压缩性能超过了 H.265。和文献[29]类似,Lu 等在文献[44]中提出了一个基于经典编码框架的端到端视频压缩系统,如图 4 所示。

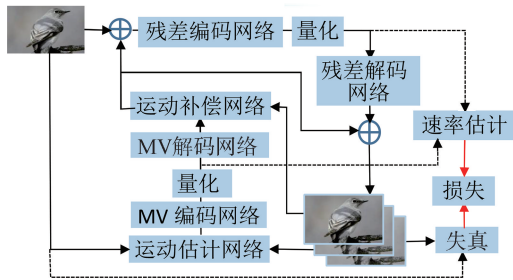


图 4 E2E 视频压缩框架

该系统采用从光流网络中学习到的像素级运动信息,并通过 AE 网络进一步压缩保存,用两个 AE 类神经网络对相应的运动和残差信息进行压缩。所有模块都通过一个损失函数进行联合优化。该方法在 PSNR 指标下优于 H.264,在 MS-SSIM 指标下与 H.265 相当。

4.3 基于 CNN 和 RNN 的编码系统

(1) 基于 CNN 的 E2E 系统

针对 CNN 不能满足梯度下降优化方法要求

E2E 率失真函数整体可微的问题,Ballé 等^[33]在 2016 年首次提出一种通过施加均匀分布噪声模拟量化过程的率失真函数松弛方法,使得整个编码框架是可微的。如图 5 所示,E2E 框架包括由卷积网络组成的分析模块(编码)与生成模块(解码),分别负责从图像到紧凑表示的映射及其重建的逆过程,GDN 为其中的激活函数,取得了与 JPEG 2000 相当的编码性能。

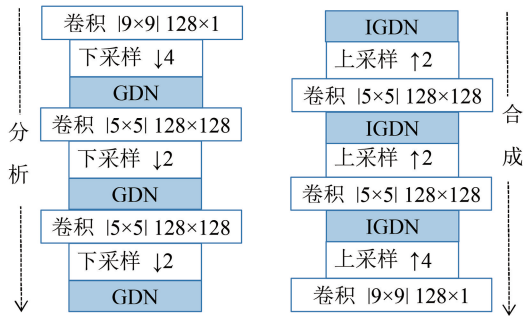


图 5 分析/合成模块

(2) 基于 RNN 的 E2E 系统

不同于 CNN,RNN 的神经元不仅与下一层的神经元相连,同时还与自身相连,对信号和梯度传导具有一定“记忆”功能。2018 年,Johnston 等^[45]提出将隐藏初始状态引入 RNN,使用 SSIM 加权的 loss 函数,并用空间自适应比特率来进一步改进基于 RNN 的算法。这个算法在 Kodak 数据集上,MS-SSIM 的指标要比 BPG 好。

4.4 基于 GAN 的编码系统

2014 年由 Goodfellow 等^[46]提出的生成对抗网络(Generative Adversarial Network, GAN)可以生成和输入图像类似的高质量图像,尤其是高主观质量图像。因此,有可能在 GAN 的基础上实现高压压缩率的编解码系统。但它和典型的 GAN 生成器不太一样,作为压缩系统,要求 GAN 生成的图像要尽量和原始图像一致。这就意味着 GAN 的图像生成需受到一定的控制。

Rippel 等^[47]首次将 GAN 引入到图像压缩中,其网络结构如图 6 所示,特征提取模块与生成模块分别作为编码器和解码器。在率失真目标函数中引入对抗损失函数进行 E2E 训练。实验结果显示该方案在率失真性能和复杂度两方面都取得很好效果:低码率下的主观重建质量超过 HEVC 帧内编码。

Kim 等^[48]提出了一个使用条件 GAN(CGAN)的视频编码框架。它有两个编解码器:一个是标准

的视频编解码器,用于压缩关键帧,另一个是生成低层软边缘映射(Low-Level Soft Edge Maps),用于压缩其他帧。对于解码,使用一个标准的视频解码器解码关键帧,提取边缘,以此作为条件,用关键帧训练 CGAN 解码其他帧。用该方案训练一个生成解码器,只需要从单个视频中获取少量的关键帧和边缘信息,而不需要任何插值。视频压缩实验表明,基于 CGAN 的压缩方法性能良好,尤其可以在非常低的比特率下实现高质量的重建视频。

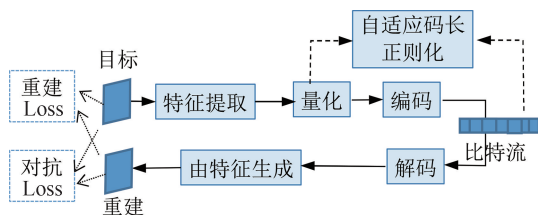


图6 基于GAN的E2E编码框图

4.5 基于非线性变换的编码系统

结合参数化非线性变换^[49]的隐变量概率分布建模,是促使 LVC 方法的性能超过传统方法的一个重要因素。最近,Ballé 等^[50]的文献是第一篇比较全面分析非线性变换编码的隐变量 RD 优化的论文,将引导视频编码从线性处理走向非线性处理。

至此,我们回顾了 E2E 图像/视频编码的多个方面,欣喜地看到了这类技术的迅速发展。尽管 E2E-NVC 刚刚起步,但其快速增长的编码效率预示着可能有良好的前景,随着 E2E 系统性能的不断改进而逐步进入实用。

5 压缩性能评价

视频编解码系统的性能主要取决于这三方面的因素:重建图像的失真,压缩后的码率和编解码系统的复杂度。理想的目标是以最小的复杂度、最低的码率获得最小失真的解压图像。当然,实际上三者同时达到最小是不可能的,因此,视频编码问题是一个根据实际条件和应用需求进行多方博弈的结果。除了这3个因素外,还有一些需要考虑的指标,如鲁棒性、灵活性等。目前基于学习的视频编码处于初始阶段,有些问题往往难以顾及,主要考虑的是失真 D 和速率 R 之间的最优权衡。

对视频压缩系统的评价,大家最为关注的就是对重建视频(图像)质量的评价。压缩视频的质量评价指标和用什么方法来实现压缩没有什么关系,即,不管是传统的编码还是基于学习的编码,其评价

方法是一样的。视频的质量评价包括主观评价和客观评价两种方法,它们的最基本思想都是对原始视频和重建视频进行比较。主观方式比较权威,如 MOS 分评价等,但费事耗时,一致性较差;客观方式简单易行,但有时会 and 人的主观感受有出入。

对 LVC 而言,在目前的研究、探索阶段,最方便的是以下两类可计算的客观评价方法:一类是峰值信噪比(Peak Signal-to-Noise Ratio, PSNR),它考虑的是原始图像和重建图像对应像素的均方误差。另一类是结构相似度(Structural SIMilarity, SSIM)评价方法,它不只简单地考虑对应像素的误差,还考虑人类视觉特性,从亮度、对比度和结构这三方面的相似性进行比较。SSIM 的值在 0 到 1 之间,且越接近 1,重建图像的质量越好。

6 研究与展望

基于学习的视频编码具有以下三个方面的优势:首先,它是一种数据驱动方式,不需要人工制定编码参数,具有强大的学习能力和非线性信号处理的能力。其次,对于 E2E 系统,可以整体优化,到达更高的压缩比和更好的解码质量;对于模块工具方式,易于和传统的编码框架融合,提升局部模块性能,从而提高整体编码性能。最后,有可能将感兴趣区间、语义分析和压缩处理结合起来获得更好的压缩效果。

神经网络的使用,在改进视频编码性能的同时也带来了不少问题,当前需要解决的主要是编码系统自身的问题,标准化问题和产业化问题。

6.1 训练和数据问题

ANN 的理论基础和工作过程尚不完全清楚,不少情况下,近乎“黑盒”操作,使人们对于编码系统的控制和改进受到很大的限制。此外,基于学习的方式常常需要大量的训练数据以及繁杂的训练工作才可能完成,使得训练数据的来源和编码系统的稳定性都成问题。另外,神经网络层数的加深,大大增加了编解码系统的实现复杂度(处理能力、存储容量等),尤其是硬件实现的复杂度,影响了处理的实时性,增加了投入实际应用的困难。

6.2 标准化趋势

和经典视频编码一样,基于学习的视频编码要走向实用,国际标准化是不可或缺的一项重要工作。近年来,多个国际标准化组织已经启动这方面技术的标准化工作。

ISO/IEC MPEG 和 ITU-T VCEG 的联合视频专

家组(Joint Video Experts Team, JVET)对基于学习的视频编码技术非常重视。2017年发布了HEVC后续研究的提案征集(Call for Proposals, CFP)后,2018年就收到了10多项基于学习的编码工具^[51]提案。2019年, JVET的一个相关组(Ad-Hoc Group 9, AHG9)研究基于神经网络的编码工具的压缩性能和复杂度,报告了完整的评估测度方法和测评结果^[52]。2020年,随着VVC的完成, JVET建立了另一个相关组,对基于学习的编码工具和端到端系统的视频编码继续调研,挖掘其性能改进潜力。在JVET2021的输入文档^[53]中可以看到多个有关基于学习的视频编码提案。

IEEE数据压缩标准委员会的未来视频编码研究组(Future Video Coding Study Group, FVC SG)2021年12月正式发布了提案征集,成为对基于学习的视频编码技术的标准化工作的开端。

近年来,我国在AVS3标准的制定的同时,基于学习的视频编码也在研究之中,已提出基于CNN的环内滤波、帧内/帧间预测等工具,并取得了附加的编码增益。

国际AOM在AV1之后探讨下一代AV2的视频编码工具期间,若干基于学习的工具已经提出。

6.3 为产业化服务

(1) 为机器视觉服务

随着人工智能的发展,需要为机器/计算机提供大量的视频,以及为此服务的视频压缩技术。和普通的消费视频压缩大不相同,它往往要求能够被机器接收、分析或解释,具有检测、识别、分类、跟踪、理解等功能。基于学习的视频编码方法比传统方法更能适应这些新的需求。例如,有可能需要打破视频的常规时空特征而自动地进行压缩。再如,有些任务有望能够在压缩域里完成,无需比特解码和像素重建。为此,2019年ISO/IEC MPEG成立了“用于机器的视频编码”(Video Coding for Machines, VCM)的工作组,研究服务于机器视觉的任务和标准,包括目标检测、目标跟踪、实例分割、姿态估计等,探索可同时用于人类感知和机器智能的视频压缩方案。2020年发布了VCM评估框架草案。

(2) 为产业实践服务

基于深度学习的视频编码研究,这种跨学科之间的融合带来多方面的技术改善。并且已经开始向实际的产业应用进行了尝试,展现了基于学习的视频编码带来的一流效率和质量。

6.4 待解决的技术问题

(1) 模型生成

对千变万化的视频内容和各种不同的应用目标,要生成网络模型是非常困难的。目前,大多数基于学习的视频压缩技术使用的是有监督学习,常常需要大量的已标注视频数据。为此,需开发大规模的数据集,如ImageNet等。还可以应用一些新技术来缓解模型生成中训练样本不足的困境,例如少镜头学习^[54]、无监督学习等。

(2) 复杂性

视频信号的空间维度高,持续时间长,实时性要求高,形成压缩处理中难以承受的复杂性。例如,常用的视频编解码器仅需要几十kB的片上存储,而大多数学习算法需要若干MB、甚至TB的存储空间。另一方面,尽管推断可能很快,但训练可能要几小时、几天,甚至几周才能收敛成为可靠的模型。这些问题已成为市场采纳学习工具的严重障碍,特别是对于能源效率敏感的移动平台更是如此。一种有效的办法就是设计专门的硬件来加速训练^[55],例如,神经网络处理单元(NPU)已经引起了大家的注意,深度学习算法有望在装备NPU的器件上大规模地开发。

(3) QoE测度

视频QoE测度和HVS有很好的相关,它不仅适合于质量评估,还适合用作神经网络视频压缩中的损失控制。在QoE质量评价中,已开发出若干新的方法,如仅可见失真(JND)^[56]和VMAF^[57]等,其中有一些已经应用于压缩算法和产品的评估中。另外,现在的基于神经网络的视频编码能够自适应地优化一个预先定义好的损失函数,例如MSE、SSIM、对抗损失^[41]以及VGG基于特征的语义损失等。然而,这些损失函数没有一个呈现出明显的优势。因此,对基于学习的视频编码,需要研究建立一个统一的、HVS驱动的QoE测度。

(4) 率失真优化

率失真理论在目前基于神经网络压缩的任务中尚未得到很好的研究。如何建立一个适合不同神经网络、不同压缩目标的全局率失真优化方法是一个值得研究解决的重要问题。

7 结束语

神经网络尤其是近十年来深度神经网络的发展,为传统的视频编码技术带来了新的契机。本文上述的种种基于学习的视频编码技术的研究和探

索,就是这一契机所带来的初步成果。目前,基于学习的视频编码技术主要有两类:一类简称“模块化工具”,用神经网络取代或协助传统编码框架中的某些功能模块,获得局部的改进,从而得到系统性能的提高;另一类简称“端到端系统”,用神经网络实现视频编码所有功能,形成全学习、优化的编码系统,获得系统性能的提高。按理说,后者的性能应该更好,但目前还不是如此。因为前者起步较早,实现相对容易,实验结果的性能指标相对较好。但是,我们还是相信,完整的端到端网络模型具有更大的潜力,更能发挥神经网络的长处,提供更大的性能改善,同时衍生出更多的功能。

对于基于学习的视频压缩技术,必须认识到,具有一定规模的研究充其量不到 10 年时间。因此,还有太多的困难需要去克服。困惑人们的有神经网络本身的理论和实践问题,有视频信号本身的统计特性、感知特性和表示特性等问题。这种困惑,犹如 20 世纪八、九十年代的人们困惑于视频编码技术能否够到达今天的水平。但是,由于 LVC 具有良好的压缩性能,再加上巨大的市场需求,有信心期待,基于学习的视频编码技术一定会逐步成熟壮大。

参考文献:

- [1] Cisco. Cisco visual networking index: forecast and trends 2017–2022 [R]. 2018.
- [2] ITU, ISO/IEC. Video codec for audiovisual services at p×64 kbit/s; ITU-T Rec. H.261 [S]. 1993.
- [3] ITU, ISO/IEC. High efficiency video coding; ITU-T Rec. H.265 [S]. 2013.
- [4] ITU, ISO/IEC. Versatile video coding; ITU-T Rec. H.266 [S]. 2020.
- [5] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. *Nature*, 2015, 521(7553): 436–444.
- [6] LIU D, LI Y, LIN J P, et al. Deep learning-based video coding [J]. *ACM Computing Surveys*, 2021, 53(1): 1–35.
- [7] MA S W, ZHANG X F, JIA C M, et al. Image and video compression with neural networks: a review [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(6): 1683–1698.
- [8] DING D D, MA Z, CHEN D, et al. Advances in video compression system using deep neural network: a review and case studies [J]. *Proceedings of the IEEE*, 2021, 109(9): 1494–1520.
- [9] AFONSO M, ZHANG F, BULL D R. Video compression based on spatio-temporal resolution adaptation [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019, 29(1): 275–280.
- [10] JIANG F, TAO W, LIU S H, et al. An end-to-end compression framework based on convolutional neural networks [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018, 28(10): 3007–3018.
- [11] KIRMEMIS O, TEKALP A M. A practical approach for rate-distortion-perception analysis in learned image compression [C] // *Picture Coding Symposium (PCS)*. 2021: 1–5.
- [12] KUANG W, CHAN Y L, TSANG S H, et al. DeepSCC: deep learning-based fast prediction network for screen content coding [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(7): 1917–1932.
- [13] WANG Y, FAN X P, LIU S H, et al. Multi-scale convolutional neural network-based intra prediction for video coding [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(7): 1803–1815.
- [14] BRAND F, SEILER J, KAUP A. Intra-frame coding using a conditional autoencoder [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2020, 15(2): 354–365.
- [15] YAN N, LIU D, LI H Q, et al. Convolutional neural network-based fractional-pixel motion compensation [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2019, 29(3): 840–853.
- [16] MAO J, YU L. Convolutional neural network based bi-prediction utilizing spatial and temporal information in video coding [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(7): 1856–1870.
- [17] YANG R, MENTZER F, VAN GOOL L, et al. Learning for video compression with recurrent auto-encoder and recurrent probability model [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2021, 15(2): 388–401.
- [18] LIU D, MA H, XIONG Z, et al. CNN-based DCT-like transform for image compression [C] // *International Conference on Multi Media Modeling (MMM)*. 2018: 61–72.
- [19] YANG K, LIU D, WU F. Deep learning-based nonlinear transform for HEVC intra coding [C] // *IEEE International Conference on Visual Communications and Image Processing (VCIP)*. 2020: 387–390.
- [20] WANG H K, YU S J, ZHANG Y, et al. Hard-decision quantization algorithm based on deep learning in intra video coding [C] // *Data Compression Conference (DCC)*. 2019: 607.
- [21] TSUBOTA K, AIZAWA K. Comprehensive comparisons

- of uniform quantizers for deep image compression[C] // IEEE International Conference on Image Processing (ICIP). 2021: 2089–2093.
- [22] MA C Y, LIU D, PENG X L, et al. Convolutional neural network-based arithmetic coding for HEVC intra-predicted residues[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(7): 1901–1916.
- [23] JIA C M, WANG S Q, ZHANG X F, et al. Content-aware convolutional neural network for in-loop filtering in high efficiency video coding[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2019, 28: 3343–3356.
- [24] ZHANG S F, FAN Z H, LING N, et al. Recursive residual convolutional neural network-based in-loop filtering for intra frames[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(7): 1888–1900.
- [25] MA D, ZHANG F, BULL D R. MFRNet: a new CNN architecture for post-processing and in-loop filtering[J]. IEEE Journal of Selected Topics in Signal Processing, 2021, 15(2): 378–387.
- [26] XU M, LI T Y, WANG Z L, et al. Reducing complexity of HEVC: a deep learning approach[J]. IEEE Transactions on Image Processing, 2018, 27(10): 5044–5059.
- [27] ZHU S P, LIU C, XU Z Y. High-definition video compression system based on perception guidance of salient information of a convolutional neural network and HEVC compression domain[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(7): 1946–1959.
- [28] TODERICI G, O'MALLEY S M, HWANG S J, et al. Variable rate image compression with recurrent neural networks[EB/OL]. [2021–11–10]. <https://arxiv.org/abs/1511.06085>.
- [29] LU G, OUYANG W L, XU D, et al. DVC: an end-to-end deep video compression framework[C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 10998–11007.
- [30] RIPPEL O, NAIR S, LEW C, et al. Learned video compression[C] // IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 3453–3462.
- [31] LIU H J, SHEN H, HUANG L C, et al. Learned video compression via joint spatial-temporal correlation exploration[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11580–11587.
- [32] CHOI Y, EL-KHAMY M, LEE J. Variable rate deep image compression with a conditional autoencoder[C] // IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 3146–3154.
- [33] BALLÉ J, LAPARRA V, SIMONCELLI E P. End-to-end optimized image compression[EB/OL]. [2021–11–10]. <https://arxiv.org/abs/1611.01704>.
- [34] BALLÉ J. Efficient nonlinear transforms for lossy image compression[C] // Picture Coding Symposium (PCS). 2018: 248–252.
- [35] MENTZER F, AGUSTSSON E, TSCHANNEN M, et al. Conditional probability models for deep image compression[C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 4394–4402.
- [36] CHEN T, LIU H J, MA Z, et al. End-to-end learnt image compression via non-local attention optimization and improved context modeling[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2021, 30: 3179–3191.
- [37] LI M, ZUO W M, GU S H, et al. Learning convolutional networks for content-weighted image compression[C] // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018: 3214–3223.
- [38] BALLÉ J, MINNEN D, SINGH S, et al. Variational image compression with a scale hyperprior[EB/OL]. [2021–11–10]. <https://arxiv.org/abs/1802.01436>.
- [39] HU Y Y, YANG W H, LIU J Y. Coarse-to-fine hyperprior modeling for learned image compression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11013–11020.
- [40] AGUSTSSON E, TSCHANNEN M, MENTZER F, et al. Generative adversarial networks for extreme learned image compression[C] // IEEE/CVF International Conference on Computer Vision (ICCV). 2019: 221–231.
- [41] LIU H J, CHEN T, SHEN Q, et al. Deep image compression via end-to-end learning[EB/OL]. [2021–11–10]. <https://arxiv.org/abs/1806.01496>.
- [42] BALLÉ J, LAPARRA V, SIMONCELLI E P. End-to-end optimization of nonlinear transform codes for perceptual quality[C] // Picture Coding Symposium (PCS). 2016: 1–5.
- [43] GOLIŃSKI A, POURREZA R, YANG Y, et al. Feedback recurrent autoencoder for video compression[C] // Computer Vision (ACCV). 2021.
- [44] LU G, ZHANG X Y, OUYANG W L, et al. An end-to-end learning framework for video compression[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(10): 3292–3308.
- [45] JOHNSTON N, VINCENT D, MINNEN D, et al. Improved lossy image compression with priming and spatially adaptive bit rates for recurrent networks[C] // IEEE/CVF Conference on Computer Vision and Pattern

- Recognition (CVPR). 2018: 4385–4393.
- [46] GOODFELLOW J, ABADIE J P, MIRZA M, et al. Generative adversarial nets [C] // Advances in Neural Information Processing Systems (NIPS). 2014: 2672–2680.
- [47] RIPPEL O, BOURDEV L. Real-time adaptive image compression[C] // International Conference on Machine Learning (ICML). 2017:2922–2930.
- [48] KIM S, PARK J S, BAMPIS C G, et al. Adversarial video compression guided by soft edge detection[C] // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2020: 2193–2197.
- [49] IGNATOV A, TIMOFTE R, CHOU W, et al. AI benchmark: running deep neural networks on android smartphones[C] // Computer Vision(ECCV).2019.
- [50] BALLÉ J, CHOU P A, MINNEN D, et al. Nonlinear transform coding[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 15(2): 339–353.
- [51] LIU D, CHEN Z Z, LIU S, et al. Deep learning-based technology in responses to the joint call for proposals on video compression with capability beyond HEVC [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(5): 1267–1280.
- [52] LI Y, LIU S, KAWAMURA K. Methodology and reporting template for neural network coding tool testing: JVET-M1006[R]. 2019.
- [53] ALSHINA E, LIU S, SEGALL A, et al. Neural network-based video coding: JVET-X0011[R].2021.
- [54] WANG Y Q, YAO Q M, KWOK J T, et al. Generalizing from a few examples [J]. ACM Computing Surveys, 2021, 53(3): 1–34.
- [55] HENNESSY J L, PATTERSON D A. A new golden age for computer architecture [J]. Communications of the ACM, 2019, 62(2): 48–60.
- [56] YUAN D, ZHAO T S, XU Y W, et al. Visual JND: a perceptual measurement in video coding [J]. IEEE Access, 2019, 7: 29014–29022.
- [57] Netflix, Inc. VMAF: perceptual video quality assessment based on multi-method fusion[EB/OL].[2022-01-20]. <https://github.com/Netflix/vmaf>.

(责任编辑:李小溪)