

基于深度学习的视频编码发展现状与未来展望

张新峰

中国科学院大学 北京 100049

摘要 近年来,深度学习技术不仅在人工智能领域取得了巨大成功,也为视频编码领域带来了新的发展机遇。文章从两个方面介绍了深度学习技术在视频编码领域的发展现状,即传统编码框架下深度学习视频编码工具和以深度学习模型为基础的视频编码新框架,并对相关代表性工作进行了详细介绍和性能分析。最后,对深度学习视频编码技术面临的挑战和未来发展方向做了分析和展望。

关键词 视频编码;人工智能;深度神经网络;视频预测

引言

视频数据量的爆炸式增长使其已经成为网络大数据的主体,思科公司预测,到2022年全球网络视频数据量将占全部网络数据流量的82%^[1]。视频编码是大数据时代视频数据高效存储、传输和应用的重要技术保障,是推动数字多媒体产业持续发展的关键动力。随着网络带宽的增加和用户体验追求的不断提升,超高清视频、虚拟现实视频、全景视频以及智能化监控视频等的快速发展,正在引领多媒体产业的下一轮颠覆性变革,也给视频编码技术提出了更大的挑战。

从1984年CCITT颁布第一个视频编码标准H.120以来^[2],国际标准化组织ITU和MEPG,以及中国音视频编码组织AVS相继制定了一系列视频编码标准,其中以MPEG-2^[3]、H.264/AVC^[4]、HEVC^[5]和AVS-1^[6]为代表在实际中广泛应用,但是这些标准都是以预测/变换为核心的混合视频编码框架为基础,以视频信号时空域的先验统计模型为出发点设计和优化视频编码工具来消除视频的时空冗余、统计冗余和视觉冗余等。传统编码框架

下手工设计的编码工具性能已经趋于极限,因此,如何打破视频编码技术的性能瓶颈,进一步提高视频数据的压缩效率,是当前亟需解决的关键问题。

近年来,深度学习技术在人工智能和信号处理等众多领域都取得了巨大的成功,也为视频编码技术的发展注入了新的活力。本文从以下两方面分别介绍深度学习技术在视频编码中的代表性工作,并展望了其面临的机遇和挑战。

1 混合视频编码框架下的深度学习编码工具

图1以HEVC视频编码标准为例,展示了以预测/变换为核心的混合视频编码框架的主要技术模块,输入的视频图像序列按照光栅扫描顺序以块为基本编码单元,采用不同大小的编码单元适应图像纹理结构的变化。编码单元经过帧内和帧间预测模块去除视频空域和时域冗余,预测残差经过变换模块改变信号统计特性使其便于后续量化和熵编码压缩。除此之外,现代视频编码框架广泛引入了环路滤波技术^[7],通过去除压缩噪声来提

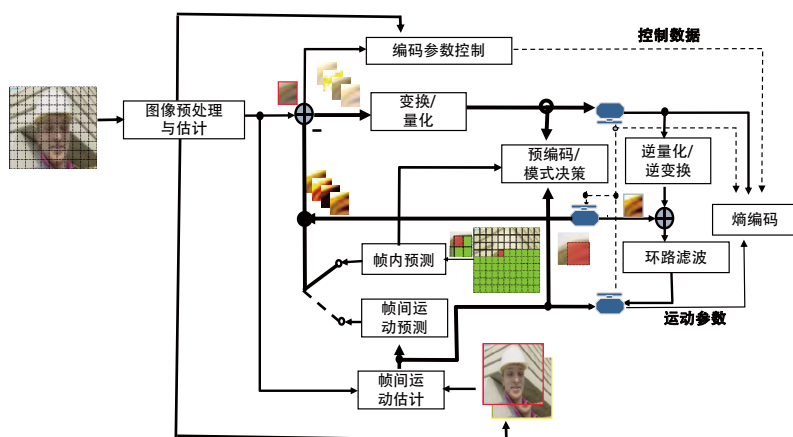


图1 HEVC混合视频编码框架示意图

高解码视频重构质量，同时为后续编码视频图像提供高质量的预测信号，进而提高视频压缩效率。

1.1 帧内预测编码

目前，混合视频编码主要技术模块都是采用手工设计的预测方式、变换核等，例如HEVC中帧内预测包含了DC预测、Planar预测和33种方向预测。如图2所示，这些方向预测主要是基于近邻相似的先验统计特性，按照近似均匀的方向划分，沿着直线对块内像素进行预测。这种方式以增加编码模式信息为代价来提高视频空域预测的自适应性，但是直线型的预测缺乏对编码单元内纹理结构的适应性，限

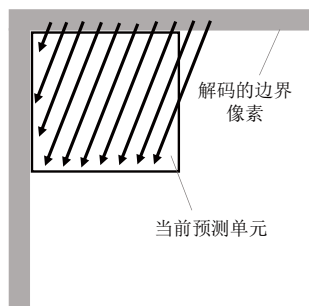


图2 HEVC中帧内方向预测示意图

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行和列的解码像素，通过端到端的训练方式来实现结构自适应的帧内预测。Hu等人提出了一种渐进的空间循环神经网络(Progressive Spatial Recurrent Neural Network, PS-RNN)帧内预测方法^[8]，如图3所示，该方法通过预处理卷积层将更多行和列的参考像素以及当前待编码的预测单元变换到特征空间，以此降低参考像素中的量化噪声，在特征空间通过级联的循环神经网络(Recurrent Neural Network, RNN)单元来生成预测单元的特征，最后通过两层的重构卷积层将预测特征变换到像素空间。该方法通过叠加多个RNN单元来增大感受野，从而使网络结构可以预测任意方向的复

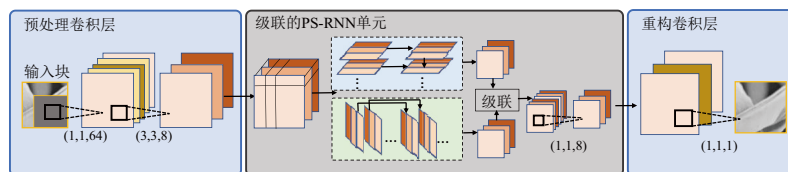


图3 基于深度学习的帧内预测网络结构：渐进的空间循环卷积网络

杂纹理结构，在HEVC视频编码平台上，该方法平均取得了2.65%的码率节省，但是却付出了超过200倍解码时间增加的成本。

除了直接增加周围行列的参考像素外，Meyer等人提出分别训练亮度预测网络和色度预测网络，其中色度预测网络的输入除了包含自身周围的参考像素外，对应位置的重构亮度重构像素也作为参考像素来提高色度预测性能^[9]。Li等人借鉴超分辨率网络结构，在视频编码的帧内预测中对编码单元进行下采样编码，用重构的低分辨率图像块来预测高分辨率图像块，由于深度超分辨率网络较好的重构性能使得该模式为帧内编码在HEVC平台上带来了平均9.0%的码率节省^[10]。

1.2 帧间预测编码

视频帧间预测是指利用已经编码的重构视频帧(称作参考帧)来预测当前编码帧，传统编码方法一般在参考帧上通过运动估计找到最优的预测块，将参考帧信息、运动向量和预测残差一起编码。Zhao等人借鉴帧率提升的思路，在双向预测编码结构下，通过帧率提升网络利用当前编码帧时序紧邻的前后两个参考帧生成当前编码帧的虚拟参

考帧^[11]，如图4所示，虚拟参考帧生成网络的输入分别如下。

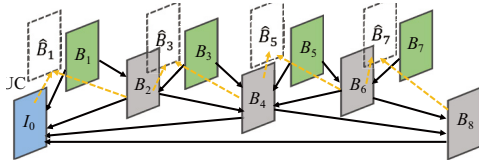


图4 虚拟参考帧生成的依赖关系

$$\begin{cases} \hat{B}_1 = F(\text{Rec}(I_0), \text{Rec}(B_2)) \\ \hat{B}_3 = F(\text{Rec}(B_2), \text{Rec}(B_4)) \\ \hat{B}_5 = F(\text{Rec}(B_4), \text{Rec}(B_6)) \\ \hat{B}_7 = F(\text{Rec}(B_6), \text{Rec}(B_8)) \end{cases} \quad (1)$$

其中 $\text{Rec}(I_0)$ 表示 I_0 对应的解码重构帧，由于帧率提升网络是在原始无压缩的视频帧上训练得到，而解码视频帧中不可避免地混入了压缩噪声，因此Zhao等人又增加了虚拟参考帧增强网络 (VRF Enhancement CNN, VECNN)来降低压缩噪声的负面效应。VECNN网络结构如图5所示，第一个卷积层采用 5×5 的卷积核，其余卷积层均采用 3×3 卷积核。该方法在随机访问编码配置 (Random Access, RA)下取得了平均5.5%的码率节省，但却带来高达700倍的解码时间的增加。

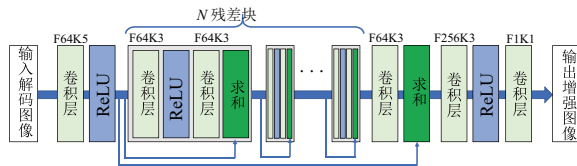


图5 虚拟参考帧增强网络结构示意图

除此之外，Choi等人进一步拓展了上述工作，利用深度学习模型实现双向虚拟参考帧和单向虚拟参考帧的生成^[12]。Jimbo等人利用当前编码块时序上前后帧空间相同位置的图像块来构建当前块的预测，通过卷积神经网络学习4个图像块的变换矩阵，将前后图像块变换后融合为新的预测块^[13]。Zhao等人通过卷积神经网络来学习视频编码双向帧间预测的自适应权重，从而提高了HEVC的编码效率^[14]。

1.3 变换编码

离散余弦变换(DCT)由于具有较好的“能量集中”特性，变换后预测残差信号主要集中在低频部分，便于

后续系数扫描和压缩，因此广泛应用于传统视频编码中。但是这种通用的正交变换核在不同统计特性的预测残差信号上性能也有明显的差异，因此均方误差意义下最优变换KLT的多种近似版本被提出并在视频编码中取得了显著的性能增益^[15]。与最小化均方误差求解最优变换不同，Liu等人利用卷积神经网络以码率和失真联合代价为目标函数，训练编码意义下最优的变化核^[16]。深度卷积神经网络的损失函数为：

$$\{f_E^*, f_Q^*, f_D^*\} = \arg \min_{f_E, f_Q, f_D} (D + \lambda R) \quad (2)$$

其中 λ 是拉格朗日因子， f_E 、 f_Q 和 f_D 分别表示正变换、量化和反变换。和离散余弦变换不同，该深度学习的变换核中引入了ReLU函数，使得正反变换是非线性变换，论文中该方法仅实现了 32×32 的变换核，因此性能上还是明显落后于HEVC。

1.4 熵编码

熵编码是一个无损压缩过程，通过利用上下文信息消除信号间的统计冗余，上下文自适应二进制算术编码器(Context-Adaptive Binary Arithmetic Coding, CABAC)是目前视频编码标准中广泛采用的熵编码工具^[4-5]，其性能提升主要源自于上下文模型构建的准确性。Song等人对HEVC中帧内预测模式信息的熵编码进行改进，将HEVC中手工方式设计的上下文模型替换为卷积神经网络对预测模式概率进行估计，利用当前编码块左边、上边和左上块，以及最有可能预测模式信息作为输入，将概率估计转化为模式预测回归问题进行训练^[17]。Ma等人提出了以稠密连接的卷积神经网络(Densely connected convolutional networks, DenseNet)为基础改进的帧间边信息的CABAC^[18]。该方法利用当前编码块周围的运动向量、参考帧索引等作为输入，主要针对Merge模式的标志、索引、参考帧索引、运动向量差和运动向量预测的索引进行编码，在HEVC平台上取得了平均0.3%的码率节省。

1.5 环路滤波

环路滤波技术目前广泛应用于视频编码标准中，主

要包含去块效应滤波器、样本自适应偏移和自适应环路滤波^[7], 后两者具有较好的内容自适应性和滤波性能, 但是由于需要编码较多的滤波参数信息导致其带来的编码增益有限。Jia等人基于像素统计特性对环路滤波器性能的显著影响^[19], 提出了像素分类和滤波联合的卷积神经网络自适应环路滤波器^[20], 其分类网络和滤波网络结构如图6、图7所示。对于每个最大编码单元(Largest Coding Unit, LCU), 该方法利用一个轻量级的AlexNet^[21]网络对其中的像素进行分类, 然后利用对应类别的卷积神经网络对像素进行滤波。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个卷积滤波器和近似最优的像素分类, 其中分类网络和滤波网络结构。该方法虽然有效地提高了环路滤波器的内容自适应性, 可以平均取得多达6.0%的码率节省, 但是也显著增加了存储和计算的复杂度, 当N=6时, 需要1GB以上的GPU内存, 解码时间增加达到130倍以上。

Zhang等人将残差学习的思想引入到视频编码环路滤波中, 提出了Residual Highway Convolutional Neural Network(RHCNN)^[22]。该深度学习模型通过级联多个highway单元和卷积层组成, 在highway单元中,

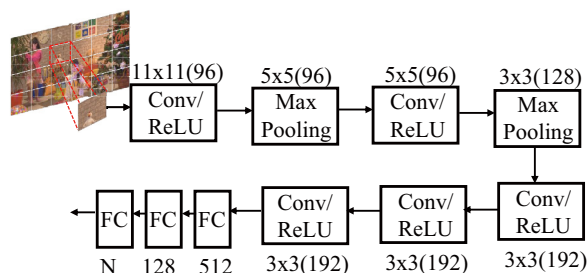


图6 分类网络结构示意图

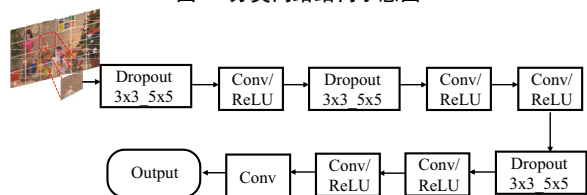


图7 滤波网络结构示意图

信息可以跨越多个网络层, 而且存在从输入直接到输出的跳跃连接, 这不仅可以有助于复原图像结构, 同时也有助于避免训练过程中梯度消失的问题。为了解决深度学习环路滤波中需要根据不同的量化参数训练多组滤波器的问题, Song等人提出了单一深度学习模型的环路滤波器, 将量化参数赋值给和图像相同维度的矩阵, 将其和解码重构图像联合作为滤波器的输入^[23], 从而有效地降低了环路滤波器数量和存储消耗。

2 基于深度学习的视频编码框架

深度学习利用大数据样本通过端到端训练的方式可以实现视频数据的紧致模型化表示, 从而构建端到端的视频压缩框架。目前, 基于深度学习的端到端图像压缩已经取得了显著的性能提升, 甚至超过传统框架下的图像编码或者视频的帧内编码方式。但是基于深度学习的视频压缩框架还处于研究的起步阶段, 性能上还是低于基于混合视频编码框架的新一代视频编码标准平台, 比如HEVC。研究人员基于不同的网络结构, 针对视频的空域和时域冗余特性进行了一系列的设计和优化, 提出了不同的基于深度学习的视频编码框架。

2.1 基于卷积神经网络的视频编码

相比图像压缩, 视频不仅具有空域相关性, 也存在很强的时域相关性, 需要同时去除空域和时域的冗余才能有效提高视频压缩效率。Cheng等人将其基于卷积神经网络的端到端图像压缩扩展到视频压缩, 在视频编码环路内增加了视频插帧模块来去除时域冗余, 实现时域的紧致表示^[24]。具体来讲, 该方法采用Autoencoder的网络结构, 引入两个卷积滤波器作为下采样单元, 构成压缩正变换和反变换模块, 并利用加性均匀噪声模型来近似压缩噪声对编解码器联合训练; 对于帧内编码, 图像直接经过正变换、量化和算术编码器进行压缩; 帧间编码利用前后编码重构帧插值网络^[25]生成预测帧, 预测残差作为编码输入, 完整的编码框架如图8所示。该框架与混合编码框架比较相似, 整体相对简单。以

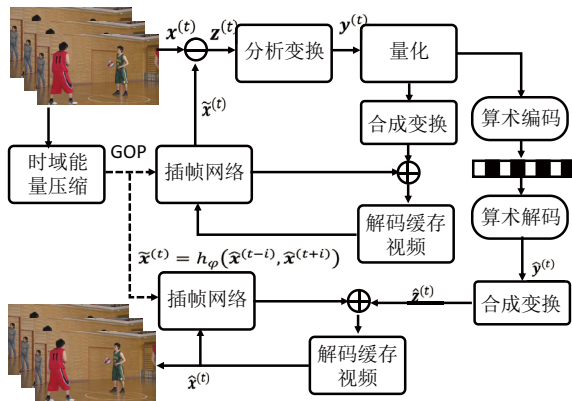


图8 基于卷积神经网络的视频压缩框架

MS-SSIM作为质量度量准则时，该方法在帧内编码上性能略高于HEVC，但是帧间编码性能上还是明显弱于HEVC的性能。

Habibian等人采用时空域联合的3D卷积网络以及视频质量和码率消耗联合的损失函数设计了新的端到端深度学习视频编码框架。其中编解码器是具有残差连接的全连接卷积网络模型，视频质量采用MS-SSIM，码率消耗采用基于PixelCNN^[26]估计的当前帧编码后隐变量的条件概率对数似然表示，该方法在一些特殊视频序列上取得了超过HEVC的性能，但是对于普通视频序列性能还是逊于HEVC。

2.2 基于深度生成网络的视频编码

近年来深度生成网络模型在图像压缩领域取得了显著的性能提升，Han等人将深度生成模型应用于视频编码中，提出了一种无监督的深度学习来进行视频编码，该方法同时学习最优的变换模型将视频在低维紧致空间进行表示和高

效的预测模型估计不同视频片段的概率，从而有助于实现相同信息损失下最小的编码长度^[27]。如图9所示，该方法对每帧视频通过局部状态网络进行编码消除空域冗余，而全局状态网络编码对视频片段进行编码去除时域冗余，编码器可以表示为在给定视频序列条件下预测其对应的隐变量和状态的分布。

$$q_{\theta}(z_{1:T}, f | x_{1:T}) \quad (3)$$

$$= q_{\theta}(f | x_{1:T}) \prod_{t=1}^T q_{\theta}(z_t | x_t)$$

其中， $x_{1:T} = (x_1, \dots, x_T)$ 表示原始视频序列， $z_{1:T} = (z_1, \dots, z_T)$ 是原始视频序列对应的局部隐变量， f 是原始序列对应的全局状态变量。该方法目前对于一些特定的低分辨率视频(64x64)在使用PSNR和MS-SSIM作为视频质量指标下可以取得超过HEVC和VP9编码器的

性能。

Pan等人利用生成对抗网络提出了一种基于颜色生成的视频编码方法，将输入视频变换到YCbCr空间，其中Y通道采用传统的编码方法HEVC进行压缩，然后利用解码的Y通道和生成网络模型生成Cb和Cr通道信号，但是目前生成网络得到的像素级保真度不高，该方法仅仅是超过了HEVC的帧内编码性能^[28]。

2.3 多网络模型联合的视频编码

除了上述端到端的视频编码网络外，一些研究人员也借鉴混合视频编码帧内、帧间预测的模块结构提出了多网络模型联合的视频编码框架。Chen等人借鉴混合视频编码的分块编码策略，对32x32的图像分别采用不同的卷积神经网络进行预测和残差编码，提出了新的深度学习视频编码框架DeepCoder^[29]。但是在帧内和帧间编码的预测选择上，该方法仅仅根据帧间预测的残差大小进行判断，同时编码器训练仅以输入图像和其重构图像的误差最小为目标，缺少码率的约束，因此该编码器性能还存在很大提升空间，目前其性能仅与x264默认配置相当。

Lu等人利用传统视频编码框

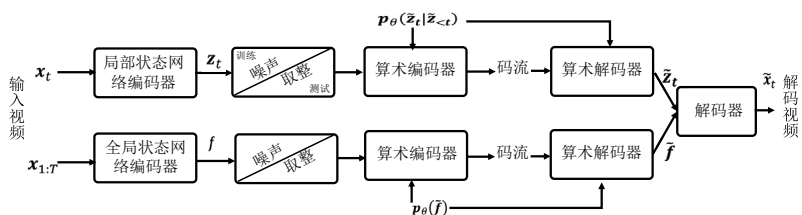


图9 基于深度生成网络的端到端视频编码

架和神经网络高效的非线性表达能力,对运动估计、运动补偿、运动压缩、残差压缩等多个模块联合学习,提出了端到端的深度视频压缩模型(Deep Video Compression, DVC)^[30],如图10所示。该方法主要采用卷积神经网络估计运动光流场、对依照光流场产生的预测图像进行滤波得到高质量的运动补偿视频帧以及采用非线性神经网络进行残差变换。为了进一步优化视频编码性能,网络训练采用率失真优化损失函数:

$$\lambda D + R = \lambda d(x_t, \hat{x}_t) + (H(\hat{m}_t) + H(\hat{y}_t)) \quad (4)$$

其中, x_t 和 \hat{x}_t 分别表示原始视频帧和解码重构视频帧, \hat{m}_t 和 \hat{y}_t 分别表示变换、量化后的运动向量和预测残差。该方案在通测序列上部分码率范围内取得了超过 very fast 配置下 x265 的性能,但还是明显低于 HEVC 的最优性能配置。

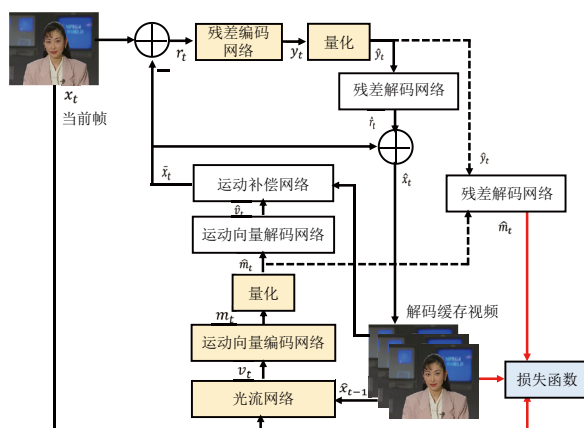


图10 多网络模型的端到端视频编码框架

3 人工智能时代深度学习视频编码的机遇和挑战

深度学习技术以数据为驱动、目标为导向在图像压缩领域已经取得了显著的性能提升,也为视频编码带来了新的契机。视频编码主要是通过预测模型来消除空域和时域的冗余,通过变换和熵编码去除信号相关性和统计冗余,在这些方面深度学习模型具有较强的优势。首先,深度学习模型可以通过视频大数据的训练有效地适应不同视频多样的空域和时域结构特性,提高视频预测性能。这在视频超分辨率、帧率提

升等问题中已经有了较好的证明,特别是视频生成网络模型在未来视频预测编码中可能具有较大的潜力;其次,深度学习模型可以具有多层级联和非线性连接方式,能够高效地抽象并加工信息,通过直接最小化信号失真和码率联合约束来实现最优的空间变换,相比早期基于字典的稀疏编码具有更多的自由度和灵活性。此外,深度学习模型对于视频纹理信息的抽象与表示在计算机视觉领域也取得了显著进步,因此未来基于深度学习模型可以通过对多任务的联合学习(比如:视频理解和压缩任务),实现对计算机视觉友好的视频信号的紧凑表达,从而可以在保障人工智能应用性能的前提下,提高视频观看质量和压缩效率。

但是,目前深度学习视频编码也面临了诸多挑战和亟需解决的问题。首先是压缩效率问题,目前深度学习模型在视频编码上的压缩效率还显著落后于 HEVC 编码标准,更大幅落后于下一代视频编码标准 VVC^[31]。主要问题在于如何通过深度学习模型提高帧间预测效率以及运动向量的编码,传统的编码方法通过精细粒度地图像块划分、多参考帧和运动向量预测等方法有效地去除了帧间相关性,然而目前基于深度学习的视频编码框架通常以帧作为预测单位,运动向量、多参考帧等时域相关性未能有效去除。其次是在视频质量评价方法上,传统的视频编码以 PSNR 为质量标准进行优化,可是深度学习模型的优势在于语义和视觉结构特性的保持,在像素级的表示和预测上并不具有明显优势,虽然研究人员对于 PSNR 在衡量视频质量上的不足有较高的共识,但是还没有形成被广泛认可的视频质量评价准则,导致目前面向视频压缩的深度学习模型在损失函数设计上缺乏理论支撑。此外,计算复杂度也是制约深度学习视频编码技术在工业界应用的重要因素,目前的深度学习视频编码模型普遍存在计算复杂度过高的问题,即使采用了高效并行计算平台 GPU,其编解码时间也远远超过了传统混合视频编解码器,因此设计低复杂度的轻量级深度学习模型也是深度学习视频编码技术从研究领域走向工业界亟需解决的问题之一。

4 结语

以信号处理技术为核心的视频编码经历数十年的发展已经取得了巨大的成功,但是现有视频编码框架正面临着压缩效率的瓶颈问题。在深度学习技术在人工智能领域取得巨大成功的背景下,本文重点介绍了传统混合视频编码框架下深度学习编码工具和基于深度学习模型的新型视频编码框架在近几年的发展状况,并总结了深度学习技术在视频编码中的潜力以及在目前阶段面临的挑战和亟需解决的问题。基于深度学习的视频编码技术目前还处于起步阶段,相信随着研究的深入和计算能力的提升,深度学习技术有可能成为下一代视频编码技术的突破口。

参考文献

- [1] Cisco.Cisco visual networking index: Forecast and trends 2017–2022[R].2018
- [2] ITU-T Recommendation H.120.Codecs for videoconferencing using primary digital group transmission[S].1993
- [3] MPEG-2.Information technology - Generic coding of moving pictures and associated audio information: Video[S].1994
- [4] Wiegand T,Sullivan G J,Bjøntegaard G,et al.Overview of the H.264/AVC video coding standard[J].IEEE Transactions on Circuits and Systems for Video Technology:2003,13(7):560–576
- [5] Sullivan G J,Ohm J R,Han W J,et al.Overview of the high efficiency video coding (HEVC) standard[J].IEEE Transactions on Circuits and Systems for Video Technology,2012,22(12):1649-1668
- [6] Liang F, Ma S, Wu F. Overview of AVS video standard[C]. IEEE International Conference on Multimedia and Expo, 2004:423-426
- [7] Zhang X, Xiong R, Lin W, et al. Low-Rank-Based Nonlocal Adaptive Loop Filter for High-Efficiency Video Compression[J]. IEEE Transactions on circuits and systems for video technology, 2017, 27(10): 2177-2188
- [8] Hu Y, Yang W, Li M, et al. Progressive Spatial Recurrent Neural Network for Intra Prediction[J]. IEEE Transactions on Multimedia, 2019, 21(12): 3024-3037
- [9] Meyer M, Wiesner J, Schneider J, et al. Convolutional Neural Networks for Video Intra Prediction Using Cross-component Adaptation[C]. IEEE International Conference on Acoustics, Speech and Signal Processing, 2019: 1607-1611
- [10] Li Y, Liu D, Li H, et al. Convolutional Neural Network-Based Block Up-Sampling for Intra Frame Coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(9): 2316-2330
- [11] Zhao L, Wang S, Zhang X, et al. Enhanced Motion-Compensated Video Coding With Deep Virtual Reference Frame Generation[J]. IEEE Transactions on Image Processing, 2019, 28(10): 4832-4844
- [12] Choi H, Bajić I V. Deep Frame Prediction for Video Coding. IEEE Transactions on Circuits and Systems for Video Technology[J/OL]. [2020-04-05]. <https://arxiv.org/abs/1901.00062>
- [13] Jimbo S, Wang J, Yashima Y. Block adaptive CNN/HEVC interframe prediction for video coding. International Workshop on Advanced Image Technology[C]. 2019
- [14] Zhao Z, Wang S, Wang S, et al. Enhanced Bi-Prediction With Convolutional Neural Network for High-Efficiency Video Coding. IEEE Transactions on Circuits and Systems for Video Technology[J]. 2019, 29(11): 3291-3301
- [15] Zhao X, Zhang L, Ma S, et al. Video Coding With Rate-Distortion Optimized Transform. IEEE Transactions on Circuits and Systems for Video Technology[J]. 2012, 22(1): 138-151
- [16] Liu D, Ma H, Xiong Z, et al. CNN-based DCT-like transform for image compression. International Conference on Multimedia Modeling[C]. 2018
- [17] Song R, Liu D, Li H, et al. Neural network-based arithmetic coding of intra prediction modes in HEVC. IEEE Visual Communications and Image Processing[C]. 2017: 1-4
- [18] Ma C, Liu D, Peng X, et al. Neural Network-Based Arithmetic Coding for Inter Prediction Information in HEVC. IEEE International Symposium on Circuits and Systems[C]. 2019: 1-5
- [19] Zhang X, Wang S, Zhang Y, et al. High-Efficiency Image Coding via Near-Optimal Filtering. IEEE Signal

- Processing Letters[J].2017,24(9):1403-1407
- [20] Jia C,Wang S,Zhang X,et al.Content-Aware Convolutional Neural Network for In-Loop Filtering in High Efficiency Video Coding. IEEE Transactions on Image Processing[J].2019,28(7):3343-3356
- [21] Krizhevsky A,Sutskever I,Hinton G E,et al.ImageNet classification with deep convolutional neural networks. Advances in neural information processing systems [C].2012:1097-1105
- [22] Zhang Y,Shen T, Ji X,et al.Residual Highway Convolutional Neural Networks for in-loop Filtering in HEVC. IEEE Transactions on Image Processing[J]. 2018,27(8):3827-3841
- [23] Song X,Yao J,Zhou L,et al.A Practical Convolutional Neural Network as Loop Filter for Intra Frame[C].IEEE International Conference on Image Processing,2018: 1133-1137
- [24] Cheng Z,Sun H,Takeuchi M,et al.Learning Image and Video Compression through Spatial-Temporal Energy Compaction[C].IEEE Conference on Computer Vision and Pattern Recognition,2019:10071-10080
- [25] Niklaus S,Mai L,Liu F.Video Frame Interpolation via Adaptive Separable Convolution[C].IEEE International Conference on Computer Vision,2017:261-270
- [26] Oord A,Kalchbrenner N,Espeholt L,et al.Conditional Image Generation with PixelCNN Decoders[C]. Advances in neural information processing systems, 2016:4790-4798
- [27] Han J,Lombardo S,Schroers C,et al.Deep Generative Video Compression[C].Advances in Neural Information Processing Systems,2019
- [28] Pan Z,Yuan F,Lei J,et al.Video Compression Coding via Colorization:A Generative Adversarial Network (GAN)-Based Approach[EB/OL].[2020-04-05].[https://arXiv preprint arXiv:1912.10653](https://arXiv.org/abs/1912.10653)
- [29] Chen T,Liu H,Shen Q,et al.DeepCoder:A deep neural network based video compression[C].IEEE Visual Communications and Image Processing,2017:1-4
- [30] Lu G,Ouyang W,Xu D,et al.DVC:An End-to-end Deep Video Compression Framework[C].IEEE Conference on Computer Vision and Pattern Recognition, 2019:10998-11007
- [31] Sullivan G J.Video Coding Standards Progress Report: Joint Video Experts Team Launches the Versatile Video Coding Project[J].SMPTE Motion Imaging Journal, 2018,127(8):94-98

作者简介



张新峰

工学博士，中国科学院大学计算机科学与技术学院助理教授，主要研究方向是视频和图像编码、处理和质量评价，目前已经发表国际学术期刊和会议论文100余篇，4次获得学术期刊和会议的最佳论文奖。

Review and Prospect on Deep Learning based Video Coding

Zhang Xinfeng

University of Chinese Academic of Sciences, Beijing 100049, China

Abstract In recent years, deep learning technologies have not only achieved great success in the field of artificial intelligence, but also brought new development opportunities to the field of video coding. This paper introduces the development status of video coding with deep learning technologies from two aspects, deep learning based video coding tools within traditional coding frameworks and new video coding frameworks based on deep learning models. The related representative works are introduced in detail and their performances are also analyzed. Finally, this paper analyzes and looks forward to the challenges and future development trends for deep learning based video coding technologies.

Keywords Video Coding; Artificial Intelligence; Deep Neural network; Video Prediction