



专题：智能通信技术的应用

基于神经网络的图像视频编码

贾川民, 赵政辉, 王苦社, 马思伟
(北京大学, 北京 100871)

摘要: 深度神经网络近年来在人工智能领域进展显著, 并引发广泛深入研究神经网络的热潮, 近期基于神经网络的图像视频编码也成为热点研究问题之一。系统梳理了基于神经网络的图像视频编码技术及进展, 对基于多层感知机、随机神经网络、卷积神经网络、循环神经网络、生成对抗网络等框架的图像压缩, 以及基于深度学习的各类视频编码工具进行了综述介绍, 同时对神经网络编码的未来发展趋势进行了分析与展望。

关键词: 神经网络; 深度学习; 卷积神经网络; 图像压缩; 视频编码

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2019142

Neural network based image and video coding technologies

JIA Chuanmin, ZHAO Zhenghui, WANG Shanshe, MA Siwei
Peking University, Beijing 100871, China

Abstract: Deep neural networks have achieved tremendous success in artificial intelligence, which makes the broad and in-depth research of neural network resurgence in recent years. Recently, the neural network based image and video coding has become one of the front-edge topics. A systematic and comprehensive review of neural network based image and video coding approaches based on network structure and coding modules were provided. The development of neural network based image compression, e.g. multi-layer perceptron, random neural network, convolutional neural network, recurrent neural network and generative adversarial network based image compression methods and neural network based video compression tools were introduced respectively. Moreover, the future trends in neural network based compression were also envisioned and discussed.

Key words: neural network, deep learning, CNN, image compression, video coding

1 引言

图像视频编码是多媒体信号处理中重点研究

的问题之一, 属于信源编码领域, 其理论基础为十九世纪四五十年代建立的信息论与编码理论^[1-2]。

图像视频编码的本质在于通过算法消除图像视频

收稿日期: 2019-04-10; 修回日期: 2019-05-08

基金项目: 国家自然科学基金资助项目 (No.61632001); 国家重点基础研究发展计划 (“973” 计划) 基金资助项目 (No.2015CB351800); 北京大学高性能计算平台资助项目

Foundation Items: The National Natural Science Foundation of China (No. 61632001), The National Key Technology R&D Program(973 Program) (No. 2015CB351800), Peking University High Performance Computing Platform

信号中存在的各类冗余信息（如空域、时域、视觉、统计等），以达到信号压缩的目的。在早期编码研究中，研究重心主要集中在对信源信号的无损压缩上，如哈夫曼编码、哥伦布编码以及算术编码等编码方法，这一类方法又统称熵编码。20世纪70年代是图像压缩技术的飞速发展时期，出现了将空域信号变换到频域信号的变换编码技术，即通过线性变换聚集能量到低频区域去除空域冗余，再使用系数扫描方法和熵编码对低频系数进行编码，从而将对像素的压缩转变为对低频系数的压缩，并根据不同的变换基函数分为傅里叶变换、哈达玛变换以及离散余弦变换域(discrete cosine transform, DCT)等经典变换编码技术。此后随着采集与显示设备的不断发展，对图像视频内容的高效压缩需求日益增长。为了高效消除视频信号中相邻帧之间的时域冗余，Brofferio等人^[3]提出基于块的运动估计和运动补偿预测编码，达到了平均1比特每像素的压缩效果。20世纪70年代末，Netravali等人^[4]提出预测与变换相结合的混合编码框架，在此后40年中混合编码框架

技术不断更迭，支持了更多精细化编码工具，如高精度分像素插值、多假设帧间预测以及环路滤波等模块，如图1所示。这一框架在图像视频编码研究领域取得了巨大的成功，形成了一系列编码标准，如JPEG、JPEG2000、MPEG1/2/4、H.120/261/2/3、H.264/AVC、HEVC/H.265、AVS-1/+2等，推动了数字媒体产业的变革^[5]，带动了超高清数字电视、IPTV、沉浸式多媒体服务以及互联网视频等应用的发展^[6]。

神经网络源于认知神经科学与数学的交叉研究，通过多层交替级联神经元与非线性激活函数构建的多层感知机(multi-layer perceptron, MLP)结构能够以足够小的误差实现对任意连续函数的逼近。神经网络的学习方法则经历了由20世纪60年代提出感知机学习算法^[7]，到20世纪80年代通过链式法则和反向传播算法建立的MLP学习过程^[8]，再到如今被广泛使用的随机梯度下降方法。然而，受限于训练数据和算力的匮乏，神经网络的发展在反向传播算法提出之后迎来了长时间沉寂。在1998年，LeCun等人^[9]提出

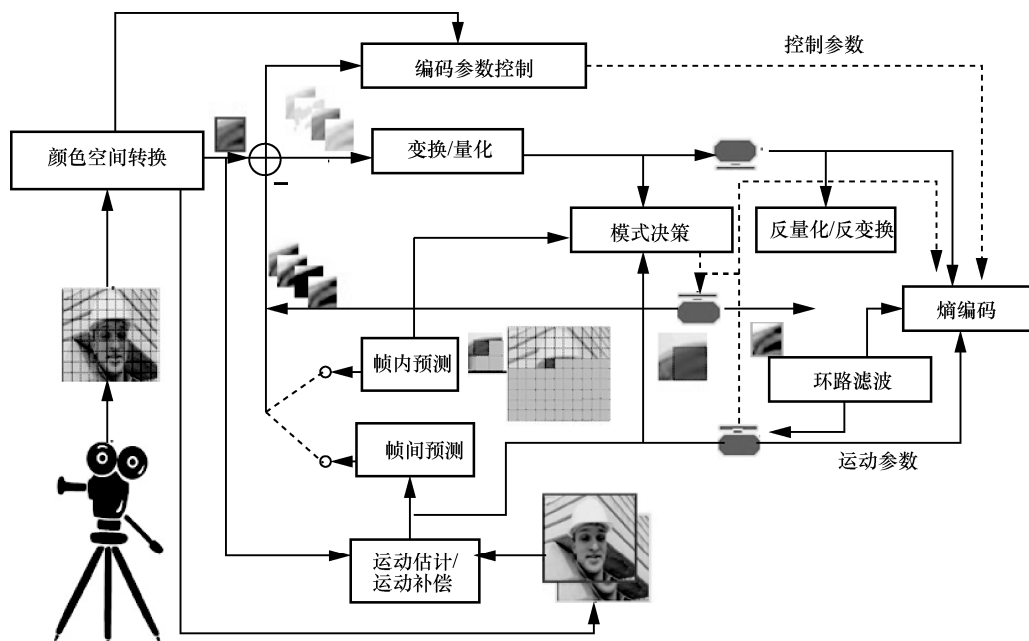


图1 预测/变换混合视频编码框架



了一种基于参数共享的轻量级卷积神经网络用于手写字识别任务。同期,为了解决时域信号梯度计算复杂度过高以及信号依赖问题,Hochreiter等人^[10]提出长短期记忆(long short-term memory, LSTM)结构,通过循环网络结构控制梯度传递实现对序列信号的高效学习。Nair等人^[11]通过对受限玻兹曼机(restricted Boltzmann machine, RBM)的每一层进行分层预训练,使得深层的神经网络训练变得可能。在解释了MLP结构具有更为优异的特征学习能力的同时,MLP在训练上的复杂度还可以通过逐层初始化和预训练来有效缓解。从此具有多隐含层的MLP结构研究再次成为热点,而神经网络也有了一个新的名称——深度学习(deep learning, DL)。

神经网络作为优化算法以及信号紧凑表征形式,与图像视频压缩结合的交叉研究早已有之。本文梳理了相关技术的发展历史和现状,以神经网络结构为主线对基于神经网络的图像压缩进行了系统性的介绍,并对各类神经网络编码工具和基于神经网络的端到端视频编码进行了详细介绍。同时展望了图像和视频编码领域目前面临的挑战以及未来的发展方向与机遇。

2 基于神经网络的图像压缩

在图像压缩中应用神经网络技术最早起源于20世纪80年代,如图2所示。此后伴随神经网络结构的发展,多层感知机网络、随机神经网络、卷积神经网络、循环神经网络以及生成对抗网络等网络结构都对基于神经网络的图像编码发展起到了很大的推动作用,将编码优化问题从基于手工特征和规则的局部优化推向了端到端的全局优化,从简单拟合率失真模型推向高度非线性复杂问题的优化建模与求解。尤其是近年来基于生成器—判别器的对抗学习范式引发了关于生成模型的研究热潮,此类基于生成对抗网络的压缩方法在提升压缩重建图像主观质量方面取得显著效果。

2.1 基于多层感知机的图像压缩

多层感知机也通常被称为人工神经网络(artificial neural network, ANN),其主要网络结构由稠密连接的输入层、隐含层和输出层组成。由于其能够实现对任意连续函数以足够小的误差进行逼近^[9],这为高维数据的降维和压缩提供了基础。1988年,Chua等人^[12]提出基于ANN的端到端图像压缩框架,该框架主要利用了ANN高密

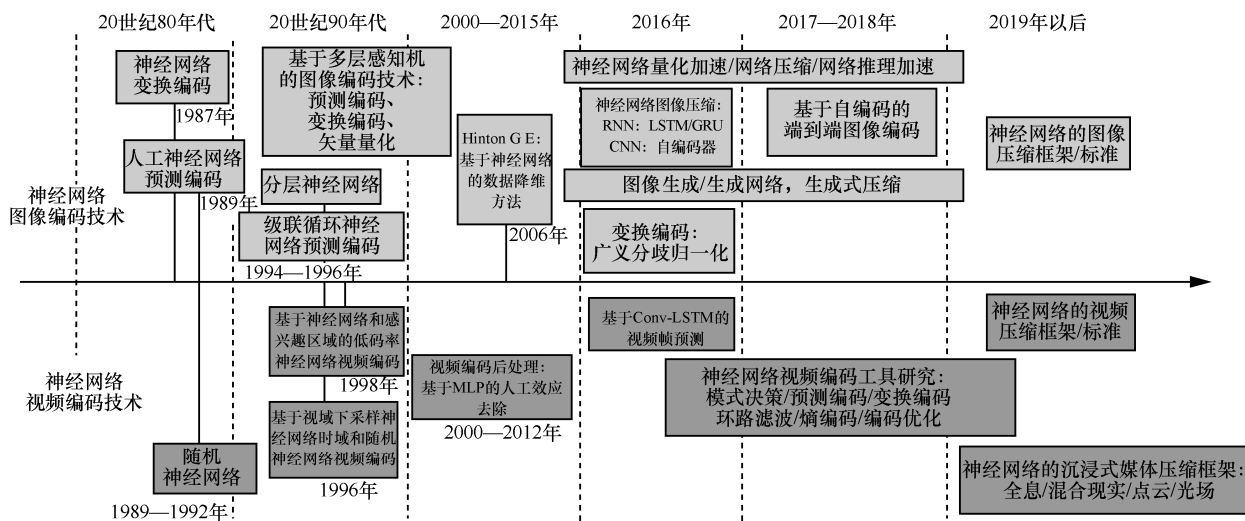


图2 基于神经网络的图像和视频编码相关技术发展历程

度并行计算的特性以及高效的信号紧凑表示能力, 将 ANN 作为变换编码模块的优化求解工具, 建模传统图像压缩中变换、量化以及二进制编码的过程, 再通过求解二进制整数解得到压缩图像的码流。Sonehara^[13]将图像压缩、量化和熵编码分成独立的模块, 采用类似于自编码器 (auto-encoder) 的结构设计 ANN 用于图像压缩, 其中瓶颈层 (bottleneck layer) 的维度明显小于输入与输出层, 从而实现图像降维和压缩。为了加速训练并降低运算复杂度, 完整的图像被分割为互不重叠的小图像块进行训练, 同时对不同位置的小图像块训练不同子网络以适应空域纹理多样性。然而这种直接将 ANN 用于像素变换编码的率失真性能不高, 泛化性也较差。

为了解决编码效率较低的问题, 参考文献[14]进一步利用图像空域像素作为上下文信息用于预测编码。如图 3 (a) 所示, 位于当前像素 X 左侧 (A)、上侧 (C) 以及左上 (B) 的像素能够被充分利用, 对当前像素值进行非线性预测; 再设计具有 30 个节点隐含层的 ANN 建模这一预测过程, 如图 3 (b) 所示, 通过反向传播算法最小化输入像素与输出像素的最小均方误差来训练网络参数。为进一步提升预测准确率, Manikopoulos^[15]通过神经网络建立了广义自回归 (auto-regressive) 模型, 以得到复杂纹理边缘和轮廓等图像内容的高阶预测。

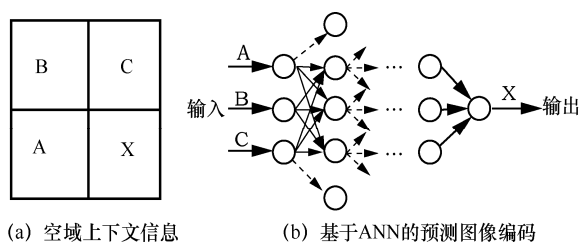


图 3 空域上下文信息和基于 ANN 的预测图像编码

2.2 基于随机神经网络的图像压缩

随机神经网络 (random neural network) 的概念诞生于 1989 年^[16], 与 ANN 通过梯度下降法处

理连续信号不同, 它的处理对象为脉冲信号, 在网络内部神经元之间的交互被建模为泊松过程。随机神经网络在图像编码中的应用主要集中于结合相对成熟的基于 ANN 的图像编码框架, 比如, Gelenbe 等人^[16]结合随机神经网络与图像压缩, 提出一种基于编码器—解码器结构的前馈神经网络。在此基础上, Cramer 等人^[17]扩展到基于块的处理并同时训练不同压缩比的自网络, 从而实现多种码率的图像压缩。

2.3 基于卷积神经网络的图像压缩

近年来, 卷积神经网络 (convolutional neural network, CNN) 技术在诸多领域进展显著, 如图像分类、目标检测等^[18]。与 ANN 相比, CNN 采用参数共享的方式大幅降低了所需要的参数量, 这种结构同时能够在一定程度上避免过拟合。然而, 直接训练端到端的 CNN 图像编码框架却是困难的, 这是由于在采用梯度下降法优化端到端编码中的率失真函数时需要目标函数具有全局可微性。而图像码率的计算则是基于量化后的离散系数, 从而使率失真函数无法满足整体连续条件。为了解决这个问题, Ballé 等人^[19]在 2016 年首次提出一种通过均匀分布的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分的。如图 4 所示, 框架主要分为由卷积层组成的分析模块与生成模块, 分别负责图像到紧凑表示的映射及其重建逆过程, 其中的激活函数为广义分歧归一化函数 (generalized divisive normalization, GDN)。该方法作为将 CNN 与端到端图像编码相结合的首创性工作, 取得了与 JPEG2000 相当的编码性能, 为基于 CNN 的图像编码注入新活力。在之后的两年中, 端到端图像编码框架持续演进, 新的码率估计方法被不断提出^[20-21], 编码性能也不断提升, 获得了与 HEVC 相当的性能^[22]。

2.4 基于循环神经网络的图像压缩

循环神经网络 (recurrent neural network, RNN) 能够处理前后具有依赖关系序列信号和时

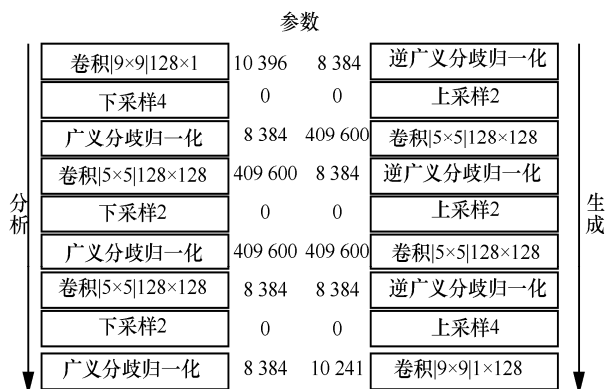


图4 Ballé等人^[19]提出的基于广义分歧归一化的卷积神经网络端到端图像编码框架

域信号。对信号和梯度传导具有一定“记忆”功能，这是RNN与CNN最大的不同点。具体指在神经网络连接时，各神经元不仅与下一层的神经元相连，同时还与自身相连，从而实现信息存储以及依赖。两种典型的RNN结构，LSTM和门控循环单元（gated recurrent unit，GRU）^[23]在时序建模上具有其优越性。Toderici等人^[24]于2017年提出一种迭代式RNN图像编码方法，通过限制每一次迭代的比特数来实现码率控制，再迭代处理RNN解码的信号和原始信号的残差来提升重建图

像细节和质量。基于RNN的端到端图像编码框架如图5所示，每一次迭代中具有3个主要模块，分别为编码网络（E）、二值化模块（B）以及解码网络（D）。在此基础上，Minnen等人^[25]进一步探索图像内部空域相关性，在RNN结构之前引入空域预测模块，实现了与JPEG2000相当的编码性能。

2.5 基于生成对抗网络的图像压缩

生成对抗式网络（generative adversarial network，GAN）是深度学习领域近年来引起最关注的方向，其学习方式包含了对抗与博弈的思想。这一学习范式最早出现于2014年^[26]，在GAN模块中，生成器和判别器可通过端到端学习的方法同时优化，基于深度神经网络的判别器来判断生成器得到的样本是否为真，同时生成器通过在目标函数中加入对抗损失函数不断根据采样信号生成高质量样本数据，使得判别器无法区分生成样本之真实性。在图像压缩任务中，GAN的这一特性能够用于提升生成图像的主观质量，从而提升压缩性能。其中具有代表性的工作便是Rippel和

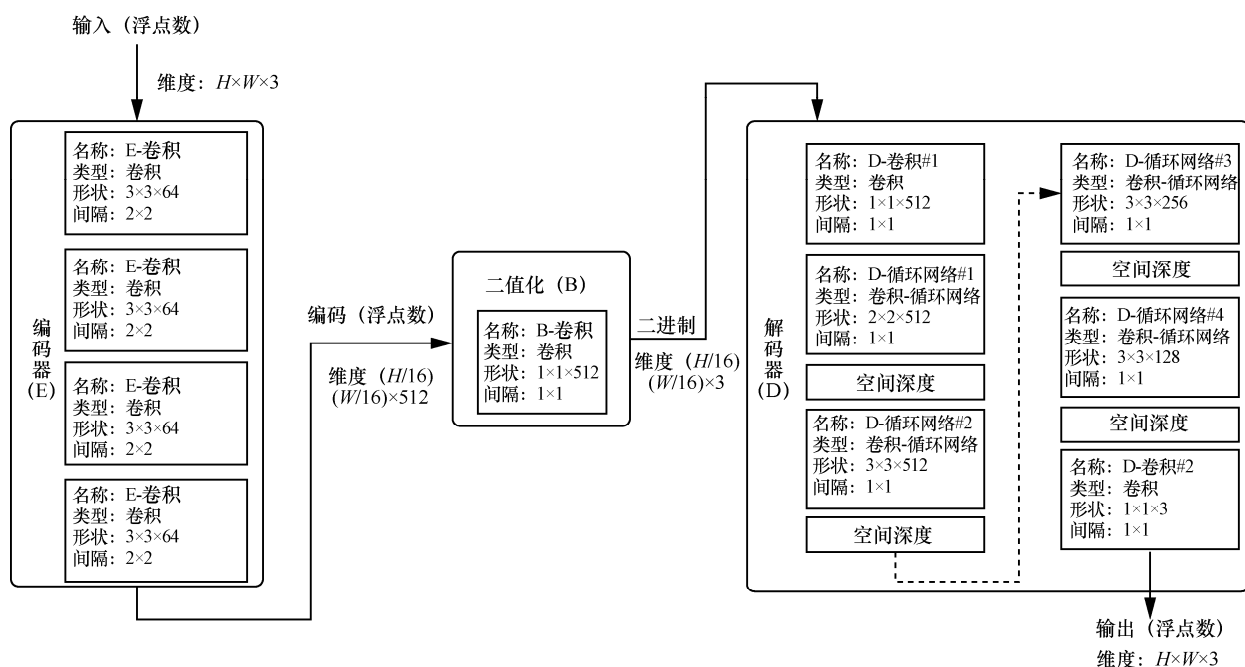


图5 基于RNN的端到端图像编码框架^[24]

Bourdev^[27]在 2017 年探索基于 GAN 的图像编码器,通过在率失真目标函数中引入对抗损失函数进行端到端训练。该编码框架在率失真性能和模型复杂度两方面都取得优异效果:低码率下的主观重建质量大幅超越 JPEG、JPEG2000、WebP 和 HEVC 等主流图像编码标准;结合 GPU 实现了实时编解码。对应的网络结构如图 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结构,与 RNN 与 CNN 结构不同的是,通过对抗学习训练方式得到的生成器能够生成逼真的纹理细节。通过进一步主观测试可以发现,基于 GAN 的编码器能够重建具有更好视觉效果的原因是生成了更丰富的高频细节。近两年 GAN 在图像压缩技术中的探索持续演进,诸如极端码率约束下的图像压缩^[28]和保持图像语义信息的生成式压缩^[29]等领域同样取得显著进展。

3 基于神经网络的视频压缩

基于神经网络的视频编码的研究主要集中于 3 个主要方面:混合式神经网络视频编码(即将神经网络代替传统编码模块嵌入到视频编码框架)、神经网络率失真优化编码以及端到端视频编码。

3.1 混合式神经网络视频编码

(1) 预测编码

基于神经网络的预测编码分别从帧内预测与帧间预测开展研究。在帧内预测方面,虽然上述基于神经网络的图像压缩算法均可用于帧内预测

编码,但是目前从编码性能和模型结构特性来说却均未能展现出具有替代 HEVC 帧内编码模式的能力,直接复用是不合适的。一方面是由于以 HEVC 为代表的视频编码算法通过数十年标准的迭代局部优化确实取得了优异的编码性能;另一方面也说明神经网络图像编码仍然具有较大的性能增长空间。基于此,研究人员转而探索将神经网络直接与视频编码帧内预测的高效结合,进一步挖掘帧内预测编码增益。Cui 等人^[30]提出一种帧内预测卷积神经网络(intra prediction convolutional neural network, IPCNN),对 HEVC 中经率失真决策得到的 8×8 预测块使用 IPCNN 进行预测增强处理,并联合每一个 8×8 预测块的左侧、上侧以及左上的 8×8 块作为帧内预测上下文信息联合输入神经网络。IPCNN 不仅继承了 HEVC 率失真预测的结果,并且结合额外的空域上下文信息得到更精准的预测。与 IPCNN 的思路类似, Li 等人^[31]使用了 MLP 网络训练一种全连接网络帧内预测(intra prediction fully connected network, IPFCN)模式,并与 HEVC 已有的 35 种模式进行率失真优化决策,如图 7 所示,IPFCN 的输入仅为当前预测块左上方的重建值,相较于 HEVC 该方法取得了 3% 的编码性能。此方法优势在于无需在 HEVC 帧内预测模块的基础之上进行,然而,由于 IPFCN 采用全连接网络结构,参数量过大造成计算复杂度过高,虽然能够取得平均 3% 的性能,却得不偿失。通过减少一半以上的参数量,IPFCN 依旧有超过 2% 的编码性能,但复杂度仍然

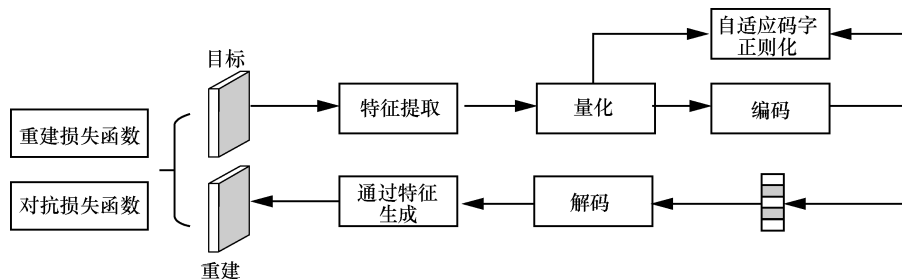
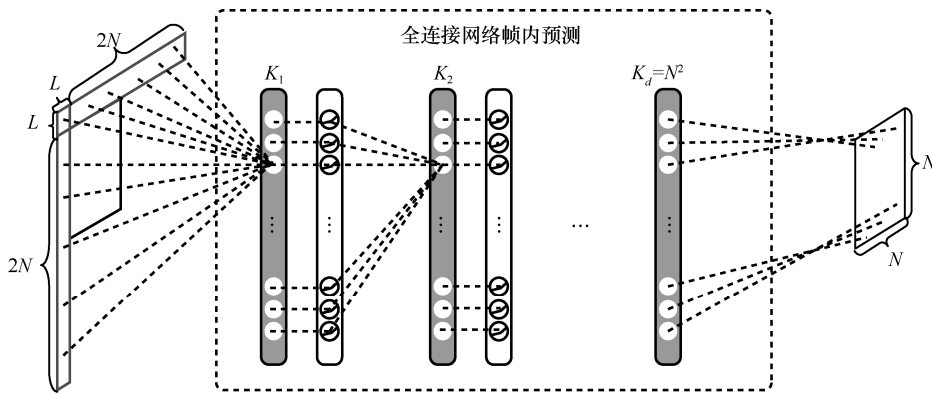


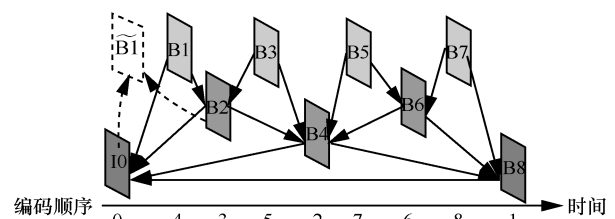
图 6 基于 GAN 的端到端图像编码框架^[27]

图7 Li 等人^[31]提出的全连接网络帧内预测模式

较高。类似的做法也同样被扩展到基于 RNN 的帧内预测上, Hu 等人^[32]用当前编码块的空域重构像素作为自回归模型的上下文信息, 采用 RNN 的方法渐进式生成帧内预测信息, 解决帧内预测中当前块左上与右下侧上下文信息不对称的问题。神经网络帧内编码的另一种思路则是采用上采样—下采样编码框架, 结合神经网络对编码块进行不同分辨率的采样操作, 从而有效地提升编码性能^[33]。最后, 针对帧内预测模式索引的高效编码, Pfaff 等人^[34]采用了直接对帧内模式索引进行预测编码的方法, 结合 MLP 结构实现对帧内模式的预测。

在帧间预测方面, 神经网络与视频编码的结合更加广泛。Huo 等人^[35]结合视频编码的时域运动补偿和空域相关性提出基于神经网络的运动补偿增强 (convolutional neural network based motion compensation refinement, CNNMCR)。CNNMCR 获得增益的方式主要是对帧间运动补偿的结果进行噪声抑制和细节增强。对于运动估计模块, 基于图像超分辨率的 CNN 模块可被用于生成更高质量的分像素以提升运动估计性能^[36]。受益于计算机视觉领域中帧率提升技术的发展, Zhao 等人^[37]利用帧间编码的参考帧训练 CNN 并合成出虚拟参考帧作为帧间预测的候选, 如图 8 所示, 该方法能够有效捕捉运动的连续性以及解决场景遮挡

问题, 从而显著提升帧间预测效率。不仅如此, Zhao 等人^[38]探究了 CNN 在双向预测中的融合特性, CNN 能够通过监督学习的方法学习到帧间编码中双向预测的高阶融合方式。该方法较之于 HEVC 中简单线性组合方法能够大幅提升编码性能。

图8 Zhao 等人^[37]提出的基于 CNN 的帧率提升方法用于虚拟参考帧生成

(2) 熵编码

HEVC 采用自适应上下文二进制算术编码 (context adaptive binary arithmetic coding, CABAC) 作为熵编码工具。Song 等人^[39]针对 HEVC 设计了 CNN 以预测帧内编码最可能模式 (most probable mode, MPM) 的概率分布, 在熵编码模块取得了显著性能提升。与传统基于规则的上下文概率估计规则不同, 该方法能够直接对帧内预测模式的概率进行直接估计, 从而提升熵编码的效率。Ma 等人^[40]在此基础之上扩展了概率估计模块, 将基于 CNN 的上下文预测模块替换传统的基于规则的上下文概率预测模型, 并应用于

HEVC 中 DC 系数的编码,取得了平均 1.6%的率失真性能提升。

(3) 环路滤波

深度神经网络在图像重建领域飞速发展,为基于神经网络的环路滤波算法提供了新的思路。传统视频编码的环路滤波包含去块效应滤波和样本自适应偏移,而这两个编码工具均基于信号处理理论并由人工设计相关滤波参数。近年来出现了多种 CNN 环路滤波结构通过端到端训练滤波重建过程,不仅取得明显的编码增益同时提升重建视频主观质量。Zhang 等人^[41]设计了一种跳跃式残差神经网络,并对不同的量化参数(quantization parameter, QP)以及帧类型训练对应的神经网络模型,在低码率随机访问配置下取得了超过 4%编码性能;由于时域信息为提升编码重建质量能够提供更多的先验知识,通过在环路滤波中融合时域与空域信息,Jia 等人^[42]提出了时空域残差网络(spatial-temporal residue network, SRResNet);不仅如此,Jia 等人^[43]进一步基于内容特性迭代训练最优 CNN 模型以适应不同纹理特性的环路滤波需求,在通用测试条件下取得了最高超过 10%的编码性能。

基于 CNN 的视频后处理滤波同样引起了众多研究者的关注,如去人工效应去除的卷积神经网络(artifact removal convolutional neural network, ARCNN)^[44]、可变尺寸卷积残差网络(variable-size residue convolutional neural network, VRCNN)^[45]和质量增强卷积网络(quality enhancement convolutional neural network, QE-CNN)^[46]等。

3.2 神经网络率失真优化编码

神经网络结构具有高度非线性特点,可通过监督学习训练成为高效的判别模型与分类器,为神经网络在视频编码中编码优化提供了新方向。编码模式决策的快速算法是其中的代表,不同于传统基于手工特征进行编码单元(coding unit,

CU)和预测单元(prediction unit, PU)划分,基于神经网络的 CU 划分通过分析其纹理特性和图像细节,并辅之以编码参数,不仅能够实现软件层面的快速编码端决策,同时在硬件编码器设计中也能取得令人满意的编码性能与编码速度的折中^[48],如图 9 所示^[47],CNN 与编码 QP 相结合用于图像纹理分析并将模式决策问题转化为分类问题,实现了对四叉树—二叉树(quad-tree plus binary-tree, QTBT)结构的编码端帧间快速决策算法。Xu 等人^[49]结合 CNN 与 LSTM 对多种 HEVC 编码配置的划分结构设计了端到端决策网络,实现了神经网络完全替换传统模式决策。

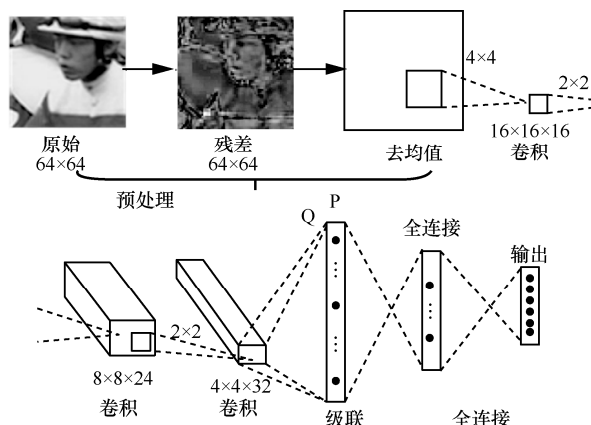


图9 基于CNN和MLP的快速帧间编码决策算法^[47]

神经网络不仅是高效判别模型,同时也是复杂问题建模的优化方法之一。针对传统视频编码中经典的率失真优化模型,Xu 等人^[50]提出了基于CNN的失真度量模型和CNN与MLP级联的码率估计模型,用于近似率失真模型,实现高效的码率分配方案。

3.3 端到端视频编码

与图像压缩类似,基于神经网络的端到端视频压缩旨在通过神经网络实现完整编码框架。参考文献^[51]通过用CNN替换传统混合框架中的各个模块实现了整体基于深度学习的视频编码框架。Chen 等人^[52]设计了(pixel motion CNN, PMCNN)以实现传统编码中的帧间运动估计模



块, 再结合 LSTM 模块与 CNN 模块对连续的图像组 (group of picture, GOP) 进行三维卷积, 最后利用基于神经网络的熵编码模块得到压缩视频流, 取得了与 H.264/AVC 接近的主观重建质量。参考文献[53]提出了一种端到端的视频编码框架, 通过 CNN 图像编码关键帧并结合视频时域帧插值的方法生成未编码的非关键帧。由于端到端视频编码框架相较于神经网络图像编码拥有时域信息和复杂的运动特性, 相关高效编码方法仍处于探索阶段, 已有工作为未来该领域的发展提供一些有益启发。

4 神经网络图像视频编码技术的机遇与挑战

图像视频编码旨在寻求一种在信号保真测度下的紧凑表示。神经网络编码方法为该领域注入的活力主要体现在 3 个方面: 首先, 神经网络对于图像视频丰富的内容特性具有极高的自适应能力。这是由于神经网络参数来源于数据驱动方式进行的优化学习, 同时通过调整优化过程中超参数保持泛化性能, 因而对图像视频编码的纹理多样性具有较好的支持; 其次, 神经网络通过多层级联的方式构建信息通路, 抽象并加工信息, 以获得高度非线性结构和广阔的感受区域, 较之传统图像视频编码技术中使用局部重构像素具有更丰富的上下文信息, 从而取得更高的编码效率; 再次, 神经网络对于纹理信息的抽象与表示是对图像视频信号的高层紧凑表达, 因此它不仅能够用于纹理信息本身的重建, 同时可服务于基于图像视频内容的分析应用, 从这个角度看来, 实现对人类视觉和机器视觉的联合编码优化是可行的。

神经网络图像视频编码技术同样面临着诸多挑战和未解决的问题。首先是信号保真度量层面, 神经网络编码技术的优势在于语义保持和视觉优化特性, 对质量评价和信号失真度的评测方法需要辩证看待; 其次, 图像视频编码技术的基础为率失真理论, 探索神经网络优化与率失真优化的交叉研

究也是未来神经网络编码技术需要突破的基础研究; 再次, 由于图像视频编码对实时性和高效性的要求, 轻量级和低复杂度的神经网络推理计算是神经网络编码技术从研究领域走向标准化的必经之路, 也是神经网络领域亟需解决的问题之一。

5 结束语

神经网络与图像视频编码领域虽然各自经过了数十年的发展, 但随着算力的提升和数据量的指数型增长, 新技术和应用仍然不断涌现。本文以神经网络结构和图像视频编码模块为主线梳理了神经网络技术在图像和视频编码过去 30 年的演进。一方面, 神经网络在图像编码中主要通过预测编码与非线性变换实现图像的高效紧凑表达和压缩, 其中经历了由模块局部优化到整体端到端学习的演进, 同时取得可喜进展; 另一方面, 基于神经网络的视频编码在目前主流视频编码标准的基础之上取得了显著的性能提升, 或许是未来视频编码架构发展的一个重要方向; 不仅如此, 由于深度神经网络强大的表征学习和抽象表达能力, 图像视频信号层面保真度的压缩与语义层面的内容分析应用不再是独立研究的个体, 通过神经网络的紧凑表示可以建立起高效图像视频压缩与媒体分析的桥梁, 在为图像视频压缩带来理论研究和设计原理突破的同时, 有可能服务于全新的视频应用。

参考文献:

- [1] SHANNON C E. A mathematical theory of communication[J]. Bell System Technical Journal, 1948, 27(3): 379-423.
- [2] BERGER T. Rate-distortion theory[J]. Wiley Encyclopedia of Telecommunications, 2003 (4).
- [3] BROFFERIO S, CAFFORIO C, RE P D, et al. Redundancy reduction of video signals using movement compensation[J]. Alta Frequenza, 1974, 43(10): 572-579.
- [4] NETRAVALI A N, STULLER J A. Motion-compensated transform coding[J]. Bell System Technical Journal, 1979, 58(7): 1703-1718.
- [5] 周建同, 杨海涛, 刘东, 等. 视频编码的技术基础及发展方向[J]. 电信科学, 2017, 33(8): 16-25.
ZHOU J T, YANG H T, LIU D, et al. Trends and technologies of

- video coding [J]. Telecommunications Science, 2017, 33(8): 16-25.
- [6] 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.
 - [7] ROSENBLATT F. Principles of neurodynamics: perceptrons and the theory of brain mechanisms[R]. 1961.
 - [8] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533.
 - [9] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
 - [10] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.
 - [11] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]//The 27th international conference on machine learning (ICML-10), June 21-24, 2010, Haifa, Israel. Norristown: Omnipress, 2010: 807-814.
 - [12] CHUA L O, LIN T. A neural network approach to transform image coding[J]. International journal of circuit theory and applications, June 18-22, 1988, 16(3): 317-324.
 - [13] SONEHARA N. Image data compression using a neural network model[C]//International 1989 Joint Conference on Neural Networks, June 18-22, 1989, Washington, DC, USA. Piscataway: IEEE Press, 1989.
 - [14] DIANAT S A, NASRABADI N M, VENKATARAMAN S. A non-linear predictor for differential pulse-code encoder (DPCM) using artificial neural networks[C]//International Conference on Acoustics, Speech, and Signal Processing, April 14-17, 1991, Toronto, USA. Piscataway: IEEE Press, 1991: 2793-2796.
 - [15] MANIKOPOULOS C N. Neural network approach to DPCM system design for image coding[J]. IEEE Proceedings I (Communications, Speech and Vision), 1992, 139(5): 501-507.
 - [16] GELENBE E. Random neural networks with negative and positive signals and product form solution[J]. Neural Computation, 1989, 1(4): 502-510.
 - [17] CRAMER C, GELENBE E, BAKIRCIOGLU I. Video compression with random neural networks[C]// International Workshop on Neural Networks for Identification, Control, Robotics and Signal/Image Processing, August 23-26, 1996, Venice, Italy. Piscataway: IEEE Press, 1996: 476-484.
 - [18] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436.
 - [19] BALLÉ J, LAPARRA V, SIMONCELLI E P. End-to-end optimized image compression[J]. arXiv:1611.01704, 2016.
 - [20] THEIS L, SHI W, CUNNINGHAM A, et al. Lossy image compression with compressive autoencoders[J]. arXiv: 1703.00395, 2017.
 - [21] AGUSTSSON E, MENTZER F, TSCHANNEN M, et al. Soft-to-hard vector quantization for end-to-end learning compressible representations[J]. arXiv:1704.00648, 2017.
 - [22] BALLÉ J, MINNEN D, SINGH S, et al. Variational image compression with a scale hyperprior[J]. arXiv:1802.01436, 2018.
 - [23] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv:1412.3555, 2014.
 - [24] TODERICI G, VINCENT D, JOHNSTON N, et al. Full resolution image compression with recurrent neural networks[C]//The IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, USA. Piscataway: IEEE Press, 2017: 5306-5314.
 - [25] MINNEN D, TODERICI G, COVELL M, et al. Spatially adaptive image compression using a tiled deep network[C]//2017 IEEE International Conference on Image Processing (ICIP), Sept 17-20, 2017, Beijing, China. Piscataway: IEEE Press, 2017: 2796-2800.
 - [26] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Advances in neural information processing systems, Dec 8-13, 2014, Montreal, Canada. Cambridge: MIT Press, 2014: 2672-2680.
 - [27] RIPPEL O, BOURDEV L. Real-time adaptive image compression[Z]. 2017.
 - [28] AGUSTSSON E, TSCHANNEN M, MENTZER F, et al. Generative adversarial networks for extreme learned image compression[J]. arXiv:1804.02958, 2018.
 - [29] GREGOR K, BESSE F, REZENDE D J, et al. Towards conceptual compression[C]//Advances in Neural Information Processing Systems, Dec 5-10, 2016, Barcelona, Spain. [S.l.: s.n.], 2016: 3549-3557.
 - [30] CUI W, ZHANG T, ZHANG S, et al. Convolutional neural networks based intra prediction for HEVC[J]. arXiv:1808.05734, 2018.
 - [31] LI J, LI B, XU J, et al. Fully connected network-based intra prediction for image coding[J]. IEEE Transactions on Image Processing, 2018, 27(7): 3236-3247.
 - [32] HU Y, YANG W, XIA S, et al. Enhanced intra prediction with recurrent neural network in video coding[C]//2018 Data Compression Conference, Mar 27-30, 2018, Snowbird, USA. Piscataway: IEEE Press, 2018: 413.
 - [33] 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.
 - [34] PFAFF J, HELLE P, MANIRY D, et al. Neural network based intra prediction for video coding[Z]. 2018.
 - [35] HUO S, LIU D, WU F, et al. Convolutional neural network-based motion compensation refinement for video coding[C]//2018 IEEE International Symposium on Circuits and Systems (ISCAS), May 26-29, 2018, Sapporo, Japan. Piscataway: IEEE Press, 2018: 1-4.
 - [36] YAN N, LIU D, LI H, et al. Convolutional neural network-based fractional-pixel motion compensation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018: 840-853.



- [37] ZHAO L, WANG S, ZHANG X, et al. Enhanced CTU-level inter prediction with deep frame rate up-conversion for high efficiency video coding[C]//2018 25th IEEE International Conference on Image Processing (ICIP), Oct 7-10, 2018, Athens, Greece. Piscataway: IEEE Press, 2018: 206-210.
- [38] ZHAO Z, WANG S, WANG S, et al. Enhanced bi-prediction with convolutional neural network for high efficiency video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018: 1.
- [39] SONG R, LIU D, LI H, et al. Neural network-based arithmetic coding of intra prediction modes in HEVC[C]//2017 IEEE Visual Communications and Image Processing (VCIP), Dec 10-13, 2017, Petersburg, USA. Piscataway: IEEE Press, 2017: 1-4.
- [40] MA C, LIU D, PENG X, et al. Convolutional neural network-based arithmetic coding of DC coefficients for HEVC intra coding[C]//2018 25th IEEE International Conference on Image Processing (ICIP), Oct 7-10, 2018, Athens, Greece. Piscataway: IEEE Press, 2018: 1772-1776.
- [41] ZHANG Y, SHEN T, JI X, et al. Residual highway convolutional neural networks for in-loop filtering in HEVC[J]. IEEE Transactions on Image Processing, 2018, 27(8): 3827-3841.
- [42] JIA C, WANG S, ZHANG X, et al. Spatial-temporal residue network based in-loop filter for video coding[C]//2017 IEEE Visual Communications and Image Processing (VCIP), Nov 10-13, 2017, St. Petersburg, USA. Piscataway: IEEE Press, 2017: 1-4.
- [43] JIA C, WANG S, ZHANG X, et al. Content-aware convolutional neural network for in-loop filtering in high efficiency video coding[J]. IEEE Transactions on Image Processing, 2019: 1.
- [44] DONG C, DENG Y, LOY C C, et al. Compression artifacts reduction by a deep convolutional network[C]// the IEEE International Conference on Computer Vision, June 7-12, 2015, Boston, USA. Piscataway: IEEE Press, 2015: 576-584.
- [45] DAI Y, LIU D, WU F. A convolutional neural network approach for post-processing in HEVC intra coding[C]//International Conference on Multimedia Modeling, Jan 4-6, 2017, Reykjavik, Iceland. Heidelberg: Springer, 2017: 28-39.
- [46] YANG R, XU M, WANG Z. Decoder-side HEVC quality enhancement with scalable convolutional neural network[C]//2017 IEEE International Conference on Multimedia and Expo (ICME), July 10-14, 2017, Hongkong, China. Piscataway: IEEE Press, 2017: 817-822.
- [47] WANG Z, WANG S, ZHANG X, et al. Fast QTBT partitioning decision for interframe coding with convolution neural network[C]//2018 25th IEEE International Conference on Image Processing (ICIP), Oct 7-10, 2018, Athens, Greece. Piscataway: IEEE Press, 2018: 2550-2554.
- [48] LIU Z, YU X, GAO Y, et al. CU partition mode decision for HEVC hardwired intra encoder using convolution neural network[J]. IEEE Transactions on Image Processing, 2016, 25(11): 5088-5103.
- [49] XU M, LI T, WANG Z, et al. Reducing complexity of HEVC: a deep learning approach[J]. IEEE Transactions on Image Processing, 2018, 27(10): 5044-5059.
- [50] XU B, PAN X, ZHOU Y, et al. CNN-based rate-distortion modeling for H. 265/HEVC[C]//2017 IEEE Visual Communications and Image Processing (VCIP), Dec 10-13, St. Petersburg, USA. Piscataway: IEEE Press, 2017: 1-4.
- [51] CHEN T, LIU H, SHEN Q, et al. Deepcoder: a deep neural network based video compression[C]//2017 IEEE Visual Communications and Image Processing (VCIP), Dec 10-13, St. Petersburg, USA. Piscataway: IEEE Press, 2017: 1-4.
- [52] CHEN Z, HE T, JIN X, et al. Learning for video compression[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019 (14).
- [53] WU C Y, SINGHAL N, KRAHENBUHL P. Video compression through image interpolation[C]//Proceedings of the European Conference on Computer Vision (ECCV), Sep 8-14, 2018, Munich, Germany. Heidelberg: Springer, 2018: 416-431.

[作者简介]



贾川民 (1993-), 男, 北京大学博士生, 主要研究方向为图像视频编码与处理。



赵政辉 (1993-), 男, 北京大学博士生, 主要研究方向为图像视频编码与处理。



王苦社 (1981-), 男, 博士, 北京大学助理研究员, 主要研究方向为视频编码与视频处理。

马思伟 (1979-), 男, 博士, 北京大学教授、博士生导师, 主要研究方向为视频编码与处理。