

中图分类号: TP3 文献标识码: A 文章编号: 1006-8961(2021)06-1179-22

论文引用格式: Jia C M, Ma H C, Yang W H, Ren W Q, Pan J S, Liu D, Liu J Y and Ma S W. 2021. Video processing and compression technologies. Journal of Image and Graphics 26(06): 1179-1200(贾川民, 马海川, 杨文瀚, 任文琦, 潘金山, 刘东, 刘家瑛, 马思伟. 2021. 视频处理与压缩技术. 中国图象图形学报 26(06): 1179-1200 [DOI: 10. 11834/jig. 200861]

视频处理与压缩技术

贾川民¹, 马海川², 杨文瀚³, 任文琦⁴, 潘金山⁵, 刘东², 刘家瑛⁶, 马思伟^{1*}

1. 北京大学信息科学技术学院, 北京 100871; 2. 中国科学技术大学信息科学技术学院, 合肥 230027;
3. 香港城市大学计算机科学系, 中国香港 999077; 4. 中国科学院信息工程研究所, 北京 100196;
5. 南京理工大学计算机科学与工程学院, 南京 210094; 6. 北京大学王选计算机研究所, 北京 100871

摘要: 视频处理与压缩是多媒体计算与通信领域的核心主题之一, 是连接视频采集传输和视觉分析理解的关键桥梁, 也是诸多视频应用的基础。当前“5G + 超高清 + AI”正在引发多媒体计算与通信领域的新一轮重大技术革新, 视频处理与压缩技术正在发生深刻变革, 亟需建立视频大数据高效紧凑表示理论和方法。为此, 学术研究机构 and 工业界对视频大数据的视觉表示机理、视觉信息紧凑表达、视频信号重建与恢复、高层与低层视觉融合处理方法及相应硬件技术等前沿领域进行了广泛深入研究。本文从数字信号处理基础理论出发, 分析了当前视频处理与压缩领域的热点问题和研究内容, 包括基于统计先验模型的视频数据表示模型及处理方法、融合深度网络模型的视频处理技术、视频压缩技术以及视频压缩标准进展等领域。详细描述了视频超分辨率、视频重建与恢复、视频压缩技术等面临的前沿动态、发展趋势、技术瓶颈和标准化进程等内容, 对国际国内研究内容和发展现状进行了综合对比与分析, 并展望了视频处理与压缩技术的发展与演进方向。更高质量视觉效果和高效率视觉表达之间将不再是单独研究的个体, 融合类脑视觉系统及编码机理的视频处理与压缩技术将是未来研究的重要领域之一。

关键词: 多媒体技术; 视频信号处理; 视频压缩; 人工智能; 深度学习

Video processing and compression technologies

Jia Chuanmin¹, Ma Haichuan², Yang Wenhan³, Ren Wenqi⁴, Pan Jinshan⁵,
Liu Dong², Liu Jiaying⁶, Ma Siwei^{1*}

1. School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China;
2. School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China;
3. Department of Computer Science, City University of Hong Kong, Hong Kong 999077, China;
4. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100196, China;
5. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China;
6. Wangxuan Institute of Computer Technology, Peking University, Beijing 100871, China

Abstract: Video processing and compression are the most fundamental research areas in multimedia computing and communication technologies. They play a significant role in bridging video acquisition, video streaming, and video delivery together with the visual information analysis and visual understanding. Video processing and compression are also the foundations of

收稿日期: 2020-12-31; 修回日期: 2021-01-12; 预印本日期: 2021-01-19

* 通信作者: 马思伟 swma@pku.edu.cn

基金项目: 国家自然科学基金项目(61772483, 61632001); 中国博士后科学基金项目(2020M680238)

Supported by: National Natural Science Foundation of China (61772483, 61632001)

applicational multimedia technologies and support various down-stream video applications. Digital videos are the largest big data in our contemporary modern society. The multimedia industry is the core component of the intellectual information era. The human kind steps into the intellectual information era with the continuous development of artificial intelligence and new generation of information revolution. Many emerging interdisciplinary research topics interact and fuse. Currently, the 5G plus ultra-high definition plus artificial intelligence invokes a novel trend of massive technology revolution in the context of multimedia computing and communication. The video processing and compression techniques also face challenging and intensive reform given this background. The demands for the theoretical and applicational breakthrough research on the compact video data representations, the highly efficient processing pipelines, and the high-performance algorithms are increasing. To address these issues, the academic and industrial society have already made extensive contributions and studies into several cutting-edge research areas and contents, including visual signal representation mechanism of video big data, compact visual information expression, video signal restoration and reconstruction, high-level and low-level vision fusion methods, and their hardware implementations. Based on fundamental theories in discrete signal processing, the active research topics as well as the corresponding state-of-the-art methodologies in the field of video processing and compression are systematically reviewed and analyzed. A comprehensive review of research topics, namely, statistical prior model-based video data representation learning and its processing methods, deep network-based video processing and compression solutions, video coding techniques, and video compression standardization process is provided. More importantly, the challenges of these research areas, the future developing tendency, the state-of-the-art approach as well as the standardization process are also provided from top to bottom. Specifically, the video processing algorithms, including model-based and deep learning based video super-resolution and video restoration solutions are initially reviewed. The video super-resolution contains spatial super-resolution and temporal super-resolution methods. The video restoration focuses on video deblurring and video deraining. The prior model based approaches and neural approaches are reviewed and compared. Subsequently, this paper presents the review of video compression methods from two aspects, namely, conventional coding tool development and learning-based video coding approaches. The former focuses on the modular improvements on predictive coding, transform and quantization, filtering, and entropy coding. With the development of multiple next-generation video coding standards, the scope and depth for the coding tool research in conventional hybrid coding framework are extensively broadened. The latter introduces the deep learning based video coding methods, not only for hybrid coding framework but also for end-to-end coding framework. Deep neural network based coding would definitely become the next jump of high-dimensional multimedia signal coding. For both parts, the detailed technology and standardization are described to shape the overall development of video compression. In addition, the extensive comparative study on these areas between oversea community and domestic community is conducted and analyzed, providing the evidence for the difference and similarity in the current situation. Finally, the future work on theoretical and application studies in video processing and compression is envisioned. In particular, the research between high quality visual effects and high efficiency visual representation would not be separate areas. The fusion of brain-like visual system and encoding mechanism for video processing and compression is a key direction of future research.

Key words: multimedia technology; video signal processing; video compression; artificial intelligence; deep learning

0 引言

自1940年代计算机诞生以来,经过数次革命性的发展,计算的概念从单纯解决科学与工程中的数学问题扩展到数据处理和智能多媒体计算领域。视频作为承载了海量非结构化数据和应用最广泛的多媒体数据格式已与人们的生活密不可分,是人类获得信息的重要途径之一。与视频相关的各种

技术问题都得到了广泛关注。在视频技术的发展过程中,从模拟到数字的转换是一次伟大的技术革新,由此带来了数字信号处理等理论的构建,为该领域的持续发展奠定了基础。继数字化、高清化之后,视频技术正在经历由超高清和智能化等新一轮技术革新带来的跨越式发展。

1940—1950年是视频技术基础理论发展的关键期。香农提出的《通信中的数学理论》和信息论三大定理奠定了信息论与编码研究的理论基础,其

中,信源编码和信道编码理论至今依然对高效通信系统具有巨大影响。20世纪70年代是视频技术研究的快速发展时期。1979年,Netravali等人发表了《Motion—Compensated Transform Coding》,标志着基于块的预测变换混合编码框架正式建立。随着电荷耦合器件(charge coupled device, CCD)和互补金属氧化物半导体(complementary metal oxide semiconductor, CMOS)等成像技术的发展,视频处理领域具备了坚实的硬件基础。

20世纪80年代是视频处理领域发展的萌芽期。Tsai和Huang(1984)开创了视频超分辨率研究。在神经网络研究方面,提出了多层神经网络及其反向传播算法,开启了视频处理与神经网络研究的热潮。在视频压缩领域,国际电报电话咨询委员会(Consultative Committee of International Telegraph and Telephone, CCITT)于1984年颁布了首个视频压缩国际标准H.120。

20世纪90年代至21世纪初,针对视频处理技术,包括基于对象的统计先验模型构建方法及在视频重建和分辨率提升等领域的应用,主流的研究手段是建立高秩空间与低秩空间的局部线性嵌入表示。这段时期是视频压缩标准研究的黄金时代,国际标准化组织(International Organization for Standardization, ISO)、国际电工委员会(International Electrotechnical Commission, IEC)和国际电信联盟电信标准分局(International Telecommunication Union—Telecommunication Standardization Sector, ITU-T)先后发布了MPEG(moving picture expert group)系列标准,H.261、H.263和H.263+等国际标准。2006年以后,随着算力和成对标注数据规模的提升,深度学习逐渐成为主流方法,深度神经网络在视频处理和视频压缩任务中均大幅超越传统的基于统计模型规则的方法,并陆续提出了多种深度网络模型结构及其优化方法,将视频处理和压缩任务从局部优化迁移到端到端整体优化。

随着智能化与信息化时代的来临,以数字视频为核心内容的多媒体技术蓬勃发展,移动互联网的飞速演进为数字视频创造了更广阔的应用空间。视频技术是智慧城市、视觉通讯等领域的核心技术,也是智能交通、智慧医疗和远程教育等行业应用中不可或缺的关键模块。据统计,视频相关产业规模已超过数万亿,是促进国家产业转型与升级、关乎国计

民生的重要支柱。

视频行业的核心需求是将视频实时高清地呈现给用户,技术基础是视频处理和压缩。一方面,视频在获取过程中不可避免地会引入信号失真,为提高视频用户体验,对视频进行实时高效的处理,提高视频质量非常必要。另一方面,产业规模的扩大和应用场景的拓宽衍生了海量非结构化视频数据,高效地传输和存储是确保视频产业健康发展的基础,而视频压缩技术正是实现视频紧致表达的关键。面向视频产业的发展需求,本文将归纳视频压缩和处理的国内外发展现状,分析该领域的关键技术,并展望未来发展空间。

1 国际研究现状

1.1 视频处理技术

1.1.1 视频超分辨

视频超分辨率算法分为传统基于信号处理的方法和基于深度学习的方法。在传统基于信号处理的方法中,Liu和Sun(2013)提出一种贝叶斯方法可以同时估计底层运动、模糊核和噪声水平,并重构高分辨率帧。Ma等人(2015)采用期望最大化(expectation maximization, EM)方法估计模糊核,并指导高分辨率帧的重建。这类方法由于采用固定的解决方案,不能适应视频中的各种场景。在基于深度学习的超分辨率算法中,出现了许多基于深度神经网络的超分辨率方法,包括基于卷积神经网络(convolutional neural networks, CNN)、生成对抗网络(generative adversarial network, GAN)和递归神经网络(recurrent neural network, RNN)的视频超分辨方法。

从是否使用对齐类方法的角度介绍国际视频超分辨的研究现状。

1) 对齐超分辨算法。国外视频超分辨率对齐方法大多采用运动补偿和运动估计技术。运动估计的目的是提取帧间的运动信息,而运动补偿是根据帧间的运动信息进行帧间的矫正操作,使一帧与另一帧对齐。大部分运动估计技术是使用光流方法(Dosovitskiy等,2015)完成的。Kappeler等人(2016)提出的视频超分网络(video super resolution network, VSRnet)由3个卷积层组成,除最后一个外,每个卷积层后面都有一个非线性激活单元(rectified linear unit, ReLU)。VSRnet使用多个连续帧,

这些连续帧都是补偿帧。目标帧与补偿帧之间的运动信息由 Druleas 算法计算得出 (Drulea 和 Nedevschi 2011)。此外, VSRnet 提出了滤波器对称增强 (filter symmetric enhancement, FSE) 机制和自适应运动补偿机制, 分别用于加速训练和减少冗余的补偿帧影响, 从而提高视频的超分辨率性能。Caballero 等人 (2017) 提出的视频子像素卷积网络 (video efficient sub-pixel convolutional neural network, VESPCN) 设计了一种空间运动补偿变压器 (motion compensation transformer, MCT) 模块用于运动估计和补偿, 并将补偿后的帧送入一系列卷积层进行特征提取和融合, 最后通过亚像素卷积层得到超分辨率结果。MCT 模块采用卷积神经网络提取运动信息, 进行运动补偿, 使用由粗到细的方法来计算图像序列的光流。Sajjadi 等人 (2018) 提出了 FRVSR (frame-recurrent video super-resolution) 网络, 主要特点在于帧间的对齐方式, 该网络不会直接矫正目标帧的前一帧 (低分辨率帧), 而是扭曲矫正前一帧对应的高分辨率帧。受反投影算法 (Haris 等, 2018; Irani 和 Peleg, 1991, 1993) 的启发, Haris 等人 (2019) 提出了递归反投影网络 (recursive back projection network, RBPN), 由特征提取模块、投影模块和重构模块组成。特征提取模块包含提取目标帧特征和提取相邻帧特征两个操作, 并计算从相邻帧到目标帧的光流, 然后进行隐式地对齐。Bare 等人 (2019) 提出实时视频超分辨率 (real time video super resolution, RTVSR) 通过运动卷积核估计网络, 使用编解码结构估计目标框架与相邻帧之间的运动, 产生一对与当前目标帧和相邻帧对应的 1 维卷积核, 然后利用估计的卷积核对相邻帧进行矫正, 使其与目标帧对齐。

2) 非对齐超分辨算法。与对齐方法不同, 未对齐方法在重建前不进行帧对齐, 分为空间未对齐和时空未对齐两种。空间非对齐方法不需要帧间的运动估计和运动补偿等对齐操作, 如视频超分辨残差网络 (video super resolution residual network, VSRResNet) (Lucas 等, 2019)。时空未对齐方法的特点是同时利用输入视频中的时空信息进行超分辨任务, 如动态滤波器 (dynamic upsampling filter, DUF) 和 3 维超分网络 (3-dimensional super resolution network, 3DSRnet)。VSRResNet 通过对抗性训练解决视频超分辨率问题, 由鉴别器决定输出的是生成图像还是真实图像, 使生成器产生更接近真实图像的

结果。DUF (Jo 等, 2018) 采用动态滤波器结构, 执行滤波和上采样操作, 与 3 维卷积学习的时空信息相结合, 避免了运动估计和运动补偿。为了增强超分辨率结果的高频细节, DUF 使用网络估计目标帧的残差图, 一帧的最终输出结果是残差映射和经过动态上采样滤波器处理的帧的总和。此外, DUF 提出一种基于时间轴的视频数据增强, 对不同时间间隔内按顺序或倒序采样视频帧, 得到不同运动速度和方向的视频。3DSRnet (Kim 等, 2019) 使用 3D 卷积提取连续视频帧之间的时空信息用于视频超分辨率任务。随着网络的深入, 3 维卷积后的特征图的深度也会变浅。为了保留深度和时序信息, 3DSRNet 采用扩展操作, 即在连续帧的开始和结束分别增加一帧。同时提出一种实际应用中场景变化的方法, 采用浅分类网络判断输入的连续帧, 有效解决了场景变化导致的性能下降问题。非对齐网络比较简单, 超分辨重建效果也很有限, 说明在利用多帧进行视频超分辨的过程中, 帧间信息融合非常重要, 对帧间融合的方法尚需进一步研究。

3) 视频插帧。在视频插帧研究方面, Meyer 等人 (2018) 基于相位的运动表示提出了 PhaseNet 结构, 对运动模糊或闪动变化产生了比较鲁棒的结果, 但不能有效重建详细的纹理。核方法在 2017 年是一个研究热点, Niklaus 等人 (2017a, b) 连续在 ICCV (IEEE International Conference on Computer Vision) 和 CVPR (IEEE Conference of Computer Vision and Pattern Recognition) 会议上提出了基于核的方法, 为每个像素估计一个自适应卷积核。基于核的方法可以产生合理的结果, 但是不能处理大运动场景。为了有效利用运动信息, Niklaus 和 Liu (2018) 使用前向变形技术从连续的两帧中生成了中间帧, 然而前向扭曲矫正存在像素缺失和重叠。因此, 大多数基于流的算法都是基于反向扭曲矫正的。为了使用反向扭曲, 需要估计中间运动 (即中间帧的运动向量)。

1.1.2 视频恢复

视频恢复是视频处理的关键任务之一, 对视频主客观质量提升和下游视觉分析任务具有至关重要的作用。从成像设备捕捉到的降质图像中恢复出富有细节的清晰场景图像是一个值得长期研究的问题, 降质模型包括模糊、噪声和天气效应等。Garg 和 Nayar (2005, 2007) 最早尝试从视频中去除雨痕, 并提出直接增加曝光时间或减小摄像机景深的方

法,但无法处理靠近摄像机的快速移动物体,并且如果控制视频质量不发生显著下降,则无法通过此方法调整摄像机设置(Tripathi 和 Mukhopadhyay, 2014)。在过去的几年中,用于从静态/动态场景的视频恢复算法已经探索并形式化描述了降质模型的很多固有特性。这些算法主要分为4类:基于时域的算法、基于频域的算法、基于低秩和稀疏性的算法以及基于深度学习的算法。前3类算法使用手工设计的框架模拟降质环境,视为基于模型驱动的方法。第4类算法遵循数据驱动的方法,可以从预先收集的成对训练数据中自动学习特征(Wei 等 2017; Yang 等 2018)。

1) 基于时域的方法。Garg 和 Nayar 是该领域的先驱,从时域方面对雨痕对成像系统的视觉影响进行了全面分析(Garg 和 Nayar 2004),设计了视频雨痕检测和去除算法,利用时空相关模型描述雨水的动态特性,基于物理的运动模糊模型解释雨的光度特性,通过利用连续帧之间的差异消除雨痕(Tripathi 和 Mukhopadhyay, 2014)。Park 和 Lee (2008) 提出估计像素的强度,通过卡尔曼滤波器递归恢复视频,在具有固定背景的视频中表现良好。Brewer 和 Liu (2008) 借鉴成像过程中的光学和物理特性,提出首先识别出显示短时强度峰值的受雨影响区域,然后用连续帧中的平均值替换受雨影响的像素,能够区分由雨引起的灰度变化与由场景运动引起的灰度变化,然而对检测重叠的形状各异的雨痕造成的大雨并不适用。为了同时处理动态背景和相机运动,Bossu 等人(2011) 利用高斯混合模型(Gaussian mixture model, GMM) 和几何矩估计雨痕的方向直方图。受贝叶斯理论启发, Tripathi 和 Mukhopadhyay (2011) 依靠降雨的时间特性建立雨痕去除概率模型。由于受雨影响的像素和无雨像素的灰度变化因波形的对称性而异,因此使用灰度波动范围和扩展不对称性两个统计特征区分雨痕与无雨运动物体。由于没有关于雨滴形状和大小的任何假设,因此该方法对雨天条件是鲁棒的。为了进一步减少连续帧的使用, Tripathi 和 Mukhopadhyay (2012) 转向采用时空过程,与依靠降雨的时间特性建立的雨痕去除概率模型(Tripathi 和 Mukhopadhyay 2011) 相比,检测准确度较低,但具有更好的感知质量。

2) 基于频域的方法。频率特性也是视频恢复研究的重点内容。Barnum 等人(2007, 2010) 展示了

一种基于时空频率物理和统计模型的全局检测雨/雪的方法,利用模糊的高斯模型近似估计了雨滴和频域滤波导致的模糊效果,模拟雨/雪导致的能见度下降。这个想法仍然适用于具有场景运动和摄像机运动的视频,并且可以有效地分析重复的降雨模式。但是,模糊的高斯模型不能总是覆盖不够锐利的雨痕。此外,当雨痕的频率成分不按顺序排列时,基于频率的检测方式常常会出错(Tripathi 和 Mukhopadhyay 2011)。

3) 基于低秩和稀疏的方法。在近10年的研究中,研究人员对低秩和稀疏方法进行广泛研究以实现视频恢复。为了克服基于朴素最大值后验方法的局限性,同时避免启发式边缘预测步骤,一些学者设计了有效的图像先验,如归一化稀疏性先验(Krishnan 等 2011) 和 L0 正则先验(Zoran 和 Weiss 2011) 以去除视频中的模糊效应。Chen 和 Hsu (2013) 从雨痕的相似性和重复性出发,提出了一个从矩阵到张量结构的低秩模型,以捕获时空相关的雨痕特性。为了处理高度动态的场景(Garg 和 Nayar 2004; Park 和 Lee 2008), Chen 和 Hsu (2013) 设计了基于动态场景运动的分割算法,利用光度和色度约束进行雨水检测以及雨痕动态特性建模,可以在像素复原过程中自适应地利用空间和时间信息,但是仍然无法很好地适应相机抖动(Ren 等, 2017)。Kim 等人(2015) 提出从当前帧中减去时域上对齐的帧以获得初始降雨图,然后通过支持向量机将其分解为两种基本矢量。为了处理有显著噪声和离群值的低质视频, Hu 等人(2018b) 改进了现有方法以提升重建视频处理离群值时的鲁棒性。

4) 基于深度网络的方法。Chen 等人(2018a) 提出了一个 CNN 框架用于视频去雨,可以处理带有不透明雨痕遮挡的暴雨。该方法采用超像素作为基本处理单元,能够对具有高度复杂和动态场景的视频内容进行对齐和遮挡消除。

1.2 视频压缩技术

1.2.1 混合框架中的传统技术

传统视频编码采用基于块划分的混合编码框架,包括帧内预测、帧间预测、变换、量化、熵编码和环路滤波等技术模块。这些模块经过几十年的发展已逐渐成熟。

1) 编码块划分结构。基于块的编码结构一直是传统视频编码框架的核心。先进视频编码(ad-

vanced video coding, H. 264/AVC) 采用基于 16×16 宏块的划分结构。为了进一步提升块划分的灵活性和自适应性, 高效视频编码 (high efficiency video coding, H. 265/HEVC) 中采用了四叉树划分结构。在 HEVC 与 VVC (versatile video coding) 两代标准技术的间隔期, 四叉树联合二叉树的划分结构得到广泛研究。在 VVC 标准中, 采纳了高通公司 (Chen 等人, 2018a) 提出的四叉树、三叉树和二叉树联合的多级划分方式, 有效提高了编码框架的灵活性。

2) 帧内预测。帧内预测主要是利用邻近块之间的空域相关性来消除空域冗余。具体来讲, 帧内预测利用当前帧中已重构的像素, 导出当前块的预测值。预测块的导出规则是帧内预测的核心。HEVC 中的帧内预测包含平面 (planar) 预测、直流 (direct current, DC) 预测和角度预测等模式。为了提升压缩性能, Said 等人 (2016) 提出基于位置的帧内预测组合 (position dependent intra prediction combination, PDPC), 使用边界参考像素以及滤波后的边界参考像素对传统平面预测、直流预测和角度预测模式的结果进行修正, 以降低预测失真。

3) 帧间预测。在帧间预测中, 运动矢量编码消耗的比特数制约着压缩性能, 如何导出高质量、低开销的运动矢量一直是行业关注的焦点。Zhang 等人 (2018a) 提出一种运动矢量精度的自适应算法, 与 HEVC 中固定 $1/4$ 像素的运动矢量精度不同, 该方法为运动矢量设置了整像素、 4 像素和 $1/4$ 像素等多种可选项, 通过率失真优化的方法进行编码端决策, 提升运动矢量的表示能力。Xiu 等人 (2018) 提出一种高级时域运动矢量预测技术 (advanced temporal motion vector prediction, ATMVP), 对编码单元进行细分, 针对每个编码单元在每个参考方向上导出多组运动矢量预测信息, 通过这些运动信息对当前编码单元内的子块分别进行运动补偿, 提升预测的准确度。

4) 变换。变换的主要作用是去除残差信号的统计相关性。HEVC 中对残差信号进行一次变换, 主要包含整数离散余弦变换和整数离散正弦变换两种变换方式。Koo 等人 (2019) 在 VVC 中提出一种二次变换技术——低频非分离变换 (low-frequency non-separable transform, LFNST), 可以使残差信号能量进一步集中, 有效减少残差编码消耗的比特数。

5) 量化。量化是变换后对变换系数的处理, 也

是压缩失真的主要来源。将变换系数划分为不同的区间, 每个区间用一个标号代表, 标号数量小于原始数据量, 由此达到压缩的目的。Schwarz 等人 (2018) 提出一种上下文依赖量化 (dependent quantization, DQ) 方式, 此方法与 HEVC 中常用的独立标量量化相比, 将变换系数在向量空间中进行更加密集地压缩, 可以有效减小原始变换系数与量化后系数之间的差值, 从而降低压缩失真。

6) 熵编码。熵编码模块用于去除统计域的冗余, 将编码控制数据、量化变换系数、帧内预测数据、运动数据和滤波器控制数据等编码为二进制数进行存储和传输。不同于传统的系数组概念, Lyu 等人 (2020) 提出一种基于扫描区域的系数编码方案, 使用一组常数控制量化系数非零的位置, 使用率失真优化方法选择最优扫描区域, 码率和失真之间达到较好的平衡。同时, 编码区域内非零系数时, 采取从右下到左上的反 Z 扫描, 采用分层编码和多套上下文提升压缩效率, 上下文模型根据系数在扫描区域的相对位置、扫描区域面积和通道进行分组。

7) 环路滤波。滤波是去除压缩失真的关键技术, 可以明显提高重建视频的主客观质量, 提高视频压缩的效率。在国际标准 VVC 中, Karczewicz 等人 (2017) 提出基于几何变换的自适应环路滤波技术, 根据图像亮度分量的梯度信息进行聚类, 对每种类别的像素点使用临近位置的点进行自适应滤波, 从而最小化原始图像与重建图像之间的均方误差。除此之外, Misra 等人 (2019) 探究视频不同分量之间的相关性, 利用维纳滤波原理, 提出跨分量的自适应环路滤波技术, 利用结构信息保存更加完好的亮度分量, 对色度分量进行滤波, 大幅提升了色度分量的压缩性能。

1.2.2 混合框架中的深度学习技术

随着混合编码框架的发展完善, 压缩效率不断提高, 已成为支撑视频产业不断繁荣的基础。但传统视频编码遇到了性能提升的瓶颈, 原因在于各个模块一般基于人工设计的统计先验模型, 而实际应用中的视频通常内容多样, 无法通过简单设计的模块高效表达。而深度学习技术可以从大量数据中学习复杂特征, 拟合复杂的非线性函数, 进行精准预测。因此, 基于深度学习技术设计和优化视频编码模块具有巨大潜力。

1) 帧内预测。Pfaff 等人 (2018) 采用全连接网

络进行帧内预测,利用当前块的上下文信息为不同的预测模式训练不同的网络,并且根据上下文,额外训练一个网络预测不同模式的概率。

2) 帧间预测。基于深度学习的帧间预测主要研究如何高效利用视频帧间的时域相关性以及如何将时域与空域进行融合。Choi 和 Bajić (2020) 将基于神经网络的插帧技术用于视频编码的帧间预测,使用已经编码的帧通过神经网络预测出当前帧,并将该预测帧作为一个虚拟参考帧使用,从而高效地去除时间冗余。

3) 上下采样。随着视频分辨率的提高,数据量成倍增长,给视频传输系统提出了巨大挑战。当传输带宽受到限制时,通常的做法是降低编码前的视频分辨率,并提高解码后的视频分辨率。这种做法称为基于下采样和上采样的编码技术。Afonso 等人 (2018) 提出了一种联合空间和像素值进行下采样再使用卷积神经网络进行上采样的方法,在编码器端,通过传统方法实现空间下采样,用支持向量机决定是否对每帧执行下采样。在解码器端用卷积网络将解码出的低分辨率视频上采样到原始分辨率。

4) 熵编码。深度学习熵编码研究始于2017年,Puri 等人 (2017) 提出使用卷积神经网络估计不同变换模式的概率分布,提高了变换模式编码的效率。随后,Pfaff 等人 (2018) 进一步在 VVC 标准技术研究过程中提出使用全连接神经网络分析视频的纹理特征,估计帧内不同预测模式的概率分布,提高了帧内预测模式的编码效率。

5) 滤波。基于神经网络的滤波方法可以显著提高编码效率。滤波方法根据是否影响后续编码分为环内滤波技术和后处理技术。国际上对环内滤波技术进行了广泛研究,Park 和 Kim (2016) 提出了一个3层卷积神经网络用于压缩视频环路滤波,针对高低码率训练了两个神经网络模型并根据量化步长选择使用不同的模型。Kang 等人 (2017) 提出使用块划分信息提高针对 HEVC 编码视频的环路滤波效率。后处理方面,Kim 等人 (2020) 指出在实际应用中压缩视频的量化步长通常是未知的,为此设计了神经网络用于估计压缩视频的量化步长,并据此选择对应的神经网络模型进行滤波,明显提高了后处理效率。Huang 等人 (2020) 提出使用更大的图像块进行训练以提高后处理神经网络的性能,并使用块划分和帧内预测模式作为辅助信息,进一步增强

后处理的效果。

6) 编码优化。基于深度学习的编码工具又称为编码优化工具,作用是编码加速和码率控制等,目标是提高编码效率。国际上对编码优化的研究主要集中于将深度网络模型与编码单元划分决策相结合。Kim 等人 (2019)、Paul 等人 (2020)、Galpin 等人 (2019) 和 Su 等人 (2019a) 使用深度神经网络加快视频编码过程中的 CU (coding unit) 划分过程。Su 等人 (2019b) 使用神经网络快速选择变换核以加速 AV1 的编码过程。

1.2.3 端到端深度学习视频压缩

从2017年开始,国际上越来越多的研究人员开始致力于构建端到端的深度学习视频压缩方案。该框架所有模块都是基于深度神经网络实现,可以直接端到端优化率失真目标函数,更容易实现全局最优。端到端视频压缩根据应用场景分为两类。

第1类面向随机切入场景,主要基于帧内插的方式进行运动补偿。Wu 等人 (2018) 提出了第一个基于帧内插模型的视频压缩方案。该方案的关键帧(I帧)通过端到端图像压缩模型进行压缩,剩下的帧(B帧)通过基于帧内插的模型以分层的方式进行压缩,框架性能与 H.264 相当。Djelouah 等人 (2019) 提出一个更加高效的基于帧内插的视频压缩方案。首先使用光流网络提取前后向参考帧相对于当前帧的原始运动场,然后将原始运动场与对齐的参考帧(warped frame)和原始帧一起输入自编码器进行压缩,解码得到的运动场和融合系数(blending coefficients)通过内插方式得到预测帧。然后在隐空间(latent space)进行残差的提取和编码。该方案在峰值信噪比(peak-signal-noise-ratim, PSNR)上的压缩性能与 H.265 相当。Park 和 Kim (2019) 使用基于帧内插的双向编码结构,并且在编码端和解码端同时引入运动估计网络,从而不再传输运动信息,在低分辨率视频上的压缩性能与 H.265 相当。Yang 等人 (2020a) 提出包含3个质量层的分层双向预测视频压缩方案,结合使用回归的质量增强网络,取得了与 H.265 相近的压缩性能。Yilmaz 和 Tekalp (2020) 提出一个端到端优化的分层双向预测视频压缩方案,在 PSNR 上的压缩性能接近 H.265。里斯本大学 Pessoa 等人 (2020) 的方案将视频压缩问题看成一个“空一时”域自编码器的率失真优化问题,从而避免了显式的运动估计和补偿。Habibian

等人(2019)提出基于3D自编码器的模型,结合基于自回归先验的熵模型同时压缩一段连续的视频帧,整个系统针对率失真损失函数进行端到端优化,在多尺度结构相似度(multi-scale structural similarity, MS-SSIM)上的压缩性能接近H.265。

第2类面向低延时场景,主要基于帧外插的方式进行运动补偿。2018年,Rippel等人(2019)提出首个面向低延时的端到端视频压缩方案。在解码器端使用隐状态(latent state)保存历史帧的信息,并使用生成网络生成运动场和残差的重建,设计了空域码率控制方法减小误差累积,在MS-SSIM上性能与H.265相当。Agustsson等人(2020)使用自编码器基于前一个参考帧和当前原始帧提取原始的尺度空间流(scale-space flow),并进行压缩编码,解码得到的尺度空间流结合参考帧的高斯金字塔通过三线性插值(trilinear interpolation)操作得到当前帧的预测,剩下的残差使用另外一个自编码器进行压缩,该方案的压缩性能与H.265相当。Yang等人(2020b)提出使用回归的自编码器和熵模型进行视频压缩,取得了与H.265相近的压缩性能。Golinski等人(2020)提出一个基于自编码器的端到端视频压缩方案,在解码端使用反馈回归模块将提取的历史隐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。

1.2.4 标准进展

国际标准组织ITU和ISO制定了一系列基于混合编码框架的视频压缩标准。2015年,运动图像专家组和视频压缩专家组联合成立JVET(joint video exploration/expert team)工作组,开启了新一代视频压缩标准制定工作,参与单位包括高通、联发科、三星、华为和腾讯等公司,高校和科研机构包括德国范霍夫海因里希赫兹研究所、北京大学、中国科技大学、浙江大学和亚琛大学等。2018年4月,JVET正式将下一代视频压缩标准命名为多用途视频编码(versatile video coding,VVC),并于2020年7月正式发布标准草案,在PSNR指标下,压缩效率相比于上一代国际标准H.265/HEVC提升约36.6%。VVC标准针对混合编码框架的每一个模块都设计了全新的压缩技术,显著提升了压缩效率。

同时,MPEG组织中的三星、华为、高通和Divideon等公司牵头制定了MPEG-5 EVC(essential

video coding)标准,主要面向超高清、高动态范围和广色域视频内容。EVC标准的制定方法尝试了不同于以前的标准化制定过程。1)选择专利期限超过20年的技术或有免版税声明的技术来定义该标准;2)在基本工具集上定义了一些其他工具,每个工具在压缩性能方面都有显著改进;3)每一个附加的工具都是隔离的,可以独立于其他工具进行开关控制;4)鼓励技术赞助者提交与专利许可或出版有关的自愿声明;5)定义了分析机制,以便允许不同模式可以包含不同的工具。

此外,为了适应互联网应用以及设计开源和专利政策友好的编解码器,多家科技巨头联合成立了开放媒体联盟(alliance for open-media,AOM),致力于推广和研发多媒体的视频编解码技术,为下一代多媒体体验创造新机遇。AOM联盟于2018年初正式推出了AV1视频压缩标准。AOM联盟多数成员单位是与视频行业紧密相连的互联网公司、硬件设备厂商、内容供应商和主流浏览器厂商等,行业优势使得AV1基本做到了主流平台的全覆盖,形成从内容端、产品端到芯片端的完整生态链。

1.3 多视点/立体视频压缩

立体视频在近年来得到了广泛应用。3DoF(degree of freedom)是全景视频,3DoF+视频是在全景之外,还支持用户在3个空间维度上有限范围的变化,一个典型的例子就是坐在椅子上观看场景,允许头部在一定范围内运动。增加运动自由度后,更符合人感知场景的真实体验,同时带来双目立体视差。

从H.264/AVC时期就开始有多视点视频编码(multiview video coding,MVC)的研究,发展出了一些视间预测的技术,使得多个视点之间可以相互参考。3D-HEVC是一个具有标志性意义的3维视频编码标准,采纳了许多可以有效提升编码效率的关键技术。纹理图编码工具包括邻块视差矢量导出(neighbor block derived vector,NBDV)、视间残差预测、视点合成预测(view synthesis prediction,VSP)等。深度图编码工具包括视差预测、深度建模模式(depth modeling model,DMM)等。在2018年的3DTV会议上,Fachada等人(2018)提出一种基于深度图的虚拟视点合成技术,可以运用于6自由度(6DoF)和360视频(3DoF+)的立体全景视频中,通过增加参考视图的数量克服了诸如遮挡、相机轴的切向曲面和低质量深度图中的瑕疵等问题。

2 国内研究进展

2.1 深度学习视频处理技术

2.1.1 视频超分辨

1) 对齐超分辨算法。对齐视频超分辨算法通过网络提取运动信息,使相邻帧与目标帧进行对齐,然后再进行后续重构。该方法主要采用运动补偿和可变形卷积两种常用的帧间对齐技术。

Tao 等人(2017)提出面向细节增强的深度视频超分网络(detail-revealing deep video super-resolution, DRVSR),提出一种亚像素运动补偿层(sub-pixel motion compensation, SPMC),可以根据估计的光流信息对相邻输入帧同时进行上采样和运动补偿操作。此外,融合模块采用卷积长短期记忆单元(convolutional long-short term memory, ConvLSTM)模块(Shi 等,2015)来处理时空信息。Wang 等人(2018b)提出光流超分的视频分辨率提升网络(super-resolving optical flow for video super resolution, SOFVSR),采用一种由粗到细包含光流网络(optical flow network, OFNet)的方法估计帧间的光流,得到高分辨率光流。然后通过转换机制(space-to-depth)将高分辨率(high resolution, HR)光流转换为低分辨率(low resolution, LR)光流。相邻帧通过 LR 光流矫正,与目标帧对齐。Xue 等人(2019)提出面向任务驱动的光流网络(task-oriented flow, TOFlow),将光流估计网络与重构网络相结合,共同进行训练,采用空域金字塔网络(spatial pyramid network, SpyNet)(Ranjan 和 Black, 2017)作为光流估计网络,然后采用空间变换网络根据计算出的光流对相邻帧进行变形。然而,使用最先进的光流估计网络也不容易获得足够高质量的运动估计;使用精确的运动场,运动补偿也会在图像结构周围产生伪影,这些伪影可能会传播到最终重建的 HR 帧中。

可变形卷积网络最早由 Dai 等人(2017a)提出,随后 Zhu 等人(2019b)提出了改进版本。在普通的卷积神经网络中,惯例是对一层中的所有核使用固定的大小,这限制了网络建模几何变换的能力。而可变形卷积则可以缓解这一限制。将偏移量加入到传统的卷积核中,得到可变形的卷积核,再与输入的特征图进行卷积,得到输出的特征图。采用可变形卷积的超分辨方法主要有香港中文大学提出的增强

可变卷积网络(enhanced deformable convolutional networks, EDVR)(Wang 等,2019d)、华南理工大学提出的可变形非局部网络(deformable non local network, DNLN)(Wang 等,2019a)以及清华大学和美国罗切斯特大学提出的时间形变对齐网络(time deformable alignment network, TDAN)(Tian 等,2020)。EDVR 方法由对齐模块、融合模块和重建模块3部分组成。对齐模块对输入帧进行对齐,然后通过融合模块将对齐后的帧进行融合,并将融合后的结果送入重构模块细化特征,通过上采样得到一幅 HR 图像,称为残差图像。最后将残差图像添加到直接上采样目标帧中得到最终结果。DNLN 方法设计了基于可变形卷积和非局部网络(Wang 等,2018d)的对齐模块和非局部注意模块。对齐模块使用原始可变形卷积内的层次化特征融合模块(hierarchical feature fusion block, HFFB)(Hui 等,2020)生成卷积参数。此外,DNLN 通过级联的方式利用了多个可变形卷积,使得帧间对齐更加精确。非局部注意模块以目标帧和对齐的相邻帧为输入,通过非局部操作生成引导注意的特征。TDAN 方法对目标帧和邻近帧进行形变卷积,得到相应的偏移量。然后根据偏移量对相邻帧进行变形,使其与目标帧对齐。

2) 非对齐超分辨算法。与对齐分辨方法不同,非对齐分辨方法在重建前不进行帧对齐。空间非对齐方法直接将输入帧输入到2维卷积网络中进行空间特征提取、融合和超分辨率操作,不进行帧间的运动估计和运动补偿等对齐操作。

循环卷积神经网络(recurrent convolutional neural network, RCNN)在处理时序数据(如自然语言、视频和音频等)时具有很强的时间依赖性。一种直接的方法是使用 RCNN 处理视频序列。基于这一思想,提出了双向循环卷积网络(bidirectional recurrent convolutional network, BRCN)(Huang 等,2018)、残差递归卷积网络(residual recursive convolution network, RRCN)(Li 等,2019a)、残差可变速空卷积(residual invertible spatio-temporal network, RISTN)(Zhu 等,2019b)和时空卷积网络(spatial-temporal convolution network, STCN)(Guo 和 Chao,2017)等 RCNN 方法实现视频超分辨率。

BRCN 利用 RCNN 适合处理序列数据的优势处理视频中连续帧之间的依赖关系,包含前向和后向两个网络,两个网络结构相似但处理顺序不同。前

向网络中, 隐含层各节点的输入有 3 个, 分别是前一层当前时刻 i 、上一时刻 $i-1$ 的输出和当前层中前一个节点的输出, 分别用于提取空间相关性和连续帧之间的时间相关性。RRCN 是一个学习残差图像的双向递归神经网络, 提出一种非同步的全循环卷积网络, 非同步指多个连续视频帧输入但只有中间帧需要超分辨率重建。采用局部—全局—全变量组合 (global local total variation, GLG-TV) 方法 (Drulea 和 Nedevschi 2011) 对目标帧及其相邻帧进行运动估计和补偿。RISTN 基于可逆网络 (Jacobsen 等, 2018) 设计了残差可逆块用于提取空间信息的有效视频帧, 利用 LSTM 提取时空特性, 由时间模块、空间模块和重构模块 3 部分组成。空间模块主要由多个平行残差可逆单元组成, 其输出作为时态模块的输入。在时间模块中, 提取时空信息后, 采用稀疏融合策略有选择地融合特征。最后, 在重构模块中通过反卷积重建目标帧得到高分辨结果。STCN 是一种不需要运动估计和运动补偿的端到端视频超分辨率方法, 利用 LSTM 算法提取帧内的时间信息 (Hochreiter 和 Schmidhuber, 1997), 与 RISTN 类似, 该网络由空间模块、时间模块和重构模块 3 部分组成。武汉大学、哈尔滨工业大学和鹏城实验室联合提出的渐进非局部残差融合 (progressive fusion of non local network, PFNL) (Yi 等 2019) 采用非局部残差提取时空特征, 采用渐进融合残差块 (progressive fusion of residual block, PFRB) 融合时空特征。最后, 通过跳接操作将双三次插值向上采样的输入帧和亚像素卷积层的输出进行接合, 得到最终的超分辨率结果。复旦大学提出一种网络, 与普通的运动估计和运动补偿技术不同, 将低分辨率的未对齐视频帧和前一帧的高分辨率输出直接作为网络的输入以恢复高频细节, 保持时间一致性 (Yan 等 2019a)。

2.1.2 视频恢复

1) 基于时域的方法。Zhang 和 Patel (2017) 嵌入雨的时域和色度特性, 并利用 K-means 聚类从视频中区分背景和雨痕。这个方法在处理小雨和大雨以及对焦/不对焦雨时效果很好。然而, 由于背景的时间平均, 该方法经常趋于使图像模糊。Zhao 等人 (2008) 利用雨痕的时空特征设计了用于检测和去除降雨的直方图模型, 该模型嵌入了简洁、低复杂度的 K-means 聚类算法 (Zhang 等 2006)。

2) 基于低秩和稀疏的方法。考虑到大雨和动

态场景, Ren 等人 (2017) 将雨痕分为稀疏层和稠密层, 并将运动物体和稀疏降雨条纹的检测形式转化为多标签马尔可夫随机场, 将稠密的雨痕建模高斯分布。Jiang 等人 (2017, 2019) 提出一种新颖的基于张量的视频雨痕去除方法, 全面分析雨痕和无雨视频的判别固有特性。具体来说, 雨滴沿雨滴方向稀疏且平滑, 干净的视频沿雨水垂直的方向具有平滑度, 沿时间方向具有全局和局部相关性。Wei 等人 (2017) 将雨层编码为基于块的高斯混合, 通过整合运动对象的时空平滑配置和背景场景的低秩结构, 提出一种简洁的模型去除雨痕, 能够适应更大范围的降雨变化。

受 Zhang 和 Patel (2017) 工作的启发, Li 等人 (2018b) 考虑了视频中雨痕的两个内在特征, 即重复的局部图案稀疏地散布在视频的不同位置和由于它们出现在距摄像机的距离不同的位置上导致的多尺度特性, 提出了多尺度卷积稀疏编码模型 (multi scale-convolutional sparse coding, MS-CSC), 类似于 Wei 等人 (2017) 的工作, 分别使用 L1 和总变分 (total variation, TV) 规范特征图的稀疏性和运动对象层的平滑度。这种编码方式使得模型能够从雨视频中适当地提取自然雨痕。

3) 基于深度网络的方法。通过探索利用视频中的时间冗余信息, Liu 等人 (2018) 建立了一个混合降雨模型描述雨痕和遮挡, 利用深度 RNN (recurrent neural network) 设计了联合递归去雨和重建网络 (joint recurrent rain removal and restoration network, J4R-Net), 无缝集成了降雨分类、基于空间纹理外观的去雨和基于时序一致性的背景帧细节重建, 通过动态检测视频上下文环境实现去雨, 设计了动态路由残差循环网络 (dynamic recurrent rain removal network, D3R-Net) 和用于去除视频雨水的空间时间残差学习模块。Yang 等人 (2019b) 建立了一个双流正则化约束的两级递归网络, 以进行视频合成降雨合成模型的逆恢复过程。同时, 考虑到相邻帧高度相关可视为同一场景的不同版本且雨痕沿时间维度随机分布, Yang 等人 (2019a) 构建了一个两阶段的自学习去雨网络, 充分利用时域相关性和一致性来消除雨痕。

2.2 视频压缩技术

2.2.1 混合框架中的传统技术

1) 块划分。区别于之前的对称四叉树、二叉树

划分, Wang 等人(2019b) 在四叉树、二叉树划分的基础上提出了一种非对称的三叉树划分方式以适应更高分辨率、更多样的视频内容。Fu 等人(2019) 提出了非对称划分方式 (unsymmetrical quad tree, UQT), 可以划分出四叉树、二叉树 (quad tree plus binary tree, QTBT) 和扩展四叉树 (extended quad tree, EQT) 无法划分出的形状。为了更好地拟合编码单元中信号的变化规律, 清华大学提出了一种“帧内导出树”划分技术 (Wang 等, 2019c), 将一个帧内编码单元划分成两个或四个预测单元, 同时规定了相应预测单元对应的变换单元的划分方式。

2) 帧内预测。在帧内预测部分, 预测像素滤波技术得到了较多关注。Fan 等人(2019) 提出了一种帧内预测滤波方式, 通过对一定范围内的边界像素使用左边界或上边界参考像素对预测值进行加权滤波, 有效减少了预测噪声带来的影响。腾讯公司提出了一种多层次帧内像素滤波技术 (Wang 等, 2019e), 有别于传统的针对参考像素的滤波, 该方法根据视频信号位置、与参考像素距离决定滤波方式。除此之外, 跨分量预测技术也有较多研究。Li 等人(2020a) 利用亮度参考边界和色度参考边界信息构建线性预测模型, 有效利用分量间的相关性以降低分量间的冗余。针对屏幕内容视频, Xu 等人(2016) 提出了帧内块拷贝技术, 允许编码单元将当前帧内的已编码块作为预测块, 大幅提升编码效率。

3) 帧间预测。运动补偿对视频压缩帧间预测性能有着至关重要的影响。以往的帧间技术模块仅考虑平移的运动模型。Li 等人(2015) 提出一种仿射变换运动补偿预测方法, 对每个子块计算其在仿射运动场中的运动矢量, 从而进行子块的运动补偿, 可以有效提升具有缩放、旋转、透视和其他不规则运动场景的编码性能。基于此, Zhang 等人(2020a) 提出一种交织仿射变换运动补偿方法, 根据两种不同的划分模板将编码块划分为子块并进行仿射运动补偿, 然后加权求和得到最终预测块。Li 等人(2019b) 通过对已编码区域的运动信息建模, 提出一种基于历史信息的运动矢量预测方式, 利用非局部相似性的原理获取更多非局部的运动矢量候选。Fang 等人(2020) 提出一种基于子块的运动矢量角度预测方法, 预设了5种角度, 按照定义好的角度方向寻找每个子块对应的运动矢量。Chen 等人

(2018b) 提出的解码端运动矢量修正技术对双向预测的两个运动矢量通过双边匹配进行进一步优化。Jin 等人(2007) 提出一种帧间帧内联合预测方式, 最终的预测块由两种预测方式加权得到, 融合帧内预测和帧间预测的结果可以达到更优的像素预测。

4) 变换。针对帧间预测残差的分布特性, 华为公司提出一种子块变换技术 (Zhao 等, 2019b), 假设预测残差分布在残差块的局部区域, 仅针对残差块的一个子块进行变换处理, 从而提升压缩性能。Wang 等人(2018a) 对帧间预测残差分布特性进行分析, 提出基于位置的变换技术, 考虑到编码单元中不同位置的帧间预测残差分布特性, 对行变换和列变换分别使用两种变换核, 这种基于位置的绑定方式能有效适应不同位置帧间残差的分布规律。Zhang 等人(2020b) 提出一种隐择变换技术, 为当前残差块预设了两种变换核, 通过率失真优化决策过程选择更优的变换核, 选择结果并不在码流中传输, 而是通过符号隐藏方式标识。

5) 滤波。混合框架中的滤波技术与传统图像处理领域中基于先验模型的算法进行了大量交叉研究, 性能取得了显著提升。Zhang 等人(2015) 提出一种非局部结构相似性滤波技术, 通过统计局部多个像素发生失真得出的规律特性推断更优的全局优化策略, 虽然复杂度较高但效果明显, 全局优化处理能够带来明显的效率提升。Jian 等人(2020) 提出一种基于像素纹理方向特性的滤波技术, 通过纹理特性得到更精细的分类, 在分类结果上进行不同的偏移值补偿, 从而有效降低量化失真。

2.2.2 混合框架中的深度学习技术

1) 帧内预测。深度学习帧内预测是近年来深度学习编码研究的热点领域之一。Li 等人(2018c) 提出一种全连接网络用于帧内预测, 对于当前块, 使用其左侧、上侧以及左上重建像素作为帧内预测上下文信息输入神经网络, 预测出当前块, 并将该网络预测作为一种新的模式, 与 HEVC 已有的35种帧内模式进行率失真优化。Cui 等人(2017) 提出一种帧内预测卷积神经网络, 对 HEVC 编码器的预测块进行预测增强处理, 并联合每一个预测块的左侧、上侧以及左上的区域作为帧内预测上下文信息联合输入神经网络。与卷积网络不同, Hu 等人(2019) 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,采用循环神经网络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。

2) 帧间预测。帧间预测方面的研究主要集中在卷积神经网络对帧间预测运动补偿以及参考帧的扩展研究。Huo 等人(2018)在时域运动补偿的基础上结合空域相关性,提出使用卷积神经网络对帧间预测结果进行增强,使用待预测块周围的重建信息作为上下文可以提高帧间预测质量。Yan 等人(2017)使用卷积神经网络进行分像素插值,对如何获得分像素训练数据进行研究,提出了先对高分辨率图像进行模糊然后从模糊图像中采样像素的获取数据方法。并在后续工作(Yan 等 2018)中提出可以将分像素运动补偿视为视频帧间的回归问题,用分像素运动估计对齐当前帧和参考帧,将当前帧作为分像素,参考帧作为整像素。

3) 上下采样。得益于深度学习在图像超分辨率领域取得的显著进展, Li 等人(2018a)和 Lin 等人(2019)分别提出了块自适应分辨率编码(block-based adaptive resolution coding, BARC)框架用于 I 帧和 P、B 帧,将下采样后编码和直接编码进行率失真优化决策。Li 等人(2019d)提出一种全新的正则化损失,并验证了正则化损失在联合训练图像编码上下采样网络时非常有效。

4) 熵编码。Song 等人(2017)提出通过分析视频纹理估计不同帧内预测模式的概率分布的方法提高帧内预测模式编码的效率。与 Pfaff 等人(2018)工作的不同点在于 Song 等人(2017)使用的是卷积神经网络而不是全连接神经网络。Ma 等人(2018a)使用神经网络预测变换系数的概率分布,特别是直流变换系数的概率分布,提高了变换系数编码的效率。使用卷积神经网络分析周围块的纹理特征,可以比传统视频编码更加准确地估计当前块变换系数的概率分布。

5) 滤波。深度学习滤波是提升编码性能的有效手段。Dai 等人(2017b)提出一个 4 层的可变尺寸重建卷积神经网络(variable-kernel restoration convolution neural net, VRCNN)用于 HEVC 帧内编码的后处理。VRCNN 采用可变尺寸的滤波器设计,并采用残差连接以简化训练过程,很好地提高了 HEVC 编码的效率。Yang 等人(2019b)提出 HEVC 的后处理模型,并针对帧内模式和帧间模式使用不同的神

经网络模型,进一步提高了后处理的性能。Jin 等人(2018)提出在均方误差损失函数之外使用判别器损失函数训练神经网络模型以提高主观质量。Li 等人(2017)提出传递边信息到解码端以提高后处理神经网络性能的方法。Yang 等人(2018)和 Wang 等人(2018c)提出联合使用相邻帧提高当前帧的后处理效果。He 等人(2018)提出使用块划分信息提高后处理神经网络的性能。Ma 等人(2018b)指出将预测信息和残差信息分别输入神经网络可以提高后处理的性能。Song 等人(2018)提出将量化步长与解码帧一起输入神经网络可以提高后处理的性能。

与后处理滤波类似,基于深度神经网络的环路滤波技术研究同样能取得主观和客观质量的双重提升。Meng 等人(2018)在 HEVC 的去块效应滤波和样本自适应偏移模块后使用长短时记忆神经网络解码视频帧进行滤波,同时联合使用了块划分信息。Zhang 等人(2018b)针对帧内模式编码帧、单向预测模式编码帧、双向预测模式编码帧以及不同的量化步长范围训练的神经网络进行环路滤波。Dai 等人(2018)在原始 VRCNN 的基础上增加了网络的深度,并应用于 HEVC 的环路滤波,明显提高了 HEVC 压缩视频的主客观质量。Jia 等人(2019)训练了多个神经网络用于 HEVC 的环路滤波,并训练了解码端分类网络用于模型的选择,避免了传输模型选择信息。

6) 编码优化。神经网络为解决复杂编码优化问题提供了新思路,已广泛用于编码过程中的模式决策问题。Kuang 等人(2020)和 Zhao 等人(2019a)在屏幕视频编码场景下,使用深度神经网络预测编码模式,大幅加快了屏幕编码速度。Chen 等人(2020b)使用深度神经网络提取局部方向纹理特征预测块划分和帧内预测模式,实现了 75% 的编码加速。Song 等人(2017)提出根据纹理信息和量化步长估计帧内预测模式,从而加速 HEVC 的帧内编码过程。Zhang 等人(2019)通过分析纹理特征加速块划分过程。Li 等人(2020b)提出使用剪枝的方法加快神经网络的运行,并使用加快后的神经网络加速 HEVC 的编码单元(coding unit, CU)划分过程,实现了实时编码。Li 等人(2019c)提出使用强化学习的方法加速块划分。Tang 等人(2019)和 Jin 等人(2017)使用神经网络加速 VVC 的块划分过程。

Ren 等人(2019)、Liu 等人(2016)和 Xu 等人(2018)使用神经网络加速 HEVC 的块划分。Huang 等人(2019)使用神经网络同时预测编码和预测单元的划分方式实现加速。

除上述关于编码单元模式决策的研究之外,深度学习编码研究还可应用于码率控制。Li 等人(2017)依据 R-lambda 模型,提出使用 CNN 预测 HEVC 中编码树(coding tree unit, CTU)级别的码率控制参数,从而实现更加精确的码率控制。Lu 等人(2020c)依据 R-QP(rate-quantization parameter)模型,提出使用深度神经网络估计编码树单元级量化步长,实现更精准的码率控制。Xing 等人(2019)提出使用深度神经网络实现码率控制。Hu 等人(2018a)提出使用强化学习的方法预测 HEVC 帧内编码的量化步长。

2.2.3 端到端深度学习视频压缩

国内端到端视频压缩方案的研究也是从 2017 年左右开始,并且取得了一系列研究成果。方案主要集中在面向低延时的应用场景。Chen 等人(2017)首次提出基于神经网络的视频压缩方案,但压缩性能不及 H.264。后来提出另一个基于块的视频压缩方案(Chen 等,2020a),性能与 H.264 相当。Lu 等人(2019)提出了第一个端到端视频压缩方案。该方案可以看成是传统视频压缩方案的深度学习版本,在 PSNR 和 MS-SSIM 上的压缩性能分别与 H.264 和 H.265 相当。由于灵活性和高性能,该方案成为本领域基线方案,常在后续工作中引用。Liu 等人(2019)提出使用空域和时域的先验去除帧内纹理、运动场和残差的空域和时域冗余,构建了一个面向低延时场景的端到端视频压缩方案,并在后续工作(Liu 等,2020a)中结合了时域先验、空域自回归先验和空域超先验来压缩运动信息,在 MS-SSIM 指标压缩性能上超过了 Lu 等人(2019)的方法。之后,Liu 等人(2020b)在上一个工作的基础上,提出使用多尺度的运动补偿模块,进一步提升了压缩性能。Lin 等人(2020)在 Lu 等人(2019)方案的基础上,提出了基于多参考帧预测的 4 个模块(运动场预测模块、运动场改善模块、运动补偿模块和残差改善模块)来提升帧间预测性能,相比 Lu 等人(2019)的方法,该方案在同样的码率下,PSNR 提高了近 1 dB,达到了与 H.265 相当的水平。最近,Lu 等人(2020a,2020b)在 Lu 等人(2019)的基础上扩展了

两个工作,一是使用自适应量化层减小了可变码率编码的模型参数,同时使用运动场和残差改善模块来压缩运动场和残差,进一步提升了压缩性能;二是使用多帧的率失真损失函数来减小误差传播以及设计了在线编码端参数更新算法。这两个策略都有效提升了压缩性能。Hu 等人(2020)在 Lu 等人(2019)的方案上引入帧级和块级的分辨率自适应运动场编码模式,为每个视频帧自适应地在帧级和块级选择合适的分辨率编码。除了高校以外,国内的图鸭信息科技有限公司也参与了 CVPR2020 Workshop 中的 P 帧编码竞赛,并提出一个端到端 P 帧压缩方案(Wu 等,2020),使用自编码器基于参考帧、当前帧和原始运动场同时提取、压缩并解码出重建的运动场和残差,然后使用空域偏移卷积(spatially-displaced convolution)生成当前帧的预测,最后与残差相加得到重建帧。考虑到低码率端的量化失真,使用了后处理模块提升压缩性能。

2.2.4 标准进展

国内标准方面,数字音视频编解码技术标准工作组(audio video coding standard, AVS)于 2017 年 12 月决定开展下一代视频编码标准(即 AVS3 标准)的制定。制定工作分为两个阶段,第 1 阶段从 2018 年 3 月到 2019 年 6 月,制定面向复杂度优先的应用,性能相较于 2014 年制定完成的上一代视频编码标准 AVS2 提升约 30%。第 2 阶段从 2019 年 6 月到 2021 年 3 月,目标是编码效率超过 VVC 标准。AVS3 标准采纳了包括块划分、帧内帧间预测、变换量化、系数编码、环路滤波、针对屏幕内容的技术以及针对监控视频的技术等 40 余项技术。

2.3 多视点/立体视频压缩

多视点视频压缩同样面临分辨率提升、动态范围提升和全景视角等新兴需求和挑战。AVS 工作组针对这些需求建立了 AVS-3D 视频编码框架,具体在编码端编码稀疏的若干视点,在解码端通过视点合成技术生成任意数量的虚拟视点。针对立体视频各视点之间存在较强的内容相关性这一特点,以及深度视频通常呈现分片均匀、边缘锐利的特征,提出研究高效的视间预测编码技术去除视间冗余,并提出面向视点合成的深度编码技术,提升合成视点质量,从而进一步提升压缩效率。

3 国内外研究进展比较

3.1 视频处理技术国内外研究进展比较

近年来国内外的学术机构和工业界对视频处理的关注程度显著提高,一方面是由于视频大数据的应用场景和用户需求正在发生巨变,另一方面是由于人工智能技术的飞速发展引领了新一轮技术革新。视频处理领域的早期工作主要由国外学者和研究机构展开,2016年前后,国内学者逐渐发力,相关研究开始逐渐占据主流并贡献了大量的优越方法。特别是随着深度学习在视频处理领域的广泛应用,国内部分团队已实现国际领先水平,同时针对不同前沿学科之间的交叉研究进行了深入探索。

3.2 视频压缩技术国内外研究进展比较

当前国际国内研究单位针对混合视频压缩框架的研究主要集中于局部模块优化以及性能提升。随着国内各公司与高校不断加大投入和布局,现有视频压缩研究局面已发生根本性改变,由欧美日韩公司主导部分国内企业与高校辅助参与变为华为、高通、三星、腾讯、字节跳动和德国赫兹研究所等机构齐头并进的局面。不仅如此,近年来国内外华人参与视频压缩技术研究的比例逐年上升,北京大学、中国科学技术大学、浙江大学和哈尔滨工业大学等高校对视频压缩新技术和新方向的探索同样越发深入,视频压缩技术研究的重心逐渐由欧美地区转向东亚地区。

视频压缩标准方面,新一代混合框架的视频压缩标准目前呈现“三足鼎立”的态势,主要有国际电信联盟和国际标准化组织推出的 H. 26x 系列标准、中国 AVS 工作组研制的 AVS 系列标准以及 AOM 联盟提出的 AV1 标准,这些标准技术框架具有一定相似性但技术细节各有千秋,主要差异体现在对各自的应用场景设计了独特的压缩算法和优化方法。图 1 展示了不同视频压缩标准的客观性能对比测试结果,图中所示为采用 6 个 4 K 超高清测试序列的平均性能,测试配置为随机访问。可以看出,相较于 HEVC, VVC、EVC、AVS3(HPM) 和 AV1 等 4 个标准在客观性能上均有较大提升,AVS3 与 VVC 相比还存在性能差距,仍有继续探索和提升的空间。

深度学习视频压缩是当前国内外公司和高校研究的热点领域。对于混合框架中的深度学习编码,

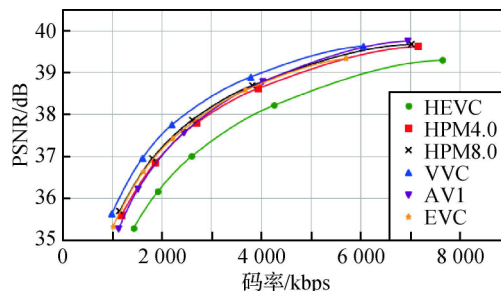


图 1 不同视频编码标准的性能对比

Fig. 1 Comparison of the rate-distortion performance among video coding standards

国内研究机构是该方向的前驱者,在各个编码模块均进展显著,并由此引发了工业界和学术研究机构重新审视深度学习在视频压缩领域的应用。高通、Technicolor、联发科和夏普等公司持续跟进了深度学习视频压缩以及标准化进程。从 2017 年开始,端到端视频压缩研究工作逐年增加,国内研究结构主要针对压缩效率提升开展工作,较少涉及复杂度优化。而国外研究机构主要针对可变码率模型和模型压缩、模型定点化等方面进行探索。

4 发展趋势与展望

随着物质生活水平的提高,人民对精神文化生活的的需求不断增长,渴望更加便捷高效地交流、学习、工作、娱乐和消费体验。另外,通信技术和硬件计算能力的不断进步,使得视频消费的成本逐步下降。在这两方面作用下,视频产业在近几年获得了巨大的发展机遇,以视频为主要载体的新应用场景和商业模式不断出现,提出了对视频处理与压缩技术的新要求,进而吸引了产业界和学术界的大力投入和研究。

在视频处理领域,深度学习工具全面颠覆了传统方法,开创了视频处理的新纪元。通过引入具有全场景自适应性和泛化特性的深度网络模型,视频的主客观质量得到了显著提升,在多种任务上突破了传统方法固有的系统性效率瓶颈和性能缺陷,取得了飞跃式发展。

在视频压缩方面,新一代视频压缩标准的制定大幅提高了视频压缩效率,下一阶段将着重开展针对新一代标准的优化工作。基于深度学习的视频编码将成为视频压缩效率提高的最大助力,面向下游

机器视觉任务的视频编码将成为研究热点, 联合面向视频信号保真度的压缩处理与视频语义层面的内容分析可以推动视频产业的发展迈向新高峰。更高质量视觉效果和高效率视觉表达之间将不再是单独研究的个体, 融合类脑视觉系统及编码机理的视频处理与压缩技术将是未来研究的重要领域之一。

致 谢 本文由中国图象图形学学会多媒体专业委员会牵头组织撰写, 并得到了北京大学、中国科学技术大学、中国科学院信息工程研究所、香港城市大学、南京理工大学和北京航空航天大学等相关研究团队的大力支持, 中国图象图形学学会评审专家也为本文提出了宝贵意见, 在此表示感谢。

参考文献(References)

- Afonso M, Zhang F and Bull D R. 2018. Video compression based on spatio-temporal resolution adaptation. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(1): 275-280 [DOI: 10.1109/TCSVT.2018.2878952]
- Agustsson E, Minnen D, Johnston N, Ballé J, Hwang S J and Toderici G. 2020. Scale-space flow for end-to-end optimized video compression//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 8503-8512 [DOI: 10.1109/CVPR42600.2020.00853]
- Bare B, Yan B, Ma C X and Li K. 2019. Real-time video super-resolution via motion convolution kernel estimation. *Neurocomputing*, 367: 236-245 [DOI: 10.1016/j.neucom.2019.07.089]
- Barnum P C, Kanade T and Narasimhan S. 2007. Spatio-temporal frequency analysis for removing rain and snow from videos//*Proceedings of the 1st International Workshop on Photometric Analysis for Computer Vision*. Rio de Janeiro, Brazil: HAL: 1-8
- Barnum P C, Narasimhan S and Kanade T. 2010. Analysis of rain and snow in frequency space. *International Journal of Computer Vision*, 86(2/3): #256 [DOI: 10.1007/s11263-008-0200-2]
- Bossu J, Hautière N and Tarel J P. 2011. Rain or snow detection in image sequences through use of a histogram of orientation of streaks. *International Journal of Computer Vision*, 93(3): 348-367 [DOI: 10.1007/s11263-011-0421-7]
- Brewer N and Liu N J. 2008. Using the shape characteristics of rain to identify and remove rain from video//*Proceedings of 2008 Joint IAPR International Workshops on Structural, Syntactic, and Statistical Pattern Recognition*. Orlando, USA: Springer: 451-458 [DOI: 10.1007/978-3-540-89689-0_49]
- Caballero J, Ledig C, Aitken A, Acosta A, Totz J, Wang Z H and Shi W Z. 2017. Real-time video super-resolution with spatio-temporal networks and motion compensation//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA: IEEE: 2848-2857 [DOI: 10.1109/CVPR.2017.304]
- Chen H, Chen J, Chernyak R and Esenlik S. 2018a. Description of SDR, HDR and 360° Video Coding Technology Proposal by Huawei, GoPro, HiSilicon, and Samsung-General Application Scenario, in Joint Video Experts Team (JVET) of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-J0025, 10th Meeting. San Diego, USA: [s. n.]
- Chen J, Tan C H, Hou J H, Chau L P and Li H. 2018b. Robust video content alignment and compensation for rain removal in a CNN framework//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 6286-6295 [DOI: 10.1109/CVPR.2018.00658]
- Chen T, Liu H J, Shen Q, Yue T, Cao X and Ma Z. 2017. DeepCoder: a deep neural network based video compression//*Proceedings of 2017 IEEE Visual Communications and Image Processing*. St. Petersburg, USA: IEEE: 1-4 [DOI: 10.1109/VCIP.2017.8305033]
- Chen Y L and Hsu C T. 2013. A generalized low-rank appearance model for spatio-temporally correlated rain streaks//*Proceedings of 2013 IEEE International Conference on Computer Vision*. Sydney, Australia: IEEE: 1968-1975 [DOI: 10.1109/ICCV.2013.247]
- Chen Y W and Chien W J. 2018b. Description of SDR, HDR and 360° Video Coding Technology Proposal by Qualcomm and Technicolor Low and High Complexity Versions. JVET Document, JVET-J0021
- Chen Z B, He T Y, Jin X and Wu F. 2020a. Learning for video compression. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(2): 566-576 [DOI: 10.1109/TCSVT.2019.2892608]
- Chen Z B, Shi J and Li W P. 2020b. Learned fast HEVC intra coding. *IEEE Transactions on Image Processing*, 29: 5431-5446 [DOI: 10.1109/TIP.2020.2982832]
- Choi H and Bajić I V. 2020. Deep frame prediction for video coding. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(7): 1843-1855 [DOI: 10.1109/TCSVT.2019.2924657]
- Cui W X, Zhang T, Zhang S P, Jiang F, Zuo W M, Wan Z L and Zhao D B. 2017. Convolutional neural networks based intra prediction for HEVC//*Proceedings of 2017 Data Compression Conference*. Snowbird, USA: IEEE: 436-436 [DOI: 10.1109/DCC.2017.53]
- Dai J F, Qi H Z, Xiong Y W, Li Y, Zhang G D, Hu H and Wei Y C. 2017a. Deformable convolutional networks//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 764-773 [DOI: 10.1109/ICCV.2017.89]
- Dai Y Y, Liu D and Wu F. 2017b. A convolutional neural network approach for post-processing in HEVC intra coding//*Proceedings of the 23rd International Conference on Multimedia Modeling*. Reykjavik, Iceland: Springer: 28-39 [DOI: 10.1007/978-3-319-51811-4_3]
- Dai Y Y, Liu D, Zha Z J and Wu F. 2018. A CNN-based in-loop filter

- with CU classification for HEVC//Proceedings of 2018 IEEE Visual Communications and Image Processing. Taichung, China: IEEE: 1-4 [DOI: 10.1109/VCIP.2018.8698616]
- Djelouah A, Campos J, Schaub-Meyer S and Schroers C. 2019. Neural inter-frame compression for video coding//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 6421-6429 [DOI: 10.1109/ICCV.2019.00652]
- Dosovitskiy A, Fischer P, Ilg E, Häusser P, Hazirbas C, Golkov V, van der Smagt P, Cremers D and Brox T. 2015. FlowNet: learning optical flow with convolutional networks//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 2758-2766 [DOI: 10.1109/ICCV.2015.316]
- Drulea M and Nedeveschi S. 2011. Total variation regularization of local-global optical flow//Proceedings of the 14th International IEEE Conference on Intelligent Transportation Systems. Washington, USA: IEEE: 318-323 [DOI: 10.1109/ITSC.2011.6082986]
- Fachada S, Bonatto D, Schenkel A and Lafruit G. 2018. Depth image based view synthesis with multiple reference views for virtual reality//Proceedings of 2018 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video. Helsinki, Finland: IEEE: 1-4 [DOI: 10.1109/3DTV.2018.8478484]
- Fan K, Wang R G, Li G and Gao W. 2019. Efficient prediction methods with enhanced spatial-temporal correlation for HEVC. IEEE Transactions on Circuits and Systems for Video Technology, 29(12): 3716-3728 [DOI: 10.1109/TCSVT.2018.2885002]
- Fang S, Sun Y, Chen F and Wang L. 2020. CE3-3: Motion Vector Angular Prediction. AVS-Doc, M5143
- Fu T L, Zhang K, San L Z, Liu H B, Wang S S and Ma S W. 2019. Unsymmetrical quad-tree partitioning for audio video coding standard-3 (AVS-3) //Proceedings of 2019 Picture Coding Symposium (PCS). Ningbo, China: IEEE: 1-5 [DOI: 10.1109/PCS48520.2019.8954558]
- Galpin F, Racapé F, Jaiswal S, Bordes P, Le Lannec F and François E. 2019. CNN-based driving of block partitioning for intra slices encoding//Proceedings of 2019 Data Compression Conference. Snowbird, USA: IEEE: 162-171 [DOI: 10.1109/DCC.2019.00024]
- Garg K and Nayar S K. 2004. Detection and removal of rain from videos//Proceedings of 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE: #1 [DOI: 10.1109/CVPR.2004.1315077]
- Garg K and Nayar S K. 2005. When does a camera see rain? //Proceedings of the 10th IEEE International Conference on Computer Vision. Beijing, China: IEEE: 1067-1074 [DOI: 10.1109/ICCV.2005.253]
- Garg K and Nayar S K. 2007. Vision and rain. International Journal of Computer Vision, 75(1): 3-27 [DOI: 10.1007/s11263-006-0028-6]
- Golinski A, Pourreza R, Yang Y, Sautiere G and Cohen T S. 2020. Feedback recurrent autoencoder for video compression [EB/OL]. [2021-01-11]. <https://arxiv.org/pdf/2004.04342.pdf>
- Guo J and Chao H Y. 2017. Building an end-to-end spatial-temporal convolutional network for video super-resolution//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI: 4053-4060
- Habibian A, van Rozendaal T, Tomczak J and Cohen T S. 2019. Video compression with rate-distortion autoencoders//Proceedings of 2019 IEEE International Conference on Computer Vision. Seoul, Korea (South): IEEE: 7033-7042 [DOI: 10.1109/ICCV.2019.00713]
- Haris M, Shakhnarovich G and Ukita N. 2018. Deep back-projection networks for super-resolution//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 1664-1673 [DOI: 10.1109/CVPR.2018.00179]
- Haris M, Shakhnarovich G and Ukita N. 2019. Recurrent back-projection network for video super-resolution//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 3892-3901 [DOI: 10.1109/CVPR.2019.00402]
- He X Y, Hu Q, Zhang X Y, Zhang C Y, Lin W Y and Han X T. 2018. Enhancing HEVC compressed videos with a partition-masked convolutional neural network//Proceedings of the 25th IEEE International Conference on Image Processing. Athens, Greece: IEEE: 216-220 [DOI: 10.1109/ICIP.2018.8451086]
- Hochreiter S and Schmidhuber J. 1997. Long short-term memory. Neural Computation, 9(8): 1735-1780 [DOI: 10.1162/neco.1997.9.8.1735]
- Hu J H, Peng W H and Chung C H. 2018a. Reinforcement learning for HEVC/H.265 intra-frame rate control//Proceedings of 2018 IEEE International Symposium on Circuits and Systems. Florence, Italy: IEEE: 1-5 [DOI: 10.1109/ISCAS.2018.8351575]
- Hu Y Y, Yang W H, Li M D and Liu J Y. 2019. Progressive spatial recurrent neural network for intra prediction. IEEE Transactions on Multimedia, 21(12): 3024-3037 [DOI: 10.1109/TMM.2019.2920603]
- Hu Z, Cho S, Wang J and Yang M H. 2018b. Deblurring low-light images with light streaks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(10): 2329-2341
- Hu Z H, Chen Z B, Xu D, Lu G, Ouyang W L and Gu S H. 2020. Improving deep video compression by resolution-adaptive flow coding//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 193-209 [DOI: 10.1007/978-3-030-58536-5_12]
- Huang H Y, Schiopu I and Munteanu A. 2020. Frame-wise CNN-based filtering for intra-frame quality enhancement of HEVC videos. IEEE Transactions on Circuits and Systems for Video Technology (99): #3018230 [DOI: 10.1109/TCSVT.2020.3018230]

- Huang Y, Song L and Izquierdo E. 2019. CNN accelerated intra video coding, where is the upper bound? //Proceedings of 2019 Picture Coding Symposium. Ningbo, China: IEEE: 1-5 [DOI: 10.1109/PCS48520.2019.8954494]
- Huang Y, Wang W and Wang L. 2018. Video super-resolution via bidirectional recurrent convolutional networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40 (4): 1015-1028 [DOI: 10.1109/TPAMI.2017.2701380]
- Hui Z, Li J, Gao X B and Wang X M. 2020. Progressive perception-oriented network for single image super-resolution [EB/OL]. [2021-01-11]. <https://arxiv.org/pdf/1907.10399.pdf>
- Huo S, Liu D, Wu F and Li H Q. 2018. Convolutional neural network-based motion compensation refinement for video coding//Proceedings of 2018 IEEE International Symposium on Circuits and Systems. Florence, Italy: IEEE: 1-4 [DOI: 10.1109/ISCAS.2018.8351609]
- Irani M and Peleg S. 1991. Improving resolution by image registration. CVGIP: Graphical Models and Image Processing, 53(3): 231-239
- Irani M and Peleg S. 1993. Motion analysis for image enhancement: resolution, occlusion, and transparency. Journal of Visual Communication and Image Representation, 4(4): 324-335
- Jacobsen J H, Smeulders A and Oyallon E. 2018. i-RevNet: deep invertible networks [EB/OL]. [2021-01-11]. <https://arxiv.org/pdf/1802.07088.pdf>
- Jia C M, Wang S Q, Zhang X F, Wang S S, Liu J Y, Pu S L and Ma S W. 2019. Content-aware convolutional neural network for in-loop filtering in high efficiency video coding. IEEE Transactions on Image Processing, 28 (7): 3343-3356 [DOI: 10.1109/TIP.2019.2896489]
- Jian Y, Zhang J, Luo F, Wang S and Ma S. 2020. CE6-4 related: An improved SAO filtering algorithm. AVS-Doc, M5374
- Jiang T X, Huang T Z, Zhao X L, Deng L J and Wang Y. 2017. A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 4057-4066 [DOI: 10.1109/CVPR.2017.301]
- Jiang T X, Huang T Z, Zhao X L, Deng L J and Wang Y. 2019. Fast-derain: a novel video rain streak removal method using directional gradient priors. IEEE Transactions on Image Processing, 28(4): 2089-2102 [DOI: 10.1109/TIP.2018.2880512]
- Jin X, Ngan K N and Zhu G X. 2007. Combined inter-intra prediction for high definition video coding//2007 Picture Coding Symposium. Lisbon, Portugal: EURASIP: 1-4
- Jin Z P, An P, Shen L Q and Yang C. 2017. CNN oriented fast QTBT partition algorithm for JVET intra coding//Proceedings of 2017 IEEE Visual Communications and Image Processing. St. Petersburg, USA: IEEE: 1-4 [DOI: 10.1109/VCIIP.2017.8305020]
- Jin Z P, An P, Yang C and Shen L Q. 2018. Quality enhancement for intra frame coding via CNNs: an adversarial approach//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. Calgary, Canada: IEEE: 1368-1372 [DOI: 10.1109/ICASSP.2018.8461356]
- Jo Y, Oh S W, Kang J and Kim S J. 2018. Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3224-3232 [DOI: 10.1109/CVPR.2018.00340]
- Kang J H, Kim S and Lee K M. 2017. Multi-modal/multi-scale convolutional neural network based in-loop filter design for next generation video codec//Proceedings of 2017 IEEE International Conference on Image Processing. Beijing, China: IEEE: 26-30 [DOI: 10.1109/ICIP.2017.8296236]
- Kappeler A, Yoo S, Dai Q Q and Katsaggelos A K. 2016. Video super-resolution with convolutional neural networks. IEEE Transactions on Computational Imaging, 2 (2): 109-122 [DOI: 10.1109/TCI.2016.2532323]
- Karczewicz M, Zhang L, Chien W J and Li X. 2017. Geometry transformation-based adaptive in-loop filter//2016 Picture Coding Symposium. Nuremberg, Germany: IEEE: 1-5 [DOI: 10.1109/PCS.2016.7906346]
- Kim J H, Sim J Y and Kim C S. 2015. Video deraining and desnowing using temporal correlation and low-rank matrix completion. IEEE Transactions on Image Processing, 24(9): 2658-2670 [DOI: 10.1109/TIP.2015.2428933]
- Kim K and Ro W W. 2019. Fast CU depth decision for HEVC using neural networks. IEEE Transactions on Circuits and Systems for Video Technology, 29(5): 1462-1473 [DOI: 10.1109/TCSVT.2018.2839113]
- Kim S Y, Lim J, Na T and Kim M. 2019. Video super-resolution based on 3D-CNNs with consideration of scene change//Proceedings of 2019 IEEE International Conference on Image Processing. Taipei, China: IEEE: 2831-2835 [DOI: 10.1109/ICIP.2019.8803297]
- Kim Y, Soh J W, Park J, Ahn B, Lee H S, Moon Y S and Cho N I. 2020. A pseudo-blind convolutional neural network for the reduction of compression artifacts. IEEE Transactions on Circuits and Systems for Video Technology, 30(4): 1121-1135 [DOI: 10.1109/TCSVT.2019.2901919]
- Koo M, Salehifar M, Lim J and Kim S. 2019. CE6: Reduced Secondary Transform (RST) (CE6-3.1). JVET-N0193
- Krishnan D, Tay T and Fergus R. 2011. Blind deconvolution using a normalized sparsity measure//Proceedings of 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE: 233-240 [DOI: 10.1109/CVPR.2011.5995521]
- Kuang W, Chan Y L, Tsang S H and Siu W C. 2020. DeepSCC: deep learning-based fast prediction network for screen content coding. IEEE Transactions on Circuits and Systems for Video Technology, 30(7): 1917-1932 [DOI: 10.1109/TCSVT.2019.2929317]
- Li C, Song L, Xie R and Zhang W J. 2017. CNN based post-processing

- to improve HEVC//Proceedings of 2017 IEEE International Conference on Image Processing. Beijing, China: IEEE: 4577-4580 [DOI: 10.1109/ICIP.2017.8297149]
- Li D Y, Liu Y and Wang Z F. 2019a. Video super-resolution using non-simultaneous fully recurrent convolutional network. IEEE Transactions on Image Processing, 28(3): 1342-1355
- Li J H, Li B, Xu J Z, Xiong R Q and Gao W. 2018c. Fully connected network-based intra prediction for image coding. IEEE Transactions on Image Processing, 27(7): 3236-3247 [DOI: 10.1109/TIP.2018.2817044]
- Li J R, Wang M, Zhang L, Zhang K, Liu H B, Wang S Q, Ma S W and Gao W. 2019b. History-based motion vector prediction for future video coding//Proceedings of 2019 IEEE International Conference on Multimedia and Expo. Shanghai, China: IEEE: 67-72 [DOI: 10.1109/ICME.2019.00020]
- Li J R, Wang M, Zhang L, Zhang K, Wang S Q, Wang S S, Ma S W and Gao W. 2020a. Sub-sampled cross-component prediction for chroma component coding//Proceedings of 2020 Data Compression Conference. Snowbird, USA: IEEE: 203-212 [DOI: 10.1109/DCC47342.2020.00028]
- Li L, Li H Q, Lyu Z Y and Yang H T. 2015. An affine motion compensation framework for high efficiency video coding//Proceedings of 2015 IEEE International Symposium on Circuits and Systems. Lisbon, Portugal: IEEE: 525-528 [DOI: 10.1109/ISCAS.2015.7168686]
- Li M H, Xie Q, Zhao Q, Wei W, Gu S H, Tao J and Meng D Y. 2018b. Video rain streak removal by multiscale convolutional sparse coding//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 6644-6653 [DOI: 10.1109/CVPR.2018.00695]
- Li N, Zhang Y, Zhu L W, Luo W H and Kwong S. 2019c. Reinforcement learning based coding unit early termination algorithm for high efficiency video coding. Journal of Visual Communication and Image Representation, 60: 276-286 [DOI: 10.1016/j.jvcir.2019.02.021]
- Li T Y, Xu M, Deng X and Shen L Q. 2020b. Accelerate CTU partition to real time for HEVC encoding with complexity control. IEEE Transactions on Image Processing, 29: 7482-7496 [DOI: 10.1109/TIP.2020.3003730]
- Li Y, Li B, Liu D and Chen Z B. 2017. A convolutional neural network-based approach to rate control in HEVC intra coding//Proceedings of 2017 IEEE Visual Communications and Image Processing. St. Petersburg, USA: IEEE: 1-4 [DOI: 10.1109/VCIP.2017.8305050]
- Li Y, Liu D, Li H Q, Li L, Li Z and Wu F. 2019d. Learning a convolutional neural network for image compact-resolution. IEEE Transactions on Image Processing, 28(3): 1092-1107 [DOI: 10.1109/TIP.2018.2872876]
- Li Y, Liu D, Li H Q, Li L, Wu F, Zhang H and Yang H T. 2018a. Convolutional neural network-based block up-sampling for intra frame coding. IEEE Transactions on Circuits and Systems for Video Technology, 28(9): 2316-2330 [DOI: 10.1109/TCSVT.2017.2727682]
- Lin J P, Liu D, Li H Q and Wu F. 2020. M-LVC: multiple frames prediction for learned video compression//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 3546-3554 [DOI: 10.1109/CVPR42600.2020.00360]
- Lin J P, Liu D, Yang H T, Li H Q and Wu F. 2019. Convolutional neural network-based block up-sampling for HEVC. IEEE Transactions on Circuits and Systems for Video Technology, 29(12): 3701-3715 [DOI: 10.1109/TCSVT.2018.2884203]
- Liu H J, Chen T, Lu M, Shen Q and Ma Z. 2019. Neural video compression using spatio-temporal priors [EB/OL]. [2021-01-11]. <https://arxiv.org/pdf/1902.07383.pdf>
- Liu C and Sun D Q. 2013. On Bayesian adaptive video super resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(2): 346-360 [DOI: 10.1109/TPAMI.2013.127]
- Liu H J, Lu M, Ma Z, Wang F, Xie Z H, Cao X and Wang Y. 2020b. Neural video coding using multiscale motion compensation and spatiotemporal context model [EB/OL]. [2021-01-11]. <https://arxiv.org/pdf/2007.04574.pdf>
- Liu H J, Shen H, Huang L C, Lu M, Chen T and Ma Z. 2020a. Learned video compression via joint spatial-temporal correlation exploration//Proceedings of the 34th AAAI Conference on Artificial Intelligence AAAI 2020, the 32nd Innovative Applications of Artificial Intelligence Conference, IAAI 2020, the 10th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020. New York, USA: AAAI: 11580-11587
- Liu J Y, Yang W H, Yang S and Guo Z M. 2018. Erase or fill? Deep joint recurrent rain removal and reconstruction in videos//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 3233-3242 [DOI: 10.1109/CVPR.2018.00341]
- Liu Z Y, Yu X Y, Gao Y, Chen S L, Ji X Y and Wang D S. 2016. CU partition mode decision for HEVC hardwired intra encoder using convolution neural network. IEEE Transactions on Image Processing, 25(11): 5088-5103 [DOI: 10.1109/TIP.2016.2601264]
- Lu G, Cai C L, Zhang X Y, Chen L, Ouyang W L, Xu D and Gao Z Y. 2020b. Content adaptive and error propagation aware deep video compression//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 456-472 [DOI: 10.1007/978-3-030-58536-5_27]
- Lu G, Ouyang W L, Xu D, Zhang X Y, Cai C L and Gao Z Y. 2019. DVC: an end-to-end deep video compression framework//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 10998-11007 [DOI: 10.1109/CVPR.2019.01126]

- Lu G, Zhang X Y, Ouyang W L, Chen L, Gao Z Y and Xu D. 2020a. An end-to-end learning framework for video compression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*: #2988453 [DOI: 10.1109/TPAMI.2020.2988453]
- Lu X, Zhou B X, Jin X S and Martin G. 2020c. A rate control scheme for HEVC intra coding using convolution neural network (CNN) // *Proceedings of 2020 Data Compression Conference*. Snowbird, USA: IEEE: 382-382 [DOI: 10.1109/DCC47342.2020.00055]
- Lucas A, López-Tapia S, Molina R and Katsaggelos A K. 2019. Generative adversarial networks and perceptual losses for video super-resolution. *IEEE Transactions on Image Processing*, 28(7): 3312-3327 [DOI: 10.1109/TIP.2019.2895768]
- Lyu Z Y, Piao Y J, Wu Y, Choi K and Choi K P. 2020. Scan region-based coefficient coding in avs3 // *Proceedings of 2020 IEEE International Conference on Multimedia and Expo Workshops*. London, UK: IEEE: 1-5 [DOI: 10.1109/ICMEW46912.2020.9105993]
- Ma C Y, Liu D, Peng X L and Wu F. 2018a. Convolutional neural network-based arithmetic coding of DC coefficients for HEVC intra coding // *Proceedings of the 25th IEEE International Conference on Image Processing*. Athens, Greece: IEEE: 1772-1776 [DOI: 10.1109/ICIP.2018.8451166]
- Ma L, Tian Y and Huang T. 2018b. Residual-based video restoration for HEVC intra coding // *Proceedings of the 4th IEEE International Conference on Multimedia Big Data*. Xi'an, China: IEEE: 1-7 [DOI: 10.1109/BigMM.2018.8499072]
- Ma Z, Liao R J, Tao X, Xu L, Jia J and Wu E H. 2015. Handling motion blur in multi-frame super-resolution // *Proceedings of 2015 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Boston, USA: IEEE: 5224-5232 [DOI: 10.1109/CVPR.2015.7299159]
- Meng X D, Chen C, Zhu S Y and Zeng B. 2018. A new HEVC in-loop filter based on multi-channel long-short-term dependency residual networks // *Proceedings of 2018 Data Compression Conference*. Snowbird, USA: IEEE: 187-196 [DOI: 10.1109/DCC.2018.00027]
- Meyer S, Djelouah A, McWilliams B, Sorkine-Hornung A, Gross M and Schroers C. 2018. PhaseNet for video frame interpolation // *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 498-507 [DOI: 10.1109/CVPR.2018.00059]
- Misra K, Bossen F and Segall A. 2019. On cross component adaptive loop filter for video compression // *2019 Picture Coding Symposium*. Ningbo, China: IEEE: 1-5 [DOI: 10.1109/PCS48520.2019.8954547]
- Niklaus S and Liu F. 2018. Context-aware synthesis for video frame interpolation // *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 1701-1710 [DOI: 10.1109/CVPR.2018.00183]
- Niklaus S, Mai L and Liu F. 2017a. Video frame interpolation via adaptive separable convolution // *Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 261-270 [DOI: 10.1109/ICCV.2017.37]
- Niklaus S, Mai L and Liu F. 2017b. Video frame interpolation via adaptive convolution // *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA: IEEE: 2270-2279 [DOI: 10.1109/CVPR.2017.244]
- Park W and Kim M. 2019. Deep predictive video compression with bi-directional prediction [EB/OL]. [2021-01-11]. <https://arxiv.org/pdf/1904/1904.02909.pdf>
- Park W J and Lee K H. 2008. Rain removal using Kalman filter in video // *Proceedings of 2008 International Conference on Smart Manufacturing Application*. Goyangi, Korea (South): IEEE: 494-497 [DOI: 10.1109/ICSMA.2008.4505573]
- Park W S and Kim M. 2016. CNN-based in-loop filtering for coding efficiency improvement // *Proceedings of the 12th IEEE Image, Video, and Multidimensional Signal Processing Workshop*. Bordeaux, France: IEEE: 1-5 [DOI: 10.1109/IVMSPW.2016.7528223]
- Paul S, Norkin A and Bovik A C. 2020. Speeding up VP9 intra encoder with hierarchical deep learning-based partition prediction. *IEEE Transactions on Image Processing*, 29: 8134-8148 [DOI: 10.1109/TIP.2020.3011270]
- Pessoa J, Aidos H, Tomás P and Figueiredo M A T. 2020. End-to-end learning of video compression using spatio-temporal autoencoders // *Proceedings of 2020 IEEE Workshop on Signal Processing Systems*. Coimbra, Portugal: IEEE: 1-6 [DOI: 10.1109/SiPS50750.2020.9195249]
- Pfaff J, Helle P, Maniry D, Kaltenstadler S, Samek W, Schwarz H, Marpe D and Wiegand T. 2018. Neural network based intra prediction for video coding // *Proceedings of SPIE 10752, Applications of Digital Image Processing XLII*. San Diego, USA: SPIE: 1075213 [DOI: 10.1117/12.2321273]
- Puri S, Lasserre S and Le Callet P. 2017. CNN-based transform index prediction in multiple transforms framework to assist entropy coding // *Proceedings of the 25th European Signal Processing Conference*. Kos, Greece: IEEE: 798-802 [DOI: 10.23919/EUSIPCO.2017.8081317]
- Ranjan A and Black M J. 2017. Optical flow estimation using a spatial pyramid network // *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA: IEEE: 2720-2729 [DOI: 10.1109/CVPR.2017.291]
- Ren W H, Tian J D, Han Z, Chan A and Tang Y D. 2017. Video desnowing and deraining based on matrix decomposition // *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA: IEEE: 4210-4219 [DOI: 10.1109/CVPR.2017.303]
- Ren W P, Su J, Sun C and Shi Z P. 2019. An IBP-CNN based fast block partition for intra prediction // *2019 Picture Coding Symposium*. Ningbo, China: IEEE: 1-5 [DOI: 10.1109/PCS48520.

2019. 8954522]
- Rippel O, Nair S, Lew C, Branson S, Anderson A and Bourdev L. 2019. Learned video compression//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 3453-3462 [DOI: 10.1109/ICCV.2019.00355]
- Said A, Zhao X, Karczewicz M, Chen J L and Zou F. 2016. Position dependent prediction combination for intra-frame video coding//Proceedings of 2016 IEEE International Conference on Image Processing. Phoenix, USA: IEEE: 534-538 [DOI: 10.1109/ICIP.2016.7532414]
- Sajjadi M S M, Vemulapalli R and Brown M. 2018. Frame-recurrent video super-resolution//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 6626-6634 [DOI: 10.1109/CVPR.2018.00693]
- Schwarz H, Nguyen T, Marpe D and Wiegand T. 2018. Non-CE7: Alternative Entropy Coding for Dependent Quantization. JVET-K0072. [s. n.]
- Shi X J, Chen Z R, Wang H, Yeung D Y, Wong W K and Woo W C. 2015. Convolutional LSTM network: a machine learning approach for precipitation nowcasting//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada: MIT Press: 802-810
- Song R, Liu D, Li H Q and Wu F. 2017. Neural network-based arithmetic coding of intra prediction modes in HEVC//Proceedings of 2017 IEEE Visual Communications and Image Processing. St. Petersburg, USA: IEEE: 1-4 [DOI: 10.1109/VCIP.2017.8305104]
- Song X D, Yao J B, Zhou L L, Wang L, Wu X Y, Xie D and Pu S L. 2018. A practical convolutional neural network as loop filter for intra frame//Proceedings of the 25th IEEE International Conference on Image Processing. Athens, Greece: IEEE: 1133-1137 [DOI: 10.1109/ICIP.2018.8451589]
- Su H, Chen M L, Bokov A, Mukherjee D, Wang Y Q and Chen Y. 2019b. Machine learning accelerated transform search for AV1//Proceedings of 2019 Picture Coding Symposium. Ningbo, China: IEEE: 1-5 [DOI: 10.1109/PCS48520.2019.8954514]
- Su H, Tsai C Y, Wang Y Q and Xu Y W. 2019a. Machine learning accelerated partition search for video encoding//Proceedings of 2019 IEEE International Conference on Image Processing. Taipei, China: IEEE: 2661-2665 [DOI: 10.1109/ICIP.2019.8803311]
- Tang G W, Jing M G, Zeng X Y and Fan Y B. 2019. Adaptive CU split decision with pooling-variable CNN for VVC intra encoding//Proceedings of 2019 IEEE Visual Communications and Image Processing. Sydney, Australia: IEEE: 1-4 [DOI: 10.1109/VCIP47243.2019.8965679]
- Tao X, Gao H Y, Liao R J, Wang J and Jia J Y. 2017. Detail-revealing deep video super-resolution//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 4482-4490 [DOI: 10.1109/ICCV.2017.479]
- Tian Y P, Zhang Y L, Fu Y and Xu C L. 2020. TDAN: temporally-deformable alignment network for video super-resolution//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 3360-3369 [DOI: 10.1109/CVPR42600.2020.00342]
- Tripathi A K and Mukhopadhyay S. 2011. A probabilistic approach for detection and removal of rain from videos. IETE Journal of Research, 57(1): 82-91 [DOI: 10.4103/0377-2063.78382]
- Tripathi A K and Mukhopadhyay S. 2012. Video post processing: low-latency spatiotemporal approach for detection and removal of rain. IET Image Processing, 6(2): 181-196 [DOI: 10.1049/iet-ipr.2010.0547]
- Tripathi A K and Mukhopadhyay S. 2014. Removal of rain from videos: a review. Signal, Image and Video Processing, 8(8): 1421-1430 [DOI: 10.1007/s11760-012-0373-6]
- Tsai R Y and Huang T S. 1984. Multiple frame image restoration and registration//Proceedings of 1984 Advances in Computer Vision and Image Processing. Greenwich, UK: JAI Press: 317-339
- Wang H, Su D W, Liu C C, Jin L C, Sun X F and Peng X Y. 2019a. Deformable non-local network for video super-resolution. IEEE Access, 7: 177734-177744 [DOI: 10.1109/ACCESS.2019.2958030]
- Wang L G, Guo Y L, Lin Z P, Deng X P and An W. 2018b. Learning for video super-resolution through HR optical flow estimation//Proceedings of the 14th Asian Conference on Computer Vision. Perth, Australia: Springer: 514-529 [DOI: 10.1007/978-3-030-20887-5_32]
- Wang L Q, Cao X R, Niu B B, Yu Q H, Zheng J H and He Y. 2019b. Derived tree block partition for AVS3 intra coding//2019 Picture Coding Symposium. Ningbo, China: IEEE: 1-5 [DOI: 10.1109/PCS48520.2019.8954542]
- Wang L Q, Niu B B, Lin Y B, Yu Q H, Zheng J H and He Y. 2018a. Texture and position based multiple transform for inter-predicted residue coding//Proceedings of 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference. Honolulu, USA: IEEE: 31-35 [DOI: 10.23919/APSIPA.2018.8659520]
- Wang M, Li J R, Zhang L, Zhang K, Liu H B, Wang S Q, Kwong S and Ma S W. 2019c. Extended coding unit partitioning for future video coding. IEEE Transactions on Image Processing, 29: 2931-2946 [DOI: 10.1109/TIP.2019.2955238]
- Wang T T, Xiao W H, Chen M J and Chao H Y. 2018c. The multi-scale deep decoder for the standard HEVC bitstreams//Proceedings of 2018 Data Compression Conference. Snowbird, USA: IEEE: 197-206 [DOI: 10.1109/DCC.2018.00028]
- Wang X L, Girshick R, Gupta A and He K M. 2018d. Non-local neural networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 7794-7803 [DOI: 10.1109/CVPR.2018.00813]
- Wang X T, Chan K K C, Yu K, Dong C and Loy C C. 2019d. EDVR: video restoration with enhanced deformable convolutional networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and

- Pattern Recognition Workshops. Long Beach, USA: IEEE: 1954-1963 [DOI: 10.1109/CVPRW.2019.00247]
- Wang Y, Xu X, Li Y and Liu S. 2019e. CE1-related: A Rediction Pixel Filtering Method of Intra Coding. AVS-Doc, M5079
- Wei W, Yi L X, Xie Q, Zhao Q, Meng D Y and Xu Z B. 2017. Should we encode rain streaks in video as deterministic or stochastic? //Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2516-2525 [DOI: 10.1109/ICCV.2017.275]
- Wu C Y, Singhal N and Krahenbühl P. 2018. Video compression through image interpolation//Proceedings of the 15th European Conference on Computer Vision. Munich, Bavaria, Germany: Springer: 416-431 [DOI: 10.1007/978-3-030-01237-3_26]
- Wu X J, Zhang Z W, Feng J, Zhou L and Wu J M. 2020. End-to-end optimized video compression with MV-residual prediction//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, USA: IEEE: 156-157 [DOI: 10.1109/CVPRW50498.2020.00086]
- Xing H F, Zhou Z C, Wang J L, Shen H F, He D L and Li F. 2019. Predicting rate control target through a learning based content adaptive model//2019 Picture Coding Symposium. Ningbo, China: IEEE: 1-5 [DOI: 10.1109/PCS48520.2019.8954541]
- Xiu X, He Y, Ye Y, Jang H, Nam J, Kim S, Lim J, Chen H, Yang H and Chen J. 2018. CE4-related: One Simplified Design of Advanced Temporal Motion Vector Prediction (ATMVP). JVET-K0346
- Xu M, Li T Y, Wang Z L, Deng X, Yang R and Guan Z Y. 2018. Reducing complexity of HEVC: a deep learning approach. IEEE Transactions on Image Processing, 27(10): 5044-5059 [DOI: 10.1109/TIP.2018.2847035]
- Xu X Z, Liu S, Chuang T D, Huang Y W, Lei S M, Rapaka K, Pang C, Seregin V, Wang Y K and Karczewicz M. 2016. Intra block copy in HEVC screen content coding extensions. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 6(4): 409-419 [DOI: 10.1109/JETCAS.2016.2597645]
- Xue T F, Chen B A, Wu J J, Wei D L and Freeman W T. 2019. Video enhancement with task-oriented flow. International Journal of Computer Vision, 127(8): 1106-1125 [DOI: 10.1007/s11263-018-01144-2]
- Yan B, Lin C M and Tan W M. 2019. Frame and feature-context video super-resolution//Proceedings of the 33rd AAAI Conference on Artificial Intelligence, AAAI 2019, the 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019, the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019. Honolulu, USA: AAAI: 5597-5604
- Yan N, Liu D, Li H Q and Wu F. 2017. A convolutional neural network approach for half-pel interpolation in video coding//2017 IEEE International Symposium on Circuits and Systems. Baltimore, USA: IEEE: 1-4 [DOI: 10.1109/ISCAS.2017.8050458]
- Yan N, Liu D, Li H Q, Xu T, Wu F and Li B. 2018. Convolutional neural network-based invertible half-pixel interpolation filter for video coding//Proceedings of the 25th IEEE International Conference on Image Processing. Athens, Greece: IEEE: 201-205 [DOI: 10.1109/ICIP.2018.8451286]
- Yang R, Mentzer F, van Gool L and Timofte R. 2020a. Learning for video compression with hierarchical quality and recurrent enhancement//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 6628-6637 [DOI: 10.1109/CVPR42600.2020.00666]
- Yang R, Mentzer F, van Gool L and Timofte R. 2020b. Learning for video compression with recurrent auto-encoder and recurrent probability model [EB/OL]. [2021-01-11]. <https://arxiv.org/pdf/2006.13560.pdf>
- Yang R, Xu M, Liu T, Wang Z L and Guan Z Y. 2019b. Enhancing quality for HEVC compressed videos. IEEE Transactions on Circuits and Systems for Video Technology, 29(7): 2039-2054 [DOI: 10.1109/TCSVT.2018.2867568]
- Yang R, Xu M, Wang Z L and Li T Y. 2018. Multi-frame quality enhancement for compressed video//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 6664-6673 [DOI: 10.1109/CVPR.2018.00697]
- Yang W H, Liu J Y and Feng J S. 2019a. Frame-consistent recurrent video deraining with dual-level flow//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 1661-1670 [DOI: 10.1109/CVPR.2019.00176]
- Yi P, Wang Z Y, Jiang K, Jiang J J and Ma J Y. 2019. Progressive fusion video super-resolution network via exploiting non-local spatio-temporal correlations//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 3106-3115 [DOI: 10.1109/ICCV.2019.00320]
- Yilmaz M A and Tekalp A M. 2020. End-to-end rate-distortion optimization for bi-directional learned video compression//Proceedings of 2010 IEEE International Conference on Image Processing. Abu Dhabi, United Arab Emirates: IEEE: 1311-1315 [DOI: 10.1109/ICIP40778.2020.9190881]
- Zhang H and Patel V M. 2017. Convolutional sparse and low-rank coding-based rain streak removal//Proceedings of 2017 IEEE Winter Conference on Applications of Computer Vision. Santa Rosa, USA: IEEE: 1259-1267 [DOI: 10.1109/WACV.2017.145]
- Zhang J, Jia C M, Ma S W and Gao W. 2015. Non-local structure-based filter for video coding//2015 IEEE International Symposium on Multimedia. Miami, USA: IEEE: 301-306 [DOI: 10.1109/ISM.2015.90]
- Zhang K, Zhang L, Liu H B, Xu J Z, Deng Z P and Wang Y. 2020a. Interweaved prediction for video coding. IEEE Transactions on Image Processing, 29: 6422-6437 [DOI: 10.1109/TIP.2020.2987432]
- Zhang X P, Li H, Qi Y Y, Leow W K and Ng T K. 2006. Rain removal

in video by combining temporal and chromatic properties//Proceedings of 2006 IEEE International Conference on Multimedia and Expo. Toronto , Canada: IEEE: 461-464 [DOI: 10.1109/ICME.2006.262572]

Zhang Y , Han Y , Chen C C , Hung C H , Chien W J and Karczewicz M. 2018a. CE4. 3. 3: Locally Adaptive Motion Vector Resolution and MVD Coding. JVET-K0357

Zhang Y B , Shen T , Ji X Y , Zhang Y , Xiong R Q and Dai Q H. 2018b. Residual highway convolutional neural networks for in-loop filtering in HEVC. IEEE Transactions on Image Processing , 27(8) : 3827-3841 [DOI: 10.1109/TIP.2018.2815841]

Zhang Y F , Wang G , Tian R , Xu M and Kuo C C J. 2019. Texture-classification accelerated CNN scheme for fast intra CU partition in HEVC//Proceedings of 2019 Data Compression Conference. Snowbird , USA: IEEE: 241-249 [DOI: 10.1109/DCC.2019.00032]

Zhang Y H , Zhang K , Zhang L , Liu H B , Wang Y , Wang S S , Ma S W and Gao W. 2020b. Implicit-selected transform in video coding//Proceedings of 2020 IEEE International Conference on Multimedia and Expo Workshops. London , UK: IEEE: 1-6 [DOI: 10.1109/ICMEW46912.2020.9105962]

Zhao L L , Wei Z W , Cai W T , Wang W Y , Zeng L Y and Chen J W. 2019a. Efficient screen content coding based on convolutional neural network guided by a large-scale database//Proceedings of 2019 IEEE International Conference on Image Processing. Taipei , China: IEEE: 2656-2660 [DOI: 10.1109/ICIP.2019.8803294]

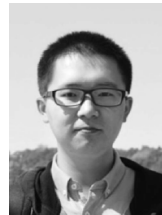
Zhao X D , Liu P , Liu J F and Tang X L. 2008. The application of histogram on rain detection in video//Proceedings of the 11th Joint International Conference on Information Sciences. [s. l.]: Atlantis Press: 382-387 [DOI: 10.2991/jcis.2008.65]

Zhao Y , Gao H , Yang H and Chen J. 2019b. Sub-block Transform (SBT) for Inter Blocks. JVET-M0140

Zhu X Z , Hu H , Lin S and Dai J F. 2019. Deformable ConvNets V2: More deformable , better results//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach , USA: IEEE: 9300-9308 [DOI: 10.1109/CVPR.2019.00953]

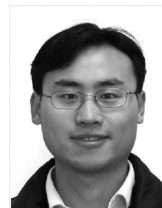
Zoran D and Weiss Y. 2011. From learning models of natural image patches to whole image restoration//Proceedings of 2011 IEEE International Conference on Computer Vision. Barcelona , Spain: IEEE: 479-486 [DOI: 10.1109/ICCV.2011.6126278]

作者简介



贾川民 ,1993 年生 ,男 ,博士后 ,主要研究方向为图像视频编码与处理。

E-mail: cmjia@pku.edu.cn



马思伟 通信作者 ,男 ,教授 ,主要研究方向为视频编码与传输、多媒体信号处理与理解。

E-mail: swma@pku.edu.cn

马海川 男 ,博士研究生 ,主要研究方向为图像与视频编码、信号处理、机器学习。E-mail: hcma@mail.ustc.edu.cn

杨文瀚 男 ,博士后 ,主要研究方向为图像视频增强、恶劣环境降质复原、人机协同编码。

E-mail: wyang34@cityu.edu.hk

任文琦 男 ,副研究员 ,主要研究方向为计算机视觉与图像处理。E-mail: renwenqi@iie.ac.cn

潘金山 男 ,教授 ,主要研究方向为图像/视频处理。

E-mail: jspan@njust.edu.cn

刘东 男 ,特任教授 ,主要研究方向为多媒体信息处理。

E-mail: dongeliu@ustc.edu.cn

刘家瑛 女 ,副教授 ,主要研究方向为多媒体信号处理与视觉理解。E-mail: liujiaying@pku.edu.cn