

文献引用格式: 严昊. 基于深度学习的视频压缩综述 [J]. 电视技术, 2024, 48(1): 216–219.

YAN H. An overview of video compression based on deep learning [J]. Video Engineering, 2024, 48(1): 216–219.

中图分类号: TP391.41; TP18

文献标识码: A

DOI: 10.16280/j.videoe.2024.01.060

基于深度学习的视频压缩综述

严 昊

(福州大学 物理与信息工程学院, 福建 福州 350108)

摘要: 近年来, 深度神经网络以其强大的非线性建模能力, 促进了视频压缩领域的发展, 促使基于深度学习的视频压缩成为一个备受关注的研究领域。介绍基于深度学习的端到端视频压缩技术及其发展, 重点介绍和总结现有的端到端视频压缩框架, 并分析未来的端到端视频压缩发展趋势。

关键词: 深度学习; 视频压缩; 神经网络

An Overview of Video Compression Based on Deep Learning

YAN Hao

(College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China)

Abstract: In recent years, deep neural network, with its powerful non-linear modeling ability, has promoted the development of video compression, which makes video compression based on deep learning become a research field. This paper introduces the end-to-end video compression technology based on deep learning and its development, especially introduces and summarizes the existing end-to-end video compression framework, and analyzes the future trend of end-to-end video compression.

Keywords: deep learning; video compression; neural network

0 引言

自 20 世纪 80 年代以来, 高效视频压缩一直是多媒体领域的一项重要课题。影像设备的更新迭代带来了更佳的观看体验, 传输和存储的数据量激增, 因此越来越需要拥有更高效的压缩方案。传统视频编码标准, 如动画专家组 (Moving Picture Experts Group, MPEG)、高级视频编码 (Advanced Video Coding, AVC)、高效视频压缩编码 (High Efficiency Video Coding, HEVC) 和通用视频编码 (Versatile Video Coding, VVC) 等在视频压缩任务取得了令人瞩目的性能效果^[1]。近年来, 神经网络已经在各领域获得了广泛应用。在压缩编码领域, 神经网络能够将像素域的信息转换成更紧凑的潜在表征, 从而提高视频的压缩性能。此外, 神经网络中的模型参数可以通过大量的视频样本学习得到, 减轻模型对手工设计的依赖。目前, 基于深度学习

的视频压缩已经成为一种拥有巨大潜力的视频压缩方案。

1 视频压缩框架概述

图像视频压缩的本质是运用算法剔除图像与视频信号中的多种多余信息, 如空间上的重复、时间上的冗余、视觉上的非必要元素以及编码层面的额外信息等。相比于图像压缩只在空间域减少冗余, 视频压缩需要利用连续帧之间的时间相关性减少时域冗余。传统的视频编码标准, 如 HEVC 是基于块结构的混合编码框架。混合编码框架主要包括帧内预测、帧间预测、变换量化、熵编码和去块滤波等模块^[2]。在视频压缩框架中应用深度学习技术时, 可以将基于深度学习的视频压缩分为两类, 分别为基于深度学习的混合式视频压缩和基于深度学习的端到端视频压缩。其中: 基于深度学习的混合式视频压缩是在传统混合编码框架的基础上, 利用神经网络

络模型替换原有的手工模块;基于深度学习的端到端视频压缩则是全部利用神经网络来进行视频的压缩与解压。

2 基于深度学习的混合式视频压缩

基于深度学习的混合式视频编码是将深度学习与传统视频编码器相结合的方法。根据将深度学习应用于视频编码框架的不同位置,将基于深度学习的混合视频编码分为基于深度学习的编码器优化和基于深度学习的后处理两类。

2.1 基于深度学习的编码器优化

基于深度学习的编码器优化是针对传统视频编码框架中的特定模块,利用神经网络模型进行优化。YAN 等人^[3]揭示了分数插值问题中的可逆性特性,即分数插值滤波器不仅应该从整数样本生成分数样本,还应该以可逆的方式从分数样本中恢复整数样本,提出了基于卷积神经网络(Convolutional Neural Network, CNN)的可逆性驱动插值滤波器来提升编码性能。ZHANG 等人^[4]则通过引入可变形卷积层和生成对抗网络,进一步提升了插值滤波器的性能。DING 等人^[5]通过使用 CNN 生成更高质量的虚拟参考帧,实现了更高效的帧间预测,由此提升压缩性能。HUANG 等人^[6]提出了一种基于神经网络的逐帧滤波方法,用于提升帧内编码的性能,并消除帧内压缩伪影。HU 等人^[7]提出了一种分辨率自适应的光流编码框架,利用全局和局部多分辨率光流图的表示,智能选择最佳分辨率,以实现视频的高效压缩。LIU 等人^[8]提出了多模式视频压缩框架,能够适应不同的运动模式和熵模型,通过基于块的模式集成和双模熵编码方案,实现了更好的视频压缩性能。此外,还有一些方法将深度学习应用于环内滤波器模块^[9-10]。通过使用 CNN 建立重建视频和未压缩视频之间的映射关系,由此明显改善压缩视频的质量。

2.2 基于深度学习的后处理

基于深度学习的后处理将深度学习用于视频编码的后处理阶段,通过去噪、增强、修复等操作,减少压缩带来的失真,提高解码后的视频质量。YANG 等人^[11]将基于深度学习的方法应用于解码端,以提高单帧视频的质量。近年出现了许多基于多帧的质量增强方法。DENG 等人^[12]提出使用一

个时空可变形融合方案,对目标帧及其相邻的多个参考帧进行融合的方法,实现了对压缩视频的增强。ZHAO 等人^[13]提出了一种递归可变形时空注意力模块,递归地将前一帧估计的隐藏特征与当前特征相结合。LUO 等人^[14]提出了一种新的端到端网络,称为粗到精细的时空信息融合,通过预测更大接受域的更好的偏移量来提高压缩视频的质量。然而,这些方法没有考虑邻域帧的质量高低,因此许多高质量多帧增强方法被提出。这些方法在根据定位峰值质量帧,利用相邻的高质量帧增强低质量帧,大大降低了压缩视频帧间的质量波动^[15-16]。

3 基于深度学习的端到端视频压缩

基于深度学习的图像编码成功推动了基于深度学习的视频压缩框架的发展。与图像编码相似,基于深度学习的端到端视频编码旨在借助深度学习实现完整的编解码框架。2019 年,LU 等人^[17]提出了一项开创性的工作,利用卷积神经网络替换了传统编码框架的各个模块,实现了第一个端到端视频压缩框架。近年来,人们获得了越来越多的端到端视频压缩研究成果,可以将这些研究分为两类,分别为基于残差编码的端到端视频压缩和基于条件编码的端到端视频压缩。

3.1 基于残差编码的端到端视频压缩

残差编码源自传统的视频编码标准编码器,先生成运动补偿预测,后对其与当前帧的残差进行编码。YILMAZ 等人^[18]提出了一种学习型分层双向视频编解码器,结合了分层运动补偿预测和端到端优化的优点。LIU 等人^[19]先将每个视频帧用低维的潜空间变量表示,设置了一个卷积长短期记忆网络,用于利用视频帧序列的时间相关性来预测未来的潜变量表示,并通过生成式对抗网络恢复视频的重建帧。HU 等人^[20]提出了一种特征空间视频编码网络,通过在特征空间内执行所有关键操作,包括运动估计、运动压缩、运动补偿和残差压缩实现视频编码。ZHAO 等人^[21]提出了一种端到端的深度视频编解码器,将视频压缩模块和视频后处理模块联合进行优化。

3.2 基于条件编码的端到端视频压缩

基于条件编码的端到端视频压缩将时序上的相邻帧或者特征用作当前帧的编码条件信息来提高

压缩效率。THEO 等人^[22]首次使用条件编码来编码前景内容。LI 等人^[23]正式提出了一个深度上下文视频压缩框架,利用特征域的上下文信息作为条件,有助于视频高频内容的重建,并实现了从预测编码到条件编码的转变。此外,该作者后续的工作通过引入特征传播和多尺度时间上下文进一步提高了压缩比^[24]。考虑到大多数现有的视频编解码器直接使用来自图像编解码器的现有熵模型来编码残差或运动矢量,并未充分利用视频中的时空特征,LI 等人^[25]提出了一个强大的神经视频编解码器熵模型,有效捕捉了视频的时空特征,推动了基于深度学习视频编解码器的发展。

4 总结与展望

基于深度学习的视频压缩技术的研究在过去几年取得了显著进展。通过利用大规模视频数据集训练得到的神经网络模型替换手工设计的编码器模块,不仅提高了编码器的压缩性能,而且提升了自适应的泛化能力,降低了其对手动设置参数的依赖。此外,将深度学习应用于压缩视频后处理阶段,更有利于提取整个视频序列时空域信息来减少视频压缩造成的失真。但是,针对编码模块优化和压缩视频后处理的深度学习视频压缩方案是基于传统视频编码框架进行改进的,神经网络模型与传统编码器模块间并不能完全适配,因此压缩性能的提升存在一定的限制。

基于此,目前出现了许多基于深度学习的端到端视频编码方案。这类方法通过深度学习构建整个编解码框架,通过权衡码率与重建视频质量的损失函数,训练得到一个最优的视频压缩模型,大大提升了压缩性能。得益于深度学习在提取更紧凑的潜在向量表示的强大能力,这类方法具有巨大的发展潜力。然而,这类方法为了提升视频压缩性能增加了网络模型的参数量,牺牲了一定的时间效率,未能达到实时解压。此外,端到端的视频压缩方案通常训练多个模型来适配不同的码率要求,成倍增加了模型的训练时间。

未来基于深度学习的视频压缩趋势不仅要进一步提升压缩性能,而且要考虑时间复杂度。随着第五代移动通信技术(5th Generation Mobile Communication Technology, 5G)和边缘计算的不断发展,基于深度

学习的视频压缩需要更好地满足低延迟和实时性的要求,以支持实时视频传输和互动应用。

5 结语

近年来,爆炸性增长的视频数据量给传输和存储带来了更大压力,使得视频压缩研究的重要性越来越凸显,而深度学习的蓬勃发展给视频压缩提供了技术支持。本文介绍视频压缩的框架,讨论视频压缩的分类,详细阐述基于深度学习的混合式视频压缩和于深度学习的端到端视频压缩的现有工作。其中,基于深度学习的混合式视频压缩可分为基于深度学习的编码器优化和基于深度学习的后处理,基于深度学习的端到端视频压缩可分为基于残差编码的端到端视频压缩和基于条件编码的端到端视频压缩。文章对现有方法进行了总结,分析了目前研究存在的不足,以期对基于深度学习的视频压缩的未来发展提供借鉴。

参考文献:

- [1] 贾川民, 赵政辉, 王苦社, 等. 基于神经网络的图像视频编码[J]. 电信科学, 2019, 35 (5) : 32-42.
- [2] 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.
- [3] YAN N, LIU D, LI H, et al. Invertibility-driven interpolation filter for video coding[J]. IEEE Transactions on Image Processing, 2019, 28 (10) : 4912-4925.
- [4] ZHANG Q, HUANG X, YIN H, et al. Deformable convolution network based invertibility-driven interpolation filter for HEVC[C]//2021 Data Compression Conference (DCC), 2021.
- [5] DING Q, SHEN L, YANG H, et al. VRFCNN: virtual reference frame generation network for quality SHVC[J]. IEEE Signal Processing Letters, 2020, 27 : 2049-2053.
- [6] HUANG H, SCHIOPU I, MUNTEANU A. Frame-wise CNN-based filtering for intra-frame quality enhancement of HEVC videos[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31 (6) : 2100-2113.
- [7] HU Z, CHEN Z, XU D, et al. Improving deep video compression by resolution-adaptive flow coding[C]//2020 European Conference on Computer Vision (ECCV), 2020.
- [8] LIU B, CHEN Y, MACHINENI R C, et al. MMVC: learned

- multi-mode video compression with block-based prediction mode selection and density-adaptive entropy coding[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.
- [9] DING D, KONG L, CHEN G, et al. A switchable deep learning approach for in-loop filtering in video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30 (7) :1871-1887.
- [10] PAN Z, YI X, ZHANG Y, et al. Efficient in-loop filtering based on enhanced deep convolutional neural networks for HEVC[J]. IEEE Trans Image Process, 2020, 29:5352-5366.
- [11] YANG R, XU M, LIU T, et al. Enhancing quality for HEVC compressed videos[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 29 (7) 2039-2054.
- [12] DENG J, WANG L, PU S, et al. Spatio-temporal deformable convolution for compressed video quality enhancement[C]// Association for the Advancement of Artificial Intelligence (AAAI), 2020.
- [13] ZHAO M, XU Y, ZHOU S. Recursive fusion and deformable spatiotemporal attention for video compression artifact reduction[C]//Proceedings of the 29th ACM International Conference on Multimedia, 2021.
- [14] LUO D, YE M, LI S, et al. Coarse-to-fine spatio-temporal information fusion for compressed video quality enhancement[J]. IEEE Signal Processing Letters, 2022, 29: 543-547.
- [15] YANG R, XU M, WANG Z, et al. Multi-frame quality enhancement for compressed video[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
- [16] GUAN Z, XING Q, XU M, et al. MFQE 2.0 : a new approach for multi-frame quality enhancement on compressed video[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43 (3) :949-963.
- [17] LU G, OUYANG W, XU D, et al. DVC : an end-to-end deep video compression framework[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.
- [18] YLMAZ M A, TEKALP A M. End-to-end rate-distortion optimized learned hierarchical bi-directional video compression[J]. IEEE Transactions on Image Processing, 2021, 31 :974-983.
- [19] LIU B, CHEN Y, LIU S, et al. Deep learning in latent space for video prediction and compression[C]//Proceedings of the IEEE/CVF Conference On Computer Vision And Pattern Recognition, 2021.
- [20] HU Z, LU G, XU D. FVC : a new framework towards deep video compression in feature space[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021.
- [21] ZHAO T, FENG W, ZENG H, et al. Learning-based video coding with joint deep compression and enhancement[C]// Proceedings of the 30th ACM International Conference on Multimedia, 2022.
- [22] THEO L, PIERRICK P, WASSIM H, et al. Optical flow and mode selection for learning-based video coding[C]//22nd IEEE International Workshop on Multimedia Signal Processing, 2020.
- [23] LI J, LI B, LU Y. Deep contextual video compression[J]. Advances in Neural Information Processing Systems, 2021, 34 :18114-18125.
- [24] SHENG X, LI J, LI B, et al. Temporal context mining for learned video compression[J]. IEEE Transactions on Multimedia, 2023, 25 :7311-7322.
- [25] LI J H, LI B, LU Y. Hybrid spatial-temporal entropy modelling for neural video compression[C]//Proceedings of the 30th ACM International Conference on Multimedia, 2022.

编辑:王倩