

基于最邻近帧质量增强的视频编码参考帧列表优化算法

霍俊彦, 邱瑞鹏, 马彦卓, 杨付正

(西安电子科技大学 ISN 国家重点实验室, 陕西 西安 710071)

摘要: 帧间预测是视频编码的核心模块, 其利用参考帧的重建样本来预测当前图像样本, 从而通过传输少量预测残差数据表示复杂视频内容。在有损视频编码中, 参考帧质量受到量化失真的影响, 导致预测精度较差, 影响编码性能。针对低时延视频业务, 提出一种基于最邻近帧质量增强的参考帧列表优化算法, 通过基于深度学习的卷积神经网络增强与当前帧最邻近参考帧的质量, 并将增强后的高质量帧整合到当前帧的参考帧列表中, 提高了帧间预测精度。以高效视频编码 H.265/HEVC 参考软件平台 HM16.22 为参考基准, 所提算法在 Y、Cb、Cr 这 3 个分量上可分别节省 9.06%、14.92%、13.19% 的编码码率。

关键词: H.265/HEVC; 深度学习; 帧间预测; 参考帧列表

中图分类号: TN911.7

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2022185

Reference frame list optimization algorithm in video coding by quality enhancement of the nearest picture

HUO Junyan, QIU Ruipeng, MA Yanzhuo, YANG Fuzheng

State Key Laboratory of Integrated Services Network, Xidian University, Xi'an 710071, China

Abstract: Interframe prediction is a key module in video coding, which uses the samples in the reference frames to predict those in the current picture, thus helps to represent the complex video by transmitting a small amount of the prediction residual. In lossy video coding, the qualities of reference frames are affected by the quantization distortion, which lead to poor prediction accuracy and performance degradation. Targeted at the low latency video services, a reference frame list optimization algorithm was proposed, which enhanced the quality of the nearest reference frame by a deep learning-based convolutional neural network, and integrated the enhanced reference frame into the reference frame list to improve the accuracy of interframe prediction. Compared with H.265/HEVC reference software HM16.22, the proposed algorithm provides BD-rate savings of 9.06%, 14.92% and 13.19% for Y, Cb and Cr components, respectively.

Keywords: H.265/HEVC, deep learning, interframe prediction, reference frame list

0 引言

近年来, 短视频的广泛普及使视频数据量进入持续爆发式增长阶段。同时, 直播电商作为一种新的购物消费方式开始上线。受新冠疫情影响, 视频会议的应用需求也呈现爆炸式增长。为保证用户间的实时交互, 这些视频业务对时延有严格的要求。

针对低时延、大数据量的特点, 在低时延视频应用场景下, 提高视频编码效率对高效开展视频会话具有积极的意义。与此同时, 基于深度学习的卷积神经网络 (CNN, convolutional neural network) 在视频处理等方面取得了令人瞩目的成果, 利用深度学习提升视频编码效率是未来的研究趋势。

高效视频编码 H.265/HEVC^[1]标准采用传统的

收稿日期: 2022-06-21; 修回日期: 2022-09-13

通信作者: 马彦卓, yzma@mail.xidian.edu.cn

基金项目: 国家自然科学基金资助项目 (No.62101409, No.62171353)

Foundation Item: The National Natural Science Foundation of China (No.62101409, No.62171353)

视频混合编码框架，使用灵活的块划分结构，并在预测、变换和熵编码等各个环节增加多种新算法。相比于先进视频编码 H.264/AVC^[2]标准，在相同的视频质量条件下，H.265/HEVC 可降低约 50% 的编码码率。针对日益增长的低时延视频业务，视频通用测试条件设计了低时延 P 帧(LDP, low-delay only P)配置，其仅利用单向参考帧进行帧间预测。在此配置下，视频序列的编码顺序与播放顺序一致，每帧编码时仅参考播放顺序在当前帧之前的已编码帧。在以往的研究中，基于深度学习的方法已经应用到了视频编码框架的主要模块中，如帧内预测、帧间预测和环路滤波等。本文算法的设计目标是借助深度学习的方法辅助帧间预测，提高 LDP 配置下 H.265/HEVC 的编码效率。

帧间预测是视频编码框架的关键模块，其利用视频的时间相关性提高编码效率。针对每个待编码块，帧间预测利用多个参考帧的重建样本，运用多种运动补偿技术构造预测样本，预测样本与原始样本之间的差值经过变换、量化、熵编码等模块处理后被送入码流进行传输。

目前，基于深度学习的帧间预测包含亚像素插值、预测值生成和参考帧列表(RFL, reference frame list)优化等方案，其中，参考帧列表优化使用神经网络生成额外的虚拟参考帧，用于待编码帧的预测参考。虚拟参考帧的生成涉及内插(两帧之间插入)和外插(多帧之外插入)2种方式。在 LDP 配置下，为了保证低时延，参考帧列表的图像皆为先于待编

码帧的图像。此时，虚拟参考帧的生成仅能由当前时刻之前的参考帧外插得到。相比于内插算法，外插算法需要生成参考帧之外的新内容，预测精度较低，对帧间编码性能的提高非常有限。

本文针对 LDP 配置提出基于最邻近帧质量增强的参考帧列表优化算法，优化框架如图 1 所示。将待编码帧前向参考帧列表中的第一个参考帧，即最邻近参考帧，送入提出的参考帧质量增强网络进行质量增强，并将其作为额外的参考帧整合到待编码帧的参考帧列表中。与现有虚拟参考帧算法相比，本文提出的参考帧增强算法的输出帧不包含虚构内容，其构造算法较简单，且性能优于已有虚拟参考帧算法。

综上，本文提出一种新型的参考帧列表优化算法，设计一种改进的视频帧质量增强网络对最邻近帧进行质量增强。首先将可变形卷积网络(DCN, deformable convolutional network)和光流相结合，应用光流引导可变形卷积的偏移预测，然后将网络生成的高质量参考帧插入参考帧列表参与帧间预测，提高预测准确度，进而提高编码效率。

1 相关研究

1.1 基于神经网络的视频帧质量增强

视频帧质量增强旨在减弱视频经过有损压缩引入的失真，基于深度学习的方法在视频帧质量增强方面取得了显著的成效。该类方法大致可分为单帧质量增强和多帧辅助质量增强。单帧质量增强方

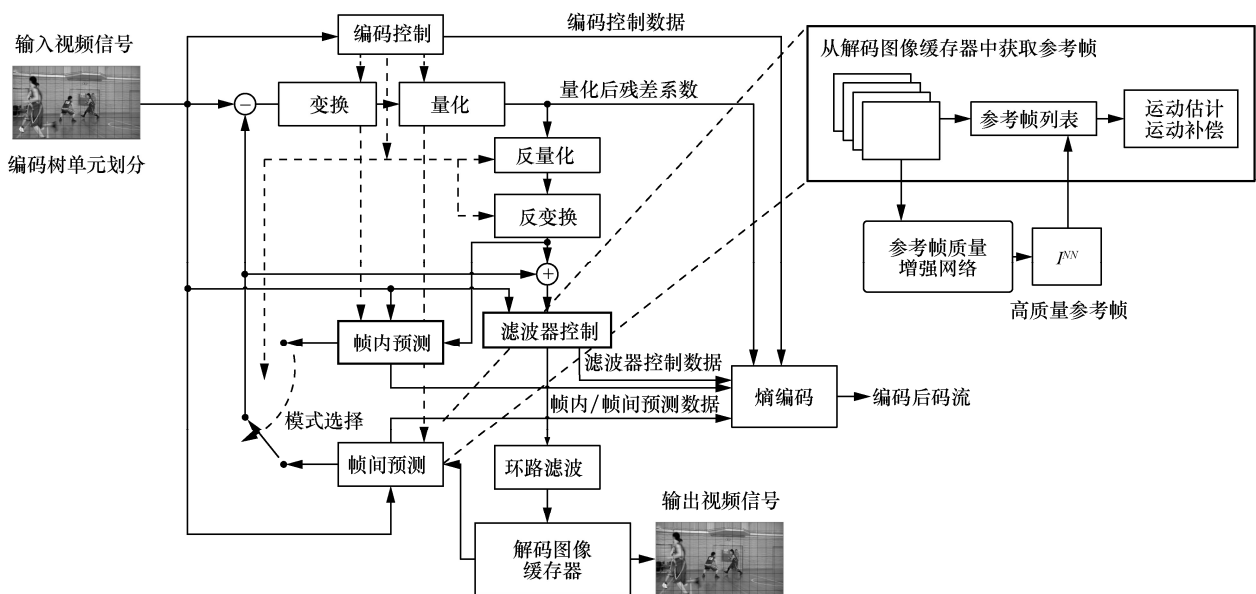


图 1 基于最邻近帧质量增强的参考帧列表优化框架

法基于深度学习的卷积神经网络来提取图像的空域信息和特征,通过强化这些信息对图像进行增强。多帧辅助质量增强方法则利用一个或多个相邻帧的时域和空域信息进行质量增强,是目前视频帧质量增强的主要手段。

1.1.1 单帧质量增强方法

受基于神经网络的图像超分辨率算法的启发,面向减弱失真的卷积神经网络(ARCNN, artifact reduction CNN)^[3]率先针对 JPEG 重建图像引入 CNN 滤波方法,实现对单帧图像的质量增强。基于卷积神经网络的环内滤波(IFCNN, in-loop filtering using CNN)技术^[4]使用三层卷积网络代替 H.265/HEVC 中的样点自适应补偿滤波模块。残差高速卷积神经网络(RHCNN, residual highway CNN)^[5]引入由残差高速单元和卷积层组成的神经网络,作为 H.265/HEVC 中额外的环内滤波器。非对称卷积残差网络(ACRN, asymmetric convolutional residual network)^[6]利用密集结构来提取重建帧的层次性特征,利用非对称卷积块提取纹理的方向性特征来恢复纹理。递归残差卷积神经网络(RRCNN, recursive residual CNN)^[7]作为 H.265/HEVC 中额外的环内滤波器,在编码树单元(CTU, coding tree unit)级自适应选择使用此滤波器,同时采用递归残差模块实现残差模块的参数共享。基于变滤波器尺寸的残差学习卷积神经网络(VRCNN, variable-filter-size residue-learning CNN)^[8]应用 Inception 模块的方式,利用多个小尺度卷积层并行组合来代替大尺度卷积层。多级注意力卷积神经网络(MACNN, multi-stage attention CNN)^[9]采用改进的 Inception 模块和自注意力机制,使用较浅的卷积网络获得全局信息能力。基于压缩激发模块的滤波卷积神经网络(SEFCNN, squeeze-and-excitation filtering CNN)^[10]利用特征提取和特征增强 2 个子网组合来更获取通道间的非线性关系,并结合压缩激发(SE, squeeze-and-excitation)模块^[11]来建立通道间注意力机制。

有些工作将编码信息应用到网络中,帮助网络快速收敛并进一步提高重建帧的质量。多模型/多尺度卷积神经网络(MMSCNN, multi-modal/multi-scale CNN)^[12]将 H.265/HEVC 的编码单元和变换单元的分区分区映射作为网络的输入。分区掩盖卷积神经网络^[13]深入分析编码单元分区信息的生成方法和帧融合方法。引入残差的卷积神经网络滤波器(CNNF-R,

CNN filter using residual)^[14]将编码后的残差信息作为辅助信息输入网络。深度残差卷积神经网络(DRCNN, deep residual CNN)^[15]采用归一化量化参数(QP, quantization parameter)作为网络输入,帮助网络学习 QP 与输入重建帧压缩失真的关系。

1.1.2 多帧辅助质量增强方法

空-时残差网络(STResNet, spatial-temporal residue network)^[16]通过同时输入当前块和同位块提高重建帧的质量。深度卡尔曼滤波网络(DKFN, deep Kalman filtering network)^[17]融合卡尔曼滤波模型和神经网络的高度非线性映射的优势,提供更准确的时间信息,同时还利用预测残差作为先验信息,从而产生更高质量的恢复结果。质量门控卷积长短期记忆(QG-ConvLSTM, quality-gated convolutional long short-term memory)网络^[18]根据不同质量帧的重要性不同,通过网络学习使每帧的信息得到合理和充分的使用。上述方法直接使用相邻同位块或参考帧作为网络的输入,未考虑运动信息,难以利用神经网络学习跨多帧的依赖关系。

考虑运动信息的多帧辅助质量增强一般分为帧对齐、帧融合和质量增强 3 个步骤。基于光流对齐和基于可变形卷积对齐是目前最常用的帧对齐方法。帧融合是将对齐的帧与待增强的帧进行融合,学习时域和空域相关性来生成融合特征以增强图像质量。质量增强是充分挖掘融合特征中的互补信息,生成增强目标帧。

基于光流的帧对齐通过估计帧之间的运动信息,将邻近帧变形映射(warp),使之与待增强帧对齐。质量增强网络(QENet, quantity enhancement network)^[19]、基于学习的多帧视频质量增强(LMVE, learning-based multi-frame video quality enhancement)^[20]和多帧引导的注意力网络(MGANet, multi-frame guided attention network)^[21]利用光流网络^[22]及使用金字塔、映射和代价量的卷积神经网络(PWC-Net)^[23]估计的光流得到对齐帧,并与当前帧一起作为网络的输入。多帧质量增强(MFQE, multi-frame quality enhancement)^[24]和 MFQE 2.0^[25]由基于光流的运动补偿子网络和质量增强子网络组成,将与待增强帧最接近的 2 个高质量关键帧作为输入。

现有的光流估计算法对有遮挡场景内容的稳健性不强,在运动幅度大的情况下不能保证运动信息的准确性。基于光流的帧对齐在每个特征位置只学习一个偏移量,而 DCN^[26]和 DCNv2^[27]引入多个

偏移量进行帧对齐。空-时可变形融合 (STDF, spatio-temporal deformable fusion)^[28]和递归融合及可变形空-时注意力 (RFDA, recursive fusion and deformable spatial-temporal attention)^[29]学习一种新的时空可变形卷积来聚合时间信息的同时进行帧融合, 并利用由密集连接的卷积层来实现质量增强。基于可变形卷积的帧对齐仍然存在一些问题, 如快速运动不准确、训练不稳定。由于可变形偏移与基于光流的偏移关系密切, 光流引导的可变形对齐网络 (FDAN, flow-guided deformable alignment network)^[30]和基本视频超分 (BasicVSR++, basic video super-resolution) 算法^[31]提出了新型的对齐算法, 将可变形卷积和光流相结合。

帧融合是将对齐后的相邻帧与当前帧进行融合, 生成融合特征以增强图像质量。RFDA^[29]提出一种递归融合模块, 在一段长时间范围内构建时间依赖性。LMVE^[20]和基于预测的多帧视频增强 (PMVE, prediction-based multi-frame video enhancement)^[32]展示了直接融合、早期融合和慢速融合 3 种不同的融合方法, 并通过实验验证了慢速融合在效率上的优势。但是这些方法忽略了不同位置对增强帧的影响不同。为此, 基于增强可变形卷积网络的视频恢复 (EDVR, video restoration with enhanced deformable convolutional network) 算法^[33]引入了时间和空间注意力来帮助跨多个对齐特征聚合信息, 但是没有区分多个相邻帧的重要性, 忽略了它们的不同特征。

1.2 基于神经网络的虚拟参考帧生成

上述单帧或多帧辅助质量增强方法可作为后处理技术应用在视频编码系统中, 也可用于环内滤波处理。但这些方式多以提高主观质量为目标, 对提高编码效率帮助有限。同时, 基于神经网络的环内滤波处理不可避免的一个问题是多次增强, 即多帧之间存在参考依赖, 质量提升效果会随着帧之间的传递逐渐减小。质量增强之后的帧较平滑, 多次滤波后平滑效果更加明显, 这时直接作为输出会影响主观质量。

出于以上原因, 本文聚焦于基于神经网络的帧间预测技术, 此类技术不仅致力于提高重建图像的质量, 而且利用增强技术可提高帧间预测的精确度, 从而明显提高视频编码效率, 同时也避免了多次增强的问题。

目前, 基于神经网络的帧间预测方法主要应用

在亚像素插值、预测值生成和参考帧列表优化等环节。亚像素插值方法^[34-37]通过神经网络利用整像素来推导亚像素位置。预测值生成方法^[38-39]通过神经网络来代替线性加权预测, 可更好地处理复杂运动。使用网络生成额外的虚拟参考帧是目前基于深度学习的参考帧列表优化的主要方法。虚拟参考帧生成一般采用内插、外插 2 种方式。Lin 等^[40]采用拉普拉斯金字塔状的生成式对抗网络, 利用待编码帧的前 4 个已编码的重建帧生成额外的具有高质量的参考帧。Zhao 等^[41]从 2 个重建参考帧中生成高质量虚拟帧放入参考帧列表, 并设计 CTU 级编码模式。之后 Zhao 等^[42]进一步利用神经网络将合成帧整合到参考帧列表中进行运动估计。Lee 等^[43]提出使用之前已编码的重建帧来合成虚拟参考帧, 具有更高的时间相关性, 可以同时进行视频帧内插和外插, 并自适应地修改了高级运动向量预测 (AMVP, advanced motion vector prediction) 模式和 Merge 模式的预测机制。Choi 等^[44]利用非线性变换和自适应空间变化滤波器, 使用扩张的卷积和减少滤波器长度以减少网络模型参数, 为单向预测和双向预测 2 种方式设计模型。此类方法在实时性的 LDP 配置下编码时, 仅能采用外插虚拟帧方法。而外插虚拟帧方法准确度低, 很难生成之后时刻尚不存在的信息, 导致性能损失。针对该问题, 本文提出利用深度学习实现参考帧质量增强, 并将增强后的参考帧作为额外的参考帧插入参考帧列表, 以提高帧间预测准确度, 从而提高视频编码效率。

2 基于神经网络的参考帧列表优化算法

2.1 网络结构

在常见的基于深度学习的视频帧质量增强网络中, 一般仅对一个分量, 即 Y 分量进行质量增强。H.265/HEVC 的通用测试序列格式为 YCbCr 4:2:0 格式, Y、Cb、Cr 这 3 个分量在尺寸、分布范围、纹理特征等方面各不相同。本文提出的网络结构设计对 3 个分量分支路单独处理。由于色度分量纹理相对平坦, 因此色度分量支路网络结构为亮度分量支路的简化版本, 并且 Cb 和 Cr 分量支路共享网络参数, 以避免色度分量的过拟合。

在 LDP 配置下, 待编码帧的参考帧列表一般存在 4 个可用参考帧, 故网络输入设置为 4 个参考帧, 其中参考帧列表中第一帧 (即最邻近帧) 为待增强的参考帧。帧质量增强网络旨在估计出一个高质量

的参考帧放入参考帧列表中。假设参考帧列表中的参考帧对应的时刻依次为 t, t_1, t_2, t_3 ，网络生成的高质量参考帧 \hat{I}_t 表示为

$$\hat{I}_t = f_{\text{RQE}}(I_t, I_{t_1}, I_{t_2}, I_{t_3}) \quad (1)$$

其中， f_{RQE} 为本文所提出的参考帧质量增强网络， $I_t, I_{t_1}, I_{t_2}, I_{t_3}$ 为输入的参考帧。

如图 2 所示，首先，参考帧被送入偏移和掩码预测（OMP, offset and mask prediction）模块。偏移信息为参考帧对齐提供参考帧之间的运动偏移量，掩码信息为帧融合提供融合权重。

然后，利用 OMP 模块提供的偏移信息和掩码信息，本文提出的网络模型使用 DCNv2^[26] 来对齐并同时融合待增强参考帧和其他参考帧。

最后，将融合得到的特征送入质量增强（QE, quality enhancement）模块进一步处理，计算出增强的结果。为了利用残差学习的优势^[45]，QE 模块的输出与待增强参考帧相加得到增强后的参考帧。生成的高质量参考帧为

$$\begin{aligned} \hat{I}_t &= I_t + f_{\text{QE}}(f_{\text{DCN}}(I, f_{\text{OMP}}(I))) \\ I &= [I_t, I_{t_1}, I_{t_2}, I_{t_3}] \end{aligned} \quad (2)$$

其中， f_{OMP} 、 f_{DCN} 和 f_{QE} 分别为 OMP 模块、可变形卷积 DCNv2 模块和 QE 模块； I 为参考帧集合，包含 4 个参考帧。OMP 模块生成 t_1, t_2, t_3 时刻参考

帧与 t 时刻参考帧之间的偏移信息和掩码信息，如式(3)所示。

$$[o_{t_i}^{\text{DCN}}, m_{t_i}^{\text{DCN}}] \leftarrow f_{\text{OMP}}(I_t, I_{t_i}) \quad (3)$$

其中， $o_{t_i}^{\text{DCN}}$ 和 $m_{t_i}^{\text{DCN}}$ 分别代表 I_{t_i} 对应的偏移信息和掩码信息， $i \in \{1, 2, 3\}$ 。

2.1.1 偏移和掩码预测

不同参考帧之间存在不同程度的运动偏移，准确的帧对齐可以更有效地挖掘多个视频帧的信息。考虑到光流引导的可变形对齐的优异性能^[31]，如图 3 所示，本文使用光流引导偏移量预测（FGOP, flow guided offset prediction）子模块来获得基本偏移量。进一步地，使用特征学习子模块（FLSN, feature learning sub-net）获得残差偏移信息和掩码信息，进而为输入帧的每个位置提供最终偏移信息和融合权重。

如图 3 所示，以 I_t 和 I_{t_i} 为例，首先将这 2 个重建帧的亮度分量 I_t^Y 和 $I_{t_i}^Y$ 经过光流预测得到两帧亮度分量之间的光流信息 $\text{flow}_{t \rightarrow t_i}^Y$ ，再将其和 $I_{t_i}^Y$ 经过 warp 操作得到对应的 warp 图像 $I_{t_i}^{Y-w}$ 。在色度分量支路中，使用亮度分量光流信息的 $\frac{1}{2}$ 下采样作为两帧色度分量之间的光流信息，具体如式(4)所示。

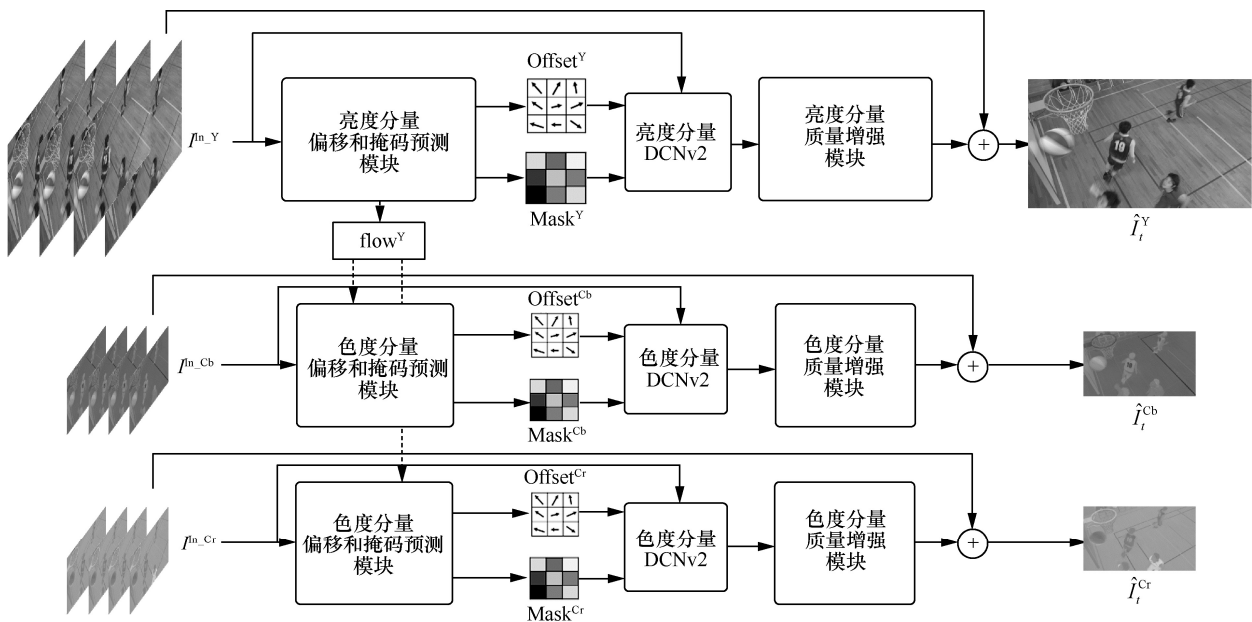


图 2 基于深度学习的参考帧质量增强网络模型框架

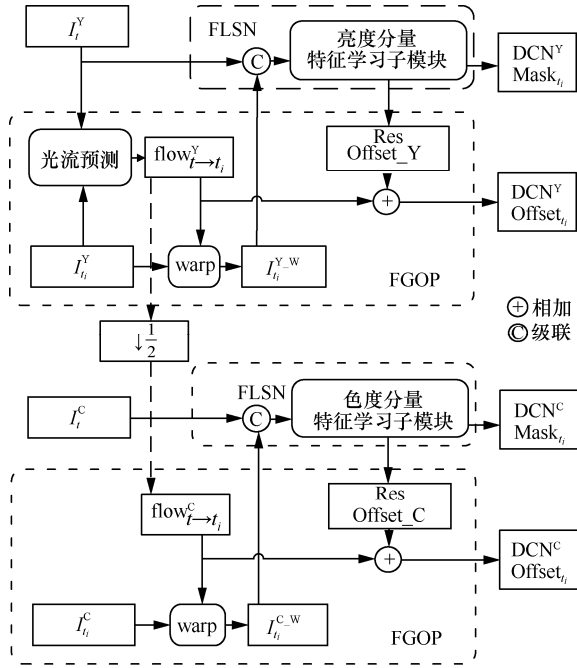


图 3 带有光流引导的偏移和掩码预测模块

$$\begin{aligned}
 \text{flow}_{t \rightarrow t_i}^Y &= \text{Flow}(I_t^Y, I_{t_i}^Y) \\
 I_{t_i}^{Y-W} &= \text{warp}(I_t^Y, \text{flow}_{t \rightarrow t_i}^Y) \\
 I_{t_i}^{C-W} &= \text{warp}(I_t^C, \text{flow}_{t \rightarrow t_i}^C)
 \end{aligned} \quad (4)$$

其中, $I_{t_i}^{Y-W}$ 为 I_t^Y 亮度分量的 warp 图像, $I_{t_i}^{C-W}$ 为 I_t^C 色度分量 Cb 或 Cr 的 warp 图像。本文使用简单有效的 SpyNet^[46]来计算光流。金字塔结构的 SpyNet 通过平均池化得到不同尺度图, 利用深度学习来处理不同尺度下的运动生成不同尺度下的光流信息并进一步细化, 模型参数更小, 在准确度和计算速度方面都有一定的优势。

基于编码-解码网络结构, 本节中的亮度分量特征学习子模块的网络结构如图 4 所示。在该结构中, 跨度为 2 的卷积层和反卷积层分别用于下采样和上采样, 亮度分量分别进行三次卷积下采样和反卷积上采样。对于跨度为 1 的卷积层, 使用零填充来保留特征大小。为了简单起见, 所有(反)卷积层的通道数为 32。这样, 每一层的内部特征的通道数是相同的。在 FLSN 中, 将输入帧经过图 4 所示的一系列处理得到残差偏移信息和掩码信息。

以亮度分量支路为例, 将 I_t^Y 和 $I_{t_i}^{Y-W}$ 送入特征学习子模块, 经计算得到亮度分量的残差偏移信息和掩码信息, 如式(5)所示。

$$[o_{t_i}^{Y-\text{res}}, m_{t_i}^{Y-\text{DCN}}] = f_{\text{FLSN}}(I_t^Y, I_{t_i}^{Y-W}) \quad (5)$$

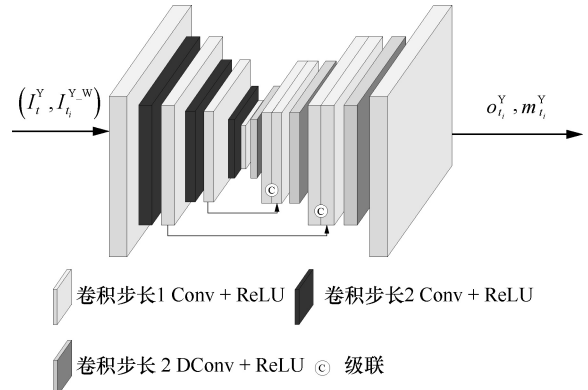


图 4 亮度分量特征学习子模块的网络结构

其中, f_{FLSN} 表示特征学习子模块。最终的偏移信息为

$$o_{t_i}^{Y-\text{DCN}} = \text{flow}_{t \rightarrow t_i}^Y + o_{t_i}^{Y-\text{res}} \quad (6)$$

色度分量支路的特征学习子模块为亮度分量的简化版本, 其网络结构如图 5 所示。

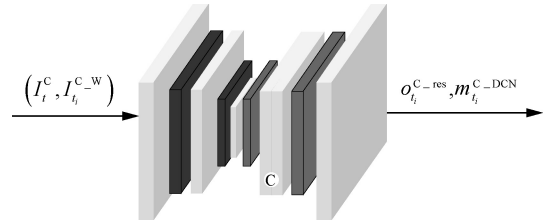


图 5 色度分量特征学习子模块的网络结构

图 5 中, I_t^C 表示图像的 Cb 或 Cr 分量, $o_{t_i}^{C-\text{res}}$ 和 $m_{t_i}^{C-\text{DCN}}$ 分别表示色度分量的偏移残差和掩码信息。

2.1.2 质量增强模块

QE 模块将帧对齐并融合后得到的特征作为输入, 充分挖掘特征图中包含的信息, 进一步提高生成帧的质量。

通过 DCNv2 得到的融合特征 (FF, fusion feature) 如式(7)所示。

$$\begin{cases}
 \text{FF}^Y = f_{\text{DCN}}(I_t^Y, [o_{t_i}^{Y-\text{DCN}}, m_{t_i}^{Y-\text{DCN}}]) \\
 \text{FF}^C = f_{\text{DCN}}(I_t^C, [o_{t_i}^{C-\text{DCN}}, m_{t_i}^{C-\text{DCN}}])
 \end{cases} \quad (7)$$

其中, FF^Y 表示亮度分量的融合特征, FF^C 表示色度分量 Cb 或 Cr 的融合特征。经过 QE 模块之后将得到最终的高质量参考帧 \hat{I}_t , 各分量的处理如式(8)所示。

$$\begin{cases}
 \hat{I}_t^Y = I_t^Y + f_{\text{QE}}^Y(\text{FF}^Y) \\
 \hat{I}_t^C = I_t^C + f_{\text{QE}}^C(\text{FF}^C)
 \end{cases} \quad (8)$$

其中, \hat{I}_t^Y 表示输出的质量增强图像的亮度分量, \hat{I}_t^c 表示输出的质量增强图像的 Cb 或 Cr 分量。

本文采用 8 层卷积网络作为亮度的质量增强模块, 使用 4 层卷积网络作为色度的质量增强模块, 分别如图 6 和图 7 所示。

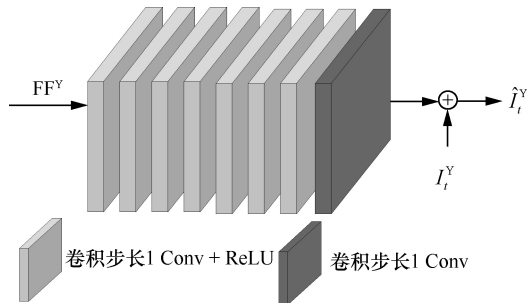


图 6 亮度分量质量增强

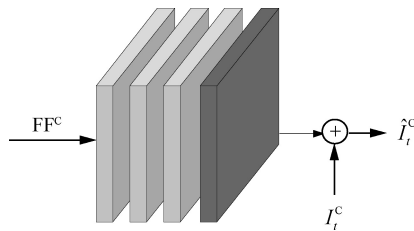


图 7 色度分量质量增强

2.2 高质量参考帧的整合

如图 1 所示, 解码图像缓存器 (DPB, decoded picture buffer) 用于存放解码重建完成的图像。每个片 (slice) 的帧间预测参考帧都由 RFL 来管理, RFL 的图像实际存储于 DPB 中。在编解码过程中, 不同时刻会为 DPB 中存在的每帧重建图像赋予一个不同的标记, 标记有两层含义, 首先区分其是否作为参考帧, 如果作为参考帧, 再区分其是长期参考帧还是短期参考帧。当图像不再作为其他图像的参考帧时, 将其从 DPB 中移除。

图 8 给出了 LDP 配置下图像之间参考关系以及各位置图像与其质量对应关系。通常, 图像组 (GoP, group of picture) 大小设置为 8, 图像间的编码顺序、解码顺序与显示顺序均相同。图 8 中的箭头方向表示参考关系, 每帧图像仅参考显示顺序排在其之前的图像, 包括其最邻近图像和 3 个前序 GoP 中质量最高的图像。图 8 中的图像使用不同的 QP 进行编码, 因此具有不同的重建质量。一般来说, QP 越小质量越高, 图 8 中白色框表示对应图像使用的 QP 最小, 其质量最高; 图像标识颜色越深, 表示其对应的 QP 越大, 质量越差。

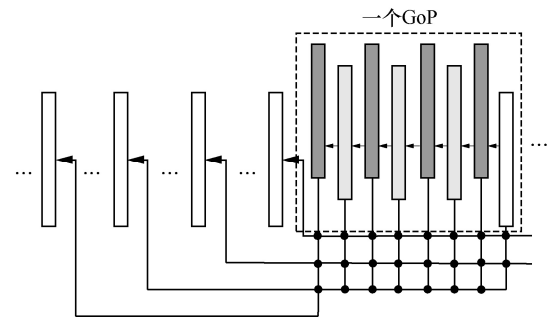


图 8 LDP 配置下图像之间参考关系以及各位置图像与其质量对应关系

观察图 8 可知, 由于采用了不同 QP, 与待编码图像时间上最接近、内容最相似的参考帧质量往往不是最好的。基于此, 本文提出对最邻近参考帧进行质量增强从而得到高质量参考帧。

本节将本文算法集成到 H.265/HEVC 参考软件平台 HM (HEVC model) 16.22 版本上, 可分为以下几个步骤。

步骤 1 在待编码帧编码之前创建一个增强参考帧缓存 I^{NN} , 并将其放入 DPB 中。本文所提出的算法对参考帧列表中第一个参考帧进行质量增强, 即图像顺序计数 (POC, picture order count) 值最大的参考帧作为待增强帧。将 I^{NN} 的 POC 值设置为 DPB 中最大的 POC 值, 同时设置一个标志信息与原参考帧进行区分。

步骤 2 LDP 配置下待编码帧仅含有前向参考帧列表, 如图 9 所示, 将前向参考帧列表的 4 个参考帧输入本文提出的网络, 进行参考帧质量增强。将增强后的参考帧赋给 I^{NN} , 并将 I^{NN} 整合到参考帧列表中。

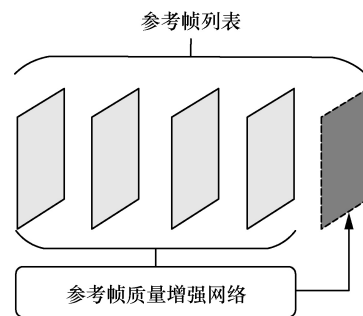


图 9 LDP 配置编码下参考帧列表优化方案

步骤 3 对待编码帧进行编码。在当前帧编码完成之后, 将 I^{NN} 从 DPB 中移除, 以免影响后续的输出及编码。

需要注意的是, 本文提出的参考帧质量增强网络的输入均为存储在 DPB 中的重建帧, 编码端和解码端的操作保持一致, 所以编码端不需要向比特流中传输任何额外的信息来标记创建的参考帧。

3 实验结果

3.1 训练细节

本文使用的数据集选自 Xiph 和 VQEG 中的 106 个序列，包含 280p~1 080p 不同分辨率。使用 H.265/HEVC 参考软件 HM16.22 对每个序列分别使用 4 个典型的 QP {22,27,32,37} 进行 LDP 配置下的编码。

常用的损失函数包括均方差 (MSE, mean squared error)、平均绝对误差 (MAE, mean absolute error) 等。本文使用式(9)所示的 MAE 作为损失函数，为方便整数运算，最终选取式(10)作为最终的损失函数。

$$MAE(I, \hat{I}) = \frac{\|I - \hat{I}\|}{WH + \frac{WH}{2} + \frac{WH}{2}} = \frac{\|Y - \hat{Y}\| + \|Cb - \hat{Cb}\| + \|Cr - \hat{Cr}\|}{\frac{3}{2}WH} = \frac{4}{6}MAE_Y + \frac{1}{6}MAE_{Cb} + \frac{1}{6}MAE_{Cr} \quad (9)$$

$$loss_{RQE} = 4MAE_Y + MAE_{Cb} + MAE_{Cr} \quad (10)$$

其中， \hat{Y} 、 \hat{Cb} 、 \hat{Cr} 分别表示本文所提出的网络输出的高质量帧各分量值； Y 、 Cb 、 Cr 作为监督信号，表示输入待增强帧所对应原始帧的各分量值。各分量的 MAE 以 4:1:1 的比例加权得到总 MAE 作为损失函数，可以更好地体现各分量的重要性。

在训练过程中，本文使用 Adam 优化算法对网络进行更新，其相应参数设置为 $\alpha = 0.0001$ ， $\beta_1 = 0.9$ ， $\beta_2 = 0.999$ ， $\epsilon = 10^{-8}$ 。

3.2 与标准参考软件的对比

本文在 H.265/HEVC 的 LDP 配置^[47]下进行了相关的测试，具体测试了 A~E 类共计 18 个视频序列，测试仍采用 4 个典型的 QP，即 {22,27,32,37}。

本节实验使用 BD-rate^[48]对测试方法进行客观评价，以 HM16.22 作为基准方案，BD-rate 值代表在相同重建视频质量下的码率节省量，若 BD-rate 为负值，则表示测试方法优于基准方案；若 BD-rate 为正值，则表示测试方法相比于基准方案需要传输更多的编码码率，引入了编码损失。

表 1 给出了相应的测试结果，其中包含三类

表 1 LDP 配置编码下引入增强参考帧在 HM16.22 下的测试结果

类别	测试序列名称	BD-rate											
		第一位			第二位			最后位			替换第一位		
		Y	Cb	Cr	Y	Cb	Cr	Y	Cb	Cr	Y	Cb	Cr
A	PeopleOnStreet	-5.29%	-11.13%	-4.14%	-6.90%	-13.56%	-7.46%	-6.11%	-14.09%	-8.64%	-3.98%	-9.01%	-1.14%
	Traffic	-6.26%	-8.92%	-5.30%	-8.53%	-11.25%	-8.39%	-8.89%	-12.11%	-9.47%	-2.46%	-4.55%	-0.28%
	平均	-5.78%	-10.02%	-4.72%	-7.72%	-12.40%	-7.92%	-7.50%	-13.10%	-9.06%	-3.22%	-6.78%	-0.71%
B	Kimono	-8.28%	-15.60%	-3.92%	-9.12%	-16.32%	-5.21%	-8.57%	-15.87%	-5.89%	-7.34%	-14.45%	-1.97%
	ParkScene	-3.35%	-8.66%	-2.57%	-5.13%	-10.56%	-4.93%	-5.19%	-11.16%	-5.85%	-1.20%	-5.26%	0.38%
	Cactus	-9.62%	-18.73%	-14.65%	-11.25%	-20.21%	-16.45%	-11.13%	-20.24%	-16.82%	-5.34%	-14.80%	-10.32%
	BasketballDrive	-5.40%	-17.14%	-13.80%	-7.36%	-19.09%	-15.84%	-6.92%	-18.27%	-15.46%	-3.41%	-15.66%	-10.96%
	BQTerrace	-13.13%	-24.33%	-22.21%	-14.21%	-25.09%	-23.51%	-14.13%	-25.60%	-23.82%	-11.60%	-23.68%	-22.05%
平均	-7.96%	-16.89%	-11.43%	-9.41%	-18.25%	-13.19%	-9.19%	-18.23%	-13.57%	-5.78%	-14.77%	-8.98%	
C	BasketballDrill	-6.93%	-15.53%	-16.37%	-8.29%	-17.03%	-18.21%	-8.12%	-16.81%	-17.27%	-5.20%	-14.03%	-14.09%
	BQMall	-5.87%	-13.16%	-12.40%	-7.93%	-15.06%	-14.24%	-7.58%	-15.07%	-14.32%	-2.92%	-10.03%	-8.53%
	PartyScene	-6.17%	-13.32%	-10.87%	-7.25%	-14.73%	-12.26%	-7.02%	-14.74%	-12.47%	-5.80%	-12.72%	-9.59%
	RaceHorses	-3.09%	-7.08%	-9.80%	-5.02%	-9.67%	-13.12%	-4.37%	-9.21%	-12.64%	-0.55%	-3.19%	-4.84%
	平均	-5.52%	-12.27%	-12.36%	-7.13%	-14.12%	-14.46%	-6.77%	-13.96%	-14.18%	-3.62%	-9.99%	-9.26%
D	BasketballPass	-6.75%	-17.39%	-14.50%	-8.23%	-19.56%	-15.68%	-7.00%	-18.61%	-15.16%	-4.73%	-16.29%	-11.93%
	BQSquare	-9.48%	-12.02%	-15.88%	-11.01%	-12.61%	-17.53%	-10.83%	-13.07%	-18.59%	-9.64%	-9.59%	-15.09%
	BlowingBubbles	-6.31%	-14.14%	-10.94%	-7.74%	-15.50%	-13.47%	-7.67%	-15.46%	-14.29%	-4.21%	-12.10%	-8.28%
	RaceHorses	-4.15%	-7.01%	-10.07%	-6.40%	-9.85%	-12.13%	-5.54%	-9.00%	-12.37%	-1.19%	-2.52%	-6.97%
	平均	-6.67%	-12.64%	-12.85%	-8.35%	-14.38%	-14.70%	-7.76%	-14.03%	-15.10%	-4.94%	-10.12%	-10.57%
E	FourPeople	-10.45%	-14.50%	-14.32%	-11.99%	-15.91%	-16.12%	-12.09%	-15.89%	-16.08%	-6.29%	-10.51%	-9.77%
	Johnny	-12.65%	-8.54%	-9.01%	-15.21%	-12.55%	-11.35%	-15.02%	-13.76%	-12.08%	-8.11%	-3.33%	-3.34%
	KristenAndSara	-8.95%	-6.43%	-8.90%	-11.48%	-10.03%	-11.46%	-11.74%	-11.25%	-12.03%	-4.52%	-0.21%	-2.42%
	平均	-10.68%	-9.82%	-10.75%	-12.89%	-12.83%	-12.98%	-12.95%	-13.63%	-13.40%	-6.31%	-4.68%	-5.17%
总平均	-7.34%	-12.98%	-11.09%	-9.06%	-14.92%	-13.19%	-8.77%	-15.01%	-13.51%	-4.92%	-10.11%	-7.84%	

内容,即不同测试序列对应的类别、名称、3种颜色分量的性能。为充分分析增强参考帧放在参考帧列表不同位置造成的影响,本文方案将增强参考帧分别放到参考帧列表的第一位、第二位和最后位进行测试。同时,为了验证不改变参考帧列表长度下引入增强参考帧的性能,本文也测试了将增强参考帧替换最邻近参考帧的性能。

从表 1 中可以看出,基于最邻近帧质量增强的参考帧列表优化算法将增强参考帧放到参考帧列表第二位时,获得的性能最高。在 Y、Cb、Cr 这 3 个分量上平均得到-9.06%、-14.92%、-13.19%的 BD-rate 增益。同时,在所有测试序列上都得到了增益,从 Y 分量来看,可得到-15.21%~5.02%的 BD-rate 增益。在不改变参考帧列表长度的情况下,将增强参考帧替换最邻近参考帧也可带来一定的 BD-rate 增益,在 Y、Cb、Cr 这 3 个分量分别得到-4.92%、-10.11%、-7.84%的 BD-rate 增益。在相同视频质量前提下,所提算法均可降低编解码率。

在测试时,本文深度学习网络部分使用 GPU

进行运算,编解码器的其他部分均使用 CPU。表 2 给出了增强参考帧放到参考帧列表中不同位置时的编解码时间比。

表 2 增强参考帧放到参考帧列表中不同位置时的编解码时间比

时间比	第一位	第二位	最后位	替换第一位
编码时间比	125%	124%	127%	117%
解码时间比	5 464%	5 464%	5 465%	5 500%

由表 2 可知,将增强参考帧放到参考帧列表中不同位置时的编解码时间接近,其中,编码时间约为 HM16.22 的 1.2 倍,解码时间约为 HM16.22 的 55 倍。

基于以上结果,本文选择将质量增强后的视频帧插入待编码帧的参考帧列表中的第二个位置。

3.3 与现有方法的对比

本节将所提算法与文献[39,44]方法在相同的条件下进行测试和比较。由于 HM16.20 的 LDP 配置编码使用的 GoP 为 4,与 HM16.22 (GoP 为 8)有较大差异,本文将参考帧质量增强网络在 HM16.20 上重新训练并测试,如表 3 所示。

表 3 与现有方法的编解码性能比较

类别	测试序列名称	BD-rate								
		文献[39]方法			文献[44]方法			本文算法		
		Y	Cb	Cr	Y	Cb	Cr	Y	Cb	Cr
A	PeopleOnStreet	-5.8%	-4.0%	-4.4%	-8.2%	-10.3%	-11.8%	-7.2%	-13.8%	-7.5%
	Traffic	-3.6%	-3.8%	-3.1%	-5.8%	-11.0%	-11.8%	-8.3%	-8.7%	-6.3%
	平均	-4.7%	-3.9%	-3.8%	-7.0%	-10.7%	-11.8%	-7.8%	-11.2%	-6.9%
B	Kimono	-5.1%	-7.3%	-3.7%	-8.9%	-22.2%	-4.6%	-8.5%	-12.5%	-3.7%
	ParkScene	-3.3%	-3.9%	-3.8%	-4.7%	-17.4%	-5.5%	-4.7%	-9.4%	-2.8%
	Cactus	-6.9%	-10.3%	-6.0%	-10.7%	-19.5%	-9.4%	-10.4%	-17.1%	-11.7%
	BasketballDrive	-2.8%	-5.6%	-3.7%	-4.3%	-12.3%	-12.0%	-7.1%	-18.1%	-16.1%
	BQTerrace	-4.2%	-3.6%	-1.4%	-9.6%	-9.1%	-16.3%	-14.6%	-16.4%	-12.6%
平均	-4.5%	-6.4%	-3.7%	-7.6%	-16.1%	-9.6%	-9.0%	-14.7%	-9.4%	
C	BasketballDrill	-2.9%	-5.2%	-3.0%	-5.5%	-9.3%	-9.3%	-9.5%	-19.4%	-16.9%
	BQMall	-6.0%	-7.3%	-6.7%	-9.3%	-19.8%	-20.9%	-8.0%	-14.6%	-13.0%
	PartyScene	-3.3%	-3.9%	-3.8%	-4.9%	-12.2%	-11.7%	-6.4%	-11.9%	-8.7%
	RaceHorses	-1.1%	-1.5%	-1.9%	-1.9%	-4.3%	-5.8%	-5.3%	-9.0%	-11.7%
	平均	-3.3%	-4.5%	-3.9%	-5.4%	-11.4%	-11.9%	-7.3%	-13.7%	-12.6%
D	BasketballPass	-4.4%	-5.7%	-4.6%	-5.9%	-10.5%	-8.6%	-6.9%	-16.8%	-13.4%
	BQSquare	-3.2%	-0.6%	1.5%	-6.8%	-4.1%	-11.3%	-9.8%	-9.4%	-10.9%
	BlowingBubbles	-4.1%	-5.2%	-5.3%	-5.9%	-13.0%	-12.5%	-7.5%	-13.9%	-10.6%
	RaceHorses	-1.8%	-2.7%	-3.2%	-2.8%	-5.6%	-6.9%	-6.3%	-9.6%	-11.8%
	平均	-3.4%	-3.6%	-2.9%	-5.4%	-8.3%	-9.8%	-7.6%	-12.4%	-11.7%
E	FourPeople	-10.1%	-11.9%	-11.6%	-12.2%	-13.1%	-17.2%	-12.6%	-12.6%	-12.4%
	Johnny	-8.6%	-9.1%	-7.5%	-11.3%	-9.2%	-6.9%	-15.5%	-5.1%	-8.1%
	KristenAndSara	-8.7%	-22.2%	-9.1%	-12.4%	-12.0%	-14.4%	-11.9%	-6.3%	-9.2%
	平均	-9.1%	-14.4%	-9.4%	-12.0%	-11.4%	-12.8%	-13.3%	-8.0%	-9.9%
总平均	-4.8%	-5.7%	-4.5%	-7.3%	-12.0%	-10.9%	-8.9%	-12.5%	-10.4%	

由表 3 可知, 所提算法在 Y、Cb、Cr 这 3 个分量分别获得了-8.9%、-12.5%、-10.4%的 BD-rate 增益。将所提算法与文献[39,44]方法的最优性能进行对比, 可见其在不同分辨率的序列上都取得了最优的性能。

正如前文所述, 在 LDP 配置下, 文献[39,44]方法都利用当前时刻之前的重建帧进行外插来得到虚拟参考帧, 旨在生成与待编码帧内容一致的参考帧。然而, 因为无法避免运动遮挡问题, 虚拟参考帧中往往存在部分虚构内容, 从而影响帧间预测的准确性。而本文算法仅对已有的最邻近参考帧进行质量增强, 避免了因虚构内容带来的预测误差, 同时经过增强的参考帧内容往往具有较高的预测精度, 有助于提高帧间编码效率。

文献[39,44]方法经过神经网络处理得到的帧可直接作为编码块的预测值, 从而减少了大量运动信息的传输, 这是本文算法无法达到的效果。后续研究可以考虑综合这 2 种方法的优势进行组合设计, 达到更高的压缩性能。

4 结束语

本文提出了一种基于最邻近参考帧质量增强的参考帧列表优化算法, 以提高在 LDP 配置下的视频编码效率。具体做法是先应用光流引导可变形卷积的偏移预测, 再结合多参考帧通过 QE 网络对最邻近参考帧进行质量增强, 将增强后的参考帧作为额外的高质量参考帧插入参考帧列表。本文实验结果表明, 所提算法可以提高低时延场景下视频帧间编码的效率。

本文将基于神经网络的视频帧质量增强应用于参考帧列表优化算法中, 为视频编码标准的研究提供了可行的研究方向。然而将基于深度学习的网络模型与传统编码框架相结合大大增加了编解码复杂度, 下一步可考虑对神经网络进行压缩和简化以降低复杂度。同时, 随着软硬件计算能力的提升, 设计更高效的质量增强网络应用到参考帧列表优化中, 也可进一步提升视频编码性能。

参考文献:

[1] SULLIVAN G J, OHM J R, HAN W J, et al. Overview of the high efficiency video coding (HEVC) standard[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2012, 22(12): 1649-1668.

[2] 毕厚杰. 新一代视频压缩编码标准—H.264/AVC[M]. 北京: 人民邮

电出版社, 2005.

- BI H J. A new generation of video compression and coding standards-H.264 /AVC[M]. Beijing: Posts and Telecommunications Press, 2005.
- [3] DONG C, DENG Y B, LOY C C, et al. Compression artifacts reduction by a deep convolutional network[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Piscataway: IEEE Press, 2015: 576-584.
- [4] PARK W S, KIM M. CNN-based in-loop filtering for coding efficiency improvement[C]//*Proceedings of 2016 IEEE 12th Image, Video, and Multidimensional Signal Processing Workshop*. Piscataway: IEEE Press, 2016: 1-5.
- [5] ZHANG Y B, SHEN T, JI X Y, et al. Residual highway convolutional neural networks for in-loop filtering in HEVC[J]. *IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society*, 2018, 27(8): 3827-3841.
- [6] XIA J Y, WEN J T. Asymmetric convolutional residual network for AV1 intra in-loop filtering[C]//*Proceedings of 2020 IEEE International Conference on Image Processing*. Piscataway: IEEE Press, 2020: 1291-1295.
- [7] ZHANG S F, FAN Z H, LING N, et al. Recursive residual convolutional neural network-based in-loop filtering for intra frames[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(7): 1888-1900.
- [8] DAI Y Y, LIU D, WU F. A convolutional neural network approach for post-processing in HEVC intra coding[C]//*IEEE International Conference on Multimedia Modeling*. Piscataway: IEEE Press, 2017: 28-39.
- [9] LAI P R, WANG J S. Multi-stage attention convolutional neural networks for HEVC in-loop filtering[C]//*Proceedings of 2020 2nd IEEE International Conference on Artificial Intelligence Circuits and Systems*. Piscataway: IEEE Press, 2020: 173-177.
- [10] DING D D, KONG L Y, CHEN G Y, 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.
- [11] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011-2023.
- [12] KANG J H, KIM S, LEE K M. Multi-modal/multi-scale convolutional neural network based in-loop filter design for next generation video codec[C]//*Proceedings of 2017 IEEE International Conference on Image Processing*. Piscataway: IEEE Press, 2017: 26-30.
- [13] HE X Y, HU Q, ZHANG X Y, et al. Enhancing HEVC compressed videos with a partition-masked convolutional neural network[C]//*Proceedings of 2018 25th IEEE International Conference on Image Processing*. Piscataway: IEEE Press, 2018: 216-220.
- [14] LI D W, YU L. An in-loop filter based on low-complexity CNN using residuals in intra video coding[C]//*Proceedings of 2019 IEEE Interna-*

- tional Symposium on Circuits and Systems. Piscataway: IEEE Press, 2019: 1-5.
- [15] ZHU H, XU X Z, LIU S. Residual convolutional neural network based in-loop filter with intra and inter frames processed respectively for Avs3[C]//Proceedings of 2020 IEEE International Conference on Multimedia & Expo Workshops. Piscataway: IEEE Press, 2020: 1-6.
- [16] JIA C M, WANG S Q, ZHANG X F, et al. Spatial-temporal residue network based in-loop filter for video coding[C]//Proceedings of 2017 IEEE Visual Communications and Image Processing. Piscataway: IEEE Press, 2017: 1-4.
- [17] LU G, OUYANG W L, XU D, et al. Deep Kalman filtering network for video compression artifact reduction[C]//Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2018: 568-584.
- [18] YANG R, SUN X Y, XU M, et al. Quality-gated convolutional LSTM for enhancing compressed video[C]//Proceedings of 2019 IEEE International Conference on Multimedia and Expo. Piscataway: IEEE Press, 2019: 532-537.
- [19] LU M, CHENG M, XU Y L, et al. Learned quality enhancement via multi-frame priors for HEVC compliant low-delay applications[C]//Proceedings of IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2019: 934-938.
- [20] TONG J C, WU X L, DING D D, et al. Learning-based multi-frame video quality enhancement[C]//Proceedings of 2019 IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2019: 929-933.
- [21] MENG X D, DENG X, ZHU S Y, et al. A robust quality enhancement method based on joint spatial-temporal priors for video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(6): 2401-2414.
- [22] DOSOVITSKIY A, FISCHER P, ILG E, et al. FlowNet: learning optical flow with convolutional networks[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2016: 2758-2766.
- [23] SUN D Q, YANG X D, LIU M Y, et al. PWC-net: CNNs for optical flow using pyramid, warping, and cost volume[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 8934-8943.
- [24] YANG R, XU M, WANG Z L, et al. Multi-frame quality enhancement for compressed video[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 6664-6673.
- [25] GUAN Z Y, XING Q L, 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.
- [26] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 764-773.
- [27] ZHU X Z, HU H, LIN S, et al. Deformable ConvNets V2: more deformable, better results[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 9300-9308.
- [28] DENG J N, WANG L, PU S L, et al. Spatio-temporal deformable convolution for compressed video quality enhancement[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 10696-10703.
- [29] ZHAO M Y, XU Y, ZHOU S G. Recursive fusion and deformable spatiotemporal attention for video compression artifact reduction[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM Press, 2021: 5646-5654.
- [30] LIN J Y, HUANG Y, WANG L. FDAN: flow-guided deformable alignment network for video super-resolution[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 1-10.
- [31] CHAN K C K, ZHOU S C, XU X Y, et al. BasicVSR++: improving video super-resolution with enhanced propagation and alignment[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 5972-5981.
- [32] DING D D, WANG W Y, TONG J C, et al. Biprediction-based video quality enhancement via learning[J]. IEEE Transactions on Cybernetics, 2022, 52(2): 1207-1220.
- [33] WANG X T, CHAN K C K, YU K, et al. EDVR: video restoration with enhanced deformable convolutional networks[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2019: 1954-1963.
- [34] YAN N, LIU D, LI H Q, et al. A convolutional neural network approach for half-pel interpolation in video coding[C]//Proceedings of 2017 IEEE International Symposium on Circuits and Systems. Piscataway: IEEE Press, 2017: 1-4.
- [35] ZHANG H, SONG L, LUO Z Y, et al. Learning a convolutional neural network for fractional interpolation in HEVC inter coding[C]//Proceedings of 2017 IEEE Visual Communications and Image Processing. Piscataway: IEEE Press, 2017: 1-4.
- [36] YAN N, LIU D, LI H Q, et al. Convolutional neural network-based fractional-pixel motion compensation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(3): 840-853.
- [37] YAN N, LIU D, LI H Q, et al. Convolutional neural network-based invertible half-pixel interpolation filter for video coding[C]//Proceedings of 2018 25th IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2018: 201-205.
- [38] ZHAO Z H, WANG S Q, WANG S 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, 2019, 29(11): 3291-3301.

- [39] CHOI H, BAJIĆ I V. Deep frame prediction for video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(7): 1843-1855.
- [40] LIN J P, LIU D, LI H Q, et al. Generative adversarial network-based frame extrapolation for video coding[C]//Proceedings of 2018 IEEE Visual Communications and Image Processing. Piscataway: IEEE Press, 2018: 1-4.
- [41] ZHAO L, WANG S Q, ZHANG X F, et al. Enhanced CTU-level inter prediction with deep frame rate up-conversion for high efficiency video coding[C]//Proceedings of 2018 25th IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2018: 206-210.
- [42] ZHAO L, WANG S Q, ZHANG X F, et al. Enhanced motion-compensated video coding with deep virtual reference frame generation[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2019, 28(10): 4832-4844.
- [43] LEE J K, KIM N, CHO S, et al. Deep video prediction network-based inter-frame coding in HEVC[J]. IEEE Access, 2020, 8: 95906-95917.
- [44] CHOI H, BAJIĆ I V. Affine transformation-based deep frame prediction[J]. IEEE Transactions on Image Processing, 2021, 30: 3321-3334.
- [45] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 1646-1654.
- [46] RANJAN A, BLACK M J. Optical flow estimation using a spatial pyramid network[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 4161-4170.
- [47] BOSSEN F. Common HM test conditions and software reference configuration[R]. 2013.
- [48] BJØNTEGAARD G. Calculation of average PSNR differences between RD-curves[R]. 2001.

[作者简介]



霍俊彦(1982-),女,山西晋中人,博士,西安电子科技大学副教授,主要研究方向为多媒体通信、虚拟现实、智能信息处理。



邱瑞鹏(1996-),男,河南上蔡人,西安电子科技大学硕士生,主要研究方向为视频压缩。



马彦卓(1980-),女,河北深州人,博士,西安电子科技大学副教授,主要研究方向为视频编码与视频传输。



杨付正(1977-),男,山东德州人,博士,西安电子科技大学教授、博士生导师,主要研究方向为新一代视频压缩标准、基于深度学习的视频处理和虚拟现实。