

文章编号: 2095-4980(2020)02-0291-07

## 结合卷积神经网络的 HEVC 帧内编码压缩改进算法

王 婷, 何小海\*, 孙伟恒, 熊淑华, Karn Pradeep

(四川大学 电子信息学院, 四川 成都 610065)

**摘要:** 近年来, 卷积神经网络深度学习已在图像处理、目标检测等领域取得巨大成功。受其启发, 将卷积神经网络(CNN)应用于传统视频压缩标准已成为一个新的研究热点。本文提出一种集成卷积神经网络的高效视频编码(HEVC)压缩改进算法, 将下采样过程、HEVC 的编解码过程、上采样及质量增强过程集成为一体。为高效提取视频帧的结构特征, 在所提压缩算法中集成了两个卷积神经网络。提出了一种下采 CNN(DwSCNN)代替双三次下采, 在有效降低分辨率的同时保留细节信息, 得到更为紧凑的低分辨率视频序列, 将此低分辨率视频序列通过 HEVC 帧内编码进行进一步的数据量压缩, 通过提出一个质量增强 CNN(PPCNN)来改善解码后恢复到原始分辨率的降质视频序列。实验结果显示, 本文压缩改进算法在低码率段与标准 HEVC 相比, 能达到更好的质量重建, 并且在接近一致的 PSNR 值时, 能节省 39.46% 的时间和 11.04% 的比特率, 本文算法的视频压缩性能优于 HEVC 标准算法和相关文献方法。

**关键词:** 卷积神经网络; 集成; 高效视频编码; 下采样

**中图分类号:** TN919.81

**文献标志码:** A

**doi:** 10.11805/TKYDA2019065

## Improved HEVC intra coding compression algorithm combined with convolutional neural network

WANG Ting, HE Xiaohai\*, SUN Weiheng, XIONG Shuhua, Karn Pradeep

(College of Electronics and Information Engineering, Sichuan University, Chengdu Sichuan 610065, China)

**Abstract:** In recent years, deep learning of convolutional networks has achieved great success in the fields of image processing and target detection. It has become a new research hotspot to apply Convolutional Neural Networks(CNN) to traditional video compression standards. An improved High Efficiency Video Coding(HEVC) compression algorithm integrated with convolutional neural network is proposed, which integrates down-sampling, HEVC codec, up-sampling and quality enhancement process. In order to extract the structural features of video frames efficiently, two convolutional neural networks are integrated in the proposed compression algorithm. Down Sampling CNN(DwSCNN) replaces bicubic down-sampling, which preserves the detailed information while reducing the resolution, obtaining a more compact low resolution video sequence. The low-resolution video sequence is further compressed by HEVC intra coding, and a quality-enhanced Post Processing CNN(PPCNN) is proposed to improve the degraded video sequence that is restored to the original resolution after decoding. The experimental results show that the proposed compression improvement algorithm can achieve better quality reconstruction than the standard HEVC in the low code rate segment, and can save time by 39.46% and bit rate by 11.04% when the PSNR value is close to the same. The video compression performance of the algorithm is superior to the HEVC standard algorithm and other related literature methods.

**Keywords:** Convolutional Neural Network; integrated; HEVC; down-sampling

收稿日期: 2019-03-01; 修回日期: 2019-04-03

基金项目: 国家自然科学基金资助项目(61871279,61471248); 成都市产业集群协同创新项目资助(2016-XT00-00015-GX)

作者简介: 王 婷(1994-), 女, 在读硕士研究生, 主要研究方向为图像处理与网络通信。email:793681717@qq.com

\*通信作者: 何小海 email:hxh@scu.edu.com

近年来,智能移动终端的普及使得人们对视频分辨率的要求逐渐提高。在此背景下,国际电信联盟 ITU-T 与国际标准化组织 ISO/IEC 合作发布了高效视频编码标准 H.265/HEVC。H.265/HEVC 包含最新的视频编码技术,相较于 H.264/AVC(Advanced Video Coding)在同等编码质量下能节省 50%左右的码率<sup>[1]</sup>。现今,应用于卫星、智能家居等领域的高清数字电影的拍摄,均采用帧内编码方式进行压缩,HEVC 帧内编码的应用日益广泛,受终端设备编解码方式和其固定的带宽限制,迫切需要在视频帧内压缩标准上进一步实现视频高性能压缩。为实现高质量的视频压缩,结合下采样与超分辨率重建工作也越来越多。文献[2]采用一种基于块自适应下采样和超分辨率重建方法改善低码率图像编码方法。文献[3]通过找到空域下采样和视频编码引起失真的最优下采样比,实现低码率段视频最优压缩性能。文献[2]和文献[3]能一定程度地优化视频压缩性能,但受传统算法约束,并不能带来较大突破。近几年,卷积神经网络已在目标检测、图像处理领域取得了成功,实验证明其能更好地提取视频帧的结构特征和细节信息,因此也有学者提出了可通过卷积神经网络来改善和增强 HEVC 编码视频帧质量。文献[4]提出一种基于帧的动态元数据嵌入的后处理方案,将视频内容复杂度作为辅助信息嵌入比特流指导 CNN 重建解码后的视频,平均减少 2.4%的 BD-rate。文献[5]提出多帧质量增强方法,利用视频特性,使用相邻的高质量帧来增强低质量帧。这些采用下采样和超分辨率重建的工作取得了一定成果,随着技术的推进和硬件设备的完善,开始将字典学习用于优化 HEVC 压缩性能,文献[6]设计了编码视频帧的压缩,通过训练好的字典重建高质量视频。以上提到的这些方法往往下采、编码和超分辨率重建三个过程分离,处理过程较为繁杂。即使最终获得压缩性能的提升,但处理过程并不高效。受文献[6]和文献[7]启发,本文提出一种将下采、HEVC 编码、上采和后处理集成于一体的端到端压缩,在提高 HEVC 帧内编码压缩效率的同时,改进压缩编码数据处理过程。

## 1 结合卷积神经网络的 HEVC 帧内编码压缩改进算法

视频经 HEVC 压缩后其质量将严重衰减,尤其是在低码率段。因此,在 HEVC 解码端进行视频视觉质量的提升十分有必要<sup>[8]</sup>。如图 1 所示,本文提出了一种结合卷积神经网络的 HEVC 帧内编码压缩改进算法,能有效压缩传输数据量并改善压缩图像的质量。所提压缩改进算法由两个 CNN 网络和 HEVC 编解码器组成,首先使用提出的下采卷积神经网络(DwSCNN)代替双三次下采,尽可能地保留细节信息,得到更为紧凑的低分辨率视频序列。然后将此低分辨率视频序列通过 HEVC 进行编码,最后通过所提出的质量增强卷积神经网络(PPCNN)来改善解码后恢复到原始空间分辨率的降质视频序列。

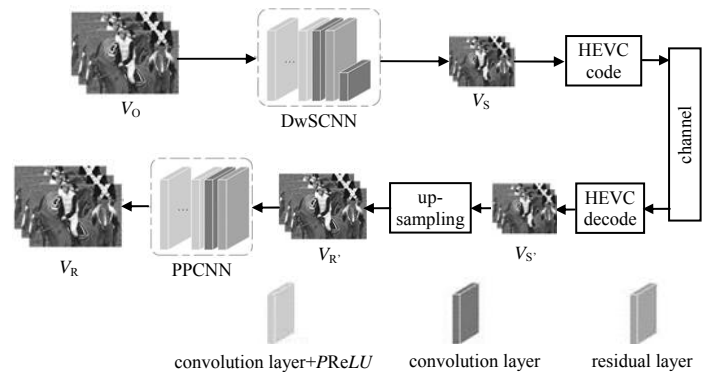


Fig.1 Framework of the proposed method

图 1 本文方法框架

在编码前,将原视频帧  $V_0$  通过离线训练好的最佳下采模型下采,得到更为丰富的低空间分辨率视频帧  $V_S$ ,其长宽均被采样为原视频图像的 1/2。将得到的更为紧凑的小视频  $V_S$  进行 HEVC 编码,得到小视频编码码流。然后对小视频编码码流进行解码,解码得到的小视频  $V_{S'}$  先进行双三次上采,恢复到原始分辨率的低质大视频  $V_{R'}$ 。最后将质量损失严重的  $V_{R'}$  视频帧通过训练好的最佳后处理模型进行质量增强。

## 2 下采卷积神经网络

### 2.1 下采卷积神经网络结构

对高清视频的追求意味着传输较高的数据量,在对视频进行编码前将其进行下采样,在保留原始视频信息的同时传输更少的数据码流。下采网络的性能将很大程度决定重建视频质量,因此需要设计合适的下采网络捕获更多视频帧的结构信息与细节信息。

如图 2 所示,本文提出的基于卷积神经网络的下采网络 DwSCNN 共有 8 个卷积层。前 6 个卷积核分别紧跟一个参数化修正线性单元(PReLU)作为非线性激活函数,用于提取帧内编码特征;第 7 个卷积层用于增强特征映射,提取更能表达重建图像的特征;为确保输入视频帧和重建残差大小一致,需要在尺寸下采前采用残差学

习，残差学习的引入可加快网络收敛，并进一步获得准确的图像高频信息；其中，第 8 个卷积层步长为 2，用于实现分辨率的降低。

DwSCNN 的配置如表 1 所列，DwSCNN 的公式表示为：

$$F_0(Y) = Y \quad (1)$$

$$F_i(Y) = PReLU(W_i \bullet F_{i-1}(Y) + B_i), i \in \{1, 2, 3, 4, 5, 6\} \quad (2)$$

$$F_7(Y) = W_7 \bullet F_6(Y) + B_7 \quad (3)$$

$$F_8(Y) = F_0(Y) + F_7(Y) \quad (4)$$

$$F_9(Y) = W_9 \bullet F_8(Y)_{\downarrow 2} + B_9 \quad (5)$$

式中： $Y$  为原始输入视频帧； $W_i$  和  $B_i$  分别为第  $i$

层的权重和偏置； $\downarrow_2$  表示该层卷积步长为 2，旨在最后实现空间分辨率的下采，减少高频信息的丢失。值得说明的是， $PReLU$  是被应用到 1~6 层卷积层的非线性激活函数。

表 1 DwSCNN 的配置

Table 1 Configuration of DwSCNN

layer	Conv1	Conv2	Conv3	Conv4	Conv5	Conv6	Conv7	Conv8
filter size	3×3	3×3	3×3	3×3	3×3	3×3	3×3	9×9
filter number	64	64	64	64	64	64	1	1
stride	1	1	1	1	1	1	1	2
W learning rate	10 <sup>-4</sup>	10 <sup>-4</sup>	10 <sup>-4</sup>	10 <sup>-4</sup>	10 <sup>-4</sup>	10 <sup>-4</sup>	10 <sup>-4</sup>	10 <sup>-5</sup>
B learning rate	10 <sup>-5</sup>	10 <sup>-5</sup>	10 <sup>-5</sup>	10 <sup>-5</sup>	10 <sup>-5</sup>	10 <sup>-5</sup>	10 <sup>-5</sup>	10 <sup>-5</sup>

通常认为构建较深的网络能获得更好的性能提升，例如在图像感知领域著名的 GooLeNet<sup>[9]</sup>和 ResNet<sup>[10]</sup>。但是最新成果表明，网络最终性能取决于结构和应用场景。如图 2 相似的结构，在相同条件下将总卷积层数设置为 4,6,8,10,12，其中 8 个卷积层为本文所提结构，增加和减少的卷积层为 64×3×3 的卷积层。用以上所述 5 个不同深度的 DwSCNN 训练后，RaceHorses\_416×240 序列用 5 个深度下的下采模型采样到低分辨率视频，再进行双三次上采恢复到原始分辨率，从图 3 可知 5 个深度下的迭代次数和此时重建视频质量之间的关系，8 层的 DwSCNN 在训练期间收敛较快并且能获得最高的 PSNR。更浅的深度虽然能较快收敛，但是达不到更高的质量重建；在增加了网络深度后，带来的代价是收敛变慢并且最终重建帧质量也并不高，甚至由于样本库有限而模型比较复杂，10 层和 12 层的 DwSCNN 重建时性能并不稳定，出现了较为明显的过拟合现象，在某些迭代次数上出现质量的急剧下降。图 3 直观表明，在处理视频帧空间分辨率降低时，选择以上介绍的含 8 个卷积层的 DwSCNN 能保留更加丰富的视频信息。

### 2.2 下采卷积神经网络训练算法

为实现重建视频帧与原始视频帧之间最小的损失，采用均方误差(Mean Squared Error, MSE)作为网络的损失函数。 $\{X_n\}_{n=1}^N$  为原始视频帧的图像块， $\{Y_n\}_{n=1}^N$  是相对应的增加空间分辨率的视频帧图像块。 $\{Y_n\}_{n=1}^N$  是输入增大空间分辨率的样本， $\{X_n\}_{n=1}^N$  为对应的目标标签。 $F(\bullet)$  为 DwSCNN 的输出函数，网络损失函数表示为：

$$L(\theta_1) = \frac{1}{N} \sum_{n=1}^N \|F(Y_n; \theta_1)_{\downarrow 2} - X_n\|_2^2 \quad (6)$$

式中： $\theta_1 = \{W_i, B_i\}$  表示 DwSCNN 的权重和偏置； $\downarrow_2$  表示在网络输出前采用了一个步长为 2 的卷积层。

训练集和验证集样本来自于 BSDS500 数据库<sup>[11]</sup>，将 BSDS500 里的训练图库和测试图库共同用于 DwSCNN 的训练图库。为了准备训练和验证数据，将 400 张原始训练图和 100 张验证图以缩放因子为 2 上采样为高分辨率图像。为扩大样本数和更加精确恢复图像结构特征，将 2 个图库进行 4 个方向的旋转和 3 个尺度的缩放，再

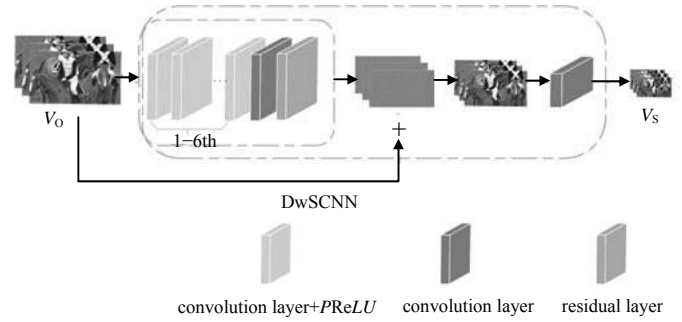


Fig.2 Down-sampling CNN(DwSCNN)  
图 2 下采卷积神经网络(DwSCNN)

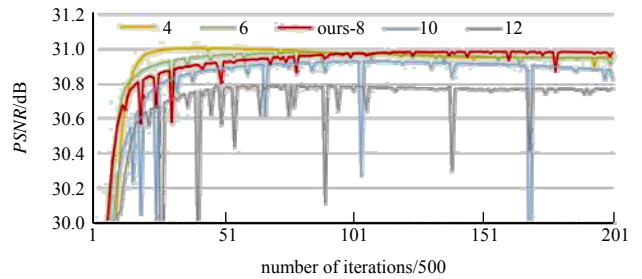


Fig.3 Average PSNR over the RaceHorses\_416×240 by DwSCNN with various depths down-sampling and bicubic up-sampling  
图 3 在 RaceHorses\_416×240 上采用 DwSCNN 在不同深度中下采和双三次上采后的平均 PSNR

把扩展的原始图库和高分辨率图库进行分割,生成训练样本和验证样本。考虑到 CTU 的大小为 64,32,16,但是学习 64×64 大小的标签没有成功<sup>[12]</sup>,因此 DwSCNN 中以步长 32 将高分辨率图像切割成 32×32,原始图像切割成 16×16 的样本。按照以上处理方法,将获得 1 920 000 张训练样本和 480 000 张验证样本。实验时使用深度学习 CAFFE<sup>[13]</sup>来训练网络。采用 Adam 算法<sup>[14]</sup>,基础学习率为 0.000 1,权重衰减因子为 0.000 1,动量设为 0.9。

2.3 下采卷积神经网络性能

通过降低传输视频的空间分辨率来减少传输数据量,同时保证传送信息不丢失。双三次下采无法自适应地保留视频序列结构纹理。基于深度学习的 DwSCNN 相较于双三次下采能获取视频帧更多的高频信息,图 4 为使用 DwSCNN 和双三次下采于 HEVC 压缩框架中的性能效果。图 4 给出了在 mobile 和 BQMall 两个视频序列中使用 DwSCNN 或双三次下采,再进行 HEVC 帧内编解码和双三次上采压缩过程后的率失真曲线。

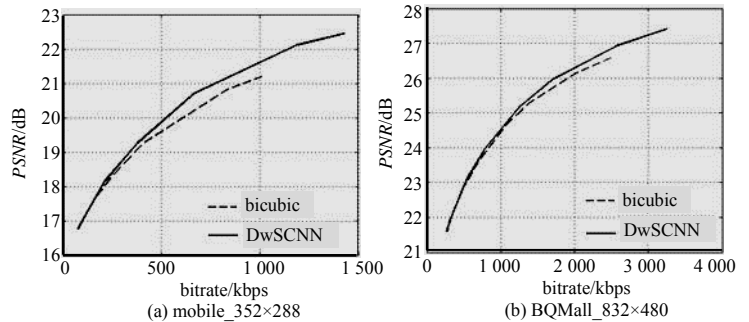


Fig.4 Rate distortion curves of the two sequences  
图 4 两个序列的率失真曲线

显而易见,DwSCNN 应用于帧内编码压缩中的率失真曲线在双三次下采应用于帧内编码压缩中的率失真曲线之上,说明 DwSCNN 相较于双三次下采能保留更多的视频序列信息。并且,在使用相同量化参数(Quantization Parameter, QP)范围进行压缩,DwSCNN 在压缩中的率失真曲线,即图中实线的 bitrate 坐标范围更宽,进一步说明使用 DwSCNN 下采后的低空间分辨率视频含有更多信息,使得传送的数据量略有增加。

3 质量增强卷积神经网络

3.1 质量增强卷积神经网络结构

HEVC 压缩编码后的视频序列,尤其是在低码率段将产生严重的块效应、振铃效应。为了减少这些失真,后处理技术十分重要。受 CNN 在计算机视觉领域取得的巨大成功启发,HEVC 编码后处理的研究中也广泛采用 CNN。基于此本文提出质量增强卷积神经网络 PPCNN,如图 5 所示,用于增强编解码后的视频质量。

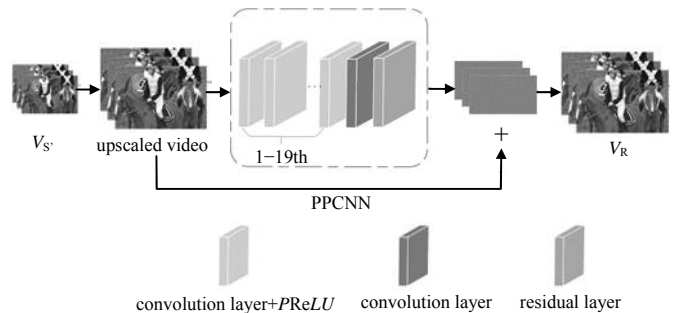


Fig.5 Post Processing CNN(PPCNN)  
图 5 质量增强卷积神经网络(PPCNN)

基于一种非常深的甚深超分辨率(Very Deep Super-Resolution, VDSR)网络<sup>[15]</sup>提出一种质量增强卷积神经网络 PPCNN,使其更适用于本文所提压缩框架。其依旧是一个 20 层的全卷积网络,如文献[15]中 4.1 节所述,可利用 20 层全卷积网络的高非线性实现特征映射。

在大多数情况下,网络性能随着深度的增加而增加,但在处理相同样本库时卷积层的增加最终会达到数据处理的饱和甚至造成过拟合,20 层的全卷积网络是综合考虑收敛速度和网络性能的最优选择。其前 19 个卷积核分别紧跟一个参数化修正线性单元(PReLU)作为非线性激活函数,用于帧内编码特征提取;第 20 个卷积层用于增强特征映射,完成图像重建;为加速收敛和解决梯度消失问题,采用残差学习。

PPCNN 的配置如表 2,PPCNN 的公式表示为:

表 2 PPCNN 的配置

$$F_0(Y) = Y \tag{7}$$

$$F_i(Y) = PReLU(W_i \cdot F_{i-1}(Y) + B_i), i \in \{1, 2, \dots, 19\} \tag{8}$$

$$F_{20}(Y) = W_{20} \cdot F_{19}(Y) + B_{20} \tag{9}$$

$$F_{21}(Y) = F_0(Y) + F_{20}(Y) \tag{10}$$

layer	Conv1	Conv2	Conv3	...	Conv19	Conv20
filter size	3×3	3×3	3×3	3×3	3×3	3×3
filter number	64	64	64	64	64	1
W learning rate	10 <sup>-4</sup>	10 <sup>-4</sup>	10 <sup>-4</sup>	10 <sup>-4</sup>	10 <sup>-4</sup>	10 <sup>-4</sup>
B learning rate	10 <sup>-5</sup>	10 <sup>-5</sup>	10 <sup>-5</sup>	10 <sup>-5</sup>	10 <sup>-5</sup>	10 <sup>-5</sup>

式中: Y 为原始输入视频帧; W<sub>i</sub>和 B<sub>i</sub>分别为第 i 层的权重和偏置。

3.2 质量增强卷积神经网络训练算法

与下采网络训练算法相似,为实现重建视频帧与原始视频帧之间最小的损失,采用均方误差 MSE 作为网络的损失函数。

$\{X_n\}_{n=1}^N$  为原始视频帧的图像块,  $\{Y_n\}_{n=1}^N$  是原始视频帧采用提出的 DwSCNN 下采, 分 QP 编解码再进行双三次上采恢复原始空间分辨率的视频帧图像块。因此,  $\{Y_n\}_{n=1}^N$  是输入样本,  $\{X_n\}_{n=1}^N$  为对应的目标标签。  $F(\bullet)$  为 PPCNN 的输出函数, 因此, 网络损失函数表示为:

$$L(\theta_2) = \frac{1}{N} \sum_{n=1}^N \|F(Y_n; \theta_2) - X_n\|_2^2 \quad (11)$$

式中  $\theta_2 = \{W_i, B_i\}$  表示 PPCNN 的权重和偏置。为能更精准地重建 DwSCNN 下采时捕获的结构信息, PPCNN 采用和 DwSCNN 一样的原始训练图库和验证图库。为了准备训练和验证数据, 将 400 张原始训练图和 100 张验证图先通过 DwSCNN 下采, 然后在 HEVC 帧内模式配置 encoder\_intra\_main.cfg 下当 QP=35,40,45,50 时分别进行编码, 最后采用双三次上采恢复到原始空间分辨率, 以此得到 4 个降质的图像库。考虑 PPCNN 需要更精准的重建性能, 并且其为两个相同空间分辨率大小的图库, 样本数据将少于 DwSCNN, 因此将两个图库进行 4 个方向的旋转和 5 个尺度的缩放。把原始图库和降质图库进行分割生成训练样本和验证样本, 以步长 32 将两个图库图像切割成 32×32。按照以上处理方法, 将获得 779 200 张训练样本和 194 800 张验证样本。实验时使用深度学习 CAFFE 来训练网络。采用 Adam 算法, 基础学习率为 0.000 1, 权重衰减因子为 0.000 1。

训练出 4 个 QP 的最佳模型, 用于提出的压缩改进算法, 改善 HEVC 帧内编码带来的视频失真。同时, 结合分层编码结构和扩大所提方法普适性, 将质量增强网络训练好的 4 个 QP 下最佳模型应用于 QP = 32 ~ 51 的范围内, 即 QP = 35 的最佳模型用于 QP = (32,37] 时压缩视频帧的重建, QP = 40 的最佳模型用于 QP = (37,42], 以此类推, QP = 45 的最佳模型用于 QP = (42,47], QP = 50 的最佳模型用于 QP = (47,51]。

#### 4 实验结果与分析

本文所有的实验均使用 Matlab R2014a 和 Visual Studio 2013 工具在 CPU 为 i5-2320 @3.00 GHz 和 GPU Nvidia GeForce GTX 1080 环境运行。实验过程采用标准 HM16.0 测试模型进行实验, 配置文件用的 encoder\_intra\_main.cfg, 量化 QP 分别为 35,37,42,45,47,51, 通过比较标准算法和所提方法的实验结果, 可以验证所提方法的有效性。表 3 给出了 6 个不同视频序列分别采用标准 HM16.0 算法和本文所提方法进行处理后的码率及相应的重建视频峰值信噪比(Peak Signal to Noise Ratio, PSNR)。

通过表 3 分析, Mobisode2 序列在采用 H.265/HEVC 标准 HM16.0 压缩, 当码率为 88.992 kbps 时, 重建视频 PSNR 值为 38.4673 dB; 仅采用所提下采网络 DwSCNN, 当码率为 83.64 kbps 时, 重建视频 PSNR 值为 38.894 2 dB; 采用结合下采网络 DwSCNN 和质量增强网络 PPCNN 的本文所提端对端压缩改进算法, 当码率为 83.64 kbps 时, 重建视频 PSNR 值为 39.119 9 dB。可以看出, 在 Mobisode2 序列中, 此时几乎同等码率下, 甚至在所提码率略低的情况下, 下采网络能带来 0.43 dB 的客观质量提升, 质量增强网络 PPCNN 能进一步带来 0.22 dB 的客观质量提升。同样地, 在 QP = 32 ~ 51 间, 其他视频序列也能取得类似的质量提升。

图 6 给出了视频序列 Akiyo\_352×288, Mobisode2\_416×240, RaceHorses\_416×240 和 RaceHorses\_832×480 的率失真曲线。可以看出本文所提方法在不同内容视频序列和同一内容不同分辨率视频序列上的率失真曲线均在 H.265/HEVC 标准上方, 在低码率段能达到更优压缩性能, 重建视频保留更多细节信息, 获得更优质量。

结合卷积神经网络的算法会损失一定的时间训练最佳模型, 但训练后进行离线重建速度却很快。为说明本文所提方法在离线重建时的效率, 定义变量  $\Delta B$  和  $\Delta T$ , 其中  $\Delta B$  表示改进算法相对于 H.265/HEVC 标准方法在低码率段的比特率变化,  $\Delta T$  表示算法复杂度的变化, 公式定义为:

$$\Delta B = \frac{B_{\text{proposed}} - B_{\text{HM16.0}}}{B_{\text{HM16.0}}} \times 100\% \quad (12)$$

$$\Delta T = \frac{T_{\text{proposed}} - T_{\text{HM16.0}}}{T_{\text{HM16.0}}} \times 100\% \quad (13)$$

在两种方法 PSNR 值几乎一致时, 对表 3 中所列 6 个视频序列进行实验, 结果如表 4 所示。由表 4 知, 本文所提改进算法平均能降低 11.04% 的比特率和 39.46% 的时间, 在 6 个视频序列里最高能降 17.66% 的比特率和 56.31% 的时间。改进算法之所以能有这样的效率, 一方面, 所提压缩框架下仅对空域下采样低分辨率视频进行压缩编码, 有效地减少编码数据量, 同时花费更少时间; 另一方面, 应用卷积神经网络最佳模型对视频序列进行离线重建速率较快, 并且随着硬件设备性能提高, 网络重建速度还会越来越快。这充分说明本文所提改进算法在低码率段的优越性。

表 3 采用 HM16.0 标准和本文方法的码率及 PSNR 参数对比  
Table3 Comparison of the bitrate and PSNR between HM16.0 standard and the proposed method

video sequences	H.265/HEVC standard			DwSCNN+HM16.0+bicubic-up			the proposed method		
	QP	bitrate/kbps	PSNR/dB	QP	bitrate/kbps	PSNR/dB	QP	bitrate/kbps	PSNR/dB
Akiyo 352×288	35	504.552	36.6281	35	274.884	32.0627	35	274.884	32.542
	37	416.520	35.2941	37	222.936	31.2368	37	222.936	31.6698
	42	247.512	31.9225	42	121.920	28.6366	42	121.920	28.9836
	45	177.552	30.0744	45	84.180	27.1291	45	84.180	27.3234
	47	134.988	28.8607	47	62.832	26.0218	47	62.832	26.1644
Bowling 352×288	51	83.280	26.9777	51	39.900	24.3867	51	39.900	24.4911
	35	575.112	35.7648	35	260.292	31.1526	35	260.292	31.2831
	37	453.852	34.3409	37	179.352	30.1436	37	179.352	30.2776
	42	206.040	30.7132	42	93.936	28.0876	42	93.936	28.1797
	45	132.480	29.2237	45	62.772	26.6398	45	62.772	26.6787
Mobisode2 416×240	47	102.000	28.1889	47	40.812	25.2716	47	40.812	25.4266
	51	57.876	26.2791	51	23.784	23.9941	51	23.784	24.1168
	35	158.220	41.8517	35	83.64	38.8942	35	83.640	39.1199
	37	135.384	40.8899	37	71.592	36.4768	37	71.592	38.2588
	42	88.992	38.4673	42	45.384	34.1534	42	45.384	35.8057
RaceHorses 416×240	45	68.976	36.9267	45	34.752	31.9128	45	34.752	34.3946
	47	55.116	35.8524	47	28.944	33.1729	47	28.944	33.4832
	51	37.428	34.0010	51	20.208	31.2844	51	20.208	31.6006
	35	1251.48	31.9739	35	609.504	28.2369	35	609.504	28.6471
	37	980.676	30.7273	37	478.752	27.4551	37	478.752	27.8431
Crew 704×576	42	543.852	27.9743	42	248.520	25.1524	42	248.520	25.5091
	45	375.288	26.3694	45	162.204	23.6796	45	162.204	23.9206
	47	284.256	25.3602	47	117.360	22.7171	47	117.360	22.9561
	51	156.132	23.4016	51	62.076	20.9281	51	62.076	21.0592
	35	2438.856	36.2867	35	1139.64	33.3181	35	1139.640	33.4238
RaceHorses 832×480	37	1895.712	35.3637	37	882.552	32.5271	37	882.552	32.6404
	42	1005.840	32.9678	42	433.752	30.3442	42	433.752	30.4546
	45	668.736	31.4829	45	276.792	29.0331	45	276.792	29.0334
	47	492.456	30.5282	47	197.424	28.1118	47	197.424	28.1855
	51	264.888	28.7521	51	99.744	26.5341	51	99.744	26.6087
RaceHorses 832×480	35	4005.516	32.7366	35	1783.992	28.9148	35	1783.992	29.1312
	37	3030.072	31.3416	37	1273.236	26.8381	37	1273.236	28.2806
	42	1370.532	28.3442	42	604.656	24.8976	42	604.656	26.3769
	45	915.336	27.1019	45	401.964	23.0678	45	401.964	25.1122
	47	691.200	26.3117	47	295.320	24.0842	47	295.320	24.3089
51	391.596	24.7274	51	158.544	22.4910	51	158.544	22.6540	

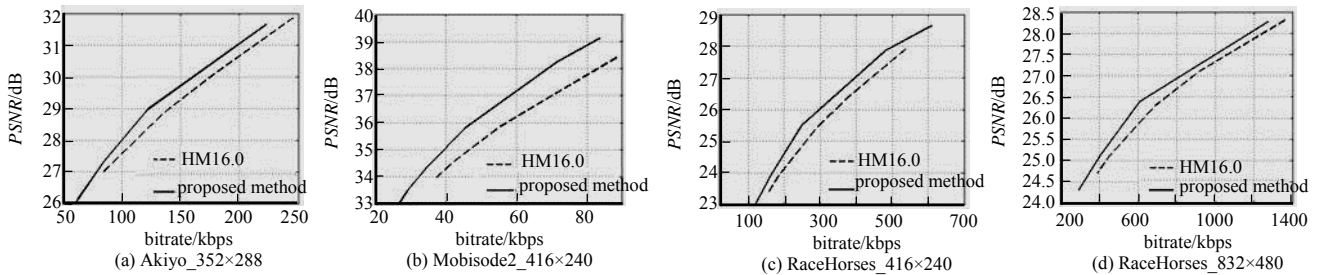


Fig.6 Rate distortion curves of the four sequences

图 6 四个序列的率失真曲线

表 4 本文方法与 HM16.0 标准码率节省和时间节省对比  
Table4 Comparison of bitrate saving and time saving between the proposed method and HM16.0 standard

video sequences	Akiyo 352×288	Bowling 352×288	Mobisode2 416×240	RaceHorses 416×240	Crew 704×576	RaceHorses 832×480	average
$\Delta B(100\%)$	-9.68	-7.91	-17.66	-11.97	-11.92	-7.1	-11.04
$\Delta T(100\%)$	-33.86	-37.41	-28.10	-30.37	-56.31	-50.72	-39.46

为了进一步表明本文所提压缩改进算法性能的优势，将其与相关提升压缩性能文献[6]在相同配置下进行对比。图 7 给出了视频 Johnny\_1 280×720 序列应用两种方法所得到的率失真曲线，可以看出本文所提方法率失真曲线明显高于文献[6]所给方法。

可以看出本文方法性能优于文献[6]所提方法, 虽然均是在空域上对视频序列进行压缩设计, 但本文方法不但在训练好最佳模型后有较高效的处理过程, 还有更优的压缩性能。文献[6]主要针对超高清视频强相关性来降低视频高压压缩比下的码率, 而本文方法具有更好的普适性。

## 5 结论

本文提出一种结合卷积神经网络的 HEVC 帧内编码压缩改进算法。结合一个下采网络 DwSCNN 和一个质量增强网络 PPCNN, DwSCNN 更加精准地捕捉原始视频的细节结构和高频信息, PPCNN 再进一步增强重建视频的质量。通过两个网络训练出最佳的网络模型, 将 HEVC 压缩过程的下采、编解码和上采及质量增强集成于一体。与 H.265/HEVC 标准编解码方法相比, 本文方法能达到更好质量重建, 并且在高压压缩比、同一视频质量前提下, 能平均减少 39.46%的比特率和 11.04%的时间。所提压缩改进算法, 能改善双三次压缩重建视频质量并且使得视频压缩处理过程更为高效。进一步的工作是探索适用于更宽码率范围内的高性能压缩改进算法。

### 参考文献:

- [ 1 ] SULLIVAN G J, OHM J, HAN W, 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 ] CHEN H G, HE X H, MA M L, et al. Low bit rates image compression via adaptive block down-sampling and super resolution[J]. Journal of Electronic Imaging, 2016, 25(1):013004.
- [ 3 ] DONG J, YE Y. Adaptive down-sampling for high-definition video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2014, 24(3):480–488.
- [ 4 ] LI C, SONG L, XIE R, et al. CNN based post-processing to improve HEVC[C]// 2017 IEEE International Conference on Image Processing. Beijing, China: IEEE, 2017:4577–4580.
- [ 5 ] YANG R, XU M, WANG Z L, et al. Multi-frame quality enhancement for compressed video[C]// 2018 IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018:6664–6673.
- [ 6 ] LI X Q, HE X H, QING L B, et al. Research on compression performance of ultrahigh-definition videos[J]. Journal of Electronic Imaging, 2017, 26(6):063008.
- [ 7 ] TAO W, JIANG F, ZHANG S P, et al. An end-to-end compression framework based on convolutional neural networks[C]// 2017 Data Compression Conference. Snowbird, UT, USA: IEEE, 2017:463.
- [ 8 ] YANG R, XU M, LIU T, et al. Enhancing quality for HEVC compressed videos[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(7):2039–2054.
- [ 9 ] CHRISTIAN S, LIU W, JIA Y Q, et al. Going deeper with convolutions[C]// 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA: IEEE, 2015:1–9.
- [ 10 ] HE K M, ZHANG X Y, REN S Q, et al. Identity mappings in deep residual networks[C]// 2016 European Conference on Computer Vision. Amsterdam, Netherland: IEEE, 2016:630–645.
- [ 11 ] ARBELAEZ P, MAIRE M, FOWLKES C, et al. Contour detection and hierarchical image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5):898–916.
- [ 12 ] LEE Y W, KIM J H, CHOI Y J, et al. CNN-based approach for visual quality improvement on HEVC[C]// 2018 IEEE International Conference on Consumer Electronics. Las Vegas, NV, USA: IEEE, 2018:1–3.
- [ 13 ] JIA Y Q, SHELHAMER E, DONAHUE J, et al. Caffe: convolutional architecture for fast feature embedding[C]// 2014 ACM International Conference on Multimedia. Orlando, Florida, USA: ACM, 2014:675–678.
- [ 14 ] KINGMA D P, BA J L. Adam: a method for stochastic optimization[C]// 2015 International Conference for Learning Representations. San Diego, CA, USA: [s.n.], 2015:1–15.
- [ 15 ] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016:1646–1654.

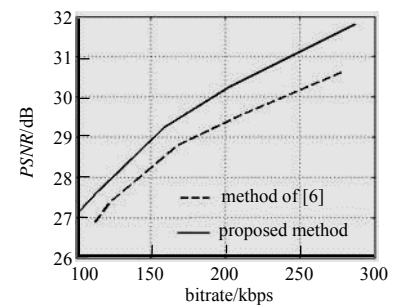


Fig.7 Rate distortion curves for Johnny with two methods

图 7 Johnny 序列两种方法率失真曲线