基于卷积神经网络的视频编码优化算法

陆 宇，陈 谦，殷海兵

（杭州电子科技大学 通信工程学院，杭州 310018）

摘 要：为进一步提高高效视频编码（HEVC）的压缩效率，使其更好地适用于高清视频的压缩，利用深度学习对视频特征强大的挖掘能力，提出了一种多输入的多尺度残差卷积神经网络和网络迭代训练方法，显著提高了HEVC环路滤波的性能；提出了一种新颖的分像素插值滤波方法，进一步提高编码的压缩效率。实验结果表明，本文算法在RA编码模式下平均可以减少7.47%的BD-rate。与现有的两种编码优化算法相比，本文提出的优化算法有效地提升了压缩效率，同时提高了视频质量。

关键词：卷积神经网络；环路滤波；高效视频编码；分像素插值滤波

中图分类号：TN919.81 **文献标志码**：A **文章编号**：1671-5497(2024)11-3296-06

DOI：10.13229/j.cnki.jdxbgxb.20230934

Optimization algorithm for video coding based on convolutional neural networks

LU Yu，CHEN Qian，YIN Hai-bing

（*Hangzhou Dianzi University*，*School of Communication Engineering*，*Hangzhou* 310018，*China*）

Abstract：In order to further improve the compression efficiency of Efficient Video Coding（HEVC）and make it more suitable for high-definition video compression. By utilizing the powerful mining ability of deep learning for video features，this paper proposes a multi input multi-scale residual convolutional neural network and network iterative training method，which significantly improves the performance of HEVC loop filtering. And a novel pixel based interpolation filtering method was proposed to further improve the compression efficiency of the encoding. The experimental results show that the algorithm proposed in this paper can reduce BD rate by an average of 7.47% in RA encoding mode. Compared with the two existing encoding optimization algorithms，the optimization algorithm proposed in this paper effectively improves compression efficiency while enhancing video quality.

Key words：convolutional neural network；loop filtering；efficient video encoding；pixel based interpolation filtering

0 引 言

根据其形式视频可划分为模拟视频和数字视频两种[1]，前者由模型相机逐行或隔行扫描生成，主要用于模拟电视系统；后者由数字相机拍摄生成或由模拟视频生成，日常生活所涉及视频多为数字视频。传统编码技术已经无法满足当下数字视频压缩、存储、传输等方面的要求，由此，高效视

收稿日期：2023-09-03.

基金项目：浙江省教育厅科研项目（Y202249588）；国家自然科学基金项目（61972123）.

作者简介：陆宇（1977-），男，副教授，博士.研究方向：智能视频信息处理方法.E-mail：luyu20230@163.com

频编码(high efficiency video coding,HEVC)应运而生[2],HEVC是为满足数字视频有线和无线传输需求而开发的视频编码标准。

经过HEVC编码/解码过程后,重构的帧会通过HEVC环路滤波器进行后处理,以消除伪影。HEVC和其他标准都存在两种主要的压缩失真,这是由基于块的预测、变换和有限精度的量化引起的。最常见的失真是块效应。在HEVC中,帧首先被划分为块(CTUs/CUs)作为基本的编码单元。这些块在预测、变换和量化方面的编码相对独立。由于变换和量化过程中会引入一些损失,编码块只能提供原始帧的近似表示,因此这些近似之间的差异可能导致块边界出现不连续性,从而产生块效应。在变换和量化过程中,高频信号会丢失,解码过程很难恢复这种信息丢失,因此,会导致图像严重失真,并出现振铃效应。

近年来,针对HEVC的编码优化方法成为研究热点。例如,采用分像素插值方法提高HEVC的压缩效率。Pan等[4]提出一种基于增强型深度卷积神经网络(EDCNN)的环内滤波算法,使用多种损失函数的组合,显著提高HEVC中环内滤波的性能;Sun等[5]提出一种非局部环路滤波框架,将基于CNN的压缩噪声估计方法插入环路滤波框架中,可以在不预先拟合噪声和量化信息之间关系的情况下,实现更准确的结果;Wang等[6]提出一种自适应插值滤波算法,通过滤波器系数对称优化,降低码流所需滤波器系数和解码计算的复杂度;基于机器学习的方法,Lu等[7]提出低复杂度的高效视频编码方法。

利用深度学习对视频特征强大的挖掘能力,本文提出了一种基于卷积神经网络的视频编码优化方法。

1　基于卷积神经网络的环路滤波算法

1.1　网络结构

本文提出的多尺度残差卷积神经网络是基于ResNet[8]的改进,其主要结构如图1所示。

为了尽可能提高当前帧的质量,本文提出使用额外的先验信息——高质量参考图像分量来提高网络增强性能。

在四叉树编码结构下,由于量化参数(QP)值不同,参考帧列表中的参考帧质量会发生波动。在帧间预测过程中,时间最近的帧具有最高的被选为参考帧的概率。然而,时间最近的参考帧并不一定是当前参考列表中质量最好的帧,因为参考列表由时间最近的帧和几个高质量帧组成。如果直接使用参考帧作为输入分量,有可能使从参考中提取有用信息非常困难。最近的帧是最相似的,而质量最高的帧具有最小的失真。

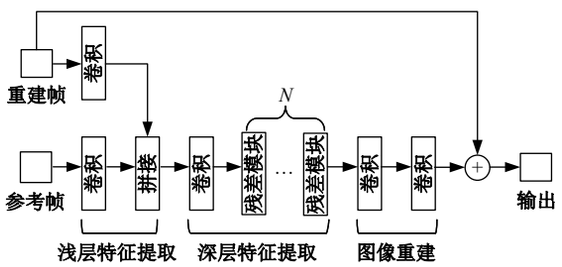


图1　本文卷积神经网络结构

Fig. 1　Proposed network structure

为了消除质量波动,采用参考帧列表中PSNR最高的帧作为补充输入。高质量参考帧可以提供更多有用的高质量信息来提高重建帧的质量。

下面分别介绍所提网络的各个模块:

(1)浅层特征提取模块

如图1所示,本模块的作用是对输入的参考图像和重建图进行浅层特征的提取,以便后续更深层次特征的提取。

为了同时处理这些输入分量,使用了两个对称的分支。在每个分支中,首先分别将这两个分量输入一个卷积核,提取特定的特征图,然后将它们拼接并输入下一模块。卷积核的大小设置为3×3,通道数设置为16,可以用式(1)表示:

$$F_1=\sigma(Conv^1_{3\times3}(x_h))\otimes\sigma(Conv^2_{3\times3}(x_r)) \quad (1)$$

式中:F_1表示预处理模块提取的特征;x_h表示输入的高质量参考帧分量;x_r表示输入的未经滤波的重建分量;$Conv^n_{k\times k}$表示第n个$k\times k$卷积核的卷积计算;σ表示LeakyReLU激活函数;$\otimes$表示张量通道上的拼接操作。

(2)深层特征提取模块

本模块的作用是利用所提出的多尺度残差块进行深度特征的提取,采用更深的网络进一步增强表达能力。

本模块由一个卷积层和N个多尺度残差块串联而成。本文网络所使用的多尺度残差块的结构如图2所示。在每个多尺度残差块中,输入的

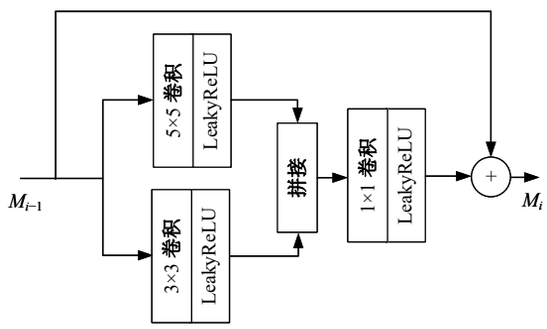


图2 多尺度残差块结构

Fig. 2 Multi-scale residual block structure

特征图分别经过5×5和3×3的卷积核,得到的张量进行拼接操作,然后进行1×1的卷积操作,并采用跳过连接将输出与输入特征相加。有N个残差块串联,为了平衡编码时间和编码质量,本文取N=10,即10个多尺度残差块进行串联,以提取深层特征。此模块可以用式(2)表示:

$$\begin{cases} C_i^1=\sigma(Conv_{5\times5}^{i}(M_{i-1})) \\ C_i^2=\sigma(Conv_{3\times3}^{i+2}(M_{i-1})) \\ M_i=M_{i-1}+C_i^1=\sigma(Conv_{1\times1}^{i}(C_i^1\otimes C_i^2)) \end{cases} \tag{2}$$

式中:C_i^1表示5×5卷积层的输出;C_i^2表示3×3卷积层的输出;M_i表示多尺度残差块的输出,i表示第i个多尺度残差块;σ表示LeakyReLU激活函数。

不同尺寸大小的卷积核可以获取不同尺度的特征[9],在提出的多尺度残差块结构中,大卷积核更擅长提取大尺度的轮廓特征,小卷积核则更擅长提取细节区域的特征。

(3)图像重建模块

如图1所示,本模块包含两个卷积层,输入为深层特征提取模块提取的特征图。这两个卷积层用于由上述增强特征重建残差图像,可以用式(3)表示:

$$O=Conv_{3\times3}^{N+4}(Conv_{3\times3}^{N+3}(M_N)) \tag{3}$$

式中,O表示图像重建模块的输出。

最后,将重建帧和图像重建模块的输出相加,得到网络的最终输出,使网络训练生成残差图像,减轻网络训练的负担。最终输出可以用式(4)表示:

$$y=x_r+O \tag{4}$$

2.2 问题分析

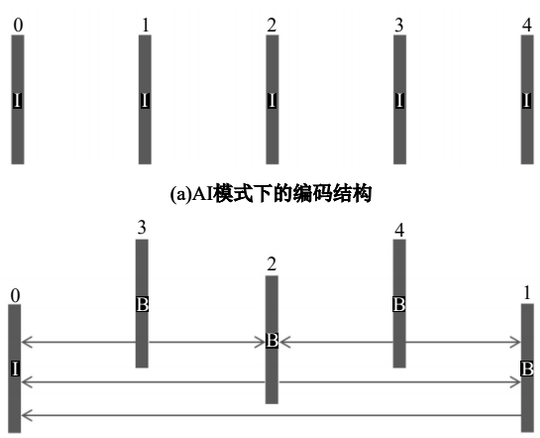
在文献[4,5]方法中,所有的网络都是基于HEVC的HM编码器生成的未经滤波的图像进行训练的。然后,基于神经网络的环路滤波滤波器同时应用于I帧和B帧。然而,在这个过程中可能存在一些问题。

全帧内(All intra ,AI)模式和随机访问(Random access ,RA)模式的编码结构如图3所示。对于AI模式,帧之间的预测过程中没有任何依赖关系(如图3(a)中的第0到第4帧)。换句话说,当前重建帧的质量不会对下一个编码帧的未经滤波的图像质量产生影响。然而,对于RA和LD(Low Delay)模式,如果当前重建帧的质量提高,下一个编码帧的未经滤波图像质量也会有所提高,因为当前帧将为下一帧提供更高质量的图像。在图3(b)中,箭头指向参考帧方向。例如,如果第0帧的重建质量提高,第1帧的重建质量也会提高。

基于上述分析,若当前帧启用基于神经网络的循环滤波器,下一个编码帧的未经滤波图像质量会得到提高。然而,用于下一个编码帧的基于神经网络的循环滤波器是基于质量较低的未经滤波图像进行训练的。因此,最终的测试过程会与训练过程存在不一致的结果。

1.3 网络迭代训练方法

为了解决这个不一致的问题,本文提出了一种迭代训练方法。在整个迭代流程中,有多个训练操作。初始训练过程与传统训练过程类似,由基于HM编码器(关闭环路滤波模块)编码生成的未经滤波的图像组成训练集进行网络训练。然后,将初始训练生成的神经网络环路滤波器集成到HM编码器再次生成训练集,基于这个新的训练集继续训练网络,并且网络的初始模型参数与



(a)AI模式下的编码结构

(b)RA模式下的编码结构

图3 AI和RA模式下的编码结构

Fig. 3 Coding structures under AI and RA configurations

初始训练得到的网络参数相同。重新训练过程将以迭代的方式进行,直到不能进一步改进性能或达到目标性能。

2　HEVC分像素插值滤波算法

2.1　相关工作

由于HEVC固有的插值滤波器无法根据视频内容自适应减小帧间预测误差,出现编码效率低的问题。本文提出了一种新颖的HEVC分像素插值滤波算法,并在编码器中采用率失真优化的方法,为每个预测单元选择最好的插值滤波器。

2.2　HEVC分像素插值滤波算法

在视频的每帧中新增一组与HEVC传统固定插值滤波器g^1、g^2和g^3相对应的自适应插值滤波器,记为g^4、g^5和g^6,抽头数量分别为7、8和7,其中,g^4对应搜索得到(1/4,0)或(0,1/4)位置分像素,g^5对应搜索得到(1/2,0)或(0,1/2)位置分像素,g^6对应搜索得到(3/4,0)或(0,3/4)位置分像素。

记录当前帧中全部分像素运动向量为(1/2,0)或(0,1/2)的编码单元,对应原始像素为y_i,$i=1,2,\cdots,M$,y_i对应预测像素记为q_i,计算方式如下所示:

$$q_i=\sum_{j=1}^{8}d_j^i g_j^5 \tag{5}$$

式中:j为抽头;d_j^i为参考帧中对应的整像素。

定义预测误差E如下所示:

$$E=\sum_{i=1}^{M}\left(y_i-q_i\right)^2 \tag{6}$$

将式(5)代入式(6),得到预测误差E如下所示:

$$E=\sum_{i=1}^{M}\left(y_i-\sum_{j=1}^{8}d_j^i g_j^5\right)^2 \tag{7}$$

自适应插值滤波器g^5需与HEVC传统固定插值滤波器g^2一样保持对称性,即$g_i^5=g_{9-i}^5$,$i=1,2,3,4$,则由此可将式(5)转变为如下形式:

$$\begin{bmatrix}q_1\\q_2\\\vdots\\q_M\end{bmatrix}=\begin{bmatrix}d_1^1&d_2^1&\cdots&d_8^1\\d_1^2&d_2^2&\cdots&d_8^1\\\vdots&\vdots&&\vdots\\d_1^M&d_2^M&\cdots&d_8^M\end{bmatrix}\begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\\0&0&0&1\\0&0&1&0\\0&1&0&0\\1&0&0&0\end{bmatrix}\begin{bmatrix}g_1^5\\g_2^5\\g_3^5\\g_4^5\end{bmatrix} \tag{8}$$

式(8)用矩阵表示,D为等式右侧第一项,K为右侧第二项,g为右侧第三项,则式(7)如下所示:

$$E=\left\|\boldsymbol{y}-q\right\|^2=\left\|\boldsymbol{y}-DKg\right\|^2 \tag{9}$$

式中:$\boldsymbol{y}$表示原始像素矩阵;$\boldsymbol{q}$表示预测像素矩阵。对上式求其最小优化解,得到自适应插值滤波器g^5如下所示:

$$g^5=\left[(DK)^{\mathrm{T}}(DK)\right]^{-1}(DK)^{\mathrm{T}}\boldsymbol{y} \tag{10}$$

记录当前帧中全部分像素运动向量为(1/4,0)或(0,1/4)的编码单元,对应原始像素为y_i',y_i'对应预测像素记作q_i',如下所示:

$$\begin{bmatrix}q_1'\\q_2'\\\vdots\\q_M'\end{bmatrix}=\begin{bmatrix}d_1^1&d_2^1&\cdots&d_7^1\\d_1^2&d_2^2&\cdots&d_7^1\\\vdots&\vdots&&\vdots\\d_1^M&d_2^M&\cdots&d_7^M\end{bmatrix}\begin{bmatrix}g_1^4\\g_2^4\\\vdots\\g_7^4\end{bmatrix} \tag{11}$$

引入矩阵D,得到自适应插值滤波器g^4预测误差E'如下所示:

$$\begin{cases}E'=\left\|\boldsymbol{y}'-\boldsymbol{q}'\right\|^2=\left\|\boldsymbol{y}'-Dg^4\right\|^2\\\boldsymbol{D}=\begin{bmatrix}d_1^1&d_2^1&\cdots&d_7^1\\d_1^2&d_2^2&\cdots&d_7^1\\\vdots&\vdots&&\vdots\\d_1^M&d_2^M&\cdots&d_7^M\end{bmatrix}\end{cases} \tag{12}$$

式中:$\boldsymbol{y}'$表示原始像素矩阵;$\boldsymbol{q}'$表示预测像素矩阵。

由此可得到自适应插值滤波器g^4如下所示:

$$g^4=\left(D^{\mathrm{T}}D\right)^{-1}D^{\mathrm{T}}\boldsymbol{y}' \tag{13}$$

通过与g^4相同的方法可获取到自适应插值滤波器g^6,区别在于记录的是分像素运动向量为(3/4,0)或(0,3/4)的编码单元。自适应插值滤波器g^6的表达式如下所示:

$$g^6=\left(D^{\mathrm{T}}D\right)^{-1}D^{\mathrm{T}}\boldsymbol{y}' \tag{14}$$

若采用当前帧所得滤波器插值当前帧图像,则需要二次编码当前帧,造成计算复杂度大幅度增加。因此,本文采用参考帧的分像素插值滤波器对当前帧图像插值以提高编码效率。

3　实验与性能分析

3.1　HEVC编码方法

本文采用率失真优化(RDO)策略,从基于神经网络的环路滤波器和HEVC环路滤波器中自适应选择,使用一个帧级标记位来表示采用何种

环路滤波器。如果帧级标记为0,当前帧的所有CTU都不会应用所提出的环路滤波器;如果帧级标志为1,则会通过CTU级的标志表示是否采用本文提出的环路滤波器。

本文采用率失真优化(RDO)的策略,从HEVC固定滤波器和分像素滤波器中进行自适应选择,以实现最佳的编码性能,如图5所示。

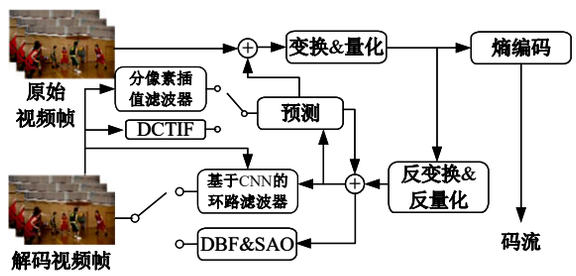


图5 本文采用的HM编码器

Fig. 5 proposed HM encoder

3.2 实验结果对比分析

为了测试本文方法的率失真性能,本文使用18个不同分辨率和不同运动情况的数字视频序列,依据分辨率将数字视频序列划分为5个不同组别。

为验证本文方法的有效性,将本文方法与一些最新的文献方法[4,5]进行比较。使用BD-BR[13]来评估编码性能,代表在相同PSNR下的比特率减少量。BD-BR的负值意味着当前算法优于参考算法。BD-BR的正值意味着在相同PSNR下比特率增加,即性能下降。在RA配置下,本文方法与参考文献[4,5]的方法比较结果如表1所示,可见,本文方法与HM16.9相比可以将BD-rate最多减少14.52%。在RA配置下,BD-rate平均减少可以达到7.47%。与其他两个方法相比,本文方法可以实现最多的码率节省。这表明本文方法能获得较好的压缩效率。通过分析表1中的数据还发现,本文方法对于一些具有复杂纹理和快速运动的序列,性能表现一般。将来会研究并利用更先进的先验信息来缩小性能优良序列和性能较差序列之间的差距。

表1 编码性能比较

Table 1 Coding performance comparison %

序列	文献[4]方法	文献[5]方法	本文方法
Traffic	−5.97	−6.24	−8.21
PeopleOnStreet	−9.17	−6.23	−7.84
Kimono1	−3.68	−5.73	−7.19
Cactus	−5.29	−7.92	−8.92
BQTerrace	−5.08	−13.80	−11.43
ParkScene	−3.85	−1.26	−4.36
BasketballDrive	−5.99	−7.34	−6.91
RaceHorses	—	−4.53	−7.89
BQMall	−7.37	−4.81	−8.76
PartyScene	−9.28	−1.68	−3.23
BasketballDrill	−10.65	−3.55	−4.99
RaceHorses	−3.52	−3.85	−6.27
BQSquare	−10.99	−1.68	−3.61
BlowingBubbles	−7.54	−0.89	−3.09
BasketballPass	−6.34	−3.10	−6.15
FourPeople	−10.53	−10.69	−10.43
Jonny	−11.22	−14.55	−14.52
KristenAndSara	−9.00	−11.91	−10.73
平均	−6.62	−6.09	−7.47

3.3 主观效果分析

为了展示不同算法的视频主观质量,选择两个视频序列进行比较,分别是“BasketballDrive”和BQMall”。在每个序列中选取128×128大小的图像块进行比较,然后在QP32,RA模式下对序列进行编码。这两个序列的比较结果如图6所示,最左侧是原始序列图像,从左到右分别是原始图像块、文献[4]方法、文献[5]方法和本文方法的主观结果。在图6中,可以看到与其他方法相比,使用本文的方法编码的序列在主观质量上取得了优势,几乎没有伪影,并且保留了更多的细节。

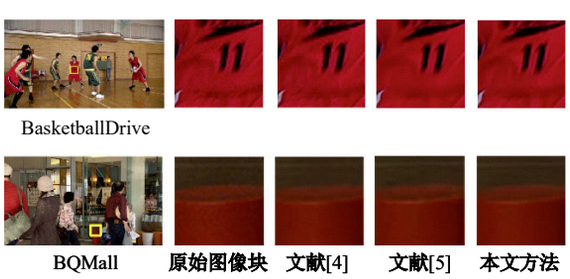


图6 视频主观质量比较本文方法与其他方法的效果对比

Fig. 6 Subjective quality comparison

5 结束语

本文提出了一种基于卷积神经网络的视频编码优化算法。该方法包括一种基于卷积神经网络的环路滤波方法和一种提高环路滤波质量的迭代训练方法,并通过新颖的HEVC分像素插值滤波算法进一步提高HEVC的压缩效率。此外,本文采用高质量参考帧作为神经网络的额外输入,将其和当前重建帧输入到基于CNN的环路滤波网络中,以生成更高质量的重建帧。

参考文献:

[1] 韩丽, 王华东. 动态视频多帧连续图像形变特征重构方法研究[J]. 计算机仿真, 2022, 39(12): 245-248.
Han Li, Wang Hua-dong. Research on deformation feature reconstruction of dynamic video multi-frame continuous image[J]. Computer Simulation, 2022, 39(12): 245-248

[2] Sullivan G J, Ohm J R, Han W J, et al. Overview of the high efficiency video coding standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(12): 1649-1668.

[3] 惠超, 蒋林, 朱筠, 等. HEVC中分像素插值算法的动态可重构实现[J]. 计算机工程与设计, 2022, 43(3): 764-770.
Hui Chao, Jiang Lin, Zhu Jun, et al. Dynamic reconfigurable implementation of pixel interpolation algorithm in HEVC[J]. Computer Engineering and Design, 2022, 43(3): 764-770.

[4] Pan Z, Yi X, Zhang Y, et al. Efficient in-loop filtering based on enhanced deep convolutional neural networks for HEVC[J]. IEEE Transactions on Image Processing, 2020, 29: 5352-5366.

[5] Sun W, He X, Chen H, et al. A nonlocal HEVC in-loop filter using CNN-based compression noise estimation[J]. Applied Intelligence, 2022, 52(15): 17810-17828.

[6] 王刚, 陈贺新, 陈绵书. 基于HEVC的自适应插值滤波算法[J]. 吉林大学学报: 理学版, 2018, 56(2): 320-328.
Wang Gang, Chen He-xin, Chen Mian-shu. Adaptive interpolation filtering algorithm based on HEVC[J]. Journal of Jilin University (Science Edition), 2018, 56(2): 320-328.

[7] Lu Y, Huang X, Liu H, et al. Fast SHVC inter-coding based on bayesian decision with coding depth estimation[J]. Journal of Real-Time Image Processing, 2021, 18(6): 2269-2285.

[8] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770-778.

[9] Li J C, Fang F M, Mei K F, et al. Multi-scale residual network for image super-resolution[C]//Proceedings of the European Conference on Computer Vision, Munichi, Germany, 2018: 527-542.

[10] Agustsson E, Timofte R. Ntire 2017 challenge on single image super-resolution: dataset and studyp[C]// IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, USA, 2017: 1122-1131.

[11] Ma D, Zhang F, Bull D. BVI-DVC: a training database for deep video compression[J]. IEEE Transactions on Multimedia, 2021, 24: 3847-3858.

[12] Kingma D P, Ba J. Adam: a method for stochastic optimization[J]. Arxiv Preprint, 2014, 9: 14126980.

[13] Bjøntegaard G. Calculation of average PSNR differences between RD-curves[J]. ITU-T VCEG-M33, 2001, 4: 2-4.