

林湧, 杨郑龙, 罗亦茜, 等. 面向 H.266/VVC 的 $R - \lambda$ 帧内码率控制算法[J]. 智能计算机与应用, 2024, 14(4): 60-67. DOI: 10.20169/j.issn.2095-2163.240408

面向 H.266/VVC 的 $R - \lambda$ 帧内码率控制算法

林 湧, 杨郑龙, 罗亦茜, 刘欣昱

(上海工程技术大学 城市轨道交通学院, 上海 201620)

摘要: 在 $R - \lambda$ 帧内码控中, 提出基于卷积神经网络(Convolutional Neural Networks, CNN)的最佳比特分配和最优拉格朗日因子 λ 选择。首先, 探索编码树单元(Coding Tree Unit, CTU)的码率与失真(Rate-Distortion, $R - D$)及码率与拉格朗日因子 λ ($R - \lambda$, $R - \lambda$)的关系特性, 设计具有四输出的 CNN 预测 $R - D$ 和 $R - \lambda$ 曲线的关键参数; 然后, 建立帧级 λ 和目标码率的优化方程, 反演得到最佳 CTU 码率分配; 最后, 根据 CTU 码率分配和先知的 $R - \lambda$ 曲线, 得到最优 CTU 级 λ 。实验表明, 算法在保持 4.76% 控制精度下, 比 VTM13.0 默认码控算法提高 0.31 dB 的编码质量。

关键词: 卷积神经网络; 帧内码率控制; 码率失真和 $R - \lambda$ 关系; 率失真优化

中图分类号: TP919.81 文献标志码: A 文章编号: 2095-2163(2024)04-0060-08

Optimal $R - \lambda$ intra rate control in H.266/Versatile Video Coding

LIN Yong, YANG Zhenglong, LUO Yixi, LIU Xinyu

(School of Urban Railway Transportation, Shanghai University of Engineering Science, Shanghai 201620, China)

Abstract: Optimal bit allocation and optimal Lagrange factor λ selection based on Convolutional Neural Networks (CNN) are proposed in $R - \lambda$ intra-frame code control. Firstly, we explore the properties of the coding tree unit (CTU) in terms of the relationship between the code rate and distortion ($R - D$) and the code rate and the Lagrange factor λ ($R - \lambda$, $R - \lambda$), and design the key parameters of the CNN with four outputs for predicting the $R - D$ and $R - \lambda$ curves; and then, we establish the optimization of the frame-level λ and target code rate equations and invert them to obtain the optimal CTU code rate allocation; finally, the optimal CTU level λ is obtained based on the CTU code rate allocation and the prophet's $R - \lambda$ curve. experiments show that the algorithm improves the coding quality by 0.31 dB over the VTM13.0 default code-control algorithm while maintaining 4.76% control accuracy.

Key words: convolutional neural network; intra frame rate control; rate distortion and $R - \lambda$ relationship; rate distortion optimization

0 引言

视频服务和应用已在人们日常生活中占据重要地位, 为提供更高质量的视频服务和适应更多的应用场合, 视频编码技术亦不断地被改进与更新。相比于上一代高效视频编码^[1](High Efficiency Video Coding, HEVC), 新一代视频编码标准(Versatile Video Coding, VVC)^[2]的编码性能提高一倍, 其能够适应 8 K、虚拟现实视频的编码和传输需求。VVC 中采用多种新的视频编码技术, 如四叉树二叉

树编码结构^[3]、仿射运动估计^[4], 以及增加帧内角度预测^[5]。码率控制 VVC 编码标准的重要组成部分, 目前有多种经典的模型算法, 如 $R - \rho$ 模型^[6]和统一码率-量化(Unified Rate-Quantization, URQ)模型^[7]。由于在视频编码过程中, 率失真优化(Rate Distortion Optimization, RDO)^[8]用以平衡码率与编码失真, Li 等^[9]提出 λ 域的码率控制算法, 其将拉格朗日因子 λ 作为码控的关系因素, 达到很好编码效果。随后, λ 域码率控制算法^[10]被 HEVC 采纳。

由于帧内码控编码质量将影响后续编码, 有必

基金项目: 张江国家自主创新示范区专项发展资金重点项目(ZJ2020-ZD-009)。

作者简介: 林 湧(1998-), 男, 硕士研究生, 主要研究方向: 人工智能, 视频编解码技术; 罗亦茜(1999-), 女, 硕士研究生, 主要研究方向: 人工智能, 视频编解码技术; 刘欣昱(1999-), 女, 硕士研究生, 主要研究方向: 人工智能, 图像处理。

通讯作者: 杨郑龙(1978-), 男, 博士, 讲师, 主要研究方向: 人工智能, 视频编解码技术。Email: 949713872@qq.com

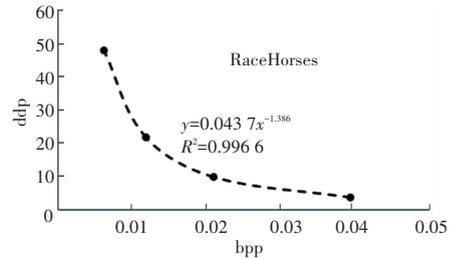
收稿日期: 2023-3-17

要提高帧内码控性能。Karczewicz 等^[11]对帧内编码使用绝对变换差值之和作为码率分配的标准。Wang 等^[12]提出码率分配的依据是由编码树单元 (Coding Tree Unit, CTU) 的梯度项计算所得的场景复杂度。对于基于感知的帧内码率控制, Gao 等^[13]根据结构相似性 (结构相似性指数测量) 建立率失真关系, 以获得每个 CTU 的最佳感知码率分配。因为 λ 域码率控制具有良好的编码性能^[14], 其也被 VVC 标准所采用。为了进一步提高码率控制的性能, 大多数研究人员都注意到了帧内码控的重要性。Liu 等^[15]提出了一种自适应的 λ 比率估计, 以平衡编码失真和比特成本。Mao^[16]认为转换系数是最重要的参数, 用复合 Cauchy 分布对该系数进行建模, 以提供更高的 $R-D$ 特性准确度。Li^[17]提出了分层编码结构的质量依赖系数, 以提高编码间的质量。与有大量时间信息参考的帧间码控不同, 帧内码控没有先验信息可依赖, 很难选择合适的初始参数进行编码。在 VVC 帧内码控中, CTU 的码率分配是根据纹理复杂度计算, 编码参数为预先设定。然而, 预定义的编码参数并不总能适合所分配的比特, 由于不必要的编码而降低了比特率。因此, 为了得出帧内码率控制参数, Chen 等^[18]提出使用二次函数建模 $R-D$ 特性。然而, 不同的 CTU 具有不同的编码纹理, 初始二次函数不能满足复杂的 CTU 特征, 导致编码效率低。由于卷积神经网络 (Convolutional Neural Network, CNN)^[19]通过提取图像纹理获得先验信息, 可以准确提供初始编码信息。Li^[20]使用 CNN 提取与视频内容相关的帧的特征来预测帧内码率控制参数。Li 等^[21], 提出了一种基于 CNN 的 $R-\lambda$ 帧内码率控制模型, 每个 CTU 的 $R-\lambda$ 曲线参数可由 CNN 预测所得。同样, Wei 等^[22]用 CNN 预测 $R-\lambda$ 曲线, 并建立了 CTU 级 λ 和帧级 λ 之间的关系。然而, 只考虑 λ 选择无法达到最佳编码性能, 码率分配同样是影响帧内码控性能的另一个重要因素。

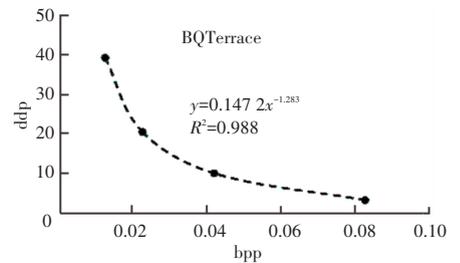
本文考虑帧内码控的 CTU 编码特性, 基于 CNN 提出一种 CTU 最佳码率分配和 λ 选择方案。首先, 探索不同的帧内 CTU 的 $R-D$ 和 $R-\lambda$ 特性, 设计了具有四输出的 CNN 网络结构, 预测 CTU 的 $R-D$ 和 $R-\lambda$ 曲线的关键参数, 以预知 CTU 编码特征。然后, 将帧的失真转化为所有 CTU 的总失真, 通过求解率失真优化问题, 结合 CTU 的率失真特性, 推演每个 CTU 的最佳码率分配表达式; 最后, 通过预测的 $R-\lambda$ 关系和分配的码率, 得到最优 CTU 级 λ , 在编码失真和码率分配间达到最佳动态平衡。

1 $R-D$ 与 $R-\lambda$ 关系

在没有码率控制的情况下, 多次运行 VVC 测试模型 (VTM13.0), 同时观察不同 CTU 在所有帧内配置下 $R-D$ 和 $R-\lambda$ 特性。实验结果如图 1、图 2 所示。



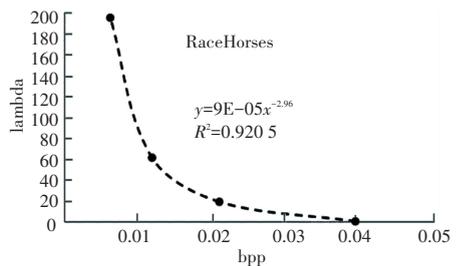
(a) RaceHorses 的第一帧第一个 CTU 的 $R-D$ 关系



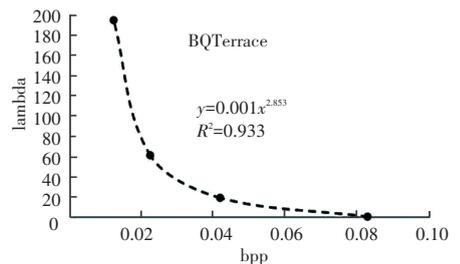
(b) BQTerrace 的第一帧第一个 CTU 的 $R-D$ 关系

图 1 不同 CTU 的 $R-D$ 特性

Fig. 1 Different CTUs $R-D$ characteristic



(a) RaceHorses 的第一帧第一个 CTU 的 $R-\lambda$ 关系



(b) BQTerrace 的第一帧第一个 CTU 的 $R-\lambda$ 关系

图 2 不同 CTU 的 $R-\lambda$ 特性

Fig. 2 Different CTUs $R-\lambda$ characteristic

从图 1 中可以看出, 帧内编码的 $R-D$ 关系符合双曲线函数, 如式 (1) 所示:

$$D = a \cdot R^{-b} \quad (1)$$

其中, D 是 CTU 的失真; R 是 CTU 的码率; a 和 b 是率失真关键参数。

图 2 中的 $R - \lambda$ 关系也符合双曲线函数规律, 如式(2)所示:

$$\lambda = \mu \cdot R^{-\gamma} \quad (2)$$

其中, λ 是 CTU 的拉格朗日因子, μ 和 γ 是关键参数。

从图 1 和图 2 中可以看出, 不同的 CTU 具有不同的 $R - D$ 和 $R - \lambda$ 特性。另一方面, 如果在编码前能得到每个 CTU 准确的 $R - D$ 特性, 就可以推演出每个 CTU 最优码率分配。结合先知的 $R - \lambda$ 特性, 只要分配的码率点位于 $R - \lambda$ 曲线上, 就能达到最佳的编码性能。

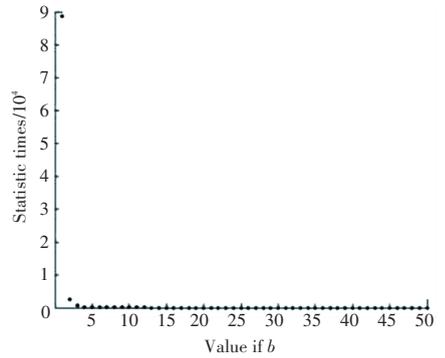
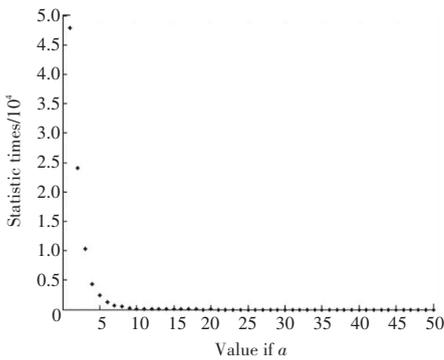


图 3 $R - D$ 曲线的关键参数 a 和 b

Fig. 3 Key parameters a and b of the $R - D$ curve

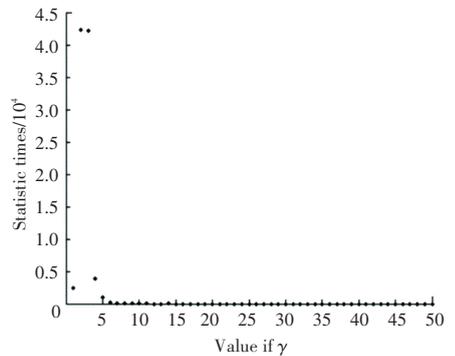
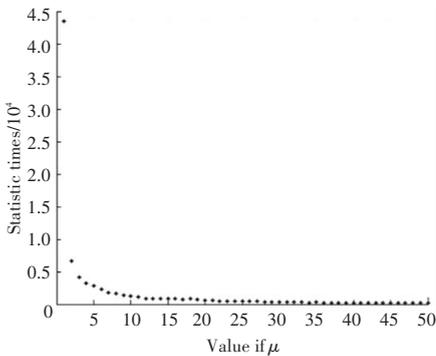


图 4 $R - \lambda$ 曲线的关键参数 μ 和 γ

Fig. 4 Key parameters μ and γ of the $R - \lambda$ curve

图 3 中, 横轴分别代表 a 和 b 的值, 纵轴代表获得该值的统计次数。可以看出, a 的值范围大多位于 $[0, 20]$ 之间, 而 b 的值范围大多位于 $[0, 5]$ 之间。因此, 将真实标签中 a 和 b 的取值范围分别限制在 $[0, 20]$ 和 $[0, 5]$ 。同样, 如图 4 所示, 将真实标签中 μ 和 γ 的值限制在 $[0, 15]$ 和 $[0, 5]$ 范围内。

本文的 CNN 的结构如图 5 所示。

2 基于 CNN 的最佳帧内码率控制算法

2.1 基于 CNN 的关键参数预测

本文关闭 VTM13.0 码率控制功能, 使用帧内编码模式获得真实标签数据。其中, 量化参数 QP 分别设置为 22、27、32、和 37, 可以得到一帧中所有 CTU 的 4 个 $R - D$ 点和 4 个 $R - \lambda$ 点; 然后, 用双曲线函数拟合 $R - D$ 曲线和 $R - \lambda$ 曲线, 并获取曲线的关键参数; 最后, 将关键参数与原 CTU 图像一起制作成数据集。本文从不同种类的序列中选择了 100 帧, 制作出 100 000 个 $R - D$ 曲线和 100 000 条 $R - \lambda$ 曲线, 观察关键参数 a 、 b 、 μ 和 γ 的取值范围, 其结果如图 3、图 4 所示。

图 5 中, CNN 结构由 5 个卷积层 (Conv)、7 个线性激活层 (ReLU)、两个最大池化层 (Pool Max) 和 3 个全连接层 (FC) 组成。网络损失函数如式 (3) 所示:

$$Loss = \frac{1}{N} \sum_{i=1}^N \| a_i - A_i + b_i - B_i + \gamma_i - X_i + \mu_i - Y_i \|^2 \quad (3)$$



图 5 用于预测 a, b, μ 和 γ 的 CNN 结构

Fig. 5 CNN structure used to predict a, b, μ and γ

其中, N 是用于训练的 CTU 数量; a_i, b_i, γ_i 和 μ_i 为 CNN 对第 i 个 CTU 的输出, A_i, B_i, X_i 和 Y_i 为第 i 个 CTU 的真值数据标签。采用随机梯度下降和反向传播进行训练。对于 CNN 网络训练, 其中训练集 CTU 数目为 90 000, 测试集 CTU 数目为 20 000 个。训练和测试的损失如图 6 所示。

行求导优化可得式(6):

$$\frac{\partial \sum_{i=1}^{N_{CTU}} d_{CTU}(i)}{\partial r_{CTU}(z)} + \lambda_F \cdot \frac{\partial \left(\sum_{i=1}^{N_{CTU}} r_{CTU}(i) - T \right)}{\partial r_{CTU}(z)} = 0 \quad (6)$$

将式(1)代入式(6)可得式(7):

$$\frac{\partial \sum_{i=1}^{N_{CTU}} a(i) \cdot r_{CTU}(i)^{-b(i)}}{\partial r_{CTU}(z)} + \lambda_F \cdot \frac{\partial \left(\sum_{i=1}^{N_{CTU}} r_{CTU}(i) - T \right)}{\partial r_{CTU}(z)} = 0$$

$$z \in [1, N_{CTU}] \quad (7)$$

由于帧内编码 CTU 间互不影响, 式(7)可以转换为式(8):

$$-b(i) \cdot a(i) r_{CTU}(i)^{-b(i)-1} + \lambda_F = 0, \quad i \in [1, N_{CTU}] \quad (8)$$

其中, $a(i)$ 和 $b(i)$ 是 $R - D$ 曲线的关键参数, 其可由 CNN 网络模型预测所得。根据公式(8), 码率分配 $r_{CTU}(i)$ 可以表示为式(9):

$$r_{CTU}(i) = \left(\frac{\lambda_F}{b(i) \cdot a(i)} \right)^{\frac{1}{-b(i)-1}} = \left(\frac{\lambda_F}{\alpha(i)} \right)^{\frac{1}{\beta(i)}} \quad (9)$$

其中, $\alpha(i) = a(i) \cdot b(i), \beta(i) = -b(i) - 1$ 。由于帧内 CTU 总码率等于当前帧的码率。因此可得式(10):

$$\left(\frac{\lambda_F}{\alpha(1)} \right)^{\beta(1)} + \left(\frac{\lambda_F}{\alpha(2)} \right)^{\beta(2)} + \dots + \left(\frac{\lambda_F}{\alpha(N_{CTU})} \right)^{\beta(N_{CTU})} = T \quad (10)$$

通过式(10)可以求得 λ_F 。最后, 将 λ_F 反向带入到式(9), 即可得到每个 CTU 的最佳码率分配。最后, 根据式(2), 可以得到最优 $\lambda_{CTU}(i)$, 即

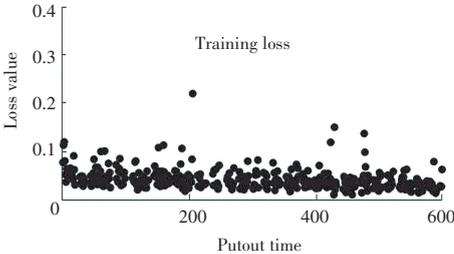
$$\lambda_{CTU}(i) = \mu(i) \cdot r_{CTU}(i)^{-\gamma(i)} \quad (11)$$

在式(11)中, $\mu(i)$ 和 $\gamma(i)$ 是 $R - \lambda$ 曲线的关键参数, 可由 CNN 网络模型预测所得。

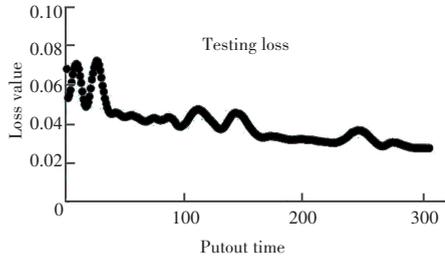
3 实验结果与分析

在 VTM13.0 编码平台上, 使用通测条件进行实验。实验使用的测试序列如下:

- A 类: (1) PeopleOnStreet (2) Traffic;
- B 类: (1) BasketballDrive (2) BirdsInCage (3) BQTerrace (4) Cactus (5) Cafe (6) Kimon (7)



(a) 训练损失



(b) 测试损失

图 6 CNN 的训练和测试损失

Fig. 6 CNN's training and testing losses

从图 6 可以看出, 训练和测试的损失结果逐渐趋于稳定, 并且保持在 0.015 和 0.026。因此, 网络预测结果将达到较高精度水平。

2.2 最优 CTU 码率分配和 λ 选择

建立帧内码控编码率失真关系模型如式(4)所示:

$$\min_{\{r_{CTU}(i)\}_{i=1}^{N_{CTU}}} D_F = \sum_{i=1}^{N_{CTU}} d_{CTU}(i) \quad \text{s.t.} \quad R_F = \sum_{i=1}^{N_{CTU}} r_{CTU}(i) = T \quad (4)$$

其中, D_F, R_F 分别为帧的失真和码率; d_{CTU}, r_{CTU} 分别为 CTU 的失真和码率; N_{CTU} 为帧中 CTU 的数量; T 为目标比特率。将率失真约束关系转换为等效的无约束关系, 如式(5)所示:

$$\min_{\{r_{CTU}(i)\}_{i=1}^{N_{CTU}}} \sum_{i=1}^{N_{CTU}} d_{CTU}(i) + \lambda_F \cdot \left(\sum_{i=1}^{N_{CTU}} r_{CTU}(i) - T \right) \quad (5)$$

其中, λ_F 是该帧的拉格朗日乘数。对式(5)进

ParkScene;

C类:(1) BasketballDrill(2) BQMall;

D类:(1) RaceHorses(2) BQSquare;

E类:(1) FourPeople(2) KristenAndSara;

F类:(1) SlideShow (2) ChinaSpeed (3)

BasketballDrillText。

本文选择VTM13.0默认的CTU级帧内码率控制(模式1)、VTM13.0默认的帧级帧内码率控制(模式2)、文献[20]的帧内码率控制以及文献[22]的帧内码率控制作为比较算法。实验采用本文的数据

集对文献[20]和文献[22]中的CNN网络进行训练。每个测试序列只编码第一帧。码率控制的精度定义为

$$M = |T - T_{\text{Actual}}| / T \quad (12)$$

其中, T 是目标码率, T_{Actual} 是实际码率。

Δ PSNR 定义为

$$\Delta\text{PSNR} = \text{PSNR} - \text{PSNR}_{\text{Anchor}} \quad (13)$$

其中, $\text{PSNR}_{\text{Anchor}}$ 是模式1的客观编码质量, PSNR 是其它码率控制算法的客观编码质量。码率控制精度的对比结果见表1。

表1 不同算法的比特率精度对比

Table 1 Comparison of bit rate accuracy of different algorithms

Sequences	T (kb/s)	模式1		本文		文献[20]		文献[21]		模式2	
		Actual	$M/\%$								
People on Street	12 000	12 033.36	0.28	12 309.12	2.58	11 921.52	0.65	11 895.12	0.87	10 980.72	8.49
	30 000	30 015.84	0.05	30 565.92	1.89	29 985.36	0.05	29 941.44	0.20	23 574.24	21.42
	50 000	49 658.88	0.68	50 480.16	0.96	50 058.96	0.12	49 948.80	0.10	38 032.80	23.93
Traffic	100 000	100 381.68	0.38	106 389.60	6.39	100 650.24	0.65	100 766.16	0.77	87 779.28	12.22
	150 000	150 452.16	0.30	158 346.72	5.56	151 142.16	0.76	151 094.64	0.73	131 082.00	12.61
	300 000	290 238.72	3.25	301 774.80	0.59	295 257.84	1.58	294 558.24	1.81	242 257.44	19.25
	Basketball	1 500	1 513.20	0.88	1 584.40	5.63	1 605.20	7.01	1 641.60	9.44	1 214.80
Drive	2 000	1 995.20	0.24	2 076.00	3.80	2 904.80	4.74	2 110.40	5.52	1 621.20	18.94
	2 500	2 504.80	0.19	2 604.80	4.19	2 655.60	6.22	2 725.20	9.01	2 032.40	18.70
	Birds in Cage	3 000	3 039.55	1.32	3 270.91	9.03	3 085.63	2.85	3 069.12	2.30	2 626.94
	4 000	4 039.10	0.98	4 309.25	7.73	4 111.10	2.78	4 118.21	2.96	3 445.44	13.86
	6 000	6 050.88	0.85	6 560.06	9.33	6 211.39	3.52	6 099.46	1.66	5 218.75	13.02
	BQ Terrace	14 000	14 118.72	0.85	14 311.68	2.23	12 966.72	7.38	13 152.96	6.05	117 391.68
	30 000	30 194.40	0.65	31 734.72	5.78	29 834.40	0.55	29 881.92	0.39	229 470.24	664.90
	40 000	40 182.72	0.46	41 247.36	3.12	40 080.00	0.20	40 112.16	0.28	275 037.12	587.59
	Cactus	6 000	6 022.80	0.38	6 316.40	5.27	6 273.20	4.55	6 464.00	7.73	5 252.00
	12 000	11 952.40	0.40	12 493.20	4.11	11 788.80	1.76	11 833.20	1.39	10 600.00	11.67
	14 000	13 943.60	0.40	14 505.20	3.61	13 827.20	1.23	13 868.40	0.94	12 395.20	11.46
	Cafe	3 000	3 009.22	0.31	3 162.05	5.40	3 387.07	12.90	3 379.58	12.65	2 659.01
	4 000	3 983.04	0.42	4 065.98	1.65	4 344.77	8.62	4 365.50	9.14	3 384.96	15.38
	5 000	4 973.76	0.52	5 079.17	1.58	5 467.39	9.35	5 431.30	8.63	4 173.31	16.53
	Kimon 1	1 000	1 021.63	2.16	1 085.38	8.54	1 654.66	65.47	1 690.37	69.04	1 225.73
	2 000	2 023.10	1.16	2 178.24	8.91	2 946.43	47.32	3 019.01	50.95	2 379.07	18.95
	3 000	3 030.14	1.00	3 174.53	5.82	4 044.29	34.81	4 194.05	39.80	3 338.50	11.28
Average			0.75		4.74		9.38		10.10		96.52

表1中,模式1、本文算法、文献[20]所提算法、文献[22]中算法、模式2的平均比特率准确率结果分别为0.75%、4.74%、9.38%、10.10%和96.52%。结果表明,模式1有时无法达到目标码率,导致码控不够精确。例如,在"BasketballDrillText"中,目标比特率是2 000、4 000和6 000 kb/s。然而,实际的比特率是1 940、3 817.6和5 780.8 kb/s。同样的情况在"

PeopleOnStreet"的50 000,"Cactus"的1 200和1 400 kb/s, BasketballDrill的3 000、4 000和8 000 kb/s以及"RaceHorses"的5 000、8 000和1 000 kb/s中均有出现。相比之下,本文算法由于采用较高的码率情况,导致码控不够精确。但这同时也是本文算法提高帧内码控编码性能的主要原因。由于帧内码控是一项极具挑战性的任务,模式

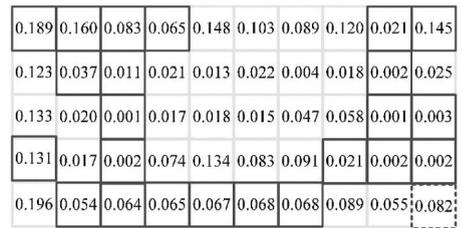
1 和本文算法的码率控制精度都保持在高精度水平。虽然文献[20]和文献[22]算法的码率控制性能较弱,但是其控制精度性能依然具有较高水平。

而对于模式 2 来说,其控制性能相比其它算法差很多,已经不能满足码率控制的需要。对于编码质量的比较见表 2。

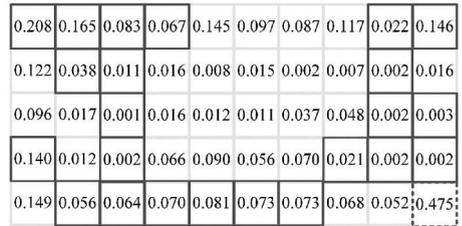
表 2 不同算法的实验比较结果
Table 2 Experimental comparison results of different algorithms

Sequences	T (kb/s)	Model PSNR	本文			文献[20]		文献[22]		模式 2	
			PSNR	Δ PSNR							
People on Street	12 000	33.28	33.36	0.08	33.11	-0.17	33.08	-0.20	32.89	-0.38	
	30 000	37.87	37.92	0.05	37.82	-0.05	37.81	-0.06	36.70	-1.17	
	50 000	40.50	40.60	0.10	40.53	0.02	40.52	0.01	39.08	-1.43	
Traffic	100 000	44.35	44.44	0.09	44.28	-0.07	44.29	-0.07	43.74	-0.61	
	150 000	46.38	46.48	0.10	46.34	-0.04	46.32	-0.05	45.71	-0.67	
	300 000	50.95	51.35	0.39	51.11	0.15	51.07	0.12	49.59	-1.36	
Basketball Drive	1 500	31.96	32.15	0.19	32.25	0.29	32.32	0.36	31.05	-0.91	
	2 000	33.10	33.21	0.11	33.24	0.14	33.35	0.24	32.29	-0.81	
	2 500	33.93	34.13	0.20	34.18	0.26	34.31	0.38	33.13	-0.80	
Birds In Cage	3 000	40.01	40.20	0.19	39.99	-0.02	39.98	-0.03	39.70	-0.31	
	4 000	40.76	40.89	0.13	40.75	-0.01	40.75	0	40.43	-0.32	
	6 000	41.71	41.86	0.15	41.72	0.02	41.66	-0.05	41.45	-0.26	
BQ Terrace	14 000	29.89	29.89	0.01	29.39	-0.50	29.46	-0.42	40.31	10.42	
	30 000	33.24	33.41	0.17	33.08	-0.16	33.10	-0.14	46.35	13.11	
	40 000	34.58	34.62	0.04	34.45	-0.13	34.44	-0.13	48.11	13.54	
Cactus	6 000	30.81	30.89	0.09	30.88	0.07	31.01	0.20	30.40	-0.40	
	12 000	33.47	33.60	0.13	33.34	-0.13	33.36	-0.11	33.13	-0.35	
	14 000	34.08	34.23	0.15	33.98	-0.10	33.98	-0.10	33.71	-0.37	
Cafe	3 000	41.56	41.77	0.20	42.07	0.51	42.03	0.47	41.09	-0.47	
	4 000	42.66	42.73	0.07	42.96	0.30	42.96	0.30	42.04	-0.62	
	5 000	43.44	43.46	0.03	43.74	0.30	43.69	0.25	42.79	-0.64	
Kimon 1	1 000	32.38	32.56	0.17	34.54	2.15	34.70	2.32	33.23	0.84	
	2 000	35.59	35.85	0.26	37.55	1.96	37.72	2.12	36.51	0.92	
	3 000	37.70	37.84	0.14	39.14	1.44	39.36	1.65	38.23	0.52	
Average				0.14		0.26		0.29		1.14	

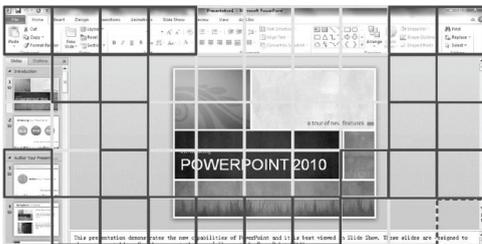
从表 2 中可以看出,与模式 1 相比,本文算法、文献[20]、文献[22]和模式 2 的编码质量分别提高 0.14、0.26、0.29 和 1.14 dB。由于本文算法通过 CNN 网络准确先知 $R - D$ 关系曲线,通过帧级率失真优化策略推导 CTU 最佳码率分配,然后根据 $R - \lambda$ 曲线,获取最优拉格朗日因子,使得码率与失真达到最佳的平衡状态。因此,本文算法在保证较高的码率控制精度情况下,能有效提高码控编码质量。尽管文献[20]、文献[22]也提高了编码性能,但都以牺牲码率精度为代价。模式 2 的编码性能虽然也有提高,但是其控制精度的性能太差。对于每个 CTU 分配码率情况如图 7 所示。



(b) 本文算法的实际码率分布



(c) 模式 1 的实际码率分布



(a) ground truth 图像

图 7 Ground truth 图像及不同算法的实际码率分布

Fig. 7 Ground truth images and actual code rate distributions of different algorithms

SlideShow 中 $T = 1\,000$ kb/s 时不同算法的每个 CTU 的实际 bpp 分布如图 7 所示。其中,黑色标记的 CTU 表示本文算法比模式 1 分配的码率少,灰色的

标记的 CTU 表示本文算法比模式 1 分配的码率多。从图中可以看出,本文算法在纹理复杂区域,分配的码率多于模式 1 分配的码率,而黑色的 CTU 具有相对平坦的纹理,本文算法对其节省码率,再将节省的码率分配给其它纹理复杂区域。最后一个 CTU 使用黑色虚线标识,可以看到模式 1 为了控制精度,在最后一个 CTU 上分配过多的码率,使得总体的编码中可使用的码率大量减少,从而影响到编码质量。而本文算法总体码率分配较为合理,使得码率控制性能得到提高。主观质量的比较如图 8、图 9 所示。

轮廓丢失严重。然而,本文算法的同区域相对清晰。从图 9 中可以看到,图 9(c-1) 中的篮球区域明显变形,而图 9(b-1)、9(d-1) 和 9(e-1) 中是较为完整的。在图 9(b-2)-9(e-2) 4 幅图中可以看到,图 9(c-2) 中的插座区域几乎消失,而图 9(d-2) 和图 9(e-2) 中的其他区域比图 9(b-2) 更模糊。

综上所述,本文所提出的算法相比其他算法有更好的主观表现效果。由于帧内码率控制将影响后续的编码质量,因此在目标码率下提高帧内编码质量非常重要。

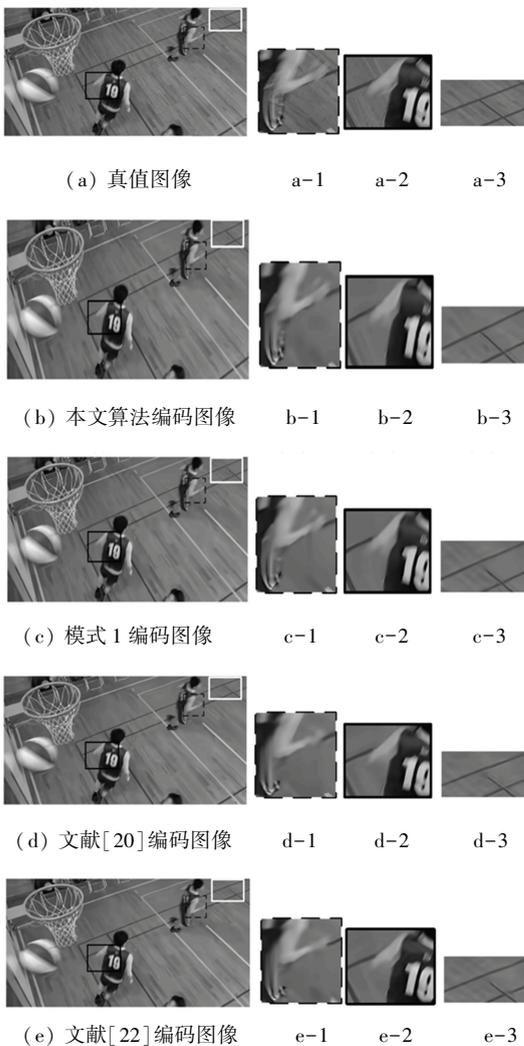


图 8 BasketballDrill 上不同算法的主观表现图 ($T=2\ 000\ \text{kps}$)

Fig. 8 Subjective representations of different algorithms on Basketball Drill ($T=2\ 000\ \text{kps}$)

由图 8 可见,图 8(b-1) 中的手臂和手部区域比图 8(c-1)、8(d-1) 和 8(e-1) 中的区域更为清晰;图 8(c-2) 中肘部区域的主观效果不如其他算法模型。此外,在白线框区域,明显可见图 8(c-3)、8(d-3) 和 8(e-3) 中白色区域十分模糊,甚至外形

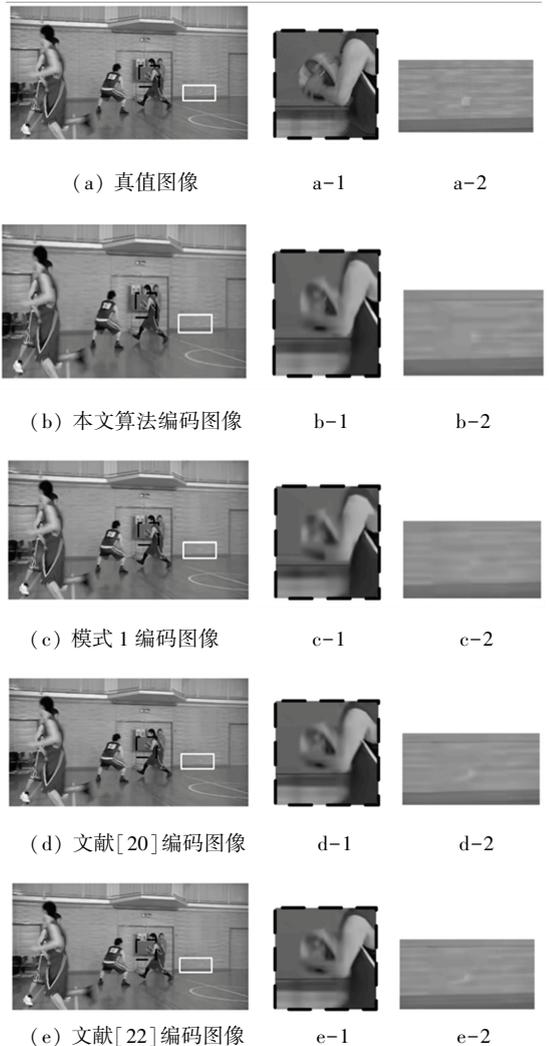


图 9 BasketballDrive 上不同算法的主观表现图 ($T=2\ 500\ \text{kps}$)

Fig. 9 Subjective performance of different algorithms on Basketball Drive ($T=2\ 500\ \text{kps}$)

4 结束语

本文在 VVC 中提出了一种基于 CNN 的用于帧内码率控制的 CTU 初始最优码率分配和 λ 选择。首先,将 $R-D$ 和 $R-\lambda$ 关系建模为双曲线函数,并

设计了一个具有4个输出的CNN来预测 $R-D$ 和 $R-\lambda$ 曲线的关键参数。由于提出的CNN可以有效地提取图像特征,预测的 $R-D$ 和 $R-\lambda$ 曲线符合编码特征。将一帧的失真分解为所有CTU的失真之和,通过推导一帧的RDO过程,得到与帧 λ 和 $R-D$ 参数相关的最佳码率分配表达式。最后,根据分配的码率和 $R-\lambda$ 函数,计算出CTU的最佳 λ ,这样就可以有效地利用码率进行编码。实验结果表明,本文算法可以显著提高帧内编码质量,并具有良好的码率控制精度。

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, 2013, 22(12): 1649-1668.
- [2] BROSS, CHEN J, LIU S. Versatile video coding (Draft 5): JVET - N1001 - v10 [S]. Geneva: Joint Expert Group on International Video Encoding and Decoding Standards, 2019.
- [3] HUANG Y W, HSU C W, CHEN C Y, et al. A VVC proposal with quaternary tree plus binary-ternary tree coding block structure and advanced coding techniques[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(5): 1311-1325.
- [4] PARK S H, KANG J W. Fast affine motion estimation for versatile video coding (VVC) encoding [J]. IEEE Access, 2019, 7: 158075-158084.
- [5] YANG H, SHEN L, DONG X, et al. Low-complexity CTU partition structure decision and fast intra mode decision for versatile video coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(6): 1668-1682.
- [6] HE Z, MITRA S K. Optimum bit allocation and accurate rate control for video coding via ρ -domain source modeling[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2002, 12(10): 840-849.
- [7] CHOI H, YOO J, NAM J, et al. Pixel-wise unified rate-quantization model for multi-level rate control[J]. IEEE Journal of Selected Topics in Signal Processing, 2013, 7(6): 1112-1123.
- [8] SULLIVAN G J, WIEGAND T. Rate-distortion optimization for video compression[J]. IEEE Signal Processing Magazine, 1998, 15(6): 74-90.
- [9] LI B. Encoding with fixed Lagrange multipliers[C]//Proceedings of the 11th Meeting on JCTVC-J0242, JCTVC of ISO/IEC and ITU-T. IEEE, 2012.
- [10] LI B. Rate control by $R-\lambda$ model for HEVC [C]//Proceedings of the 11th Meeting on JCTVC-K0103, JCTVC of ISO/IEC and ITU-T. IEEE, 2012.
- [11] KARCZEWICZ M, WANG X. Intra frame rate control based on SATD[C]//Proceedings of the 13th Meeting Document JCTVC-M0257. IEEE, 2013.
- [12] WANG M, NGANK N, LI H. An efficient frame-content based intra frame rate control for high efficiency video coding[J]. IEEE Signal Processing Letters, 2014, 22(7): 896-900.
- [13] GAO W, KWONG S, ZHOU Y, et al. SSIM-based game theory approach for rate-distortion optimized intra frame CTU-level bit allocation[J]. IEEE Transactions on Multimedia, 2016, 18(6): 988-999.
- [14] LI Y. "Rate control for VVC" in Joint Video Explor[Z]. Geneva: JVET, 2018.
- [15] LIU Z Z, LI Y M, CHEN Z Z. "Adaptive lambda ratio estimation for rate control in VVC," in Joint Video Explor [Z]. Geneva: JVET, 2018.
- [16] MAO Y, WANG M, WANG S, et al. High efficiency rate control for versatile video coding based on composite Cauchy distribution [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(4): 2371-2384.
- [17] LI Y, LIU Z, CHEN Z, et al. Rate control for versatile video coding [C]//Proceedings of 2020 IEEE International Conference on Image Processing (ICIP). IEEE, 2020: 1176-1180.
- [18] CHEN Y, KWONG S, ZHOU M, et al. Intra frame rate control for versatile video coding with quadratic rate-distortion modelling [C]//Proceedings of ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020: 4422-4426.
- [19] DU X, CAI Y, WANG S, et al. Overview of deep learning [C]//Proceedings of the 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). IEEE, 2016: 159-164.
- [20] LI Y, LI B, LIU D, et al. A convolutional neural network-based approach to rate control in HEVC intra coding [C]//Proceedings of 2017 IEEE Visual Communications and Image Processing (VCIP). IEEE, 2017: 1-4.
- [21] LI Y, LIU D, CHEN Z B. "CNN-based lambda-domain rate control for intra frames [C]//Proceedings of the 13th Meeting on JVET-M0215. Marrakech: JVET, 2019.
- [22] WEI L, YANG Z, WANG Z, et al. A CNN-based optimal CTU λ decision for HEVC intra rate control [J]. IEICE Transactions on Information and Systems, 2021, 104(10): 1766-1769.