

数据驱动的 AVS3 像素域最小可觉差预测模型

李兰兰, 刘晓琳, 吴珂欣, 林丽群, 魏宏安, 赵铁松

(福州大学物理与信息工程学院, 福建省媒体信息智能处理与无线传输重点实验室, 福州 350108)

摘要: AVS3 作为中国第三代国家数字音视频编码技术标准, 在消除视频时域/空域冗余信息方面发挥了重要的作用, 但在消除感知冗余方面仍存在进一步优化的空间。本文提出一种数据驱动的 AVS3 像素域最小可觉差 (Just noticeable distortion, JND) 预测模型, 在尽量保证视觉主观质量的前提下, 对 AVS3 视频编码器进行优化。首先基于主流的大型 JND 主观数据库, 获取符合人眼视觉特性的像素域 JND 阈值; 然后基于深度神经网络构建像素域 JND 预测模型; 最后通过预测的像素域 JND 阈值建立残差滤波器, 消除 AVS3 的感知冗余, 降低编码比特率。实验结果表明, 与 AVS3 的标准测试模型 HPM5.0 相比, 在人眼主观感知质量几乎无损的情况下, 所提出的像素域 JND 模型最高可节省 21.52% 的码率, 平均可节省 5.11% 的码率。

关键词: 视觉感知特性; 最小可觉差预测模型; AVS3; 残差滤波

中图分类号: TP391.4 **文献标志码:** A

Just Noticeable Distortion Prediction Model of Data-Driven AVS3 Pixel Domain

LI Lanlan, LIU Xiaolin, WU Kexin, LIN Liqun, WEI Hongan, ZHAO Tiesong

(Fujian Key Lab for Intelligent Processing and Wireless Transmission of Media Information, College of Physics and Information Engineering, Fuzhou University, Fuzhou 350108, China)

Abstract: The hybrid coding framework of the third generation audio and video coding standard (AVS3) plays an important role in eliminating redundant information in the video time domain/space domain, but needs to be further improved in eliminating perceptual redundancy and further improving coding performance. This paper proposes a just noticeable distortion (JND) prediction model of data-driven pixel domain to optimize AVS3 video encoder under the premise of ensuring the subjective quality of vision. Firstly, based on the current large subjective database of JND, the threshold of perceptive perception distortion in the pixel domain is obtained according to the human eye characteristics. Secondly, the pixel domain JND prediction model based on deep neural network is constructed. Finally, the residual filter established by the predicted pixel domain JND threshold is used to eliminate perceptual redundancy in AVS3 and reduce coding bitrate. The experimental results show that compared with the AVS3 standard test model HPM5.0, the proposed JND model can save up to 21.52% bitrate and an average of 5.11% bitrate.

Key words: visual perception characteristics; just noticeable distortion (JND) model; advanced audio and video coding standard (AVS3); residual filter

基金项目: 国家自然科学基金面上(61671152)资助项目; 国家自然科学基金青年科学基金(61901119)资助项目; 福建省自然科学基金(2019J01222)资助项目。

收稿日期: 2020-07-15; **修订日期:** 2020-09-30

引 言

大数据时代促进多媒体技术和现代通信技术的迅速发展,也为视频技术带来新的机遇与挑战。视频能够最大程度地表达内容,是人们生活中不可缺少的一部分。视频压缩技术主要是通过去除视频数据中的各种冗余信息来提升视频压缩比,便于视频的储存传输。目前的视频编码标准针对视频数据进行压缩编码,去除了大量时域冗余和空域冗余,但仍存在部分的视觉感知冗余。为了使视频的压缩性能得以进一步提升,研究人员通过研究人眼视觉感知系统特性(Human visual system, HVS),将视觉感知冗余数量化,并将其运用到视频编码标准中。

基于HVS构建的最小可觉差(Just noticeable distortion, JND)模型对视觉感知冗余的量化估计较为准确,对感知视频编码优化效果较好。JND模型是从心理学的角度出发,通过模拟人类视觉系统的感知特性得到人眼的最低视觉门限阈值,当变化值低于阈值时,人眼将无法觉察出该变化。对于JND模型的研究,文献[1]提出了JND阈值的产生原因是潜在的时空敏感性和空间掩蔽效应。文献[2]介绍了无序隐藏对JND估计的影响。文献[3]提出通过使用视觉显著性来调整JND阈值,消除更多的感知冗余。文献[4]提出了一种结合视觉显著性和JND量化控制算法,为每个编码树单元分配自适应的量化参数,节省比特率。文献[5]提出了一种基于感知失真度量的高效视频编码早期SKIP模式决策方法,节省了时间复杂度。

JND模型主要分为像素域JND模型和变换域JND模型,根据二者不同的特性作用在不同的领域。变换域的JND模型适用于边缘区域和纹理区域的JND阈值估计,因此在图像/视频处理中应用较为广泛^[6-7]。文献[8]提出了一种更广义的离散余弦变换(Discrete cosine transform, DCT)域中的可见性模型,可以估计DCT核任意大小且残差变换系数任意分布的失真可见性程度,实现了比特率降低的显著性能改善。文献[9]通过结合HVS的双目特性、背景亮度和每个DCT分量的空间频率,提出了一种DCT域的三维最小可觉差模型,可以在保持图像主观质量损失较低情况下节省比特率。文献[10]通过将视频压缩中的量化操作应用到PVCJND建模中,提出了一种显著量化失真模型,用于任何视频压缩方案之前的预处理。

像素域JND模型由于其阈值表示的意义更加直观而常被用于视频编码的运动估计、率失真优化以及预测残差块的自适应滤波。文献[11]提出将JND感知模型应用于H.264/AVC编解码器,通过JND阈值自适应地控制编码器的残差量化,从而提高率失真性能和整体压缩效率。文献[12]提出一种基于视觉注意力的像素域JND模型,通过自适应调节拉格朗日乘子实现根据视觉冗余度调节率失真的优化过程。文献[13]通过建立JND模型对原始的率失真优化算法进行改进,利用JND值自适应修正量化参数和调节朗格朗日乘数,实现编码比特和失真的感知平衡。文献[14]提出了一种基于变换域与像素域结合的即时差分JND模型,对于HEVC的变换跳过模式采用了现有的像素域JND模型,对于变换非跳过模式则采用变换域JND模型,以进一步减少感知冗余。

目前的AVS3视频编码标准^[15]主要是基于客观度量标准来编码视频数据,在消除视频时域/空域冗余信息方面发挥了很大的作用,但在消除感知冗余方面仍存在进一步优化的空间。在AVS3视频编码中应用JND计算视觉感知冗余,可以进一步提高视频编码性能。本文提出通过主流的JND大型主观数据集来获取人眼真实的像素域JND阈值,基于深度神经网络构建符合主观感受的像素域JND预测模型,在AVS3中嵌入像素域JND阈值构建的自适应残差滤波器进行编码优化,可以在保证人眼主观感知体验质量几乎不变的情况下节省更多的码率,优化AVS3视频编码器。

1 像素域 JND 预测模型

1.1 基于数学统计的像素域 JND 模型

像素域 JND 模型能够更加简单直观地数量化感知冗余。像素域 JND 阈值影响因素主要有两个:亮度自适应掩蔽和纹理掩蔽。亮度自适应掩蔽指的是人类视觉系统对相对亮度更加敏感。大量研究表明,亮度自适应掩蔽阈值与背景亮度之间满足一种如图 1 所示的非线性“U”形关系^[16]。纹理掩蔽效应是指人眼对于纹理复杂区域所能容忍的 JND 阈值要高于纹理平坦区域,因此现有的大部分像素域 JND 模型的基础框架都是将亮度自适应与纹理掩蔽进行非线性叠加,如式(1)所示。

$$T_{\text{JND}} = T_b + T_c - p \times \min(T_b, T_c) \quad (1)$$

式中: T_{JND} 表示像素域 JND 阈值; T_b 表示亮度自适应掩蔽; T_c 表示纹理掩蔽; p 为可以补偿掩蔽因子之间重叠的增益减少因子,通常取经验值 0.3。

大部分像素域 JND 模型主要是通过分析各种人眼视觉特性,优化如式(1)所示的像素域 JND 基础模型,通过数学公式推导出像素域 JND 阈值。然而各种研究表明人眼视觉特性非常复杂且受很多因素影响,数学模型推导的像素域 JND 阈值无法考虑到所有影响因素,与人眼真实视觉感知存在一定的偏差,无法完全符合人眼主观感知特性,较难保证不同视频编码场景的适应性。

1.2 基于数据驱动的像素域 JND 模型

深度学习的发展为图像视频编码领域带来了新的发展潜力。与基于数学统计方法构建的 JND 模型不同,通过深度神经网络构建合适的像素域 JND 模型,能够实现适应更多场景和更符合人眼真实视觉感知体验的像素域 JND 阈值预测。因此本文提出基于现有的大型视频主观数据库 VideoSet^[17]获取符合人眼真实的像素域 JND 阈值,通过获取的像素域 JND 阈值和 VideoSet 视频集来构建像素域 JND 预测模型。由于 VideoSet 主观数据库涵盖了多种场景下的 JND 统计,因此本文基于该数据库构建的像素域 JND 预测模型能够适用于更多的视频编码场景。

1.2.1 基于主观数据库的像素域 JND 阈值

VideoSet 主观数据库是根据 JND 理论测量的大规模主观测试数据集,视频集包括 220 个 5 s 序列,每个序列具有 4 个分辨率(1 920 像素×1 080 像素,1 280 像素×720 像素,960 像素×540 像素和 640 像素×360 像素),共有 880 个未压缩的视频序列。通过 ITU-R BT.2022^[18]建立主观实验环境进行大量的主观实验,统计得出每个原始视频所对应的最小可觉差的量化参数 Q_{JND} 。

本文通过对原始视频与对应的压缩视频(量化参数为 Q_{JND})组成的视频对进行处理,获得每个原始视频的像素域 JND 阈值。为了增加数据的多样性和数量,在神经网络进行训练时有更好的泛化性能。本文将原始视频和对应压缩视频切割成大小为 64 像素×64 像素的编码宏块(Large coding block, LCU)。通过计算每个原始视频 LCU 与对应压缩视频 LCU 的平均绝对误差(Mean absolute error, MAE),将其作为原始视频 LCU 块像素点的像素域 JND 阈值,如式(2)所示。

$$\text{MAE} = \frac{\sum_{i=0}^m \sum_{j=0}^m |l(i, j) - l'(i, j)|}{m \times m} \quad (2)$$

式中: $l(i, j)$ 为原始视频像素值; $l'(i, j)$ 为对应的压缩视频(量化参数为 Q_{JND})像素值; m 为 LCU 的边长 64。

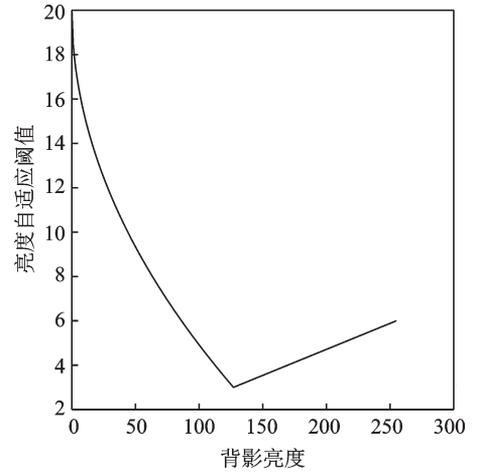


图1 亮度自适应阈值与背景亮度关系图

Fig.1 Relation between luminance adaptive threshold and background luminance

对计算所得的每个LCU的像素域JND阈值进行统计分析,分布规律为纹理复杂的LCU像素域JND阈值较大,纹理简单的LCU像素域JND阈值较小,符合像素域JND阈值分布的一般规律。原始图像与加入计算所得的像素域JND阈值噪声后的失真图像对比以及局部细节图对比如图2所示。从对比图中几乎无法察觉到原始图像与失真图像之间的区别,因此本文利用VideoSet主观数据库获得的像素域JND阈值在很大程度上符合人眼视觉主观感受。

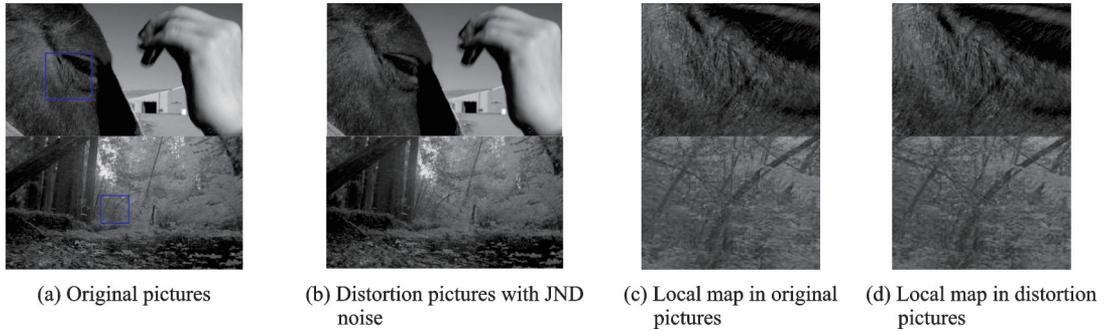


图2 原始图像与失真图像对比
Fig.2 Comparison of original and distorted pictures

1.2.2 J-VGGNet网络构建

由于像素域的JND阈值产生原因主要与背景亮度和纹理分布相关,并且大量研究表明人眼对于亮度敏感而对于色度并不敏感,因此本文将每个原始视频LCU块的灰度图矩阵作为神经网络的输入,而目标变量为每个LCU块的像素域JND阈值 $T_{JND-LCU}$,表达式为

$$T_{JND-LCU} = f \begin{pmatrix} p(0, 0), \dots, p(0, m) \\ p(1, 0), \dots, p(1, m) \\ \vdots \\ p(m, 0), \dots, p(m, m) \end{pmatrix} \quad (3)$$

式中 $p(x, y)$ 为原始视频LCU块内每个像素点灰度值; $m=64$ 。预测函数基于改进后的J-VGGNet网络实现。

本文提出的J-VGGNet网络主要是以VGG16^[19]网络模型作为基本网络结构,其结构如图3所示。

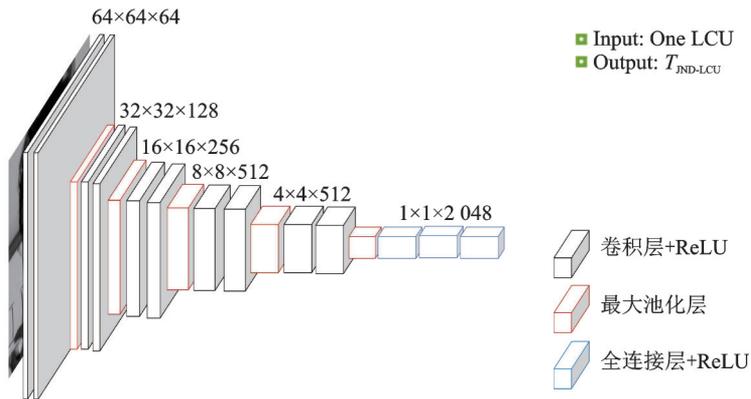


图3 J-VGGNet结构图
Fig.3 Framework of J-VGGNet

VGG16网络拥有较深的网络结构和较小的卷积核与池化核,使其能在参数个数一定的情况下获取更多图像特征,因此本文所提出的像素域 JND 预测模型通过深度神经网络来提取图像特征,与数学统计的 JND 模型人工提取图像特征的方式相比表现更佳。

VGG16 网络每个卷积层之后紧跟 Relu 激活函数可增加其非线性特性,因此本文构建的深度像素域 JND 预测模型可以充分拟合复杂的非线性关系。由于本文主要将 VGG 网络用于提取图像特征后进行非线性预测,与原来的 VGG16 相比,所提出的 J-VGGNet 对于卷积层做了精简并去掉了用于分类的 Softmax 激活层,减少了网络训练参数,训练速度得到提升。因此本文的 J-VGGNet 网络主要结构包括 10 个卷积层和 3 个全连接层。

本文所提出的 J-VGGNet 网络的优化算法主要是采用了随机梯度下降法 (Stochastic gradient descent, SGD),并结合了动量梯度下降法,动量参数设为默认值 0.9。由于该网络主要是用于进行非线性预测,故选用均方误差 (Mean square error, MSE) 作为损失函数,计算公式为

$$MSE = \frac{\sum_{i=1}^n (y_i - y_i')^2}{n} \tag{4}$$

训练样本为 VideoSet 视频集中未进行压缩的原始视频 LCU,样本标签为通过 VideoSet 数据库所获得的 LCU 像素域 JND 阈值。将处理之后的样本集分为训练集与验证集,并按照 9:1 分布,训练集由 257 400 对数据组成,验证集由 28 600 对数据组成。深度学习框架选择 Pytorch,迭代次数设为 50 次, Batch_size 设定为 64。

2 自适应 JND 残差滤波

通过以上构建的 JND 模型可以预测出与人眼视觉特性相符合的像素域 JND 阈值。本文基于该模型所预测的像素域 JND 阈值构建自适应 JND 残差滤波模型,并将其嵌入 AVS3 标准的高性能编码平台 HPM 中进行编码,如图 4 所示。先将视频信号信息传入已经训练好的 JND 模型中,对每帧视频

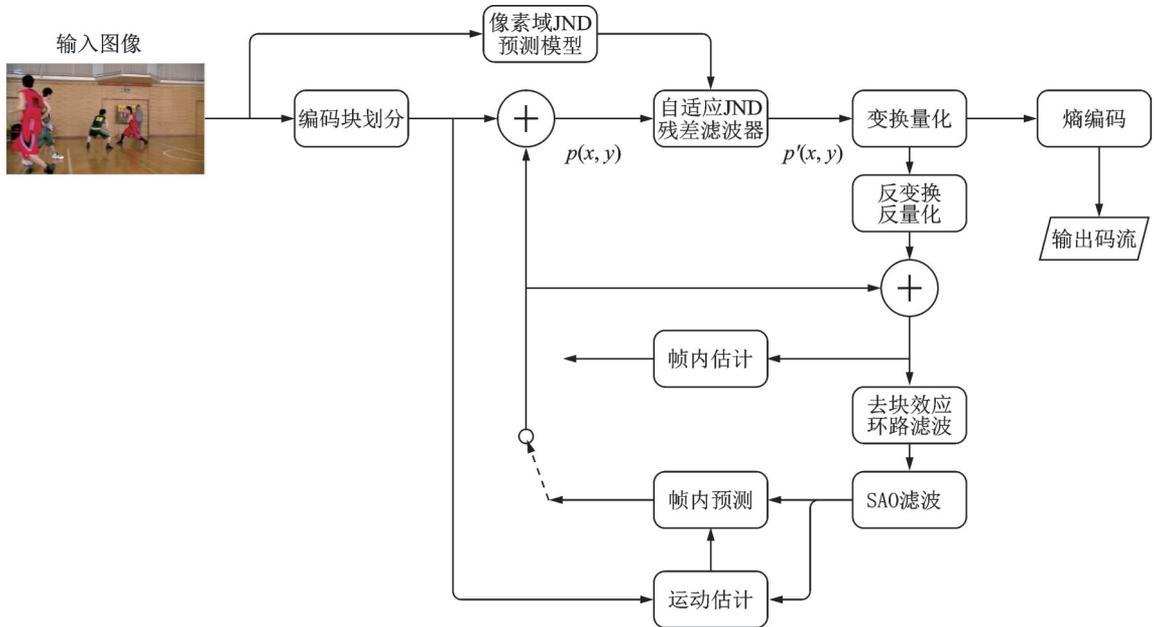


图4 本文所提算法框图

Fig.4 Framework of the proposed algorithm

信号进行分块得到LCU,再预测出每个LCU块的像素域JND阈值,基于文献[20]的残差滤波公式构建自适应JND残差滤波器,将其嵌入运动补偿与残差块变换量化之间的位置。自适应JND残差滤波计算公式为

$$p'(x, y) = \begin{cases} p(x, y) + T_{\text{JND-LCU}} & p(x, y) - \overline{p_{\text{LCU}}} < -T_{\text{JND-LCU}} \\ \overline{p_{\text{LCU}}} & |p(x, y) - \overline{p_{\text{LCU}}}| < T_{\text{JND-LCU}} \\ p(x, y) - T_{\text{JND-LCU}} & p(x, y) - \overline{p_{\text{LCU}}} > T_{\text{JND-LCU}} \end{cases} \quad (5)$$

式中: $p(x, y)$ 为原始残差像素值; $p'(x, y)$ 为经过滤波后的残差像素值; $\overline{p_{\text{LCU}}}$ 为残差块的平均像素值。当残差块像素点值与平均像素点值之差小于 $T_{\text{JND-LCU}}$ 时,将该残差像素值设为平均值 $\overline{p_{\text{LCU}}}$,若差值绝对值大于 $T_{\text{JND-LCU}}$,则对残差像素值进行自适应调整。

3 实验结果及分析

为验证本文提出的JND模型对AVS3视频编码的优化效果,将基于JND阈值所构建的自适应残差滤波器引入AVS3的高性能编码平台HPM中,并且以HPM5.0为参考进行测试。本次测试采用的是AVS3标准的通用测试序列,共6组,其中测试序列Basketball和Cactus分辨率为1 920像素×1 080像素,其他4组测试序列分辨率为1 280像素×720像素。HPM5.0的配置文件为encode_RA.cfg;编码使用的量化参数Q为22、27、32和37;最大LCU尺寸为64像素×64像素,最大划分深度为4。

表1描述了嵌入本文算法模型的AVS3视频编码框架与AVS3标准视频编码框架HPM5.0的编码比特数和Y分量峰值信噪比Y-PSNR值的对比,图5给出了本文算法模型与HPM5.0的测试序列编码比特数的直观对比。图6为HPM5.0和所提算法模型解码视频的主观感知对比。表1中 ΔB 为节省比特率,其计算公式如式(6)所示。其中 B_{ref} 为HPM5.0编码比特数, B_{pro} 为所提算法模型编码比特数。

$$\Delta B = \frac{B_{\text{ref}} - B_{\text{pro}}}{B_{\text{ref}}} \times 100\% \quad (6)$$

编码时间复杂度计算公式如式(7)所示,其中 T_{ref} 为HPM5.0编码时间, T_{pro} 为所提算法模型编码时间。

$$\Delta T = \frac{T_{\text{pro}}}{T_{\text{ref}}} \times 100\% \quad (7)$$

结合表1和图5可以看出,本文所提出的模型算法与AVS3标准视频编码框架相比,节省了最高21.52%、平均5.11%的码率,且编码时间没有增加。由此可说明在保证客观质量差异较小的情况下,本文提出算法模型在一定程度上可以提升视频压缩率。

多尺度结构相似性(Multi-scale structural similarity, MS-SSIM)是结构相似性(Structural similarity, SSIM)评估方法的改进,用来衡量两幅图像多种尺度下的结构相似性平均值,最大值为1。MS-SSIM平均差异值表示HPM5.0标准算法的解码视频与本文算法模型解码视频之间的相似度差异值。由于MS-SSIM考虑到了人眼视觉系统特性,可以较好地反映人眼的主观感受,与主观评价相似度高,因此可作为客观评价指标PSNR的补充。目前,MS-SSIM已被AVS3标准采纳为质量评价算子。因此本文采用MS-SSIM对HPM5.0标准算法和所提出的模型算法进行主观质量评价,如表2所示。结合表1、2可以看出,当码率平均节省5.11%时,所提算法模型解码视频与HPM5.0解码视频的Y分量MS-SSIM平均差异值仅为0.000 429,几乎可以忽略不计。

传统图像质量客观评价标准PSNR不能很好地反映人眼感知失真,为了验证所提算法主观感知质量与比特节省率关系,本文进行了主观质量测试实验,实验结果如图7所示。所有的测试人员均在相对

表 1 本文所提模型与 HPM5.0 的码率对比

Table 1 Bitrate comparison of the proposed model and HPM5.0

测试序列	Q	编码比特/(kb·s ⁻¹)		比特节省/%	Y-PSNR / dB		Y-PSNR 差值/dB	编码时间复杂度/%
		HPM5.0	所提模型		HPM5.0	所提模型		
City	22	12 489.53	9 801.55	21.52	40.599 7	39.579 5	-1.020 2	88.83
	27	4 638.96	3 856.94	16.86	38.498 1	37.995 6	-0.502 5	92.95
	32	2 051.49	1 866.54	9.01	36.867 2	36.598 6	-0.268 6	99.61
	37	1 075.50	1 045.08	2.83	35.199 8	35.092 8	-0.107 0	103.47
Vidyo1	22	2 465.87	2 295.63	6.90	44.467 4	44.182 6	-0.284 8	94.63
	27	1 121.69	1 087.76	3.02	43.213 5	43.053 7	-0.159 8	97.68
	32	608.53	602.53	0.99	41.922 9	41.836 3	-0.086 6	100.73
	37	376.96	374.66	0.61	40.491 6	40.424 8	-0.066 8	102.53
Vidyo3	22	3 504.01	3 187.70	9.03	43.879 8	43.472 2	-0.407 6	97.07
	27	1 607.02	1 542.30	4.03	42.569 2	42.356 1	-0.213 1	99.62
	32	879.03	862.17	1.92	41.269 7	41.127 7	-0.142 0	98.43
	37	540.84	534.79	1.12	39.734 1	39.622 9	-0.111 2	101.39
Crew	22	7 584.20	6 684.38	11.86	41.876 7	41.454 1	-0.422 6	100.68
	27	2 740.25	2 590.11	5.48	40.440 4	40.266 2	-0.174 2	100.97
	32	1 246.81	1 223.66	1.86	39.317 1	39.213 8	-0.103 3	107.20
	37	661.13	652.65	1.28	38.055 9	37.972 1	-0.083 8	100.67
Basketball	22	22 513.89	21 616.29	3.99	39.976 8	39.783 5	-0.193 3	95.83
	27	9 128.15	9 018.69	1.20	38.789 7	38.653 8	-0.135 9	100.85
	32	4 552.58	4 504.42	1.06	37.723 3	37.596 0	-0.127 3	102.45
	37	2 573.39	2 544.82	1.11	36.462 8	36.358 0	-0.104 8	102.21
Cactus	22	26 456.31	23 643.56	10.63	39.056 2	38.667 7	-0.388 5	88.80
	27	9 788.90	9 339.42	4.59	37.828 5	37.659 3	-0.169 2	97.14
	32	4 559.14	4 518.58	0.89	36.689 1	36.551 2	-0.137 9	101.46
	37	2 546.04	2 527.00	0.75	35.368 9	35.252 7	-0.116 2	102.59
平均值				5.11			-0.230 3	99.07

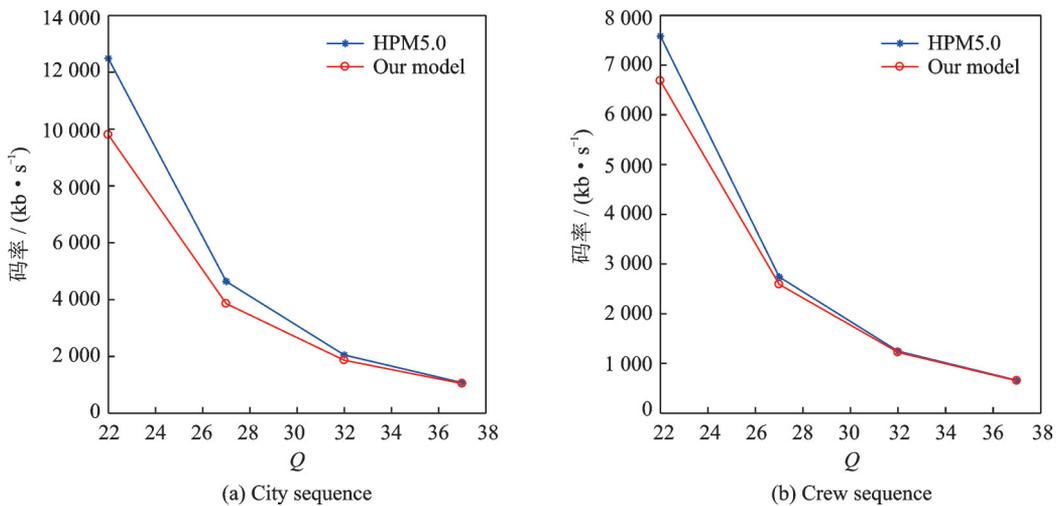


图 5 本文所提模型与 HPM5.0 测试序列编码比特数对比

Fig.5 Bitrate comparison of the proposed model and HPM5.0 on test sequences

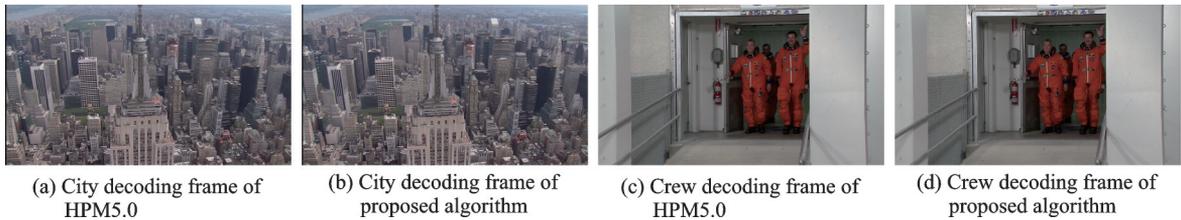
图6 测试视频序列解码视频对比($Q = 27$)Fig.6 Comparisons of the reconstructed frame of the test sequence($Q = 27$)

表2 本文模型与 HPM5.0 的 Y 分量 MS-SSIM 平均值对比

Table 2 Comparison of Y-MS-SSIM of the proposed model and HPM5.0

分辨率/(像素×像素)	测试序列	Y分量 MS-SSIM 平均值		Y分量 MS-SSIM 平均
		HPM5.0	所提模型	差异值
1 280×720	City	0.989 341	0.988 620	-0.000 721
	vidyo1	0.992 787	0.992 655	-0.000 132
	vidyo3	0.993 315	0.993 194	-0.000 121
	Crew	0.984 967	0.984 532	-0.000 435
1 920×1 080	BasketballDrive	0.984 142	0.983 576	-0.000 566
	Cactus	0.985 094	0.984 497	-0.000 597
平均值				-0.000 429

安静的环境中进行主观实验,实验中要求每个测试者观看 2 个连续显示的编码视频,并且比较所提算法的编码视频与 HPM5.0 编码视频哪个感知质量更好。主观实验中 27 位测试人员对每个序列编码视频对的主观感知结果统计表明:认为本文所提算法编码视频质量更好的总人数与认为 HPM5.0 编码视频质量更好的总人数各占 50% 左右。由此可以认为,测试人员感受不出 HPM5.0 编码视频与所提算法编码视频的感知差异,即本文所提出的像素域 JND 预测模型编码视频与 HPM5.0 编码视频相比可认为是感知无损的。

结合表 1 和图 7 主观实验结果可以说明,所提算法可以在节省一定码率的情况下保证主观感知质量无损,从而达到消除感知冗余、优化 AVS3 视频编码器的目的,体现了所提算法在编码性能上的先进性。

4 结束语

本文提出了一种基于神经网络和大型 JND 主观数据库构建的像素域 JND 预测模型,并将该模型引入 AVS3 视频编码器中进行编码优化,一定程度上消除了编码过程中的感知冗余。为了验证所提算法模型性能,本文设计了客观实验和主观实验。实验结果显示,与 HPM5.0 标准算法相比,所提模型在降低编码码率的同时,主观感知质量几乎没有损失,对 AVS3 的编码优化效果较好。下一步将考虑结合时域影响因素对 JND 预测模型进行调整,消除更多感知冗余,进一步提升 AVS3 编码效果。

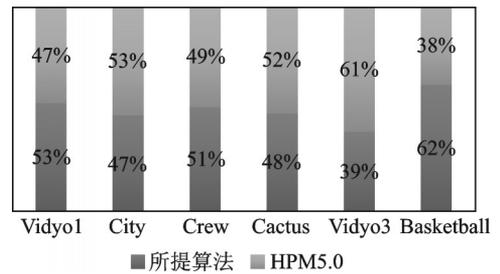


图7 主观实验结果

Fig.7 Subjective experimental results

参考文献:

- [1] ZHANG X H, LIN W S, XUE P. Improved estimation for just-noticeable visual distortion[J]. *Signal Processing*, 2005, 85(4): 795-808.
- [2] WU J, SHI G, LIN W S, et al. Just noticeable difference estimation for images with free-energy principle[J]. *Multimedia*, 2013, 15(7): 1705-1710.
- [3] WANG H, WANG H, HU X, et al. Perceptual video coding based on saliency and just noticeable distortion for H.265/HEVC [C]// *Proceedings of 2014 International Symposium on Wireless Personal Multimedia Communications (WPMC)*. Sydney, NSW: IEEE, 2014: 106-111.
- [4] NAMI S, PAKDAMAN F, HASHEMI M R. Juniper: A JND-based perceptual video coding framework to jointly utilize saliency and JND[C]// *Proceedings of 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*. London, UK: IEEE, 2020: 1-6.
- [5] KIM J, IZQUIERDO E. Early SKIP mode decision method in HEVC based on perceptual distortion measure[C]// *Proceedings of 2019 27th European Signal Processing Conference (EUSIPCO)*. A Coruna, Spain: IEEE, 2019: 1-5.
- [6] 尹肖波. 基于 MATLAB 实现的车牌识别系统[D]. 杭州: 浙江工业大学, 2017.
YIN Xiaobo. Car license recognition system design based on MATLAB[D]. Hangzhou: Zhejiang University of Technology, 2017.
- [7] 李淑芝, 龙香玉, 邓小鸿, 等. 结合纹理复杂度和 JND 模型的图像水印算法[J]. *计算机应用研究*, 2019, 36(2): 528-531, 542
LI Shuzhi, LONG Xiangyu, DENG Xiaohong, et al. Image watermarking algorithm combining texture complexity and JND model[J]. *Application Research of Computers*, 2019, 36(2): 528-531, 542
- [8] BAE S, KIM J, KIM M. HEVC-based perceptually adaptive video coding using a DCT-based local distortion detection probability model[C]// *Proceedings of IEEE Transactions on Image Processing*. [S.l.]: IEEE, 2016: 3343-3357.
- [9] JABALLAH S, LARABI M, TAHAR J B. Asymmetric DCT-JND for luminance adaptation effects: An application to perceptual video coding in MV-HEVC[C]// *Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Calgary, AB: IEEE, 2018: 1797-1801.
- [10] KI S, KIM M, KO H. Just-noticeable-quantization-distortion based preprocessing for perceptual video coding[C]// *Proceedings of 2017 IEEE Visual Communications and Image Processing (VCIP)*. St. Petersburg, FL: IEEE, 2017: 1-4.
- [11] NACCARI M, PEREIRA F. Decoder side just noticeable distortion model estimation for efficient H.264/AVC based perceptual video coding[C]// *Proceedings of Proceedings of the International Conference on Image Processing, ICIP 2010*. Hong Kong, China: IEEE, 2010.
- [12] 曾志鹏. JND 模型及其在 HEVC 感知视频编码的应用[D]. 厦门: 华侨大学, 2019.
ZENG Zhipeng. JND model and its application in HEVC perceptual video coding[D]. Xiamen: Huaqiao University, 2019.
- [13] 方少杰, 杨静. 基于恰可察觉失真模型的率失真优化算法[J]. *电视技术*, 2019, 43(2): 11-13.
FANG Shaojie, YANG Jing. Rate distortion optimization algorithm based on just noticeable distortion model[J]. *Video Engineering*, 2019, 43(2): 11-13.
- [14] KIM J, BAE S, KIM M. An HEVC-compliant perceptual video coding scheme based on JND models for variable block-sized transform kernels[C]// *Proceedings of IEEE Transactions on Circuits and Systems for Video Technology*. [S.l.]: IEEE, 2015: 1786-1800.
- [15] LI J, WANG M, ZHANG L, et al. History-based motion vector prediction for future video coding[C]// *Proceedings of 2019 IEEE International Conference on Multimedia and Expo (ICME)*. Shanghai, China: IEEE, 2019: 67-72.
- [16] 肖贺. HEVC 关键技术研究暨并行技术研究[D]. 北京: 北京邮电大学, 2018.
XIAO He. Research on key technology of HEVC and parallel technology[D]. Beijing: Beijing University of Posts and Telecommunications, 2018.
- [17] WANG H, KATSAVOUNIDIS I, ZHOU J, et al. VideoSet: A large-scale compressed video quality dataset based on JND

measurement[J]. *Journal of Visual Communication and Image Representation*, 2017, 46: 292-302.

- [18] International Telecommunication Union. ITU-R BT.2022: General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays[M]. [S.l.]: ITU, 2012.
- [19] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04) [2020-03-10]. <https://arxiv.org/abs/1409.1556>.
- [20] WU J, LI L, DONG W, et al. Enhanced just noticeable difference model for images with pattern complexity[J]. *Image Processing, IEEE Transactions on*, 2017, 26(6): 2682-2693.

作者简介:



李兰兰(1986-),女,博士,副教授,硕士生导师,研究方向:人工智能医学图像处理、高速光通信解决方案等, E-mail: 282520711@qq.com。



刘晓琳(1996-),女,硕士研究生,研究方向:智能视频编码、深度学习等。



吴珂欣(1997-),女,硕士研究生,研究方向:人工智能医学图像处理和智能视频编码与通信。



林丽群(1980-),女,讲师,研究方向:图像/视频信号处理、视觉质量评估和视频编码等。



魏宏安(1978-),通信作者,男,高级实验师,硕士生导师,研究方向:智能视频编码与通信、图像处理与计算机视觉和全媒体融合, E-mail: weihongan@fzu.edu.cn。



赵铁松(1984-),男,博士,教授,博士生导师,研究方向:图像处理与计算机视觉、智能视频编码与通信和触感信息与触感网。

(编辑:张黄群)