



研究与开发

一种基于编码单元快速划分的VVC帧内编码方法

钟辉, 陆宇, 殷海兵, 黄晓峰

(杭州电子科技大学通信工程学院, 浙江 杭州 310018)

摘要: 相比于高效视频编码 (high efficiency video coding, HEVC) 标准, 新一代编码标准多功能视频编码 (versatile video coding, VVC) 引入了很多新的技术, 其中包括四叉树 (quadtree, QT) 和多类型树 (multi-type tree, MTT) 划分, MTT划分由 HEVC 中的 QT 划分延伸而来。新划分方法提高了压缩效率, 但导致编码时间急剧增加。为了降低编码复杂度, 提出了一种结合深度学习方法和 MTT 方向早期判决的快速帧内编码算法。首先使用轻量级的卷积神经网络 (convolutional neural network, CNN) 对 QT 和部分 MTT 进行预测划分, 其余 MTT 则采用提前预测 MTT 划分方向的方法作进一步的优化。实验结果表明, 所提方法能够大幅降低编码复杂度, 相比于原始编码器的编码时间减少了 74.3%, 且只有 3.3% 的码率损失, 性能优于对比的方法。

关键词: VVC; 帧内编码; 卷积神经网络; 快速编码; 四叉树; 多类型树

中图分类号: TP393

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2024196

A VVC intra coding method based on fast partition for coding unit

ZHONG Hui, LU Yu, YIN Haibing, HUANG Xiaofeng

College of Communication Engineering, Hangzhou Danzi University, Hangzhou 310018, China

Abstract: Compared to the high efficiency video coding (HEVC) standard, the latest generation coding standard, versatile video coding (VVC) has introduced many new technologies, including quadtree (QT) and multi-type tree (MTT) partitioning. MTT partition is extended from QT partition in HEVC. The new partition method increases encoding complexity, leading to a sharp increase in encoding time. To reduce encoding complexity, a fast intra coding method combining deep learning methods and early decision in the MTT direction was proposed. Firstly, a light-weight convolutional neural network (CNN) network was used to predict partition for QT and partial MTT. Then, an early prediction for MTT partition direction method was adopted for further optimization of residual MTT. Experi-

收稿日期: 2024-04-03; 修回日期: 2024-08-02

基金项目: 国家自然科学基金资助项目 (No.61972123); 科技部重点研发课题项目 (No.2023YFB4502800); 浙江省教育厅科研项目 (No.Y202249588)

Foundation Items: The National Natural Science Foundation of China (No.61972123), The Key Research and Development Project Funding Program of Ministry of Science and Technology (No.2023YFB4502800), The Research Project of Zhejiang Provincial Department of Education (No.Y202249588)



mental results show that the proposed method can significantly reduce encoding complexity, with a 74.3% reduction in encoding time compared to the original encoder with only 3.3% rate loss. Moreover, the performance of proposed method is superior to other comparative algorithms.

Key words: VVC, intra coding, convolutional neural network, fast coding, quadtree, multi-type tree

0 引言

随着互联网的快速发展,人们对视频质量的要求逐渐提高,为了满足日益增长的需求,新一代视频编码标准由联合视频专家小组(Joint Video Experts Team, JVET)开发,并于2020年7月被最终确定作为新一代国际视频编码标准^[1]。与上一代高效视频编码(high efficiency video coding, HEVC)标准^[2]相比,新一代视频编码标准多功能视频编码(versatile video coding, VVC)显著降低了比特率,但同时增加了复杂性。为了降低帧内编码的复杂度,近年来涌现了很多VVC快速帧内编码方法。这些方法大致可以分为两类,其中一类是基于传统方式的快速编码方法,该方法主要是利用均值、方差、梯度等图像特征,以及编码比特、率失真代价等编码信息,分别设定相应的阈值来提前终止编码单元(coding unit, CU)划分或者跳过不必要的CU划分。Fan等^[3]指出均匀区域倾向于对整体较大的CU直接进行帧内预测,而复杂区域倾向于划分为较小的CU,并且子CU通常具有不同的纹理,基于此特点,首先利用CU方差预测CU是否划分,其次利用Sobel算子计算水平和垂直方向的梯度比值,进一步判断CU是否执行QT划分或MTT划分,最后计算子CU方差来对MTT划分进行预测,该方法能够直接预测出CU的划分模式,但该方法只针对32×32(单位为像素,下文同)大小的CU块,因此,能降低的编码复杂度是有限的。Li等^[4]提出利用原始像素亮度和预测像素亮度的差来获取CU失真,并利用失真建立两个早期跳过决策模型,分别预测CU是否跳过二叉

树(binary-tree, BT)或三叉树(ternary-tree, TT)划分,但也只针对多类型树(multi-type tree, MTT)的划分进行了加速,四叉树(quadtree, QT)划分并没有得到加速效果。Zhang等^[5]提出了一种基于跳过冗余编码块结构和VVC中不必要的划分方向的快速帧内编码算法,通过分析VVC中帧内编码的CU划分结果,发现存在一些冗余的划分,于是设计了跳过冗余划分的方法,还利用图像纹理方向等提前跳过VVC中不必要的划分方向,但是,该方法借助了一些人为观察的特征,导致加速效果不是很明显。Ni等^[6]提出了一种基于纹理分析的TT和BT划分策略,以及一种基于梯度的帧内模式决策方法,可分别加速TT和BT划分以及帧内模式预测,采用纹理检测方法计算CU的纹理复杂度和预测方向,但是只对TT和BT进行了加速,对QT划分没有进行处理。Li等^[7]提出了一种VVC帧内编码CU划分早期判决算法,基于像素之间的变化趋势,通过Schar算子提取图像的水平 and 垂直梯度特征,用于CU划分提前终止、CU划分方向提前判决和TT划分模式的早期跳过。Zhao等^[8]提出利用高斯拉普拉斯(Laplacian of Gaussian, LOG)算子提取编码帧中的边缘,分为带边缘的复杂纹理区域和不带边缘的简单纹理区域,CU在水平和垂直方向上被平均划分为4个子块,分别计算水平和垂直边缘的特征值,根据得到的特征值可以提前跳过不必要的划分,对于没有边缘的CU,则根据相邻CU的深度信息决定是否终止划分。另一类VVC快速帧内编码方法是基于学习的方法,分为基于机器学习和基于深度学习两种。基于机器学习的方法利用决策树、随机森

林、支持向量机等实现快速编码,基于深度学习的方法则通过构建深度神经网络来实现快速编码。这两种基于学习的快速编码方法通过训练数据集获得预测模型,再将预测模型导入编码器实现对CU划分的分类和预测。尤其是基于深度学习的快速编码方法利用深度神经网络训练大量的数据集来提高CU划分的准确度,这种以数据驱动的编码方法可以自动调整特征并具有很强的灵活性和适应性。相比于基于传统方式的快速编码方法,基于学习的快速编码方法的编码加速效果更加显著,但是需要更多的计算资源。Si等^[9]提出了结合随机森林分类器对CU划分类型进行快速判断,并融入了传统快速自适应CU划分决策算法,该算法能够针对不同的CU尺寸进行划分决策,并完全绕过率失真优化(rate-distortion optimization, RDO)的复杂过程,从而显著减少了编码时间。Wu等^[10]提出了一种基于支持向量机的VVC帧内编码快速CU划分算法,该算法利用纹理信息预测CU划分来提前终止冗余划分,针对不同大小的CU分别训练了分类器,以提高CU划分的准确性以及控制分类器本身的复杂性,并且为每个分类器设置了不同的阈值,以实现编码复杂度和RD性能之间的权衡。He等^[11]提出了基于随机森林的算法来降低CU划分的复杂性,首先将CU分为3类,即简单CU、模糊CU和复杂CU,对于简单和复杂的CU,训练一个随机森林分类器来直接预测最佳划分模式,对于模糊CU,训练另一个分类器来预测当前划分过程是否终止。Saldanha等^[12]提出了一种使用轻量级梯度提升机(light gradient boosting machine, LGBM)分类器的划分方法,为了避免全部遍历RDO过程,该方法跳过不太可能被选为最佳划分类型的CU划分判决,利用纹理、编码和上下文信息等有效特征,对5个分类器进行了离线训练,并针对不同的场景有不同的配置,在节省时间和编码效率之间进行不同的权衡。为了进一步降低VVC

帧内编码复杂度,Abdallah等^[13]使用了基于卷积的神经网络对帧内的QT和MTT划分进行决策,从而确定每个128×128编码树单元(coding tree unit, CTU)的QTMTT的划分,但提出的算法仅对32×32大小CU的BT深度和64×64大小的QT深度进行了快速划分。Tissier等^[14]提出了一种基于卷积神经网络(convolutional neural network, CNN)的编码复杂度降低方法,CNN以大小为64×64块的亮度CU作为输入,并生成一个向量,该向量反映了在块的4×4边界处具有边缘的概率,编码器进一步利用这个概率向量来跳过不太可能的CU划分,从而实现编码复杂度的降低。Li等^[15]设计了一个自适应损失函数来训练MSE-CNN模型,并采用早期退出机制来确定CU划分,并且提出了一种多门限判决方案,实现了编码复杂度和RD性能之间的折中。Zan等^[16]采用U-Net作为网络的基本框架,该方法能够有效地提取纹理信息,并加入量化参数(quantization parameter, QP)来调节对CU划分的影响,采用后处理算法来平衡复杂度和编码性能。Xu等^[17]提出一种基于CNN的帧内编码快速CU划分算法,利用纹理信息来预测划分模式并提前终止冗余划分过程来加速CU划分,针对不同的CU尺寸设计了相应的分类器,以提高预测精度,该算法还将错误分类导致的性能损失引入损失函数,以确保在加速编码的同时保持较低的编码失真。HoangVan等^[18]提出了一种新的基于深度学习的快速QTMTT划分方法,使用深度学习CNN模型来预测CU映射,再将其输入VVC编码器以提前终止块分割过程。Feng等^[19]使用CNN模拟CU递归划分的搜索过程,用于预测最优的CU划分图,该划分图包括QT深度图、MTT深度图和方向图,并采用后处理方法调整预测的划分图,使其符合VVC编码标准,从而实现快速编码。Chen等^[20]提出了一种快速帧内编码的框架,该框架能在编码性能和计算复杂度之间进行权衡,同时设



计了一种能有效提取图像纹理特征的CNN模块,该框架还结合了帧内模式决策,能在保持较低编码失真的同时,大幅降低编码复杂度。

目前许多基于CNN模型预测的方法针对大部分CU深度作划分。由于VVC的CU划分类型较多,随着编码深度的增加,需要预测的CU块的数量激增,这大大增加了CNN模型的复杂度并且会降低CU划分预测的准确性。此外,在VVC帧内编码中,128×128块直接划分为64×64块,而视频分辨率的多样性又造成一帧图像中往往无法包含整数个64×64块,那些非64×64块如果也用CNN预测划分,其准确度必然不高,应该单独采用其他方法来作快速划分。基于此,本文提出联合CNN预测CU划分和提前预测MTT划分方向的方法来实现快速编码。受文献[19]的启发,本文构造了一个轻量级的CNN,对所有QT和部分MTT划分进行预测得到最优划分,然后对其他的MTT则提前预测MTT的划分方向,从而取得了相比于现有的方法更高的时间效率。

1 VVC帧内编码的CU划分的复杂度分析

上一代编码标准HEVC中,帧内划分的过程只有QT划分,新一代标准中则引入了更加灵活的BT和TT划分,两者合称为MTT划分,MTT划分的引入更加符合图像的纹理特征,因此能够有效地提高编码效率。在HEVC中,允许的最大CTU大小为64×64,并且只存在QT划分,因此,CTU可以逐级向下划分为32×32、16×16以及8×8大小的块,这些即可能存在的划分大小。为满足目前高分辨率4K、8K等视频的编码需求,VVC中支持比HEVC更大的CTU。VVC中允许的CTU大小从HEVC中的64×64变为128×128,并引入了水平二叉树(horizontal binary tree, BT_H)、垂直二叉树(vertical binary tree, BT_V)、水平三叉树(horizontal ternary tree,

TT_H)和垂直三叉树(vertical ternary tree, TT_V)这4种新的CU结构,从而出现了更多不同尺寸的CU,VVC帧内编码的CU划分类型见表1。在进行最佳CU划分判断的递归过程中,采用了穷举搜索的方法,需要不断地计算每种划分的率失真代价,因此产生了大量的计算过程,使得编码的复杂性大大增加。

表1 VVC帧内编码的CU划分类型

高/ px	宽/px					
	128	64	32	16	8	4
128	QT	—	—	—	—	—
64	—	QT	—	—	—	—
32	—	—	QT、BT、TT	BT、TT	BT	BT_H
16	—	—	BT、TT	QT、BT、TT	TT_H	TT_H
8	—	—	BT、TT_V	BT、TT_V	BT	BT_H
4	—	—	BT_V、TT_V	BT_V、TT_V	BT_V	—

为了更加直观地体现VVC中各种划分所带来的时间开销,随机选取了BQTerrace、BQMall、BasketballPass、FourPeople这4个不同分辨率的视频序列进行验证并分别在量化参数QP22与QP37下作编码测试,统计了QT、BT、TT的几种划分占据总编码时间的比例。QP22和QP37下各划分类型占总编码时间的比例分别如图1和图2所示。由图1和图2可知,无论是在QP22还是QP37的情况下,帧内划分所需要的时间都占据了总编码时间的95%以上,此外,对于VVC中引入的MTT划分占据了总编码时间的90%左右,因此,有必要对快速CU划分判决的方法进行研究以降低帧内编码的复杂度。

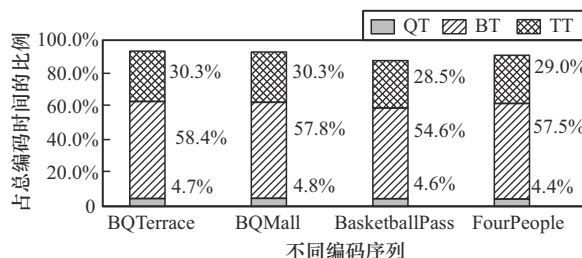


图1 QP22下各划分类型占总编码时间的比例

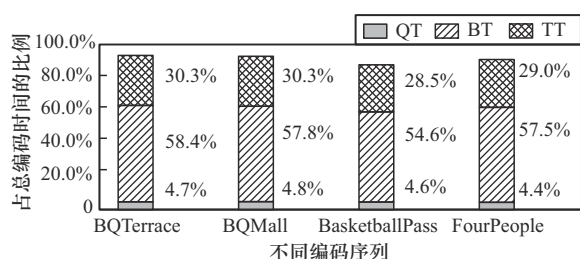


图2 QP37下各划分类型占总编码时间的比例

2 VVC快速帧内编码方法

2.1 基于CNN预测CU划分的快速编码方法

基于CNN直接预测CU的划分类型,可有效避免通过RDO递归寻找最优划分的过程,从而降低了编码时间。在VVC帧内编码的CU划分中,QT划分是相对简单的,仅涉及深度而不涉及方向,且QT划分后的子块完全相等,便于打标签作数据训练。相比之下,MTT划分不仅涉及不同深度的划分,还包含不同方向的划分,且MTT划分得到的子块大小不同,这给数据训练带来了困难,同时,对不同深度和方向的预测增加了神经网络预测的难度。如果完全采用CNN预测所有划分,将导致预测准确率降低,造成较高的比特率损失。此外,在JVET通用测试条件(common test condition, CTC)中,指定了在全Intra模式下,MTT划分的最大深度为3,这意味着一个块的MTT划分次数不超过3次。因此,为了降低CNN复杂度,仅使用CNN模型预测部分CU的划分,采用对QT全部划分作预测,并只对CU的第一次MTT划分作预测。

为了降低CNN复杂度,本文采用了轻量级的

CNN,网络结构如图3所示,CNN输入端为 64×64 的图像块,再加上左侧和顶部各4行的参考像素组成 68×68 大小的输入块。网络的输出涵盖了QT划分深度、MTT划分深度以及划分方向。该CNN结构主要由QT-Net中的多层池化模块、纹理特征模块和残差模块3个部分组成,具体如图4所示。

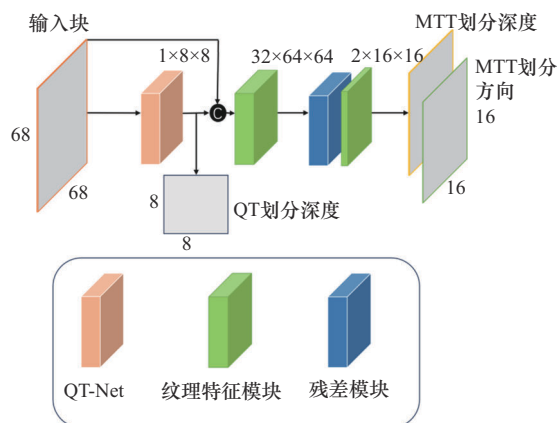


图3 本文采用的轻量级CNN结构

- 多层池化模块:在不同层次对特征图进行最大池化和插值,可以获得多个尺寸的特征图,这有助于模型在不同尺度上捕获图像中的信息,提高了模型对图像中不同大小物体的表示能力。最大池化的操作有效减少了数据量,节省了后续计算所需要的资源,有助于降低模型的计算负担,从而提高计算效率。通过不同的最大池化和插值操作,模型能够学习到不同层次的空间上下文信息。拼接不同尺寸的特征图,增加了特征的丰富性,使模型更好地适应不同尺度和空间



图4 CNN结构的模块组成



结构的特征。将由4个不同大小的最大值池化操作得到的特征图进行拼接，并送入下一层网络，如下所示。

$$x = \text{Concat}(x_1, x_2, x_3, x_4) \quad (1)$$

其中， x_1 、 x_2 、 x_3 、 x_4 分别表示经过不同大小的池化操作之后的输出，Concat表示将 x_1 、 x_2 、 x_3 、 x_4 在指定的维度上面拼接起来。

- 纹理特征模块：由3个不同形状堆叠的卷积模块构成，采用了3个不同卷积核大小的卷积操作^[20]。这些差异化的卷积核操作能够捕获不同CU大小的特征，提取这些不同的特征并将它们拼接在一起，有助于从数据中更准确地提取纹理特征，这种设计能够更有效地处理图像中的复杂特征。
- 残差模块：单个残差模块结构如图5所示，残差模块的引入旨在增强模型的学习能力。随着网络深度的增加，模型的复杂度上升，训练变得更加困难，同时可能出现梯度消失、梯度爆炸以及深层网络性能退化等问题。实践证明，增加网络层数并不能一直提升模型的精度，这是因为模型的学习能力会达到饱和点。残差模块的引入有效地解决了这些问题，加速了网络的训练过程。对于单个残差模块来说，它的输出由原始输入和经过相应卷积操作的和计算得到，如式(2)所示。随后，通过多次残差模块的堆叠，紧接着连接一个卷积模块，可以调整网络的输出形式，以确保网络输出与所需要的标签形式保持一致。

$$H(x) = x + F(x) \quad (2)$$

其中， x 表示残差块的输入， $F(x)$ 表示经过卷积之后的输出， $H(x)$ 表示残差块的输入和经过卷积操作之后的输出之和。

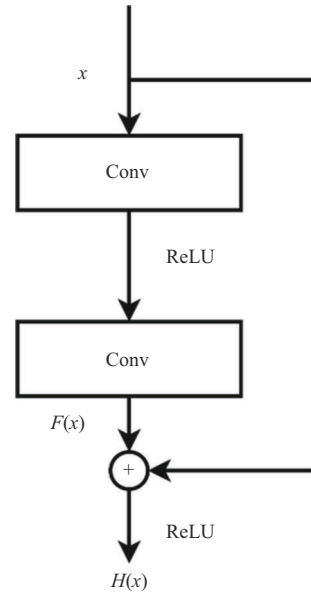


图5 单个残差模块结构

为了评估模型的学习程度，采用L2范数作为损失函数，L2范数是平方损失的总和，其具有良好的平滑性，使得优化过程更加稳定，并且对噪声具有很好的鲁棒性。损失函数表达式如式(3)和式(4)所示。

$$\text{LossQ} = \sum_{i=1}^N \sum_{j=1}^N (q_{ij} - \bar{q}_{ij})^2 \quad (3)$$

$$\text{LossFM} =$$

$$\sum_{i=1}^M \sum_{j=1}^M (\text{Dep}_{ij} - \overline{\text{Dep}}_{ij})^2 + \sum_{i=1}^M \sum_{j=1}^M (\text{Dir}_{ij} - \overline{\text{Dir}}_{ij})^2 \quad (4)$$

其中，LossQ为预测QT划分的损失值，LossFM预测第一次MTT划分的损失值， q_{ij} 、 Dep_{ij} 、 Dir_{ij} 分别表示模型输入的真实值， $\bar{q}_{ij}$ 、 $\overline{\text{Dep}}_{ij}$ 、 $\overline{\text{Dir}}_{ij}$ 分别表示模型的预测值， N 为8， M 为16。

2.2 针对MTT划分的快速编码方法

上述CNN的输入以大小为 64×64 的块作为一个基本单元，并通过网络进行预测以得到相应的输出，对于不满足 64×64 大小的CU块则不能用CNN进行预测。不能用CNN预测CU划分的区域示意图如图6所示，图6的右边方框和底部方框

中存在很大一部分不是 64×64 大小的区域,不能用CNN预测其CU划分。前文中还提到,仅有部分MTT用了CNN来预测划分,因此还需要对剩下的MTT用其他快速方法进行划分。

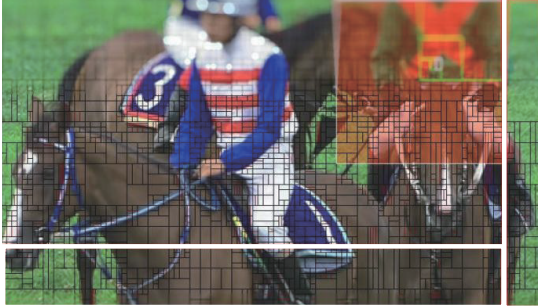


图6 不能用CNN预测CU划分的区域示意图

针对这两部分的MTT,本文采用提前预测MTT划分方向的方法来实现快速编码。在VVC帧内编码中,TT划分与BT划分之间存在显著的相关性。当BT划分为水平时,TT划分更有可能是水平方向,反之亦然。这意味着TT划分的方向更倾向于与BT划分的方向一致。因此,利用BT和TT划分方向之间的关系,使用BT的划分方向来决定是否提前终止TT的划分方向。在QTMTT中,CU首先进行水平二叉树划分,接着进行垂直二叉树划分,然后尝试进行水平三叉树划分,最后进行垂直三叉树划分。最佳的BT划分方向能被提前获取,利用这一条件,可通过式(5)判断TT划分的方向:

$$TT_{mode} = \begin{cases} TT_H, J(B_H) < J(B_V) \\ TT_V, \text{其他} \end{cases} \quad (5)$$

其中, B_H 表示水平二叉树划分, B_V 表示垂直二叉树划分, $J(B_H)$ 表示BT水平划分的率失真代价, $J(B_V)$ 表示BT垂直划分的率失真代价。当 $J(B_H)$ 小于 $J(B_V)$ 时,根据式(6)可知,更优的TT划分方向是代价值较小的水平划分TT_H,反之则为垂直划分。

$$J = D + \lambda R \quad (6)$$

其中, J 为率失真代价, D 为原始块和预测块之

间计算的失真, R 为特定模式的比特开销, λ 为拉格朗日乘子。

2.3 整体算法流程

本文方法流程如图7所示,在开始CTU编码后,由两部分算法组成,算法A即本文第2.1节提出的快速编码方法,使用CNN预测全部QT和部分MTT的划分,其余MTT则采用算法B来预测划分方向。算法B即本文第2.2节提出的快速编码方法,首先在RDO过程中得到最优的BT划分方向的率失真代价,如果BT_H是最优的BT划分,且当前正要进行TT_H划分,那么直接跳过TT_V划分的计算,仅计算TT_H划分的率失真代价。同理,如果BT_V是最优的BT划分,且当前正要进行TT_H划分,那么直接跳过这次计算,仅计算TT_V划分的率失真代价。

3 实验结果

3.1 数据集的建立

实验中使用了不同的分辨率图像来建立训练需要的数据集,采用了来自CPIV^[15]数据集的视频和图像,使用的视频分辨率包括 3840×2160 、 2048×1080 、 1920×1080 、 1280×720 。在准备这些视频作为训练集前,为了防止训练过程中模型有太多相同的帧,对视频序列每隔8帧编码一帧,使用FF_MPEG软件将图像裁剪成不同的尺寸作为数据源,同样使用FF_MPEG软件将抽取的图像合成YUV视频序列,这样就得到了不同内容和分辨率的视频序列,再将这些视频序列经过VTM10.0,用4个不同量化参数22、27、32、37进行编码,从而得到对应的标签并保存到训练集。

3.2 实验环境配置

为了证明本文方法的可行性并评价其性能,将所提出的算法在VVC标准的参考编码模型VTM10.0上实现,以VTM中规定的通用测试条件CTC作为测试标准^[21],采用的配置文件为en-

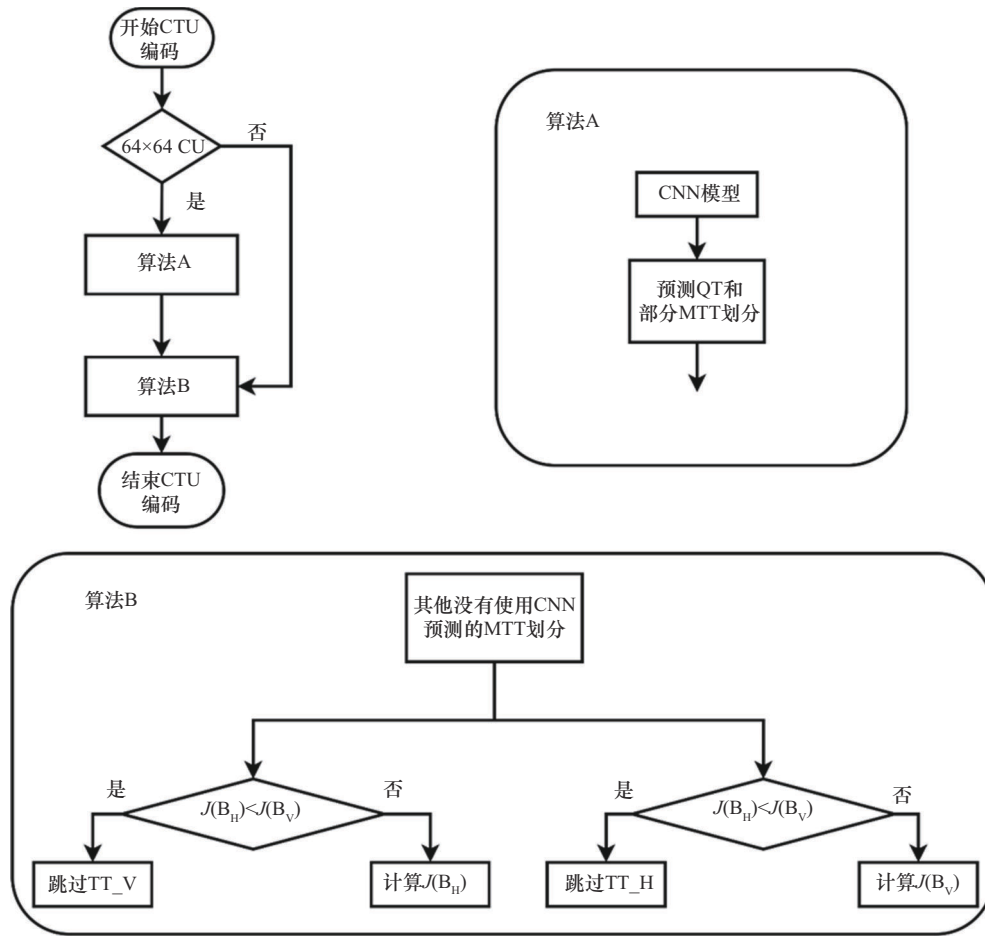


图7 本文方法流程

coder_intra_vtm.cfg, 分别在视频分辨率 3 840×2 160、3 840×2 160、1 920×1 080、832×480、416×240 和 1 280×720 下, 使用 A1、A2、B、C、D 和 E 类视频序列进行实验评估。在全帧内 (all intra, AI) 配置下编码, 其中 4 个量化参数值为 22、27、32 和 37, 实验环境是在一台配有 Intel i5 型号的中央处理单元 (central processing unit, CPU), 内存大小为 32 GB 并配有一张 RTX3060 显卡的 Windows 11 主机上进行测试, 模型推理使用的是 PyTorch, 且在 VTM 编码的过程中不需要使用显卡来进行加速。用来衡量方法性能的指标是 BDBR (Bjontegaard delta bit rate)^[22] 和编码时间节省的百分比。BDBR 越大则表示编码的压缩效率损失越大, 编码时间节省的百分比越大则表示编码速度越快。编码时间节省的百分比是 4 个

QP 下节省的编码时间的平均值, 表达式如下。

$$TS = \frac{1}{4} \sum_{QP_i} \frac{T_{org(i)} - T_{prop(i)}}{T_{org(i)}} \times 100\% \quad (7)$$

其中, TS 为编码时间节省的百分比, $T_{org(i)}$ 表示在一个 QP 下的原始编码时间, $T_{prop(i)}$ 表示本文方法的编码时间。

3.3 结果

基于分析 CNN 预测 CU 划分的快速编码方法性能见表 2, 相比于原始编码器, 该方法平均节省了 68.5% 的时间, 却只有 2.8%BDBR 的损失。针对 MTT 划分的快速编码方法性能见表 3, 该方法平均节省了 28.7% 的编码时间, 却只有 0.8%BDBR 的损失, 压缩损失很小。本文方法与其他方法的性能比较见表 4, 本文提出的方法能够平均节省 74.3% 的编码时间, 而有 3.3%BDBR

表2 基于CNN预测CU划分的快速编码方法性能

类	视频序列	BDBR	平均 BDBR	TS	平均 TS
A1	Tango2	4.5%	3.6%	63.3%	67.0%
	FoodMarket4	3.3%		66.6%	
	Campfire	3.0%		71.2%	
A2	CatRobot1	3.9%	3.0%	72.9%	73.6%
	DaylightRoad2	2.8%		72.2%	
	ParkRunning3	2.3%		75.7%	
B	CaCtus	2.7%	2.6%	73.9%	71.7%
	BasketballDrive	3.0%		71.0%	
	BQTerrace	2.3%		71.3%	
	MarketPlace	2.1%		71.4%	
	RitualDance	2.8%		70.9%	
C	BasketballDrill	4.3%	2.7%	69.7%	71.5%
	BQMall	2.7%		72.5%	
	PartyScene	1.6%		71.2%	
	RaceHorsesC	2.1%		72.5%	
D	BasketballPass	2.5%	1.8%	57.6%	55.3%
	BQSquare	1.4%		53.3%	
	BlowingBubbles	1.3%		52.9%	
	RaceHorses	2.1%		57.4%	
E	FourPeople	2.9%	3.4%	74.5%	73.2%
	Johnny	4.1%		71.9%	
	KristenAndSara	3.3%		73.2%	
	平均值	2.8%	2.8%	68.5%	68.5%

的损失。文献[15]的方法平均节省了63.8%的编码时间，损失了3.2%的BDBR。虽然本文方法在BDBR上的损失比文献[15]增加了0.1%，但是节省的平均编码时间高出了10.5%。文献[12]方法的平均BDBR增加了2.4%，节省的平均编码时间为61.4%。与文献[12]相比，虽然本文方法在BDBR损失上比其高出0.9%，但是节省的平均编码时间比其多了12.9%。此外，根据文献[19]提供的分类视频编码结果，本文方法与文献[19]方法的性能比较见表5。不包含D类视频序列的比较结果，本文方法与文献[19]相比，平均BDBR上升了0.8%，但是平均时间效率提高了2.3%。文献[19]对于D类视频序列是作单独比较的，与文献[19]相比，本文方法的时间效率提高了

8.8%，却只有0.6%的BDBR上升。由此可见，本文提出的方法能够有效地加快VVC的帧内编码并控制压缩效率的损失。

表3 针对MTT划分的快速编码方法性能

类	视频序列	BDBR	平均 BDBR	TS	平均 TS
A1	Tango2	0.4%	0.5%	24.2%	26.0%
	FoodMarket4	0.5%		25.3%	
	Campfire	0.7%		28.4%	
A2	CatRobot1	0.9%	0.8%	28.4%	29.6%
	DaylightRoad2	0.9%		28.0%	
	ParkRunning3	0.5%		32.4%	
	CaCtus	0.9%	0.8%	30.9%	29.0%
B	BasketballDrive	1.0%		27.9%	
	BQTerrace	0.8%		27.9%	
	MarketPlace	0.5%		28.1%	
	RitualDance	0.9%		30.3%	
	BasketballDrill	1.3%	0.9%	29.6%	29.9%
C	BQMall	1.1%		31.7%	
	PartyScene	0.6%		30.8%	
	RaceHorsesC	0.6%		27.5%	
	BasketballPass	1.0%	0.7%	17.8%	27.9%
D	BQSquare	0.5%		33.7%	
	BlowingBubbles	0.7%		28.7%	
	RaceHorses	0.6%		31.2%	
	FourPeople	1.2%	1.1%	30.4%	29.4%
E	Johnny	1.2%		28.6%	
	KristenAndSara	1.0%		29.3%	
	平均值	0.8%	0.8%	28.7%	28.7%

4 结束语

针对VVC帧内编码中，QTMTT划分占据了大部分的编码时间而造成编码复杂度上升的问题，本文提出了一种联合CNN预测CU划分和提前预测MTT划分方向的VVC快速帧内编码方法，能够极大地减少RDO遍历过程，有效地降低编码复杂度。实验结果表明，相对于原始VVC编码器，本文提出的方法能够平均节省74.3%的编码时间，性能优于对比的方法。



表4 本文方法与其他方法的性能比较

类	视频序列	本文方法		文献[15]方法		文献[12]方法	
		BDBR	TS	BDBR	TS	BDBR	TS
A1	Tango2	4.7%	67.3%	3.5%	54.4%	1.6%	62.0%
	FoodMarket4	3.6%	70.9%	2.8%	57.9%	1.5%	56.8%
	Campfire	3.5%	76.5%	4.2%	66.5%	2.0%	59.6%
A2	CatRobot1	4.5%	77.2%	4.9%	62.9%	2.1%	57.3%
	DaylightRoad2	3.3%	77.3%	2.8%	65.6%	2.5%	66.8%
	ParkRunning3	2.7%	80.1%	2.7%	63.0%	1.0%	52.1%
B	CaCtus	3.3%	78.7%	2.9%	67.7%	2.6%	64.8%
	BasketballDrive	3.7%	76.3%	3.9%	68.5%	2.6%	66.9%
	BQTerrace	2.8%	76.3%	2.6%	64.4%	2.4%	62.9%
	MarketPlace	2.4%	77.7%	1.9%	65.2%	1.4%	67.7%
	RitualDance	3.5%	80.7%	2.7%	63.0%	2.5%	65.4%
C	BasketballDrill	5.4%	74.8%	4.7%	61.0%	4.4%	56.3%
	BQMall	3.4%	77.2%	3.1%	67.5%	3.3%	63.1%
	PartyScene	2.1%	75.9%	1.9%	64.6%	2.1%	61.2%
	RaceHorsesC	2.5%	77.6%	2.5%	64.6%	2.0%	61.0%
D	BasketballPass	3.2%	66.0%	3.7%	62.6%	3.2%	59.9%
	BQSquare	1.7%	63.1%	2.0%	62.5%	1.6%	56.4%
	BlowingBubbles	1.8%	63.6%	2.4%	61.9%	2.3%	57.2%
	RaceHorses	2.6%	67.0%	2.9%	61.3%	2.1%	56.9%
E	FourPeople	3.7%	78.5%	3.3%	66.9%	4.0%	66.5%
	Johnny	5.0%	76.0%	5.1%	64.4%	3.5%	66.7%
	KristenAndSara	3.9%	76.8%	3.9%	66.1%	2.9%	63.1%
	平均值	3.3%	74.3%	3.2%	63.8%	2.4%	61.4%

表5 本文方法与文献[19]方法的性能比较

类	本文方法		文献[19]方法	
	BDBR	TS	BDBR	TS
A1	3.9%	71.7%	3.2%	73.9%
A2	3.5%	78.2%	2.8%	76.0%
B	3.1%	78.0%	2.5%	74.0%
C	3.4%	76.4%	2.5%	71.7%
E	4.2%	77.1%	3.3%	75.5%
以上5类平均值	3.6%	76.4%	2.8%	74.1%
D	2.3%	64.9%	1.7%	56.1%

参考文献:

- [1] BROSS B, WANG Y K, YE Y, et al. Overview of the versatile video coding (VVC) standard and its applications[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3736-3764.
- [2] 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.
- [3] FAN Y B, CHEN J A, SUN H M, et al. A fast QTMT partition decision strategy for VVC intra prediction[J]. IEEE Access, 2020 (8): 107900-107911.
- [4] LI Y, YANG G B, SONG Y, et al. Early intra CU size decision for versatile video coding based on a tunable decision model[J]. IEEE Transactions on Broadcasting, 2021, 67(3): 710-720.
- [5] ZHANG Z H, FU C H, XIE K, et al. Fast VVC intra coding by skipping redundant coding block structures and unnecessary directional partition[C]//Proceedings of the 2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval (MIPR). Piscataway: IEEE Press, 2022: 84-89.
- [6] NI C T, LIN S H, CHEN P Y, et al. High efficiency intra CU

- partition and mode decision method for VVC[J]. IEEE Access, 2022(10): 77759-77771.
- [7] LI Q, MENG H, LI Y. Texture-based fast QTMT partition algorithm in VVC intra coding[J]. Signal, Image and Video Processing, 2023, 17(4): 1581-1589.
- [8] ZHAO S, SHANG X W, WANG G Z, et al. A fast algorithm for intra-frame versatile video coding based on edge features[J]. Sensors, 2023, 23(13): 6244.
- [9] SIL N, ZHU W D, ZHANG Q W. Fast adaptive CU partition decision algorithm for VVC intra coding[J]. IEEE Access, 2023(11): 119766-119778.
- [10] WU G Q, HUANG Y, ZHU C, et al. SVM based fast CU partitioning algorithm for VVC intra coding[C]//Proceedings of the 2021 IEEE International Symposium on Circuits and Systems (ISCAS). Piscataway: IEEE Press, 2021: 1-5.
- [11] HE Q, WU W X, LUO L, et al. Random forest based fast CU partition for VVC intra coding[C]//Proceedings of the 2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB). Piscataway: IEEE Press, 2021: 1-4.
- [12] SALDANHA M, SANCHEZ G, MARCON C, et al. Configurable fast block partitioning for VVC intra coding using light gradient boosting machine[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(6): 3947-3960.
- [13] ABDALLAH B, BELGHITH F, ALI BEN AYED M, et al. Fast mode decision for CU coding based on CNN for the VVC standard[J]. Journal of Electronic Imaging, 2022, 31(5): 053004.
- [14] TISSIER A, HAMIDOUCE W, VANNE J, et al. CNN oriented complexity reduction of VVC intra encoder[C]//Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2020: 3139-3143.
- [15] LI T Y, XU M, TANG R Z, et al. DeepQTMT: a deep learning approach for fast QTMT-based CU partition of intra-mode VVC[J]. IEEE Transactions on Image Processing, 2021(30): 5377-5390.
- [16] ZAN Z, HUANG L L, CHEN S S, et al. Fast QTMT partition for VVC intra coding using U-net framework[C]//Proceedings of the 2023 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2023: 600-604.
- [17] XU J, WU G Q, ZHU C, et al. CNN-based fast CU partitioning algorithm for VVC intra coding[C]//Proceedings of the 2022 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2022: 2706-2710.
- [18] HOANGVAN X, NGUYENQUANG S, DINHBAO M, et al. Fast QTMT for H.266/VVC intra prediction using Early-Terminated Hierarchical CNN model[C]//Proceedings of the 2021 International Conference on Advanced Technologies for Communications (ATC). Piscataway: IEEE Press, 2021: 195-200.
- [19] FENG A L, LIU K, LIU D, et al. Partition map prediction for fast block partitioning in VVC intra-frame coding[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2023(32): 2237-2251.
- [20] CHEN Z B, SHI J, LI W P. Learned fast HEVC intra coding[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2020(29): 5431-5446.
- [21] BOSSEN F, BOYCE J, LI X, et al. JVET common test conditions and software reference configurations for SDR video[J]. Joint Video Experts Team (JVET) of ITU-T SG, 2019(16): 19-27.
- [22] BJONTEGAARD G. Calculation of average PSNR differences between RD-curves[J]. ITU-T VCEG-M33, 2001(4): 2-4.

[作者简介]



钟辉 (1998-), 男, 杭州电子科技大学通信工程学院硕士生, 主要研究方向为视频压缩编码。



陆宇 (1977-), 男, 博士, 杭州电子科技大学通信工程学院副教授, 主要研究方向为图像压缩和视频编码。



殷海兵 (1974-), 男, 博士, 杭州电子科技大学通信工程学院教授, 主要研究方向为数字视频编解码。



黄晓峰 (1988-), 男, 博士, 杭州电子科技大学通信工程学院副教授, 主要研究方向为图像压缩和视频编码。