



研究与开发

一种基于图神经网络和统计分析的VVC帧内编码快速算法

黎天送¹, 刘昊坤¹, 崔少国¹, 刘姝岑¹, 陈艳¹, 王鸿奎²

(1. 重庆师范大学, 重庆 401331;

2. 杭州电子科技大学, 浙江 杭州 310018)

摘要: 多功能视频编码 (versatile video coding, VVC) 作为最新一代的视频编码标准, 通过引入多种高效的编码工具进一步提升了视频编码性能。然而, VVC 标准引入了四叉树加多类型树 (quadtree plus multi-type tree, QTMT) 划分结构, 并将帧内预测模式从 35 种扩展到 67 种, 导致编码复杂度急剧上升。为降低 VVC 的帧内编码复杂度, 首先, 提出了一种基于图神经网络的帧内编码单元 (coding unit, CU) 划分快速算法, 该算法利用高效的图神经网络模型直接预测 CU 的最优划分模式, 从而跳过冗余的 CU 划分遍历。其次, 提出了一种基于空间相关性和纹理特征的帧内模式选择快速算法, 该算法利用平均方向方差和 Sobel 梯度算子确定纹理方向, 并跳过部分角度预测模式, 同时结合预测模式间的相关性精简率失真模式列表。实验结果表明, 该算法能够在 BDBR (bjontegaard delta bit rate) 上升 2.29% 的代价下, 节省 64.04% 的编码时间。

关键词: 多功能视频编码; 帧内编码; 编码单元划分; 帧内角度模式

中图分类号: TP391

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2024213

A fast VVC intra-coding algorithm based on graph neural network and statistical analysis

LI Tiansong¹, LIU Haokun¹, CUI Shaoguo¹, LIU Shucen¹, CHEN Yan¹, WANG Hongkui²

1. Chongqing Normal University, Chongqing 401331, China

2. Hangzhou Dianzi University, Hangzhou 310018, China

Abstract: VVC as the latest generation of video coding standards, further improves video compression quality by introducing a variety of efficient coding tools. However, the VVC standard introduces the QTMT division structure and expands the intra prediction modes from 35 to 67, resulting in a sharp increase in coding complexity. Firstly, a fast algorithm for intra-frame coding unit (CU) division based on graph neural network was proposed, in order to reduce the complexity of intra-frame coding of VVC. An efficient graph neural network model was used to directly predict the

收稿日期: 2024-03-27; 修回日期: 2024-09-11

基金项目: 重庆市科技局自然科学基金项目 (No.CSTB2022NSCQ-MSX1231); 重庆市教委青年项目 (No.KJQN202200519); 重庆师范大学人才基金项目 (No.21XLB031)

Foundation Items: The Natural Science Foundation Project of Chongqing Science and Technology Bureau (No. CSTB2022NSCQ-MSX1231), The Science and Technology Research Program of Chongqing Municipal Education Commission(No. KJQN202200519), The Talents Fund Project of Chongqing Normal University (No.21XLB031)



optimal partition mode of CU, thus skipping redundant CU partition traversal. Secondly, a fast algorithm for intra-frame mode selection based on spatial correlation and texture features was proposed. The average direction variance and Sobel gradient operator were used to determine the texture direction, some angle prediction modes were skipped, and the correlation between prediction modes to streamline the rate-distortion mode list were combined. Experimental results show that this algorithm can save 64.04% of encoding time at the cost of increasing BDBR by 2.29%.

Key words: VVC, intra coding, coding unit partition, intra-angel mode

0 引言

随着通信技术的突破和视频应用的爆发式增长,视频已成为数字时代主要的信息传输媒体。新兴视频技术,如虚拟现实(virtual reality, VR)、增强现实(augmented reality, AR)、4K和8K分辨率、高动态范围(high dynamic range, HDR)、广色域和高帧率视频等,在为人们带来沉浸式体验的同时,也使得视频数据量呈指数级增加,传输视频所需带宽开销激增。为满足当今新一代视频编码的需求,联合视频探索专家组(Joint Video Exploration Team, JVET)制定了新一代视频编码标准H.266/VVC^[1]。VVC为提升视频压缩性能,引入了众多先进编码工具,如四叉树加多类型树的划分结构^[2]、跨分量线性模型^[3]、基于位置信息的帧内联合预测^[4]、多参考行预测^[5]、帧内子块预测^[6]和67种帧内预测模式等。这些新兴编码技术的应用,使VVC能够在主观视觉质量保持不变的情况下提升50%以上的编码效率^[7],然而编码时间相对也增加了18倍以上^[8],这种极高的编码复杂度为VVC标准的实时编码应用带来巨大困难。因此,在保证编码性能的同时,如何显著降低VVC的编码复杂度成为当前的研究热点。

目前,VVC快速算法可大致分为两类:基于统计的方法和基于学习的方法。在基于统计的方法中,Li等^[9]将编码单元(coding unit, CU)划分的早期跳过与否视作二分类问题,通过为每个CU的BT和TT分区建立决策模型,以判断当前模式是否被跳过。Dong等^[10]提出了一种包含自适应模式剪枝和与模式相关终止的快速帧内决策

算法,前者用于跳过被预测为不可能的预测模式,后者用于消除基于所选最佳模式深度对剩余深度的不必要预测。Tang等^[11]提出了一种用于帧内编码和帧间编码的快速块划分算法。在帧内编码中,该方法采用基于块级的Canny边缘检测器来提取边缘特征,以便跳过垂直或水平划分模式。在帧间预测中,通过三帧差法判断当前块是否为运动对象,从而提前结束划分流程。陈燕辉等^[12]依据纹理复杂度判别CU是否不进行划分,并通过CU的中间区域与边缘区域复杂度的不同精简划分模式列表。

在基于学习的方法中,Wu等^[13]提出了一种基于支持向量机(SVM)的帧内编码快速CU划分算法。该方法为每个尺寸的CU训练不同的分类器,以便提前跳过冗余的划分模式。Liu等^[14]采用了提取与CU划分相关的视频帧特征的方法来表示CU复杂度,随后使用SVM作为分类模型,用于区分不同CU的复杂度。该方法通过提前结束划分流程,实现加速编码的目的。Park等^[15]通过开发与TT相关的显示VVC特征和派生VVC特征,运用轻量级神经网络模型根据这些特征来判断,以决定是否终止四叉树划分后的多类型树划分。Zhang等^[16]提出了一种基于全局卷积网络(global convolutional network, GCN)的快速CU划分方法。该方法通过预测最优划分模式,以跳过低概率的划分模式,从而降低VVC的计算复杂度。Wu等^[17]用层次网络图来表示CU的QTMT划分结构,并提出层次网格全卷积网络,用于预测当前CU及其子CU的划分信息。程栋梁等^[18]提出了一种基于DenseNet+FPN的CU快速

划分算法, 利用多尺度信息和自适应质量复杂度均衡损失函数实现CU划分结构的精准预测。Zhang等^[19]利用卷积神经网络预测每个 64×64 块中 4×4 块的边缘为划分边界的概率, 从而直接得到当前CU最优划分结构, 以实现降低编码复杂度的目的。

虽然上述算法在一定程度上能够降低VVC的编码复杂度, 但仍存在一些问题。基于统计的方法严重依赖人为手动设计的特征, 难以挖掘CU的深层次特征。在基于学习的方法中, 机器学习方法预测能力有限, 难以实现优异的编码性能优化。而深度学习方法往往基于卷积神经网络(convolutional neural network, CNN), 其局部特征提取的特性使得对CU纹理方向和纹理复杂度等全局信息的提取能力欠佳。此外, 编码单元划分过程由于具有较高的计算冗余, 机器学习和深度学习方法的强大预测能力能够有效降低划分过程的编码复杂度。而模式选择过程与块划分过程相比编码时间占比更少, 基于学习的方法由于推

理时间长、参数量庞大, 其额外的计算开销导致并不能高效降低模式选择过程的编码复杂度。鉴于上述问题, 本文提出了一种基于图神经网络和统计分析的VVC帧内编码快速算法, 包括基于图神经网络的帧内CU划分快速算法和基于空间相关性和纹理特征的帧内模式选择快速算法。该算法利用CU的亮度分量和量化参数(quantization parameter, QP)跳过非必要的划分模式, 并根据模式间相关性和纹理方向进行帧内预测模式的快速决策。实验结果表明, 本文算法在可忽略的编码性能损失下, 有效降低编码复杂度。

1 基于图神经网络和统计分析的VVC帧内编码快速算法

本节首先对VVC帧内编码的复杂度进行了详细分析。然后, 介绍了基于图神经网络的CU划分快速算法。最后, 介绍了基于空间相关性和纹理特征的模式选择快速算法。算法流程如图1所示, 后文将对具体细节进行详细介绍。

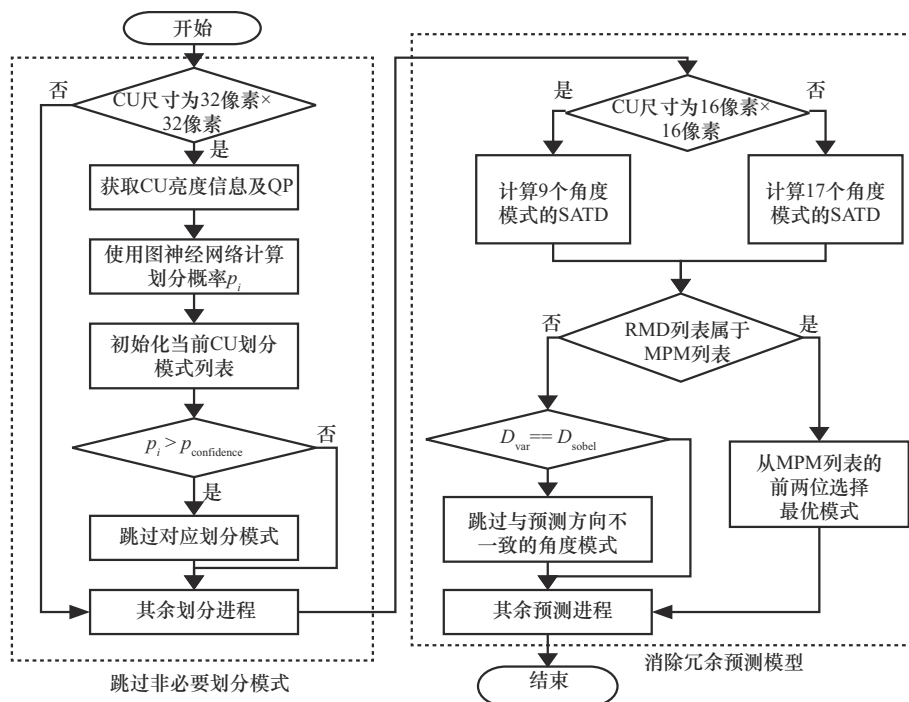


图1 算法流程

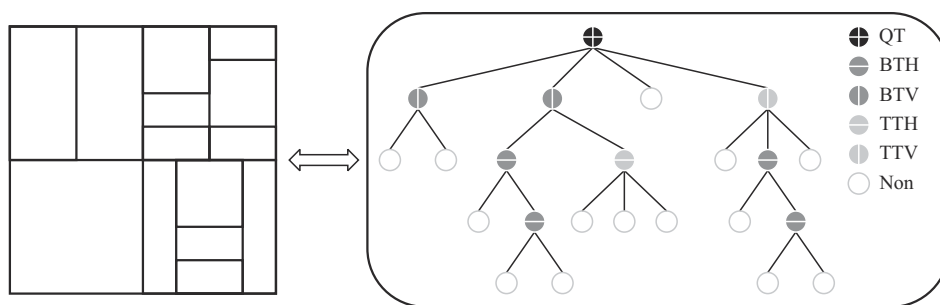


图2 QTMT划分结构及其对应的多叉树表示

1.1 复杂度分析

在CU块划分流程中，每个CU首先进行四叉树（QT）划分，然后在QT划分的叶节点尝试进行多叉树（multi-type tree, MTT）递归划分，QTMT划分结构及其对应的多叉树表示如图2所示。其中QTMT划分结构主要包含四叉树划分（QT）、二叉树划分（binary tree, BT）和二叉树划分（ternary tree, TT）。BT和TT划分可被进一步分为垂直二叉树划分（SPLIT_VT_VER, BTV）、水平二叉树划分（SPLIT_BT_HOR, BTH）、垂直二叉树划分（SPLIT_TT_VER, TTV）和水平二叉树划分（SPLIT_TT_HOR, TTH）。除了前述提及的划分模式外，CU还能够不进行划分。这种灵活而复杂的划分方式能够更好地贴合图像纹理，但却带来了巨大的计算冗余，导致编码复杂度显著增加。

在模式选择过程中，VVC帧内编码通过遍历67种预测模式，选择率失真代价最低的模式作为帧内预测的最优模式。这个过程包括粗选过程（rough mode decision, RMD）、最优可能列表（most probable mode, MPM）获取过程和率失真优化（rate distortion optimization, RDO）过程。然而，由于采用暴力搜索来确定最优模式，该过程具有较高的计算复杂度。为了分析模式选择中的编码复杂度，分别选择B、C、D、E类中的一个序列，并统计其在4个QP下第一轮RMD过程、第二轮RMD过程和RDO过程的平均编码时间比例，帧内编码模式选择过程中

各部分时间占比如图3所示。可以观察到，RDO过程占据了90%以上的编码时间，而RMD过程所占时间相对较少，均在10%左右。MPM过程仅占编码时间的1%~2%，几乎可以忽略不计。根据上述统计数据，可以得出模式选择过程的编码复杂度主要集中在第一轮RMD和RDO过程。因此，关键问题在于减少第一轮RMD过程和RDO过程的计算冗余，以降低整体编码复杂度。

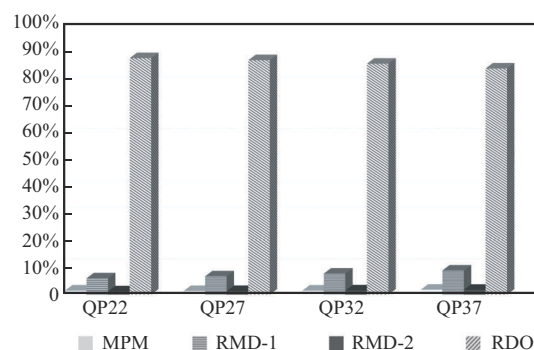


图3 帧内编码模式选择过程中各部分时间占比

在确定最优预测模式的RMD过程和MPM过程中，VVC使用残差的绝对变换差之和（sum of absolute transformed differences, SATD）计算失真，以选择最佳模式。在最优划分的递归过程和RDO过程中，VVC通过比较率失真代价（rate-distortion cost, RDcost）来得到最优模式MBest，SATD和RDcost的计算式为：

$$D_{SATD} = \sum_N \sum_N |HXX| \quad (1)$$

$$RDcost = D + \lambda \cdot R \quad (2)$$

其中, N 表示编码块尺寸, X 表示残差信号, H 表示归一化后的Hadamard变换矩阵, R 和 D 分别表示编码所消耗的比特数和失真程度, 为控制率失真性能的拉格朗日乘子。

1.2 基于图神经网络的块划分快速算法

1.2.1 划分过程建模与数据集建立

在VVC的CU划分流程中, 如果能直接预测编码单元的划分结构, 从而跳过非必要划分模式的率失真代价计算过程, 即通过纹理特性直接确定当前CU划分结构, 便能够有效降低编码复杂度。为此, 本文算法将最优CU划分结构预测视作图像分类任务, 将QTMT中的6种划分结构作为6个类别, 以获取当前CU的6种划分结构的预测概率。将高概率的划分模式直接视作最优划分模式, 减少块划分过程中需要遍历的划分模式数量, 达到减少计算冗余的目的。

为了训练出能够精准预测CU划分结构的深度学习模型, 构建一个充足且高质量的数据集是非常必要的, 因此, 本文建立了一个针对VVC视频编码CU划分的数据集。由于使用同一视频序列制作的训练数据样本基本一致, 不能使模型得到充足训练, 同时考虑帧内编码与图像压缩任务的相似性, 可以使用自然图像来进一步扩充训练数据, 因此, 本文选择RAISE^[20]作为数据来源。

通过将图像制作为视频序列, 并采用VVC标准参考编码器VTM-10.0, 在全帧内编码(all-intra)配置下, 对训练集中的视频序列进行编码以完成数据标注。每个训练样本包含 32×32 的CU亮度像素值、在4个QP下的最终划分结果和对应的QP值。

1.2.2 CU划分预测模型总体结构

基于图神经网络的CU最优划分结构预测模型结构如图4所示, 该模型主要由特征提取模块和预测模块组成, 其中当前CU亮度像素值和QP值作为网络输入, 而网络输出则是每种划分模式的概率。

首先, 将当前CU亮度分量块送入预测网络, 随后通过特征提取模块提取图像特征。考虑VVC帧内编码中最小编码树单元(coding tree unit, CTU)的尺寸为 4×4 以及划分过程的非重叠特性, 在初始卷积后设置了一个大小为 4×4 、步长为4的卷积层, 以适应VVC标准中CU划分的特点。

为更好地学习CU更高层次和不同大小范围的特征, 该模块后续采用两个GCN Block模块进行特征提取。GCN Block包括图卷积操作和用于特征变换的前馈神经网络两个部分。GCN Block来源于ViG^[21], 通过引入图神经网络进行计算机视觉任务, 模型能够更好地获取图像中不同区域

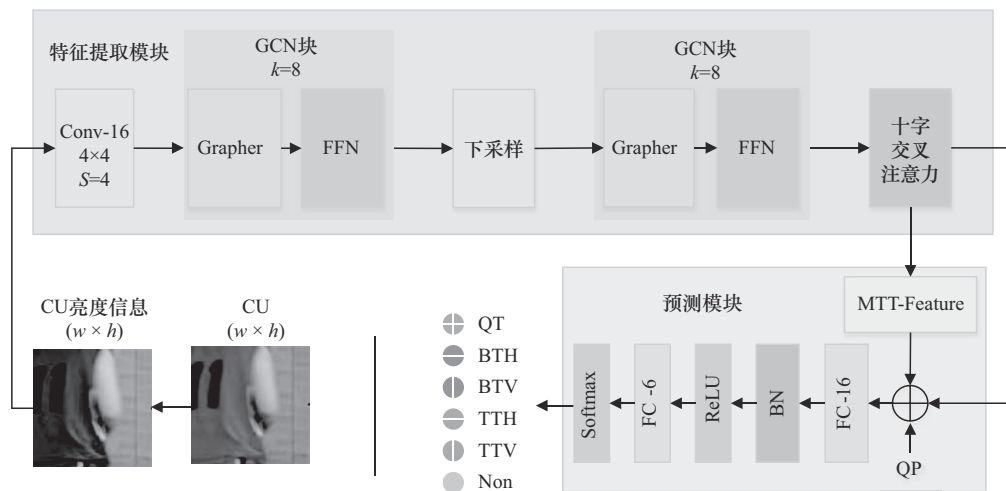


图4 基于图神经网络的CU最优划分结构预测模型结构



间的关系，从而获取全局特征。

本文在特征提取模块中首先引入了GCN Block模块，利用图卷积提取图像纹理特征。在图卷积操作的构图过程中，邻居节点数 K 是控制聚集范围的参数，邻居过少会降低信息交换的幅度，而邻居过多会导致节点间区分度下降。为找到适用于预测CU划分的最佳邻居节点数，比较了不同 K 取值下的模型预测精度。不同 K 取值下模型预测精度如图5所示。实验结果表明，将 K 设置为6时，模型能够达到最优预测性能。其次，在特征提取模块的尾部使用CCNet^[22]中的十字交叉注意力（criss-cross attention, CCA），以提升远距离像素间的关系。CCA通过采用十字交叉的方式，使得特征图中每个像素点周围呈十字形的相邻像素进行信息交互。这有助于实现水平和垂直方向上的较远点像素与当前像素点关系的建模，从而显著提升模型预测能力。

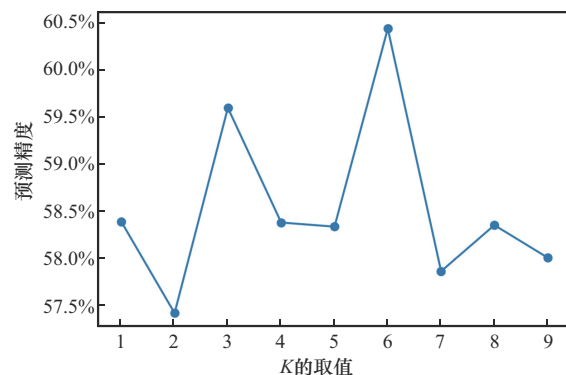


图5 不同 K 取值下模型预测精度

此外，为了提升CU划分结构预测模型的性能，设计了一个多叉树特征提取模块（MTT-Feature）。该模块通过引入与QTMT相关的先验知识来获取MTT相关特征，多叉树特征提取模块结构如图6所示。MTT特征模块设计了5个分支，每个分支都对应QTMT划分结构中的一种划分。通过不同分支得到与多叉树划分对应的特征向量，在进行拼接操作后使用一维卷积进行特征融合。最终，将提取的特征向量与特征提取模块

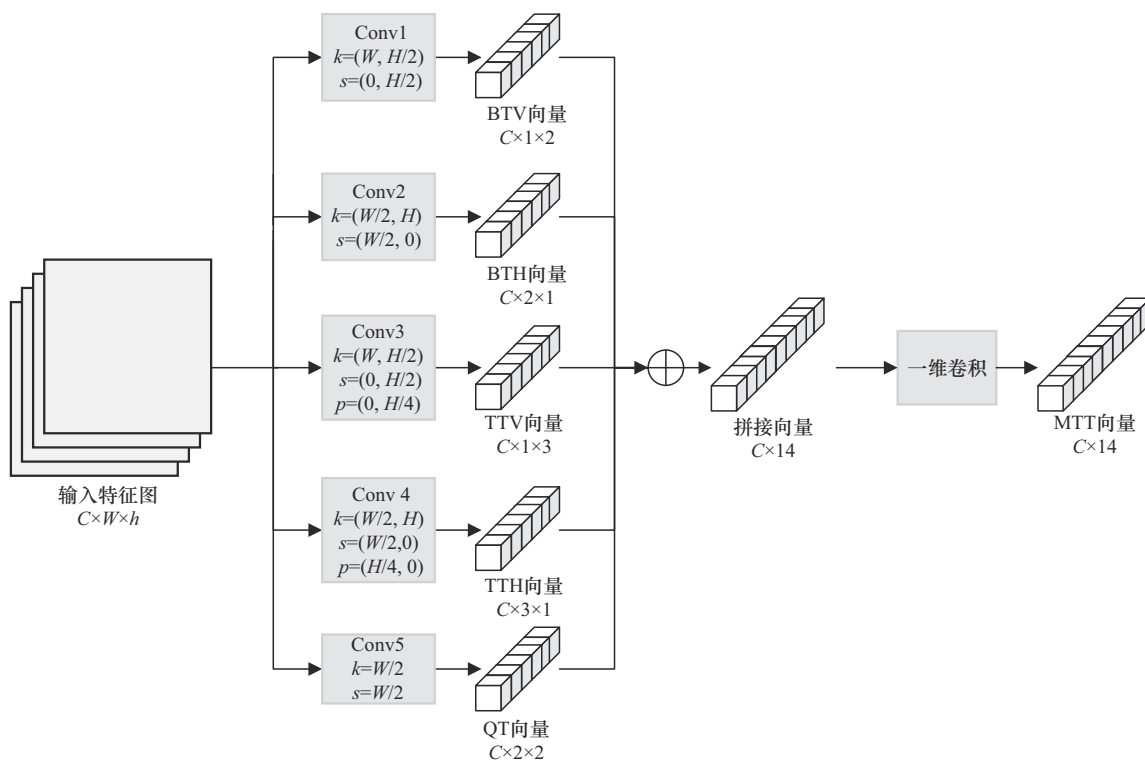


图6 多叉树特征提取模块结构

得到的特征向量一同送入预测模块。

预测模块的输入包括来自特征提取模块输出的特征图、MTT特征模块的特征向量和经过归一化的QP值。随后，将这三者进行拼接，并经过全连接层、批归一化和激活函数，得到长度为6的预测向量，表示6种QTMT划分结构的预测概率。

除预测模块最后一层使用Softmax激活函数外，其余激活函数都是用GELU激活函数，相比于常见的Sigmoid激活函数，GELU更加平滑，有助于提高训练过程的收敛速度和性能，且能够更好地解决梯度消失问题。

为计算预测损失，所提算法使用多分类交叉熵来定义损失函数 L ，其计算式如下。

$$L(y, \hat{y}) = - \sum_{i=1}^C y_i \log(\hat{y}_i) \quad (3)$$

其中， C 表示类别的数量， y_i 实际类别为第 i 类的概率， $\hat{y}_i$ 表示模型对第 i 类的预测概率。

不同视频编码场景下对编码的需求也存在差异。在医学图像和卫星遥感等场景中，通常需要获得无损或极高编码质量的压缩视频；而在日常使用的自然视频中，则更注重较高的压缩性能以满足视频传输需求。为了满足不同编码场景下的需求，提出了一种基于预测概率和划分模式数量的多阈值决策方案，以实现编码性能和编码质量之间的平衡。

具体来说，决策方案包含两个关键要素：概率置信度阈值 τ_p 和编码单元划分数量（CU partition mode, CPM）数量阈值 τ_m 。概率置信度阈值 τ_p 代表着对应的CPM被视为最优CPM的可接受预测概率阈值。只有当预测概率高于该阈值时，才被认可为最优划分模式，这有助于避免MTT-GNN输出的预测概率相近导致的错误预测。该阈值取值范围为 $[0,1]$ ，表示预测概率的可信程度。而CPM数量阈值 τ_m 则表示允许被选择的CPM数

量。通过扩充被选中的模式数量，可以在稍微增加编码时间的代价下，维持编码性能。该阈值取值范围为 $\{1,2,3,4,5,6\}$ ，表示可以选择的最大CPM数量。

通过妥善调节 τ_m 和 τ_p 两个阈值，可以实现不同场景下的编码需求，从而实现编码性能与编码复杂度之间的权衡。为此，本文提出了多阈值决策方案，并设计了“效率优先”和“质量优先”两组阈值，多阈值决策方案参数见表1。

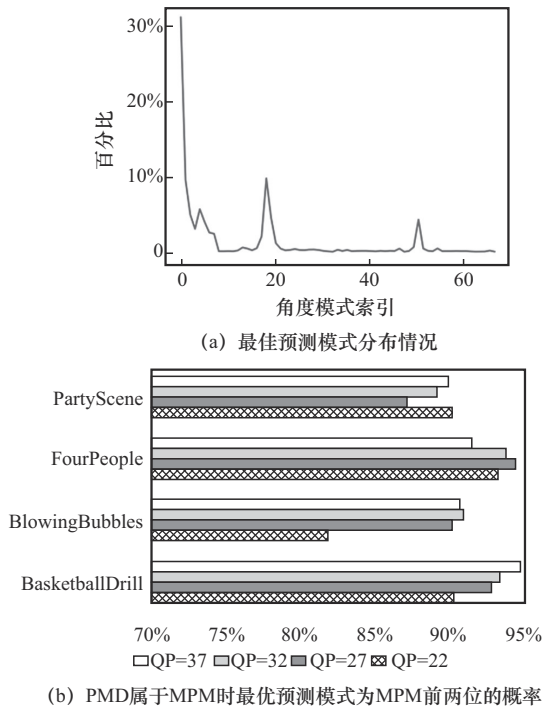
表1 多阈值决策方案参数

阈值	τ_m	τ_p
效率优先	1	1.0
质量优先	3	0.3

在“效率优先”方案中，只选择当前预测概率最高的作为最优划分结构，而不考虑预测概率的约束，以最大限度地减少编码时间。而在“质量优先”方案中，首先对预测概率进行约束，去除概率低于0.3的划分结构，然后选择概率最高的3种模式作为最优参考划分结构。此时编码器能够实现高编码质量下的编码复杂度降低。此外，针对具体应用场景，可以通过调整多阈值决策方案中的 τ_m 和 τ_p 阈值参数来满足不同编码场景的需求。

1.3 基于空间相关性和纹理特征的模式选择快速算法

为深入探索帧内模式选择的分布规律，对帧内最优预测模式进行了预实验，所使用的测试序列与图3中的序列一致。模式选择快速算法预实验如图7所示。图7（a）展示了在QP=22时，统计序列最佳预测模式的平均分布情况。从图7中可以看出，Planar模式、DC模式、垂直模式和水平模式所占比例明显大于其余预测模式。基于这一分布规律，可以在最佳预测模式分布的特性基础上进行帧内模式选择的优化，保留这4种预测模式，以在维持编码性能的同时降低编码复杂度。



此外, 模式选择流程通过依次更新RMD过程、MPM过程和RDO过程的最优模式选择候选列表, 然后从最终的候选模式列表中依据RDcost选择最优预测模式。若能在编码复杂度高的RDO过程前, 提前利用模式间相关性判别最优模式, 将有效降低编码复杂度。为此, 统计了当RMD得到的预测模式列表属于MPM得到的相邻预测模式列表时, 最优预测模式属于MPM列表的前两个预测模式的比例, 使用的序列与先前实验一致, 如图7(b)所示。可以观察到, 在RMD列表属于MPM列表时, MPM列表的前两个预测模式以90%左右的概率成为最优预测模式。因此, 可以直接比较MPM列表的前两个模式的率失真代价值, 以确定最优预测模式。

基于上述统计分析, 本文提出了基于空间相关性和纹理特征的帧内模式选择快速算法。该算法在通过优化RMD过程、RMD和MPM列表合并过程以及最终的RDO过程, 实现了编码流程的高效加速。模式选择快速算法流程

如图8所示。

在RMD过程的第一阶段中, 若当前CU尺寸小于16, 则每隔4个角度模式建立初始角度模式列表; 反之, 每隔8个角度模式建立初始角度模式列表。这一优化使得原先需遍历的角度模式数量从33缩减到17或9, 同时初始角度模式列表都包含了占比最高的垂直模式和水平模式, 与图7中的统计结果保持一致。

在RMD和MPM列表合并过程中, 若RMD得到的候选列表存在于MPM列表中, 直接选取MPM列表中前两个预测模式作为候选最优模式, 通过计算并选择其中率失真代价较低的模式作为最优预测模式, 避免较高复杂度的RDO过程, 从而降低了模式选择中的编码复杂度。

当RMD列表不属于MPM列表时, 由于预测角度模式的选择与纹理方向具有强相关性, 考虑通过垂直、水平方向的平均方差来跳过RDO过程中方向不一致的角度模式, 其计算式为:

$$\text{MeanVar}_{\text{Hor}} = \frac{1}{H} \sum_{j=1}^H \sum_{i=1}^W (I_{i,j} - \bar{I}_i)^2 \quad (4)$$

$$\text{MeanVar}_{\text{Ver}} = \frac{1}{W} \sum_{i=1}^W \sum_{j=1}^H (I_{i,j} - \bar{I}_j)^2 \quad (5)$$

其中, W 和 H 为当前CU的宽和高, $I_{i,j}$ 为 (i, j) 位置的像素值, $\bar{I}_i$ 和 $\bar{I}_j$ 为当前行或列的像素平均值。由于标准差进行纹理方向判别过于粗糙, 本文还引入Sobel梯度算子来辅助纹理方向判别, 其中水平和垂直Sobel绝对梯度值的计算表达式为:

$$D_x = \sum_{i=1}^W \sum_{j=1}^H |D_x(i,j)| \quad (6)$$

$$D_y = \sum_{i=1}^W \sum_{j=1}^H |D_y(i,j)| \quad (7)$$

其中, W 和 H 分别为CU的宽和高, D_x 和 D_y 为单个像素由Sobel算子计算得到的梯度。在得到由平均方差和Sobel梯度两种方法判别的纹理方向后, 仅当二者判断方向一致时, 才会依据CU纹

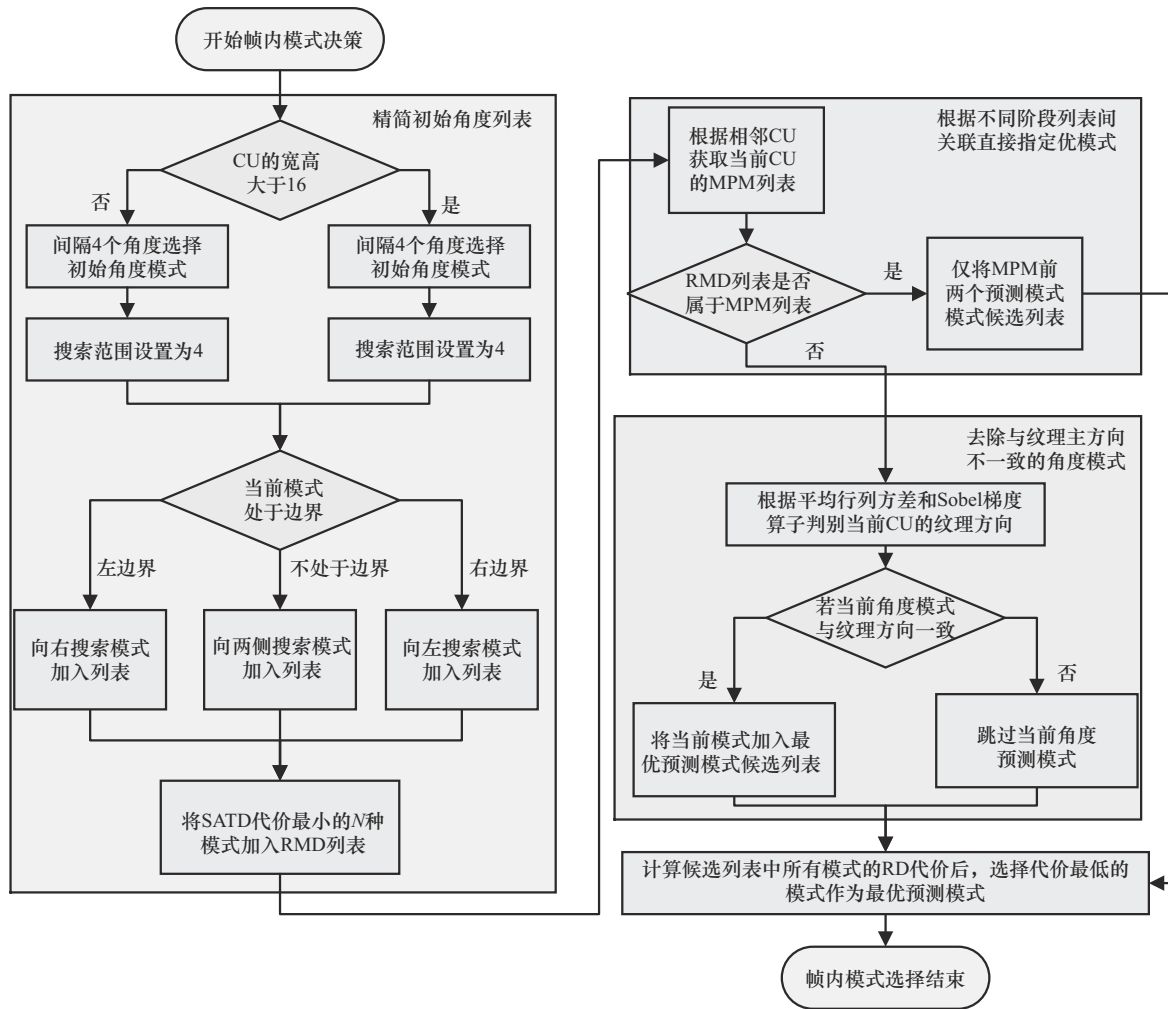


图8 模式选择快速算法流程

理方向跳过与方向不一致的角度预测模式，这有助于减少错误判断的可能性，进一步保障了编码性能。

2 实验与性能分析

为验证本文所提算法的有效性，将块划分快速算法与模式选择快速算法进行集成，嵌入VVC标准测试软件VTM-10.0中进行编码测试。本文实验环境的硬件配置为3.70 GHz的Intel i9-10900X CPU，运行内存为128 GB，在64位Windows 10操作系统下进行测试。深度学习模型的实现框架使用Pytorch 1.13.1+cuda。所提算法在模型训练阶段使用GPU进行加速，在性能评估阶

段仅使用CPU进行运算。测试序列为VVC通测条件（common test condition, CTC）下的B、C、D、E共4类序列，对每个序列进行了全帧内编码（all intra, AI）配置，并测试了各序列下的编码性能。

在性能评估指标方面，本文采用VCEG-M33^[23]提出的BDBR方法，主要评估指标分别为BDBR和 ΔT 。其中，BDBR表示相同图像质量下的相对比特率，数值越小说明提出的方法在编码效率方面表现更优。 ΔT 表明编码时间缩短量，包含了深度学习模型的推理时间，其计算式为：



表2 本文算法与VTM-10.0原始平台的对比实验结果

类型	序列	文献[12]方法		文献[15]方法		文献[18]方法		文献[19]方法		本文方法	
		ΔT	BDBR	ΔT	BDBR	ΔT	BDBR	ΔT	BDBR	ΔT	BDBR
B	BasketballDrive	50.47	1.31	28.00	0.61	47.68	0.76	63.74	2.03	67.75	2.26
	BQTerrace	38.38	0.93	23.00	0.49	40.88	0.87	81.10	1.39	66.81	3.81
	Cactus	40.53	1.10	27.00	0.48	43.28	1.17	76.06	2.12	65.86	2.96
	MarketPlace	44.82	1.13	33.00	0.32	—	—	51.61	1.55	61.24	2.97
	RitualDance	43.51	2.19	29.00	0.47	—	—	58.98	1.88	65.74	2.45
C	BasketballDrill	38.40	1.91	25.00	0.68	49.78	1.01	34.40	2.32	67.15	3.73
	BQMall	44.86	0.95	23.00	0.60	45.21	0.67	36.82	1.64	64.96	1.60
	PartyScene	29.38	0.48	25.00	0.34	35.67	0.85	27.34	0.76	59.18	0.80
	RaceHorses	30.64	0.65	26.00	0.31	40.08	0.51	39.39	1.03	63.04	1.49
D	BasketballPass	41.26	1.10	25.00	0.43	—	—	23.50	1.03	67.79	2.23
	BlowingBubbles	35.03	0.67	22.00	0.23	31.01	0.76	16.60	0.56	63.83	1.21
	BQSquare	38.99	0.91	24.00	0.32	34.69	0.69	17.94	0.58	57.40	2.63
	RaceHorses	33.72	0.72	24.00	0.29	32.78	0.71	23.21	0.80	58.87	0.91
E	FourPeople	45.60	1.51	24.00	2.69	35.71	0.59	54.98	2.67	67.83	2.64
	Johnny	39.57	1.85	26.00	0.62	39.90	0.67	55.54	3.25	64.79	2.95
	KristenAndSara	44.89	1.12	25.00	0.53	33.09	0.73	50.79	2.43	62.33	2.04
	平均	40.00	1.16	25.56	0.59	39.21	0.77	44.50	1.63	64.04	2.29

表3 块划分算法、模式选择算法与集成综合算法的编码结果

类型	序列	CU 划分		模式选择		集成综合	
		ΔT	BDBR	ΔT	BDBR	ΔT	BDBR
B	BasketballDrive	62.28	2.39	10.47	-0.03	67.75	2.26
	BQTerrace	63.60	3.79	11.53	0.25	66.81	3.81
	Cactus	63.54	2.74	8.58	0.31	65.86	2.96
	MarketPlace	57.54	2.88	10.49	0.19	61.24	2.97
	RitualDance	64.78	2.10	9.08	0.32	65.74	2.45
C	BasketballDrill	64.65	3.81	8.73	-0.28	67.15	3.73
	BQMall	62.83	1.78	7.18	-0.31	64.96	1.60
	PartyScene	56.09	0.66	7.93	0.83	59.18	0.80
	RaceHorses	61.61	1.52	8.48	-0.11	63.04	1.49
D	BasketballPass	64.22	2.34	9.24	-0.09	67.79	2.23
	BlowingBubbles	57.91	1.02	9.05	0.46	63.83	1.21
	BQSquare	55.63	2.85	9.88	-0.31	57.40	2.63
	RaceHorses	57.26	0.96	8.30	-0.04	58.87	0.91
E	FourPeople	64.40	2.38	10.00	0.60	67.83	2.64
	Johnny	61.96	2.71	9.81	0.33	64.79	2.95
	KristenAndSara	59.99	2.07	8.75	-0.04	62.33	2.04
	平均	61.14	2.25	9.22	0.13	64.04	2.29

$$\Delta T = \frac{T_{\text{org}} - T_{\text{pro}}}{T_{\text{org}}} \times 100\% \quad (8)$$

其中, T_{org} 和 T_{pro} 表示 VVC 原始编码器 VTM 和本文算法在相同测试条件下得到的编码时间。

本文算法与 VTM-10.0 原始平台的对比实验结果见表 2, 块划分算法、模式选择算法与集成综合算法的编码结果见表 3。实验结果表明, 本文算法能够在 BDBR 平均为 2.29% 的情况下, 实现平均 64.66% 的复杂度降低。在不同类别的视频序列中, 性能表现接近, 这表明所提算法具有良好的泛化性能。在 BasketballPass 序列中, 本文算法取得了最大幅度的时间节省, 减少了 67.83% 的编码时间。这是由于所提算法能够准确分类 CU 划分结构, 并跳过低概率的划分模式, 同时在模式选择中跳过非必要的预测模式。关于本文算法中 CU 划分的优化部分, 能够在 2.25%BDBR 提升的前提下, 平均节省 61.14% 的编码时间。其中, RitualDance 序列实现的时间节省最为显著, 达到 64.78%, 这是由于提出的深度学习模型能够精准预测 CU 划分结构, 并基于得到的预测概率跳过了非必要的划分结构。关于预测模式选择的优化部分, 本文算法在 0.13%BDBR 提升的前提下, 实现平均 9.22% 的时间缩短, 这表明所提算法能够基于方向方差和 Sobel 梯度判别纹理方向, 并依据 RMD 与 MPM 的模式间相关性跳过冗余预测模式的计算, 有效降低模式选择过程中的编码复杂度。然而, 对于个别序列, 如 BQSquare, 时间节省仅为 57.40%, 这是由于在该序列上对 CU 划分结构的预测能力稍差, 倾向于将 CU 进一步划分, 这使得算法未能有效降低 CU 划分过程中的编码复杂度, 使得综合算法节省时间较少。

与其他先进算法相比, 本文算法在平均编码时间节省方面表现更加出色。陈燕辉等^[12]通过子块纹理复杂度选出复杂 CU, 再根据子块方向特性决定 CU 划分方向, 再利用相邻子块间差异决策是

否跳过二叉树划分, 该方法依赖于对 CU 纹理复杂度和子块划分方向特性的分析, 并不能带来良好的编码性能。Park 等^[15]通过使用帧内预测期间得到显式和派生的 VVC 特征作为轻量级神经网络的输入, 以决定是否跳过 CU 二叉树划分。然而, 这种基于有限特征的预测无法捕获 CU 深层次特征, 降低了所提算法的预测精度。程栋梁等^[18]提出一种基于纹理复杂度的 CU 分类模型, 利用多尺度信息和自适应质量复杂度均衡损失函数实现划分结构的精准预测, DenseNet 和 FPN 带来的相对较高的模型推理时间, 导致该方法虽能实现优异的编码性能损失, 但仅能实现较少程度的编码时间降低。Zhang 等^[19]通过使用基于 CNN 架构的 DenseNet 获取 CU 深层次特征, 但相较于 GNN, 其全局特征获取能力较低。且该算法主要针对 4×4 块的划分边界进行预测, 并根据预测的边界概率间接推断划分结构, 然而, 由于该方法并不是直接预测当前 CU 的划分结构, 因此难以充分发挥深度学习在预测性能方面的优势。并且上述方法并未优化帧内模式选择过程, 而本文所提算法则考虑帧内模式选择过程的优化, 因此在减少编码复杂度方面表现更为出色。但本文算法中对划分模式和预测模式的预测并非完全准确, 导致本文算法结果中 BDBR 相较 Zhang^[19]有所增加。

为了直观展现所提算法的主观质量评价, 本文算法与原始编码器的解码帧比较如图 9 所示, 展示了 BQTerrace、BlowingBubbles 和 BasketballDrive 序列在 QP=22 时的解码帧图像。从图 9 中可以看出, 所提算法在解码帧的整体与细节部分表现出与原始平台相媲美的性能。这表明所提算法能够在实现 64.04% 的编码时间减少和 2.29%BDBR 的情况下, 以较高编码效率实现可观的高编码质量。进一步证明了所提算法在显著降低编码复杂度的同时, 成功避免了解码后视频主观质量的降低。

此外, 为更直观展示本算法的率失真性能, 本



图9 本文算法与原始编码器的解码帧比较

文绘制了BDBR最优和损失最多的两个序列的RD曲线, 本文算法与原始编码器的RD性能比较如图10所示。图10(a)展示了BQTerrace序列的RD曲线, 本文算法的峰值信噪比(peak signal-to-noise ratio, PSNR)与原平台基本一致, 但比特率略有升高。图10(b)展示了PartyScene序列的RD曲线, 该序列对应RD曲线与原始编码器的RD曲线几乎完全重合, 表明在实验的4个QP下, 所提算法在RD性能上与VTM 10.0保持一致。这进一步说明了所提算法具有良好的率失真性能。因此, 所提算法能够在编码损失可忽略的前提下, 显著降低编码复杂度。

3 结束语

本文提出了一种基于图神经网络和统计分析的VVC帧内编码快速算法。首先, 利用图神经网络预测CU最优划分结构, 并据此跳过冗余的划分模式。同时, 考虑CU空间相关性和纹理特征, 优化了帧内模式选择过程, 避免了冗余的预测模式遍历。实验结果表明, 相较于原始平台VTM-10.0, 该算法能够减少64.04%的编码时间, BDBR仅略微增至2.29%。未来研究将进一步致力于设计深度学习模型, 实现对更多尺寸的CU最优划分模式的准确预测。同时, 将深入研究深

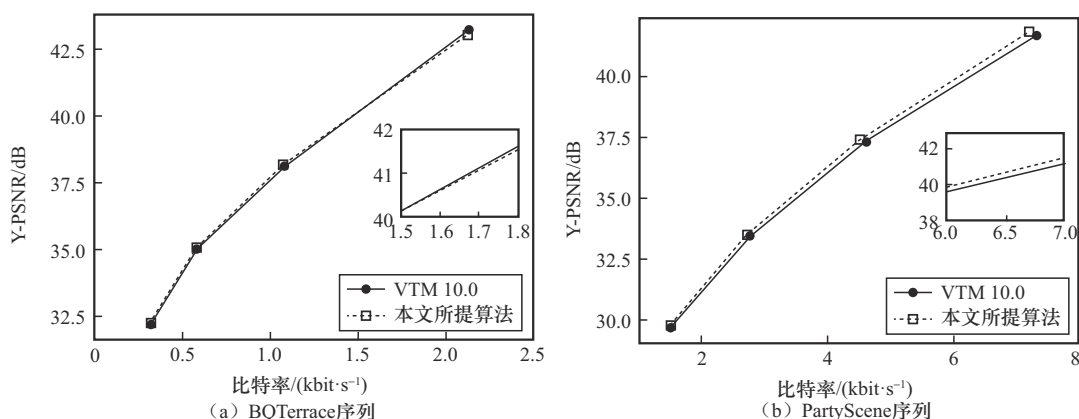


图10 本文算法与原始编码器的RD性能比较

度学习方法在帧间快速算法中的应用, 以进一步提高编码速度。

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] HUANG Y W, AN J C, HUANG H, et al. Block partitioning structure in the VVC standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(10): 3818-3833.
- [3] MAX, YANG H, CHEN J. CE3: CCLM/MDLM using simplified coefficients derivation method(Test 5.6. 1, 5.6. 2 and 5.6. 3)[J]. Joint Video Experts Team(JVET) of ITU-T SG, 2018(16): 3-12.
- [4] CHEN J, YE Y, KIM S. Algorithm description for versatile video coding and test model 5(VTM 5)[EB]. 2019.
- [5] JIANLE C. Algorithm description for versatile video coding and test model 4(VTM 4)[EB]. 2019.
- [6] SANTIAGO D L H. CE3: Intra sub-partitions coding mode[EB]. 2019.
- [7] BROSS B, CHEN J L, OHM J R, et al. Developments in international video coding standardization after AVC, with an overview of versatile video coding(VVC) [J]. Proceedings of the IEEE, 2021, 109(9): 1463-1493.
- [8] PAKDAMAN F, ALI A M, GABBOUJ M, et al. Complexity analysis of next-generation VVC encoding and decoding[C]// Proceedings of the 2020 IEEE International Conference on Image Processing(ICIP). Piscataway: IEEE Press, 2020: 3134-3138.
- [9] 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.
- [10] DONG X C, SHEN L Q, YU M, et al. Fast intra mode decision algorithm for versatile video coding[J]. IEEE Transactions on Multimedia, 2022(24): 400-414.
- [11] TANG N, CAO J, LIANG F, et al. Fast CTU partition decision algorithm for VVC intra and inter coding[C]//Proceedings of the 2019 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS). Piscataway: IEEE Press, 2019: 361-364.
- [12] 陈燕辉, 李强, 董阳, 等. 基于 CU 特征差异的 VVC 帧内快速划分算法[J]. 计算机应用研究, 2024, 41(3): 939-943.
- CHEN Y H, LI Q, DONG Y, et al. Fast partitioning algorithm for VVC intra coding based on the characteristic differences[J]. Application Research of Computers, 2024, 41(3): 939-943.
- [13] WU G, 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. Piscataway: IEEE Press, 2021: 1-5.
- [14] LIU X, LI Y, LIU D, et al. An adaptive CU size decision algorithm for HEVC intra prediction based on complexity classification using machine learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 29(1): 144-155.
- [15] PARK S, KANG J W. Fast multi-type tree partitioning for versatile video coding using a lightweight neural network[J]. IEEE Transactions on Multimedia, 2020(23): 4388-4399.
- [16] ZHANG S, FENG S, CHEN J, et al. A GCN-based fast CU partition method of intra-mode VVC[J]. Journal of Visual Communication and Image Representation, 2022, 88: 103621.
- [17] WU S, SHI J, CHEN Z. HG-FCN: hierarchical grid fully convolutional network for fast VVC intra coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(8): 5638-5649.
- [18] 程栋梁, 张吟龙, 周道先, 等. 基于 DenseNet+FPN 网络的视频帧内 CU 快速划分算法[J]. 光电子·激光, 2024: 1-8.
- CHENG D L, ZHANG Y L, ZHOU D X, et al. A DenseNet+FPN network based fast partitioning algorithm for CU intra coding[J]. Journal of Optoelectronics-Laser, 2024: 1-8
- [19] ZHANG Q, GUO R, JIANG B, et al. Fast CU decision-making algorithm based on DenseNet network for VVC[J]. IEEE Access, 2021(9): 119289-119297.
- [20] DANG-NGUYEN D T, PASQUINI C, CONONTTER V, et al. Raise: a raw images dataset for digital image forensics[C]// Proceedings of the 6th ACM Multimedia Systems Conference. New York: ACM Press, 2015: 219-224.
- [21] HAN K, WANG Y, GUO J, et al. Vision gnn: an image is worth graph of nodes[J]. Advances in Neural Information Processing Systems, 2022(35): 8291-8303.
- [22] HUANG Z, WANG X, HUANG L, et al. CCNet: criss-cross attention for semantic segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. Piscataway: IEEE Press, 2019: 603-612.
- [23] BJONTEGAARD G. Calculation of average PSNR differences between RD-curves[J]. ITU SG16 Doc. VCEG-M33, 2001.



[作者简介]



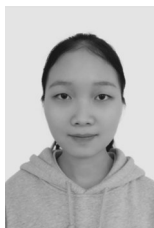
黎天送 (1987-), 男, 博士, 重庆师范大学讲师, 主要研究方向为图像/视频编码、多视点视频编码、多媒体信号处理、人工智能。



刘姝岑 (1999-), 女, 重庆师范大学硕士生, 主要研究方向为3D-HEVC 视频编码、深度学习。



刘昊坤 (1999-), 男, 重庆师范大学硕士生, 主要研究方向为H.266/VVC 视频编码、深度学习。



陈艳 (2000-), 女, 重庆师范大学硕士生, 主要研究方向为H.266/VVC 视频编码、360°全景视频编码、深度学习。



崔少国 (1974-), 男, 重庆师范大学教授, 主要研究方向为大数据与人工智能、计算机视觉(CV)与自然语言处理(NLP)、医学影像智能分析。



王鸿奎 (1990-), 男, 现就职于杭州电子科技大学丽水研究院, 主要研究方向为标准视频压缩、感知视频压缩以及智能视频压缩。