研究与开发

一种基于深度可分离卷积的 VVC 帧内编码快速块划分算法

叶振[1]，王国相[1]，宋俊锋[1]，刘昊坤[2]，黎天送[2]
（1. 丽水学院，浙江 丽水 323000；2. 重庆师范大学，重庆 401331）

摘　要：最近，联合视频探索工作组（JVET）将通用视频编码（VVC）作为新一代视频编码标准，它利用复杂的四叉树加多类型树（QTMTT）划分结构有效地提升了编码性能，但也导致编码复杂度急剧攀升，大幅地增加了编码时间。为解决上述问题，提出了一种基于深度可分离卷积的 VVC 帧内编码快速块划分算法，将编码单元（CU）的原始像素值作为输入，利用轻量化的深度可分离卷积神经网络提取 CU 纹理信息特征指导 CU 的划分模式选择，实现精准的划分模式预测。该方案通过跳过低概率的划分模式，减少 CU 划分模式的遍历，大幅地降低编码器的复杂度。实验结果表明，所提算法在 VTM 15.2 平台上实现了 18%～48%的编码时间节省，仅仅带来了平均 0.15%的性能损失，并且轻量化的深度可分离卷积计算带来的额外复杂性也可以忽略不计。

关键词：视频编码；深度学习；帧内编码；编码单元划分
中图分类号：TP391
文献标志码：A
doi: 10.11959/j.issn.1000−0801.2023132

A fast block partitioning algorithm for VVC intra coding based on depthwise separable convolution

YE Zhen[1], WANG Guoxiang[1], SONG Junfeng[1], LIU Haokun[2], LI Tiansong[2]
1. Lishui University, Lishui 323000, China
2. Chongqing Normal University, Chongqing 401331, China

Abstract: The joint video exploration team (JVET) proposed versatile video coding (VVC) as a new video coding standard, and its quadtree plus multi-type tree (QTMTT) partition structure brings effective coding performance improvements. However, it brings about a sharp increase in encoding complexity, which greatly increases the encoding time. In order to solve the above problems, a fast block partitioning algorithm for VVC intra coding based on depthwise separable convolution was proposed. The pixel of coding unit (CU) was used as input, and the texture information feature of CU was extracted through depth-separable convolution. Therefore, accurate partition mode prediction was realized in the QTMT structure in VVC, and the complexity of the encoder was reduced by skipping

收稿日期：2023−03−28；**修回日期：**2023−06−13
通信作者：黎天送，tiansongli@cqnu.edu.cn
基金项目：重庆市科技局自然基金项目（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.CSTB2022 NSCQ-MSX1231), The Science and Technology Research Program of Chongqing Municipal Education Commission (No.KJQN20220051), The Talents Fund Project of Chongqing Normal University (No.21XLB031)

low-probability partition modes. Experimental results show that the proposed algorithm saves 18% to 48% of encoding time on the VTM 15.2, and only brings an average performance loss of 0.15%. And the additional complexity brought by the lightweight depthwise separable convolution calculation is also negligible.

Key words: video coding, deep learning, intra coding, coding unit division

0 引言

随着多媒体技术的快速发展，人们对高清和超高清视频的需求正在不断增加。然而，随着4K/8K、360°视频、高动态范围和宽广色域等新的视频数据格式的不断涌现，前一代视频标准——高效视频编码（high efficiency video coding，HEVC/H.265[1]）面对日益增长的编码效率需求显得力不从心。为了解决这个问题，国际电信联盟（International Telecommunications Union，ITU）和国际标准化组织（International Organization for Standardization，ISO）/国际电工委员会（International Electrotechnical Commission，IEC）合作开发了通用视频编码（versatile video coding，VVC/H.266[2]）标准。相较于 HEVC，VVC 具有更好的网络适应性、并行处理能力和更高的压缩效率，能够提升约 30%的编码性能。

VVC 采用了与 HEVC 类似的混合编码框架，主要包含帧内预测、帧间预测、变换与量化、环路滤波及熵编码等模块。帧内预测主要用于去除图像的空域冗余；帧间预测主要用于去除图像的时域冗余；变换与量化通过对残差数据的处理去除频域相关性，实现有损压缩；环路滤波通过削减甚至去除重构图像方块效应提高图像的主观质量并提高压缩率[3]。相较于 HEVC，VVC 引入了许多新技术提升编码性能，尤其在帧内预测方面。这些新技术包括嵌套多类型树的四叉树（quadtree with nested multi-type tree，QTMTT）、跨分量线性模型（cross-component linear model，CCLM）[4]、基于位置信息的帧内联合预测（position dependent prediction combination，PDPC）[5]、多参考行（multiple reference line，MRL）预测[6]、帧内子块划分（intra sub-partitions，ISP）[7]和矩阵权重帧内预测（matrix weighted intra prediction，MIP）[8]，以及增加到 67 种的帧内模式选择。这些帧内预测新工具的引入大幅提升了编码性能，但同时也导致了编码复杂度的急剧提升。VVC 帧内编码相较于 HEVC 提升了约 25 倍的复杂度，其中块划分技术对编码过程产生了较大的影响。VVC 块划分中新技术的提出带来了 6.67%编码性能的提升，但同时也增加了 10.87 倍的编码复杂度。可以看出，块划分在编码效率的提升和编码复杂度的增加上都占据了相当大的比例。因此，如何在尽量减少编码性能损失的前提下有效降低帧内预测的编码复杂度成为众多学者研究的重点。

目前，针对 HEVC 和 VVC 中的快速块划分问题，研究主要可分为 3 类：启发式方法、基于机器学习的方法以及基于深度学习的方法。启发式方法通常通过阈值或率失真代价等人为设计的决策规则进行编码加速。卢嘉彬等[9]提出了一种面向 VVC 的帧内编码快速算法，通过自适应阈值和纹理方向去除非最优划分模式，同时使用划分模式指导编码单元（coding unit，CU）进行角度模式的预测；陶浩然等[10]提出一种基于梯度幅值相似度的 CU 快速划分方法，通过当前 CU 的子 CU 的平均梯度幅值相似度偏差来判断是否进行四叉树划分，通过子块方差值来确定二叉树和三叉树的最优化分；Tang 等[11]提出了一种用于帧内和帧间编码的快速块分割算法，通过边缘检测在帧内编码中跳过一些划分模式，在帧间编码中则通过三帧差分法实现；Dong 等[12]提出了一种综合决策策略来排序候选模型，以选择合适的预测模型；Li 等[13]通过为 CU 的二叉树（binary tree，BT）划分和三叉树（ternary tree，TT）划分建立决策

模型，以跳过部分划分模式。虽然这些启发式方法可以加速编码，但是它们很难找到具有普遍适用性的规则，因为不同视频序列之间的差异较大，算法的鲁棒性较低。为了克服这个问题，一些学者将机器学习应用于视频编码领域，将块划分视为分类问题，将块划分的可能结果分为多个类别，并使用高级特征来预测分类结果。这些方法可以使用灵活的表示能力提前终止或剪枝操作，以减少编码时间。熊丹祺等[14]使用决策树加速帧内 CU 划分，通过将各深度下从 CU 提取的表征复杂度的特征输入决策树模型预测 CU 划分结果；Zhao 等[15]提出了一种基于支持向量机（support vector machine，SVM）的快速 CU 划分决策算法，该算法选择更可靠的相关特征并决定是否划分，使用最大标准偏差（standard deviation，SD）和边缘点比（edge point ratio，EPR）；Wu 等[16]提出了一种基于 SVM 的方法，该方法利用纹理信息预测 CU 划分；Liu 等[17]提出了一种利用 SVM 对 CU 复杂度进行分析并构建分类模型的方法。虽然这些机器学习方法取得了不错的成果，但由于获取的特征能力有限，算法的性能仍有提升空间。因此，近年来，一些学者开始采用深度学习方法解决这个问题。Zhang 等[18]提出了全局卷积网络（global convolutional network，GCN）来感知全局信息，从而更好地利用网络预测划分模式；Tech 等[19]提出使用卷积神经网络（convolutional neural network，CNN）预测限制编码块高和宽的两个参数，以跳过部分预测模式，从而减少编码时间；WU 等[20]使用层次全卷积网络（hierarchical grid fully convolutional network，HG-FCN）进行一次推理即可获得当前 CU 和子 CU 的划分情况；Fang 等[21]通过结合特征分析和深度学习加速 VVC 中的块划分，其中特征分析方法基于像素的方差，使用卷积神经网络进行分类。随着深度学习轻量化研究的发展，以深度可分离卷积为代表的轻量化模型有着参数量小、推理速度快且对硬件要求低的特征，这使得深度学习能够应用在实际生活场景中。而推理速度快、硬件要求低正是 CU 划分预测所需要的，因此深度可分离卷积对于帧内块划分任务是行之有效的。

针对现有方法存在的不足，本文提出了一种基于深度可分离卷积的 VVC 帧内编码快速块划分算法。该算法利用了深度学习的特征表征和分类能力，通过精准预测待编码块的划分模式，减少 CU 划分模式的遍历。实验结果表明，本文提出的方法能够在保持编码性能几乎不变的情况下，高效地大幅减少编码时间。

1　VVC 中的 CU 划分结构

编码块是视频编码的基础单元，HEVC 和 VVC 都是基于块的视频编码标准。在 HEVC 中，块划分模式仅有四叉树（quadtree，QT）划分，从尺寸为 64 pixel×64 pixel 的 CU 开始逐步按照四叉树划分为 4 个大小相同的正方形子 CU，同时深度值加 1。故在 HEVC 中 CU 尺寸与 CU 深度一一对应，可以通过直接预测 CU 深度推导出对应 CU 尺寸，HEVC 块划分示例如图 1 所示。在进行块划分后，CU 会被进一步划分为预测单元和变换单元以进行预测和变换过程及后续编码流程。

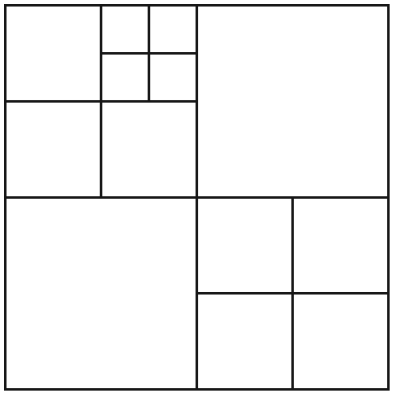
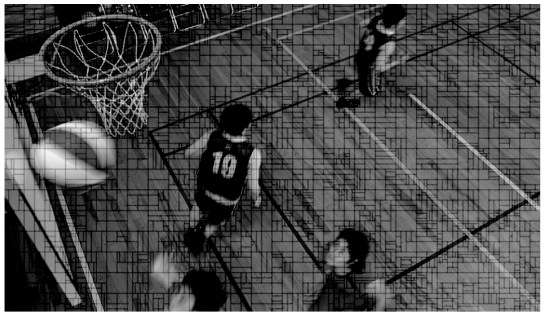


图 1　HEVC 块划分示例

VVC 引入了四叉树加多类型树（quadtree plus multi-type tree，QTMTT）划分结构，在四叉树划分的基础上增加了水平和竖直的二叉树和三叉树

划分，VVC 块划分实例和 QTMTT 中的 6 种划分模式如图 2 所示。在 VVC 中编码树单元（coding tree unit，CTU）尺寸增大到 128 pixel×128 pixel，首先会被划分为 4 个 64 pixel×64 pixel 的 CU，之后各子 CU 会递归按照多叉树划分直至结束。多叉树划分包括竖直二叉树（vertical binary tree，BTV）、水平二叉树（horizontal binary tree，BTH）、竖直三叉树（vertical ternary tree，TTV）和水平三叉树（horizontal ternary tree，TTH）。得益于灵活的块划分结构能够适应不同的局部纹理特征，VVC 能够更加准确地进行块划分供后续帧内/帧间编码。但此划分过程同样导致了编码复杂度的成倍提升，因为划分决策需要进行率失真代价计算，而每次计算需要完整的预测、变换、量化、熵编码等过程。因此在编码过程中能够在块划分流程中进行优化加速，可以有效减少编码时间，降低编码复杂度。



VVC划分实例

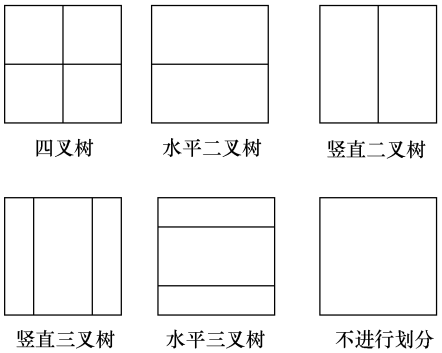


图 2　VVC 划分结构实例和 QTMTT 中的 6 种划分模式

每个 CU 的划分方式初始化在 CU 块划分模式列表（CU partition mode list）中，根据 VVC 中的划分规则和限制条件，将所有能够计算率失真代价的划分模式加入列表中，在后续的编码过程中遍历列表中所有的划分方式及子块的不同划分模式组合，在分别计算率失真代价后确定最优块划分方式，未添加到列表中的划分方式与组合将不再计算代价。因此在快速块划分算法中，可以通过将可能性较低的划分模式从列表中删除，跳过部分率失真代价的计算以达到降低编码时间的目的。

2　基于深度可分离卷积的 VVC 帧内编码快速块划分算法

2.1　数据集构建

深度学习模型会受到数据集种类和数量的显著影响，构建分布均匀、数量庞大的数据集将得到更高效的训练结果。为保证实验的准确性，本文将 RAISE[22]数据集的 2 000 张图像制作为可供编码器使用的视频序列，通过逐帧编码该序列获取各帧图像的所有编码单元的划分结构。随后将各帧图像的亮度分量图像裁剪成与先前编码结果相对应的编码单元，从而制作出可供神经网络学习的数据集。此数据集共包含近 59 万个尺寸为 32 pixel×32 pixel 的训练样本和 15 万个尺寸为 64 pixel×64 pixel 的训练样本。

在亮度通道上不同尺寸下 CU 划分结构的比例不同，例如在宽度大于长度的 32 pixel×16 pixel、32 pixel×8 pixel、16 pixel×8 pixel 等尺寸的块中，垂直划分的比例远远超过水平划分的比例，即编码器此时更倾向于对该类尺寸的块进行垂直划分，相同的情况也会出现在长度大于宽度的尺寸的块中。对于模型的训练而言，各类数据的分布较为平均，可以避免模型偏移的问题。而在长宽一致的尺寸中，过小尺寸的 CU 各划分结构的比例相当不均衡，故选择 32 pixel×32 pixel 和 64 pixel× 64 pixel 的尺寸作为训练数据，可以有效避免数据不平衡的问题。同时，可以使用焦点损失（focal loss）提高数量稀少种类的权重来解决

样本不平衡问题，通过给较低比例的划分结构更大的权重以降低不平衡带来的偏移问题。

2.2　基于深度可分离卷积的深度学习模型构建

为使得深度学习模型相比传统卷积网络更加轻量化，本文使用高效的深度可分离卷积替换了传统卷积，使得模型在保持原有性能的同时有效减少参数量，且通道注意力机制和空间注意力机制的加入使得模型能够学习更贴合 CU 划分任务的特征，实现对编码单元划分结构的精准预测。本文构建的基于深度可分离卷积的深度学习模型的网络结构如图 3 所示。模型的输入为当前待编码块的原始像素，尺寸为 32 pixel×32 pixel。模型的输出为 QTMTT 6 种划分模式的概率值。下面对模型的结构进行详细介绍。

本文提出的深度学习模型共分为 3 个模块，依次为输入预处理模块、特征提取模块和预测分类模块。待编码块的亮度矩阵为单通道，直接运用在特征提取模块无法得到较好的效果，需要将矩阵处理为张量并提升维度。具体而言，输入预处理模块接收大小为32 pixel×32 pixel或64 pixel×64 pixel 的编码单元的原始亮度信息，将亮度矩阵转换为深度学习模型运算的张量，并通过卷积将张量映射到高维空间供后续特征提取模块进行操作。特征提取模块在得到输入预处理模块的张量后，通过若干卷积操作进行特征提取。为加快推理速度和降低模型参数量，本文方法使用深度可分离卷积替换传统卷积操作，并将之用于特征提取与特征融合，再通过池化操作和残差连接获取不同阶段的特征，最后在由若干全连接层组成的预测分类模型将特征图映射到各划分模式的概率上，并运用此概率指导后续编码的块划分。

深度可分离卷积作为传统卷积层的替换可以大大减少运算量和参数数量，由 DW（depthwise）卷积和 PW（pointwise）卷积组成。传统卷积可以同时改变特征图尺寸和通道数。而在深度可分离卷积中，DW 卷积用来改变特征图尺寸，PW 卷积用来改变特征图通道数。DW 卷积通过对每一个输入通道运用一个一维的卷积，卷积运算会减小特征图的尺寸，输入特征图的通道数等于输出特征图的通道数。PW 卷积为 1×1 的点卷积，此时卷积操作不改变特征图尺寸，输出特征矩阵的通道数等于卷积核的个数。故将 DW 卷积和 PW 卷

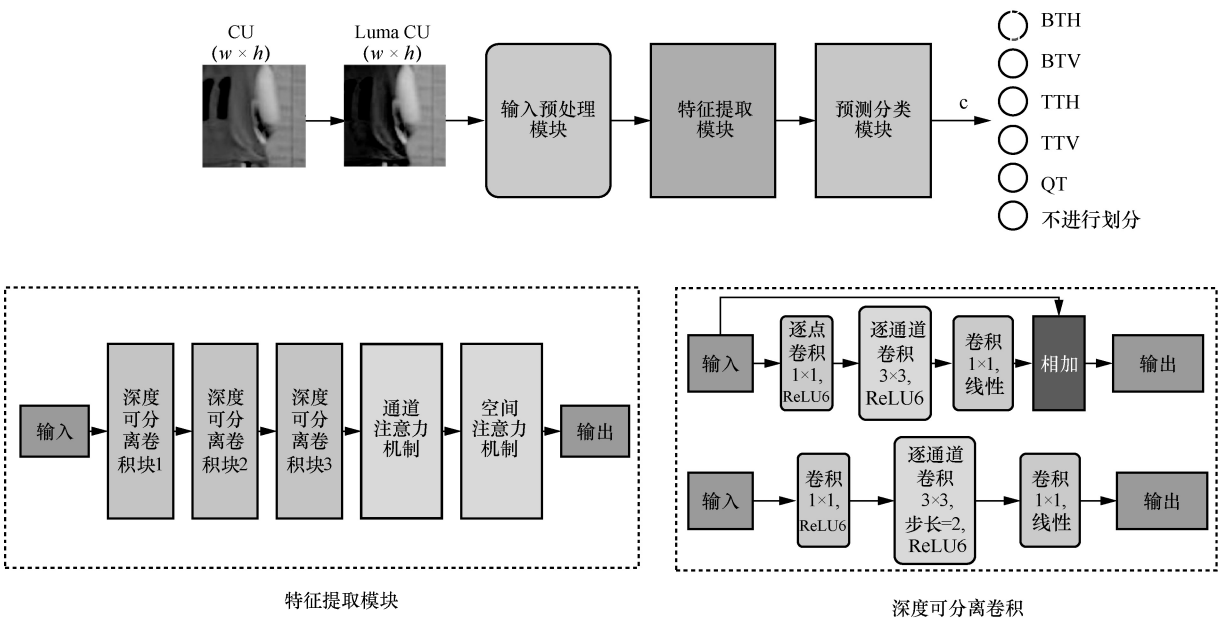


图 3　基于深度可分离卷积的深度学习模型的网络结构

积组合起来，前者用于滤波而后者用于合并。这种方法大大减少了计算量，减小了模型尺寸，并带来了良好的推理速度优化。

传统卷积通过在图片上以一定的步长划动，每次划动将原图上的像素与卷积核的对应位置上的值相乘后相加得到一个输出，故输出特征图中一个值的运算量是 $D_K \times D_K \times M$，总计需要计算 $D_F \times D_F \times N$ 次，故传统卷积的总运算量为：

$$C_{conv} = D_K \times D_K \times M \times N \times D_F \times D_F \quad (1)$$

其中，D_K 是卷积核尺寸，M 为输入特征图的通道数，N 为输出特征图的通道数，D_F 为特征图尺寸。深度可分离卷积可分为 DW 卷积和 PW 卷积两个部分。DW 卷积中一个通道只被一个卷积核卷积。其中输出特征图中一个值的运算量为 $D_K \times D_K \times M$，总计需要计算 $D_F \times D_F$ 次。PW 卷积用来处理通道间信息的交互，会将上一层的特征图在通道上进行加权组合，输出新的特征图。这种卷积可以通过较少的计算量改变输出数据的维度。在输出特征图的一维中，总共需计算 $M \times D_F \times D_F$ 次，而输出特征图维数为 N。综上，深度可分离卷积的运算量为：

$$C_{DW} = D_K \times D_K \times M \times D_F \times D_F \quad (2)$$

$$C_{PW} = M \times N \times D_F \times D_F \quad (3)$$

$$C_{ALL} = C_{DW} + C_{PW} = D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F \quad (4)$$

$$\frac{C_{ALL}}{C_{conv}} = \frac{D_K \times D_K \times M \times D_F \times D_F + M \times N \times D_F \times D_F}{D_K \times D_K \times M \times N \times D_F \times D_F} = \frac{1}{N} + \frac{1}{D_K^2} = \frac{1}{N} + \frac{1}{9} \quad (5)$$

深度可分离卷积是由一层深度卷积和一层逐点卷积组合而成的，相比传统卷积可以在保证精度不变的情况下大幅减少参数和计算量。将深度可分离卷积和传统卷积的计算量相除，可以得到理论上的计算量减少。理论上标准卷积的计算量是深度可分离卷积的 8～9 倍。

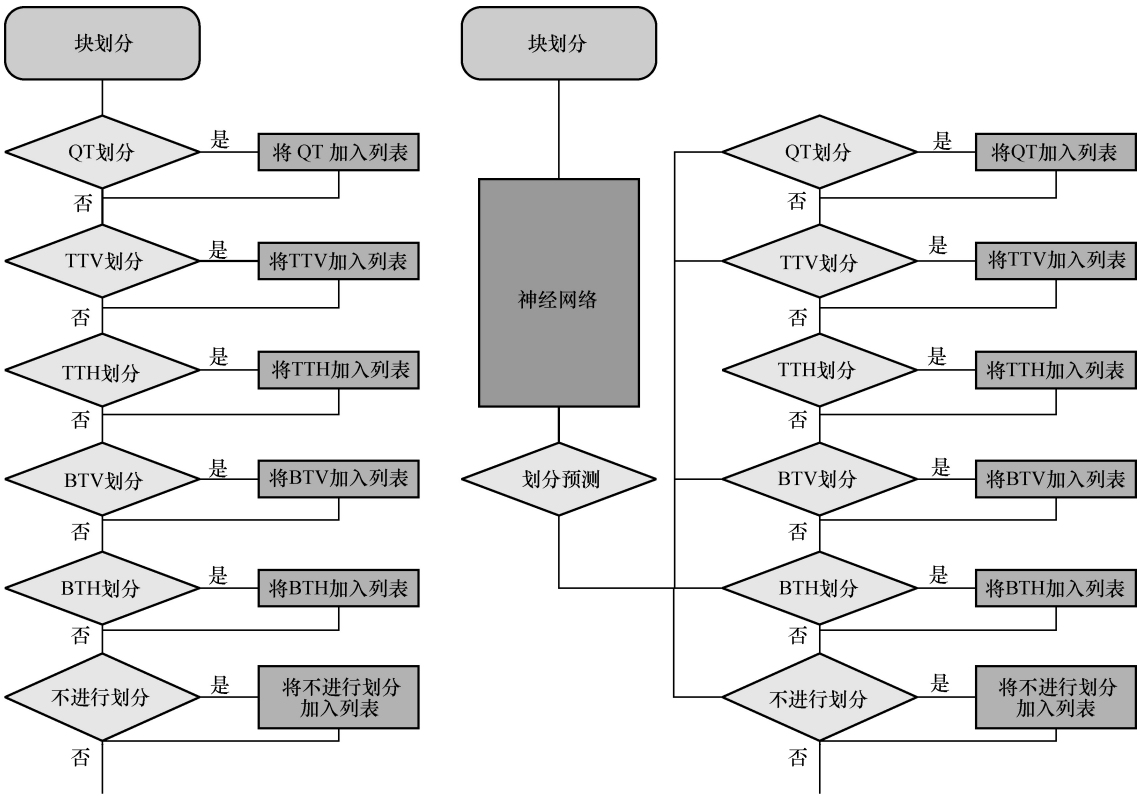
2.3 基于深度学习的 VVC 帧内编码快速块划分方案

本文提出的基于深度学习的 VVC 帧内编码快速块划分与原编码器中块划分的对比如图 4 所示。首先，在读取待编码视频序列时，需要通过对视频进行解码，以获取视频中的原始像素信息。在这个过程中，需要加载训练好的深度学习模型，供后续编码过程中使用模型进行划分预测。其次，在进行帧内编码块划分之前，需要将待编码的视频帧分成多个编码块。划分的目的是将视频帧分成多个块，每个块的像素信息相对独立，这样可以对每个块进行单独的压缩和编码，以实现更加高效的压缩率。在这个过程中，可以利用深度学习模型进行划分模式的预测，通过预测可以确定每个块的划分模式，以达到更加高效的编码效果。划分模式可以根据每个块的亮度像素信息，得到当前编码块的最优划分模式，后续通过跳过低概率的划分模式来有效降低编码时间，且深度学习的模型运行时间几乎可以忽略不计，这证明了本文方法实现帧内预测快速块划分的可行性。

3 实验与性能分析

3.1 实验环境与编码配置

为验证本文方法的性能，需要在性能强力的服务器环境下进行训练数据的生成和深度学习模型的训练。本文的实验环境使用 Windows 10 操作系统，CPU 为 Intel i9 10900X@3.70 GHz 处理器，显卡为 NVIDIA GeFore RTX 3090，实验使用 Python 3.7 编程语言，采用 PyTorch 1.10.1 版本的深度学习计算框架。模型训练过程中采用了 Adam 训练且并采用了 0.000 1 的权重衰退，训练批次设为 512，学习率设为 0.001，共计训练 100 轮。为验证本文方法的性能表现，采用全帧内（all intra-frame）编码模式在 VVC 测试标准平台 VTM 15.2 上进行测试。编译软件为 Visiual Studio 2019，



(a) 原编码器中的块划分　　(b) 基于深度学习的VVC帧内编码快速块划分

图 4　基于深度学习的 VVC 帧内编码快速块划分与原编码器中块划分的对比

配置文件为 encoder_intra_vtm.cfg。

3.2　编码性能分析

本文方法通过 TS 和 BD-BR（Bjøntegaard delta bit-rate）两个指标来评价提出方法的性能。其中，TS 表示编码时间降低率：

$$\mathrm{TS}=\frac{1}{4}\sum_{\mathrm{QP}}\frac{T_{\mathrm{R}}(\mathrm{QP})-T_{\mathrm{C}}(\mathrm{QP})}{T_{\mathrm{R}}(\mathrm{QP})}\times 100\% \tag{6}$$

其中，QP=[22,27,32,37]，T_{R} 为 VTM 15.2 的编码时间，T_{C} 为本文方法的编码时间。进行实验的视频序列为 JVET 测试集中的视频序列（未包含在训练集中），通过对编码时间和 BD-BR 进行性能分析，在 4 个 QP 下求得一个序列的平均编码时间。之后 Bjøntegaard 比特率增量（即相同质量下的比特率变化百分比）和实际编码时间减少。得出的编码时间中还包括模型的推理时间，故计算实际编码时间减少时需要减去该时间。

BD-BR 是一种用于比较视频编码器性能的指标，它可以帮助评估不同编码器在相同质量下的比特率差异。其计算方式是将两个编码器在相同质量下的比特率之差平均化，得到 BD-BR 值。该值越小，表示两个编码器在相同质量下的差异越小，即编码器性能越接近；该值越大，表示两个编码器在相同质量下的差异越大，即编码器性能差异越大。

在 VTM 15.2 平台和全帧内配置下，本文方法与其他方法的比较见表 1，本文方法的编码结果见表 2。所有实验结果均在本文实验环境以及码参数配置下得出。

表 1 比较了本文方法与文献[18]方法和文献[21]方法性能。可以看出，本文方法与其他方法在相同条件下能够带来更优的 BD-BR 和 TS。以 C 序列中的 BasketballDrill 为例，编码时间的减少从 25.32%提升至 48.24%，编码时间缩短近一半，而

表 1　本文方法与其他方法的比较

类别	视频序列	文献[18]方法		文献[21]方法		本文方法	
		BD-BR	TS	BD-BR	TS	BD-BR	TS
B	BasketballDrive	0.18%	13.41%	1.79%	33.78%	0.05%	27.05%
1 920 pixel×1 080 pixel	BQTerrace	0.38%	12.52%	0.58%	30.27%	0.02%	18.31%
	Cactus	0.27%	13.46%	0.81%	30.11%	0.03%	27.80%
C	BasketballDrill	0.12%	25.32%	0.92%	29.93%	0.11%	48.24%
832 pixel×480 pixel	BQMall	−0.03%	16.22%	1.09%	32.63%	0.11%	47.26%
	PartyScene	0.00%	17.97%	0.22%	25.50%	0.04%	35.18%
	RaceHorses	0.19%	17.30%	0.45%	31.64%	0.08%	46.38%
D	BasketballPass	0.32%	13.29%	1.13%	29.19%	0.14%	46.36%
416 pixel×240 pixel	BlowingBubbles	0.12%	14.42%	0.23%	23.93%	0.04%	40.55%
	BQSquare	0.32%	20.93%	0.08%	19.95%	0.02%	26.23%

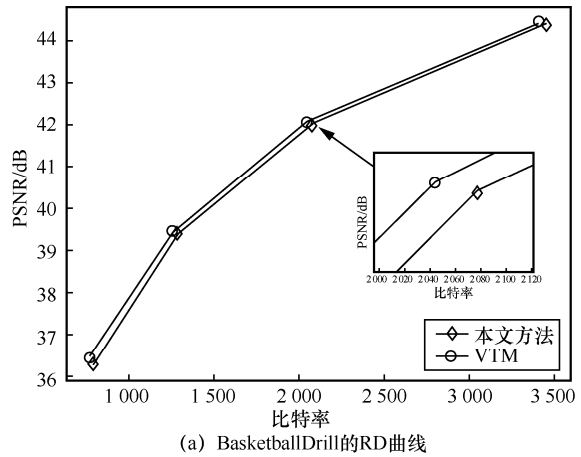
对于 BD-BR 的增加，相差不大。这表明本文方法优于现有方法。

表 2 展示了本文方法在 B、C、D、E 类序列下的实验结果，本文方法能够在几乎不降低编码质量的情况下有效减少编码时间。本文方法平均 BD-BR 增量为 0.08%，这可以说是忽略不计的编码质量下降，而平均缩短的编码时间为 37.12%，对于 BasketballDrive 序列甚至最高能够达到 50.24%的编码时间减少而仅带来 0.15%BD-BR 增加。这说明本文方法针对编码器的块划分算法起到了优化作用。

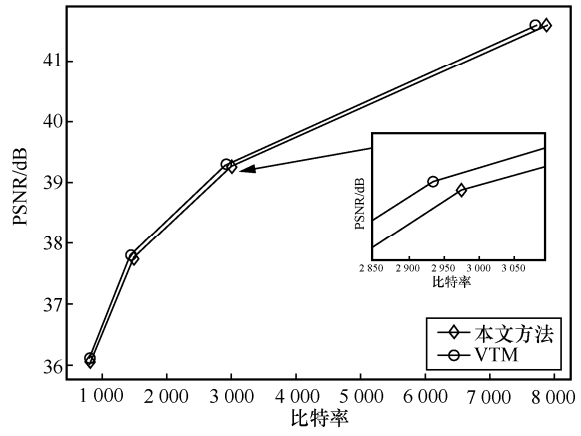
此外，本文方法与 VTM 的性能对比如图 5 所示。图 5（a）～图 5（b）分别展示了本文方法在 B 类序列的 BasketballDrill 和 E 类序列 FourPeople 上的率失真（rate distortion，RD）曲线。如图 5（a）～图 5（b）所示，本文方法能够展现出不错的 RD 性能，这表明本文方法可以很好地适应不同的比特率。而且在高低分辨率序列中，二者几乎能够保持编码质量，产生的 RD 损失可以忽略不计。图 5（c）展示了本文方法在不同 QP 下的编码时间降低曲线，可以看到本文方法在不同 QP 下均能够实现编码时间降低。

表 2　本文方法的编码结果

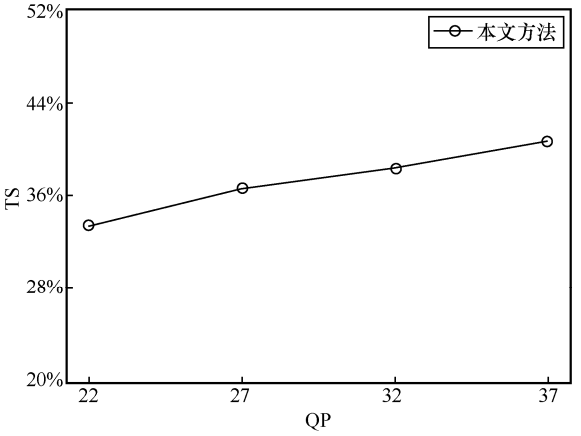
类别	视频序列	本文方法	
		BD-BR	TS
B	BasketballDrive	0.05%	27.05%
1 920 pixel×	BQTerrace	0.02%	18.31%
1 080 pixel	Cactus	0.03%	27.80%
	Kimono1	0.03%	27.48%
	ParkScene	0.03%	20.29%
C	BasketballDrill	0.11%	48.24%
832 pixel×	BQMall	0.11%	47.26%
480 pixel	PartyScene	0.04%	35.18%
	RaceHorses	0.08%	46.38%
D	BasketballDrive	0.14%	46.36%
416 pixel×	BQTerrace	0.04%	40.55%
240 pixel	Cactus	0.02%	26.23%
	Kimono1	0.06%	40.86%
E	BasketballDrive	0.15%	50.24%
1 280 pixel×	BQTerrace	0.24%	46.30%
720 pixel	Cactus	0.14%	45.54%



(a) BasketballDrill的RD曲线



(b) FourPeople的RD曲线



(c) 各QP下的平均TS

图 5　本文方法与 VTM 的性能对比

4　结束语

本文针对 VVC 中块划分新技术的引入所带来的编码效率降低的问题，提出了一种基于深度可分离卷积的 VVC 帧内编码快速块划分算法，该方法通过轻量化的深度可分离卷积从待编码块 CU 的原始像素中提取纹理信息特征，并将之用于精确预测 QTMTT 划分结构。在得到各划分结构的预测概率后，仅对高概率的划分结构进行率失真代价计算而跳过低概率的划分结构。首先，构建了供深度学习模型学习的 CU 及其对应划分的数据集。其次，提出基于深度可分离卷积的 CU 预测划分模型，通过轻量化的深度学习模型进行精准的划分结构预测。最后，将深度学习模型嵌入编码器通过跳过低概率的划分结构减少编码时间。实验结果表明，本文方法在编码性能和编码质量方面有可观表现，本文方法在 BD-BR 平均提高了 0.15%的情况下实现了 18%～48%的编码时间节省。本文方法利用深度学习对尺寸为 32 pixel×32 pixel 的块的划分过程进行优化加速，后续将设计更为高效轻量的网络模型，并在其他尺寸的块上进行划分结构预测，通过降低块划分过程的计算复杂度来进一步提升视频压缩效率。

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, 2012, 22(12): 1649-1668.

[2] 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.

[3] 万帅, 杨付正. 新一代高效视频编码 H.265/HEVC: 原理、标准与实现[M]. 北京: 电子工业出版社, 2014.
WAN S, YANG F Z. New generation efficient video coding H.265/HEVC: principles, standards, and implementation[M]. Beijing: Publishing House of Electronics Industry, 2014.

[4] XIANG M. CE3: CCLM/MDLM using simplified coefficients derivation method[EB]. 2018.

[5] CHEN J, YE Y, KIM S. Algorithm description for versatile video coding and test model 5 (VTM 5)[EB]. 2019.

[6] JIANLE C. Algorithm description for versatile video coding and test model 4 (VTM 4)[EB]. 2019.

[7] SANTIAGO D L H. CE3: Intra sub-partitions coding mode[EB]. 2019.

[8] JONATHAN P, BJÖRN S. CE3: affine linear weighted intra prediction[EB]. 2019.

[9] 卢嘉彬, 彭宗举, 束争杰, 等. 面向 VVC 帧内编码的快速

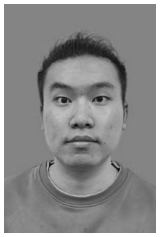
CU 划分和角度模式决策[J]. 光电子·激光, 2021, 32(11): 1171-1179.
LU J B, PENG Z J, SHU Z J, et al. Fast CU partition and angle mode decision for VVC intra coding[J]. Journal of Optoelectronics · Laser, 2021, 32(11): 1171-1179.
[10] 陶浩然, 路锦正, 李意弦. 一种 VVC 帧内编码单元快速划分算法[J]. 小型微型计算机系统, 2021, 42(7): 1470-1474.
TAO H R, LU J Z, LI Y X. Fast division algorithm of VVC intra-coding unit[J]. Journal of Chinese Computer Systems, 2021, 42(7): 1470-1474.
[11] TANG N, CAO J, LIANG F, et al. Fast CTU partition decision algorithm for VVC intra and inter coding[C]//Proceedings of 2019 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS). Piscataway: IEEE Press, 2020: 361-364.
[12] DONG X C, SHEN L Q, YU M, et al. Fast intra mode decision algorithm for versatile video coding[J]. IEEE Transactions on Multimedia, 2021, 24: 400-414.
[13] 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.
[14] 熊丹祺, 高伟, 滕国伟. 基于决策树的 H.266/VVC 帧内预测 CU 快速划分算法[J]. 工业控制计算机, 2021, 34(7): 88-90, 92.
XIONG D Q, GAO W, TENG G W. Decision tree accelerated CU partition algorithm for intra prediction in H.266/VVC[J]. Industrial Control Computer, 2021, 34(7): 88-90, 92.
[15] ZHAO J C, WU A B, ZHANG Q W. SVM-based fast CU partition decision algorithm for VVC intra coding[J]. Electronics, 2022, 11(14): 2147.
[16] WU G Q, HUANG Y, ZHU C, et al. SVM based fast CU partitioning algorithm for VVC intra coding[C]//Proceedings of 2021 IEEE International Symposium on Circuits and Systems (ISCAS). Piscataway: IEEE Press, 2021: 1-5.
[17] LIU X G, LI Y Y, LIU D Y, 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, 2019, 29(1): 144-155.
[18] ZHANG S P, FENG S X, CHEN J W, et al. A GCN-based fast CU partition method of intra-mode VVC[J]. Journal of Visual Communication and Image Representation, 2022, 88: 103621.
[19] TECH G, PFAFF J, SCHWARZ H, et al. Fast partitioning for VVC intra-picture encoding with a CNN minimizing the rate-distortion-time cost[C]//Proceedings of 2021 Data Compression Conference (DCC). Piscataway: IEEE Press, 2021: 3-12.
[20] WU S L, SHI J, CHEN Z B. 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.
[21] FANG J T, LIU B H, CHANG P C. Fast coding unit partitioning algorithms for versatile video coding intra coding[J]. Journal of Visual Communication and Image Representation, 2022, 87: 103542.
[22] DANG-NGUYEN D T, PASQUINI C, CONOTTER 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.

[作者简介]

叶振（1984− ），男，博士，丽水学院计算机系讲师，主要研究方向为人工智能、视频图像处理。

王国相（1989− ），男，丽水学院助教，主要研究方向为信号处理、人工智能、视频图像处理。

宋俊锋（1984− ），男，丽水学院高级实验师，主要研究方向为虚拟现实、人工智能。



刘昊坤（1999− ），男，重庆师范大学计算机与信息科学学院硕士生，主要研究方向为 H.266/VVC 视频编码、深度学习。

黎天送（1987− ），男，博士，重庆师范大学计算机与信息科学学院讲师，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。