

一种快速的多尺度多输入编码树单元互补分类网络

唐述* 周广义 谢显中 赵瑜 杨书丽

(重庆邮电大学计算机科学与技术学院 重庆 400065)

摘要: 深度神经网络(DNN)已被广泛应用到高效视频编码(HEVC)编码树单元(CTU)的深度划分中,显著降低了编码复杂度。然而现有的基于DNN的CTU深度划分方法却忽略了不同尺度编码单元(CU)间的特征相关性和存在着分类错误累积等缺陷。基于此,该文提出一种多尺度多输入的互补分类网络(MCCN)来实现更高效且更准确的HEVC帧内CTU深度划分。首先,提出一种多尺度多输入的卷积神经网络(MMCNN),通过融合不同尺度CU的特征来建立CU间的关联,进一步提升网络的表达能力。然后,提出一种互补的分类策略(CCS),通过结合二分类和三分类,并采用投票机制来决定CTU中每个CU的最终深度值,有效避免了现有方法中存在的错误累积效应,实现了更准确的CTU深度划分。大量的实验结果表明,该文所提MCCN能够更大程度降低HEVC编码的复杂度,同时实现更准确的CTU深度划分:仅以增加3.18%的平均增量比特率(BD-BR)为代价,降低了71.49%的平均编码复杂度。同时,预测 32×32 CU和 16×16 CU的深度准确率分别提升了0.65%~0.93%和2.14%~9.27%。

关键词: 深度神经网络; 帧内高效视频编码; 特征表示; 编码树单元深度划分; 多尺度多输入; 互补分类

中图分类号: TN939.1; TP391.4

文献标识码: A

文章编号: 1009-5896(2024)09-3646-08

DOI: 10.11999/JEIT240223

A Multi-scale-multi-input Complementation Classification Network for Fast Coding Tree Unit Partition

TANG Shu ZHOU Guangyi XIE Xianzhong ZHAO Yu YANG Shuli

(College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400064, China)

Abstract: Deep Neural Networks (DNN) have been widely applied to Coding Tree Unit(CTU) partition of intra-mode High Efficiency Video Coding(HEVC) for reducing the HEVC encoding complexity, however, existing DNN-based CTU partition methods always neglect the correlation of features between Coding Units (CU) at different scales and suffer from the accumulation of classification errors. Therefore, in this paper, a Multi-scale-multi-input Complementation Classification Network (MCCN) for faster and more accurate CTU partition is proposed. First, a Multi-scale Multi-input Convolutional Neural Network (MMCNN) is proposed, which builds up the correlation of features between CUs at different scales by fusing multi-scale CU features. Therefore, our MMCNN possess more powerful representation abilities. Second, a Complementary Classification Strategy (CCS) is proposed, in which the final depth prediction results for each CU are determined by combining the results of multi-classification with the results of binary classification and triplex classification with the voting mechanism. The proposed CCS avoids the accumulation of classification errors and achieves more accurate CTU partition. Extensive experiments demonstrate that our MCCN achieves lower HEVC encoding complexity and more accurate CTU partition: reduce the average encoding complexity by 71.49% only at the cost of a 3.18% average Bjøntegaard Delta Bit-Rate(BD-BR). And the average accuracies of 32×32 CU depth prediction and 16×16 CU depth prediction are increased by 0.65%~0.93% and 2.14%~9.27% respectively.

收稿日期: 2024-03-29; 改回日期: 2024-07-21; 网络出版: 2024-08-03

*通信作者: 唐述 tangshu@cqupt.edu.cn

基金项目: 国家自然科学基金(61601070), 重庆市自然科学基金面上项目(CSTB2023NSCQ-MSX0680), 重庆市教育委员会科学技术研究重大项目(KJZD-M202300101), 重庆邮电大学博士研究生创新人才项目(BYJS202217)

Foundation Items: The National Natural Science Foundation of China (61601070), Chongqing Natural Science Foundation General Project (CSTB2023NSCQ-MSX0680), The Major Project of Science and Technology Research of Chongqing Education Commission (KJZD-M202300101), The Innovative Talents Project for Doctoral students of Chongqing University of Posts and Telecommunications (BYJS202217)

Key words: Deep Neural Networks(DNN); Intra-mode High Efficiency Video Coding(HEVC); Features Representation; Coding Tree Unit(CTU) partition; Multi-scale-multi-input; Complementation classification

1 引言

高效的视频编码(High Efficiency Video Coding, HEVC)^[1]作为新一代的混合编码模型提出了编码树单元(Coding Tree Unit, CTU)和35种预测模式,虽然能够在同等感知视频质量的基础上实现比H.264/AVC视频编码标准^[2]降低约50%的视频码率,但同时也导致HEVC的编码复杂度极大,尤其是针对CTU深度的划分^[3]。现有提升CTU深度划分效率的方法主要分为两大类:基于启发式的方法和基于深度学习的方法。启发式的方法是通过人为提取图像中的纹理特征或者人为设置某些规则来提前终止CTU的深度划分,从而达到降低复杂度的目的。但是这种人为的方式是几乎不可能找到一种普适性的规则来满足所有编码单元(Coding Unit, CU)的。

近年来,卷积神经网络(Convolutional Neural Network, CNN)已经被广泛用于优化HEVC帧内模式的CTU深度划分中。虽然基于CNN的深度预测方法能够通过自主学习CTU中CU的结构来做出对应的划分决策。但是,一方面,现有的基于CNN的深度预测方法几乎都是采用串联的二分类策略(Serial Binary Classification Strategy, SBCS)来决定CTU的划分结果,不可避免地会导致严重的错误累积效应:上层的分类错误会直接影响到下层的分类结果。另一方面,现有方法几乎都只考虑了单一尺度的输入,忽略了不同尺度CU之间特征的相关性。

基于上述分析,针对现有方法的缺陷,本文提出一种多尺度多输入的互补分类网络(Multi-scale-multi-input Complementation Classification Network, MCCN)来实现更高效和更准确的HEVC帧内CTU深度划分。本文所提MCCN能够被端到端地训练,其贡献主要体现在以下几个方面:

(1)本文提出一种新颖的多尺度多输入卷积神经网络(Multi-scale Multi-input Convolutional Neural Network, MMCNN)。同时将多个量化参数(Quantization Parameter, QP)值(22, 27, 32和37)和多个不同尺度的CU图像块作为输入,通过融合不同尺度CU之间的特征信息,进一步提升了网络的特征提取能力和表达能力,极大提升了HEVC帧内CTU深度划分的效率。

(2)不同于现有方法的SBCS,本文提出一种互补的分类策略(Complementary Classification Strategy, CCS)。首先,创建两个MMCNN来分别

预测CTU中每个 32×32 的CU的深度(三分类)和判断每个 16×16 的CU是否需要划分(二分类),然后,综合考虑三分类和二分类的结果,并采用投票机制来决定每个CU的最终深度结果。通过结合多分类和投票机制,CCS中的三分类和二分类很好地实现了分类结果的相互互补,很好地避免了现有方法中存在的错误累积效应,较大提升了CTU中CU深度预测的准确性。

2 相关工作

为了有效降低HEVC帧内编码的复杂度,人们提出了大量的基于启发式的和基于深度学习的提升CTU深度划分效率的方法。

Zhao等人^[4]采用低成本的哈达马成本(Hadamard cost)来作为终止CU划分的标准。Kim等人^[5]提出了一种高效的基于CU中高频关键点数量的CU划分方法。Zhang等人^[6]提出了一种基于梯度的机器学习粗模式决策和RD优化算法。Jamali等人^[7]提出了一种针对帧内模式的率失真优化(Rate Distortion Optimization, RDO)成本预测方法和一种基于梯度的模式筛选方法。Jamali等人采用Prewitt算子来从候选模式列表中筛选出方向相关的模式。Amna等人^[8]使用了一种基于支持向量机(Support Vector Machines, SVM)的方法,通过设置相关特征以有效预测CU的深度划分,降低了HEVC的编码复杂度。Werda等人^[9]提出基于SVM和梯度的快速CU分区和帧内预测模块。

近年来, CNN已经被广泛用于优化HEVC帧内模式的CTU深度划分中。Yu等人^[10]提出了一种基于CNN的二分类CTU深度划分方法,并且对所有深度的CU采用相同的网络架构。Li等人^[11]针对不同深度的CU设计了不同的神经网络以实现更具针对性的CU深度预测。Xu等人^[12]提出了一种早期停止卷积网络(Early Terminated Hierarchical CNN, ETH-CNN)来进行分等级的CU深度地图预测以降低复杂度。Li等人^[13]提出了一种基于CNN的速率失真优化算法。Qin等人^[14]提出了一种改进的视觉几何小组网络(Visual Geometry Group Network, VGGNet)来替代暴力搜索策略,有效降低了帧内编码复杂度。Feng等人^[15]提出了一种面向HEVC帧内编码的CNN快速CTU深度划分方法。Hari等人^[16]采用不同大小的卷积核来预测CTU的深度地图。Lorkiewicz等人^[17]提出了一种预训练的神经网络来预测CTU的深度地图。Feng等人^[18]提出了一种基

于CNN的HEVC帧内深度范围预测网络, 并仅在预测的深度范围内执行率失真代价(Rate Distortion cost, RDcost)计算来减少RDcost的计算次数。Li等人^[19]提出了一种端到端的快速算法, 以辅助帧内编码中CTU的结构划分决策。Imen等人^[20]基于LeNet-5和AlexNet, 有效降低了检查所有块决策候选者的计算时间。Yao等人^[21]构建了一种基于CNN的双网络模型, 在确定最优CU划分模式的同时获得CU分区的最佳RD估计。Linck等人^[22]提出了3个基于VGGNet的CNN来预测HEVC的CU块的四叉树深度。Amna等人^[23]提出了一种基于CNN的方法以降低HEVC编码的复杂性。Jia等人^[24]提出一种编码单元划分和特征向量预测的多层特征传递卷积神经网络, 降低了视频编码的复杂度。Zuo等人^[25]使用线性回归网络预测RDcost, 创建了一个分区模式的映射集, 显著降低了编码复杂性。Wu等人^[26]引入基于频率的神经网络, 通过在频域学习的方式来降低HEVC编码复杂度。

3 本文所提多尺度多输入互补分类网络

本文采用HEVC帧内编码中的通用规定: CTU的大小为 64×64 , 而CTU中CU的大小则根据其深度的不同分别为 64×64 , 32×32 , 16×16 和 8×8 , 其对应的深度分别用0, 1, 2和3来标记。

3.1 互补分类策略(CCS)

为了实现更准确的深度划分, 本文提出一种新颖的互补分类策略(CCS)。首先分别将每个 64×64 的CTU进行 16×16 和 32×32 两种不同的划分, 得到两幅不同的分块图(如图1所示): 本文用 $\text{Image}_{\text{block}16}$ 来表示以 16×16 为单元的划分图; 用 $\text{Image}_{\text{block}32}$ 来表示以 32×32 为单元的划分图。然

后, 针对 $\text{Image}_{\text{block}16}$, 采用本文提出的二分类MMCNN(详见3.2节)对其中的每个 16×16 CU进行是否需要被进一步划分的二分类预测。针对 $\text{Image}_{\text{block}32}$, 采用本文提出的三分类MMCNN(详见3.2节)对其中的每个 32×32 CU进行三分类的深度预测。

在得到了每个 16×16 CU的二分类结果和每个 32×32 CU的三分类结果之后, 本文综合考虑二分类和三分类的结果, 采用投票机制来决定 64×64 的CTU中每个CU的最终深度。本文投票机制的算法流程为: 对于每个 64×64 的CTU而言, (1)首先参考 $\text{Image}_{\text{block}32}$ 中每个 32×32 CU的三分类结果, 如果其中的3个或3个以上 32×32 CU的深度都被预测为0, 则认为这整个 64×64 的CTU的深度也为0, 即不需要划分; 否则判定该 64×64 的CTU会被划分为4个 32×32 的CU。(2)如果该 64×64 的CTU被判定为需要划分, 那么针对 $\text{Image}_{\text{block}32}$ 中深度分类为0或1的 32×32 而言, 本文将再查看 $\text{Image}_{\text{block}32}$ 中对应位置上4个 16×16 CU的二分类结果, 如果4个 16×16 块的二分类结果中有且仅有1个 16×16 块的结果是需要被划分, 则忽略该 16×16 块的结果, 判定该 32×32 的CU不需要被划分; 否则判定该 32×32 的CU需要被划分为4个 16×16 的CU, 且其中每个 16×16 CU是否需要被划分也是由二分类MMCNN中相同位置上 16×16 CU的二分类结果来决定。(3)如果该 64×64 的CTU被判定为需要被划分, 那么针对 $\text{Image}_{\text{block}32}$ 中深度预测为2的 32×32 的CU而言, 则判定该 32×32 的CU是肯定要被划分为4个 16×16 的CU, 且其中每个 16×16 CU是否需要被划分也是由二分类MMCNN中相同位置上 16×16 CU的二分类结果来决定。

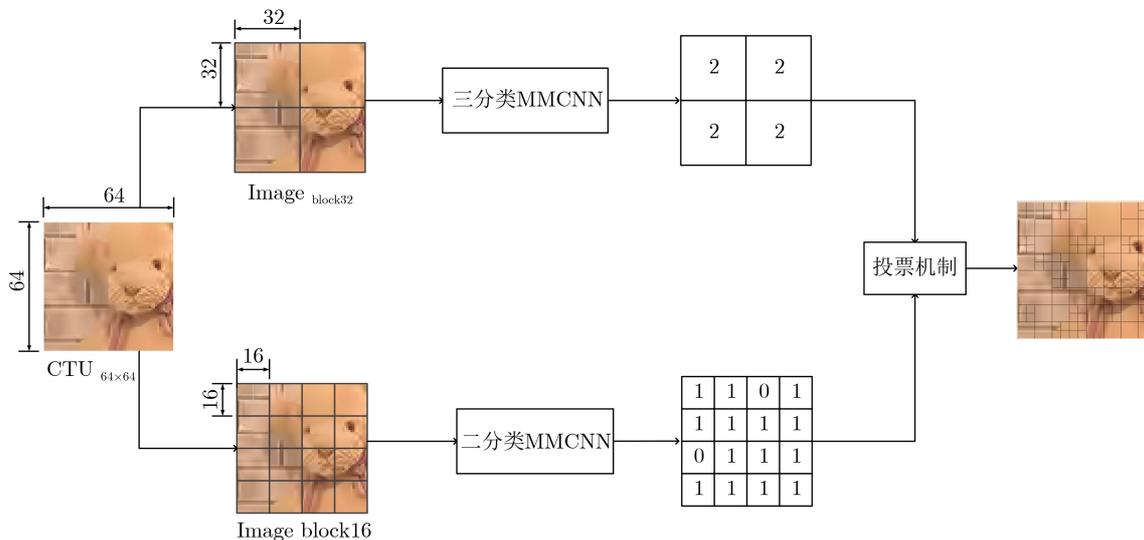


图1 本文所提CCS

由以上分析可知，本文提出的CCS采用了投票机制来融合二分类和三分类的预测结果，很好实现了多种分类结果的互补，可以综合利用这两种分类信息来有效降低单个网络的分类错误对最终结果的影响，很好地避免了现有方法中存在的错误累积效应，较大提升了CTU深度划分的准确性。

3.2 多尺度多输入卷积神经网络(MMCNN)

除了以上的分类策略之外，现有的基于CNN的CTU深度划分方法还忽略了不同尺度CU之间的特征关联。基于此，本文提出一种多尺度多输入卷积神经网络(MMCNN)，并配合CCS，设计了两种MMCNN：针对 16×16 CU的二分类MMCNN(如图2所示)和针对 32×32 CU的三分类MMCNN(如图3所示)。接下来，将以二分类MMCNN为例进行详细的论述。

如图2所示，首先，在输入端二分类MMCNN采用了多尺度多输入的策略：共3级输入，每一级分别对应不同尺寸的CU：将 64×64 的CU输入到第1级，同时将该 64×64 CU中所有的 32×32 CU和所有的 16×16 CU分别输入到第2级和第3级。并且在第1级的输入中，本文还将不同的QP分别与 64×64 的CU进行拼接：每个QP值均与同一个 64×64 的CU进行通道维度的拼接，组成一个 $64 \times 64 \times 2$ 的张量输入到第1级中。本文同时考虑了4个不同的

QP值，即每个 64×64 的CU分别与不同的QP值进行通道维度的拼接，依次输入到网络的第1级中。

其次，对于具体的网络结构而言，在第1级中，针对 64×64 的CU，本文首先采用了一个步长为1、卷积核为 5×5 的卷积层来提取 64×64 CU中的浅层特征，并在该卷积层之后执行1次 2×2 的最大池化操作来提取每个 2×2 区域内最显著的特征，在保留显著特征的同时实现降维，将特征图的分辨率降为 32×32 ；然后再采用2个 3×3 的卷积层来进一步提取细节特征，其中，第1个 3×3 卷积层的步长为1，第2个 3×3 卷积层的步长为2，再次将特征图的分辨率降为 16×16 。在第2级中，首先采用一个步长为1的 3×3 的卷积层来提取 32×32 CU的特征，并且同样在卷积层之后执行1次 2×2 的最大池化操作，将第2级的特征图的分辨率降为 16×16 ，并将提取的 32×32 CU中的显著特征与第1级的 16×16 特征图在通道维度上进行拼接，然后，将拼接后的特征图再经过1个 3×3 的卷积层和1个 2×2 的最大池化操作，将拼接特征图的分辨率降为 8×8 ，实现 64×64 CU特征与 32×32 CU特征的融合。在第3级中，首先采用一个步长为1的 3×3 卷积层和一个 2×2 的最大池化层将输入的 16×16 CU映射为 8×8 的特征图，并与第2级中的 8×8 特征图再次进行通道维度上的拼接，然后将再次拼接后的特征图经过2个

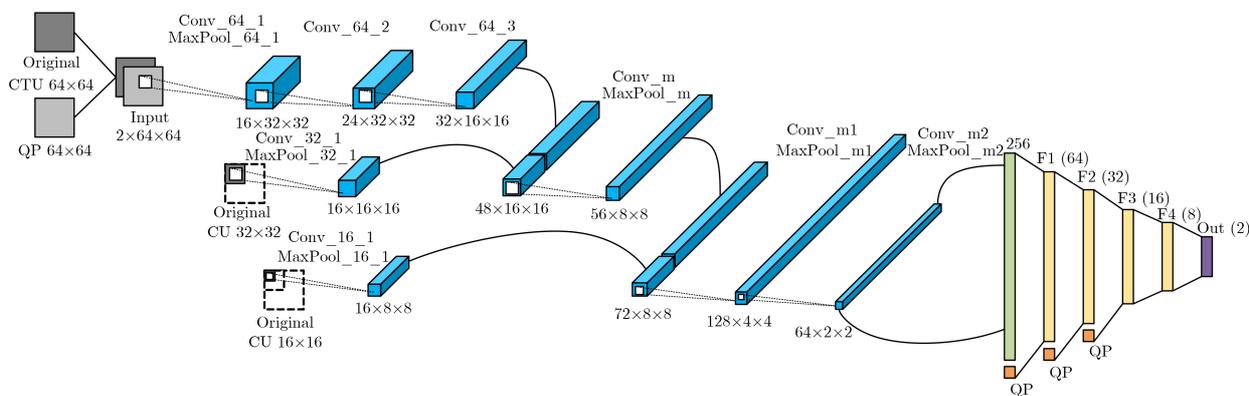


图2 二分类MMCNN

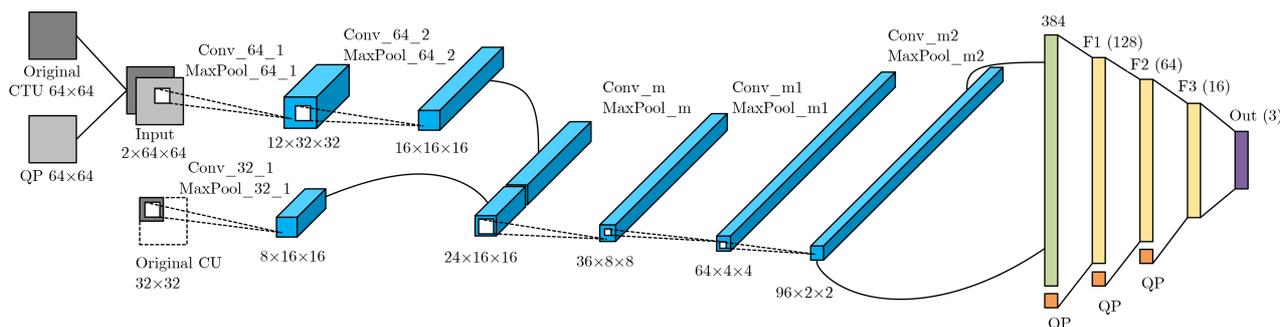


图3 三分类MMCNN

步长为1的 3×3 卷积层和2个 2×2 的最大池化操作,实现 64×64 CU特征、 32×32 CU特征和 16×16 CU特征的融合,最终得到一个维度为 $2\times 2\times 64$ 的特征图。在以上基于卷积的特征提取过程中,每个卷积层都采用参数化修正单元(Parametric Rectified Linear Unit, PReLU)来作为非线性激活函数来抑制特征中的负值成份。在完成了基于卷积的3级特征提取和融合之后,本文采用5个连续的全连接层对 $2\times 2\times 64$ 的特征图进行进一步的特征提取和输出最终的二分类结果。5个全连接层由4个隐藏层和1个输出层组成,针对前3个隐藏层,将输入的QP值与每个隐藏层的输出特征拼接到一起,一并输入到下一个隐藏层。

三分类MMCNN与二分类MMCNN十分相似,其区别仅是三分类MMCNN面向的是 32×32 的CU,同时仅采用了4个全连接层且最后的输出为三分类。二分类MMCNN和三分类MMCNN均采用交叉熵作为其损失函数,如式(1)所示

$$L = -\frac{1}{N} \sum_{n=1}^N [y_i^n \log_2(p_i^n)] \quad (1)$$

其中,在二分类MMCNN中, N 表示 16×16 CU的数量, y_i^n 和 p_i^n 分别表示每个 16×16 CU真实的类别和网络输出的类别。在三分类MMCNN中, N 表示 32×32 CU的数量, y_i^n 和 p_i^n 分别表示每个 32×32 CU真实的类别和网络输出的类别。对于二分类MMCNN, $i\in[0,1]$,对于三分类MMCNN, $i\in[1,2,3]$ 。

4 实验分析

4.1 数据集和评价指标

本文的实验中采用了数据集CPIH和JCT-VC,并考虑了4种不同的QP值:22,27,32和37。本文采用分类准确率、编码时间节省率 ΔT 、增量比特率(Bjontegaard Delta Bit-Rate, BD-BR)和增量峰值信噪比(Bjontegaard Delta Peak Signal Noise Ratio, BD-PSNR)4种指标来定量评估本文所提MCCN的性能。

ΔT 表示本文所提方法相对于HEVC的时间节省率, ΔT 的计算如式(2)所示

$$\Delta T = \frac{T_{\text{HEVC}} - T_{\text{test}}}{T_{\text{HEVC}}} \quad (2)$$

其中, T_{HEVC} 是HEVC编码所需的时间, T_{test} 是本文采用方法所需的时间。

BD-BR的计算如式(3)所示

$$\text{BD-BR} = \frac{1}{\Delta D} \int_{D_1}^{D_h} (r_{\text{test}} - r_{\text{HEVC}}) dD \quad (3)$$

其中, D_h 和 D_1 分别为亮度范围的最大值和最小值, $\Delta D=D_h-D_1$; r_{test} 和 r_{HEVC} 分别表示测试方法和原始HEVC方法的对应比特率。

BD-PSNR的计算如式(4)所示

$$\text{BD-PSNR} = \frac{1}{\Delta r} \int_{r_1}^{r_h} (D_{\text{test}} - D_{\text{HEVC}}) dr \quad (4)$$

其中, r 表示码率, $\Delta r=r_h-r_1$; D_{test} 和 D_{HEVC} 分别表示测试方法和HEVC方法的对应PSNR值

4.2 实验设置

在训练阶段,针对二分类MMCNN和三分类MMCNN采用相同的实验设置:训练共执行200次epoch,批次大小设置为1024,采用随机梯度下降优化器(Stochastic Gradient Descent, SGD),动量设置为0.9,初始学习率为0.01,并在每个epoch之后以固定步长 0.5×10^{-4} 进行下降。本文的所有实验都是在Intel(R) Core(TM) i7-7820X CPU @ 3.60 GHz和NVIDIA GeForce RTX 2080 Ti的windows10 64位操作系统的计算机,以及Pytorch深度学习框架上进行训练和测试的。

4.3 本文所提方法的消融实验

MMCNN的有效性消融实验:为了能够准确评估本文所提MMCNN的有效性,本文创建了两个模型:MMCNN-NoQP和MMCNN-OneScale。其中,MMCNN-NoQP不会将QP值输入到网络中;MMCNN-OneScale则表示在二分类MMCNN-OneScale和三分类MMCNN-OneScale中均只考虑第1级的输入。

本文在JCT-VC的18个标准视频序列上将MCCN-NoQP,MCCN-OneScale和MCCN的平均BD-BR和平均 ΔT 进行比较实验,其实验结果如表1所示。如表1中的第1行和第2行所示,MCCN-NoQP在平均BD-BR和平均 ΔT 两个方面均落后于MCCN:不仅平均BD-BR增加了6.62%,而且需要消耗更多的编码时间:平均 ΔT 减少了12.89%。如表1中的第1行和第3行所示,MCCN-OneScale在平均BD-BR和平均 ΔT 两个方面也都落后于MCCN:平均BD-BR增加了8.32%,而平均 ΔT 反而减少了7.19%。表1很好地证明了本文所提MMCNN的有效性。

CCS的有效性消融实验:移除CCS,将现有的

表1 MMCNN的有效性消融实验(%)

模型	平均BD-BR	平均 ΔT
MCCN	3.18	71.49
MCCN-NoQP	9.80	58.60
MCCN-OneScale	11.50	64.30

串联二分类策略(SBCS)应用到MMCNN中：采用图2的二分类MMCNN来分别依次对64×64, 32×32和16×16大小的CU进行是否需要划分的二分类决策：首先训练一个针对64×64CU的二分类MMCNN(命名为：二分类MMCNN64)来判断64×64的CU是否需要划分，然后再训练一个针对32×32 CU的二分类MMCNN32来判断32×32的CU是否需要划分，最后训练一个针对16×16 CU的二分类MMCNN16来判断16×16的图像块是否需要划分。将二分类MMCNN+SBCS命名为MCCN-SBCS，因此，比较MCCN-SBCS与MCCN之间的性能差异能够准确评估本论文提出的CCS的有效性。

在CPIH的100张测试集上，将本文提出的MCCN与MCCN-SBCS在64×64 CU, 32×32 CU和16×16 CU上的深度划分准确率的平均值进行了比较。如表2所示，本文提出的MCCN在64×64 CU, 32×32 CU和16×16 CU上的深度划分平均准确率都高于MCCN-SBCS的平均准确率，分别提升了：2.25%，1.04%和3.98%。由此可见，本文提出的CCS通过结合多分类和投票机制，很好地实现了三分类和二分类结果的相互互补，有效避免了现有方法中存在的错误累积效应，能够较大提升深度预测的准确性。

除了深度划分的准确率之外，本文还在JCT-VC上将所提MCCN与MCCN-SBCS在平均BD-BR和平均ΔT上进行了比较，如表3所示。由表3可见，MCCN-SBCS在平均BD-BR和平均ΔT两个方面也都落后于MCCN：平均BD-BR增加了4.68%，平均ΔT反而减少了4.49%。由此可见，本文提出的CCS不仅能够有效提升深度预测的准确性，还能有助于提升HEVC帧内CTU深度划分的效率。表2和表3很好地证明了本文提出的CCS和投票机制的有效性。

4.4 本文所提方法与前沿方法的比较

为了进一步验证本文所提方法的优越性，本节

首先将提出的MCCN与文献[12]和文献[15]在100张CPIH测试集上进行深度划分准确率的比较(如表4所示)；然后，再将提出的MCCN与近几年最具代表性的7种方法：文献[11–17]在JCT-VC上进行BD-BR和ΔT的比较(如表5所示)。

如表4所示，虽然本文提出的MCCN对64×64 CU的深度预测平均准确率略低于文献[12]和文献[15]，分别低了0.68%和0.88%，但是在32×32 CU的深度预测和16×16 CU的深度预测上，本文提出MCCN的平均准确率均高于文献[12]和文献[15]，分别提升了0.65%~0.93%和2.14%~9.27%。因此，该现象恰恰很好的证明了本文提出的MCCN中的投票机制能够有效避免现有方法中存在的错误累积效应：大块的分类结果不会影响到对应小块的深度划分。

表5所示为文献[11–17]的方法和本文提出的MCCN在JCT-VC所有的18个标准视频测试集上的平均BD-BR和平均比较。由表5可见，本文提出的MCCN具有最低的平均编码复杂度：仅以增加3.18%的平均BD-BR为代价，降低了71.49%的平均编码复杂度。特别地，首先本文提出的MCCN在平均BD-BR和平均ΔT两方面均优于文献[16]，不仅平均ΔT提高了1.1%，而且BD-BR也下降了1.09%。其次，与文献[11–15, 17]相比，虽然本文提出的MCCN的编码复杂度实现了大幅的下降，平均ΔT分别提高了9.24%，9.64%，7.44%，11.78%，5.94%和10.86%。表4和表5很好地证明了本文提出的MCCN的优越性。

表 3 MCCN和MCCN-SBC的平均BD-BR和平均ΔT比较(%)

模型	平均BD-BR	平均ΔT
MCCN	3.18	71.49
MCCN-SBCS	7.86	67.00

表 4 平均准确率比较(其中最好的性能已加粗标记)(%)

	本文	文献[12]	文献[15]
64 × 64 CU	90.30	90.98	91.18
32 × 32 CU	87.55	86.62	86.90
16 × 16 CU	89.69	80.42	87.55

表 2 消融实验：不同大小CU的深度划分平均准确率比较(%)

	MCCN	MCCN-SBCS
64 × 64 CU	90.30	88.05
32 × 32 CU	87.55	86.51
16 × 16 CU	89.69	85.71

表 5 所有方法在JCT-VC标准视频测试集上的平均BD-BR和平均ΔT比较(其中最好的性能已加粗标记)(%)

指标	本文方法	2017年	2018年	2023年	2022年	2021年	2022年	2021年
		文献[11]	文献[12]	文献[13]	文献[14]	文献[15]	文献[16]	文献[17]
BD-BR	3.18	2.21	2.25	1.94	2.04	2.02	4.27	1.81
ΔT	71.49	62.25	61.85	64.05	59.71	65.55	70.39	60.63

5 结束语

本文提出了一种多尺度多输入的互补分类网络(MCCN),能够在极大降低HEVC帧内编码复杂度的同时实现更准确的CTU深度划分。不同于现有的基于神经网络的学习方法,首先将多个QP值和多个不同尺度的CU图像块同时输入到网络中,提出了一种MMCNN来实现更高效的CTU深度划分,同时还拥有更强的QP值泛化能力。其次,提出了一种CCS,通过结合多分类和投票机制,很好地避免了现有方法中存在的错误累积效应,实现了对不同尺寸CU更准确的深度预测。大量的实验结果表明,本文提出的MCCN仅以增加3.18%的平均比特率为代价就降低了71.49%的平均编码复杂度,同时,预测的 32×32 CU和 16×16 CU的深度准确率分别平均提升了0.65%~0.93%和2.14%~9.27%。将本文提出的方法扩展到其它的编码方法中,如:H.266/VVC等,将是本研究小组未来的工作重点。

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. doi: [10.1109/TCSVT.2012.2221191](https://doi.org/10.1109/TCSVT.2012.2221191).
- [2] WIEGAND T, SULLIVAN G J, BJONTEGAARD G, *et al.* Overview of the H. 264/AVC video coding standard[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2003, 13(7): 560–576. doi: [10.1109/TCSVT.2003.815165](https://doi.org/10.1109/TCSVT.2003.815165).
- [3] POURAZAD M T, DOUTRE C, AZIMI M, *et al.* HEVC: The new gold standard for video compression: How does HEVC compare with H. 264/AVC?[J]. *IEEE Consumer Electronics Magazine*, 2012, 1(3): 36–46. doi: [10.1109/MCE.2012.2192754](https://doi.org/10.1109/MCE.2012.2192754).
- [4] ZHAO Liang, FAN Xiaopeng, MA Siwei, *et al.* Fast intra-encoding algorithm for high efficiency video coding[J]. *Signal Processing: Image Communication*, 2014, 29(9): 935–944. doi: [10.1016/j.image.2014.06.008](https://doi.org/10.1016/j.image.2014.06.008).
- [5] KIM N, JEON S, SHIM H J, *et al.* Adaptive keypoint-based CU depth decision for HEVC intra coding[C]. 2016 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), Nara, Japan, 2016: 1–3. doi: [10.1109/BMSB.2016.7521923](https://doi.org/10.1109/BMSB.2016.7521923).
- [6] ZHANG Tao, SUN Mingting, ZHAO Debin, *et al.* Fast intra-mode and CU size decision for HEVC[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2017, 27(8): 1714–1726. doi: [10.1109/TCSVT.2016.2556518](https://doi.org/10.1109/TCSVT.2016.2556518).
- [7] JAMALI M and COULOMBE S. Fast HEVC intra mode decision based on RDO cost prediction[J]. *IEEE Transactions on Broadcasting*, 2019, 65(1): 109–122. doi: [10.1109/TBC.2018.2847464](https://doi.org/10.1109/TBC.2018.2847464).
- [8] AMNA M, IMEN W, NACIR O, *et al.* SVM-Based method to reduce HEVC CU partition complexity[C]. 2022 19th International Multi-Conference on Systems, Signals & Devices (SSD), Sétif, Algeria, 2022: 480–484. doi: [10.1109/SSD54932.2022.9955731](https://doi.org/10.1109/SSD54932.2022.9955731).
- [9] WERDA I, MARAOUI A, SAYADI F E, *et al.* Fast CU partition and intra mode prediction method for HEVC[C]. 2022 IEEE 9th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), Hammamet, Tunisia, 2022: 562–566. doi: [10.1109/SETIT54465.2022.9875798](https://doi.org/10.1109/SETIT54465.2022.9875798).
- [10] YU Xianyu, LIU Zhenyu, LIU Junjie, *et al.* VLSI friendly fast CU/PU mode decision for HEVC intra encoding: Leveraging convolution neural network[C]. 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, Canada, 2015: 1285–1289. doi: [10.1109/ICIP.2015.7351007](https://doi.org/10.1109/ICIP.2015.7351007).
- [11] LI Tianyi, XU Mai, and DENG Xin. A deep convolutional neural network approach for complexity reduction on intra-mode HEVC[C]. 2017 IEEE International Conference on Multimedia and Expo (ICME), Hong Kong, China, 2017: 1255–1260. doi: [10.1109/ICME.2017.8019316](https://doi.org/10.1109/ICME.2017.8019316).
- [12] XU Mai, LI Tianyi, WANG Zulin, *et al.* Reducing complexity of HEVC: A deep learning approach[J]. *IEEE Transactions on Image Processing*, 2018, 27(10): 5044–5059. doi: [10.1109/TIP.2018.2847035](https://doi.org/10.1109/TIP.2018.2847035).
- [13] LI Huayu, WEI Geng, WANG Ting, *et al.* Reducing video coding complexity based on CNN-CBAM in HEVC[J]. *Applied Sciences*, 2023, 13(18): 10135. doi: [10.3390/app131810135](https://doi.org/10.3390/app131810135).
- [14] QIN Liming, ZHU Zhongjie, BAI Yongqiang, *et al.* A complexity-reducing HEVC intra-mode method based on VGGNet[J]. *Journal of Computers*, 2022, 33(4): 57–67. doi: [10.53106/199115992022083304005](https://doi.org/10.53106/199115992022083304005).
- [15] FENG Aolin, GAO Changsheng, LI Li, *et al.* Cnn-based depth map prediction for fast block partitioning in HEVC intra coding[C]. 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 2021: 1–6. doi: [10.1109/ICME51207.2021.9428069](https://doi.org/10.1109/ICME51207.2021.9428069).
- [16] HARI P, JADHAV V, and RAO B K N S. CTU partition for intra-mode HEVC using convolutional neural network[C]. 2022 IEEE International Symposium on Smart Electronic Systems (ISES), Warangal, India, 2022: 548–551. doi: [10.1109/ISES54909.2022.00120](https://doi.org/10.1109/ISES54909.2022.00120).
- [17] LORKIEWICZ M, STANKIEWICZ O, DOMANSKI M,

- et al.* Fast selection of INTRA CTU partitioning in HEVC encoders using artificial neural networks[C]. 2021 Signal Processing Symposium (SPSymo), LODZ, Poland, 2021: 177–182. doi: [10.1109/SPSymo51155.2020.9593483](https://doi.org/10.1109/SPSymo51155.2020.9593483).
- [18] FENG Zeqi, LIU Pengyu, JIA Kebin, *et al.* HEVC fast intra coding based CTU depth range prediction[C]. 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, China, 2018: 551–555. doi: [10.1109/ICIVC.2018.8492898](https://doi.org/10.1109/ICIVC.2018.8492898).
- [19] LI Yixiao, LI Lixiang, FANG Yuan, *et al.* Bagged tree and ResNet-based joint end-to-end fast CTU partition decision algorithm for video intra coding[J]. *Electronics*, 2022, 11(8): 1264. doi: [10.3390/electronics11081264](https://doi.org/10.3390/electronics11081264).
- [20] IMEN W, AMNA M, FATMA B, *et al.* Fast HEVC intra-CU decision partition algorithm with modified LeNet-5 and AlexNet[J]. *Signal, Image and Video Processing*, 2022, 16(7): 1811–1819. doi: [10.1007/s11760-022-02139-w](https://doi.org/10.1007/s11760-022-02139-w).
- [21] YAO Chao, XU Chenming, and LIU Meiqin. RDNet: Rate-distortion-based coding unit partition network for intra-prediction[J]. *Electronics*, 2022, 11(6): 916. doi: [10.3390/electronics11060916](https://doi.org/10.3390/electronics11060916).
- [22] LINCK I, GOMEZ A T, and ALAGHBAND G. CNN quadtree depth decision prediction for block partitioning in HEVC intra-mode[C]. 2023 Data Compression Conference (DCC), Snowbird, USA, 2023: 352. doi: [10.1109/DCC55655.2023.00054](https://doi.org/10.1109/DCC55655.2023.00054).
- [23] AMNA M, IMEN W, and EZAHRA S F. Deep learning for intra frame coding[C]. 2021 International Conference on Engineering and Emerging Technologies (ICEET), Istanbul, Turkey, 2021: 1–4. doi: [10.1109/ICEET53442.2021.9659742](https://doi.org/10.1109/ICEET53442.2021.9659742).
- [24] 贾克斌, 崔腾鹤, 刘鹏宇, 等. 基于深层特征学习的高效率视频编码中帧内快速预测算法[J]. *电子与信息学报*, 2021, 43(7): 2023–2031. doi: [10.11999/JEIT200414](https://doi.org/10.11999/JEIT200414).
- JIA Kebin, CUI Tenghe, LIU Pengyu, *et al.* Fast prediction algorithm in high efficiency video coding intra-mode based on deep feature learning[J]. *Journal of Electronics & Information Technology*, 2021, 43(7): 2023–2031. doi: [10.11999/JEIT200414](https://doi.org/10.11999/JEIT200414).
- [25] ZUO Yanchen, GAO Changsheng, LIU Dong, *et al.* Learned rate-distortion cost prediction for ultrafast screen content intra coding[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024, 34(3): 1976–1980. doi: [10.1109/TCSVT.2023.3296515](https://doi.org/10.1109/TCSVT.2023.3296515).
- [26] WU Yi and CHEN Lei. Fast algorithm for HEVC using frequency-based convolutional neural networks[C]. 2023 3rd International Conference on Electronic Information Engineering and Computer (EIECT), Shenzhen, China, 2023: 559–563. doi: [10.1109/EIECT60552.2023.10442731](https://doi.org/10.1109/EIECT60552.2023.10442731).
- 唐述：男，副教授，研究方向为视频信号处理、低水平视觉任务、图像超分辨率重建、模糊图像复原。
- 周广义：男，硕士，研究方向为视频信号处理。
- 谢显中：男，教授，研究方向为信号与信息处理、计算机通信方向、通信与信息系统。
- 赵瑜：男，硕士，研究方向为视频信号处理。
- 杨书丽：女，博士，研究方向为图像超分辨率重建。
- 责任编辑：余蓉