一种快速的多尺度多输入CTU互补分类网络

摘 要: 近年来，为了有效降低HEVC的编码复杂度，深度神经网络（deep neural networks：DNN）被广泛应用到HEVC编码树单元（Coding Tree Unit：CTU）的深度划分中，并取得了较大的进步。然而现有的基于DNN的CTU深度划分方法却忽略了不同尺度CU间的特征相关性和存在着分类错误累积等缺陷，因此，现有方法的性能仍有较大提升空间。基于此，本论文提出了一种多尺度多输入的互补分类网络（Multi-scale-multi-input Complementation Classification Network：MCCN）来实现更高效且更准确的HEVC帧内CTU深度划分。首先，我们提出了一种多尺度多输入的卷积神经网络（Multi-scale Multi-input Convolutional Neural Network：MMCNN），将多个不同的QP值和多个不同尺度的图像块同时输入到MMCNN中，并通过融合不同尺度CU的特征信息建立不同尺度CU间特征的关联，进一步提升网络的特征提取能力和表达能力，达到极大降低HEVC编码复杂度的目的。然后，我们提出了一种互补的分类策略（Complementary Classification Strategy：CCS）。在提出的CCS中，我们构建了两种MMCNN来分别对CTU中的每个的编码单元（Coding Unit：CU）进行三分类的深度预测和对每个的CU进行是否需要划分的二分类预测。通过结合二分类和三分类的结果，并采用投票机制来决定CTU中每个CU的最终深度值。CCS中的三分类和二分类充分实现了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，因此实现了更准确的CTU深度划分。两个MMCNN和CCS组成了本论文的MCCN。大量的实验结果表明，与近几年极具代表性的方法相比，本论文提出的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BD-BR（Bjøntegaard delta bit-rate）为代价，降低了71.49%的平均编码复杂度。同时，预测的CU和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HEVC; CTU深度划分;多尺度多输入;互补分类

中图法分类号: TN919.8

A Multi-scale-multi-input Complementation Classification Network for Fast CTU Partition

**Abstract**: Recently, deep neural networks (DNN) have been widely applied to Coding Tree Unit (CTU) partition of intra-mode 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. So, the performance of existing methods still has rooms for improvement. Therefore, in this paper, we propose a multi-scale-multi-input complementation classification network (MCCN) for faster and more accurate CTU partition at intra-mode HEVC. First, we propose a multi-scale multi-input convolutional neural network (MMCNN), which takes multiple quantization parameters (QP) and multi-scale image patches as inputs, and builds up the correlation of features between CUs at different scales by fusing multi-scale CU features. Therefore, our MMCNN possess more powerful abilities of features capture and representation, and can reduce encoding complexity of HEVC greatly. Second, we propose a complementary classification strategy (CCS), in which we construct two MMCNNs to predict the depth of each CU (multi-classification) and determine whether to split a CU (binary classification), respectively. And the final depth prediction results for each CU are determined by combining the results of multi-classification with the results of binary classification with the voting mechanism. In our CCS, binary classification and multi-classification play complementary roles from each other, which avoids the accumulation of classification errors and achieves more accurate CTU partition. Two MMCNNs and the CCS compose our MCCN. Extensive experiments demonstrate that compared with the state-of-the-art (SOTA) methods, our proposed 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 CU depth prediction and CU depth prediction are increased by 0.65%-0.93% and 2.14%-9.27% respectively.

**Key words**: deep neural networks; intra-mode HEVC; CTU Partition; multi-scale-multi-input; complementation classification.;

# 引言

高效的视频编码（High Efficiency Video Coding，HEVC）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，通过递归地计算所有深度上所有图像块的率失真代价（Rate-Distortion cost：RD cost）来决定每个编码单元（Coding Unit ：CU)的深度。虽然如此的方式能够使得HEVC在同等感知视频质量的基础上比H.264/AVC视频编码标准降低约50%的视频码率，但同时也导致HEVC的编码复杂度变得极大，尤其是针对CTU深度的划分[1]-[4]。因此，如何能够在保证较低BD-BR（Bjøntegaard delta bit rate）的同时尽可能提升CTU深度划分的效率已经成为了近年来人们研究的热点。

现有的提升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（Heuristic-based Methods）和基于深度学习的方法（Deep-learning-based Methods）。对于启发式的方法而言，几乎都是通过人为提取图像中的纹理信息特征或者人为设置某些规则来提前终止CTU的深度划分，以避免暴力递归的RDO（Rate Distortion Optimization）搜索，从而达到降低深度划分的复杂度的目的，例如：利用CTU中每个CU的纹理特征以及其与相邻CU的空间相关性来决定CTU的划分结果[5]-[18]。虽然这类方法能够在一定程度上降低HEVC编码的复杂度，但是自然图像中的内容是千变万化的，不同CU之间的纹理特征以及它们之间的关系是十分复杂的，这种人为提取CU纹理特征并人为设置相关规则的方式是几乎不可能找到一种普适性的规则来满足所有CU的。其次，虽然基于启发式的方法能够在一定程度上降低HEVC的编码复杂度，但其性能仍有较大提升空间。

近年来，深度学习神经网络，尤其是卷积神经网络（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了一种串联的二分类策略（Serial Binary Classification Strategy：SBCS）来决定CTU的划分结果，即对每个深度层的CU分别训练一个二分类网络：首先针对的图像块训练一个二分类网络来判断的图像块是否需要划分，然后再训练一个针对图像块的二分类网络来判断的图像块是否需要划分，最后再训练一个针对图像块的二分类网络来判断的图像块是否需要划分。显而易见，如此的串联二分类策略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[19]-[27]。另一方面，现有的基于CNN的CTU深度划分方法几乎都只考虑了单一尺度的输入，而忽略了不同尺度CU之间特征的相关性[23][24]，然而，众所周知，小尺度CU的结构特征会对包含了该小尺度CU的大尺度CU的深度划分产生极大的影响。因此，现有方法的性能仍有较大提升空间。

基于上述分析，针对现有方法的缺陷，本论文提出了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。本论文提出的MCCN能够被端到端的训练，其贡献主要体现在以下三个方面：

1、本论文提出了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）。提出的MMCNN同时考虑了将多个QP值（本论文考虑了22、27、32和37四个不同的QP值）和多个不同尺度的CU图像块作为输入，通过对不同尺度CU之间特征信息的有效融合，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、不同于现有方法的串联二分类策略，本论文将多分类融入到CU的深度预测中，提出了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中每个的CU的深度（三分类）和判断每个的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的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。

3、大量的实验结果表明，与近年来极具代表性的先进方法相比，本论文提出的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BD-BR（Bjøntegaard delta bit-rate）为代价，降低了71.49%的平均编码复杂度。同时，CU和CU的深度预测准确率分别提升了0.65%-0.93%和2.14%-9.27%。

# 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。基于启发式的方法通常是通过人为提取CTU中CU的特征或者人为设置某些规则来提前终止CTU的深度划分过程，以避免暴力递归的RDO搜索，从而达到降低深度划分复杂度的目的[5]-[16]。在文献[5]中，Gweon等人通过在亮度和色度的所有编码块标志Coded Block Flags（CBF）为零的情况下跳过对当前CU深度的RDcost计算来减少计算复杂度。Choi等人通过判断当前CU深度下的最佳预测模式是否为SKIP来提前结束对该CU深度的划分[6]。Yang等人通过采用差分运动矢量和编码块标志两个特征作为CU是否划分的标准来降低编码器的复杂度[7]。Kim等人使用了具有低复杂度和低RDcost的贝叶斯决策规则来对CU的划分做出判决[8]。Zhang等人采用时空相关性进行自适应的CU深度搜索，从而降低CTU划分的复杂度[9]。Zhao等人采用低成本的Hadamard cost来作为终止CU划分的标准[10]。Kim等人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[11]。Khan等人提出了一种基于视频帧内容自适应的快速CU决策方法：根据视频帧中不同局部区域的内容，自适应的将较小的CU组合成较大的CU[12]。Xiong等人提出了一种基于金字塔运动发散的模型来快速决定CU是否需要划分[13]。在文献[14]中，Zhang等人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。为了实现快速的CU划分决策，Zhang等人提出了两个以深度差和RD代价为特征的线性支持向量机来进行早期CU分裂和早期CU终止决策。Hu等人基于离散余弦变换系数透明复合模型，将帧内编码模式判决问题转换成贝叶斯判决问题，提出了一种基于离群点的快速帧内模式判决算法[15]。Liu等人提出了一种基于支持向量机的三分类器来判断CU是否需要划分[16]。Jamali等人提出了一种针对帧内模式的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[18]。由以上分析可知，现有的基于启发式的方法几乎都是通过人为手动的对图像中某些成分的特征（如边缘、纹理）来设置一些规则，使得满足这些特征的CU可以提前终止划分，来达到降低HEVC帧内编码复杂度的目的。然而自然图像中边缘和纹理是千变万化的，不同CU之间的边缘/纹理特征以及它们之间的关系是十分复杂的，这种人为设置相关规则的方式是几乎不可能满足所有的CU的。因此，虽然启发式的方法能够在一定程度上降低HEVC的编码复杂度，但其性能仍有较大提升空间。

近年来，深度学习神经网络[26]-[32][35]-[39]，尤其CNN，因为其强大的特征提取能力和非线性表达能力，已经被广泛用于优化HEVC帧内模式的CTU深度划分中[40]-[49]。Yu等人提出了一种基于CNN的二分类CTU深度划分方法：从初始的图像块开始，判断其是否需要划分，如需要划分则继续对下一层的每个图像块进行二分类判断，如不需要划分则直接终止对该层CU的划分，并且对所有深度的CU采用相同的网络架构[22]。在文献[23]中，Li等人针对不同深度的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，但是Li等人却忽略了不同尺寸CU间的关联性。Xu等人提出了一种ETH-CNN（Early Terminated Hierarchical-CNN）来进行分等级的CU深度地图预测，以降低帧内模式HEVC的复杂度[24]。Kuanar等人将空间模式分为四类，提出了一种CNN来检测CU中的纹理和对象形状[25]。Li等人提出了一种基于CNN的速率失真优化算法，通过结合混合注意力机制模块，以优化HEVC中CTU划分的RDO。首先，Li等人设计了一个与当前CU大小相兼容的CNN来预测CU的深度划分。其次，他们设计了一个卷积块，以增强CU块之间的信息交互，并将卷积块注意力模块（Convolution Block Attention Module:CBAM）引入到CNN中，构建了CNN-CBAM。该模块能够正确关注目标对象，提高CU划分的准确率[26]。针对CU深度划分的复杂递归计算，Qin等人同样提出了一种基于CNN的CU划分方法：设计了一种改进的VGGNet来替代暴力搜索策略，有效降低了帧内编码的计算复杂度[29]。Fan等人提出了一种采用块划分神经网络来预测CTU的块划分结构的方法。Fan等人采用自适应阈值技术来准确管理CNN的预测误差[31]。Wang等人提出了一种基于密集连接卷积神经网络（Densely Connected Convolution Neural Network:D-CNN）来的CU深度划分方法。Wang等人首先设计了一个密集连接块来充分提取CTU的像素特征；然后，Wang等人采用有效的通道注意力（Efficient Channel Attention: ECA）和自适应卷积核大小应来捕获D-CNN卷积通道的特征信息；最后，Wang等人制定了一种阈值优化策略，通过为每个深度选择一个最佳阈值来平衡视频编码的计算复杂性和RD性能[35]。Feng等人提出了一种面向HEVC帧内编码的CNN快速CTU深度划分方法，但是该方法仅以单个尺度的CTU作为输入，导致网络难以学习到小尺寸CU的特征[36]。Hari等人采用不同大小的卷积核（、、）来预测CTU的深度地图[37]。Lorkiewicz等人提出了一种预训练的神经网络（Artificial Neural Network：ANN）来预测CTU的深度地图，但Lorkiewicz等人同样只考虑了单一尺度的CTU输入[38]。Ren等人提出了一种基于CNN的帧内块划分（intra block partition CNN：IBP-CNN）网络，通过学习图像块的纹理特征来实现块的深度划分[43]。Feng等人提出了一种基于CNN的HEVC帧内深度范围（HEVC intra depth range CNN：HIDR-CNN）预测网络，通过预测CTU可能的深度范围，并仅在预测的深度范围内执行RDcost计算来减少RDcost的计算次数，从而达到降低编码复杂度的目的[45]。Li等人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[46]。Imen等人基于LeNet-5和AlexNet，有效降低了检查所有块决策候选者的计算时间，从而深度优化了HEVC编码单元划分模块的全帧内配置[48]。Yao等人[49]构建了一种基于CNN的双网络模型：预测网络和目标网络。通过基于RD函数的强化学习方法，在确定最优CU划分模式的同时获得CU分区的最佳RD估计。

由以上分析可知，虽然现有的基于CNN的深度预测方法能够通过自主学习CTU的结构来做出对应的深度划分决策，但是，一方面，这些方法几乎都是仅采用了一种串联的二分类策略来决定CTU的划分结果，存在着明显的分类错误累积影响。另一方面，现有的基于CNN的CTU深度划分方法几乎都只考虑了单一尺度的输入，而忽略了不同尺度CU之间的特征关联。

综上所述，针对现有方法的缺陷，本论文提出了一种多尺度多输入的互补分类网络（MCCN）来实现更高效和更准确的HEVC帧内CTU深度划分。接下来，我们就将对本论文提出的MCCN进行详细的论述。

# 本论文提出的多尺度多输入互补分类网络（MCCN）

众所周知，在HEVC的帧内编码中规定：编码树单元（CTU）的大小为，而CTU中编码单元（CU）的大小则根据其深度的不同分别为、、和，其对应的深度分别用0、1、2和3来标记。因此，在本论文提出的MCCN中，我们同样采用该项规定来对HEVC的帧内CTU进行深度划分。如前所述，本论文提出的MCCN的主要贡献为：互补的分类策略（CCS）和多尺度多输入卷积神经网络（MMCNN）。因此，接下来我们就将对提出的贡献点进行详细的论述。

## 互补分类策略（CCS）

现有的基于CNN的CTU深度划分方法几乎都是从的图像块开始，采用一种从大到小的串联的二分类策略来决定CTU的划分结果。然而，如此的串联二分类策略不仅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[22]-[27]，而且还需要对不同大小的块分别训练不同的网络：需要训练出3种不同的二分类网络来分别预测、和是否需要划分。那么，能否同时实现更少的训练网络和更准确的深度划分呢？基于此，本论文提出了一种新颖的互补分类策略（CCS）。我们首先将每个的CTU分别以和两种不同的单元进行划分，得到两幅不同的分块图（如图1所示）：以为单元时，每个的CTU被划分成16个的分块图，我们用来表示；以为单元时，每个的CTU被划分成4个的分块图，我们用来表示。然后，针对，采用本论文提出的二分类MMCNN（详见3.2小节）对其中的每个CU进行是否需要被进一步划分的二分类预测：1表示需要将该的CU划分为4个的CU，0则表示该的CU不需要分割。同时，针对，采用本论文提出的三分类MMCNN（详见3.2小节）对其中的每个CU进行三分类的深度预测：0表示当前这个CU的深度为0，1则表示当前这个CU的深度为1，2则表示当前这个CU的深度为2。

在得到了每个CU的二分类结果和每个CU的三分类结果之后，我们综合考虑二分类和三分类的结果，采用投票机制来决定的CTU中每个CU的最终深度。本论文投票机制的算法流程为：对于每个的CTU而言，（1）、首先参考中每个CU的三分类结果，如果其中的3个或3个以上CU的深度都被预测为0，则认为这整个的CTU的深度也为0，即不需要划分；否则判定该的CTU需要被划分，会被划分为4个的CU。（2）、如果该的CTU被判定为需要划分，那么针对中深度分类为0或1的而言，我们将再查看中对应位置上4个CU的二分类结果，如果4个块的二分类结果中有且仅有一个块的分类结果是需要被划分，则忽略该块的分类结果，判定该的CU不需要被划分；否则判定该的CU需要被划分为4个的CU，且其中每个CU是否需要被划分也是由二分类MMCNN中相同位置上CU的二分类结果来决定。（3）、如果该的CTU被判定为需要被划分，那么针对中深度预测为2的的CU而言，则判定该的CU是肯定要被划分为4个的CU，且其中每个CU是否需要被划分也是由二分类MMCNN中相同位置上CU的二分类结果来决定。本论文投票机制的伪代码如算法1所示。

由以上分析可知，通过结合多分类和投票机制，本论文提出的CCS不仅只需要训练两种不同的网络（一个针对CU的二分类网络和一个针对CU的三分类网络），而且两种分类网络之间很好地实现了分类结果的相互互补，因此也就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。本论文提出的CCS如图1所示。

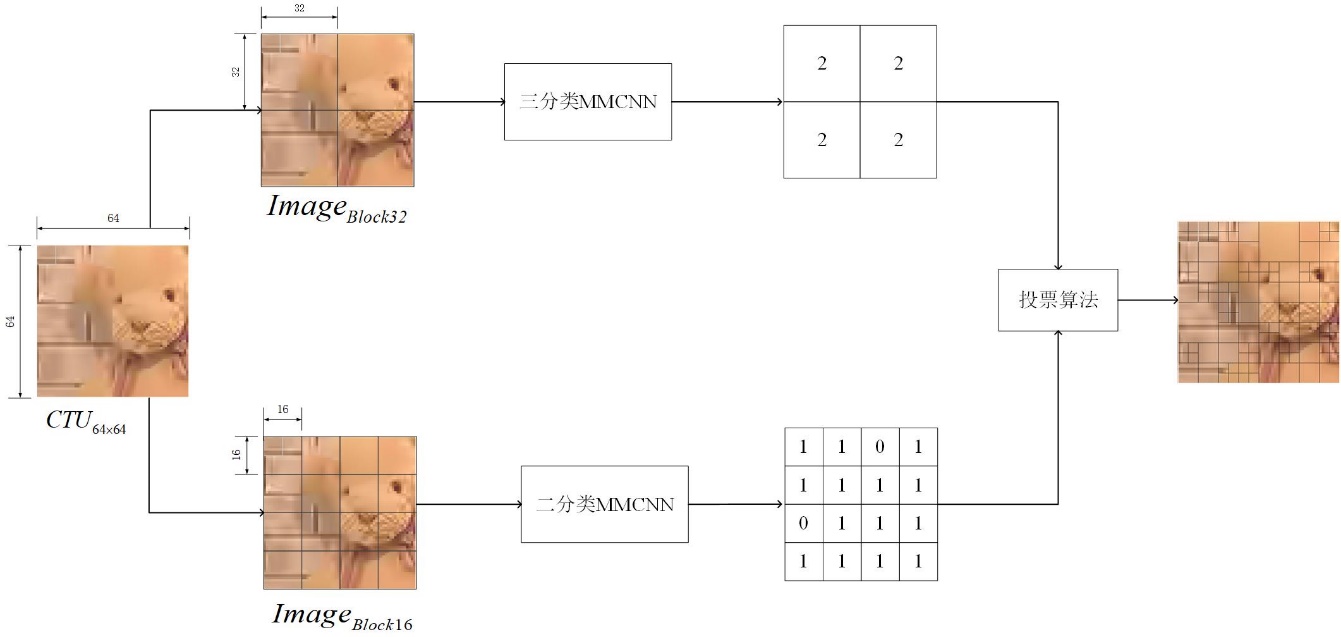


图1 本论文提出的CCS

|  |
| --- |
| 算法**1 投票算法** |
| **输入：**三分类网络预测结果矩阵*PreInfo32*和二分类网络预测结果*PreInfo16*  **输出：**的CTU深度信息矩阵*DepInfo*  # 判断的CTU是否需要划分  1) IF *PreInfo32*矩阵内等于0的数量>=3  2) *DepInfo*内都赋值为0  # 判断的CU是否需要划分  3) ELSE  4) FOR *i*=0,1 DO  5) FOR *j* =0,1 DO  6) IF *PreInfo32* [*i*][*j*]=0 OR *PreInfo32* [*i*][*j*]=1 AND *PreInfo16*矩阵内对应位置等于1的个数<=1  7) *DepInfo*内对应位置赋值为1  # 判断的CU是否需要划分  8) ELSE *PreInfo2* [*i*][*j*] =2  9) FOR *m*=*2i*, *2i+1* DO  10) FOR *n*=*2j*, *2j+1* DO  11) IF *PreInfo16* [*m*][*n*] =0  12) *DepInfo*内对应位置赋值2  13） ELSE  14) *DepInfo*内对应位置赋值3  15) END IF  16) END FOR  17) END FOR  18) END FOR  19) END FOR  20) END FOR  21) END IF |

## 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，本论文提出了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CU的二分类MMCNN（如图2所示）和针对CU的三分类MMCNN（如图3所示）。接下来，我们将以二分类MMCNN为例进行详细的论述。

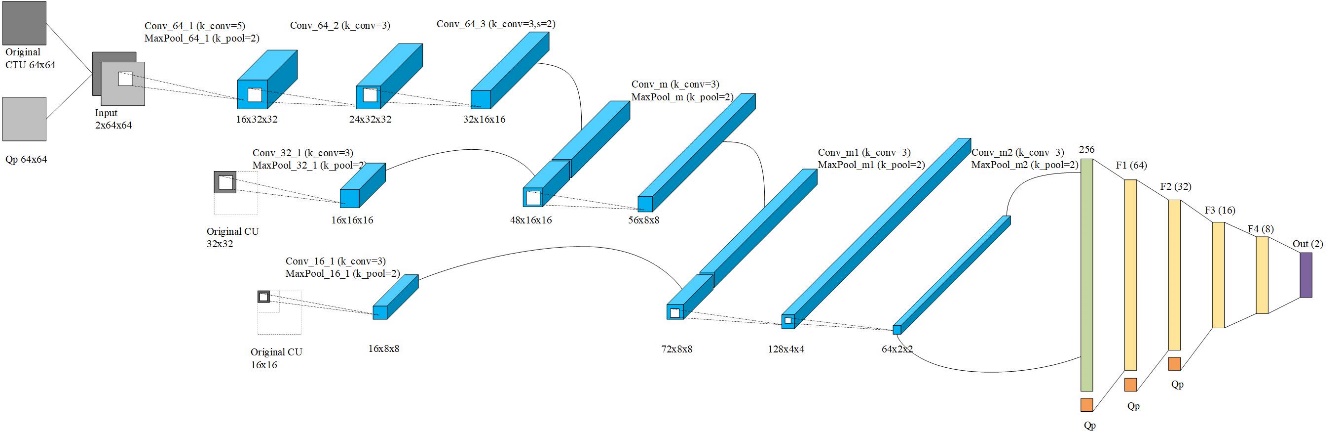


图2 二分类MMCNN

图示

描述已自动生成

图3 三分类MMCNN

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：共三级输入，每一级分别对应不同尺寸的CU：将的CU输入到第一级，同时将该CU中所有的CU和所有的CU分别输入到第二级和第三级。并且在第一级的输入中，我们还将不同的QP分别与的CU进行拼接：每个QP值均与同一个的CU进行通道维度的拼接，组成一个的张量输入到第一级中。本论文同时考虑了四个不同的QP值， 即每个的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一级中。

其次，对于网络结构而言，在第一级中，针对的CU，我们首先采用了1个步长为1的卷积层（k\_conv=5）来提取其中的特征信息，并在该卷积层之后执行1次的最大池化操作（k\_pool=2），在保留显著特征的同时实现降维，将特征图的分辨率降为；然后再采用2个的卷积层来进一步提取细节特征，其中，第1个卷积层的步长为1，第2个卷积层的步长为2，再次将特征图的分辨率降为。在第二级中，我们首先采用1个步长为1的的卷积层来提取CU的特征，并且同样在卷积层之后执行1次的最大池化操作，将第二级的特征图的分辨率降为，并与第一级的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个的卷积层和1个的最大池化操作，将拼接特征图的分辨率降为。在第三级中，我们首先采用一个步长为1的卷积层和一个的最大池化层将输入的CU映射为的特征图，并与第二级中的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个步长为1的卷积层和2个的最大池化操作，最终得到一个维度为的特征图。在以上基于卷积的特征提取过程中，每个卷积层都采用PReLU来作为非线性激活函数。在完成了基于卷积的三级特征提取和融合之后，我们采用5个连续的全连接层对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，其区别仅是输入由二分类MMCNN的三级变为两级，因为三分类MMCNN面向的是 的CU，同时仅采用了4个全连接层且最后的输出为三分类。表1和表2分别为二分类MMCNN和三分类MMCNN的具体网络参数配置。无论是二分类MMCNN还是三分类MMCNN，我们均采用交叉熵作为其损失函数，如公式(1)所示：

(1)

其中，在二分类MMCNN中，N表示CU的数量，和分别表示每个CU真实的类别和网络输出的类别。在三分类MMCNN中，*N*表示CU的数量，和分别表示每个CU真实的类别和网络输出的类别。*i*表示类别数，对于二分类MMCNN，*i*∈[0,1]，对于三分类MMCNN，*i*∈[0,1,2]。由以上分析可知，一方面，本论文提出的MMCNN通过将不同尺度的CU同时输入到网络中，并通过卷积、池化和拼接等操作，实现了多尺度CU特征信息的有效融合，也就将不同尺度的CU特征自适应的关联了起来，从而进一步提升了网络的特征提取能力和表达能力，在较大程度上提升了CTU深度划分的效率，达到有效降低HEVC编码复杂度的目的。另一方面，通过在训练阶段将多个不同的QP值作为网络的输入，使得我们的MMCNN只需训练一次即可适用于多个不同的QP值。二分类MMCNN和三分类MMCNN的具体参数配置如表1和表2所示。

表1 二分类MMCNN的参数配置

|  |  |  |
| --- | --- | --- |
| Layer | Filter Size | Number of Filters |
| Conv\_64\_1 | 5x5 | 16 |
| Conv\_64\_2 | 3x3 | 24 |
| Conv\_64\_3 | 3x3 | 32 |
| Conv\_32\_1 | 3x3 | 16 |
| Conv\_16\_1 | 3x3 | 16 |
| Conv\_m | 3x3 | 56 |
| Conv\_m1 | 3x3 | 128 |
| Conv\_m2 | 3x3 | 64 |
| F1 | 1x1 | 64 |
| F2 | 1x1 | 32 |
| F3 | 1x1 | 16 |
| F4 | 1x1 | 8 |
| Out | 1x1 | 2 |

表2 三分类MMCNN的参数配置

|  |  |  |
| --- | --- | --- |
| Layer | Filter Size | Number of Filters |
| Conv\_64\_1 | 5x5 | 12 |
| Conv\_64\_2 | 3x3 | 16 |
| Conv\_32\_1 | 3x3 | 8 |
| Conv\_m | 3x3 | 36 |
| Conv\_m1 | 3x3 | 64 |
| Conv\_m2 | 3x3 | 96 |
| F1 | 1x1 | 128 |
| F2 | 1x1 | 64 |
| F3 | 1x1 | 16 |
| Out | 1x1 | 3 |

# 实验分析

在本章中，我们进行了大量的实验来验证本论文提出的MCCN的有效性和优越性，同时我们还进行了消融实验，以此来评估本论文提出的贡献点的有效性。

## 数据集和评价指标

在本论文的实验中，我们采用了数据集CPIH[23]，该数据集共包含2000张分辨率为的自然图片，我们将其中的1700张图片作为训练集，200张图片作为验证集，100张图片作为测试集，用于测试CTU深度划分的准确率。此外，我们还把在CPIH上训练好的模型直接用到JCT-VC（Joint Collaborative Team on Video Coding）的18个标准视频序列[34]中进行编码复杂度的评估。在标签的制作方面，我们采用HEVC参考软件HM16.25（HEVC Test Model），其中所有的帧内配置模式都采用默认配置文件encoder\_intra\_main.cfg[33]，并考虑了四种不同的QP值：22,27,32和37，来对CPIH和JCT-VC进行帧内编码，将得到的CTU深度划分结果作为每个CU的真实深度值（Ground Truth），即标签。本论文采用分类准确率（即：在网络输出的所有CU深度值中，判断正确的占比，其值越高越好）、BD-BR（Bjφntegaard delta bit-rate）[33]表示在同样的客观质量下，两种方法的码率节省情况，BD-BR的值越小越好）和编码时间节省率∆T，三种指标来定量评估本论文提出的MCCN的性能，其中的计算如公式(2)所示：

(2)

其中，是HM16.25所需的时间，是本论文方法所需的时间。因此，越大越好，表示节省的编码时间越多，即编码复杂度越低。

## 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）[25]，动量设置为0.9，初始化学习率为0.01，并在每个epoch之后以固定步长进行下降。本论文的所有实验都是在 Intel(R) Core(TM) i7-7820X CPU @ 3.60GHz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。

## 消融实验

如前所述，本文提出的贡献点为：（1）多尺度多输入卷积神经网络（MMCNN）和（2）互补分类策略（CCS）。因此，在本章节中，我们将在CPIH和JCT-VC上进行消融实验，以此来验证本论文提出的MMCNN和CCS的有效性。消融实验中所有模型的设置和训练细节都一致。

MMCNN的有效性消融实验：

如前文所述，本论文提出的MMCNN的创新点在于：同时考虑了多个QP值：22、27、32和37，以及多个不同尺度的CU，并通过融合不同尺度CU的特征信息来提升HEVC帧内CTU深度划分的效率。因此，为了能够准确评估本论文提出的MMCNN的有效性，我们创建了两个模型：MMCNN-NoQP和MMCNN-OneScale。其中，MMCNN-NoQP不会将QP值输入到网络中，在二分类MMCNN-NoQP和三分类MMCNN-NoQP中均是仅有的CU作为第一级的输入，而其余的网络结构和配置与二分类MMCNN和三分类MMCNN一致。而MMCNN-OneScale则表示在二分类MMCNN-OneScale和三分类MMCNN-OneScale中均只考虑第一级的输入，而忽略第二级和第三级的输入，因此，在二分类MMCNN-OneScale中，没有第二级的Conv\_32\_1（k\_conv=3）、对应的MaxPool\_32\_1（k\_pool=2）和拼接操作，以及第三级的Conv\_16\_1（k\_conv=3）、对应的MaxPool\_16\_1（k\_pool=2）和拼接操作，而其余的网络结构和配置与二分类MMCNN一致。在三分类MMCNN-OneScale中，则没有第二级的Conv\_32\_1（k\_conv=5）、MaxPool\_32\_1（k\_pool=2）和拼接操作，而其余的网络结构和配置与三分类MMCNN一致。MMCNN-NoQP和MMCNN-OneScale也都采用CCS。我们将MMCNN-NoQP+CCS命名为MCCN-NoQP，将MMCNN-OneScale+CCS命名为MCCN-OneScale。因此，比较MCCN-NoQP和MCCN-OneScale与MCCN之间的性能差异是能够公平和准确评估多QP值输入和多尺度CU特征融合的有效性的。

我们在JCT-VC的18个标准视频序列上将MCCN-NoQP、MCCN-OneScale和MCCN的平均BD-BR和平均进行比较实验，其实验结果如表3所示。如表3中的第二行和第三行所示，MCCN-NoQP在平均BD-BR和平均两个方面均落后于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∆T减少了12.89%。这是因为没有了QP值的输入，MCCN-NoQP就需要为每个QP单独训练一个网络（4个QP值需要训练出4个模型），从而使得HEVC在对同一视频序列下的不同QP值分别进行训练时会造成对同一CTU的划分采用不同的策略：对于数值较大的QP，HEVC会倾向于将CTU划分成较大的CU，而对于数值较小的QP，HEVC则会倾向于将CTU划分为较小的CU，也就很难学习到一种能够适用于所有QP值的划分策略。然而本论文提出的MCCN将所有的QP值都输入到网络中来引导网络的学习和参数的更新，因此只需要训练出一个模型就能够让网络学习到适用于不同QP的划分策略。

如表3中的第二行和第四行所示，MCCN-OneScale在平均BD-BR和平均两个方面也都落后于MCCN：平均BD-BR增加了8.23%，而平均反而减少了7.19%。由此可见，执行不同尺度CU特征的融合，是能够极大提升HEVC帧内CTU深度划分的效率的。表3很好地证明了本论提出的MMCNN的有效性。

表3 MMCNN的有效性消融实验

|  |  |  |
| --- | --- | --- |
| Model | 平均BD-BR | 平均 |
| MCCN | 3.18% | 71.49% |
| MCCN-NoQP | 9.8% | 58.6% |
| MCCN-OneScale | 11.5% | 64.3% |

CCS的有效性消融实验：

如前文所述，本论文提出的CCS能够很好地避免现有方法中存在的错误累积效应，提升深度预测的准确性。因此在CCS的有效性消融实验中，我们移除CCS，将现有的串联二分类策略（SBCS）应用到我们的MMCNN中：采用图2的二分类MMCNN来分别依次对、和大小的CU进行是否需要划分的二分类决策：首先训练一个针对CU的二分类MMCNN（命名为：二分类MMCNN64）来判断的CU是否需要划分；然后再训练一个针对 CU的二分类MMCNN32来判断的CU是否需要划分；最后训练一个针对CU的二分类MMCNN16来判断的图像块是否需要划分，因此大块的分类结果会直接影响到对应小块的划分结果。我们将二分类MMCNN+SBCS命名为MCCN-SBCS，因此，比较MCCN-SBCS与MCCN之间的性能差异能够公平和准确评估本论文提出的CCS的有效性。

我们在CPIH的100张测试集上，将本论文提出的MCCN与MCCN-SBCS在CU、CU和CU上的深度划分准确率的平均值进行了比较。如表4所示，本论文提出的MCCN在CU、CU和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，本论文提出的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。

我们在CPIH的100张测试集上，将本论文提出的MCCN与MCCN-SBCS在CU、CU和CU上的深度划分准确率的平均值进行了比较。如表4所示，本论文提出的MCCN在CU、CU和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，本论文提出的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。

表4 在100张CPIH测试集上CCS的有效性消融实验：CU、CU和CU的深度划分的平均准确率比较。

|  |  |  |
| --- | --- | --- |
| 方法 | MCCN | MCCN-SBCS |
| CU | 90.30% | 88.05% |
| CU | 87.55% | 86.51% |
| CU | 89.69% | 85.71% |

除了深度划分的准确率之外，我们还在JCT-VC上将本论文提出的MCCN与MCCN-SBCS在平均BD-BR和平均上进行了比较，如表5所。由表5可见，MCCN-SBCS在平均BD-BR和平均两个方面也都落后于MCCN：平均BD-BR增加了10.62%，平均∆T反而减少了4.49%。由此可见，本论文提出的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4和表5很好地证明了本论提出的CCS的有效性。

表5 在JCT-VC上MCCN和MCCN-SBC的平均BD-BR和平均比较

|  |  |  |
| --- | --- | --- |
| Model | 平均BD-BR | 平均 |
| MCCN | 3.18% | 71.49% |
| MCCN-SBCS | 13.8% | 67% |

## 与前沿方法的比较

为了验证本论文提出方法的优越性，在本小节中，我们首先将提出的MCCN与方法[18]和[38]在100张CPIH测试集上进行深度划分准确率的比较（如表6所示）；然后，我们再将提出的MCCN与近几年最具代表性的8种方法：[18]、[23]、[24]、[26]、[29]、[36]、[37]和[38]在JCT-VC上进行BD-BR和的比较（如表7所示）。其中，[23]、[24]、[26]、[29]、[36]、[37]和[38]是基于神经网络学习的方法，而[18]是属于启发式的方法。

如表6所示，虽然本论文提出的MCCN对 CU的深度预测平均准确率略低于方法[24]和[36]，分别低了0.68%和0.88%，但是在CU的深度预测和CU的深度预测上，本论文提出MCCN的平均准确率均高于方法[24]和[36]，分别提升了0.65%-0.93%和2.14%-9.27%。通过分析表6可知，虽然本论文提出的MCCN在最大的CU上的准确率略低，但是对其中CU和CU的深度预测准确率却反而提升较大，因此，该现象恰恰很好的证明了本论文提出的MCCN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接影响到对应小块的深度划分。

表7所示为方法[18]、[23]、[24]、[26]、[29]、[36]、[37]、[38]和本论文提出的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7可见，本论文提出的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。特别的，首先本论文提出的MCCN在平均BD-BR和平均两方面均优于方法[37]，不仅平均提高了1.1%，而且BD-BR也下降了1.09%。其次，与方法[18]、[23]、[24]、[26]、[29] 、[36]和[38]相比，虽然本论文提出的MCCN在平均BD-BR方面分别增加了1.81%、0.97%、0.93%、1.24、1.14、1.16%和1.37%，但是MCCN的编码复杂度却实现了大幅的下降，平均分别提高了24.19%、9.24%、9.64%、7.44、11.78、5.94%和10.86% 。因此，本论文提出的MCCN明显优于现有的极具代表性的方法[18]、[23]、[24]、[26]、[29]、[36]、[37]和[38]。表6和表7很好地证明了本论文提出的MCCN的优越性。

表6 方法[24]、[36 ]和本论文提出的MCCN在100张CPIH测试集上的平均准确率比较（其中最好的性能已加粗标记）

|  |  |  |  |
| --- | --- | --- | --- |
| 方法 | Proposed | [24] | [36] |
| CU | 90.30% | 90.98% | **91.18%** |
| CU | **87.55%** | 86.62% | 86.90% |
| CU | **89.69%** | 80.42% | 87.55% |

表**7** 方法[23]、[24]、[26]、[29] 、[36]、[37]、[38]、[18]和本论文提出的MCCN在JCT-VC所有18个标准视频测试集上的平均BD-BR和平均比较（其中最好的性能已加粗标记）

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 出版年份 | Ours | 2023 | 2022 | 2022 | 2021 | 2021 | 2018 | 2018 | 2017 |
| 方法 | Proposed | [26] | [29] | [37] | [36] | [38] | [18] | [24] | [23] |
| BD-BR | 3.18% | 1.94% | 2.04% | 4.27% | 2.02% | 1.81% | **1.37%** | 2.25% | 2.21% |
|  | **71.49%** | 64.05% | 59.71% | 70.39% | 65.55% | 60.63% | 47.30% | 61.85% | 62.25% |

# 总 结

本论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，我们首先将多个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，仅需要两个分类网络（二分类MMCNN和三分类MMCNN）就实现了更高效的CTU深度划分，而且还拥有了更强的QP值泛化能力。其次，我们还提出了一种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（、和）的更准确的深度预测。大量的实验结果表明，本论文提出的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CU和CU的深度准确率分别平均提升了0.65%-0.93%和2.14%-9.27%。将本论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。

References:

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. 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.
3. Lainema J, Bossen F, Han W J, et al. Intra coding of the HEVC standard[J]. IEEE transactions on circuits and systems for video technology, 2012, 22(12): 1792-1801.
4. 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.
5. Gweon R, Lee Y L. Early termination of CU encoding to reduce HEVC complexity[J]. IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences, 2012, 95(7): 1215-1218.
6. Choi K, Park H M, Jang E S. JCTVC-F092 Coding tree pruning based CU early termination[J]. Joint Collaborative Team on Video Coding (JCTVC) of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11 Torino July, 2011: 1-11.
7. Yang J, Kim J, Won K, et al. Early SKIP detection for HEVC. JCT-VC of ITU-T SG16 WP3 and ISO[R]. IEC JTC1/SC29/WG11, Document JCTVC-G543, 2011.
8. Cho S, Kim M. Fast CU splitting and pruning for suboptimal CU partitioning in HEVC intra coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(9): 1555-1564.
9. Zhang Y, Wang H, Li Z. Fast coding unit depth decision algorithm for interframe coding in HEVC[C]//2013 Data Compression Conference. IEEE, 2013: 53-62.
10. Zhao L, Fan X, Ma S, et al. Fast intra-encoding algorithm for high efficiency video coding[J]. Signal Processing: Image Communication, 2014, 29(9): 935-944.
11. 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). IEEE, 2016: 1-3.
12. Khan M U K, Shafique M, Henkel J. An adaptive complexity reduction scheme with fast prediction unit decision for HEVC intra encoding[C]//2013 IEEE International Conference on Image Processing. IEEE, 2013: 1578-1582.
13. Xiong J, Li H, Wu Q, et al. A fast HEVC inter CU selection method based on pyramid motion divergence[J]. IEEE transactions on multimedia, 2013, 16(2): 559-564.
14. Zhang T, Sun M T, Zhao D, et al. Fast intra-mode and CU size decision for HEVC[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 27(8): 1714-1726.
15. Hu N, Yang E H. Fast mode selection for HEVC intra-frame coding with entropy coding refinement based on a transparent composite model[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 25(9): 1521-1532.
16. Liu D, Liu X, Li Y. Fast CU size decisions for HEVC intra frame coding based on support vector machines[C]//2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). IEEE, 2016: 594-597.
17. Bakkouri S, Elyousfi A. Machine learning-based fast CU size decision algorithm for 3D-HEVC inter-coding[J]. Journal of Real-Time Image Processing, 2021, 18: 983-995.
18. Jamali M, Coulombe S. Fast HEVC intra mode decision based on RDO cost prediction[J]. IEEE transactions on broadcasting, 2018, 65(1): 109-122.
19. Chiang J C, Peng K K, Wu C C, et al. Fast intra mode decision and fast CU size decision for depth video coding in 3D-HEVC[J]. Signal Processing: Image Communication, 2019, 71: 13-23.
20. Huang Z, Sun J, Guo X. FastCNN: Towards Fast and Accurate Spatiotemporal Network for HEVC Compressed Video Enhancement[J]. ACM Transactions on Multimedia Computing, Communications and Applications, 2023, 19(3): 1-22.
21. Huang Y, Song L, Xie R, et al. Modeling acceleration properties for flexible INTRA HEVC complexity control[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(11): 4454-4469.
22. Yu X, Liu Z, Liu J, 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). IEEE, 2015: 1285-1289.
23. Li T, Xu M, Deng X. A deep convolutional neural network approach for complexity reduction on intra-mode HEVC[C]//2017 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2017: 1255-1260.
24. Xu M, Li T, Wang Z, et al. Reducing complexity of HEVC: A deep learning approach[J]. IEEE Transactions on Image Processing, 2018, 27(10): 5044-5059.
25. Kuanar S, Rao K R, Bilas M, et al. Adaptive CU mode selection in HEVC intra prediction: A deep learning approach[J]. Circuits, systems, and signal processing, 2019, 38: 5081-5102.
26. Li H, Wei G, Wang T, et al. Reducing Video Coding Complexity Based on CNN-CBAM in HEVC[J]. Applied Sciences, 2023, 13(18): 10135.
27. Bouaafia S, Khemiri R, Maraoui A, et al. CNN-LSTM learning approach-based complexity reduction for high-efficiency video coding standard[J]. Scientific Programming, 2021, 2021: 1-10.
28. Wang Z, Li F. Convolutional neural network based low complexity HEVC intra encoder[J]. Multimedia Tools and Applications, 2021, 80: 2441-2460.
29. Qin L M, Zhu Z J, Bai Y Q, et al. A Complexity-Reducing HEVC Intra-Mode Method Based on VGGNet[J]. Journal of Computers, 2022, 33(4): 57-67.
30. Grellert M, da Silva Cruz L A, Zatt B, et al. Coding mode decision algorithm for fast HEVC transrating using heuristics and machine learning[J]. Journal of Real-Time Image Processing, 2021: 1-16.
31. Fan J, Song L. Fast Intra-frame Prediction Algorithm for HEVC Based on Neural Networks and Adaptive Threshold[C]//Proceedings of the 2022 6th International Conference on Video and Image Processing. 2022: 127-134.
32. Xu M, Jeon B. Learning-Based Efficient Quantizer Selection for Fast HEVC Encoder[J]. IEEE Transactions on Broadcasting, 2023: 1-13.
33. Bossen F. Common test conditions and software reference configurations, document JCTVC-L1100[J]. JCT-VC, San Jose, CA, 2012.
34. Ohm J R, Sullivan G J, Schwarz H, et al. Comparison of the coding efficiency of video coding standards—including high efficiency video coding (HEVC)[J]. IEEE Transactions on circuits and systems for video technology, 2012, 22(12): 1669-1684.
35. Wang T, Wei G, Li H, et al. A Method to Reduce the Intra-Frame Prediction Complexity of HEVC Based on D-CNN[J]. Electronics, 2023, 12(9): 2091.
36. Feng A, Gao C, Li L, 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). IEEE, 2021: 1-6.
37. Hari P, Jadhav V, 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). IEEE, 2022: 548-551.
38. 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 (SPSympo). IEEE, 2021: 177-182.
39. Zhao R, Huang H, Zhang R, et al. Fast HEVC Intra CTU Partition Algorithm Based on Lightweight CNN[C]//Signal and Information Processing, Networking and Computers: Proceedings of the 8th International Conference on Signal and Information Processing, Networking and Computers (ICSINC). Singapore: Springer Nature Singapore, 2022: 996-1003.
40. Paul S, Norkin A, Bovik A C. Speeding up VP9 intra encoder with hierarchical deep learning-based partition prediction[J]. IEEE Transactions on Image Processing, 2020, 29: 8134-8148.
41. 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.
42. Chen Z, Shi J, Li W. Learned fast HEVC intra coding[J]. IEEE Transactions on Image Processing, 2020, 29: 5431-5446.
43. Ren W, Su J, Sun C, et al. An IBP-CNN based fast block partition for intra prediction[C]//2019 Picture Coding Symposium (PCS). IEEE, 2019: 1-5.
44. Zhang Q, Wang Y, Huang L, et al. Fast CU partition and intra mode decision method for H. 266/VVC[J]. IEEE Access, 2020, 8: 117539-117550.
45. Feng Z, Liu P, Jia K, et al. HEVC fast intra coding based CTU depth range prediction[C]//2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC). IEEE, 2018: 551-555.
46. Li Y, Li L, Fang Y, 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.
47. Zaki F, Mohamed A E, Sayed S G. CtuNet: A deep learning-based framework for fast CTU partitioning of H265/HEVC intra-coding[J]. Ain Shams Engineering Journal, 2021, 12(2): 1859-1866.
48. 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.
49. Yao C, Xu C, Liu M. RDNet: Rate–Distortion-Based Coding Unit Partition Network for Intra-Prediction[J]. Electronics, 2022, 11(6): 916.