

## 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

贾克斌\* 崔腾鹤 刘鹏宇 刘畅

(北京工业大学信息学部 北京 100124)

(先进信息网络北京实验室 北京 100124)

**摘要:** 高效视频编码(HEVC)标准相对于H.264/AVC标准提升了压缩效率,但由于引入的编码单元四叉树划分结构也使得编码复杂度大幅度提升。对此,该文提出一种针对HEVC帧内编码模式下编码单元(CU)划分表征矢量预测的多层特征传递卷积神经网络(MLFT-CNN),大幅度降低了视频编码复杂度。首先,提出融合CU划分结构信息的降分辨率特征提取模块;其次,改进通道注意力机制以提升特征的纹理表达性能;再次,设计特征传递机制,用高深度编码单元划分特征指导低深度编码单元的划分;最后建立分段特征表示的目标损失函数,训练端到端的CU划分表征矢量预测网络。实验结果表明,在不影响视频编码质量的前提下,该文所提算法有效地降低了HEVC的编码复杂度,与标准方法相比,编码复杂度平均下降了70.96%。

**关键词:** 高效视频编码; 复杂度降低; 深度学习; 帧内编码

中图分类号: TN919.81

文献标识码: A

文章编号: 1009-5896(2021)07-2023-09

DOI: [10.11999/JEIT200414](https://doi.org/10.11999/JEIT200414)

## Fast Prediction Algorithm in High Efficiency Video Coding Intra-mode Based on Deep Feature Learning

JIA Kebin CUI Tenghe LIU Pengyu LIU Chang

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

(Beijing Laboratory of Advanced Information Networks, Beijing 100124, China)

**Abstract:** Compared to H.264/AVC coding standard, High Efficiency Video Coding (HEVC) improves the compression efficiency, but the consequent disadvantage is the significant increase in encoding complexity by using the quad-tree partition. A Multi-Layer Feature Transfer Convolutional Neural Network (MLFT-CNN) for Coding Unit (CU) division and characterization vector prediction in HEVC intra coding mode is proposed, which greatly reduces the complexity of video coding. Firstly, a reduced-resolution feature extraction module incorporating CU partition structure information is proposed. Then, the channel attention mechanism is improved for a better texture expression performance of the feature. After that, the feature transfer mechanism is designed to use the feature division of high-depth coding unit to guide the division of low-depth coding unit. Finally, the target loss function represented by the segmented feature is established, and the end-to-end CU division represents the vector prediction network. The experimental results show that the proposed algorithm effectively reduces the encoding complexity of HEVC without affecting the video coding quality. Specifically, compared to the standard method, the encoding complexity on the standard test sequence is reduced by 70.96% on average.

**Key words:** High Efficiency Video Coding(HEVC); Complexity reduction; Deep learning; Intra coding

收稿日期: 2020-05-26; 改回日期: 2020-12-15; 网络出版: 2021-01-05

\*通信作者: 贾克斌 kebinj@bjut.edu.cn

基金项目: 国家自然科学基金(61672064), 国家重点研发计划(2018YFF01010100), 青海省基础研发计划(2020-ZJ-709)

Foundation Items: The National Natural Science Foundation of China (61672064), The National Key Research and Development Project of China (2018YFF01010100), The Basic Research Program of Qinghai Province (2020-ZJ-709)

## 1 引言

随着生活水平的不断提高,人们对信息化服务的质量要求也越来越高,多视点及高清甚至超高清视频服务已成为人们追逐的目标。面对视频服务的新需求,2013年由视频编码联合协作小组(Joint Collaborative Team on Video Coding, JCT-VC)提出的着眼于“更低码率”和“更高画质”的高效视频编码标准H.265/HEVC(High Efficiency Video Coding)<sup>[1]</sup>,为视频内容从高清到超高清的发展提供了技术保障,并成为最成熟的新一代视频编码标准。与上一代高级视频编码标准H.264/AVC(Advanced Video Coding)<sup>[2]</sup>相比,H.265/HEVC通过引入编码单元(Coding Unit, CU)二叉树划分结构<sup>[3]</sup>,多角度帧内预测等技术,使压缩效率提高了1倍<sup>[1]</sup>。

但随着新编码技术的引入,H.265/HEVC的编码复杂度相对于H.264/AVC增加了253%,其中,CU的二叉树划分过程需要遍历所有可能的CU划分方式,并基于所有的划分方式进行率失真优化(Rate Distortion Optimization, RDO)的计算,这一过程占据了编码时间的80%<sup>[4]</sup>,使得H.265/HEVC很难满足目前高实时性多媒体系统的应用要求<sup>[5]</sup>。然而CU作为预测、变换、量化和熵编码等处理的基本单元,是保证编码性能的关键所在。因此,对CU划分结构预测的研究引起了国际同行学者的广泛关注。目前,针对CU划分结构预测,国内外学者在传统方法和机器学习方法都有了大量的探索与尝试。其中传统CU划分结构预测方法主要用于CU划分二叉树的剪枝处理,这些方法大多是基于统计CU的时间/空间相关性所提出的<sup>[6,7]</sup>。但传统方法存在着对先验信息要求高,对复杂情况判断准确性较低的缺点。进一步,有学者采用机器学习方法来预测CU是否划分,早期研究方法主要是基于概率密度分布,依靠设计良好的手工特征进行一些二分类问题的处理,其中利用SVM分类器最为经典<sup>[8-10]</sup>。近年来,以数据驱动的深度技术凭借其在数据表征上优秀的泛化能力,在大量视觉任务中取得了令人瞩目的成就。一些学者也开始尝试将其引入到新一代视频编码标准的优化过程中,以解决传统机器学习方法在复杂情况下判别大幅度失准的问题,例如,Feng等人<sup>[11]</sup>与Liu等人<sup>[12]</sup>基于不同的网络模型学习图像帧的纹理复杂度并自适应判断深度范围,然而,这两种方法是剪枝思想的延续,还需额外RDO计算才能确定CU划分结构;文献<sup>[13]</sup>则采用简化的卷积神经网络,针对不同尺寸CU设计对应的网络,其工程量巨大,且泛化性能差。文

献<sup>[14]</sup>通过数据降维手段,使得单个网络能够针对不同尺寸CU进行预测,但其网络结构简单,且过程中还需进行额外的RDO计算。

针对上述方法存在的不足,本文提出了一种基于HEVC编码框架的CU划分表征矢量预测网络,意在实现对CU划分结构端到端的预测,省去CU划分结构确定过程中的复杂RDO计算过程,从而降低编码复杂度。第2节简述HEVC帧内编码结构及CU划分表征矢量;第3节提出CU划分表征矢量预测的多层特征传递卷积神经网络模型的设计方案;第4节验证提出模型的有效性;第5节对全文进行总结。

## 2 HEVC帧内编码结构及CU划分表征矢量

### 2.1 HEVC帧内编码结构

在HEVC中,每帧图像会被分成若干个编码树单元(Coding Tree Units, CTU)<sup>[2]</sup>。每个CTU可以包含单个CU,也可以根据二叉树结构递归各个深度(depth)分割成多个较小的CU,图1展示了一种CTU划分结构及其对应二叉树结构。在HEVC默认配置中,CU大小可为 $8 \times 8$ ,  $16 \times 16$ ,  $32 \times 32$ ,  $64 \times 64$ 。

此外,CTU二叉树划分过程不仅包括自上而下的率失真代价计算过程,还包括自下而上的率失真代价比较过程。图2显示了父CU和子CU之间率失真代价计算和比较过程。在计算过程(如图2(a)所示)中,需要首先计算父CU的率失真代价 $RDCost^{parent}$ ,然后分别计算其子CU的率失真代价 $RDCost_i^{child}$  ( $i \in \{0, 1, 2, 3\}$ )。在比较过程(如图2(b)所示)中,比较父CU的率失真代价和4个子CU率失真代价总和的大小,来判别该父CU是否需要划分。若 $RDCost^{parent} > \sum_{i=0}^3 RDCost_i^{child}$ ,则父CU需要划分;反之,则不需划分。通过遍历CTU所有可能的划分方式,选取率失真代价最小的划分结构为该CTU最优划分结构。

需要说明的是,CTU最优划分结构是通过递归计算实现的。在一个 $64 \times 64$ 的CTU中,随着CU尺寸的减小,共需计算 $85(4^0+4^1+4^2+4^3)$ 个CU的率失真代价。其率失真代价的计算过程均需编码器对该CU进行预编码,在此过程中还需对所有可能的预测模式和变换方式进行编码,故时间开销巨大。然而,CTU最优划分结构在限定编码参数条件下具有唯一性。故需要最少计算1个(如果 $64 \times 64$  CU没有被分割),最多64个(如果整个CTU被分割为 $8 \times 8$  CU)CU率失真代价来完成编码需求。因此,通过对CU划分的准确预测可以避免最多84个,最少21个CU的预编码过程。

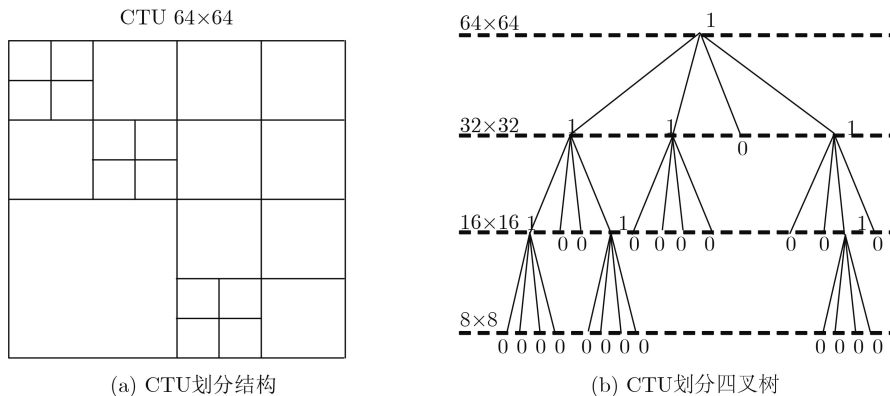


图 1 CTU划分结构示意图

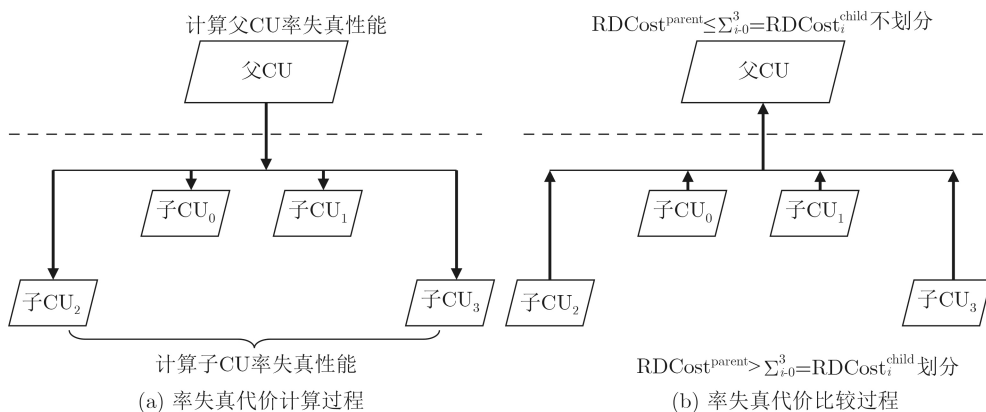


图 2 父CU和子CU之间率失真代价计算和比较过程

### 2.2 CU划分表征矢量

根据HEVC中CTU划分结构可知，当CU尺寸大于等于 $16 \times 16$ ，只存在划分和不划分两种情况。所以可利用其划分标志位来描述CU划分结构，其中0代表不划分，1代表划分。基于此，本文提出了由21位二进制数字组成的CU划分表征矢量，可表示CTU所有的划分可能。其21位分别对应于从浅层到深层的CU，依次按照对应从左到右，从上到下的CU的划分标志位，其CU划分表征矢量对应位置如图3所示。

### 3 CU划分表征矢量预测的多层特征传递卷积神经网络模型

MLFT-CNN模型结构如图4所示，可分为泛化特征提取、深度差异特征提取，以及全连接特征分类3部分。从降低编码复杂度的整体需求出发，考虑到亮度分量表征视觉信息能力远大于色度分量，网络输入为从原始序列中提取的大小为 $64 \times 64$ 的亮度信息。泛化特征提取部分对数据进行全局降维和多维特征提取。深度差异特征提取部分从CU划分的3个深度出发进行CU划分表征矢量端到端的训练。全连接特征分类部分用高深度CU划分特征指导低深度CU划分过程，强化低深度CU的划分标识符的全局表征能力。最终拼接得到CU划分表征矢

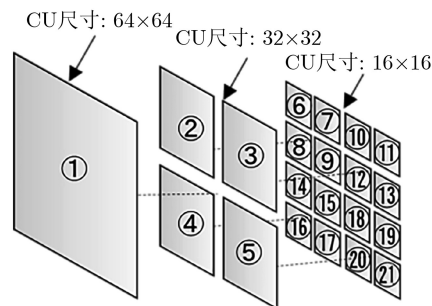


图 3 CU划分表征矢量对应位置示意图

量。下面对模型各个模块结构及整体模型学习过程进行详细介绍。

#### 3.1 降分辨率特征提取模块

在基于深度学习的CU划分表征矢量预测任务中，若采用传统的卷积操作训练，则忽略了划分过程的边界限制条件，各层感受野与CU的划分尺寸之间缺少关联信息，使得深层特征之间存在较多信息冗余，这无疑会扩大网络体积，拖慢前向传播速度，在实际视频编码过程中得不偿失。故本文在降分辨率特征提取过程中，设计了针对CU四叉树划分过程的卷积过程，提出了应用边长为 $2^n$  ( $n=1, 2$ )的卷积核，并进行步长即步进距离的卷积操作。这使得CU划分尺寸与卷积得到的特征图感受野尺

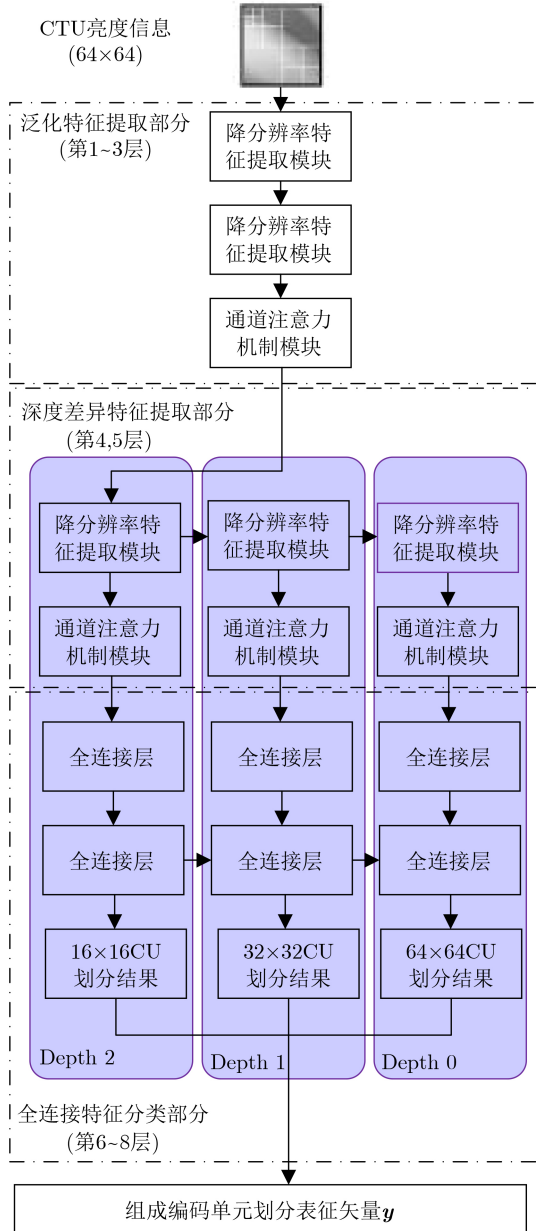


图4 整体网络模型图

寸相耦合，在卷积过程中融入了视频编码的先验信息，故更适用于CU划分的降分辨率特征的提取。

其中在全局特征提取部分，对输入亮度CU单元 $S^0$ 进行边长为 $s$ 的全局多核卷积降分辨率特征提取，可用式(1)表示；在深度差异特征提取部分中将全局泛化特征专一化，将高深度特征作为输入来获取只属于该深度的降分辨率特征，可用式(2)来表示。

$$S_c^{i,j}[x,y] = \sigma \left( b_c^{i,j} + \sum_{x',y',c'} w_{cc'}^{i,j}[x',y'] \times S_c^{i-1,j}[sx+x',sy+y'] \right) \quad (1)$$

$$S_c^{i,j-1}[x,y] = \sigma \left( b_c^{i,j} + \sum_{x',y',c'} w_{cc'}^{i,j}[x',y'] \times S_c^{i,j}[sx+x',sy+y'] \right) \quad (2)$$

其中， $x$ 和 $y$ 为特征图位置信息， $x'$ 和 $y'$ 为卷积核位置信息 ( $x' \in \{0, 1, \dots, s-1\}$ ,  $y' \in \{0, 1, \dots, s-1\}$ )， $w$ 为权值系数， $b$ 为偏置系数， $c$ 和 $c'$ 分别对应当前层与前一层的通道数目， $*$ 表示卷积运算， $S^{i,j}$ 表示第 $i$ 层深度为 $j$ 的特征图， $\sigma(\cdot)$ 表示非线性变换函数，即 $\sigma(\cdot) = \max(\cdot, 0) + 0.2 \max(-\cdot, 0)$ 。

降分辨率特征提取过程中， $s$ 间接反映了分辨率的缩放系数， $s$ 越大分辨率降低得越多，为了更快地使特征图感受野与CU划分尺寸相匹配，进一步减少降分辨率网络所需的网络训练参数，加速CU划分结构预测过程， $s$ 的设置在各层有所不同。其中在 $i=1$ 时， $s$ 设置为4，在 $i=2, 4$ 时， $s$ 设置为2。此外，经过降分辨率特征提取模块的次数越多，网络所能表达的结构特征越复杂，其特征也越能够反映全局特征，对于大尺寸的CU具有更好的描述效果。卷积结果 $S^{i,j}$ 还需经批归一化处理得到对应的降分辨率特征 $H^{i,j}$ ，从而提高所提取降分辨率特征的鲁棒性，使其对于CU单元在不同层次上的结构表征具有更好的泛化能力。

### 3.2 通道注意力机制模块

卷积运算是对于空间信息进行多尺度融合，虽然在空间上能够有效地进行特征融合，但忽略了不同特征通道对于关键信息描述的差异性。因此，本文引入了计算机视觉中的注意力机制。利用网络学习各通道特征重要性的差异，通过赋予不同特征通道不同的权值将降分辨率提取到的多通道特征进行非均匀融合，以提升融合特征的信息表达能力。本文参考SENet<sup>[15]</sup>模型，提出了一种改进的通道注意力机制，通过设计一个全局卷积操作代替其原本的全局平均池化操作来实现数据的压缩，鼓励模型在学习时充分考虑空间和通道信息，而非对两种信息进行割裂的处理。对于降分辨率特征 $H^{i,j}$ 进行全局卷积操作可表示为

$$G_c^{i,j} = \sigma \left( b_c^{i,j} + \sum_{x,y} w^{i,j}[x,y] \times H_c^{i,j}[x,y] \right) \quad (3)$$

其中， $w^{i,j}$ 是该过程的权值系数， $b_c^{i,j}$ 是该过程的偏置系数， $G_c^{i,j}$ 表示第 $i$ 层深度为 $j$ 的全局通道卷积特征。

全局卷积操作只实现了各个通道的全局独立表征，未学习通道之间的重要性。因此，需要通过两层的全连接层来学习各个通道之间的非线性关系，其过程可表示为

$$T_c^{i,j} = \sigma \left( w_{cc'} * \sigma \left( \sum_c w_{c'c} \times G_c^{i,j} + b_{c'} \right) + b_c \right) \quad (4)$$

其中， $w_{c'c}$ ， $b_{c'}$ ， $w_{cc'}$ ， $b_c$ 分别对应于两个全连接层的权值系数和偏置系数。

通过此方法得到用于表示各通道重要性的矢量  $T_c^{i,j}$ ，它与降分辨率特征  $H_c^{i,j}[x,y]$  进行点乘操作，得到融合特征，如式(5)所示

$$A_c^{i+1,j}[x,y] = T_c^{i,j} \cdot H_c^{i,j}[x,y] \quad (5)$$

在模型第5层中，得到的CU划分深度对应的3层融合特征  $A_c^{i,j}$  ( $j \in \{0, 1, 2\}$ )，最后将它用于各层CU划分结构的预测，这种参考了空间信息、通道信息的融合特征，在编码单元编码结构描述上具有更好的纹理表达性能。

### 3.3 全连接判别模块

全连接判别模块由3个深度支路中各自的3层全连接层(2层隐层，1层输出层)和本文所提出的作用于不同深度支路间的特征传递机制构成，其中两层隐层特征分别用  $\{f^{6,l}\}_{l=0}^2$  和  $\{f^{7,l}\}_{l=0}^2$  表示。特征传递机制是基于不同深度划分信息高度相关性设计的，利用高深度CU划分特征指导低深度CU划分的划分过程。其过程如式(6)所示

$$f_c^{7,j-1} = \sigma \left( b_c^{6,j-1} + \sum_{c'} w_{cc'}^{6,j-1} \times f_{c'}^{6,j-1} \right) + f_c^{7,j} \quad (6)$$

其中， $w_{cc'}^{6,j-1}$ ， $b_c^{6,j-1}$  为模型中的权值系数和偏置系数。

此外，将归一化的QP信息以数值的形式拼接在全连接层的预测矢量上，实现QP在网络中的非线性融合，强化CU划分结构预测网络在不同QP值下的预测性能。拼接3级CU划分标识符得到CU划分表征矢量  $\hat{y}$ ，对其进行优化处理(若父CU划分标识符为0，则其子CU划分标识符将被置为0)后作为网络判别结果输出，并被用到后续编码过程中。

### 3.4 模型损失函数设计

为了更好地训练MLFT-CNN，通过整合CU划分深度预测的先验信息以及QP对于视频编码质量的影响，创造性地设计了目标损失函数，用于协同训练网络模型。其目标损失函数  $J$  可定义为

$$\begin{aligned} J = & \alpha_0 \sum_{n=1}^{N_{\text{Depth}0}} [y_n \cdot \ln(\hat{y}_n) + (1 - y_n) \cdot \ln(1 - \hat{y}_n)] \\ & + \alpha_1 \sum_{n=N_{\text{Depth}0}+1}^{N_{\text{Depth}0}+N_{\text{Depth}1}} [y_n \cdot \ln(\hat{y}_n) + (1 - y_n) \\ & \cdot \ln(1 - \hat{y}_n)] \cdot \max[\text{sgn}(\hat{y}_{\text{ceil}[(n-1)/4]} - \text{th}_0), 0] \\ & + \alpha_2 \sum_{n=N_{\text{Depth}0}+N_{\text{Depth}1}+N_{\text{Depth}2}}^{N_{\text{Depth}0}+N_{\text{Depth}1}+N_{\text{Depth}2}} [y_n \cdot \ln(\hat{y}_n) + (1 - y_n) \\ & \cdot \ln(1 - \hat{y}_n)] \cdot \max[\text{sgn}(\hat{y}_{\text{ceil}[(n-5)/4]} - \text{th}_1), 0] \\ & \cdot \max[\text{sgn}(\hat{y}_{\text{ceil}[(n-5)/16]} - \text{th}_0), 0] \quad (7) \end{aligned}$$

其中， $a_j$  ( $j \in \{0, 1, 2\}$ ) 表示深度为  $j$  的损失函数的权重值， $\text{th}_0$  和  $\text{th}_1$  为在CU深度为0, 1时，判别其是

否划分的阈值，其中  $y_n$  表示第  $n$  位划分标识符的真实值。

## 4 实验与结果讨论

### 4.1 实验数据与实验环境

本文实验环境对于实验数据的获取有重大影响，本文的实验是基于HEVC参考软件HM16.5<sup>[4]</sup>开展的，所有配置采用默认帧内编码配置，即采用给定的encoder\_intra\_main.cfg<sup>[16]</sup>进行实验数据的测定。本文提出的网络是在CPH-Intra<sup>[17]</sup>数据集上训练并验证的，该数据集是由2880×1920, 1536×1024以及768×512 3种分辨率各500张图像视频序列在4个QP{22, 27, 32, 37}值下的110, 105, 784条帧内编码样本构成的。整个数据集中，CU划分与不划分的样本数量分别占总数的49.2%和50.8%。其中各QP值的85%构成训练集，10%构成测试集，5%构成验证集。此外，为了更好地展示本文网络对于不同分辨率的测试序列都有着较好的预测性能，本文对JCT-VC给出的5类(A, B, C, D, E)不同分辨率下的10个标准序列<sup>[18]</sup>在4个QP值(22, 27, 32, 37)进行了验证实验。为了对编码复杂度有一致的衡量标准，所有的实验是在Intel(R) Core(TM) i5-8300H CPU @2.3 GHz, 8.00 GB RAM以及64位Windows 10操作系统的计算机上进行的。

由于MLFT-CNN是针对HEVC帧内编码模式的快速编码算法，本文选取了2种目前在帧内编码具有代表性的算法进行比较。其中文献[9]是目前机器学习方法中最具代表性的算法，文献[14]是采用卷积神经网络的代表性算法。在检验所提模型性能时，采用提出算法相对于原始HM16.5编码算法的时间节省百分比来衡量编码复杂度的降低程度，可用式(8)表示，并采用了Bjontegaard比特率差值(Bjontegaard Delta Bit Rate, BDBR)和Bjontegaard峰值信噪比差值(Bjontegaard Delta Peak Signal-to-Noise Rate, BD-PSNR)对视频编码性能进行评定<sup>[19]</sup>。其中，BDBR表示在同样的客观质量下，两种方法的码率节省情况，BD-PSNR表示在给定的同等码率下，两种方法的亮度峰值信噪比的差异。

$$\Delta T = \frac{T_{\text{Proposed}} - T_{\text{HM-16.5}}}{T_{\text{HM-16.5}}} \times 100\% \quad (8)$$

此外，本文是基于Tensorflow深度学习框架实现所有深度学习方法。在模型训练中使用Adam优化算法<sup>[20]</sup>进行参数优化，并设置初始学习率为  $3 \times 10^{-4}$ ，0.9的1阶矩估计的指数衰减率和0.999的2阶矩估计的指数衰减率为默认超参数，同时采用学习率衰减和Dropout优化技术避免模型过拟合。非线性激活函数使用Leaky-ReLU函数。

## 4.2 结果分析与讨论

表1和表2分别展示了MLFT-CNN在CPH-Intra测试集和HEVC标准测试序列上的编码复杂度下降程度和编码性能差异。由表1可知,在QP值为22, 27, 32, 37时, MLFT-CNN在编码复杂度上分别平均降低了68.51%, 73.34%, 74.28%, 76.56%, 与文献[9]和文献[14]方法相比, MLFT-CNN算法编码复杂度下降得更多。需要说明的是, 文献[14]虽然也为利用深度卷积神经网络的编码复杂度降低算法, 且其在网络模型的构建上, 参数更少, 但由于其在判别是否需要从 $64 \times 64$ 分裂到 $32 \times 32$ 时, 需要进行

RDO搜索, 所以其在编码复杂度降低性能上不如MLFT-CNN。此外, 文献[14]和本文所提出的MLFT-CNN对于不同的序列, 其 $\Delta T$ 的标准差都小于 $\Delta T$ 平均值的1/10, 可以证明在降低编码复杂度性能上两种算法都有着较好的稳定性。为了更加直观地对比3种算法在标准测试序列上的性能, 图5(a)展示了3种方法在不同QP值下的编码时间节省百分比(中间值为平均值, 上下的偏置为标准差)。从表1中展示出来的算法性能差异在表2中得到进一步证实, 在不同QP值, 不同标准视频测试序列上, 平均降低了70.96%的编码复杂度。

表1 图像测试序列结果

训练集	分辨率	方法	BD-BR(%)	BD-PSNR(dB)	$\Delta T(\%)$			
					QP=22	QP=27	QP=32	QP=37
CPH-Intra	768×512	文献[9]	5.113	-0.343	-59.43	-54.70	-48.74	-44.83
		文献[14]	2.885	-0.210	-54.97	-58.78	-61.78	-64.41
		本文算法	1.71	-0.116	-65.18	-72.01	-72.07	-74.83
	1536×1024	文献[9]	6.002	-0.374	-58.94	-54.85	-50.57	-50.95
		文献[14]	3.134	-0.208	-55.84	-59.46	-62.43	-64.17
		本文算法	1.63	-0.113	-66.98	-72.21	-71.10	-74.18
	2880×1920	文献[9]	4.035	-0.207	-57.03	-52.79	-52.31	-59.51
		文献[14]	2.130	-0.115	-59.95	-63.14	-68.07	-69.46
		本文算法	1.3278	-0.075	-70.47	-74.67	-75.82	-77.59
	4928×3264	文献[9]	4.630	-0.209	-58.02	-62.74	-65.30	-67.46
		文献[14]	1.863	-0.086	-61.43	-65.27	-68.70	-71.00
		本文算法	2.085	0.080	-71.42	-74.89	-78.11	-79.63
标准差	文献[9]	0.831	0.088	1.06	4.41	7.52	9.89	
	文献[14]	0.604	0.064	3.13	3.07	3.64	3.49	
	本文算法	0.312	0.093	2.93	1.55	3.27	2.53	
最优值	文献[9]	4.035	-0.207	-59.43	-62.74	-65.30	-67.46	
	文献[14]	1.863	-0.086	-61.43	-65.27	-68.70	-71.00	
	本文算法	1.328	-0.075	-71.42	-74.89	-78.11	-79.63	
平均值	文献[9]	4.945	-0.284	-58.36	-56.27	-54.23	-55.69	
	文献[14]	2.353	-0.155	-58.05	-61.66	-65.25	-67.26	
	本文算法	1.688	-0.096	-68.51	-73.45	-74.28	-76.56	

表2 HEVC标准测试序列结果

类别	序列名称	方法	BD-BR(%)	BD-PSNR(dB)	$\Delta T(\%)$			
					QP=22	QP=27	QP=32	QP=37
A	PeopleOnStreet	文献[9]	9.627	-0.492	-52.12	-50.63	-37.79	-34.81
		文献[14]	3.969	-0.209	-50.79	-53.87	-56.58	-61.15
		本文算法	3.679	-0.216	-63.91	-67.38	-68.78	-70.86
	Traffic	文献[9]	6.411	-0.304	-37.11	-25.36	-19.63	-33.38
		文献[14]	4.945	-0.240	-53.86	-59.08	-63.54	-66.88
		本文算法	3.225	-0.178	-75.45	-77.96	-79.7	-81.12

续表 2

类别	序列名称	方法	BD-BR(%)	BD-PSNR(dB)	$\Delta T(\%)$			
					QP=22	QP=27	QP=32	QP=37
B	Cactus	文献[9]	7.533	-0.248	-38.37	-40.83	-43.61	-51.23
		文献[14]	6.021	-0.208	-58.18	-61.01	-64.94	-67.78
		本文算法	3.634	-0.141	-69.24	-74.67	-74.12	-73.69
	ParkScene	文献[9]	3.630	-0.149	-41.69	-44.79	-59.98	-64.92
		文献[14]	3.417	-0.135	-60.27	-65.10	-68.57	-70.16
		本文算法	2.561	-0.113	-65.03	-70.62	-70.45	-71.46
C	BQMall	文献[9]	9.646	-0.486	-52.62	-42.97	-35.52	-37.12
		文献[14]	8.077	-0.468	-47.08	-51.15	-53.26	-57.05
		本文算法	6.14	-0.395	-62.09	-65.89	-65.86	-69.1
	RaceHorses	文献[9]	7.220	-0.379	-46.46	-40.13	-41.49	-50.28
		文献[14]	4.422	-0.264	-50.52	-59.30	-59.81	-63.15
		本文算法	3.228	-0.217	-64.44	-71.22	-70.17	-72.47
D	BasketballPass	文献[9]	10.054	-0.546	-43.69	-41.03	-37.46	-36.69
		文献[14]	8.401	-0.457	-60.24	-62.89	-64.31	-66.67
		本文算法	4.489	-0.264	-74.99	-77.29	-77.81	-79.36
	BlowingBubbles	文献[9]	6.178	-0.373	-57.15	-42.45	-25.73	-22.81
		文献[14]	8.328	-0.463	-54.62	-60.45	-62.55	-65.48
		本文算法	5.217	-0.315	-61.68	-65.97	-62.99	-66.43
E	FourPeople	文献[9]	9.077	-0.480	-53.52	-40.88	-26.12	-24.34
		文献[14]	8.002	-0.439	-54.79	-59.79	-64.39	-67.17
		本文算法	4.298	-0.258	-65.21	-69.51	-70.94	-71.98
	Johnny	文献[9]	12.182	-0.474	-58.29	-60.21	-63.98	-70.70
		文献[14]	7.956	-0.307	-62.92	-65.51	-67.71	-70.05
		本文算法	4.162	-0.176	-72.02	-74.84	-75.35	-76.12
方差	文献[9]	2.444	0.127	7.68	8.76	14.24	16.18	
	文献[14]	2.013	0.127	5.04	4.51	4.78	4.08	
	本文算法	1.048	0.084	5.16	4.48	5.20	4.50	
最优值	文献[9]	3.63	-0.149	-58.29	-60.21	-63.98	-70.70	
	文献[14]	3.417	-0.135	-62.92	-65.51	-68.57	-70.16	
	本文算法	2.561	-0.113	-75.45	-77.96	-79.7	-81.12	
平均值	文献[9]	8.156	-0.393	-48.10	-42.93	-39.13	-42.63	
	文献[14]	6.354	-0.319	-55.33	-59.82	-62.57	-65.56	
	本文算法	4.063	-0.227	-67.41	-71.54	-71.62	-73.26	

由于编码复杂度的降低往往是通过牺牲编码质量实现的, 评价降低编码复杂度算法的好坏, 还要进行视频编码质量的比较。从表1和表2可以看出, 本文所提MLFT-CNN尽管在图像测试序列上BD-BR平均增加1.688%, BD-PSNR平均造成-0.096 dB的性能下降。在HEVC标准测试序列BD-BR平均增加 4.063%, BD-PSNR平均造成-0.227 dB的性能下降, 但在两种测试序列上性能都要优于文献[9]和文献[14]。图5(b)和(c)分别给出了标准测试序列不同类别BD-BR和BD-PSNR对比图。

综上所述, MLFT-CNN要优于上述两种算法。之所以有这样好的性能表现, 是因为本文提出的网络结构, 基于CU划分过程的先验信息能够更好地拟合CU四叉树划分过程, 因此能更好地学习CU划分过程。

此外, 在实验过程中发现各算法在图像测试序列上的性能都要优于在标准测试序列上的测试结果, 原因在于CPH-Intra数据集中由风景、人物等几类风格相似的图像组成的、纹理相似的CU会出现在训练、测试以及验证集中, 而在预测标准测试

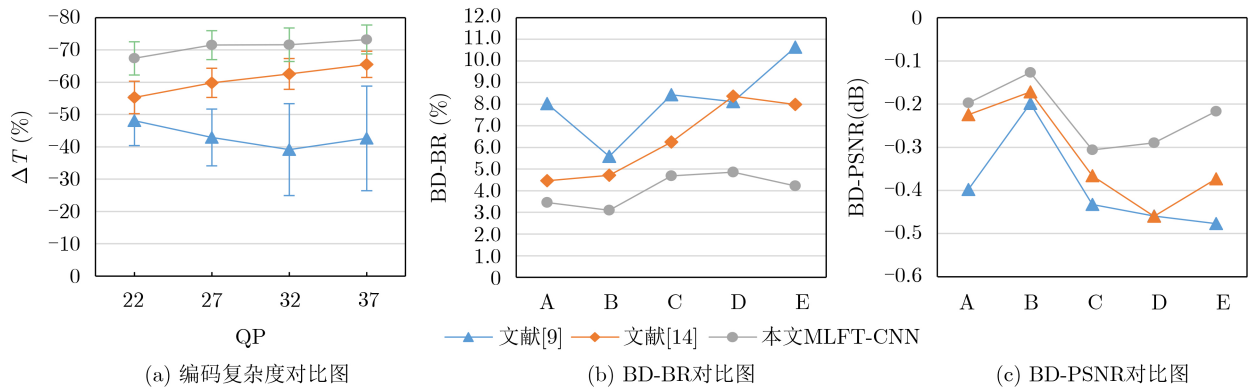


图5 标准测试序列编码性能对比图

序列中一些室内以及近距离场景图像的纹理信息时预测准确度降低。因此,各算法在数据集上的表现要优于HEVC标准测试序列。不过,MLTF-CNN在标准测试序列上也有上佳的表现,证明了MLTF-CNN能够更好地学习CU划分的深层特征,具有更好的鲁棒性。

## 5 结束语

本文提出了一种针对HEVC帧内编码模式下CU划分表征矢量预测的MLFT-CNN网络,为高实时性多媒体系统的视频编码应用要求提供了切实可行的技术方案。主要贡献在于利用视频编码过程的先验信息设计模型,其中提出的降分辨率特征提取模块可有效减少模型参数,改进的通道注意力机制提升了特征的纹理表达能力,设计的特征传递机制强化了不同深度CU特征之间的关联性,并在此基础上设计了具有分段表达能力的目标损失函数,多管齐下共同加速了HEVC的帧内编码过程。通过实验结果可以看出,本文提出的MLFT-CNN在图像视频序列和HEVC不同级别的标准测试序列上都有着卓越的性能表现,在标准测试序列上,编码复杂度更是平均下降了70.96%,证明了本文算法的有效性。此外,通过与其他算法比较,表明了本文所提出的算法是一种有效深度融合编码过程先验信息的CU划分结构预测模型,可有效缓解HEVC编码复杂度与编码质量的内在矛盾,对于HEVC编码标准在多媒体设备上的应用有着重要意义。

## 参考文献

- [1] 王莉,曹一凡,杜高明,等.一种低延迟的3维高效视频编码中深度建模模式编码器[J].电子与信息学报,2019,41(7):1625-1632. doi: 10.11999/JEIT180798.  
WANG Li, CAO Yifan, DU Gaoming, et al. A low-latency depth modelling mode-1 encoder in 3D-high efficiency video coding standard[J]. *Journal of Electronics & Information Technology*, 2019, 41(7): 1625-1632. doi: 10.11999/JEIT180798.
- [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.
- [3] KIM I K, MIN J, LEE T, et al. Block partitioning structure in the HEVC standard[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2012, 22(12): 1697-1706. doi: 10.1109/TCSVT.2012.2223011.
- [4] JCT-VC.HMSoftware[EB/OL].[https://hevc.hhi.fraunhofer.de/svn/svn\\_HEVCSoftware/tags/HM-16.5/](https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.5/), 2014.
- [5] CORREA G, ASSUNCAO P, AGOSTINI L, et al. Performance and computational complexity assessment of high-efficiency video encoders[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2012, 22(12): 1899-1909. doi: 10.1109/TCSVT.2012.2223411.
- [6] 齐美彬,陈秀丽,杨艳芳,等.高效率视频编码帧内预测编码单元划分快速算法[J].电子与信息学报,2014,36(7):1699-1705. doi: 10.3724/SP.J.1146.2013.01148.  
QI Meibin, CHEN Xiuli, YANG Yanfang, et al. Fast coding unit splitting algorithm for high efficiency video coding intra prediction[J]. *Journal of Electronics & Information Technology*, 2014, 36(7): 1699-1705. doi: 10.3724/SP.J.1146.2013.01148.
- [7] 汤进,彭勇.基于时空相关与纹理特性的HEVC编码单元快速划分算法[J].计算机与数字工程,2019,47(7):1753-1756,1782. doi: 10.3969/j.issn.1672-9722.2019.07.038.  
TANG Jin and PENG Yong. Fast coding unit partition algorithm for HEVC based on temporal-spatial correlation and texture property[J]. *Computer and Digital Engineering*, 2019, 47(7): 1753-1756, 1782. doi: 10.3969/j.issn.1672-9722.2019.07.038.
- [8] BOUAAFIA S, KHEMIRI R, SAYADI F E, et al. Fast CU partition-based machine learning approach for reducing HEVC complexity[J]. *Journal of Real-Time Image Processing*, 2020, 17(1): 185-196. doi: 10.1007/s11554-019-

- 00936-0.
- [9] LIU Deyuan, LIU Xingang, and LI Yayong. 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, Auckland, New Zealand, 2016: 594–597.
- [10] LIU Xingang, LI Yayong, LIU Deyuan, *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. doi: [10.1109/TCSVT.2017.2777903](https://doi.org/10.1109/TCSVT.2017.2777903).
- [11] FENG Zeqi, LIU Pengyu, JIA Kebin, *et al.* Fast intra CTU depth decision for HEVC[J]. *IEEE Access*, 2018, 6: 45262–45269. doi: [10.1109/ACCESS.2018.2864881](https://doi.org/10.1109/ACCESS.2018.2864881).
- [12] LIU Zhenyu, YU Xianyu, CHEN Shaolin, *et al.* CNN oriented fast HEVC intra CU mode decision[C]. 2016 IEEE International Symposium on Circuits and Systems, Montreal, Canada, 2016: 2270–2273.
- [13] LI Xin and GONG Na. Run-time deep learning enhanced fast coding unit decision for high efficiency video coding[J]. *Journal of Circuits, Systems and Computers*, 2020, 29(3): 2050046. doi: [10.1142/S0218126620500462](https://doi.org/10.1142/S0218126620500462).
- [14] LIU Zhenyu, YU Xianyu, GAO Yuan, *et al.* CU partition mode decision for HEVC hardwired intra encoder using convolution neural network[J]. *IEEE Transactions on Image Processing*, 2016, 25(11): 5088–5103. doi: [10.1109/tip.2016.2601264](https://doi.org/10.1109/tip.2016.2601264).
- [15] HU Jie, SHEN Li, SUN Gang, *et al.* Squeeze-and-excitation networks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 7132–7141.
- [16] XU Mai, DENG Xin, LI Shengxi, *et al.* Region-of-interest based conversational HEVC coding with hierarchical perception model of face[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2014, 8(3): 475–489. doi: [10.1109/jstsp.2014.2314864](https://doi.org/10.1109/jstsp.2014.2314864).
- [17] 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).
- [18] 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. doi: [10.1109/tcsvt.2012.2221192](https://doi.org/10.1109/tcsvt.2012.2221192).
- [19] BJONTEGARD G. Calculation of average PSNR differences between RD-curves[C]. The 13th Video Coding Experts Group Meeting, Austin, USA, 2001: VCEG-M33.
- [20] KINGMA D P and BA J. Adam: A method for stochastic optimization[C]. The 3rd International Conference on Learning Representations, San Diego, USA, 2015: 1–15.

贾克斌：男，1962年生，教授，研究方向为多媒体信息系统、模式识别。

崔腾鹤：男，1996年生，硕士生，研究方向为视频编码。

刘鹏宇：女，1979年生，副教授，研究方向为多媒体信息系统。

刘 畅：女，1994年生，博士生，研究方向为3D视频编码。

责任编辑：马秀强