采用轻量级卷积神经网络的 H.266/通用视频编码跨分量预测

邹承益[1]，万帅[1,2]，朱志伟[1]，尹宇杰[1]
（1. 西北工业大学电子信息学院，710129，西安；
2. 皇家墨尔本理工大学工程学院，VIC3001，澳大利亚墨尔本）

摘要：为提高新一代通用视频编码标准（H.266/VVC）中色度帧内预测的准确度，提出了采用轻量级卷积神经网络的跨分量预测方法。设计了亮度模块和边界模块，从亮度和色度参考样本中提取特征。设计了注意力模块，构建当前亮度参考样本和边界亮度参考样本之间的空间关系，并应用于边界色度参考样本生成色度预测样本。为降低编解码复杂度，设计网络在二维完成特征融合和预测，优化了现有的同组参数处理不同块大小的训练策略。并且，引入宽度可变卷积，根据不同的块大小调整网络参数。实验结果表明：与 H.266/VVC 测试模型 VTM18.0 相比，所提网络在 Y（亮度分量）、Cb（蓝色色度分量）、Cr（红色色度分量）上分别实现了 0.30%、2.46%、2.25%的码率节省。与其他基于卷积神经网络的跨分量预测方法相比，有效地降低了网络参数和推理复杂度，分别节省了约 10%的编码时间和 19%的解码时间。

关键词：通用视频编码；跨分量预测；轻量级卷积神经网络；注意力机制；宽度可变卷积
中图分类号：TN919.8 **文献标志码**：A
DOI：10.7652/xjtuxb202502018 **文章编号**：0253-987X(2025)02-0180-09

Cross-Component Prediction for H.266/Versatile Video Coding Based on Lightweight Convolutional Neural Network

ZOU Chengyi[1], WAN Shuai[1,2], ZHU Zhiwei[1], YIN Yujie[1]
(1. School of Electronic and Information, Northwestern Polytechnical University, Xi'an 710129, China;
2. School of Engineering, Royal Melbourne Institute of Technology, Melbourne VIC3001, Australia)

Abstract: To improve the accuracy of intra chroma prediction in H.266/versatile video coding (VVC), a cross-component prediction method based on lightweight convolutional neural network was proposed in this paper. The luma module and chroma module were designed to extract features from luma and chroma reference samples, and the attention module was designed to leverage the attention mechanism to construct the spatial correlation between the current luma reference samples and the boundary luma reference samples. Finally, the attention mask was applied to the boundary chroma reference samples to generate chroma prediction value. To reduce the encoding and decoding complexity, the feature fusion and prediction in the network were achieved in two dimensions, the existing training strategy with shared parameters to handle variable block sizes was improved, and slimmable convolutions were introduced to adjust network

收稿日期：2024-05-09。 作者简介：邹承益（1996—），男，博士生；万帅（通信作者），女，教授，博士生导师。 基金项目：陕西省自然科学基础研究计划资助项目（2024JC-YBMS-463）；TCL 科技创新基金资助项目。
网络出版时间：2024-09-12 网络出版地址：https://link.cnki.net/urlid/61.1069.T.20240912.1458.004

parameters according to different block sizes. The experimental results show that the proposed algorithm achieved 0.30%/2.46%/2.25% BD-rate reduction on the Y/Cb/Cr component, respectively, compared with the H.266/VVC test model VTM18.0. Compared with other convolutional neural networks-based cross-component prediction methods, the proposed method effectively reduced the network parameters and inference complexity, saving 10% encoding time and 19% decoding time.

Keywords: versatile video coding; cross-component prediction; lightweight convolutional neural network; attention mechanism; slimmable convolution

近年来，随着互联网的迅速发展，4K/8K 超高清视频、高帧率视频、大动态范围视频、全景视频等视频业务层出不穷，相关应用对高性能视频编码技术的需求与日俱增。如何进一步降低编码比特率，同时保持高视频质量，一直是学术界和工业界的研究热点。新一代视频编码标准 H.266/通用视频编码(versatile video coding，VVC)是最新的视频编码标准之一，相比于上一代视频编码标准 H.265/高效视频编码(high efficiency video coding，HEVC)[1]，在保证相同视频图像质量的前提下，可节省近 50%的码率，满足了对高效视频压缩的需求，并支持当今更广泛的媒体内容和新兴应用[2-4]。

H.266/VVC 提高帧内编码效率的方法之一是为亮度分量和色度分量引入独立的划分结构。由于亮度块是在色度块之前编码的，所以编码的亮度信息可以用于色度预测，以进一步减少分量之间的冗余。为实现这一点，H.266/VVC 利用了重构的亮度信息进行跨分量预测。跨分量线性模型(cross-component linear model，CCLM)[5]是一种有效的色度分量编码工具，它建立了一个跨分量线性预测模型，并将该模型应用于下采样的亮度重建样本来预测色度。然而，简单的线性映射无法准确地表示分量之间的关系，在预测复杂的情况下性能有限。

近年来，卷积神经网络(convolutional neural network，CNN)已成功用于跨分量预测[6-15]。文献[6]首次提出了一种用于跨分量预测的混合神经网络。基于此，文献[7]提出了变换域损失以获得更高的预测性能。文献[8]研究了超参数的影响，并实现了计算效率和压缩性能之间的平衡。文献[9]提出在编码树单元(coding tree unit，CTU)级别使用深度神经网络进行跨分量预测。自注意力机制可用于评估特定输入变量对输出的影响，文献[10]将这个概念成功地扩展到跨分量预测，以评估每个边界参考像素对当前块像素的影响。针对文献[10]中的方法，文献[11]提出了使用单个模型预测不同块大小数据的训练方法以减少模型数量。文献[12]提出了空间信息细化，进一步提高了编码性能。文献[13]提出了一个轻量级卷积神经网络实现了更低的编解码复杂度。基于文献[13]中的方法，文献[14]提出了样本自适应方法进一步提高性能，并提出了简化卷积的方法进一步减少推理参数。

然而，上述方法中的网络是在高维空间完成亮度和色度特征的融合，需要额外的卷积层将融合后的特征映射到二维色度空间，复杂度较高。而且，上述方法均采用单模型训练所有块大小，该策略虽然可以节省总参数量，但大块和小块采用同样的参数量并不合理。相对于大块，小块通常内容更为简单且容易被预测，应采用更少的参数。本文提出了一种基于注意力的轻量级卷积神经网络用于跨分量预测，在保证网络预测性能的前提下显著降低了网络参数量。并且，改变现有的同组参数处理不同块大小的训练策略，提出了利用宽度可变卷积根据不同的块大小调整网络参数，减少了小块的推理复杂度。实验结果表明，与 H.266/VVC 测试模型 VTM18.0 相比，在相同重建视频质量下，本文所提出的方法在在 Y(亮度分量)、Cb(蓝色色度分量)、Cr(红色色度分量)分量上可分别节省 0.30%、2.46%和 2.25%的码率。与其他基于卷积神经网络的跨分量预测方法相比，本文网络在提高色度编码性能的同时，实现了更少的编解码时间。

1　相关研究

1.1　跨分量线性模型

YCbCr 颜色空间的 3 个分量间存在相关性，去除分量间的冗余可以进一步提高视频编解码性能。视频编码中重建的亮度像素可以用于色度帧内预测。文献[5]中首次引入了亮度和色度分量之间的线性模型，其中线性模型的参数可以通过显式传输或隐式推导获得。基于高效视频编码框架，文献[15]设计了几种改进的色度亮度预测算法，称为跨分量预测，具有优异的性能。文献[16]建议仅使用左邻居或上邻居采样来推断线性模型，文献[17]中的工作提出

了一种自适应模板选择方法来扩展相邻像素采样，其中 Cr 分量可以从 Y 分量或 Cb 分量预测，或者从 Y 分量和 Cb 分量进行预测。在文献[18]中，相邻像素的加权和被用作当前色度像素的预测，其中加权因子由亮度和色度之间的相关性确定。文献[19]对同一位置的亮度块和色度块进行模板匹配，然后使用匹配的亮度块的同一位置色度块作为当前色度块的预测器。在文献[20]中，引入了一种使用亮度残差信号来补偿色度残差信号的去相关方法。

为进一步提高跨分量线性模型的预测能力，在文献[21-22]中，提出了多模型的 CCLM(multi-model based cross-component linear model，MMLM)方法。该方法首先将当前块的重建亮度值划分为组，并将线性模型应用于每组。其次，设计了更多的亮度下采样滤波器。最后，提出了一种结合角度帧内预测和 CCLM 预测的预测方法。文献[23]提出了多种梯度下采样滤波器，利用亮度梯度构建与色度之间的梯度线性模型。文献[24]提出了卷积跨分量模型，使用相邻模板构建亮度和色度之间多对一的关系。

1.2 基于神经网络的跨分量预测

与简单的线性模型无法准确描述亮度和色度分量之间的相关性相比，神经网络善于建模更复杂的非线性映射，如图 1 所示。

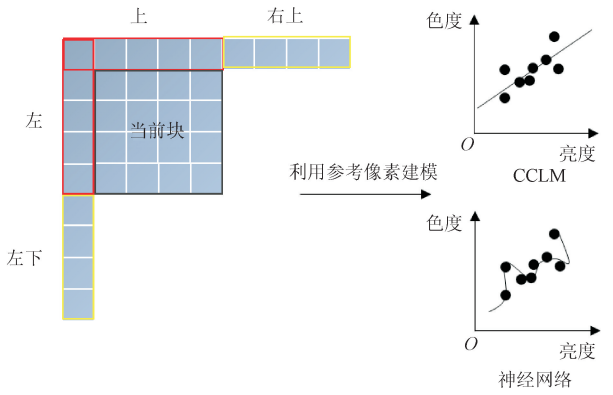


图 1 线性模型和神经网络建模对比

Fig. 1 Comparison between linear modeling and neural network modeling

文献[6]首次将混合神经网络引入色度帧内预测，该网络利用卷积神经网络和全连接网络分别从重构的亮度样本和相邻重构的亮度和色度样本中提取特征。然后，将这两个特征进行融合以预测色度样本。基于这种方法，文献[7]提出了一个变换域损失函数，有利于熵编码过程。此外，文献[7]还研究了超参数的影响，以实现计算效率和压缩性能之间的平衡。文献[8]提出了一种不同的卷积神经网络，它使用亮度和色度样本作为网络的两个模块的输入。此外，针对基于网络的帧内预测模式，提出了一种新的信令方案，并证实了绝对变换差分之和损失函数对训练预测网络的好处。文献[9]提出通过更深的神经网络和更多的参考像素来预测 CTU 级别的色度，小于 CTU 的块将复制位于同一位置的预测。CCLM 的预测值被生成作为色度初始化，并且编码失真水平被引入作为网络的输入。然而，这种方法需要修改 VVC 的编码和解码流程，因为该网络是在 CTU 级别应用的，并且预测值需要传输到编码单元级别。

注意力网络被广泛用于不同类型的深度学习任务，以提高神经网络的性能。文献[10]首次提出一种基于注意力的色度预测神经网络，该网络使用注意力机制来建模参考样本和预测样本之间的空间关系。该网络由 4 个模块组成。前两个模块(跨分量边界模块和亮度卷积模块)从相邻的重构样本和并置的重构亮度块中提取跨分量信息和亮度空间信息。前两个模块的输出特征由第 3 个基于注意力的模块融合，最后一个预测头模块产生色度预测值。基于这项工作，文献[11]提出了一个多模型用于处理可变的块大小并简化推理过程。此外，文献[11]还提出了几种简化方案来进一步降低原始多模型的复杂性，包括卷积运算的复杂性降低框架、使用稀疏自动编码器的简化跨分量边界模块以及具有整数精度近似的算法。为进一步改进色度预测，文献[12]提出了两种基于现有注意力结构的空间信息细化方案，包括增加下采样模块和位置图。文献[13]提出了面向轻量级的卷积神经网络，并将 Cb 和 Cr 分量分别进行预测，然而这样大大提高了编解码复杂度。基于文献[13]中的网络，文献[14]提出了样本自适应方法利用更多的边界参考样本以进一步提高预测性能，还提出了简化卷积的方法进一步减少推理参数以降低编解码复杂度。然而，上述方法的网络都是在高维空间完成特征融合，再通过预测模块降维以完成最终预测，且大块和小块均采用相同的参数进行训练和推理，这导致网络设计和针对不同块大小的训练策略都有进一步的简化空间。

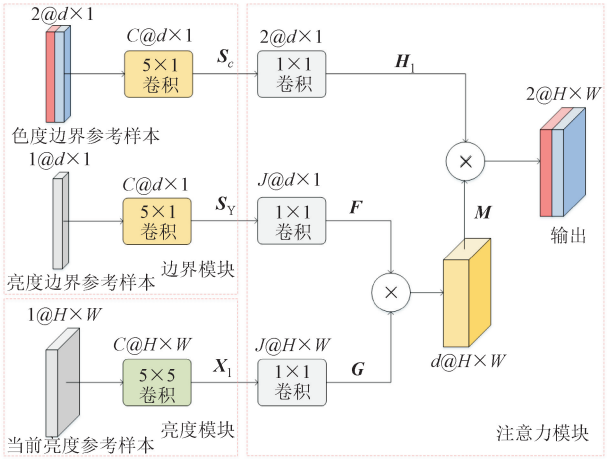
2 采用轻量级卷积神经网络的 H.266/VVC 跨分量预测

本节首先详细介绍了所提出的跨分量预测网络，然后提出了一种利用可变卷积训练不同块大小

的策略，以降低复杂度。

2.1　基于注意力的轻量级卷积神经网络

基于注意力的轻量级卷积神经网络框架如图 2 所示。该网络由边界模块、亮度模块、注意力模块组成。与传统色度预测模式[25]不同，Cb 和 Cr 分量的边界样本一起输入。这样，两个色度分量只需要预测一次，而不需要预测两次，可以显著降低推理复杂度。此外，与其他色度预测网络不同，亮度和色度边界样本是单独输入的，与文献[10-12]相比，这可以更好地从不同的参考样本中提取特征，同时节省了注意力模块的复杂度。与文献[13]和[14]相比，在提取特征后直接映射到二维预测，不需要额外的预测模块得到 Cb 和 Cr 分量的预测值，从而大大降低了网络复杂度。



S_c、S_Y、X_1、H_1、F、G、M— 特征图；$C@H\times W$— 输出特征图大小为通道 C、高 H、宽 W；d— 边界长度。

图 2　基于注意力的轻量级卷积神经网络框架

Fig. 2　Attention-based lightweight convolutional neural network framework

由于在视频编码中经常使用 YCbCr 4∶2∶0 格式，因此在预测同一位置的色度块之前对重建的亮度块进行下采样。值得注意的是，这里使用与 CCLM 相同的传统下采样操作，并且下采样的亮度块 $\boldsymbol{X}\in\mathbb{R}^{H\times W}$ 的大小设为 $H\times W$。在色度预测的过程中，拼接当前块的上边界和左边界作为亮度和色度边界参考样本 $\boldsymbol{B}_Y\in\mathbb{R}^{d}$、$\boldsymbol{B}_c\in\mathbb{R}^{2\times d}$，其中 d 随着编码顺序和块大小变化而变化，Y 是亮度分量，$c\in\{\mathrm{Cb},\mathrm{Cr}\}$ 是当前色度分量。

边界模块和亮度模块并行地从边界参考样本和同一位置的亮度参考样本中提取特征。边界模块被分为两个分支，分别从亮度和色度边界参考样本中提取亮度和色度特征。这种方法清楚地分离不同的分量，直到最后的注意力模块，然后可以根据亮度图更有效地包裹色度特征。每个边界分支的输出特征图 $\boldsymbol{S}\in\mathbb{R}^{C\times d}$ 可以表示为

$$\boldsymbol{S}(\boldsymbol{B},\boldsymbol{W},\boldsymbol{b})=\mathrm{ReLU}(\boldsymbol{W}\boldsymbol{B}+\boldsymbol{b}) \tag{1}$$

式中：$\boldsymbol{W}\in\mathbb{R}^{5\times C\times D}$ 和 $\boldsymbol{b}\in\mathbb{R}^{C}$ 代表 5×1 卷积的权重和偏置，D 为输入通道数；ReLU 为激活函数修正线性单元。

亮度模块并行地提取位于同一位置的亮度参考样本 $\boldsymbol{X}$ 上的空间信息。使用步长为 1 的 C 维 5×5 卷积层来获得亮度特征图 $\boldsymbol{X}_1\in\mathbb{R}^{C\times H\times W}$，其在网络中具有最大的卷积核大小。映射函数可以写成

$$\boldsymbol{X}_1(\boldsymbol{X},\boldsymbol{W}_Y,\boldsymbol{b}_Y)=\mathrm{ReLU}(\boldsymbol{W}_Y\boldsymbol{X}+\boldsymbol{b}_Y) \tag{2}$$

式中：$\boldsymbol{W}_Y\in\mathbb{R}^{5\times5\times C}$ 和 $\boldsymbol{b}_Y\in\mathbb{R}^{C}$ 是 5×5 卷积的权重和偏置。

注意力模块通过注意力机制对所有输入特征进行融合，以获得融合特征。与文献[10-12]中融合当前亮度块特征和跨分量边界特征的注意力模块不同，注意力图是根据当前亮度块和亮度边界的特征生成的。这种方法不仅可以更好地预测色度，还减少了对齐特征维度（即 1×1 卷积）的计算需求。与文献[13-14]中的方法类似，注意力模块将亮度参考特征通过 1×1 卷积将映射到更小的空间，从而得到 $\boldsymbol{F}\in\mathbb{R}^{J\times b}$ 和 $\boldsymbol{G}\in\mathbb{R}^{J\times H\times W}$，不同的是色度边界的特征会直接映射到 2 维空间，从而得到 $\boldsymbol{H}\in\mathbb{R}^{2\times b}$，这样可以在注意力模块直接完成最终的预测，而不需要多余的卷积再来实现多维特征到 2 维的映射，从而大大减少了参数量。然后，将 $\boldsymbol{F}$ 和 $\boldsymbol{G}$ 相乘得到注意力图 $\boldsymbol{A}=\boldsymbol{F}^{\mathrm{T}}\boldsymbol{G}$，$\boldsymbol{A}\in\mathbb{R}^{b\times H\times W}$。通过对 $\boldsymbol{A}$ 应用 softmax 函数以生成注意力掩码 $\boldsymbol{M}\in\mathbb{R}^{b\times H\times W}$，表示每个边界位置对当前块的影响。$\boldsymbol{A}$ 中每个值的计算方式为

$$\alpha_{j,i}=\frac{\exp(m_{j,i})}{\sum_{i=0}^{b-1}\exp(m_{j,i})} \tag{3}$$

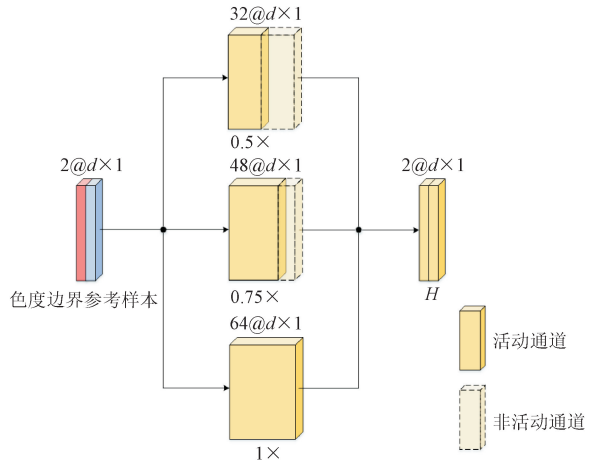
式中：$j=0,\cdots,H\times W-1$ 代表了预测块的样本位置；$i=0,\cdots,b-1$ 代表参考样本位置；$m_{j,i}$ 代表 $\boldsymbol{M}$ 中每个值。通过运算 $\hat{\boldsymbol{X}}=\boldsymbol{H}\boldsymbol{M}$，将掩码作用在色度边界特征上，得到色度预测值 $\hat{\boldsymbol{X}}\in\mathbb{R}^{2\times H\times W}$。

2.2　宽度可变卷积

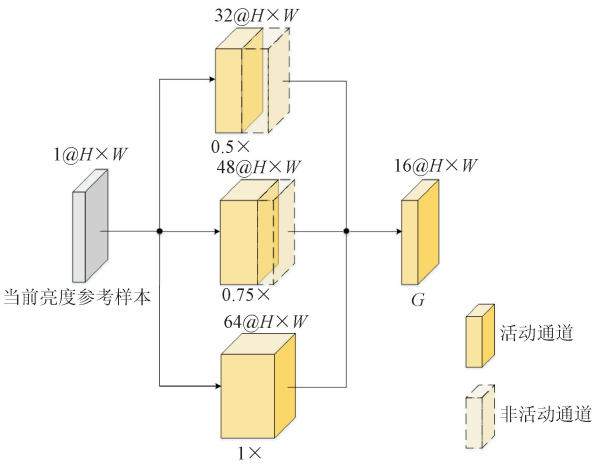
文献[11]提出了块独立的训练策略，可以使用单个网络预测 3 种不同大小的块，文献[14]进一步将这种方法运用在所有尺寸的块上。通常越小的块越容易预测，所以预测小块可以使用比预测大块更少的参数。文献[26]首次提出了宽度可变卷积的概

念,针对不同硬件设备需要训练部署不同模型的问题,只需要训练一个网络就可以根据实际硬件设备的资源限制动态调整卷积宽度。文献[27-28]分别将宽度可变卷积用于基于神经网络的图像压缩和视频压缩,以改变码率失真的权衡并控制复杂度。受这些文献的启发,本文提出采用宽度可变卷积的方法来控制不同块大小所使用的参数量,在保证性能的基础上极大减少了预测小块的推理复杂度。

图 3 展示了具有不同数量活动通道的宽度可变卷积。以色度边界分支和亮度分支为例,同样的模型可以以多种不同的宽度(活动通道数量)运行,并且模型变体的参数是共享的,允许实现精度和效率之间的平衡。本文采用 3 种不同的宽度,1×代表使用所有的通道,0.75×代表使用前 75%的通道,0.5×代表使用前一半的通道。



(a)边界模块的宽度可变卷积



(b)亮度模块的宽度可变卷积

图 3 具有不同活动通道数量的宽度可变卷积

Fig. 3 Slimmable convolutions with different numbers of active channels

本文网络完整的训练过程伪代码如下。

输入:

$\{\boldsymbol{X}_m^N, \boldsymbol{B}_m^N, \boldsymbol{Z}_m^N\}$:当前亮度块,边界参考样本,基准真实值。其中,$m \in [0, M]$ 代表数据数量,$N \in \{16, 32, 64, 128, 256, 512, 1\,024\}$ 代表待预测块的面积。

C:可变宽度卷积的活动通道数

$\theta_N(\boldsymbol{W}^{(t)})$:$N$ 个网络共享权重 $\boldsymbol{W}^{(t)}$

$L_{\text{reg}}^{(t)}$:训练 t 步的目标函数

optimizer$(g^{(t)})$:优化器函数

过程:

$t \leftarrow 0$ 初始化步长

while $\theta_N(\boldsymbol{W}^{(t)})$ 未收敛 do:

for $m \in [0, M)$ do

for $N \in \{16, 32, 64, 128, 256\}$ do

if $16 \leqslant N < 64$ do

$C = 32$

if $64 \leqslant N < 256$ do

$C = 48$

if $256 \leqslant N \leqslant 1\,024$ do

$C = 64$

$t \leftarrow t + 1$

$L_{\text{reg}}^{(t)} \leftarrow \text{MSE}(Z_m^N, \theta_N(\boldsymbol{X}_m^N, \boldsymbol{B}_m^N, \boldsymbol{W}^{(t-1)}))$

$g^{(t)} \leftarrow \nabla_W L_{\text{reg}}^{(t)}$(获取 t 步的梯度)

$\boldsymbol{W}^{(t)} \leftarrow \text{optimizer}(g^{(t)})$

end for

end for

end while

在训练第 t 步时,首先根据预测块的面积确定可变宽度卷积的活动通道数,本文最多使用 64 个活动通道,最少使用 32 和活动通道,即 1×=64,0.75×=48,0.5×=32。确定活动通道数后即可更新网络参数,获得一组新的权重。为了让预测值更接近于原始值,使用最小化均方误差(mean-square error, MSE)作为目标函数来更新网络参数,MSE 公式如下

$$\text{MSE}(y_i, \hat{y}_i) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4}$$

式中:y_i 为真实值;$\hat{y}_i$ 为预测值;n 为像素总数。

本文网络的参数如表 1 所示。由于采用了宽度可变卷积,边界模块和亮度模块的输出通道及注意力模块的输入通道会根据块的大小发生变化。

表 1　所提网络参数

Table 1　Parameters of the proposed model

模块	输入通道数	卷积核大小	输出通道数
边界	2（色度）	5×1	64，48，32
	1（亮度）		64，48，32
亮度	1	5×5	64，48，32
预测	64，48，32	1×1	2
	64，48，32		16
	64，48，32		

2.3　网络训练

训练数据集由来自 DIV2K 数据集[29]的 800 张图像和来自 BVI-DVC 数据集[30]的 800 张图像组成。DIV2K 数据集是一个高质量的自然图像数据集，包含 800 个训练样本和 100 个验证样本。DIV2K 数据库中图像的最大分辨率为 2K((2 040×648)～(2 040×2 040)像素)。该数据库还提供了 3 个较低分辨率版本((1 020×324)～(1 020×1 020)像素，(680×216)～(680×680)像素，(510×162)～(510×510)像素)，它们通过双线性和未知滤波器按因子 2、3、4 进行下采样。对于每个数据实例，随机选择一个分辨率。值得注意的是，选择高分辨率图像的概率被设置为高于低分辨率图像的选择概率，并且图像需要从 PNG 格式转换 YCbCr 4∶2∶0 格式。BVI-DVC 是一个 YCbCr 格式的代表性视频数据库，旨在训练基于 CNN 的视频压缩算法，其中包含 270p～2 160p 的各种空间分辨率的 800 个序列。对于每个序列，随机选择一个帧，然后选择多个 $H\times W$ 大小的块。所有像素值都需要转换为浮点数，并归一化到[0，1]范围内。所有尺寸的块一起输入到网络中训练。

在本文中，Tensorflow/Keras 被用于训练所提出的网络。批量大小设置为 16。所提出的网络使用 Adam 优化器，并以 10^{-4} 的学习率开始训练。为了公平比较复杂度，本文只在训练网络时使用 GPU，在 VTM 推理时使用 CPU。所使用 CPU 为 Intel Core i9-12900K，GPU 为 NVIDIA GeForce RTX 3090。

2.4　H.266/VVC 集成

将所提出的方法作为一种新模式加入到 VTM18.0[31]中，与传统色度预测模式（平面、直流、角度预测、CCLM）进行率失真竞争。集成后，共有 9 种候选预测模式。在编码端，所有候选预测模式通过计算率失真性能代价进行模式选择，选出损失最小的模式。这不仅需要修改预测过程，而且需要额外的二进制块级语法标志来指示给定块是否使用所提出的方法。如图 4 所示，如果最佳模式是新模式，则设置为 1。否则，指定为 0。这样，新模式只需要一个标志，而其他模式有一个以上的标志。与传统模式不同，所提出的预测模式在编码和解码过程中只需一次执行就可以获得两个色度分量（Cb 和 Cr）的预测值，这大大降低了编码和解码的复杂度。

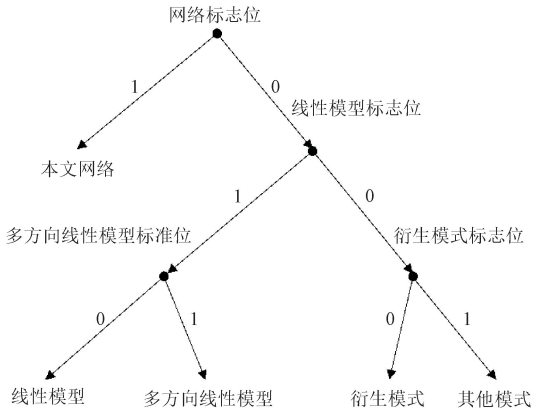


图 4　模式标志方法

Fig. 4　Mode signaling method

3　实验与分析

3.1　编码性能评估

编码性能测试采用全帧内配置配置，量化参数（quantization parameter，QP）设置为通用测试条件（common test conditions，CTC）[32]所建议的 22、27、32、37。采用 BD-rate（Bjøntegaard delta bit rate，用符号 ΔR 表示）来评估相对于 VTM18.0 的编码性能。测试序列包括 CTC 推荐的 22 个不同分辨率的视频序列，分为 5 个类别，分别为 class A（6 个视频序列）、class B（5 个视频序列）、class C（4 个视频序列）、class D（4 个视频序列）和 class E（3 个视频序列）。表 2 总结了本文网络和其他网络在 VTM18.0 的 BD-rate 对比结果。实验结果表明：在相同重建视频质量下，本文网络在 Y、Cb、Cr 上分别平均节省了 0.30%、2.46%和 2.25%的编码码率。这表明本文网络可提高压缩效率，尤其是针对色度分量，并且与文献[11]和文献[14]相比，编码性能更高。

表 2 不同网络的编码性能对比

Table 2 Coding performance comparison of different algorithms

测试序列集	ΔR/%								
	文献[11]网络			文献[14]网络			本文网络		
	Y	Cb	Cr	Y	Cb	Cr	Y	Cb	Cr
class A1	−0.28	−3.20	−1.85	−0.21	−1.56	−2.50	−0.27	−2.47	−2.18
class A2	−0.25	−3.11	−1.54	−0.43	−1.94	−2.18	−0.34	−2.68	−2.07
class B	−0.26	−2.28	−2.33	−0.27	−1.89	−2.59	−0.24	−2.58	−2.52
class C	−0.30	−1.92	−1.57	−0.34	−1.51	−1.47	−0.33	−2.23	−1.95
class D	−0.29	−1.70	−1.77	−0.37	−1.38	−1.20	−0.39	−2.50	−2.00
class E	−0.13	−1.59	−1.45	−0.19	−1.78	−3.12	−0.20	−2.25	−2.78
平均值	−0.26	−2.25	−1.80	−0.30	−1.68	−2.14	−0.30	−2.46	−2.25

3.2 复杂度分析

表 3 展示了网络参数数量以分析网络复杂性。

表 3 不同网络的参数数量比较

Table 3 Comparison of model parameter quantity for different algorithms

网络	参数数量		
	$16 \leqslant N < 64$	$64 \leqslant N < 256$	$256 \leqslant N \leqslant 1\,024$
文献[11]	7 074	7 074	7 074
文献[14]	5 538	5 538	5 538
本文	**4 962**	**2 498**	**1 266**

注:表中加粗数据为最优值。

从表 3 可以看出:对于面积大于等于 256 的块,本文网络所使用的训练参数数量(4 962)小于文献[11]和文献[14]中的网络参数数量(7 074 和 5 538);对于面积小于 256 的块,本文网络所使用的训练参数数量(2 498 和 1 266)远小于文献[11,14]中的网络参数数量(7 074 和 5 538)。这意味着本文提出的神经网络比文献[11,14]中的网络更轻量,且能根据块大小灵活调整参数。

表 4 总结了本文网络和其他网络在 VTM18.0 的编解码复杂度对比结果。其中,文献[14]为目前最先进的卷积神经网络跨分量预测方法之一。实验结果表明:相比于 VTM18.0,本文网络编码和解码时间分别增加了 20%和 297%。与文献[11,14]中的网络相比,本文网络在提高编码性能的同时,实现了更少的编码时间和解码时间。相比文献[14]中的网络,本文网络编码和解码时间分别降低了 10%和 19%。

表 4 不同网络的的编解码复杂度对比

Table 4 Encoding and decoding complexity comparison of different algorithms

测试序列集	文献[11]网络		文献[14]网络		本文网络	
	相对编码时间/%	相对解码时间/%	相对编码时间/%	相对解码时间/%	相对编码时间/%	相对解码时间/%
class A1	172	1 478	140	408	130	314
class A2	176	1 501	143	605	126	476
class B	160	1 551	133	569	121	482
class C	158	943	126	372	113	317
class D	146	1 035	126	466	110	365
class E	159	1 144	138	575	125	428
平均值	160	1 249	133	490	120	397

4 结 论

本文提出了一种基于注意力的轻量级神经网络用于跨分量预测,该网络能够有效地对参考样本和预测样本之间的空间关系进行建模。为降低复杂度,设计网络在二维完成参考样本的特征融合和 Cb 与 Cr 两个色度分量的预测。并且,改变了现有跨分量预测网络的训练策略,采用宽度可变卷积根据块

大小调整网络参数。实验结果表明，新的预测模式有效，相对于最先进的卷积神经网络跨分量预测方法，在节省了比特率的同时，大大减少了编解码复杂度。

在未来的工作中，为进一步提高编码性能，可考虑使用更多的参考信息，并根据内容设计不同的网络进行预测。

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, CHEN Jianle, OHM J R, et al. Developments in international video coding standardization after AVC, with an overview of versatile video coding (VVC) [J]. Proceedings of the IEEE, 2021, 109(9): 1463-1493.

[3] 万帅，霍俊彦，马彦卓，等. 新一代通用视频编码标准 H.266/VVC：现状与发展 [J]. 西安交通大学学报，2024，58(4)：1-17.
WAN Shuai, HUO Junyan, MA Yanzhuo, et al. The new-generation versatile video coding standard H.266/VVC: state-of-the-art and development [J]. Journal of Xi'an Jiaotong University, 2024, 58(4): 1-17.

[4] 万帅，霍俊彦，马彦卓，等. 新一代通用视频编码 H.266/VVC：原理、标准与实现 [M]. 北京：电子工业出版社，2022.

[5] LEE S H, CHO N I. Intra prediction method based on the linear relationship between the channels for YUV 4:2:0 intra coding [C]//2009 16th IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE, 2009: 1037-1040.

[6] LI Yue, LI Li, LI Zhu, et al. A hybrid neural network for chroma intra prediction [C]//2018 25th IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE, 2018: 1797-1801.

[7] LI Yue, YI Yan, LIU Dong, et al. Neural-network-based cross-channel intra prediction [J]. ACM Transactions on Multimedia Computing Communications and Applications, 2021, 17(3): 77.

[8] MEYER M, WIESNER J, SCHNEIDER J, et al. Convolutional neural networks for video intra prediction using cross-component adaptation [C]//2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ, USA: IEEE, 2019: 1607-1611.

[9] ZHU Linwei, ZHANG Yun, WANG Shiqi, et al. Deep learning-based chroma prediction for intra versatile video coding [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(8): 3168-3181.

[10] BLANCH M G, BLASI S, SMEATON A, et al. Chroma intra prediction with attention-based CNN architectures [C]//2020 IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE, 2020: 783-787.

[11] BLANCH M G, BLASI S, SMEATON A F, et al. Attention-based neural networks for chroma intra prediction in video coding [J]. IEEE Journal of Selected Topics in Signal Processing, 2021, 15(2): 366-377.

[12]]ZOU Chengyi, WAN Shuai, JI Tiannan, et al. Spatial information refinement for chroma intra prediction in video coding [C]//2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). Piscataway, NJ, USA: IEEE, 2021: 1422-1427.

[13] ZOU Chengyi, WAN Shuai, MRAK M, et al. Towards lightweight neural network-based chroma intra prediction for video coding [C]//2022 IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE, 2022: 1006-1010.

[14] ZOU Chengyi, WAN Shuai, JI Tiannan, et al. Chroma intra prediction with lightweight attention-based neural networks [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(1): 549-560.

[15] KHAIRAT A, NGUYEN T, SIEKMANN M, et al. Adaptive cross-component prediction for 4∶4∶4 high efficiency video coding [C]//2014 IEEE International Conference on Image Processing (ICIP). Piscataway, NJ, USA: IEEE, 2014: 3734-3738.

[16] ZHANG Xingyu, GISQUET C, FRANÇOIS E, et al. Chroma intra prediction based on inter-channel correlation for HEVC [J]. IEEE Transactions on Image Processing, 2014, 23(1): 274-286.

[17] ZHANG Tao, FAN Xiaopeng, ZHAO Debin, et al. Improving chroma intra prediction for HEVC [C]//2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). Piscataway, NJ, USA: IEEE, 2016: 1-6.

[18] LEE S H, MOON J W, BYUN J W, et al. A new intra prediction method using channel correlations for the

H. 264/AVC intra coding [C]//2009 Picture Coding Symposium. Piscataway, NJ, USA: IEEE, 2009: 1-4.

[19] YEO C, TAN Y, LI Zhengguo, et al. Chroma intra prediction using template matching with reconstructed luma components [C]//2011 18th IEEE International Conference on Image Processing. Piscataway, NJ, USA: IEEE, 2011: 1637-1640.

[20] KIM W S, PU Wei, KHAIRAT A, et al. Cross-component prediction in HEVC [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30 (6): 1699-1708.

[21] ZHANG Kai, CHEN Jianle, ZHANG Li, et al. Multi-model based cross-component linear model chroma intra-prediction for video coding [C]//2017 IEEE Visual Communications and Image Processing (VCIP). Piscataway, NJ, USA: IEEE, 2017: 1-4.

[22] ZHANG Kai, CHEN Jianle, ZHANG Li, et al. Enhanced cross-component linear model for chroma intra-prediction in video coding [J]. IEEE Transactions on Image Processing, 2018, 27(8): 3983-3997.

[23] KUO Chewei, LI Xinwei, XIU Xiaoyu, et al. Gradient linear model for chroma intra prediction [C]//2023 Data Compression Conference (DCC). Piscataway, NJ, USA: IEEE, 2023: 13-21.

[24] ASTOLA P. AHG12: convolutional cross-component model (CCCM) for intra prediction: JVET-Z0064 [EB/OL]. (2022-04-13) [2024-05-01]. https://jvet-experts.org/.

[25] BROSS B, WANG Yekui, YE Yan, 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.

[26] YU Jiahui, YANG Linjie, XU Ning, et al. Slimmable neural networks [C]//7th International Conference on Learning Representations. New York, USA: ICLR, 2019:1-12.

[27] YANG Fei, HERRANZ L, CHENG Yongmei, et al. Slimmable compressive autoencoders for practical neural image compression [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2021: 4996-5005.

[28] LIU Zhaocheng, HERRANZ L, YANG Fei, et al. Slimmable video codec [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2022: 1742-1746.

[29] MA Di, ZHANG Fan, BULL D R. BVI-DVC: a training database for deep video compression [J]. IEEE Transactions on Multimedia, 2022, 24: 3847-3858.

[30] TIMOFTE R, AGUSTSSON E, GOOL L V, et al. NTIRE 2017 challenge on single image super-resolution: methods and results [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2017: 1110-1121.

[31] BROWNE A, YE Y, KIM S. Algorithm description for versatile video coding and test model 18(vtm 18): JVET-AB2002 [EB/OL]. (2022-10-28) [2024-05-01]. https://jvet-experts.org/.

[32] BOYCE J, SUEHRING K, LI L, et al. JVET common test conditions and software reference configurations: JVET-J1010 [EB/OL]. (2018-04-20) [2024-05-01]. https://jvet-experts.org/.

(编辑 陶晴 武红江)