基于轻量级全连接网络的 H.266/VVC 分量间预测

霍俊彦[1]，王丹妮[1]，马彦卓[1]，万帅[2]，杨付正[1]
（1. 西安电子科技大学 ISN 国家重点实验室，陕西 西安 710071；2. 西北工业大学电子信息学院，陕西 西安 710072）

摘 要：新一代视频编码标准 H.266/VVC 引入分量间线性模型（CCLM）预测提高压缩效率。针对亮度色度分量存在相关性却难以建模的问题，提出基于神经网络的分量间预测算法。该算法根据待预测像素与参考像素的亮度差遴选出相关性强的参考像素构成参考子集，然后将参考子集送入轻量级全连接网络获得色度预测值。实验结果表明，与 H.266/VVC 测试模型版本 10.0（VTM10.0）相比，所提算法可提高色度预测准确度，在 Y、Cb 和 Cr 上可分别节省 0.27%、1.54%和 1.84%的码率。所提算法具有不同块尺寸和编码参数均可使用统一网络结构的优点。

关键词：H.266/VVC；色度帧内预测；分量间预测；神经网络
中图分类号：TN911.7
文献标志码：A
DOI: 10.11959/j.issn.1000−436x.2022031

Efficient cross-component prediction for H.266/VVC based on lightweight fully connected networks

HUO Junyan[1], WANG Danni[1], MA Yanzhuo[1], WAN Shuai[2], YANG Fuzheng[1]
1. State Key Laboratory of Integrated Services Network, Xidian University, Xi'an 710071, China
2. School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China

Abstract: Cross-component linear model (CCLM) prediction in H.266/versatile video coding (VVC) can improve the compression efficiency. There exists high correlation between luma and chroma components while the correlation is difficult to be modeled explicitly. An algorithm for neural network based cross-component prediction (NNCCP) was proposed where reference pixels with high correlation were selected according to the luma difference between the reference pixels and the pixel to be predicted. Based on the high-correlated reference pixels and the luma difference, the predicted chroma was obtained based on lightweight fully connected networks. Experimental results demonstrate that the proposed algorithm can achieve 0.27%, 1.54%, and 1.84% bitrate savings for luma and chroma components, compared with the VVC test model 10.0 (VTM10.0). Besides, a unified network can be employed to blocks with different sizes and different quantization parameters.

Keywords: H.266/VVC, chroma intra prediction, cross component prediction, neural network

0 引言

随着信息互联网迅猛发展和智能移动终端广泛普及，海量视频信息不断涌现。4K/8K 超高清视频、全景视频、短视频等新兴视频业务层出不穷。视频业务的蓬勃发展给传输带宽带来了巨大的挑战。以提高视频压缩效率为目标的视频压缩编码技术一直是学术界和工业界研究的热点。近期，远程办公和在线教育需求激增，在有限的网络带宽条件下提供高质量视频服务尤其迫切。

2020 年 8 月，由隶属于 ISO/IEC 的 MPEG 组和隶属于 ITU-T 的 VCEG 组成立的联合视频专家组（JVET, joint video experts team）完成了新一代视频编码标准 H.266/通用视频编码（VVC, versatile video

收稿日期：2021−11−19；**修回日期**：2022−01−24
基金项目：国家自然科学基金资助项目（No.62101409, No.62171353）
Foundation Item: The National Natural Science Foundation of China (No.62101409, No.62171353)

coding）[1]的制定。除传统视频外，该标准还可实现超高清视频[2]、360 视频[3]和宽动态视频[4]的高效压缩编码。相比于上一代视频编码标准 H.265/高效视频编码（HEVC, high efficiency video coding）[5]，H.266/VVC 在保证相同视频图像质量的前提下，可节省近 50%的码率，代表了当前视频编码技术的最高水平。

H.266/VVC 沿用基于块的预测编码、变换编码和熵编码的混合编码框架，其卓越的压缩性能归因于在各个模块引入了大量新的编码技术[6-7]。其中，在块划分上，H.266/VVC 扩大编码树单元（CTU, coding tree unit）尺寸，并允许采用二叉树、三叉树和四叉树对 CTU 进行迭代划分得到编码单元（CU, coding unit）[8]，同时，支持亮度色度独立划分 CU[6]。对于帧内预测模块，H.266/VVC 扩展了角度预测模式，并新增了基于矩阵的帧内预测（MIP, matrix-based intra prediction）技术[9]、帧内子块划分技术[10]及分量间线性模型（CCLM, cross-component linear model）预测技术[11]。对于帧间预测模块，H.266/VVC 引入仿射运动补偿技术[12]、几何划分技术[13]、双向光流补偿技术[1]等新技术。上述技术旨在提高预测块的准确度，降低预测残差。之后在对预测残差进行处理的模块中，H.266/VVC 通过新增多核变换[14]、低频不可分变换[15]和子块变换[14]等技术优化变换模块，同时通过扩展量化参数范围、增添依赖量化技术来优化量化模块。此外，为了进一步改善视频编码质量，在环路滤波模块中，H.266/VVC 引入自适应环路滤波技术[16]。

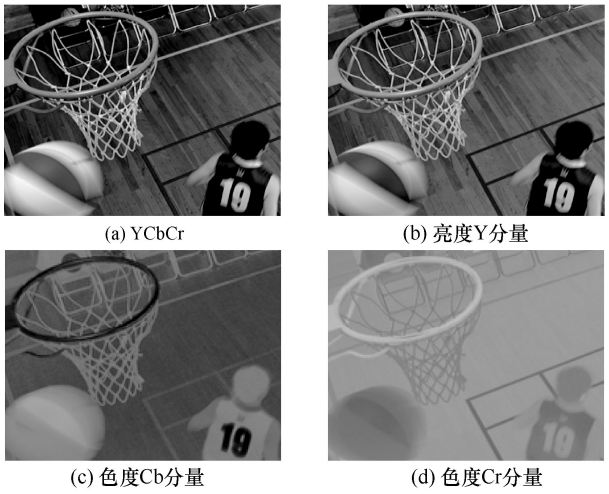
众所周知，基于神经网络（NN, neural network）的算法在计算机视觉任务上取得了巨大成功，已广泛应用于图像分类[17]、目标检测[18]、图像增强[19]等领域。近年来，NN 逐渐渗透到视频编码领域，成为进一步提高压缩效率的有效手段[20-21]。基于 NN 的视频编码大致可分为 2 个方向。第一类是基于 NN 建立全新的视频编码框架，具体分为基于 NN 的图像编码和基于 NN 的视频编码。Minnen 等[22]提出的基于自编码器的图像压缩网络是典型的面向图像的编码方案，通过变换网络及熵模型网络进行图像压缩，有效去除空间冗余。基于 NN 的视频编码方案采用网络实现运动估计和补偿。以深度视频压缩（DVC, deep video compression）模型为例，该模型采用光流估计网络获取帧间运动信息，通过基于自编码器的网络对运动信息和残差信息压缩，达到有效去除时空冗余的目的。第二类是在传统编码框架内利用 NN 设计新的编码工具，具体可针对现有框架中的帧内预测、帧间预测、分量间预测、概率分布预测、变换、环路滤波、上/下采样等技术进行改进，取代传统框架中对应的工具或引入新的工具，实现更高的压缩效率。深度学习视频编码（DLVC, deep learning video coding）模型通过在传统编码框架中引入多项深度编码工具提升了传统框架的压缩效率。上述算法可利用现有 NN，如多层感知器（MLP, multi layer perceptron）、随机神经网络、卷积神经网络（CNN, convolutional neural network）、递归神经网络（RNN, recurrent neural network）和生成式对抗网络（GAN, generative adversarial network）等，根据视频编码的特性进行网络架构设计，已展现出了在视频编码领域的可期前景。总之，一方面，NN 具有强大的非线性拟合能力，可有效提高视频压缩效率；另一方面，NN 计算复杂度相当高，在与传统编码框架结合时需要在编码性能与复杂度之间进行优化和折中，如 H.266/VVC 已采纳的 MIP 技术就是源于 NN 设计并合理简化后的帧内预测算法，这为在传统视频编码框架下开展基于 NN 的算法设计提供了可行思路。本文重点针对 NN 与分量间预测结合的内容展开研究，已有算法的详细介绍见 1.2 节。

本文面向 H.266/VVC，提出一种基于轻量级 NN 的分量间预测（NNCCP, neural network based cross-component prediction）算法，通过 NN 构造准确度高的色度预测值，从而提高视频压缩效率。通常，在图像的局部区域内，若像素间亮度差值越小，像素相关性越强，其色度相关性也越强。基于该现象，本文利用亮度差值，从参考区域中提取固定数量的参考像素组成参考子集。进一步将该参考子集与待预测像素的亮度差向量和该参考子集的参考色度向量输入色度预测模块构造色度预测值。由于参考子集的元素数量固定，色度预测模块可针对 H.266/VVC 各种尺寸的 CU 使用统一的神经网络进行处理。将 NNCCP 集成至 H.266/VVC 参考软件 VTM10.0[23]，并通过实验验证其编码性能的提升。实验结果表明，NNCCP 算法可提高色度预测准确度，有效提升 H.266/VVC 的压缩效率。

1　相关研究

1.1　H.266/VVC 高效色度帧内预测

H.266/VVC 的色度帧内预测算法大致可分为三类，第一类为默认传统预测模式，包括 PLANAR、DC、水平和垂直 4 种模式，原理是根据前面已编码块的重建色度预测当前块的色度分量；第二类为亮度推导模式，该模式借用对应位置亮度的帧内预测模式作为色度的帧内预测模式；第三类为 H.266/VVC 新引入的 CCLM 模式，该模式利用同位置的重建亮度值通过线性模型计算色度预测值。通常，YCbCr 颜色空间的各个分量之间存在较强的相关性，如图 1 所示。因此，利用分量间相关性设计算法是提高压缩效率的有效手段，上述第二类和第三类色度帧内预测模式皆是基于分量间相关性所设计的。



(a) YCbCr　(b) 亮度Y分量　(c) 色度Cb分量　(d) 色度Cr分量

图 1　BasketballDrill 序列 YCbCr 图像和三分量

H.266/VVC 的 CCLM 技术是基于图像局部区域内亮度与色度呈线性关系的假设提出的，CCLM 预测过程如图 2 所示，对于当前 CU，其参考区域为当前 CU 的上、右上、左和左下的像素。在对当前 CU 进行编码之前，参考区域的亮度和色度均已重建。CCLM 首先利用参考区域的重建亮度和重建色度建立线性模型；然后根据该线性模型和当前 CU 的重建亮度信息求解出当前 CU 的色度预测值。需要指出的是，针对 YCbCr 4:2:0 采样格式的视频，亮度图像需进行下采样，从而与色度图像的分辨率一致。

CCLM 算法在 H.266/VVC 中的整体性能[24]如表 1 所示。在相同重建视频质量下，CCLM 可为 Y、Cb、Cr 分量分别节省 1.54%、13.89%、14.76%的编码码率。在复杂度方面，H.266/VVC 官方提供了使用 CCLM 算法与不使用 CCLM 算法的时间占比，编码时间和解码时间几乎相同。该数据进一步验证了分量间预测的简单有效性。

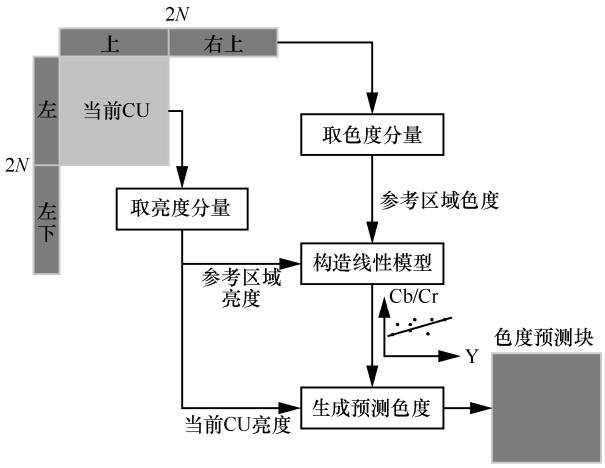


图 2　CCLM 预测过程

表 1　CCLM 算法在 H.266/VVC 中的整体性能

指标	数值
Y 的编码率节省量	1.54%
Cb 的编码率节省量	13.89%
Cr 的编码率节省量	14.76%
编码时间（使用与不使用 CCLM 算法的时间占比）	100%
解码时间（使用与不使用 CCLM 算法的时间占比）	101%

为了充分利用分量间相关性，CCLM 算法历经数次演进。Lee 等[11]揭示了 YCbCr 4:2:0 采样格式的分量间仍存在冗余，可使用分量间线性模型设计色度预测算法。考虑到编码块内容的多样性，Zhang 等[25]提出了基于多分段线性模型的分量间预测技术。出于对复杂度和编码性能的综合考虑，H.266/VVC 采纳了单线性模型的 CCLM 技术[26]。为了提高 CCLM 预测准确度，H.266/VVC 引入了 3 种线性模型[27]，通过率失真准则为当前 CU 选取最优线性模型。同时，H.266/VVC 在保证编码性能的前提下对 CCLM 进行了复杂度优化。Laroche 等[28]提出基于参考区域的最大/最小值构造线性模型参数的算法，该算法可显著降低复杂度。进一步，笔者前期在分析像素欧氏距离与相关性关系的基础上，提出基于参考像素位置构造模型参数的算法[29]。该算法在降低 CCLM 计算复杂度的同时引入少量编码增益，被 H.266/VVC 采纳。

1.2 基于神经网络的色度帧内预测

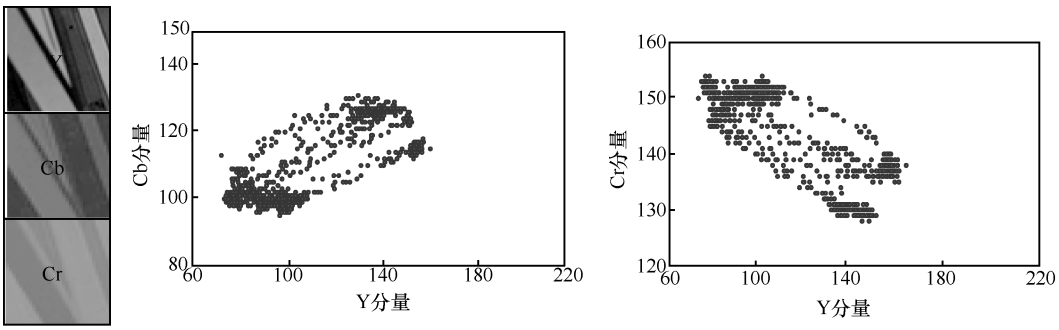
事实上，亮度与色度之间的关系往往是复杂的。图 3 以 BasketballDrill 序列中 2 个不同内容的图像块为例，给出了亮度分量与色度分量的对应关系。由图 3 可以看出，图像块内亮度与色度分量的关系复杂，仅利用简单的线性模型很难处理图像块内所有的情况。同时，由于图像内容多样，不同内容图像块的亮度与色度的相关性也不同，且与内容有关。因此，亮度进行简单的映射通常不能实现准确的色度预测。得益于 NN 强大的建模能力，基于 NN 的分量间预测成为提高色度压缩效率的研究热点。

Blanch 等[30]通过引入基于卷积网络的注意力模块来建立参考像素和待预测像素之间的关系。Zhu 等[31]提出以 CTU 为单位的色度预测方法，充分利用空间和分量间相关性，同时将量化参数（QP, quantization parameter）作为边信息输入，进一步提高预测准确度，降低预测误差。Li 等[32]通过基于卷积网络和全连接网络的混合神经网络改进色度预测性能。纵览以上各个方案，基于 NN 构造色度预测值，尤其是 CNN，可更好地使用非线性函数表示亮度与色度之间的映射关系。然而，现有算法通常需要针对不同的编码参数（例如 QP、编码尺寸等）训练不同的网络参数，这在实际视频编码系统中是难以应用的。此外，基于 CNN 的色度预测算法通常具有极高的复杂度，其解码端复杂度相比于传统预测算法成倍增加。对编码参数的依赖和极高的复杂度导致基于 CNN 的分量间预测算法在实际应用上受到了极大限制。

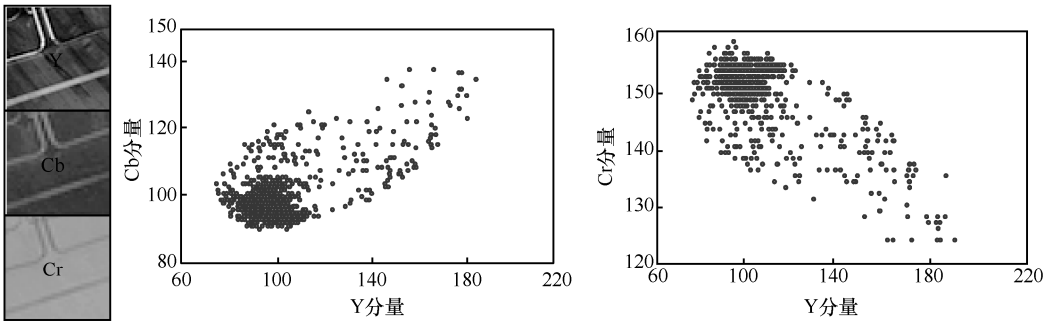
近期，继 H.266/VVC 标准发布后，JVET 着手开展基于 NN 视频编码的探索性研究。实用化的基于 NN 的分量间预测算法也属于其中一个重要的议题：一方面，成熟的 CCLM 算法通过建立简单的线性模型来表示整个 CU 亮度和色度之间的关系，存在较大误差，性能提升潜力有限；另一方面，已有的基于 CNN 的色度预测准确度高，但存在复杂度过高的问题，现阶段难以实用。

2 基于轻量级 NN 的分量间预测算法

在视频编码中，亮度与色度之间的关系通常局部化到一个编码块内讨论。不同于采用线性模型的 CCLM 算法，基于 NN 的预测方法通过数据驱动建立亮度与色度之间的非线性映射。笔者通过研究发现，来自局部近邻的已编码块的亮度和色度信息能够为网络提供非常重要的先验信息。因此，本文提出一种基于轻量级 NN 的分量间预测算法，借助亮度差从参考区域中提取参考子集，从而缩小参与建模的像素规模，即仅采用数个与待预测像素具有较小亮度差的像素进行建模，最终利用轻量级全连接



(a) 示例块1



(b) 示例块2

图 3 YUV 空间下亮度分量与色度分量的散点图图像（YCbCr 4:2:0 格式）

网络实现色度预测。具体来说，遴选出少量且有效的参考像素，利用网络对相关性强的像素赋予大权重，而相关性弱的像素则赋予很小或零权重，进而为色度预测提供有效信息，达到降低预测误差、提高色度压缩效率的目的，同时满足视频编码对低复杂度的需求。

2.1　NNCCP 算法框架

本文提出的 NNCCP 算法框架如图 4 所示，该框架包含数据预处理模块和色度预测模块。其中，数据预处理模块以当前 CU 的上、右上、左、左下参考区域的参考像素和当前像素的亮度值作为输入，经过提取参考子集后输出 $M\times1$ 的亮度差向量和参考色度向量；色度预测模块将预处理模块输出的亮度差向量和参考色度向量作为输入，通过全连接网络构造色度预测值。下面介绍 NNCCP 算法的详细过程。

图 4(a)所示的数据预处理模块包含 3 个步骤：向量化、求亮度差和提取。首先将参考区域的像素（包括重建亮度值和重建色度值）进行一维向量化；然后对每个待预测像素 i，求解 i 与参考像素的亮度差；最后从 $4N$ 个亮度差值中提取出 M 个亮度差绝对值较小的像素组成参考子集，并得到亮度差向量 $|\Delta \boldsymbol{Y}|_i$，其色度值组成参考色度向量 $\boldsymbol{C}_i$。若参考子集的像素数不足 M 个，$|\Delta \boldsymbol{Y}|_i$、$\boldsymbol{C}_i$ 各自使用固定值进行填充，补足 M 个元素。

统计发现，在局部区域内，像素间的亮度差值越小，其相关性越强，色度相关性也越强。本文设计的数据预处理模块从参考区域中提取出亮度差值小的像素构成参考子集，参考子集中的像素具有与待预测像素相关性强的优点。

图 4(b)所示的色度预测模块以 $|\Delta \boldsymbol{Y}|_i$ 和 $\boldsymbol{C}_i$ 作为输入，其中，$|\Delta \boldsymbol{Y}|_i$ 通过 L 层全连接网络后得到权重向量 $\boldsymbol{W}_i$。最终，当前像素的色度预测值 C_i^{p} 为

$$C_i^{\mathrm{p}} = \boldsymbol{W}_i^{\mathrm{T}}\boldsymbol{C}_i \ ,\ \ \boldsymbol{W}_i = \boldsymbol{F}\left(\left|\Delta \boldsymbol{Y}\right|_i\right) \tag{1}$$

其中，$\boldsymbol{F}(\cdot)$表示通过 NN 学习的映射函数，向量 $\boldsymbol{W}_i$、$\boldsymbol{C}_i$ 和 $|\Delta \boldsymbol{Y}|_i$ 的定义为

$$\boldsymbol{W}_i = \begin{pmatrix} W_{i,1} \\ W_{i,2} \\ \vdots \\ W_{i,M} \end{pmatrix}, \boldsymbol{C}_i = \begin{pmatrix} C_{i,1} \\ C_{i,2} \\ \vdots \\ C_{i,M} \end{pmatrix}, \left|\Delta \boldsymbol{Y}\right|_i = \begin{pmatrix} \left|\Delta Y_{i,1}\right| \\ \left|\Delta Y_{i,2}\right| \\ \vdots \\ \left|\Delta Y_{i,M}\right| \end{pmatrix} \tag{2}$$

对于图 4(b)中的全连接网络，网络从输入到输出的描述如下：M 维向量 $|\Delta \boldsymbol{Y}|_i$ 作为第一层的输入，之后每层进行非线性加权后作为下一层的输入。非线性加权可表示为

$$y_{lj} = g_l\left(k\sum w_{ljk}x_{lk}\right), j=1,\cdots,M,\ \ l=1,\cdots,L \tag{3}$$

其中，M 是每层的神经元个数，L 是网络层数，x_{lk} 是第 l 层的第 k 个输入，w_{ljk} 是第 l 层的第 j 个神经元对第 k 个输入的权值，$g_l(\cdot)$是第 l 层的激活函数，y_{lj} 是第 l 层的第 j 个神经元的输出结果。由于网络输出层的输出结果为 0~1，本文网络最后一层的激活函数采用归一化指数函数 Softmax，其他层的激活函数均采用修正线性单元 ReLU。

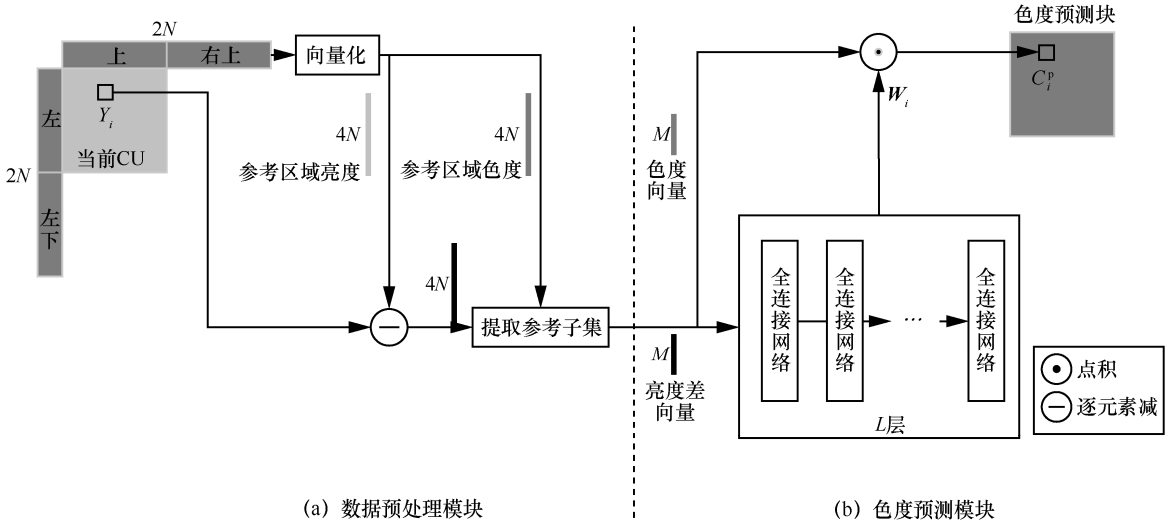


图 4　NNCCP 算法框架

2.2 损失函数

理论上，色度预测模块中全连接网络的损失函数是输出权重 $\boldsymbol{W}$ 与真实权重 $\boldsymbol{W}^{\mathrm{o}}$ 之间的差距，即

$$\mathrm{Loss}=\left\|\boldsymbol{W}-\boldsymbol{W}^{\mathrm{o}}\right\|_2 \tag{4}$$

事实上，真实权重难以估计。考虑到 NNCCP 的目的是得到更准确的预测结果，因此可将色度预测量 $\boldsymbol{C}^{\mathrm{p}}$ 与色度原始量 $\boldsymbol{C}^{\mathrm{o}}$ 的平方误差和作为损失函数。

$$\mathrm{Loss}=\left\|\boldsymbol{C}^{\mathrm{p}}-\boldsymbol{C}^{\mathrm{o}}\right\|_2 \tag{5}$$

在视频编码中，$\boldsymbol{C}^{\mathrm{p}}-\boldsymbol{C}^{\mathrm{o}}$ 为预测残差，需将其经处理后写入压缩码流。

为了提高压缩效率，通常将预测残差变换至频域进行量化和熵编码。为了有效与视频编码结合，接近真实的编码损失，本文使用离散余弦变换（DCT, discrete cosine transform）。具体地，对预测残差进行 DCT，将其绝对误差和作为损失函数

$$\mathrm{Loss}=\left\|\mathrm{DCT}(\boldsymbol{C}^{\mathrm{p}}-\boldsymbol{C}^{\mathrm{o}})\right\|_1 \tag{6}$$

2.3 网络训练

本文所提 NNCCP 算法基于 Pytorch 深度学习框架，实验环境为 64 位 Windows10 操作系统，模型采用 Adam 优化器，初始学习率和批大小设置为 1×10^{-4} 和 128。

为了验证本文算法的色度预测性能和泛化性，采用公开的 DIV2K[33]数据集作为训练和验证的数据来源。具体地，将 800 张训练图片、100 张验证图片均统一裁剪为 4×4 的块作为训练集和验证集。最优模型参数根据其在验证集上的色度预测性能来选取。

2.4 H.266/VVC 集成

H.266/VVC 设计成多种使用前面已编码块的色度预测当前块的色度分量的模式，如 DC、PLANAR、各种角度模式等。为了有效应用于视频编码框架，进一步提升编码性能，本文将离线训练好的 NNCCP 模型作为一种新增的色度帧内预测模式，以 C++语言集成到 H.266/VVC 编解码器中，与已有的预测模式共存，是对现有色度预测模式的有效补充。集成后，H.266/VVC 色度帧内预测模式包括 PLANAR、DC、水平、垂直、亮度推导色度模式、NNCCP 和 3 种 CCLM，共 9 种预测模式。

在编码器侧，所有候选预测模式通过率失真性能度量准则进行模式选择，选出率失真代价最小的模式，并将最优预测模式编号传输。最优预测模式可表示为

$$J^{*}=\underset{s}{\mathrm{argmin}}\left\{D_s+\lambda R_s\right\} \tag{7}$$

其中，s 是色度帧内预测候选模式，D_s 和 R_s 分别是采用不同色度帧内预测模式时的编码失真和编码比特数，λ是拉格朗日因子。

在解码器侧，通过码流解析得到的每个 CU 的最优预测模式序号，并根据最优预测模式进行色度预测值的构造。

2.5 算法的通用性

NNCCP 算法不仅易于集成至 H.266/VVC 编码框架中，还具有良好的通用性。

1) H.266/VVC 采用了灵活的块划分技术，支持二叉树、三叉树和四叉树划分，同一视频中存在多种尺寸的方形块和矩形块。这样一来，参考区域的参考像素数量随着块尺寸的变化而变化。针对参考区域像素数量不固定的问题，NNCCP 算法设计了统一的数据预处理方法，从不同数量的参考像素集合中选取固定数量的相关性强的参考像素组成参考子集。因送入色度预测模块的参考子集像素数量固定，对于任意尺寸 CU，NNCCP 算法均可使用相同的神经网络，不需要针对不同尺寸 CU 单独设计。

2) 视频编码往往需要根据带宽自适应调整编码参数，例如 QP。经测试，NNCCP 架构适用于不同的 QP 配置。因此，对于不同的 QP，NNCCP 算法可使用统一的网络结构及网络参数。该特性优于为不同 QP 设计不同网络结构或训练不同网络参数的方案。

3) NNCCP 算法适用于不同的颜色分量。在 YCbCr 颜色空间上，色度分量包含 Cb 和 Cr 这 2 个分量，所提 NNCCP 算法针对 Cb 和 Cr 分量共享一组网络参数。此外，本文以 YCbCr 4:2:0 采样格式为例，所提算法同样适用于 YCbCr 4:4:4、YCbCr 4:2:2 和 RGB 等其他颜色空间。

3 实验结果

为了充分验证 NNCCP 算法在视频编码上的性能，本节从 4 个方面对其性能进行分析与评估，包括 NNCCP 超参数选择、色度预测性能评估、编码性能评估及 NNCCP 选中比例分析。

3.1 NNCCP 超参数选择

NNCCP 算法的网络结构存在 2 个关键的超参

数，即 M 和 L。参考子集的像素数 M 决定色度预测模块的输入，网络层数 L 决定神经网络的学习能力，二者均影响色度预测的准确性。为了确定 NNCCP 算法的最佳 M 和 L，本文进行了超参数选择实验。实验分为 2 个方案：方案 1 通过固定 L 数值改变 M 的方式探究 M 对预测结果的影响，从而选取最佳的 M；方案 2 是在方案 1 的基础上通过给定的 M 改变 L 的方式确定最佳的 L。

图 5 给出了超参数取不同数值时 NNCCP 算法在 DIV2K 验证集上的 DCT 域损失曲线，每条曲线的标签为（M, L），其中，图 5(a)~图 5(c)是固定 L、改变 M 的损失函数曲线，图 5(d)是固定 M、改变 L 的损失函数曲线。图 5(a)~图 5(c)展示了方案 1 的 3 组实验结果。可以观察到，在相同 L 下，随着 M 值的增加，损失数值逐渐降低，并且损失数值的降低幅度呈减小趋势。在相同 L 下，M=16 的损失数值最小，M=8 次之，M=4 的损失数值最大且明显大于 M=8 和 M=16 的损失数值。为了保证预测效果，本文的 M 在 8 和 16 中选取。

在色度预测模块中，神经网络的运算次数 CN 可定义为

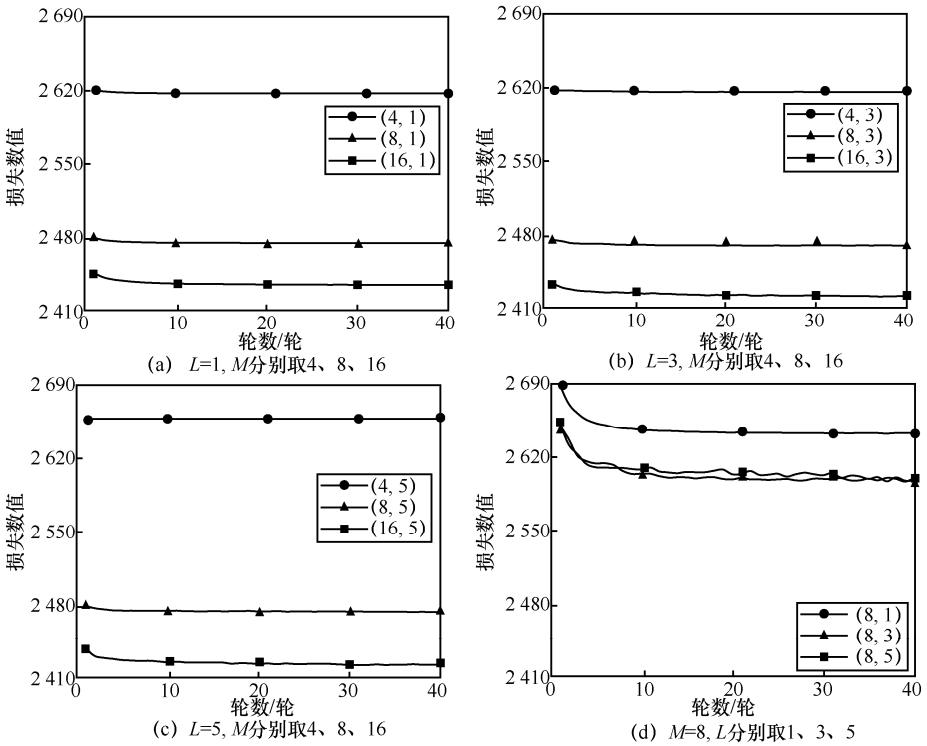
$$\mathrm{CN} = 2M^2L \tag{8}$$

式(8)表明，运算次数受 M 和 L 影响，且 M 起主要影响。随着 M 增加，运算次数急剧增加。同时，M 对数据预处理模块的处理速度起决定性作用，并且数据预处理模块中提取操作的复杂程度随 M 值的增加而增加。基于上述分析，为了在获得较小损失的前提下不引入极高的复杂度，本文将 M 定为 8。

图 5(d)进一步展示了 M=8 时 3 种 L 取值对应的损失曲线。对比发现，损失数值随着 L 的增加而降低，且当 L 增加至 3 层时，损失数值的降幅变缓。图 5(d)的实验结果表明，L=3 或 L=5 的损失数值非常接近并且均小于 L=1 的损失数值。由式(8)可知，L=5 时网络的运算次数高于 L=3 时网络的运算次数。基于此，本文将 L 定为 3，即采用 3 层全连接网络。

综上所述，在复杂度和预测准确度的权衡之下，所提 NNCCP 算法选定 M=8，L=3。表 2 列出了 NNCCP 算法的网络结构参数。其中，全连接网络层数为 3 层，每层节点数量为 8，第一、二层使用 ReLU 激活函数，最后一层使用 Softmax 函数。

由表 2 可以看出，本文使用的神经网络的神经元总数量仅为 24，网络参数较少，内存占用很少，



(a) L=1, M分别取4、8、16　(b) L=3, M分别取4、8、16

(c) L=5, M分别取4、8、16　(d) M=8, L分别取1、3、5

图 5　超参数取不同数值时 NNCCP 算法在 DIV2K 验证集上的 DCT 域损失曲线

是一个轻量级网络，集成至视频编解码框架中具有较小的复杂度。后续的性能测试实验均基于此网络结构开展。

表 2 NNCCP 算法的网络结构参数

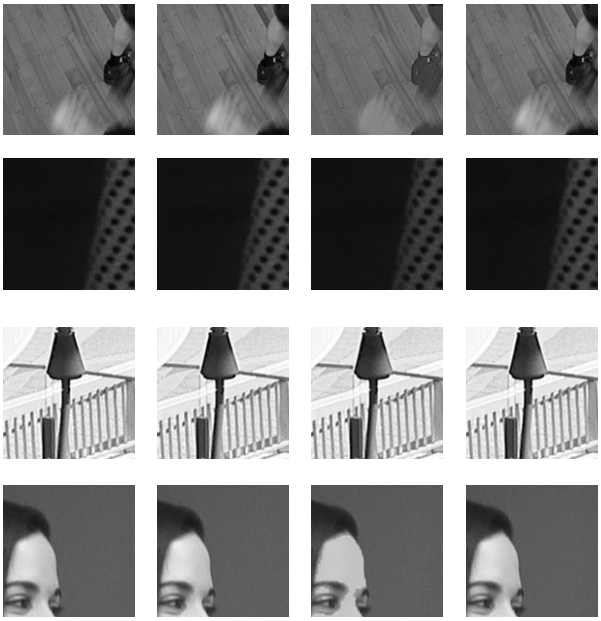
网络层	含义类型	节点数量	激活函数
一	FC	8	ReLU
二	FC	8	ReLU
三	FC	8	Softmax

3.2 色度预测性能评估

为了比较 CCLM 算法和 NNCCP 算法的性能，本节对这 2 种算法构造出的色度预测块进行对比。实验过程如下，将图像分成固定尺寸的块，利用参考区域的原始像素、当前块的原始亮度值分别通过 CCLM 算法和 NNCCP 算法进行色度预测，并将预测结果与原始色度进行对比。需要指出，本节实验均在色度预测块上进行，并非直接用于观看的解码重建块，在实际视频编码中，为了保证观看视频质量，还需将此色度预测块与原始色度块的残差进行编码，并传输至解码端，得到最终用于观看的重建块。

首先，本节从 H.266/VVC 通用测试条件（CTC, common test condition）[34]推荐的 BasketballDrill、BQMall、MarketPlace 和 Tango2 视频中分别选取 64×64 的色度块，通过 CCLM 算法和 NNCCP 算法进行色度预测。图 6 展示了 2 种算法构造的色度预测块的效果对比，其中，图 6(a)~图 6(d)依次为原始块、亮度块、采用 CCLM 算法构造的色度预测块和采用 NNCCP 算法构造的色度预测块。分析发现，在复杂纹理区域或存在边界内容区域上，对图像块中所有像素使用单一线性模型的 CCLM 算法产生了较大的预测误差，而 NNCCP 算法由于提取与当前像素相关性强的参考像素，减少了参考区域中不相关信息的干扰，提高了预测准确性，主观效果更自然，同时色度上也更接近原始图像色度。

进一步，本节在 DIV2K 验证集上分别计算通过 CCLM 算法、文献[30]算法及 NNCCP 算法进行预测的峰值信噪比（PSNR, peak signal to noise ratio），从而客观评估每种算法的预测效果。PSNR 根据色度预测值与色度原始值计算得到。为了验证不同尺寸块上不同算法的性能，实验对 16×16、8×8 及 4×4 尺寸的色度块进行了离线预测的 PSNR 测试，实验结果如表 3 所示。从表 3 可以发现，随着块尺寸的减小，参考区域和图像块的相关性增强，CCLM 算法、文献[30]算法及 NNCCP 算法下的 PSNR 都随之提高。对比上述 3 种算法在同尺寸图像块上的离线预测 PSNR，NNCCP 算法的预测 PSNR 在每种尺寸的图像块上都是最高的，文献[30]算法次之，CCLM 算法最低。特别地，在同尺寸的图像块上，NNCCP 算法的预测 PSNR 相较于 CCLM 算法的提升量为 6 dB 左右，相较于文献[30]也有 1 dB 左右的提升量，同时，随着块尺寸变化，PSNR 提升量基本恒定。上述实验数据说明对不同尺寸块使用统一的网络结构及网络参数的 NNCCP 算法均有优秀的色度预测准确度，进一步验证了 NNCCP 算法的通用性。



(a) 原始块 (b) 亮度块 (c) CCLM 算法 (d) NNCCP 算法

图 6 CCLM 算法与 NNCCP 算法色度预测效果对比

3.3 编码性能评估

为了揭示 NNCCP 算法对视频编码性能的影响，本节将其集成至 H.266/VVC 参考软件 VTM10.0。在编码过程中，NNCCP 算法与传统色度预测模式竞争，通过率失真准则确定最优预测模式，并在码流中传输对应模式序号。本节采用 H.266/VVC 通用测试条件推荐的 21 个测试序列，涵盖了不同类型的视频内容和不同的分辨率。本实验采用全帧内配置（AI, all intra）编码。编码性能采用 BD-rate[35]作为评估标准。当 BD-rate 为负值时，表示在获得相同视频图像质量的前提下所提算法可节省的编码码率。

表 3　　色度预测性能比较

算法	PSNR/dB		
	16×16	8×8	4×4
CCLM	28.40	30.12	32.45
文献[30]	33.48	35.40	37.29
NNCCP	34.47	36.48	38.42
NNCCP 相对 CCLM 的提升量	6.07	6.36	5.97

将 NNCCP 算法集成到 VTM10.0 进行编码测试，并与 VTM10.0 的编码性能进行对比。表 4 给出了详细的 BD-rate 对比结果，并分别计算了 Y、Cb、Cr 和 YCbCr 分量的 BD-rate 值。实验结果表明，在相同重建视频质量下，NNCCP 算法在 Y、Cb、Cr 上分别平均节省 0.27%、1.54%和 1.84%的编码码率。这证明了所提算法可提高压缩效率，尤其是针对色度分量。为了综合衡量所提算法的编码性能，YCbCr 分量的综合 PSNR 由 Y、Cb、Cr 分量加权计算[23]得到。相比于 VTM10.0，NNCCP 算法在综合 PSNR 下可平均节省 0.46%的编码码率，有效提高了编码性能。同时，由表 4 列出的各个序列的编码性能可以看出，NNCCP 算法具有良好的序列一致性。

表 4　　NNCCP 与 VTM10.0 编码性能比较

类	序列	Y	Cb	Cr	YCbCr
A1	Tango2	−0.66%	−4.48%	−4.93%	−1.12%
	FoodMarket4	−0.21%	−1.22%	−1.80%	−0.43%
	Campfire	−1.30%	0.14%	−4.23%	−1.24%
A2	CatRobot	−0.34%	−2.06%	−2.36%	−0.65%
	DaylightRoad2	−0.04%	−1.58%	−1.10%	−0.14%
	ParkRunning3	−0.04%	−0.45%	−0.41%	−0.27%
B	MarketPlace	−0.45%	−2.83%	−1.72%	−0.77%
	RitualDance	−0.30%	−2.13%	−3.76%	−0.64%
	Cactus	−0.08%	−0.94%	−0.76%	−0.19%
	BasketballDrive	−0.12%	−1.54%	−1.49%	−0.29%
	BQTerrace	−0.04%	−1.70%	−1.53%	−0.13%
C	BasketballDrill	−0.78%	−3.69%	−3.48%	−1.22%
	BQMall	−0.15%	−1.44%	−1.42%	−0.33%
	PartyScene	−0.12%	−1.15%	−1.14%	−0.25%
	RaceHorses	−0.15%	−0.63%	−1.03%	−0.26%
E	FourPeople	−0.02%	−0.59%	−0.49%	−0.08%
	Johnny	−0.03%	−0.52%	−0.72%	−0.11%
	KristenAndSara	−0.05%	−0.89%	−0.77%	−0.16%
所有序列的平均结果		−0.27%	−1.54%	−1.84%	−0.46%

为了进一步评估 NNCCP 算法在基于 NN 的色度预测算法中的性能，本节在 CPU 平台上将 NNCCP 算法与文献[30]算法、文献[31]算法从多方面进行比较。观察表 5 中不同算法的网络结构和网络参数，文献[30]算法针对不同尺寸的编码块训练 3 个网络模型，文献[31]算法与 NNCCP 算法都实现了统一的网络模型，其中文献[31]算法的网络层数是最多的，并且需要存储上百万的网络参数量，这对视频编码和解码器提出了很高的要求。由表 5 的数据可以看出，NNCCP 算法的网络层数最少，同时需要存储的总参数量远低于文献[30]算法和文献[31]算法，所需存储开销最少。

表 5　不同算法的网络结构和网络参数比较

算法	网络类型	总参数量	网络层数	模型数量
文献[30]	CNN	238 995	9	3
文献[31]	CNN	2 390 832	26	1
NNCCP	FC	192	3	1

图 7 进一步展示了 AI 配置下 VTM 算法、NNCCP 算法、文献[30]算法和文献[31]算法的编码性能和解码复杂度的对比情况，其中 x 轴为相对于 VTM 的解码时间增加量；y 轴为相对于 VTM 的码率节省量，值越小，表明码率节省量越多。由图 7 可以看出，文献[30]算法可节省 0.20%的编码码率，NNCCP 算法可节省 0.46%的编码码率，文献[31]算法可节省 3.6%的编码码率。在解码时间方面，文献[30]算法的解码时间相对于 VTM 算法增加了 874%，文献[31]算法的解码时间相对于 VTM 算法增加了 834%，而 NNCCP 算法的解码时间相对于 VTM 算法仅增加了 34%。视频编码和解码是工业界的典型应用，视频相关应用对压缩效率和实时性都有非常高的要求。综合考虑上述 3 种算法的编码性能和算法复杂度，虽然文献[31]算法的编码性能最佳，但其上百万参数量的存储需求和 834%的解码时间增加量在现阶段难以实际应用。文献[30]算法的网络参数量相比文献[31]算法大幅降低，但其编码性能增益有限，解码时间增加量也很高。因此，相比于文献[30]算法和文献[31]算法，本文提出的 NNCCP 算法的解码复杂度大幅降低，在解码时间增加 34%和网络参数量仅有 192 的前提下可节省 0.46%的编码码率，采用极低复杂度，有效节省了编码码率，提高了视频编码的压缩性能。

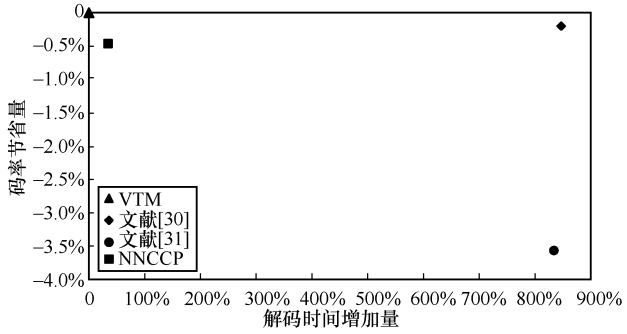


图 7　AI 配置下几种算法的编码性能和解码复杂度的对比情况

3.4　NNCCP 选中比例分析

为了进一步分析 NNCCP 对视频编码的影响，本节对 NNCCP 算法的分布情况进行分析。图 8 为采用 CatRobot（3840×2160）、Tango2（3840×2160）、ParkRunning3（3840×2160）、MarketPlace（1920×108）、KristenAndSara（1280×720）、BasketballDrill（832×480）6 个典型测试序列以 QP22 编码的第一帧、重新缩放到相同的分辨率下的可视化结果，并展示了 NNCCP 模式作为最优预测模式的编码块。从图 8 可以观察到，选中 NNCCP 的编码块尺寸多样，既有方形块，也有矩形块。

为进一步挖掘选中 NNCCP 的规律，采用评价线性拟合程度的指标 R^2 来定量分析编码块采用线性模型的预测值与原始值的拟合程度。通常，R^2 范围为[0,1]，越逼近 1，拟合程度越高，线性模型的可靠性就越高。对图 8 中 NNCCP 编码块的 R^2 值进行统计，其分布曲线如图 9 所示。观察各个序

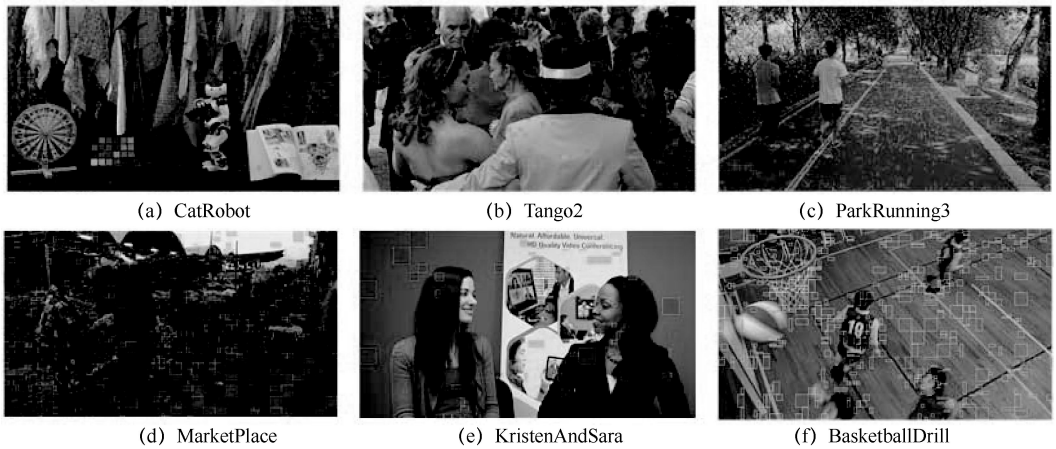


图 8　NNCCP 选中区域分布展示

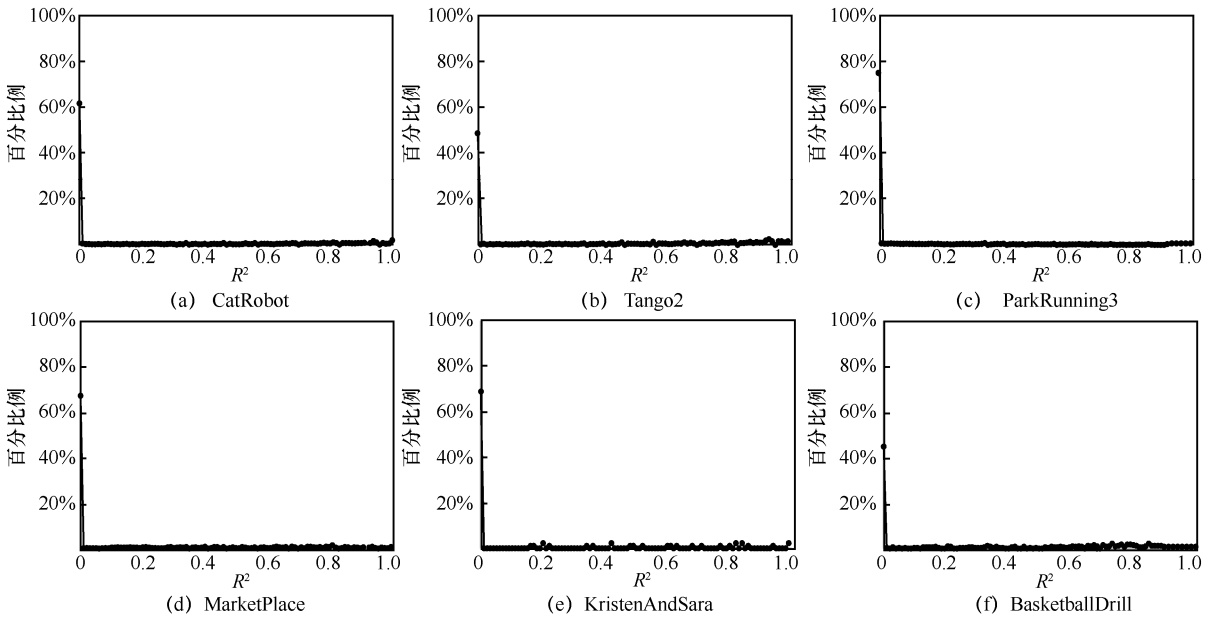


图 9　选中 NNCCP 编码块的 R^2 的统计分布曲线

列 R^2 的分布情况可以明显发现，NNCCP 编码块的 R^2 大量集中在 0 值附近，即 NNCCP 编码块的 R^2 普遍较小。

图 10 进一步给出了各类序列 NNCCP 编码块和 CCLM 编码块 R^2 的平均值。对比发现，在 A1~E 序列下，NNCCP 编码块的 R^2 平均值皆明显小于 CCLM 的 R^2 平均值。对于所有序列，NNCCP 编码块的平均 R^2 仅为 0.27，而 CCLM 中的平均 R^2 高达 0.69。因此，对于使用线性模型拟合程度较差的编码块，即 R^2 较小时，通常会选择 NNCCP 模式；对于采用线性模型拟合程度较好的编码块，即 R^2 较大时，NNCCP 和 CCLM 都可较好地进行预测，通常会选择 H.266/VVC 的 CCLM 模式。

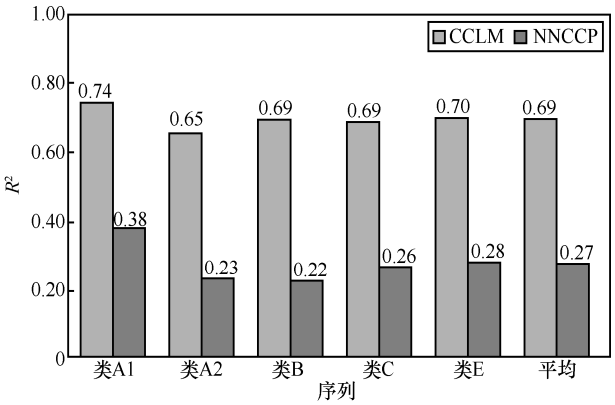


图 10　各类序列 NNCCP 编码块和 CCLM 编码块 R^2 的平均值

表 6 进一步给出了测试序列在不同编码 QP 下选中 NNCCP 模式的像素数占比。对于不同的 QP 参数，NNCCP 的选中像素数占比可达 13.36%~16.57%。不同 QP 下像素数占比并未存在明显差异。观察各个序列 NNCCP 的选中比例可以发现，对于内容简单的序列而言，H.266/VVC 中原有的色度预测模式已足以对此类内容实现准确的预测，此时 NNCCP 的选中比例较低，如图 8 中 KristenAndSara 序列；当序列内容丰富、复杂时，NNCCP 的选中比例较高，如图 8 中的 CatRobot、Tango2 和 MarketPlace 序列。对于内容丰富的序列，原有色度预测模式不足以应对纹理复杂多样的编码块，在此情况下存在较多 R^2 偏小的内容，此时 NNCCP 作为一种新的预测模式，通过率失真准则筛选为最佳预测模式。NNCCP 改善了 R^2 偏小的编码块的预测效果，对内容丰富的序列可以达到较准确的预测。

表 6　测试序列在不同编码 QP 下选中 NNCCP 模式的像素数占比

类	序列	QP			
		22	27	32	37
A1	Tango2	20.59%	25.12%	27.89%	27.93%
	FoodMarket4	11.49%	13.77%	15.31%	14.66%
	Campfire	19.39%	28.11%	30.94%	28.26%
A2	CatRobot	13.46%	15.64%	16.59%	17.27%
	DaylightRoad2	10.40%	8.89%	8.66%	8.84%
	ParkRunning3	13.47%	14.08%	14.30%	12.69%
B	MarketPlace	21.53%	27.19%	29.38%	26.54%
	RitualDance	17.05%	21.02%	22.05%	19.93%
	Cactus	12.37%	13.31%	13.76%	13.76%
	BasketballDrive	8.88%	9.43%	9.25%	8.82%
	BQTerrace	15.32%	14.50%	13.10%	11.58%
C	BasketballDrill	20.46%	26.18%	30.63%	32.03%
	BQMall	15.06%	15.41%	14.30%	13.44%
	PartyScene	17.15%	19.85%	18.99%	16.57%
	RaceHorses	6.15%	8.54%	13.75%	15.10%
E	FourPeople	5.10%	6.58%	8.16%	8.73%
	Johnny	5.82%	5.10%	4.34%	4.57%
	KristenAndSara	6.79%	5.84%	6.82%	6.90%
所有序列的平均结果		13.36%	15.48%	16.57%	15.98%

4　结束语

本文提出了一种基于轻量级全连接网络的色度预测算法。相比于现有基于神经网络的色度预测方法，NNCCP 算法在获得编码性能提升的同时，使用了轻量级的全连接神经网络，其在解码端引入的时间复杂度大幅低于现有网络算法。所提网络可适用于不同尺寸的编码块和不同 QP，网络设计时充分考虑了视频编码的特点。

本文将 NNCCP 算法作为一种新的色度帧内预测模式集成到 H.266/VVC 软件平台。实验结果表明，该算法在 H.266/VVC 基础上还可获得 0.46% 的码率节省量，尤其可有效提高色度分量的压缩效率。

NNCCP 算法可有效应用于 H.266/VVC，同时该算法也为下一代视频编码标准提供了可行的研究思路。一方面，可考虑进一步降低 NNCCP 算法

的复杂度；另一方面，随着计算能力的不断发展，可设计更高效的网络结构进行分量间预测，进一步提升视频压缩效率。

参考文献：

[1] ITU-T. ITU-T Recommendation H.266 and ISO/IEC 23090-3 VVC standard[S]. 2020.

[2] ALBRECHT M, BARTNIK C. Description of SDR, HDR, and 360° video coding technology proposal by Fraunhofer HHI[R]. JVET-J0014, 2018.

[3] YE Y, BOYCE J M, HANHART P. Omnidirectional 360° video coding technology in responses to the joint call for proposals on video compression with capability beyond HEVC[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(5): 1241-1252.

[4] FRANÇOIS E, SEGALL C A, TOURAPIS A M, et al. High dynamic range video coding technology in responses to the joint call for proposals on video compression with capability beyond HEVC[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(5): 1253-1266.

[5] ITU-T. ITU-T Recommendation H.265 and ISO/IEC 23008-2 HEVC standard. High efficiency video coding[S]. 2013.

[6] BROSS B, CHEN J L, 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.

[7] 朱秀昌, 唐贵进. H.266/VVC: 新一代通用视频编码国际标准[J]. 南京邮电大学学报(自然科学版), 2021, 41(2): 1-11.
ZHU X C, TANG G J. H.266/VVC: versatile video coding international standard[J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science), 2021, 41(2): 1-11.

[8] HUANG Y W, HSU C W, CHEN C Y, et al. A VVC proposal with quaternary tree plus binary-ternary tree coding block structure and advanced coding techniques[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(5): 1311-1325.

[9] SCHÄFER M, STALLENBERGER B, PFAFF J, et al. Efficient fixed-point implementation of matrix-based intra prediction[C]//Proceedings of 2020 IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2020: 3364-3368.

[10] PFAFF J, SCHWARZ H, MARPE D, et al. Video compression using generalized binary partitioning, trellis coded quantization, perceptually optimized encoding, and advanced prediction and transform coding[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30(5): 1281-1295.

[11] 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]//Proceedings of 2009 16th IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2009: 1037-1040.

[12] ZHANG K, CHEN Y W, ZHANG L, et al. An improved framework of affine motion compensation in video coding[J]. IEEE Transactions on Image Processing, 2019, 28(3): 1456-1469.

[13] GAO H, ESENLIK S, ALSHINA E, et al. Geometric partitioning mode in versatile video coding: algorithm review and analysis[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(9): 3603-3617.

[14] NASER K, POIRIER T, LEANNEC L F. Non-CE6: shape adaptive transform selection for ISP, SBT and MTS[R]. JVET-N0388-v5, 2019.

[15] KOO M, SALEHIFAR M, LIM J, et al. Low frequency non-separable transform (LFNST)[C]//Proceedings of 2019 Picture Coding Symposium (PCS). Piscataway: IEEE Press, 2019: 1-5.

[16] TSAI C Y, CHEN C Y, YAMAKAGE T, et al. Adaptive loop filtering for video coding[J]. IEEE Journal of Selected Topics in Signal Processing, 2013, 7(6): 934-945.

[17] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.

[18] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[19] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 1646-1654.

[20] LIU D, LI Y, LIN J P, et al. Deep learning-based video coding[J]. ACM Computing Surveys, 2021, 53(1): 1-35.

[21] MA S W, ZHANG X F, JIA C M, et al. Image and video compression with neural networks: a review[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(6): 1683-1698.

[22] MINNEN D, BALLÉ J, TODERICI G. Joint autoregressive and hierarchical priors for learned image compression[J]. arXiv Preprint, arXiv: 1809.02736, 2018.

[23] CHEN J, YE Y, KIM S. Algorithm description for versatile video coding and test model 10 (VTM 10)[R]. JVET-S2002, 2020.

[24] CHIEN W J, BOYCE J, CHEN Y W, et al. JVET AHG report: tool reporting procedure (AHG13)[R]. JVET-T0013, 2020.

[25] ZHANG K, CHEN J, ZHANG L, 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.

[26] MA X, YANG H, CHEN J. Tests of cross-component linear model in BMS1.0[R]. JVET-K0190, 2018.

[27] MA X, YANG H, CHEN J. CE3: CCLM/MDLM using simplified coefficients derivation method (Test 5.6.1, 5.6.2 and 5.6.3)[R]. JVET-L0340, 2018.

[28] LAROCHE G, TAQUET J, GISQUET C, et al. CE3: cross-component linear model simplification (Test 5.1)[R]. JVET-L0191, 2018.

[29] HUO J Y, MA Y Z, WAN S, et al. CE3-1.5: CCLM derived from four neighbouring samples[R]. JVET-N0271, 2019.

[30] BLANCH M G, BLASI S, SMEATON A, et al. Chroma intra prediction with attention-based CNN architectures[C]//Proceedings of 2020 IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2020: 783-787.

[31] ZHU L W, ZHANG Y, WANG S Q, 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.

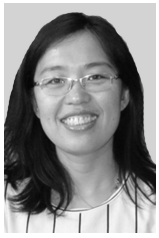
[32] LI Y, LI L, LI Z, et al. A hybrid neural network for chroma intra prediction[C]//Proceedings of 2018 25th IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2018: 1797-1801.

[33] TIMOFTE R, AGUSTSSON E, GOOL L V, et al. NTIRE 2017 challenge on single image super-resolution: methods and results[C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2017: 1110-1121.

[34] BOYCE J, SUEHRING K, LI L, et al. JVET common test conditions and software reference configurations[R]. JVET-J1010, 2018.

[35] BOSSEN F. On reporting combined YUV BD rates[R]. JVET-N0341, 2019.

[作者简介]



霍俊彦（1982− ），女，山西晋中人，博士，西安电子科技大学副教授，主要研究方向为多媒体通信、视频编码、智能信息处理。

王丹妮（1996− ），女，陕西西安人，西安电子科技大学硕士生，主要研究方向为视频压缩编码。

马彦卓（1980− ），女，河北深州人，博士，西安电子科技大学副教授，主要研究方向为视频编码与视频传输。

万帅（1979− ），女，河南洛阳人，博士，西北工业大学教授、博士生导师，主要研究方向为视频编码、点云压缩及多媒体通信。

杨付正（1977− ），男，山东德州人，博士，西安电子科技大学教授、博士生导师，主要研究方向为新一代视频压缩标准、基于深度学习的视频处理和虚拟现实。