arXiv:2301.10455v1 [eess.IV] 25 Jan 2023

面向视频编码的码率-感知优化预处理

Chengqian Ma、Zhiqiang Wu、Chunlei Cai、Pengwei Zhang、Yi Wang、Long Zheng、Chao Chen、Quan Zhou 哔哩哔哩 上海，中国{
machengqian01, wuzhiqiang01, caichunlei, zhangpengwei, wangyi, zhenglong, chenchao02, zhouquan}@bilibili.com



图 1. **左侧：** RPP 的部署流程：对输入帧进行单次处理，并适用于所有标准视频编解码器。**右侧：** 相同多尺度结构相似性下 H.265 与 RPP + H.265 的帧片段。放大以查看细节。

摘要

过去几十年来，视频压缩领域取得了诸多进展，涵盖传统视频编解码器和基于学习的视频编解码器。然而，很少有研究关注利用预处理技术提升率失真性能。本文提出一种率失真感知优化预处理（*RPP*）方法。我们首先引入一种自适应离散余弦变换损失函数，既能节省比特率，又能保留关键的高频分量。此外，我们还将低层视觉领域的多项先进技术融入该方法，例如高阶退化模型、高效的轻量级网络设计以及图像质量评估模型。通过联合运用这些强大的技术，我们的 *RPP* 方法在多种质量指标下，配合 *AVC*、*HEVC* 和 *VVC* 等不同视频编码器，平均可节省 *16.27%* 的比特率。在部署阶段，*RPP* 方法简单高效，无需更改视频编码、流媒体和解码的设置。每个输入帧只需经过 *RPP* 一次，即可送入视频编码器。此外，在主观视觉质量测试中，*87%* 的用户认为，使用 *RPP* 处理的视频优于或不逊于仅使用编解码器压缩的视频，而前者平均可节省约 *12%* 的比特率。我们的 *RPP* 框架已集成到视频转码服务的生产环境中，该服务每天为数百万用户提供服务。论文被接收后，我们将发布代码和模型。

1. 引言

近年来，对高清视频在线视频流媒体的需求迅速增长，预计未来几年还将持续增长。这些高清视频流媒体消耗了大量带宽，占消费者互联网流量总量的 80% 以上 [14]。因此，构建高效的视频压缩系统，以在给定带宽预算下提供更好的视频质量至关重要。为此，过去几十年间，人们制定了许多视频编码标准，例如 H.264 [46], H.265 [38], H.266 [8], 以及 AOMedia Video 1（AV1） [11]。这些传统方法由许多人工设计的模块构成，例如块划分、离散余弦变换（DCT） [2], 和帧内/帧间预测等。尽管这些人工设计的方法取得了良好的率失真性能，受深度神经网络在其他图像处理领域取得成功的启发，基于学习的视频压缩方法 [10, 28, 29] 仍然日益受到关注。这些基于学习的方法声称，其性能可与传统编解码器相媲美，甚至更胜一筹。然而，大多数现有的基于学习的视频压缩方法都增加了编码器和解码器两端的复杂度。这种计算开销巨大的解

码器使部署难以实现，尤其是在手机和笔记本电脑等终端用户设备上。一些研究尝试将标准混合视频编码器设计中的关键组件转化为可训练框架，以便对视频编 [29, 49]码器中的所有模块进行端到端优化。然而，鲜有研究尝试使用预处理方法来提升视频压缩系统的率失真性能。

本文提出一种率失真感知优化预处理器（RPP），能够在单次独立前向传播中高效地同时优化码率和视觉质量。具体而言，我们将自适应离散余弦变换（DCT）损失引入 RPP 的训练阶段。此外，我们还将全参考图像质量评价模型：多尺度结构相似性 [45] 用于训练，以优化模型的感知质量。同时，我们设计了一种带有注意力机制的轻量级全卷积神经网络，以提升效率。

我们的工作贡献可总结如下：

- 我们首先引入自适应离散余弦变换（DCT）损失，该损失能够减少空间冗余，同时保留内容中的重要高频分量。实验结果表明，在训练中引入自适应离散余弦变换损失，可以显著降低视频的比特率，同时保持其视觉质量。

- 我们提出了一种率失真感知优化预处理器（RPP），它是一种轻量级的全卷积神经网络，并采用了注意力机制。RPP模型通过同时利用自适应离散余弦变换损失和基于参考的IQA损失函数，在感知与失真之间取得平衡。我们还在训练阶段引入高阶退化模型，以提升预处理帧的视觉质量。

- 我们的方法可以轻松集成到任何标准视频编解码器的预处理流程中，例如 AVC、HEVC、AV1 或 VVC。借助我们的方法，这些标准视频编解码器无需进行任何改动，也无需在视频编码和解码方面作出牺牲，便能获得更优的 BD-rate 表现。与最先进的视频编解码方法相比，我们的模型在多种质量指标下平均可将 BD-rate 降低约 16.27%。此外，我们的 RPP 模型效率极高，在推理时可达到 1080p@87FPS，远超实时处理效率。

2. 相关工作

2.1. 图像压缩

过去几十年来，人们提出了许多传统图像压缩方法，例如 JPEG [40], JPEG2000 [13]和 BPG [6]。这些方法通过利用手工设计技术，有效减小图像尺寸，并取得了出色的性能。这些手工设计方案中最重要的部分之一是 DCT 等变换。DCT 将像素线性映射到频域。DCT 的一个优点是能够集中能量，从而便于减少图像的空间冗余。变换后，这些方法会对相应的系数进行量化，然后执行熵编码。近年来，得益于深度神经网络，基于学习的图像压缩方法 [3, 4, 31] 已取得与传统图像压缩编码相当甚至更优的性能。

2.2. 视频压缩

视频压缩方法的发展历史悠久。在过去几十年中，人们提出了多种视频编码标准，并将其广泛应用于实际场景，例如 H.264 [46], H.265 [38], H.266 [8],和 AOMedia Video 1（AV1） [11]。随着视频编码标准的不断发展，这些传统视频压缩方法性能强劲，并取得了显著进步。凭借硬件支持，这些方法也适用于在线视频流、数字电视等现实应用。近年来，人们提出了许多基于深度神经网络的方法，应用于视频编码的各个环节，例如帧内预测、残差编码 [10], 、模式决策 [28],、熵编码等。这些方法用于提升传统视频编解码器中特定模块的性能。与替换传统视频压缩编解码器中的特定组件不同，一些方法专注于端到端优化的视频压缩框架 [29, 49]。此外，A. Chadha 等人 [9]尝试将标准视频编码器设计中的核心组件转换为可训练框架，并以端到端的方式，通过该可微框架联合优化预处理器。

2.3. 指标

在过去几十年中，峰值信噪比（PSNR）一直是评估视频保真度和质量时使用最广泛的全参考方法，并且在评估视频压缩算法方面持续发挥着根本性作用。然而，事实证明，PSNR 与人类感知的相关性较差 [17,43]。因此，人们提出了各种全参考图像质量评估（IQA）或视频质量评估（VQA）方法 [5, 25, 36, 44]。对于

例如，结构相似性（SSIM）指数 [44]通过考虑结构信息来估计感知失真，而其变体多尺度结构相似性（MS-SSIM） [45] 则通过融入多尺度分辨率处理，提供更出色的性能和更高的灵活性。视频多方法评估融合（VMAF） [25,26]是另一种主流评估指标，在现实产业中，Netflix 等众多知名商业公司 [26], Meta [34],TikTok [48], Intel [22]等企业以及 AO-Media 等标准化组织 [12] 都采用该指标进行视频编解码器评估。VMAF 结合了三种质量特征：视觉信息保真度（VIF）、 [36], 细节损失指标（DLM） [23], 和运动特征，用于训练支持向量机（SVM）回归器[15] 以预测视频质量的主观评分。大量研究表明，与结构相似性（SSIM）和峰值信噪比（PSNR）相比，VMAF 与平均意见分（MOS）的相关性显著更高 [5,33,47]。

2.4. 图像增强

图像增强因其在各类视觉应用中具有重要的实际价值，长期以来一直是一个研究课题。近年来，随着深度学习技术的发展，例如网络设计和基于梯度的优化方法，基于学习的方法 [1,42] 在超分辨率、去噪、去模糊等图像增强领域展现出良好的性能。一些方法 [20, 27] 旨在借助精心设计的轻量级卷积神经网络实现实时图像超分辨率，并以有限的计算开销取得更好的结果。其他方法 [16,42] 则专注于设计退化模型，以模拟图像复杂的退化过程。Wang 等人 [42] 采用高阶退化过程来模拟复杂的真实世界退化。尽管图像增强领域已涌现出许多优秀成果，但鲜有研究利用视频编码的方法和技术。

3. 提出的方法

3.1. 概述

本节将简要介绍我们提出的码率与感知优化预处理（RPP）方法。我们的预处理模型旨在通过可学习的预处理神经网络，提供同时针对码率与感知进行优化的预处理输入帧。具体而言，为了在码率与失真之间实现平衡，我们设计了一种自适应离散余弦变换损失，既能减少空间冗余，又能保留对感知至关重要的高频分量。此外，在感知优化方面，我们使用全参考图像质量评价模型：结构相似性（SSIM），以提升预处理后输入帧的感知质量。训练过程中，我们将该图像质量评价模型用作损失函数，

此外，我们结合高阶退化建模过程来模拟真实世界的复杂退化 [42]。通过这种高阶退化方法生成成对训练数据，我们的预处理网络能够学会应对真实世界中的一些复杂退化，同时提升网络输出的感知质量。再者，为兼顾性能与效率，我们构建了一种轻量级全卷积神经网络，并采用通道注意力机制 [18]。在部署框架中，对于给定的视频帧 f_i，只需通过 RPP 网络进行一次前向传播。随后，RPP 网络输出的处理后帧 f_o 可由标准视频编解码器编码，例如 AVC [46],HEVC [38], VVC [8],或 AV1 [11]编码器。

3.2. 自适应离散余弦变换损失

尽管离散余弦变换最早应用于图像/视频压缩算法至今已有多年，但由于其高效且易于使用，类离散余弦变换仍是当今的主流变换。通常，二维（2D）离散余弦变换的基函数可表示为：

$$B_{h,w}^{i,j} = cos\frac{h\pi}{H}(i+\frac{1}{2})cos\frac{w\pi}{H}(j+\frac{1}{2}) \tag{1}$$

那么，二维离散余弦变换可表示为：

$$F_{h,w} = \sum_{i=0}^{H-1}\sum_{j=0}^{W-1} f_{i,j}B_{h,w}^{i,j} \tag{2}$$

$$s.t. \quad h \in \{0,1,\cdots,H-1\}, w \in \{0,1,\cdots,W-1\}$$

其中，$F \in \mathbb{R}^{\mathbb{H}\times\mathbb{W}}$ 是二维离散余弦变换频谱，$f \in \mathbb{R}^{\mathbb{H}\times\mathbb{W}}$是输入帧，$H$ 是 f的高度，W 是 f的宽度。通常，高度和宽度相同。在大多数常见情况下，H和 W 通常表示为 N。

输入帧 f后，它会将像素块转换为大小相同的频率系数块。正如我们所提到的，离散余弦变换有一个关键特性：频率系数块会将高频分量与低频分量分离。在图像中，大部分能量集中在较低频率上，因此，传统压缩算法会直接丢弃较高频率的频率系数，以减少空间冗余。然而，一些高频分量对整帧的视觉质量也起着非常重要的作用。因此，我们首先提出了用于视频预处理的自适应离散余弦变换损失。首先，我们使用离散余弦变换将帧 f转换到频域。其次，我们使用ZigZag顺序遍历选出属于高频分量的频率系数 I。公式可写为：

$$F'_{h,w} = F_{h,w} * I_{h,w} \tag{3}$$

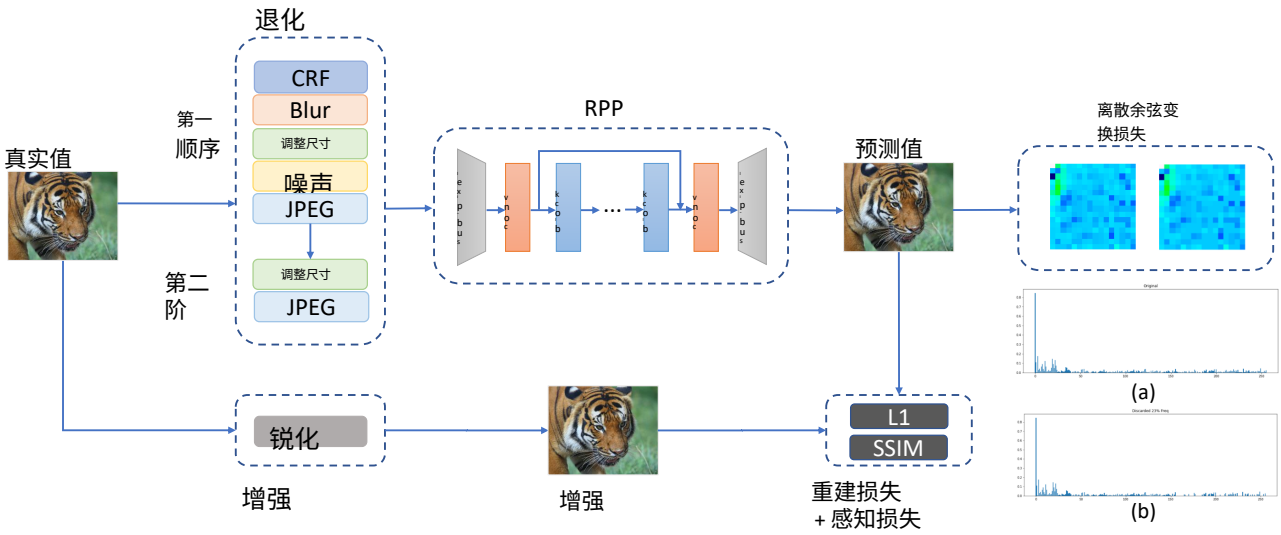


图2. 示例框架 用于训练 RPP。（a）为预测帧的频率系数直方图。（b）为直方图 of
经过频率系数滤波的 b 通过自适应离散余弦变换函数

$$where \quad I_{h,w} = \begin{cases} 0 & \text{if } (h+w) < S \\ 1 & \text{if } (h+w) \geq S \end{cases} \tag{4}$$
$$S \in \{0, 1, \cdots, (H-1)(W-1)\}$$

在离散余弦变换频域中，频率系数的值表示该频率分量在整帧中所含能量的大小。若某个频率分量的能量较低，就意味着重建帧时该频率分量相对不那么重要。因此，我们希望丢弃一些系数值相对较小的高频分量。在这种情况下，我们对所选系数 $F'_{h,\ w}$ 的绝对值取平均，得到阈值 T，其公式可表示为：

$$T = \frac{1}{H \cdot W} \sum_{h=i}^{H-1} \sum_{w=j}^{W-1} (|F'_{h,w}|) \tag{5}$$
$$where \quad i + j \geq N$$

如果 $|F'_{h,w}|$ 小于阈值 T，这意味着它对帧重建的影响小于平均水平。然后，我们将其选入另一组系数 $F''_{h,w}$。最后，我们计算滤波后的离散余弦变换频率系数 $F''_{h,w}$ 与零之间的平均绝对误差，可表示为：

$$\mathcal{L}_{dct} = \sum_{h=i}^{H-1} \sum_{w=j}^{W-1} |F''_{h,w} - 0|, \tag{6}$$
$$F''_{h,w} \in \{|F'_{h,w}| < T\} \quad and \quad i + j \geq N$$

在模型训练中使用该损失函数，模型将得到优化，以保留重要的高频分量并舍弃一些不重要的高频分量。经过这种优化后，模型处理的帧可以使视频编码器为这些重要的高频分量（例如边缘和对比度区域）分配更多比特率。同时，由于自适应DCT损失函数会将一些不重要的高频分量滤为零，因此也有助于熵编码过程 [19, 35]，连续的零将大幅降低比特率。

3.3. 网络与图像退化

受图像增强领域轻量级网络架构的启发，我们借鉴了其中的一些思路 [20, 27]。具体而言，我们以 RFDB [27], 等特征提取模块为基础，在模块中加入通道注意力机制 [18]，使网络能够更加关注不同通道的频率。此外，我们采用了一种高效的亚像素卷积，该方法最初由 Shi 等人 [37] 提出，用于缩小和放大特征图的分辨率。整体网络架构如图 4 所示。

对训练数据退化方式进行建模对于提升网络训练过程中的视觉质量至关重要。我们在退化模型中加入了一些通用退化方法 [16]，例如模糊、噪声、调整尺寸和 JPEG 压缩。对于模糊，我们使用各向同性和各向异性的高斯滤波器对模糊退化进行建模。我们选取两种常用噪声类型——高斯噪声和泊松噪声——来模拟噪声退化。对于调整尺寸，我们采用多种缩放算法进行上采样和下采样，包括区域插值、双线性插值和双三次插值。由于在真实应用中，

我们框架的输入帧大多是从压缩视频中解码得到的，因此我们加入了视频压缩退化，它可能会在空间域和时间域引入块效应和振铃伪影。正如前文所述，高阶退化建模 [42] 已被提出，以更好地模拟复杂的真实世界退化。我们也将这一理念应用于图像退化模型。通过使用这些退化模型生成训练对，我们旨在使模型能够去除常见噪声和压缩噪声，从而优化码率，因为视频编解码器无法有效编码噪声。

3.4. 损失函数

我们的目标是通过同时优化码率和感知来训练预处理网络。为了针对ˆ重建帧 f 相对于输入帧 f的码率和感知进行优化，我们将自适应离散余弦变换损失 $\mathcal{L}_{dct}$、重建损失 $\mathcal{L}_r$ 和感知损失 $\mathcal{L}_p$ 结合起来优化模型。 $\mathcal{L}_{dct}$ 是我们提出的方法，用于优化式中的码率和失真。6。对于重建损失 $\mathcal{L}_r$，我们希望确保模型具备基本的重建能力，因此采用 L1 距离作为重建损失，其公式如下：

$$\mathcal{L}_r = \frac{1}{HW}\sum_{i=0}^{H-1}\sum_{j=0}^{W-1}|f_{i,j}^{GT} - \hat{f}_{i,j}| \tag{7}$$

其中 f^{GT}是 f的处理后真实值。众所周知，高频区域中的对比度或边缘与人类感知的相关性更高。多尺度结构相似性（MS-SSIM） [45]已被证明善于保留高频区域中的结构信息和对比度。因此，我们采用 MS-SSIM 作为感知损失部分，其公式如下：

$$\mathcal{L}_p = 1 - \mathcal{L}_{MS-SSIM}(\hat{f}, f^{GT}) \tag{8}$$

结合 $\mathcal{L}_{dct}$、$\mathcal{L}_r$ 和 $\mathcal{L}_p$，总体损失函数可表示为：

$$\mathcal{L}_{all} = \lambda_1\mathcal{L}_{dct} + \lambda_2\mathcal{L}_p + \mathcal{L}_r \tag{9}$$

其中， λ_1 和 λ_2 分别为码率系数和感知系数。

4. 实验

4.1. 实验设置

数据集。 我们采用 DIV2K 和 Flickr2K 数据集来训练 RPP 模型，其中 DIV2K 包含 1000 张高清 2K 分辨率图像，Flickr2K 包含 2650 张 2K 分辨率图像。为评估所提出方法的性能，我们在 UVG 数据集、HEVC 标准 1080p 测试序列和 MCL-JCV 数据集上进行测试。凭借丰富多样的内容，这些数据集被广泛用于评估视频压缩算法的性能。

实现细节。 我们分两个阶段训练 RPP 模型。第一阶段为预热阶段，使用 Adam 优化器 [21] ，以初始学习率 1×10^{-3}，分别将 β_1 设为 0.9、 β_2 设为 0.999，并基于重建损失 $\mathcal{L}_r$ 训练模型。小批量大小设为 32。训练图像的分辨率为 128×128，图像从数据集中的原始图像随机裁剪得到。经过 600K 次预热训练迭代后，我们使用总体损失函数 $\mathcal{L}_{all}$ 进行训练，其中λ_1 设为 10， λ_2 设为 0.1，并将学习率调整为 1×10^{-4}。具体而言，在自适应离散余弦变换损失的设置中，我们同时使用 N=8和 N=16训练网络，因为传统视频编解码器中最常见的宏块尺寸为 8 和 16。采用此设置，我们将 RPP 模型再训练 700K 次迭代，使其收敛。两个训练阶段所用的训练数据均通过我们的二阶图像退化模型进行增强。整个训练框架基于 PyTorch [32] 实现，使用两块 NVIDIA GeForce RTX3090 训练网络仅需约 1 天。在部署阶段，输入帧首先会送入已部署的 RPP 模型进行预处理。我们在此设置一个超参数 α ，用于控制该方法的预处理强度，以适应某些无需进行强预处理、且对视频中的所有高频信息都很敏感的情况。 α 的取值根据实验经验确定。预处理帧可表示为：

$$f_p = \alpha f_o + (1-\alpha) f_i \tag{10}$$

其中， f_o 是 RPP 模型的输出帧， f_i 是输入帧。随后，预处理帧将由标准视频编解码器进行编码。得益于我们的网络设计，RPP 模型部署在单张 NVIDIA GeForce RTX3090 上并通过 TensorRT [39]运行时，处理 1080p视频的推理性能可达 87.7 FPS。对于 720p 和 4 K 视频，推理性能分别为 185 FPS 和 22.6 FPS。

评估方法。 为了衡量所提出方法的性能，我们采用两种评估指标：多尺度结构相似性和 VMAF。多尺度结构相似性是学术视频编解码领域最常用的指标，而 VMAF 是视频流媒体行业主要的感知导向指标。我们使用 AVC/H.264、HEVC/H.265、VVC/H.266 和 AV1 对所提出的方法进行测试，这些编解码器涵盖了所有主流标准视频编解码器。

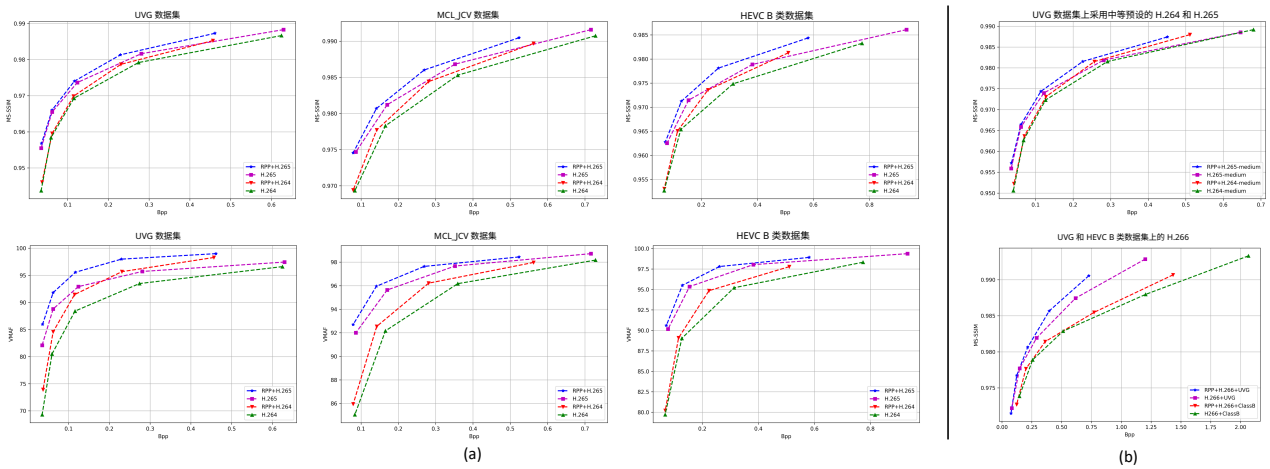


图 3. (a) UVG 数据集、MCL JCV 数据集和 HEVC B 类数据集在多尺度结构相似性和 VMAF 上的率失真曲线。图中绘制了标准编解码器和 RPP + 1, 标准编解码器的曲线。对于每个数据集, 2 3, 我们提出的方法对应的 BD 码率分别列于表格中。(b) 上：UVG 数据集采用中等预设时在多尺度结构相似性上的率失真曲线。下：UVG 数据集和 HEVC B 类数据集的 H.266 在多尺度结构相似性上的率失真曲线。

4.2. 实验结果

本节展示标准视频编解码器与我们的 RPP +标准视频编解码器之间的比较实验结果。我们为 HEVC 数据集和 MCL JCV 数据集固定 $\alpha = 0.5$（见公式）10，为 UVG 数据集固定 $\alpha = 1$。图 3(a) 以及表 1、2、3 的结果表明，在全部三个数据集上，与标准编解码器相比，我们的方法均能显著改善两种指标的 BD 码率。RPP + H.264 在三个数据集上的平均节省量为：VMAF 下 18.21%，多尺度结构相似性下 8.73%。RPP +H.265 在三个数据集上的平均节省量为：VMAF 下 24.62%，多尺度结构相似性下 13.51%。一些基于学习的视频编码器 [29] 已证明，只有在“非常快”预设下才能优于传统标准编解码器。为展示我们方法的通用性，我们还使用“中等”预设测试了 RPP 方法。如图 3(b) 上图所示，我们的方法仍优于标准编解码器，这与图 3(a) 中“非常快”预设下的结果一致。此外，我们还在 UVG 数据集和 HEVC B 类数据集上使用 H.266 测试了 RPP 方法。如图 3(b) 下图所示，RPP +H.266 在这两个数据集上的平均节省量为多尺度结构相似性下的 8.42%。正如预期，与所有主流标准编解码器结合使用时，我们的方法可以取得显著收益。此外，在相同量化参数（QP）下，我们的方法比标准编解码器具有更低的比特率，这体现了该方法节省比特的能力。

	VMAF	多尺度结构相似性
RPP+H.264（非常快）	-26.92	-4.86
RPP+H.265（非常快）	-39.77	-8.70
RPP+H.264（中等）	-27.30	-5.60
RPP+H.265（中等）	-39.24	-9.58

表 1. RPP+H.264 和 RPP+H.265 在 UVG 数据集上的 BD 码率，采用“非常快”和“中等”预设

	VMAF	多尺度结构相似性
RPP+H.264	-15.88	-9.59
RPP+H.265	-19.14	-11.93

表2. RPP+H.264 和 RPP+H.265 在 MCL JCV 数据集上的 BD 码率

	VMAF	多尺度结构相似性
RPP+H.264	-11.84	-11.75
RPP+H.265	-14.94	-19.90

表3. RPP+H.264 和 RPP+H.265 在 HEVC ClassB数据集上的 BD 码率

4.3. 消融研究与分析

自适应 DCT 损失的有效性。为探究自适应DCT损失函数的影响，我们设置 $\lambda_1 = 0$in $\mathcal{L}_{all}$，使自适应DCT损失函数不会影响训练优化。我们在 UVG 数据集上进行了这项消融研究。如图4(a) 所示，在多尺度结构相似性指标下，自适应DCT损失使 H.264 的 BD-rate 节省达到 3.05%，H.265 达到 6.04%，效果非常显著，这一结果

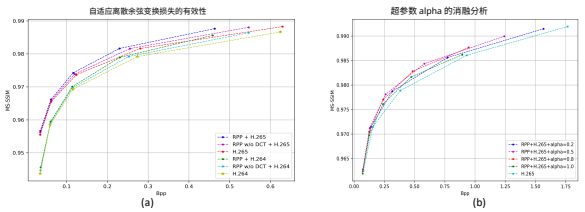


图 4。(a) UVG 数据集上自适应离散余弦变换损失的消融研究 (b) HEVC B 类和 MCL JCV 数据集上超参数 α 的消融研究

令人印象深刻。与表 1 中的 BD-rate 节省相比，它对整个方法的码率节省贡献超过 60%。

超参数的选择与分析 α 我们在 HEVC B 类数据集和 MCL JCV 数据集上测试了所提方法，并在公式中设置了不同的 α 值（0.2、0.5、0.8、1.0）。10。从图 4(b)中可以看出，$\alpha = 0.5$ 对应的率失真曲线优于其他 α值。正如前文所述，α 用于控制所提方法的预处理强度。从我们的角度来看，我们需要使用超参数控制强度，原因有二。首先，我们的模型是在一个固定设置下，使用规模较小的公开数据集训练的，这意味着数据多样性不足。其次，有些视频对高频分量极为敏感，而采用固定设置的预训练模型可能会对其进行过度预处理。

5. 结论

本文提出了一种率感知优化预处理（RPP）方法，通过神经网络生成经过码率优化且感知质量增强的帧，用于视频编码。在部署阶段，我们的 RPP 方法可即插即用地应用于标准视频编解码器，无需更改编码和解码设置。此外，我们提出的方法效率很高，性能远超实时处理。如实验结果所示，我们的 RPP 方法在不同指标上均能与所有主流标准视频编解码器实现显著且稳定的增益。

参考文献

[1] Eirikur Agustsson 和 Radu Timofte。NTIRE 2017 单幅图像超分辨率挑战赛：数据集与研究。载于*IEEE* 计算机视觉与模式识别会议（*CVPR*）研讨会，2017 年 7 月。3，5[2] N. Ahmed、T. Natarajan 和 K.R. Rao。离散余弦变换。*IEEE*《计算机汇刊》，C-23(1)：90–93，1974 年。1[3] Johannes Ball´e、Valero Laparra 和 Eero P Simoncelli。端到端优化图像压缩。*arXiv* 预印本 *arXiv:1611.01704*，2016 年。2

[4] Johannes Ball´e、David Minnen、Saurabh Singh、Sung Jin Hwang 和 Nick Johnston。基于尺度超先验的变分图像压缩。arXiv 预印本 arXiv:1802 01436，2018 2。.[5] Nabajeet Barman、Steven Schmidt、Saman Zadtootaghaj、Maria G Martini 和 Sebastian M¨oller。被动游戏视频流的视频质量评估指标评估。载于第 23 届分组视频研讨会论文集，第 7–12 页，2018。2，3[6] Fabrice Bellard。BPG 图像格式（2014 2，2016 1 2）。卷，：。[7] Frank Bossen 等。通用测试条件和软件参考配置。JCTVC-L1100，12(7)，2013。5[8] Benjamin Bross、Ye-Kui Wang、Yan Ye、Shan Liu、Jianle Chen、Gary J Sullivan 和 Jens-Rainer Ohm。通用视频编码（VVC）标准及其应用概述。IEEE《视频技术电路与系统汇刊》，31(10):3736–3764，2021。1，2，3[9] Aaron Chadha 和 Yiannis Andreopoulos。面向视频编码的深度感知预处理。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 14852–14861 页，2021。2[10] Tong Chen、Haojie Liu、Qiu Shen、Tao Yue、Xun Cao 和 Zhan Ma。DeepCoder：一种基于深度神经网络的视频压缩方法。载于 2017 年 IEEE 视觉通信与图像处理（VCIP），第 1–4 页。IEEE，2017。1，2[11]Yue Chen、Debargha Mukherjee、Jingning Han、Adrian Grange、Yaowu Xu、Sarah Parker、Cheng Chen、Hui Su、Urvang Joshi、Ching-Han Chiang 等。AV1 编码工具概述：开放媒体联盟推出的首款视频编解码器。APSIPA 信号与信息处理汇刊，9，2020。1，2，3[12]Yue Chen、Debargha Murherjee、Jingning Han、Adrian Grange、Yaowu Xu、Zoe Liu、Sarah Parker、Cheng Chen、Hui Su、Urvang Joshi 等。AV1 视频编解码器核心编码工具概述。载于 2018 年图像编码研讨会（PCS），第 41–45 页。IEEE，2018。3[13] Charilaos Christopoulos、Athanassios Skodras 和 Touradj Ebrahimi。JPEG2000 静态图像编码系统概述。IEEE《消费电子学汇刊》，46(4):1103–1127，2000。2[14] U Cisco。思科年度互联网报告（2018–2023）白皮书。思科：美国加利福尼亚州圣何塞，2020。1[15] Corinna Cortes 和 Vladimir Vapnik。支持向量网络。《机器学习》，20(3):273–297，1995。3[16] Michael Elad 和 Arie Feuer。从多幅模糊、含噪且欠采样的测量图像中恢复单幅超分辨率图像。IEEE《图像处理汇刊》，6(12):1646–1658，1997。3，4[17] Bernd Girod。均方误差有什么问题？《数字图像与人类视觉》，第 207–220 页，1993。2[18] Jie Hu、Li Shen 和 Gang Sun。压缩与激励网络。载于 IEEE 计算机视觉与模式识别会议论文集，第 7132–7141 页，2018。3，4[19]David A Huffman。最小冗余码的构造方法。《IRE 会议录》，40(9):1098–1101，1952。4

[20] Zheng Hui、Xinbo Gao、Yunchu Yang 和 Xiumei Wang。采用信息多重蒸馏网络的轻量级图像超分辨率。载于第 27 届 ACM 多媒体国际会议论文集，第 2024–2032 页，2019。3，4[21] Diederik P Kingma 和 Jimmy Ba。Adam：一种随机优化方法。arXiv 预印本 arXiv:1412 6980，2014 5。.[22] Faouzi Kossentini、Hassen Guermazi、Nader Mahdi、Chekib Nouira、Amir Naghdinezhad、Hassene Tmar、Omar Khlif、Phoenix Worth 和 Foued Ben Amara。SVT-AV1 编码器：概述、特性及速度—质量权衡。《数字图像处理应用 XLIII》，11510:469–490，2020。3[23] Songnan Li、Fan Zhang、Lin Ma 和 King Ngi Ngan。通过分别评估细节损失和加性失真来进行图像质量评估。IEEE《多媒体汇刊》，13(5):935–949，2011。3[24]Yawei Li、Kai Zhang、Radu Timofte、Luc Van Gool、Fangyuan Kong、Mingxi Li、Songwei Liu、Zongcai Du、Ding Liu、Chenhui Zhou 等。NTIRE 2022 高效超分辨率挑战赛：方法与结果。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 1062–1102 页，2022。[25] Zhi Li、Anne Aaron、Ioannis Katsavounidis、Anush Moorthy 和 Megha Manohara。迈向实用的感知视频质量指标。Netflix 技术博客，6(2)，2016。2，3[26] Zhi Li、Christos Bampis、Julie Novak、Anne Aaron、Kyle Swanson、Anush Moorthy 和 JD Cock。VMAF：持续演进。Netflix 技术博客，25，2018。3[27] Jie Liu、Jie Tang 和 Gangshan Wu。用于轻量级图像超分辨率的残差特征蒸馏网络。载于欧洲计算机视觉会议论文集，第 41–55 页。Springer，2020。3，4[28] Zhenyu Liu、Xianyu Yu、Yuan Gao、Shaolin Chen、Xiangyang Ji 和 Dongsheng Wang。使用卷积神经网络为 HEVC 硬连线帧内编码器进行 CU 分区模式决策。IEEE《图像处理汇刊》，25(11):5088–5103，2016。1，2[29] Guo Lu、Wanli Ouyang、Dong Xu、Xiaoyun Zhang、Chunlei Cai 和 Zhiyong Gao。DVC：端到端深度视频压缩框架。载于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 11006–11015 页，2019。1，2，6[30] Alexandre Mercat、Marko Viitanen 和 Jarno Vanne。UVG 数据集：用于视频编解码器分析与开发的 50/120 fps 4K 序列。载于第 11 届 ACM 多媒体系统会议论文集，第 297–302 页，2020。5[31] David Minnen、Johannes Ball´e 和 George D Toderici。用于学习式图像压缩的联合自回归与层次先验。《神经信息处理系统进展》，31，2018。2[32] Adam Paszke、Sam Gross、Francisco Massa、Adam Lerer、James Bradbury、Gregory Chanan、Trevor Killeen、Zeming Lin、Natalia Gimelshein、Luca Antiga 等。PyTorch：一种命令式风格、高性能的深度学习库。《神经信息处理系统进展》，32，2019。5

[33] Reza Rassool。VMAF 可复现性：验证一种实用的感知视频质量指标。载于《2017年 IEEE 宽带多媒体系统与广播国际研讨会（BMSB）论文集》，第1–2页。IEEE，2017年。3[34] Shankar L Regunathan、Haixiong Wang、Yun Zhang、Yu Ryan Liu、David Wolstencroft、Srinath Reddy、Cosmin Stejerean、Sonal Gandhi、Minchuan Chen、Pankaj Sethi等。Facebook 视频生态系统中的大规模高效质量测量。载于《数字图像处理应用 XLIII》，第11510卷，第69–80页。SPIE，2020年。3[35]J. Rissanen 和 G. G. Langdon。算术编码。IBM《研究与开发期刊》，第23卷第2期：149–162，1979年。4[36] Hamid R Sheikh 和 Alan C Bovik。图像信息与视觉质量。IEEE《图像处理汇刊》，15(2)：430–444，2006年。2, 3[37] Wenzhe Shi、Jose Caballero、Ferenc Husz´ár、Johannes Totz、Andrew P Aitken、Rob Bishop、Daniel Rueckert 和 Zehan Wang。使用高效子像素卷积神经网络实现实时单图像和视频超分辨率。载于《IEEE计算机视觉与模式识别会议论文集》，第1874–1883页，2016年。4[38] Gary J Sullivan、Jens-Rainer Ohm、Woo-Jin Han 和 Thomas Wiegand。高效视频编码（HEVC）标准概述。IEEE《视频技术电路与系统汇刊》，22(12)：1649–1668，2012年。1, 2, 3[39] Han Vanholder。使用 TensorRT 进行高效推理。载于GPU技术大会，第

1，　2, 2016　5 页。

[40] Gregory K Wallace。JPEG 静态图像压缩标准。ACM通讯，34(4)：30–44，1991年。2[41] Haiqiang Wang、Weihao Gan、Sudeng Hu、Joe Yuchieh Lin、Lina Jin、Longguang Song、Ping Wang、Ioannis Katsavounidis、Anne Aaron 和 C-C Jay Kuo。MCL-JCV：基于刚可觉差的 H.264/AVC 视频质量评估数据集。载于《2016年 IEEE国际图像处理会议（ICIP）论文集》，第1509–1513页。IEEE，2016年。5[42] Xintao Wang、Liangbin Xie、Chao Dong 和 Ying Shan。Real-ESRGAN：使用纯合成数据训练面向真实场景的盲超分辨率模型。载于《IEEE/CVF计算机视觉国际会议论文集》，第1905–1914页，2021年。3, 5 – .[43] Zhou Wang 和 Alan C Bovik。均方误差：是取是舍？重新审视信号保真度度量。IEEE《信号处理杂志》，26(1)：98–117，2009年。2[44] Zhou Wang、Alan C Bovik、Hamid R Sheikh 和 Eero P Simoncelli。图像质量评估：从误差可见性到结构相似性。IEEE《图像处理汇刊》，13(4)：600–612，2004年。2, 3[45] Zhou Wang、Eero P Simoncelli 和 Alan C Bovik。用于图像质量评估的多尺度结构相似性。载于《第37届阿西洛马会议：信号、系统与计算，2003年》，第2卷，第1398–1402页。IEEE，2003年。2, 3, 5[46] Thomas Wiegand、Gary J Sullivan、Gisle Bjontegaard 和 Ajay Luthra。H.264/AVC 视频编码标准概述。IEEE《视频技术电路与系统汇刊》，13(7)：560–576，2003年。1, 2, 3

[47] Fan Zhang、Angeliki V Katsenou、Mariana Afonso、Goce Dimitrov 和 David R Bull。使用客观和主观评估比较 VVC、HEVC 与 AV1。*arXiv* 预印本 *arXiv:2003.10282*，2020年。3[48] Han Zhang、Jizheng Xu 和 Li Song。基于视频多方法评价融合的可变视频编码率失真优化。载于*2021*年 *IEEE*国际图像处理会议（*ICIP*），第2064–2068页。IEEE，2021年。3[49] Saiping Zhang、Marta Mrak、Luis Herranz、Marc G´orriz Blanch、Shuai Wan 和 Fuzheng Yang。DVC-P：采用感知优化的深度视频压缩。载于*2021*年视觉通信与图像处理国际会议（*VCIP*），第1–5页。IEEE，2021年。2