

北京航空航天大学学报
*Journal of Beijing University of Aeronautics and Astronautics*
ISSN 1001-5965,CN 11-2625/V

# 《北京航空航天大学学报》网络首发论文

题目：频率感知上下文建模的图像编码算法
作者：李玲雪，夏凡鑫，刘美琴，姚超，赵耀
DOI：10.13700/j.bh.1001-5965.2026.0489
收稿日期：2026-07-12
网络首发日期：2026-11-02
引用格式：李玲雪，夏凡鑫，刘美琴，姚超，赵耀．频率感知上下文建模的图像编码算法[J/OL]．北京航空航天大学学报．
https://doi.org/10.13700/j.bh.1001-5965.2026.0489



**网络首发**：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

**出版确认**：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

# 频率感知上下文建模的图像编码算法

李玲雪[1,2]，夏凡鑫[1,2]，刘美琴[1,2,✉]，姚超[3]，赵耀[1,2]

(1．北京交通大学 信息科学研究所，北京 100044；
2．北京交通大学 视觉智能交叉创新教育部国际合作联合实验室，北京 100044；
3．北京科技大学 计算机与通信工程学院，北京 100083)
*通信作者 E-mail：mqliu@bjtu.edu.cn

摘要 基于深度学习的图像编码算法通过神经网络端到端优化率失真损失函数，学习自然图像的高阶统计特征，获得良好的压缩性能，成为多媒体处理领域的核心发展方向。然而，由于现有算法未充分利用自然图像固有的频谱分布先验，导致对图像频谱结构信息利用不足、高频纹理区域的建模精度不足。因此，本文提出了一种频率感知上下文建模的图像编码算法。该算法设计 DCT 初始化的空频变换机制，将图像映射至频域，解耦低频结构与高频纹理信息。该算法还构建了频率感知上下文模块，提取频域统计特征先验信息，通过残差融合注入空间上下文，使模型自适应地区分高低频区域。此外，本文构建了带感知门控熵模型，通过轻量门控机制调制高斯分布参数，对不同频带系数进行差异化建模，解决统一建模适配性差的问题。实验结果表明，该算法在 Kodak 测试集上较图像编码标准 VVC (4:4:4) 节省 8.33% 的码率，参数量仅 12.51M，重建了更加清晰的高频纹理区域。

关键词 图像编码；深度学习；频率感知；熵编码；上下文建模
中图分类号 TP391 文献标志码 A DOI: 10.13700/j.bh.1001-5965.2026.0489

# Frequency-aware context modeling for learned image compression

LI Lingxue[1,2]，XIA Fanxin[1,2]，LIU Meiqin[1,2,✉]，YAO Chao[3]，ZHAO Yao[1,2]

(1．Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China；
2．Visual Intelligence+X International Cooperation Joint Laboratory, Beijing Jiaotong University, Beijing 100044, China；3．School of Computer and Communication Engineering, University of Science and Technology Beijing，Beijing 100083, China)
*E-mail: mqliu@bjtu.edu.cn

**Abstract** Learned image coding, which optimizes rate-distortion loss end-to-end and captures high-order statistics of natural images, has become a key direction in multimedia compression. However, existing methods insufficiently exploit spectral structures and lack accuracy in modeling high-frequency textures. To this end, we propose a frequency-aware context-based image compression method. It employs a DCT-initialized spatial-frequency transform to decouple low-frequency structures and high-frequency textures, uses a frequency-aware context module to fuse frequency priors and spatial context, and introduces a gated entropy model to separately model coefficients of different frequency bands. Experiments on the Kodak dataset show that our approach outperforms VVC (4:4:4) by 8.33% bitrate saving with only 12.51M parameters, and yields sharper high-frequency reconstructions at equivalent bitrates.
**Key words** image coding; deep learning; frequency domain analysis; entropy coding; context modeling

图像编码技术是降低数据存储成本、提升传输效率的核心支撑，也是多媒体信号处理领域的长期研究热点。传统图像编码标准（如 JPEG[1]、JPEG2000[2]和 VVC(4:4:4)[3]等）基于手工设计的算法矩

收稿日期：2026-07-12
基金项目：国家自然科学基金（62372036, 62120106009）
Fund：National Natural Science Foundation of China (62372036, 62120106009)
网络首发时间：2026-11-02 10:30:08 网络首发地址：https://link.cnki.net/urlid/11.2625.V.20261030.1718.003

阵，难以适配复杂多样的自然图像内容。基于深度学习的端到端图像编码算法凭借强大的特征学习与非线性拟合能力实现了性能突破，逐步成为图像压缩领域的主流研究方向。这类算法通过神经网络端到端联合优化率失真损失，自动学习自然图像的高阶统计特征，替代了传统编码中固定的变换、量化与熵建模模块，在多个公开测试集上的率失真性能已逐步超越传统编码标准。现有研究主要围绕变换网络优化与熵模型改进两大核心方向推进。在变换网络层面，从最初的单尺度卷积架构逐步发展为多尺度分解、残差增强、可逆网络等多种形态，不断提升隐特征的能量聚集效率；在熵模型层面，从早期的因子化熵模型，到超先验架构、自回归上下文模型，再到基于 Transformer 的长程依赖建模，上下文信息的利用维度与表征深度持续拓展，熵模型的概率预测精度稳步提升。尽管现有深度学习图像编码方法已取得显著进展，但其大多在像素域开展特征建模与熵编码，未能充分挖掘自然图像固有的频域能量聚集先验。自然图像的能量主要集中于低频分量，承载图像的主体结构信息，而高频分量对应纹理细节，具有稀疏性与重尾分布特性，二者的统计规律存在显著差异。现有编码算法对全图所有区域采用统一的特征建模与参数预测策略，导致稀疏高频系数的建模精度严重不足，最终造成重建图像高频细节模糊、纹理质感丢失。

将经典频域先验引入深度学习编码框架，是突破当前性能瓶颈的可行研究路径。已有部分研究尝试将离散余弦变换与深度网络结合，或直接在频域开展编码优化，但多数方案仍存在明显局限：一类方法采用固定的离散余弦变换 (Discrete Cosine Transform, DCT) 基函数实现频域映射，无法根据图像内容自适应调整变换核，难以匹配复杂自然图像的多样化统计分布；另一类方法虽构建了空频联合编码结构，但仅通过简单的通道拼接实现特征融合，频域信息与空间上下文未能实现深度耦合，频域先验难以有效赋能熵参数估计过程；还有部分频域感知编码方案仅将频域特征作为辅助信息浅层融入网络，未在上下文建模与熵模型设计层面建立频带自适应的差异化建模机制，高频系数的建模精度提升十分有限。整体而言，如何实现频域先验与深度学习编码框架的端到端深度融合，构建可自适应区分高低频区域的上下文建模与熵估计机制，仍是当前亟待解决的关键问题。

因此，本文提出了一种频率感知上下文建模的图像编码算法，主要贡献点如下：

(1) 设计空频变换机制完成像素域到频率域的映射，在保留能量聚集特性的同时，适配自然图像的统计特性。

(2) 构建了频率感知的上下文模块，将频谱特征先验与空间上下文特征残差融合，实现空频跨尺度的联合上下文表征。

(3) 提出了带感知门控熵模型，利用频谱特征生成轻量级门控图，逐通道、逐位置地调制概率分布参数，对不同频带系数进行差异化建模。

## 1 相关工作

### 1.1 基于变换网络优化的图像编码

变换网络是端到端图像编码的基础模块，通过将像素域图像转换为能量高度聚集的隐特征域，为后续熵编码提供更易压缩的特征表示。Ballé等人[4]首次提出了包含分析变换、合成变换、熵模型的端到端压缩框架，用多层卷积网络替代了传统编码中固定的离散余弦变换 DCT，为后续深度学习图像编码工作奠定了基础框架。然而，该工作采用单尺度卷积变换，对图像多尺度特征的表征能力有限，能量聚集效率仍有较大提升空间。

针对单尺度变换的局限，研究者们围绕变换网络的多尺度分解与特征表征能力展开了持续优化。Johnston 等人[5]提出了多尺度变换网络架构，通过层级下采样与上采样实现全局结构与局部细节的联合建模，显著提升了变换网络的表征能力。在此基础上，Li 等人[6]将残差块、通道注意力模块引入变换网络，进一步提升了隐特征的能量聚集效率。Liu 等人[7]采用 Transformer 与卷积混合架构以兼顾长程依赖建模与局部特征提取，但参数量与计算量显著上升。

为了降低变换网络的参数量与计算开销，研究者进一步从轻量化网络、高效长程建模及模型压缩等角度开展优化。Tu 等人[8]基于多尺度可逆神经网络构建了低复杂度变换框架，Chen 等人[9]利用内

容自适应的状态空间建模增强空间距离较远但内容相关特征之间的交互，在保持较高计算效率的同时提升全局冗余建模能力。Chen 等人[10]进一步从模型压缩角度引入知识蒸馏，通过分阶段知识迁移降低学习式图像编码模型的计算开销。上述方法主要从模型表征与计算效率角度进行优化，对自然图像固有频域统计特性的利用仍较有限。尽管已有少数工作[11][12]尝试将 DCT 变换引入深度学习编码框架，借助频域先验辅助特征学习，却采用固定的 DCT 基函数，难以匹配复杂自然图像的多样化统计分布。

### 1.2 基于熵模型优化的图像编码

熵模型的核心目标是依托已解码上下文精准预测编码符号的概率分布，进而最小化熵编码码率；上下文模型为熵模型提供高精度先验信息，是提升熵模型预测精度的关键。

Ballé等人[4]采用了因子化熵模型，假设隐变量的各通道、各空间位置之间相互独立，仅对单变量的分布进行建模，无法利用特征间的相关性。针对因子化熵模型的独立假设缺陷，Ballé 等人[13]提出了 Hyperprior 超先验架构，通过超网络建模隐变量全局通道相关性，显著降低了码率开销，却无法捕捉局部空间邻域相关信息，熵模型预测精度有限。为了充分利用空间邻域相关性，Minnen 等人[14]首次将自回归模型与超先验架构结合，同步建模全局通道与局部空间相关性，却受限于因果卷积网络，易产生棋盘伪影和未来信息泄露的问题。针对这一问题，He 等人[15]提出了棋盘掩码自回归上下文模型，保证自回归因果约束。

在此基础上，研究者们进一步拓展上下文建模的感受野与表征能力。Qian 等人[16]引入 Transformer 建模长程特征依赖，但参数量与解码延迟显著增加。Zeng 等人[17]将状态空间模型引入上下文建模，并结合窗口局部注意力增强通道与空间相关性建模，以兼顾长程依赖建模能力和计算效率。He 等人[18]提出的空间-通道上下文自适应编码框架，通过上下文融合模块与增强的变换网络，扩大了上下文感受野。Hu 等人[19]提出的多参考熵模型框架，构建了多尺度层级上下文模型，实现跨尺度上下文信息融合。Jiang 等人[20]进一步增强多参考上下文与全局相关性建模，提高隐变量概率估计精度。

少数工作尝试引入频域特征辅助熵建模，Li 等人[21]提出的频域感知 Transformer 压缩框架将频域特征与空间特征简单拼接后输入熵模型，实现了频域与空间特征的浅层融合，但高频建模精度提升有限。

## 2 本文算法

### 2.1 整体网络架构

本文提出的频率感知上下文建模的图像编码算法 FLIC，框架如图 1 所示。在编码阶段，输入图像 $x$ 经空频变换机制从像素域图像映射至频率域，在保留图像的频谱结构信息的同时得到频域系数图 $Y_{\text{coef}} \in \mathbb{R}^{B\times 3\times H\times W}$ 。随后，频域系数图 $Y_{\text{coef}}$ 输入多尺度频率特征提取网络，通过层级的卷积操作与频域可逆单元完成多尺度分解，生成 5 个不同尺度的隐变量切片 $y=(y_1,y_2,y_3,y_4,y_5)$ ，实现从粗到细的渐进式编码。对于每个隐变量切片，算法首先利用已解码的低尺度切片生成通道上下文特征，捕捉全局跨尺度相关性；再基于已解码的锚点系数，通过频率感知上下文模块生成融合频谱信息的空间上下文特征，捕捉局部空间依赖关系并融入频域统计先验信息。将两类特征拼接后输入熵参数预测网络，由带感知门控熵模型生成均值与方差，对残差系数进行量化与熵编码，最终获得当前切片的码流。

在解码阶段，该算法基于码流与熵模型逐切片完成解码，通过可逆变换的逆过程重建频率系数图，再经过可学习空频变换的逆变换得到像素域重建图像，最终经后处理网络优化输出最终图像 $\hat{x}$ 。此外，该算法构建了增益单元和逆增益单元，分别用于编码端调整隐变量系数的动态范围和解码端恢复系数的原始动态范围。

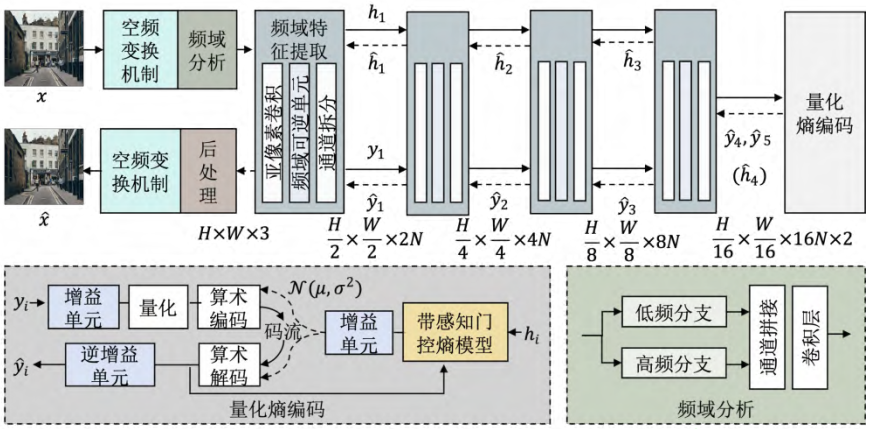



图 1 频域感知上下文建模的图像编码算法结构图
Fig.1 The architecture of the proposed FLIC

## 2.2 空频变换机制

自然图像在频域具有明显的能量紧致特性，其能量主要集中于低频分量。传统编码标准普遍采用固定 DCT 变换实现像素域到频域的映射，但其固定基函数缺乏自适应性，难以匹配复杂多样的图像内容。为此，本文提出 DCT 初始化的空频变换机制，基于 DCT 频域先验对变换基进行优化。本文所提可学习块变换通过二维可分离正交变换矩阵 $Q$ 实现。为保证变换矩阵在训练过程中始终满足正交约束，采用 QR 分解对其进行参数化。在初始化阶段，使用标准 $8\times8$ 维 DCT 正交矩阵对可学习参数矩阵 $\boldsymbol{W}$ 赋值，为模型注入可靠的频域先验。训练阶段，对参数矩阵 $\boldsymbol{W}$ 执行 QR 分解，可表示为：

$$\boldsymbol{W}=\boldsymbol{Q}\boldsymbol{U} \quad (1)$$

式中: $\boldsymbol{Q}$ 表示正交矩阵，$\boldsymbol{U}$ 表示上三角矩阵。由于正交矩阵满足 $\boldsymbol{Q}^{-1}=\boldsymbol{Q}^{\mathrm{T}}$ ，逆变换可直接采用转置矩阵实现，无需额外计算逆矩阵。

对于给定输入特征图 $Y\in\mathbb{R}^{B\times C\times H\times W}$ ，前向变换首先对特征图 $Y$ 的高、宽方向进行反射填充，使填充后特征图的尺寸为块大小的整数倍，以减轻块效应。随后将填充后的特征图划分为若干 $8\times8$ 的非重叠块，得到分块张量 $\mathbb{R}^{B\times C\times H_b\times W_b\times b\times b}$ ，其中 $H_b$ 、 $W_b$ 分别为分块后的高和宽。之后利用二维可分离正交变换对每个块计算得到频域系数块，再经重排恢复为与输入同尺寸的系数图 $Y_{\text{coef}}$ 。

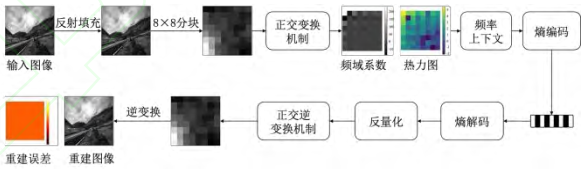



图 2 空频变换机制结构图
Fig.2 Illustration of the frequency and bandwidth conversion mechanism

## 2.3 频率感知的上下文模型

图像经块变换得到频率系数后，在不同空间位置的频域能量分布存在差异。平滑区域能量集中于低频分量，纹理区域高频分量占比显著提升。因此，本文提出频率感知的上下文模型，结构如图 3 所示。该模型将频谱统计特征显式融入自回归上下文预测过程，实现更精准的系数预测。

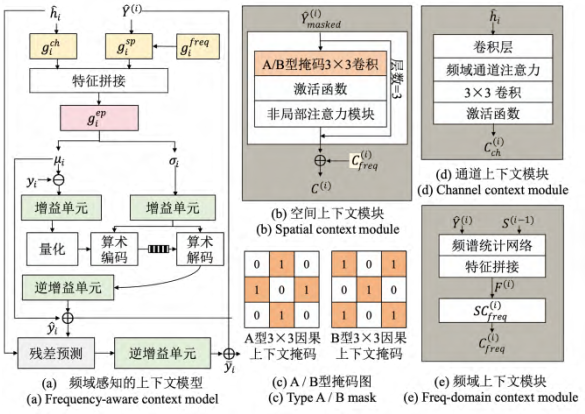



图 3 频域感知的上下文模型结构图
Fig.3 Illustration of the frequency-aware context model

具体地，该模块首先构造 12 维频谱特征图 $F^{(i)}$ ，刻画当前隐变量的频域特性，可表示为：

$$F^{(i)} = \mathrm{Concat}\left(F_{\mathrm{stat}}\left(\hat{Y}^{(i)}\right), F_{\mathrm{stat}}\left(S^{(i-1)}\right)\right) \quad (2)$$

式中：$\mathrm{Concat}(\cdot)$ 表示沿通道维度的特征拼接操作；$F_{\mathrm{stat}}(\cdot)$ 表示 6 维频谱统计提取函数；$\hat{Y}^{(i)}$ 表示当前层用于上下文计算的已解码隐变量特征图，区别于切片重建值 $\hat{y}_i$ ；上标 $(i)$ 表示当前上下文建模层级，$S^{(i-1)}$ 表示前一已解码低尺度的超先验解码特征，最粗尺度时赋值为零。该统计函数提取特征的绝对值均值、标准差等频域统计量，表征局部区域的频域能量分布。

在空间上下文建模阶段，首先通过 A/B 型掩码卷积提取原始空间上下文特征 $C_{\mathrm{sp}}^{(i)}$ ，表示为：

$$C_{\mathrm{sp}}^{(i)} = SC^{(i)}\left(\hat{Y}_{\mathrm{masked}}^{(i)}\right) \quad (3)$$

式中：$SC^{(i)}(\cdot)$ 表示掩码卷积网络；$\hat{Y}_{\mathrm{masked}}^{(i)}$ 表示掩码处理后的已解码隐变量。随后引入频谱上下文网络，将 12 维频谱特征映射为匹配维度的频谱感知特征 $C_{\mathrm{freq}}^{(i)}$ ，可表示为：

$$C_{\mathrm{freq}}^{(i)} = SC_{\mathrm{freq}}^{(i)}\left(F^{(i)}\right) \quad (4)$$

式中：$SC_{\mathrm{freq}}^{(i)}$ 表示轻量级频谱上下文网络，仅由 $1\times1$ 卷积构成；$F^{(i)}$ 表示 12 维频谱特征图。

为保留空间上下文的有效信息，该模块采用残差融合策略注入频域先验得到融合后的上下文特征 $C^{(i)}$ ，可表示为：

$$C^{(i)} = C_{\mathrm{sp}}^{(i)} \oplus C_{\mathrm{freq}}^{(i)} \quad (5)$$

式中：“ $\oplus$ ”表示逐元素相加。最后将融合特征与通道上下文特征 $C_{\mathrm{ch}}^{(i)}$ 拼接，输入熵参数预测网络生成高斯分布参数 $G^{(i)}$ ，可表示为：

$$G^{(i)} = g_i^{\mathrm{ep}}\left(\mathrm{Concat}\left(C_{\mathrm{ch}}^{(i)}, C^{(i)}\right)\right) \quad (6)$$

式中：$g_i^{\mathrm{ep}}$ 表示熵参数预测网络.

## 2.4 带感知门控熵模型

不同频带对应的频域系数统计特性存在显著差异，如低频系数分布平滑且近似高斯分布，高频系数呈现稀疏性与重尾特性。传统熵模型采用统一网络，难以适配频带间的统计差异，建模精度受限。针对这一问题，本文提出带感知门控熵模型，利用频谱特征通过轻量级门控机制实现频带自适应参数调制，结构如图 4 所示。

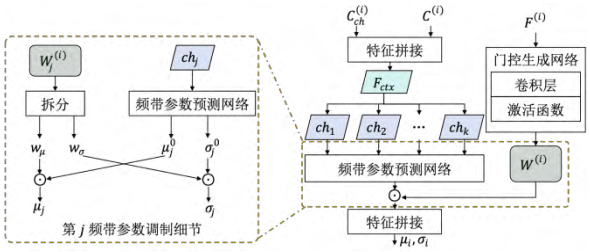



图 4 带感知门控熵模型结构图
Fig.4 Illustration of the frequency-aware gated entropy model

随后，对融合上下文特征按通道维度进行分组拆分，将完整通道划分为 $k$ 个互不重叠的通道组 $ch_1, ch_2, \ldots, ch_k$ ，每个通道组对应一个独立的频带，拆分过程可表示为：

$$[ch_1, ch_2, \ldots, ch_k] = \mathrm{Split}(F_{\mathrm{ctx}}, C/k) \tag{7}$$

式中：$\mathrm{Split}(\cdot)$ 表示通道拆分操作；$C$ 表示道数；$k$ 为频带分组数。

在生成频带组的同时，该模型基于 12 维频谱特征图生成轻量级频谱门控图 $W^{(i)}$ ，可表示为：

$$W^{(i)} = \mathrm{Sigmoid}\left(h_{\phi}^{(i)}\left(F^{(i)}\right)\right) \tag{8}$$

式中：$\mathrm{Sigmoid}(\cdot)$ 表示 Sigmoid 激活函数，用于将门控值约束在 $[0,1]$ 区间；$h_{\phi}^{(i)}(\cdot)$ 表示轻量级 $1\times1$ 卷积网络；$F^{(i)}$ 表示 12 维频谱特征图。随后，将门控图按频带分组，得到每个频带对应的门控权重，用于后续参数调制。接下来，为每个频带通道组 $ch_j$ 设计共享权重的轻量子网络，对每个频带的上下文特征进行独立建模，生成该频带的初始均值 $\mu_j^0$ 与方差参数 $\sigma_j^0$ ，可表示为：

$$\left[\mu_j^0, \sigma_j^0\right] = \mathbb{N}_j(ch_j) \tag{9}$$

式中：$\mathbb{N}_j$ 表示频带参数预测子网络。

生成初始参数后，利用对应频带的门控权重对初始参数进行逐元素调制，实现带感知的参数优化，得到第 $j$ 个频带最终的均值 $\mu_j$ 和方差 $\sigma_j$ ，可表示为：

$$\mu_j = \mu_j^0 \otimes W_{j,\mu}^{(i)},\quad \sigma_j = \sigma_j^0 \otimes W_{j,\sigma}^{(i)} \tag{10}$$

式中:“ $\otimes$ ”表示逐元素相乘操作；$W_{j,\mu}^{(i)}$ 、 $W_{j,\sigma}^{(i)}$ 分别表示第 $j$ 个频带对应均值、方差的门控权重。最后，将所有频带的均值和方差参数沿通道维度拼接，得到第 $i$ 尺度隐变量的高斯分布参数 $\mu_i$ 和 $\sigma_i$ 。

## 2.5 损失函数

本文 FLIC 算法的优化仍遵循率失真损失 $L$ ，可表示为：

$$L = D + \lambda R \tag{11}$$

式中：$D$ 表示重建图像与原始图像的失真损失；$\lambda$ 表示率失真平衡系数，用于调整失真与码率的优化权重；$R$ 表示熵编码的总码率损失。

此外，本文针对频域感知建模的设计，引入频域失真辅助损失函数 $D_{\mathrm{freq}}$ 用于保证频域系数的重建精度，保证频域能量一致性，可以表示为：

$$D_{\mathrm{freq}} = E\left[||Y_{\mathrm{coef}} - \hat{Y}_{\mathrm{coef}}||_2^2\right] \tag{12}$$

式中：$Y_{\mathrm{coef}}$ 表示原始图像经空频变换得到的频域系数；$\hat{Y}_{\mathrm{coef}}$ 表示解码重建的频域系数。将核心率失真损失与频域辅助损失加权融合，得到总损失函数 $L_{\mathrm{total}}$ 可表示为：

$$L_{\mathrm{total}} = D + \lambda R + \alpha D_{\mathrm{freq}} \tag{13}$$

式中：$\alpha$ 表示频域损失权重系数，用于平衡主损失与频域约束。

# 3 实验结果与分析

## 3.1 实验设置

### 3.1.1 训练设置

训练集采用 ImageNet 数据集[22]，训练过程采用 2 张 NVIDIA GeForce RTX 3090 Ti GPU 进行并行训练，全局批次大小为 8，共训练 1000 个轮次，完整训练时间约为 60 h，训练期间单张 GPU 显存占用约为 8 GB。初始学习率为 $1\times10^{-4}$，在轮次 675 和 720 时，学习率分别衰减为原来的 0.1 倍。训练过程中，拉格朗日乘子 $\lambda$ 的取值为{0.0018, 0.0035, 0.0067, 0.0130, 0.0250, 0.0483, 0.0932, 0.1806}，实现多码点的联合优化。模型采用 Adam 优化器，梯度裁剪的最大范数为 1.0。训练过程中，对输入图像进行随机裁剪得到 256×256 的图像块，通过随机水平翻转进行数据增强，最终将像素值归一化到 [0,1] 的范围。损失函数使用均方误差（Mean Squared Error, MSE）和多尺度结构相似性（Multi-Scale Structural Similarity Index Measure, MS-SSIM）两种失真度量，分别用于提升编码模型的峰值信噪比（Peak Signal-to-Noise Ratio, PSNR）和 MS-SSIM 。

### 3.1.2 测试设置

本文采用 Kodak 数据集与 CLIC 数据集作为测试数据集。其中，Kodak 数据集包含 24 张分辨率为 512×768 或 768×512 的自然图像。CLIC 数据集包含 41 张分辨率为 2048×1370 的高分辨率自然图像，用于验证算法在高分辨率图像上的编码性能。

实验采用 PSNR 与 MS-SSIM 作为客观评价指标，其中，由于 MS-SSIM 指标集中于 0.9 到 1 之间，从而难以区分，本文通过 $-10\times log_{10}(1-\mathrm{MS-SSIM})$ 将其转换为分贝（dB）形式，便于结果对比。为定量评估算法的率失真性能，本文采用 Bjøntegaard Delta (BD) 指标，以 VVC(4:4:4)[3] 作为基准算法。BD-PSNR 表示相同码率条件下，测试算法相对基准算法的平均 PSNR 变化量（单位：dB），数值越大性能越优；BD-Rate 表示相同重建质量条件下，测试算法相对基准算法的平均码率变化率（单位：%），数值越低码率节省效果越显著。

除率失真性能外，本文进一步从模型复杂度和运行效率两个方面评价各算法。其中，模型复杂度采用参数量和 FLOPs 进行衡量，运行效率采用编码时间和解码时间进行衡量。各算法均在相同硬件和软件环境下进行测试，运行时间在 NVIDIA GeForce RTX 3090 Ti GPU 上统计，并取测试集全部图像的平均编码时间和解码时间。

## 3.2 对比实验

对比算法包括传统编码标准 JPEG[1]、JPEG2000[2]、BPG[23]、VVC(4:4:4)[3]和深度学习编码算法（Entroformer[16]，MLIC[19]，FTIC[21]，GMM[24]，GLLMM[25]，TCM[7]，HPCM[26]，DCAE[27]，LALIC[28]，SFTNet[29]等）。评价指标采用峰值信噪比（PSNR）、多尺度结构相似性（MS-SSIM），以及 BD-Rate（VVC(4:4:4)[3]为基准）。

### 3.2.1 客观对比结果

率失真性能对比结果如图 5 所示。在 Kodak 数据集上，FLIC 算法在 PSNR 与 MS-SSIM 指标上均展现出明显优势，整体性能优于最新的传统编码标准 VVC(4:4:4)[3]。与 FTIC[21]和 SFTNet[29]等频域感知编码算法相比，FLIC 算法在多数码率区间表现出更好的率失真性能。与 DCAE[27]、LALIC[28]和 HPCM[26]等基于学习的图像编码算法相比，FLIC 算法在低码率区间的率失真性能相对较低，但随着码率提高，其相对性能逐渐提升。

各算法在 Kodak 数据集上的 BD-Rate 指标与模型复杂度对比结果如表 1 所示。FLIC 算法的参数量为 12.51 M，在所有对比算法中最小，具有明显的轻量化优势。率失真性能方面，FLIC 算法相较 VVC(4:4:4)[3]节省 8.33%的码率。与频域感知编码算法 SFTNet[29]相比，FLIC 算法在取得更低 BD-Rate 的同时，参数量和 FLOPs 也进一步降低。另一方面，DCAE[27]、LALIC[28]和 HPCM[26]算法的整体 BD-Rate 优于 FLIC 算法，但在保证编解码时间相当的前提下，其模型规模和计算复杂度也明显更高。因此，FLIC 算法可以在压缩性能与模型复杂度之间取得平衡。

为进一步分析 FLIC 算法与各算法之间的差异，本文列出了 Kodak 数据集上的分段 BD-Rate 对比结果，如表 2 所示。可以看出，DCAE[27]和 LALIC[28]等基于学习的编码算法性能优势主要集中在低码率区间。随着码率提高，FLIC 算法的相对性能逐渐提升，其中在中、高码率区间的 BD-Rate 较 DCAE[27]算法分别低 0.53%和 0.88%，较 LALIC[28]算法分别低 1.33%和 1.71%。对于同样引入频域信息的 SFTNet[29]算法，FLIC 算法在全码率区间的 BD-Rate 由−6.55%降低至−8.33%，并在中、高码率区间分别进一步降低 4.80%和 7.33%。这些结果验证了所提 FLIC 算法在中、高码率条件下的率失真性能优势。

在 CLIC 测试集上，FLIC 算法同样在中高码率区间取得了较好的率失真性能。进一步的定量结果如表 3 所示，以 VVC (4:4:4) [3]为基准，FLIC 算法的 BD-PSNR 为 0.604 dB，BD-Rate 为-13.49%，即在相同重建质量下平均节省 13.49%的码率。整体率失真性能优于 Entroformer[16]、GLLMM[25]和 SFTNet[29]等算法，但低于算法 DCAE[27]、LALIC[28]和 HPCM[26]。FLIC 算法仅包含 12.51 M 参数，明显低于算法 DCAE[27]和 LALIC[28]，表明本文算法在高分辨率图像上仍能够以较低模型复杂度获得具有竞争力的编码性能。

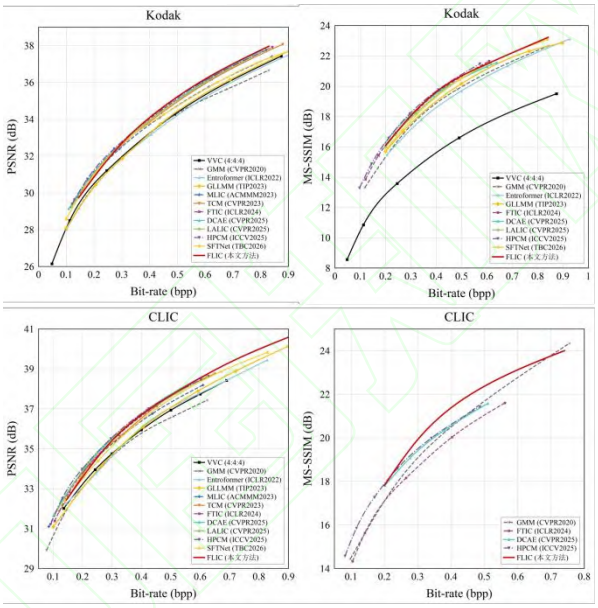



图 5 率失真性能对比结果

Fig.5 The RD performance on Kodak and CLIC dataset

表 1 Kodak 数据集上各深度学习编码算法的模型复杂度与 BD-Rate 对比结果

Table 1 Comparison results of model complexity and BD-Rate for learned image compression algorithms on Kodak dataset

| 对比算法 | 参数量/M | FLOPs/G | 编码时间/ms | 解码时间/ms | BD-PSNR/dB | BD-Rate/% |
|---|---|---|---|---|---|---|
| GMM | 13.18 | 30 | 3708 | 8658 | -0.21 | 4.94 |
| GLLMM | 15.71 | -- | 385260 | 387620 | -0.07 | 2.55 |
| Entroformer | 187.69 | -- | 4768 | 85919 | -0.19 | 4.70 |
| MLIC | 83.27 | 443 | 190 | 268 | 0.41 | -7.16 |
| TCM | 75.9 | 415 | 154 | 107 | 0.34 | -6.54 |
| FTIC | 45.9 | -- | 223 | 96 | 0.71 | -12.85 |
| HPCM | 68.50 | 88 | 143 | 191 | 0.77 | -14.30 |
| DCAE | 119.4 | 273 | 133 | 146 | 0.70 | -13.57 |
| LALIC | 63.24 | 286 | 274 | 150 | 0.63 | -12.16 |
| SFTNet | 45.78 | 208 | 116 | 91 | 0.31 | -6.55 |
| FLIC（本文） | 12.51 | 61 | 153 | 191 | 0.56 | -8.33 |

表 2 Kodak 数据集上各深度学习编码算法的分段BD-Rate 对比结果

Table 2　Comparison of BD-Rate Results Across Different Bitrate Ranges for Learned Image Compression Algorithms on the Kodak Dataset

| 对比算法 | 全区间 BD-Rate/% | 低码率 BD-Rate/% | 中码率 BD-Rate/% | 高码率 BD-Rate/% |
|---|---|---|---|---|
| GMM | 4.94 | 3.48 | 2.81 | -0.18 |
| GLLMM | 2.55 | 2.78 | -1.09 | -5.58 |
| Entroformer | 4.70 | 3.90 | 5.84 | 4.28 |
| MLIC | -7.16 | -12.94 | -8.73 | -4.50 |
| TCM | -6.54 | -1.17 | -3.38 | -6.05 |
| FTIC | -12.85 | -15.29 | -14.21 | -11.68 |
| HPCM | -14.30 | -16.95 | -15.60 | -10.77 |
| DCAE | -13.57 | -22.30 | -8.80 | -9.31 |
| LALIC | -12.16 | -19.85 | -8.00 | -8.48 |
| SFTNet | -6.55 | -12.26 | -4.53 | -2.86 |
| FLIC（本文） | -8.33 | -5.30 | -9.33 | -10.19 |

表 3 CLIC 数据集上各深度学习编码算法的 BD-Rate 对比结果

Table 3　Comparison results of BD-Rate for learned image compression algorithms on CLIC dataset

| 对比算法 | BD-PSNR/dB | BD-Rate/% |
|---|---|---|
| GMM | -0.25 | 6.55 |
| GLLMM | -0.02 | 0.16 |
| Entroformer | -0.04 | 0.96 |
| MLIC | 0.48 | -11.78 |
| TCM | 0.51 | -11.80 |
| FTIC | 0.59 | -13.62 |
| HPCM | 0.76 | -17.54 |
| DCAE | 0.70 | -16.17 |
| LALIC | 0.70 | -16.20 |
| SFTNet | 0.40 | -9.44 |
| FLIC（本文） | 0.604 | -13.49 |

### 3.2.2 主观对比结果

为了评估 FLIC 算法的主观重建性能，本文选取传统编码标准（JPEG[1]、JPEG2000[2]、BPG[23]和 VVC(4:4:4)[3]），以及 ELIC[18]、GMM[24]深度学习算法作为对比对象，在近似码率条件下对比各算法的图像细节重建效果，并对边缘纹理丰富的区域进行放大展示，如图 6 所示。本次实验选取 Kodak 数据集中包含天空低频平滑区域、建筑边缘、岩石海浪高频纹理区域的图像。图中每个算法下方括号内数值分别表示 bpp / PSNR (dB) / MS-SSIM。

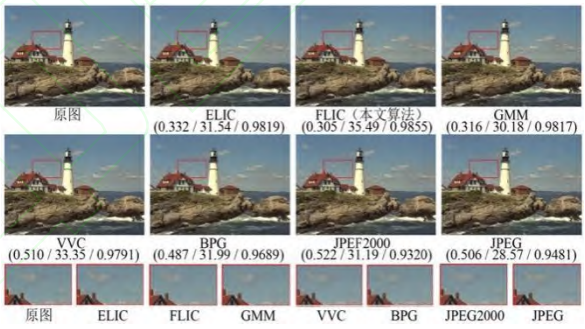



图 6 不同算法的重建图像视觉质量对比结果

Fig.6 Comparison of visual quality of reconstructed images of different algorithms

可以看出，JPEG[1] 重建图像出现了严重的块效应与色彩失真，建筑边缘与云朵纹理完全模糊。JPEG2000[2] 与 BPG[23] 虽缓解了块效应问题，但仍存在细节丢失与振铃伪影，高频纹理区域出现明显的模糊虚化。VVC(4:4:4)[3]也无法完全还原云朵的纹理层次，天空区域仍存在轻微的压缩噪点。GMM[24] 和 ELIC[18]对云朵纹理复原存在轻微模糊。FLIC 算法在云朵纹理、建筑边缘、岩石海浪等高频细节区域未出现模糊、锯齿现象。

### 3.3 消融实验

为验证本文提出的三个核心模块的有效性，本文在 Kodak 数据集上进行了消融实验，结果如表 4 所示。以基线模型为基准，逐个添加所提模块，测试 BD-Rate 指标。模型 1 为基线模型，采用固定

离散余弦变换，未引入本文提出的模块；模型 2 仅添加空频变换机制；模型 3 进一步引入频域感知上下文模块；FLIC 为本文完整模型，集成全部三个核心模块。

可以看出，相较于模型 1，模型 2 凭借可学习空频变换机制节省了 1.33% 的码率，验证了该机制相比固定变换的性能优势。这一增益主要得益于空频变换对频域先验与数据自适应能力的结合。其中，正交约束保证了变换过程的稳定性，DCT 初始化提供了较好的频域能量聚集特性。同时，可学习变换基能够通过端到端训练进一步适应图像内容分布，从而获得更加有效的频率表征。在此基础上，模型 3 进一步引入频率感知上下文模块，相比模型 2 进一步降低 1.68% 码率。该模块提取频谱统计特征，并将其与空间上下文进行融合，使熵参数预测能够同时利用空间邻域信息和频率分布信息，从而提高概率估计精度并减少编码冗余。进一步地，FLIC 算法加入带感知门控熵模型，相较模型 3 再次节省了 1.20% 的码率。该机制依据频谱特征对不同频带的分布参数进行自适应调制，降低不同频带采用统一建模方式带来的概率失配，从而进一步提升熵编码效率。综上，本文提出的空频变换机制、频域感知上下文模块与带感知门控熵模型三者协同作用，共节省了 4.21% 的码率，充分验证了各核心模块设计的合理性与必要性，在保证模型轻量化的同时，显著提升了率失真性能。

表 4 核心模块的消融实验结果

Table 4 Ablation study results on Kodak dataset

| 对比模型 | 空频变换机制 | 频域感知上下文 | 带感知门控熵模型 | BD-Rate/% |
|---|---|---|---|---|
| 模型 1 | DCT | | | 0.00 |
| 模型 2 | √ | | | -1.33 |
| 模型 3 | √ | √ | | -3.01 |
| FLIC | √ | √ | √ | -4.21 |

带感知门控熵模型中，频带分组数 k 值的选取十分关键，它决定了频带差异化建模的粒度。本文对比了 $k$ =2、4、8 时的模型性能，结果如表 5 所示。表中 BD-Rate(%) 以 VVC (4:4:4) 为基准计算。可以看出，当 $k$ =2 时，频带划分粒度过粗，无法充分捕捉不同频带的统计差异，性能最差。当 $k$ =4 时，模型性能达到最优，能够节省 8.33%的码率。当 $k$ =8 时，频带划分粒度过细，导致模型参数量显著增加，同时出现过拟合现象，性能反而下降。因此，本文选择 $k$ =4 作为最优频带分组数，可以在建模精度和计算复杂度之间取得良好平衡。

表 5 Kodak数据集上不同频带分组数的性能对比结果

Table 5 Performance comparison of different frequency band numbers on Kodak dataset

| 频带分组数 $k$ | BD-Rate/% | 参数量/M | 解码时间/ms |
|---|---|---|---|
| 2 | -7.12 | 12.38 | 187.2 |
| 4 | -8.33 | 12.51 | 191.0 |
| 8 | -7.89 | 12.87 | 198.5 |

## 4 结论

1)本文提出了一种频率感知上下文建模的图像编码算法 FLIC，以频域先验建模为核心思路，设计 DCT 初始化的空频变换机制，在保留频域能量聚集特性的同时，自适应匹配自然图像的统计分布。

2)算法构建频率感知上下文模块，将频谱统计特征与空间上下文残差融合，实现空频联合特征表征。本文算法的带感知门控熵模型完成对不同频带系数的参数调制。

3)实验结果表明，该文算法在 Kodak 测试集上相较 VVC(4:4:4) 节省了 8.33% 的码率，重建图像的高频纹理区域更加清晰。

4)未来，本文的频域感知建模机制可拓展至视频编码领域，利用帧间频谱相关性进一步提升压缩

效率。还可以结合生成式模型优化高频细节重建，进一步提升在低码率下的图像编码性能。

## 参考文献（References）


[1] Wallace G K.The JPEG still picture compression standard[J].Communications of the ACM,1991,34(4):30-44.

[2] Taubman D S,Marcellin M W.The JPEG2000 image compression fundamentals,standards and practice[J].Journal of Electronic Imaging,2002,11(2):286.

[3] JVET.Versatile video coding (VVC)reference software:VVC test model (VTM)[R].Geneva:Joint Video Experts Team,2019.

[4] Ballé J,Laparra V,Simoncelli E P.End-to-end optimized image compression[C]//Proceedings of the International Conference on Learning Representations.Toulon,France,2017.

[5] Johnston N,Vincent D,Minnen D,et al.Improved lossy image compression with priming and spatially adaptive bit rates for recurrent networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Salt Lake City,USA,2018:4385-4393.

[6] Li W G,Sun W Y,Zhao Y D,et al.Deep image compression with residual learning[J].Applied Sciences,2020,10(11):1-12.

[7] Liu J M,Sun H M,Katto J.Learned image compression with mixed Transformer-CNN architectures[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Vancouver,Canada,2023:14388-14397.

[8] Tu H Y,Wu S Q,Li L,et al.Multi-scale invertible neural network for wide-range variable-rate learned image compression[J].IEEE Transactions on Multimedia,2025,27:859-872.

[9] Chen Y N,Lyu Z Z,He B,et al.Content-aware Mamba for learned image compression[C]//Proceedings of the 14th International Conference on Learning Representations.Rio de Janeiro,Brazil,2026.

[10] Chen Y N,Lyu Z Z,He B,et al.Knowledge distillation for learned image compression[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision.Honolulu,USA,2025:4996-5006.

[11] Ehrlich M,Davis L.Deep residual learning in the JPEG transform domain[C]//Proceedings of the IEEE International Conference on Computer Vision.Seoul,Republic of Korea,2019:3484-3493.

[12] Guo L N,Shi X J,He D L,et al.Practical learned lossless JPEG recompression with multi-level cross-channel entropy model in the DCT domain[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.New Orleans,USA,2022:5862-5871.

[13] Ballé J,Minnen D,Singh S,et al.Variational image compression with a scale hyperprior[C]//Proceedings of the International Conference on Learning Representations.Vancouver,Canada,2018.

[14] Minnen D,Ballé J,Toderici G D.Joint autoregressive and hierarchical priors for learned image compression[C]//Proceedings of the Advances in Neural Information Processing Systems.Montreal,Canada,2018,31.

[15] He D L,Zheng Y Y,Sun B C,et al.Checkerboard context model for efficient learned image compression[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Virtual Event,USA,2021:14771-14780.

[16] Qian Y C,Sun X Y,Lin M,et al.Entroformer:a Transformer-based entropy model for learned image compression[C]//Proceedings of the International Conference on Learning Representations.Virtual Event,Online,2022.

[17] Zeng F H,Tang H,Shao Y H,et al.MambaIC:State space models for high-performance learned image compression[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.Nashville,USA,2025:18041-18050.

[18] He D L,Yang Z M,Peng W K,et al.ELIC:efficient learned image compression with unevenly grouped space-channel contextual adaptive coding[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.New Orleans,USA,2022:5718-5727.

[19] Jiang W,Yang J Y,Zhai Y Q,et al.MLIC:multi-reference entropy model for learned image compression[C]//Proceedings of the ACM International Conference on Multimedia.Ottawa,Canada,2023:7618-7627.

[20] Jiang W,Zhai Y Q,Yang J Y,et al.MLICv2:Enhanced multi-reference entropy modeling for learned image compression[J].ACM Transactions on Multimedia Computing,Communications,and Applications,2026,22(4):1-23.

[21] Li H,Li S H,Dai W R,et al.Frequency-aware Transformer for learned image compression[C]//Proceedings of the International Conference on Learning Representations.Vienna,Austria,2024:1-14.

[22] Deng J,Dong W,Socher R,et al.ImageNet:A large-scale hierarchical image database[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Miami,USA,2009:248-255.

[23] 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.

[24] Cheng Z X,Sun H M,Takeuchi M,et al.Learned image compression with discretized Gaussian mixture likelihoods and attention modules[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Seattle,USA,2020:7939-7948.

[25] Fu H S,Liang F,Lin J P,et al.Learned image compression with Gaussian-Laplacian-Logistic mixture model and concatenated residual modules[J].IEEE Transactions on Image Processing,2023,32:2063-2076.

[26] Li Y Q,Zhang H T,Li L,et al.Learned image compression with hierarchical progressive context modeling[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision.Honolulu,USA,2025:18834-18843.

[27] Lu J B,Zhang L H,Zhou X Y,et al.Learned image compression with dictionary-based entropy model[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Nashville,USA,2025:12850-12859.

[28] Feng D H,Cheng Z X,Wang S,et al.Linear attention modeling for learned image compression[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Nashville,USA,2025:7623-7632.

[29] Liu M Q,Li L X,Liang J M,et al.SFTNet:Spatial-Frequency Transformer Network for Learned Image Compression[J].IEEE Transactions on Broadcasting,2026,72(1):247-259.

[30] Li C,Yin S Z,Jia C M,et al.Multirate progressive entropy model for learned image compression[J].IEEE Transactions on Circuits and Systems for Video Technology,2024,34(8):7725-7741.