基于多维度高阶非线性变换的高效可学习图像信源编码方法

李超[1,2]，谭文[1]，孟凡阳[2]，柳伟[4]，梁永生[1,3]

（1.哈尔滨工业大学（深圳）信息科学与技术学院，广东 深圳 518060；2.鹏城国家实验室宽带通信部，广东 深圳 518055；3.深圳技术大学大数据与互联网学院，广东 深圳 5181182；4.深圳信息职业技术大学计算机与软件学院，广东 深圳 518172）

摘 要：为提升图像压缩效率以更好支撑带宽受限通信场景下的图像传输，提出一种基于多维度高阶非线性变换的可学习图像信源编码方法。首先，将现有变换方法中的线性与非线性算子串行堆叠结构，形式化为高阶非线性组合函数的一般表达，并从泰勒展开的视角分析了现有变换方法在特征表达能力方面的局限；其次，设计了由多路径组成的多维度高阶非线性变换模块。其中通道路径通过构建高维通道域并引入乘法融合操作，形成跨通道高阶特征项，空间路径结合方向卷积与乘法融合操作，以增强结构与纹理建模，从而提升空间维度的高阶表达能力；最后，将通道路径与空间路径的特征按类泰勒展开方式进行高阶组合，并将该模块嵌入可学习图像压缩框架中进行端到端训练。实验表明，所提方法在提升率失真性能的同时，保持了较低的参数量与计算量。

关键词：信源编码；可学习图像压缩；非线性变换；高阶函数；乘法融合

中图分类号：TP183

文献标志码：A

DOI：10.11959/j.issn.1000

An Efficient Learned Image Source Coding Method Based on Multidimensional High-Order Nonlinear Transform

LI Chao[1,2], TAN Wen[1], MENG Fanyang[2], LIU Wei, LIANG Yongsheng[1,3]

1. School of Information Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518060, China
2. Communication Department, Pengcheng Laboratory, Shenzhen 518055, China
3. College of Big Data and Internet, Shenzhen Technology University, Shenzhen 518118, China

Abstract: To enhance compression efficiency for image transmission in bandwidth-limited communication scenarios, a learned image source coding method based on multidimensional high-order nonlinear transformations is proposed. First, the serially stacked linear and nonlinear operators in learned image compression are formalized as a general representation of high-order nonlinear composite functions, and the limitations of existing transformation methods are analyzed from the perspective of Taylor expansion. Second, a multidimensional high-order nonlinear transformation module composed of multiple pathways is designed. In this module, the channel pathway constructs a high-dimensional channel domain and introduces multiplicative fusion to form cross-channel high-order feature terms, while the spatial pathway integrates directional convolutions with multiplicative fusion to enhance structure and texture modeling, thereby improving high-order spatial expressiveness. Finally, the features from the channel and spatial pathways are combined in a Taylor-like manner, and the module is incorporated into a learned image compression framework for end-to-end training. Experimental results demonstrate that the proposed method achieves notable improvements in rate–distortion perfor-

收稿日期：2026–01–05；修回日期：2026–04–03

通信作者：梁永生，liangys@hit.edu.cn

基金项目：国家自然科学基金资助项目（No.62571160, No.62031013）；广东省重点建设学科科研能力提升项目（No.2022ZDJS117）；广东省普通高校工程技术研究(开发)中心（No.2024GCZX004）

Foundation Items: The National Natural Science Foundation of China (No. 62571160, No.62031013), The Guangdong Province Key Construction Discipline Scientific Research Capacity Improvement Project (No.2022ZDJS117), Engineering Technology R&D Center of Guangdong Provincial Universities (No.2024GCZX004).

mance while maintaining lower parameter and computational costs.`

Keywords: source coding, learned image compression, nonlinear transform, higher-order function, multiplicative fusion

0 引言

随着移动互联网、物联网及智能终端的快速发展，图像等视觉数据在无线通信系统中的传输频率和数据量持续增长，已成为占用通信带宽和系统资源的主要信息类型之一[1-4]。尤其在带宽受限、时延敏感及边缘计算等典型通信场景中，作为图像通信链路的起始环节，高效的图像信源编码对于降低传输负载、提高通信效率具有重要意义[5]。图像信源编码（即图像压缩）的目标是在尽可能保证重建质量的前提下，最小化图像传输所需的比特数。传统图像压缩方法通常采用人工设计的模块化结构构建完整的编码流程，在率失真（rate-distortion, R-D）性能方面表现优异。然而，这种人工设计且各模块相互独立的特性，在一定程度上也限制了其编码性能的进一步提升。近年来，随着深度学习的快速发展，可学习图像压缩（learned image compression, LIC）依托神经网络强大的非线性表达能力与端到端优化机制，在R-D性能上已超越最先进的传统压缩方法，成为突破现有编码效率瓶颈的有效途径[5]。

当前大多数先进的可学习图像压缩方法通常遵循文献[7]提出的变分自编码器（variational autoencoder, VAE）框架，其整体结构由变换、量化和熵编码三个核心环节构成。变换网络用于在输入图像与其潜在特征表示之间建立双向映射。量化步骤将变换后的连续潜在特征表示离散化，以满足后续编码的需求。在熵编码阶段，量化后的潜在特征表示首先由熵模型进行概率分布建模与估计，以计算其熵下界；随后依据所估计的分布执行熵编码，将特征压缩为二进制码流，从而近似实现最优的压缩效率。近年来，学术界和工业界的相关研究也主要围绕上述三个核心环节展开，重点在于构建表达能力更强的变换网络[8-11]、设计更高效的量化策略[12-15]，以及建立更精确的熵模型[16-21]。在整个压缩流程中，变换模块作为率失真优化的起点，不仅决定潜在特征表示的紧凑性，还直接影响量化的稳定性和熵编码的建模精度[22-23]。因此，变换网络已成为提升R-D性能的关键环节，也是当前LIC领域持续关注的研究方向。

根据网络构建方式的不同，现有变换网络研究大致可分为整体结构设计和模块化结构设计两个层面。整体结构设计侧重于网络框架的宏观布局，例如通过对编码器与解码器进行对称[24]或非对称建模[8]，构建多尺度特征结构[9-10]，以及引入更先进的特征流动机制[25-27]来提升压缩性能。此类方法通过整体优化网络的表征层级，能够在更大感受野和更丰富的多尺度结构上建模图像统计特性，因此在全局与局部特征的联合表达方面具有明显优势。然而，此类方法通常需要对网络架构进行大规模重构，设计过程复杂，且在不同压缩框架间的迁移性和泛化性较弱。

相比之下，模块化结构设计更关注变换网络中非线性变换模块的细粒度建模与功能强化，例如通过引入注意力机制、仿射变换或增强的激活函数等方式提升特征表达能力。文献[28-29]延续了广义除法归一化（generalized divisive normalization, GDN）的设计理念，通过引入多样化的特征提取模块构建增强型 GDN 以提升压缩性能。文献[30-32]将注意力机制及其变体引入非线性变换模块中，选择性强化关键信息并抑制冗余特征，有效增强了变换网络的非线性表达能力。然而，无论是 GDN 系列还是基于注意力机制的非线性变换模块，其设计大多依赖于卷积神经网络（convolutional neural network, CNN），受限于局部感受野，难以有效建模全局上下文信息，从而限制了压缩性能的进一步提升。因此，研究人员从多个方向展开探索，主要通过设计 Transformer 分支[33]、Mamba 单元[34]以及融合结构[35-36]来增强非线性变换模块的全局建模能力。例如，文献[35]提出了一种双分支的 Transformer-CNN 混合模块，将CNN的局部建模能力与Transformer的全局感知能力结合，从而显著提升了率失真性能。此外，研究者还从空间域和频域两个层面提出了一系列空频融合的非线性变换方法[37-38]，通过整合不同域的表征，进一步提升了网络的去冗余能力与整体压缩效率。尽管现有的模块化结构设计方法在R-D性能上取得了持续进展，但绝大多数仍遵循线性变换和非线性激活的串行堆叠方式。从函数逼近的角度来看，此类结构的表达能力依赖于有

限深度下的非线性阶数与算子多样性，其对目标映射函数的建模通常受限于低阶非线性组合，难以充分刻画图像信号中复杂而高度相关的统计结构。此外，为弥补表达能力的不足，此类方法往往通过增加通道数或堆叠更多层来提升性能，导致参数规模与计算开销显著上升，因此，在压缩效率与模型复杂度之间的权衡方面仍有进一步的提升空间。

针对上述问题，本文从泰勒展开的角度出发，提出一种面向可学习图像压缩的多维度高阶非线性变换方法。相较于现有 LIC 方法及近期相关工作所遵循的线性变换与非线性激活串行堆叠的复杂非线性变换架构，本文方法的本质区别在于：采用轻量化多路径结构与乘法融合，显式生成不同维度的高阶项，在较低参数量与计算开销下，有效提升变换的非线性阶数与算子表达多样性，为压缩网络提供更具表达性且更紧凑的潜在特征表示，进而在 R-D 性能与模型复杂度之间取得更优平衡。

本文主要贡献如下。

1) 重新审视可学习图像压缩中的非线性变换的建模形式，将传统的线性-非线性堆叠结构重新表述为非线性组合函数建模问题，并基于泰勒展开分析了现有方法在特征表达方面的局限性。

2) 提出一种多维度高阶非线性变换方法，通过高维通道建模与乘法机制引入跨通道高阶依赖，并结合方向卷积与乘法机制生成空间高阶项，最终以类泰勒展开的方式融合不同维度的高阶项，从而显著增强变换网络的非线性表达能力。

3) 实验结果表明，所提出的非线性变换模块可即插即用于LIC方法中，在显著提升率失真性能的同时保持较低的参数规模与计算开销，从而在压缩性能与模型复杂度之间实现更优的权衡。

1　相关工作

1.1　可学习图像压缩

近年来，基于VAE的LIC方法在率失真优化方面取得了显著成果。其整体框架通常包括变换、量化与熵编码三个关键阶段。具体而言，对于给定的输入图像x，分析变换网络首先将其转换为紧凑的潜在表示y，随后将其量化为离散值$\hat{y}$。熵模型估计$\hat{y}$的概率分布，并执行熵编码，计算比特率。最后，合成变换网络将熵解码后的潜在特征表示映射为重建图像$\hat{x}$。该过程的优化目标是最小化率失真损失函数，可以表示为

$$\begin{aligned}\mathcal{L} &= R(\hat{y}) + \lambda \cdot D(x,\hat{x}) \\ &= \mathbb{E}_{x\sim p_x}\left[-\log_2 p_y(\hat{y}) \right] + \lambda \cdot \mathbb{E}_{x\sim p_x}\left[d(x,\hat{x}) \right]\end{aligned} \tag{1}$$

其中，R表示比特率，D表示输入图像与重建图像之间的失真度量，λ是用于平衡二者的超参数。一些研究者们致力于设计变换模块或编解码器，以增强变换网络非线性建模能力，生成紧凑的潜在表示y，并提升率失真性能。早期工作主要包括整体残差结构[39]、多尺度特征提取机制[9]以及注意力结构等。此外，研究者们也致力于设计更加合理或具有更好平滑性的量化策略，以优化y到$\hat{y}$的离散过程，从而有效缓解信息过度损失。文献[12]通过近似率失真损失函数，提出了一种新颖的量化策略，使压缩网络在量化阶段能够有效学习灵活的量化步长，从而实现了可扩展的率失真性能。文献[13]提出了一种基于解析率失真优化的通用渐进式 LIC 方法，在潜在特征中引入死区量化器以替代传统均匀量化，从而更高效地实现嵌入式量化与渐进式编码。此外，研究者们还探索了更先进的量化器设计方案，如自适应量化策略[14]和基于采样的可学习非均匀量化[15]，进一步提升了压缩性能与模型灵活性。在最后一个阶段，研究者们聚焦于设计更为先进的熵模型，以更准确地估计式 (1) 中$\hat{y}$的概率分布。这类模型通常融合多层次或多维度的上下文信息，以提升对潜在特征概率分布的估计精度，从而进一步改善压缩性能。文献[7]首次在 LIC 框架中引入基于特征的辅助信息，显著提升了熵模型的概率估计精度。随后，该思路被扩展至空间自回归模型[16-17]和通道自回归模型[18-20]，分别通过空间顺序建模与通道并行建模进一步增强了分布估计能力。文献[21]则结合棋盘式上下文与通道上下文，在提升率失真性能的同时有效降低了解码延迟。

1.2　LIC 中的非线性变换模块

在 LIC 的整体执行流程中，变换过程主要发生于编码的起始阶段和解码的末端阶段，其对紧凑特征的表达能力将直接影响后续的压缩效果。而这种表达能力在很大程度上取决于编解码器中非线性变换模块的建模能力。文献[7]首次在 LIC 框架中引入GDN，通过可学习的参数对特征图各通道的激活值进行自适应归一化，有效增强了变换网络的去冗余能力。在此基础上，后续研究进一步将 GDN 与其他结构相结合，以构建非线性表达能力

更强的变换模块。文献[28]提出了残差GDN结构，通过在残差块中堆叠更多GDN层以进一步提取紧凑的特征表示。文献[10]则采用多尺度残差块和可变形残差块，更有效地捕捉潜在特征的空间相关性，从而提升整体压缩性能。文献[46]提出了一种基于残差块的高效通道分组变换方法，该方法在空间维度上进行卷积核因式分解，并在通道维度上扩展网络宽度，使得模型具备更强的特征表示能力。在注意力机制方面，文献[20]提出了一种面向 LIC 的边缘注意力机制，通过多分支滤波器增强高频信息提取能力，从而改善率失真性能，并借助参数重参数化策略在推理阶段有效降低计算开销。文献[30]则在变换网络中引入基于 Transformer 的窗口式注意力机制，更有效地捕捉空间相邻区域之间的相关性。近来，文献[40]提出了一种基于能量集中的非线性变换。该方法通过渐进式的分组卷积堆叠，实现了特征能量的主动聚集，为熵编码提供先验知识更多的特征。文献[42]和文献[45]从特征调制角度出发，分别提出了基于幅度调制与基于特征调制的非线性变换方法，通过引入调制机制动态调整特征的空间分布与通道响应，有效增强了网络的非线性表达能力。文献[43]则提出了一种基于特征维度分离的非线性变换结构，利用轻量化卷积单元构建多分支变换路径以丰富特征表示，并通过加法融合各分支输出，进一步提升了率失真性能。

综上所述，现有大多数非线性变换方法仍采用线性变换与非线性激活函数的串行组合形式。从函数表达的角度来看，此类结构实质上等价于对目标映射函数的低阶泰勒近似，表达能力存在一定局限。此外，为提升非线性表达能力，通常需要增加网络层数，从而导致模型参数量和计算开销显著增加，进一步压缩了其在压缩效率与模型复杂度之间的权衡空间。

2 算法设计

2.1 问题分析

在LIC中，变换网络的核心目标是构建具有足够表达能力的非线性映射函数，以从输入图像中提取紧凑的潜在特征表示。近年的可解释性研究[47]表明，深度神经网络可视为一种通用非线性函数逼近器，其由卷积、激活函数及其他特征变换构成的组合形式，本质上可对应于泰勒展开中的不同阶非线性项。从函数逼近角度看，网络的参数化变换过程可视为对多维泰勒多项式各阶系数的可学习拟合，网络的深度与宽度在一定程度上决定了可逼近的非线性阶数。然而，现有LIC方法普遍采用线性变换（如卷积）与非线性激活串联的结构，其非线性表达能力仍局限于局部且阶数固定的变换形式。在建模具有复杂高阶依赖关系的映射时，该类结构通常依赖深层堆叠以提升有效非线性阶数，从而降低整体表达效率。为进一步解释这一现象，本文从泰勒展开的角度分析该类结构的表达局限性。

设$f^*(\boldsymbol{x})$为目标映射函数，表示从输入图像$\boldsymbol{x}\in\mathbb{R}^d$到潜在空间的非线性变换。根据泰勒展开理论，$f^*(\boldsymbol{x})$在基点$\boldsymbol{x}_0$可展开为

$$\begin{aligned} f^*(\boldsymbol{x}) &= f^*(\boldsymbol{x}_0) + \sum_i \frac{\partial f}{\partial x_i}(x_i - x_{0,i}) + \\ &\frac{1}{2}\sum_{i,j}\frac{\partial^2 f}{\partial x_i \partial x_j}(x_i - x_{0,i})(x_j - x_{0,j}) + \\ &\frac{1}{6}\sum_{i,j,k}\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(x_i - x_{0,i})(x_j - x_{0,j})(x_k - x_{0,k}) + \cdots \end{aligned} \tag{2}$$

其中，零阶项对应常数偏置，一阶项反映局部线性响应；二阶项描述输入变量之间的二阶交互关系；而三阶及以上项则刻画更复杂的高阶非线性耦合。从表达能力角度看，一个具有足够容量的神经网络应能够以较高效率逼近这些高阶项所代表的非线性结构。目前多数变换网络结构及其中的非线性变换模块通常依赖不同形式的卷积算子或其它特征提取单元，通过线性结构与非线性激活函数的级联来构建映射。其一般形式可写作

$$\begin{aligned} f(\boldsymbol{x}) = \sigma_n\Big(\boldsymbol{W}_n\sigma_{n-1}\Big(\boldsymbol{W}_{n-1}\cdots\sigma_1\big(\boldsymbol{W}_1\boldsymbol{x}+\boldsymbol{b}_1\big)+\boldsymbol{b}_2\Big)+ \\ \cdots+\boldsymbol{b}_n\Big) \end{aligned} \tag{3}$$

其中，$\boldsymbol{W}_i$为线性变换，σ_i为非线性激活函数。每一层的输出为

$$f_i(\boldsymbol{x}) = \sigma\big(\boldsymbol{W}_i f_{i-1}(\boldsymbol{x}) + \boldsymbol{b}_i\big) \tag{4}$$

从函数逼近的角度看，这类串行结构依赖逐层累积的方式提升非线性能力，单层仅能引入一阶及有限形式的二阶非线性。要刻画复杂映射中的高阶变量耦合关系，必须增加网络深度才能间接逼近，从而带来明显的参数与计算成本，对压缩任务的效率造成限制。近年来，为提升特征多样性，不少方法引入了并行分支结构，如 CNN-Transformer 并联或CNN-Mamba并联等，其整体流程的表达形式可

写为

$$f(\boldsymbol{x}) = f_1(\boldsymbol{x}) + f_2(\boldsymbol{x}) + \cdots + f_K(\boldsymbol{x}) \tag{5}$$

其中每个分支$f_K(\boldsymbol{x})$仍然是由线性变换与激活函数串联构成的, 例如，在TCM[35]模型中，CNN 分支通常采用残差块结构，而Transformer分支则采用类似式(3)所示的嵌套变换形式。这种结构通过并行路径提升特征多样性，但从函数展开角度来看，其整体表达仍然只是多个低阶近似项的线性叠加。设每个分支在输入点处的泰勒展开为

$$\begin{aligned} f_k(\boldsymbol{x}) &= f_k(\boldsymbol{x}_0) + J_k(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) \\ &+ \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}_0)^T H_k(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + \cdots \end{aligned} \tag{6}$$

其中$J_k(\boldsymbol{x}_0)$表示第k个分支在点$\boldsymbol{x}_0$处的雅可比矩阵，$H_k(\boldsymbol{x}_0)$表示第k个分支在点$\boldsymbol{x}_0$处的海森矩阵。则整体函数展开为

$$f(\boldsymbol{x}) = \sum_{k=1}^{K}\left[f_k(\boldsymbol{x}_0) + J_k(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0) + \cdots \right] \tag{7}$$

尽管并行分支能够提升特征多样性，其整体表达本质上仍是由各分支的低阶局部近似线性组合而成，高阶耦合主要来自分支内部，无法在跨分支层面显式构造多变量间的乘性交互。这种缺乏显式交叉项建模的结构，在处理真实映射中普遍存在的二阶及更高阶非线性关系时表达能力受限，只能依赖增加深度或宽度来间接累积高阶特征。因此，在保持结构简洁与计算负担尽可能小的前提下，引入能够直接刻画高阶耦合的建模机制，已成为进一步提升非线性变换模块表达力的关键。

2.2　多维度高阶非线性变换

针对大多数现有非线性变换结构难以显式刻画高阶耦合的问题，本文提出了一种多维度高阶非线性变换（multidimensional high-order nonlinear transform, MHNT）。该模块在轻量化结构中引入显式的高阶特征交互，并在特征、通道与空间等多个维度构建跨变量的高阶耦合项，从而减少对传统串行或并行网络低阶叠加机制的依赖。通过显式建模交叉项，本方法能够更有效地捕获真实映射中广泛存在的二阶及更高阶非线性关系，实现更高效且更具表达力的非线性变换。

本文提出的MHNT的结构如图1右上所示，给定一个中间特征表示为$\boldsymbol{X} \in \mathbb{R}^{C\times H\times W}$，其中$C$表示通道数，$H$和$W$分别为特征图的高度与宽度。该特征分别输入到空间高阶项路径、通道高阶项路径以及原始特征路径中。具体而言，空间路径用于捕捉局部邻域中像素的空间交叉关系。该路径通过两个方向的深度可分离卷积（depthwise convolution, DW）分别提取水平与垂直特征，并采用逐元素乘法进行显式交互，从而形成空间高阶响应，该过程可以表示为

$$\mathcal{F}_{spatial}(\boldsymbol{X}) = \boldsymbol{Conv}^{\boldsymbol{DW}}_{1\times 9}(\boldsymbol{X}) \odot \boldsymbol{Conv}^{\boldsymbol{DW}}_{9\times 1}(\boldsymbol{X}) \tag{8}$$

其中$\boldsymbol{Conv}^{\boldsymbol{DW}}_{1\times 9}$和$\boldsymbol{Conv}^{\boldsymbol{DW}}_{9\times 1}$分别定义为方向分离的卷积，分别提取的是水平方向$g_x$和垂直方向$g_y$的特征，$\odot$表示逐元素乘法，融合两个方向的空间响应。该设计通过方向解耦的方式提升了空间高阶路径的特征多样性，并能够更有效地提取图像中具有方向性结构与纹理模式的特征表示。二者输出通过逐像素乘法操作实现交叉融合，乘法操作具有将特征投影到极高维隐式特征空间的能力[44]，类似于多项式核函数，可以进一步构造了空间维度的高阶项，有助于建模不同方向特征之间的非线性交互，从而增强模型对空间结构的表达能力。

从混合偏导形式来看，在连续域中，图像函数$f(x,y)$的方向导数组合可以表示

$$\begin{aligned} \frac{\partial^2 f}{\partial x \partial y} &\approx f(x+1,y+1) - f(x+1,y) \\ &- f(x,y+1) + f(x,y) \end{aligned} \tag{9}$$

该项本质上是图像在x和y两个方向上的混合二阶导的有限差分近似。在离散空间中，空间路径中构造的交叉项$g_x \cdot g_y$实际上可以近似为

$$\left(\sum_{i=-4}^{4} \boldsymbol{w}_i x_{h,w+i}\right) \cdot \left(\sum_{j=-4}^{4} \boldsymbol{v}_j x_{h+j,w}\right) = \sum_{i,j} \boldsymbol{w}_i \boldsymbol{v}_j x_{h,w+i} x_{h+j,w} \tag{10}$$

其中$\boldsymbol{w}_i$和$\boldsymbol{v}_j$分别表示水平方向卷积核和垂直方向卷积核的权值。可以看出，空间路径构造的是空间方向上的双变量组合项，即$x_{h+i,w} \cdot x_{h,w+j}$可视为二维高阶导数项的离散逼近。此外，空间路径中所有DW卷积均在每个通道内独立操作，与通道路径进行显式解耦，这种解耦有助于提升整体特征建模的多样性与表达能力。

通道高阶路径侧重于建模不同通道之间的高阶组合关系。该路径首先通过两个并行的1×1卷积支路对输入特征进行通道升维，从而将原始低维特征映射到更高维的通道空间。具体而言，原始输入特征属于一个较低维度的空间$\mathbb{R}^C$，升维后被映射至

更高维的空间 $\mathbb{R}^{K\cdot C}$，在这个更大的特征空间中，通道间特征的可分性更强，能够表达更复杂的组合模式。为进一步增强通道高维空间的非线性建模能力，区别于空间分支，在1×1卷积之后引入Leaky-ReLU 激活函数。在完成升维与非线性激活之后，通道路径中得到的两个中间表示 $\boldsymbol{F}_1,\boldsymbol{F}_2\in\mathbb{R}^{K\cdot C\times H\times W}(K=3)$ 通过逐元素乘法进行交互融合，即

$$\mathcal{F}_{channel}(\boldsymbol{x})=\boldsymbol{Conv}_{down}^{1\times 1}\left(\phi(U_1\boldsymbol{x})\odot\psi(U_2\boldsymbol{x})\right) \tag{11}$$

其中 $U_1,U_2\in\mathbb{R}^{C'\times C}$ 分别表示两个升维卷积核；ϕ,ψ 为

LeakyReLU 激活函数；$\odot$ 表示逐元素乘法，$\boldsymbol{Conv}_{down}^{1\times 1}$ 为降维卷积操作。该操作实现了通道维度上的显式非线性交互建模。与传统的加法或拼接方式不同，逐元素乘法能够引入特征之间的乘积项耦合关系，等价于构造通道维度上的高阶组合特征，使得模型能够捕捉更复杂的通道依赖结构。具体而言，取某空间位置处的特征向量 $x_{hw}\in\mathbb{R}^C$，则第 i 个升维通道的两个支路输出分别为

$$z_i^{(1)}=\phi\left(\sum_j U_1^{(i,j)}x_j\right),\ z_i^{(2)}=\psi\left(\sum_k U_2^{(i,k)}x_k\right) \tag{12}$$

其乘积输出为

$$y_i=z_i^{(1)}\cdot z_i^{(2)}=\phi\left(\sum_j U_1^{(i,j)}x_j\right)\cdot\psi\left(\sum_k U_2^{(i,k)}x_k\right) \tag{13}$$

考虑 LeakyReLU 的分段非线性特性，设

$$z_1=\phi(a^{\top}\boldsymbol{x}),\ z_2=\psi(b^{\top}\boldsymbol{x}) \tag{14}$$

则其乘积输出为

$$y=z_1\cdot z_2=\phi(a^{\top}\boldsymbol{x})\cdot\psi(b^{\top}\boldsymbol{x}) \tag{15}$$

由于 ϕ,ψ 为分段线性函数，可近似为分段多项式展开

$$y\approx\sum_{j,k}a_jb_kx_jx_k+\sum_{j,k,l}\gamma_{jkl}x_jx_kx_l+\cdots \tag{16}$$

其中 $x_jx_kx_l$ 表示输入 $\boldsymbol{x}$ 中的第 j,k,l 个通道分量，并以此类推，a_j，b_k 分别是升维卷积核 a,b 的第 j 和第 k 个权重；γ_{jkl} 由非线性激活函数的高阶展开所引入的三阶组合系数。因此，通道路径通过升维映射、非线性激活与乘法组合，构造了多变量的高阶组合项；数学上等价于通道维度上的分段多项式函数族。随后，通过一个1×1卷积完成降维操作，将通道数恢复至原始尺度，从而对高阶特征表示进行压缩并增强特征选择性。最后，将两种高阶项与原始特征路径进行融合，整体结构可表示为

$$f(\boldsymbol{x})=\mathcal{F}_{channel}(\boldsymbol{x})+\mathcal{F}_{spatial}(\boldsymbol{x})+\boldsymbol{x} \tag{17}$$

其中 $\mathcal{F}_{channel}(\boldsymbol{x})$ 表示通道路径构造的通道高阶非线性项；$\mathcal{F}_{spatial}(\boldsymbol{x})$ 表示空间路径构造的空间高阶组合项；$\boldsymbol{x}$ 为原始输入残差项，提供恒等映射。该结构等价于一个结构化近似的多阶泰勒展开形式，即

$$\begin{aligned}f(\boldsymbol{x})\approx f(\boldsymbol{x}_0)+\sum_i\frac{\partial f}{\partial x_i}(x_i-x_{0,i})+\\ \sum_{i,j}\frac{\partial^2 f}{\partial x_i\partial x_j}(x_i-x_{0,i})(x_j-x_{0,j})+\cdots\end{aligned} \tag{18}$$

其中 $\boldsymbol{x}_0$ 可视为中心点（例如恒等映射），而 $\mathcal{F}_{spatial}(\boldsymbol{x}),\mathcal{F}_{channel}(\boldsymbol{x})$ 分别提供了显式的二阶及部

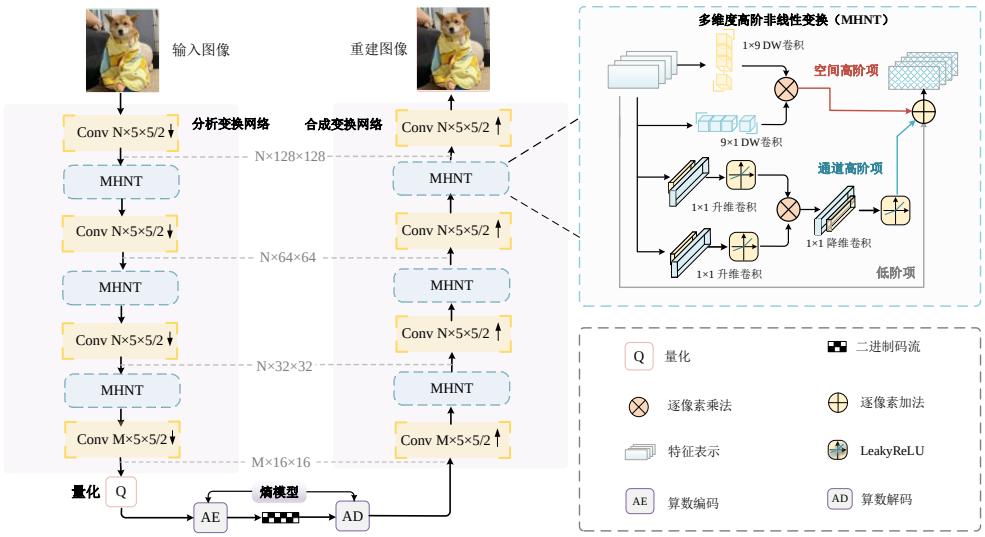

图1　LIC 整体框架及所提出的多维度高阶非线性变换模块的示意图

分高阶项的近似表示。现有的非线性变换方法需要通过层层堆叠逐步逼近这些高阶项；相比之下，本方法通过结构化设计在浅层即显式引入多个高阶项的近似表达，这些项在结构上与多维泰勒展开中的一阶、二阶及更高阶项对应，其表达形式可视为对泰勒展开项的结构化近似。因此，MHNT 的构造原则与泰勒展开的多阶非线性项具有理论一致性，能够为网络提供显式的高阶非线性建模能力。此外，通道与空间路径的解耦设计使模型能够分别捕捉通道维度中的跨通道非线性关系与空间维度中的方向性交互结构，形成互补的表征机制。需要说明的是，由于 MHNT 显式构建的高阶项数量有限，其数学形式可视为多维泰勒展开在第 K 阶处的有限项近似，可以表示为

$$f(x)=\sum_{k=0}^{K}\frac{1}{k!}\nabla^{k}f(x_0)(x-x_0)^{k}+R_{K+1}(x) \quad (19)$$

由于仅构建到第 K 阶，高阶项被截断后会引入理论上的截断误差 $R_{K+1}(x)$，可以表示为

$$R_{K+1}(x)=\frac{1}{(K+1)!}\nabla^{K+1}f(\xi)(x-x_0)^{K+1} \quad (20)$$

其中 ξ 位于 x 与 x_0 之间。尽管理论上存在截断误差，但深度模型在训练过程中通常能够通过迭代优化学习到对目标映射的有效近似。此外，关于神经网络内部的可解释性研究仍未形成统一共识，因此本文不对其是否显式补偿截断误差作进一步推断。基于此，本文在理论部分仅给出截断误差的数学形式，以明确有限项高阶结构的构造条件；后续章节的实验结果表明，所采用的有限阶结构已能够有效刻画图像信号中的主要非线性特征，因而在工程实践中是充分且有效的。更为深入的理论分析与完备推导将留待后续工作进一步研究。

2.3　实例化

在明确多维度高阶非线性变换的基本结构后，本节进一步给出其在 LIC 框架下的具体实例化过程，从而为后续实验与性能评估提供完整实现。

实例化的整体结构如图 1 所示，由分析变换网络、量化与熵模型、以及合成变换网络三部分组成。在编码端，输入图像首先经过分析变换网络处理。该网络由三组 5×5、步长为 2 的下采样卷积层与 MHNT 模块交替堆叠而成，通过逐层降低空间分辨率并提取更紧凑的特征。MHNT 引入显式的高阶特征交互，在通道和空间建立分别高阶项，有效增强了网络的非线性建模能力和特征表达能力。经过三次卷积与 MHNT 处理后，图像分辨率依次降至 N×128×128、N×64×64、N×32×32，并最终得到尺寸为 M×16×16 的潜在特征表示。其中，N 表示变换网络的通道数，M 则对应熵模型的通道数。具体而言，本文遵循文献[19][21]中的参数设置，将 N:M 设置为 192:320。随后，模型采用可微量化近似完成量化过程，并利用熵模型对潜在表示的概率分布进行估计，从而实现精确的码率估计。在解码端，量化后的潜在表示经熵解码恢复后进入合成变换网络。该网络与编码端结构对称，由三组上采样卷积和 MHNT 模块构成。得益于 MHNT 的强非线性建模能力，解码过程能够在重建阶段充分挖掘纹理与细粒度结构信息，从而弥补编码端下采样导致的信息损失，提升重建图像的细节保真度。

3　实验分析

3.1　实验配置

为验证所提方法的有效性，本文在统一实验设置下完成模型的训练与测试。训练阶段使用可学习图像压缩挑战赛（challenge on learned image compression, CLIC）的官方训练集，并随机裁剪得到 247,576 张尺寸为 3×256×256 的训练样本。测试阶段选用 KodaK 和 Tecnick 两个标准数据集：其中 Kodak 包含 24 张 3×512×768/768×512 的自然图像，Tecnick 包含 100 张分辨率为 3×1200×1200 的自然图像。

本文遵循 LIC 领域的通用评测设置[21]，聚焦信源编码的方法设计与率失真性能评价，因此实验中未纳入无线信道与传输链路建模。在训练策略方面，所有模型均基于 CompressAI 框架实现，训练轮数设为 100，batch size 为 16。初始学习率为 1×10^{-4}，并在第 40 和第 80 个 epoch 分别衰减至原来的 0.5 和 0.2。按照 LIC 的常规设置，本文针对均方误差（mean squared error, MSE）和多尺度结构相似性（multi-scale structural similarity, MS-SSIM）两类失真指标分别训练多组模型。其中，MSE 的 λ 取值为 {0.0018, 0.0035, 0.0067, 0.0130, 0.0300, 0.0483}，MS-SSIM 的 λ 取值为 {2.4, 4.58, 8.73, 16.64, 31.73, 60.50}。性能评价方面，综合采用每像素比特率（bits per pixel, bpp）、峰值信噪比（peak signal-to-noise ratio, PSNR）和 MS-SSIM 进

行对比分析。此外，图像的压缩倍数可通过 bpp 值换算得出，计算公式为原图每像素位数与 bpp 的比值，其中原图位数通常取 24 bit（即 8 bit 每通道 × 3 通道）。

3.2　率失真性能分析

为验证所提方法的有效性，本文以 He2022[21] 作为基线网络，将提出的非线性变换嵌入其中构建“本文方法”。随后在 Kodak 与 Tecnick 两个经典图像数据集上开展 R-D 实验，并分别从 PSNR 与 MS-SSIM 两个指标对结果进行对比分析。实验方法涵盖传统编解码器、代表性和近期的 LIC 方法（包括 Ballé2018[7]、Minnen2018[14]、Cheng2020[39]、Chen2021[11]、Minnen2020[19]、Hu2020[41]、Hu2021[24]、He2022[21]、Zuo2022[30]、Tang2022[8]、Liu2023-S[35]、Fu2023[18]、Li2024[20]、Bao2025[37]、Tan2025[42]）。

鉴于 He2022 的非线性变换由三层连续堆叠的 ResBlock 及若干注意力模块构成，为保证比较的公平性将单层非线性变换中的 MHNT 数量固定为 2。基于其他熵模型的 R-D 性能以及更全面的对比结果将于表 1 和图 4 中进一步呈现。

图 2（a）与图 2（b）分别展示了 Kodak 数据集上的 PSNR 与 MS-SSIM 随 bpp 变化的曲线；图 2（c）与图 2（d）则给出了 Tecnick 数据集上的对应结果。如图 2（a）所示，在高码率区间，本文方法相较于所有对比方法取得了最优的 R-D 性能；随着比特率降低，虽然性能提升幅度略有下降，但整体仍保持与先进方法相竞争的水平。从其他图中也可以观察到类似的现象。原因可能在于：在低码率区域，可用比特极少，主要用于表示基础结构信息，复杂纹理与高频细节难以有效编码，从而限制了具备高阶表达能力的 MHNT 的发挥。同时，He2022 的熵模型在空间与通道维度上具有较强的建模能力，已对特征分布进行了充分压缩，使前端

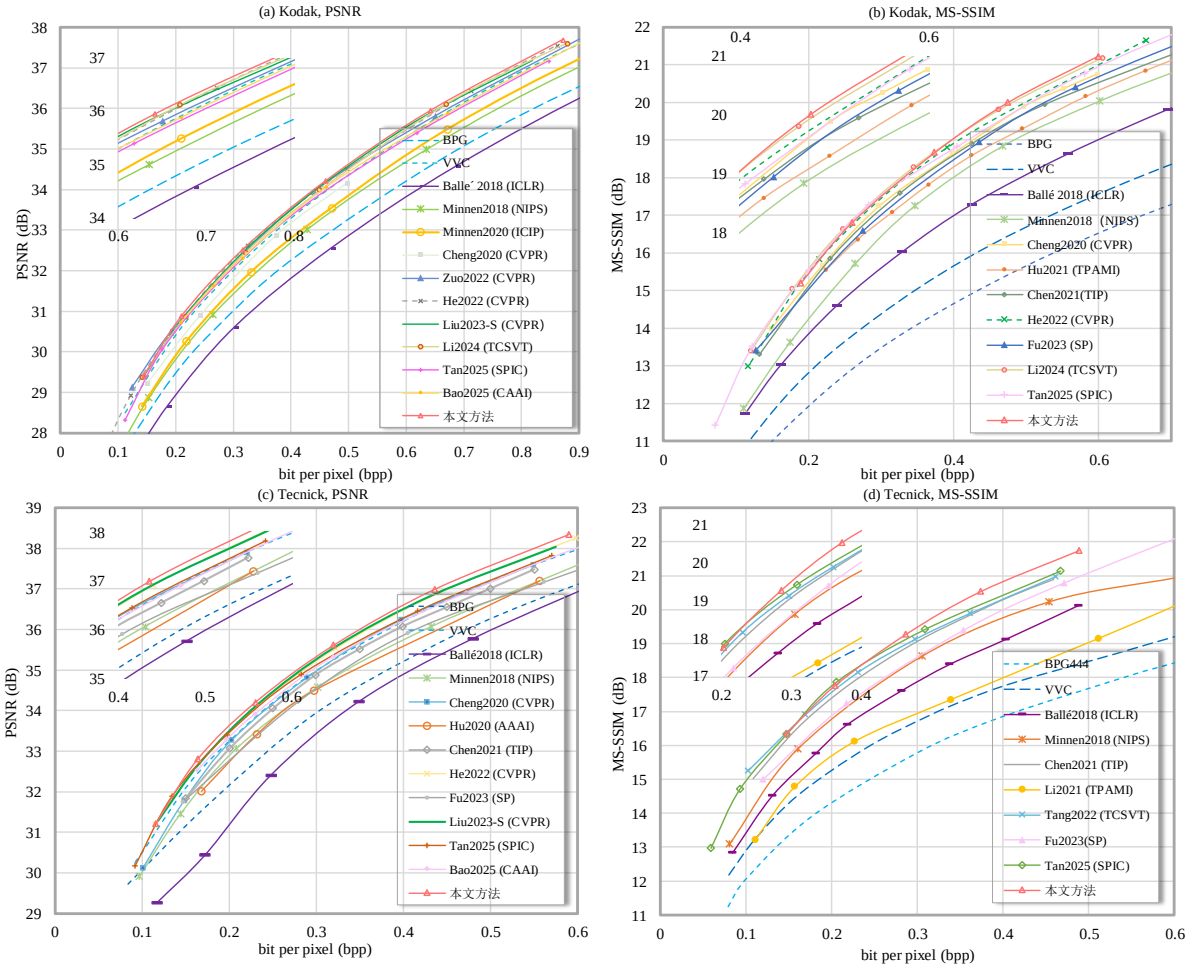

图 2　Kodak 和 Tecnick 数据集上的 R-D 曲线

变换的提升空间较小，因而整体增益相对有限。但随着码率提升，比特数约束减弱，MHNT 通过空间高阶项与通道高阶项的双向融合结构能够更充分地发挥其高阶非线性表达能力，对复杂纹理和细节结构的建模更加精准，因此在高码率区域表现出更显著的性能优势。

在压缩倍数方面，以 Tecnick 数据集中典型分辨率 1200×1200×3 (RGB, 8 bit/通道) 的图像为例，根据图 2(c)中的 bpp 数据换算可知：在最高码率点设置下，Liu2023-S 相较于原图实现了约 41.8 倍的压缩，而本文方法的压缩倍数为 40.6 倍。需要说明的是，在 MSE 优化的模型中，λ= 0.0018 的低码率设置下，本文方法与 He2022 复杂熵模型联合训练时训练稳定性下降，具体表现为梯度振荡加剧、率失真损失在迭代过程中出现显著波动。在 100 代的训练迭代中，上述不稳定性会逐步累积，并在部分迭代中多次触发 NaN，导致模型无法稳定收敛。造成该现象的主要原因在于：低码率约束使潜变量分布被进一步压缩并趋于尖峰或稀疏，从而显著增加复杂熵模型的先验拟合难度，使码率项（负对数似然）的梯度更易出现放大与不稳定；同时，MHNT 采用多分支乘法融合以显式建模空间和通道高阶项，乘法单元的交叉梯度会放大分支激活不均衡带来的梯度波动，进而引发数值溢出并导致训练

表 1　　LIC 非线性变换的高效性指标比较

熵模型	方法	参数量	BD-Rate	计算量	总编码时间	总解码时间	T 编码时间	E 编码时间
Hyper[7]	ResBlock[39]	663.936 K	−9.327%	10.872 G	89.0 ms	126.7 ms	20.9 ms	68.1 ms
	WSA[30]	906.342 K	−8.822%	14.823 G	116.9 ms	127.4 ms	49.1 ms	67.8 ms
	EASN[29]	1624.06 K	−13.483%	25.971 G	106.7 ms	146.8 ms	40.7 ms	66.0 ms
	IQ[45]	536.640 K	−12.672%	8.758 G	94.5 ms	132.3 ms	26.4 ms	68.1 ms
	SMCCT[42]	738.048 K	−12.93%	12.08 G	93.0 ms	128.7 ms	23.7 ms	69.3 ms
	本文方法	336.960 K	-14.065%	5.492 G	88. 5 ms	125.7 ms	20.2 ms	68.3 ms
MPEM[20]	ResBlock[39]	663.936 K	-18.810%	10.872 G	79.5 ms	132.1 ms	26.1 ms	53.4 ms
	本文方法	336.960 K	-21.061%	5.492 G	80.6 ms	132.7 ms	25.6 ms	55.0 ms
Informer[17]	ResBlock[39]	663.936 K	-24.393%	10.872 G	3508.2 ms	8213.1 ms	21.2 ms	3487.0 ms
	本文方法	336.960 K	-26.276%	5.492 G	3432.3 ms	8195.9 ms	20.4 ms	3411.9 ms

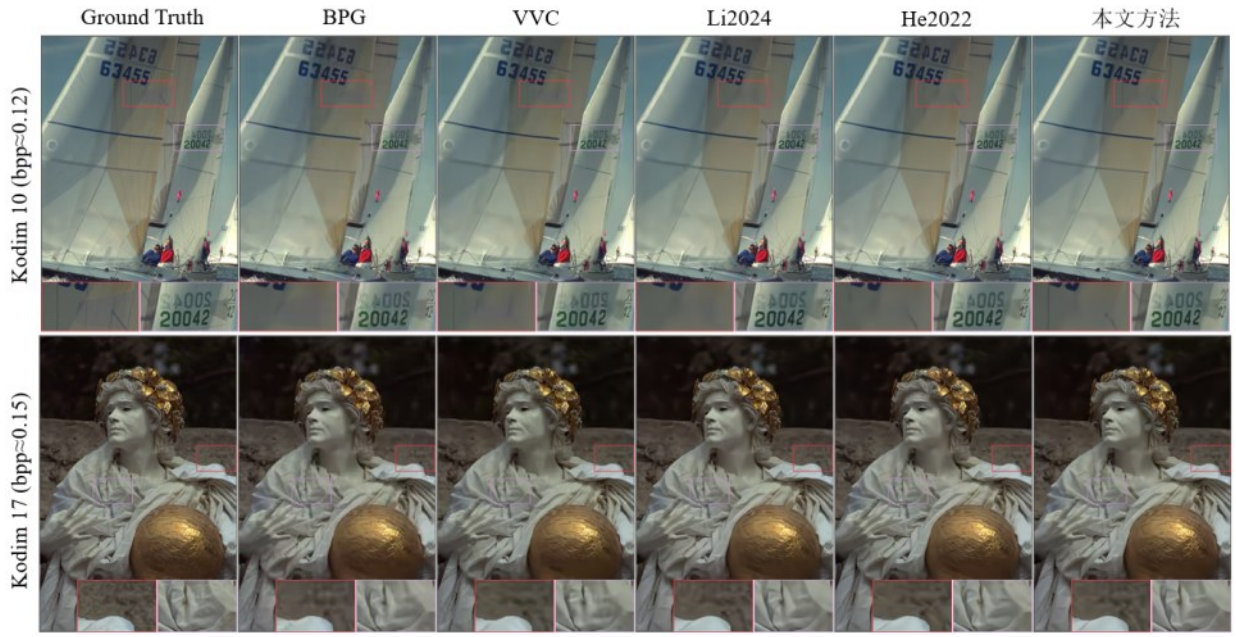

图 3　重建图像的主观质量比较

失败。

为了获得稳定且可复现的低码率结果，最低码率点采用了 LIC 中常见的相邻码率点迁移微调策略[35]：以 λ= 0.0035 的收敛模型为初始化，对 λ= 0.0018 进行 20 代微调。该策略虽较独立训练的最优结果略有性能折损，但由于微调迭代次数较少，可显著提升低码率点的收敛稳定性。因此，图 2(a) 和图 2(c) 中本文方法最低码率点的 PSNR 略低于部分对比方法。在 MS-SSIM 方面，尽管采用相同的相邻码率点微调策略（由 λ=4.58 微调 λ=2.0），训练过程中仍会触发 NaN。该现象主要源于失真项数值性质的差异：以 PSNR 为目标时通常采用 MSE 作为逐像素失真项，梯度形式简单且幅值受控，微调阶段参数更新更平滑稳定；相比之下，MS-SSIM 对局部均值、方差、协方差等统计量高度敏感，极低码率下重建图像易出现局部对比度降低、方差接近零或近似常量块增多等退化，使其进入数值敏感区间并导致梯度放大与数值不稳定。基于实验稳定性考虑，本文在图 2(b) 和图 2(c) 中未报告 λ=2.0 对应的 MS-SSIM 结果，更稳健的损失形式与训练策略也是本文后续工作的研究目标。

从图 2(b) 和 (d) 的 PSNR-bpp 与 MS-SSIM-bpp 对比结果可以看出，所提方法在大多数码率范围内均优于对比方法。上述结果表明，本文方法在 R-D 性能上取得了稳定且显著的提升，进一步证明了其有效性。此外，若将本文方法用于无线场景，无论是传输压缩码流，还是直接传输潜在特征表示，在高移动性与高密集干扰条件下，或信道存在锐利衰落与路径损耗等效应时，可能引入误码或丢包，进而导致解码失效或重建质量下降。相应的鲁棒传输通常需借助差错控制、重传机制实现，或在联合信源信道编码（joint source-channel coding, JSCC）等端到端框架下进行联合设计。因此，结合 JSCC 框架的扩展研究也是本文后续工作的重要方向。

3.3 复杂度分析

为评估所提出 MHNT 在计算复杂度与参数量方面的优势，本文对不同非线性变换结构的参数量、乘加运算次数（Multiply-Accumulate Operations, MACs）以及实际推理时间进行了对比分析。其中，为直接比较各非线性变换模块的参数量与 MACs，选取编码器中的第一个非线性变换模块进行测算，输入张量大小固定为 192×128×128。为进一步体现整体运行效率，推理时间实验在相同的 NVIDIA Tesla V100 硬件平台上，使用 Kodak 数据集进行测试，并按照标准流程进行预热，最终报告 20 次推理的平均时间，同时给出了细化的时间统计结果，即分析变换网络的编码时间“T 编码时间”和熵模型编码时间“E 编码时间”。需要说明的是，由于机器状态等不可控因素的影响，平均测量仍可能存在细微波动，这属于可接受范围。BD-Rate 的计算以 Ballé2018[7] 为锚点，实验结果如表 1 所示。

从表中可以观察到，MHNT 在参数量和计算量方面较大多数非线性变换取得显著优势，并在 Hyper 熵模型下实现了最高的 R-D 性能增益（-14.065%）。其主要原因在于：MHNT 采用深度可分离卷积与线性投影结构，有效降低了卷积核的空间计算开销，使整体 MACs 相比基于标准卷积堆叠的非线性变换方法大幅减少。同时，MHNT 的高阶通道项与空间项均采用轻量化结构实现特征融合，与多层 ResBlock 和 SMCCT 模块相比，参数规模分别减少约 49.2% 和 54.3%，计算复杂度分别降低约 49.5% 和 54.5%，从而在复杂度与性能之间实现了更优平衡。从实际推理时间来看，尽管深度可分离卷积在理论上具有更低的计算量，但由于其高度依赖逐通道计算，与现有 GPU 的并行架构匹配度较低，硬件支持不充分；此外，1×1 升降维卷积仍占据主要计算开销，使推理时延未能与 MACs 的下降呈线性对应。

3.4 主观质量分析

为进一步验证所提出 MHNT 在主观视觉质量方面的优势，本文选取 Kodak 数据集中的 Kodim9 与 Kodim17 作为测试样例，两幅图像分别包含平滑区域、纹理细节以及复杂边缘等典型结构。实验对比了传统方法 BPG 与 VVC，以及 LIC 方法 He2022[21] 和 Li2024[20]。对于 LIC 方法（包含本文方法），均使用相同的 λ = 0.0035；同时根据各方法在压缩过程中的实际 bpp，对 BPG 与 VVC 的质量因子进行调整，使所有方法在相近的 bpp 下进行公平比较。其中，Kodim9 的 bpp 约为 0.12，Kodim17 的 bpp 约为 0.15。在满足相近 bpp 的条件下，对比了各方法的重建结果，并采用 MulimgView 工具对关键区域进行局部放大，以便更直观地观察纹理与结构还原效果。相应的主观对比结果如图 3 所示。

实验结果表明，在Kodim10图像中，所提方法能够更准确地恢复帆船帆面区域的层次与纹理；相比之下，BPG出现明显的过度平滑，He2022和Li2024也存在细节不足的情况。在Kodim17图像中，所提方法在衣服褶皱与墙面等复杂细节区域的纹理保留方面同样优于其他方法。综上所述，MHNT在提升重建图像的主观视觉质量方面展现出显著优势。

4　消融实验分析

4.1　泛化性分析

为验证MHNT在不同熵模型框架下的适用性与泛化能力，本文将其分别集成到多种主流熵模型中，包括基于空间自回归的Informer[17]和采用多速率概率估计的MPEM[20]。需要说明的是，基于He2022[21]的实验结果已在图2中给出，此处不再重复。在各类熵模型中，本文均以MHNT替换原有结构中的GDN[7]或常见ResBlock[39]非线性变换，并在统一的模型配置与训练策略下进行公平对比。除特别说明外，所有消融实验均在Kodak数据集上完成评测。实验结果如图4所示。

从图中可以观察到，在所有熵模型中引入MHNT后均获得了稳定的率失真性能提升。例如，在MPEM框架下，加入MHNT能够在全码率范围内显著提高PSNR。以GDN方法作为锚点计算BD-rate，相较于ResBlock，所提方法实现约−2.3%的性能增益。进一步地，在更复杂的Informer熵模型中，MHNT在高码率区域的细节重建方面仍保持明显优势，使其性能曲线在整个码率区间均优于对应的基线模型和ResBlock版本。然而，在低码率区

间，其增益幅度相较高码率区域略有下降。这一现象与图2中的趋势一致，主要原因在于低码率下潜变量的空间依赖建模能力受限，使非线性变换在特征表达上的优势被部分削弱，从而使MHNT的高阶特征建模能力无法得到充分发挥。尽管如此，所提方法在低码率区域仍实现了持续的性能改善，说明其结构改进在不同码率条件下均具备有效性。

上述现象表明，MHNT所增强的多维度高阶特征表达能力能够在潜变量中构建更具判别性的上下文依赖关系，从而为不同类型的熵模型提供更加精细的潜在表示，使概率估计更精确。这充分证明了所提MHNT在不同压缩框架中的良好兼容性与泛化能力。

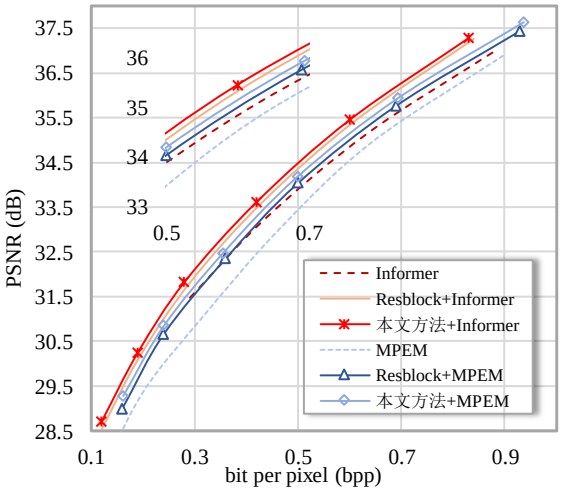

图4　不同熵模型上的消融实验

4.2　组件影响分析

为进一步验证MHNT各子模块（或分支）的有效性与合理性，本文对其关键组件开展了系统性的消融实验。以超先验熵模型[7]为基础框架，并结合第2.2节中所述的MHNT设计理念，即高阶项的表达形式，宏观的消融设置包括：仅使用通道高阶项、仅使用空间高阶项、同时启用通道与空间高阶项（完整MHNT），以及移除特征低阶项四种配置。

图5给出了不同消融配置下的R-D曲线。从图中可以看出，相较于基线模型，仅保留通道高阶项即可在各码率下取得稳定的性能提升。此外，在仅保留空间高阶项的设置下，尽管仍能带来一定幅度的性能改善，但在高码率阶段整体PSNR甚至低于基线模型。这主要是由于空间高阶项采用通道独立的DW卷积结构，其跨通道信息交互能力有限；

而在高码率条件下，压缩系统需要依赖于充分的跨通道相关性建模以准确恢复细节，仅依靠空间维度的高阶建模难以捕获这些复杂依赖，因而导致性能受限。当同时引入通道与空间高阶项时，即构成完整的MHNT，整体性能进一步提升。双分支结构能够同时建模跨通道与空间维度的高阶依赖关系，使得完整MHNT实现出更高的R-D性能。当移除特征低阶项时，MHNT的性能出现明显下降。这表明从特征表达的角度来看，低阶项在整个非线性变换中承担基础映射的作用，为高阶分支提供稳

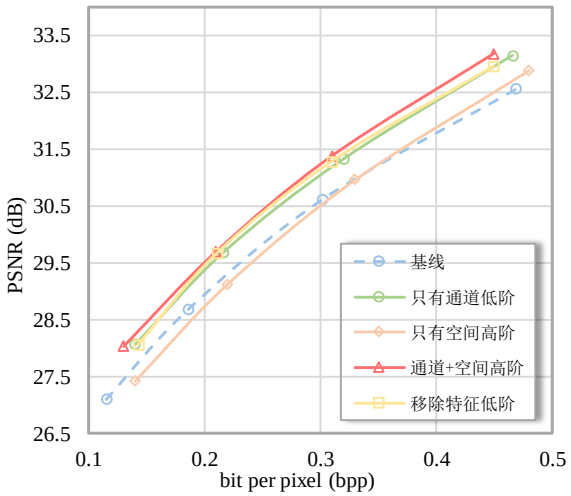

图 5　不同宏观配置下的消融实验

定且连

续的特征基底，对完整的特征表达仍不可或缺。

在验证了通道与空间高阶分支在宏观结构上的有效性之后，本文进一步对 MHNT 的关键设计进行了更细粒度的模块化消融分析。具体而言，本文从更微观的角度分析若干结构因素对性能的影响，包括是否采用 DW 分解、DW 卷积的类型选择、卷积核大小对高阶建模能力的影响，以及通道升维模块对特征表达能力的贡献。

实验结果如图 6 所示，可以观察到，当不采用 DW 分解、而是直接使用标准卷积时，R-D 性能相较于 DW 分解仅获得了小幅度提升。然而，这一差距在 MHNT 轻量化、多分支的整体结构下几乎可以忽略不计。与此同时，标准卷积会带来成倍增长

的计算量与参数量，计算成本显著高于 DW 卷积。综合计算复杂度、参数量与性能增益三者的权衡， DW 分解卷积在 MHNT 中更具优势。此外，当将 1×9 与 9×1 的 DW 卷积替换为 1×5 与 5×1

的较小卷积核时，性能出现了明显下降。这说明由于 DW 卷积本身缺乏跨通道的特征交互能力，因此其空间感受野应尽可能保持较大，以弥补结构上的表达限制；当卷积核尺寸减小后，其特征建模能力受到削弱，从而影响最终压缩性能。此外，当不采用 1×1 卷积进行通道升维时，率失真性能也出现了明显下降。原因在于，通道高阶分支需要在更高的通道维度中进行多项式式高阶特征建模；若不进行升维，分支的表示容量受到限制，高阶项之间的通道混合与组合能力不足，从而削弱了该分支对潜变量复杂通道依赖关系的建模能力，最终导致整体压缩性能下降。

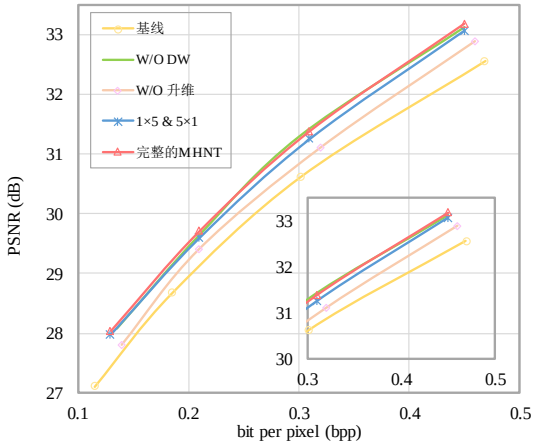

图 6　不同微观配置下的消融实验

综上所述，宏观结构消融表明，通道与空间高阶分支均能提升性能，而双分支融合的完整 MHNT 效果最佳，展示了其高阶建模能力的有效互补。微观模块消融进一步证明了 DW 分解、大核卷积以及通道升维等设计均对性能有正向贡献。整体结果共同说明 MHNT 的结构设计在各层面均具有合理性与稳定的性能贡献。

4.3　融合方式分析

融合方式是 MHNT 构建高阶表示的关键环节之一，不同的融合策略会直接影响通道与空间高阶特征的交互方式及最终的表达能力。本小节将针对多种融合配置进行比较，以评估它们对率失真性能的影响。如图 2 中 MHNT 的整体流程所示，融合操作分为两个阶段：第一阶段为同一高阶分支内部的特征融合，用于构建分支内的高阶关系；第二阶段为通道与空间分支之间的跨分支融合，用于实现不同高阶维度的联合建模。

根据表 2 的实验结果可以观察到，在第一分阶段中采用乘法融合方式明显优于加法融合。加法融合本质上是线性叠加，而乘法融合可视为在不同通道之间执行成对特征相乘的核函数操作，形式上类似于多项式核。当其在网络中多层堆叠时，会带来隐式表示维度的指数式扩展，从而增强模型对高阶依赖关系的建模能力。然而，当第二阶段仍使用乘法融合时，性能却显著下降。可能原因在于，从泰勒展开的角度来看，高阶表示应在基础特征上逐级

叠加加性高阶项，而非继续放大已有的乘法关系；

若第二阶段仍采用乘法，会使前一阶段的高阶项被

进一步放大，导致隐式高阶项迅速膨胀、展开结构被破坏，进而造成特征表达不稳定。在第二阶段采用通道维度拼接并配合 1×1 卷积压缩通道后，其性能与加法类似，主要得益于额外引入了可学习的线性变换，但这种方式带来了额外的参数和计算量，因此整体性价比不及加法融合。

表 2　　MHNT 融合方法的消融实验

方法	率失真损失	Bpp	PSNR
1 阶段加法	0.478	0.20	29.555 dB
2 阶段乘法	0.527	0.21	28.828 dB
2 阶段卷积	0.479	0.20	29.591 dB
本文方法	0.471	0.21	29.703 dB

综上所述，第一阶段适合采用乘法以增强高阶建模能力，而第二阶段采用加法更能保持结构稳定性与计算效率，是当前设计下的最优融合策略。

4.4　特征可视化分析

MHNT 的设计初衷是在保持结构简洁的前提下，引入更高效的高阶建模机制，以进一步增强非线性变换对特征的表达与重构能力。由于“高阶表达能力”本身较为抽象、难以直接量化，本文从可观测的统计特性出发对其进行间接验证。具体而言，在 LIC 中，潜在特征在通道维度上的能量集中特性常被用作衡量变换网络表达能力的重要指标：若变换网络能够将信息更有效地聚合到少数通道，则能量分布将呈现“少数通道占据主要能量、多数通道能量接近零”的特征，这通常意味着潜在表示更稀疏、更易建模，从而有利于后续熵模型建模与熵编码效率提升。为刻画这一特性，按照标准定义和做法[45]，本文首先对非线性变换前后各通道能量占比的分布进行统计分析，用以验证 MHNT 是否能够实现更强的能量集中，从而间接反映其高阶特征建模与表达能力的提升。

如图 7 所示，为直观刻画不同非线性变换模块的能量聚合能力，本文在编码器中选取第二次非线性变换前后的特征，计算各通道的能量占比，并按

照能量大小对通道进行降序排序后绘制前 60 个通道的能量谱。因此，横轴的通道索引 k 表示排序后的序号，而非原始通道编号。从图中可以观察到，与 Hu2023[43] 相比，本文方法在 Kodim02 与 Kodim06 上均呈现更显著的能量集中特性：其能量谱在前面通道形成更高峰值，并表现出更快的衰减率，说明更多能量被压缩到少数主导通道，而尾部

通道的能占比接近于零。与 Bao2022[45] 相比，尽管该方法在头部通道也实现了较高能量，但其能量在第二、第三等后续主导通道中仍有明显分布，而本文方法能够将能量进一步集中，使尾部通道的能量几乎完全消失，体现出更强的能量压缩能力。上述结果从统计特性上验证了本文在潜在表示上的能量聚合与特征表达能力更强，从而降低了熵模型的建模难度，有助于进一步提升编码效率。

5　结束语

本文提出了一种基于多维度高阶非线性变换的可学习图像信源编码方法。该方法通过将传统的线性与非线性堆叠映射重新建模为高阶非线性组合函数，并据此构建多维度高阶非线性变换模块，从而在通道与空间维度以类泰勒展开的方式显式构造多

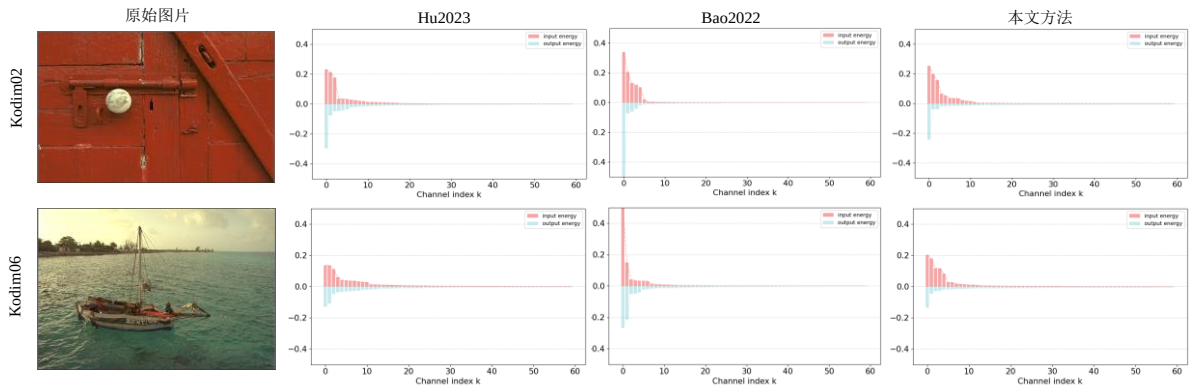

图 7　能量集中特性比较

阶特征项。该模块与可学习图像压缩框架实现端到端集成，显著提升了模型的表达能力与非线性拟合能力。本文在 Kodak 和 Tecnick 标准图像数据集上进行测试，结果表明，本文方法在保持较低参数量和计算开销的前提下，为多种 LIC 框架带来了稳定的率失真性能增益。尽管该方法在平衡性能与复杂度方面表现优异，但其设计仍停留在有限阶高阶结构层面，难以直接拓展至更复杂的 LIC 结构和视频编码场景中。此外，高阶特征项的内部机理尚不明确，仍有待深入的归因分析。未来将从结构泛化与模型可解释性两个维度展开研究，以提升该方法在更广泛压缩场景中的适用性与理论完备性。

一寸照片

谭文（1999-），男，湖南衡阳人，哈尔滨工业大学（深圳）博士生，主要研究方向为深度学习、图像/视频编码。

孟凡阳（1986- ），男，河南南阳人，博士，鹏城实验室副研究员，主要研究方向为信源信道编码、智能视频编码。

柳伟（1973- ），男，湖南长沙人，博士，深圳信息职业技术大学教授，主要研究方向为人工智能、视觉媒体处理。

梁永生（1971- ），男，黑龙江肇东人，博士，哈尔滨工业大学教授，主要研究方向为软硬件协同优化、信源-信道-网络联合优化编码。

参考文献：

[1] 杨栩, 朱策, 郭红伟, 等. 基于球域失真空-时依赖的全景视频编码[J]. 通信学报, 2023, 44(10): 58-71.
Yang X, Zhu C, GUO H W, et al. Panoramic video coding based on spherical distortion with spatio-temporal dependency[J]. Journal on Communications, 2023, 44(10): 58-71.

[2] 杨舒涵, 申滨, 黄晓舸. 基于深度灵活编码策略与性能预测的无线语义图像传输系统[J]. 通信学报, 2025, 46(05): 29-46.
Yang S H, Shen B, Huang X G. Wireless semantic image transmission system based on deep flexible coding strategy and performance prediction[J]. Journal on Communications, 2025, 46(05): 29-46.

[3] 郭红伟, 朱策, 杨栩, 等. 基于失真反向传播的时域依赖率失真优化[J]. 通信学报, 2022, 43(12): 222-232.
Guo H W, Zhu C, Yang X, et al. Temporal dependent rate-distortion optimization based on distortion backward propagation[J]. Journal on Communications, 2022, 43(12): 222-232.

[4] 高文, 田永鸿, 王坚. 数字视网膜: 智慧城市系统演进的关键环节[J]. 中国科学: 信息科学, 2018, 48(8): 1076-1082.
Gao W, Tian Y H, W J. Digital retina: revolutionizing camera systems for the smart city[J]. Science China Information Science, 2018, 48(8): 1076-1082.

[5] 赵琛, 马思伟, 张新峰, 等. 基于云数据的高效图像编码方法[J]. 计算机学报, 2017, 40(11): 2433-2447.
Zhao C, Ma S W, Zhang X F, et al. An efficient image compression method based on cloud data[J]. Chinese Journal of Computers, 2017, 40 (11): 2433-2447.

[6] 贾川民, 赵政辉, 王苫社, 等. 基于神经网络的图像视频编码[J]. 电信科学, 2019, 35(05): 32-42.
Jia C M, Zhao Z H, Wang S S, et al. Neural network based image and video coding technologies[J]. Telecommunications Science, 2019, 35 (05): 32-42.

[7] Ballé J, Minnen D, Singh S, et al. Variational image compression with a scale hyperprior[J]. arXiv:1802.01436, 2018.

[8] Tang Z S, Wang H L, Yi X K, et al. Joint graph attention and asymmetric convolutional neural network for deep image compression[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 33 (1): 421-433.

[9] Tu H Y, Wu S Q, Li L, et al. Multi-scale invertible neural network for wide-range variable-rate learned image compression[J]. arXiv: 2503.21284, 2025.

[10] Fu H S, Liang F, Liang J, et al. Asymmetric learned image compression with multi-scale residual block, importance scaling, and post-quantization filtering[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(8): 4309-4321.

[11] Chen T, Liu H, Ma Z, et al. End-to-end learnt image compression via non-local attention optimization and improved context modeling[J]. IEEE Transactions on Image Processing, 2021, 30: 3179-3191.

[12] El-Nouby A, Muckley M J, Ullrich K, et al. Image compression with product quantized masked image modeling[J]. arXiv: 2212.07372, 2022.

[13] Li S H, Li H, Dai W R, et al. Learned progressive image compression with dead-zone quantizers[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 33(6): 2962-2978.

[14] Ge Z Q, Ma S W, Gao W, et al. NLIC: non-uniform quantization-based learned image compression[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(10): 9647-9663.

[15] Li S H, Dai W R, Kan N W, et al. Learnable non-uniform quantization with sampling-based optimization for variable-rate learned image compression[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(8): 8314-8329.

[16] Minnen D, Balle J, Toderici G. Joint autoregressive and hierarchical priors for learned image compression[C]//2018 NeurIPS Conference (NIPS). New York: Curran Associates, 2018: 10794-10803.

[17] Kim J H, Heo B H, Lee J S. Joint global and local hierarchical priors for learned image compression[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 5992-6001.

[18] Fu H, Liang F. Learned image compression with generalized octave convolution and cross-resolution parameter estimation[J]. Signal Processing, 2023, 202: 108778.

[19] Minnen D, Singh S. Channel-wise autoregressive entropy models for learned image compression[C]//2020 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2020:

3339-3343.

[20] 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.

[21] He D, Yang Z, Peng W, et al. ELIC: efficient learned image compression with unevenly grouped space-channel contextual adaptive coding [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 5718-5727.

[22] 柏园超, 刘文昌, 江俊君, 等. 深度神经网络图像压缩方法进展综述[J]. 电子与信息学报, 2025, 47(11): 4112-4128.
Bai Y C, Liu W C, Jiang J J, et al. Advances in deep neural network based image compression: a survey[J]. Journal of Electronics & Information Technology, 2025, 47(11): 4112-4128.

[23] 贾川民, 马海川, 杨文瀚, 等. 视频处理与压缩技术[J]. 中国图象图形学报, 2021, 26(06): 1179-1200.
Jia C M, Ma H C, Yang W H, et al. Video processing and compression technologies[J]. Journal of Image and Graphics, 2021, 26(06): 1179-1200.

[24] Hu Y Y, Yang W H, Ma Z, et al. Learning end-to-end lossy image compression: a benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(8): 4194-4211.

[25] Ma H C, Liu D, Xiong R Q, et al. iWave: CNN-based wavelet-like transform for image compression[J]. IEEE Transactions on Multimedia, 2019, 22(7): 1667-1679.

[26] Xie Y Q, Cheng K L, Chen Q F, et al. Enhanced invertible encoding for learned image compression[C]//2021 ACM International Conference on Multimedia (ACMMM). New York: ACM Press, 2021: 162-170.

[27] Wang D Z, Yang W H, Hu Y, et al. Neural data-dependent transform for learned image compression[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 17379-17388.

[28] Akbari M, Liang J, Han J, et al. Learned variable-rate image compression with residual divisive normalization[C]//2020 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE Press, 2020: 1-6.

[29] Shin C J, Lee H M, Son H B, et al. Expanded adaptive scaling normalization for end-to-end image compression[C]//2022 European Conference on Computer Vision (ECCV). Cham: Springer, 2022: 390-405.

[30] Zou R, Song C, Zhang Z. The devil is in the details: window-based attention for image compression[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 17492-17501.

[31] Duan W H, Chang Z, Jia C M, et al. Learned image compression using cross-component attention mechanism[J]. IEEE Transactions on Image Processing, 2023, 32: 5478-5493.

[32] Feng D H, Cheng Z X, Wang S, et al. Linear attention modeling for learned image compression[C]//2025 IEEE/CVF Computer Vision and Pattern Recognition Conference (CVPR). Piscataway: IEEE Press, 2025: 7623-7632.

[33] Zhu Y H, Yang Y, Cohen T. Transformer-based transform coding[C]// 2022 International Conference on Learning Representations (ICLR). Piscataway: IEEE Press, 2022: 1-10.

[34] Zeng F H, Tang H, Shao Y H, et al. MambaIC: state space models for high-performance learned image compression[C]//2025 Computer Vision and Pattern Recognition Conference (CVPR). Piscataway: IEEE Press, 2025: 18041-18050.

[35] Liu J M, Sun H M, Katto J. Learned image compression with mixed transformer-CNN architectures[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 14388-14397.

[36] Li H, Li S H, Dai W R, et al. Frequency-aware transformer for learned image compression[J]. arXiv:2310.16387, 2023.

[37] Bao Y N, Tan W, Li M, et al. SFNIC: hybrid spatial-frequency information for lightweight neural image compression[J]. CAAI Transactions on Intelligence Technology, 2025, 10(6): 1717-1730.

[38] Tan W, Meng F Y, Li C, et al. Exploring high-dimensional feature space with channel-spatial nonlinear transforms for learned image compression[J]. CAAI Transactions on Intelligence Technology, 2025, 10 (4):1235-1253.

[39] Cheng Z, Katto J. Learned image compression with discretized Gaussian mixture likelihoods and attention modules[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 7936-7945.

[40] Li C, Li T Y, Meng F Y, et al. One is all: a unified rate-distortion-complexity framework for learned image compression under energy concentration criteria[J]. IEEE Transactions on Multimedia, 2025,27(2): 3992-4007.

[41] Hu Y, Yang W H, Liu J. Coarse-to-fine hyper-prior modeling for learned image compression[C]//2020 AAAI Conference on Artificial Intelligence (AAAI). Palo Alto: AAAI Press, 2020: 11013-11020.

[42] Tan W, Bao Y N, Meng F Y, et al. Adaptive cross-channel transformation based on self-modulation for learned image compression[J]. Signal Processing: Image Communication, 2025: 117325.

[43] Hu Y T, Tan W, Meng F Y, et al. A decoupled spatial-channel inverted bottleneck for image compression[C]//2023 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE Press, 2023: 1740-1744.

[44] Ma X, Dai X, Bai Y, et al. Rewrite the Stars[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2024: 5694-5703.

[45] Bao Y N, Meng F Y, Li C, et al. Nonlinear transforms in learned image compression from a communication perspective[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 33(4): 1922-1936.

[46] Tan W, Bao Y N, Meng F Y, et al. Grouped Transform for Ultra-Low-Complexity Learned Image Compression[C]// 2025 IEEE International Symposium on Circuits and Systems (ISCAS). Piscataway: IEEE Press, 2025: 1-5.

[47] Xiao T, Zhang W, Cheng Y, et al. HOPE: High-Order Polynomial Expansion of Black-Box Neural Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 7924-7939.

李超（1996− ），男，广西北海人，哈尔滨工业大学（深圳）博士生，主要研究方向为深度学习、视频编码、神经网络结构优化等。