arXiv:2603.06766v1 [eess.IV] 6 Mar 2026

HiDE：面向学习式图像压缩的分层字典熵建模

Haoxuan Xiong[1] , **Yuanyuan Xu**[1*]、 **Kun Zhu**[2] , **Yiming Wang**[3] , **Baoliu Ye**[4]河海大学[2]南京航空航天大学[3]南京审计大学[4]南京大学{haoxuan x, yuanyuan xu}@hhu.edu.cn, zhukun@nuaa.edu.cn, wangyiming@nau.edu.cn, yebl@nju.edu.cn,

摘要

学习式图像压缩（LIC）已实现显著的编码效率，其中熵建模通过利用信息先验来降低比特率，发挥着关键作用。现有方法主要利用输入图像中的内部上下文，却未能充分利用蕴含于大规模训练数据中的丰富外部先验。近期，基于字典的熵模型取得了进展，表明融入外部先验能够显著提升压缩性能。然而，当前方法将异质外部先验组织在单层字典中，导致利用不均衡且表征能力有限。此外，有效的熵建模不仅需要具有表达力的先验，还需要能够对其进行解读的参数估计网络。为应对这些挑战，我们提出了**HiDE**，一种面向学习式图像压缩的**分层字**典**熵**建模框架。HiDE通过级联检索，将外部先验分解为全局结构字典和局部细节字典，从而实现对外部信息的结构化、高效利用。此外，我们还引入了一种采用并行多感受野设计的上下文感知参数估计器，以自适应地利用异质上下文，实现精确的条件概率估计。实验结果表明，在Kodak、CLIC和Tecnick数据集上，HiDE相较于VTM-12.1的BD-Rate分别降低了**18.5%**、**21.99%**和**24.01%**。

1 引言

图像压缩仍是多媒体通信中的一项基本挑战，旨在降低存储和传输成本，同时保持重建质量。近年来，学习式图像压缩（LIC）取得了长足进展，在率失真（RD）性能方面超越了 JPEG [Wallace, 1991]和 VVC-Intra [Bross 等人的传统标准。LIC 遵循变分自编码器

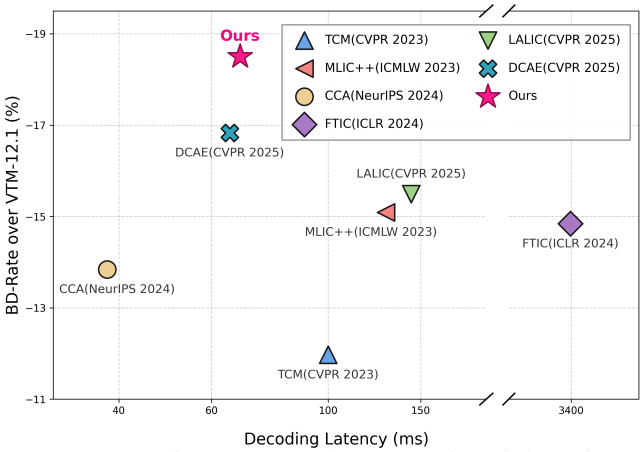


图1：Kodak 数据集上的 BD-Rate 与解码延迟（左上方表示性能更佳）。

框架，将图像转换为潜在表示，再对其进行量化和熵编码。在这一框架中，熵模型对压缩效率起着决定性作用，因为它对潜在表示的概率分布进行建模，而该分布决定了熵编码后的比特率。

为了尽可能降低比特率，降低潜在表示的不确定性至关重要。熵模型通过基于信息先验进行概率估计来实现这一目标，从而降低每个潜在符号的条件熵。开创性工作 [Ball´e 等，2017年；Minnen 等，2018a] 引入了超先验以捕捉空间依赖性，并引入自回归模型以利用因果上下文。后续研究着重增强上下文模型的表达能力，以更有效地捕捉潜在符号之间的相关性。相关方法从通道切片 [Minnen 和 Singh， 2020] 和棋盘格机制 [He 等， 2021] 逐步发展到复杂的空间-通道上下文 [He 等，2022年；Jiang 等，2023年；Jiang 和 Wang，2023]。值得注意的是，MLIC系列 [Jiang 等，2023年；Jiang 和 Wang，2023] 采用先进的注意力机制，联合整合局部、全局和逐通道上下文，取得了出色的性能。然而，这些方法——

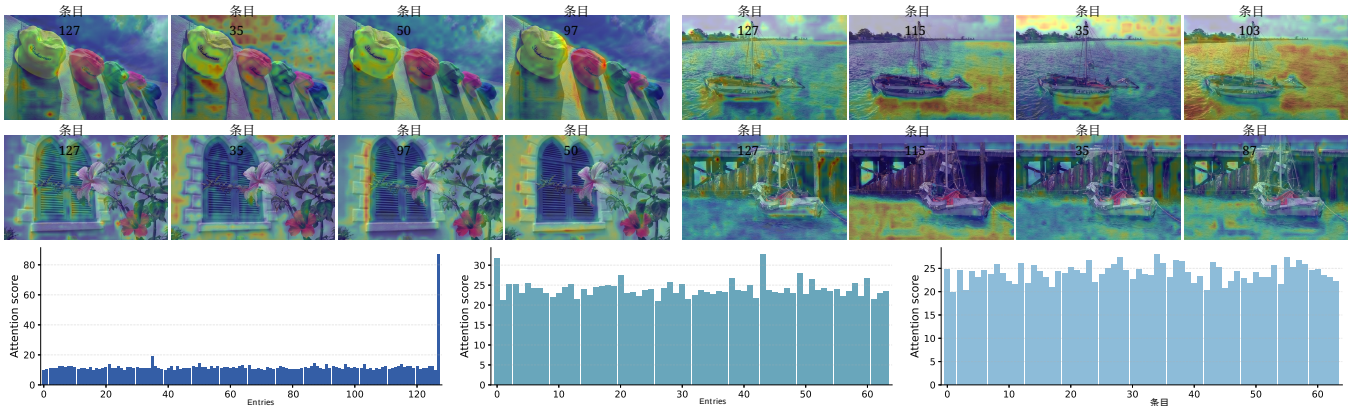


图 2：词典条目利用率可视化。前两行展示了 DCAE 在 Kodak 图像上的注意力热图，凸显检索过程中占主导地位的条目。底行展示了各词典条目在 Kodak 数据集上的平均注意力分数。DCAE 的单层词典（左）呈现出高度偏斜的使用情况，而我们提出的**全局**（中）和**细节**（右）词典则实现了更均衡的利用。

——完全依赖从输入中提取的内部上下文，而忽略了大规模训练数据中蕴含的丰富统计模式。

近期，基于词典的交叉注意力熵模型（DCAE）[Lu等， 2025]通过引入学习得到的词典作为外部先验，解决了这一局限。DCAE 利用交叉注意力检索相关词典条目，将其作为熵建模的外部先验，从而降低潜在符号的不确定性。即使与简单的逐通道上下文相结合，DCAE 仍实现了最先进的性能，凸显了相较于日益复杂的内部上下文建模，外部先验的有效性。

使用有限词典来表示多样化的视觉内容，天生容易导致表示坍塌，这是向量量化生成模型中一个广为人知的问题 [Oord 等，2017年；Esser 等，2021年；Zhu 等， 2025]。在这种情况下，少数词典条目会被频繁选中，而大多数条目则很少得到利用。为进一步研究这一现象，我们分析了 DCAE 中的词典利用模式。如图 2 所示，词典条目的注意力热图呈现出明显的赢家通吃倾向：少数通用模式在不同图像的检索过程中占据了不成比例的主导地位。具体而言，条目 127 持续响应复杂的结构区域，而条目 35 和 115 则主要在较平滑的区域激活。图 2 左下方的直方图显示出明显的偏斜分布：大多数词典条目的注意力分数相对较低，而少数条目的激活值则显著更高。这种不均衡表明，外部先验未得到均衡利用，退化为静态偏差，而非发挥动态的、内容自适应参照的作用，从而对熵建模造成表示瓶颈。

此外，仅有丰富的先验并不能保证准确的条件概率估计。有效的熵建模需要一个参数估计网络，能够将异构的上下文先验转换为适当的熵参数。现有方法 [Minnen 和 Singh，2020年；Liu 等，2023年；Lu 等，2025]通常采用具有固定感受野的浅层卷积估计器，以整合超先验、自回归上下文和基于词典的先验。随着上下文信息多样性的增加，这类有限的架构会限制对更丰富先验的有效利用，并制约模型性能。

为应对上述挑战并充分挖掘外部先验建模的潜力，我们提出了**HiDE**，这是一种用于学习式图像压缩的分层字典熵模型，并配备了上下文感知参数估计网络。我们的主要贡献总结如下：

- 我们提出了一种分层字典熵框架，将外部先验分解为全局结构字典和局部细节字典，从而能够以结构化且高效的方式利用外部信息。
- 我们设计了一种上下文感知参数估计网络，其中包含多感受野上下文提取器，能够自适应地利用多种上下文，从而更准确地估计条件概率。
- 全面的实验表明，HiDE在多个基准数据集上始终优于现有最先进方法，同时具备有竞争力的解码速度，如图1和表1所示。

2 相关工作

2.1 学习式图像压缩

LIC 采用非线性变换框架 [Ball´e等人,2017, 2021], 通常由卷积神经网络 (CNNs)、视觉变换器 (ViTs) 或结合二者的混合架构参数化 [Cheng 等人, 2020年; Zhong等人, 2020 年 ; Zhu 等人 , 2022; Liu 等人 , 2023 年 ; Li 等人 , 2024; Feng 等人 , 2025 年 ; Zeng 等人 , 2025]。这些研究主要聚焦于架构改进，以获得更紧凑的潜在表示。然而，仅靠网络架构带来的性能提升，最终会受到潜在空间表征能力的限制。

为降低编码成本，熵模型旨在缩小预测概率分布与潜变量实际分布之间的交叉熵。超先验模型 [Ball´e等人，2018]引入辅助信息来估计条件熵并改进概率建模。在此基础上，自回归模型 [Minnen 等人, 2018a] 利用先前解码的元素作为因果上下文，显著提高了预测准确性。为平衡压缩效率与解码并行度，后续方法提出了通道切片 [Minnen 和 Singh, 2020]以及棋盘式 [He等人, 2021] 上下文模型，将潜变量划分为多个组以便并行处理。

近期的研究进一步通过多参考先验 [He等，2022年; Jiang 等，2023年; Jiang 和 Wang， 2023] 或通过通道级因果调整损失 [Han等， 2024]丰富了上下文模型。比如， [Jiang等，2023] 引入了复杂的空间-通道上下文建模，以捕捉潜在空间中的组件间及组件内相关性。尽管取得了这些进展，这些上下文模型仍完全依赖从输入中提取的内部信息，忽视了大规模训练数据中蕴含的丰富外部先验，而这些先验本可进一步增强熵建模。

2.2 字典学习

字典学习为利用外部数据先验提供了一种有效范式。在生成建模中，向量量化（VQ）方法 [Oord等，2017年; Esser 等，2021]表明，学习得到的字典能够通过将图像表示为离散编码条目的组合来概括复杂的视觉模式。

在图像压缩领域，早期方法 [Minnen等，2018年b]采用了静态非可学习字典，缺乏适应多样化外部知识的灵活性。近年来，条件潜编码（CLC） [Wu 等，2025] 构建特征字典，为潜变量调整提供条件参考；而基于掩码的选择性补偿（MSC） [Kuang等， 2025]则检索补偿向量以纠正残差误差。尽管这些方法验证了外部知识库的优势，但其关注重点主要在重建，而非熵估计。

基于字典的交叉注意力熵模型（DCAE） [Lu等， 2025] 通过以下方式解决了这一局限：
将字典先验与内部先验相结合，以提高概率估计的准确性。尽管效果显著，DCAE 仍采用单层词典来表示所有视觉模式。正如我们的分析所指出的，这种扁平设计容易导致表示坍塌，从而造成字典利用不均，并限制模型表达能力。

2.3 熵参数估计

虽然上下文模型能够捕捉潜变量之间的依赖关系，但参数估计网络在将这些先验映射为条件潜变量分布的参数方面发挥着关键作用。高斯尺度混合（GSM）模型 [Ball´e 等人， 2018] 预测潜变量的尺度；而 [Minnen 等人，2018a] 则扩展了 GSM，可联合估计均值和尺度，从而实现更准确的密度建模。潜变量残差预测（LRP） [Minnen

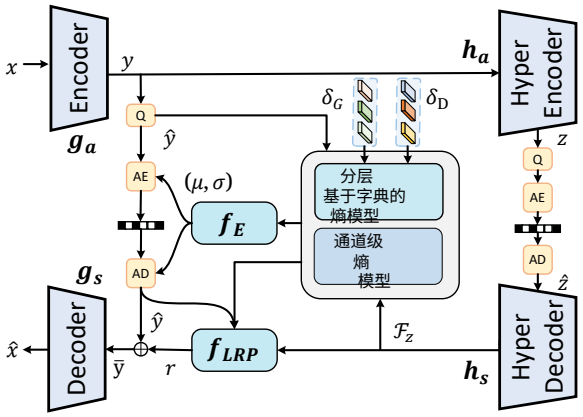


图3：所提出的 HiDE 框架概览。HiDE 整合分层字典 (δ_G， δ_D) 以检索外部先验，并将其与超先验 $\mathcal{F}_z$ 及按通道自回归上下文融合。聚合后的上下文引导熵参数估计器 f_E 和潜变量残差预测器 f_{LRP}，以提升重建质量并提高熵建模的准确性。

和 Singh， 2020]进一步改进了量化误差预测。这些研究确立了这样一种做法：预测均值、尺度，以及可能的残差，以进行熵编码。

尽管上下文模型日趋复杂，参数估计网络的架构却基本没有变化。大多数方法都采用感受野固定的浅层卷积估计器，不考虑输入先验的异质性。然而，现代熵模型融合了高度异质的上下文来源，包括编码全局统计信息的超先验、捕获局部因果依赖的自回归上下文，以及近年来出现的外部字典先验 [Ball´e等，2018年; Minnen 等，2018b；Minnen 和 Singh，2020年; He 等，2021年、2022年；Jiang 等，2023年；Lu 等，2025]。对这类异质上下文应用固定尺度的卷积，会限制参数估计器充分利用它们的互补特性。从这一角度看，问题不在于上下文不足，而在于参数估计过程中对上下文的提取和利用不够充分。

3 方法

3.1 预备知识

如图3所示，编码器 g_a 将输入图像 x 转换为连续的潜在表示 $y = g_a(x)$。为支持熵编码，超先验模块提取辅助信息 $z = h_a(y)$，并将其量化为 $\hat{z} = Q(z)$后编码到比特流中。此处， Q表示舍入操作。基于解码后的超先验 $\hat{z}$，超解码器 h_s 生成超先验上下文特征$\mathcal{F}_z = h_s(\hat{z})$。给定预测均值 μ 后，将 y 量化为 $\hat{y} = Q(y - \mu) + \mu$ ，并对 $Q(y - \mu)$进行无损压缩。潜变量残差预测（LRP）估计量化误差（残差） $r \approx y - \hat{y}$ ，并将潜在表示细化为 $\bar{y} = \hat{y} + r$。最后，解码器 g_s 重建图像 $\hat{x} = g_s(\bar{y})$。该框架遵循率失真

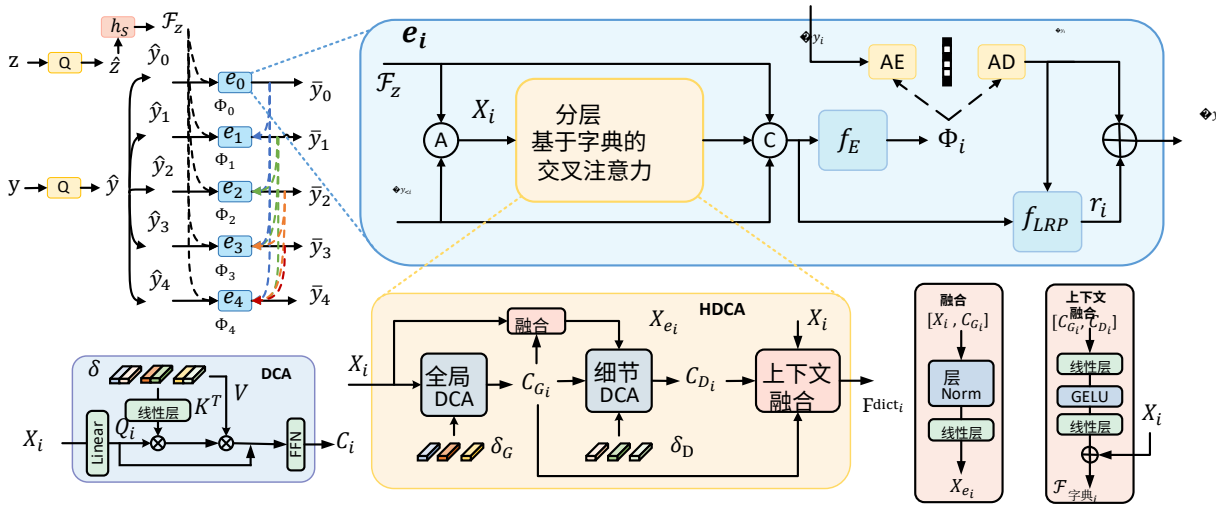


图4：所提出的基于层次字典的熵模型概览，该模型由基于层次字典的上下文模型（HD）和上下文感知参数估计（CaPE）组成。左上：切片熵模型，聚合超先验 $\mathcal{F}_z$ 与自回归上下文 $\bar{y}_{<i}$。右上（蓝色框 e_i）：第 i个切片网络的内部结构，采用层次字典交叉注意力（HDCA）模块检索外部先验，用于参数估计（f_E）和残差预测（f_{LRP}）。下方（黄色框）：HDCA执行两阶段检索：首先查询**全局字典** δ_G 以获取全局结构先验$\mathcal{C}_{G_i}$，然后在全局上下文条件下查询**细节字典** δ_D 以获取局部细节纹理先验 $\mathcal{C}_{D_i}$。右下（粉色框）：融合阶段生成中间 X_{e_i} 与最终字典感知上下文$\mathcal{F}_{dict_i}$。

优化，

$$\mathcal{L} = \mathcal{R}(\hat{y}) + \mathcal{R}(\hat{z}) + \lambda \mathcal{D}(x, \hat{x}), \tag{1}$$

其中 $\mathcal{R}(\cdot)$使用熵模型估计比特率， $\mathcal{D}(\cdot)$衡量重建失真。由于HiDE的主要贡献在于熵建模，我们对 g_a、 g_s、h_a 和 h_s采用DCAE [Lu 等人， 2025] 的骨干架构。

比特率 $\mathcal{R}(\hat{y})$ 主要取决于条件概率 $p(\hat{y}|\cdot)$ 的准确性。根据高斯假设 [Ball´e 等人，2018年；Minnen 等人，2018a]，，条件密度建模为 $p_{\hat{y}}(\hat{y}|\hat{z}) = \prod_i \left(\mathcal{N}(\mu_i, \sigma_i) * \mathcal{U}(-\frac{1}{2}, \frac{1}{2})\right)(\hat{y}_i)$，其中 $\mathcal{N}(\mu_i, \sigma_i)$ 是均值为 μ_i、标准差为 σ_i 的高斯分布，而 $*$ 表示与单位均匀分布 $\mathcal{U}(-\frac{1}{2}, \frac{1}{2})$ 的卷积，用于处理量化噪声。为捕获通道间的依赖关系，将 y 划分为 s 个通道切片 $\{y_0, \ldots, y_{s-1}\}$。对于每个切片 i，上下文包括超先验特征$\mathcal{F}_z$ 和先前解码的切片 $\bar{y}_{<i}$。参数估计网络 f_E 预测 $\Phi_i = (\mu_i, \sigma_i)$，潜变量残差预测器 f_{LRP} 则估计量化残差r_i：

$$\begin{aligned} \mu_i, \sigma_i &= f_E(\mathcal{F}_z, \bar{y}_{<i}, \mathcal{F}_{dict_i}), \\ r_i &= f_{LRP}(\mathcal{F}_z, \bar{y}_{<i}, \mathcal{F}_{dict_i}, \hat{y}_i), \end{aligned} \tag{2}$$

其中 $0 \leq i < s$，且 $\mathcal{F}_{dict_i}$ 表示下文介绍的分层字典上下文。图4给出了所提出的分层基于字典的熵模型概览，该模型包含两个关键组件：基于层次字典的上下文模型（HD）和上下文感知参数估计模块（CaPE）。

3.2 基于层次字典的上下文建模

为有效利用外部先验并缓解字典表征坍缩，我们提出基于层次字典的上下文模型（HD），该模型将外部知识分解为以由粗到细方式检索的互补全局与局部分量。

构建两个可学习字典，并由编码器和解码器共享。全局结构字典 $\delta_G \in \mathbb{R}^{N_G \times C_d}$ 旨在捕获全局模式和长程依赖，而局部细节字典 $\delta_D \in \mathbb{R}^{N_D \times C_d}$ 则专注于细粒度纹理和局部依赖。 N_G， N_D 表示条目数量，C_d 表示每个字典条目的通道维度。两个字典均在学习式图像压缩中联合优化。

在切片级上下文模型中，潜在表示 y被划分为通道级切片 y_i。对于第 i 个切片，输入上下文 X_i 由超先验特征 $\mathcal{F}_z$ 和先前解码的切片 $\bar{y}_{<i}$聚合而成。该聚合方式后接 [Lu等人， 2025]。

层次化检索分为两个连续阶段，通过逐步细化外部先验来完成。第一阶段使用 Q_{G_i} 查询全局字典，借助交叉注意力检索全局上下文特征 $\mathcal{C}_{G_i}$。全局字典同时充当键和值，提供粗粒度的结构参考。我们采用多头交叉注意力机制，

$$Q_{G_i} = X_i W_Q^G, \quad K_G = \boldsymbol{\delta}_G W_K^G, \quad V_G = \boldsymbol{\delta}_G, \tag{3}$$

$$\mathcal{C}_{G_i} = \text{Softmax}\left(\frac{Q_{G_i} K_G^T}{\tau_i}\right) V_G, \tag{4}$$

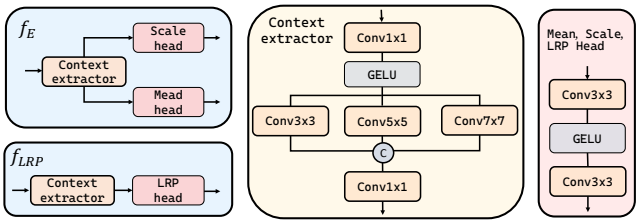


图 5：所提出的上下文感知参数估计（CaPE）模块示意图。左：CaPE 用于熵模型中的两个模块。首先，CaPE 作为参数估计器 f_E 提取共享上下文表示，以预测 μ 和σ。然后， f_{LRP} 预测量化残差 r。中：上下文提取器的内部结构。右：用于均值、尺度和残差预测的任务特定头，由轻量级堆叠 3×3 卷积和 GELU 激活构成。

其中 W_Q^G和 W_K^G是可学习的投影矩阵。 τ_i是一个可学习的温度参数，用于控制注意力分布的尖锐程度。通过提供结构先验，该阶段在连贯的框架内降低了后续检索中的不确定性。

第二阶段在全局上下文的条件下检索细节纹理。如融合模块（图 4，右下）所示，增强后的查询X_{e_i} 通过线性投影后接层归一化运算符，将原始上下文 X_i 与全局先验 $\mathcal{C}_{G_i}$ 融合而成：

$$X_{e_i} = \text{LayerNorm}([X_i, \mathcal{C}_{G_i}]W_{proj}). \tag{5}$$

以全局上下文为条件进行细节检索，可将纹理选择限制在结构一致的范围内，从而实现更稳定且更具判别性的字典利用。细节注意力计算如下：

$$Q_{D_i} = X_{e_i}W_Q^D, \quad K_D = \delta_D W_K^D, \quad V_D = \delta_D, \tag{6}$$

$$\mathcal{C}_{D_i} = \text{Softmax}\left(\frac{Q_{D_i}K_D^T}{\tau_i}\right)V_D. \tag{7}$$

最后，将检索到的全局上下文和细节上下文进行整合，形成字典感知表示。如图 4（右下）所示，我们首先采用轻量级线性层来融合异构字典特征$[\mathcal{C}_{G_i}, \mathcal{C}_{D_i}]$。为确保这些外部先验在发挥细化作用的同时不丢失原始内部上下文，我们通过从输入 X_i引入残差连接，对这种融合进行数学形式化表示：

$$\mathcal{F}_{\text{dict}_i} = \phi([\mathcal{C}_{G_i}, \mathcal{C}_{D_i}]W_1)W_2 + X_i, \tag{8}$$

其中， W_1， W_2 为投影矩阵， ϕ 为高斯误差线性单元激活函数。这种残差设计有助于梯度有效传播，并确保模型明确学习如何利用外部知识丰富内部上下文。

3.3 上下文感知参数估计

为了有效利用并解读来自超先验、通道级自回归上下文和字典上下文的多样先验信息，我们引入了上下文感知参数估计（CaPE）模块，如图 5 所示。CaPE 对传统参数估计器进行了增强，采用具有多感受野设计的并行分支，动态捕获不同类型上下文之间的相关性。

给定聚合特征 $\mathcal{S} = [\mathcal{F}_z, \bar{y}_{<i}, \mathcal{F}_{dict_i}]$，CaPE 首先应用 1×1 卷积，将 $\mathcal{S}$ 投影到较低维的特征空间 $\mathcal{S}_{proj}$，随后通过三个并行卷积分支，其卷积核尺寸为 $k \in \{3,5,7\}$:

$$F_k = \phi(\text{Conv}_{k\times k}(\mathcal{S}_{proj})), \quad k \in \{3,5,7\}, \tag{9}$$

其中 ϕ表示 GELU 激活函数。这些分支的输出同时捕获局部和全局依赖关系，并通过另一个 1×1 卷积进行拼接与融合：

$$\mathcal{F}_{ctx} = \text{Conv}_{1\times1}(\text{Concat}([F_3, F_5, F_7])). \tag{10}$$

融合表示 $\mathcal{F}_{ctx}$ 作为参数估计的共享上下文。具体而言， f_E 通过两个轻量级任务特定头 $\mathcal{H}_\mu$ 和 $\mathcal{H}_\sigma$预测高斯分布参数（μ, σ）。对于潜在残差预测， f_{LRP}采用另一个上下文提取器及LRP头 $\mathcal{H}_{lrp}$ 来估计量化残差 r:

$$\mu = \mathcal{H}_\mu(\mathcal{F}_{ctx}), \quad \sigma = \mathcal{H}_\sigma(\mathcal{F}_{ctx}), \quad r = \mathcal{H}_{lrp}(\mathcal{F}_{ctx}). \tag{11}$$

得益于基于并行分支多感受野的上下文提取器与任务特定预测头，CaPE能够实现更准确的熵参数预测和残差校正，并在与分层字典先验相结合时显著提升压缩性能。

4 实验

4.1 实验设置

我们的模型使用从OpenImage数据集 [Krasin 等人，2017]中采样的30万张图像进行训练。训练期间，我们随机裁剪 256×256 图像块，批大小为16。优化采用Adam优化器，初始学习率为 1×10^{-4}，共训练80个epoch，随后学习率衰减至 1×10^{-5}，再训练20个epoch。为获得不同的率–失真（RD）权衡，式（1）中的加权参数 λ 在$\{0.0018, 0.0035, 0.0067, 0.013, 0.025, 0.05\}$之间变化。训练在两块NVIDIA RTX 4090 GPU上进行，每个比特率约需16天。六个码率点通过针对 $\lambda \in \{0.0035, 0.05\}$ 训练独立模型，并从第95个epoch的预训练检查点初始化、再对额外模型微调5个epoch而获得。更多实现配置与消融结果见补充材料。

4.2 与最先进方法的比较

我们在三个广泛使用的基准测试上评估所提出的 HiDE 框架：Kodak [kodak、 1993]，Tec- nick [Asuni 等、 2014]，以及 CLIC 专业验证数据集 [CLIC, 2021]。比较对象包括传统编解码器 VVC （ VTM-12.1 ） [Dominguez 和 Rao,2022]，以及近期最先进的学习式图像压缩（LIC）模型，即 TCM [Liu 等, 2023]，MLIC+ [Jiang等，2023]，MLIC++ [Jiang 和 Wang，2023]，CC A [Li

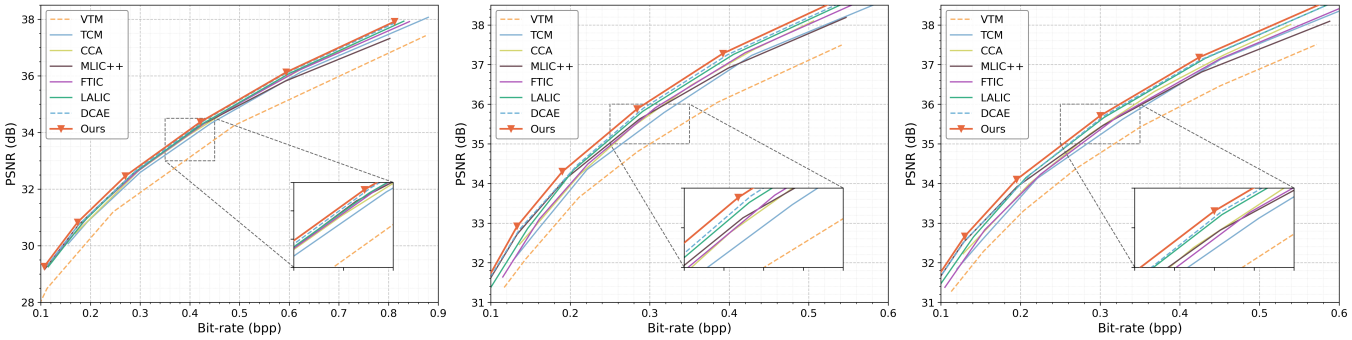


图 6：率失真 曲线在三个基准数据集上的表现：Kodak、Tecnick 和 CLIC专业验证集（从左到右 t).

模型	相对于 VTM-12.0 的 BD-Rate (%)				延迟（毫秒）			参数量（M）	GFLOPs
	Kodak		Tecnick	CLIC	总计	Enc.	Dec.		
	PSNR	多尺度结构相似性	PSNR	PSNR					
TCM（CVPR 2023）	-11.97	–	-11.95	-11.96	220	120	100	75.9	700.65
MLIC+ (ACM MM 2023)	-13.19	–	-17.47	-16.45	–	–	–	–	–
MLIC++ (ICMLW'23)	-15.09	–	-18.68	-16.84	235.4	106.5	128.9	116.7	615.93
FTIC (ICLR'24)	-14.84	-54.30	-15.24	-13.58	6846.6	1727.7	3391.9	69.78	245.46
CCA (NeurIPS'24)	-13.84	–	-15.34	-14.67	110	72	38	64.9	615.93
LALIC (CVPR'25)	-15.49	-54.00	-18.50	-18.08	210	143.9	66.1	66.13	303.18
DCAE (CVPR'25)	-16.83	-55.66	-21.28	-19.59	128	63	65	119.4	426.92
HiDE（本文方法）	**-18.50**	**-56.49**	**-24.01**	**-21.99**	134	66	68	134.9	447.64

表1：与最先进压缩方法的比较。相对于 VTM-12.0 报告 BD-Rate。延迟、参数量和 GFLOPs 均在使用一块 NVIDIA RTX 4090 GPU 的 Kodak 数据集上测得。

等，2024], FTIC [Han 等，2024], LALIC [Feng等,2025], 以及 DCAE [Lu 等，2025]。

图6展示了所有数据集上的率失真（RD）曲线，表1报告了相应的 BD-Rate [Bjontegaard、 2001] 节省率和模型复杂度指标。总体而言，HiDE 在三个基准测试中始终取得最低的 BD-Rate，以显著优势超越 DCAE 及其他竞争方法。在 Tecnick（1K）和 CLIC（2K）等高分辨率数据集上，这一性能优势尤为明显，凸显了层次化先验建模在捕获全局结构和精细纹理方面的益处。在计算效率方面，HiDE 仅以参数量和 GFLOPs 的小幅增加为代价实现了这些性能提升，同时保持了相当的延迟。

4.3 消融实验

为考察各个提出组件的贡献，我们针对基于分层词典的交叉注意力（HD）和上下文感知参数估计（CaPE）模块开展消融实验。所有消融模型均基于 DCAE [Lu等人, 2025]构建，并在由 VOC [Everingham 等人，2010] 和 DIV2K [Agustsson 和 Timofte，2017]提供的 1.4 万张图像组成的精简数据集上训练。为确保公平比较，每个模型均以 8 的批量大小训练 300 个 epoch，并在 Kodak 数据集上进行评估。

模型	BD-Rate (%)	参数量 (M)
+HD	-1.35	139.4
+CaPE	-2.82	114.4
分层字典 + CaPE（HiDE）	-3.81	134.9

表 2：在 Kodak 数据集上对分层字典（HD）和上下文感知参数估计（CaPE）进行消融实验。DCAE [Lu 等， 2025] 用作基线。

所提出组件的有效性

表 2 量化了各模块的影响。将 DCAE 中的单层词典替换为我们的分层设计（+HD），相较于 DCAE，BD-rate 降低了 1.35%，这表明将外部先验分解为全局字典和细节字典能够缓解表征竞争。用所提出的 CaPE 模块（+CaPE）替代标准固定尺度估计器，可进一步将压缩效率提升 2.82%，同时将参数量从 119.4M 降至 114.4 M。如图 2 所示，与 DCAE 的扁平结构相比，我们的分层字典利用更加均衡，证实了表征多样性有所提升。结合这两个组件后，HiDE 实现了最大的整体增益 3.81%，验证了二者具有很强的互补性。

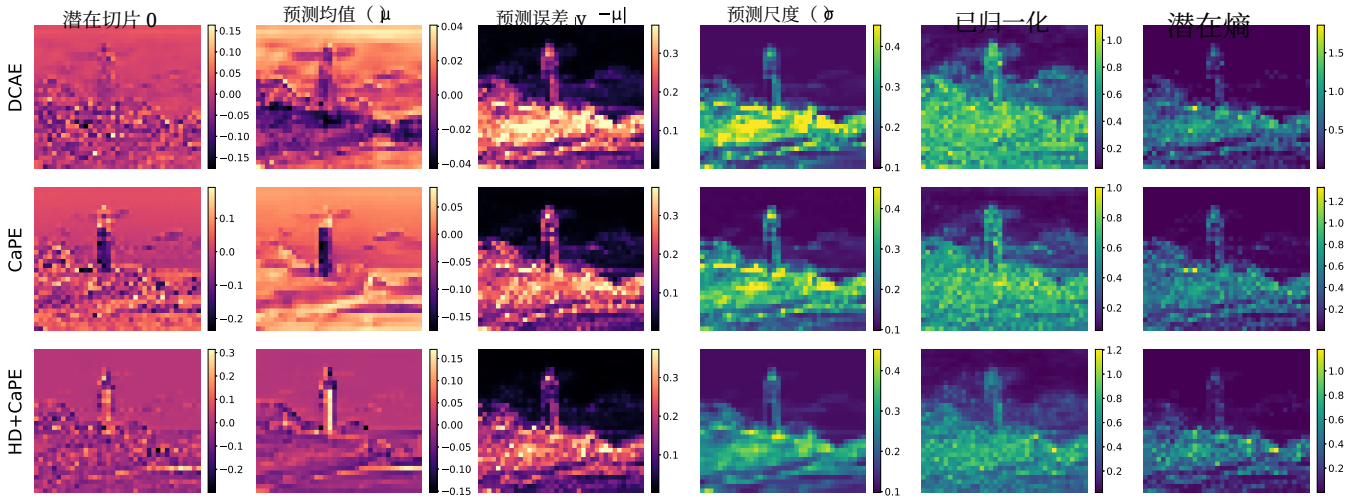


图 7：潜在表示和预测分布参数的可视化。从上到下依次为：基线 DCAE、仅 CaPE 变体和完整 HiDE（HD+CaPE）。从左到右依次为：潜在切片、预测均值、绝对预测误差（$y - \mu$）、预测尺度 σ、归一化残差（$y - \mu$）$/\sigma$以及潜在熵图。

G-D	BD-Rate (%)	BD-PSNR (dB)
32–96	-1.26	0.060
64–64	**-1.35**	**0.065**

表 3：在总字典容量固定的情况下，不同全局–细节（**G–D**）字典大小配置的影响。DCAE [Lu 等人，2025] 用作基线。

模型	BD-Rate（%）	参数量（M）	延迟（ms）
TCM	-5.3	45.2	93
TCM + CaPE	-5.6	55.0	94

表 4：将上下文感知参数估计（CaPE）模块应用于 TCM [Liu 等人的轻量版 2023] 框架时的泛化能力。结果以 [Cheng 等人的方案 2020] 为基准进行报告。

字典大小的影响

为分析全局字典和细节字典之间容量分配的影响，我们在保持条目总数不变的情况下调整两者的大小。如表 3 所示，两种配置的表现均优于基线 DCAE，BD-Rate 分别降低了 1.26% 和 1.35%。不同配置的表现相近，说明性能提升主要源自层次分解，而非特定的字典大小比例。因此，我们采用均衡配置（64–64）作为 HiDE 的默认设置。

参数估计的泛化能力

为评估 CaPE 的普适性，我们将其集成到 TCM [Liu 等人, 2023] 模型的小型版本中，以替换其原有的参数估计模块。如表 4 所示，CaPE 在延迟开销具有竞争力的同时，额外降低了 0.3% 的 BD-Rate，这表明其优势并不局限于基于字典的架构。尽管 TCM 的提升幅度小于 DCAE，但这符合预期，因为 TCM 仅依赖通道切片内部上下文，而 DCAE 这类基于字典的模型则能从上下文感知估计中获益更多，因为它们具有更丰富的上下文交互。

4.4 学习式图像压缩中的参数估计分析

图 7 展示了 DCAE、仅 CaPE 变体和 HiDE（HD+CaPE）预测的分布参数。我们展示了 Kodak 图像 *kodim21* 中熵最高的潜在切片。尽管不同模型的潜在映射 y 和预测均值 μ （第一列和第二列）在视觉上相似，但预测误差（$y - \mu$）（第三列）显示，HiDE 的残差幅度明显降低。这一改进伴随着更小的预测尺度 σ（第四列），反映出不确定性降低、估计更加可靠。我们还通过第五列中的归一化残差 $\frac{(y-\mu)}{\sigma}$ 评估建模能力。对于基线 DCAE，归一化域中仍存在较强的结构相关性，灯塔的轮廓清晰可见。相比之下，我们的方法显著减弱了这些结构依赖，表明空间去相关性有所增强。最后一列展示了比特率的空间分配情况。HiDE 生成的表示最为紧凑，凸显了准确参数估计对于优化编码效率的关键作用。

5 结论

本文提出了 **HiDE**，一种分层字典式熵建模框架，能够有效利用外部先验进行学习式图像压缩。HiDE 将外部先验组织为全局字典和细节字典，用于建模粗粒度结构模式和细粒度纹理，同时缓解表征冲突。级联检索

机制结合全局条件调节，确保语义一致性，并促进对外部先验的均衡利用。此外，本文引入了一种上下文感知参数估计网络，以克服单尺度卷积估计器的局限性，从而提高条件分布预测的准确性。实验结果表明，在多个基准数据集上，HiDE 始终优于现有最先进方法，验证了分层外部先验建模在高效熵估计方面的有效性。

参考文献

Eirikur Agustsson 和 Radu Timofte。NTIRE 2017 单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。

Nicola Asuni、Andrea Giachetti 等。Testimages：用于测试视觉设备和基础图像处理算法的大型档案库。第 63–70 页，2014年。

JohannesBall´e、Valero Laparra 和 Eero P.Simoncelli。端到端优化的图像压缩。载于第五届国际学习表示会议论文集，法国土伦，2017年4月。

Johannes Ball´e、David Minnen、Saurabh Singh、Sung Jin Hwang 和 Nick Johnston。采用尺度超先验的变分图像压缩。载于第六届国际学习表示会议论文集，温哥华，不列颠哥伦比亚省，加拿大，2018年4月。

Johannes Ball´e、Philip A. Chou、David Minnen、Saurabh Singh、Nick Johnston、Eirikur Agustsson、SungJin Hwang 和 George Toderici。非线性变换编码。*IEEE* 信号处理精选主题期刊，15(2):339–353，2021年。

Gisle Bjontegaard。率失真曲线之间平均PSNR差异的计算。*ITU-T*第*16*研究组，文档 *VCEG-M33*，2001年。

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* 电路与系统视频技术汇刊，31(10):3736–3764，2021年。

Zhengxue Cheng、Heming Sun、Masaru Takeuchi 和 Jiro Katto。采用离散高斯混合似然和注意力模块的学习式图像压缩。载于*2020*年*IEEE/CVF*计算机视觉与模式识别会议论文集，美国华盛顿州西雅图，第7936–7945页。计算机视觉基金会 / IEEE，2020年6月。

CLIC。学习式图像压缩和多类图像分类研讨会及挑战赛，2021年。

Humberto Ochoa Dominguez 和 Kamisetty Ramamohan Rao。《通用视频编码》。River publishers，2022年。

Patrick Esser、Robin Rombach 和 Bj¨orn Ommer。用于高分辨率图像合成的变换器驯化方法。载于*IEEE*计算机视觉与模式识别会议论文集，线上，第 12873–12883 页。计算机视觉基金会 / IEEE，2021年6月。

Mark Everingham、Luc Van Gool、Christopher K. I. Williams、John Winn 和 Andrew Zisserman。Pascal 视觉目标类别（VOC）挑战赛，2010年。Donghui Feng、Zhengxue Cheng、Shen Wang、Ronghua Wu、Hongwei Hu、Guo Lu 和 Li Song。面向学习式图像压缩的线性注意力建模。载于*IEEE/CVF*计算机视觉与模式识别会议论文集，美国田纳西州纳什维尔，第 7623–7632 页。计算机视觉基金会 / IEEE，2025年6月。Minghao Han、Shiyin Jiang、Shengxi Li、Xin Deng、Mai Xu、Ce Zhu 和 Shuhang Gu。面向学习式图像压缩的因果上下文调整损失。载于神经信息处理系统进展第 *38* 卷：*2024*年神经信息处理系统年会，加拿大不列颠哥伦比亚省温哥华，2024年12月。

Dailan He、Yaoyan Zheng、Baocheng Sun、Yan Wang 和 Hongwei Qin。用于高效学习式图像压缩的棋盘格上下文模型。载于《*IEEE*计算机视觉与模式识别会议论文集，线上》，第14771–14780页。计算机视觉基金会 / IEEE，2021年6月。

Dailan He、Ziming Yang、Weikun Peng、Rui Ma、Hongwei Qin 和 Yan Wang。ELIC：采用不均匀分组空域-通道上下文自适应编码的高效学习式图像压缩。载于《*IEEE/CVF*计算机视觉与模式识别会议论文集，新奥尔良，路易斯安那州，美国》，第5708–5717页。计算机视觉基金会 / IEEE，2022年6月。

Wei Jiang 和 Ronggang Wang。MLIC++：用于学习式图像压缩的线性复杂度多参考熵建模。国际机器学习大会 *2023*年 神经压缩研讨会：从信息论到应用，2023年。

Wei Jiang、Jiayu Yang、Yongqi Zhai、Peirong Ning、Feng Gao 和 Ronggang Wang。MLIC：用于学习式图像压缩的多参考熵模型。载于论文集 *f*

第*31*届 *ACM* 多媒体国际会议，加拿大安大略省渥太华，第7618–7627页。ACM，2023年。

伊士曼柯达。Kodak 无损真彩色图像套件。数据集可从 *https://r0k.us/graphics/kodak/* 获取，1993 年。

Ivan Krasin、Tom Duerig、Neil Alldrin、Vittorio Ferrari、Sami Abu-El-Haija、Alina Kuznetsova、Hassan Rom、Jasper Uijlings、Stefan Popov、Andreas Veit 等。OpenImages：一个用于大规模多标签和多类别图像分类的公共数据集。数据集可从 *https://github.com/openimages* 获取，2(3):18，2017年。

Haowei Kuang、Wenhan Yang、Zongming Guo 和 Jiaying Liu。基于掩码补偿信息的学习式图像压缩跨粒度在线优化。载于*IEEE/CVF* 国际计算机视觉会议论文集，第 16514–16523 页，2025年10月。

Han Li、Shaohui Li、Wenrui Dai、Chenglin Li、Junni Zou 和 Hongkai Xiong。用于学习式图像压缩的频率感知变换器。载于第 *12* 届国际学习表征会议论文集，奥地利维也纳，2024年5月。

Jinming Liu、Heming Sun 和 Jiro Katto。采用混合 Transformer-CNN 架构的学习式图像压缩。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，加拿大不列颠哥伦比亚省温哥华，第 14388–14397 页。计算机视觉基金会 / IEEE，2023年6月。

Jingbo Lu、Leheng Zhang、Xingyu Zhou、Mu Li、Wen Li 和 Shuhang Gu。基于字典的熵模型的学习式图像压缩。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，纳什维尔，田纳西州，美国，第 12850–12859 页。计算机视觉基金会 / IEEE，2025年6月。

David Minnen 和 Saurabh Singh。用于学习式图像压缩的逐通道自回归熵模型。载于*IEEE* 国际图像处理会议论文集，阿布扎比，阿拉伯联合酋长国，第 3339–3343 页。IEEE，2020年10月。

David Minnen、Johannes Ball´e 和 George Toderici。用于学习式图像压缩的联合自回归与分层先验。载于神经信息处理系统进展第*31*卷：*2018*年神经信息处理系统年会论文集，*NeurIPS 2018*，*2018*年*12*月*3*日至*8*日，加拿大蒙特´利尔，第 10794–10803 页，2018年。

David Minnen、George Toderici、Saurabh Singh、Sung Jin Hwang 和 Michele Covell。用于学习式图像压缩的图像依赖型局部熵模型。载于《*2018*年*IEEE*国际图像处理会议论文集》，希腊雅典，第430–434页。IEEE，2018年10月。

Aaron Van Den Oord、Oriol Vinyals 和 Koray Kavukcuoglu。神经离散表示学习。载于《神经信息处理系统进展第*30*卷：神经信息处理系统年会论文集》，美国加利福尼亚州长滩，第6306–6315页，2017年。

Gregory K. Wallace。JPEG静态图像压缩标准。《*ACM*通讯》，34(4):30–44，1991年。

Siqi Wu、Yinda Chen、Dong Liu 和 Zhihai He。用于深度图像压缩的条件潜编码与可学习合成参考。载于《*AAAI-25*论文集》，由人工智能促进协会主办，美国宾夕法尼亚州费城，第12863–12871页。AAAI出版社，2025年。

Fanhu Zeng、Hao Tang、Yihua Shao、Siyu Chen、Ling Shao 和 Yan Wang。Mambaic：用于高性能学习式图像压缩的状态空间模型。发表于*IEEE/CVF* 计算机视觉与模式识别会议论文集，纳什维尔，田纳西州，美国，第 18041–18050 页，2025年6月。

Zhisheng Zhong、Hiroaki Akutsu 和 Kiyoharu Aizawa。用于深度图像压缩的通道级可变量化网络。载于第二十九届国际人工智能联合会议论文集，第467–473页。ijcai.org，2020年。

Yinhao Zhu、Yang Yang 和 Taco Cohen。基于 Transformer 的变换编码。载于 第十届国际学习表征会议，线上活动，2022年4月。

Yongxin Zhu、Bocheng Li、Yifei Xin、Zhihua Xia 和 Linli Xu。用一个线性层解决向量量化模型中的表示坍塌问题。发表于*IEEE/CVF* 国际计算机视觉会议论文集，第 22968–22977 页，2025年。