统一的基于感兴趣区域的图像压缩范式与广义高斯模型

Kai Hu、Junfu Tan、Fang Xu、Ramy Samy、Yu Liu

arXiv:2602.01325v1 [eess.IV] 1 Feb 2026

摘要——**基于感兴趣区域（ROI）的图像压缩根据不同区域的语义重要性不均匀地分配比特。这种差异化编码通常会产生尖峰且具有重尾特征的分布。从数学上讲，这种分布特性需要使用形状参数可调的概率模型进行准确描述。然而，现有方法通常使用高斯模型拟合该分布，导致编码性能下降。为系统分析该分布对感兴趣区域编码的影响，我们构建了统一率失真优化理论范式。在此范式基础上，我们提出一种新颖的广义高斯模型（GGM），以实现对潜变量分布的灵活建模。为支持 GGM 的稳定优化，我们引入了有效的可微函数，并进一步提出动态下界，以缓解训练-测试不匹配。此外，我们引入有限差分，以解决 GGM 拟合分布后的梯度计算问题。在 COCO2017 上的实验表明，我们的方法在 ROI重建和下游任务（例如分割、目标检测）中均达到了最先进的性能。此外，与经典概率模型相比，我们的 GGM 能够更精确地拟合特征分布，并实现更优的编码性能。项目页面：https://github.com/hukai-tju/ROIGGM。**

索引词—**基于感兴趣区域的图像压缩、广义高斯模型、动态下界、训练-测试不匹配、梯度近似估计**

I. 引言

THE 数字时代带来了高分辨率视觉数据前所未有的激增。这些内容中有许多存在内部冗余，真正关键的内容只占很小一部分。感兴趣区域（ROI）图像压缩不像传统全局压缩那样均匀分配比特，而是在固定比特预算下，为信息区域分配更多比特，为不太重要的区域分配更少比特。传统标准（如 JPEG2000）通过小波变换和优化截断嵌入式块编码（EBCOT）算法实现这一原则，从而使感兴趣区域编码系统化 [1]。同样，基于ROI的神经图像压缩（NIC）利用数据驱动方法，实现更灵活、高效的编码。这一范式推动图像压缩从 均匀保真度转向语义驱动、按需保真 ，这对于对局部信息完整性有严格要求的应用尤为重要，例如医学诊断、视频监控与会议、自动驾驶感知以及遥感监测。

Kai Hu、Junfu Tan 和 Yu Liu 任职于中国天津市 300072 的天津大学（邮箱：kaihu@tju.edu.cn）。Fang Xu 任职于中国重庆的西南大学。Ramy Samy 任职于埃及空间技术中心通信部。通讯作者：*Yu Liu*、*Fang Xu*（邮箱：liuyu@tju.edu.cn、2011204017@tju.edu.cn）。

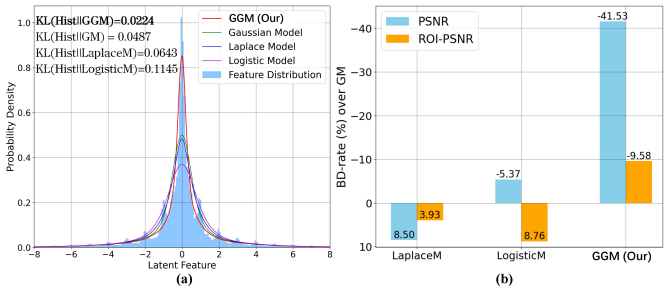


图 1. (a) 潜在特征直方图与拟合分布的比较，以及对应的 Kullback-Leible (KL) 散度值。我们的 GGM 能够更准确地拟合潜变量的尖峰重尾分布，达到最低 KL 散度 0.0224。(b) 各概率模型相较于 GM 基线所实现的 BD-rate 节省。对于整幅图像和目标区域，我们的 GGM 在 BD-rate 方面均实现了最佳编码性能，显著优于经典概率模型。

近年来，NIC凭借其出色的编码性能 [2]–[6] 迅速发展。Ballé等人提出的端到端学习框架 [2], 极大推动了该领域的发展，这类框架通常由分析变换、合成变换、量化器和熵模型组成。他们后续的工作 [7] 引入了尺度超先验网络，利用条件高斯先验对空间依赖关系建模。在此基础上，Minnen等人 [3] 采用自回归空间熵模型，利用因果上下文进行预测。随着这些关键进展，基于感兴趣区域的神经图像压缩也已成为一个重要且快速发展的研究方向 [8]–[13]。Li等人 [8], [12] 和 Mentzer等人 [9] 通过生成显式重要性图来引导自适应比特分配。Cai等人 [11]在此基础上 [10] ，通过点积将ROI掩码应用于多尺度特征表示。为了解决量化前掩码的性能局限，近期研究 [14]–[16]提出了更精细的技术。Li等人 [14]在量化前将潜在特征分解为基本层和增强层，并在增强层中应用显著性掩码，有选择地将背景区域置零。我们近期的工作 [15], [16]实现了联合隐式比特分配和双解码器，无需依赖显式的量化前掩码，即可实现高质量的区域自适应编码。

然而，我们观察到，基于 ROI 的压缩中潜变量分布存在显著的空间异质性，感兴趣区域和非感兴趣区域的统计特性似乎存在根本差异。这种分布特性亟需采用形状参数可变的分布进行数学描述。如图1 所示，全局潜变量分布具有尖峰和

厚尾特性。零值附近的突出峰值主要由背景特征构成，而两侧的尾部几乎完全由感兴趣区域特征主导。现有基于ROI 的压缩方法 [13]–[15] 通常假设采用单一概率模型（例如高斯分布）。这种简化假设无法准确捕捉潜变量分布的内在特性，进而导致熵编码效果欠佳，并降低整体压缩性能。一些神经图像压缩方法 [17]–[19] 采用了更灵活的先验分布，以突破单一高斯模型（GM）的局限。此类方法包括高斯混合模型（GMM） [17]和高斯–拉普拉斯–逻辑斯谛混合模型（GLLMM） [18]。Zhang 等人 [19]采用广义高斯模型（GGM），并设计了复杂程度各异的形状参数，以增强对自然图像分布的建模能力。然而，这些方法并非针对感兴趣区域编码而设计，忽略了其固有的空间异质性。此外，混合模型 [17] [18] 往往计算复杂度较高，而所采用的 GGM [19] 则忽视了优化尺度和形状参数对于准确建模潜变量分布的重要性。

ROI压缩中的潜变量分布由率失真（RD）函数和网络架构共同决定。因此，本文首先构建了一个统一率失真优化（RDO）理论范式。该范式不仅整合了以往针对ROI编码的设计选择（例如隐式/显式比特分配策略 [12], [15]和区域差异化损失），还着重关注概率模型。作为GM的扩展，GGM [20] 提供了两个近似正交的控制参数——尺度 α和形状 β。基于上述框架，我们进一步提出一种新型GGM，充分利用学习得到的 α 和学习得到的 β ，以准确建模复杂的空间异质分布。随后，我们针对各参数的特性设计了专用且可微的激活函数：用于 β 的*Softplus* 函数，以及用于 α的类 *Huber* 函数。这些函数不仅有助于稳定优化GGM，还融入了有意义的信息，从而提高潜在分布建模的准确性。训练与测试不匹配主要源于训练期间的加性均匀噪声与测试期间的硬舍入之间的差异。此外，我们注意到，较小的 α 值会加剧训练-测试不匹配，并降低编码性能 [21]。为缓解这一问题，我们针对 α引入了一种动态下界策略。此外，我们还针对GGM [22] 中正则化下不完全伽马函数 $P(a,\ b)$的梯度计算在解析上的难处理性，提出了一种高效的有限中心差分方案，从而实现稳定的端到端优化。

我们的主要贡献总结如下：

- 我们为感兴趣区域编码构建了统一的率失真优化理论范式。该范式聚焦于核心的比特分配策略、概率模型和率失真函数。
- 基于上述范式，我们为基于感兴趣区域的图像压缩引入了一种精细的 GGM。实验结果表明，我们的 GGM 对潜变量分布的拟合比经典概率模型更为精确——即高斯分布、拉普拉斯分布和逻辑斯谛模型及其混合模型——同时取得了更优的编码性能。
- 为解决GGM训练过程中的稳定性优化问题，我们分别为学习得到的尺度 α 和形状 β设计了有效且可微的激活函数。此外，我们提出了一种动态下界，以缓解尺度 α 的训练-测试不匹配。
- 为解决GGM中 $P(a,\ b)$函数梯度计算涉及的复杂数值积分问题，我们设计了一种高效且稳健的梯度近似估计方案，从而实现端到端的稳定优化。

本文其余部分安排如下。第II节简要回顾感兴趣区域压缩以及图像压缩概率模型的相关工作。第III节为感兴趣区域编码建立率失真优化理论范式。第IV节详细介绍我们的GGM。最后，第V节和VI 节分别给出实验结果和结论。

II. 相关工作

A. 基于感兴趣区域的图像压缩

Tahoces等人 [1] 针对JPEG2000标准采用了一种选择性系数掩码移位编码方法，以选择性地偏移不同分辨率子带的ROI系数，从而实现对感兴趣区域与背景之间相对质量的精细控制。随着卷积神经网络的发展，基于ROI的数据驱动图像压缩已成为新的研究重点。Li等人 [8], [12] 为关键区域引入了映射子网络，以生成显式的重要性映射，从而根据图像内容自适应地分配比特。Mentzer等人 [9] 同样使用重要性图来指导比特分配。基于 [10],Cai等人 [11] 使用ROI掩码对多尺度表示进行点积，这主要使更多比特能够分配给图像中的前景。Ma等人 [13]采用三维掩码对潜在表示进行剪枝，以改善背景区域的比特分配，并采用生成对抗网络对整个编码网络进行监督训练。Li等人 [23] 将ROI掩码集成到压缩网络的不同层中，通过修改拉格朗日乘子 λ 在不同区域实现空间适应性与可变码率压缩。Li等人 [14]开发了一种面向显著性分割的深度压缩模型，采用显著性分割模型生成掩码。针对不同图像区域的比特分配，他们在量化之前将潜在特征分解为基本层和增强层，并应用掩码将增强层中的背景区域局部置零，随后使用双尺度熵模块进行熵编码。

Peng等人 [24] 采用显著性图引导的图像压缩网络来处理机器视觉编码任务。Jin等人 [25] 提出了一种可定制的基于感兴趣区域的深度图像压缩范式，并融入了用户提供的文本信息。他们提出的框架既支持可定制的ROI压缩，也支持用户调整ROI与非ROI重建之间的质量权衡。Hu等人 [15], [16] 将隐式比特分配和双解码器应用于基于ROI的神经图像压缩模型，实现了高质量的区域自适应编码。它还

证实隐式方法比显式比特分配更加灵活，且效果更好。

B. 图像压缩的概率分布模型

概率分布模型是准确预测潜在特征分布的基础，这对于实现高效压缩至关重要。Ballé等人 [2] 使用分段线性函数，以非参数概率模型对每个量化潜在张量的边缘分布进行建模，从而避免由人工先验引起的模型假设偏差。Ballé等人 [7]引入尺度超先验，将潜变量张量建模为条件高斯尺度模型（GSM），从而能够有效建模潜变量张量之间的空间依赖关系。Minnen等人 [3]将GSM扩展为GMM，同时预测每个潜变量张量的均值与尺度参数。随后使用自回归模型对潜变量张量的因果上下文进行建模，以提高编码性能。Cheng等人 [4] 使用多组划分高斯混合模型对潜变量张量的分布进行建模，而非单一高斯分布，从而提供更精细的概率建模。每个潜变量张量由多个加权高斯分量组成，能够灵活建模多峰和非对称等复杂分布。Fu等人提出了一种结合高斯分布、拉普拉斯分布和逻辑分布的更灵活的混合模型， [18] 以准确拟合所有图像或图像内不同区域的特征分布。

Peng等人 [26] 使用广义高斯分布进行点云属性压缩，并通过形状参数 β 控制该分布，从而更准确地拟合潜在张量。同时，还采用动态似然区间来改进现有方法在算术编码方面存在的问题。为在压缩性能与复杂度之间实现更好的平衡，Zhang 等人 [27]将带有均值与尺度参数的高斯混合模型扩展为GGM。他们还设计了不同复杂度的形状参数，并采用了一些改进的训练方法，以实现自然图像的高质量编码。Jiang等人 [28] 使用不同的概率模型，准确地对屏幕内容图像中的高频和低频分量分布进行建模。

III. 感兴趣区域编码范式

本节将系统分析基于感兴趣区域的图像压缩中的率失真优化。基于感兴趣区域编码的特征分布特性，我们提出了一种建立在GGM基础上的统一理论范式。

A. 空间异质性约束下的率失真优化

感兴趣区域编码中潜在特征的分布并非固有属性，而是由率失真目标函数和网络架构共同塑造的 [11], [12], [15]。给定卷积神经网络针对输入图像 $\mathbf{x}$生成的二值重要性图 $\mathbf{M} \in \{0,1\}^{H\times W}$，其中 $M_i = 1$ 表示像素 i 属于感兴趣区域，我们对感兴趣区域和非感兴趣区域中的重建误差分别施加不同的惩罚因子。一种简单而有效的加权策略是

$$w_i = \begin{cases} w_{\text{roi}}, & M_i = 1. \\ w_{\text{nonroi}}, & M_i = 0. \end{cases}$$ 在约束 $w_{\text{roi}} > w_{\text{nonroi}}$下，这种加权方式能够实现差异化的内容重建。令 ϕ 和 θ 分别表示条件编码器 $g_a(\mathbf{x}, \mathrm{F}(\mathbf{M}); \phi)$ 和解码器 $g_s(\hat{\mathbf{y}}; \theta)$ 的参数。这里，$\mathrm{F}(\mathbf{M})$ 是作用于重要性图 $\mathbf{M}$的变换，用于对编码器内的潜在特征 $\mathbf{y}$ 实施特定的比特分配策略（例如隐式/显式比特分配 [11], [12], [15]）。

在这些约束下，编码器 g_a 被驱动生成高方差、信息丰富的潜在特征，以降低感兴趣区域中的加权失真；同时在背景区域生成低方差、接近零的特征，以降低比特率，如图 1所示。这种空间差异化优化使全局潜在特征分布呈现出显著的空间异质性，以及尖峰和重尾特性。基于标准变分自编码器（VAE）的压缩框架通常假设，潜变量的后验分布$q_\phi(\hat{\mathbf{y}}|\mathbf{x})$可以由参数化的高斯先验$p_\theta(\hat{\mathbf{y}}) \sim \mathcal{N}(\mu, \sigma^2)$ [29]–[31]近似。然而，在感兴趣区域加权失真的影响下，真实的后验分布 $q_\phi(\hat{\mathbf{y}}|\mathbf{x})$会发生严重的形状变形 [15]。因此，直接采用高斯先验会在KL散度中产生不可约的下界，即理论上的变分间隙。例如，如图1(a)所示，使用高斯混合模型时，KL散度为0.0487，而我们的GGM仅为0.0224。

B. 统一率失真优化范式

为弥合这一变分间隙，我们引入GGM作为先验，并将其与率失真函数和比特分配策略相结合，从而建立面向ROI压缩的统一率失真优化理论范式。根据 [7], [32], 通用率失真优化函数可通过引入拉格朗日乘子 λ来表示：

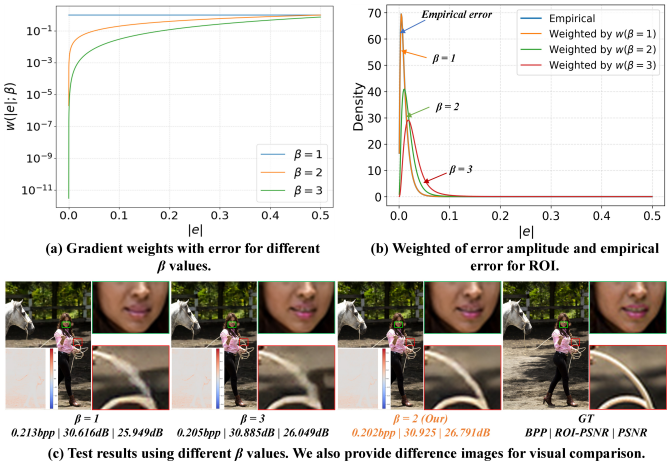
$$\mathcal{L}(\phi,\theta) = \underbrace{\mathbb{E}_{q_\phi(\hat{\mathbf{y}}|\mathbf{x})}[-\log p_\theta(\mathbf{x}|\hat{\mathbf{y}})]}_{Distortion} + \lambda \cdot \underbrace{D_{KL}[q_\phi(\hat{\mathbf{y}}|\mathbf{x})||p_\theta(\hat{\mathbf{y}})]}_{Rate}, \tag{1}$$

其中 $p_\theta(\mathbf{x}|\hat{\mathbf{y}})$表示似然（解码器 g_s输出分布）， $p_\theta(\hat{\mathbf{y}})$ 表示先验概率模型。$D_{KL}[\cdot||\cdot]$ 表示KL散度，$\hat{\mathbf{y}}$ 为量化结果。我们力求利用GGM推导失真项和码率项，以处理感兴趣区域编码中不同的统计特性。失真项对应于给定潜变量条件下解码结果的负对数似然。为刻画感兴趣区域编码中潜在的分布特性，我们将解码器的输出分布 $p_\theta(\mathbf{x}|\hat{\mathbf{y}})$建模为均值 $g_s(\hat{\mathbf{y}}; \theta)$、尺度 α'和形状 β'的GGM：

$$p_\theta(\mathbf{x}|\hat{\mathbf{y}}) = \frac{\beta'}{2\alpha'\,\Gamma(1/\beta')} \exp\left(-\left|\frac{\mathbf{x}-g_s(\hat{\mathbf{y}};\theta))}{\alpha'}\right|^{\beta'}\right), \tag{2}$$

其中 $\Gamma(\cdot)$ 是伽马函数。具体而言，我们假设α' 和 β' 均为常数。式（$\mathcal{D}$ （1）中的失真项定义如下：

$$\begin{aligned}\mathcal{D} &= \mathbb{E}_{q_\phi(\hat{\mathbf{y}}|\mathbf{x},\mathcal{F}(M))}\left[-\log\frac{\beta'}{2\alpha'\,\Gamma(1/\beta')} + \left|\frac{\mathbf{x}-g_s(\hat{\mathbf{y}};\theta)}{\alpha'}\right|^{\beta'}\right] \\ &= \mathbb{E}_{q_\phi(\hat{\mathbf{y}}|\mathbf{x},\mathcal{F}(\mathbf{M}))}\left[\frac{|\mathbf{x}-g_s(\hat{\mathbf{y}};\theta)|^{\beta'}}{\alpha'^{\beta'}}\right] + \log\frac{2\alpha'\,\Gamma(1/\beta')}{\beta'},\end{aligned} \tag{3}$$



(a) Gradient weights with error for different β values.

(b) Weighted of error amplitude and empirical error for ROI.

(c) Test results using different β values. We also provide difference images for visual comparison.

图 2. 形状参数 β 的分析。(a) 失真项 D 的有效梯度权重 $\omega(|e|; \beta) = \beta|e|^{\beta-1}$。设定 $\beta = 1$ 时权重为常数； $\beta = 2$ 会产生线性缩放； $\beta = 3$ 则会放大较大误差，形成二次放大效应。(b) COCO2017 上的重建误差谱。$\beta = 1$ 对较大误差惩罚不足，导致残差呈重尾分布； $\beta = 3$ 过度关注离群值； $\beta = 2$ 则实现了最佳平衡。(c) 率失真性能。采用 $\beta = 2$ 的模型能以更低的比特率实现更优的重建质量。

因此，在优化过程中，第二项可以视为常数。此外，因子 $\alpha'^{-\beta'}$ 可以并入拉格朗日乘子 λ。因此，感兴趣区域编码的最小失真 D 近似为：

$$\mathcal{D} = \mathbb{E}_{q_\phi(\hat{\mathbf{y}}|\mathbf{x},\mathcal{F}(\mathbf{M}))}\left[\sum_{i=1}^{N} w_i \, |x_i - g_{i,s}(\hat{\mathbf{y}};\theta)|^{\beta'}\right], \tag{4}$$

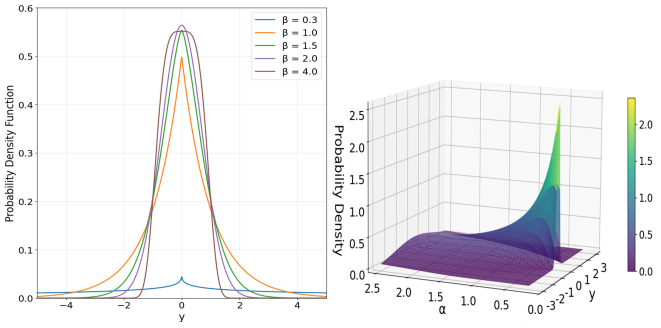
完整的式（1）可表示为

$$\begin{aligned}\mathcal{L}(\phi,\theta) = {} & \mathbb{E}_{q_\phi(\hat{\mathbf{y}}|\mathbf{x},\mathcal{F}(\mathbf{M}))}\left[\sum_{i=1}^{N} w_i \, |x_i - g_{i,s}(\hat{\mathbf{y}};\theta)|^{\beta'}\right] \\ & + \lambda \cdot \mathbb{E}_{q_\phi(\hat{\mathbf{y}}|\mathbf{x},\mathcal{F}(\mathbf{M}))}\left[\log q_\phi(\hat{\mathbf{y}}|\mathbf{x},\mathcal{F}(\mathbf{M})) - \log p_\theta(\hat{\mathbf{y}})\right],\end{aligned} \tag{5}$$

根据最大熵分布原理 [33]—该原理指出，在均值和方差固定的条件下，高斯分布的熵最大—我们在 GGM 的失真项 D 中设定 $\beta' = 2$。我们还尝试使用$\beta' = 1$、2 和 3 进行分析（图 2），发现 $\beta' = 2$ 可实现最佳率失真性能。根据 [7], [17]，码率项通过添加均匀噪声来放宽量化限制 $U(-\frac{1}{2}, \frac{1}{2})$，从而得到 $q_\phi(\hat{\mathbf{y}}|\mathbf{x}, \mathbf{M}) = \prod_i U(\hat{\mathbf{y}}_i|\mathbf{y}_i - \frac{1}{2}, \mathbf{y}_i + \frac{1}{2})$。因此，$\log q_\phi(\hat{\mathbf{y}}|\mathbf{x}, \mathbf{M})$ 为常数，可以省略。

因此，我们推导出基于 GGM 的最终统一率失真优化编码范式如下：

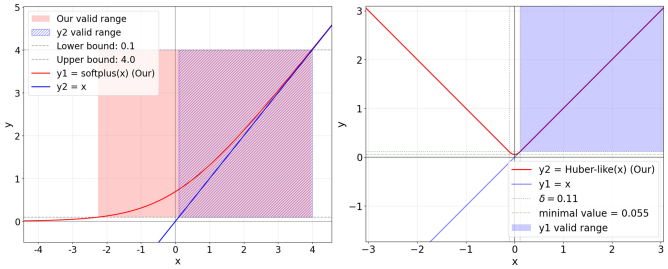
$$\begin{aligned} & p_\theta(\hat{\mathbf{y}}) = \prod_i \left[\mathcal{N}_\beta(\mu_{\hat{\mathbf{y}},i}, \alpha^\beta_{\hat{\mathbf{y}},i}) * U(-\frac{1}{2}, \frac{1}{2})\right](\hat{\mathbf{y}}_i), \\ & \mathcal{L}(\phi,\theta) = \underbrace{\mathbb{E}_{q_\phi(\hat{\mathbf{y}}|\mathbf{x},\mathcal{F}(\mathbf{M}))}\left[\sum_{i=1}^{N} w_i \, \|x_i - g_{i,s}(\hat{\mathbf{y}};\theta)\|^2\right]}_{Distortion} \\ & \quad - \lambda \cdot \underbrace{\mathbb{E}_{q_\phi(\hat{\mathbf{y}}|\mathbf{x},\mathcal{F}(\mathbf{M}))}\left[\log p_\theta(\hat{\mathbf{y}})\right]}_{Rate}, \end{aligned} \tag{6}$$



(a) The shape of the Probability Density Function varies with β (α =1.0).

(b) The shape of the Probability Density Function varies with α (β =2.0).

图 3. 在均值固定为 $\mu = 0$ 时，尺度参数（α）和形状参数（β）对 GGM 概率密度函数的影响。



(a) Different functions used for the shape β

(b) Different functions used for the scale α

图 4. 用于尺度 α 和形状 β 的函数。形状 β 和尺度 α 的下界需要为正值。已知我们的函数在定义域内可微，且有效输入范围更广。

该框架整合了通用比特分配和 GGM 下的区域自适应损失，从而实现更优的感兴趣区域编码。

IV. 我们的方法

A. 经典广义高斯模型

GGM 记为 $\mathbf{Y} \sim \mathrm{N}\beta(\mu, \alpha^\beta)$，它通过引入形状参数 β 扩展了标准 GM。这个额外参数使 GGM 能够自适应地对具有不同尾部行为的分布进行建模，因此特别适合捕捉潜变量张量的复杂统计特性，如图 3 所示。对称 GGM 的概率密度函数（PDF）为 [20], [34] $f_\beta(\mathbf{y}) = \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left(-\left(\frac{|\mathbf{y}-\mu|}{\alpha}\right)^\beta\right)$。当 $\mu = 0$ 和 $\alpha = 1$ 时，~~PDF 的~~标准形式~~及其~~对应的累积分布函数（CDF）为：

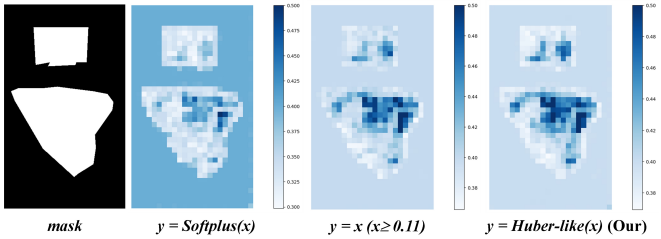
$$\begin{aligned} f_\beta(\mathbf{y}) &= \frac{\beta}{2\alpha\Gamma(1/\beta)} \exp\left(-|\mathbf{y}|^\beta\right), \\ c_\beta(\mathbf{y}) &= \int_{-\infty}^{\mathbf{y}} f_\beta(u)du = \frac{1}{2} + \frac{\mathrm{sgn}(\mathbf{y})}{2} P\left(\frac{1}{\beta}, |y|^\beta\right), \end{aligned} \tag{7}$$

其中，$c_\beta(\mathbf{y})$ 是 GGM 的累积分布函数，而 $P(\cdot,\cdot)$ 表示正则化下不完全伽马函数 [22]。不完全伽马函数 $\gamma(a,b)$ 和伽马函数$\Gamma(a)$ 定义如下

$$\begin{aligned} P(a,b) &= \frac{\gamma(a,b)}{\Gamma(a)}, \\ \Gamma(a) &= \int_0^\infty t^{a-1}e^{-t}dt; \gamma(a,b) = \int_0^b t^{a-1}e^{-t}dt, \end{aligned} \tag{8}$$



(a) 使用不同函数得到的形状参数 β 特征图。



(b) 使用不同函数得到的尺度参数 α 特征图。

图 5. 形状参数 β 和尺度参数 α 使用不同函数得到的特征图。

这里，我们设定 $a = \frac{1}{\beta}$ 和 $b = |\mathbf{y}|^{\beta}$。GGM 由两个近似正交的可学习参数表征：尺度α 和形状 β。这种双参数表述是其区别于高斯分布和拉普拉斯分布等单参数模型的关键所在。尺度 α 控制分布的离散度（宽度），并直接影响熵：较小的 α 表示更高的可预测性和更低的比特率，而较大的α 则意味着更大的不确定性和更高的比特率。形状参数 β 决定峰度，影响尖峰程度和尾部厚度。经验上，复杂纹理区域用较小的 β 建模效果更好，而较平滑的区域则对应较大的 β。值得注意的是，GGM 将高斯分布（$\beta = 2$，其中 $\sigma = \alpha/\sqrt{2}$）和拉普拉斯分布（$\beta = 1$）作为其特例。

B. 我们的概率分布模型

如图 1所示，基于 ROI 的压缩会产生空间异质的潜在分布。来自感兴趣区域（细节丰富）的潜在特征往往服从尖峰、重尾分布，而来自非感兴趣区域（较平滑）的潜在特征则更集中，呈轻尾分布。这种分布特性从数学上要求采用具有可变形状参数的分布来准确描述。为对此类异质性建模，我们引入了一种精细的 GGM，其中包含学习得到的尺度（α）和形状（β）参数，从而实现精确的分布拟合。然而，要稳定有效地优化 α 和 β 并非易事，但这对实现高质量编码至关重要。为解决这一挑战，我们为 α 和 β精心设计了优化策略。

形状参数 β： 为确保形状参数 β保持为正且整个模型可微，我们引入了*Softplus* 函数。

$$f(\beta) = Softplus(\beta) = \log(1 + e^{\beta}) \tag{9}$$

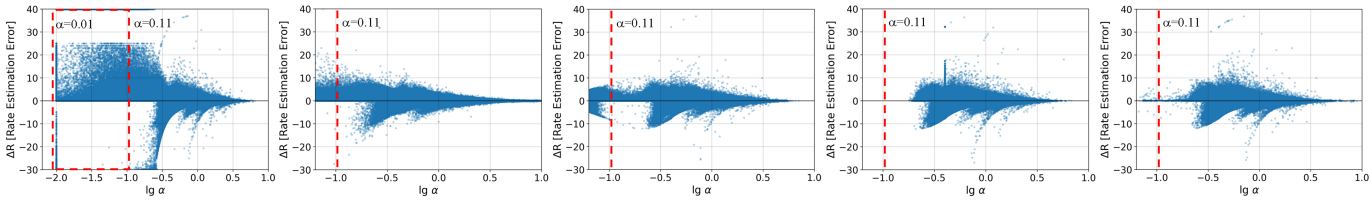
该*Softplus* 函数将输入映射到正数域 (0,∞)。与线性函数 $y = x$不同，其导数是一个 *Sigmoid*函数，可确保梯度平滑且稳定。如图4所示，*Softplus*函数显著扩大了可学习 β 值的有效范围，同时在优化过程中避免梯度爆炸或梯度消失，从而有助于实现稳定训练并学习有意义的分布形状。此外，按照 [27], ，我们将 *Softplus* 函数的输出限制在区间 [0内。1, 4]。这一实用边界使 β 能够更好地捕捉本任务中潜在特征分布的稀疏性和尖峰程度特征。图 5a展示了该函数的效果，其中比较了使用不同函数得到的 β 特征图。与简单的线性函数相比，我们的*Softplus*函数能够更准确地描绘前景和背景区域中的高频信息，从而提高潜在空间的建模精度。

尺度参数 α： 为确保尺度参数 α 保持正值且可微，并避免取值过小，我们设计了一种类*Huber*激活函数，其灵感源自Huber损失 [36]。

$$f(\alpha) = \begin{cases} \frac{\alpha^2}{2\delta} + \frac{\delta}{2} & \text{if } |\alpha| \le \delta \\ |\alpha| & \text{if } |\alpha| > \delta \end{cases}$$

根据 [21], [29], [37], 我们设定 $\delta = 0。11$。该函数在 $|\alpha|$ 较小时呈二次函数特性，在 $|\alpha|$较大时呈线性特性（见图4），因此在整个参数范围内都能提供稳定的梯度。相比之下，先前工作中用于高斯混合模型的线性函数$y = x$（[21], [29], [37]）缺乏这种自适应特性。如图 α 所示，尺度在接近零时高度敏感。图6对此进行了说明。虽然*Softplus* 函数在训练期间会使较小的 α 值趋近于零，从而加剧训练-测试不匹配，但我们的类*Huber*函数有效缓解了这一问题，并提高了优化稳定性。此外，图5b 展示了由不同函数生成的 α 特征图。我们观察到，*Softplus* 函数过度强调极小值（例如背景区域中的值），因而忽视了目标区域的重要性。相比之下，线性函数$y = x$无法充分映射关键目标区域。我们的函数有效平衡了前景与背景的重要性，同时确保完整映射目标区域，从而能够准确建模潜在空间分布。

尺度的动态下界 α： 训练期间使用加性均匀噪声，而测试期间采用硬舍入，两者之间的差异会导致训练-测试不匹配。尽管失真项中的这种不匹配通常会得到处理，但码率项中的不匹配依然存在。当尺度 α 变得非常小时，这一问题会更加严重，如图6(a)-(c)所示。测试期间，区间 [−0. 5，0.5] 上的概率质量趋近于 1.0，误导模型认为对符号“0”进行编码几乎不需要比特。然而，在训练期间，模型需要估计整个连续区间的比特率，因而系统性地高估码率。此外，基于 ROI 的压缩中的潜在分布呈现出强烈的空间异质性。这种结构性差异带来了独特的优化挑战，超出了先前界限所处理的一般训练-测试不匹配问题（例如，高斯模型的固定界限 [21], [29], [37]，或理论推导的 β依赖于 [27] GGM 的界限。具体而言，若没有针对这种异质性设置合适的约束，α 与 β 的联合优化就会变得不稳定。如图8(a)所示，许多较小的 α 值错误地聚集在 β相对较大的区域中。这种不适定的参数耦合不仅加剧了码率估计误差，更重要的是，还会降低



(a) When the lower bound of the scale α is 0.01. (b) When the scale α no *β -dependent dynamic lower bound.* (c) When the scale α adopts the function *y = softplus(x).* (d) When the scale α adopts the function *y = x (x ≥ 0.11).* **(e) When the scale α adopts the function *y = Huber-like(x).(Our)***

图6. 码率估计误差 $\Delta R = R_{train} - R_{test}$ 作为尺度 α 的函数，由100幅图像和 8×10^6 个样本计算得到（Pascal VOC2012 [35]）。虽然 ΔR 通常较小且集中在零附近，但当 α 非常小时（例如 < 0.11），它会变得很大且为正，表明存在严重的训练时码率高估。

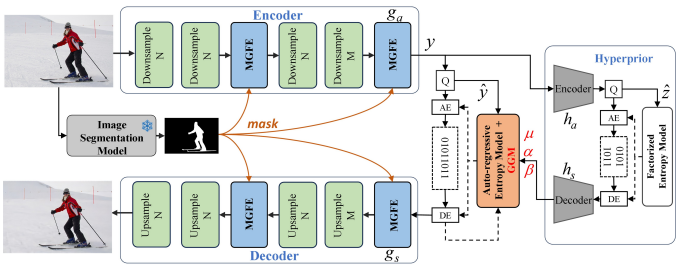
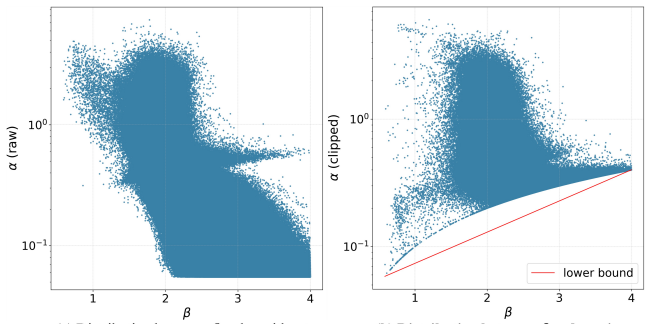


图7. 我们的整体框架。MGFE表示掩码引导特征增强模块，如 [16], 所示，用于执行隐式比特分配和区域自适应特征增强。AE和AD为算术编解码器。

模型准确表征前景和背景区域各自统计特性的能力，从而限制整体编码性能。



(a) Distribution between β and α without dynamic lower bound optimization. **(b) Distribution between β and α using β-dependent dynamic optimization.**

图8. 尺度 α的动态下界示意图。这些分布主要来自Pascal VOC2012数据集中的100幅图像和一组 1×10^7 个样本。

为应对这种不稳定性，同时保持对分布形状的自适应，我们为 α引入动态下界，通过一个简单而有效的线性约束将其与 β 显式关联：

$$\alpha_{min} = \max\{\alpha, \zeta \cdot \beta\}, \tag{10}$$

其中， ζ 是一个超参数。此设计确保 α 允许的最小值不再是固定常数，而是根据 β所指示的分布形状进行调整。鉴于基于感兴趣区域的编码中的潜在空间分布通常具有尖峰和厚尾的特征，较小的 β （对应于尾部更厚的分布）允许相应地采用较低的 α，从而准确地建模聚集在零附近的值。相反，较大的 β 需要较高的 α，以防止 GGM 生成不切实际的狭窄分布，从而加剧训练-测试不匹配。如图 8（b）所示，应用此动态约束后， α 值与β呈正相关，过小的 α 值则得到有效抑制。此机制减少了训练测试差异，提升了整体编码性能。

C. 网络框架概述

基于式（6）中提到的率失真优化范式，我们开发了一种高效的基于感兴趣区域的图像压缩模型，如图 7所示。该模型采用 [16]作为网络骨干，但仅使用单个解码器同时重建图像的前景和背景，从而降低模型复杂度。原始图像 $\mathbf{x} \in \mathbb{R}^{H \times W \times 3}$ 首先输入编码器网络 g_a和预训练分割模型。掩码可由预训练分割模型生成，例如语义分割模型 [38]。为了利用生成的掩码 $\mathbf{M}$ 进行隐式比特分配，我们采用了 MGFE模块（详细说明见附录 C）。该模块能够精准区分前景与背景区域，从而实现高质量的区域自适应编码。为更准确地捕捉基于感兴趣区域的压缩中的潜在空间分布，我们引入 GGM 来取代高斯混合模型。

编码器网络 g_a表示的潜在特征 $\mathbf{y}$ 随后经超先验编码器 h_a进一步转换为超先验潜在特征 $\mathbf{z}$。特征 $\mathbf{y}$ 和 $\mathbf{z}$ 输入量化器 Q ，得到结果 $\hat{\mathbf{y}}$ 和 $\hat{\mathbf{z}}$。我们采用简单的因子化熵模型对量化后的 $\hat{\mathbf{z}}$进行编码。随后，解码后的 $\hat{\mathbf{z}}$ 输入超先验解码器 h_s ，以推导出三个参数，例如均值 μ、尺度 α 和形状 β。为便于使用 GGM 进行熵编码，我们对 $\lfloor \mathbf{y} - \mu \rceil + \mu$采用零中心量化。得益于隐式比特分配策略，量化潜在特征 $\hat{\mathbf{y}}$ 保留了完整的背景信息。最后，我们使用解码器 g_s 重建高质量图像 $\hat{\mathbf{x}}$。

D. 训练方法

对 $P(a, b)$ 函数进行可微优化： 通过梯度下降优化参数 μ α, β 时，$c_\beta(\mathbf{y})$ 的导数可以通过以下方式计算

$$\begin{aligned} \frac{\partial c_\beta(\mathbf{y})}{\partial \mathbf{y}} &= f_\beta(\mathbf{y}), \\ \frac{\partial c_\beta(\mathbf{y})}{\partial \beta} &= \frac{\operatorname{sgn}(\mathbf{y})}{2}\left(-\frac{1}{\beta^2}P' + \frac{|\mathbf{y}|\ln|\mathbf{y}|}{\Gamma(1/\beta)}e^{-|\mathbf{y}|^\beta}\right), \end{aligned} \tag{11}$$

根据式（8），可以得到

$$P' = \frac{\partial P}{\partial a}\frac{da}{d\beta} + \frac{\partial P}{\partial b}\frac{db}{d\beta}, \tag{12}$$

其中，两个主要的导数解分别为 $\frac{\partial P}{\partial b} = \frac{b^{a-1}e^{-b}}{\Gamma_a}$ 和 $\frac{\partial P}{\partial a}$（$\frac{\partial}{\partial a}\frac{\gamma(a,b)}{\Gamma(a)}$）$da$。应用分数阶求导法则后，$\frac{\partial P}{\partial a}(a,b)$ 被分解为

$$\frac{\partial P(a,b)}{\partial a} = \frac{\frac{\partial \gamma(a,b)}{\partial a}\Gamma(a) - \gamma(a,b)\frac{d\Gamma(a)}{da}}{\Gamma(a)^2}, \tag{13}$$

为了计算该表达式，我们需要导数 $\partial\gamma(a,b)/\partial a$ 和 $d\Gamma(a)/da$。$d\Gamma(a)/da = \Gamma(a)\psi(a)$，其中 $\psi(\cdot)$ 是双伽马函数。对于伽马函数 $\gamma(a,b)$，$\frac{\partial \gamma(a,b)}{\partial a} = \int_0^b t^{a-1}e^{-t}\ln(t)dt$。该积分不存在闭式解，这给稳定且高效的数值计算带来了挑战 [22]。因此，在 PyTorch 等自动微分框架中，函数 $P(a,b)$ 关于参数 a 不可微。

为实现基于梯度的稳定优化，我们避免直接进行数值积分。相反，我们提出一种高效的中心差分格式，用于近似 $\gamma(a, b)$ 的梯度，从而恢复训练过程中的可微性。

$$\frac{\partial}{\partial a}\gamma(a,b) \approx \frac{\gamma(a+\epsilon,b) - \gamma(a-\epsilon,b)}{2\epsilon}, \tag{14}$$

步长为 $\epsilon = 1\times10^{-5}$。具体而言，我们的有限差分法具有恒定的 $O(1)$ 复杂度，而数值积分通常需要 $O(N)$ 次运算，其中 N 是积分步数（详细的误差界分析见附录B）。GGM 的有限差分梯度计算详见算法1。

训练损失：基于 ROI 的神经图像压缩本质上是一个率失真优化问题。其目标是最小化码率项与失真项的加权和，并分别针对前景和背景进行评估。根据式（6），我们任务中的码率项可表示为：

$$\mathcal{R} = \mathbb{E}_{p(\mathbf{x})}\mathbb{E}_{q_\phi(\hat{\mathbf{y}},\hat{\mathbf{z}}|\mathbf{x})}[-\log p_\theta(\hat{\mathbf{y}}|\hat{\mathbf{z}}) - \log p_\theta(\hat{\mathbf{z}})], \tag{15}$$

其中，$p_{\hat{\mathbf{y}}}$ 和 $p_{\hat{\mathbf{z}}}$ 分别量化潜变量 $\hat{\mathbf{y}}$ 和 $\hat{\mathbf{z}}$ 的概率分布。相应地，失真项 D 分解为前景分量 D_{FG} 和全局分量 D_G。因此，我们的总体损失函数定义为

$$\mathcal{L} = (\mathcal{D}_{FG} + k_1\cdot\mathcal{D}_G + \mathcal{L}_{b_rec}) + \lambda\cdot\mathcal{R}, \tag{16}$$

L_{b_rec} 表示全局图像的深度感知损失。

为实现高保真的前景重建，同时保持令人满意的背景质量，我们采用区域特定的损失函数。失真项定义如下

$$\begin{aligned}\mathcal{D}_{FG} &= \mathrm{MSE}_{ROI}(\mathbf{x}\times\mathbf{M}, \hat{\mathbf{x}}\times\mathbf{M}),\\ \mathcal{D}_G &= \text{MS-SSIM}(\mathbf{x}, \hat{\mathbf{x}}),\end{aligned} \tag{17}$$

其中，MSE_{ROI} 表示感兴趣区域内的均方误差损失。为契合人类视觉感知并保留细节，

算法 1通过有限差分计算 GGM 的梯度

要求： 输入潜变量张量$\mathbf{x}$；形状参数 β；有效区间 $(\beta_{\min}, \beta_{\max})$；最小值 ε；步长 ϵ_{fd}

函数 前向$(\mathbf{x}, \beta)$ // 计算累积分布函数
  // 确保其为正数且处于有效范围内
  $\beta \leftarrow$ 截断 (Softplus(β), $\beta_{\min}$, $\beta_{\max}$)
  $x_{\text{abs}} \leftarrow$ 最大值$(|x|, \varepsilon)$
  $a \leftarrow 1/\beta$; $b \leftarrow x_{\text{abs}}^{\beta}$ // 参数变换
  累积分布函数 $\leftarrow \frac{1}{2}(1 +$符号函数$(x)\cdot P(a, b))$ // 输出累积分布函数
  保存 $(x_{\text{abs}}, \beta, a, b)$ // 用于反向传播
  返回 累积分布函数
结束函数
函数 反向传播(梯度_输出) // 计算 x, β 梯度
  加载 $(\mathbf{x}_{\text{abs}}, \beta, a, b)$
  $\Gamma_a \leftarrow \exp(\text{lgamma}(a))$ // 伽马函数值
  $\frac{\partial P}{\partial b} \leftarrow \frac{b^{a-1}e^{-b}}{\Gamma_a}$; $\frac{db}{dx} \leftarrow \beta\mathbf{x}^{\beta-1}$ 绝对值
  // 输出 x 梯度
  grad_$\mathbf{x} \leftarrow$ 梯度_输出 $\cdot \frac{\text{sign}(\mathbf{x})}{2}\left(\frac{\partial P}{\partial b}\cdot\frac{db}{dx}\right)$
  $v_+ \leftarrow P(a+\epsilon_{\text{fd}}, b)\cdot\exp(\text{lgamma}(a+\epsilon_{\text{fd}}))$
  $v_- \leftarrow P(a-\epsilon_{\text{fd}}, b)\cdot\exp(\text{lgamma}(a-\epsilon_{\text{fd}}))$
  // 通过中心有限差分进行近似
  $\frac{\partial\gamma(a,b)}{\partial a} \approx \frac{v_+ - v_-}{2\epsilon_{\text{fd}}}$
  $\frac{\partial P}{\partial a} \leftarrow \frac{1}{\Gamma_a}\frac{\partial\gamma(a,b)}{\partial a} - P(a,b)\,\psi(a)$
  $\frac{da}{d\beta} \leftarrow -\beta^{-2}$; $\frac{db}{d\beta} \leftarrow b\cdot\log(\mathbf{x}_{\text{abs}})$
  $\frac{\partial P}{\partial\beta} \leftarrow \frac{\partial P}{\partial a}\frac{da}{d\beta} + \frac{\partial P}{\partial b}\frac{db}{d\beta}$
  grad_$\beta \leftarrow$ grad_out $\cdot \frac{\text{sign}(\mathbf{x})}{2}\cdot\frac{\partial P}{\partial\beta}$ // 输出 β 梯度
  返回(grad_$\mathbf{x}$, grad_β)
结束函数

全局损失 $\mathcal{D}_G$ 整合了多尺度结构相似性（多尺度结构相似性）损失 [39]。此外，我们还引入感知损失 $\mathcal{L}_{per}$和风格损失 $\mathcal{L}_{sty}$，遵循 [40] 以提升输出图像的自然度和一致性：

$$\mathcal{L}_{b_rec} = k_2\cdot\mathcal{L}_{per} + k_3\cdot\mathcal{L}_{sty}, \tag{18}$$

其中超参数设置为 $k_1 = 100$, $k_2 = 0.02$，且为$k_3 = 50$。

V. 实验

A. 实验详情

实验数据。我们使用 COCO2017 数据集 [43],开展实验，该数据集提供训练集、验证集和测试集。我们从约 118,000 张图像的训练集中随机选取 15,000 张图像，并将每张图像裁剪为 256×256 像素用于训练。通过在像素级对实例分割标注进行二值化，生成相应的前景掩码。对于训练和测试，我们选取前景占图像面积 8% 至 80% 的掩码-图像对，最终从包含 5,000 张图像的验证集中得到 3,443 对用于评估。

为评估泛化能力，我们还采用了 HRSOD 数据集 [44], ，这是一个用于高分辨率显著目标检测的基准数据集，包含 8,398 张训练图像和 2,100 张测试图像（分辨率介于 1200 至 2560 像素之间）。按照类似的筛选流程，我们在相同的前景比例范围 [0.08,0.8] 内随机抽取 115 对掩码-图像对。

表 I平均 BD-rate（%）节省、BD-PSNR（dB）增益，以及相较于 Zou2022 [37] 在 COCO2017 和 HRSOD 数据集上的 BD-mAP（%）增益。符号（“-”）表示 BD-rate 增益优于 Zou2022，而对于 BD-mAP 和 BD-PSNR 增益则表示相反。

方法	COCO2017 数据集						HRSOD 数据集			
	感兴趣区域峰值信噪比		分割准确率		检测准确率		PSNR		感兴趣区域峰值信噪比	
	BD-rate ↓	BD-PSNR ↑	BD-rate ↓	BD-mAP ↑	BD-rate ↓	BD-mAP ↑	BD-rate ↓	BD-PSNR ↑	BD-rate ↓	BD-PSNR ↑
Zou2022 [37]	–	–	–	–	–	–	–	–	–	–
Ballé2018 [7]	136.02	-3.416	75.44	-3.215	76.76	-2.995	150.63	-4.488	152.08	-4.608
BPG(YUV444)	44.16	-1.343	78.30	-3.215	75.44	-3.674	38.17	-1.209	40.57	-1.325
Cheng2020 [4]	19.32	-0.666	-3.29	0.042	-2.55	-0.049	-1.65	0.054	-0.18	-0.001
通用视频编码 [41]	5.95	-0.261	26.25	-1.199	24.08	-1.330	13.11	-0.464	15.05	-0.504
Liu2023 [5]	-6.55	0.260	5.55	-0.241	-0.57	-0.261	-10.40	0.422	-9.49	0.399
Liu2024 [42]	-10.52	0.403	-14.84	0.636	-14.61	0.699	-6.04	0.231	-4.92	0.195
Hu2025 [16]	-81.98	5.897	-89.63	5.032	-87.49	5.725	-10.14	0.309	-86.94	7.531
Our	**-85.00**	**6.294**	**-99.78**	**5.443**	**-99.99**	**6.298**	**-28.46**	**0.949**	**-89.89**	**7.900**

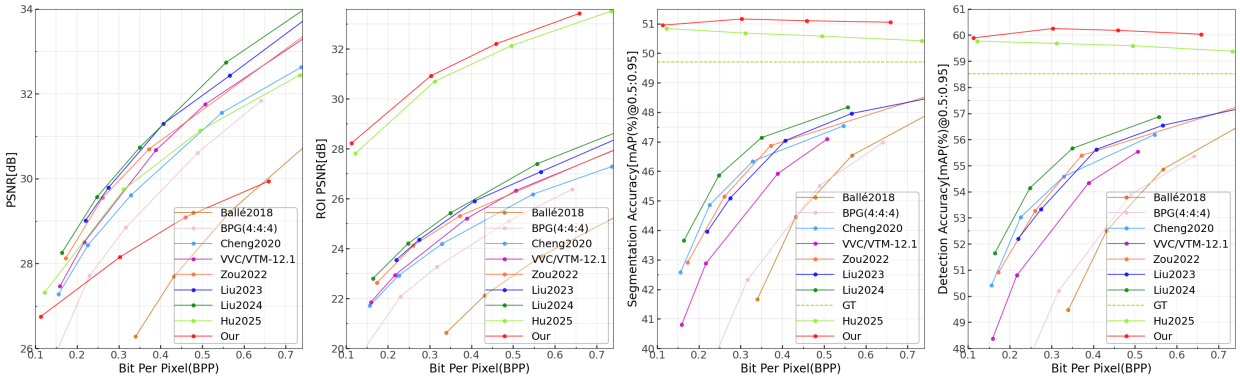


图9. COCO2017 数据集上 PSNR、感兴趣区域峰值信噪比和 mAP@0.50:0.95 的平均率失真性能　　t.

实验设置。 在我们的实验中，不同模型采用的量化策略有所不同：GMM 和 GLLMM 使用非零中心量化（$\lfloor y \rceil$），而包括 GGM 在内的所有其他概率模型均使用零中心量化（$\lfloor \mathbf{y} - \mu \rceil + \mu$）。所有模型均使用 Adam 优化，采用均方误差损失，批量大小为 8，并训练 450 轮。我们通过改变拉格朗日乘子 $\lambda \in \{32, 64, 128, 512\}$，针对多个码率点训练模型。学习率初始为$1 \times 10^{-4}$，并在最后 100 轮衰减至 1×10^{-5}。压缩网络中的通道数 N 和 M 分别设为 192 和 320。超参数 ζ 设为 0.1。我们的实现基于 CompressAI PyTorch 库 [45]，，实验在单个 NVIDIA V100 GPU 上进行。

我们的性能基准包括神经图像压缩方法（Ballé2018 [7], Cheng2020 [4], Zou2022 [37], Liu2023 [5],Liu2024 [42]）、经典编码器（BPG [46], VVC/VTM-12.1 [41]）以及近期的感兴趣区域方法 Hu2025 [16]。我们使用率失真曲线评估重建质量，以每像素比特数（BPP）为横轴，分别绘制整体峰值信噪比和感兴趣区域峰值信噪比。基于 Bjøntegaard-Delta方法 [47]，使用 BD-rate 和 BD-PSNR 指标对性能进行定量比较。除了像素保真度，我们还通过评估目标检测和实例分割性能，衡量压缩图像在机器视觉（MV）中的效用。平均精度均值（mAP）采用交并比阈值 0.50:0.95，作为主要的机器视觉指标。我们绘制每像素比特数与 mAP 的关系，并采用类似的 BD-mAP 和 BD-rate 指标进行定量比较。

B. 压缩性能

为验证压缩性能，我们绘制了各种方法的率失真曲线。如图 9所示，我们的方法在感兴趣区域中实现了最佳率失真性能，优于近期最先进方法（Hu2025 [16]）和传统神经图像压缩方法。这一优势可归因于我们精心设计的 GGM 和隐式比特分配方法。值得注意的是，由于采用单解码器框架专注于感兴趣区域的保真度，我们的方法在整幅图像上的性能相对较低。相比之下，Hu2025 采用双解码器架构，在前景和背景之间实现了更均衡的重建。图 10提供了视觉对比。我们的方法采用感知损失与多尺度结构相似性损失联合优化，重建图像保留了较高的视觉质量。尤其是，感兴趣区域内的重建效果明显优于其他方法，且最接近原始未压缩图像（真实值）。虽然 Hu2025 方法也展现出良好的保真度，但其对高频纹理的还原以及色彩准确度仍不够精确。

为了进行定量评估，我们采用 BD-rate 和 BD- PSNR 指标，并以 Zou2022 [37] 作为基线（表I）。

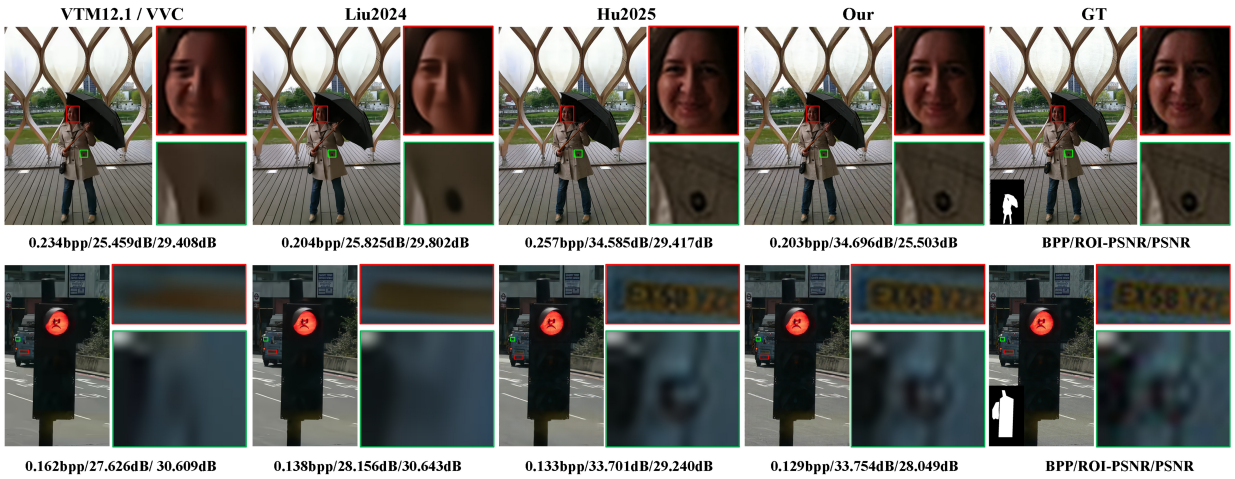


图 10. COCO2017 数据集上使用不同方法的压缩结果视觉对比。我们还在 GT 图像上叠加了掩码，以直观比较重要区域的重建效果。

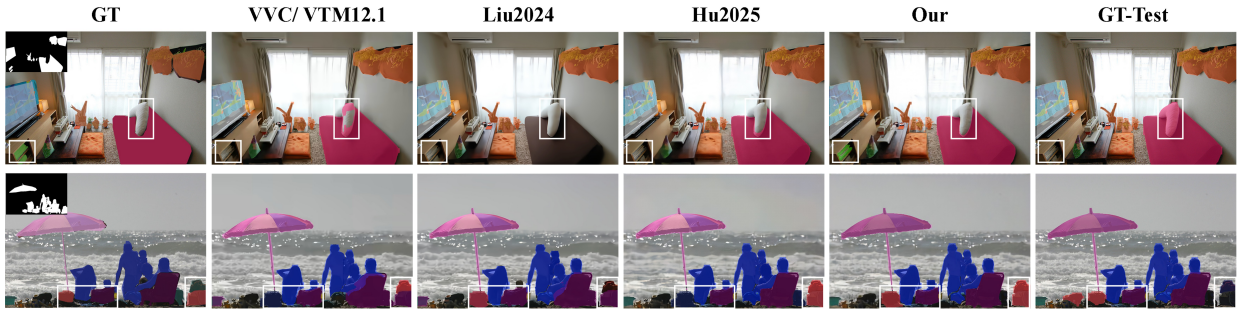


图 11. 使用不同压缩方法在 COCO2017 上得到的分割结果。白框标示出主要错误（分割错误/漏检）。GT-test 表示原始图像经过测试后的分割结果。GT 图像包含其标注掩码，以供参考。

我们的方法使 BD-rate 降低了 85%，并使 BD- PSNR 提高了 6.294 dB，显著优于所有对比方法。相较于我们的方法，Hu2025 的 BD-rate 降幅低 2.25%，BD-PSNR 增益小 0.311 dB，进一步证实了我们的方法在感兴趣区域编码方面的有效性。基础方法 Balle2018 [7]在此评估中，BD-rate 增加了 136.02%，BD-PSNR 则下降了 3.416 dB。我们还在高分辨率 HRSOD 数据集 [44]（表I）上进一步评估了泛化能力。与 Zou2022 方法 [37]相比，我们的 GGM 方法再次取得最佳表现：全图的 BD-rate 降幅为 28.46%，目标区域的降幅为 89.89%。这主要归因于 HRSOD 数据集的分布特征：除前景物体外，背景区域中的高频纹理相对较少。这些结果一致表明，我们的 GGM 在基于 ROI 的压缩方面具有优越性和广泛的适用性。

C. 机器视觉性能

本研究的主要目标是提升压缩图像在多视图任务中的实用性。因此，我们评估了不同压缩方法对两项具有代表性的多视图任务——目标检测和实例分割——的影响。鉴于 YOLOv11 [48]，性能出色且提供可用的预训练变体，我们将其用作多视图任务网络。

为了直观比较性能，我们绘制了两项任务的识别准确率与每像素比特数曲线，如图 9所示。我们的重建图像在不同比特率下均取得最高的平均精度均值，优于近期提出的基于感兴趣区域的方法 Hu2025 [16]，甚至优于 GT-test 图像。这一优势源于我们的方法能够在感兴趣区域内实现高保真重建，并抑制容易造成干扰的背景纹理，从而减少误识别和漏检。图 11 展示了不同压缩方法的分割结果。我们的方法生成了最准确的分割掩码，分类错误和漏检更少，同时对不同尺寸的目标均保持较高的准确率。尽管 Hu2025 方法和 GT-test 也取得了不错的效果，但它们出现了更明显的分割错误和漏检。

为进行定量比较，我们计算 BD-rate 和 BD-mAP 指标，并以 Zou2022 [37] 作为基线（表I）。对于分割任务，我们的方法实现了 99.78% 的 BD-rate 降低和 5.443% 的 BD-mAP 增益，显著超越所有其他方法。值得注意的是，其 BD-rate 降幅比 Hu2025 方法高出 10.15%，BD-mAP 增益高出 0.411%。同样，对于目标检测，我们的方法取得了 99.99% 的 BD-rate 降低和 6.298% 的 BD-mAP 增益，分别比 Hu2025 高出 12.5% 和 0.573%。这些结果一致表明，我们的方法最能

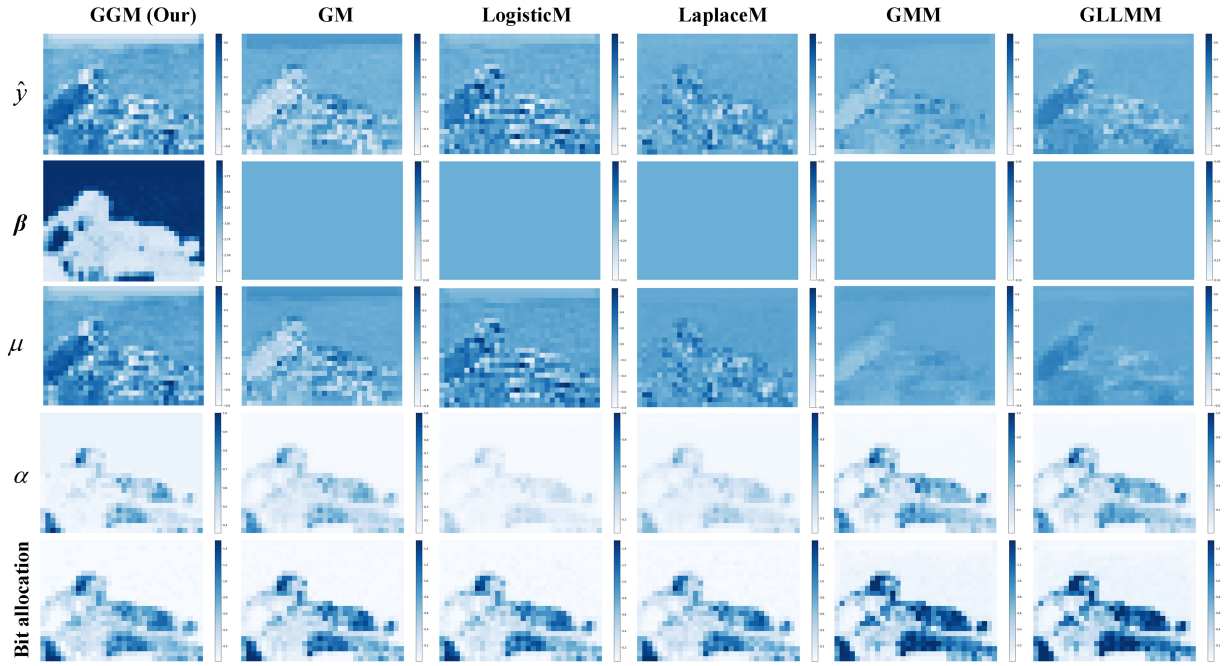


图 12. Pascal VOC2012 数据集上不同 PDM 的形状参数 β、均值 μ、尺度参数 α、量化潜在特征 $\hat{y}$ 及比特分配图的可视化结果。

保留对下游 MV 任务至关重要的信息。

D. 不同概率分布模型的比较

为分析不同概率分布模型（PDM）对编码性能的影响，我们通过将压缩网络中的 GGM 替换为几种经典 PDM，开展了对比实验：高斯混合模型（GM）、拉普拉斯模型（LaplaceM）、逻辑斯蒂模型（LogisticM）、GMM 和 GLLMM。我们从三个方面评估性能：PSNR、感兴趣区域峰值信噪比和模型复杂度。我们采用 GM 作为基线，计算并比较各模型在 BD-rate、BD-PSNR、参数数量和乘加运算（MACs）等指标上的压缩性能。实验结果（表 II）表明，我们的 GGM 通过学习得到的形状参数 β，能够准确地对潜在分布建模，并取得最佳编码性能。与 GM 基线相比，我们的 GGM 对整幅图像的 BD-rate 降低了 41.53 %，BD-PSNR 提升了 1.131 dB；对感兴趣区域的 BD-rate 降低了 9.58%，BD-PSNR 提升了 0.303 dB。LaplaceM 和 LogisticM 的性能略逊于 GM，表明它们对实际潜在分布的拟合效果较差。GMM 和 GLLMM 旨在通过混合分量对复杂分布建模。然而，尽管理论上具有近似复杂分布的能力，但在我们的特定任务中，它们面临优化困难。我们的潜在特征主要呈现尖峰重尾分布，导致混合模型在端到端训练期间的参数估计不稳定，无法有效收敛到最优分布。具体而言，GMM 和 GLLMM 的 BD-rate 分别增加了 419.83% 和 236.90%，BD-PSNR 分别降低了 8.323 dB 和 7.772 dB。就复杂度而言，我们的网络采用 GM 时有 40.599M 个参数，且每像素需要 173.712G MACs。GLLMM 的复杂度最高，增加了 6.395M 个参数和每像素 0.576G MACs，显著高于基于 GGM 的模型。

表 II COCO2017 数据集上不同 PDM 相较于高斯混合模型的平均性能提升。

PDMs	PSNR		感兴趣区域峰值信噪比		模型复杂度	
	BD-rate ↓	BD-PSNR ↑	BD-rate ↓	BD-PSNR ↑	△ 参数量(MB)	△ 每像素乘加运算量(G)
GM	–	–	–	–	40.599	173.712
LaplaceM	8.50	-0.194	3.93	-0.123	0.000	0.000
LogisticM	-5.37	0.129	8.76	-0.269	0.000	0.000
GMM	419.83	-8.323	139.12	-2.899	1.843	0.472
GLLMM	236.90	-7.772	119.77	-3.076	12.698	3.253
GGM（我们的方法）	**-41.53**	**1.131**	**-9.58**	**0.303**	**5.395**	**0.576**

我们进一步可视化不同概率分布模型（PDMs）的特征图、量化后的潜在特征$\hat{y}$和比特分配图（图12）。我们的 GGM 潜在特征保留了更丰富的信息，有助于实现更优的重建。学习得到的 β 参数能够准确地对分布进行建模。在比特分配图中（颜色深浅表示比特分配量：颜色越深 = 表示比特越多），基于 GGM 的比特分配图整体色调更浅，消耗的比特最少，并为许多非重要区域分配零比特或极少比特。总之，GGM 特别适用于基于感兴趣区域的图像压缩。它能够准确地对潜在分布的特征进行建模，同时保持较低的复杂度，从而提升编码网络的整体压缩性能。

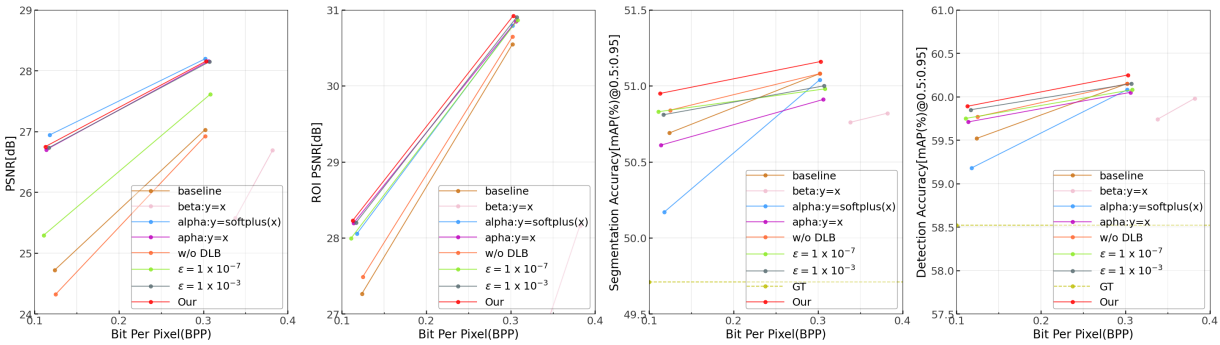


图13。 COCO2017 数据集上不同变体在峰值信噪比、感兴趣区域峰值信噪比和 mAP@0.50:0.95 方面的率失真性能 t.

E. 消融实验

我们针对方法的关键组成部分进行了消融研究，主要关注 GGM 的核心参数设置。对于形状 β，我们将其替换为简单的线性函数 y =x。对于尺度 α，我们测试了两种替代方案：(a) 采用线性函数 y = x，并设置固定下界 $x \geq 0.11$（遵循 [21] 中的 GM 策略）；(b) 采用 y = Softplus(x) 函数。我们还单独测试了移除尺度 α 的动态下界（DLB）。随后，我们建立了一个全面的基线：GGM 对 β 和 α 均采用 y = x，并移除 DLB。此外，我们改变用于 $P(a, b)$ 中心差分近似的相对步长 ϵ，将其设为 1×10^{-3} 和 1×10^{-7}，以考察其对梯度稳定性和最终性能的影响。

形状参数设置分析 β。形状参数设置的定量影响如表 III 所示。使用线性函数 $y = x$ 来设置 β 会降低重建质量，尤其是在感兴趣区域中。与综合基线相比，此设置使 BD-rate 增加了 162.77%，并使 BD-PSNR 降低了 4.731 dB。相比之下，我们的方法实现了 21.66% 的 BD-rate 节省和 0.791 dB 的 BD-PSNR 提升，性能显著更优。我们还可视化了它们的率失真曲线

表 III 使用 COCO2017 数据集时，GGM 中相关变体相对于基线变体的平均 BD-rate (%) 节省和 BD-PSNR (dB) 增益。

变体	PSNR		感兴趣区域峰值信噪比	
	BD-rate ↓	BD-PSNR ↑	BD-rate ↓	BD-PSNR ↑
基线	–	–	–	–
形状参数 β				
w $y = x$	68.11	-2.115	162.77	-4.731
尺度参数 α				
w $y = Softplus(x)$	-58.92	1.726	-15.93	0.588
w $y = x$	-55.01	1.605	-19.52	0.709
无 DLB	9.68	-0.264	-3.92	0.146
ϵ 对 P的取值(a, b)				
w $\epsilon = 1 \times 10^{-7}$	-24.71	0.680	-18.19	0.651
w $\epsilon = 1 \times 10^{-3}$	-54.58	1.594	-19.04	0.703
Our	**-56.45**	**1.642**	**-21.66**	**0.791**

见图13。当形状参数 β 采用线性函数$y = x$建模时，无法准确表示潜在分布，导致总体比特率大幅增加且率失真性能较差。图 5a 展示了由不同函数生成的 β 特征图。在这些图中，颜色较深的区域（较大的 β 值）对应以低频内容为主的区域，而颜色较浅的区域（较小的 β 值）则表示存在高频信息。由我们的 βSoftplus 函数生成的 *Softplus* 图能够准确描绘图像中的频率分布。得益于 Softplus 的特性，优化后的 β 能够有效识别并表示背景区域中存在的稀疏高频细节。相比之下，线性函数 β 生成的图 $y = x$ 并不准确，严重削弱了 GGM 精确建模潜在分布的能力。

尺度参数的设置分析 α。为评估尺度参数 α的影响，我们对不同函数进行了定量比较（表III）和率失真曲线对比（图13）。尽管线性函数 $y = x$ （下限为 0.11）和 *Softplus* 函数取得了有竞争力的全局峰值信噪比，但其感兴趣区域性能明显逊于我们的*Huber* 类函数。具体而言，与我们的函数相比，它们的 BD-rate 节省分别低 2.14% 和 5.73%。此外，使用这些替代函数重建的图像会导致下游 MV 任务的性能显著下降。图5b 展示了 α的特征图。更深的区域（更大的α）对应包含高频信息的区域，而较亮的区域（较小的 α）则表示低频内容。*Softplus* 函数往往会过度强调非常小的 α 值，尤其是在背景区域。这种偏差会妨碍关键区域的准确分布建模，并加剧训练-测试不匹配，从而降低整体性能。尽管线性函数 $y = x$ 通过固定下限避免数值过小，但也对 α的动态范围施加了严格限制。因此， α 无法灵活地适应真实潜在分布的建模。

相比之下，我们的类 *Huber*函数有效缓解了这两个问题。它避免了极小值带来的训练与测试不匹配，同时不限制 α 参数的优化范围。通过更准确地拟合潜变量分布，它在所有评估指标上始终展现出更优的编码性能。

动态下界设置的分析（针对 α）。为量化其对训练-测试不匹配的影响，我们对尺度参数 α 采用不同的下界，并计算码率估计误差 ΔR。如图 6所示，较小的下界（例如 0.01）会导致更明显的 ΔR。 *Softplus* 函数

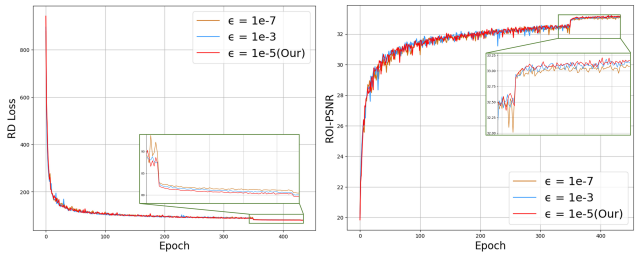


图 14。 ϵ 值对压缩网络优化和编码性能的影响。

以及不使用动态下界（“w/o DLB”）的情况，在 ΔR 低于约 0.11 时也会出现明显的 α，从而加剧失配问题。图 8展示了 α 值的分布。没有 DLB 约束时，α 值会逐渐降低，其最小值集中在形状参数较大的区域 β，从而加剧训练-测试不匹配。相比之下，应用我们的 DLB 可使 α 在整个 β范围内始终高于 0.11，从而有效缓解失配问题。性能影响汇总于表III和图 13中。缺少 DLB 会显著降低压缩性能。与完整方法相比，“w/o DLB”变体使全局图像的 BD-PSNR降低 1.906 dB，使 ROI 的 BD-PSNR降低 0.645 dB。此外，该变体在下游 MV 任务上的准确率也较低。

函数 ϵ 中的 $P(a,\ z)$值设置分析。我们评估了步长 ϵ 的不同取值（表 III、图 13 和 14）。结果表明，过大或过小的 ϵ都会降低率失真性能。较大的 ϵ（例如 1×10^{-3}）会导致梯度不准确，使 GGM 无法准确地对潜在分布进行建模。相反，过小的ϵ（例如 1×10^{-7}）会在梯度计算中引入数值不稳定性和噪声，同样会严重损害性能。适当选择的 ϵ 能够为 $P(a,\ b)$提供准确且稳定的梯度近似，使 GGM 能够有效捕捉潜在特征尖峰状和重尾特性。这反过来提升了感兴趣区域重建质量以及多视图任务的准确性。

总之，我们对形状 β （通过 *Softplus* 函数）和尺度 α （通过 类 *Huber* 函数及 DLB）所做的设计选择至关重要。二者共同实现了准确的潜在分布建模，缓解了训练-测试不匹配，并最终提升了压缩性能。

VI. 结论

本文提出了一种统一的基于感兴趣区域的图像压缩率失真优化理论范式。为克服高斯混合模型在建模感兴趣区域编码所特有的尖峰和重尾分布方面的局限性，我们基于该理论范式提出了一种精细的 GGM，其学习得到的形状为 β，尺度为 α。我们为 α 设计了专用的激活函数和动态下界，以确保优化过程稳定。此外，研究验证了有限中心差分方案是一种高效且具有足够稳定性的替代方法，可用于处理 GGM 中 $P(a,b)$函数梯度计算所涉及的复杂数值积分问题。在多个数据集上的实验结果表明，我们的方法能够重建高保真感兴趣区域，并在压缩和下游 MV 任务中均实现最优性能。与经典概率模型相比，我们的 GGM 在 BD-rate 和 BD-PSNR 方面取得了显著提升，展现出更高的编码效率。未来的工作将探索将 GGM 扩展至其他图像压缩任务。

参考文献

[1] P. G. Tahoces, J. R. Varela, M. J. Lado, and M. Souto,《图像压缩: JPEG2000 中的 Maxshift 感兴趣区域编码选项》，《计算机视觉与图像理解》，第 109 卷，第 2 期，第 139–145 页，2008 年。[2] J. Ballé, V. Laparra, and E. P. Simoncelli, 《端到端优化图像压缩》，arXiv 预印本 arXiv:1611.01704，2016 年。[3] D. Minnen, J. Ballé, and G. D. Toderici, 《学习式图像压缩的联合自回归先验与层次先验》，《神经信息处理系统进展》，第 31 卷，2018 年。[4] Z. Cheng and J. Katto, 《采用离散化高斯混合似然和注意力模块的学习式图像压缩》，载于 IEEE计算机视觉与模式识别会议研讨会（CVPRW）论文集，2020 年，第 7936–7945 页。[5] J. Liu, H. Sun, and J. 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（CVPR）论文集，2023 年，第 14 388–14 397 页。[6] S. Li, H. Li, W. Dai, C. Li, J. Zou, and H. Xiong, 《采用死区量化器的学习式渐进图像压缩》，《IEEE电路与系统视频技术汇刊》，第 33 卷，第 6 期，第 2962–2978 页，2023 年。[7] J. Ballé, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, 《带尺度超先验的变分图像压缩》，载于国际学习表征会议（ICLR）论文集，2018 年。[8] M. Li, W. Zuo, S. Gu, D. Zhao, and D. Zhang, 《学习用于内容加权图像压缩的卷积网络》，载于 IEEE计算机视觉与模式识别会议（CVPR）论文集，2018 年，第 3214–3223 页。[9] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. Van Gool, 《深度图像压缩的条件概率模型》，载于 IEEE计算机视觉与模式识别会议（CVPR）论文集，2018 年，第 4394–4402 页。[10] C. Cai, L. Chen, X. Zhang, and Z. Gao, 《采用多尺度分解网络的高效可变码率图像压缩》，《IEEE电路与系统视频技术汇刊》，第 29 卷，第 12 期，第 3687–3700 页，2018 年。[11] C. Cai, L. Chen, X. Zhang, and Z. Gao, 《端到端优化的感兴趣区域图像压缩》，载于《IEEE图像处理汇刊》，第 29 卷，2019 年，第 3442–3457 页。[12] M. Li, W. Zuo, S. Gu, J. You, and D. Zhang, 《学习内容加权深度图像压缩》，《IEEE模式分析与机器智能汇刊》，第 43 卷，第 10 期，第 3446–3461 页，2020 年。[13] Y. Ma, Y. Zhai, C. Yang, J. Yang, R. Wang, J. Zhou, K. Li, Y. Chen, and R. Wang, 《针对视觉质量优化的可变码率感兴趣区域图像压缩》，载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集，2021 年，第 1936–1940 页。[14] Y. Li, W. Gao, G. Li, and S. Ma, 《面向显著性分割、采用新颖比特分配的深度图像压缩》，《IEEE图像处理汇刊》，2024 年。[15] K. Hu, H. Wang, Y. Lyu, R. Liu, and Y. Liu, 《文本引导的基于感兴趣区域的神经图像压缩》，《IEEE广播汇刊》，第 1–10 页，2025 年。[16] K. Hu, H. Wang, R. Liu, Z. Li, S. Song, and Y. Liu, 《采用隐式比特分配的基于感兴趣区域的深度图像压缩》，2025 年。 [在线]。可查阅: https://arxiv.org/abs/ 2511 08918 。Z. [17] Cheng, H. Sun, M. Takeuchi, and J. Katto, 《采用离散化高斯混合似然和注意力模块的学习式图像压缩》，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2020 年，第 7936–7945 页。[18] H. Fu, F. Liang, J. Lin, B. Li, M. Akbari, J. Liang, G. Zhang, D. Liu, C. Tu, and J. Han, 《采用高斯-拉普拉斯-逻辑斯蒂混合模型和串联残差模块的学习式图像压缩》，《IEEE图像处理汇刊》，第 32 卷，第 2063–2076 页，2023 年。[19] H. Zhang, L. Li, and D. Liu, 《学习式图像压缩的广义高斯模型》，《IEEE图像处理汇刊》，第 34 卷，1965，2025 年 1950pp. – 。[20] T. Wang, H. Li, Z. Li, and Z. Wang, 《广义高斯分布的快速参数估计》，载于第 8 届国际信号处理会议论文集，第 1 卷。IEEE，2006 年。[21] H. Zhang, L. Li, and D. Liu, 《关于学习式图像压缩的均匀标量量化》，arXiv 预印本 arXiv:2309.17051, 2023 年。

[22] U. Blahak, 《云模型应用中不完全伽马函数的高效近似》, 《地球科学模型开发》, 第 3 卷, 第 2 期, 第 329–336 页, 2010 年。[23] B. Li, J. Liang, H. Fu, and J. Han, 《采用Swin Transformer的基于感兴趣区域的深度图像压缩》, 载于 IEEE国际声学、语音与信号处理会议（ICASSP）论文集, 2023 年, 第 1–5 页。[24] B. Peng, T. Lin, D. Jin, Z. Pan, and J. Lei, 《显著图引导的面向机器的端到端图像编码》, 《IEEE信号处理快报》, 第 31 卷, 第 1755–1759 页, 2024 年。[25] J. Jin, F. Xia, F. Ding, X. Zhang, M. Liu, Y. Zhao, W. Lin, and L. Meng, 《可定制的基于感兴趣区域的深度图像压缩》, 《IEEE电路与系统视频技术汇刊》, 2025 年。[26] C. Peng, 《采用动态似然区间的点云属性压缩广义高斯熵模型》, 载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集, 2025 年, 第 11 779–11 788 页。[27] H. Zhang, L. Li, and D. Liu, 《学习式图像压缩的广义高斯模型》, 《IEEE图像处理汇刊》, 2025 年。[28] S. Jiang, T. Ren, H. Yuan, J. Huo, and X. Lu, 《Omr-net+: 一种面向高效屏幕内容图像压缩的频率感知特征细化与熵建模方法》, 《IEEE信号处理快报》, 2025 年。[29] D. He, Z. Yang, W. Peng, R. Ma, H. Qin, and Y. Wang, 《Elic: 采用不均匀分组的空域-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（CVPR）论文集, 2022 年, 第 5718–5727 页。[30] Y. Kim, H. Jeong, J. Yu, Y. Kim, J. Lee, S. Y. Jeong, and H. Y. Kim, 《面向视频编码模型的端到端可学习多尺度特征压缩》, 《IEEE电路与系统视频技术汇刊》, 第 34 卷, 第 5 期, 第 3156–3167 页, 2024 年。[31] X. Fang, X. Wu, F. Li, Y. Duan, and X. Tao, 《兼顾机器视觉与人类视觉的分组图像压缩》, 《IEEE电路与系统视频技术汇刊》, 第 35 卷, 第 3 期, 第 2820–2831 页, 2025 年。[32] Y. Bai, X. Liu, K. Wang, X. Ji, X. Wu, and W. Gao, 《用于无损和近无损图像压缩的深度有损加残差编码》, 《IEEE模式分析与机器智能汇刊》, 第 46 卷, 第 5 期, 第 3577–3594 页, 2024 年。[33] T. M. Cover, 《信息论基础》。John Wiley & Sons, 1999 年。[34] A. Dytso, R. Bustin, H. V. Poor, and S. Shamai, 《广义高斯分布的解析性质》, 《统计分布与应用杂志》, 第 5 卷, 第 1 期, 第 6 页, 2018 年。[35] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, 《PASCAL视觉目标类别挑战赛2012（VOC2012）结果》,
h
t
t
p://
www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html。[36] K. Gokcesu and H. Gokcesu, 《用于鲁棒学习的广义Huber损失及其在稳健统计中的高效最小化》, arXiv 预印本 arXiv:2108.12627, 2021 年。[37] R. Zou, C. Song, and Z. Zhang, 《细节决定成败: 用于图像压缩的基于窗口的注意力》, 载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集, 2022 年, 第 17 492–17 501 页。[38] X. Wang, R. Zhang, T. Kong, L. Li, and C. Shen, 《Solov2: 动态快速实例分割》, 《神经信息处理系统进展》, 第 33 卷, 第 17 721–17 732 页, 2020 年。[39] Z. Wang, E. Simoncelli, and A. 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、系统与计算论文集, 第 2 卷, 2003 年, 第 1398–1402 页。[40] X. Wang, Y. Li, H. Zhang, and Y. Shan, 《借助生成式人脸先验迈向真实场景盲人脸复原》, 载于 IEEE/CVF计算机视觉与模式识别会议（CVPR）论文集, 2021 年, 第 9164–9174 页。[41]《多功能视频编码参考软件版本 12.1（vtm-12.1）》, 2021 年。*[在线]*。
https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftwareVTM/-/tags/VTM-12.1。[42] Y. Liu, W. Yang, H. Bai, Y. Wei, and Y. Zhao, 《结合分割先验的区域自适应变换图像压缩》, 载于欧洲计算机视觉会议（ECCV）论文集, 2024 年, 第 181–197 页。[43] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, 《Microsoft COCO: 上下文中的常见物体》, 载于欧洲计算机视觉会议（ECCV）论文集, 2014 年, 第 740–755 页。[44] Y. Zeng, P. Zhang, J. Zhang, Z. Lin, and H. Lu, 《迈向高分辨率显著目标检测》, 载于 IEEE/CVF国际计算机视觉会议（ICCV）论文集, 2019 年, 第 7234–7243 页。[45] J. Bégaint, F. Racapé, S. Feltman, and A. Pushparaja, 《CompressAI: 用于端到端压缩研究的PyTorch库和评估平台》, arXiv 预印本 arXiv:2011.03029, 2020 年。[46] F. Bellard, 《BPG图像格式》, 2015 年。*[在线]*。https://bellard.org/bpg。[47] G. Bjontegaard, 《计算RD曲线之间的平均峰值信噪比差异》, ITU SG16文件 VCEG-M33, 2001 年。[48] R. Khanam and M. Hussain, 《YOLOv11: 关键架构改进概述》, arXiv 预印本 arXiv:2410.17725, 2024 年。

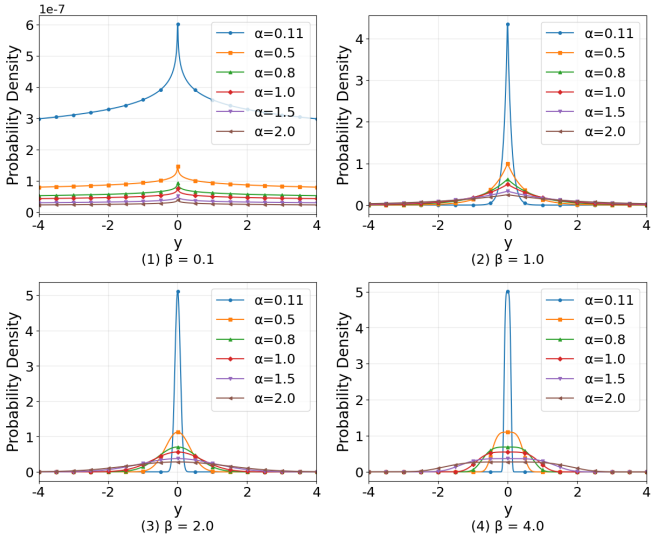


图 15. GGM 的概率密度函数形状参数与尺度参数之间的关系 α 在不同形状参数 β 下。

附录

基于广义高斯模型的统一的基于感兴趣区域的图像压缩范式

A. 经典广义高斯模型

我们可视化了尺度参数和形状参数对 GGM 对潜在空间分布建模的影响，如图所示。15。可以观察到，尺度参数 α 控制分布的宽度或离散度，而形状参数 β 决定分布的峰度，即其“尖峰程度”和“尾部厚度”。

B. 中心差分格式的误差界分析

为实现基于梯度的稳定优化，我们提出一种高效的中心差分格式来近似 $\gamma(a,\ b)$ 的梯度，从而在训练过程中恢复可微性。

$$\frac{\partial}{\partial a}\gamma(a,b) \approx \frac{\gamma(a+\epsilon,b)-\gamma(a-\epsilon,b)}{2\epsilon}, \tag{19}$$

步长为 $\epsilon = 1\times10^{-5}$。为分析用于替代数值积分的中心差分格式的误差大小，我们使用泰勒展开推导公式（19）。根据泰勒定理可知,

$$\begin{aligned} f(a+\epsilon) &= f(a)+f'(a)\epsilon+\frac{f''(a)}{2!}\epsilon^2+O(\epsilon^3),\\ f(a-\epsilon) &= f(a)-f'(a)\epsilon+\frac{f''(a)}{2!}\epsilon^2-O(\epsilon^3) \end{aligned} \tag{20}$$

两式相减得到中心差分:

$$\frac{f(a+\epsilon)-f(a-\epsilon)}{2\epsilon} = f'(a)+O(\epsilon^2) \tag{21}$$

由此可见，中心差分方案的截断误差为 $O(\epsilon^2)$，而标准前向差分的截断误差为 $O(\epsilon)$。因此，我们的方案能够实现精度显著更高的近似。所得误差可控，且满足有效训练所需的精度。

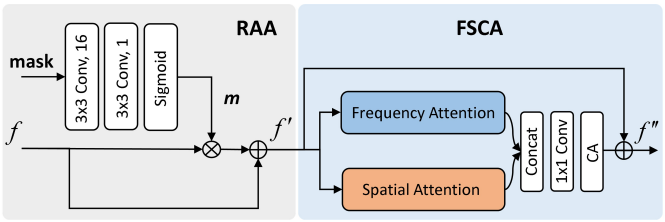


图16. MGFE模块示意图。MGFE模块主要由区域自适应注意力（RAA）块和频率-空间协同注意力（FSCA）块组成。FSCA块中的频率注意力（FA）分支和空间注意力（SA）分支

C. 掩码引导特征增强模块

为了利用生成的掩码 $\mathbf{M}$ 进行隐式比特分配，我们采用了 MGFE 模块，如图16,$f'' = \mathrm{MGFE}(\mathbf{M}, f)$。MGFE模块主要由区域自适应注意力（RAA）块和频率-空间协同注意力（FSCA）块组成。RAA 块用作变换函数 F (·)，以实现重要比特的差异化分配。我们逐层集成 MGFE 模块，取代初始的窗口注意力模块，从而精确区分前景和背景区域，实现高质量的区域自适应编码。具体而言，MGFE 模块可表示为

$$\begin{aligned} m &= Sigmoid(Conv_{3\times3}(Conv_{3\times3}(\mathbf{M}))), \\ f' &= f \times m + f, \\ f_{att} &= \mathrm{Concat}\left(FreqAtt(f'), SpatAtt(f')\right), \\ f'' &= f' + CA\left(Conv_{1\times1}\left(f_{att}\right)\right). \end{aligned} \tag{22}$$

与传统的仅空间注意力不同，FSCA块结合了频域注意力和空间域注意力，在两个正交维度上协同增强关键特征，从而获得高质量的图像表示。FSCA块首先将RAA块的输出特征$\boldsymbol{f}'$ 作为输入。随后，输入通过频率注意力（FA）分支$FreqAtt(\cdot)$和空间注意力（SA）分支 $SpatAtt(\cdot)$并行处理。将这两个分支的输出沿通道维度进行拼接（Concat）。随后，一个 1×1 卷积层对拼接结果进行特征融合和降维。接着，将融合后的特征输入通道注意力（CA）块，以自适应地重新校准不同特征通道的权重。最后，通过残差连接将注意力特征与FSCA块的输入 $\boldsymbol{f}'$ 相加，得到最终的特征图 $\boldsymbol{f}''$。

D. HRSOD 数据集上的压缩性能

我们引入了一个包含高分辨率图像的 HRSOD 数据集，用于评估其率失真性能。如图 17所示，我们绘制了不同压缩方法的率失真曲线。我们的方法在感兴趣区域和非感兴趣区域均取得了最佳的整体压缩性能，并优于最新的最先进方法 Hu2025。这进一步证实了我们的 GGM 在基于感兴趣区域的压缩任务中的优越性和泛化能力。

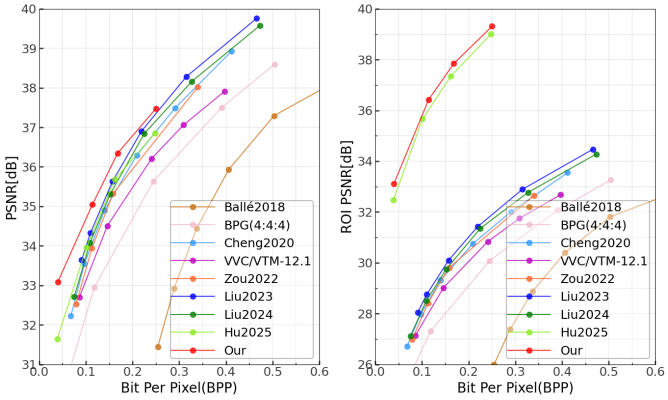


图 17. HRSOD 数据集上的峰值信噪比和感兴趣区域峰值信噪比平均率失真性能。

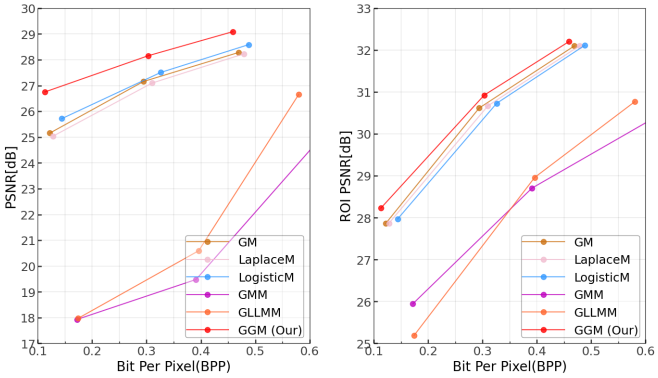


图 18. 使用不同概率密度函数的率失真性能。

E. 不同概率分布模型的比较

为分析不同概率分布模型（PDM）对编码性能的影响，我们通过将压缩网络中的 GGM 替换为几种经典的概率分布模型，开展了对比实验。如图 18所示，我们绘制了不同概率分布模型的率失真曲线。我们的 GGM 通过学习得到的形状参数 β，能够准确地建模潜变量张量的分布，在整幅图像和感兴趣区域上均实现了最佳编码性能。使用 GMM 和 GLLMM 不适合建模其真实分布，且容易导致模型冗余和过拟合。