QVRF：面向学习式图像压缩的量化误差感知可变码率框架

Kedeng Tong[1*]，*Yaojun Wu*[2], *Yue Li*[2], *Kai Zhang*[2], *Li Zhang*[2], *Xin Jin*[1]

[1]清华大学深圳国际研究生院，中国深圳，518055 [2]字节跳动多媒体实验室，美国加利福尼亚州圣迭戈，92122

摘要

学习式图像压缩展现出良好的压缩性能，但在较宽范围内实现可变比特率仍然面临挑战。最先进的可变码率方法以牺牲模型性能为代价，并且需要大量额外参数。本文提出一种量化误差感知可变码率框架（QVRF），利用单变量量化调节器 a，在单个模型中实现宽范围的可变码率。具体而言，QVRF 定义了一个量化调节器向量，并将其与预定义的拉格朗日乘子相结合，以控制离散可变码率下所有潜在表示的量化误差。此外，重参数化方法使 QVRF 与舍入量化器兼容。大量实验表明，现有的固定码率基于变分自编码器的方法在配备 QVRF 后，可以在单个模型中实现宽范围的连续可变码率，且不会造成显著的性能下降。此外，QVRF 仅需少量额外参数，其率失真性能便优于当前的可变码率方法。代码见 https://github.com/bytedance/QRAF。

索引词— 学习式 图像压缩、可变码率、量化调节器， 熵编码

1. 引言

近年来，采用变分自编码器（VAE）框架的蓬勃发展的图像压缩技术 [1-11] ，与 JPEG [12], JPEG 2000 [13], HEVC [14], 和 VVC [15] 等混合编解码器相比，展现出出色且更优的压缩性能。这些大感受野学习式压缩方法在大型数据集上训练，以最小化率失真（RD）代价 R D+λ，并针对各种内容全局优化网络参数。这里，λ 是用于控制 RD 权衡的拉格朗日乘子。这些方法克服了混合编解码器中区域受限预测和模块独立优化的缺点。因此，学习式图像压缩被认为是图像压缩的未来，而 JPEGAI 致力于制定基于学习的图像编码标准 [16]。

然而，在较宽范围内实现可变比特率仍是学习式图像压缩面临的一项重大挑战。大多数学习式方法通过训练多个对应不同 λ 取值的模型来覆盖较宽的比特率范围，而每个模型都只能采用固定码率。使用多个固定码率模型会带来高昂的训练成本和内存需求，而且无法像混合编解码器那样灵活调整比特率；混合编解码器可以在一个编解码器中实现宽范围的可变码率和码率自适应，且不会造成性能下降（例如，HEVC 支持 51 个级别）[17-19]。

为解决这一问题，一些技术采用单个模型实现可变码率。条件自编码器方案 [20-24] 将以 λ 的先验信息为输入的子网络纳入卷积层，从而实现可变码率。然而，这些方案显著增加了计算复杂度和模型大小。Choi 等人 [24] 提出调整通用量化中均匀噪声的间隔大小，将固定码率模型扩展到较窄范围的连续可变码率（例如 ±0.1 bpp），且不会造成显著的性能下降。幅度调制范式 [25-28] 定义缩放系数或子网络，以改变潜变量的幅度，从而间接控制量化误差并实现可变比特率。对缩放系数进行指数插值 [26] 有助于实现连续可变码率。然而，这种范式会影响概率建模的分布，并且在宽范围可变比特率下会导致压缩性能下降。

本文提出量化误差感知可变码率框架（QVRF），该框架利用单变量量化调节器 a 控制潜在表示的整体量化误差，从而在单个模型中实现宽范围可变码率。具体而言，QVRF 为离散可变码率定义了一个与预定义拉格朗日乘子集相耦合的量化调节器向量。此外，我们提出了一种重参数化方法，使 QVRF 能够与舍入量化器兼容地运行。我们发现，a 与 λ 的平方根呈线性关系，并且在量化和熵过程中改变 a 的值

[*] Kedeng Tong 的这项工作是在字节跳动实习期间完成的。

足以在 QVRF 中实现连续可变码率，且不会造成性能下降。为展示其通用性，我们将 QVRF 集成到三种学习式图像压缩方法 [8-10] 中。配备 QVRF 的基于变分自编码器的模型能够在单个模型中实现宽范围连续可变码率。值得注意的是，在率失真性能、额外参数量和计算复杂度方面，QVRF 均优于最先进的可变码率方法（SOTA）。

2. 所提方法

本节首先回顾基于变分自编码器的压缩的表述，然后介绍所提出的 QVRF 和重参数化方法。

2.1. 基于变分自编码器的模型表述

基于变分自编码器的压缩优化问题可表述为一个拉格朗日优化过程：

$$\begin{aligned}\mathcal{L} &= \mathcal{R}(\hat{y}) + \mathcal{R}(\hat{z}) + \lambda \cdot \mathcal{D}(x, \hat{x}) \\ &= \mathbb{E}[-\log_2(p_{\hat{y}|\hat{z}}(\hat{y}\,|\,\hat{z}))] + \mathbb{E}[-\log_2(p_{\hat{z}}(\hat{z}))] \\ &\quad + \lambda \cdot \mathcal{D}(x, \hat{x}\,|\,\hat{y})\end{aligned}, \tag{1}$$

其中 $\mathcal{R}(\cdot)$、$\mathcal{D}(\cdot,\cdot)$、$p_{\hat{y}|\hat{z}}$ 和 $p_{\hat{z}}$ 分别表示压缩 bitrate、原始图像 x 与重建图像 $\hat{x}$ 之间的失真、量化尺寸信息 $\hat{z}$ 的概率质量函数，以及在 $\hat{z}$ 条件下对量化潜在表示 $\hat{y}$ 进行的高斯熵估计。由于无损算术编码采用取整量化，量化潜变量 $\hat{y}$ 的熵模型表示为：

$$\begin{aligned}p_{\hat{y}|\hat{z}}(\hat{y}\,|\,\hat{z}) &= \prod_i p_{\hat{y}|\hat{z}}(\hat{y}_i\,|\,\hat{z}) \\ p_{\hat{y}|\hat{z}}(\hat{y}_i\,|\,\hat{z}) &= \mathcal{N}(u_i, \sigma_i) * \mathbb{U}(-\frac{1}{2}, \frac{1}{2})(\hat{y}_i)\end{aligned}, \tag{2}$$

其中 $\hat{y}_i$、u_i 和 σ_i 分别是 $\hat{y}$ 的估计值、均值和估计高斯分布的尺度。在式 1 中固定 λ 时的优化问题用于固定码率范式。

2.2. 量化误差感知码率自适应框架

与固定 λ [24] 下通用量化的区间调整不同，我们提出将 λ 和 a（一种单变量量化调节器）相结合，以控制整体潜在表示的量化误差并实现宽范围的可变码率。所提出的 QVRF 框架如图 1 所示，其中 Q 和 AE/AD 分别表示量化和算术熵编码。如图所示，QVRF 在量化和熵编码过程中使用单变量 a，以控制潜在表示 y 的整体量化误差，从而实现可变码率。QVRF 保持概率建模网络的输入不变，使对 y 进行的高斯熵估计基本保持固定。量化调节器 a 用于调整量化区间大小。对于 QVRF 中由编码器生成的潜在表示 y_i 的每个元素，式 2 中的累积质量函数定义为：

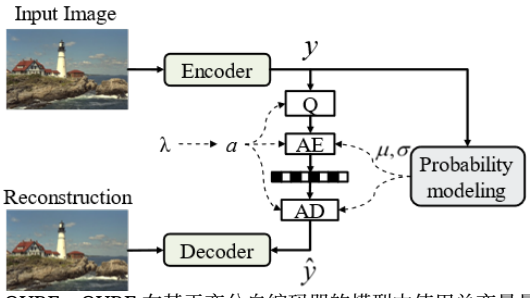


图 1。QVRF。QVRF 在基于变分自编码器的模型中使用单变量量化调节器 a，以控制量化与熵编码过程中所有 y 的量化误差，同时保持概率建模几乎不变。

$$\begin{aligned}p_{\hat{y}|\hat{z}}(\hat{y}_i\,|\,\hat{z}, a) &= \int_{y_i-\frac{1}{2a}}^{y_i+\frac{1}{2a}} \frac{1}{\sqrt{2\pi}\sigma_i} e^{\frac{(y_i-u_i)^2}{2\sigma_i^2}} dy_i \\ &= \mathcal{N}(u_i, \sigma_i) * \mathbb{U}(-\frac{1}{2a}, \frac{1}{2a})(\hat{y}_i)\end{aligned}. \tag{3}$$

式 3 中的量化为 quantization error QE between y and $\hat{y}$

$$QE = y - \hat{y} = \mathbb{U}(-\frac{1}{2a}, \frac{1}{2a}). \tag{4}$$

通过调整 a 的值，可以改变 y 与 $\hat{y}$ 之间的整体量化误差，并影响存储和传输所需的比特数。我们提出将 λ 与 a 耦合起来，使二者建立一一对应关系。先前的工作 [25, 26] 表明，在训练中使用多个 λ 对于实现离散可变码率是必要的。因此，定义一个由 $\{a_1, ..., a_n\}$ 构成的量化调节器向量 $A \in R^n$，并在训练过程中将其组装起来，使其映射到包含 n 个元素的预定义拉格朗日乘子集 Λ。此外，将式 1 中的率失真优化函数改写为：

$$\begin{aligned}\mathcal{L} &= \sum_{\lambda \in \Lambda} \mathcal{R}(\hat{y}) + \mathcal{R}(\hat{z}) + \lambda \cdot \mathcal{D}(x, \hat{x}) \\ &= \sum_{j=1}^{n} \mathbb{E}[-\log_2(p_{\hat{y}|\hat{z}}(\hat{y}\,|\,\hat{z}, a_j))] + \mathbb{E}[-\log_2(p_{\hat{z}}(\hat{z}))] \\ &\quad + \lambda_j \cdot \mathcal{D}(x, \hat{x}\,|\,\hat{y})\end{aligned}. \tag{5}$$

2.3. 重参数化方法

提出重参数化方法，使 QVRF 与现有的学习式压缩方法 [8-10] 兼容；这些方法利用舍入量化器进行量化和算术编码。重参数化方法通过对 y_i、u_i 和 σ_i 执行缩放操作来替换式 3：

$$\begin{aligned}p_{\hat{y}|\hat{z}}(\hat{y}_i\,|\,\hat{z}) &= \int_{ay_i-\frac{1}{2}}^{ay_i+\frac{1}{2}} \frac{1}{\sqrt{2\pi}(a\sigma_i)} e^{\frac{(ay_i-au_i)^2}{2(a\sigma_i)^2}} day_i \\ &= \mathcal{N}(au_i, a\sigma_i) * \mathbb{U}(-\frac{1}{2}, \frac{1}{2})(a\hat{y}_i)\end{aligned}. \tag{6}$$

因此，尺度 y_i 量化，同时包含 $1/a$ 的量化误差，缩放后的量化值用于熵编码。请注意，用于熵编码的缩放量化值应当

按照严格的倒数原则进行反向缩放，以准确获得 $\hat{y}$ 。

3. 实验结果

为展示我们提出的算法的有效性，我们进行了实验和对比。

3.1. 实验设置

遵循先前的工作 [4], , 我们从 ImageNet 训练集中选取了 8,000 张图像的子集 [29] ，并从 CLIC2020 训练集中选取了 584 张图像 [30] 作为训练数据。公式 5 中的失真采用均方误差（MSE）。我们选取了三种具有代表性的基于变分自编码器的方法作为基线方法：基础的基于变分自编码器的方法 [10], 、基于变分自编码器的自回归模型 [9], 以及带有注意力模块的高级网络 [8], 。我们将 QVRF 与每种方法结合。为覆盖从低比特率到高比特率的范围，我们将预定义的拉格朗日乘子集 Λ 设为 (0.0018, 0.0035, 0.0067, 0.0130, 0.0250, 0.0483, 0.0932, 0.1800)。根据经验， Λ 中的第一个值设为参考值 $\lambda_{\text{参考值}}$ ，并将 A 初始化为 $\sqrt{\lambda / \lambda_{\text{参考值}}}$ ，其中 a_1 设为常数 1。训练过程使用 Adam优化器，批量大小为 8，初始学习率设为 0.0001。我们采用分阶段训练策略，前 2000 个 epoch 在 λ=0.18 上优化网络参数。随后，将 A 与噪声近似联合优化 500 个 epoch，再通过直通估计 [5] 训练 500 个 epoch。比特率以每像素比特数（bpp）衡量，指率失真曲线中的熵编码实际比特数。峰值信噪比（PSNR）和多尺度结构相似性指数（MS-SSIM）[16]是衡量图像失真的客观指标。

我们在多个数据集上测试所提出的算法，包括包含24张图像的 Kodak 数据集 [31],包含41张高分辨率图像的 CLIC 验证集[30], 以及包含16张高质量图像的 JEPGAI 测试数据集[16]。我们纳入了来自 [32] 的 Ballé 等人 [10] 和 Minnen 等人 [9] 的固定码率模型。我们在相同的训练数据集上重新训练了 Cheng 等人 [8], 的8种固定码率模型，其中卷积通道 N 和M均为192，以作比较。为确保公平比较，我们按照原论文的设置，在基线模型上实现了三种最先进的可变码率方法 [21, 26, 27] 。

3.2. 离散和连续可变比特率

如图2所示，与以 Cheng 方法为基准、经重新训练的八个固定码率模型相比，QVRF 在单个模型中即可实现离散和连续可变比特率（离散/连续可变比特率），且性能下降可以忽略不计[8] 。通过最小化损失函数 Eq. 5，量化调节器向量A与网络参数联合优化。因此，对于离散可变比特率，训练模型可从A 中选择适当的a值来实现离散可变码率。得益于a和网络参数对宽范围取值的兼容性，在a的最小值至最大值的扩展范围内更改该值（该值在编码和解码过程中共用），即可在

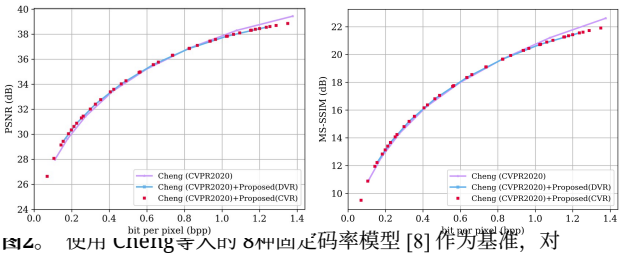


图2。 使用 Cheng等人的 8种固定码率模型 [8] 作为基准，对 QVRF 在离散和连续可变码率方面的性能进行评估。

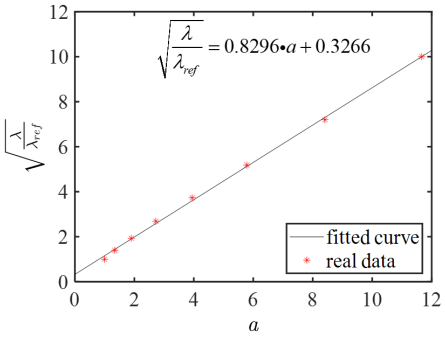


图3。单变量量化调节器 a与拉格朗日乘子 λ的平方根之间的线性相关性。真实数据来自 QVRF 的重新训练模型和 Cheng等人 [8]。与参考拉格朗日乘子$_{ref}$ λ 对应的a值设为常数 1。

从A的最小值到最大值的扩展范围内，实现连续可变比特率，如图 2 所示。

如图3所示，在采用 QVRF 和 Cheng 等人 [8]的训练模型中，可以观察到a与 λ 的平方根之间存在线性关系。该分析基于从训练后的A 和 Λ 中获得的真实数据。a_1 对应于$_{ref}$λ 为常数1。拟合a与λ 平方根之间关系的线性函数可表示如下：

$$\sqrt{\frac{\lambda}{\lambda_{ref}}} = 0.8296 \cdot a + 0.3266 \quad . \tag{7}$$

这揭示了 QVRF 中可变码率下a 与 λ 的平方根之间存在明确的线性关系，并可能有助于码率自适应。

3.3. 可变码率方法的比较

如图 4 所示，对于所有三种基线模型，QVRF 的表现均优于现有可变比特率技术 [21, 26, 27] 。Theis 等人 [27] 在低比特率和高比特率下均出现了明显的性能下降。Song 等人 [21] 基于 Ballé 等人 [10]的率失真性能在低比特率下优于固定码率模型，但在中等比特率下的表现逊于 QVRF。Cui 等人 [26] 在低比特率和

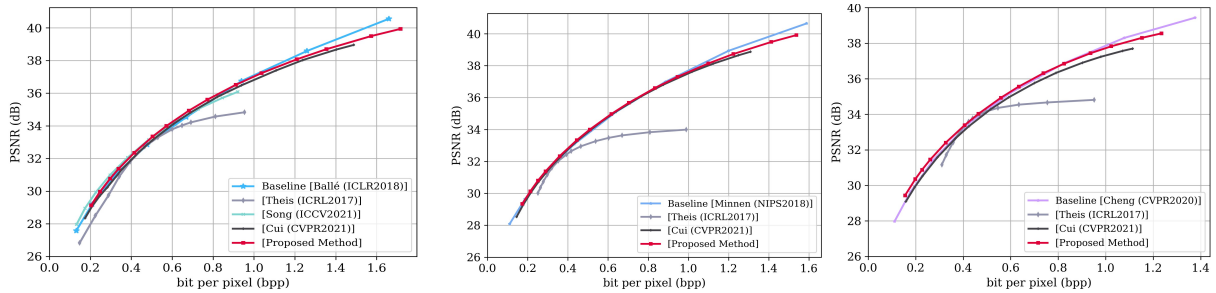


图 4。结果　可变码率方法 [22, 27, 28] 和 QVRF 与 Kodak 数据集上的三种基线方法的结果。基线模型包括 Ballé 等人 [1　　0], Minnen 等人[9], 和 Cheng 等人 [8]，从左到右。

中等比特率下的压缩性能与固定码率模型相当，但其性能下降比 QVRF 更为明显。

我们评估了 Cui 等人 [26] 和 QVRF 在 BD-Rate 和多失真权衡方面的表现，以展示 QVRF 的优越性。表 1 展示了 QVRF 和 Cui 等人 [26] 在 Kodak 数据集上与固定码率方法 [8, 10] 相比的 BD-Rate 结果。BD-Rate 根据从低、中、高比特率中选取的四个 RD 点计算得出。与 Ballé 等人 [10], 、Minnen 等人 [9], 和 Cheng 等人[8], 的固定码率模型相比，Cui 等人 [26] 的比特率分别增加了 2.58%、1.29% 和 4.87%。得益于低比特率下的改进以及高比特率下轻微的性能下降，QRVF 在三个基础模型上分别实现了 0.25%、-0.08% 和 1.21% 的比特率降低。对于多失真权衡，我们以 Minnen等人. [9] 为基线模型，在 CLIC 验证集 [30] 和 JEPGAI 测试数据集 [16] 上测试 QVRF 和 Cui 等人 [26] 。如图 5 和图 6 所示，与固定码率模型相比，就峰值信噪比和多尺度结构相似性而言，Cui 等人 [26] 在没有性能下降的情况下实现的可变比特率范围比 QVRF 更窄。

除了压缩性能之外，额外参数和复杂度也是衡量可变码率方法可行性的重要指标。我们比较了可变码率方法 [21, 26, 27] 和 QVRF 在 n 设为 8 时的额外参数 参数和计算复杂度浮点运算次数。该比较结果以相对于单个固定码率模型的百分比（%）表示，列于表 2。值得注意的是，QVRF 所需的额外参数极少，在这方面的规模几乎比 Theis 等人 [27] 和 Cui 等人[26] 小 200 倍。此外，QVRF 的额外浮点运算次数百分比与 Theis 等人 [27] 和 Cui 等人 [26], 的相近，但其规模几乎比 Song 等人 [21]小 300,000 倍。

4. 结论

通过将单变量量化调节器 a纳入量化和熵编码过程，所提出的 QVRF 扩展了现有的基于变分自编码器的压缩方法，使单个模型能够实现连续可变码率。重参数化方法使 QVRF 与舍入量化器兼容。我们证明了拉格朗日乘子的平方根与 a 呈线性关系，改变 a 的值即可实现连续可变码率。

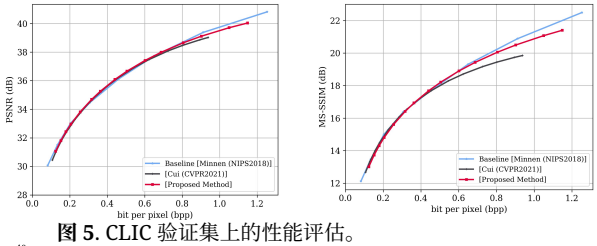


图 5. CLIC 验证集上的性能评估。

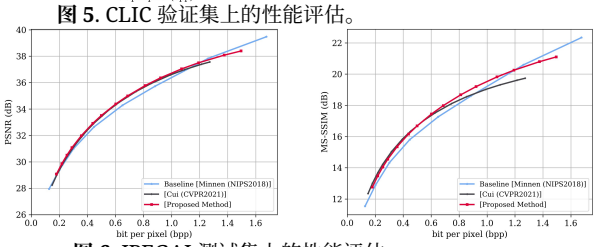


图 6. JPEGAI 测试集上的性能评估。

表 1. Cui等人[26] 和 QVRF 与固定码率模型 [8, 10] 在 Kodak 数据集上的平均 BD-Rate 结果。

可变方法	固定码率模型		
	Ballé 等人 [10]	Minnen 等人 [9]	Cheng 等人 [8]
固定码率模型 + Cui 等人 [26]	2.58%	1.29%	4.87%
固定码率模型 + QVRF	**-0.25%**	**0.08%**	**-1.21%**

表 2。与单一固定码率模型相比，可变码率方法 [22, 27, 28] 和 QVRF 的额外参数及复杂度百分比。

基础模型	Theis 等人 [27]		Song 等人 [21]		Cui 等人 [26]		QVRF	
	参数。	浮点运算次数	参数。	浮点运算次数	参数。	浮点运算次数	参数。	浮点运算次数
Ballé 等等。[10]	0.02168	0.00060	251.77958	268.37507	0.06932	0.00062	**0.00007**	0.00089
Cheng 等人等。[8]	0.00691	0.00015	-	-	0.02073	0.00016	**0.00003**	0.00022

实验结果表明，QVRF 能够使现有学习式方法在单一模型中实现宽范围的连续可变码率，同时保持与固定码率模型相当的性能。此外，在率失真性能、额外参数量和复杂度方面，QVRF 均优于当前的可变码率方法。

5. 参考文献

[1] R. Zou、C. Song 和 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论文集，2022 年，第 17492–17501 页。[2] J.-H. Kim、B. Heo 和 J.-S. Lee，《用于学习式图像压缩的联合全局与局部层级先验》，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2022 年，第 5992–6001 页。[3] D. He、Z. Yang、W. Peng、R. Ma、H. Qin 和 Y. Wang，《Elic：采用不均匀分组空域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，2022 年，第 5718–5727 页。[4] D. He、Y. Zheng、B. Sun、Y. Wang 和 H. Qin，《用于高效学习式图像压缩的棋盘格上下文模型》，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2021 年，第 14771–14780 页。[5] Z. Guo、Z. Zhang、R. Feng 和 Z. Chen，《先软后硬：重新思考神经图像压缩中的量化》，载于国际机器学习会议，2021：PMLR，第 3920–3929 页。[6] Y. Qian 等，《利用全局参考学习精准熵模型以实现图像压缩》，arXiv 预印本 arXiv:2010.08321，2020。[7] D. Minnen 和 S. Singh，《用于学习式图像压缩的逐通道自回归熵模型》，载于 2020 年 IEEE国际图像处理会议（ICIP），2020：IEEE，第 3339–3343 页。[8] Z. Cheng、H. Sun、M. Takeuchi 和 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 7939–7948 页。[9] D. Minnen、J. Ballé 和 G. D. Toderici，《用于学习式图像压缩的联合自回归与层级先验》，《神经信息处理系统进展》，第 31 卷，2018。[10] J. Ballé、D. Minnen、S. Singh、S. J. Hwang 和 N. Johnston，《采用尺度超先验的变分图像压缩》，arXiv 预印本 arXiv:1802.01436，2018。[11] K. Tong、X. Jin、C. Wang 和 F. Jiang，《SADN：采用空域-角度去相关的学习式光场图像压缩》，载于 ICASSP 2022-2022 IEEE声学、语音与信号处理国际会议（ICASSP），2022：IEEE，第 1870–1874 页。[12] G. K. Wallace，《JPEG 静态图像压缩标准》，《ACM通讯》，第 34 卷，第 4 期，第 30–44 页，1991。[13] A. Skodras、C. Christopoulos 和 T. Ebrahimi，《JPEG 2000 静态图像压缩标准》，《IEEE信号处理杂志》，第 18 卷，第 5 期，第 36–58 页，2001。[14] G. J. Sullivan、J.-R. Ohm、W.-J. Han 和 T. Wiegand，《高效视频编码（HEVC）标准概述》，《IEEE视频技术电路与系统汇刊》，第 22 卷，第 12 期，第 1649–1668 页，2012。[15] J. C. B. Bross、S. Liu 和 Y.-K. Wang，《多功能视频草案（草案 10）》，JVET-S2001，2020。[16]《JPEG AI》。https://jpeg.org/jpegai/（访问于。[17] Y. Li、Z. Liu、Z. Chen 和 S. Liu，《多功能视频编码的码率控制》，载于 2020 年 IEEE国际图像处理会议（ICIP），2020：IEEE，第 1176–1180 页。

[18] B. Li、H. Li、L. Li 和 J. Zhang，“λ 域码率控制算法在高效视频编码中的应用”，IEEE《图像处理汇刊》，第 23 卷，第 9 期，第 3841–3854 页，2014 年。[19] H. Choi、J. Yoo、J. Nam、D. Sim 和 I. V. Bajić，“用于多级码率控制的逐像素统一码率量化模型”，IEEE《信号处理选题汇刊》，第 7 卷，第 6 期，第 1112–1123 页，2013 年。[20] Z. Sun 等，“插值可变码率图像压缩”，载于第 29 届 ACM 多媒体国际会议论文集，2021 年，第 5574–5582 页。[21] M. Song、J. Choi 和 B. Han，“通过空间自适应特征变换实现可变码率深度图像压缩”，载于 IEEE/CVF国际计算机视觉会议论文集，2021 年，第 2380–2389 页。[22] J. Lin、D. Liu、J. Liang、H. Li 和 F. Wu，“用于可变码率视频压缩的深度调制方案”，载于 2021 年 IEEE国际图像处理会议（ICIP），2021 年：IEEE，第 3722–3726 页。[23] F. Yang、L. Herranz、J. Van De Weijer、J. A. I. Guitián、A. M. López 和 M. G. Mozerov，“使用调制自编码器实现可变码率深度图像压缩”，IEEE《信号处理快报》，第 27 卷，第 331–335 页，2020 年。[24] Y. Choi、M. El-Khamy 和 J. Lee，“使用条件自编码器实现可变码率深度图像压缩”，载于 IEEE/CVF国际计算机视觉会议论文集，2019 年，第 3146–3154 页。[25] S. Yin、C. Li、Y. Bao、Y. Liang、F. Meng 和 W. Liu，“通用高效可变码率神经图像压缩”，载于 ICASSP 2022—2022 IEEE声学、语音与信号处理国际会议（ICASSP），2022 年：IEEE，第 2025–2029 页。[26] Z. Cui、J. Wang、S. Gao、T. Guo、Y. Feng 和 B. Bai，“具有连续码率自适应的非对称增益深度图像压缩”，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2021 年，第 10532–10541 页。[27] L. Theis、W. Shi、A. Cunningham 和 F. Huszár，“使用压缩自编码器进行有损图像压缩”，arXiv预印本 arXiv:1703.00395，2017 年。[28] C. Jia、Z. Ge、S. Wang、S. Ma 和 W. Gao，“神经图像压缩的率失真特性建模”，载于 2022 年数据压缩会议（DCC），2022 年：IEEE，第 202–211 页。[29] J. Deng、W. Dong、R. Socher、L.-J. Li、K. Li 和 L. Fei-Fei，“ImageNet：大规模分层图像数据库”，载于 2009 年 IEEE计算机视觉与模式识别会议，2009 年：IEEE，第 248–255 页。[30] “学习式图像压缩挑战赛。” https://clic.compression.cc/（访问日期。[31] “柯达无损真彩色图像套件。” https://r0k.us/graphics/kodak/（访问日期。[32] J. Bégaint、F. Racapé、S. Feltman 和 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。

附录

QVRF 中可变比特率示意图及重参数化方法

如图 A1 所示，使用值 $a>1$ 会降低量化值出现的概率，因此需要增加熵编码所使用的比特数。

QVRF 在高比特率下的性能提升

仅仅依靠更大的预定义拉格朗日乘子集合，就能显著提升 QVRF 在高比特率下的性能。如图 A2 所示，当所提出的组合模型采用预定义的拉格朗日乘子集合（0.0018、0.0035、0.0067、0.0130、0.025、0.0483、0.0932、0.18、0.36、0.72、1.44）时，在所有比特率下，其压缩性能均可与 λ 取值范围为（0.0018、0.0035、0.0067、0.0130、0.0250、0.0483、0.0932、0.1800）的 8 种固定码率模型相媲美。

事实上，与固定码率模型相比，组合模型的比特率降低了 0.4%。

比特消耗分析

根据表 R1 中给出的比特率消耗和压缩性能结果，可以发现 QVRF 仅影响潜在表示的压缩，而辅助信息需要固定数量的比特。

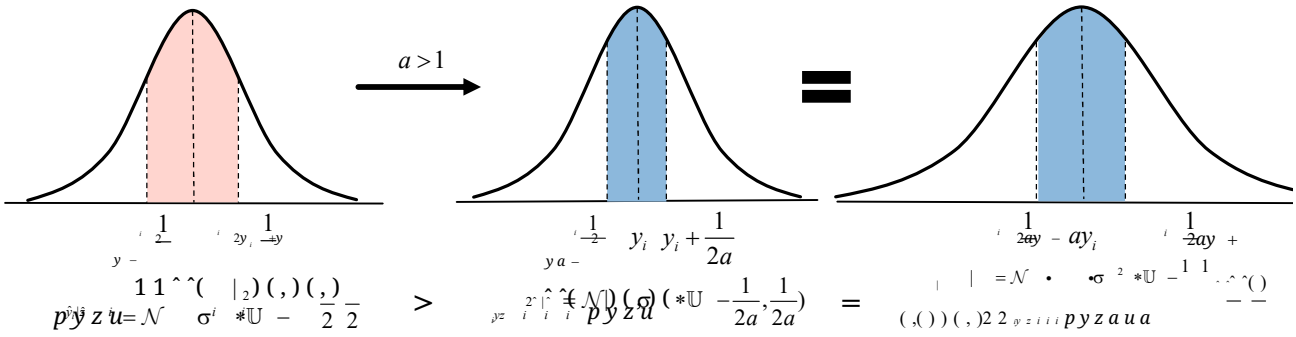


图 A1 对较大的a，熵编码需要为量化值分配更多比特。

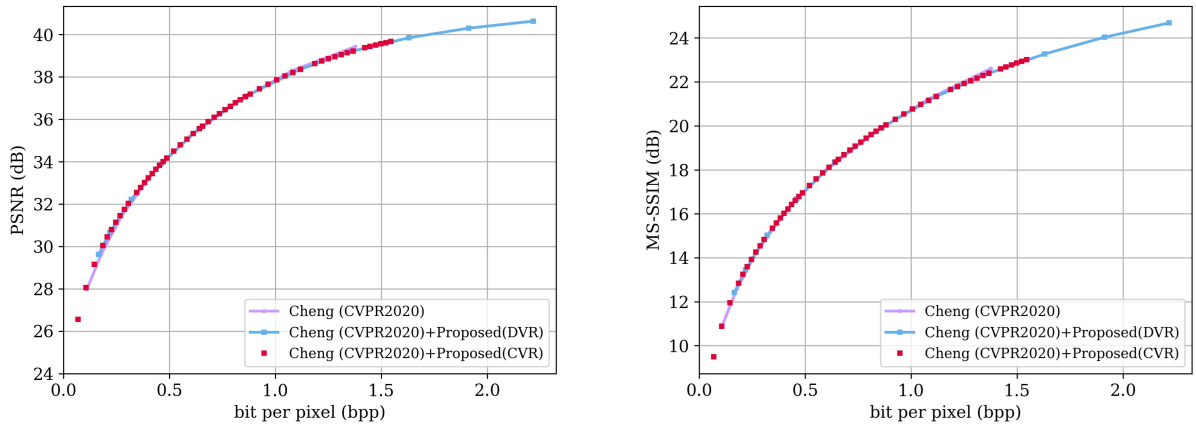


图 A2 使用 Cheng等人的 8 种固定码率模型作为基准，对 QVRF 进行离散/连续可变码率性能评估。 [8]

表 R1。 使用 QVRF 与 Cheng 等人提出的单一模型，在离散比特率下压缩 kodim01 时的比特率消耗和压缩性能结果。

总比特数（bpp）	潜在表示比特数（bpp）	辅助信息比特数（bpp）	峰值信噪比（dB）	多尺度结构相似性
0.24348	0.24112	**0.00236**	26.727	0.93857
0.30664	0.37215	**0.00236**	28.242	0.95894
0.55623	0.55387	**0.00236**	29.892	0.97330
0.79915	0.79679	**0.00236**	31.770	0.98339
1.0766	1.0743	**0.00236**	33.563	0.98944
1.3663	1.3640	**0.00236**	35.029	0.99280
1.6540	1.6516	**0.00236**	36.103	0.99467
1.9502	1.9479	**0.00236**	36.870	0.99576