arXiv:2303.09112v1 [eess.IV] 16 Mar 2023

SIGVIC：空　　间重要性引导的可变码率图像压缩

梁家铭[1] 刘美琴[1] 姚超[2] 林春雨[1] 赵耀[1]

[1] 中国北京100044，北京交通大学信息科学研究所[2] 中国北京100083，北京科技大学计算机与通信工程学院

摘要

可变码率机制提高了基于学习的图像压缩的灵活性和效率；这类方法针对不同的率失真权衡训练多个模型。可变码率最常见的方法之一，是按通道或在空间上均匀缩放内部特征。然而，空间重要性的多样性对图像压缩中的比特分配具有指导意义。本文提出空间重要性引导的可变码率图像压缩（SigVIC），并设计空间门控单元（SGU），用于自适应地学习空间重要性掩码。随后，空间缩放网络（SSN）利用该掩码引导特征缩放和可变码率比特分配。此外，为提高解码图像的质量，我们通过浅层特征融合模块（SFFM）选取 Top-K 浅层特征，以优化解码特征。实验表明，我们的方法优于其他基于学习的方法（无论是否采用可变码率机制）和传统编解码器，同时具有更低的存储开销和更高的灵活性。

索引词— 可变码率，图像压缩，空间重要性，缩放因子，浅层特征

1. 引言

随着图像存储和共享需求的激增，图像压缩逐渐成为一个重要的研究领域。它能够在可接受的失真范围内，减少图像所需的存储空间和传输成本。近年来，许多基于学习的方法 [1, 2, 3, 4, 5] 已超越传统编解码器（如 JPEG[6], JPEG2000[7], BPG[8]）。然而，这些方法依赖于训练多个具有固定率失真（RD）权衡的模型，导致存储占用相应增加，并且在实际应用中缺乏灵活性。因此，传统编解码器中已有的可变码率机制对于减少存储空间、提高灵活性至关重要。

一些方法探索了基于学习的可变比特率图像压缩，通常通过缩放编码器中的特征来获得更粗或更细的量化特征，并在解码器中进行反向缩放。Yang 等人[9]设计了一种调制自编码器，通过缩放内部特征来适应不同的率失真权衡。Choi 等人[10]提出了一种类似的条件自编码器，使用量化区间大小实现连续比特率。然而，不恰当地选择区间大小可能会导致率失真性能下降 [11]。Yin 等人[12]仅缩放瓶颈特征，但这种缩放策略不稳定，在端到端方式训练时难以适应广泛的率失真权衡 [9]。值得注意的是，所有这些方法都只涉及通道间关系，而没有利用图像的空间重要性。

当图像编码的比特率受限时，图像的空间重要性对于比特分配和码率控制具有指导意义。已有研究 [11, 13] 尝试利用空间重要性，但需要为数据集中的每幅源图像生成额外的质量图作为输入。此外，Song 等人[11]手动预先定义了用于图像压缩的统一质量图，未能充分利用空间上不同的重要性。

本文提出了一种名为 SigVIC 的空间重要性引导的可变码率图像压缩方法，无需额外输入即可适应任意比特率。具体而言，我们设计了一个空间门控单元（SGU），用于自适应地生成图像特征的空间重要性掩码。随后，我们提出了一种空间缩放网络（SSN），利用空间重要性掩码引导可变码率机制生成空间缩放因子。此外，为了提高重建图像的质量，我们提出了一种 Top-K 浅层特征融合策略，通过浅层特征融合模块（SFFM）优化解码特征。

我们在 Kodak 和 CLIC数据集上分别使用均方误差（MSE）和多尺度结构相似性（MS-SSIM）评估了所提方法。实验结果表明，与其他可变码率方法、传统编解码器以及一些知名的学习式压缩方法相比，所提方法性能更优。具体而言，根据 Kodak 数据集上的 MSE 结果计算 BD-Rate，所提方法相比 BPG[8],Song 等人[11]和 Cheng 等人[5],分别节省了 17.84%、10.85% 和 2.78% 的比特率。

2. 所提方法

我们的方法概览如图1所示，其中编码器和解码器网络由多个阶段组成。在每个阶段 t，我们设计了一个空间门控单元（SGU）和一个空间缩放网络（SSN），用于引入空间重要性

通讯作者：刘美琴。

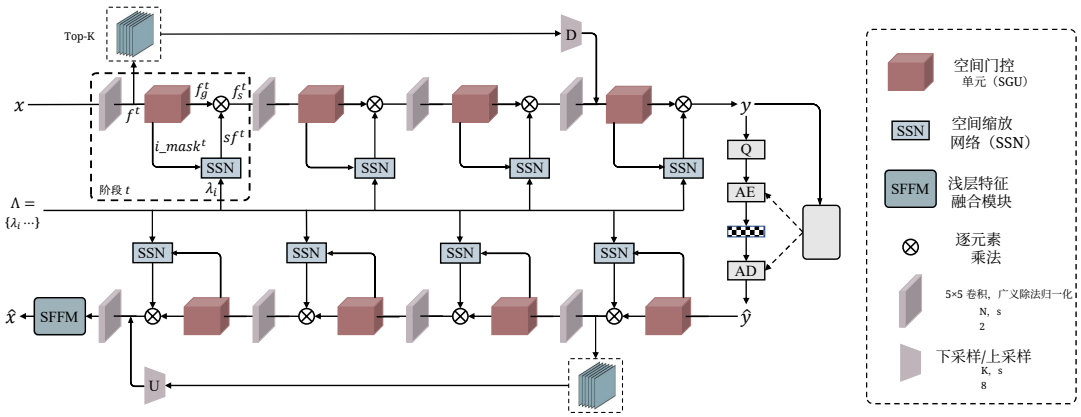


图 1。概览　　我们提出的空间重要性引导的可变码率图像压缩方法 SigVIC。

用于引导可变码率的率失真优化。此外，在解码器末端采用浅层特征融合模块（SFFM），以细化解码特征。

2.1. 空间门控单元

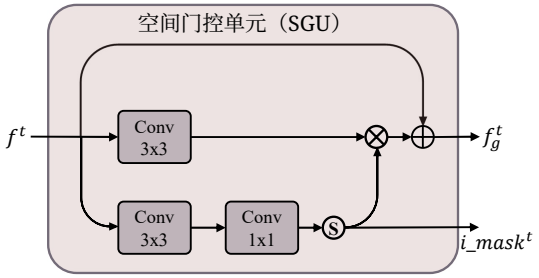


图 2。空间门控单元（SGU）的细节。

对于 f^t 阶段 t的输入特征，我们设计了一个空间门控单元（SGU），用于自适应地生成空间重要性掩码 $i\ mask^t$，如图2所示。为了生成$i\ mask^t$，采用 3×3卷积从 f^t中初步提取空间特征。随后，采用带有 sigmoid 操作的 1×1 卷积来生成 $i\ mask^t$的权重。通过 sigmoid 激活，$i\ mask^t$ 中针对 f^t的不同像素会产生不同的权重，且大多数权重趋近于 0 或 1。

$$i_mask^t = \text{Sigmoid}(C^1(C^3(f^t))) \tag{1}$$

通过与另一分支中使用 3×3 卷积提取的对应空间特征相乘，$i\ mask^t$ 可以发挥门控作用，抑制 f^t中的不重要特征。与恒等特征 f^t相加后，即可得到门控特征 f_g^t:

$$f_g^t = i_mask^t \cdot C^3(f^t) + f^t \tag{2}$$

也会将 $i\ mask^t$ 传入我们的空间缩放网络（SSN），以引导空间缩放因子的生成。

2.2. 空间缩放网络

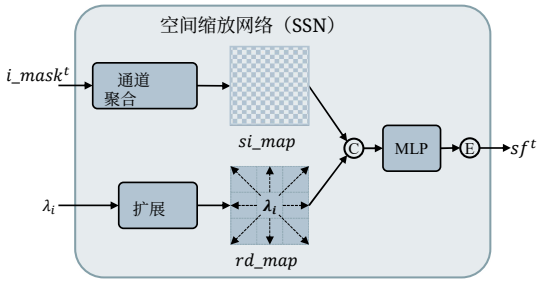


图 3。空间缩放网络（SSN）的结构。

所提出的空间缩放网络（SSN）如图3所示。为引入用于指导特征缩放和率失真优化的空间重要性掩码，SSN 将自适应学习得到的$i\ mask^t$ 和率失真权衡 λ_i 作为联合输入。首先，应将这两个输入的空间分辨率转换为与门控特征 f_g^t相同的分辨率。我们聚合 $i\ mask^t \in \mathbb{R}^{H\times W\times N}$ 的通道信息，并生成空间重要性图$si\ map \in \mathbb{R}^{H\times W\times 1}$。由于 λ_i 和H 会随输入图像尺寸的不同而变化，因此无法通过学习直接转换常数 W。我们的解决方案是将 λ_i 平铺并扩展为 $rd_map \in \mathbb{R}^{H\times W\times 1}$。随后，我们按通道拼接这些输入，并将其输入多层感知机网络，以生成空间缩放因子 sf^t，该因子同时与空间重要性和比特率相关。多层感知机由两个包含 64 个隐藏单元的全连接层组成，中间使用 ReLU。其后应用指数函数以确保输出为正，这有利于训练过程 [9]。

$$sf^t = \exp(MLP(Concat(si_map, rd_map))) \tag{3}$$

通过与 sf^t 进行空间逐点相乘，f_g^t 的不同位置会按不同精细度进行缩放，从而得到缩放后的特征 f_s^t。这样，不同

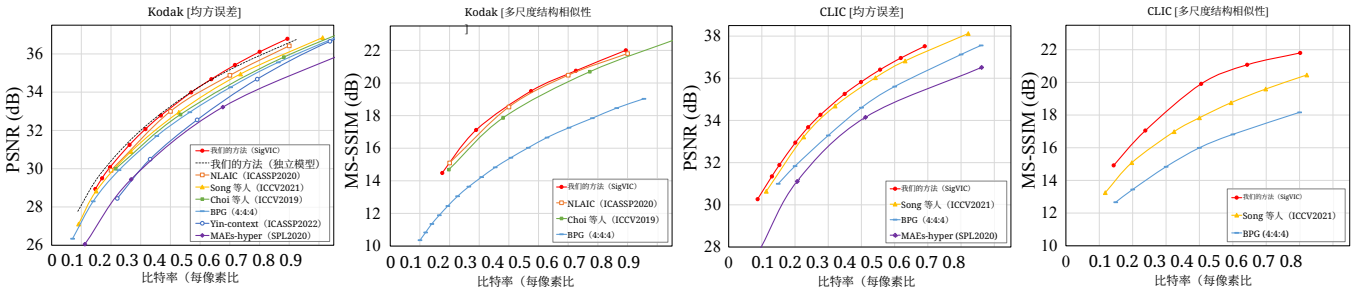


图5。在 Kodak 和 CLIC 数据集上，以 MSE 和多尺度结构相似性作为损失函数的失真项得到的率失真曲线。

比特率的比特分配便与量化后的空间重要性相关。

2.3. Top-K浅层特征融合模块

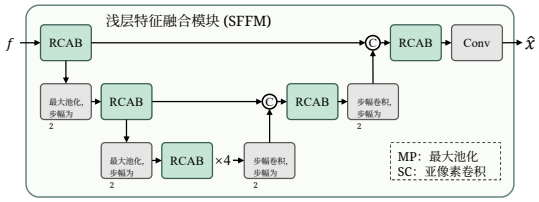


图4。浅层特征融合模块（SFFM）的结构。

为丰富解码图像 $\hat{x}$的细节，我们选取富含边缘和纹理信息的 Top-K 浅层特征，并将其融入编码特征 y。通过浅层特征融合模块（SFFM），解码器中会恢复这些特征，以细化解码特征。由于特征中的大部分信息仅由少数通道贡献[14], ，因此我们选取信息量最大的 K 特征图，以平衡性能和计算量。

SFFM采用了如图所示的U形网络。4多个残差通道注意力块（RCAB）[15]用于学习通道信息。此外，SFFM仅需少量参数，即可在整个网络中进行端到端训练。

2.4. 损失函数

我们的网络接收可变的 λ_i，以适应任意码率，同时自适应地完成空间重要性引导。损失函数定义如下:

$$L = R(\hat{y}, \hat{z}, \lambda_i) + \lambda_i D(x, \hat{x}, \lambda_i) \ \ (\lambda_i \in \Lambda) \qquad (4)$$

其中，Λ 表示两个选定边界值之间的范围，该范围包含 λ_i 的所有可能取值。

3. 实验

我们使用 ImageNet[16] 的一个子集训练模型，其中包含 13600 张图像，并将图像随机裁剪成图块

表1。Kodak 和 CLIC 上与 BPG 的 BD 结果对比

方法	Kodak		CLIC	
	BD-PSNR	BD-Rate	BD-PSNR	BD-Rate
MAEs-hyper[9]	-0.96 分贝	21.35%	-0.59 分贝	14.88%
Yin-context[12]	-0.87 分贝	17.30%	--	--
Choi 等人[10]	0.22 分贝	-4.96%	--	--
Lee 等人[4]	0.22 分贝	-5.05%	0.56 分贝	-12.25%
Song 等人[11]	0.37 分贝	-8.06%	1.05 分贝	-21.73%
NLAIC[18]	0.71 分贝	-14.50%	--	--
Cheng 等人[5]	0.78 分贝	-16.43%	1.21 分贝	-25.82%
我们的方法（SigVIC）	**0.92 分贝**	**-17.84%**	**1.23 分贝**	**-26.16%**

分辨率为 256×256。批量大小设为 8，模型使用 Adam 优化器[17]训练 $2M$ 次迭代。学习率初始设为 1×10^{-4}，并在最后 $0.1M$ 次迭代中降低至 1×10^{-5}。我们使用均方误差和多尺度结构相似性损失训练模型，为覆盖广泛的比特率，所选 Λ 的边界分别为 (0.0016, 0.045) 和 (5, 120)。经验上，选取的浅层特征数量 K 设为 32，网络的滤波器数量 N 设为 192。

3.1. 率失真（RD）性能

我们将本方法的率失真曲线与其他可变码率压缩方法进行比较，包括基于学习的方法[9, 10, 11, 12, 18] 和传统编解码器 BPG[8]。如图5所示，在 Kodak $-10log_{10}$（1 −多尺度结构相似性）（为便于说明，数值已转换为分贝）的峰值信噪比和多尺度结构相似性方面，本方法均优于其他所有方法。我们在 Kodak[19] 和 CLIC[20] 数据集上进行了测试。

我们进一步将 BD 结果（即 BD-PSNR 和 BD-Rate）与上述方法以及一些需要针对不同比特率训练多个模型的知名基于学习的压缩方法进行比较[4, 5] 。更高的 BD-PSNR 和更低的 BD-Rate 表示更好的率失真性能，对应于更大的 PSNR 增益和比特率节省。如表1所示，所有结果均以 BPG 为基准计算。显然，我们的方法在 Kodak 数据集上实现了最高的 0.92 分贝 PSNR 增益，并节省了最多的比特率（17.84%），优于其他所有方法。在分辨率更高的 CLIC 数据集上，我们的方法获得了更高的 BD-PSNR（1.23 分贝），并节省了更多比特率（26.16%）。

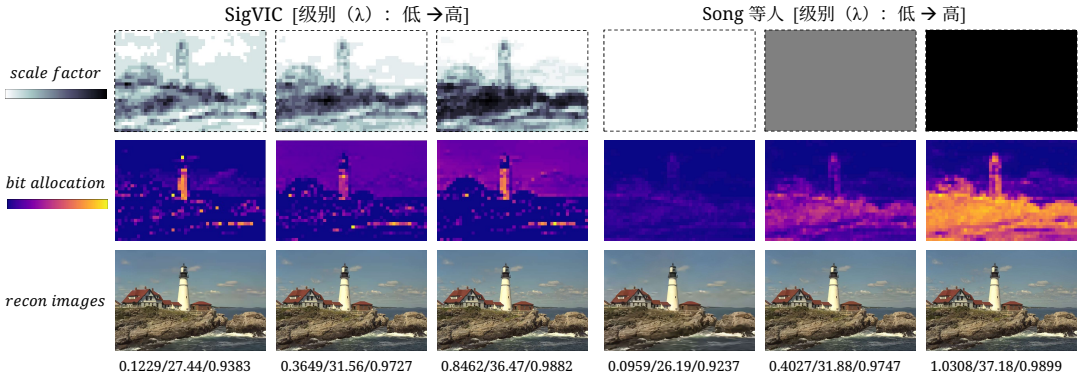


Fig. 7。我们的 SigVIC 和 Song 等人的可视化（包括缩放因子和比特分配）

此外，我们的结果略优于 Cheng 等人[5]，其性能与通用视频编码[21]相当。

为了测试可变比特率机制的鲁棒性，我们还训练了 6 个不使用 SSN 的独立模型。如*Kodak[均方误差]* 结果（见图5）所示，我们的 SigVIC 达到了与相应独立模型相当的性能。这表明，我们的可变比特率机制在提升压缩网络效率和灵活性的同时，不会影响其性能。

3.2. 主观性能

图6 展示了一些重建图像，用于评估我们方法的主观性能。它们的比特率均相近，约为每像素 0.22 比特。从放大区域可以看出，我们的方法能够还原更多细节，例如头盔和裤子上的文字、摩托车轮胎和挡泥板的纹理等。

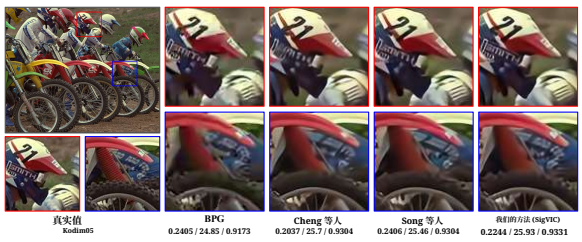


图 6。重建图像的视觉质量。

3.3. 可视化

为了评估空间重要性引导的特征缩放和比特分配的有效性，我们可视化了在若干率失真权衡水平下的缩放因子及其对应的比特分配图 λ。如图7所示，我们的缩放因子能够自适应地学习图像的空间重要性，并以不同权重缩放不同区域的特征，而 Song 等人的质量图[11]则使用相同权重缩放不同区域。因此，在比特率受限时，我们的方法能够在空间重要性的引导下，优先为信息丰富的区域分配更多比特。

3.4. 消融实验

我们研究了 SGU 和 SFFM 在所提方法中的有效性，结果如表2所示。方案A在可变码率机制中采用统一缩放因子，使用 SSN，但不使用 SGU。与方案A相比，方案B的比特率降低了 3.56%，这得益于 SGU 利用空间重要性指导比特分配和率失真优化。此外，方案C进一步节省了 6.92% 的比特率。这表明，采用 SFFM 的 Top-K 浅层特征融合策略能够进一步提升解压缩图像的质量和率失真性能。

表 2。消融实验结果

方案	SSN	SGU	SFFM	BD-Rate	参数
A	✓			0%	15.3M
B	✓	✓		-3.56%	23.21M
C	✓	✓	✓	-6.92%	24.85M

4. 结论

本文提出了一种空间重要性引导的可变码率图像压缩方法，称为 SigVIC。具体而言，我们设计了空间门控单元（SGU）和空间缩放网络（SSN），利用空间重要性引导特征缩放和可变码率的比特分配。此外，我们采用浅层特征融合模块（SFFM），利用 Top-K 浅层特征细化解码特征。实验结果表明，我们的方法取得了最先进的性能，同时显著节省存储空间，并提升了图像压缩方法的灵活性。

5. 致谢

本研究部分得到了国家自然科学基金委员会中国基金项目 61972028、基金项目 61902022 和基金项目 62120106009 的资助。

6. 参考文献

[1] Johannes Ball´e、Valero Laparra 和 Eero P Simoncelli，《端到端优化图像压缩》，*arXiv* 预印本 *arXiv:1611.01704*，2016年。

[2] Johannes Ball´e、David Minnen、SaurabhSingh、Sung Jin Hwang 和 Nick Johnston，《基于尺度超先验的变分图像压缩》，*arXiv* 预印本 *arXiv:1802.01436*，2018年。

[3] 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》，*arXiv* 预印本 *arXiv:1809.02736*，2018年。

[4]Jooyoung Lee、Seunghyun Cho 和 Seung-Kwon Beack，《面向端到端优化图像压缩的上下文自适应熵模型》，*arXiv* 预印本 *arXiv:1809.10452*，2018年。

[5] Zhengxue Cheng、Heming Sun、Masaru Takeuchi 和 Jiro 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（*CVPR*），2020年，第7939–7948页。

[6] Gregory K Wallace，《JPEG 静态图像压缩标准》，*IEEE*消费电子学汇刊，第38卷，第1期，第xviii–xxxiv页，1992年。

[7]Athanassios Skodras、Charilaos Christopoulos 和 Touradj 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年。

[8] Gary J Sullivan、Jens-Rainer Ohm、Woo-Jin Han 和 Thomas 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年。

[9] Fei Yang、Luis Herranz、Joost Van De Weijer、Jos´e A Iglesias Guiti´an、Antonio M L´opez 和 Mikhail G Mozerov，《采用调制自编码器的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。

[10] Yoojin Choi、Mostafa El-Khamy 和 Jungwon Lee，《采用条件自编码器的可变码率深度图像压缩》，载于*IEEE* 国际计算机视觉会议论文集（*ICCV*），2019年，第 3146–3154 页。

[11]Myungseo Song、Jinyoung Choi 和 Bohyung Han，“通过空间自适应特征变换实现可变码率深度图像压缩”，载于论文集 *IEEE*国际计算机视觉会议（*ICCV*），2021年，第2380–2389 页。[12] Shanzhi Yin、Chao Li、Youneng Bao、YongshengLiang、Fanyang Meng 和 Wei Liu，“通用高效可变码率神经图像压缩”，载于*IEEE*声学、语音与信号处理国际会议（*ICASSP*）论文集，2022年，第 2025–2029 页。

[13]Rushil Gupta、Suryateja BV、Nikhil Kapoor、Rajat Jaiswal、Sharmila Reddy Nangi 和 Kuldeep Kulkarni，《用户引导的可变码率学习式图像压缩》，载于*IEEE/CVF* 计算机视觉与模式识别会议论文集（*CVPR*），2022年，第 1753–1758 页。

[14] Dailan He、Ziming Yang、Weikun Peng、Rui Ma、Hong-wei Qin 和 Yan 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论文集（*CVPR*），2022年，第 5718–5727 页。

[15] Yulun Zhang、Kunpeng Li、Kai Li、Lichen Wang、Bi- neng Zhong 和 Yun Fu，“利用超深残差通道注意力网络进行图像超分辨率”，载于欧洲计算机视觉会议（*ECCV*）论文集，2018年，第 286–301 页。

[16] Jia Deng、Wei Dong、Richard Socher、Li-Jia Li、Kai Li 和 Li Fei-Fei，“ImageNet：大型层次化图像数据库”，载于*IEEE/CVF*计算机视觉与模式识别会议（*CVPR*）论文集，2009年，第 248–255 页。

[17]Diederik P Kingma 和 Jimmy Ba，“Adam：一种随机优化方法”，*arXiv*预印本 *arXiv:1412.6980*，2014年。

[18] Tong Chen 和 Zhan Ma，“采用质量缩放因子的可变码率图像压缩”，载于*IEEE*国际声学、语音与信号处理会议（*ICASSP*）论文集，2020年，第 2163–2167 页。

[19] “KodakPhotoCD 数据集。 [在线],” 2013年，获取地址：http://r0k.us/graphics/kodak/。

[20] “CLIC数据集。 [在线],” 2019年，获取地址：http://www.compression.cc/。

[21] Jens-Rainer Ohm 和 Gary J Sullivan，“通用视频编码——迈向新一代视频压缩”，载于图像编码研讨会（*PCS*）论文集，2018年，第2018卷。