DOI: 10.3979/j.issn.1673-825X.2020.05.009

基于注意力机制和离散高斯混合模型的端到端图像压缩方法

朱　俊[1,2],高陈强[1,2],陈志乾[1,2],谌　放[1]
(1.重庆邮电大学 通信与信息工程学院,重庆 400065; 2.信号与信息处理重庆市重点实验室,重庆 400065)

摘　要: 图像压缩是图像处理领域重要的基础支撑技术之一。近年来,深度学习被用于解决图像压缩任务。潜在表示特征的冗余和概率估计的不准确往往会限制压缩性能的进一步提高。为了改善这类问题,提出一种基于注意力机制和离散高斯混合模型的端到端图像压缩方法。将全局上下文注意力模块嵌入到编码器,旨在构造紧凑的潜在表示特征。同时,将潜在表示特征建模为参数化的离散高斯混合模型,用于提高码率估计的准确度。实验结果表明,提出的算法无论在峰值信噪比(peak signal noise rate,PSNR)还是多尺度结构相似度(multi-scale structural similarity,MS-SSIM)指标上都高于传统方法。在视觉感知上,提出的图像压缩算法能产生更令人满意的压缩图像。

关键词: 图像压缩; 自编码器; 卷积神经网络; 深度学习
中图分类号: TN919.8; TP391.4　　**文献标志码:** A　　**文章编号:** 1673-825X(2020)05-0769-10

End-to-end image compression method based on attention modules and discretized Gaussian mixture model

ZHU Jun[1,2], GAO Chenqiang[1,2], CHEN Zhiqian[1,2], CHEN Fang[1]
(1. School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, P. R. China;
2. Chongqing Key Laboratory of Signal and Information Processing, Chongqing 400065, P. R. China)

Abstract: Image compression is one of the important basic technologies in the image processing field. In recent years, deep learning is used to handle image compression task. However, the redundancy of the latent representation feature and inaccurate probability estimation usually limit the compression performance. To address these problems, this paper proposes an end-to-end image compression method based on attention mechanism and discrete Gaussian mixture model. Firstly, this paper embeds the global context attention module into the encoder to construct compact latent representational features. Besides, this paper models the latent features as a parameterized discrete Gaussian mixture model to improve the accuracy of rate estimation. Experimental results demonstrate that the proposed method outperforms traditional method in terms of peak signal noise rate (PSNR) and multi-scale structural similarity (MS-SSIM). In terms of visual perception, the proposed method is able to produce more satisfying visual results.

Keywords: image compression; auto encoder; convolutional neural network; deep learning

收稿日期: 2020-06-30　**修订日期:** 2020-09-14　**通讯作者:** 朱　俊　iszhujun@qq.com
基金项目: 国家自然科学基金(61571071, 61906025); 重庆市科委自然科学基金(cstc2018jcyjAX0227)
Foundation Items: The National Natural Science Foundation of China (61571071, 61906025); The Natural Science Foundation Project of Chongqing Science and Technology (cstc2018jcyjAX0227)

0 引 言

图像压缩是信息技术领域的重要基础支撑技术之一,也是计算机视觉领域的研究热点之一。图像压缩的目的是用尽可能紧凑的形式表示原有图像数据,以节省传输和存储开销。

传统的图像压缩方法,如联合图像专家组(joint photographic experts group, JPEG)[1],JPEG 2000[2],WebP[3],BPG(better portable graphics)[4]等广泛应用于我们日常工作生活中的各个方面。然而,这些方法采用人工设计的编解码器,各个模块之间有大量的参数需要单独调优,难以根据图像内容进行自适应调节,限制了性能的进一步提升。因此,研究更好的图像压缩算法具有广泛的现实意义和重要的应用价值。

近年来,深度神经网络(deep neural networks, DNNs)在图像分类、图像修复、图像质量增强等多个领域取得明显的效果。这给有损图像压缩方法提供了新的研究思路,基于深度学习的有损图像压缩成为了最新的研究热点。

到目前为止,相关研究人员开展了大量的研究工作。一部分工作[5-6]使用递归神经网络(recurrent neural networks, RNNs)对残差信息进行递归压缩,实现可分级编码。该类方法的图像压缩比是由潜在表示特征的大小和网络迭代次数共同决定,属于渐进式编码方式。另外一些工作[7-8]用生成对抗网络(generative adversarial networks, GANs)来处理图像压缩任务。这类方法可以在生成极低压缩码流的同时,获得不错的重建图像质量。然而,使用GANs的这类方法难以控制生成图像细节,容易丢失原始图像的部分细节信息。除了上述两类方法外,更具代表性的方法[9-14]致力于使用自编码器结构的神经网络来提升图像压缩性能。该结构首先将图像像素转换到一个具有恢复图像关键信息且冗余度较低的潜在表示空间,然后依次对潜在特征进行编码和解码。在这些方法中,转换编解码、量化、熵编码等模块以端到端训练的方式进行联合优化。因此,图像压缩系统的设计与改进也集中在这3个方面。对于量化来说,由于端到端的训练要求在反向传播中所有模块可微,文献[10]提出在训练时加入均匀噪声扰动来替代原本不可微的取整量化方式。文献[13]提出一种从软到硬量化的方式来近似不可微的量化过程。

对于转换编解码的网络设计,大多数工作通过堆叠卷积块来实现[9-10,15-16],然而在感受野受限的情况下执行局部卷积,难以兼顾到特征之间的全局关联性。进一步,由于卷积操作共享特征,所有特征都被视为同等重要,并没有考虑到人眼对于不同内容的视觉敏感度,难以形成紧凑的潜在表示空间。Li等[14]提出了基于内容加权的图像压缩方法,通过在图像压缩框架中训练一个3层卷积网络分支来学习不同内容的重要性映射,该重要性映射用于指导图像码率分配。但这种显式的学习内容重要权重增加了计算开销,而且对于深层特征难以自适应分配比特。近年来,注意力机制在特征重要性的自适应学习方面表现出了很大的优势,在自然语言处理[17]和语义分割[18-20]等任务上都取得了明显的效果。为此,本文提出将全局上下文注意力模块(global context attention block, GCAB)嵌入到转换编码器中,使特征具有全局自适应性响应,强化重要特征,弱化不重要特征,从而隐式地学习特征重要性映射,同时不显著增加额外计算开销,进一步提升压缩性能。

对于熵编码,精确的熵率估计是提升图像压缩性能的关键之一。Ballé等[9]提出使用线性分段函数构建的熵模型进行码率估计。随后,进一步引入高斯模型作为潜在特征的先验分布,并通过超先验网络得到熵模型的尺度参数[10]。得益于更加准确的码率估计,该工作取得了更好的压缩性能。Lee等[15]利用了2种类型的上下文,即消耗比特的上下文和无消耗比特的上下文,这使得该模型可以用更广义的近似模型形式更加准确地估计潜在表示空间的分布。Minnen等[16]将具有超先验的高斯模型和自回归掩码卷积相结合,进一步提高了基于学习的图像压缩性能上限。因此,精确的熵模型有助于提高码率估计的准确性,从而极大地提升图像压缩性能。由于高斯混合模型具有强大的分布近似能力,受到文献[16]的启发,本文引入离散高斯混合模型来构建一个更加精确和灵活的熵模型。

综上所述,本文引入了全局上下文注意力模块[18]和离散高斯混合模型来进一步改善图像压缩性能。具体来说,本文将全局上下文注意力模块嵌入到转换编码中,用于产生具有全局空间自适应激活的潜在特征。同时,引入一个参数化的离散高斯混合模型用于构建更加精确和灵活的熵模型。实验结果表明,与传统图像压缩方法和与基于深度学习的方法[10]相比,本文提出的方法具有更高的性能指标和更好的主观视觉感受。

1　基于端到端优化的图像压缩框架

本文方法的框图如图 1,其基本思路是在编码端利用深度卷积自编码器对图像进行压缩编码,得到潜在表示特征,然后将潜在表示特征进行无损熵编码后得到用于传输或存储的压缩码流。解码端利用对称的网络结构重建出原始图像。

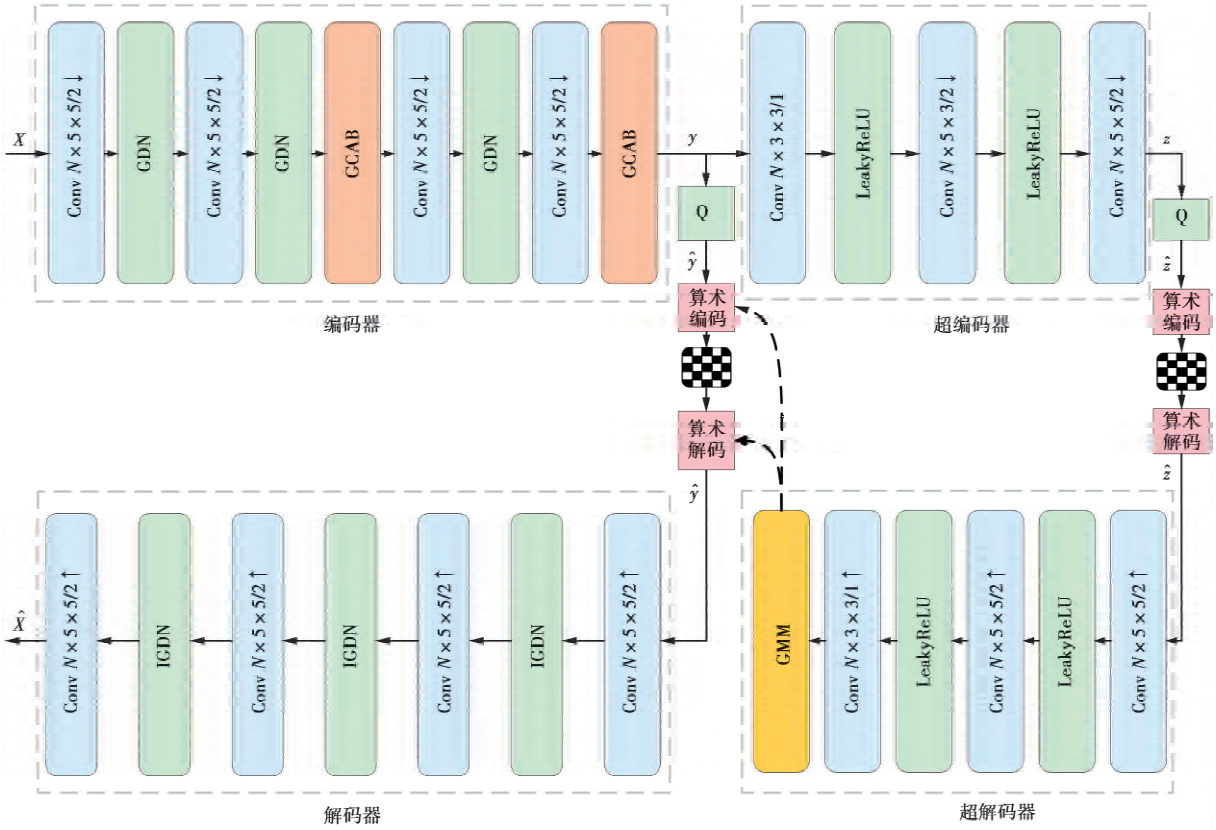


图 1　本文方法的框图

Fig.1　Framework of the proposed method

具体来说,给定一组训练图像 X,我们希望学习一个包含编码器、量化器和解码器的图像压缩系统。编码器 E 用于学习一个更好的转换编码器以较少原始图像的冗余,输出潜在表示特征 $y:=E(X)$。量化器 Q 将 y 离散化,$\hat{y}:=Q(y)$。由于 $\hat{y}$ 是离散值的,因此,可以使用熵编码技术(如游程编码、算术编码等)对其进行无损压缩,形成二进制码流后进行传输。在编码阶段,我们希望 $\hat{y}$ 在形式上是尽可能紧凑的,这样在无损熵编码时就可以用较少的比特去表示。解码器从压缩码流中接收到信号 $\hat{y}$,通过合成变换 D 得到重建图像 $\hat{X}:=D(\hat{y})$。在解码阶段,我们希望原始图像 X 和重建图像 $\hat{X}$ 的差异性 $d(X,\hat{X})$ 尽可能小,大多数工作中使用均方误差(mean square error, MSE)或多尺度结构相似度(multi-scale structural similarity,MS-SSIM)表示失真程度 d。从信息论的角度来说,一方面当编码特征越集中时,码字数量会降低,信息熵会进一步降低,但网络的表示能力会受影响,导致重建图像的质量降低;另一方面,当编码特征的信息熵越大,说明用于恢复图像的信息越多,重建后的图像质量越高,但相应的码字数量会增多。因此,需要在熵编码后的编码码率与图像重建质量之间做权衡,在图像压缩中,这种权衡称为率-失真优化。通常通过率-失真优化构建的损失函数对自编码图像压缩网络进行训练。损失函数定义为

$$Loss = d(X,\hat{X}) + \lambda H(\hat{y}) \tag{1}$$

(1)式中:λ 是用于调节权衡失真和码率的权重;H 表示编码 $\hat{y}$ 平均码长的理论下界,即 $\hat{y}$ 的信息熵。

本文把编码器和解码器建模为卷积神经网络。由于卷积核的感受野是局部的,要经过累积很多层之后才能把整个图像不同部分的区域关联起来。为了捕获特征之间的局部和全局依赖关系,强化重要特征,本文提出将全局上下文注意力模块集成到编

码端中。与此同时,为了获得更精确的熵模型,本文把潜在表示空间建模为一个参数化的离散高斯混合模型,并利用超先验网络预测模型参数。

1.1 具有注意力机制的转换编码器

传统方法中通过改进转换编解码器获得更好的压缩性能。转换编码器的设计通常通过堆叠卷积层来实现。基于学习的图像压缩中,通过卷积神经网络学习一个具有较少冗余信息又具有关键重构信息的转换编码器是获得更好压缩性能的关键之一。然而,在卷积神经网络中,卷积层只能建立局部范围中像素之间的关系,难以考虑到全局像素之间的关联,增加网络的深度可以获得更深层的依赖关系,但显著增加了模型的参数量和计算开销。视觉注意力机制从人类视觉机制出发,通过自适应学习不同特征的权重,获取需要重点关注的区域。为此,本文在转换编码阶段引入注意力机制,通过注意力机制建模特征之间的全局依赖关系,从而在转换编码端形成更加紧凑的图像潜在表示特征。

受到文献[18]启发,本文提出,在图像编码端嵌入全局上下文注意力模块来隐形地捕捉重要的特征,弱化不重要的特征。本文的编码器和解码器由卷积和非线性函数组合。其中,本文使用广义分歧归一化(generalized divisive normalization, GDN)变换和逆广义分歧归一化(inversed generalized divisive normalization, IGDN)变换作为非线性函数。GDN/IGDN 已被验证适合于概率建模和图像压缩任务[9-10]。本文在变分自编码的网络结构中引入 GCAB 用于捕捉局部和全局的像素之间的依赖关系,同时强化重要特征的响应值,弱化不重要特征的响应值,从而在转换编码端形成紧凑特征结构。在转换解码端,需要尽可能恢复潜在表示特征中的信息,并不需要对特征进行选择性增强或弱化,解码端所有特征信息视为同等重要,因此,本文在解码端没有加入对称的注意力模块。

全局上下文注意力模块的结构如图 2。图 2 中,x 表示 GCAB 的一个输入特征实例,$x=\{x_i\}_{i=1}^{N_p}$。其中,x_i 表示 x 在位置 i 处的特征值,N_p 是特征的数量。同理,z_i 表示输出注意力掩码 z 在位置 i 处的注意力值。z_i 公式化可表示为

$$z_i = W_{v2}\delta\left[\left(W_{v1}\sum_{j=1}^{N_p}\frac{e^{W_k x_j}}{\sum_{m=1}^{N_p}e^{W_k x_m}}x_j\right)\right] \tag{2}$$

(2)式中:W_k 表示图 2 上下文建模单元中的 1×1 卷积操作,通过上下文建模单元进行全局注意力池化,用于全局上下文建模;W_{v1},W_{v2}分别表示图 2 转换单元中的 1×1 卷积操作;$\delta(\cdot)$表示进行层归一化操作(layer normalization,LN)和线性整流函数(rectified linear unit,ReLU)的非线性映射操作。通过转换单元捕获特征不同通道间的依赖,输出特征 x 对应的注意力掩码 z。具有全局上下文信息的注意力特征为 x',由(3)式得到。将原始输入特征 x 和注意力掩码 z 按照广播机制相加就得到了具有全局上下文信息的注意力特征 x',其中,x 和 x'具有相同的维度。

$$x' = x + z \tag{3}$$

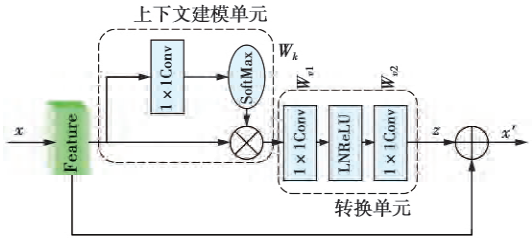


图 2 全局上下文注意力模块

Fig.2 Global context attention module

1.2 量化器

在图像压缩中,量化带来了信息的损失,不利于解码端图像恢复。但是量化减少了信息熵,有助于压缩成更小的比特流。本文中,量化方法遵循文献[9-10]在训练时采用加入服从均匀分布的噪声扰动去替代标准的取整量化操作,在测试时采用标准的取整量化。

1.3 码字估计模块

在为了达到提升压缩效率的目的,需要用码字估计模块在训练中对码率进行约束。码字估计通常利用先验概率模型对编码特征分布进行准确估计,保证估计的编码特征分布与实际分布尽可能接近,然后通过计算 $\hat{y}$ 的信息熵的结果估算理想熵编码后的码字大小。在文献[10]中通过引入边信息 z 来捕捉 $\hat{y}$ 的空间依赖关系。对于量化后的潜在表示特征 $\hat{y}$,其中每个元素表示为 $\hat{y}_i$,$\hat{y}_i$被建模为均值为 μ_i,方差为 σ_i 的条件高斯分布,如

$$p_{\hat{y}|\hat{z}}(\hat{y}\mid\hat{z}) = \prod_i \left[N(\mu_i,\sigma_i^2) * U(-0.5,0.5)\right](\hat{y}_i) \tag{4}$$

(4)式中,假定均值 μ_i 为 0;方差 σ_i 由 z_i 作为其先

验预测得到; $\hat{z}$ 称为超先验; $U(-0.5,0.5)$ 表示参数为 $(-0.5,0.5)$ 的均匀分布,用于模拟取整的量化操作引入的噪声扰动 "$*$"表示卷积运算。因为 $\hat{z}$ 的先验信息不存在,$\hat{z}$ 被建模为一个非参数、完全因式分解的模型,其公式化描述可写为

$$p_{\hat{z}|\psi}(\hat{z}\mid\psi)=\prod_i p_{z_i|\psi_i}(\psi_i) * u(-0.5,0.5)(\hat{z}_i) \tag{5}$$

(5) 式中: ψ_i 表示每个单变量分布 $p_{z_i|\psi_i}(\psi_i)$ 的参数; $\hat{z}_i$ 用于预测方差 σ_i,同时占用少量的比特数作为边信息传递到解码端。本文用信息熵来估计无损熵编码时的最小码流,即

$$R_{\hat{y}}=-\sum_i \mathrm{lb}\left(p_{\hat{y}_i|\hat{z}_i}(\hat{y}_i\mid\hat{z}_i)\right) \tag{6}$$

$$R_{\hat{z}}=-\sum_i \mathrm{lb}\left(p_{\hat{z}_i|\psi_i}(\hat{z}_i\mid\psi_i)\right) \tag{7}$$

最终的压缩后码率表示为

$$R=R_{\hat{y}}+R_{\hat{z}} \tag{8}$$

虽然单一的高斯熵模型与 Ballé 等之前的工作[10]相比,压缩性能有提升,但是单一的高斯模型表达能力始终有限。高斯混合模型相比单一的高斯模型更加灵活,具有更强大的数据分布近似能力。通过增加高斯模型中子高斯模型的数量,可以近似任何连续的概率分布。与此同时,率失真优化的图像压缩框架中,码率估计的准确性影响着整个端到端系统的参数优化。显然,越准确的码率估计能够学习到更合适的模型参数,最终得到性能更好的压缩模型。因此,本文对文献[10]中的单一高斯熵模型进行改进,提出使用离散高斯混合模型进行码率估计,进一步地提高图像压缩系统的性能。所以,(4) 式可以被改写为

$$p_{\hat{y}|\hat{z}}(\hat{y}\mid\hat{z})=\prod_i\left[\sum_{k=1}^{K}\omega_i^{(k)}N(\mu_i^k,\sigma_i^{2(k)})*U(-0.5,0.5)\right](\hat{y}_i) \tag{9}$$

(9) 式中: i 表示特征图的位置; ω_i 表示不同高斯模型的权重; μ_i 表示高斯模型的均值; σ_i 表示高斯模型的方差; K 表示混合模型中子高斯模型的数量。因为可以获得大量的训练数据以及神经网络具有强大的非线性拟合能力,所以设计了一个卷积神经网络模块用于预测离散混合高斯模型中的 3 种参数 ω_i, μ_i, σ_i。其模型框图如图 3。该模块由 3 个卷积层和带泄露线性整流函数(leaky rectified linear unit,LeakyReLu)层堆叠而成。

在本文实验中,K 设置为 3,N 表示图 1 中的通道数量,所以离散高斯混合模型的输出通道数为 $9N$。其中,前 $3N$ 个通道用来预测每个符号的高斯模型的权重参数,后 $6N$ 个通道分别用来预测均值和方差。为了正确预测每个高斯模型的权重,对前 $3N$ 个通道添加 sigmoid 层,保证对于同一个符号不同高斯模型的权重之和为 1。为了选择最合理的超参数 K,讨论了 K 分别为 2,3 和 4 时模型的收敛性,如图 4。可以看到,K 为 3 时,模型收敛时的总损失更小; K 为 4 时,模型反而收敛到更大的损失; 而当 K 为 1 时,模型退化为单一的高斯模型。因此,选择 K 为 3 作为最优超参数。

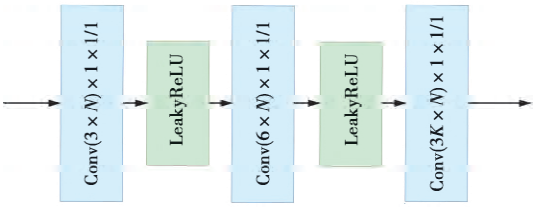


图 3　高斯混合模型模块结构

Fig.3　Structure of Gaussian mixture model module

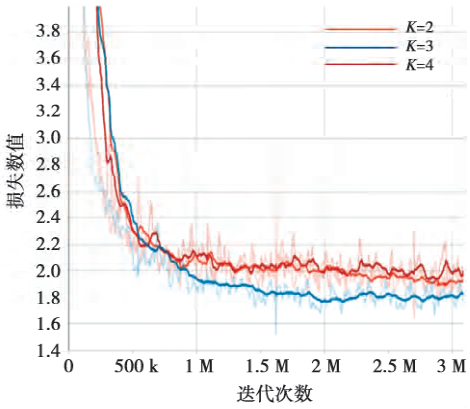


图 4　不同 K 值的模型收敛性比较

Fig.4　Comparison of model convergence with different K values

1.4　损失函数

本文实验中分别采用 MSE 作为图像失真度量,表示为

$$d(X,\hat{X})=\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[X(i,j)-\hat{X}(i,j)]^2 \tag{10}$$

(10) 式中,m 和 n 分别表示图像块的长和宽,实验中都设置为 256。损失函数为

$$Loss=R+\lambda d(X,\hat{X}) \tag{11}$$

(11) 式中,R 由潜在表示空间码字估计和用于超先验估计的边信息相加而得; λ 是用于权衡图像失真

和图像压缩比的权重参数，不同 λ 对应不同的压缩比。在训练时，通过设置不同的 λ 来得到具有不同压缩比的模型。图 5 给出了 λ 为 2 048 时压缩模型的损失(loss) 收敛过程。

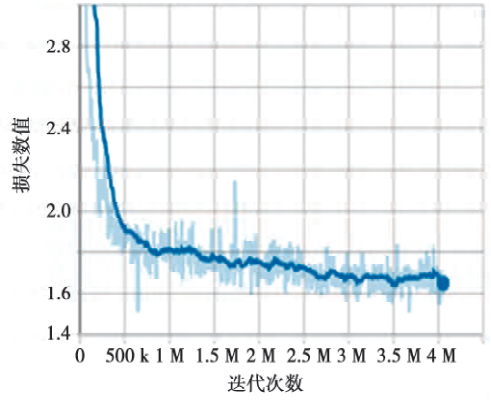


图 5 损失(loss) 的收敛性

Fig.5 Convergence of loss

2 实验与结果分析

2.1 实现细节

1) 实验环境。本文的实验环境是 Linux 系统，Pytorch [22] 深度学习框架，所有实验均在一块显存为 11 GByte 的 GeForce GTX 1080Ti 的 GPU 上进行。

2) 实验数据。由于图像压缩任务属于无监督任务，除了图像文件外不需要额外的标注文件，因此，大多数图像压缩工作 [7,9-10,23] 通过网页抓取高清图像数据或者使用公开的图像数据集(如 ImageNet [24] ，Cityscapes [25] 等) 作为模型训练集。为了验证本文方法的有效性，我们下载了来源于 Flick.com 网站的 20 745 张高质量的图像。和文献 [10] 的数据预处理过程保持一致，本文将原始图像像素随机裁剪为 256×256 的图像块，一共得到 864 216 张固定大小的图像，其中 860 000 张图片作为训练集，其余 4 216 张作为验证集。为了评估图像压缩模型性能，本文采用被广泛使用的 Kodak Photo CD 图像数据集 [26] 作为测试集。

3) 对比方法。本文将提出的方法与已知的标准图像压缩方法进行比较，如 JPEG [1] ，JPEG 2000 [2] ，WebP [3] 和 BPG [4] ，以及近期具有代表性的基于深度学习的图像压缩方法，如文献 [10]。JPEG 压缩使用 python 中的 PIL 库 [27] 实现。JPEG 2000 压缩使用配置为 YUV 420 的官方测试模型 OpenJPEG 2000 [28] 实现。对于 BPG，使用 BPG 软件 [4] 分别在 YUV440 和 YUV420 的格式下测试了压缩性能。YUV440 的表现要优于 YUV420，因为它能防止色彩空间转换时色彩成分丢失。文献 [10] 的数据来自于其发布的结果①。

4) 本文方法。文本所有模型在单个 GPU 上训练，批量大小为 4，模型通道数 N 设置为 192，采用 Adam [29] 进行参数优化。本文训练权重参数 λ 为 8 192 时对应的高比特模型作为预训练模型，其他 λ 则直接在预训练模型上进行微调。在训练 λ 为 8 192 的模型时，初始学习率设置为 1×10^{-4}，迭代 2×10^{6} 次后学习率衰减至 1×10^{-5}，继续迭代 5×10^{5} 次。对于其他模型，选取不同 λ，学习率设置为 1×10^{-5}，保持其他参数不变，对预训练模型进行微调，迭代 5×10^{5} 次。为了绘制率-失真性能曲线，选取了 8 个不同的 λ，并分别训练对应的模型。每一个率-失真点代表一个模型在测试数据集上的平均性能。将这些点用折线连接起来表示本文的图像压缩方法的整体性能趋势。文献 [10] 中，模型大小为 20.3 MByte。由于本文加入了全局上下文注意力模块和离散高斯混合模型模块，在参数量方面相较于文献 [10] 有略微增加，最终模型大小为 23.3 MByte。全局注意力模块几乎没有增加模型的参数量，参数量的增加主要由离散高斯混合模型模块引起。在训练时间上，保持上述同样的参数设置，文献 [10] 中在单 GPU 上训练一个良好的模型需要花费一个星期的时间。本文首先训练了高比特模型，需要的时间约为 192 h，剩余的其他模型在高比特模型上微调，只需要48 h 就可以收敛。

5) 评价指标。为了评估率-失真性能，使用平均每个像素需要的比特数量(bit per pixel，bpp) 来反映图像压缩比。对于固定长宽的图像，bpp 越小，相应的压缩比越高。灰度图像为 8 bpp，一张未经压缩的 RGB 图像为 24 bpp。对于压缩图像的失真度量，分别使用常见的峰值信噪比(peak signal to noise ratio，PSNR) 和 MS-SSIM 指标来评估。PSNR 指标可以反映图像的像素级失真程度，MS-SSIM 指标与人眼视觉感知有更高的相关性。为了更清晰地观察到不同方法压缩性能的差异，将 MS-SSIM 的值转换为分贝($-10\lg(1-\text{MS-SSIM})$) 为单位来绘制率-失真性能曲线。

① https://github.com/tensorflow/compression/tree/master/results/image_compression/kodak

2.2　压缩性能

图 6 分别绘制了本文方法和不同对比方法在 PSNR 指标(见图 6a)和 MS-SSIM 指标(见图 6b)下的率-失真性能曲线。可以看到,同样的压缩比下,无论是 PSNR 还是 MS-SSIM 指标,提出的方法都优于 JPEG,JPEG 2000,WebP,BPG 和文献[10]的方法。因此,本文提出的方法能有效提高图像压缩的客观指标,并且具有较高的灵活性。为了进一步验证模型的泛化性能,在 Tecnick 数据集[30]上测试了码率为 0.15 附近的平均 PSNR 和 MS-SSIM 指标,具体结果如表 1[30]。可以看到,本文提出的方法在 PSNR 和 MS-SSIM 指标上都优于传统方法,与文献[10]比较,在码率略低的同时,PSNR 指标依旧高于文献[10]。

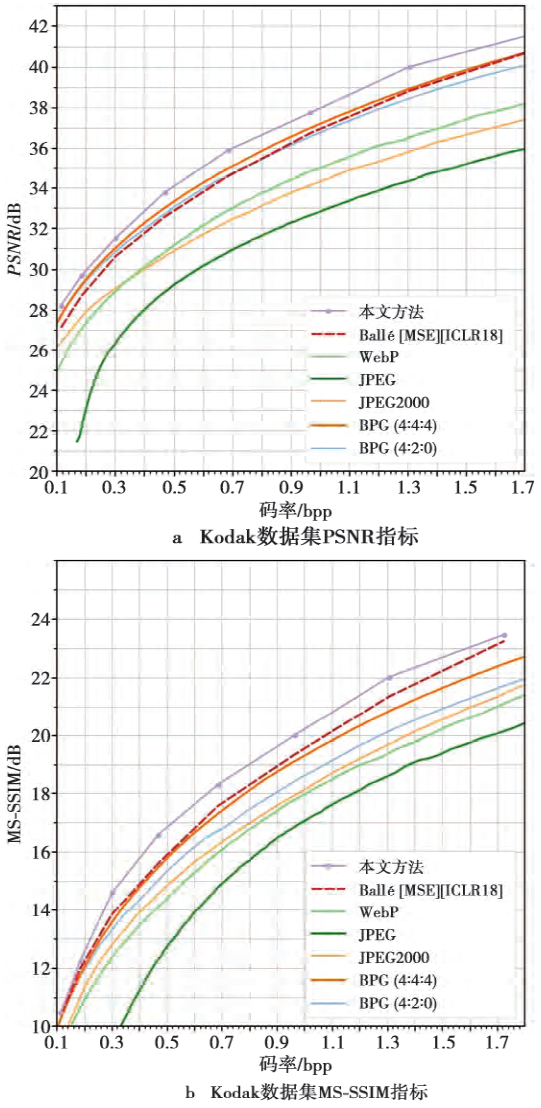
表 1　不同压缩方法在 Tecnick 数据集的性能

Tab.1　Performance of different compression methods on Tecnick datasets

模型	PSNR/dB	MS-SSIM	码率/bpp
JPEG	26.012	0.880	0.150
JPEG 2000	29.213	0.922	0.150
WebP	28.026	0.934	0.151
BPG(4:4:4)	31.107	0.956	0.150
文献[10]	30.862	0.957	0.152
本文方法	31.285	0.957	0.150

2.3　视觉比较

进一步地,图 7 展示了使用不同压缩方法对 Kodak Photo CD 图像数据集[24]中的部分图像压缩后的结果。由于基于神经网络的图像压缩方法难以严格限定压缩图像大小,压缩图像的大小会在给定目标比特附近微小波动,因此,图 7 中给出了不同压缩方法在较相近压缩比下的压缩图像比较。可以看到,本文方法在 bpp 略小于对比方法的同时,依然能获得较高的客观指标。在主观视觉感知上,传统的图像压缩方法中,特别是 JPEG,可以看到明显的图像失真现象。JPEG 2000 也存在明显的块效应。BPG 表现不错,在图 7b,图 7c 中,很容易观察到部分区域模糊和振铃效应。文献[10]是基于深度学习的方法,没有明显的压缩伪影产生。但是,在第 1 行绿色方框区域,可以看到墙壁的纹理信息有丢失;在第 4 行青色方框鹦鹉羽毛的区域有明显的模糊存在;在第 5 行绿色方框帽子区域的纹理模糊不清。相比之下,本文方法由于引入注意力机制和更准确的概率模型,对于纹理区域有更细腻的视觉效果。



a　Kodak数据集PSNR指标

b　Kodak数据集MS-SSIM指标

图 6　不同方法的率-失真性能曲线比较

Fig.6　Comparison of the rate-distortion curves by different methods

此外,可以看到第 1,3,5 行,本文方法产生了更加好的重建质量的同时,编码所需要的比特数相比 Ballé[10]消耗更少的比特数。一个可能的解释是 GCAB 学习了一个空间自适应的特征权重,导致编码时更多的比特分配到边缘或者纹理细节区域,较少的比特分配到平滑区域,以致于整体的编码方法使用更少的码字获得质量更好的重建图像。总体而言,在压缩比相近的情况下,本文方法有更好的人眼视觉感知质量。

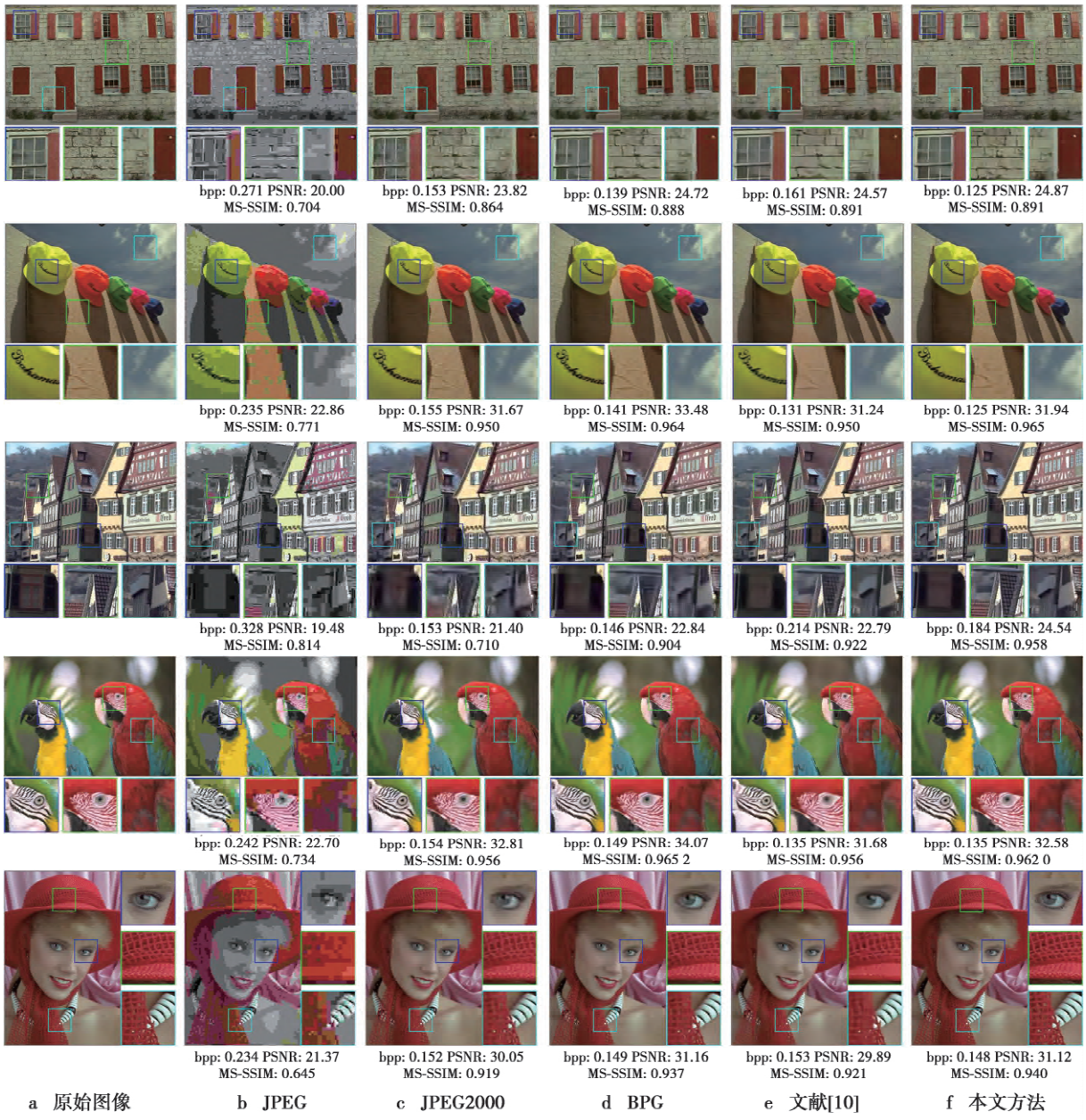


图7　不同压缩方法在不同压缩率下的图像比较

Fig.7　Image comparison by different compression methods at different compression rates

3　结束语

本文提出了一种基于全局上下文注意力机制和离散高斯混合模型的图像压缩方法。引入了全局上下文注意力模块来关注重点特征，弱化不重要特征以生成具有较少冗余的潜在表示特征，提高编码效率。此外，更准确的概率估计已被验证有助于压缩性能的提升，有助图像边缘和纹理区域细节信息更好地重建。因此，我们构建了一个参数化的离散高斯混合模型，并使用超先验网络预测离散高斯混合模型的参数。本文对网络模型进行了端到端的学习，各个模块的误差通过前向传播反映到损失函数，再通过反向传播优化参数，减少总误差。通过率-失真框架优化参数得到具有不同压缩比的模型，绘制了率-失真性能曲线和对不同的压缩方法进行了比较。实验结果表明，相较于JPEG，JPEG 2000，WebP，BPG等传统的编码标准，本文方法大幅提升了压缩性能。相较于现有的基于深度学习的图像压缩方法，如文献[10]，本文方法也有更高的指标和更好的视觉质量。

参考文献:

[1] WALLACE G K. The JPEG still picture compression standard [J]. IEEE transactions on consumer electronics, IEEE, 1992, 38(1): xviii-xxxiv.

[2] RABBANI M, JOSHI R. An overview of the JPEG 2000 still image compression standard [J]. Signal processing: Image communication, Elsevier, 2002, 17(1): 3-48.

[3] RABBAT R. A new image format for the Web | WebP | Google Developers [EB/OL]. (2020-05-28) [2020-06-29].https://developers.google.com/speed/webp/.

[4] BELLARD F. BPG Image format [EB/OL]. (2018-04-28) [2020-06-29].https://bellard.org/bpg/.

[5] TODERICI G, VINCENT D, JOHNSTON N, et al. Full resolution image compression with recurrent neural networks [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: CVPR, 2017: 5306-5314.

[6] JOHNSTON N, VINCENT D, MINNEN D, et al. Improved lossy image compression with priming and spatially adaptive bit rates for recurrent networks [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: ICCV,2018: 4385-4393.

[7] AGUSTSSON E, TSCHANNEN M, MENTZER F, et al. Generative Adversarial Networks for Extreme Learned Image Compression [C] //Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: ICCV, 2019: 221-231.

[8] RAMAN S K,RAMESH A, NAGANOOR V, et al. CompressNet: Generative Compression at Extremely Low Bitrates [C] //The IEEE Winter Conference on Applications of Computer Vision. [S.l.]: WCACV,2020: 2325-2333.

[9] BALLÉ J, LAPARRA V, SIMONCELLI E. End-to-end optimized image compression [C] //5th International Conference on Learning Representations. [S.l.]: ICLR, 2017.

[10] BALLÉ J, MINNEN D, SINGH S, et al. Variational image compression with a scale hyperprior [C] //International Conference on Learning Representations. [S.l.]: ICLR, 2018.

[11] THEIS L, SHI W, CUNNINGHAM A, et al. Lossy image compression with compressive autoencoders [EB/OL]. (2017-03-01) [2020-06-29]. https://arxiv.org/abs/1703.00395.

[12] RIPPEL O, BOURDEV L. Real-Time Adaptive Image Compression [C] //International Conference on Machine Learning. [S.l.]: ICML,2017: 2922-2930.

[13] AGUSTSSON E,MENTZER F, TSCHANNEN M, et al. Soft-to-hard vector quantization for end-to-end learning compressible representations [C] //Advances in Neural Information Processing Systems. [S.l.]: NeurIPS,2017: 1141-1151.

[14] LI M, ZUO W, GU S, et al. Learning convolutional networks for content-weighted image compression [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: CVPR 2018: 3214-3223.

[15] LEE J, CHO S, BEACK S K. Context-adaptive Entropy Model for End-to-end Optimized Image Compression [C] //International Conference on Learning Representations. [S.l.]: ICLR,2018.

[16] MINNEN D, BALLÉ J, TODERICI G D. Joint autoregressive and hierarchical priors for learned image compression [C] //Advances in Neural Information Processing Systems. [S.l.]: NeurIPS,2018: 10771-10780.

[17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C] //Advances in neural information processing systems. [S.l.]: NeurIPS,2017: 5998-6008.

[18] CAO Y, XU J, LIN S, et al. Gcnet: Non-local networks meet squeeze-excitation networks and beyond [C] //Proceedings of the IEEE International Conference on Computer Vision Workshops.Seoul: IEEE, 2019.

[19] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks [C] //Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.]: CVPR,2018: 7794-7803.

[20] HU J,SHEN L,SUN G. Squeeze-and-excitation networks [C] //Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. [S.l.]: CVPR, 2018: 7132-7141.

[21] HUANG Z, WANG X, HUANG L, et al. Ccnet: Criss-cross attention for semantic segmentation [C] //Proceedings of the IEEE International Conference on Computer Vision. [S.l.]: ICCV,2019: 603-612.

[22] PASZKE A, GROSS S, MASSA F, et al. Pytorch: An imperative style, high-performance deep learning library [C] //Advances in Neural Information Processing Systems. [S.l.]: NeurIPS,2019: 8026-8037.

[23] MENTZER F, AGUSTSSON E, TSCHANNEN M, et al. Conditional probability models for deep image compression [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: CVPR, 2018: 4394-4402.

[24] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database [C] //2009 IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: CVPR,2009: 248-255.

[25] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding [C] //Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. [S. l.]: CVPR, 2016: 3213-3223.

[26] FELIX T. Kodak PhotoCD dataset [EB/OL]. http://r0k.us/graphics/kodak/. (2013-01-27) [2019-11-20] http://r0k.us/graphics/kodak/.

[27] LUNDH F, CLARK A. Python Imaging Library (PIL) [EB/OL]. (2018-01-01) [2020-03-21]. https://pillow.readthedocs.io/en/5.1.x/.

[28] UCL. JPEG2000 official software OpenJPEG [EB/OL]. (2019-10-10) [2020-03-23]. http://www.openjpeg.org/.

[29] KINGMA D P, BA L J. Adam: A Method for Stochastic Optimization [EB/OL]. (2017-01-30) [2020-06-16]. https://arxiv.org/abs/1412.6980.

[30] ASUNI N, GIACHETTI A. TESTIMAGES: a Large-scale Archive for Testing Visual Devices and Basic Image Processing Algorithms [C] //Smart Tools ans Apps for Graphics. [S.l.] STAG, 2014: 63-70.

作者简介:

朱 俊(1995-),男,云南昆明人,硕士研究生,主要研究方向为图像处理、图像压缩。E-mail: iszhujun@qq.com。

高陈强(1981-),男,重庆人,教授,博导。2009 年至今在重庆邮电大学任教。主要研究方向为图像处理、深度学习、行为识别、事件检测、小目标检测、红外图像分析。E-mail: gaocq@cqupt.edu.cn。

陈志乾(1996-),男,重庆人,硕士研究生,主要研究方向为图像压缩、深度学习。E-mail: chenzqian@yeah.net。

谌 放(1998-),男,湖北武汉人,在读本科生,主要研究方向为图像处理、篡改检测、深度学习。E-mail: cfun@outlook.com。

(编辑: 王敏琦)