计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.201100134

端到端优化的图像压缩技术进展

刘 东 王叶斐 林建平 马海川 杨闰宇
中国科学技术大学电子工程与信息科学系 合肥 230027

摘 要 图像压缩是数据压缩技术在数字图像上的应用,其目的是减少图像数据中的冗余,从而用更加高效的格式存储和传输数据。传统的图像压缩方法中,图像压缩分为预测、变换、量化、熵编码等步骤,每一步均采用人工设计的算法分别进行优化。近年来,基于深度神经网络的端到端图像压缩方法在图像压缩中取得了丰硕的成果,相比传统方法,端到端图像压缩可以进行联合优化,能够取得比传统方法更高的压缩效率。文中首先对端到端图像压缩的方法和网络结构进行了介绍;接着对端到端图像压缩中的关键技术进行了阐述,包括量化技术、概率建模和熵编码技术以及编码端码率分配技术;然后介绍了端到端图像压缩的扩展应用研究,包括可伸缩编码、可变码率压缩、面向视觉感知和机器感知的压缩;最后通过实验对端到端图像压缩方法目前可达到的压缩效率与传统方法进行了对比,展示了其压缩性能。实验结果表明,目前最新的端到端图像压缩方法的压缩效率远高于 JPEG,JPEG2000,HEVC intra 等传统图像编码方法,相比目前最先进的编码标准 VVC intra,在同样的 MS-SSIM 上节省了高达 48.40%的编码码率。

关键词:图像压缩;端到端优化;深度神经网络;压缩效率;JPEG;JPEG2000;HEVC;VVC
中图法分类号 TN919.81;TP37

Advances in End-to-End Optimized Image Compression Technologies

LIU Dong,WANG Ye-fei,LIN Jian-ping,MA Hai-chuan and YANG Run-yu
Department of Electronic Engineering and Information Science,University of Science and Technology of China,Hefei 230027,China

Abstract Image compression is the application of data compression technologies on digital images,aiming to reduce redundancy in image data,so as to store and transmit data with a more efficient format. In traditional image compression methods,image compression is divided into several steps,such as prediction,transform,quantization and entropy coding,and each step is optimized by manually designed algorithm separately. In recent years,end-to-end image compression methods based on deep neural networks have achieved fruitful results. Compared with the traditional methods,end-to-end image compression can be optimized jointly,which often achieves higher compression efficiency than the traditional methods. In this paper,the end-to-end image compression methods and network structures are introduced,and the key technologies of end-to-end image compression are described,including quantization technology,probability modeling and entropy coding technology,as well as encoder-side bit allocation technology. Then it introduces the research of extended applications of end-to-end image compression,including scalable coding,variable bit rate compression,visual perception and machine perception oriented compression. Finally,the compression efficiency of end-to-end image compression is compared with the traditional methods,and the compression performance is demonstrated. Experimental results show that the compression efficiency of the state-of-the-art end-to-end image compression method is much higher than that of the traditional image coding methods including JPEG,JPEG2000 and HEVC intra. Compared with the newest coding standard VVC intra,the end-to-end image compression method can save up to 48.40% of the coding rate while maintain the same MS-SSIM.

Keywords Image compression,End-to-end optimization,Deep neural network,Compression efficiency,JPEG,JPEG2000,HEVC,VVC

1 引言

图像压缩,也称图像编码,是将图像压缩成二进制比特流以进行后续的传输和存储的技术。一般未经压缩的原始图像数据量巨大,难以进行后续的使用。图像压缩技术可以对图像数据进行几十甚至几百倍的高效压缩,是实现图像的交互、处理和视觉任务的前端技术,同时也是支撑信息时代高速发展的基础性使能技术。

到稿日期:2020-11-19 **返修日期**:2020-12-02
基金项目:国家自然科学基金(61772483)
This work was supported by the National Natural Science Foundation of China (61772483).
通信作者:刘东(dongeliu@ustc.edu.cn)

由于对互操作性的要求，图像编码的算法研究和标准化工作早在20世纪80年代就已经开始。1992年，著名的图像编码标准JPEG诞生[1]，其通常可实现10倍以上的压缩。截止到目前，JPEG仍然是最为广泛使用的图像编码方法。2000年，JPEG2000[2]作为JPEG的后继格式被开发出来。除进一步提高压缩比之外，JPEG2000还实现了可伸缩编码和感兴趣区域编码，大大扩展了可支持的下游任务。目前性能最好的图像编码方法是BPG(Better Portable Graphics)[3]，它由视频编码标准H.265/HEVC[4]的帧内编码发展而来。另外，同等视觉质量下，正在开发的H.266/VVC[5]的帧内编码进一步实现了50%的码率节省，将发展出新的图像编码方法。

传统图像编码采用混合编码框架，一般包括预测、变换、量化、熵编码等模块。这些模块通常采用人工设计的方法得到，经过三十余年的发展，其性能很难得到进一步的提升。另一方面，尽管这些模块在功能上相互联系，但是传统方法往往单独优化设计每一个模块，无法实现整个编码框架的联合调优。更重要的是，随着计算机视觉技术的持续进步，对图像编码的需求越来越多样化，如视觉媒体时代对图像主观质量的要求、蓬勃发展的计算机视觉任务对图像语义质量的要求等。由此可见，传统图像编码方法已经难以适应时代发展的需求。

近年来，基于深度学习的端到端图像编码技术获得了学术界和工业界共同的关注。2015年谷歌公司的研究人员提出了首个端到端图像编码方法[6]，向人们展示了这一技术的可能性。此后，大量端到端图像编码方法被提出，仅5年时间，已经有不少工作超过了BPG的性能，甚至有工作超越了最新的VVC的帧内编码模式。

虽然端到端图像编码仍采用混合编码的架构，明显包括了变换、量化、熵编码等模块，但与后者又有本质的区别。首先，每个模块都可以采用学习的方法利用大量的图像数据进行优化，因此比手工设计模块更加高效。另外，由于所有的模块都采用神经网络来实现，整个框架都可以通过梯度反传的方式进行联合调优，因此更加便于协调各个模块之间的关系，进一步提高其性能。最重要的是，由于端到端编码通过给定损失函数进行训练的方法获得，因此可以通过更换损失函数来实现不同的编码目的，如针对主观质量和语义质量的编码。

为了更好地梳理端到端图像编码的发展历程，突出其技术难点和重点，探究未来可能的发展方向，本文对端到端图像编码进行了综述。本文第2节对图像编码进行了概述，特地对比了传统图像编码和端到端图像编码；第3节详细介绍了端到端图像编码，包括关键的技术和应用；第4节对比展示了端到端图像编码相比传统编码的压缩结果；最后总结全文。

2 图像压缩概述

图像压缩主要包含3个步骤，即变换、量化和熵编码，一些新的方法还引入了预测技术。变换是将图像从像素空间变换到低冗余特征空间，去除像素间的空间相关性，从而减小图像的空间冗余；量化是将变换系数进行离散化，从而实现系数多对一的映射，量化可以有效减小系数取值空间，取得更好的压缩效果，但由于量化不可逆，它也是图像失真的主要来源；熵编码是按照信息熵的原理，将量化后的系数或符号转化为可用来传输或存储的二进制码流，同时消除符号间的统计冗余。此外，一些新的压缩方法还在变换前加入了预测，预测是使用图像的空域相关性，用已编码的像素预测空域邻近的像素，这亦是一种消除空间冗余的方法。

目前，图像压缩按照框架划分主要包含两大类方法，即传统混合框架图像编码和端到端深度神经网络图像压缩，如图1所示。传统混合编码框架是使用人工设计的算法将预测、变换、量化、熵编码等模块组合起来，从而得到一套完整的图像编码器。传统方法技术成熟，复杂度低，但是非常依赖人工设计，且难以联合优化。相比传统方法，采用深度神经网络的端到端图像压缩方法，不仅可以联合优化网络参数，还避免了复杂的人工算法设计，网络可以更为智能地学习去除图像中的冗余信息，因此在压缩率和视觉质量上往往可以获得比传统方法更好的效果。

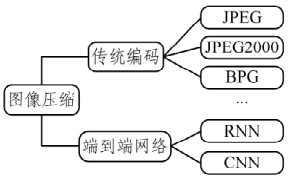


图1 图像编码方法分类

Fig. 1 Classification of image coding methods

2.1 混合框架图像编码

混合编码框架主要通过研究图像中的相关性，利用人工设计算法进行预测、变换、量化、熵编码等模块操作，去除图像中的冗余。例如，1992年通过的JPEG图像压缩标准[1]，将图像划分为8×8大小的图像块，采用的技术包括块级离散余弦变换(Discrete Cosine Transform，DCT)和哈夫曼编码[7]。JPEG至今仍是最广泛使用的图像压缩格式。

最新的一些混合编码框架图像压缩方法采用了视频编码标准中的技术。例如，BPG[3]采用了H.265/HEVC[4]视频编码标准的技术，引入了基于角度预测的帧内预测技术，预测残差采用整数离散余弦变换或整数离散正弦变换，量化系数进行二值化后，采用上下文自适应二元算术编码[8]输出码流。为了适应不同图像内容，BPG采用了四叉树图像块划分方式，图像块根据率失真代价的大小自适应选择最优的编码模式和参数。因此，BPG的压缩效率明显优于JPEG和JPEG2000。

2.2 端到端图像编码

传统图像编码经过了前期的快速发展，近年来在性能提升方面遇到了瓶颈。另外，传统图像编码一般针对PSNR等客观指标进行优化，难以适应对图像主观质量和语义质量的需求。在这种背景下，基于深度学习的端到端图像编码在近几年获得了研究者的大量关注，并取得了巨大的进步。从2015年第一个端到端图像编码方案被提出以来[6]，其性能从最初的低于JPEG发展到如今与最先进的H.266帧内编码相当[9]，端到端图像编码在过去的5年里追赶上了已经发展了30年的传统图像编码。

端到端图像编码整体上沿用了自编码器[10]的结构。原始的自编码器被提出用于数据降维，包括编码器和解码器两部分。编码器从高维的输入信号(如图像)中提取关键特征，

解码器则利用提取到的特征重建图像。为了适应压缩的需求，端到端图像编码在自编码器的基础上引入了量化[6]和熵编码[11]，用于高效编码提取到的特征。区别于传统图像编码手工设计各个模块，端到端图像编码的所有模块可以联合调优。根据香农的率失真理论[12]，编码的本质是码率和失真的权衡。在这种思想的指导下，目前端到端图像编码的优化普遍采用如下损失函数：

$$L=D+\lambda R$$

其中，D 表示失真，如均方误差(MSE)；R 表示码率，用于衡量编码特征所需的数据量；λ 用于实现 R 和 D 之间的权衡。为了实现端到端联合优化，端到端编码前期的很多工作都致力于解决两个问题，即量化不可导问题[6,13]和码率估计问题[11,13-16]。后文将对此进行详细的介绍。

基于所使用的网络结构，端到端图像编码可以分为两类：基于循环神经网络(Recurrent Neural Network，RNN)的端到端图像编码和基于卷积神经网络(Convolutional Neural Networks，CNN)的端到端图像编码。Toderici 等基于 RNN 提出了首个端到端图像编码方法[6]，该方法通过迭代调用基于 RNN 的编码器压缩图像/残差，实现了可伸缩编码。之后，其通过引入更加高效的熵编码方法、隐状态的初始化、SSIM 加权的失真度量和码率分配等方法[11,16]，基于 RNN 的端到端图像编码方法得到了进一步的发展，在 MS-SSIM 指标上超越了 BPG。Balle 等[13]最早提出了全卷积的端到端图像编码结构，该结构首次使用了后来被广泛使用的 GDN 非线性函数[17]，用于更好地移除图像内部冗余。之后，他们通过先后引入超先验模型[15]和自回归模型[14]来进行高效的熵编码。基于 CNN 的端到端编码方法在 PSNR 指标上超过了 BPG。相比 RNN，基于 CNN 的端到端图像编码因为结构简单、便于优化、性能更优等优点，被后续的大多数端到端图像编码方法所采用[9,14-15,18-20]。值得关注的是，区别于以上两种架构，文献[20]使用基于 CNN 的小波变换实现了端到端图像编码，统一了有损压缩与无损压缩，并支持可伸缩编码，为这一领域的发展提供了新的可能。

相比传统图像编码，端到端方案最大的特点在于其通过梯度反传方式进行联合调优。这允许端到端图像编码结合编码端优化方法以进一步提升其自身性能[21-23]。更重要的是，这使端到端图像编码更容易进行功能上的扩展，如针对视觉感知的编码[24-26]、针对机器任务的编码[27-28]、可伸缩编码[6,29-30]、可变码率编码[31-33]等。部分端到端图像压缩方法的开源信息如表 1 所列。

表 1　部分端到端图像压缩方法的开源信息

Table 1　Open-source information of several end-to-end image compression methods

论文	开源网址	开源类型
文献[13,15]	https://github.com/tensorflow/compression	训练和测试代码＋模型
文献[52]	https://github.com/huzi96/Coarse2Fine-ImaComp	测试代码＋模型
文献[45]	https://github.com/ZhengxueCheng/Learned-Image-Compression-with-GMM-and-Attention	测试代码＋模型
文献[20]	https://github.com/mahaichuan/Versatile-Image-Compression	测试代码＋模型
文献[37]	https://github.com/limuhit/CWIC	模型
文献[18]	https://github.com/JooyoungLeeETRI/CA_Entropy_Model	测试代码＋模型
文献[22]	https://github.com/zzs1994/CVQN	训练代码
文献[35]	https://github.com/fab-jul/imgcomp-cvpr	训练代码＋模型
文献[51]	https://github.com/limuhit/CCN	训练和测试代码＋模型
文献[19]	https://github.com/limuhit/Nonlocal-CCN	测试代码＋模型

3　核心模块

3.1　量化

与自编码器不同，为了实现编码压缩的目的，端到端图像编码必须引入量化模块。但是，一方面量化过程是不可导的，另一方面量化会引入信息损失，这都给端到端训练带来了困难。设计高效和可训练的量化方法是所有端到端图像编码方法必须解决的问题。

文献[6]首先提出采用二值量化，文献[11,16]沿用了这种做法。二值量化具有 3 个优点：1)可直接序列化，方便传输和存储；2)通过控制特征的维度可精准控制码率；3)可强制神经网络学习图像中的有效表征。为了实现端到端训练，在反传过程中梯度直接通过量化层而不进行任何修正。

文献[13-15]则使用多进制量化得到整系数。尽管整系数表征增大了熵编码的难度，但大大减小了量化的信息损失，使得端到端训练更加稳定高效。在训练阶段，文献[13]通过添加均匀噪声来模拟实际量化以保证梯度反传。类似地，文献[23]也使用多进制量化，但使用了类似于文献[6]的方法来解决梯度反传问题：在正向传播阶段运用四舍五入取整；在反向传播阶段梯度直接通过量化层而不进行修正。

文献[34]提出了 soft-to-hard 向量量化方法。在训练阶段，Christopoulos 等使用软量化(量化取值是所有量化中心的加权和)来保证梯度反传，并使用退火方法使软量化逐渐逼近硬量化(量化取值是量化中心之一)。文献[35]实现了 soft-to-hard 标量量化，与文献[34]的不同之处在于其不使用退火的方法，而是在反向传播阶段采用软量化，在正向传播阶段采用硬量化。soft-to-hard 算法使标量量化中心也是可学习的，从而可以进一步挖掘基于学习的图像编码方法的潜力。文献[36]使用 soft-to-hard 方法实现了二进制量化，并通过调节量化步长实现了不同的量化精度。

为了进一步提高量化的灵活性，文献[35,37]提出使用 IM(Importance Map)，根据不同区域的纹理特点调整不同区域的码率。从某种程度来说，IM 也可以被视为一种特殊的量化策略，即对特征层数的量化。为了优化 IM，文献[37]提出了两步量化法：首先，通过直接优化率失真损失函数，得到最优的 IM；其次，在率失真损失函数的指导下优化 IM 的生成

网络。与 IM 类似，文献[22]通过通道注意力机制和通道分层编码机制，为不同的通道使用不同的量化策略，充分利用了特征在不同通道之间重要性不同的特点。

文献[33,38]提出通过优化编码特征的方法来进一步提高压缩性能。端到端编码网络训练完毕后，针对特定图像，在率失真损失函数的指导下优化编码特征。在不改变解码复杂度的情况下，这种方法可以明显提升端到端编码方法的率失真性能。从本质上说，这与传统编码方法中的率失真优化量化(RDOQ)非常类似，但是端到端编码框架简化了 RDOQ 算法的实现过程。

除此之外，文献[39-42]提出在量化过程中结合使用缩放和偏置系数，以灵活调节量化深度，从而实现可变码率压缩。文献[43]使用 Lloyd 算法学习量化中心和量化区间，这种做法实际上是将训练编解码器的过程和训练量化器的过程分步进行。

3.2 熵编码

由于变换对图像内部冗余去除不彻底，待编码变换系数(下文简称系数)或特征本质上存在统计冗余，因此无论是传统编码还是端到端编码，都依赖高效的熵编码来进一步提升编码性能。特别是对于端到端编码，由于整个框架在率失真损失函数的指导下进行调优，码率估计的效率和精度将直接影响整个结构的优化效果。

文献[11,16]采用 RNN 实现端到端图像编码基本框架，生成的特征具有可伸缩的特点。与可伸缩特征相适应，该文献采用 RNN 结构提取特征内部以及不同迭代产生的特征之间的相关性。然而，RNN 存在难以训练的问题，因此使用得较少，大多数方法仍然通过 CNN 进行概率建模。

文献[13]采用分段线性函数拟合系数的概率分布，该方法在本质上假设系数之间独立同分布。在文献[13]提出的端到端框架的基础上，文献[15]指出系数之间存在较强的空域相关性，并通过提取额外的先验信息来利用这种相关性，从而大幅提升了编码性能。具体来说，文献[15]提出的系数由两部分组成，分别是超先验系数和主干系数。超先验系数从主干系数中分析得到，用于描述主干系数的分布特性，使用独立同分布的概率模型编码；主干系数则使用高斯概率模型，其参数从超先验系数中推断得到。文献[14,18]在文献[15]的基础上引入了空域自回归模型，以进一步利用系数的空域相关性。与提取超先验系数相比，自回归模型通过分析已编码的相邻系数来推断当前系数的概率分布，不需要传输额外的信息，因此熵编码的效率得到了进一步提高。为了更加精准地表示系数的分布，文献[9,21,44-46]使用多高斯模型替换单高斯模型作为概率函数。文献[9,19]提出利用系数的全局相关性。相似的图像块经过变换之后，系数之间具有相似性。文献[19]使用 Non-local 模块，使相似区域的相似系数可被用于上下文，以进一步提高上下文建模的精确性。文献[47]则模仿传统编码方法，提出使用二进制模型来进行更加精准的上下文概率建模。

空域自回归模型对特征通道之间的相关性利用不足。文献[35,37-48]提出使用三维上下文模型，以充分利用编码系数之间的相关性；文献[49-50]基于超先验参数结合使用三维上下文模型，通过分割卷积核和使用门卷积等操作来提高概率估计精度；文献[51]提出以 zigzag 扫描顺序进行三维上下文建模，并通过划分独立编码区域来实现高效的编解码。

由于串行编解码，自回归模型本质上具有复杂度高的缺点。文献[52]提出使用多层超先验模型来充分提取系数之间的相关性。实验结果表明，多层超先验模型相比单层超先验模型具有更好的编码性能，特别是在高分辨率图像上性能提升得更加明显。文献[53]提出了通道自回归模型。与提取空域相关性的空域自回归模型不同，通道自回归模型仅利用不同通道之间的相关性，将先编码的通道作为后编码通道的上下文使用，同一通道的不同区域可以并行处理。相比空域自回归模型，该方法同时提高了编码性能和编码速度。

3.3 编码端优化

在一些传统图像编码方法中(如 BPG)，编码器会对图像中不同区域的内容进行自适应的码率控制，如平滑图像区域可以分配较少的码率，而结构、纹理和细节非常丰富的区域则应分配较多的码率，这样可以实现在给定码率下使得图像失真尽可能小。端到端图像编码中同样引入了码率分配技术。端到端图像压缩码率分配的主要技术手段是采用重要性图(importance map)或注意力(attention)来实现，两者在本质上是一样的，都通过将图像表征乘以一个权值图来对系数进行缩放，由此等效获得不同的量化步长或损失函数权重，从而实现空域码率分配。

Li 等[23]最先提出使用重要性图来实现码率分配，如图 2 所示。首先采用一个重要性图生成网络对编码网络输出的图像特征进行进一步处理，得到一个特征重要性图，接着对重要性图做量化，最后通过量化权重来控制空域每个位置的特征码长。Liu 等[50,54]和 Chen 等[55]提出使用级联非局部注意力模块(Non-Local Attention Module)，采用级联的 1x1 卷积核与 softmax 激活层，实现对图像全局相关性和局部相关性的联合提取，从而有效提升了网络对图像空间冗余的捕捉和消除能力。Cheng 等[45]对文献[50,54]中的注意力模块进行简化，移除非局部注意力模块，只采用深度残差块就能够非常有效地捕捉到很大的感受野，从而缩短网络训练时间。Liu 等[21]进一步提出一种简单的通道注意力模块，采用通道平均池化和全连接网络提取图像特征每个通道的权重，只对通道维度进行加权，大大减小了计算量。Zhong 等[22]使用通道重要性模块计算每个特征通道的权重，对不同权重的通道采用不同的量化器和概率模型。Wu 等[56]使用重要性图指导基于生成对抗网络(Generative Adversarial Network，GAN)损失函数的图像压缩，以提升图像主观质量。

除此之外，还有一些显式获得重要性图的方法，包括基于感兴趣区域(Region of Interest，ROI)[34]的图像编码、基于语义分析以及基于图像分割的方法。Akutsu 等[57-58]使用人工设置的失真权重图来控制码率分配，对图像中感兴趣区域分配较大的失真权重，从而提升感兴趣区域的质量。Cai 等[59]使用 ROI 预测网络来计算图像多尺度特征的三维 ROI 掩膜，而不是直接在原图像上使用 ROI 掩膜，并且在训练的损失函数中增加一个 ROI 失真。Wang 等[60]采用预训练的语义分析网络生成语义重要性图，对图像中重要区域分配更多的码率。Xia 等[61]采用 DeepLab[62]图像分割网络对图像中的前景和背景进行分割，然后分别对前景和背景采用不同的熵编码引擎。

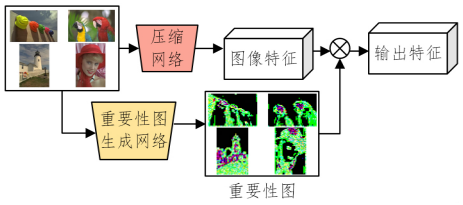


图 2　基于重要性图的码率分配

Fig. 2　Bit rate allocation based on importance map

3.4　特殊应用

一般图像编码以最小化率失真代价为目标，失真主要是以 MSE 或多尺度结构相似度（Multi-Scale Structural SIMilarity，MS-SSIM）为代表的客观质量评价指标。但是也有一些最新研究专注于特殊应用场景或需求，主要包括可伸缩编码、可变码率压缩、面向视觉感知的压缩以及面向机器感知的压缩等。

可伸缩编码主要面向传输适配，将图像特征分为基本层和增强层。当带宽较小时，只传输基本层的码流，此时解码得到一张质量较低的图像；而当带宽充足时，则传输增强层码流，从而提高图像的质量。Toderici 等[6,11]和 Johnston 等[16]基于 RNN 的端到端图像压缩方法，将每一轮编码残差作为下一轮编码的输入，每一轮网络都会输出一部份码流，循环次数越多，码率越高，图像质量就越好。Jia 等[63]也采用了同样的思想。Guo 等[29]采用多层卷积网络和下采样方法提取图像的多尺度分层特征，包括 1 层基本层和 3 层增强层，并对基本层和增强层进行去相关以减小码率，增强层用于高质量图像重建。Zhang 等[64]将图像进行比特平面分解，每个比特平面分别输入网络支路，在解码端将所有支路的重建相加，也实现了可伸缩编码。

可变码率压缩解决了端到端图像压缩中不同码率点需要不同模型的问题，即所有码率点均使用同一个网络。Toderici 等[6,11]的 RNN 端到端图像压缩本质也是一种可变码率压缩，但是只能在固定的若干码率点之间调整；Choi 等[31]使用基于拉格朗日因子 λ 条件的编解码网络和概率模型，将 λ 编码为一个 one-hot 向量，经过全连接网络分别生成增益和偏置，与图像特征一起进行线性运算；Yang 等[32]使用一种可调节编解码网络，将 λ 输入到一个调节网络中，调节网络则对编解码网络每一层输出的特征进行缩放来实现可变码率编码；Cui 等[39]和 Guo 等[33]提出 G-VAE（Gained Variational Autoencoder），用于训练若干增益向量，对每个特征通道进行加权，增益向量的选择则依据拉格朗日因子 λ。类似地，Chen 等[41]使用 λ 来训练对应的通道缩放因子函数以进行通道缩放；Zhou 等[42]采用了一种可调节量化步长的死区量化器，该量化器可以应用于不同码率。

面向视觉感知压缩的主要目标是提高图像的主观质量，主要技术手段是采用与主观质量一致的损失函数。Rippel[24]最先将 GAN 损失引入端到端图像压缩，可以在极低码率下获得主观视觉质量很好的重建图像。随后，Agustsson 等[65]、Akutsu 等[58]以及 Dash 等[66]针对极低码率下的主观质量优化均采用了 GAN 失真函数。与此同时，也有采用其他失真函数来模拟主观视觉失真的研究，例如，Chen 等[25]采用一个网络模拟 VMAF（Video Multimethod Assessment Fusion）分数，Lee 等[26]和 Patel 等[67]采用了 VGG 失真。为了解决主观视觉失真下重建图像和原始图像存在的内容不一致的问题，Kudo 等[68]采用互信息正则化，即引入一个互信息损失，并在训练过程中最大化重建图和原图之间的互信息；Mentzer[69]则基于码率-失真-感知权衡理论[70]，使用率失真感知损失函数对网络参数进行优化，实验中可以获得视觉效果优良并与原图高度一致的高分辨率的重建图像；Luo[71]提出 noise-to-compression variational antoencoder（NC-VAE），在训练中对样本进行噪声扰动，从而有效提高网络鲁棒性和重建图与原图的一致性；Lee 等[26]则采用迁移学习的方法来提升网络重建图像的质量。

面向机器感知的压缩是将图像压缩和计算机视觉任务相结合，即不再是最小化图像失真，而是优化视觉任务的准确度。Hu 等[27]使用边缘提取的方法来指导网络进行图像重建，并采用 GAN 失真可以获得视觉质量非常好的重建图，在人脸特征点检测中可以获得远高于 JPEG 的准确度。Duan 等[28]提出使用高层语义图来增强低级视觉特征，并验证了此方法可以有效提升图像压缩的码率-感知-准确度-失真表现。在实验中，该方法不仅可以减小图像失真，还有效提高了图像的主观质量以及在物体检测任务中的准确度。

4　实验

本文主要对端到端图像编码的各代表性方法[9,11,13-16,18,45,52]和传统的图像编码标准 JPEG[1]，JPEG2000[2]，HEVC-intra[4]以及 VVC-intra[5]在公开的图像数据集 Kodak[72]和 Tecnick[73]上的编码性能进行比较。图 3 和图 4 直观地展示了它们的率失真曲线。图像的失真采用峰值信噪比（Peak Signal-to-Noise Ratio，PSNR）和 MS-SSIM[74]来衡量。为了清晰地显示二者的差别，MS-SSIM 被转换成分贝（$-10\log_{10}(1-MS-SSIM)$）的形式。图像的码率采用平均每个像素的比特数（Bpp）来计算。图 3 给出了 Kodak 数据集上的率失真曲线。

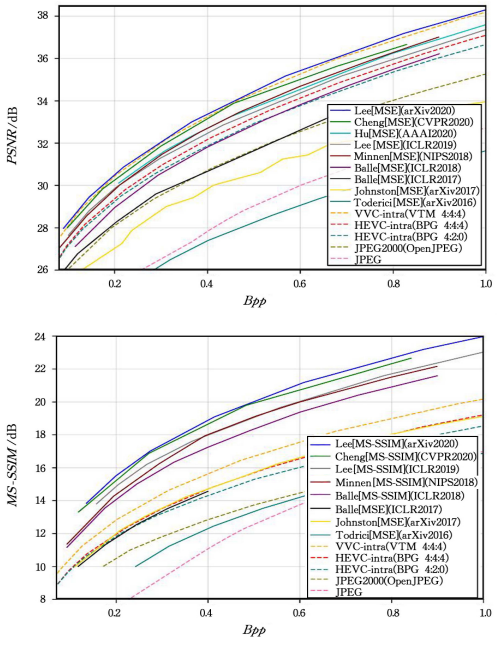


图 3　Kodak 数据集上的率失真表现

Fig. 3　Rate-distortion performance on Kodak dataset

从图中可以看出，在2016年，端到端图像编码方法Toderici(arXiv2016)的编码性能与JPEG相当，之后每年端到端图像编码方法的编码性能都以惊人的速度持续提升。截止到2020年，最新的方法Lee(arXiv2020)[9]在PSNR和MS-SSIM上的整体性能已经超过所有传统编解码标准。文献[9]称，他们的方法相比VVC-intra(VTM7.1)在同样的PSNR和MS-SSIM上分别节省了1.65%和48.40%的编码码率。如图4所示，在更高分辨率的数据集Tecnick上进行测试，结果显示最新的方法在PSNR和MS-SSIM上的编码性能也超过了VVC-intra。

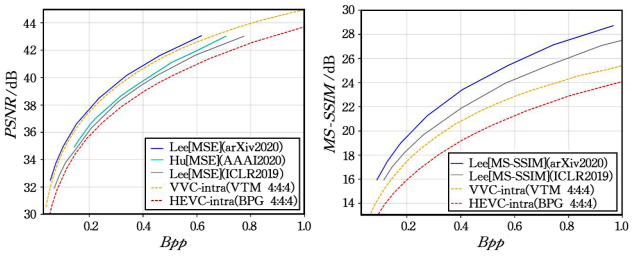


图4 Tecnick数据集上的率失真表现

Fig. 4 Rate-distortion performance on Tecnick dataset

图5给出了面向MSE优化的方法Lee(arXiv2020)[9]和Lee(ICLR2019)[18]，以及VTM7.1[5]和BPG[3]解码Kodak数据集中的图像kodim07的可视化结果。从图中可以看出，Lee(arXiv2020)的解码图像相比其他方法具有更高的视觉质量。

Lee[MSE](arXiv2020) Bpp:0.1243, PSNR:31.40 dB	VVC-intra(VTM7.1 4:4:4) Bpp:0.1248, PSNR:30.96 dB	Lee[MSE](ICLR2019) Bpp:0.1043, PSNR:29.46 dB	HEVC-intra(BPG 4:4:4) Bpp:0.1188, PSNR:29.41 dB

图5 Kodak数据集中的图像kodim07的可视化结果

Fig. 5 Visualization results of kodim07 in Kodak dataset

图6给出了面向MS-SSIM优化的方法Lee(arXiv2020)和Lee(ICLR2019)，以及VTM7.1和BPG解码Kodak数据集中的图像kodim13的可视化结果。

Lee[MS-SSIM](arXiv2020) Bpp:0.2442, MS-SSIM:0.9319	VVC-intra(VTM7.1 4:4:4) Bpp:0.2409, MS-SSIM:0.8719	Lee[MS-SSIM](ICLR2019) Bpp:0.2630, MS-SSIM:0.9313	HEVC-intra(BPG 4:4:4) Bpp:0.2760, MS-SSIM:0.8699

图6 Kodak数据集中的图像kodim13的可视化结果

Fig. 6 Visualization results of kodim13 in Kodak dataset

可以看出，面向MS-SSIM优化的端到端图像压缩方法Lee(arXiv2020)和Lee(ICLR2019)相比传统方法VVC-intra和HEVC-intra明显包含更多的纹理细节，因此具有更高的视觉质量。

结束语 本文首先对图像压缩的研究背景和意义进行了概述，然后对基于深度学习的端到端的图像编码方法进行了分析和总结。此外，本文还分别介绍了几种端到端图像压缩的核心部分，包括量化、熵编码、智能码率分配的编码器优化以及端到端图像编码对特殊应用的适配方法。最后对端到端图像压缩与传统图像压缩的编码性能进行比较。目前，端到端图像编码存在的问题、难点以及解决的思路如下。

(1)复杂度过高。在实际应用中，会对图像压缩的复杂度提出一定的要求。目前性能优异的端到端图像压缩方法通常包含几十层网络，尤其是基于PixelCNN的熵编解码部分，更是极大地提高了图像编解码的复杂度。因此，未来需要研究复杂度更低的网络结构以适应实际的应用。

(2)提升主观质量。对于图像压缩而言，提升图像主观质量的重要意义不言而喻。目前，传统的图像压缩方法通常是基于MSE进行优化的，而MSE并不能很好地度量图像的主观质量。对于端到端图像压缩，很容易通过设计损失函数等方法来提升图像的主观质量。不过人眼视觉系统是一个非常复杂的系统，目前对其的研究相当有限，因此更好地提升图像主观质量的方法值得进一步的分析与研究。这对于进一步降低码率也有着重要的意义。

(3)更精准的码率控制。对于传统图像编码而言，可以通过改变量化步长来实现精准的码率控制；而端到端的图像编码通常只能通过训练多个模型来实现多个码率点的编码。这样既浪费了计算资源，对存储也产生了很大的压力。因此，高效的可变码率端到端图像压缩值得进一步的探索与研究。

参考文献

[1] WALLACE G K. The JPEG still picture compression standard [J]. IEEE Transactions on Consumer Electronics, 1992, 38(1): 18-34.

[2] RABBANI M. JPEG2000: Image compression fundamentals, standards and practice[J]. Journal of Electronic Imaging, 2002, 11(2): 286.

[3] BPG Image format [CP/OL]. https://bellard.org/bpg, 2015.

[4] SULLIVAN G J, OHM J R, HAN W J, et al. Overview of the high efficiency video coding (HEVC) standard[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2012, 22(12): 1649-1668.

[5] Versatile video coding reference software version 7.1 (VTM-7.1) [CP/OL]. https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM/tags/VTM-7.1.

[6] TODERICI G, O'MALLEY S M, HWANG S J, et al. Variable rate image compression with recurrent neural networks [J]. arXiv: 1511.06085, 2015.

[7] HUFFMAN D A. A method for the construction of minimum-redundancy codes[J]. Proceedings of the IRE, 1952, 40(9): 1098-1101.

[8] MARPE D, SCHWARZ H, WIEGAND T. Context-based adaptive binary arithmetic coding in the H.264/AVC video compres-

sion standard[J]. IEEE Transactions on Circuits and Systems for Video Technology，2003，13(7)：620-636.

[9] LEE J，CHO S，KIM M. An end-to-end joint learning scheme of image compression and quality enhancement with improved entropy minimization[J]. arXiv：1912. 12817，2019.

[10] HINTON G E，SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks[J]. Science，2006，313(5786)：504-507.

[11] 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. 2017：5306-5314.

[12] COVER T M，THOMAS J A. Elements of information theory[M]. New York，US：John Wiley & Sons，1991.

[13] BALLÉ J，LAPARRA V，SIMONCELLI E P. End-to-end optimized image compression[J]. arXiv：1611. 01704，2016.

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

[15] BALLÉ J，MINNEN D，SINGH S，et al. Variational image compression with a scale hyperprior[J]. arXiv：1802. 01436，2018.

[16] 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. 2018：4385-4393.

[17] BALLÉ J，LAPARRA V，SIMONCELLI E P. End-to-end optimization of nonlinear transform codes for perceptual quality[C]// 2016 Picture Coding Symposium (PCS). IEEE，2016：1-5.

[18] LEE J，CHO S，BEACK S K. Context-adaptive entropy model for end-to-end optimized image compression[J]. arXiv：1809. 10452，2018.

[19] LI M，ZHANG K，ZUO W，et al. Learning context-based non-local entropy modeling for image compression[J]. arXiv：2005. 04661，2020.

[20] MA H C，LIU D，YAN N，et al. End-to-end optimized versatile image compression with wavelet-like transform[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2020 (99).

[21] LIU J，LU G，HU Z，et al. A unified end-to-end framework for efficient deep image compression[J]. arXiv：2002. 03370，2020.

[22] ZHONG Z，AKUTSU H，AIZAWA K. Channel-level variable quantization network for deep image compression[J]. arXiv：2007. 12619，2020.

[23] 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. 2018：3214-3223.

[24] RIPPEL O，BOURDEV L. Real-time adaptive image compression[J]. arXiv：1705. 05823，2017.

[25] CHEN L H，BAMPIS C G，LI Z，et al. Perceptually optimizing deep image compression[J]. arXiv：2007. 02711，2020.

[26] LEE J，KIM D，KIM Y，et al. A training method for image compression networks to improve perceptual quality of reconstructions[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020：144-145.

[27] HU Y，YANG S，YANG W，et al. Towards coding for human and machine vision：A scalable image coding approach[C]// 2020 IEEE International Conference on Multimedia and Expo (ICME). IEEE，2020：1-6.

[28] DUAN S，CHEN H，GU J. JPAD-SE：High-level semantics for joint perception-accuracy-distortion enhancement in image compression[J]. arXiv：2005. 12810，2020.

[29] GUO Z，ZHANG Z，CHEN Z. Deep scalable image compression via hierarchical feature decorrelation[C]// 2019 Picture Coding Symposium (PCS). IEEE，2019：1-5.

[30] AKBARI M，LIANG J，HAN J，et al. Learned variable-rate image compression with residual divisive normalization[C]// 2020 IEEE International Conference on Multimedia and Expo (ICME). IEEE，2020：1-6.

[31] CHOI Y，EL-KHAMY M，LEE J. Variable rate deep image compression with a conditional autoencoder[C]// Proceedings of the IEEE International Conference on Computer Vision. 2019：3146-3154.

[32] YANG F，HERRANZ L，VAN DE WEIJER J，et al. Variable rate deep image compression with modulated autoencoder[J]. IEEE Signal Processing Letters，2020，27：331-335.

[33] GUO T，WANG J，CUI Z，et al. Variable rate image compression with content adaptive optimization[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020：122-123.

[34] CHRISTOPOULOS C，ASKELOF J，LARSSON M. Efficient methods for encoding regions of interest in the upcoming JPEG2000 still image coding standard[J]. IEEE Signal Processing Letters，2000，7(9)：247-249.

[35] 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. 2018：4394-4402.

[36] ALEXANDRE D，CHANG C P，PENG W H，et al. Learned image compression with soft bit-based rate-distortion optimization[C]// 2019 IEEE International Conference on Image Processing (ICIP). IEEE，2019：1715-1719.

[37] LI M，ZUO W M，GU S H，et al. Learning content-weighted deep image compression[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2020，PP(99).

[38] CAMPOS J，MEIERHANS S，DJELOUAH A，et al. Content adaptive optimization for neural image compression[J]. arXiv：1906. 01223，2019.

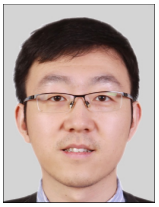
[39] CUI Z，WANG J，BAI B，et al. G-VAE：A continuously variable rate deep image compression framework[J]. arXiv：2003. 02012，2020.

[40] AKBARI M，LIANG J，HAN J，et al. Learned variable-rate image compression with residual divisive normalization[C]// 2020 IEEE International Conference on Multimedia and Expo (ICME). IEEE，2020：1-6.

[41] CHEN T，MA Z. Variable bitrate image compression with quality scaling factors[C]// 2020 IEEE International Conference on Acoustics，Speech and Signal Processing (ICASSP). IEEE，2020：2163-2167.

[42] ZHOU J，NAKAGAWA A，KATO K，et al. Variable rate image compression method with dead-zone quantizer[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern

Recognition Workshops. 2020:162-163.
[43] CAI J, ZHANG L. Deep image compression with iterative non-uniform quantization[C]// 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018:451-455.
[44] CHENG Z, SUN H, KATTO J. Low bitrate image compression with discretized Gaussian mixture likelihoods[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020:126-127.
[45] CHENG Z, SUN H, TAKEUCHI M, et al. Learned image compression with discretized Gaussian mixture likelihoods and attention modules[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020:7939-7948.
[46] WEN S, ZHOU J, NAKAGAWA A, et al. Variational autoencoder based image compression with pyramidal features and context entropy model[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019:1-4.
[47] LADUNE T, PHILIPPE P, HAMIDOUCHE W, et al. Binary probability model for learning based image compression[C]// 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020:2168-2172.
[48] GUO Z, WU Y, FENG R, et al. 3-D context entropy model for improved practical image compression[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020:116-117.
[49] LIU H, CHEN T, GUO P, et al. Gated context model with embedded priors for deep image compression[J]. arXiv: 1902.10480, 2019.
[50] LIU H, CHEN T, GUO P, et al. Non-local attention optimized deep image compression[J]. arXiv:1904.09757, 2019.
[51] LI M, MA K, YOU J, et al. Efficient and effective context-based convolutional entropy modeling for image compression[J]. IEEE Transactions on Image Processing, 2020, 29:5900-5911.
[52] HU Y, YANG W, LIU J. Coarse-to-fine hyper-prior modeling for learned image compression[C]// AAAI. 2020:11013-11020.
[53] MINNEN D, SINGH S. Channel-wise autoregressive entropy models for learned image compression[J]. arXiv: 2007.08739, 2020.
[54] LIU H, CHEN T, SHEN Q, et al. Practical stacked non-local attention modules for image compression[C]// CVPR Workshops. 2019:1-4.
[55] CHEN T, LIU H, MA Z, et al. Neural image compression via non-local attention optimization and improved context modeling[J]. arXiv:1910.06244, 2019.
[56] WU L, HUANG K, SHEN H. A GAN-based tunable image compression system[C]// The IEEE Winter Conference on Applications of Computer Vision. 2020:2334-2342.
[57] AKUTSU H, NARUKO T. End-to-end learned ROI image compression[C]// CVPR Workshops. 2019:1-5.
[58] AKUTSU H, SUZUKI A, ZHONG Z, et al. Ultra low bitrate learned image compression by selective detail decoding[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020:118-119.
[59] CAI C, CHEN L, ZHANG X, et al. End-to-end optimized ROI image compression[J]. IEEE Transactions on Image Processing, 2019, 29:3442-3457.
[60] WANG C, HAN Y, WANG W. An end-to-end deep learning image compression framework based on semantic analysis[J]. Applied Sciences, 2019, 9(17):3580.
[61] XIA Q, LIU H, MA Z. Object-based image coding: A learning-driven revisit[C]// 2020 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2020:1-6.
[62] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4):834-848.
[63] JIA C, LIU Z, WANG Y, et al. Layered image compression using scalable auto-encoder[C]// 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). IEEE, 2019:431-436.
[64] ZHANG Z, CHEN Z, LIN J, et al. Learned scalable image compression with bidirectional context disentanglement network[C]// 2019 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2019:1438-1443.
[65] 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. 2019:221-231.
[66] DASH S, KUMARAVELU G, NAGANOOR V, et al. CompressNet: Generative compression at extremely low bitrates[C]// 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2020:2314-2322.
[67] PATEL Y, APPALARAJU S, MANMATHA R. Deep perceptual compression[J]. arXiv:1907.08310, 2019.
[68] KUDO S, ORIHASHI S, TANIDA R, et al. GAN-based image compression using mutual information maximizing regularization[C]// 2019 Picture Coding Symposium (PCS). IEEE, 2019:1-5.
[69] MENTZER F, TODERICI G, TSCHANNEN M, et al. High-fidelity generative image compression[J]. arXiv: 2006.09965, 2020.
[70] BLAU Y, MICHAELI T. Rethinking lossy compression: The rate-distortion-perception tradeoff[J]. arXiv:1901.07821, 2019.
[71] LUO J, LI S, DAI W, et al. Noise-to-compression variational autoencoder for efficient end-to-end optimized image coding[C]// 2020 Data Compression Conference (DCC). IEEE, 2020:33-42.
[72] KODAK E. Kodak lossless true color image suite (1993) [DB/OL]. http://r0k.us/graphics/kodak/.
[73] ASUNI N, GIACHETTI A. TESTIMAGES: A large-scale archive for testing visual devices and basic image processing algorithms[C]// Smart Tools and Apps for Graphics - Eurographics Italian Chapter Conference. 2014:63-70.
[74] WANG Z, SIMONCELLI E P, BOVIK A C. Multiscale structural similarity for image quality assessment[C]// The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. IEEE, 2003, 2:1398-1402.



LIU Dong, born in 1983, Ph.D, professor, is a senior member of China Computer Federation. His main research interests include multimedia signal processing and so on.