

基于深度学习的图像压缩算法研究综述

于 恒,梅红岩,许晓明,贾慧萍

辽宁工业大学 电子与信息工程学院,辽宁 锦州 121001

摘 要:随着深度学习的不断发展与图像数据的爆炸式增长,如何使用深度学习来获得更高压缩比和更高质量的图像逐渐成为热点研究问题之一。通过对近几年相关文献的分析与整理,将基于深度学习的图像压缩方法按照卷积神经网络、循环神经网络、生成对抗网络进行总结与分析,对不同种方法分别列举了具有代表性的实例,并对基于深度学习的图像压缩算法的常用训练数据集、评价指标进行了介绍,根据深度学习在图像压缩领域中的优势对其未来的发展趋势进行了总结与讨论。

关键词:深度学习;图像压缩;卷积神经网络;循环神经网络;生成对抗网络

文献标志码:A **中图分类号:**TP391 **doi:**10.3778/j.issn.1002-8331.2003-0294

于恒,梅红岩,许晓明,等.基于深度学习的图像压缩算法研究综述.计算机工程与应用,2020,56(15):15-23.

YU Heng, MEI Hongyan, XU Xiaoming, et al. Survey of image compression algorithm based on deep learning. Computer Engineering and Applications, 2020, 56(15):15-23.

Survey of Image Compression Algorithm Based on Deep Learning

YU Heng, MEI Hongyan, XU Xiaoming, JIA Huiping

School of Electronics and Information Engineering, Liaoning University of Technology, Jinzhou, Liaoning 121001, China

Abstract: With the continuous development of deep learning and the explosive growth of image data, how to use deep learning to obtain higher compression ratio and higher quality images has gradually become one of the hot research issues. Through the analysis of the related literatures in recent years, the image compression method based on the deep learning is summarized and analyzed according to the Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Generative Adversarial Network (GAN). This paper enumerates the typical examples, and the image compression algorithm based on depth study of the training data set, commonly used evaluation indexes are introduced, according to the deep learning advantages in the field of image compression for its future development trend are summarized and discussed.

Key words: deep learning; image compression; Convolutional Neural Network (CNN); Recurrent Neural Network (RNN); Generative Adversarial Network (GAN)

1 引言

图像压缩一直是图形图像处理领域的基础课题,自1948年提出电视信号数字化设想后,图像压缩便登上了历史的舞台,其大概发展路程如图1所示,从图中可以看出,在有了BP算法^[1]之后,就已经有研究人员将神经网络引入图像压缩领域的先例,之后随着深度学习地不断深入研究,基于深度学习的图像压缩方法也随之被提

出。深度学习^[2]对图像特征提取、表达能力,以及高维数据的处理能力等都被认为对于图像压缩存在独有的优势,时至今日研究这一方向的人数日益增多,将深度学习应用于图像压缩逐渐成为当前的热点研究问题之一。传统的图像编码标准如:JPEG^[3]、JPEG2000^[4]和BPG已被广泛使用,传统的图像压缩多采用固定的变换方式和量化编码框架,如离散余弦变换和离散小波变换,在结

基金项目:辽宁省自然科学基金(No.20180550886, No.201602372)。

作者简介:于恒(1992—),男,硕士研究生,研究领域为深度学习,计算机图形图像处理,数据挖掘;梅红岩(1978—),通信作者,女,博士,副教授,CCF会员,研究领域为深度学习,大数据分析 with 数据挖掘,网络服务, E-mail: liao-ning_mhy@126.com; 许晓明(1996—),女,硕士研究生,研究领域为深度学习,大数据分析 with 数据挖掘,推荐系统;贾慧萍(1995—),女,硕士研究生,研究领域为深度学习,大数据分析 with 数据挖掘,推荐系统。

收稿日期:2020-03-19 **修回日期:**2020-05-18 **文章编号:**1002-8331(2020)15-0015-09

CNKI网络出版:2020-06-18, <https://kns.cnki.net/kcms/detail/11.2127.TP.20200618.0854.002.html>

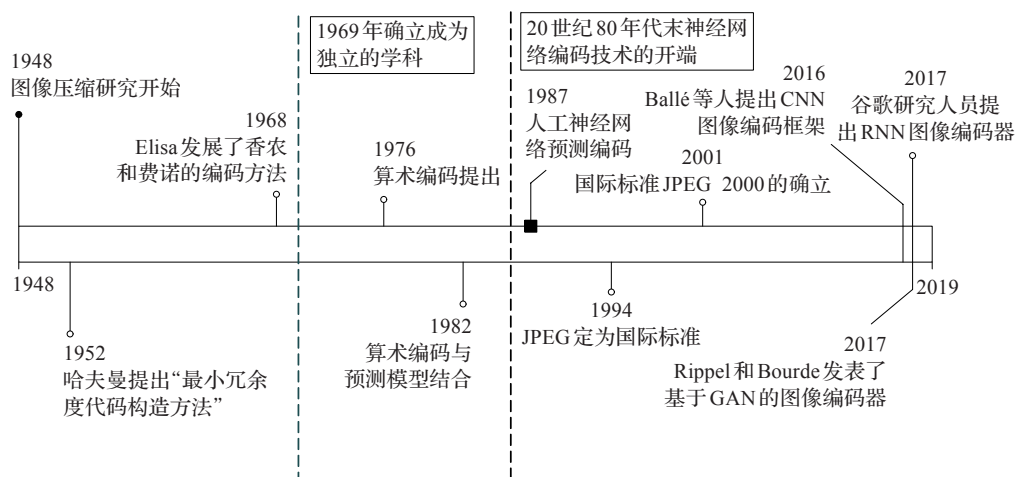


图1 图像压缩发展简图

合量化和编码器来减少图像的空间冗余,但是并非所有类型的图像都适用于这种方式,如以图像块的方式进行变换量化后会有块效应。同时在大量传输图像时由于网络带宽的限制,为了实现低比特率编码,会导致图像的模糊^[5]现象。深度学习技术可以根据自身特点优化上述问题:如在编码器的性能上,深度学习技术可以对编码器和解码器进行联合优化,不断提升编码器的性能;在图像清晰度上,基于深度学习的图像超分辨率^[6]技术,以及生成对抗网络都能使图像重建更加清晰;在面对不同类型的图像,针对不同类型的任务上,深度学习技术能够根据任务的特点对图像实现更智能、更针对性的编解码。本文将根据不同的深度学习方法对图像压缩处理取得的成果进行介绍。

2 传统图像编码背景

图像压缩的目的是通过消除数字图像像素间的冗余实现图像压缩处理。在静态图像中,空间冗余是存在的最多的冗余,物体与背景具有很强的联系,这种联系映射到像素一级上,就体现了很强的相关性,这种相关性在数字图像中就被称为数据冗余,通过压缩的方式来消除数据冗余的原理主要分为三类:预测编码、变换编码和统计编码。

预测编码的基础理论为现代统计学和控制论,其技术是建立在信号数据的相关性上。最经典的方式为DPCM法,利用当前图像的一个像素的真实值,根据相邻像素的相关性对当前像素进行预测,利用两者具有预测性的残差进行量化、编码,通过降低码流进而达到图像压缩的目的。变换编码技术的图像压缩算法主要是对图像进行函数变换,将空域信息变换到频域,之后在对频域信息进行处理,将高频信号和低频信号进行分离,按照信号的重要程度对比特位进行分配,减少信息冗余,达到压缩目的。统计编码也被称为熵编码是根据信息出现概率的分布特性进行编码。基于深度学习的图像压缩,并不是独立于传统的图像压缩方法的,更多

的方法是建立在传统的编码方式之上,对传统图像压缩的比特率和重建图像的分辨率进行提升。

在使用变换编码时,主要问题在于重建图像时存在块效应与伪影,这些问题其实并非只有深度学习能够处理,很多方法^[7-8]都可以对这些块效应进行很好的处理,但是深度学习更有能力处理这类问题。在早期,一个普通的多层感知器已经被用来直接学习一个投影从一个有噪声的图像到一个无噪声的图像,在近期,利用卷积神经网络、生成对抗网络,对图像进行超分辨率更是取得了阶段性的成果。熵编码是图像压缩框架中的一个重要组成部分。根据信息论^[9]编码信号所需的比特率受信息熵的限制,信息熵对应于表示信号的符号的概率分布。因此,在端到端学习图像压缩框架中嵌入熵编码组件来估计潜在表示的概率分布,并对熵进行约束来降低比特率。熵编码模型提供了对所有元素的可能性的估计,在熵模型中,大部分的工作都是利用算术编码对元素的符号进行无损编码。传统的图像压缩采用变换方式在配合熵编码进行图像压缩,而深度学习则是采用端到端的结构设计和不同种类的网络经过训练替代传统的固定变换方式,进而提升图像压缩。同时近些年GPU的高速发展,为更多样性网络结构的设计提供了计算保障,也为性能的提升提供了硬件支持,使基于深度学习的图像压缩在其分辨率、比特率等各方面有了提高。

3 基于深度学习的图像压缩方法

图像压缩根据对编码信息的恢复程度来进行分类,主要分为无损压缩^[10]和有损压缩^[11],基于深度学习的图像压缩方法多为有损图像压缩,依赖深度学习强大的建模能力,基于深度学习的图像压缩性能已经超过了JPEG和BPG,并且这种性能上的差距仍在逐步扩大。下面将分别对基于卷积神经网络(Convolutional Neural Network, CNN)^[12]、循环神经网络(Recurrent Neural Network, RNN)^[13]、生成对抗网络(Generative Adversarial Network, GAN)^[14]进行介绍。

3.1 基于卷积神经网络的图像压缩方法

CNN 在图像领域发展迅速,特别是在计算机视觉领域中表现出优异的性能。如目标检测、图像分类、语义分割等。CNN 卷积运算中的稀疏连接和参数共享两大特性使 CNN 在图像压缩中彰显优势。稀疏链接可以通过卷积核的大小来限制输出参数的多少,在图像中都有空间组织结构,图像中的一个像素点在空间上与周围的像素点都有紧密的关系,稀疏连接借鉴这一关系只接受相互有关联的区域作为像素点的输入,之后将所有神经元接收到的局部信息在更深层的网络进行综合,就可以得到全局的信息,从而降低了参数,也降低计算的复杂程度。权值共享是指每个神经元的参数都是相同的,在同一个卷积核的图像处理中参数都是共享的,卷积神经网络采用这种方式也会显著地降低参数的数量,并在一定程度上避免了过拟合的发生。卷积神经网络的这两大特性,更好地降低了计算的复杂程度,使训练可以向更深、更优的网络结构发展。同时这两大特性也减少图像压缩的数据量。CNN 的图像压缩多以端到端的方式进行图像压缩,通过 CNN 设计编码端与解码端,通过大量图像数据以及优化网络方式,获得高性能的压缩框架。

经典的图像压缩如 JPEG、JPEG2000 通常是将变化、量化、熵编码三个部分分别手动优化,图像码率经过量化计算后为离散系数,而基于 CNN 端到端优化采用梯度下降时要求函数全局可微,为此 Ballé 等人^[15]提出基于广义分歧归一化的卷积神经网络图像编码框架,使线性卷积和非线性更灵活的转换,这种方法将卷积层分为两个部分,一部分负责分析图像的紧凑表示,另一部分负责重建和逆过程,使用广义分歧归一化函数作为激活函数,这个方法取得了可以媲美 JPEG2000 的编码性能。之后 Ballé 等人^[16]又提出一种由非线性变换与统一量化的图像压缩方法,通过 CNN 实现非线性变换,并通过之前的广义分歧归一化实现了局部增益。这也是首次将 CNN 与图像压缩相结合,给之后基于 CNN 的端到端图像压缩的可行性奠定了基础。之前图像重建工作为了提高重建图像质量,研究的关注点多在一些图像先验模型,这些模型即使提高了重建图像的质量但多存在时效性低的问题,限制了其实际应用价值,并且忽略了图像压缩时的退化信息,为了提高重建图像的质量,Jiang

等人^[17]提出基于 CNN 的图像端到端压缩框架,其结构如图 2 所示,该方法从图像的编码器端和解码器端同时使用两个卷积神经网络将编码器与解码器进行联合,采用统一优化方法训练了两个 CNN,使其相互配合在编码器端使用一个 CNN 用于对图像进行紧凑表示后,在通过编码器进行编码,在解码器端使用一个 CNN 对解码后的图像进行高质量的复原,两个网络同时作用,通过卷积采样代替传统图像压缩以图像块为单位的变换计算,其块效应与 JPEG 相比有明显提升。

虽然 Jiang 等人端到端图像压缩同时优化两个卷积神经网络,但其在编解码前后直接连接两个卷积神经网络的近似方法并不是最优的。Zhao 等人^[18]认为这个问题最优解的直观想法是使用 CNN 来完美地替代传统的梯度反向传播编解码器端到端的方式提高重构图像的编码质量,因此他们提出使用一个虚拟编码器,在训练时使用虚拟编码器用于连接编码端和解码端,虚拟编码器也为 CNN,并通过虚拟编码器使解码端的 CNN 逼近最优解,这种方式将真实图像的有效表示信息经过虚拟编码器投影到用于重构图像的解码网络。该方法不仅得到了高质量的重建图像,也可以和端到端的网络结构一样可以兼容传统编码器,也可以推广到其他基于 CNN 的端到端图像压缩结构中,但是整个框架存在三个 CNN,经过一次训练难度相对较大,因此在训练上需要对三个网络进行分解训练,但实际应用只需要两个网络。

尽管 CNN 对于图像压缩具有优势,但是采用基于 CNN 的图像压缩仍然具有一定的困难:首先是优化问题,CNN 通常采用端到端的模式,在传统编码器的两端加入 CNN,这两个 CNN 都是需要通过训练来达到图像压缩和图像重建的目的,但是深度学习的优化问题本身就是一个难点问题,同时让两端进行联合优化,从而得到性能良好的框架也并非易事;二是传统的图像压缩方法往往能够定量地对图像压缩,如 JPEG 可以对图像进行 50:1 的压缩,但是基于 CNN 的图像压缩很少能够对图像进行固定比率的图像压缩;在压缩图像分辨率问题上,由于 CNN 方法大多采用对图像进行下采样,卷积核的感受野是有限的,如在对 1 024×1 024 的图像进行压缩时,采用的 128×128 的训练框架,往往得不到很好的效果,因而要实现全分辨率就要深化网络模型,提高框

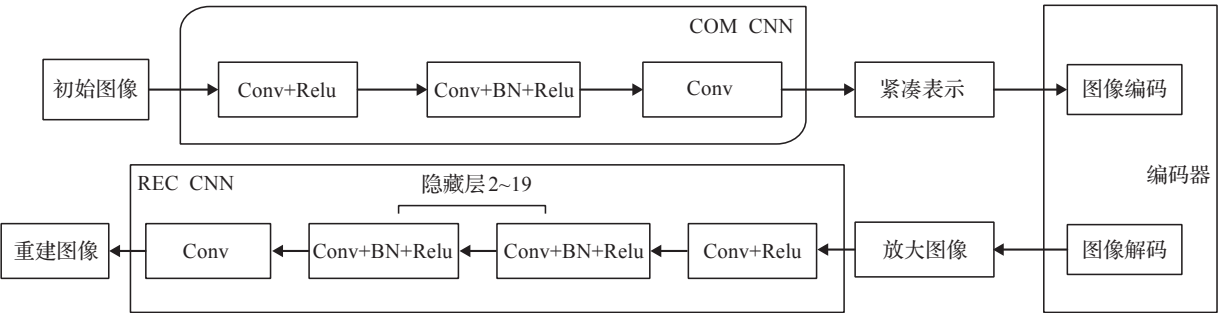


图 2 CNN 图像压缩框架图

架的能力,但同时会增加网络结构的训练难度。

3.2 基于循环神经网络的图像压缩方法

RNN 出现于 20 世纪 80 年代,RNN 最初因实现困难并没有被广泛使用,之后随着 RNN 结构方面的进步和 GPU 性能的提升使得 RNN 逐渐流行起来,目前 RNN 在语音识别、机器翻译等领域取得诸多成果。与 CNN 对比,RNN 与 CNN 一样都有参数共享的特性,不同的是 CNN 的参数共享是空间上的,而 RNN 则是时间上的,也就是序列上的,这使得 RNN 对于之前的序列信息有了“记忆”,同其训练方式是通过梯度下降的方式迭代向前计算。这两种方式一是可以提高数据的压缩程度,二是可以通过迭代的方式来控制图像的码率,都可以提高图像的压缩性能。因此应用 RNN 的图像压缩在对全分辨率图像压缩和通过码率来控制压缩比都取得了较为不错的成果,但值得注意的是在采用 RNN 时多数都需要引入 LSTM^[19]或者 GRU^[20]来解决长期依赖问题,因此在模型的训练上也会更加的复杂。

Toderici 等人^[21]首次使用了卷积 LSTM 实现了可变比特率的端到端学习图像压缩,可以说该方法是利用 RNN 进行图像压缩具有代表性的方法,它验证了任意的输入图像,在给定图像质量的情况下都能得到比目前最优压缩率更好的重建图像质量效果,但是这一效果限制在 32×32 尺寸的图像,这说明了该方法在捕捉图像依赖关系的不足,为了解决这一问题。Toderici 等人^[22]设计一种基于残差块的剩余编码器和一个熵编码器,不仅能够捕捉图像中补丁之间的长期依赖关系并结合两种可能的方法来提高给定质量的压缩率,并且实现了全分辨率的图像压缩。该方法利用 RNN 梯度下降的训练方式,提出了一种基于全分辨率的有损图像压缩方法。其结构如图 3 所示。该方法包括三个主要部分,分别为:Encoder 编码、Binarizer 二值化、Decoder 解码。首先对输入图像进行编码,然后将其转换成二进制代码,可以存储或传输到解码器。编码部分由一个 CNN 和三个 RNN 构成,Encoder 解码器网络根据接收到的二进制代

码创建原始输入图像的估计值。Binarizer 二值化部分主要通过一个 RNN 进行,Decoder 解码部分使用卷积-循环神经网络的结构对信号进行迭代来恢复原图像,在迭代的过程中权值共享,并且每次迭代都会产生一个二值化的比特数,同时在每次迭代中,网络从当前的残差中提取新的信息,并将其与存储在循环层的隐藏状态中的上下文相结合,通过这些信息实现图像的重建。该方法利用 RNN 的成功是有目共睹的,使更多人的目光转向了图像压缩。在此之后 Johnston 等人^[23]为了提高框架的压缩性能,修改了递归结构从而改善了空间扩散,使得网络能够更加高效地捕获图像信息;引入了一种空间自适应比特分配算法,它可以根据图像的复杂性动态的调整每个图像的比特率;采用了基于 SSIM 加权像素损失训练^[24-25],该方法可以更好地感知图像。

基于深度学习的图像压缩框架多采用端到端的方式,并且大多数图像压缩系统对空间块分别进行解码,而不考虑与周围块的空间依赖性,因此 Ororbria 等人^[26]没有采用端到端的压缩框架,而是关注了空间块的相关性,引入了一种有效利用因果信息和非因果信息来改进低比特率重构结构,更专注于系统的解码器,在算法的设计上采用了非线性估计作为编码器,将空间上像素的关联和非关联的相关性引入了 RNN 中,利用 RNN 的局部记忆捕捉短期的因果环境,通过 RNN 的记忆对图像斑块进行逐步改善重建,将图像压缩中重建图像的行为视为一个多步重建问题,建立一个模型使其在有限数量的通道上改进其对某些目标样本的重建效果,以逐步改善图像重建质量,达到在给定编码位数的情况下提高编码精度,并且根据不同的编码器和量化方案,寻求最优的非线性解码器,从而避开如近似、量化等问题,使其可以更好地利用开发编码器和量化操作。值得一提的是该方法可以用于任意的传统编码器中。

3.3 基于生成对抗网络的图像压缩方法

GAN 最早由 Goodfellow 等人提出^[14],目前在图像生成、图像风格迁移和视频帧生成等领域获得了很好的成

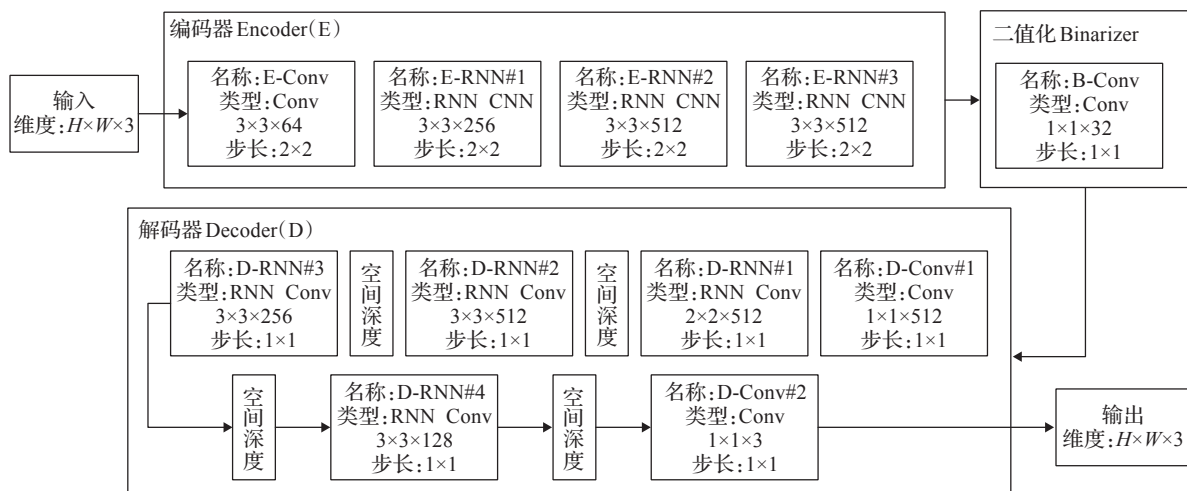


图3 RNN 图像压缩框架图

绩。近期在基于GAN的图像超分辨率^[27]也有了诸多成果。GAN的思想是对抗和博弈,在对抗中不断发展,一个生成器通过输入噪声样本进行生成数据,一个判别器用于接收生成器生成的数据和真实的数据样本,并且对输入的真实数据和生成数据做出正确的判断,通过对生成器和判别器的不断对抗,使网络架构得到优化。GAN根据这一特性,通过生成器的生成图像来不断“愚弄”判别器,使得最后得到的输出图像有更加清晰的纹理,更好的视觉感官效果。

GAN初期的发展由于其生成图像类型单一,模型训练难度大,研究人员并没有将目光投向这一算法,之后随着GPU运算效率的不断增加,Rippel等人^[28]提出了一篇实时自适应图像压缩算法,这是首次将GAN引入到图像压缩中,并且该算法在低码率条件下生成的文件要比传统的JPEG小2.5倍,通过GPU进行框架部署提高了实时性,该算法在率失真目标函数加入了一个多尺度对抗训练模型,使得重建图像与真实图像更加接近,即使在低比特率的情况下也能产生更清晰的图片,可以说该算法为基于GAN的图像压缩创建了基石;之前的基于深度学习图像压缩算法关注点多在重建图像分辨率或图像编解码结构的设计上,Santurkar等人研究的关注点与之前图像压缩算法不同,之前研究重建图像分辨率通常是对于像素目标的优化,而Santurkar等人^[29]提出了生成压缩模型,将合成变换训练成模型,替代图像重建的优化,该方法不仅能通过GAN生成高质量的图像,同时也与编码器进行了很好的结合,在编码器中加入GAN,通过不断优化网络结构得到更高质量的重建图像。但是GAN生成图像有着极大的不稳定性,在生成图像时有可能生成的图像具有清晰的纹理,很好的视觉效果,很高的分辨率和清晰度,但与原图对比却可能存在明显差异,这也就形成一种欺诈性的清晰与高分辨率。

通过GAN得到高清正确的重建图像并非易事,GAN的训练较为困难,在训练中要协调好生成器和判别器的训练程度,若判别器训练得过于优越那么会使生成器在训练时发生梯度消失等问题,而判别器训练的程度不够时,又会导致生成器无法生成理想的图像。为了得到更高分辨率的生成图像Agustsson等人^[30]提出了从语义标签映射中生成高分辨率重建图像的算法,该算法不仅在全分辨率的前提下实现了超低码率的极限压缩,同时也实现了在低码率时的高分辨率重建图像,其训练结构如图4所示,其中E和q分别表示编码器和量化, \hat{z} 则代表一个压缩表示,G和D分别为生成器与判别器,通过D来提升G的质量。他们分别采用了GAN、cGAN的生成图像压缩和具有选择性的生成压缩,生成压缩用于保留图像的整体结构,生成不同尺度的图像结构,选择性地生成压缩用于从语义标签映射中完全生成图像的各个部分,同时保留用户定义的具有高度细节的区域。在两种方式的共同作用下,保证重建图像的分辨率。

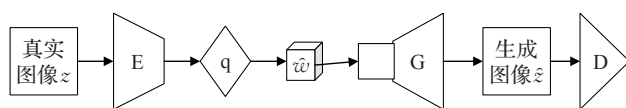


图4 GAN压缩网络结构

目前GAN研究发展火热,基于GAN的图像压缩也有很多优势:一是GAN可以对全分辨率图像进行压缩,体现了该方法有较好的适用性;二是GAN能够实现极限码率的图像压缩;三是尽管GAN生成的图像可能存在问题,但是其重建图像的分辨率和清晰度这一优点是值得被人重视的,尤其是近几年GAN的深入研究衍生了诸多基于GAN的模型,如:Denton等人^[31]提出了Lap-GAN(Laplacian Generative Adversarial Networks)将GAN和cGAN^[32]结合,并且通过图像金字塔的方式能够产生更高分辨率的图像;Radford等人^[33]将CNN与GAN结合到了一起,提出了DCGAN(Deep Convolutional GAN)的网络结构。DCGAN能有效地学习物体的特征,且在训练过程中表现更加稳定;Arjovsky等人^[34]提出WGAN(Wasserstein GAN),引入了Earth Mover(EM)距离缓解了GAN在训练时的不稳的问题。近期GAN的发展迅猛,这些衍生模型都可以尝试性地应用于图像压缩领域。

3.4 对比分析

基于深度学习的图像压缩涵盖了很多不同的算法,每种不同的算法都各有特点,CNN在提取特征方面要比传统的图像压缩变换更好,并且应用RNN和GAN处理图像压缩时也经常采用CNN进行图像特征提取;LSTM作为RNN的模型之一,LSTM模型可以很好地处理、合并空间信息,并且各种具有卷积运算的LSTM,这都使其可能更适用于图像压缩;GAN在对图像的极限压缩和提高重建图像质量,以及对图像数据实时性的压缩等方面表现良好。由于不同算法目的和评估的侧重点不同、使用数据集尺寸和类型也有所不同,文中所述所有算法多数都会给出其重建图像与JPEG或BPG的RD对比曲线,因此本文根据其文献中与JPEG或BPG图像的主观对比,以及压缩数据后对比,将对较为经典的算法进行对比与分析,表1为基于深度学习的图像压缩方法比较。

4 数据集与评价标准

图像数据集的种类与对图像进行评价的标准有很多,这一部分主要是对已经在基于深度学习技术的图像压缩方法中应用到的图像数据集和标准进行总结。

4.1 数据集介绍

使用深度学习的网络架构完成图像压缩训练时需要大量图像数据支撑,选择和采用正确图像数据集对网络结构训练的作用是至关重要的。表2主要介绍在基于深度学习的图像压缩算法应用过的数据集。

表2介绍了9种数据集,在具体的使用中需要根据

研究人员的实验目的和实验方法来合理的数据集应用。具体来讲:需要对实验的可行性进行分析时,可选用Cifar-10或LSUN中的一个场景,这类数据集数据量小,包含内容适中,训练速度快,可以满足实验设计的可行性;当实验目的定位需求在较高的重建图像分辨率时,可以采用DIV2K、Flickr这类数据集;当实验设计网络结构较深,需要大量且类型多样化的数据集时,可以采用ImageNet、Open Images V4这类数据集;在实验设计中需要一些条件特征来进行约束时,如在图像压缩使用cGAN时使用图像的语义信息来对生成器进行约束,就可以选择Cityscapes、COCO、LSUN这类带有语义注释的图像数据集。

用于测试的重建图像质量的数据集主要有以下Kodak PhotoCD^[47]、CLIC、RAISE-1k、Tecnick^[48]。这些数据集都有很高的分辨率,如:Kodak PhotoCD数据集的图像分辨率为762×512,且其像素约达40万;CLIC作为一个专门为图像压缩发起的挑战赛,其提供的图像照片的分辨率更高,手机图片的分辨率为1 913×1 361,专业相机图片的分辨率为1 803×1 175;Tecnick数据集的像素约达140万。

4.2 测试标准

图像压缩模型大多数采用端到端的形式,将深度学习技术应用到图像压缩中,当然也有采用自己独立的编码方式,因此图像评价上,多使用被压缩后的重建图像

表1 基于深度学习的图像压缩方法比较

模型名称	压缩程度	重建图像质量	其他
Generalized Normalization ^[15]	一般	一般	首次将广义分歧归一化的方法引入到了端到端的图像压缩
ComCNN+RecCnn ^[17]	一般	较好	模型较为复杂,训练难度较大,时间较长
End-to-End Learning ^[35]	一般	较好	基于ResNet的架构基础上,引入了感知损失和对抗损失以便生成更清晰的重建图像
Virtual Codec ^[18]	一般	较好	为了增加编码端与解码端的关联程度,引入了虚拟CNN进行训练
Residual CNN ^[36]	一般	一般	使用CNN进行图像特征提取,在重构图像时使用SRGAN提高重建图像质量
CNNC ^[37]	可控	较好	利用卷积与反卷积实现图像的压缩与重建,通过控制卷积层和池化层的层数控制压缩比
Variable Rate ^[21]	可控	一般	首次提出了基于LSTM的端到端图像压缩,并且可以拓展到视频中
Full resolution ^[22]	可控	优秀	实现了全分辨率的RNN图像压缩,模型结构复杂,训练较为困难
Entropy Model ^[38]	较高	优秀	使用了3D-CNN的条件概率压缩模型
Real-time ^[28]	较高	优秀	第一个采用GAN的多尺度实时性图像压缩
Generative Model ^[39]	极高	优秀	首次使用GAN对图像进行极低比特率的图像压缩
Generative compression ^[29]	较高	优秀	尽管有清晰的生成图像但是生成图像单一、不稳定,有可能与真实图像存在偏差,训练时优化困难
Extreme Learned ^[30]	极高	优秀	可以通过语义提高重建图像分辨率
Extremely Low Bitra-tes ^[40]	极高	一般	将自编码器变体与GAN结合,实现了极低比特率下的图像压缩

表2 数据集介绍

名称	图像数量及简介	备注
ImageNet ^[41]	1 500多万张图像,2万多类图像并且带有标签	ImageNet通常用于图像分类、目标检测等任务,由于其图像数据量大,因此在实际实验中可选用需要的类型即可
Yahoo Flickr	图像存储及视频托管网站,存储超过6亿张图片	Flickr图像均为用户上传图像,其图像分辨率较高,图像类型多样,图像尺寸不同
Open Images V4 ^[42]	具有对象位置注释数据集,约包含900万张图像	该数据集源自图像挑战赛,其图像类型多样,图像内容场景复杂,图像中包含对象较多,并对图像中对象位置有注释
DIV2K ^[43]	1 000张高清图,其中800张作为训练,100张作为验证,再100张作为测试	数据集拥有超高的分辨率,并且有较强的多样性,是NTIRE2018指定的训练数据集,多用于图像超分辨率任务中
Cityscapes ^[44]	5 000幅图像,500张验证图和1 525张测试图	包含50个城市的一年四季的图像,有高质量的图像注释,从27个城市手动选择5 000幅图像进行了密集的注释
Place365 ^[45]	包括365个场景的180万张图像,每个场景最多包含5 000张图像	Places365有两个版本:Places 365-standard和Places365-challenge,该数据集标有场景的语义类别,多用于场景分类任务中
COCO	包括80类超过33万张图像,其中20万张图像附有标注	COCO多用于目标识别、物体检测、对象分割等领域,其单一类别图像数据量大,并且类别丰富
Cifar-10	32×32彩色图像包含5万个训练样本,10个类别	经典的图像分类、图像识别数据集,图像分辨率不高,类别较少
LSUN ^[46]	包含10个场景类别20个对象类别,约100万个标记图像	主要包含了卧室、客厅、教室等10类场景,涵盖了基本的生活场景。可用于大规模场景理解

进行图像压缩性能的评估。

均方误差(Mean Square Error, MSE)^[49],是计算两幅相同尺寸图像像素之间的方差平方和,如式(1)所示。其中 M 和 N 分别表示图像的长与宽, $I(x, y)$ 和 $I'(x, y)$ 分别表示待评价图像与原始图像, $I(x, y)$ 表示在 (x, y) 位置的像素值。MSE 是最简单的图像评价方式,但是这只能说明两幅图像的差异,它不会考虑到图像高频像素和低频像素分量,而图像压缩就是需要能更大程度上保留和恢复低频像素,因此该方法不为常用。

$$MSE = \frac{1}{MN} \sum_{y=1}^N \sum_{x=1}^M [I(x, y) - I'(x, y)]^2 \quad (1)$$

峰值信噪比(Peak Signal to Noise Ratio, PSNR)^[50],是两个图像峰值误差的度量,如公式(2)所示, R 表示输入图像的最大值。PSNR 是全参考图像评价指标,是一种客观的评价指标,通常来讲 PSNR 大于 40 dB 说明图像质量接近原图,在 30 dB 与 40 dB 之间时图像存在失真,20 dB 到 30 dB 说明图像质量不好,低于 20 dB 时说明图像质量差。PSNR 与 MSE 通常只针对图像对应像素点间的误差,并不会考虑到人的视觉特性,因此往往存在出现评价结果与人的主观感受不一致的情况。

$$PSNR = 10 \lg \frac{R^2}{MSE} \quad (2)$$

结构相似性(Structural Similarity Index, SSIM)^[51],是用于判断两幅图像相似性的指标,SSIM 可用于测量经过图像压缩后的图像质量下降,如公式(3)所示, x 和 y 分别代表两幅图像, μ_x 、 μ_y 代表两幅图像的平均像素值, σ_{xy} 是图像 x 和图像 y 的协方差, σ_x^2 、 σ_y^2 分别代表图像 x 和图像 y 的方差。SSIM 在进行评价时考虑到了图像中的可见结构,可以很好地评估图像压缩前与图像压缩后的图像质量。

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3)$$

多层结构相似性(Multi-Scale, MS-SSIM)^[52],该方法在融合图像分辨率和观测条件变化方面比单尺度方法具有更大的灵活性,该方法通过迭代的方式对图像的各个尺度上进行 SSIM,如公式(4)所示,其中 $l_M(X, Y)$ 表示在 M 尺度下亮度比较, j 表示在第 j 尺度下进行比较, $c_j(X, Y)$ 表示对比度比较, $s_j(X, Y)$ 表示结构比较,将不同尺度下的测量结构结合起来进行 SSIM 评价作为最后的评价结果。对于指数 α_M 、 β_j 和 γ_j 可以取经过实验得出的经验值即可,该评价标准已经被证明比 SSIM 更贴近人的主观视觉。

$$MS-SSIM(X, Y) = [l_M(X, Y)]^{\alpha_M} \cdot \prod_{j=1}^M [c_j(X, Y)]^{\beta_j} [s_j(X, Y)]^{\gamma_j} \quad (4)$$

MSE 和 PSNR 是最简单和最广泛使用的全参考质

量度量,其通过平均失真和参考图像像素的平方强度差异以及峰值信噪比的相关数量来计算的。这些方法计算简单,物理意义明确,但是其评价存在客观性。SSIM 和 MS-SSIM 都为主观的评价方式,这种评价方式最大的优点就是更加贴近人的主观视觉,因此在图像压缩的实验中,研究人员多采用客观评价和主观评价结合来对重建图像进行评估,即采用 PSNR 也会使用 MS-SSIM 或 SSIM。除了在采取这两种措施的同时,也有研究会让人一部分人来观察重建图像,通过观察人员给出的结论进行统计来对图像进行评价。

5 深度学习在图像压缩领域的未来发展

进入信息时代后,人们对数字图像的质量要求越来越高,数字图像也向着更清晰、更高分辨率的方向发展^[53]。随着大数据时代的到来,图像数据量的增长速度远超存储设备和传输技术的发展速度,只增加存储容量和网络带宽并不是解决问题的根本方法,寻找更加合理的图片压缩算法是解决这一问题的有效办法之一。深度学习能够有效地提取图像的特征信息,不仅能清晰地分辨出图像的重要信息与冗余信息,同时也可以对特征信息进行很好地表达,还可以对图像信息进行高分辨率地重建,使图像在消除冗余信息的同时保持更清晰的分辨率和更好的视觉效果。因此深度学习技术将在图像压缩领域得到很好的发展。

图像压缩的目的不仅仅是追求更小单位的数据量,在图像压缩的同时也追求更好的压缩比和更高的重建图像清晰度。下面根据深度学习的特点和图像压缩领域的优势对其未来发展趋势进行总结与讨论:

(1)深度学习对图像处理任务本身就有着很强的能力,尽管目前有了诸多成果,但如何使用深度学习方法得到高层次紧凑表达;如何通过深度学习的预测能力、记忆能力对图像上下文关系信息进行更为高效编码;如何利用 GAN 等方式图像的生成能力,生成具有更为真实图像纹理信息的重建图像;如何设计更好的网络模型结构以及模型参数的调优方式来提高图像压缩的泛化性;图像压缩初期提取的高层次紧凑信息能否应用于机器视觉等其他应用,这些问题都是基于深度学习的图像压缩需要不断深入研究的热点问题。

(2)由于图像类型的多样化,不同的图像有各自的特征,如海洋图像大部分以蓝色为基调、CT 图像需要对病理区域更加清晰、多光谱图像数据量大等特征。传统的图像压缩算法不会适用所有的图像类型,深度学习技术可以根据不同图像的特征和需求设计针对性图像压缩,如孔繁铨等人^[54]提出基于 CNN 的多光谱图像压缩方法,就在保证图像信息的情况下进一步提升了压缩性能。因此根据特殊图像的需求,使用深度学习进行针对

性图像压缩也是研究方向之一。

(3) 图像编码技术需要实时性^[55], 如今基于深度学习的图像压缩框架多为深度神经网络, 尽管有着很好的重建图像分辨率, 但难免对实时性和高效有所保证, 在面对海量流数据传输时也就失去实用性的意义, 因此在研究中保障重建图像质量和高压压缩比的同时, 如何使用低复杂度的深度学习方法, 提高算法实时性, 寻求高性能、高时效性算法也将是该领域的热点研究问题之一。

(4) 基于深度学习的图像压缩对于重建图像已经有了很高的还原度, 但是目前的图像压缩评价指标多为 PSNR 和 SSIM 或 MS-SSIM, 但是这些指标并不能十分全面精准地衡量重建图像质量, 同时完全采用人的视觉来进行主观评价将耗费很大的人力与时间, 但在现实应用中, 图像更多是被人所观看、应用, 图像质量往往由人来进行评估。因此建立一个更为精准的图像评价指标也将是基于深度学习技术的图像压缩领域一个热点研究。

6 结束语

图像压缩领域经过了几十年的不断发展, 虽然已经有了十分成熟的算法与标准, 但面对 5G 时代的海量数据难免捉襟见肘, 随着深度学习技术的突破, 基于深度学习的图像压缩应用不断出现, 深度学习大概率成为图像压缩领域未来发展的助推器。本文通过对传统图像压缩算法的简述, 分析了传统方法目前存在的问题以及深度学习可以在传统方法上所作的提升。根据不同的深度学习网络结构, 分别对近几年有代表性文献进行了介绍与对比, 分析了不同深度学习方法应用于图像压缩领域的优点与不足, 最后依据深度学习的算法特点、图像压缩的实时性需要、图像压缩的评价指标对基于深度学习的图像压缩研究内容进行了讨论与展望。

参考文献:

- [1] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. Nature, 1986, 323(6088): 533-536.
- [2] 史加荣, 马媛媛. 深度学习的研究进展与发展[J]. 计算机工程与应用, 2018, 54(10): 1-10.
- [3] Wallace G K. The JPEG still picture compression standard[J]. IEEE Transactions on Consumer Electronics, 1991, 38(1): 43-59.
- [4] Rabbani M, Joshi R. An overview of the JPEG 2000 still image compression standard[J]. Signal Processing: Image Communication, 2002, 17(1): 3-48.
- [5] Zhao L, Bai H, Wang A, et al. Two-stage filtering of compressed depth images with Markov random field[J]. Signal Processing: Image Communication, 2017, 54: 11-22.
- [6] 李云红, 王珍, 张凯兵, 等. 基于学习的图像超分辨重建方法综述[J]. 计算机工程与应用, 2018, 54(15): 13-21.
- [7] Liew A W C, Yan H. Blocking artifacts suppression in block-coded images using overcomplete wavelet representation[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2004, 14(4): 450-461.
- [8] Wang C, Zhou J, Liu S. Adaptive non-local means filter for image deblocking[J]. Signal Processing: Image Communication, 2013, 28(5): 522-530.
- [9] Nyquist H. Certain factors affecting telegraph speed[J]. Transactions of the American Institute of Electrical Engineers, 1924, 43: 412-422.
- [10] Weinberger M J, Seroussi G, Sapiro G. The LOCO-I lossless image compression algorithm: principles and standardization into JPEG-LS[J]. IEEE Transactions on Image Processing, 2000, 9(8): 1309-1324.
- [11] Said A, Pearlman W A. An image multiresolution representation for lossless and lossy compression[J]. IEEE Transactions on Image Processing, 1996, 5(9): 1303-1310.
- [12] Dong C, Deng Y, Change Loy C, et al. Compression artifacts reduction by a deep convolutional network[C]// Proceedings of the IEEE International Conference on Computer Vision, 2015: 576-584.
- [13] Mikolov T, Karafiát M, Burget L, et al. Recurrent neural network based language model[C]// Eleventh Annual Conference of the International Speech Communication Association, 2010.
- [14] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]// Advances in Neural Information Processing Systems, 2014: 2672-2680.
- [15] Ballé J, Laparra V, Simoncelli E P. End-to-end optimized image compression[J]. arXiv: 1611.01704, 2016.
- [16] Ballé J, Minnen D, Singh S, et al. Variational image compression with a scale hyperprior[J]. arXiv: 1802.01436, 2018.
- [17] Jiang F, Tao W, Liu S, et al. An end-to-end compression framework based on convolutional neural networks[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 28(10): 3007-3018.
- [18] Zhao L, Bai H, Wang A, et al. Learning a virtual codec based on deep convolutional neural network to compress image[J]. Journal of Visual Communication and Image Representation, 2019, 63: 102589.
- [19] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8).
- [20] Chung J, Gulcehre C, Cho K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv: 1412.3555, 2014.
- [21] Toderici G, O'Malley S M, Hwang S J, et al. Variable rate image compression with recurrent neural networks[J]. arXiv: 1511.06085, 2015.
- [22] 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.

- [23] 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.
- [24] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
- [25] Zhao H, Gallo O, Frosio I, et al. Loss functions for image restoration with neural networks[J]. IEEE Transactions on Computational Imaging, 2016, 3(1):47-57.
- [26] Ororbia A G, Mali A, Wu J, et al. Learned neural iterative decoding for lossy image compression systems[C]//2019 Data Compression Conference(DCC), 2019:3-12.
- [27] 李诚,张羽,黄初华.改进的生成对抗网络图像超分辨率重建[J]. 计算机工程与应用, 2020, 56(4):191-196.
- [28] Rippel O, Bourdev L. Real-time adaptive image compression[C]//Proceedings of the 34th International Conference on Machine Learning-Volume 70, 2017:2922-2930.
- [29] Santurkar S, Budden D, Shavit N. Generative compression[C]//2018 Picture Coding Symposium(PCS), 2018:258-262.
- [30] 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.
- [31] Denton E L, Chintala S, Fergus R. Deep generative image models using a laplacian pyramid of adversarial networks[C]//Advances in Neural Information Processing Systems, 2015:1486-1494.
- [32] Mirza M, Osindero S. Conditional generative adversarial nets[J]. arXiv:1411.1784, 2014.
- [33] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv:1511.06434, 2015.
- [34] Arjovsky M, Chintala S, Bottou L. Wasserstein gan[J]. arXiv:1701.07875, 2017.
- [35] Liu H, Chen T, Shen Q, et al. Deep image compression via end-to-end learning[C]//IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018:2575-2578.
- [36] Deshmukh K, Pollett C. Residual CNN image compression[C]//International Symposium on Visual Computing. Cham: Springer, 2019:371-382.
- [37] 崔建良,李建飞,陈春晓,等.基于CNNC的卷积神经网络图像的压缩方法[J]. 生物医学工程研究, 2019, 38(4):415-419.
- [38] 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.
- [39] Tschannen M, Agustsson E, Lucic M. Deep generative models for distribution-preserving lossy compression[C]//Advances in Neural Information Processing Systems, 2018:5929-5940.
- [40] 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, 2020:2325-2333.
- [41] 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, 2009:248-255.
- [42] Kuznetsova A, Rom H, Alldrin N, et al. The open images dataset v4: unified image classification, object detection, and visual relationship detection at scale[J]. arXiv:1811.00982, 2018.
- [43] Agustsson E, Timofte R. Ntire 2017 challenge on single image super-resolution: dataset and study[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017:126-135.
- [44] 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, 2016:3213-3223.
- [45] Zhou B, Lapedriza A, Khosla A, et al. Places: a 10 million image database for scene recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(6):1452-1464.
- [46] Yu F, Seff A, Zhang Yinda, et al. LSUN: construction of a large-scale image dataset using deep learning with humans in the loop[J]. arXiv:1506.03365, 2015.
- [47] Franzen R. Kodak lossless true color image suite[EB/OL]. (1999-04). <http://r0k.us/graphics/kodak>.
- [48] Asuni N. Tecnick test image library[EB/OL]. (2011). <http://testimages.tecnick.com>.
- [49] Eskicioglu A M, Fisher P S. Image quality measures and their performance[J]. IEEE Transactions on Communications, 1995, 43:2959-2965.
- [50] Tanchenko A. Visual-PSNR measure of image quality[J]. Journal of Visual Communication and Image Representation, 2014, 25(5):874-878.
- [51] Wang Z, Bovik A, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[C]//IEEE Transactions on Image Processing, 2014, 13:600-612.
- [52] Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment[C]//The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003:1398-1402.
- [53] 任杰. 基于深度学习的图像压缩方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2017.
- [54] 孔繁镭, 周永波, 沈秋, 等. 基于卷积神经网络的端到端多光谱图像压缩方法[J]. 中国激光, 2019, 46(10):285-293.
- [55] 李崇. 基于神经网络的实时性图像压缩算法研究[D]. 北京: 北京邮电大学, 2019.