文章编号：1671-7848(2022)03-0536-06 **DOI:** 10.14107/j.cnki.kzgc.20210542

基于自注意力机制增强的深度学习图像压缩

展亚南[1]，施晓东[1]，孙镱诚[1]，丁阳[1]，杨万扣[2]

（1. 中国电子科技集团公司第二十八研究所，江苏 南京 210007；2. 东南大学 自动化学院，江苏 南京 211189）

摘 要：提出了一种基于自注意力机制增强的深度学习模型，用于无人机侦察图像的压缩与解压。与现有方法相比，提出的深度学习模型有两个显著特点：其一，模型由四部分组成（编码器、二值化器、量化器和解码器），并且可以通过端到端的优化提高模型的压缩和解压效率；其二，量化器是基于自注意力机制增强的多层前馈神经网络，它能充分利用图像的上下文信息对图像进行压缩。在公开数据集 Kodak 和 Tecnick 的实验结果表明，提出模型的压缩率−保真率曲线优于传统的图像压缩标准和现有的深度学习模型。对于常规大小的图像，在保持图像质量 MS-SSIM 为 85%~95%的前提下，图像压缩比 BPP 能达到 7%~15%，并且在普通 CPU 上其处理速度达 0.48 秒/张，能显著降低影像的数据大小且不牺牲处理速度。

关键词：图像压缩；深度学习；自注意力机制；端到端；多层前馈神经网络

中图分类号：TP18 **文献标识码：**A

A Self-attention Mechanism Augmented Deep Learning Model for Images Compression

ZHAN Ya-nan[1], *SHI Xiao-dong*[1], *SUN Yi-cheng*[1], *DING Yang*[1], *YANG Wan-kou*[2]

(1. The 28th Research Institute of China Electronics Technology Group Corporation, Nanjing 210007, China; 2. School of Automation, Southeast University, Nanjing 211189, China)

Abstract: A self-attention mechanism augmented deep learning model is proposed to compress and decompress the UAV reconnaissance image in this paper. Compared with the existing methods, the proposed deep learning model in this paper has two significant characteristics. Firstly, the model consists of four parts (encoder, binarizer, quantizer and decoder), and the compression and decompression efficiency of the model can be improved through end-to-end optimization. Secondly, the quantizer is a self-attention mechanism augmented multi-layer feedforward neural network, which can make full use of the context information to compress the image. Experimental results on public data sets such as Kodak and Tecnick show that the Bits Per Pixel-Peak Signalto Noise Ratial (BPP-PSNR)curve of the proposed model is better than that of the traditional image compression standards and existing deep learning models. For images with commonly used size, the compression ratio of the model, e.g. BPP, can reach 7%~15% while maintaining the MS-SSIM of 85%~95%, and the processing speed can reach 0.48 s/sheet on the ordinary CPU. The model proposed in this paper can significantly reduce the data size of the compressed image without sacrificing the processing speed.

Key words: Image compression; deep learning; self-attention mechanism; end-to-end; multi-layer feedforward neural network

1 引 言

图像压缩的目的是为了减少图像的冗余信息，以较低的比特率存储或传输图像。图像具有局部相似的特点，相邻图像块或者像素之间具有很强的相关性，从统计学的角度来说，这些相关性会存在大

收稿日期：2021-07-02；**修回日期：**2021-08-28

预研项目：装备预先研究项目(301021302)

作者简介：展亚南(1990−)，女，河南周口人，硕士，工程师，主要从事态势情报信息处理、展现等方面的科研工作（本文通信作者，Email: zhanyanan03@163.com）。

量的冗余信息。深度学习模型是目前较好的处理图像数据的机器学习模型，其在图像识别、目标检测与跟踪、图像分割等图像处理任务中有非常好的效果。深度学习的基本思路是通过一个多层的神经网络建模一组多层数据驱动的非线性变换，在这个过程中图像逐渐由像素表示转为语义表示。基于深度学习模型的图像压缩编码方法研究思路是希望在图像编码阶段得到图像的语义表示；在图像的解码阶段再通过语义表示恢复图像的像素表示。在图像存储和传输时只需保留图像的语义表示，这样能在最大程度上压缩图像所占的空间。

2 相关工作

现有图像压缩标准中（如JPEG[1]、JPEG2000[2]和Better Portable Graphics(BPG)[3]等）的编码器和解码器是分开优化的。在编码阶段，首先对图像执行一个线性变换，然后利用量化和无损熵编码来最小化压缩率。在解码阶段，通过设计译码算法和逆变换，使失真率最小化。然而，这类图像压缩方法往往存在压缩伪影的问题，特别是在低压缩率的情况下。为了解决这个问题，研究人员先后提出了几种改进的传统方法[4]和基于深度卷积神经网络(convolutional neural networks, CNN) 模型的方法[5]。Jiang等[6]提出了ComCNN，在传统编解码器对图像进行编码之前，对图像进行预处理，以及提出了RecCNN，用于对传统的解码结果进行后处理。

近年来，基于深度学习的图像压缩模型取得了一定进展[7,10]。对于无损图像压缩，深度学习模型已经达到了最先进的性能[11,13]。对于有损图像压缩，Toderici等[14,20]提出了一种循环神经网络(RNN)用来压缩32×32图像，并进一步提出一种用于图像渐进编码和解码的全分辨率压缩方法。Lin等[16]提出用RNN建模算术编码的过程，进一步压缩图像的空间冗余信息。与本文最相关的是基于卷积自动编码器的工作[17,19]。Ballé等[17]提出了除数归一化(generalized divisive normalization, GDN)，并用加性均匀噪声代替舍入量化，实现失真率损失和熵率损失的可微松弛。Theis等[18]提出了取整函数导数的光滑近似，并利用离散熵率损失的上界进行可微松弛。Agustsson等[19]提出了一种软到硬的量化处理方法，可以简化网络的训练。Rippel等[13]提出一种深度自动编码器，具有生成式对抗训练的特点，可以实时运行。

目前，基于深度学习的图像压缩的理论和方法仍然在不断发展中，现有的图像压缩方法没有解决好如下两个问题：① 图像的上下文信息在图像的编码和解码中没有得到充分利用；② 现有的图像压缩技术的各个模块是相互独立的，缺乏一个端到端的系统，不能够在训练的过程中同时优化图像压缩的编码器、量化器和解码器，从而导致图像的压缩能力有限。

本文将围绕解决面向图像压缩的卷积神经网络模型存在的两个问题展开研究。

3 本文提出的模型

本文提出的图像压缩模型由四部分组成：编码器(encoder)、二值化器(binarizer)、量化器(quantizer)和解码器(decoder)。给定输入的图像 x，基于卷积网络(CNN)的编码器定义了一组非线性变换，并输出图像编码 $E(x)$。量化器中CNN，它接收编码器的中间特征函数 $F(x)$ 作为输入，输入一组用于度量重要程度(Importance)的内容加权(content-weighted)的特征函数 $P(x)$。之后采用舍入函数(rounding Function) $Q(\cdot)$ 量化 $P(x)$，再通过掩码算子 $M(\cdot)$ 使得量化器的输入和编码器的输出 $E(x)$ 保持同维。首先将二值化器接收编码器的输出 $E(x)$ 作为输入，并对其做二值化处理 $B(\cdot)$（比如大于0.5取值为1，其它情况取值为0）得到 $B\big(E(x)\big)$；再结合量化器输出 $M\big(Q\big(P(x)\big)\big)$，得到修剪后的二进制编码 c；最后，CNN解码器所建模的重构函数 $D(\cdot)$ 接收二进制编码 c 作为输入，并输出重构后的图像 $\hat{x}$。下文将先介绍基于CNN的编码器和解码器，然后介绍二值化器和基于内容加权的量化器。本文提出的深度学习模型结构示意图，如图1所示。

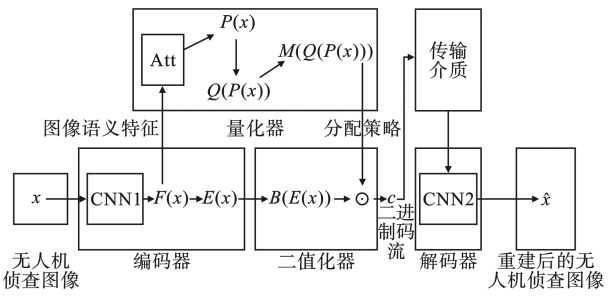

图 1 本文提出的深度学习模型结构示意图

Fig. 1 Architecture of deep learning model proposed in this paper

3.1 基于CNN的编码器和解码器

本文的编码器和解码器都是 CNN 结构，记为 CNN1 和 CNN2，它们可以通过反向传播算法训练。编码器网络由 3 个卷积层和 3 组残差模块构成。其中每一个残差模块有两个卷积层。类似于单图像的超分辨率重建模型，残差模块里去掉了批标准化(batch normalization)操作。有实验结果表明这样的设置有助于抑制光滑区域的视觉压缩伪影。编码 CNN 的输入是图像 x，其由三组卷积(Conv)操作构成，每一组卷积之后紧跟着一个采样因子为 2 的下采样(down-sampling)操作，在下采样操作之后紧接着 DenseBlock，这个模块由 7 个卷积操作构成。在编码器的最后一层，我们增加了一个带有多滤波器的卷积层。从经验上来说，滤波器的取值应该被设置为压缩模型的比特率的上限。用于编码器和解码器的 CNN 的体系结构示意图如图 2 所示。

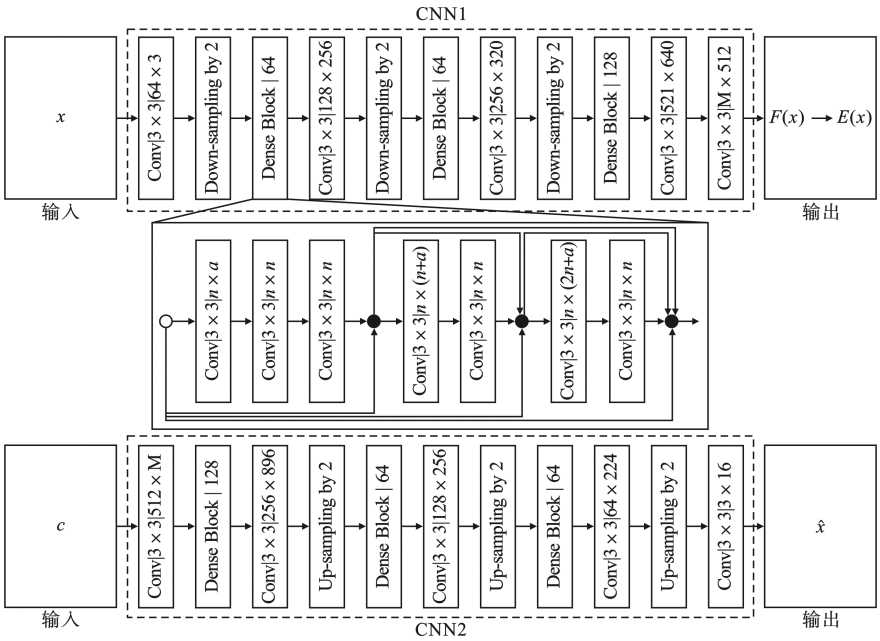

图 2 用于编码器和解码器的 CNN 的体系结构示意图

Fig. 2 Architecture of CNNs for encoder and decoder

解码器 $D(c)$ 的网络结构与编码器的网络结构是对称的，其中 c 是图像 x 二进制编码。为了得到上采样的特征图，本文采用了 Toderici 等[20]提出的“Depth-to-Space”模块。解码器的最后一个卷积层由 3 个滤波器构成，分别对应图像的 RGB3 个道。

3.2 二值化器

编码器的最后一层的激活函数是sigmoid函数，编码器的输出 $e=E(x)$ 的取值范围是[0,1]。令 e_{ijk} 表示 e 中的元素，则二值化器定义为

$$B(e_{ijk})=\begin{cases}1, & \text{if } e_{ijk}>0.5\\ 0, & \text{otherwise}\end{cases} \tag{1}$$

然而这样的二值器函数 $B(e_{ijk})$ 的导数值，除了在 $e_{ijk}=0.5$ 处为 ∞，其他情况下都为0。这使得网络在训练过程中，二值器之前的所有层的参数都不能得到更新。

基于最近的二值化神经网络(binarized neural network, BNN)的研究[21,23]，本文在这里通过引入代理函数 $\tilde{B}(\cdot)$ 来逼近 $B(\cdot)$。在前向传播的过程中函数 $B(\cdot)$、代理函数 $\tilde{B}(\cdot)$ 只是在反向传播的时候使用。受BNN的启发，这里设计了一种分段线性函数 $\tilde{B}(\cdot)$ 作为 $B(\cdot)$ 的近似。

$$\tilde{B}(e_{ijk})=\begin{cases}1, & \text{if } e_{ijk}>1\\ e_{ijk}, & \text{if } 0\le e_{ijk}\le 1\\ 0, & \text{otherwise}\end{cases} \tag{2}$$

这样便可以得到有效的梯度信息。

$$\tilde{B}'(e_{ijk})=\begin{cases}1, \text{if } 0\le e_{ijk}\le 1\\ 0, \text{otherwise}\end{cases} \tag{3}$$

3.3 基于自注意力机制的量化器

在最近发表的文献[17,18]中，量化后的编码长度是一个空间不变量，还需要熵编码(entropy coding)操作进行进一步压缩。事实上，图像不同区域的信息的压缩难度应该是不同的。平滑的区域比那些有突出物体或者丰富纹理的区域要容易压缩。因此平滑的区域应该分得较少的比特数，而那些有突出物体或者丰富纹理的区域应该分得较多的比特数。因此本文提出一种基于自注意力机制的内容加权的重要性量化器，用于比特数分配和压缩率控制。这个映射只有一个通

道，它的大小与编码器的输出一致，取值范围是(0,1)，它以编码器的中间特征函数 $F(x)$ 作为输入，通过堆叠 N 个自注意力模块Att产生输出 $p=P(F(x))$ 。对于每一个Att，其中 E_p 表示位置嵌入； Q,K,V 分别表示Query,Key,Value； A 表示每次查询的输出；Add表示求和；Norm表示规范化；Position-wise FFN表示前馈神经网络。堆叠多个自注意力模块Att能强化输入特征之间的上下文关系，提高编码的压缩效率。基于自注意力机制的量化器示意图如图3所示。

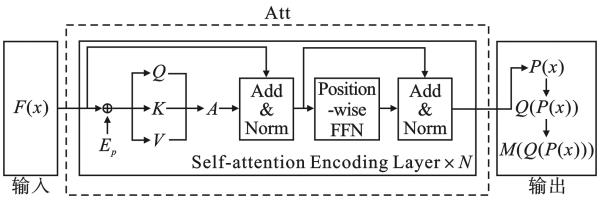

图 3 本文提出的基于自注意力机制的量化器示意图
Fig. 3 Schematic diagram of quantizer based on self attention mechanism

令 $h\times w$ 表示特征函数 $P(x)$ 的输出大小， n 表示编码器网络输出的特征函数的数目。为了指导比特数的分配，首先将 p 中的每个元素量化为不大于 n 的整数，然后生成一个大小为 $n\times h\times w$ 的重要性掩码。给定 p 中的一个元素 p_{ij} ，从量化器到重要性映射定义为

$$Q(p_{ij})\equiv l-1,\ \text{if}\ \frac{l-1}{L}\le p_{ij}\le\frac{1}{L},\ l=1,2,\cdots,L \quad (4)$$

式中， $L\in\{16,32\}$ 为重要性等级； $n \bmod L=0$ 。每一个重要性等级对应的比特数为 n/L 。由于 $p_{ij}\in(0,1)$ ，因此 $Q(p_{ij})$ 只有 L 个不同的数值，即 $0,1,\cdots,L-1$ 。注意到当 $Q(p_{ij})=0$ 时，其所对应位置不需要分配比特数，其全部信息都可以在解码阶段通过其上下文信息重建。从这个角度看，重要性特征函数不仅可以作为熵率估计的替代方法，而且可以自然地考虑上下文信息。

有了 $Q(p_{ij})$ 之后，重要性掩码可以通过式(5)计算：

$$m_{ijk}=M(p_{ij})\equiv\begin{cases}1, \text{ if } \dfrac{n}{L}Q(p_{ij})\ge k\\ 0, \text{ otherwise}\end{cases} \quad (5)$$

输入图像 x 的最终编码结果 c 可表示为

$$c=M(p)\circ B(e) \quad (6)$$

式中，$\circ$ 表示元素级的点积运算。

注意到这个编码中考虑到内容重要性，故而 $B(e)$ 中所有掩码取值为0的比特可以被移除。因此，对于每一个位置只需要 $Q(p_{ij})n/L$ 比特，而不是 n 比特。类似于式(1)的二值化器函数，式(4)的量化函数和式(5)的掩码函数使得 m 关于 p 的梯度也几乎处处为0。为了解决这个问题，首先把式(4)的量化函数和式(5)的掩码函数合并重写为

$$m_{ijk}=\tilde{M}\circ\tilde{Q}(p_{ij})\equiv\begin{cases}1, \text{ if ceiling}(\dfrac{kL}{n})<Lp_{ij}\\ 0, \text{ otherwise}\end{cases} \quad (7)$$

式中，ceiling函数表示取上整。

类似于式(3)二值化器的梯度， m 关于 p 的梯度可以写为

$$\frac{\partial m_{kij}}{\partial p_{ij}}\equiv\begin{cases}L, \text{ if } Lp_{ij}-1\le\text{ceiling}(\dfrac{kL}{n})\le Lp_{ij}+1\\ 0, \text{ otherwise}\end{cases} \quad (8)$$

4 目标函数和模型训练

一般来说，本文所提出的内容加权图像压缩可以被定义为一个率失真优化问题。这里的优化目标是最小化编码失真和编码比特组合。为了平衡编码失真和压缩率，引入了一个折衷参数。模型的目标函数定义如下：

$$L=\sum_{x\in X}L_D(c,x)+\lambda L_R(x) \quad (9)$$

式中， $L_D(c,x)$ 为编码失真损失函数； $L_R(x)$ 为编码比特代价函数。

编码失真函数用于度量输入图像和重建图像之间的失真程度，其定义如下：

$$L_D(c,x)=\|x-D(c)\|_2^2 \quad (10)$$

压缩率损失函数用于度量图像压缩后的编码长度。本文提出的修剪后的二进制编码 $n/L\sum_{i,j}Q(p_{ij})$ 可以作为压缩率损失函数，但是由于量化函数 $Q(\cdot)$ 的导数问题使得直接采用 $n/L\sum_{i,j}Q(p_{ij})$ 作为压缩率损失函数会带来训练困难的问题。因此，这里把 $Q(p)$ 放松到其连续形式 $p=P(x)$ ，并引入一个阈值 r 用于控制压缩率。压缩率损失函数定义为

$$L_R(P(x))=\begin{cases}\sum_{i,j}p_{ij}-r, \text{ if } \sum_{i,j}p_{ij}>r\\ 0, \text{ otherwise}\end{cases} \quad (11)$$

得益于松弛后的压缩率损失函数，整个基于内容加权的图像压缩模型的梯度是可以直接计算的，因此整个压缩系统可以用端到端的方式训练。这里采用的带冲量的随机梯度下降算法Adam作为模型优化器，相关的超参数设置为 $\beta_1=0.9$, $\beta_2=0.98$, $\beta_3=0.98$, $\varepsilon=10^{-9}$ 。采用可变的学习率 $\varsigma(t)=d^{-0.5}\times\min\{t^{-0.5},t\times w_s^{-0.5}\}$ ，其中， t 表示训练步长数； $d=$

$d(x)+d(p)$表示模型输出的维数；w_s=20 000表示预热步长数。从上式可以看出，在预热步长以内，学习率随步长的增加而线性增加；超过预热步长以后，学习率随步长的平方根的倒数等比例减少。

5 实　验

本文提出的内容加权图像压缩模型首先在ImageNet[24]和Open Image V4[25]上训练。在训练之前需要将这些图像大小规范到752×496，再裁剪成24个128×128的小块。经过训练之后，我们在Kodak数据集和Tecnick数据集[27]上测试了本文提出的模型，并对其有损图像压缩的压缩效果进行了度量。其中压缩率所采用的度量指标是像素平均比特率(bits per pixel, BPP)，它是用编码图像的总比特数除以总像素数来计算的。保真率采用的评价指标为：多尺度结构相似性(multi-scale structure similarity, MS-SSIM)。

为了综合评价本文提出的方法，我们与现有的图像压缩标准以及现有的基于深度学习的模型做了对比，例如JPEG[1]、JPEG2000[2]、BPG[3]和Belle[17]。图4和图5分别给出在Kodak数据集和Tecnick数据集上不同方法的压缩率-保真率曲线。从图上可以看出，本文提出的方法超过所有的Base line方法。

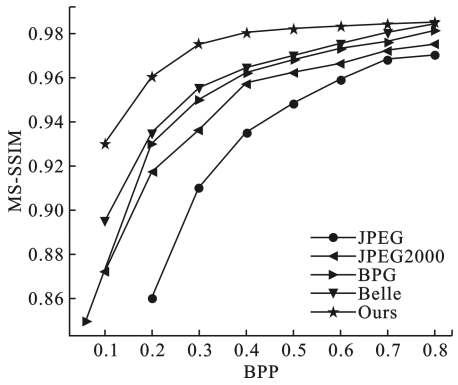

图 4 在 Kodak 数据集上不同方法的性能对比

Fig. 4 Performance comparison of different methods on Kodak dataset

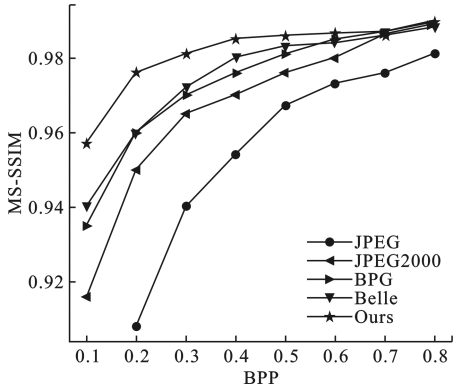

图 5 在 Tecnick 数据集上不同方法的性能对比

Fig. 5 Performance comparison of different methods on Tecnick dataset

为了测试本文所提出的方法的运行时间，我们用NVIDIA TITAN Xp显卡在Kodak数据集上测试了不同的BPP条件下的编码和解码时间，从图6可以看出，本文所提出的方法的编码和解码时间对于大小为752×496的图像在绝大多数情况下都少于1 s。

综上，图4、图5和图6的实验结果表明，本文提出的模型在保持图像质量为85%~95%的前提下图像压缩比能达到7%~15%，并且在普通的CPU上其处理速度达到0.48 秒/张。

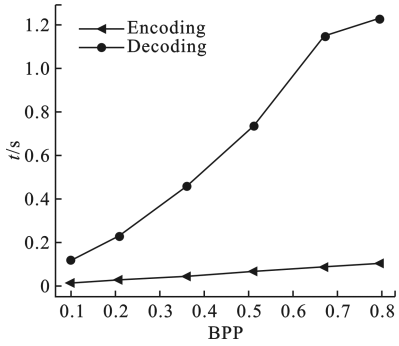

图 6. 本文提出的模型在 Kodak 数据集和 Tecnick 数据集上的平均运行时间分析

Fig. 6 The average running time analysis of the model proposed in this paper on Kodak data set and Tecnick data set

6 结　论

本文研究了基于深度卷积神经网络构建一个直接从输入图像到重建图像的端到端的图像压缩模型。由于图像不同位置的比特率是由图像的局部内容决定的，在这一思想的指导下提出了一种基于自注意力机制的内容敏感的比特率分配策略，即一种可学习的量化器。针对量化器和二值化器的离散值问题，通过引入代理函数对反向传播的二值运算进行逼近，使其具有可微性。这样传统图像压缩的编码器、量化器和解码器就可以融合在一个统一的深度学习框架之内，并可以联合起来一起优化，从而构建一个端到端的图像压缩系统。实验结果表明本文提出的方法能显著降低影像的数据大小，压缩后的图像能适用于各种平台的数据存储和传输。

参考文献(References)

[1] Wallace G K. The JPEG Still Picture Compression Standard[J]. IEEE Transactions on Consumer Electronics, 1992, 38(2): xviii-xxxiv.

[2] Christopoulos C, Skodras A, Ebrahimi T. The JPEG2000 Still Image Coding System: an Overview[J]. IEEE Transactions on Consumer Electronics, 2000, 46(11): 1103-1127.

[3] Albalawi U, Mohanty S P, Kougianos E. A Hardware Architecture for Better Portable Graphics (BPG) Compression Encoder[J]. Indore: 2015 IEEE International Symposium on Nanoelectronic and Information Systems, 2015.

[4] Rothe R, Timofte R, Van Gool L. Efficient Regression Priors for Reducing Image Compression Artifacts[C]. Quebec: IEEE International Conference on Image Processing, 2015.

[5] Dong C, Deng Y, Loy C C. Compression Artifacts Reduction by a Deep Convolutional Network[C]. New York: IEEE International Conference on Computer Vision, 2015.

[6] 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, 2018, 28(10): 3007-3018.

[7] 刘东, 王叶斐, 林建平, 等. 端到端优化的图像压缩技术进展[J]. 计算机科学, 2021, 48(3): 1-8.
Liu D, Wang Y F, Lin J P, et al. Advances in End-to-end Optimized Image Compression Technologies[J]. Computer Science, 2021, 48(3): 1-8.

[8] 曲海成, 田小容, 刘腊梅, 等. 多尺度显著区域检测图像压缩[J]. 中国图象图形学报, 2020, 25(1): 31-42.
Qu H C, Tian X R, Liu L M, et al. Image Compression Method Based on Multi-scale Saliency Region Detection[J]. Journal of Image and Graphics, 2020, 25(1): 31-42.

[9] 穆克, 李文娜. 基于模糊 C 均值聚类的医学图像压缩算法[J]. 控制工程, 2016, 23(5): 706-710.
Mu K, Li W N. The Medical Image Compression Method Based on Fuzzy C-mean Clustering[J]. Control Engineering of China, 2016, 23(5): 706-710.

[10] 郭剑, 韩崇, 施金宏, 等. 基于稀疏采样的无线多媒体传感网图像压缩算法[J]. 太原理工大学学报, 2021, 52(1): 76-82.
Guo J, Han C, Shi J H, et al. An Image Compression Algorithm Based on Sparse Sampling for Wireless Multimedia Sensor Network[J]. Journal of Taiyuan University of Technology. 2021, 52(1): 76-82.

[11] Theis L, Bethge M. Generative Image Modeling Using Spatial LSTMs[J]. arXiv eprint arXiv: 1506.0347, 2015.

[12] Oord A V, Kalchbrenner N, Kavukcuoglu K. Pixel Recurrent Neural Networks[C]. New York: 33rd International Conference on Machine Learning, 2016.

[13] Rippel O, Bourdev L. Real-time Adaptive Image Compression[C]. New York: 34th International Conference on Machine Learning, 2017.

[14] Toderici G, O'Malley S M, Hwang S J, et al. Variable Rate Image Compression with Recurrent Neural Networks[J]. San Juan: International Conference on Learning Representations, 2016.

[15] Yang X, Yumer E, Asente P, et al. Learning to Extract Semantic Structure from Documents Using Multimodal Fully Convolutional Neural Networks[C]. Nashville: IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[16] Lin C, Yao J, Chen F, et al. A Spatial RNN Codec for End-to-end Image Compression[C]. Nashville: IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.

[17] Ballé J, Laparra V, Simoncelli E P. End-to-end Optimized Image Compression[J]. arXiv preprint arXiv:1611.01704, 2016.

[18] Theis L, Shi W, Cunningham A, et al. Lossy Image Compression with Compressive Autoencoders[J]. arXiv preprint arXiv:1703.00395, 2017.

[19] Agustsson E, Mentzer F, Tschannen M, et al. Soft-to-hard Vector Quantization for End-to-end Learning Compressible Representations[J]. arXiv preprint arXiv:1704.00648, 2017.

[20] Toderici G, Vincent D, Johnston N, et al. Full Resolution Image Compression with Recurrent Neural Networks[C]. Honolulu: IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[21] Zhou S, Ni Z, Zhou X, et al. Dorefa-net: Training Low Bitwidth Convolutional Neural Networks with Low Bitwidth Gradients[J]. arXiv:1606.06160v1, 2016.

[22] Rastegari M, Ordonez V, Redmon J, et al. Xnor-net: Imagenet Classification Using Binary Convolutional Neural Networks[C]. Amsterdam: The 14th European Conference on Computer Vision, 2016.

[23] Courbariaux M, Bengio Y. Binarynet: Training Deep Neural Networks with Weights and Activations Constrained to +1 or -1[J]. arXiv: 1602.02830v3, 2016.

[24] Deng J, Dong W, Socher R, et al. Imagenet: a Large-scale Hierarchical Image Database[J]. Miami: IEEE Conference on Computer Vision and Pattern Recognition, 2009.

[25] 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.

[26] 黄迪, 刘畅. 智能决策系统的深度神经网络加速与压缩方法综述[J]. 指挥信息系统与技术, 2019, 10(2): 8-13.
Huang D, Liu C. Review of Acceleration and Compression Methods for Deep Neural Networks in Intelligent Decision Systems[J]. Command Information System and Technology, 2019, 10(2): 8-13.

[27] Asuni N, Giachetti A. Testimages: a Large Data Archive for Display and Algorithm Testing[J]. Journal of Graphics Tools, 2013, 17(4): 113-125.