

基于深度卷积神经网络的信息流增强 图像压缩方法

李志军, 杨楚哲, 刘 丹, 孙大洋

(吉林大学 通信工程学院, 长春 130012)

摘 要:在图像压缩过程中,信息利用率对图像压缩的效率起决定性作用。为了更好地提高信息的利用率,提出了一种端到端的基于深度卷积神经网络的信息流增强图像压缩方法。在编解码网络中,采用特殊的网络结构,增加了卷积层之间的前向与后向连接,与传统的前向神经网络相比,信息流的双向流动和视觉特征的循环反馈可有效实现信息流的增强,从而提高图像的压缩效率。实验表明:在相同码率下,本文算法复原图像的 MS-SSIM 分别比 JPEG、JPEG2000 和 HEVC 高 0.08、0.027 和 0.012。将基于深度神经网络的信息流增强结构用于图像压缩,可有效提高压缩效率。

关键词:图像压缩;自动编码器;卷积神经网络;信息流增强

中图分类号:TN919.81 **文献标志码:**A **文章编号:**1671-5497(2020)05-1788-08

DOI:10.13229/j.cnki.jdxbgxb20190793

Deep convolutional networks based image compression with enhancement of information flow

LI Zhi-jun, YANG Chu-xi, LIU Dan, SUN Da-yang

(College of Communication Engineering, Jilin University, Changchun 130012, China)

Abstract: In the image compression process, the compression efficiency is largely affected by the information utilization. In order to improve the utilization of information, in this paper, we propose an end-to-end image compression method with information enhancement based on deep convolutional networks. In the encoding and decoding networks, a special network structure is adopted, which can both increase the forward and backward connections between the convolutional layers. As a result, the bidirectional information flow and the cyclic feedback of visual features are realized, which can effectively reduce the number of bits required for image compression. Experimental results show that, at the same compression ratio, the multiscale structural similarity index (MS-SSIM) of the reconstructed image of the proposed method can gain 0.08, 0.027 and 0.012 higher than JPEG, JPEG2000 and HEVC, respectively. The information enhancement structure based on deep convolutional network can effectively improve the coding efficiency when it is used for image compression.

收稿日期:2019-08-08.

基金项目:国家自然科学基金面上项目(61671219).

作者简介:李志军(1971-),男,高级工程师,硕士.研究方向:无线通信与图像编码.E-mail:zhijun@jlu.edu.cn

通信作者:孙大洋(1979-),男,副教授,博士.研究方向:人工智能与计算机视觉应用.E-mail:dysun@jlu.edu.cn

Key words: image compression; auto encoder; convolution networks; information enhancement

0 引言

随着多媒体大数据和互联网+时代的到来,互联网中各种数据资源呈现爆炸式增长,给网络的存储和传输都带来了很大压力。而图像数据作为信息的一种重要载体,在网络总数据中所占的比例已超过80%,因此,如何在节省存储空间的同时,尽可能多地保留图像中的有用信息就成为了目前亟待解决的问题,这就需要利用图像压缩方法对图像进行压缩编码,减少图像所需要的码流。最初图像压缩的提出是为了解决数字电视中数据量太大导致的存储和带宽不足的问题,后来因为数字图像的大量使用,图像压缩一直是图像处理领域的研究热点^[1-3]。早期的图像压缩算法主要是通过信息熵保持的编码方法对图像数据进行无损的熵编码,这种方法虽然可以获得无损的复原图像,但能达到的压缩比非常有限;随后,逐渐采用视觉保持的有损压缩算法,目的是在保持视觉效果不受影响的前提下,获得高效的压缩,从而有效提高压缩比,有损压缩也是目前图像压缩研究的主流。图像压缩过程大体可分为编码和解码两个阶段,对于编码阶段,又可细分为变换、量化、熵编码3个主要过程,变换主要完成图像从空间像素域到变换域的转换,从而有效去除数据相关性;量化是一个有损的过程,将数据进行有限位的量化表示,从而减少编码所需码流;熵编码则是对数据进行二进制表示,获得二进制的压缩码流。相对的,解码过程包括解码、反量化、逆变换3个过程,可看作是编码的逆过程,解码后最终输出重构图像。

现有图像压缩标准主要有 JPEG^[4]、JPEG2000^[5]和 H. 265/HEVC 帧内编码^[6]。对于基于图像压缩标准的传统图像压缩方法来说,编解码的每个过程都是固定的,虽然 HEVC 采用了多模态的预测模式,但每种预测模式也是固定的,这就大大限制了压缩方法的去冗余能力。为解决这一问题,基于深度神经网络的图像压缩在 2017 年被提出,在深度神经网络中,可直接根据复原图像的效果对编码器进行优化,大大增加了压缩效率。

在基于深度神经网络的图像压缩方法中,会通过神经元或卷积层进行输入图像的特征提取,首先提取到的是低层特征,随着网络的加深,逐渐获得图像的高层特征。然而,在图像压缩中,不同层级间的特征通常也具有很大的相关性,为了更好地利用不同层级之间的相关性进行压缩,本文提出了一种基于信息流增强的深度图像压缩方法,在网络中增加由高层特征到低层特征的循环反馈,使不同的卷积层之间连接更加紧密,提高编解码过程的信息利用率。同时,为进一步提高压缩效率,采用基于条件概率的神经网络模型进行码字的熵编码,从而通过联合训练的方式对整个压缩系统进行率失真优化。

1 基于 CliqueNet 的信息增强模块

对于深度卷积神经网络来说,网络深度越深,就可以获得更高层的图像特征,然而随着网络深度的加深,在训练过程中的梯度消失的问题也逐渐凸显。为解决这一问题,He 等^[7]提出在卷积层间引入残差连接,在残差连接中,由于输入和输出是直接相加的,因此反向传播中,梯度就可以跳过残差块的内部,从而保留足够的梯度信息。但是,在残差网络中,很多层级的特征图都存在着冗余,使模型的参数利用率不高,为解决这一问题,Huang 等^[8]提出 DenseNet,将每层的输出都作为后面所有层的输入,并将所有层级输出的特征图按深度进行拼接,这样就确保了每一层都能直接建立关于损失函数的梯度,即使在网络很深的情況下也能有效的进行训练。然而,随着网络深度的增加,DenseNet 中由于其层级连接的密集性和复杂性,网络参数的数量会大幅度的增长,使计算复杂度和时间复杂度都大大增加。为优化这一问题,CliqueNet^[9]结构被提出,在 CliqueNet 中,并不是将所有层级的网络都连接起来,而是将 T 个卷积层划分为一个 Clique 模块,只有在同一个 Clique 模块中的卷积层之间才建立紧密连接。与 DenseNet 相比,由于 Clique 模块中的卷积层数量 T 远远小于整个网络的卷积层数,因此在网络深度加深时,参数增加的数量会大大减少,从而提升了算法的效率;更重要的是,在 Clique 模块中,网

络层级之间不仅有前向的链接,还有从高层特征到低层特征的后向连接,这种任意两层之间的前向和后向连接可有效提高网络中信息的利用效率。

Clique 模块示意图如图 1 所示,整个模块可分为两个阶段:①与 DenseNet 的前馈过程相同,主要是将所有前向卷积层的输出特征作为当前层的输入特征,并进行模块中各层的特征前向提取;②利用网络中其他卷积层,同时通过前向和后向连接对当前层进行特征精炼的过程,并且这一过程可以进行多次循环精炼。

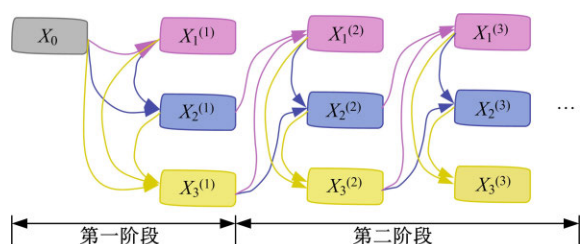


图 1 Clique 模块示意图

Fig. 1 Illustration of Clique block

整个 Clique 模块的输入用 X_0 表示,在第一阶段,每层的输出特征都会合并之前所有层的特征,然后利用合并得到的特征图获得新一层的特征。第一阶段各层获得的特征用 $X_1^{(1)}, X_2^{(1)}, \dots, X_T^{(1)}$ 表示, T 为 Clique 模块中的网络层数,整个过程可表示为:

$$X_i^{(1)} = \sigma [w_{0i} * X_0 + \sum_{l < i} (w_{li} * X_l^{(1)})] \quad (1)$$

式中: $X_i^{(1)}$ 的下标 i 为网络层数的索引,上标 (1) 为第一阶段; w_{ij} 为第 i 层到第 j 层的权重系数; σ 为激活函数。

在第二阶段,当获取本阶段第 i 层的特征图时,需要合并模块中除了第 i 层之外的所有特征。其中,对于第 i 的所有前向卷积层来说,由于其第二阶段的特征已经获取,因此使用第二阶段的特征进行合并;而对于第 i 层的所有后向卷积层来

说,由于其第二阶段的特征还未获取,因此使用第一阶段的特征进行合并。第二阶段各层获得的特征用 $X_1^{(2)}, X_2^{(2)}, \dots, X_T^{(2)}$ 表示,整个过程可表示为:

$$X_i^{(2)} = \sigma [\sum_{l < i} w_{li} * X_l^{(2)} + \sum_{m > i} w_{mi} * X_m^{(1)}] \quad (2)$$

并且,第二阶段可以循环进行多次,第 k 次循环过程可表示为:

$$X_i^{(k)} = \sigma [\sum_{l < i} w_{li} * X_l^{(k)} + \sum_{m > i} w_{mi} * X_m^{(k-1)}] \quad (3)$$

最终 Clique 模块的输出特征为最后一次循环精炼获得的 T 层特征的合并,即,当循环次数为 q 次时, Clique 模块的输出可表示为:

$$Y = \{X_1^{(q)}, X_2^{(q)}, \dots, X_T^{(q)}\} \quad (4)$$

式中: Y 为 Clique 模块的输出特征;“ $\{\cdot\}$ ”为在深度维度上的拼接操作。

由于在 Clique 模块中,不同卷积层之间的连接非常紧密,这一特征非常适合应用在图像压缩网络中。根据这一特点,以 Clique 为基础,提出了一种图像压缩网络中的信息增强模块 (Information enhancement, IE 模块),整个模块的特征获取过程如表 1 所示。

根据表 1 可以看出,在一个 IE 模块中,每两个卷积层之间都同时具有正向和反向的权重系数连接,在训练过程中,每一阶段的卷积层权重系数都是共享的,因此实现了利用高层特征对低层特征的反向优化。每个信息增强模块中的卷积层数设为 $T=4$,在训练过程中,为减少算法复杂度,在第二阶段,只进行一次循环精炼,并将第二阶段所有卷积层的特征图输出与总输入特征进行深度维度的合并,作为 IE 模块的总输出。由于在信息输出时合并了网络的输入特征,因此使用 IE 模块可有效避免梯度消失的发生。另外,与传统“单线型”网络相比,在 IE 模块中,网络的输出端合并了所有前向卷积层的信息,因此该模块还可以实现信息流的增强,以达到提高信息利用率的目的。

表 1 信息增强模块的特征获取过程

Table 1 Process of feature extraction of information enhancement block

总输入	第一阶段			第二阶段			总输出
	输入特征	权重系数	输出	输入特征	权重系数	输出	
X_0	X_0	w_{01}	$X_1^{(1)}$	$\{X_2^{(1)}, X_3^{(1)}, X_4^{(1)}\}$	$\{w_{21}, w_{31}, w_{41}\}$	$X_1^{(2)}$	$\{X_0, X_1^{(2)}, X_2^{(2)}, X_3^{(2)}, X_4^{(2)}\}$
	$\{X_0, X_1^{(1)}\}$	$\{w_{02}, w_{12}\}$	$X_2^{(1)}$	$\{X_3^{(1)}, X_4^{(1)}, X_1^{(2)}\}$	$\{w_{32}, w_{42}, w_{12}\}$	$X_2^{(2)}$	
	$\{X_0, X_1^{(1)}, X_2^{(1)}\}$	$\{w_{03}, w_{13}, w_{23}\}$	$X_3^{(1)}$	$\{X_4^{(1)}, X_1^{(2)}, X_2^{(2)}\}$	$\{w_{43}, w_{13}, w_{23}\}$	$X_3^{(2)}$	
	$\{X_0, X_1^{(1)}, X_2^{(1)}, X_3^{(1)}\}$	$\{w_{04}, w_{14}, w_{24}, w_{34}\}$	$X_4^{(1)}$	$\{X_1^{(2)}, X_2^{(2)}, X_3^{(2)}\}$	$\{w_{14}, w_{24}, w_{34}\}$	$X_4^{(2)}$	

2 基于深度神经网络的信息增强图像压缩

2.1 概述

利用深度神经网络进行图像压缩,整个网络主要分为自动编码器、量化和熵编码模型3个部分,图像压缩网络的训练框架如图2所示,其中每个模块均是由全卷积神经网络组成:编码器和解码器共同构成了自动编码器的结构,用以实现图像的恒等变换;熵编码模型由三维卷积神经网络组成,用以进行码率估计。

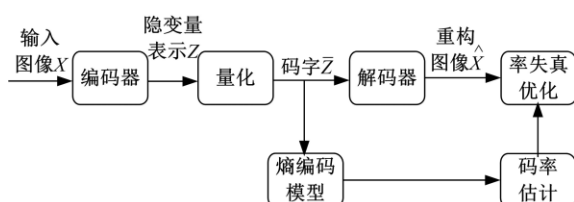


图2 基于深度神经网络的图像压缩框架

Fig. 2 Framework of deep neural network based image compression

如图2所示,自动编码器由编码器和解码器两部分组成,输入图像 X 经过编码器,将图像由空间像素域变换到隐变量空间,获得隐变量空间的特征表示 Z 。在图像压缩网络中, Z 的维度通常比 X 的维度小,这样就达到了数据压缩的目的;在获得隐空间的特征表示后,通过量化过程减少每个码字位所需的字节,量化后获得的输出即为码字 \bar{Z} ;然后通过解码器将码字解码,获得重构图像 \hat{X} 。对于自动编码器来说,其训练目标是使重构图像的失真最小,因此有:

$$L_{AE} = \min [d(\hat{X}, X)] \quad (5)$$

式中: L_{AE} 为自动编码器的损失函数; $d(\hat{X}, X)$ 为重构图像与输入图像相比产生的失真。

引入熵编码模型的目的是为了更准确地估计压缩后的码流大小,通过合适的概率模型对码字进行先验概率估计,并利用信息熵的思想计算压缩后的码流大小。对于熵编码模型来说,其训练目标是使压缩后的码流尽量小,表示为:

$$L_P = \min [r(\bar{Z})] \quad (6)$$

式中: L_P 为熵编码模型的损失函数; $r(\bar{Z})$ 为压缩码字所需的码率大小。

由于图像压缩需要在保证图像复原质量的同时尽可能地减少压缩后的码率,因此采用联合训练

的方式对自动编码器和熵编码模型进行训练,训练过程总的损失函数可表示为:

$$L = L_{AE} + L_P \quad (7)$$

2.2 基于信息增强模块的自动编码器

在自动编码器中,编码器和解码器通常由多个卷积层构成。在编码器端,输入图像经过一系列卷积层之后,相当于利用滤波器对图像进行特征提取,由于卷积层的步长不同,在编码过程中可提取到多尺度的特征。经过编码器之后,获得的是隐变量空间的特征表示,对于图像压缩来说,隐变量空间的特征要足够紧凑,才能同时满足数据维度的降低和尽量少的信息损失。自动编码器采用对称的网络结构,通过解码器逆向获得重构图像,在训练时只需比较重构图像与输入图像之间的失真,这样整个网络就可实现无监督训练过程。为更有效地利用卷积层之间的信息,获得更紧凑的码字,使用信息增强模块替代一部分卷积层,并且在多个信息增强模块之间使用残差网络的结构,形成残差-信息增强模块。编码器的网络结构如图3(a)所示,其中,每个信息增强模块(IE模块)都是一个Clique模块,其具体的网络结构如图3(b)所示。编码器的输入为待编码图像 X ,编码器的输出为图像的隐变量表示 Z 。

如图3所示,编码器主要由3个卷积层和5个残差-信息增强模块组成。使用卷积层的目的主要是通过不同步长的卷积获得多尺度的特征表示;而残差-信息增强模块的作用主要是为了在编码过程中提高信息流的使用效率。卷积层示意图的中括号中的数字表示使用的卷积核大小,后面跟随的数字是卷积核的个数,步长代表卷积操作时采用的步长大小。IE模块的结构是固定的,首先通过一个初始化的卷积层,将通道数由128降为64,这样做的目的主要有两个:①降低IE模块所需要的参数数量,减少训练过程中的内存消耗;②只有降低一定数量的通道数,才能在模块输出端,使深度维度拼接后的通道数仍然等于128,保证残差结构的实现。初始化后,经过表1所示的两个阶段完成4层特征的提取和精炼,最后将所有层的特征图进行深度维度上的拼接,获得通道数为128的IE模块输出特征。然后,与残差网络相同,将IE模块的输出与输入进行相加,形成一个完整的残差-信息增强模块。

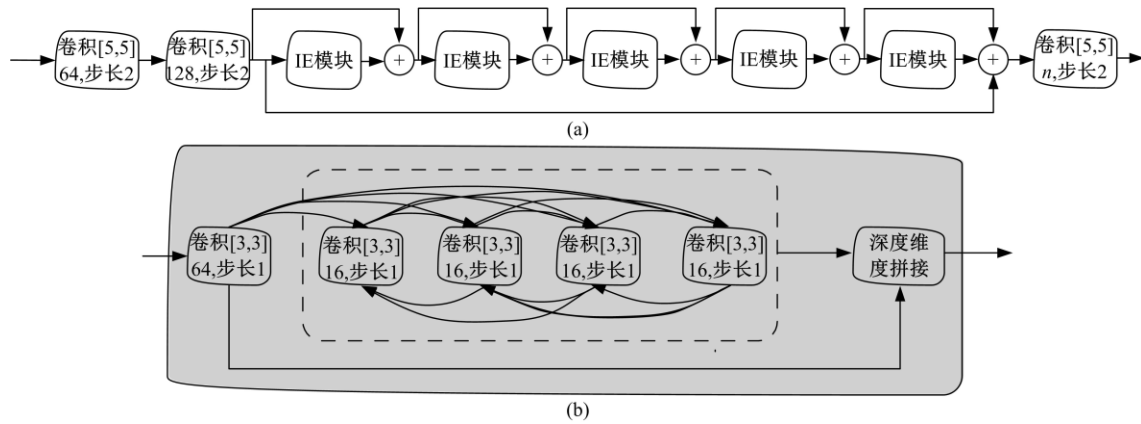


图3 编码器网络

Fig. 3 Encoder network

在编码器的最后一层,卷积核的数量为 n 个,其取值应该随着压缩后码率的不同而变化,在码率较大时, n 的取值也应该比较大,反之码率较小时, n 的取值也应该较小。在本文算法中,将 n 固定设置为30,另外,为了更好地进行码率控制,在编码器的末端除了获取码字之外,还通过另一个支路获得单通道的码字重要图^[10],其作用是根据目标码率选择性的对码字进行保留(对目标码率的说明详见2.3节)。在训练一个模型时,会根据所需的码率,通过重要图的取值自适应地保留所需码字。解码器与编码器的结构是对应的,由3个转置卷积层和5个信息增强模块组成,由于其原理与编码器基本一致,在此不再赘述,解码器的输出是重构图像。

2.3 量化与熵编码

在获得编码器的输出码字之后,需要通过量化过程减少每个码字位所需的比特数。本文采用非均匀量化的方式对码字进行量化,量化中心的数量为6个,所有量化中心均通过训练获得,量化过程可表示为:

$$\bar{z}_i = \arg \min_{q_j} \|z_i - q_j\|_2 \quad (8)$$

式中: \bar{z}_i 的下标 i 为码字索引; q_j 为第 j 个量化中心。

由于量化过程是不可逆的,因此在基于梯度下降的训练过程中,会导致反向传播无法进行。为解决这一问题,在反向传播过程中,采用可微近似的方式替代式(8)的量化过程,在前向传播中,量化方式保持不变。可微近似的量化方式表示为:

$$\hat{z}_i = \sum_{j=1}^6 \frac{\exp(-\|z_i - q_k\|_2)}{\sum_{k=1}^6 \exp(-\|z_i - q_k\|_2)} \times q_j \quad (9)$$

在量化之后,为了进行率失真优化,需要计算压缩后的码率大小。本文在该压缩系统中采用基于条件概率的神经网络模型^[11]作为熵编码模型,对码字进行熵编码。首先,将量化后的码字 \bar{z} 输入到熵编码模型中,基于条件概率的信息熵可表示为:

$$H(\bar{z}) = E_{\bar{z} \sim p(\bar{z})} \{ -\log [p(\bar{z})] \} = E_{\bar{z} \sim p(\bar{z})} \left\{ \sum_{i=1}^m -\log [p(\bar{z}_i | \bar{z}_{i-1}, \bar{z}_{i-2}, \dots, \bar{z}_1)] \right\} \quad (10)$$

式中: $H(\bar{z})$ 为 \bar{z} 的信息熵,其大小可以表示压缩后的码流大小; E 为期望; $p(\bar{z})$ 为 \bar{z} 的概率分布; m 为码字数量。

在概率模型中,利用基于卷积神经网络的条件概率估计 \bar{z} 的概率分布,设估计的概率表示为 $\hat{p}(\bar{z})$,有:

$$E_{\bar{z} \sim p(\bar{z})} \left\{ \sum_{i=1}^m -\log [\hat{p}(\bar{z}_i | \bar{z}_{i-1}, \bar{z}_{i-2}, \dots, \bar{z}_1)] \right\} \approx E_{\bar{z} \sim p(\bar{z})} \left\{ \sum_{i=1}^m -\log [p(\bar{z}_i | \bar{z}_{i-1}, \bar{z}_{i-2}, \dots, \bar{z}_1)] \right\} \quad (11)$$

在训练过程中,希望能达到更准确的概率估计,即希望 $\hat{p}(\bar{z})$ 与 $p(\bar{z})$ 越接近越好。因此,采用极大似然估计法,利用交叉熵损失函数最小化 $\hat{p}(\bar{z})$ 与 $p(\bar{z})$ 分布之间的距离,损失函数可表示为:

$$L_p = E_{\bar{Z} \sim p(\bar{Z})} \{-p(\bar{Z}) \log [\hat{p}(\bar{Z})]\} = E_{\bar{Z} \sim p(\bar{Z})} \{-\log [\hat{p}(\bar{Z})]\} \approx H(\bar{Z}) \quad (12)$$

由式(12)可以看出,在利用极大似然估计法对概率模型进行优化时,相当于间接地优化了码字的信息熵,即估计概率越接近真实概率,压缩后的码率就越小,因此可以利用该码率进行整个压缩系统的率失真优化。

熵编码网络采用轻量级的三维卷积网络实现^[11],卷积核大小为 $5 \times 5 \times 5$ (卷积核不仅在宽,高维度滑动,还在深度维度滑动),每层的卷积核个数都为30个,与编码器最后一层的卷积核数量相对应。在熵编码模型的最后一层,利用softmax函数进行归一化,输出码字在条件概率模型下的概率分布 $\hat{p}(\bar{Z})$,并利用此概率分布根据式(12)进行概率模型优化。概率模型和自动编码器的优化是联合进行的,根据式(5)(7)(12)总的损失函数为:

$$L = L_{AE} + L_p = \min [d(\bar{X}, X)] + E_{\bar{Z} \sim p(\bar{Z})} \{-\log [\hat{p}(\bar{Z})]\} \quad (13)$$

在实际训练过程中,使用的损失函数为:

$$L = L_{AE} + \max(L_p - R_t, 0) \quad (14)$$

式中: R_t 为人为设定的目标码率。

在使用式(14)作为损失函数进行训练的过程中,只有当压缩码率大于目标码率 R_t 时,才会对码率进行优化,而当压缩码率小于目标码率 R_t 时,网络只会对图像的重构质量进行优化,最终会使压缩码率在 R_t 附近达到平衡。因此,通过改变目标码率 R_t 的数值,就可训练出多个码率下的图像压缩模型。

3 实验结果与比较

3.1 训练设置及初始化

首先将输入图像的灰度值除以255,将输入数据范围归一化到 $[0, 1]$ 之间,网络输入的图块大小为 160×160 的三通道图像,利用随机裁取的方式从图像中裁取该大小的图块,并进行随机翻转。训练采用批训练(Minibatch)的方式进行,批大小设置为32;网络中卷积层和信息增强模块中采用的激活函数均为Relu函数,并且,在卷积层中加入了批归一化(Batch normalization)过程。

对于量化过程,6个量化中心的初始化值设置为 $[-2, 2]$ 之间的均匀分布。网络训练集由ImageNet中^[12]选取的40万幅图像组成,测试集为Kodak photo CD数据集^[13]。

训练的初始学习率设置为 1×10^{-4} ,每经过2轮训练,学习率就变为之前的0.1倍。整个训练过程将经过8轮,并利用NVIDIA GTX1070进行GPU加速,每个模型的训练过程大约耗时26 h。本文实验中共训练了3个压缩模型,其压缩后的码率分别为0.3、0.5、0.7比特·像素⁻¹。

3.2 评价指标及算法对比

训练过程中,用MS-SSIM^[14]作为图像质量的量化指标,代入损失函数进行计算。MS-SSIM是一种应用广泛的图像质量评价指标,能很好地反映人眼对图像的观测。当模型训练完毕后,利用不同压缩码率模型对图像进行压缩,压缩完毕后,以MS-SSIM指标为评价标准,将获得的重构图像与相同码率下JPEG、JPEG2000和HEVC帧内编码的重构图像进行对比。为了更好地衡量本文算法的性能,同样对比了主流的基于深度神经网络的图像压缩算法^[15]在Kodak测试集上的压缩结果。率失真曲线的对比结果如图4所示。

由图4可以看出,与传统压缩方法相比,在低码率时,本文算法的压缩性能明显优于JPEG、JPEG2000以及HEVC帧内编码,相同压缩比下本文算法重构图像的MS-SSIM最高比JPEG高

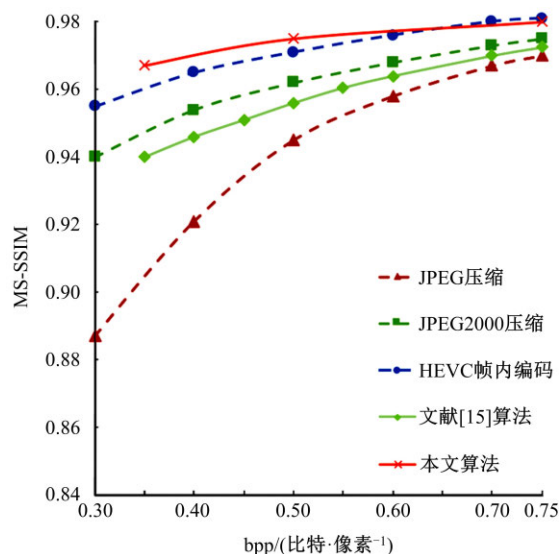


图4 率失真曲线对比

Fig. 4 Comparison of R-D curves

0.08, 比 JPEG2000 和 HEVC 分别高达 0.027 和 0.012; 在高码率下, 相同压缩比下, 本文算法的性能与 HEVC 帧内编码基本持平, 但始终优于 JPEG 和 JPEG2000。与基于深度神经网络的图像压缩方法^[15]相比, 由于本文算法在编解码网络中引入了残差-信息增强模块, 使压缩性能得到了有效提高, 相同码率下, 本文算法复原图像的

MS-SSIM 最高比文献[15]复原图像的 MS-SSIM 高 0.027, 验证了残差-信息增强模块的有效性。

复原图像的主观细节对比结果如图 5 所示, 可见, 本文算法可实现更好的图像细节复原, 并且边缘更加锐利, 即使在码率较低的情况下, 复原图像仍然可获得很好的主观效果。



图 5 复原图像局部细节对比

Fig. 5 Detail comparison between reconstructed images

4 结束语

提出了一种基于深度神经网络的图像压缩方法, 在神经网络中, 创新性地采用残差-信息增强模块进行卷积层之间的循环精炼, 有效提高了图像压缩过程中信息的利用率; 量化过程采用训练获得的量化中心对码字进行非均匀量化, 并利用条件熵模型对码字进行先验概率建模, 完成码率估计; 最后, 利用率失真优化的方法对整个网络进行联合训练。实验结果表明: 无论是在客观指标方面还是主观视觉效果方面, 本文算法都表现出了优越的性能。在相同的压缩比下, 本文算法的复原图像的 MS-SSIM 可分别超越 JPEG、JPEG2000 和 HEVC 帧内编码 0.08、0.027 和 0.012。与基于深度神经网络的压缩方法相比, 由于通过信息增强模块提高了信息利用效率, 因此相同码率下, 本文算法的复原图像能达到更高的 MS-SSIM。

参考文献:

- [1] 刘丽丽, 陈贺新, 桑爱军, 等. 基于多维矢量矩阵正交变换理论的彩色图像压缩编码[J]. 吉林大学学报: 工学版, 2011, 41(6): 1754-1759.
Liu Li-li, Chen He-xin, Sang Ai-jun, et al. Color image coding based on multidimensional vector matrix discrete cosine transform theory[J]. Journal of Jilin University(Engineering and Technology Edition), 2011, 41(6): 1754-1759.
- [2] 桑爱军, 杨树媛, 赵欣. 基于多维矢量矩阵离散余弦变换的熵编码[J]. 吉林大学学报: 工学版, 2011, 41(增刊 1): 319-324.
Sang Ai-jun, Yang Shu-yuan, Zhao Xin. Entropy code based on multidimensional vector matrix DCT [J]. Journal of Jilin University(Engineering and Technology Edition), 2011, 41(Sup. 1): 319-324.
- [3] 赵岩, 李丽, 王世刚. 结合成像几何特征的立体元图像阵列编码[J]. 光学精密工程, 2018, 26(12): 3060-3066.
Zhao Yan, Li Li, Wang Shi-gang. Elemental image

- array coding combining imaging geometry features[J]. Optics and Precision Engineering, 2018, 26(12): 3060–3066.
- [4] Wallace G K. The JPEG still picture compression standard[J]. IEEE Transactions on Consumer Electronics, 1992, 38(11): 1–17.
- [5] Skodras A, Christopoulos C, Ebrahimi T. The JPEG 2000 still image compression standard[J]. IEEE Signal Processing Magazine, 2001, 18(5): 36–58.
- [6] Sullivan G, Ohm J, Han W, 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.
- [7] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Las Vegas, USA, 2016: 770–778.
- [8] Huang G, Liu Z, Maaten L V D, et al. Densely connected convolutional networks[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017: 2261–2269.
- [9] Yang Y, Zhong Z, Shen T, et al. Convolutional neural networks with alternately updated clique[C]// IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake, 2018: 2413–2422.
- [10] Li M, Zuo W, Gu S, et al. Learning convolutional networks for content-weighted image compression [C]//IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake, 2018: 3214–3223.
- [11] Mentzer F, Agustsson E, Tschannen M, et al. Conditional probability models for deep image compression[C]//IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake, 2018: 4394–4402.
- [12] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database[C] //IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA, 2009: 248–255.
- [13] Kodak Photo CD dataset [EB/OL]. (1999–11–15). [2013–01–27]. <http://r0k.us/graphics/kodak>
- [14] Wang Z, Simoncelli E P, Bovik A C. Multiscale structural similarity for image quality assessment[C]// Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers, Pacific Grove, USA, 2003: 1398–1402.
- [15] Theis L, Shi W, Cunningham A, et al. Lossy image compression with compressive autoencoders[C] // ICLR2017, Toulon, 2017: 1–19.