

# 极低比特率图像压缩技术综述

岳爽, 陈喆, 殷福亮

(大连理工大学信息与通信工程学院, 大连 116024)

**摘要:** 图像是人们获取信息的重要途径之一。随着图像传输与存储需求的不断增加,尤其在带宽受限或云存储情况下,对图像进行极低比特率压缩,对于提高传输效率和节省存储空间具有重要意义。基于此,本文对有损图像极低比特率压缩技术进行了系统综述。首先,在基于生成对抗网络(Generative adversarial networks, GAN)的图像压缩衍生算法在高分辨率图像压缩、生成图像模糊、忽视语义信息与纹理信息等方面问题的基础上,介绍了最新的极低比特率图像压缩方法。然后,阐述了分层压缩、基于对象和感兴趣区域等其他非GAN模型的极低比特率图像压缩方法。接着,描述了常用数据集及有损压缩条件下的图像质量评价方法。最后,对极低比特率有损图像压缩技术做出总结,并对其后续的发展进行了展望。

**关键词:** 图像压缩;极低比特率;图像有损编码;深度学习;生成对抗网络

**中图分类号:** TP391 **文献标志码:** A

## Review of Very Low Bitrate Image Compression Techniques

YUE Shuang, CHEN Zhe, YIN Fuliang

(School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China)

**Abstract:** Image is one of the important ways to obtain information. With the increasing demand of image transmission and storage, especially in bandwidth limited or cloud storage situations, compressing images at extremely low bitrates is of great significance for improving transmission efficiency and saving storage space. Based on this, this paper presents a systematic review of very low bitrate compression techniques for lossy images. Firstly, on the basis of problems of image compression derivative algorithms based on generative adversarial network (GAN) in terms of high-resolution image compression, generating image blur, and neglecting semantic and texture information, the latest very low bitrate image compression methods are introduced. Then, this paper elaborates image compression methods that achieve very low bitrate using non-GAN models such as layered compression, object based, and region of interest. After that, the commonly used datasets and image quality evaluation methods under lossy compression conditions are described. Finally, a summary of very low bitrate lossy image compression techniques are made, and an outlook on their subsequent development is given.

**Key words:** image compression; very low bitrate; image lossy encoding; deep learning; generative adversarial network(GAN)

# 引言

图像信息是人类交流的重要媒介。通常,图像的数据量很大,在图像存储或传输时,都要对图像进行压缩。根据压缩过程中损失信息与否,图像压缩技术可分为无损压缩与有损压缩<sup>[1]</sup>。图像无损压缩可以完全恢复原始图像,但压缩倍数通常较低。图像有损压缩则在压缩过程中丢失一些图像细节和不可察觉的信息,以减小图像文件的大小<sup>[2]</sup>,压缩倍数通常较高,但由于其不可逆性,原始图像无法完全恢复。

在图像压缩中,图像每像素位数(Bits per pixel, BPP)表示压缩后每个像素所需的平均比特数,它是衡量图像压缩性能的重要指标,与图像的质量密切相关<sup>[3]</sup>。通常,BPP越高,表示可以获得更好的图像质量,同时也会占用更多的存储空间或带宽<sup>[4]</sup>。通过提高编码算法的压缩性能和减少图像的冗余信息来降低比特率,可以实现更低的BPP<sup>[5]</sup>。通常,原始RGB彩色图像的BPP为24 bit。图像经过大幅度压缩后,使得 $BPP < 0.1$ ,此时则称为极低比特率压缩<sup>[6]</sup>。

传统的图像压缩编解码器研究已取得重要进展,形成了JPEG<sup>[7]</sup>、JPEG2000<sup>[8]</sup>、BPG<sup>[9]</sup>、HEVC<sup>[10]</sup>和VVC<sup>[11]</sup>等标准,并在互联网传输、数码摄影及医学影像等领域得到广泛应用。但是,传统的图像压缩编解码器难以适应不同的图像内容,在极低比特率下难以生成高质量的重建图像。此外,在传统的图像压缩算法研究中,通常采用峰值信噪比(Peak signal to noise ratio, PSNR)<sup>[12]</sup>和多尺度结构相似性(Multi-scale structural similarity, MS-SSIM)<sup>[13]</sup>等指标来反映图像重构质量。

近年来,基于学习的有损图像压缩技术得到广泛关注。基于学习的有损图像压缩方法应用率失真优化准则,采用神经网络进行建模,其性能明显优于传统图像编解码器。然而,目前这些深度学习方法主要针对中高比特率( $> 0.1$  BPP)下的率失真性能进行优化,压缩倍数较低,无法满足实际应用要求。例如,在带宽受限的情况下(如无民用通信设施的野外偏远地区通信、复杂电磁环境下的战场通信、卫星通信、航空航天及深海和深空通信等极端环境),需要对图像进行极低比特率的压缩。生成对抗网络(Generative adversarial network, GAN)模型<sup>[14]</sup>能够生成高质量图像且对数据分布敏感,是一种在极低比特率下能实现图像高效压缩的解决方案<sup>[15]</sup>。由于传统的PSNR和MS-SSIM等像素级测量方法不能有效地反映极低比特率下图像的重构质量<sup>[12]</sup>,基于学习的图像极低比特率压缩方法则采用更符合人类视觉感知特性的图像质量客观评价准则,如感知相似度度量(Learned perceptual image patch similarity, LPIPS)<sup>[16]</sup>和Fréchet起始距离(Fréchet inception distance, FID)<sup>[17]</sup>等。

基于深度学习网络的图像压缩研究方法虽然已经取得了一定进展<sup>[18-19]</sup>,但在极低比特率图像压缩上性能有限。与此同时,生成对抗网络发展迅速,其主要包括生成器和判别器的对抗训练机制,通过最小最大化损失函数生成逼真数据样本,它从结构上更适合“压缩-解压缩”过程的实现,且基于GAN网络的极低比特率图像压缩技术取得了重要进展。本文旨在综述图像的有损压缩技术,总结了基于GAN模型和非GAN模型的极低比特率( $< 0.1$  BPP)图像压缩最新方法。此外,本文阐述了常用的图像压缩数据集和图像质量客观评价指标,并对极低比特率图像压缩技术的未来发展进行了讨论与展望。为了更清晰地呈现本文的内容框架,可以参考图1所示的脉络图。

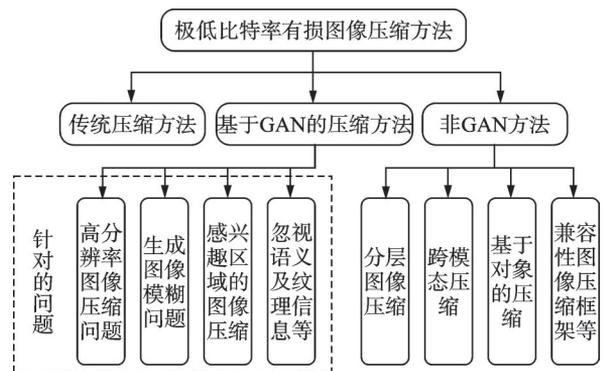


图1 本文脉络结构图

Fig.1 Structure diagram of this paper

## 1 传统图像压缩编码方法

传统图像压缩多用离散余弦变换(Discrete cosine transform, DCT)和离散小波变换(Discrete wavelet transform, DWT)去除冗余信息<sup>[18]</sup>。在压缩倍数较高时,DCT易出现方块效应;相比之下,DWT有出色的多分辨分析能力,能够有效描述图像中的概貌和轮廓信息。为此,Shapiro<sup>[20]</sup>基于DWT变换,提出了嵌入式零树编码器(Embedded zero-tree wavelet, EZW),该方法的符号冗余较大,能实现灰度图像的压缩。进一步,Brahimi等<sup>[21]</sup>提出了一种改进的EZW算法,名为IMP1EZW,减少了零树数量和符号冗余,并利用颜色组件的相互依赖性来实现彩色图像编码,在灰度和彩色图像上的低比特率压缩都有很好的效果。Said等<sup>[22]</sup>基于EZW技术,提出多级树集合分裂(Set partitioning in hierarchical trees, SPIHT)算法,但存在内存需求量大和重构图像纹理细节信息丢失的问题。为此,王亚鸽等<sup>[23]</sup>提出了一种联合Canny边缘检测和SPIHT的图像压缩算法,该算法利用Canny边缘检测提取图像的纹理边缘部分,然后通过Huffman编解码得到边缘重构图,再将其与SPIHT结合,以恢复原图像。该方法能有效地保留重构图像的高频信息,但高频信息冗余也削弱了重构图像的视觉效果。Chavan等<sup>[24]</sup>采用矢量量化技术,将带有DCT的自适应增量LBG<sup>[25]</sup>与花授粉自适应优化方法(Flower pollination adapted rider optimization algorithm, FPARA)混合来优化码本,即先将图像划分为像素块(即矢量),然后利用DCT存储图像数据,而后为每个矢量选择合适的编码词,通过编码器将编码本映射到码字来进行矢量压缩。FPARA方法使原始图像和解压缩图像之间的误差达到最小,LBG-DCT+FPARA方法有较好的压缩效果。

以上几种算法的压缩性能如表1所示,其中Brahimi等提出的IMP1EZW方法<sup>[21]</sup>在彩色图像压缩方面实现了较低的BPP,且重构质量较好。

表1 较低比特率的传统图像压缩算法性能

Table 1 Compression effect of traditional image compression algorithms with lower bitrate

方法	BPP ↓	重构质量
IMP1EZW <sup>[21]</sup>	0.015 625~0.062 5(彩色图像)	中
	0.062 5~0.5(彩色图像)	优
Canny+SPIHT <sup>[23]</sup>	0.062 5~0.125(灰度图像)	良
LBG-DCT+FPARA <sup>[24]</sup>	0.041~1.161(灰度图像)	良

注: ↓表示越小越好。

## 2 极低比特率图像压缩算法

### 2.1 基于GAN的图像压缩算法

GAN网络是一种无需大量注释训练数据就能学习深度表征的深度学习<sup>[26]</sup>方法,它由生成网络和判别网络组成,基本架构如图2所示。其中,生成网络负责生成模拟数据,判别网络负责区分真实数据和生成网络生成的模拟数据<sup>[27]</sup>。在训练过程中,生成网络通过不断优化自身来生成更加逼真的模拟数据,使得判别网络无法区分真实数据和模拟数据;同时,判别网络也通过不断优化自身使其判

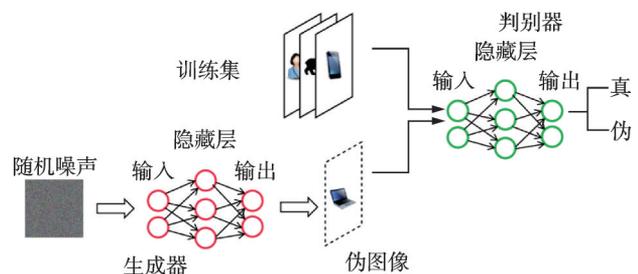


图2 GAN网络结构

Fig.2 Structure of GAN network

断更加准确,这种相互竞争、合作的关系形成了一种对抗机制<sup>[28]</sup>。最终这个竞争过程能够产生高质量、逼真的图像。

基于GAN网络的压缩算法通常使用自编码器或变分自编码器作为生成网络,同时使用判别器来检测生成的图像和原始图像之间的差异。通过精心设计的损失函数,生成网络可以生成与原始图像相似但压缩率更高的图像,而判别网络可以帮助生成网络学习有效的压缩表示形式。GAN已经在图像压缩中取得了显著的成果,并在极低比特率图像压缩方面具有很大的潜力。

### (1) 针对高分辨率图像的压缩技术

Santurkar等<sup>[26]</sup>提出基于生成模型的解码器和编码器训练方法,实现了小图像上的低码率压缩,但在极低比特率下损失了图像的重建精度和视觉效果。为此,Agustsson等<sup>[6]</sup>提出了一种基于GAN网络的新框架。该框架中包括生成压缩和选择性生成压缩两种模式,其能够保留图像结构和内容范围,实现全分辨率图像的极低比特率压缩,产生高分辨率的重建图像,但该方法使用了语义映射,这需要大量的存储比特,且通常难以获取。其网络架构如图3所示,其中 $E$ 为编码器, $q$ 为量化器, $\hat{\omega} = q(E(x))$ 表示将图像 $x$ 进行压缩的结果。生成器 $G$ 生成与原图相近的 $\hat{x}$ , $D$ 为鉴别器,用来帮助 $G$ 生成更逼真的图像。

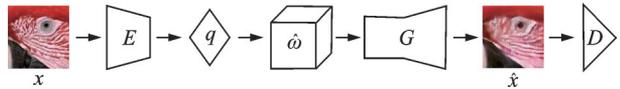


图3 基于GAN网络的压缩框架

Fig.3 Compression framework based on GAN networks

Han等<sup>[29]</sup>在文献[6]的基础上,提出了一种基于潜在表征信道间冗余的掩码压缩方法,该方法通过掩码有效地压缩潜在表征信道冗余,通过减少语义映射来降低存储需求,并使用相对速率估计替代熵模型,可适用于大图像的可变速率生成压缩。Akutsu等<sup>[30]</sup>在此基础上,使用两个解码器来实现低比特率和高感知质量,其主解码器由条件GAN训练,另一个解码器由加权MS-SSIM损失训练,能对不同区域采用不同解码器进行自动处理。

Dash等<sup>[31]</sup>提出了CompressNet网络,将GAN与自动编码器集成在一起,以更好地理解图像的全局结构。该网络引入了堆叠式自编码器(Stacked what where autoencoder, SWWAE)<sup>[32]</sup>,利用池化开关信息改进编解码器之间的数据重建,用开关预测网络来减少合并池化开关信息生成的额外数据开销。该网络在视觉质量方面表现良好,但自适应图像压缩能力较弱。

### (2) 针对生成图像模糊问题的压缩技术

目前,大多数基于学习的图像压缩方法存在模糊问题,GAN虽能改善图像使其重建清晰,但仍存在训练不稳定和产生噪音或伪影问题。IWAI等<sup>[33]</sup>采用了两阶段训练和网络插值来克服以上缺点,结合Cheng等<sup>[34]</sup>提出的离散高斯混合模型优化潜在码分布的参数化方式,有效提升了图像压缩的质量。

为了克服GAN训练不稳定和产生模糊的伪影问题,Pan等<sup>[35]</sup>提出采用预先训练的文本到图像扩散模型来建立模型,即利用现有的文本到图像扩散模型,将输入的图像压缩为极低比特率的文本嵌入,然后生成具有清晰细节的图像。该方法可以生成具有多样性的高感知质量的图像,不用在专用数据集上进行训练,且不受服务器和客户端计算资源的限制,但其计算效率略低,压缩和解压缩步骤更耗时,且在通信带宽方面非常有限。图4为Pan等<sup>[35]</sup>与Agustsson等<sup>[6]</sup>及BPG对同一图像压缩后的重构效果对比。

### (3) 基于纹理信息的压缩技术

尽管以上方法可以重建清晰的图像,但在极低比特率下难以恢复准确的纹理。为此,Iwai等<sup>[36]</sup>基于参考的图像超分辨率<sup>[37]</sup>方法,提出了一种自纹理传输模型。该模型通过从图像中裁剪出的小块送入细化的压缩网络进行压缩,以生成参考补丁,并通过自纹理传输模型,将纹理转移到已被压缩过的图像中。实验表明该方法能有效传递并恢复纹理。

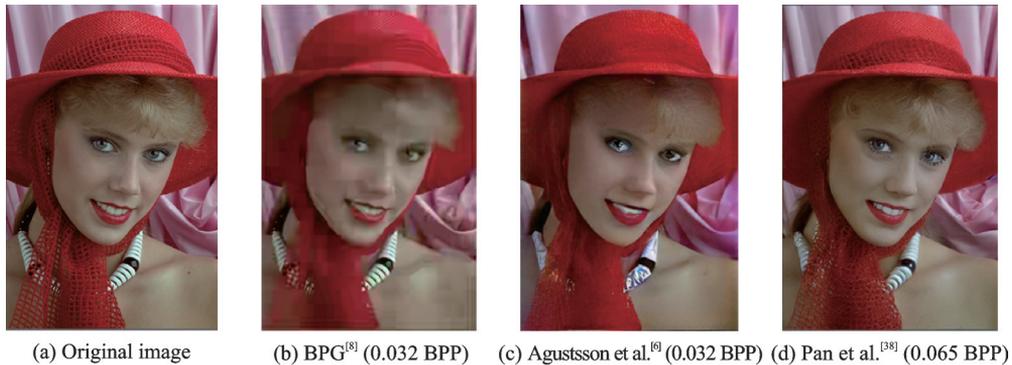


图4 重构效果对比

Fig.4 Comparison of reconstruction effects

Pan 等<sup>[38]</sup>提出了一种采用内容分支和纹理分支的极低比特率耦合压缩生成网络(Coupled compression generation network, CCGN)。该模型通过多维剩余注意机制生成极低比特流,采用增强的感知引导细化模型和多尺度融合鉴别器来获得更逼真的纹理,在遥感图像重建中可保留更多关键特征。

#### (4)合理利用语义信息的技术

语义信息对于图像的视觉表达至关重要,但在现有的基于学习的图像表征模型中常被忽略,而GAN则在保留语义结构方面有良好表现。Huang 等<sup>[39]</sup>提出了基于GAN的多尺度自动编码器极低图像编码/压缩框架,利用多尺度鉴别器来捕获局部和全局的图像统计信息,将其用于高分辨率尺度下有效压缩的先验。但该算法忽视了极低比特率场景下语义信息的视觉意义,致使客观评价受到一定的影响。

Chang 等<sup>[40]</sup>提出了语义感知视觉分解(Semantic-aware visual decomposition, STAVE)图像编码框架,利用语义分割映射来表达结构布局和语义信息,提取深度表征来建模纹理分布,并引入注意力机制和相干正则化的方法优化纹理表征空间;将跨通道熵模块用于纹理表征的准确熵估计,以改善率失真性能。该框架在极低比特率和智能视觉表示和分析等方面具有优势。

#### (5)基于矢量量化的技术

通常,GAN将生成器作为自编码器,在瓶颈层引入量化器进行压缩和重构。矢量量化(Vector quantization, VQ)在相同比特率下比标量量化的编码效率更高<sup>[41]</sup>。Agustsson 等<sup>[6]</sup>和Rippel 等<sup>[15]</sup>采用均匀标量量化的自适应算术编码技术来实现轻量级压缩网络。Lohdefink 等<sup>[42]</sup>遵循 Agustsson 等<sup>[6]</sup>和 Rippel 等<sup>[15]</sup>的架构,将对抗损失函数与矢量量化器结合,提高了压缩性能和重构图像质量。

Mao 等<sup>[43]</sup>将VQ引入生成模型,提出了矢量量化生成对抗网络(Vector quantization generative adversarial network, VGGAN)。该模型通过识别学习码本中最接近的样本来获得VQ指数,实现极高的压缩倍数;预训练的大规模码本通过K-means算法聚类为更小的码本,实现可变比特率和不同级别的重建质量。

#### (6)对图像感兴趣区域的压缩

对于感兴趣区域(Region of interest, ROI)的分割压缩同样具有重要意义。传统的方法主要包括静态图像压缩标准JPEG2000支持的ROI编码技术<sup>[8]</sup>和基于小波变换的加权算法<sup>[44]</sup>。在低比特率下,以上方法不能有效地重建ROI与背景区域,需要手工标记显著区域,且在ROI模型中会出现严重的块效应。为此,Du 等<sup>[45]</sup>利用生成模型,提出了一种基于ROI的图像压缩方案。该方法通过显著目标检测<sup>[46]</sup>区分图像的前景和背景,采用不同的损失函数对生成的压缩图像进行处理,以增强背景细节并恢复视觉注意目标的真实细节,可在极低比特率下产生视觉上令人满意的图像效果。

### (7)针对人脸图像的压缩

Mao等<sup>[47]</sup>提出基于学习的带有外部先验的人脸图像压缩框架(Peer into the sketch, PITS)。该框架将人脸图像编码为草图和缩略图,并使用可逆神经网络<sup>[48]</sup>进行编码。解码器端开发了一种两阶段生成重建的复杂解码器,结合半参数建模和从外部数据库检索指导来增强重建图像。第一阶段利用生成式解码器结合解码后的草图和缩略图进行初步人脸图像的重构。第二阶段通过评估特征域相似度来检索外部数据库中的图像。该模型编码器具有丰富的语义信息和较强的鲁棒性,但只适用于人脸图像的极低比特率压缩,且PITS依赖解码器侧的图像检索来利用外部先验。

## 2.2 基于非GAN模型的图像压缩方法

除了基于GAN的极低比特率压缩方法外,一些非GAN方法,如分层压缩、概念压缩、跨模态压缩以及基于对象压缩等极低比特率压缩方法也值得关注。

Li等<sup>[49]</sup>提出了一种基于掩码压缩模型(Masked compression model, MCM)的图像压缩方法,旨在极低比特率下实现图像的高质量重建。该方法将自监督预训练的掩码自动编码器(Masked autoencoder, MAE)与学习图像压缩(Learned image compression, LIC)框架结合,利用双自适应掩蔽方法(Dual-adaptive masking approach, DA-Mask)适应图像的结构和纹理成分,以有效去除冗余并补充信息。在掩码图像建模阶段,DA-Mask和预训练的MAE相结合为初始压缩器,用于提取语义上下文和纹理表征。该方法与LIC网络配合使用,实现了两步压缩。实验结果表明,MCM在极低比特率下具有优于先进LIC方法的评价指标和视觉质量。

概念压缩技术<sup>[50]</sup>通过提取输入视觉数据的概念表示并仅存储图像的全局信息来有效降低比特率。然而,现有的熵建模技术仍存在不足,缺乏速率优化及未充分利用纹理编码内部区域的相似性和独立性进行熵约束训练<sup>[51]</sup>。为了解决这些问题,Chang等<sup>[52]</sup>提出了基于概念编码的端到端语义先验建模方法(本文称其为Thousand to one),利用语义分割图作为结构指导,在每个语义区域内提取深度语义先验,实现更精细的纹理表示和合成。此外,引入跨通道熵模型,用于有效的先验估计和减少语义先验的通道依赖性,以在熵约束下进行更有效的率失真优化。该方法实现了超高的压缩比(1 000倍),同时提高了重建质量和内容操作的灵活性,对视觉处理和分析任务具有通用性。Chang等<sup>[53]</sup>提出了一个新的概念压缩框架,将图像内容分解为结构层和纹理层,并使用分层融合GAN(Hierarchica fusion GAN, HF-GAN)作为解码器。但在应用于语义缺口较大的数据集时,该模型无法按预期生成纹理,且会出现伪影问题。图5为Chang等<sup>[53]</sup>与JPEG、JPEG2000、BPG的重构效果对比。

为了提高实例分割网络的图像编码性能,Galteri等<sup>[54]</sup>和Fischer等<sup>[55]</sup>提出了标准混合解码器,利用显著性区域和非显著性区域来减少比特率。Chamain等<sup>[56]</sup>和Le等<sup>[57]</sup>采用神经图像压缩网络作为编解码器,在目标检测和分割任务上取得了很好的效果,但必须在需要高质量的突出区域和没有潜在对象的非突出区域之间进行隐式推导。Fischer等<sup>[58]</sup>通过引入潜在空间掩蔽网络来解决这一问题,并降低所需的比特率。Fischer等<sup>[59]</sup>利用基于VCM(Video coding for machines)技术的显著性信息,提出了分层学习压缩框架。该方法在3个不同空间分辨率的潜在空间中进行分层编码,并引入RDONet屏蔽准则来提高VCM场景下的编码性能,从而节省了大量比特数,其性能优于传统的非分层压缩网络。

Li等<sup>[60]</sup>提出了一种跨模态框架来实现语义通信,旨在增强语义信息对图像压缩的作用。该方法只保持语义的一致性,没有充分考虑信号级重构。Zhang等<sup>[61]</sup>提出了一种可扩展的跨模态压缩(Scalable cross-modality compression, SCMC)框架,其结构如图6所示。该框架包括语义层、结构层和信号层,具体而言,编解码器由语义层( $E_1$ 和 $D_1$ )、结构层( $E_2$ 和 $D_2$ )和信号层( $E_3$ 和 $D_3$ )三层组成。通过分层绘制不同模态的图像,从语义级表示到信号级重构来恢复视觉信息。在语义层仅编码来实现紧凑的表示,通过提取几何结构来弥合语义理解和信号重建之间的差距。SCMC框架在低比特率下取得了良好的率



图5 不同压缩方法重构效果对比

Fig.5 Comparison of reconstruction effects of different compression methods

失真性能,但仍需进一步适应不同分辨率的图像内容。

基于对象的图像编码(Object-based image coding, OBIC)在极低比特率通信和高级语义内容理解方面有很大潜力,但对任意形状对象的压缩效率不高,且需要大量的计算成本,因此不易广泛应用。Sikora等<sup>[62]</sup>和Li等<sup>[63]</sup>开发出形状自适应离散小波变换和余弦变换来支持任意形状,但计算量较大。Xia等<sup>[64]</sup>通过元素掩蔽机制来支持任意形状的表达,提出了LearntOBIC系统,摒弃了专用于形状自适应的工具<sup>[62-63]</sup>。LearntOBIC采用端到端的学习框架,包括对象层和背景层,其中对象层采用DeepLab技术<sup>[65]</sup>进行图像分割,背景层采用基于变分自编码器的NLAIC<sup>[66]</sup>进行并行处理掩膜图像层。与传统的JPEG2000、BPG以及基于学习的NLAIC相比,LearntOBIC在极低比特率下可为应用场景提供更好的视觉质量。但LearntOBIC中的分割通常依赖于内容,如何使分割效果更加稳定突出需要进一步研究。

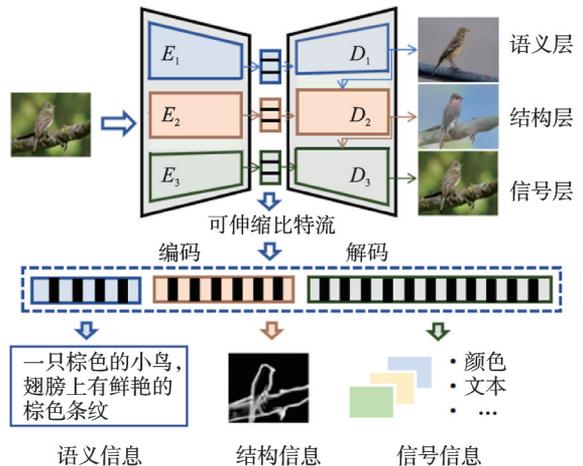


图6 可扩展的跨模态压缩网络结构

Fig.6 Structure of SCMC network

Gao等<sup>[67]</sup>提出了基于可逆图像表示的ULCompress框架,用于解决现有低比特率图像压缩框架忽略信息损失和兼容性问题。ULCompress包括可逆图像重缩放(Invertible image rescaling, IIR)模块和

压缩质量增强(Compressed quality enhancement, CQE)模块,其结构如图7所示。其中IIR模块学习高分辨率图像并压缩低分辨率图像,而CQE用于提高该框架的兼容性,在低比特率场景下明显缓解阻塞和模糊现象。该框架保证了可逆的采样过程,提高了压缩后的图像质量,在编码比特率上与标准编码器和基于学习的编解码器保持一致。相比其他方法,ULCompress关注采样过程中的信息损失问题,具有更好的通用性和实际应用前景。

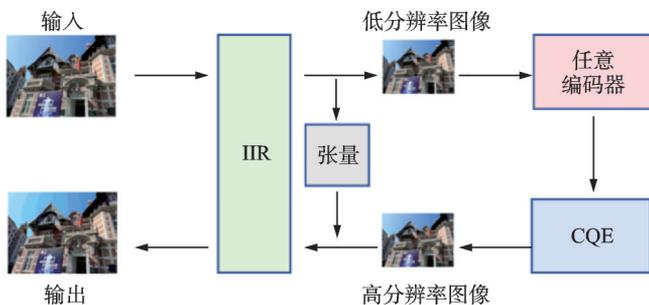


图7 ULCompress网络结构

Fig.7 Structure of ULCompress network

### 3 数据集和评价指标

#### 3.1 数据集介绍

大量的图像训练数据才能使复杂的深度学习网络模型有效。为了验证所提算法对图像压缩的效果,选择合适、有效的数据集尤为重要。以下是常见的图像压缩数据集。

(1) CLIC数据集<sup>[68]</sup>是CVPR联合举办的深度学习图像压缩挑战赛提供的用于训练和基准测试失真性能的移动数据集,包括了不同种类的PNG图片,其尺寸被调整为长边是2 048像素。

(2) Kodak PhototCD数据集<sup>[69]</sup>由Kodak公司制作,包括24张不同场景的图像(如城市、海景、人像等),分辨率为768像素 $\times$ 512像素,常用于评估图像超分辨率重建算法性能。

(3) COCO数据集<sup>[70]</sup>2015年由康奈尔大学、微软研究院和加州理工学院的研究者共同制作。数据集包含80个对象类别,91种物体类型,常用于目标检测、分割和字幕数据集。

(4) CelebAMask-HQ数据集<sup>[71]</sup>2020年由商汤研究院、香港中文大学和香港大学的研究人员合作开发。该数据集由超过30 000张高分辨率的人脸图像组成,包含细粒度的掩码标签,每张图片都带有19个面组件类别的语义掩码(例如眼睛区域、鼻子区域、嘴巴区域等),常用于人脸解析、人脸识别、人脸生成和编辑。

(5) ADE20k数据集<sup>[72]</sup>2017年由美国麻省理工学院和加拿大多伦多大学的研究人员共同开发。数据集包括150个类别,注释密集,场景丰富,包括室内、室外和自然场景等。应用于场景感知、解析、分割、多物体识别和语义理解。

(6) FFHQ(Flickr-Faces-HQ)数据集<sup>[73]</sup>是英伟达公司建立的一个高质量人脸图像数据集,主要用于GAN的训练和评估,由70 000张分辨率为1 024像素 $\times$ 1 024像素的高质量PNG图像组成。该数据集在其年龄、种族和图像背景方面比CelebAMask-HQ数据集更丰富多彩。

(7) CUB-00-2011数据集<sup>[74]</sup>由加州理工学院与加州大学圣地亚哥分校研究人员共同制作而成,该数据集包含200种鸟类的11 788张图像,这些鸟类使用相同的部件进行注释,每个图像包含15个部件位置,312个二进制属性,可用于共享部件模型的研究。

(8) Cityscape数据集<sup>[75]</sup>2015年包含50多个不同城市的5 000张街道场景图片,具有19类别的密集像素标注,并且提供算法评估接口,应用于城市场景像素级、实例级语义分割等。

(9) PASCAL VOC 2012<sup>[76]</sup>是一个世界级的机器视觉挑战赛提供的数据集。数据集共有4大类

(Vehicle, Household, Animal, Person)和20小类,常应用于图像分类、目标检测、目标分割和行为识别等多个领域。

### 3.2 评价指标

图像压缩的质量评价可以从主观和客观两个角度进行。主观质量评价是基于人眼对图像质量的感知,通过观察者的视觉感受对图像进行打分或给出意见。这种评价方法直观准确地反映了人们对图像质量的主观感受。客观质量评价则采用一系列客观的评价指标来量化图像质量。这些指标基于图像的特征和统计数据,对图像的失真程度、信息损失等进行分析。主观质量评价提供直接、全面的评估,但评价结果容易受个体差异的影响。而客观质量评价则更客观、可重复,可在大规模数据集上进行自动化评估。综合使用主观和客观评价方法可以更全面地评估图像压缩的质量。客观质量评价方法主要有以下几种:

PSNR<sup>[12]</sup>是一种常用的客观评价指标,它基于原始图像与压缩重构后图像之间对应像素点间的误差,取值范围通常在20~50 dB之间,数值越大表示图像失真越小,图像质量越高。PSNR通过计算原始图像与压缩重构后图像之间的均方根误差(Mean squared error, MSE)来量化它们之间的差异,即

$$\text{PSNR} = 10 \lg \frac{\text{MAX}^2}{\text{MSE}} = 20 \lg \frac{\text{MAX}}{\sqrt{\text{MSE}}} \quad (1)$$

$$\text{MSE} = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2 \quad (2)$$

式中:MAX表示输入图像的最大像素值; $m$ 和 $n$ 分别为图像的高度和宽度, $m \times n$ 即为原始图像 $I(i,j)$ 与压缩重构后图像 $K(i,j)$ 的尺寸, $i$ 和 $j$ 为像素的位置索引。然而,PSNR忽略了人眼对图像差异的敏感性,因此与人眼的评估结果可能会存在偏差。

结构相似性(Structural similarity, SSIM)<sup>[77]</sup>是一种全参考的图像质量评价指标,它分别从亮度、对比度、结构3方面度量图像相似性,其范围为 $[0, 1]$ ,即

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (3)$$

式中: $\mu_x$ 为 $x$ 的均值, $\mu_y$ 为 $y$ 的均值, $\sigma_x^2$ 为 $x$ 的方差, $\sigma_y^2$ 为 $y$ 的方差, $\sigma_{xy}$ 为 $x$ 和 $y$ 的协方差; $c_1 = (k_1L)^2$ , $c_2 = (k_2L)^2$ 为维持稳定的常数, $L$ 为像素值的动态范围。SSIM的数值越大,表示图像质量越好。SSIM在图像相似度的评价上优于PSNR。

MS-SSIM<sup>[13]</sup>是SSIM在多个尺度上的扩展。它将原始图像和重建图像划分为不同尺度的子图像,对每个尺度的子图像计算SSIM指数,对该SSIM指数进行加权平均后得到最终的MS-SSIM值,其范围在0到1之间,表达式为

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

式中: $M$ 表示图像尺度;采用指数 $\alpha_M$ 、 $\beta_j$ 和 $\gamma_j$ 调整亮度 $l_M$ 、对比度 $c_j$ 和结构对比 $s_j$ 三者不同分量的相对重要性。MS-SSIM的数值越大,表示图像间的相似程度越高,图像质量越好。

LPIPS是一种可学习的感知图像块相似度指标<sup>[16]</sup>,也称为感知损失,它通过训练一个神经网络模型来近似人类感知的视觉相似性。其范围通常为 $[0, 1]$ ,数值越小表示图像的感知质量越高。相比于传统的评价指标如PSNR和SSIM,LPIPS更符合人眼对于图像感知的敏感性。

FID<sup>[17]</sup>是从原始图像的计算机视觉特征的统计方面来衡量两组图像的相似度,用于计算原始图像和重建图像的特征向量之间的距离。FID数值越小越好。计算FID需要使用Inception及Net-V3模型

提取图像的特征向量,并计算两张图像特征之间的距离,表达式为

$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}) \quad (5)$$

式中: $\mu_r$ 表示真实图像的特征均值; $\mu_g$ 表示生成图像的特征均值; $\Sigma_r$ 为真实图像的协方差矩阵; $\Sigma_g$ 为生成图像的协方差矩阵; $\text{tr}(\cdot)$ 表示矩阵的迹。

### 3.3 方法比较

基于GAN和非GAN网络模型各类极低比特率图像压缩方法在BPP、重构质量及算法特点方面的比较分析分别如表2和表3所示。从表2可以看出,Chang等<sup>[40]</sup>和Mao等<sup>[47]</sup>的方法相较于其他研究者提出的图像压缩方法在BPP较小时重构质量更为优秀,但Mao等<sup>[47]</sup>的方法只适用于人脸图像压缩。从表3可以看出,Thound to one方法<sup>[52]</sup>相较于其他方法的BPP和重构质量更为优秀。通常,表2和表3内所述方法具有不同的BPP和图像重构质量,适用于不同类型的图像数据集和应用场景,因此在图像压缩应用中,需要根据具体的应用场景来进行选择。

表2 基于GAN网络模型各类极低比特率图像压缩方法比较

Table 2 Comparative analysis of various very low bitrate image compression methods based on GAN network models

作者	年份	特点	BPP ↓ (数据集)	重构质量
Iwai等 <sup>[36]</sup>	2021	在基于参考的图像超分辨率方法上提出了自纹理传输模型	0.028 7(CLIC)	次
Han等 <sup>[29]</sup>	2020	利用掩码压缩潜在表征信道冗余,通过减少语义映射降低存储需求	0.028(ADE2k)	良
Cheng等 <sup>[34]</sup>	2020	将离散化的高斯混合模型和注意力机制引入压缩模型	0.098 (Kodak)	良
Pan等 <sup>[35]</sup>	2022	首次将模型建立在预训练文本-图像扩散模型上,能从一个压缩源生成多种图像,不用在专用数据集上进行训练,计算效率低	0.070±0.008(Kodak)	良
Löhdefink等 <sup>[42]</sup>	2021	将对抗损失函数与矢量量化器结合	0.031 25~0.093 75 (Cityscape)	良
Du等 <sup>[45]</sup>	2019	采用不同的损失函数对图像的前景和背景分别进行处理,实现了极低比特率下ROI的分割压缩	0.037(MSRA-B)	良
Huang等 <sup>[39]</sup>	2019	利用多尺度鉴别器实现局部和全局的图像统计信息捕获	0.039(Cityscape) 0.047(ADE20K)	中
Dash等 <sup>[31]</sup>	2020	利用开关预测网络增强堆叠自编码器,减少数据开销	0.072 6(CLIC)	优
Pan等 <sup>[38]</sup>	2023	将GAN与卷积神经网络结合,通过融合图像内容和细节纹理达到图像的高保真压缩重构,实现了遥感图像的极低比特率压缩	0.062(DOTA dataset) 0.091(UC-Merced)	优
Chang等 <sup>[40]</sup>	2023	引入注意力机制和相干正则化去优化纹理表征空间,将跨通道熵模块用于纹理表征的准确熵估计	0.045(FFHQ) 0.018(ADE20K) 0.024(Cityscape)	优
Mao等 <sup>[47]</sup>	2023	首次提出了基于学习的带有外部先验的人脸图像压缩框架	0.0281 (CelebAMask-HQ)	优
Mao等 <sup>[43]</sup>	2023	通过识别学习码本中最接近的样本来获得VQ指数,间接压缩潜在特征	0.04~0.07(Kodak)	优

注: ↓表示越小越好。

表3 基于非GAN网络模型各类极低比特率图像压缩方法比较

Table 3 Comparative analysis of very low bitrate image compression methods based on non-GAN network models

模型	年份	特点	BPP↓(数据集)	重构质量
SCMC <sup>[61]</sup>	2023	通过语义层、结构层和信号层分层绘制不同模态的图像,将压缩问题进一步转化为表示任务,依次在3个层次上恢复视觉信息。	语义层: 0.005(CUB-200-2011) 信号层: 0.102(CUB-200-2011)	良
LearntOBIC <sup>[64]</sup>	2020	基于OBIC将分割网络和压缩网络集成在一个端到端可学习的框架	0.064 8(PASCAL VOC)	良
VCM+RDOnet <sup>[59]</sup>	2023	首次利用基于VCM的显著性信息提出分层学习压缩框架,通过3个不同空间分辨率的潜在空间中传输图像数据进行分层编码	0.059(Cityscape)	中
MCM <sup>[49]</sup>	2023	首次将MIM和LIC端到端统一起来用于极低比特率图像压缩,注重图像结构与纹理成分	0.075(COCO) 0.080 3(CelebAMask-HQ)	优(纹理细腻)
Thousand to one <sup>[52]</sup>	2021	首次提出基于概念编码的端到端语义先验建模方法	0.02~0.03(CelebAMask-HQ)	优
HF-GAN <sup>[53]</sup>	2022	将图像内容分解为结构层和纹理层,通过提取稀疏结构映射和深度纹理表示来实现概念压缩,语义缺口较大时压缩效果较差	0.031(Edges2shoes&-handbags) 0.074(CelebAMask-HQ) 0.043(Multiple seasons)	优
ULCompress <sup>[67]</sup>	2023	能够兼容标准解码器和基于学习的编解码器	0.091: ULC+JPGE(DIV2K) 0.060: ULC+HEVC(DIV2K) 0.047: CL+Cheng等 <sup>[34]</sup> (DIV2K)	优

注:↓表示越小越好。

#### 4 总结与展望

极低比特率图像压缩是一项具有挑战性的任务,它需要在保持高质量重建图像的同时,实现极低的比特率压缩。本文对基于GAN和非GAN模型的极低比特率图像压缩方法进行了分类评述。在此基础上,分析了极低比特率有损图像压缩算法的优点与不足,描述了部分常用的图像压缩数据集和客观性能评价指标。最后,针对当前极低比特率图像压缩面临的问题进行了讨论与展望。随着深度学习技术的发展,GAN网络及其他网络模型将会更广泛地应用于有损图像的极低比特率压缩,并取得更低的压缩比特率和更好的图像重建质量。下面对极低比特率图像压缩技术未来的发展趋势进行展望:

(1)基于GAN的极低比特率图像压缩方法备受关注,已取得了良好的压缩效果。但仍然面临一些技术挑战。改进生成器网络的设计很必要,以提高在极低比特率下生成图像的质量。同时,判别器网络也需要更加精确地评估生成图像和原始图像之间的差异,以帮助生成器网络改进生成图像的质量。此外,GAN是由生成器和判别器两部分组成的哑铃型网络,二者通过交替训练逐步生成高质量、逼真的数据,在图像合成、修复、生成等任务中得到广泛应用。研究人员可以探索类似GAN结构的网络,如VQ和U-net等,来完成极低比特率图像压缩任务。

(2)需要关注的另一个问题是针对特定内容的图像和文字的压缩。在许多场景中,特定内容的图

像和文字具有重要的信息含量,但目前针对这类图像压缩问题的解决方法尚不完善。因此,需要进一步研究和探索针对特定内容压缩的方法,以保证重要信息的准确传输。

(3)大多数方法都是优化图像数据传输中的压缩和解压缩过程。如何从图像的后处理入手,进一步提高重建图像的质量也将是研究的重点。

(4)未来的发展方向之一是使极低比特率图像压缩方法能够适应不同分辨率的各种图像内容,并具备通用性和可扩展性。这将使得该方法在图像和视频编码以及机器视觉任务等领域发挥更大的作用。

综上所述,通过对当前面临的问题进行讨论和思考,将有助于在极低比特率图像压缩领域取得更大的突破和进展。同时,也可以探索其他新的方法和技术,以解决当前面临的挑战,并推动图像压缩领域的创新和发展。

#### 参考文献:

- [1] HUSSAIN A J, AL-FAYADH A, RADI N. Image compression techniques: A survey in lossless and lossy algorithms[J]. *Neurocomputing*, 2018, 300: 44-69.
- [2] SAYOOD K. Introduction to data compression [M]. 5th ed. [S.l.]: Morgan-Kaufmann Publishers Inc., 2018.
- [3] ALGAZI V R, AVADHANAM N, ESTES J R R. Quality measurement and use of pre-processing in image compression[J]. *Signal Processing*, 1998, 70(3): 215-229.
- [4] SHAHID M, ROSSHOLM A, LÖVSTRÖM B, et al. No-reference image and video quality assessment: A classification and review of recent approaches[J]. *EURASIP Journal on Image and Video Processing*, 2014(1): 40.
- [5] MALLAT S, FALZON F. Analysis of low bit rate image transform coding[J]. *IEEE Transactions on Signal Processing*, 1998, 46(4): 1027-1042.
- [6] AGUSTSSON E, TSCHANNEN M, MENTZER F, et al. Generative adversarial networks for extreme learned image compression[C]//*Proceedings of IEEE International Conference on Computer Vision*. Seoul, Korea (South): IEEE, 2019: 221-231.
- [7] WALLACE G K. The JPEG still picture compression standard[J]. *Communications of the ACM*, 1991, 34(4): 30-44.
- [8] SKODRAS A, CHRISTOPOULOS C, EBRAHIMI T. The JPEG 2000 still image compression standard[J]. *IEEE Signal Processing Magazine*, 2001, 18(5): 36-58.
- [9] BELLARD F. BPG image format[EB/OL]. (2018-04-21). <https://bellard.org/bpg>.
- [10] SZE V, BUDAGAVI M, SULLIVAN G J. High efficiency video coding (HEVC): Algorithms and architectures[M]. [S.l.]: Springer, 2014.
- [11] BROSS B, WANG Y K, YE Y, et al. Overview of the versatile video coding (VVC) standard and its applications[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(10): 3736-3764.
- [12] HORÉ A, ZIOU D. Image quality metrics: PSNR vs. SSIM[C]//*Proceedings of IEEE International Conference on Pattern Recognition*. Istanbul, Turkey: IEEE, 2010: 2366-2369.
- [13] ROUSE D M, HEMAMI S S. Analyzing the role of visual structure in the recognition of natural image content with multi-scale SSIM[C]//*Proceedings of the Human Vision and Electronic Imaging XIII*. San Jose, California, USA: [s.n.], 2008: 410-423.
- [14] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[J]. *Advance in Neural Information Processing Systems*, 2014, 27: 2672-2680.
- [15] RIPPEL O, BOURDEV L. Real-time adaptive image compression[C]//*Proceedings of International Conference on Machine Learning*. Sydney: [s.n.], 2017: 2922-2930.
- [16] ZHANG R, ISOLA P, EFROS A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE, 2018: 586-595.
- [17] OBUKHOV A, KRASNYANSKIY M. Quality assessment method for GAN based on modified metrics inception score and Fréchet inception distance[C]//*Proceedings of the Software Engineering Perspectives in Intelligent Systems*. [S.l.]: Springer, 2020: 102-114.

- [18] 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[J]. 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[J]. Computer Engineering and Applications, 2020, 56(15): 15-23.
- [19] MA S, ZHANG X, JIA C, et al. Image and video compression with neural networks: A review[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(6): 1683-1698.
- [20] SHAPIRO J M. Embedded image coding using zero trees of wavelet coefficients[J]. IEEE Transactions on Signal Processing, 1993, 41(12): 3445-3462.
- [21] BRAHIMI T, LAOUIR F, BOUBCHIR L, et al. An improved wavelet-based image coder for embedded grey scale and colour image compression[J]. International Journal of Electronics and Communications, 2017, 73: 183-192.
- [22] SAID A, PEARLMAN W A. A new, fast, and efficient image codec based on set partitioning in hierarchical trees[J]. IEEE Transactions on Circuits and Systems for Video Technology, 1996, 6(3): 243-250.
- [23] 王亚鸽, 康晓东, 郭军, 等. 一种联合 Canny 边缘检测和 SPIHT 的图像压缩方法[J]. 计算机科学, 2019, 46(S1): 222-225.  
WANG Yage, KANG Xiaodong, GUO Jun, et al. Image compression method combining Canny edge detection and SPIHT [J]. Computer Science, 2019, 46(S1): 222-225.
- [24] CHAVAN P P, SHEELA RANI B, MURUGAN M, et al. An image compression model via adaptive vector quantization: Hybrid optimization algorithm[J]. The Imaging Science Journal, 2020, 68(5/6/7/8): 259-277.
- [25] SHEN F, HASEGAWA O. An adaptive incremental LBG for vector quantization[J]. Neural Networks, 2006, 19(5): 694-704.
- [26] SANTURKAR S, BUDDEN D, SHAVIT N. Generative compression[C]//Proceedings of the Picture Coding Symposium. San Francisco, CA, USA: IEEE, 2018: 258-262.
- [27] 卢奇, 龚勋. 结构约束下的生成对抗深度图修复[J]. 数据采集与处理, 2023, 38(5): 1048-1057.  
LU Qi, GONG Xun. Generate adversarial depth repair under structural constraints[J]. Journal of Data Acquisition and Processing, 2023, 38(5): 1048-1057.
- [28] 周韪鼎, 陈兆学. 基于生成对抗网络的 PPG-ECG 信号转换方法[J]. 数据采集与处理, 2023, 38(3): 608-615.  
ZHOU Weiding, CHEN Zhaoxue. Method of transferring PPG to ECG based on generative adversarial network[J]. Journal of Data Acquisition and Processing, 2023, 38(3): 608-615.
- [29] HAN C, DUAN Y, TAO X, et al. Toward variable-rate generative compression by reducing the channel redundancy[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(7): 1789-1802.
- [30] AKUTSU H, SUZUKI A, ZHONG Z, et al. Ultra low bitrate learned image compression by selective detail decoding[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops. Seattle, WA, USA: IEEE, 2020: 524-528.
- [31] DASH S, KUMARAVELU G, NAGANOOR V, et al. CompressNet: Generative compression at extremely low bitrates[C]// Proceedings of IEEE Winter Conference on Applications of Computer Vision. Snowmass, CO, USA: IEEE, 2020: 2314-2322.
- [32] ZHAO J, MATHIEU M, GOROSHIN R, et al. Stacked what-where auto-encoders[EB/OL]. (2016-02-14). <https://arxiv.org/abs/1506.02351>.
- [33] IWAI S, MIYAZAKI T, SUGAYA Y, et al. Fidelity-controllable extreme image compression with generative adversarial networks[C]//Proceedings of IEEE International Conference on Pattern Recognition. Milan, Italy: IEEE, 2021: 8235-8242.
- [34] CHENG Z, SUN H, TAKEUCHI M, et al. Learned image compression with discretized Gaussian mixture likelihoods and attention modules[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 7936-7945.
- [35] PAN Z, ZHOU X, TIAN H J A E-P. Extreme generative image compression by learning text embedding from diffusion models [EB/OL]. (2022-11-14). <https://arxiv.org/abs/2211.07793>.
- [36] IWAI S, MIYAZAKI T, SUGAYA Y, et al. Self texture transfer networks for low bitrate image compression[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Nashville, TN, USA: IEEE, 2021: 1901-1905.
- [37] YANG F, YANG H, FU J, et al. Learning texture transformer network for image super-resolution[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 5791-5800.
- [38] PAN T, ZHANG L, QU L, et al. A coupled compression generation network for remote-sensing images at extremely low bitrates[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-14.

- [39] HUANG C, LIU H, CHEN T, et al. Extreme image coding via multiscale autoencoders with generative adversarial optimization[C]//Proceedings of IEEE Visual Communications and Image Processing. Sydney, Australia: IEEE, 2019: 1-4.
- [40] CHANG J, ZHANG J, LI J, et al. Semantic-aware visual decomposition for image coding[J]. International Journal of Computer Vision, 2023, 131(9): 2333-2355.
- [41] GRAY R. Vector quantization[J]. IEEE ASSP Magazine, 1984, 1(2): 4-29.
- [42] LÖHDEFINK J, HÜGER F, SCHLICHT P, et al. Scalar and vector quantization for learned image compression: A study on the effects of MSE and GAN loss in various spaces[C]//Proceedings of International Conference on Intelligent Transportation Systems. Rhodes, Greece: IEEE, 2020: 1-8.
- [43] MAO Q, YANG T, ZHANG Y, et al. Extreme image compression using fine-tuned VQGAN models[EB/OL]. (2023-12-15). <https://arxiv.org/abs/2307.08265>.
- [44] BARTRINA-RAPESTA J, SERRA-SAGRISTA J, AULI-LLINAS F. JPEG2000 ROI coding with fine-grain accuracy through rate-distortion optimization techniques[J]. IEEE Signal Processing Letters, 2009, 16(1): 45-48.
- [45] DU Y, ZHAO N, DUAN Y, et al. Object-aware image compression with adversarial learning[C]//Proceedings of IEEE International Conference on Communications in China. Changchun, China: IEEE, 2019: 804-808.
- [46] XU X, MU N, ZHANG H, et al. Salient object detection from distinctive features in low contrast images[C]//Proceedings of IEEE International Conference on Image Processing. Quebec City, QC, Canada: IEEE, 2015: 3126-3130.
- [47] MAO Y, CHEN P, WANG S, et al. Peering into the sketch: Ultra-low bitrate face compression for joint human and machine perception[C]//Proceedings of ACM International Conference on Multimedia. New York, USA: ACM, 2023: 2564-2572.
- [48] XIE Y, CHENG K L, CHEN Q. Enhanced invertible encoding for learned image compression[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York, USA: ACM, 2021: 162-170.
- [49] LI A, LI F, HAN J, et al. You can mask more for extremely low-bitrate image compression[EB/OL]. (2023-06-27). <https://arxiv.org/abs/2306.15561>.
- [50] GREGOR K, BESSE F, JIMENEZ REZENDE D, et al. Towards conceptual compression[EB/OL]. (2016-04-29). <https://doi.org/10.48550/arXiv.1604.08772>.
- [51] CHANG J, MAO Q, ZHAO Z, et al. Layered conceptual image compression via deep semantic synthesis[C]//Proceedings of IEEE International Conference on Image Processing. Taipei, China: IEEE, 2019: 694-698.
- [52] CHANG J, ZHAO Z, YANG L, et al. Thousand to one: Semantic prior modeling for conceptual coding[C]//Proceedings of the IEEE International Conference on Multimedia and Expo. Shenzhen, China: IEEE, 2021: 1-6.
- [53] CHANG J, ZHAO Z, JIA C, et al. Conceptual compression via deep structure and texture synthesis[J]. IEEE Transactions on Image Processing, 2022, 31: 2809-2823.
- [54] GALTERI L, BERTINI M, SEIDENARI L, et al. Video compression for object detection algorithms[C]//Proceedings of IEEE International Conference on Pattern Recognition. Beijing, China: IEEE, 2018: 3007-3012.
- [55] FISCHER K, FLECKENSTEIN F, HERGLOTZ C, et al. Saliency-driven versatile video coding for neural object detection [C]//Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Toronto, Canada: IEEE, 2021: 1505-1509.
- [56] CHAMAIN L D, RACAPÉ F, BÉGAINT J, et al. End-to-end optimized image compression for machines: A study[C]//Proceedings of the Data Compression Conference. Snowbird, UT, USA: IEEE, 2021: 163-172.
- [57] LE N, ZHANG H, CRICRI F, et al. Image coding for machines: An end-to-end learned approach[C]//Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Toronto, Canada: IEEE, 2021: 1590-1594.
- [58] FISCHER K, BRAND F, KAUP A J A E-P. Boosting neural image compression for machines using latent space masking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021. DOI: <https://arxiv.org/abs/2112.08168>.
- [59] FISCHER K, BRAND F, BLUM C, et al. Saliency-driven hierarchical learned image coding for machines[C]//Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. Rhodes Island, Greece: IEEE, 2023: 1-5.
- [60] LI J, JIA C, ZHANG X, et al. Cross modal compression: Towards human-comprehensible semantic compression[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York, USA: ACM, 2021: 4230-4238.
- [61] ZHANG P, WANG S, WANG M, et al. Rethinking semantic image compression: Scalable representation with cross-modality transfer[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2023, 33(8): 4441-4445.

- [62] SIKORA T, MAKAI B. Shape-adaptive DCT for generic coding of video[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 1995, 5(1): 59-62.
- [63] LI S P, LI W P. Shape-adaptive discrete wavelet transforms for arbitrarily shaped visual object coding[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2000, 10(5): 725-743.
- [64] XIA Q, LIU H, MA Z. Object-based image coding: A learning-driven revisit[C]//*Proceedings of the IEEE International Conference on Multimedia and Expo*. London, UK: IEEE, 2020: 1-6.
- [65] 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*, 2018, 40(4): 834-848.
- [66] CHEN T, LIU H, MA Z, et al. End-to-end learnt image compression via non-local attention optimization and improved context modeling[J]. *IEEE Transactions on Image Processing*, 2021, 30: 3179-3191.
- [67] GAO F, DENG X, GAO C, et al. ULCompress: A unified low bit-rate image compression framework via invertible image representation[C]//*Proceedings of IEEE International Conference on Image Processing*. Kuala Lumpur, Malaysia: IEEE, 2023: 2095-2099.
- [68] TODERICI G, SHI W, TIMOFTE R, et al. Workshop and challenge on learned image compression[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE, 2020.
- [69] Eastman KODAK. Kodak PhotoCD dataset[EB/OL]. (1990-01-01). <https://r0k.us/graphics/kodak/>.
- [70] LIN T-Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context[C]//*Proceedings of the European Conference on Computer Vision (ECCV)*. Zurich, Switzerland: Springer, 2014: 740-755.
- [71] LEE C-H, LIU Z, WU L, et al. MaskGAN: Towards diverse and interactive facial image manipulation[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA: IEEE, 2020: 5549-5558.
- [72] ZHOU B, ZHAO H, PUIG X, et al. Scene parsing through ADE20k dataset[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA: IEEE, 2017: 633-641.
- [73] KARRAS T, LAINE S, AILA T. A style-based generator architecture for generative adversarial networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(12): 4217-4228.
- [74] WAH C, BRANSON S, WELINDER P, et al. The Caltech-UCSD birds-200-2011 dataset[EB/OL]. (2011-07-01). [https://www.vision.caltech.edu/datasets/cub\\_200\\_2011/](https://www.vision.caltech.edu/datasets/cub_200_2011/).
- [75] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding[C]//*Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA: IEEE, 2016: 3213-3223.
- [76] University of Oxford. The PASCAL visual object classes homepage[EB/OL]. (2012-09-09). <http://host.robots.ox.ac.uk/pascal/VOC/>.
- [77] SARA U, AKTER M, UDDIN M S J J O C, et al. Image quality assessment through FSIM, SSIM, MSE and PSNR—A comparative study[J]. *Journal of Computer and Communications*, 2019, 7(3): 8-18.

#### 作者简介:



岳爽(1998-),女,博士研究生,研究方向:计算机视觉、图像处理,E-mail: yueshuang@mail.dlut.edu.cn。



陈喆(1975-),通信作者,男,教授,博士生导师,研究方向:音频信号处理、图像处理和宽带无线通信,E-mail: zhechen@dlut.edu.cn。



殷福亮(1962-),男,教授,博士生导师,研究方向:音频信号处理、图像处理和宽带无线通信。

(编辑:张黄群)