

基于语义通信的低比特率图像语义编码方法

何晨光 黄声显 陈舒怡 王 喆

(哈尔滨工业大学电子与信息工程学院, 黑龙江哈尔滨 150000)

摘 要: 近年来森林火灾给环境与人类带来巨大的损失,传统基于卫星、无人机技术的探测手段效率低且成本高,而布置在山区等偏远地区的大规模传感器网络因性能受限,难以负载大数据量业务,尤其是图片等可以清楚提供现场周围环境变化的有效信息。针对以上问题,本文将基于深度学习与语义通信技术,提出一种图像语义编码方法,对火灾中传感器拍摄到的图片所包含的语义进行提取、编码、传输,在接收端利用对抗生成网络完成从语义到图片的重建,进而实现比传统图像压缩编码方法更加稳定且具有更高压缩比率,从而减轻了传感器网络的负载,其中,为了避免对复杂无线信道的学习,该方法采用成熟的 LDPC 码进行信道编码,以保证语义传输的可靠性。仿真结果表明,该方法能够实现比传统 BPG 图片压缩方法具有更低的压缩比特率和更高的清晰度。

关键词: 语义通信; 图像压缩; 对抗生成网络; 深度学习

中图分类号: TN919. 81 **文献标识码:** A **DOI:** 10. 16798/j. issn. 1003-0530. 2023. 03. 004

引用格式: 何晨光,黄声显,陈舒怡,等. 基于语义通信的低比特率图像语义编码方法[J]. 信号处理,2023,39(3): 410-418. DOI: 10. 16798/j. issn. 1003-0530. 2023. 03. 004.

Reference format: HE Chenguang, HUANG Shengxian, CHEN Shuyi, et al. A low bitrates image semantic coding method based on Semantic Communication[J]. Journal of Signal Processing, 2023, 39(3): 410-418. DOI: 10. 16798/j. issn. 1003-0530. 2023. 03. 004.

A Low Bitrates Image Semantic Coding Method Based on Semantic Communication

HE Chenguang HUANG Shengxian CHEN Shuyi WANG Zhe

(School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, Heilongjiang 150000, China)

Abstract: Wildfire has caused heavy damage to the environment and human in recent years. However, commonly used detection methods are based on the satellites or UAVs, which was inefficient and high-cost. Massive sensor network settled in remote area can not afford the big data transmission because of the limited capability, especially images transmission, which can provide useful and clear information about the change of surrounding during the disaster. To solve the above question, this paper proposed a GAN (Generative Adversarial Networks) based framework for image compression using deep learn and semantic communication, by extracting the semantic information of the captured photos and transmitting the semantic label. The generator of GAN reconstructed the corresponding image based on the received semantic information. To avoid learning the complex physical channel feature, channel coding took the convention solution like LDPC code in our method. The experimental results show that the proposed method achieves lower bitrate and less distortion than BPG, the state of the art coding format in traditional image compression.

Key words: semantic communication; image compression; generative adversarial networks; deep learning

1 引言

近年来极端高温、干旱天气频发,由此引起的山火给世界各国带来了巨大的损失,如美国、澳大利亚、中国等国家首当其冲。目前的山火检测技术仍是基于卫星云图或使用无人机进行探测,尤其是边缘地区,这种方法低效、成本高且能收集的信息有限。随着物联网技术的发展,人们逐渐开始在森林中设置大量传感器或边缘设备进行实时监测来搜集更加多样的数据。然而这些设备在完成数据采集、存储、分析、回传等功能的同时还受到功耗、性能等限制,难以图片、视频等复杂的大容量数据。与仅能提供山火预警、探测等小数据信息相比,现场环境的实时画面更能够为相关工作人员提供直观的环境信息,尤其是面对地震、山洪等可能对地形造成多次破坏的自然灾害,对后续的搜救等灾后活动提供有用的信息。

边缘设备对图像的处理通常为先压缩后传送,采用的传统图像压缩编码方法,经过数十年的发展,如哈夫曼编码,算术编码、上下文自适应的二进制编码等熵编码方法,已经逐渐逼近香农极限,基于上述编码方法的JPEG(Joint Photo graphic Experts Group),BPG^[1](Better Portable Graphics)等图像编码格式在大量工程实践中得到广泛的应用,其中,BPG格式是目前公认的表现最优秀的压缩方法,但是这些编码方法仍需要消耗较大的系统资源,致使性能有限的传感器节点仍难以负荷,并且面对大量传感器同时传送所拍摄的图片的情况,大数据业务对于传感器网络来说也是一个难以承受的网络负载量,因此,急需一种新的图像压缩方法来使上述愿景得以实现。

随着深度学习与人工智能的发展,第六代移动通信(6th Generation Mobile Communication, 6G)^[2]背景下的“人-机-物-灵”四元协作通信架构^[3]逐渐受到人们的重视,相比于5G通信技术,“灵”的概念的引入,使得通信系统得以感知用户或是业务的需求,进而传输真正对用户或是业务有用的信息,进一步减少信息冗余,实现对信息的极致压缩。而语义通信^[4]的概念的提出,使得“灵”这一概念得到理论支持。语义通信尽管并未实际脱离香农信息论的限制,但更加关注“端到端”传输过程,而无需像传统通信技术一样对每个环节进行建模处理。通信过

程中不再关注比特流编码、传输、判决,而是对信息中抽象层次更高的语义符号进行提取、传输、推断,衡量指标也从传统的比特差错转换成语义差错,差错容限得以提高,且取得了许多不错的进展。首先,文献[5]对语义通信理论进行了详细的论证,验证其理论可行性,文献[6]中对深度学习如何在语义通信中发挥作用进行详细的论证,并提出了基于深度学习的语义通信框架DeepSC,为后续研究奠定基础。文献[7]则基于联邦学习提出了一种语义感知网络,有效减小通信过程中的资源消耗,提高通信效率。文献[8]则针对语义通信在性能受限的物联网中智能化能量分配问题进行研究,基于博弈论得出在满足个体理性(Individual Rationality, IR)和激励相容(Incentive Compatibility, IC)条件下可以达到更好的通信增益效果。

由J. Ballé等人^[9]提出的一种基于神经网络的端到端图像传输框架,其定义了一个由非线性分析变换、均匀量化器和非线性合成变换组成的端到端图像压缩方法,验证了基于卷积神经网络(CNN)的端到端图像压缩框架的可行性,并在率失真性能和尺度结构相似度(Multiscale structural similarity, MS-SSIM)指标下获得了优于JPEG与JPEG2000的性能,让业界发现了语义通信在图像压缩领域中应用的可行性与优越性。文献[10]在上述基础上提出一种结合了层级先验,自回归先验,图像上下文信息等多种辅助信息的图像语义提取模型,首次实现了基于学习的图像语义压缩方法,在峰值信噪比(Peak Signal to Noise Ratio, PSNR)和MS-SSIM指标中,与BPG相比性能得到提升。另一方面,基于对抗生成网络(Generative Adversarial Network, GAN)的图像语义编码模型也被相继提出,文献[11]所提出的图-图转换多任务通用框架pix2pix结构,实现以标签图、边界图等条件图片语义生成对应真实图片,为后续基于GANs的图像生成工作奠定基础。文献[12]将语义图作为条件对抗生成网络(conditional Generative Adversarial Network, cGAN)的条件,提出多尺度的生成器和鉴别器结构,在无人工设计损失函数和预训练网络的情况下根据语义图合成出高分辨率(1024×2048 px)图像。文献[13]在前者的全局生成和多尺度鉴别器的基础上,提出一个基于GAN的基于语义的图像压缩系统,实现极高压缩率下的全分

分辨率图像的压缩与重建,且在语义图的辅助下,保留感兴趣的主体而舍弃无关背景,被舍弃部分由生成器根据学习到的数据分布进行推断产生,而只对保留区域进行编码传输,进一步提高压缩率。

截止到目前,基于深度学习的图像语义编码技术均是基于高性能计算机来进行训练和推断,得到的模型通常需要大量的计算资源来驱动,无法直接应用于边缘设备,且并未考虑信道传输的问题,而联合信源信道编码模型DeepJSCC^[14]及其改进模型DeepJSCC-f^[15]则只能对某一信噪比条件下的单一信道进行学习,面对训练信噪比与测试信噪比不一致时性能下降,也无法直接应用于真实的物理无线环境,致使语义通信技术目前也存在于仿真环境中。本文将融合语义提取压缩与传统成熟信道编码技术,在保证高压缩比的同时,使数据在实际环境进行可靠的传输。

2 系统模型

结合传统图像压缩编码技术和对抗生成网络形成了基于语义的图像编码系统框架,如图1所示,主要包括编码器,量化器和解码器和鉴别器。

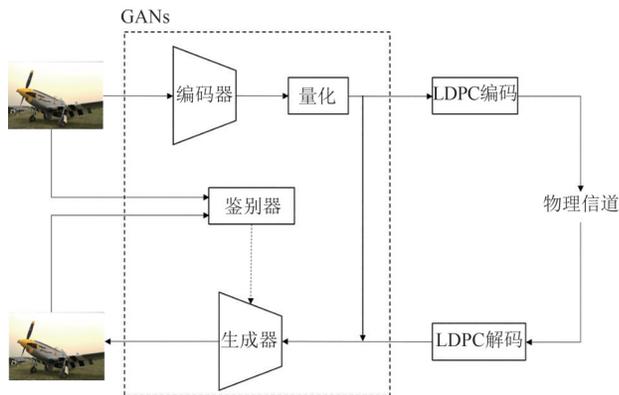


图1 系统框架

Fig. 1 System architecture

2.1 基本过程

在发射端,首先将设备捕捉到的图像 x 进行语义编码,通过卷积网络提取得到其对应的语义,即原图的特征,后经过量化得到发送数据 y ,该过程可表示为:

$$y = q(E(x; \theta)) \quad (1)$$

其中, E 和 q 分别表示信源编码过程及量化过程, θ 表示编码器网络中的可训练参数。

发送数据经过无线信道后,接收到的数据 \hat{y} 表示为:

$$\hat{y} = \mathbf{h} \cdot y + \mathbf{n} \quad (2)$$

其中 \mathbf{h} 表示信道矩阵,而 \mathbf{n} 表示加性噪声,在语义通信相关文献中,通常将其视为语义噪声,语义判决方式由接收双方的共享语义库所定义,对于本文所提模型的对抗生成网络而言,可视为生成对抗网络的生成器的泛化能力,即生成器面对新数据样本的适应能力。

而接收端的解码器由对抗生成网络模型训练而得。生成对抗网络^[16]模型(GANs)实际上为一个通过生成器与鉴别器之间利用极小化极大算法进行博弈来学习数据分布特征的模型。给定一个数据分布为 p_x 的数据集 \mathcal{X} ,以及从给定数据分布 p_z 中采样的一系列样本值 z ,GANs通过生成器学习从已知分布 p_z 到未知分布 p_x 的映射关系,记作 $G(z; \theta_g)$,同时,鉴别器用于鉴定数据样本来自于真实数据或是生成数据,记作 $D(x; \theta_d)$,其中 θ_g, θ_d 分别表示生成器与鉴别器中的可训练参数。理论上GANs同时训练生成器和鉴别器,两者收敛于纳什均衡点,但实际使用中往往难以实现,通常利用固定其中一个而训练另外一个的方法,如先对鉴别器进行训练,然后利用优化后的鉴别器指导生成器学习,再根据生成器生成的数据输入至鉴别器进行训练,循环反复达到局部最优即认为达到纳什均衡。其目标函数可以表示为:

$$\mathcal{L}_{\text{GAN}} = \min_G \max_D \mathbb{E}[f(D(x))] + \mathbb{E}[g(D(G(z)))] \quad (3)$$

其中, f 和 g 为散度函数,GANs的原始结构中 f 和 g 采用了熵的思想,即 $f(x) = \log(x), g(x) = \log(1-x)$,此时GANs生成器实际上在衡量真实数据与生成数据分布之间的JS散度,当两个数据分布不存在重叠时,JS散度恒为常数,进而梯度消失,导致训练无法继续进行。

最后,将已训练好的生成器置于接收端,根据接收到的语义对原图像进行重建,得到生成图像 \hat{x} :

$$\hat{x} = G(\hat{y}; \varphi) \quad (4)$$

其中, φ 表示生成器中的可训练参数。

2.2 LDPC码

由于无线信道的复杂多变,简单模型难以实现对物理信道所有特征进行学习,往往需要大规模网络才能够对信道进行模拟,这对于边缘设备无论是

存储空间还是计算能力都是一个极大的挑战,尽管目前已有许多针对单一信道模型如高斯白噪声信道,瑞利衰落信道等进行仿真,但尚未得到实际验证,需要进行实时信道估计,切换到目前信道所对应的模型再进行发送,随着网络规模的增大,切换开销也不断增加,无疑增加了设备负担。因此本文在信道编码部分采用目前较为成熟的低密度奇偶校验码(LDPC码)。

对于一个稀疏非系统的校验矩阵 \mathbf{H} ,可以据此定义一个 (n, j, k) 的LDPC编码规则,其中, n, j, k 分别表示校验矩阵的列数,列重和行重。校验矩阵经初等变换后得到其对应的生成矩阵 \mathbf{G} ,对于消息队列 m ,LDPC码的编码过程可表示为:

$$\mathbf{M} = m \cdot \mathbf{G} \quad (5)$$

译码规则采用消息传递算法,其中消息的传递形式的对数似然比(Log-Likelihood Ratio, LLR),对于数据中每一个比特,其最大似然比定义为:

$$r = \ln \frac{\Pr(\hat{y}|\mathbf{h}, b = 0)}{\Pr(\hat{y}|\mathbf{h}, b = 1)} \quad (6)$$

译码规则为:

$$\hat{b} = \begin{cases} 0, & r \geq 0 \\ 1, & r < 0 \end{cases} \quad (7)$$

2.3 图像质量评价指标

对于图像质量评价指标,通常可分为有参考与无参考两大类,而本文主要使用有参考的评价指标:PSNR与MS-SSIM。对于给定的两幅图像 I 和 \hat{I} ,其尺寸为 $h \times w \times c$,其中 h 为高度, w 为宽度, c 为通道数,对于常见的RGB图像,通道数为3。而本文指标计算过程中的常数均采用默认值。

PSNR是基于均方误差(MSE)的一种衡量两张图片之间单个像素之间差异的指标,其中MSE定义为:

$$\text{MSE} = \frac{1}{hwc} \sum_{k=0}^c \sum_{j=0}^w \sum_{i=0}^h (I(i, j, k) - \hat{I}(i, j, k))^2 \quad (8)$$

则峰值信噪比计算公式为:

$$\text{PSNR} = 10 \cdot \lg \left(\frac{\text{MAX}^2}{\text{MSE}} \right) (\text{dB}) \quad (9)$$

其中,MAX表示图像像素范围的最大值,对于本文所使用的图像均为位深24 bit的RGB图像,即每个颜色通道需要用8 bit来表示,像素值的最大值为 $2^8 - 1 = 255$ 。

MS-SSIM与SSIM相比于PSNR,更注重人类视觉中对图像中不同信息的敏感度,主要从亮度、对比度和结构三方面来衡量两张图片的相似性,先对SSIM做出:

$$L(I, \hat{I}) = \frac{2\mu_I \mu_{\hat{I}} + c_1}{\mu_I^2 + \mu_{\hat{I}}^2 + c_1}$$

$$C(I, \hat{I}) = \frac{2\sigma_I \sigma_{\hat{I}} + c_2}{\sigma_I^2 + \sigma_{\hat{I}}^2 + c_2}$$

$$S(I, \hat{I}) = \frac{\sigma_{\hat{I}I} + c_3}{\sigma_I \sigma_{\hat{I}} + c_3}$$

$$\text{SSIM} = L^\alpha(I, \hat{I}) C^\beta(I, \hat{I}) S^\gamma(I, \hat{I}) \quad (10)$$

其中, $L(I, \hat{I}), C(I, \hat{I}), S(I, \hat{I})$ 分别表示两张图片之间的亮度差异,对比度差异与结构差异, α, β, γ 为常数,表示三者相互权重,通常默认取1,而 μ 表示图像的均值, σ 表示图像的标准差,存在两个下标时表示协方差, c_1, c_2, c_3 为常数, L 为像素值最大值, k_1, k_2 为常数,默认值分别为0.01和0.03,对应计算方法分别为:

$$\begin{aligned} \mu_I &= \frac{1}{hwc} \sum_{k=0}^c \sum_{j=0}^w \sum_{i=0}^h I(i, j, k) \\ \sigma_I &= \sqrt{\frac{1}{hwc-1} \sum_{k=0}^c \sum_{j=0}^w \sum_{i=0}^h (I(i, j, k) - \mu_I)^2} \\ \sigma_{\hat{I}I} &= \frac{1}{hwc-1} \sum_{k=0}^c \sum_{j=0}^w \sum_{i=0}^h (I(i, j, k) - \mu_I) (\hat{I}(i, j, k) - \mu_{\hat{I}}) \\ c_1 &= (k_1 L)^2 \quad c_2 = (k_2 L)^2 \end{aligned} \quad (11)$$

而MS-SSIM即多尺度下的SSIM,计算公式如下:

$$\text{MS-SSIM} = L^{\alpha_M}(I, \hat{I}) \prod_{i=1}^M [C^{\beta_i}(I, \hat{I}) S^{\gamma_i}(I, \hat{I})] \quad (12)$$

其中 M 指的是尺度迭代次数,默认值为5,此时 $\beta=\gamma=\{0.0448, 0.2856, 0.3001, 0.2363, 0.1333\}$ 。训练模型过程中同时采用这两个衡量指标,可使模型生成的图片无论是像素差异还是观感差异都能保持与原图相似。

3 低比特率图像语义编码方法

3.1 对抗生成网络训练

由2.1中所述,原始对抗生成网络描述的是生成数据分布于原始数据之间的分布,但JS散度并不能很好的表述两个不重叠分布之间的散度。原因

在于图像在高维空间中一般呈低维的流形状态,即使两个图像的数据分布在低维空间重叠,但当映射到高维空间中进行拟合时也容易被忽略,所以其JS散度大部分情况都为常数,导致后续发生梯度弥散的问题。

本文为提高GANs训练的稳定性,利用最小二乘对抗生成网络(Least Squares Generative Adversarial Networks, LS-GAN)模型,该模型采用最小二乘损失的思想,即令式(3)中 $f(x) = (x - 1)^2$, $g(x) = x^2$,此时生成器的损失函数可改写为皮尔森卡方散度形式,即:

$$\text{Loss}(G) = \frac{1}{2} \chi_{\text{Pearson}}^2(p_x + p_z \| 2p_z) \quad (13)$$

可以发现,尽管LS-GAN并未真正解决当两个数据不重叠时所造成的梯度弥散问题,但LS-GAN能生成更加高质量的图像,原因在于上述损失函数使生成器会对一些远离决策边界的样本进行惩罚,当使用传统交叉熵作为损失函数时,只关注样本数据是否被成功分类为真实数据或是生成数据,导致被正确分类后就无法提供有效的梯度,进而产生梯度弥散现象,而最小二乘法将样本数据到决策边界的距离也作为梯度信息之一,引导生成器生成更多在决策边界附近的样本数据,进而提高生成数据的质量。

3.2 量化器设计

在量化过程中,需要将特征值映射到给定的值,如距离特征值最近的整数,使每个特征值能够用有限比特数进行表示。量化过程在传统数字编码中容易实现,然而,基于深度学习的编码器对特征值进行量化时往往会导致模型的不可微甚至不连续,导致后续迭代优化过程,反向传播及梯度下降算法无法直接应用于模型。

为了维持神经网络的可导性,我们将量化过程分为两部分,分别对应训练过程中的前向推断和反向传播过程。假设给定一个长度为 L 的量化集 $\mathcal{C} = \{c_1, c_2, c_3, \dots, c_L\} \subset \mathbf{R}$,则在前向推断过程中,传输数据样本 z 的量化过程可表示为:

$$\hat{z}_i = Q(z_i) := \arg \min_j \|z_i - c_j\| \quad (14)$$

而在反向传播过程中,则引入了指数归一化(softmax)的思想以保留其可导性,且指数求导简单,确保鉴别器中的损失函数值能够正常通过量化器,实现对编码器的优化。

$$\tilde{z} = \sum_{j=1}^L \frac{\exp(-\sigma \|z_i - c_j\|)}{\sum_{l=1}^L \exp(-\sigma \|z_i - c_l\|)} c_j \quad (15)$$

3.3 优化目标设计

为了实现更高压缩比,需要考虑熵编码部分 $H(y)$,以表示对 \hat{w} 进行编码所需要的平均比特数,即越少的比特数开销意味着更高的压缩比。然而,若是一味追求图像的高压缩比,将容易导致训练模型生成纯黑的图像,即所有像素均被压缩为0,此时比特开销为0。为了避免这种情况的发生,我们将引入一个超参数 β 来改变熵编码部分的权重,并引入原图与生成图之间的失真度来引导训练模型能够在高压缩率与失真率之间找到一个平衡,可以表示为

$$\lambda \mathbb{E}[d(x, \hat{x})] + \beta H(y) \quad (16)$$

其中 $d(x, \hat{x})$ 表示原图与生成图之间的失真函数, λ 和 β 表示失真函数项与熵编码项的权重, $\beta \in \{0, 1\}$,当 $\beta=1$ 时,模型将熵编码算法加入到所提算法后面,以提供更高的压缩比率,当 $\beta=0$ 时,模型将不适用熵编码算法,利用最多的比特数来表示图片,以保留更多细节。此外,根据熵编码部分的定义,其大小受到特征向量的维度与量化精确度的限制,可以推出其上界值,即

$$H(y) \leq \dim(y) \log_2(L) \quad (17)$$

其中, $\dim(\cdot)$ 表示特征值的维度,当且仅当图片中的所有像素两两独立等号成立。实际上,最终模型所得到的 $H(y)$ 将小于该上界值,因为图片中的像素之间存在一定的依赖关系。而根据这个上界,我们可以通过改变编码器与量化器的结构以获得不同的特征值维度与量化维度,从而实现了对图像压缩率的控制,保证本文所提算法收敛,且存在一个性能下限。

综上,整个模型的目标函数为:

$$\text{Loss} = \min_{E,G} \max_D \mathbb{E}[f(D(x))] + \mathbb{E}[g(D(G(z)))] + \lambda [d(x, \hat{x})] + \beta H(y) \quad (18)$$

3.4 多尺度鉴别器设计

对于对抗网络中的鉴别器,我们希望其能够准确的求出生成图像与原图之间的差异,尤其是在高分辨率图像的生成与鉴定中,鉴定器的感知尺寸需要同时兼顾全局特征与局部特征。大尺寸感知能够获得更高层次,抽象程度更高的全局特征,如语义信息,类别标签等,而小尺寸感知则作用于底层

次特征甚至是像素级之间的差异,保证图片局部细节保持低失真度。大尺寸感知往往需要更深的网络或是大尺寸卷积核来作为支撑,容易导致神经网络尺寸过大而难以训练甚至产生过拟合情况,对于性能有限的边缘设备来说也是巨大的能耗开销,而小尺寸感知可以节省训练开销,但同时也会忽略图片的上下文信息或是语义特征,限制模型实现更高压缩率。

为了获取合适的感知野尺寸,我们采用一种多尺度鉴别器结构,来同时兼顾生成图与原图之间的全局差异与局部差异。该鉴别器由多个独立的鉴别器构成,每个鉴别器分别衡量不同尺度下的特征差异。在实际鉴别过程中,第一个鉴别器直接作用于原尺寸的原图与生成图,以衡量两者之间的像素级特征差异,而后续对原图与生成图自称的图像对进行下采样,下采样因子为 2,即可得到长宽均为原图的一半的新图像对,其中每个像素均对应原图中的 4 个像素,将其输入到第二个鉴别器中进行鉴别,以此类推,每次利用一个独立的鉴别器对原图与生成图的下采样结果进行鉴别。本文为了简化模型结构,使用 3 个鉴别器,分别记作 D_1, D_2, D_3 ,各自拥有独立的结构与训练参数,容易得出, D_3 能够同时对原图中的 16 个像素进行鉴别,拥有最大的感知范围,可以引导模型生成与原图全局特征一致的图

片,而 D_1 引导模型生成与原图相似的细节特征。因此,在该多尺度鉴别器的指导下,系统对抗生成网络的目标函数可以修改为:

$$\min_G \max_{D_1, D_2, D_3} \sum_{k=1,2,3} \mathcal{L}_{GAN}(G, D_k) \quad (19)$$

则优化目标函数(公式(18))也可改写为:

$$\min_{E, G} \max_{D_1, D_2, D_3} \sum_{k=1,2,3} \left[\mathbb{E} \left[f(D_k(x)) \right] + \mathbb{E} \left[g(D_k(G(z))) \right] \right] + \lambda \left[d(x, \hat{x}) \right] + \beta H(y) \quad (20)$$

4 仿真结果

对于训练模型所需的数据集,我们采用 FLAME 数据集^[17],一个基于航拍的森林火灾检测图像集,由北亚利桑那大学搜集并公开,包含 254×254 px 与 3480×2160 px 两种分辨率的森林火灾图片,适用于不同训练场景。考虑到本文应用场景为森林火灾监测,传输图像的功能主要是火源定位,物人捕捉等,实际设备为低成本低性能的传感器,应用高分辨率图像的性价比不高,因此本文采用分辨率为 254×254 的相关数据子集,在数据预处理阶段将其重新采样成 256×256 px,供后续训练使用。编码器与生成器的具体架构如图 2 所示,主要由卷积块(conv)和残差网络块(Residual block)组成,每个卷积块包含一个卷积层、实例归一化层(Instance Normalization, IN)和 ReLU 激活层。其中残差网络块包含卷积层和批归一化层

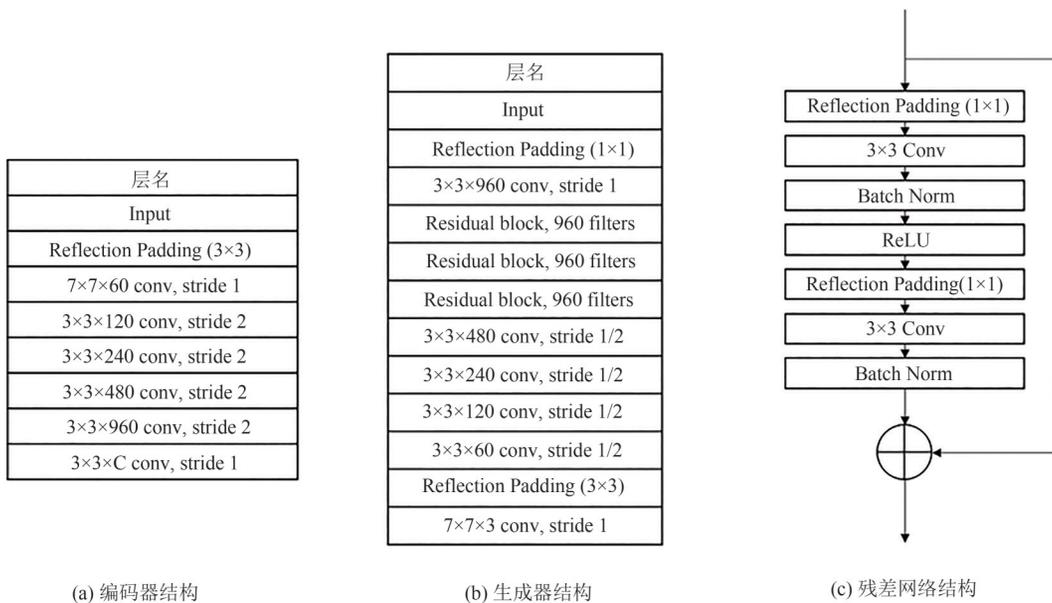


图 2 具体网络结构

Fig. 2 Network architecture

(Batch Normalization, BN)。而IN相比于BN更适用于注重单个像素点的场景,可辅助模型生成失真度更低的图片。图中编码器输出层的通道数 C 为系统结构的瓶颈层,可用于控制语义编码的压缩率。

为了获取式(17)中所讨论的熵编码的上界,不妨将熵编码项的权重 β 设为0,量化集长度为5,取 $C = \{-2, -1, 0, 1, 2\}$,则表示压缩后图像所需的比特数的上界为 $\dim(y) \cdot \log_2 5$ 。而根据编码器的结构所示,语义特征 y 的高和宽均是原图像 x 的1/16,通道数是原图的 $C/3$ 倍。对于失真函数,我们采用MSE,权重 $\lambda=10$ 。后续还可添加特征匹配损失函数和VGG损失函数,进一步提高生成图片的质量。瓶颈层数 C 分别取4,8进行仿真,并与BPG图像编码技术进行对比。

从图3可以看到,本文所提出的基于对抗生成网络的图像压缩模型明显优于BPG技术,在相同比特每像素(bit per pixel, bpp)指标下获得更高的峰值信噪比,基于GANs的压缩模型的PSNR基本可以达到28 dB,对于图片中雪地与树叶、火花之间边缘的细节也能够被较好的保留,而BPG在低bpp情况下边缘细节出现不同程度的模糊,甚至出现光晕现象。而当利用BPG压缩格式恢复出足够清晰的图片,即PSNR达到30 dB时,平均bpp为0.22,比GANs产生相近指标图片多耗费接近3倍比特开销。而随着瓶颈层 C 的增加,编码语义所需要比特开销增加,生成的图片也更加真实。当 β 不为零时,即将熵编码部

分引入到所提算法中,本文采用算术编码对瓶颈层深度为4和8所对应网络输出的语义特征进行编码,编码后的平均bpp为0.032,0.052,压缩比率分别提升11.1%与27.8%。而同样压缩比率的BPG文件经熵编码后所提升的压缩比率分别为0.042和0.079。值得注意的是,随着 β 的增大,训练结果并得到明显优化, β 对压缩率的影响仅取决于其值是否非零,原因是信息熵的值是由数据集的熵而决定的,且本文所用的算术编码为无损编码,无法对网络的更新提供有效信息,与网络结构无关。

图4显示了在不同信噪比条件的加性高斯白噪声信道和瑞利衰落信道传输过程中PSNR于MS-SSIM的变化过程。选取瓶颈层为8时所对应模型进行仿真。从峰值信噪比PSNR和MS-SSIM的指标变化可以看到,对于不同的信噪比,得益于GANs的泛化能力,使其能够有效应对噪声干扰,接收数据的部分差错尽管会影响生成器生成正确且清晰的图片,但至少能够保证二进制数据到图片的转换,而BPG等传统图像压缩技术过度依赖其固定的编解码规则,一旦数据在传输过程中遭到破坏,尤其是用于定义数据格式的头部数据,则会导致整个文件无法被正确解码恢复图像,仿真过程中默认定义格式的数据未受到破坏。

此外,信道编码中的LDPC码则保证了在不同信噪比下信息仍能成功传输,而不占据训练模型资源。对于图片等长码数据,性能更加,高码率下的LDPC相比于低码率,在低信噪比情况下也能维持保证传输质量。

对于本文所提出的语义压缩方法,其训练时间与部署时间会根据不同硬件性能而有所不同,难以将其统一到某个衡量指标上,而实际应用过程中,工作效率应更受关注,因此本文选取编解码速度作为两种方法的比较指标。表1中展示了本文所提模型与BPG之间的编解码所需总时长,编解码图片像素为 256×256 px,且模型测试过程中已禁用GPU与并行计算,两者均使用CPU资源。可以看到,BPG编码所需时间较长,而解码相对快速,平均编解码总时长达0.245 s,本文所提模型则相反,反卷积运算较慢而使解码速度变慢,平均总时长为0.180 s,快于BPG,也从侧面反映出本文模型工作所需计算资源略小于BPG压缩格式,更适用于传感器等边缘设备。

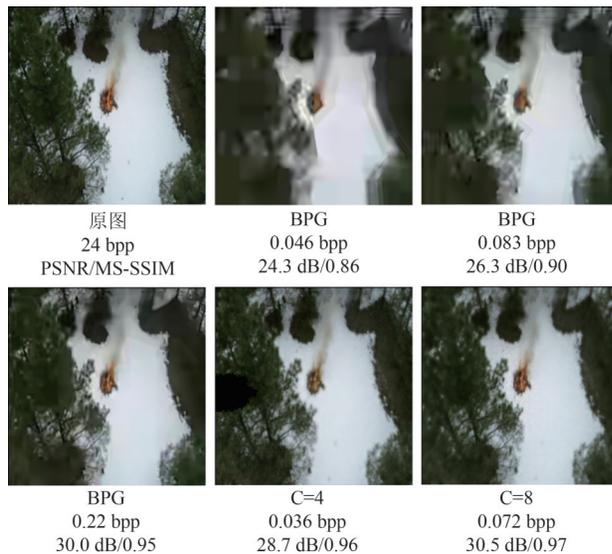


图3 本文模型与BPG对比

Fig. 3 Ours vs. BPG

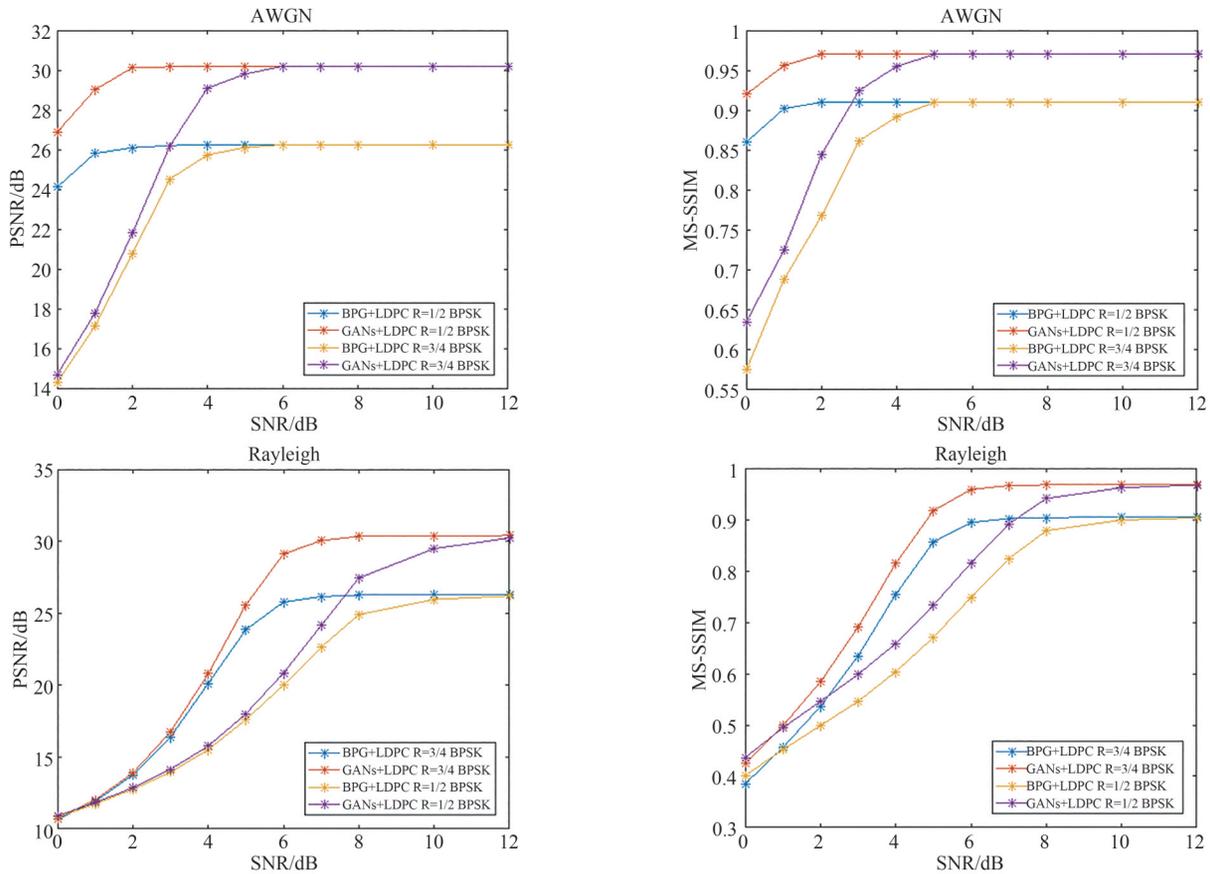


图4 不同信道下图像传输质量变化

Fig. 4 PSNR and MS-SSIM variation in AWGN and Rayleigh channel

表1 两种压缩方式时长对比

Tab. 1 Compression time comparison

时长	Ours	BPG
编码时长/s	0.024	0.193
解码时长/s	0.156	0.052

5 结论

本文针对性能有限的传感器节点或边缘设备中图像传输问题进行研究,提了一种基于语义通信的图像压缩方法,利用对抗生成网络进行语义到图像的重建,在实现比基于熵编码的传统图像压缩编码技术更高压缩比的同时,提供清晰度更高的图片。试验结果表明,相比于BPG压缩格式,同等衡量指标下,平均能够节省60%比特开销,而在同等比特开销下,能够获得将近20%的性能提升,且得益于GANs的泛化能力,能够在发生语义差错的同时保持较好的性能表现,避免了传统编码技术中数据错误尤其

是帧头被破坏后导致的整帧解码失败的情况。

参考文献

[1] BELLARD F. BPG Image format[OL]. <https://bellard.org/bpg/>.

[2] ZHANG Ping, XU Wenjun, GAO Hui, et al. Toward wisdom-evolutionary and primitive-concise 6G: A new paradigm of semantic communication networks[J]. *Engineering*, 2022, 8: 60-73.

[3] 张平, 牛凯, 田辉, 等. 6G移动通信技术展望[J]. *通信学报*, 2019, 40(1): 141-148.

ZHANG Ping, NIU Kai, TIAN Hui, et al. Technology prospect of 6G mobile communications [J]. *Journal on Communications*, 2019, 40(1): 141-148. (in Chinese)

[4] 石光明, 肖泳, 李莹玉, 等. 面向万物智联的语义通信网络[J]. *物联网学报*, 2021, 5(2): 26-36.

SHI Guangming, XIAO Yong, LI Yingyu, et al. Semantic communication networking for the intelligence of everything [J]. *Chinese Journal on Internet of Things*, 2021, 5(2):

- 26-36. (in Chinese)
- [5] ZHONG Yixin. A theory of semantic information[J]. China Communications, 2017, 14(1): 1-17.
- [6] XIE Huiqiang, QIN Zhijin, LI G Y, et al. Deep learning enabled semantic communication systems[J]. IEEE Transactions on Signal Processing, 2021, 69: 2663-2675.
- [7] SHI Guangming, XIAO Yong, LI Yingyu, et al. From semantic communication to semantic-aware networking: Model, architecture, and open problems[J]. IEEE Communications Magazine, 2021, 59(8): 44-50.
- [8] LIEW Z Q, CHENG Yanyu, LIM W Y B, et al. Economics of semantic communication system in wireless powered Internet of Things[C]//ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing. Singapore, Singapore. IEEE, 2022: 8637-8641.
- [9] BALLE J, LAPARRA E, SIMONCELLI E. End-to-end optimized image compression [C]. 5th International Conference on Learning Representations (ICLR), 2017: 1-27.
- [10] MINNEN D, BALLE J, TODERICI G. Joint Auto-regressive and hierarchical priors for learned image compression [C]. Advances in Neural Information Processing Systems 31 (NeurIPS 2018), 2018: 10771-10780.
- [11] ISOLA P, ZHU Junyan, ZHOU Tinghui, et al. Image-to-image translation with conditional adversarial networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. IEEE, 2017: 5967-5976.
- [12] WANG Tingchun, LIU Mingyu, ZHU Junyan, et al. High-resolution image synthesis and semantic manipulation with conditional GANs [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. IEEE, 2018: 8798-8807.
- [13] AGUSTSSON E, TSCHANNEN M, MENTZER F, et al. Generative adversarial networks for extreme learned image compression [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South). IEEE, 2019: 221-231.
- [14] BOURTSOULATZE E, BURTH KURKA D, GÜNDÜZ D. Deep joint source-channel coding for wireless image transmission[J]. IEEE Transactions on Cognitive Communications and Networking, 2019, 5(3): 567-579.
- [15] KURKA D B, GÜNDÜZ D. DeepJSCC-f: Deep joint source-channel coding of images with feedback[J]. IEEE Journal on Selected Areas in Information Theory, 2020, 1(1): 178-193.
- [16] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]. Advances in Neural Information Processing Systems 27 (NIPS 2014), 2014: 2672-2680.
- [17] SHAMSOSHOARA A, AFGHAH F, RAZI A, et al. Aerial imagery pile burn detection using deep learning: The FLAME dataset[J]. Computer Networks, 2021, 193: 108001.

作者简介



何晨光 男, 1981年生, 黑龙江哈尔滨人。哈尔滨工业大学电子与信息工程学院副教授, 硕士生导师, 博士。主要研究方向为移动通信技术、自组织网络等。
E-mail: hechengguang@hit.edu.cn



黄声显 男, 1998年生, 广东中山人。哈尔滨工业大学电子与信息工程学院在读硕士, 主要研究方向为语义通信、边缘计算。
E-mail: 21s105173@stu.hit.edu.cn



陈舒怡(通讯作者) 女, 黑龙江哈尔滨人。哈尔滨工业大学电子与信息工程学院助理教授, 主要研究方向为超密集组网、智能电网、空天地一体化网络等。
E-mail: chenshuyi@hit.edu.cn



王喆 男, 1982年生, 黑龙江哈尔滨人。哈尔滨工业大学硕士, 主要研究方向为电子精密机械、标准化体系建设。
E-mail: wangzhe1982@hit.edu.cn