

COSMIC：通过扩散补偿高效压缩卫星图像

Ziyuan Zhang¹ Han Qiu^{1*} Maosen Zhang¹ Jun Liu¹ Bin Chen² Tianwei Zhang³ Hewu Li¹

¹ 清华大学，中国

² 哈尔滨工业大学（深圳），中国深圳

³ 南洋理工大学，新加坡

{子源 - z23, 张姆斯 24}@mails.tsinghua.edu.cn, {邱涵, 刘军丽}@tsinghua.edu.cn 陈斌 2021@hit.edu.cn, 张天伟@ntu.edu.sg, 李和武@cernet.edu.cn

摘要

随着太空卫星数量的迅速增加以及其能力的增强，卫星收集的地球观测图像数量超过了卫星与地面链路的传输限制。尽管现有的学习图像压缩解决方案通过使用复杂的编码器提取有价值的特征作为压缩，并使用解码器进行重建，实现了出色的性能，但由于当前卫星的嵌入式 GPU 计算能力和电源有限，仍然难以直接在轨道上部署这些复杂的编码器来压缩图像。在本文中，我们提出了 COSMIC，一种简单而有效的学习压缩解决方案，用于传输卫星图像。首先，我们在卫星上设计了一个轻量级编码器（即减少 2.6 至 5 倍的浮点运算次数），以实现高图像压缩率，节省卫星与地面链路。然后，对于地面上的重建，为了应对由于简化编码器导致的特征提取能力下降，我们提出了一种基于扩散的模型，用于在解码时补偿图像细节。我们的见解是，卫星的地球观测照片不仅仅是图像，实际上是具有文本到图像配对性质的多模态数据，因为它们是在收集丰富的传感器数据（例如坐标、时间戳等）时获得的，这些数据可以作为扩散生成的条件。大量实验表明，COSMIC 在感知和失真指标方面均优于最先进的基准。代码可在 <https://github.com/Joanna-0421/COSMIC> 公开获取。

1 引言

航天工业的复兴[19, 38]，加上发射火箭成本的降低[22]，使得纳米卫星的数量呈指数级增长，导致在轨收集的图像数量大幅增加。例如，哨兵-3任务每天在卫星上最多可以收集20 TB的原始数据（主要是地球观测图像）[20]。然而，卫星与地面站之间的数据传输能力有明确的上限[18, 45, 17]。这种卫星收集图像迅速增长与向地面传输能力有限的情况，要求在将图像传输回地球之前，在卫星上进行有效的图像压缩。

当前针对卫星图像的工业压缩解决方案依赖于 JPEG [48]、JPEG2000 [46] 或 CCSDS123 [29]（例如卫星 BiSAT-1 [56]）。在各种情况下，这些解决方案的表现都逊于各种学习式压缩方法[12]。现有的学习式图像压缩方法[39、36、53]使用复杂的编码器来提取有用的特征，然后使用解码器进行解压缩[32、54]。

* Corresponding author.

尽管我们注意到在学术界[16, 17, 6, 45]和工业界（例如卫星 Phi-Sat-1 [25]、巢湖一号[3]和森林一号[4]）中，在卫星上部署嵌入式 GPU 呈现出一种新颖且有前景的趋势，这为在卫星上使用学习型压缩器带来了潜在的机会。然而，由于卫星上的 GPU（例如在森林一号上的 NVIDIA Jetson Xavier NX [4]）计算能力和电源有限[1]，其复杂的编码器仍然过于复杂，难以直接采用现有的学习型压缩解决方案。我们有两个见解来填补上述差距。（1）我们首先在卫星上设计了一个轻量级编码器，优先考虑压缩比而非特征提取能力。（2）在地面接收端，我们通过解码时补偿图像内容来处理这种简单编码器的特征提取能力下降的问题。我们选择扩散作为补偿，因为其强大的生成能力，更重要的是，卫星的地球观测照片不仅是图像，还具有多模态特性，其中丰富的实时传感器信息是对应照片的描述。例如，在图 1 中，坐标（例如纬度和经度）表示图像的位置，其描述了图像的主要类别（例如海洋、城市等），而时间戳则描述了图像类似闪电般的白昼或黑夜。

在本文中，我们提出了一种通过扩散补偿来压缩卫星图像的新方法（COSMIC），这是一种针对卫星的新型学习式图像压缩方法。COSMIC 有两个关键组成部分，即（1）用于在卫星上进行压缩的轻量级编码器，以及（2）在地面上使用解码器和扩散模型进行复杂解压缩过程的足够数量的 GPU。首先，我们设计了一种轻量级卷积架构来提取局部特征，并应用卷积以获得全局特征的注意力图，从而在浮点运算次数（FLOPs）方面实现轻量级图像压缩编码器。请注意，轻量级编码器

编码人员通常提取较少的关键特征，这增加了解码器解构的难度。

压力。因此，我们的第二个组成部分是将传感器数据作为描述进行解压缩。

西恩，有两个部分，包括解码和.....

相应的解码器，更重要的是，一个补偿模型。受卫星图像多模态特性的启发，我们旨在构建文本到图像对（如图 1 所示的图像及其传感器信息），并将扩散作为补偿模型。

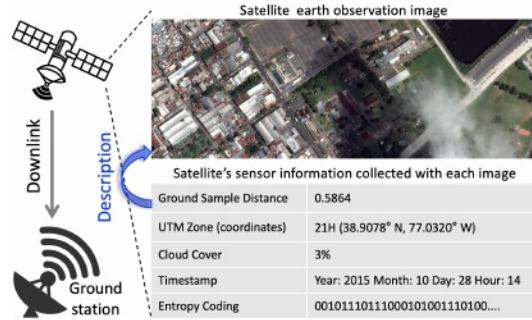


图 1：卫星地球观测图像的一个示例以及该图像的相应内容

我们将 COSMIC 与 6 个最先进的（SOTA）基准进行了比较，其中 3 个基于生成模型，同时考虑了失真和感知指标。此外，我们基于卫星图像构建了两个图像压缩测试集，考虑了卫星图像中的普通场景和独特拼贴场景。大量实验证明，COSMIC 将编码器的复杂度显著降低至少 2.6 至 5 倍的浮点运算次数（FLOPs），同时在几乎所有指标上的性能都优于基准。我们的贡献总结如下。

我们提出了一个新颖的想法，即在卫星上使用轻量级图像压缩编码器，并利用卫星图像的文本到图像配对特性在解压缩时进行补偿。

- 我们提出了一种基于稳定扩散的新型补偿模型，用于在利用卫星图像的独特传感器数据进行描述时对图像细节进行补偿。
- 我们详细分析了卫星图像的特征，并将其纳入训练和推理阶段。此外，考虑到诸如切片场景等典型的卫星图像传输任务，我们在卫星图像传输场景下构建了两个数据集。

2 背景

2.1 卫星上的地球观测任务

地球观测任务（例如NASA的陆地卫星计划[49]）涉及使用卫星照片来监测和收集数据，用于林业[5]、农业[14, 41]、土地退化[15]、土地利用和土地覆盖[40]、生物多样性[34]和水资源[35, 47]等任务。传统的地球观测任务依赖于一个管道，即卫星拍摄照片，然后将它们发送回地面站进行分析。最近，随着发射火箭和制造成本的迅速降低，

纳卫星[21, 6], 卫星上的照片数量迅速增加, 这给传输照片回地面带来了新的挑战。最近一种很有前景的方法是在卫星上部署嵌入式GPU, 以支持深度神经网络模型在传输前过滤无用数据或在卫星上进行部分处理任务。例如, 欧洲航天局的卫星Phi-Sat-1[25]首次在卫星上部署英特尔VPU, 以支持深度神经网络模型过滤无用照片(例如被云覆盖的照片), 这可以节省30%以上的传输量。OroraTech已经发射了带有NVIDIA Jetson Xavier NX的人工智能纳米卫星用于野火检测[4], 而Orbital Sidekick则使用NVIDIA Jetson AGX Xavier作为卫星边缘的AI引擎来检测天然气管道泄漏[2]。然而, 由于卫星的计算能力有限, 并且仅依靠阳光提供有限的电力(卫星上的GPU最多只能提供15瓦特[4]), 仍然需要传输一定数量的图像回地面。这使得卫星特定的图像压缩方法成为迫切需求。

2.2 神经图像压缩方法

与基于经典信息论的图像压缩方法相比, 已知的图像压缩方法实现了显著的率失真性能, 这些方法中的每一个都由一个编码器E、一个量化器Q和一个解码器D组成。编码器作为最关键的部分, 从图像中提取关键特征作为潜在表示。编码器提取的表示质量越高, 意味着压缩时内容损失越少, 在解压缩时更有可能重建出更高质量的图像。因此, 最先进的方法探索在编码器中引入更复杂的模块。[37]将变换器集成到卷积神经网络(CNN)编码器中, 并使用变换器-CNN混合模块来提取丰富的全局特征。其他方法旨在降低部署在诸如智能手机等边缘设备上的解码器的复杂性。例如, [54]采用浅层甚至线性解码变换来降低解码复杂性, 通过更强大的编码器网络和迭代编码进行补偿。

用于解压缩的生成模型。尽管基于变分自编码器(VAE)的方法已经取得了良好的性能, 但仅仅针对均方误差(MSE)进行优化可能会导致过度的图像平滑, 从而产生视觉失真。最近的工作[36, 53, 30, 24]将变分自编码器与生成模型(例如扩散)相结合, 以实现更好的视觉结果。[52]使用条件扩散模型作为图像压缩解码器, 从而改善了视觉结果。[30]和[24]将压缩任务和增强任务解耦, 将变分自编码器编解码器的输出发送到扩散模型以预测残差。

卫星图像压缩方法。有一些专门针对遥感图像的压缩方法[57, 23, 50]。[50]使用离散小波变换将图像特征分为高频特征和低频特征, 并设计一个频域编解码模块来保留高频信息, 从而提高压缩性能。[23]通过混合超先验网络探索局部和非局部冗余, 以提高熵模型估计的准确性。这些工作中很少有关关注机载部署的。[26]使用CAE模型提取图像特征并降低图像维度以实现压缩, 并将模型部署在VPU上。然而, 这种方法只考虑了图像维度的降低, 而没有考虑实际传输过程中的算术编码过程, 导致图像压缩率只能通过改变模型架构来调整。

卫星应用的局限性。上述大多数方法未考虑轻量级压缩编码器, 这使得它们在计算能力和电源受限的卫星嵌入式GPU上部署不切实际。在VPU上使用的方法不切实际, 因为图像压缩率与模型架构高度相关。此外, 它们都没有关注卫星地球观测图像的多模态特性, 也没有引入条件来进一步提高解压缩质量。

3 前言

3.1 问题表述

我们制定了学习式图像压缩的基本流程。编码器E使用非线性变换将输入图像 x 转换为潜在表示 y , 随后在学习的超先验 ζ 下通过量化Q进行离散化和熵编码。在随机高斯模型下, 给定超先验 $\zeta_i: p(\mathbf{y}_i | \zeta_i) = \mathcal{N}(\mu_i, \sigma_i^2)$, 每个离散码 \mathbf{y}_i 可以表示为均值为 μ_i 和方差为 σ_i^2 的高斯分布。解码器D重建 \hat{x}

将离散表示 \hat{y} 映射到图像 x 上。该模型可以通过损失函数（式 1）进行优化。

$$\mathcal{L}_{IC} = R + \lambda D = \mathbb{E} [-\log_2 p(\lfloor y \rfloor | \zeta) - \log_2 p(\zeta)] + \lambda \mathbb{E} [d(x, \hat{x})], \quad (1)$$

其中, R 是潜在离散编码的比特率, D 是原始图像和重建图像之间的失真（通过均方误差来衡量）, λ 控制着速率和失真之间的权衡。

3.2 扩散模型

扩散模型是一种生成模型, 能够通过多步迭代去噪从高斯噪声生成图像。这些模型包括两个马尔可夫过程。首先, 扩散过程逐渐向图像添加噪声, 直到图像被破坏并变成完全的高斯噪声。然后, 在反向阶段, 它学习将高斯噪声恢复为原始图像的过程。在推理阶段, 给定一个随机噪声样本 $x_T \sim \mathcal{N}(0, 1)$, 扩散模型可以通过 T 步去噪, 并逐渐生成一个逼真的图像 x_0 。在每一步 $t \in \{0, 1, \dots, T\}$, 中间变量 x_t 可以表示为 $x_t = 1 - \beta_t x_{t-1} + \beta_t \epsilon_t$, 其中 $\beta_t \in (0, 1)$ 是高斯分布的方差超参数, 并且满足 $\beta_1 < \beta_2 < \dots < \beta_T$; $\epsilon_t \sim \mathcal{N}(0, 1)$ 是步骤 t 的高斯噪声。在扩散模型中, 使用噪声预测网络 (ϵ_θ) 来预测步骤 t 的噪声, 并且 x_{t-1} 可以从 x_t 中得到。

扩散模型还能够理解给定条件（如文本和图像）的内容, 并生成与条件一致的图像。在这种情况下, 噪声预测网络将三个参数作为输入: 中间样本 x_t 、时间步长 t 以及给定条件 ζ 。为了保证在反向过程的第 t 步中, 由噪声预测网络预测的噪声与扩散过程中引入图像的噪声具有相同的分布, 扩散模型通常使用 L_2 来优化网络, 遵循 $\nabla_{\theta} \|\epsilon_t - \epsilon_\theta(x_t, t, \zeta)\|$ 。

稳定扩散[42]被提出以降低训练成本, 它在低维潜在空间中实现扩散过程, 同时保留原始像素空间中的高维信息以进行解码。在本文中, 我们的目标是利用稳定扩散强大的生成能力和其与给定条件保持一致的能力, 为图像压缩编码器丢失的信息提供补偿。

4 方法

4.1 COSMIC 框架

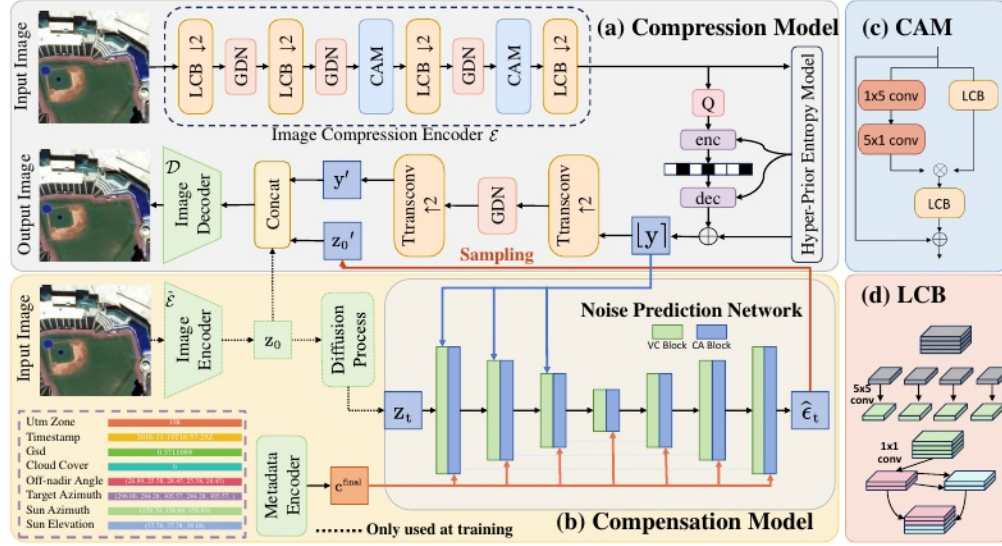
如图 2 所示, COSMIC 由两个组件组成: 一个压缩模块和一个补偿模块。为了适应卫星场景, 压缩模块包括一个部署在卫星上的轻量级图像压缩编码器 E 和一个熵模型（第 4.2 节）, 以及一个部署在地面站的图像压缩解码器 D 。为了弥补轻量级编码器导致的内容细节损失, 提出了一种补偿模块, 该模块完全部署在地面站。它有一个编码器 E ,

旨在提取补偿信息 z_0

来自原始图像, 在训练期间由解码器 D 接收, 作为对 E 提取的潜在表示 y' 的补偿（第 4.3 节）。在推理阶段, 噪声预测网络生成补偿信息 z_0' from 噪声以模拟 z_0 作为补偿（第 4.4 节）。

4.2 轻量级图像压缩编码器

为了使图像压缩编码器 E 在卫星上实用, 我们首先使其轻量化。主要思想是从浮点运算次数 (FLOPs) 的角度减少图像压缩所需的计算量。我们遵循了图像压缩编码器的经典架构[10], 该架构由步长为 2 的下采样卷积和交替排列的广义除法归一化 (GDN) [9] 组成 (图 2 (a))。由于 GDN 在通道数为 192 时对图像压缩的性能最佳[8], 卷积滤波器的增加会使卷积的计算量呈指数级增加。为了减少图像下采样所需的计算量, 我们提出了一个轻量级卷积块 (LCB), 如图 2 (d) 所示, 它使用深度卷积来替代普通卷积, 卷积核大小为 5×5 。为了在特征图的不同通道之间进行交互, 深度卷积之后跟着一个具有全通道数的 1×1 卷积。受[27]的启发, 该研究证明存在大量的



在推理阶段，训练好的扩散模型能够生成补偿信息 z_0' ，因此，我们不再需要 E_e 了。”
通过扩散模型得到的 $z_0^{\text{generated}}$ 取代了提取的 z_0 。
通过 E 来帮助图像解码器对图像进行解压缩。

4.4 用于损失补偿的条件扩散模型

如第1节所述，卫星收集的地球观测图像确实是多模态数据。在这里，我们考虑卫星图像 x_0 在传输到地面站时，包含一个离散图像编码 $\lfloor y \rfloor$ ，与表示为数值元数据 m 的传感器信息配对。对于 $m \in \mathbb{R}^M$ ，就像扩散处理时间步长 t 一样，我们使用正弦嵌入(E_{\sin})将它们编码为 $c_j \in \mathbb{R}^{1 \times d}$ ($j = 1, 2, \dots, M$)，其中 d 是片段嵌入的维度，因为我们将元数据嵌入连接在一起作为图像的描述。这个过程如公式3所示。

$$c^{\text{final}} = \text{MLP}(\text{concat}([E_{\sin}(m_1), \dots, E_{\sin}(m_M)])) \quad (3)$$

这最后的元数据条件将通过使用交叉注意力 (CA) 块纳入潜在表示中，以引导生成过程。

稳定扩散最初用于具有随机性的生成任务。在此，我们期望稳定扩散生成的图像细节不是由卫星图像编码器 E 提取的，并且仍然保留图像的整体结构。为了解决这个问题，我们将离散图像编码 $\lfloor y \rfloor$ 注入噪声预测网络的常规卷积 (VC) 块中，以提供结构信息并提高内容一致性，这可以描述为公式4。

$$f_i' = f_i + \text{projection}_i(\lfloor y \rfloor), \quad (4)$$

其中， f_i 是 U-Net 骨干网络的 i 个特征图， projection_i 是用于对齐 $\lfloor y \rfloor$ 和 f_i 之间维度的上采样卷积。以图像编码和元数据作为描述为指导，我们使用均方误差损失来最小化潜在空间中目标分布和学习分布之间的距离，如公式 5 所示。

$$\mathcal{L}_{\text{ldm}} = \mathbb{E}_{t, z_0, \epsilon \sim \mathcal{N}(0, 1)} [\|\epsilon_t - \epsilon_{\theta}(z_t, t, \lfloor y \rfloor, c^{\text{final}})\|^2] \quad (5)$$

5 实验

5.1 安装/设置

数据集。我们使用《世界地图》(fMoW) [13]作为我们的训练数据和测试数据，该数据集有 62 个类别，并且每张图像都与不同类型的元数据特征相关联。对于训练数据，我们随机裁剪图像到 256×256 像素的分辨率。对于由卫星传感器收集作为元数据的信息，我们选择由 fMoW 提供的 UTM 区域、时间戳、地面采样距离 (GSD)、云量、偏离天顶角、目标方位角、太阳方位角和太阳仰角。

对于测试数据，我们构建了两个具有不同分辨率的测试集。其中一个来自 fMoW 测试集，为了确保类别有全面的代表性，我们从每个类别中随机选择一张图像，将其裁剪为 256×256 的分辨率，并将其用作标准测试集。对于另一个测试集，我们考虑了卫星的实际场景来构建一个切片测试集。由于卫星捕获的图像是一个大的地理区域，卫星上的计算资源有限，无法直接处理大尺寸图像，因此应先将图像分割成较小的子图像，这个过程被称为切片 [31, 51, 17, 11]。在本文中，对于大小为 2306×2306 的图像，我们首先将每张图像分割成 81 个大小为 256×256 的小块，然后对每个小块单独进行压缩和解压缩，如图 5 (a) 所示。在获得所有解压缩的小块后，我们将它们重新组装成一张图像以进行进一步评估。

度量标准。我们遵循先前的研究 [39, 36, 53]，使用 4 个度量标准进行定量测量。对于失真比较，我们使用峰值信噪比 (PSNR) 和多尺度结构相似性指数 (MS-SSIM) 来验证像素保真度，并测量不同尺度的亮度、对比度和结构信息。对于感知比较，我们选择学习感知图像块相似性 (LPIPS) 和弗雷歇初始距离 (FID)。

模型训练细节。训练有两个阶段。首先，我们使用 L_{IC} 对图像压缩编码器 E 、图像编码器 E_{I} 和图像解码器 D 进行联合训练，训练 100 个周期，批次大小为 32。

其次，我们冻结在第一阶段训练的模型的参数，使用预训练的扩散模型作为噪声预测网络，并使用 λ 对其进行微调，进行10个epoch的训练，批次大小为4。所有训练实验均在 $10 \times \text{NVIDIA GeForce RTX 3090}$ 上进行，使用Adam优化器，学习率 $\text{lr} = 1 \times 10^{-4}$ ， $\lambda \in \{0.00067, 0.0013, 0.0026, 0.005\}$ 。在推理过程中，我们使用DDIM采样[43]，步长为25。

基准。我们考虑了6种基准，包括传统方法、基于变分自编码器（VAE）的方法以及基于生成模型的方法。Elic [28] 提出了一种多维熵估计模型，能够有效地降低比特率并提高编码性能。Hific [39] 更关注模型重建效果的感知，以获得视觉上令人愉悦的重建图像。基于Hific，COLIC [36] 在设计损失函数时考虑了图像的语义信息，并分别处理结构和纹理。CDC [53] 是首个将扩散模型用作图像压缩解码器的研究，在像素空间中执行扩散的逆过程以重建图像。HL_RS [50] 是一种图像压缩方法，尤其适用于遥感图像，它分别处理图像的高频部分和低频部分，以更好地保留遥感图像重要的高频特征。对于这5种基准，我们使用fMoW数据集重新训练它们的模型，以进行公平比较。此外，我们选择JPEG2000（卫星图像压缩的工业解决方案）[55] 进行比较。

5.2 与基准的比较

RD 性能。图3展示了在两个测试集上与基准的比较结果。图中的虚线表示基准，实线表示COSMIC。我们从两个角度展示了结果，即失真和感知。在所有传输速率下，COSMIC在LPIPS和FID方面都超过了基准。在低bpp下，COSMIC的MS-SSIM低于基准。这主要是因为随着bpp的降低，编码器提取的信息减少，并且在解压缩过程中，对基于扩散的生成的依赖更大（更多细节见第5.4节）。此外，由于轻量级编码器的特征提取能力下降，在低bpp下获得的特征不足以指导扩散过程生成高保真图像。随着bpp的增加，潜在编码包含更多的特征，导致COSMIC的MS-SSIM显著提高，展示了最先进的性能。请注意，在地球观测任务中，传感器数据默认传输到地面。此外，与图像相比，这些传感器数据的量可以忽略不计，因此在计算每像素比特数（bpp）时我们不考虑它们。我们在附录B中展示了更多关于更多指标和基准的结果，包括基于变分自编码器（VAE）的方法，COSMIC在所有6个感知指标上都达到了最先进的性能。

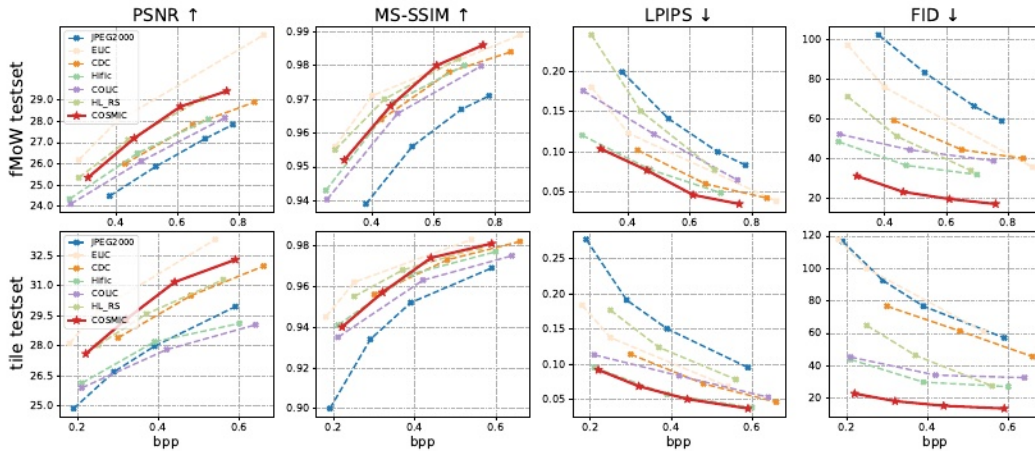


图3：在COSMIC和基准测试中比特率与不同指标之间的权衡。↑（↓）表示更好（更差）。第一行是针对fMoW测试集（图像大小 256×256 ）。第二行是通过比较拼接图像及其原始图像来针对瓷砖测试集。

视觉结果。图4展示了低比特率和高比特率的重建图像示例。为了公平比较，我们只展示了针对图像感知优化的结果。图5(b)展示了通过COSMIC和基准进行高分辨率图像重建的示例。由于分块和

在图像压缩中的拼接过程中，我们特别关注不同小块拼接形成较大图像时的接缝。JPEG2000 在图像接缝处表现出明显的不对齐。Hiflc 和 COLIC 无法准确恢复接缝的细节。例如，在橙色勾勒出的图片中，接缝处的汽车前灯被重建为一个红点。基于扩散的 CDC 重建在不同子图像之间也表现出一些颜色差异，并显示出明显的不对齐，例如红色勾勒出的图片中的接缝处的屋檐。与基准相比，COSMIC 在不同子图像之间的结构和颜色上保持了更高的相似度，从而带来了显著的视觉改进。



图 4: 解压后的 fMoW 图像（完整图像见补充材料）。1st 行: 低比特率下的比较，COSMIC 显示出更好的视觉效果。与 CDC 相比，COSMIC 在比特率更低的情况下仍能获得稍好一些的视觉重建效果。2nd 行: 高比特率下的比较。

5.3 编码器效率分析

我们在表 1 中从浮点运算次数（FLOPs）的角度对 COSMIC 的轻量级编码器进行了评估。与基准相比，COSMIC 的星上浮点运算次数（包括编码器和熵模型）显著减少了约 2.6 至 5 倍，同时在类似比特率的情况下，COSMIC 的整体性能仍优于基准。

表 1: 在图块测试集上将卫星上的浮点运算次数与基准进行比较。最佳性能以粗体突出显示。

方法	浮点运算次数 (十亿)	峰值信噪比 ↑	多尺度结构相似性 ↑	PSNR ↓	FID ↓	bpp ↓
美国疾病控制与预防中心 [53]	31.98	0.982	0.0462	45.49	0.66	
婴儿腹绞痛 [36]	26.4	29.04	0.975	0.0530	32.56	0.64
Hiflc [39]	26.4	29.11	0.977	0.0384	27.01	0.60
埃利克 [28]	21.78	33.31	0.983	0.0683	60.80	0.54
HL_RS [50]	11.87	31.30	0.979	0.0782	27.38	0.56
宇宙	4.9	32.11	0.980	0.0359	13.50	0.59

5.4 烧蚀研究

没有扩散补偿。为了展示扩散在图像重建过程中的补偿作用，我们移除了扩散补偿模块，结果如图6所示。我们移除补偿模块并重新训练模型以显示定量指标。结果表明，扩散模型在图像压缩过程中发挥着重要作用，以获得更好的感知指标。为了更清楚地展示扩散模型的补偿作用，我们直接移除扩散模型，仅使用轻量级编码器E的输出信息来重建图像。我们发现，在低比特率下，扩散补偿更为重要。由于轻量级编码器的特征提取能力不足，为了节省传输速率，许多图像内容细节被丢失，扩散需要根据编码器的有限输出以及元数据来重建大部分图像内容。随着比特率的增加，编码器提取的更多特征可以被保留，此时扩散只需要补偿编码器未能捕捉到的某些图像细节。视觉结果如图7所示。结果表明，使用扩散模型作为补偿是非常有用的，尤其是在比特率较低的情况下。

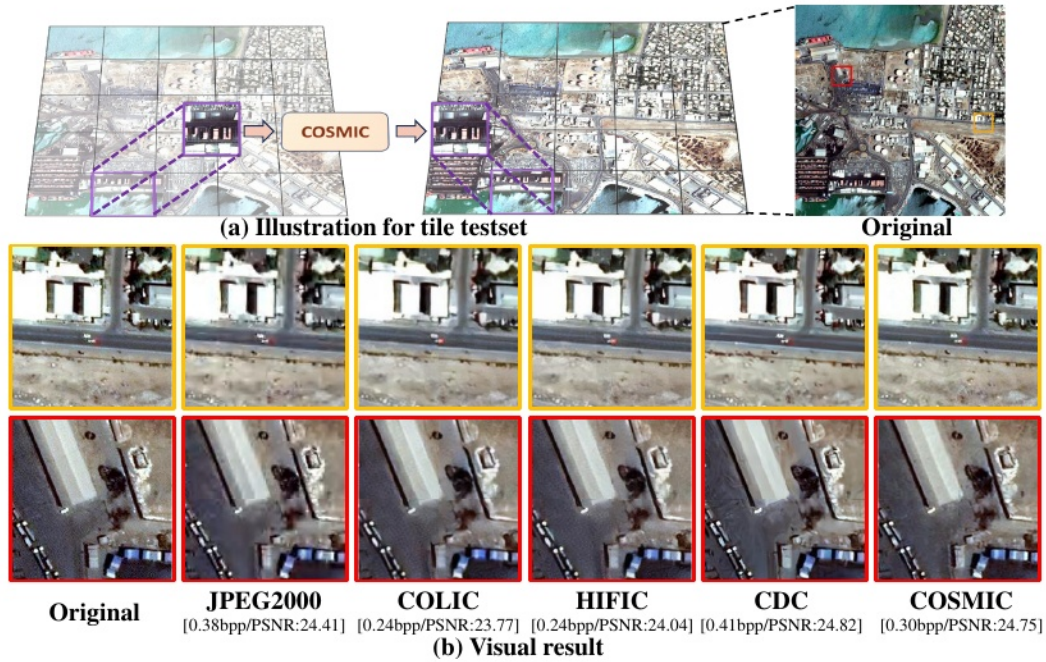


图5：（a）瓷砖测试集的图示。高分辨率图像被分割成许多小子图像（或补丁），每个子图像单独进行压缩。然后，重建的子图像被放回原始位置，并拼接在一起形成高分辨率重建图像。（b）在瓷砖测试集中，我们提供了一个拼接区域的详细视图（用橙色和红色勾勒出来）。COSMIC与基准的视觉比较显示，COSMIC在纹理对齐和颜色亮度一致性方面实现了最佳视觉效果。

表 2: 对图像分类模型的影响

课程	原始	JPEG2K [46]	埃利克 [28]	婴儿腹纹痛 [36]	HICC [39]	美国疾病控制与预防中心 [50]	宇宙
10	98.95%	-4.21%	-5.27%	-1.06%	-1.06%	-2.11%	-1.06%
15	97.92%	-4.17%	-3.84%	-0.70%	-0.70%	-1.40%	-0.70%
20	98.42%	-3.16%	-3.68%	-0.53%	-0.53%	-1.06%	-1.06%

没有 CAM 模块。为展示 CAM 模块能够捕获非局部信息，这有助于编码器 E 获得更高质量的表示。我们移除 CAM 模块，并将结果展示于图 6 中。结果表明 CAM 模块是有效的，并且能够实现更好的 RD 性能。

没有我。为了展示传感器数据能够引导扩散的产生，我们移除元数据编码器，仅使用图像编码进行扩散生成。请注意，这里主要有两种传感器数据，包括图像捕获期间使用的诸如非天顶角和目标方位等相机参数，以及关于云层覆盖、光照和地面样本距离等的的数据。图6表明，传感器数据有助于引导扩散在重建图像中的作用。

解码步骤的影响。我们进一步研究了在反向扩散过程中不同去噪步骤数对重建图像质量的影响，如图8所示。我们发现，随着去噪步骤数的增加，生成图像的感知指标逐渐改善，这与DDIM中对去噪步骤与FID分数之间关系的探索一致[43]。虽然感知结果有所改善，但像峰值信噪比这样的失真指标略有下降，展示了感知质量和失真之间的权衡。在实际部署和下游任务中，可以根据对失真和感知性能的不同强调来选择去噪步骤数，以适应实际应用。

5.5 压缩对下游任务的影响

为了确认 COSMIC 不会影响下游遥感任务，如图像分类，我们选择[7]中的图像分类任务来展示压缩所导致的效果，如表 2 所示。对相同的测试图像在低比特率级别进行压缩和解压缩，然后通过分类模型进行处理，以获得不同数量的准确率变化。

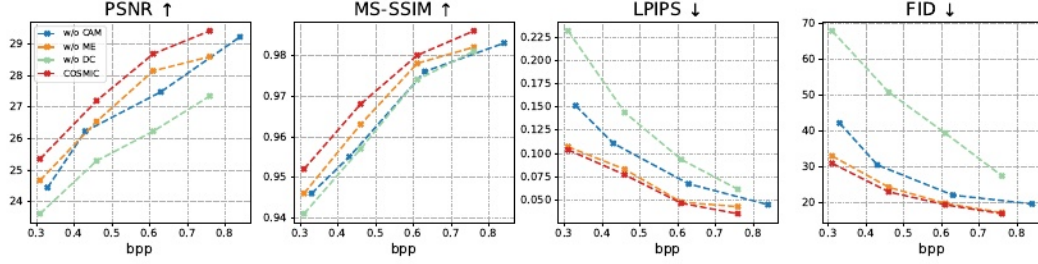


图 6: 使用 COSMIC 的不同变体的消融研究。“w/o DC”表示在解码期间未使用扩散补偿。“w/o ME”表示我们删除了元数据编码器。为了展示元数据编码器的效果，我们在图中用红色框标记了更细微的差异。

结果表明，JPEG2000 和 Elic 对精度的降低最为显著，而 COSMIC 在学习方法方面表现出色。



图 7: 对比度视觉结果。“w/ DC”表示在解码期间使用扩散进行补偿，而“w/o DC”则表示在解码期间未使用扩散补偿。

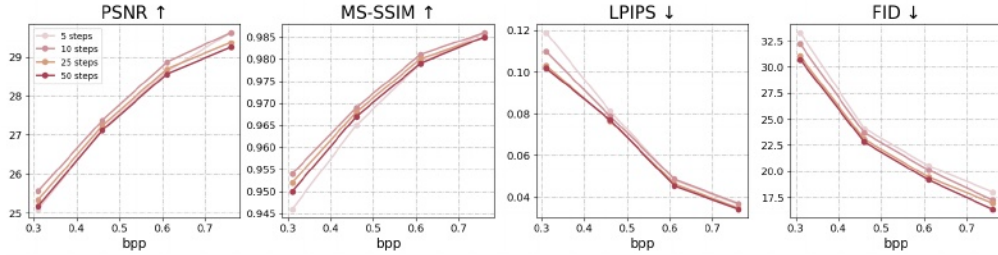


图 8: 不同解码步长的压缩性能。

6 结论

我们提出了 COSMIC，这是一种用于卫星地球观测任务图像压缩的新方法。首先，我们设计了一个轻量级编码器以适应卫星上有限的资源。然后，我们引入了一个条件潜在扩散模型，利用卫星的传感器数据作为指令来补偿由于轻量级编码器特征提取能力下降而缺失的细节。大量结果表明，COSMIC 不仅能为两个典型的卫星任务实现最先进的压缩性能，还能保证卫星图像下游任务的准确性。

局限性与未来工作。尽管 COSMIC 使用扩散来补偿编码器的限制，但在极低的 bpp（例如，小于 0.1bpp）下，潜在图像编码提供的信息可能不足以支持扩散生成高保真图像。我们认为主要原因是，我们只是对预训练的稳态扩散模型进行了微调，该模型对卫星图像缺乏足够的先验知识。作为一项前瞻性的解决方案，我们将专门为卫星图像训练一个扩散模型，或者使用历史卫星图像作为参考以进一步改进。

致谢。本研究得到了中国国家重点研发计划（2022YFB3105202）、中国自然科学基金（62106127、62301189、62132009）以及中国自然科学基金重点基金（62272266）的支持。

参考文献

- [1] 1. <https://www.nvidia.cn/autonomous-machines/embedded-systems/jetson-xavier-nx/>.
- [2] 2. <https://blogs.nvidia.com/blog/orbital-sidekick/>.
- [3] 3. <http://www.stardetect.cn/h-col-127.html>.
- [4] 4. Shoot from the stars: Startup provides early detection of wildfires from space. <https://blogs.nvidia.com/blog/ororatech-wildfires-from-space/>.
- [5] Frédéric Achard, Hans-Jürgen Stibig, Hugh D Eva, Erik J Lindquist, Alexandre Bouvet, Olivier Arino, and Philippe Mayaux. Estimating tropical deforestation from earth observation data. *Carbon Management*, 1(2):271–287, 2010.
- [6] Caleb Adams, Allen Spain, Jackson Parker, Matthew Hevert, James Roach, and David Cotten. Towards an integrated gpu accelerated soc as a flight computer for small satellites. In *2019 IEEE aerospace conference*, pages 1–7. IEEE, 2019.
- [7] Ali Bahri, Sina Ghofrani Majelan, Sina Mohammadi, Mehrdad Noori, and Karim Mohammadi. Remote sensing image classification via improved cross-entropy loss and transfer learning strategy based on deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 17(6):1087–1091, 2020.
- [8] Johannes Ballé. Efficient nonlinear transforms for lossy image compression. In *2018 Picture Coding Symposium (PCS)*, pages 248–252. IEEE, 2018.
- [9] Johannes Ballé, Valero Laparra, and Eero P Simoncelli. Density modeling of images using a generalized normalization transformation. *arXiv preprint arXiv:1511.06281*, 2015.
- [10] Johannes Ballé, David Minnen, Saurabh Singh, Sung Jin Hwang, and Nick Johnston. Variational image compression with a scale hyperprior. *arXiv preprint arXiv:1802.01436*, 2018.
- [11] Marc Bosch, Kevin Foster, Gordon Christie, Sean Wang, Gregory D Hager, and Myron Brown. Semantic stereo for incidental satellite images. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1524–1532. IEEE, 2019.
- [12] Zhengxue Cheng, Heming Sun, Masaru Takeuchi, and Jiro Katto. Learned image compression with discretized gaussian mixture likelihoods and attention modules. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7939–7948, 2020.
- [13] Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. Functional map of the world. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6172–6180, 2018.
- [14] Laura Crocetti, Matthias Forkel, Milan Fischer, František Jurečka, Aleš Grlić, Andreas Salentinig, Miroslav Trnka, Martha Anderson, Wai-Tim Ng, Žiga Kokalj, et al. Earth observation for agricultural drought monitoring in the pannonian basin (southeastern europe): current state and future directions. *Regional Environmental Change*, 20:1–17, 2020.
- [15] Rogier de Jong, Sytze de Bruin, Michael Schaepman, and David Dent. Quantitative mapping of global land degradation using earth observations. *International Journal of Remote Sensing*, 32 (21):6823–6853, 2011.
- [16] Bradley Denby and Brandon Lucia. Orbital edge computing: Nanosatellite constellations as a new class of computer system. In *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems*, pages 939–954, 2020.
- [17] Bradley Denby, Krishna Chintalapudi, Ranveer Chandra, Brandon Lucia, and Shadi Noghabi. Kodan: Addressing the computational bottleneck in space. In *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 3*, pages 392–403, 2023.

- [18] Kiruthika Devaraj, Ryan Kingsbury, Matt Ligon, Joseph Breu, Vivek Vittaldev, Bryan Klofas, Patrick Yeon, and Kyle Colton. Dove high speed downlink system. 2017.
- [19] Lauren Dreyer. Latest developments on spacex’ s falcon 1 and falcon 9 launch vehicles and dragon spacecraft. In *2009 IEEE Aerospace conference*, pages 1–15. IEEE, 2009.
- [20] Thomas Esch, Soner Üreyen, Julian Zeidler, Annekatrin Metz-Marconcini, Andreas Hirner, Hubert Asamer, Markus Tum, Martin Böttcher, S Kuchar, Vaclav Svaton, et al. Exploiting big earth data from space—first experiences with the timescan processing chain. *Big Earth Data*, 2 (1):36–55, 2018.
- [21] M Esposito, BC Dominguez, M Pastena, N Vercruyssen, SS Conticello, C van Dijk, PF Manzillo, and R Koeleman. Highly integration of hyperspectral, thermal and artificial intelligence for the esa phisat-1 mission. In *Proceedings of the International Astronautical Congress IAC, Washington, DC, USA*, pages 21–25, 2019.
- [22] Warren Frick and Carlos Niederstrasser. Small launch vehicles-a 2018 state of the industry survey. 2018.
- [23] Chuan Fu and Bo Du. Remote sensing image compression based on the multiple prior information. *Remote Sensing*, 15(8):2211, 2023.
- [24] Noor Fathima Khanum Mohamed Ghouse, Jens Petersen, Auke J Wiggers, Tianlin Xu, and Guillaume Sautiere. Neural image compression with a diffusion-based decoder. 2022.
- [25] Gianluca Giuffrida, Luca Fanucci, Gabriele Meoni, Matej Batič, Léonie Buckley, Aubrey Dunne, Chris Van Dijk, Marco Esposito, John Hefe, Nathan Vercruyssen, et al. The Ξ -sat-1 mission: The first on-board deep neural network demonstrator for satellite earth observation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2021.
- [26] Giorgia Guerrisi, Fabio Del Frate, and Giovanni Schiavon. Artificial intelligence based on-board image compression for the Ξ -sat-2 mission. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023.
- [27] Kai Han, Yunhe Wang, Qi Tian, Jianyuan Guo, Chunjing Xu, and Chang Xu. Ghostnet: More features from cheap operations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1580–1589, 2020.
- [28] Dailan He, Ziming Yang, Weikun Peng, Rui Ma, Hongwei Qin, and Yan Wang. Elic: Efficient learned image compression with unevenly grouped space-channel contextual adaptive coding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5718–5727, 2022.
- [29] Miguel Hernández-Cabronero, Aaron B Kiely, Matthew Klimesh, Ian Blanes, Jonathan Ligo, Enrico Magli, and Joan Serra-Sagrista. The ccsds 123.0-b-2 “low-complexity lossless and near-lossless multispectral and hyperspectral image compression” standard: A comprehensive review. *IEEE Geoscience and Remote Sensing Magazine*, 9(4):102–119, 2021.
- [30] Emiel Hoogeboom, Eirikur Agustsson, Fabian Mentzer, Luca Versari, George Toderici, and Lucas Theis. High-fidelity image compression with score-based generative models. *arXiv preprint arXiv:2305.18231*, 2023.
- [31] Bohao Huang, Daniel Reichman, Leslie M Collins, Kyle Bradbury, and Jordan M Malof. Tiling and stitching segmentation output for remote sensing: Basic challenges and recommendations. *arXiv preprint arXiv:1805.12219*, 2018.
- [32] Nick Johnston, Elad Eban, Ariel Gordon, and Johannes Ballé. Computationally efficient neural image compression. *arXiv preprint arXiv:1912.08771*, 2019.
- [33] Bahjat Kavar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, 35:23593–23606, 2022.

- [34] Claudia Kuenzer, Marco Ottinger, Martin Wegmann, Huadong Guo, Changlin Wang, Jianzhong Zhang, Stefan Dech, and Martin Wikelski. Earth observation satellite sensors for biodiversity monitoring: potentials and bottlenecks. *International Journal of Remote Sensing*, 35(18): 6599–6647, 2014.
- [35] Richard Lawford, Adrian Strauch, David Toll, Balazs Fekete, and Douglas Cripe. Earth observations for global water security. *Current Opinion in Environmental Sustainability*, 5(6): 633–643, 2013.
- [36] Meng Li, Shangyin Gao, Yihui Feng, Yibo Shi, and Jing Wang. Content-oriented learned image compression. In *European Conference on Computer Vision*, pages 632–647. Springer, 2022.
- [37] Jinming Liu, Heming Sun, and Jiro Katto. Learned image compression with mixed transformer-cnn architectures. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14388–14397, 2023.
- [38] Henry Martin, Conor Brown, Tristan Prejean, and Nathan Daniels. Bolstering mission success: Lessons learned for small satellite developers adhering to manned spaceflight requirements. 2018.
- [39] Fabian Mentzer, George D Toderici, Michael Tschannen, and Eirikur Agustsson. High-fidelity generative image compression. *Advances in Neural Information Processing Systems*, 33:11913– 11924, 2020.
- [40] Prem Chandra Pandey, Nikos Koutsias, George P Petropoulos, Prashant K Srivastava, and Eyal Ben Dor. Land use/land cover in view of earth observation: Data sources, input dimensions, and classifiers—a review of the state of the art. *Geocarto International*, 36(9):957–988, 2021.
- [41] George P Petropoulos, Prashant K Srivastava, Maria Piles, and Simon Pearson. Earth observation-based operational estimation of soil moisture and evapotranspiration for agricultural crops in support of sustainable water management. *Sustainability*, 10(1):181, 2018.
- [42] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.
- [43] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.
- [44] Yehui Tang, Kai Han, Jianyuan Guo, Chang Xu, Chao Xu, and Yunhe Wang. Ghostnetv2: enhance cheap operation with long-range attention. *Advances in Neural Information Processing Systems*, 35:9969–9982, 2022.
- [45] Bill Tao, Om Chabra, Ishani Janveja, Indranil Gupta, and Deepak Vasisht. Known knowns and unknowns: Near-realtime earth observation via query bifurcation in serval. In *21st USENIX Symposium on Networked Systems Design and Implementation (NSDI 24)*, pages 809–824, 2024.
- [46] David S Taubman, Michael W Marcellin, and Majid Rabbani. Jpeg2000: Image compression fundamentals, standards and practice. *Journal of Electronic Imaging*, 11(2): 286–287, 2002.
- [47] Soner Ureyen and Claudia Kuenzer. A review of earth observation-based analyses for major river basins. *Remote Sensing*, 11(24):2951, 2019.
- [48] Gregory K Wallace. The jpeg still picture compression standard. *Communications of the ACM*, 34(4):30–44, 1991.
- [49] Michael A Wulder, David P Roy, Volker C Radeloff, Thomas R Loveland, Martha C Anderson, David M Johnson, Sean Healey, Zhe Zhu, Theodore A Scambos, Nima Pahlevan, et al. Fifty years of landsat science and impacts. *Remote Sensing of Environment*, 280:113195, 2022.
- [50] Shao Xiang and Qiaokang Liang. Remote sensing image compression based on high-frequency and low-frequency components. *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

- [51] Chen Xu, Xiaoping Du, Zhenzhen Yan, and Xiangtao Fan. Cloud-based parallel tiling algorithm for large scale remote sensing datasets. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 4030–4033. IEEE, 2022.
- [52] Ruihan Yang and Stephan Mandt. Lossy image compression with conditional diffusion models. *arXiv preprint arXiv:2209.06950*, 2022.
- [53] Ruihan Yang and Stephan Mandt. Lossy image compression with conditional diffusion models. *Advances in Neural Information Processing Systems*, 36, 2024.
- [54] Yibo Yang and Stephan Mandt. Computationally-efficient neural image compression with shallow decoders. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 530–540, 2023.
- [55] Xiaqiong Yu, Jinxian Zhao, Tao Zhu, Qiang Lan, Lin Gao, and Lingzhi Fan. Analysis of JPEG2000 compression quality of optical satellite images. In *2022 2nd Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS)*, pages 500–503. IEEE, 2022.
- [56] Gokhan Yuksel, Onder Belce, and Hakan Urhan. Bilsat-1: First earth observation satellite of turkey-operations and lessons learned. In *Proceedings of 2nd International Conference on Recent Advances in Space Technologies, 2005. RAST 2005.*, pages 846–851. IEEE, 2005.
- [57] Lei Zhang, Xugang Hu, Tianpeng Pan, and Lili Zhang. Global priors with anchored-stripe attention and multiscale convolution for remote sensing images compression. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023.
- [58] Renjie Zou, Chunfeng Song, and Zhaoxiang Zhang. The devil is in the details: Window-based attention for image compression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 17492–17501, 2022.

A 关于 COSMIC 的详情

A.1 火车及推理的细节

我们在算法 1 和算法 2 中对两阶段训练和推理流程进行了更详细的解释。

Algorithm 1: Two-stage training of COSMIC.

input : A satellite image \mathbf{x} , referenced image \mathbf{x} (the same as satellite image, used only for training), metadata $\mathbf{m}_i \in \mathbb{R}^M$
output : Reconstructed image $\hat{\mathbf{x}}$
Parameters: Image encoder on satellite \mathcal{E} , image encoder of diffusion $\tilde{\mathcal{E}}$, image decoder \mathcal{D} , noise prediction network ϵ_θ

```

/* Training stage 1: Train  $\mathcal{E}$ ,  $\tilde{\mathcal{E}}$ , and  $\mathcal{D}$  together. */
1  $\mathbf{y} = \mathcal{E}(\mathbf{x})$ ;
2  $\zeta = \text{entropy\_model}(\mathbf{y})$ ;
3  $\mu_z, \sigma_z = \tilde{\mathcal{E}}(\mathbf{x}_0)$ ;
4  $\epsilon_z \sim \mathcal{N}(0, \mathbf{I})$ ;
5  $\mathbf{z}_0 = \mu_z + \sigma_z * \epsilon_z$ ;
6  $\hat{\mathbf{x}} = \mathcal{D}(\text{concat}(\mathbf{z}_0, \text{deconv}(\lfloor \mathbf{y} \rfloor)))$ ;
7 optimize the parameters of  $\mathcal{E}$ ,  $\tilde{\mathcal{E}}$ , and  $\mathcal{D}$  following:
    $\mathcal{L}_{\text{IC}} = \mathcal{R} + \lambda \mathcal{D} = \mathbb{E}[-\log_2 p(\lfloor \mathbf{y} \rfloor | \zeta) - \log_2 p(\zeta)] + \lambda \mathbb{E}[\text{d}(\mathbf{x}, \hat{\mathbf{x}})]$ 

/* Training stage 2: Freeze the parameters of  $\mathcal{E}$ ,  $\tilde{\mathcal{E}}$ , and  $\mathcal{D}$ , and only
   update the parameters of  $\epsilon_\theta$ . */
8 repeat
9   sample  $\mathbf{x}, \mathbf{m} \sim \text{dataset}$ ;
10   $\mathbf{t} \sim \mathcal{U}(1, 2, \dots, T)$ ;
11   $\epsilon_t \sim \mathcal{N}(0, \mathbf{I})$ ;
12   $\mathbf{c}^{\text{final}} = \text{metadata\_encoder}(\mathbf{m})$ ;
13  optimize the parameters of  $\epsilon_\theta$  following:
      $\mathcal{L}_{\text{ldm}} = \mathbb{E}_{\mathbf{t}, \mathbf{z}_0, \epsilon \sim \mathcal{N}(0, \mathbf{I})} [\|\epsilon_t - \epsilon_\theta(\mathbf{z}_t, \mathbf{t}, \lfloor \mathbf{y} \rfloor, \mathbf{c}^{\text{final}})\|^2]$ 
14 until converge;
```

Algorithm 2: Compress and decompress pipeline of COSMIC.

input : A satellite image \mathbf{x} , metadata $\mathbf{m}_i \in \mathbb{R}^M$
output : Reconstructed image $\hat{\mathbf{x}}$
Parameters: Image encoder on satellite \mathcal{E} , image decoder \mathcal{D} , noise prediction network ϵ_θ

```

/* Compress */
1  $\mathbf{y} = \mathcal{E}(\mathbf{x})$ ;
2  $\zeta = \text{entropy\_model}(\mathbf{y})$ ;
3  $\lfloor \mathbf{y} \rfloor = \text{Quantization}(\mathbf{y}, \zeta)$ ;
/* Decompress */
4 sample  $\mathbf{z}_T \sim \mathcal{N}(0, \mathbf{I})$ ;
5 for  $\mathbf{t} = T, \dots, 1$  do
6    $\epsilon_t \leftarrow \epsilon_\theta(\mathbf{z}_t, \mathbf{t}, \lfloor \mathbf{y} \rfloor, \mathbf{c}^{\text{final}})$ ;
7    $\mathbf{z}_{t-1} \leftarrow \text{DDIM}(\mathbf{z}_t, \mathbf{t}, \epsilon_t)$ ;
8 end
9 return  $\mathbf{z}_0$ ;
10  $\hat{\mathbf{x}} = \mathcal{D}(\text{concat}(\mathbf{z}_0, \text{deconv}(\lfloor \mathbf{y} \rfloor)))$ ;
```

表 3: E 和 D 的网络结构详情

图像编码器 E 的架构		图像解码器 D 的架构	
输入	参考图像	输入	\hat{y}, z_0'
第一层	卷积 ($3 \times 192 \times 5 \times 5$, 步长为 2), 全零填充 (192)	第一层	卷积 ($196 \times 192 \times 5 \times 5$, 步长为 1), 最大池化 (192)
第二层	卷积 ($192 \times 192 \times 5 \times 5$, 步长为 2), 全零填充 (192)	第二层	卷积 ($192 \times 192 \times 5 \times 5$, 步长为 1), 最大池化 (192)
第三层	卷积 ($192 \times 192 \times 5 \times 5$, 步长为 1), 全零填充 (192)	第三层	卷积 ($192 \times 192 \times 5 \times 5$, 步长为 2), 实例归一化 (192)
第四层	卷积 ($192 \times 8 \times 5 \times 5$, 步长为 1)	第四层	卷积 ($192 \times 3 \times 5 \times 5$, 步长为 2)
输出	z_0	输出	重建图像

A.2 COSMIC 的详细架构

COSMIC 的整体框架已在主论文中有所阐述。在此，我们分别在表 3 中提供了图像编码器 E 和图像解码器 D 的详细架构。

B 额外的率失真（感知）结果

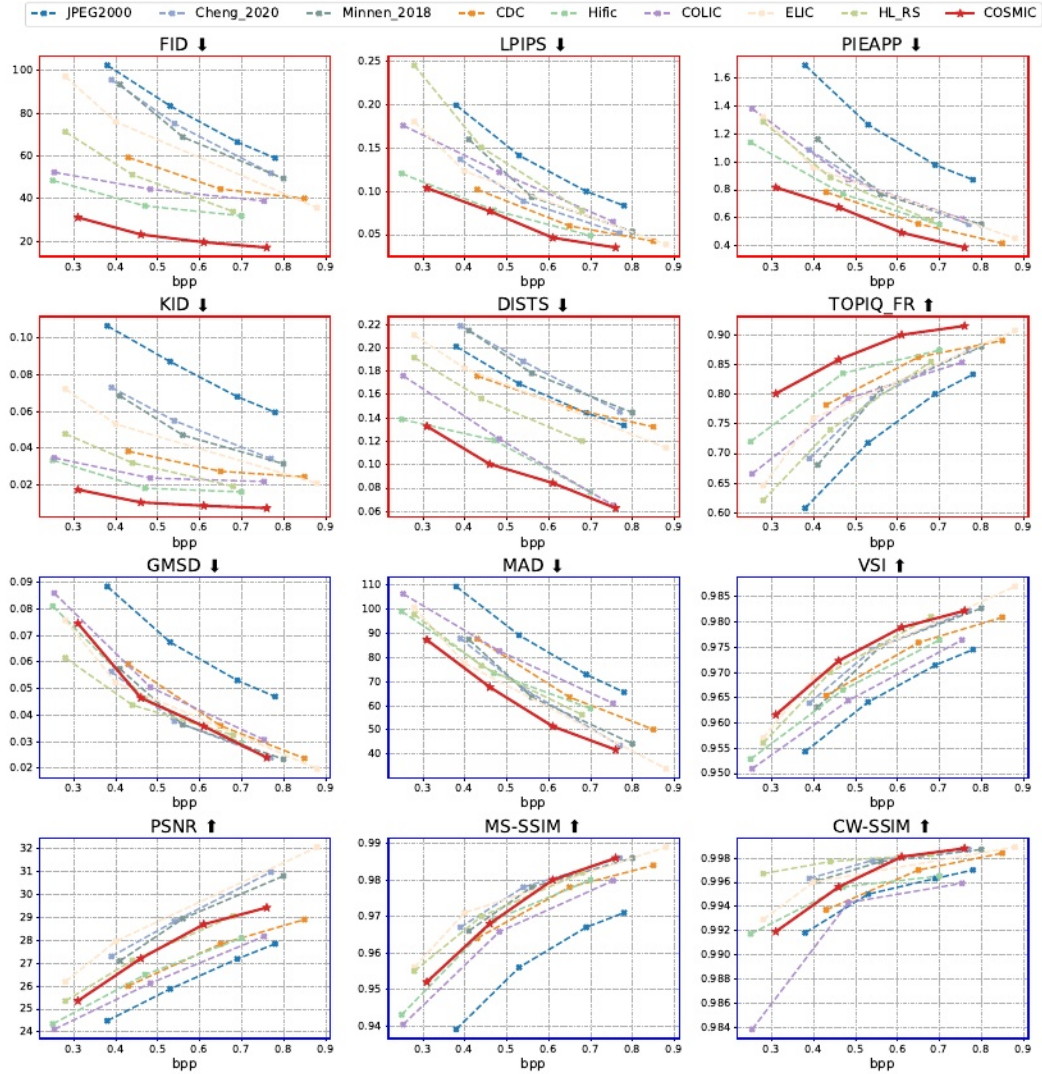


图 9: fMoW 数据集的比率失真（感知）

C 失败案例分析

稳定扩散最初被用于生成任务，它可以生成现实中不存在的纹理和内容。在COSMIC中，我们通过将离散潜在编码 y 注入噪声预测网络的Vanilla卷积（VC）块中，以提供结构信息，从而确保内容和纹理的一致性。然而，在某些情况下，特别是在极低的比特率下，潜在的离散编码可能缺乏结构信息，导致扩散生成不存在纹理。如图10（a）所示，当卫星图像捕捉到海洋场景时，低比特率下潜在编码中的信息不足，需要严重依赖扩散进行生成，导致重建图像中出现不存在纹理。在高比特率的情况下，如主论文所述，对扩散的依赖减少，在相同情况下，可以控制不存在纹理的出现。然而，在实践中，遥感下游任务更关注感兴趣区域（ROI）。例如，在海洋遥感任务中，如船舶检测，只有存在船舶的部分才是感兴趣的。如图10(b)所示，即使在低比特率下，虽然海面重建可能会引入不存在纹理，但船舶部分的重建仍然完好无损。因此，尽管COSMIC可能存在一些视觉上的缺陷，但它并不影响下游的检测任务。

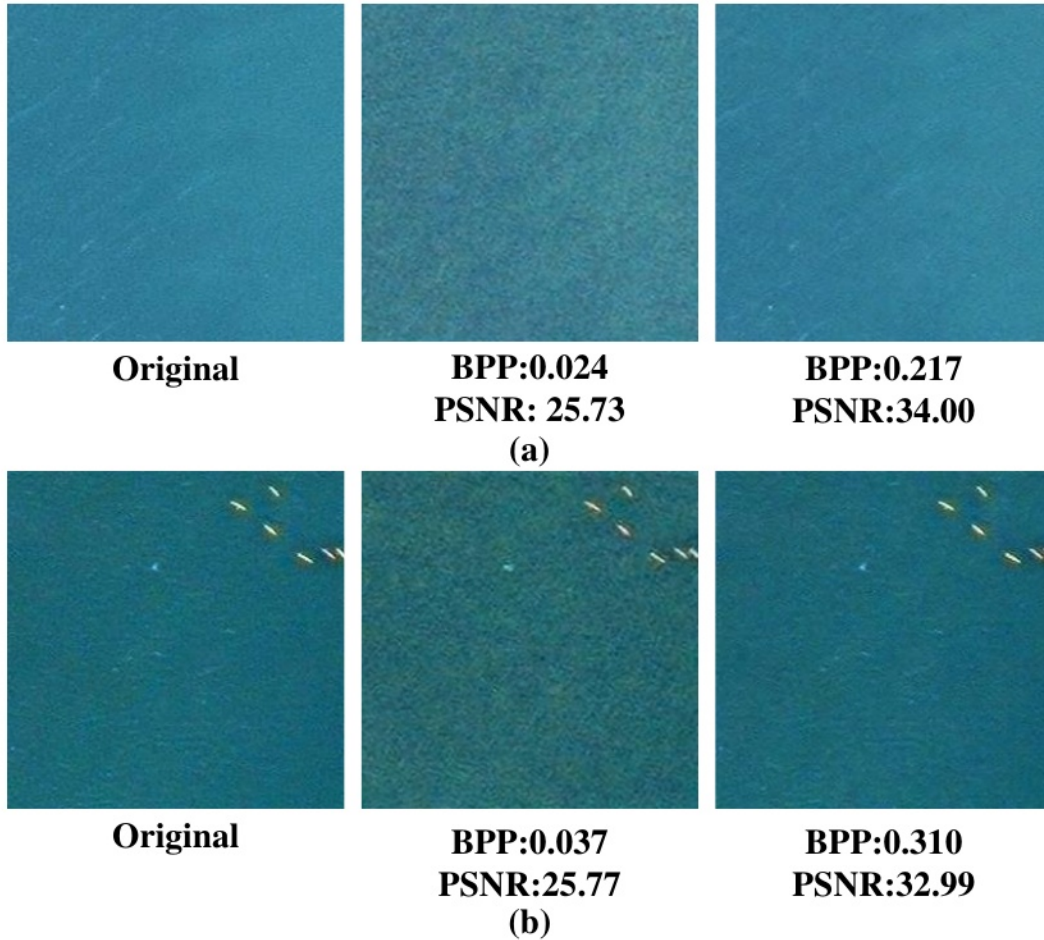


图 10：失败示例。

D 为什么不进行 artifact 校正呢？

提高卫星图像质量的最直接方法是使用JPEG2000压缩图像的伪影校正。这种方法不会给卫星带来计算负担，并且还能在一定程度上提高图像质量。我们使用DDRM[33]，这是一种图像恢复方法，用于去除JPEG2000压缩图像中的伪影。如图

表 4: fMoW 数据集不同随机种子的结果。

每秒帧数	峰值信噪比	多尺度结构相似性指数 (MSPSSIM)		首次检测时间
0.31	25.3698 \pm 0.0447	0.9521 \pm 0.0004	0.1040 \pm 0.0010	31.2750 \pm 0.4328
0.46	27.2411 \pm 0.0222	0.9682 \pm 0.0004	0.0761 \pm 0.0006	23.0243 \pm 0.2253
0.61	28.6654 \pm 0.0251	0.9799 \pm 0.0004	0.0461 \pm 0.0004	19.4264 \pm 0.1059
0.76	29.3617 \pm 0.0608	0.9852 \pm 0.0006	0.0363 \pm 0.0007	16.9090 \pm 0.1683

图11中, JPEG2000+DDRM比JPEG2000略好, 但改进有限, 与COSMIC相比仍有较大差距。我们认为, 这主要是因为基于静态压缩过程的传统压缩方法无法像神经网络那样很好地适应数据, 因此会损失大量信息, 使得图像恢复方法难以提高图像质量。因此, 有必要设计一种学习式压缩方法, 以提高卫星图像压缩质量, 同时不会给卫星带来过重的计算负担。

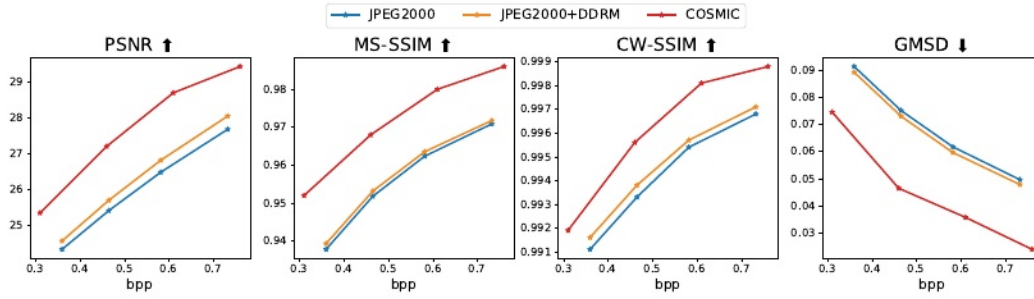


图 11: COSMIC、JPEG2000 和 JPEG2000 + DDRM 的比较结果

E 随机种子的影响

为了展示不同初始高斯噪声对重建图像质量的影响, 我们为每个比特率随机选择 5 个不同的种子, 并在表 4 中展示了 2-Sigma 误差棒的结果。结果表明, 我们的方法在各种定量指标中的方差较小, 证明了我们的方法对初始随机高斯噪声的稳健性。

F 重建图像的额外可视化

我们在主文中展示了图4和图5的全图, 并提供了更多的可视化结果, 如图12至25所示。我们从每个数据集中选择了两个不同测试集在较高比特率和较低比特率下的可视化结果。对于瓷砖测试集, 在较低比特率下, 如图5所示, 基准重建的图像在拼接接缝处表现出明显的不连续性。