arXiv:2601.18547v1 [cs.CV] 26 Jan 2026

REMAC：基于参考的火星非对称图像压缩

Qing Ding、Mai Xu*，*IEEE* 高级会员、Shengxi Li，*IEEE* 会员、Xin Deng，*IEEE* 会员、Xin Zou

摘要—为了加快火星探测进程，开发高效的火星图像压缩方法，以便通过受限的火星—地球通信信道传输图像，至关重要。尽管现有的学习式压缩方法在地球自然图像上取得了令人鼓舞的成果，但仍有两个关键问题妨碍其在火星图像压缩中的应用效果：1）这些方法忽视了火星上极其有限的计算资源；2）这些方法没有利用火星图像之间的强跨图像相似性来提升图像压缩性能。基于从纹理、颜色和语义角度对图像内及跨图像强相似性的实证分析，我们提出一种基于参考的火星非对称图像压缩（REMAC）方法，将计算复杂度从编码器转移至资源充足的解码器，同时提升压缩性能。为利用跨图像相似性，我们提出参考引导熵模块和参考解码器，从参考图像中利用有效信息，减少编码器的冗余操作，并实现更优的压缩性能。为利用图像内相似性，参考解码器采用深层多尺度架构，扩大感受野大小，以建模长程空间依赖。此外，我们还开发了潜在特征循环利用机制，以进一步缓解火星上极端的计算资源限制。实验结果表明，与当前最先进的方法相比，REMAC 将编码器复杂度降低了 43.51%，同时使 BD-PSNR 提升了 0.2664 dB。

索引词—火星、火星视觉任务、有损图像压缩、非对称自编码器、参考引导熵模型、参考感知解码器。

I. 引言

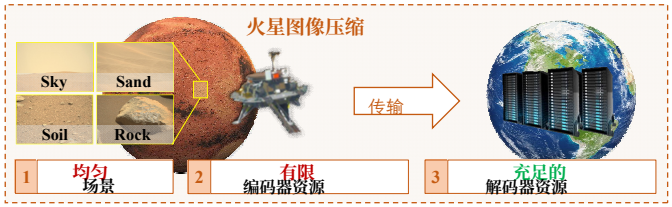
NOWADAYS，人类热衷于通过着陆巡视器收集的数据探索火星，这些数据会被传输至地球，用于进一步的科学研究和科学普及 [1]–[3]。此外，图像占火星至地球通信信道所传输数据的大多数。然而，

由于火星与地球之间距离遥远，火星探测车的资源极其有限。为了成功传输用于火星探测的高保真图像，开发一种高效的火星图像压缩方法至关重要，它

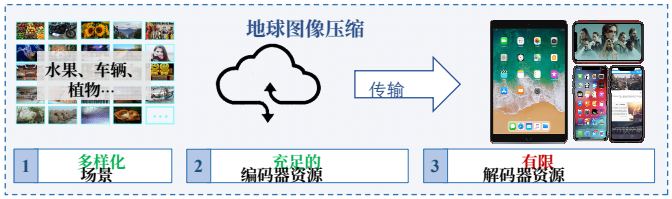
本研究得到国家自然科学基金委员会项目（62231002、62522101、62372024、62206011 和 62450131）、北京市自然科学基金项目（L223021）以及中央高校基本科研业务费的资助。

*通讯作者：M. Xu（MaiXu@buaa.edu.cn）。

Q. Ding、M. Xu、S. Li、X. Deng、X. Zou 任职于中国北京 100191 的北京航空航天大学电子与信息工程学院（电子邮箱：qingding@buaa.edu.cn；MaiXu@buaa.edu.cn；Shengx- iLi2014@gmail.com；cindydeng@buaa.edu.cn；zouxin501@163.com）。X. Zou 同时任职于中国北京 100094 的北京空间飞行器系统工程研究所。



(a) 火星图像压缩的特点



(b) 地球图像压缩的特点

图 1：火星与地球图像压缩的特征。

在实现卓越压缩性能的同时，显著降低编码器复杂度。

与地球图像相比，火星图像压缩具有鲜明的自身特点，如图 1 所示。具体而言，这些差异主要体现在以下三个方面：1）场景特征：由于火星地貌单调且均一，火星图像压缩通常处理场景均匀的图像，而地球图像压缩则必须适应多样化场景。2）编码器资源限制：火星编码过程受到着陆巡视器计算能力有限的严重制约，而地球图像压缩则可借助分布式云计算基础设施利用充足的资源。3）解码器资源限制：火星图像解码可利用地面站平台充足的计算资源；相比之下，部署在前端设备上的地球图像压缩必须在严格的资源限制下进行解码[4]。这些差异使现有压缩算法不适用于火星图像，因此有必要开发一种更高效、专门适应火星图像压缩特点的压缩方法。

现有的传统编解码器基于图像变换，例如高效视频编码（HEVC） [5], 和多功能视频编码（VVC） [6]。然而，这些编解码器无法适应图像内容，从而影响了其在火星图像压缩中的有效性。得益于深度学习的巨大成功，研究者提出了基于学习的压缩方法 [7]–[10] ，这些方法通过大量图像进行训练，

© 2025 IEEE。允许个人使用本材料。对于其他所有用途，无论是当前还是未来的媒体形式，包括为广告或宣传目的重印/再版本材料、创作新的合集作品、转售或重新分发至服务器或列表，或在其他作品中重复使用本作品的任何受版权保护部分，均须事先获得 IEEE 的许可。

因此能够适应图像内容，提升率失真（R-D）性能。由于采集到的火星图像数量庞大， [7],最先进的基于学习的图像压缩方法可用于高效压缩火星图像。尽管这些基于学习的压缩方法已取得显著进展，却忽视了上述地球图像压缩与火星图像压缩之间的差异。这导致了两个妨碍其有效进行火星图像压缩的关键问题。其一，这些方法未能充分利用火星场景均匀性这一独特特征。其二，这些方法未考虑编码器计算资源有限，而解码器计算资源充足这一情况。因此，对于实际的火星应用，需要采用轻量级编码器，同时可以在解码压缩后的火星图像时承担较高的计算复杂度。

本文提出了一种基于参考图像的火星非对称图像压缩（REMAC）方法，以降低编码器复杂度并提升压缩性能。REMAC 的设计由火星图像压缩的独特特性所指导，如图 1 所示。我们首先对火星图像进行了全面的数据分析，发现其在纹理、颜色和语义方面具有显著的 图像内和图像间相似性。基于这些观察，REMAC 旨在利用这两类相似性，将计算复杂度从资源受限的编码器转移至资源充足的解码器，同时提升压缩性能。具体而言，图像间相似性启发了参考引导熵模块和参考解码器的设计，这两者利用编码器端和解码器端均可获得的参考图像，增强了熵估计和解码重建。通过从参考图像中传递有用信息，编码器无需在输入图像上执行冗余计算，而解码器则对参考图像执行额外操作以进行补偿，从而有效实现计算非对称性。同时，图像内相似性促成了参考解码器中的深层多尺度架构，使其具备更大的感受野大小，以捕获单幅图像内部的长程空间依赖。最后，为了在推理阶段进一步降低编码器复杂度，我们提出了一种潜在特征复用机制，重新利用编码器已经提取的潜在特征进行深层参考图像选择，从而在推理过程中可以省去特定的特征提取操作。

总而言之，我们的工作有以下三方面贡献。

1) 受火星图像之间显著的图像间相似性启发，我们提出了一种新颖的参考引导熵模块和参考解码器，利用参考图像增强熵估计和解码重建，同时降低编码器复杂度。2) 受图像内显著相似性的启发，我们在参考解码器中提出了一种深层多尺度架构，以扩大感受野大小，从而建模单幅图像中的长程空间依赖。3) 我们提出了一种潜在特征循环利用机制，用于将编码器的潜在特征用于深度参考图像选择，从而省去特定的特征提取步骤，降低推理过程中的编码器复杂度。

II. 相关工作

A. 基于学习的图像压缩

现有的学习式图像压缩（LIC）方法可按三种主要的优化视角进行分类：变换导向、量化导向和熵编码导向设计。本节回顾与我们的研究最相关的变换导向和熵编码导向设计。

变换导向方法旨在改进变换模块，以提升潜在表示的质量，进而提高压缩效率。Ball´e 等人 [11] 首次提出了端到端变分自编码器压缩框架，为后续的变换导向模型奠定了基础。在此基础上，Cheng 等人 [8] 引入注意力机制，以增强变换过程中的细节保留能力。近年来，视觉 Transformer [12]的兴起催生了多种方法。Zou 等人 [9] 设计了基于窗口的注意力块，以捕获局部和全局依赖关系。Liu 等人 [13] 提出了混合Transformer-CNN块，兼取卷积神经网络和Transformer模型的优势。与此同时，针对遥感图像的多种变换导向图像压缩框架相继问世，例如采用频率分离 [14], 、马尔可夫启发式注意力[15]和多级相似性约束 [16]。此外，Ding 等人 [7] 利用非局部注意力块 [17] 进行火星图像压缩，充分利用火星图像块之间的高度相似性。

专注于熵编码的方法旨在实现更准确的熵估计，以提高压缩效率。Ball´e等人 [10] 提出了一种超先验模型，以利用紧凑潜在表示中的依赖结构来改善熵估计。Minnen 等人 [18] 在此基础上引入了空间自回归模型，以充分利用变换后的潜在表示中的上下文信息。然而，这种方法非常耗时。为了解决这一缺点，研究者提出了逐通道[19],棋盘格 [20], 以及空间-通道 [21]自回归模型，以加速这一过程。

尽管取得了显著进展，但现有的大多数学习式图像压缩方法都是为地球图像设计的，这类图像场景多样性高，且图像间冗余有限。因此，这些方法未能利用火星图像固有的高图像间相似度。此外，其中许多方法依赖计算开销巨大的编码器。这些局限共同导致压缩性能欠佳，并阻碍了其在火星上的实际部署。

B. 基于参考的方法

有损压缩在量化过程中不可避免地会丢弃图像细节，因此可以利用参考图像恢复丢失的信息，以获得更好的压缩性能。与此最接近的领域是基于参考的图像超分辨率（RefSR）。



图2：具有代表性的火星图像（左）和地球图像（右）对比。

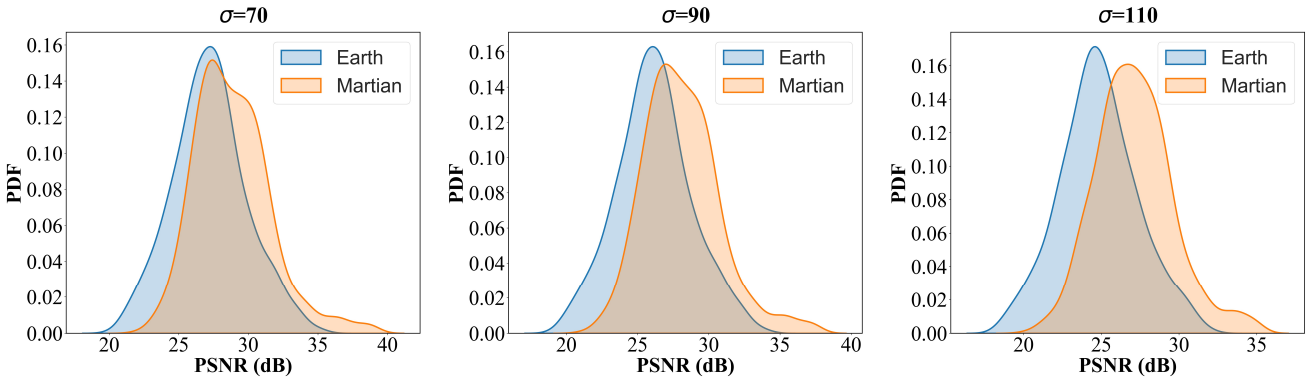


图3：地球图像和火星图像在高斯噪声水平为 $\sigma = 70$、 $\sigma = 90$和$\sigma = 110$时的峰值信噪比值分布。

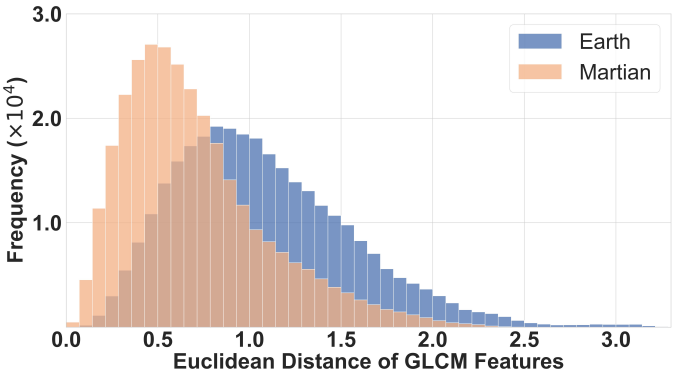


图4：火星图像和地球图像灰度共生矩阵特征的欧几里得距离直方图。

在 RefSR 中，利用高分辨率（HR）参考图像辅助低分辨率（LR）图像进行超分辨率重建。基于参考图像与目标 HR 图像非常相似这一假设，Zheng 等人 [22]提出了 CrossNet，该方法估计 LR 与 HR 图像之间的光流，并以跨尺度方式对 HR 图像进行变形。为处理参考图像相似度较低的 RefSR 任务，Zhang 等人将 RefSR 任务重新定义为神经纹理迁移问题 [23],，利用语义相关纹理进行纹理迁移。受注意力机制方面进展的启发 [24], , Zhang等人 [25]引入了纹理变换器模块，通过注意力图自适应地迁移相关纹理。然而，这些方法依赖于空间对齐或密集图像块通过匹配来定位参考图像与低分辨率图像之间的对应关系，这一过程耗时较长。为加快这一过程，Lu等人 [26]提出了 MASA-SR 方法，以由粗到细方式进行图像块匹配。尽管这些参考图像超分辨率方法在图像增强方面有效，但它们仅针对重建质量进行训练，并未考虑编码器造成的压缩伪影。将其直接应用于图像压缩时，会形成解耦式两阶段压缩流水线，其中编码器和解码器独立优化，缺乏端到端率失真优化，导致压缩性能欠佳。

目前，只有两项工作 [27], [28] 探索了基于参考图像的压缩，且都专为水下图像设计。这些方法严重依赖水下物理先验（例如使用水下散射模型），主要关注压缩效率，却忽略了计算成本，因此根本不适用于火星图像。将其应用于火星图像时存在三个关键局限：1）它们采用图像块级参考图像选择，导致计算成本高昂，且无法实现全局上下文建模。2）它们采用编码器重型架构，编码器复杂而解码器轻量，因此不适用于火星上受限的编码器资源。3）它们仅部分集成参考信息：一种方法 [27] 仅将其应用于变换模块，另一种方法 [28] 仅将其应用于熵编码模块，未能发挥联合集成的优势。

受火星图像间强烈的相互相似度启发，

我们提出了 REMAC，这是一种专为火星图像设计的基于参考图像的压缩方法。REMAC 引入了一种联合优化的架构，解决了以往工作的所有三项局限。1）采用图像级参考选择，以节省计算成本并实现全局上下文建模。2）采用以解码器为核心的设计，将计算成本从资源受限的编码器（位于火星）转移到资源充足的解码器（位于地球），适用于在火星上的实际部署。3）将参考信息全面融入熵估计和解码重建，从而能够高效利用参考图像进行图像压缩。因此，REMAC 在遵循火星图像压缩实际限制的同时，实现了高效的图像压缩，代表着对现有基于参考图像的图像压缩方法的一项重大进展。

C. 轻量级学习式图像压缩

为了在资源受限的条件下实现学习式图像压缩模型的实际部署，研究人员提出了许多轻量级学习式图像压缩方法。现有的轻量级压缩方法大致可分为两类：模型压缩技术和高效架构设计。

模型压缩技术采用网络剪枝和权重量化来减少模型参数和复杂度。为使学习式图像压缩模型适用于资源受限的前端设备，Kim 等人 [4]采用权重剪枝和滤波器剪枝来精简其解码器。Luo 等人 [29] 观察到学习式图像压缩模型中的超先验路径参数过多，因此提出对其进行剪枝，以减少模型参数。为了进一步优化超先验路径的剪枝效果，Luo 等人[30] 提出为每个每像素比特率点对应的学习式图像压缩模型搜索合适的剪枝阈值。为进一步提高学习式图像压缩模型的部署效率，Fang 等人 [31]提出了一种针对权重和激活值的有效量化方法。

高效架构设计采用专门的卷积运算和网络架构，使学习式图像压缩模型更加轻量。受自注意力模块中高效设计的启发，He 等人 [32]将高效注意力模块融入基于 Transformer 的学习式图像压缩模型中，显著降低了自注意力机制的高计算成本。为降低解码器复杂度，Fu等人 [33]和 Wang 等人 [34] 提出了非对称学习式图像压缩模型，其中解码器网络的层数少于编码器网络。最近，Bao 等人[35] 采用无参数移位运算替代大核卷积，从而降低计算成本。

尽管这些方法在一般场景中效果良好，但缺乏任务特异性，在火星图像压缩这类特定领域任务中尤其如此。这些通用设计在火星图像压缩方面有两大主要局限：(1) 未能充分利用图像特征（例如火星图像之间较高的图像间相似度）来提高压缩效率；(2) 大多数方法采用对称架构，即便少数非对称方法也只着眼于降低解码器复杂度，因此不适用于编码器资源受限的火星图像压缩场景。这些局限促使我们采用场景驱动的设计，显式利用火星图像特征和充足的解码器资源，在保持压缩性能的同时减轻编码器负担。

III. 数据分析

本节将深入分析火星图像的特征，探寻提升压缩效率和降低编码器复杂度的潜力。具体而言，我们从 DIV2K [36] 数据集中选取 784 幅地球图像，确保其宽度和高度均不小于 MIC [7] 数据集中火星图像的分辨率。同时，我们从 MIC [7] 数据集中随机选取 784幅火星图像进行分析。为确保比较公平，所有选取的地球图像均裁剪为 1600×1152 像素，与火星图像的分辨率一致。为便于直观比较这两个星球，图中展示了具有代表性的火星和地球图像2 ，以突出它们的视觉差异。如图所示，火星图像的颜色集中度较高，主要呈现四种类型：天空、沙地、土壤和岩石。相比之下，地球图像在颜色分布和场景类别方面都丰富得多。基于这些视觉观察，我们在后续的定量分析中得出了三项重要发现。

发现 1： 与地球图像相比，火星图像表现出更强的图像内和图像间纹理相似性。

分析： 为测量图像内纹理相似性，我们对含噪地球图像和火星图像均应用块匹配和三维滤波（BM3D）算法 [37]。BM3D 算法能够有效去除具有更多相似图像块的图像中的噪声。因此，去噪性能越高，表明图像内纹理相似性越强。我们向地球图像和火星图像分别添加多个水平的加性高斯噪声，随后利用 BM3D 算法去除噪声。去噪性能通过峰值信噪比（PSNR）指标进行评估。图3 展示了高斯噪声水平为 $\sigma = 70$、 $\sigma = 90$和 $\sigma = 110$时地球图像与火星图像的峰值信噪比值分布。如图所示，在多个噪声水平下，BM3D 算法在火星图像上均取得了高于地球图像的峰值信噪比值，表明火星图像具有比地球图像更强的图像内相似性。

为进一步评估图像间纹理相似性，我们使用灰度共生矩阵（GLCM） [38]提取六种纹理特征，包括对比度、不相似度、同质性、能量、相关性和角二阶矩。随后，对这些特征进行归一化并拼接成一个向量，用于表示图像纹理。接着，我们计算从不同图像中提取的这些向量之间的欧氏距离。两幅图像之间的距离越小，表明其图像间纹理相似性越高。我们还在图4中展示了这些距离的分布。如图所示，火星图像的距离分布位于地球图像分布的左侧，表明大多数距离更小。这证实，与地球图像相比，火星图像具有更强的图像间纹理相似性。

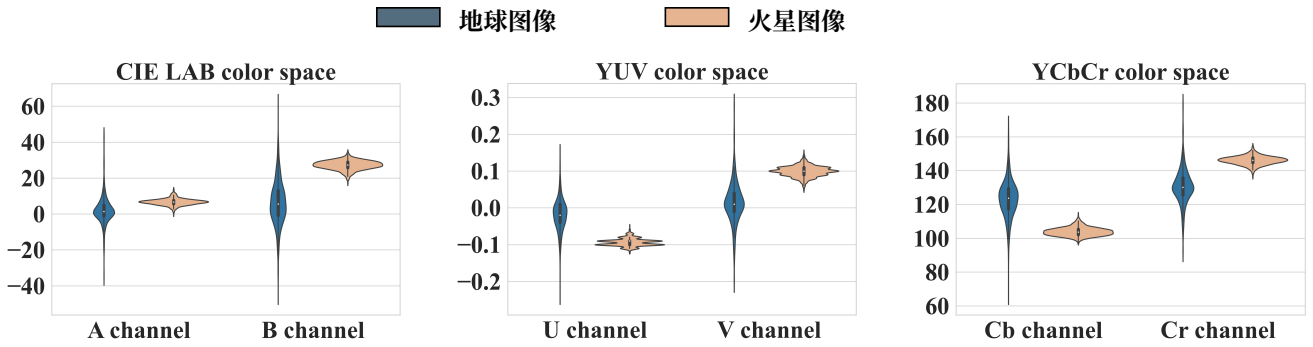


图5：不同色彩空间中各颜色通道均值的小提琴图。

表I：每种色彩空间中两个色度通道的平均标准差值。

色彩空间	色度通道	火星	地球
CIE LAB	A	3.0024	8.4402
	B	4.5352	13.9493
YUV	U	0.0171	0.0478
	V	0.0235	0.0632
YCbCr	Cb	4.4179	12.2660
	Cr	4.2587	11.4953

发现 2： 与地球图像相比，火星图像表现出更强的图像内和图像间颜色相似性。

分析： 为分析图像之间的颜色相似度，我们比较了地球图像和火星图像在 CIE LAB、YUV 和 YCbCr 色彩空间中的颜色分布。具体而言，我们首先将 RGB 图像转换到这些色彩空间，以将图像数据分解为亮度和色度分量。由于色度分量包含大部分颜色信息，我们计算每幅图像在色度通道上的标准差（STD），以评估 图像内 颜色相似度。结果见表 I。如表中所示，火星图像在各色彩空间中的平均标准差值均较低，表明火星图像具有更高的图像内颜色相似度。此外，为比较图像间 颜色相似度，我们计算了每幅图像色度通道的均值。图中绘出了地球图像和火星图像均值对应的小提琴图5。如图所示，在各色彩空间中，火星图像的色度通道均值分布更为集中，这表明火星图像之间具有更强的图像间颜色相似度。

发现 3： 与地球图像相比，火星图像表现出更强的语义相似性。

分析： 为了分析火星图像的语义相似度，我们采用广为人知的 VGG16 分类模型 [39] 从每幅图像中提取语义特征。随后计算这些特征之间的欧氏距离，以衡量语义相似程度。欧氏距离越小，表示语义相似度越高。具体而言，我们利用 VGG16 模型最后一个 ReLU层的特征来提取语义特征。

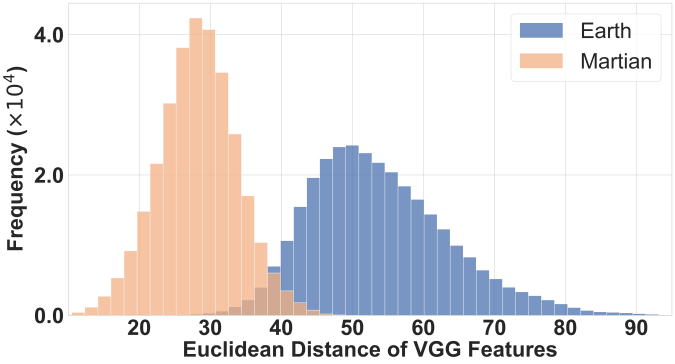


图 6：火星图像和地球图像的 VGG 特征欧几里得距离直方图。

火星图像和地球图像的欧氏距离分布如图 6所示。可以看出，火星图像之间的距离远小于地球图像之间的距离，这表明火星图像具有更高的语义相似度。

IV. 所提出的方法

基于上述发现，我们提出 REMAC 方法，以充分利用图像内 和图像间 相似性。该方法利用与输入图像相似的参考图像，充分挖掘上述火星图像特性，从而通过轻量级编码器实现高效的图像压缩。在第 IV-A 节中，我们介绍 REMAC 方法的整体架构。随后，我们将在第 IV-B、IV-C 和 IV-D 节中分别介绍所提出的参考图像选择模块、参考感知编码器（R-编码器）、解码器（R 解码器）和参考引导熵模块。第 IV-E 节介绍所提出的潜在特征循环利用机制，以及用于轻量级编码器率失真优化的相应损失函数。

A. 整体架构

我们在图7中展示了所提出的 REMAC 的整体架构，该架构由参考图像选择器、R-编码器、参考引导熵模块和 R 解码器组成。具体而言，对于参考图像选择器， N 从 MIC 数据集的训练集中随机选取图像[7] 以构建参考数据集 $\{\boldsymbol{x}_{\text{ref}}^{n}\}_{n=1}^{N}$。为了高效地

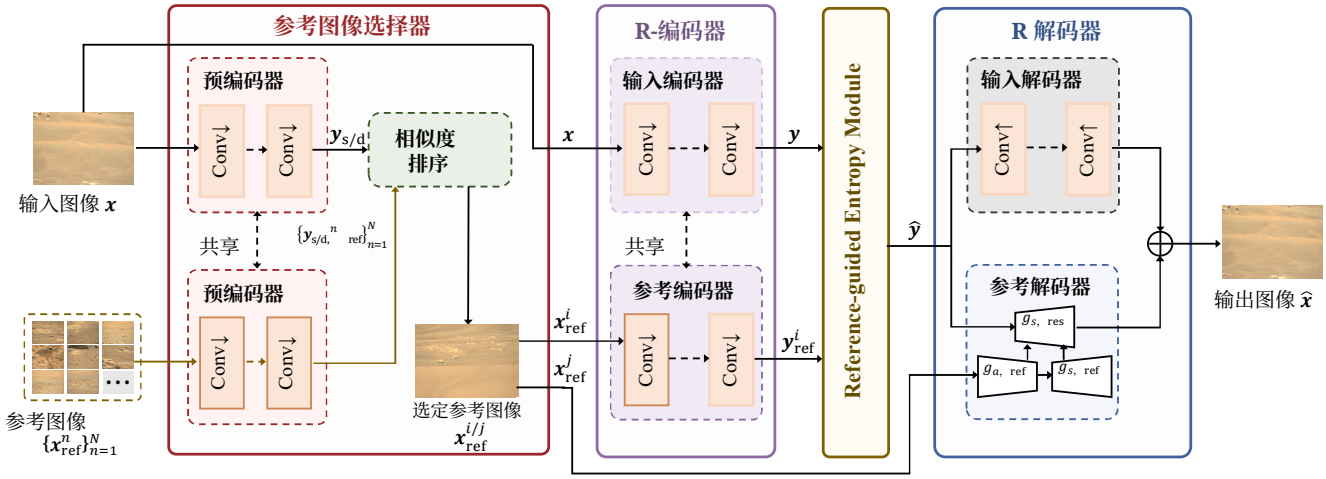


图 7：所提出的 REMAC 框架。这里，*R-Encoder* 和 *R-Decoder* 分别表示所提出的参考感知编码器和解码器。每个“Conv↓”由一个下采样卷积层和一个 GDN 层组成，但最后一个仅包含卷积层。类似地，每个“Conv↑”由一个上采样卷积层和一个逆GDN层组成，但最后一个除外。

利用参考图像，通过计算并排序由预训练编码器提取的输入特征与参考特征之间的相似度，生成输入-参考三元组。基于这些三元组，将一幅选定参考图像$\boldsymbol{x}_{\text{ref}}^{i}$ 和对应的输入图像 $\boldsymbol{x}$ 送入 R-编码器，以获得用于压缩的代表性参考特征和输入特征。随后，所提出的参考引导熵模块对这些特征进行处理，并在参考特征 $\boldsymbol{y}_{\text{ref}}^{i}$的引导下改进输入特征 $\boldsymbol{y}$ 的熵估计。然后，将另一幅选定参考图像 $\boldsymbol{x}^{j}$ 参考图像输入所提出的 R 解码器，以增强量化特征 $\hat{\boldsymbol{y}}$的重建。为进一步降低推理过程中的编码器复杂度，我们提出了一种潜在特征循环利用机制：通过对 REMAC 模型进行微调，直接重新利用其潜在特征 $\boldsymbol{y}$ 和 $\boldsymbol{y}_{\text{ref}}^{i}$ 进行深层参考选择，从而无需在参考图像选择器中进行深层特征提取。

B. 参考图像选择器

在所提出的参考图像选择器中，会选择两张参考图像来承担不同的作用：1）基于浅层纹理特征的一张图像，用于指导解码重建；2）基于深层语义特征的另一张图像，用于增强熵估计。为此，我们采用预训练编码器，从输入图像和候选参考图像中提取浅层和深层特征。对于每个候选图像，我们计算相似度并进行排序。随后，选择相似度最高的参考图像作为相应的参考图像，从而得到用于训练 REMAC 的输入图像与参考图像的初始三元组。

更具体地说，我们使用超先验压缩结构 [10],构建基础压缩模型，其中所有卷积核大小均设为 3，并结合 [19] 逐通道熵模型。该模型首先在 MIC 数据集 [7]，上训练，随后采用其预训练编码器作为我们的预编码器，如图 7所示。给定一幅输入图像$\boldsymbol{x}$ 和参考数据集 $\{\boldsymbol{x}_{\text{ref}}^{n}\}_{n=1}^{N}$，我们将它们输入预编码器，以获得浅层输入特征 $\boldsymbol{y}_{\text{s}}$、深层输入特征 $\boldsymbol{y}_{\text{d}}$、浅层参考特征 $\{\boldsymbol{y}_{\text{s, ref}}^{n}\}_{n=1}^{N}$ 和深层参考特征$\{\boldsymbol{y}_{\text{d, ref}}^{n}\}_{n=1}^{N}$。对于每个候选参考图像，我们分别计算浅层和深层特征的相似度，按降序排序并选出排名靠前的参考图像，获取其索引 j 和i，形成输入图像与参考图像的初始三元组（$\boldsymbol{x}$， $\boldsymbol{x}^{j}$ ref $\boldsymbol{x}_{\text{ref}}^{i}$）。

输入图像与参考图像特征之间的相似度通过 L1距离和 L2距离进行测量：

$$s_{\text{s}}^{n} = \|E_{\text{pre}}^{\text{s}}(\boldsymbol{x}) - E_{\text{pre}}^{\text{s}}(\boldsymbol{x}_{\text{ref}}^{n})\|_2 = \|\boldsymbol{y}_{\text{s}} - \boldsymbol{y}_{\text{s,ref}}^{n}\|_2, \quad (1)$$

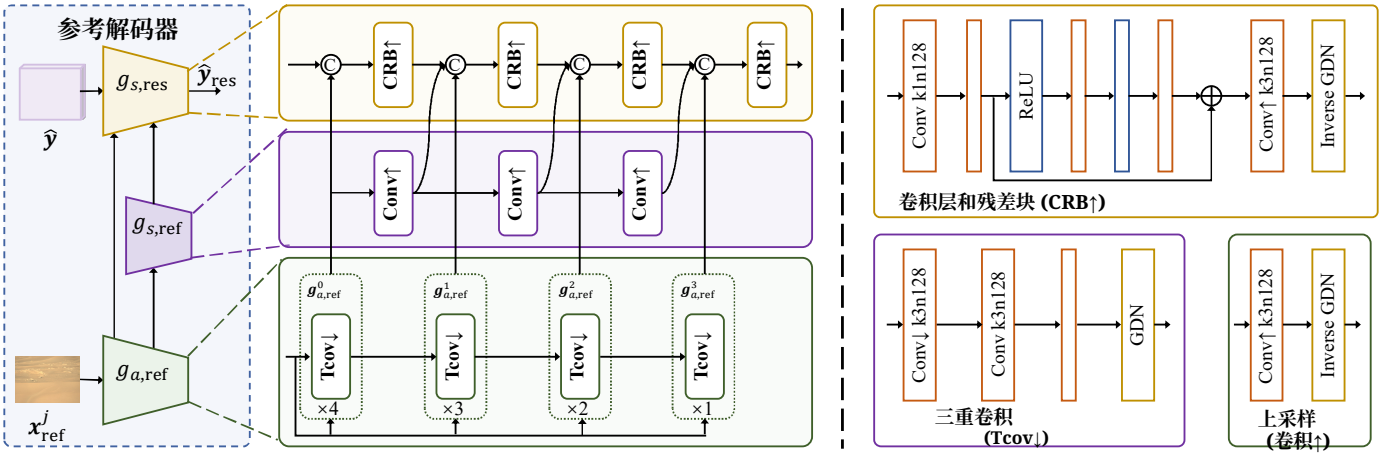
$$s_{\text{d}}^{n} = \|E_{\text{pre}}^{\text{d}}(\boldsymbol{x}) - E_{\text{pre}}^{\text{d}}(\boldsymbol{x}_{\text{ref}}^{n})\|_1 = \|\boldsymbol{y}_{\text{d}} - \boldsymbol{y}_{\text{d,ref}}^{n}\|_1, \quad (2)$$

其中， s_{s}^{n} 和 s_{d}^{n} 分别表示输入图像的浅层特征和深层特征与第 n 个参考图像之间的相似度分数，同时分别表示 L2距离和 L1距离。函数 $E_{\text{pre}}^{\text{s}}(\cdot)$和 $E_{\text{pre}}^{\text{d}}(\cdot)$分别对应预编码器的浅层和深层。数值越低表示相似度越高，并选择分数/距离最小的参考图像作为相应的参考图像。

C. 参考感知编码器和解码器

在所提出的 R-编码器中，我们利用输入编码器和参考编码器，为后续的熵估计和解码重建生成输入图像和选定参考图像的代表性特征。两个编码器均用于提取代表性特征，因此采用相同的架构和权重。它们由 4 个下采样卷积层和后续的 GDN 层组成。为降低编码器端的计算成本，每个卷积层的卷积核大小均设为 3。该过程可表示为：

$$\boldsymbol{y} = E(\boldsymbol{x}), \quad (3)$$



(a) 我们的参考解码器架构。 (b) 对应模块的详细结构。

图8：我们的参考解码器架构及对应模块的结构。此处，k 表示卷积核大小，n 表示通道数，C 表示拼接操作。未描述的矩形与前一个颜色矩形结构相同。

$$\boldsymbol{y}^{\text{i}}_{\text{ref}} = E(\boldsymbol{x}^{\text{i}}_{\text{ref}}), \tag{4}$$

其中，E 对应输入编码器和参考编码器的函数，两者共享相同的权重。同时，$\boldsymbol{y}$ 和$\boldsymbol{y}^i_{\text{ref}}$ 是用于压缩的生成特征。

在我们的 R 解码器中，我们通过输入解码器和参考解码器利用火星图像中显著的帧内和帧间图像相似性。两者共同扩大感受野大小，并整合来自参考图像的有用外部信息，从而实现更优的重建。R 解码器的架构如图7所示，其中参考解码器重建残差，作为输入解码器输出的补充；输入解码器仅利用输入信息。输入解码器旨在仅根据压缩后的输入代表性特征$\hat{}\boldsymbol{x}$重建压缩图像$\hat{}\boldsymbol{y}$。作为输入编码器的对应部分，输入解码器包含4个与逆 GDN 层结合的上采样卷积层。这些卷积的卷积核大小设为3。

为提升重建质量，本文提出参考解码器，从参考图像中提取有价值的特征，并将其与压缩输入特征进行融合，以提供有用信息。更具体地说，该解码器以多尺度方式利用有价值的参考信息，同时扩大感受野大小，以提升性能。此外，参考解码器的详细架构如图8 （a）所示。从该图可以看出，参考解码器由残差合成网络 $g_{s,\text{res}}$、参考合成网络$g_{s,\text{ref}}$以及多个多尺度参考分析网络g^p_a 、ref组成，其中 p的取值范围为 0 到 3。

为充分利用多尺度参考信息，我们首先对选定参考图像进行多尺度特征提取 $\boldsymbol{x}^j$ 参考，采用多个多尺度参考分析网络g^p_a ，参考 以及参考合成网络$g_{s,\text{ref}}$。为扩大感受野大小以实现更有效的特征提取，我们在多个参考分析网络中采用由三层卷积层和一个 GDN 层组成的模块作为基本块，并将其命名为“Tcov”g^p_a ，参考，如图8（b）所示。同时，我们还借助参考合成网络$g_{s,\text{ref}}$改进了特征提取。

给定从上述网络中提取的多尺度参考特征，我们使用残差合成网络 $\hat{}\boldsymbol{y}$将其与压缩特征进行融合$g_{s,\text{res}}$ 如图 8（a）所示。在该网络中，我们将这些特征输入拼接层，并采用两个卷积层与一个残差块相结合的结构，称为“CRB↑”，以进行进一步融合，如图 8（b）所示。最后，我们得到参考解码器的输出 $\hat{}\boldsymbol{x}_{\text{res}}$，作为输入解码器的补充。

D. 参考引导熵模块

鉴于火星图像之间具有很强的相似度，我们提出了参考引导熵模块，以充分利用参考图像，提供有价值的外部信息，从而更准确地进行熵估计。基于初始三元组，我们利用选定参考图像辅助预测熵参数。参考引导熵模块的架构如图9所示。输入特征 $\boldsymbol{y}$ 和选定参考特征$\boldsymbol{y}^i_{\text{ref}}$ 被输入参考超先验，以生成熵参数(μ， σ)。新生成的(μ， σ) 被输入各编码切片，以进行逐切片熵编码。

文献[19]提出的通道级自回归熵模型是一种广泛使用的学习式熵模型，它以超先验熵模型 [10]为基础。与超先验模型相比，通道级模型增加了逐切片熵编码。具体而言，它通过将边信息 $\boldsymbol{z}$ 传输到解码器端，预测代表性特征 $\boldsymbol{y}$ 的熵参数（μ， σ）。随后，代表性特征 $\boldsymbol{y}$被划分为 M 个切片。为执行逐切片熵编码，该模型利用预测得到的（μ， σ）估计每个切片的熵参数。此外，这些参数以先前解码的切片为条件，使模型能够利用 $\boldsymbol{y}$ 中的通道级上下文信息，从而实现高效的熵估计。

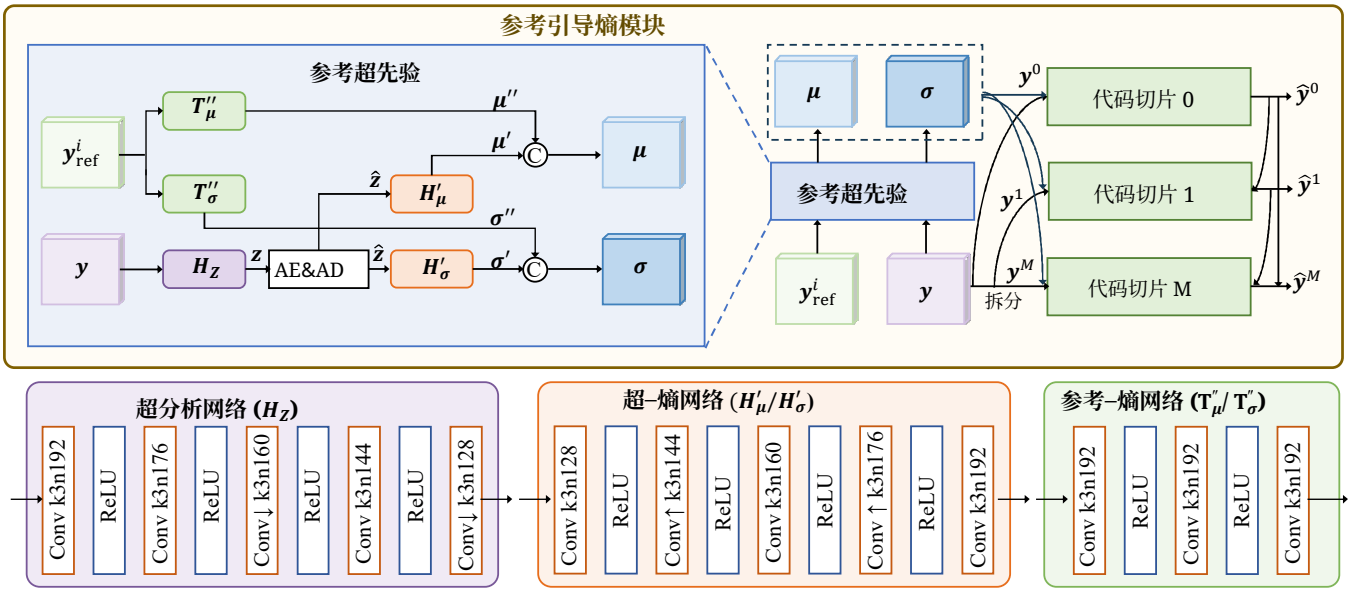


图9：我们提出的参考引导熵模块架构。这里，k 表示卷积核大小，n表示通道数，C 表示拼接操作。

然而，由于该模型未利用参考特征 $\boldsymbol{y}^i_{\text{ref}}$，而这些特征与输入特征 $\boldsymbol{y}$具有相似性，因此它不适用于火星图像压缩。为此，我们对其进行改进，并提出参考超先验（ref-hyperprior）。换言之，我们借助 $\boldsymbol{y}^i_{\text{ref}}$而非仅使用边信息 $\boldsymbol{z}$来预测熵参数（μ，σ）。由于参考图像能够提供有用的外部信息，这一过程可以在提高估计准确率的同时减少编码比特。输入特征和参考特征分别用于生成各自的熵参数，然后将这些参数拼接起来，得到更新且更准确的熵参数（μ，σ）：

$$\begin{aligned}\mu &= [\mu^{'}, \mu^{''}] = [H^{'}_{\mu}(H_z(\boldsymbol{y})), T^{''}_{\mu}(\boldsymbol{y}^i_{\text{ref}})] \\ &= [H^{'}_{\mu}(\hat{\boldsymbol{z}}), T^{''}_{\mu}(\boldsymbol{y}^i_{\text{ref}})], \qquad (5)\\ \sigma &= [\sigma^{'}, \sigma^{''}] = [H^{'}_{\sigma}(H_z(\boldsymbol{y})), T^{''}_{\sigma}(\boldsymbol{y}^i_{\text{ref}})] \\ &= [H^{'}_{\sigma}(\hat{\boldsymbol{z}}), T^{''}_{\sigma}(\boldsymbol{y}^i_{\text{ref}})]. \qquad (6)\end{aligned}$$

这里，$[\cdot]$ 表示拼接操作，$H_z(\cdot)$表示生成超先验 $\boldsymbol{z}$ 的超分析网络，
$H^{'}_{\mu}\cdot$ $H^{'}_{\sigma}\cdot$
() 和( ) 表示利用超先验 $\boldsymbol{z}$推导熵参数的超熵网络，而 $T^{''}_{\mu}(\cdot)$和 $T^{''}_{\sigma}(\cdot)$是参考熵网络，它们利用参考特征预测熵参数。为降低编码器复杂度，上述网络中的所有卷积层均采用相同的卷积核大小，即3。这些网络的详细参数见图。9。

E. 潜在特征回收与优化

为进一步应对火星上编码器资源极其有限的问题，我们提出一种潜在特征循环利用机制，通过将预先计算的深层输入特征重新用于深层参考选择，在推理过程中消除冗余的深层特征提取。在预训练阶段，我们首先按照上述图中所示的流程训练 REMAC。7，从而获得优化后的预训练 REMAC 模型。随后，在编码器端，我们存储深层参考特征$\{\boldsymbol{y}^n_{\text{ref}}\}^N_{n=1}$以及基于参考的熵参数 $\{(\mu^{''}_n, \sigma^{''}_n)\}^N_{n=1}$。在预训练 REMAC 模型的推理过程中，将输入图像 $\boldsymbol{x}$ 输入预编码器以提取用于深层参考选择的代表性特征 $\boldsymbol{y}_{\text{d}}$。为进一步降低编码器复杂度，我们不再重新提取 $\boldsymbol{y}_{\text{d}}$，而是重新利用预训练输入编码器和参考编码器提取的潜在特征 $\boldsymbol{y}$和参考特征$\{\boldsymbol{y}^n_{\text{ref}}\}^N_{n=1}$ 来进行深层参考选择，从而省去预编码器深层的特征提取。这意味着，推理过程中使用的新深层参考图像$x^q{}_{\text{ref}}$ 是根据预训练输入编码器和参考编码器提取的特征选出的，与预训练阶段使用的初始深层参考图像x^i_{ref} 不同。为确保训练与推理的一致性，我们冻结预训练输入编码器和参考编码器，然后使用新的输入-参考三元组对预训练 REMAC 模型进行微调($\boldsymbol{x}$，$\boldsymbol{x}^j{}_{\text{ref}}\boldsymbol{x}^q{}_{\text{ref}}$)。因此，编码器复杂度降低，而压缩性能略有下降。具体细节如算法1所示。

更具体地说，我们使用具有预训练参数 θ_{E} 的输入编码器和参考编码器来生成潜在输入和参考特征。首先，将输入图像 $\boldsymbol{x}^m$ 输入输入编码器，以获得潜在输入特征 $\boldsymbol{y}^m$。然后，我们计算 $\boldsymbol{s}^m_{\text{d}}$ 与参考潜在特征集合 $\{\boldsymbol{y}^n_{\text{ref}}\}^N_{n=1}$ 中的 $\boldsymbol{y}^m$ 之间的相似度分数。接下来，我们选择与最高相似度（即基于距离的分数最低）对应的参考索引q^m 。此外，我们构建新的输入-参考三元组 $\{(\boldsymbol{x}^m, \boldsymbol{x}^{j^m}{}_{\text{参考}}\boldsymbol{x}^{q^m}{}_{\text{参考}})\}^M_{m=1}$，其中 $\boldsymbol{x}^{j^m}{}_{\text{参考}}$ 表示在上述预训练阶段预先选定的浅层参考图像，并且在此微调阶段保持不变。

算法1： 潜在特征循环利用机制。

输入： 预训练的 REMAC 网络参数 $\{\theta_{\mathrm{E}}, \phi\}$，输入图像 $\{\boldsymbol{x}^m\}_{m=1}^{M}$，参考图像 $\{\boldsymbol{x}_{\mathrm{ref}}^n\}_{n=1}^{N}$，预先选定的浅层参考图像 $\{\boldsymbol{x}_{\mathrm{ref}}^{n_h}\}_{n=1}^{N}$。

输出： 微调后的 REMAC 网络参数 $\{\theta_{\mathrm{E}}, \tilde{\phi}\}$。

变量： 参考特征 $\{\boldsymbol{y}^n{}_{\text{参考}}\}_{n=1}^{N}$、输入特征 $\{\boldsymbol{y}^m\}_{m=1}^{M}$、相似度分数 $\{\boldsymbol{s}_{\mathrm{d}}^m\}_{m=1}^{M}$、选定索引 $\{q^m\}_{m=1}^{M}$。

参数： 学习率 lr、损失 L、最大训练步数 K，参考图像数量 N，输入图像数量 M。

初始化 $\tilde{\phi} \leftarrow \phi$。
当 $n < N$ **时执行**
  计算 $\boldsymbol{y}^n{}_{\text{ref}} = E(\boldsymbol{x}_{\mathrm{ref}}^n; \theta_{\mathrm{E}})$。
结束
当 $m < M$ **时执行**
  计算 $\boldsymbol{y}^m = E(\boldsymbol{x}^m; \theta_{\mathrm{E}})$。
  计算 $\boldsymbol{s}_{\mathrm{d}}^m = \|\boldsymbol{y}^m - \{\boldsymbol{y}^n{}_{\text{参考}}\}_{n=1}^{N}\|_1$。
  计算 $q^m = \underset{\text{索引}}{\text{的最小值}}\ \boldsymbol{s}_{\mathrm{d}}^m$
结束
对于 $k = 1$ **至** K，**执行**
  使用 $\{(\boldsymbol{x}^m, \boldsymbol{x}^{j^m}{}_{\text{参考}}\boldsymbol{x}^{q^m}{}_{\text{参考}})$计算 $L\}_{m=1}^{M}$。
  更新 $\tilde{\phi}^k \leftarrow \tilde{\phi}^{k-1} + lr \times \mathrm{Adam}(L)$。
结束
返回 $\tilde{\phi}$。

使用这些新的三元组，我们冻结输入编码器和参考编码器的参数（θ_{E}），然后微调参考引导熵模块和 R 解码器的参数（ϕ），以获得最终的 REMAC 模型。

对于预训练和微调过程，我们使用常用的损失函数来表示码率 R 与失真 D之间的权衡，可写为：

$$\begin{aligned} L = R + \lambda D &= R_y + R_z + R_{\mathrm{ref}} + \lambda D \\ &\approx R_y + R_z + \lambda D, \end{aligned} \tag{7}$$

其中，R_y 和 R_z 分别表示输入特征$\boldsymbol{y}$ 和超先验 $\boldsymbol{z}$的比特开销。R_{ref} 表示选定索引 i 和 j的比特开销，该开销极小，因此在此公式中忽略。失真定义为输入图像 $\boldsymbol{x}$ 与重建图像 $\hat{\boldsymbol{x}}$之间的均方误差（MSE），公式如下：

$$D = \frac{1}{N}\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|_2^2. \tag{8}$$

V. 实验

A. 实验设置

1) 数据集：为了训练所提出的 REMAC 框架，我们采用 MIC 数据集 [7]，，因为它是唯一包含大量彩色无损火星图像的数据集。为实现基于参考图像的压缩，我们首先建立参考图像数据集，并构建用于训练和验证的输入-参考三元组。具体而言，我们将由 3,088 张原始图像组成的训练集随机划分为包含 2702 张图像的训练集和包含 386 张图像的参考集。在训练阶段，我们将新加入的训练图像和参考图像裁剪为不重叠的 512×512 图像块（丢弃不完整的边界区域），并根据计算得到的相似度分数构建输入-参考图像块三元组。在验证阶段，我们采用 MIC 数据集 [7] 的验证集，并使用未经裁剪的参考图像作为参考。

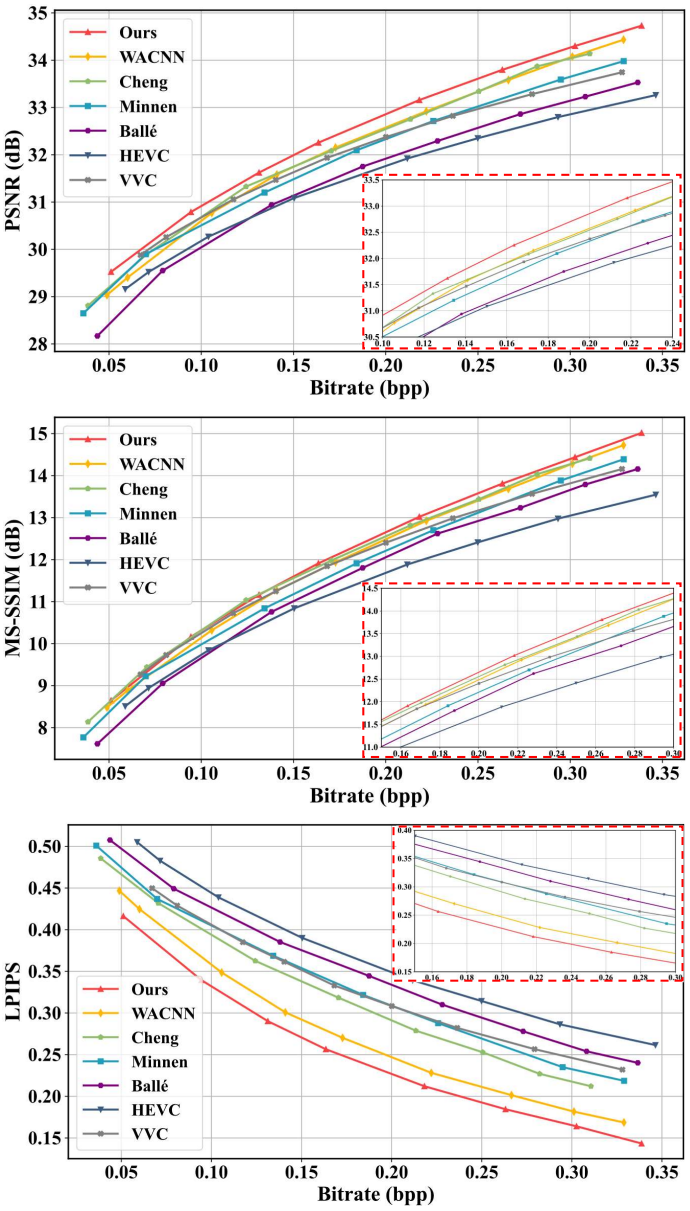


图10：我们的 REMAC 模型与其他压缩方法在峰值信噪比（PSNR）、多尺度结构相似性（MS-SSIM）和学习感知图像块相似度（LPIPS）方面的率失真曲线。

2) 训练：我们在 CompressAI 平台上实现了所提出的 REMAC 框架 [41]。为节省编码器端的计算资源，我们将潜在表示 $\boldsymbol{y}$ 和超潜在表示 $\boldsymbol{z}$ 的通道数分别设为 192 和 128。在所提出的参考引导熵模块中，我们设置 $M = 6$，这表示潜在表示 $\boldsymbol{y}$ 被划分为 6 个切片，以进行逐切片熵编码。同时，我们将公式 7 中的 lambda 值 λ 设为 {0.001、0.002、0.003、0.004、0.006、0.008、0.010、0.012}，用于训练多个模型，分别对应不同的每像素比特数（bpp）点，

表 II：率失真复杂度性能比较。

方法	编码 FLOPs (G) ↓	率失真性能					
		BD-PSNR（dB）↑	BD 码率（%）↓	BD-MSSSIM（dB）↑	BD 码率（%）↓	BD-LPIPS ↓	BD 码率（%）↓
VVC [6]	-	0.5589	-20.8681	0.6297	-19.1038	-0.0392	-24.3304
Ball´e [10]	151.33	0.0367	-1.7809	0.1868	-5.6577	-0.0194	-13.6237
Minnen [40]	362.77	0.5010	-18.3098	0.4627	-13.7433	-0.0430	-28.3645
Cheng [8]	699.55	0.6830	-23.3475	0.7615	-22.2249	-0.0555	-33.7281
WACNN [9]	1297.64	0.6560	-21.1571	0.6546	-18.3739	-0.0935	-47.6364
Ours	395.16	0.9494	-30.4268	0.8198	-22.2335	-0.1163	-55.9240

* 浮点运算次数是在压缩 1600×1152 张图像时测得的。此外，我们的测量也包括相似度计算的浮点运算次数。 **红色颜色表示最佳结果，蓝色颜色表示次佳结果。

分别对应不同的每像素比特数（bpp）点。在训练阶段，我们采用 Adam [42]优化器，并将初始学习率设为 0.0001。此外，采用*ReduceLROnPlateau*算法，在检测到验证损失趋于平稳时自动调整学习率。

3）评估指标：为全面评估重建图像的质量，我们在以下实验中采用两种客观质量指标和一种感知质量指标。首先，我们使用峰值信噪比（PSNR），这是一种压缩方法中广泛使用的指标，数值越高表示像素级保真度越好。然后，我们采用 MS-SSIM [43]，数值越高表示结构保真度越好。此外，为了放大差异，我们使用$-10 \times \log_{10}(1-x)$将 MS-SSIM 数值表示为分贝尺度。最后，我们采用 LPIPS [44], ，这是一种广泛使用的感知质量指标，数值越低表示感知质量越好。为进一步评估压缩性能，我们报告 Bjøntegaard 增量峰值信噪比（BD-PSNR） [45], 、BD-MSSSIM、BD-LPIPS 和 BD 码率，以评估率失真性能。BD 码率具体衡量在保持图像质量的同时实现的比特数减少，该质量使用 PSNR、MS-SSIM [43],和 LPIPS [44]进行评估。下文报告的所有 BD 指标均以 HEVC [5]作为锚点计算。

B. 与最先进方法的比较

在本节中，我们对传统方法和学习式图像压缩方法进行比较实验。传统方法包括通常使用的传统编解码器高效视频编码 [5] 和 VVC [6], 。我们将 RGB 图像转换到 YCbCr 色彩空间，采用 4:4:4 采样比，并使用高效视频编码和 VVC 的帧内编码配置对转换后的图像进行压缩。学习式图像压缩方法包括 Ball´e [10], Min- nen [40], Cheng [8], 和 WACNN [9] 压缩模型。为进行公平比较，我们在 MIC [7] 数据集上重新训练参与比较的模型，该数据集与我们使用的数据集相同。

率失真复杂度性能。对于火星上的实际应用，我们可能需要平衡编码器的码率、失真和计算成本。我们将这种平衡称为率失真复杂度性能。为了全面评估性能，我们采用 BD-PSNR[45],BD-MSSSIM 和 BD-LPIPS 及其相应的 BD 码率来衡量率失真性能。同时，编码器的计算复杂度通过浮点运算次数（FLOPs）进行评估。我们在表II中展示了与最先进的压缩方法的比较。由于 VVC [6] 是一种传统编解码器，因此难以获取其编码器复杂度。因此，表中未列出其编码器复杂度。如表所示，所提出的 REMAC 在峰值信噪比、多尺度结构相似性 [43],和学习感知图像块相似度 [44] 等指标上实现了更优的压缩性能，同时编码器复杂度排名第三低。

与实现最低编码复杂度的 Ball´e [10] 相比，我们的 REMAC 框架在 BD-PSNR、BD-MSSSIM 和 BD-LPIPS 指标上分别取得了显著提升，提升幅度为 +0.9127 dB、 +0.6330 dB 和 -0.0969。对于在 BD-PSNR 和 BD-MSSSIM 指标上取得次优性能的 Cheng [8]方法，我们的 REMAC 框架的编码器复杂度仅为该方法的 56.49%。与在 BD-LPIPS 指标上取得次优性能的 WACNN [9] 方法相比，我们的 REMAC 框架的编码器复杂度仅为该方法的 30.45%。因此，通过平衡码率、失真和编码器复杂度，我们的 REMAC 框架实现了最佳的率失真复杂度性能。

率失真性能。为直观比较不同方法的率失真性能，我们在 MIC 数据集的测试集上绘制了率失真曲线 [7], ，如图10所示。由图可见，大多数学习式图像压缩方法的率失真性能优于传统的高效视频编码 [5] 编解码器。这一现象可归因于火星图像固有的图像间相似度。这种相似度使擅长通过卷积神经网络挖掘有用且共有特征的学习式图像压缩方法，在率失真性能上优于传统编解码器。此外，这些率失真曲线表明，在采用峰值信噪比、多尺度结构相似性和学习感知图像块相似度指标进行评估时，我们的 REMAC 框架在多种比特率设置下总体优于最先进方法。

视觉质量。为了直观地比较重建质量，我们比较了由我们的 REMAC 以及基线方法 Cheng [8] 和 WACNN [9] 在图 11中生成的重建示例。如可视化结果所示，与其他方法相比，我们的方法取得了更优的峰值信噪比性能和更低的每像素比特数。此外，我们的 REMAC 保留了更多纹理细节。从图 11可以看出，在第一行图像中标记的红色区域内，我们的 REMAC 保留了石头的更多细节。在第二行图像中，我们的 REMAC 正确重建了地面上的三个土堆



(a) 真实值 (b) 我们的方法 (c) Cheng (d) WACNN

图11：重建图像的视觉比较。

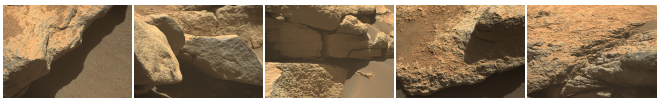


图 12：MIC 测试集中的全部五幅多阴影图像。

真值，而 Cheng [8]和 WACNN [9] 都只能恢复出两个。上述结果表明，我们的框架实现了更优的视觉质量。

多阴影图像上的率失真性能。为评估我们的方法在阴影广布的复杂光照条件下的表现，我们在多阴影图像上进行了实验。具体而言，当阴影覆盖面积超过图像总面积的 30% （ >30% ）时，我们将该图像定义为多阴影图像。按照这一标准，我们在 MIC 数据集 [7]的测试集中找到了五幅自然形成的多阴影图像。这些图像均展现了由地形结构造成的真实火星阴影模式。图 12 展示了评估中使用的全部五幅多阴影图像，它们确实都包含大面积阴影。

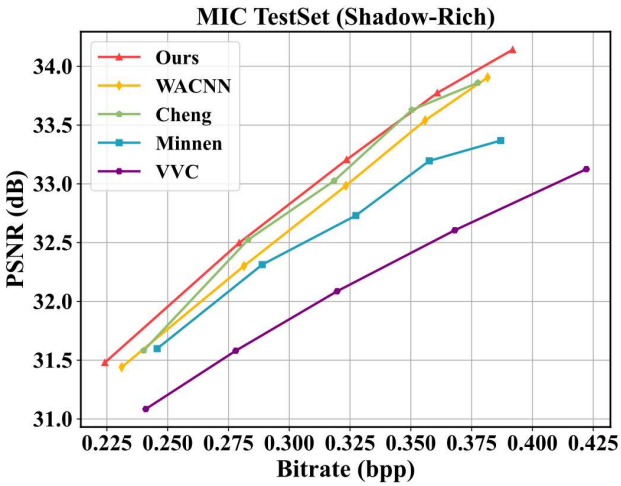


图 13：我们的 REMAC 模型与其他压缩方法在 MIC 测试集多阴影图像上的率失真曲线。

我们使用 REMAC 以及若干传统压缩方法和学习式图像压缩方法对这些图像进行压缩。率失真曲线见图 13。如图所示，REMAC 在所有码率下均持续优于 VVC [6]。此外，REMAC 的压缩性能优于 Cheng [8]，同时显著降低了编码复杂度（395.16 GFLOPs、7.78 秒，而 Cheng 为 699.55 GFLOPs、20.97 秒），因此更适合在火星任务严格的计算资源限制下进行机载部署。这些结果证实，REMAC 在火星多阴影图像压缩方面既有效又实用。

时延开销。为全面比较我们的 REMAC 与若干性能强劲的压缩基线模型之间的推理时延，我们测量并报告了编码和解码在图形处理器和中央处理器上的时延，见表 III。所有实验均在同一台固定设备上进行，该设备配备 AMD EPYC 7642 中央处理器和 GeForce RTX 3090 图形处理器。需要注意的是，由于传统编解码器 VVC 不支持 GPU 加速，因此仅在中央处理器上对其进行评估。所报告的时延为对 50 张图像取平均值的结果，图像分辨率为 [6] 。 1600×1152。

鉴于火星到地球部署场景中，编码器在火星上运行，面临严格的计算约束，而解码器在地球上运行，计算资源充足，因此我们将编码时延作为关键效率指标。如图所示，REMAC 在中央处理器和图形处理器上的编码时延均最低，展现了其在火星上进行编码的效率。至于解码，REMAC 的

得益于并行加速，其在图形处理器上的解码时间最短，但在中央处理器上的解码时延较高。这是非对称设计带来的预期权衡：将计算复杂度从编码器（位于火星）转移到解码器（位于地球）。地球上的计算资源充足，因此解码时延不会构成实际瓶颈。

内存成本。鉴于火星车的编码器资源有限，我们评估了 REMAC 以及三种压缩基线方法的编码器内存开销：广泛使用的传统编解码器 VVC [6]，以及在 MIC 数据集 [7]测试集上取得顶尖性能的三种代表性学习式图像压缩方法，包括 Minnen [40], 、Cheng [8]和 WACNN [9]。由于当前火星车仅配备基于中央处理器的系统，例如美国国家航空航天局的*Perseverance* 火星车，因此，为了进行有意义的比较，所有算法均在 Python 环境下使用仅配备中央处理器的硬件进行评估。我们从 MIC 数据集的测试集中随机选取 50 幅图像进行压缩，并报告编码过程中的峰值内存消耗。为进行更全面的比较，我们还比较了压缩性能（以 BD-PSNR 衡量）和编码时延。内存成本和编码器延迟主要取决于输入分辨率和模型架构，因此，对 50 幅图像取平均值能够提供可靠的测量结果。相比之下，BD-PSNR 在完整测试集上计算，以与上述比较中先前报告的结果保持一致。

表中列出了评估所得的内存成本、BD-PSNR 和编码器延迟。IV。根据美国国家航空航天局的火星*2020*任务[1], 毅力号火星车配备了 BAE RAD 750 处理器和 256 MB 动态随机存取存储器。为适应 256 MB 的内存限制，我们采用基于块的压缩策略，将输入图像划分为互不重叠的320×192 块进行压缩。这一策略显著降低了内存使用量。如表所示，基于块的 REMAC 内存成本最低，BD-PSNR 增益排名第二，编码器延迟最短，体现了资源效率与压缩性能之间的良好平衡。重要的是，在这种基于块的设置下，REMAC 占用 207.71 MB 内存，远低于毅力号火星车 256 MB 的动态随机存取存储器限制，证实了其能够部署于当前的火星探测硬件上。

对未见火星数据的泛化能力。为评估对未见火星数据的泛化能力，我们基于来自美国国家航空航天局火星*2020*任务[1] 的高分辨率图像构建了一个新数据集。该数据集包含由毅力号 火星车搭载的 Mastcam-Z 相机拍摄的 1,360 张原始彩色图像。所有图像均裁剪为 1600×1152 像素，并按 8:1:1 的比例划分为训练集、验证集和测试集。新数据集与 MIC 数据集的训练集之间的 Fr´echet Inception Distance（FID） [46] 为 47.49（两者均排除各自留出的参考图像），这证实了新数据集与 MIC 的分布不同，因此符合未见火星数据的定义。我们在该未见数据集上微调了 REMAC 和四种具有代表性的学习式图像压缩方法，包括 Ball´é [10],Minnen [40],Cheng [8], 和 WACNN [9], 。此外，我们还与传统压缩编解码器进行了比较，包括高效视频编码 [5] 和 VVC [6]。测试集上的率失真曲线见图14。如图所示，REMAC 在所有每像素比特数点上均持续取得更优的压缩性能，展现出对未见火星数据的强泛化能力。

[1]整个项目的详细信息请见 https://mars.nasa.gov/mars2020/。

表 III：编码和解码平均延迟的比较。

方法	图形处理器延迟 (s)↓		中央处理器延迟 (s)↓	
	编码	解码	编码	解码
VVC [6]	–	–	177.78	0.32
Minnen [40]	15.54	30.63	23.43	41.39
Cheng [8]	15.37	30.41	20.97	40.96
WACNN [9]	0.33	0.32	20.58	22.57
Ours	0.17	0.20	7.78	50.67

* 红色 表示最佳结果，蓝色 表示第二佳结果。由于部署场景为火星到地球的传输，优先考虑**编码时延**。

表 IV：编码器内存开销、BD-PSNR 和编码器延迟的比较。

方法	内存（MB）↓		
VVC [6]	493.72	0.5589	177.78
Minnen [40]	1106.84	0.5010	23.43
Cheng [8]	961.75	0.6830	20.97
WACNN [9]	1534.39	0.6560	20.58
Ours	908.52	0.9494	7.78
我们的方法（blk）	207.71	0.8170	6.50

* 红色颜色表示最佳结果，蓝色颜色表示次佳结果。**Ours (blk) 表示我们方法的基于块的变体。

C. 消融研究

在本节中，我们开展消融研究，以验证所提出的参考超先验和参考解码器的有效性。此外，我们还开展了消融实验，以探究所提出的潜在特征循环利用机制的有效性。

参考超先验的有效性。参考超先验旨在利用参考特征 $\boldsymbol{y}_{\text{ref}}^{i}$ 来获取有助于改善输入特征熵估计的有效信息$\boldsymbol{y}$。为验证其有效性，我们首先将所提出的输入编码器与通道级熵自回归模型相结合，构建基线模型 [19]。随后，我们将所提出的参考超先验纳入基线模型。鉴于参考超先验需要以参考特征作为输入，我们进一步将所提出的参考编码器加入基线模型，从而得到“基线 +参考超先验”模型。我们对它们进行训练，相应的率失真曲线如图15所示。从该图可以看出，“基线 +参考超先验”模型在较高的每像素比特数点上取得了明显的峰值信噪比增益。这验证了我们的参考超先验的有效性。此外，基线模型的 +0BD-PSNR 为 0.9558 dB，而“基线 + 参考超先验”模型

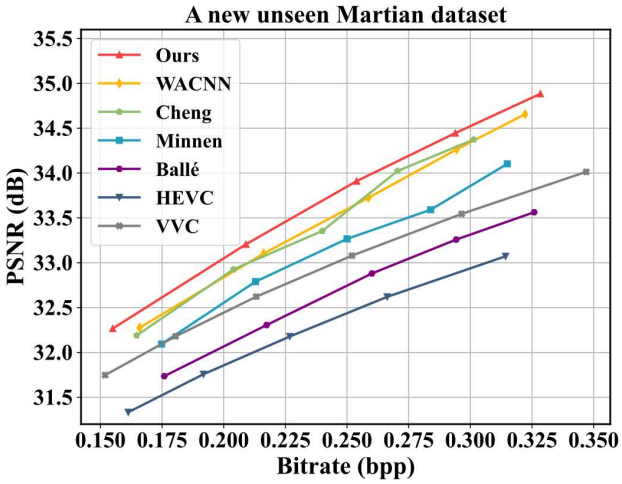


图14：我们的 REMAC 模型与其他压缩方法在一个全新的火星数据集上的率失真曲线。

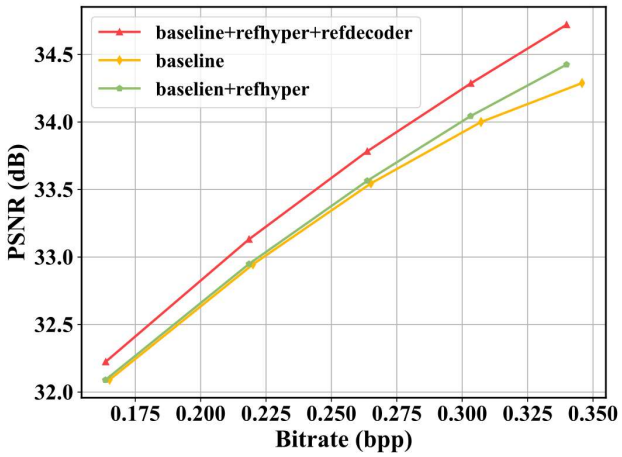


图15：基线模型、“基线模型 + 参考超先验”和“基线模型 + 参考超先验 + 参考解码器 ” 的峰值信噪比率失真曲线。

达到 +1.0005 dB。这些结果进一步表明了参考超先验的有效性。

参考解码器的有效性。提出参考解码器是为了利用参考图像 $\boldsymbol{x}^{j}$ 参考，在输出图像重建过程中提供有价值的信息$\hat{\boldsymbol{x}}$。为验证其有效性，我们将所提出的参考解码器集成到带有参考超先验的基线模型中（即“基线模型 + 参考超先验”），构成“基线模型 + 参考超先验 + 参考解码器”模型。我们对该模型进行训练，并在图15中展示其率失真曲线。如图所示，在所有每像素比特数点上，“基线模型 + 参考超先验 + 参考解码器”模型的峰值信噪比均优于“基线模型 +参考超先验”模型。这一提升量化了重建图像在像素级保真度上的增强，从而验证了所提出的参考解码器的有效性。此外，我们计算了其相对于高效视频编码的BD-PSNR。“基线模型 + 参考超先验+ 参考解码器”模型的BD-PSNR提升了 +1.2016 dB。此外，“基线模型 +参考超先验 +参考解码器”模型的峰值信噪比比 +0 高出.2011 dB

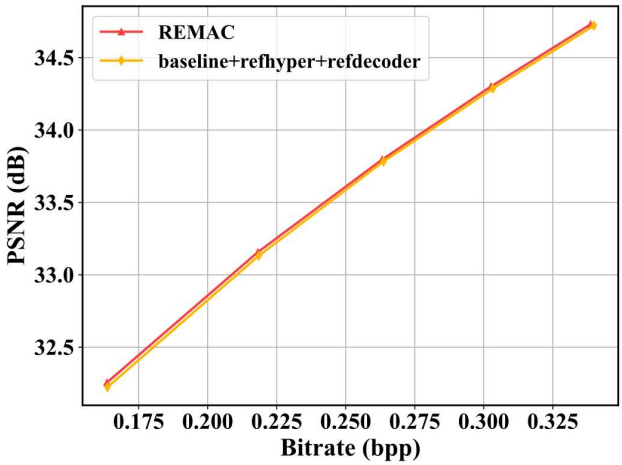


图16：“基线模型 + 参考超先验 +参考解码器”与 REMAC 的峰值信噪比率失真曲线。取得这些结果时，浮点运算次数有所减少。

“基线模型 + 参考超先验”模型。这些结果进一步表明参考解码器的有效性。

潜在特征循环利用机制的有效性。所提出的潜在特征循环利用机制旨在显著降低编码器复杂度，同时不大幅损害压缩性能。为验证其有效性，我们将该机制应用于“基线模型+ 参考超先验 + 参考解码器”模型，从而得到最终的 REMAC 模型。两个模型的率失真曲线如图16所示。从图中可以看出，与“基线模型 + 参考超先验 +参考解码器”模型相比，REMAC 模型在所有码率点上均取得了更高的峰值信噪比值。此外，“基线模型 + 参考超先验 +参考解码器”模型的编码浮点运算次数包括：对输入图像执行预编码器（共享骨干网络）的计算、输入编码器的计算，以及参考引导熵模块中涉及输入图像的组件的计算。相比之下，所有仅处理参考图像的计算均不计入其中，因为假设参考特征已在编码器端预先计算并存储。具体而言，不计入的计算包括对参考图像应用预编码器、参考编码器，以及在参考引导熵模块中生成参考专属参数(μ'', σ'') 的计算。对于 REMAC 模型的编码器复杂度，所提出的潜在特征循环利用机制会重用输入编码器生成的深层潜在特征，以进行深层参考选择，从而免去生成深层输入特征所需的预编码器计算，进一步降低编码器复杂度。在压缩 1600×1152幅图像时，“基线模型 + 参考超先验 + 参考解码器”模型的编码浮点运算次数为 406.85G，而 REMAC 模型仅需 395.16G。这使编码浮点运算次数减少了 2.87%，因此更适用于由资源受限编码器执行的火星图像压缩。这些结果证实了所提出的潜在特征循环利用机制的有效性。

非对称设计的有效性。非对称设计旨在利用解码器的计算资源，弥补轻量级

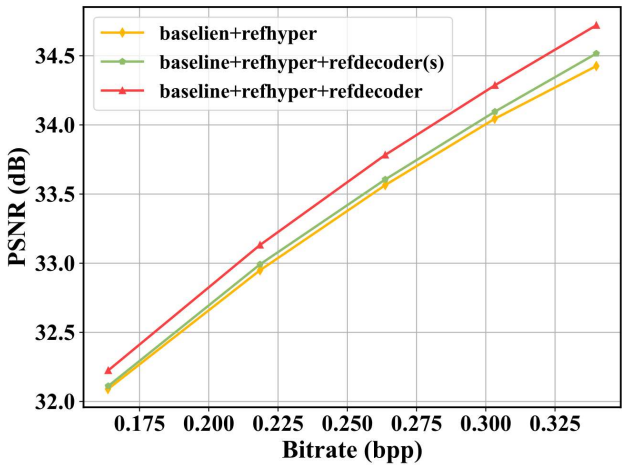


图17：对称的“baseline + ref- hyperprior”、非对称的“baseline + ref-hyperprior +ref- decoder”及其简化变体“baseline + ref-hyperprior+ ref-decoder(s)”模型的峰值信噪比率失真曲线。

编码器的性能限制。为评估其有效性，我们首先使用所提出的参考超先验（记为“基线+ 参考超先验”）扩展基线模型，从而形成对称架构。随后，我们进一步引入所提出的参考解码器，构成完整的非对称模型，记为“基线+ 参考超先验 +参考解码器”。为进一步评估解码器复杂度的影响，我们还引入了一种简化的非对称变体“基线 + 参考超先验 + 参考解码器”，通过移除若干卷积层和残差块来降低解码器深度，同时保留其多尺度结构。这些模型的率失真曲线如图 17所示。如图所示，在所有每像素比特数点上，就峰值信噪比而言，完整和简化的非对称模型均始终优于对称基线。为进行定量比较，我们以高效视频编码为锚点计算 BD-PSNR。对称基线取得了 +1.0005 dB 的 BD-PSNR 增益，而完整和简化的非对称模型分别取得了 +1.2016 dB 和 +1.0459 dB 的增益。两个非对称模型取得的 BD-PSNR 增益均高于对称基线，证明了非对称设计的有效性。此外，简化非对称模型与完整非对称模型之间的性能差距凸显了更复杂解码器的优势，表明增加解码器容量能够提升重建质量。

基于参考图像的机制的有效性。火星图像压缩中的基于参考图像的机制，源于火星图像固有的高图像间相似度。为评估该机制的普遍有效性，我们将其集成到四种具有代表性的学习式图像压缩方法中，包括 Ball´e [10], Minnen [40],Cheng [8],和 WACNN[9] 压缩模型。对于每个基线模型，我们引入了与所提出的参考引导熵模块和参考解码器相似的组件，并针对各网络架构进行了必要调整。每项调整都经过精心设计，以保留原始架构的完整性，例如保留 WACNN 中的 5×5 卷积。

原始模型及其参考增强变体的率失真曲线如图18所示。在所有每像素比特数点上，每种参考增强变体均持续获得峰值信噪比提升，这证实，显式建模图像间相似度是一种稳健且不依赖于特定架构的策略，可提升火星图像的压缩效率。

参考选择的鲁棒性。为评估我们基于特征级相似性的参考图像选择方法的鲁棒性，我们进行了全面实验，以评估其应对场景变化、传感器噪声和压缩伪影的能力。

1）对场景变化的鲁棒性：为评估跨场景鲁棒性，我们将 MIC 数据集 [7]测试集划分为四种场景类别：天空、沙地、土壤和岩石。对于每种类别，我们比较两种参考选择策略的压缩性能：（i）我们提出的基于特征级相似度的选择；（ii）随机选择。为确保统计可靠性，我们独立重复随机参考选择10次，每次构建新的输入-参考三元组，并报告平均值。如图所示，19在所有场景类型下，就峰值信噪比而言，我们提出的基于特征级相似度的选择始终比随机选择取得更好的率失真性能，证明了其对场景变化的鲁棒性。

2) 对传感器噪声的鲁棒性：为验证对传感器噪声的鲁棒性，我们添加高斯噪声，噪声水平为 $\sigma \in \{5, 10, 15\}$，以此对 MIC 数据集 [7]测试集施加不同程度的传感器退化。采用双参考选择策略时，一个参考基于浅层纹理特征（浅层选择）选取，另一个则基于深层语义特征（深层选择）选取。我们报告各选择方法相对于无噪声基线模型的准确率，以评估其鲁棒性。结果见表 V。如表所示，在所有噪声水平下，浅层选择的准确率均保持在 95.85% 以上，而深层选择的平均准确率保持在 89.70% 以上，表明其具有很强的抗传感器噪声能力。这说明浅层选择和深层选择均对传感器噪声具有鲁棒性。

3) 对压缩伪影的鲁棒性：为验证对压缩伪影的鲁棒性，我们采用广泛使用的编解码器 JPEG [47] 作为代理，以模拟常见的压缩伪影。具体而言，我们使用质量因子 $QF \in \{50, 75, 90\}$ 对 MIC 数据集[7] 测试集进行 JPEG 压缩。我们评估浅层和深层参考选择相对于无伪影基线的准确率。结果见表VI。如表所示，即使在最低质量因子（$QF = 50$）下，浅层选择的准确率仍保持在 97.93% 以上，而深层选择的平均准确率则超过 97.74%。这种很强的鲁棒性表明，在严重的块状失真下，浅层纹理特征和深层语义特征仍保有足够的判别能力。性能下降幅度极小，证实了我们的双参考选择策略对压缩伪影具有鲁棒性。

对输入-参考失配的泛化。为评估REMAC对输入-参考失配的泛化能力，我们采用以下三种参考选择策略开展了额外实验，以模拟输入图像与参考图像之间逐渐增大的差异：(1) 基于特征层相似度的策略，使用从MIC训练集中划分出的

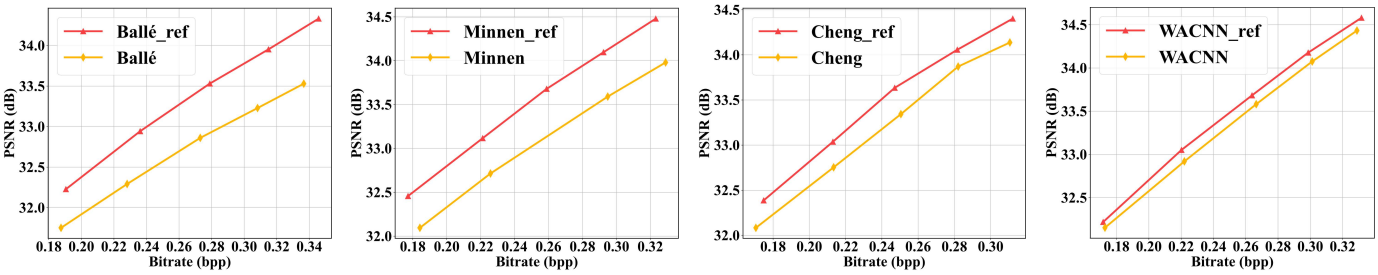


图18：现有图像压缩方法及其对应的参考增强变体在峰值信噪比方面的率失真曲线。

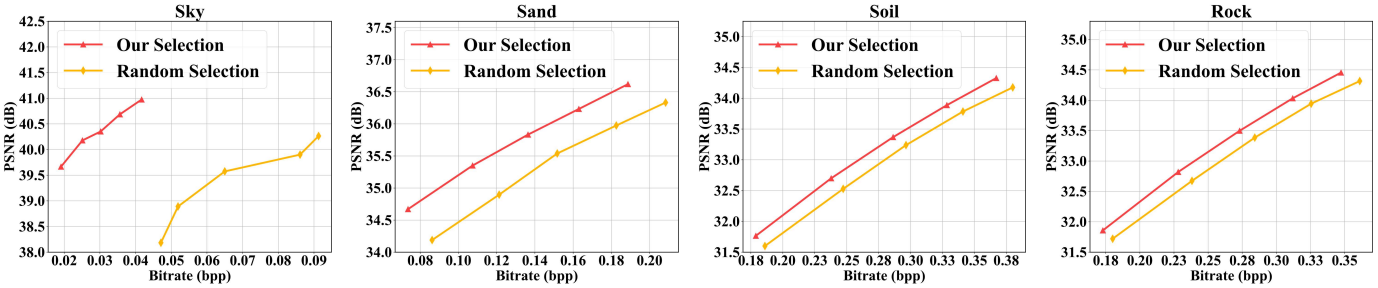


图19：我们提出的基于特征级相似度的选择方法与随机选择方法在天空、沙地、土壤和岩石四个类别上的峰值信噪比率失真曲线。

TAB表V：不同高斯噪声级别下浅层参考选择和深层参考选择的准确率 ls.

噪声水平	浅层选择（%）	深层选择（%）				
	$\lambda \in \{0.012, 0.010, 0.008, 006\}$	$\lambda = 0.012$	$\lambda = 0.010$	$\lambda = 0.008$	$\lambda = 0.006$	平均值
$\sigma = 5$	96.89	97.15	95.34	93.01	93.78	94.82
$\sigma = 10$	95.60	94.04	91.45	91.45	91.97	92.23
$\sigma = 15$	95.85	90.93	88.60	89.12	90.16	89.70

* In 浅层选择中，采用预训练特征进行相似度计算，因此不受 REMAC 的 λ 影响。
** 在 深层选择中，复用 R-编码器的特征进行相似度计算，因此受 REMAC 的 λ 影响。

TAB表 VI：不同 JPEG 质量因子下浅层和深层参考选择的准确率 rs.

噪声水平	浅层选择（%）	深层选择（%）				
	$\lambda \in \{0.012, 0.010, 0.008, 006\}$	$\lambda = 0.012$	$\lambda = 0.010$	$\lambda = 0.008$	$\lambda = 0.006$	平均值
QF = 50	97.93	97.67	98.19	97.15	97.93	97.74
QF = 75	98.70	98.70	97.41	98.45	98.70	98.32
质量因子 = 90	98.96	99.48	99.22	99.22	99.48	99.35

* 在浅层选择中，采用预训练特征进行相似度计算，因此不受 REMAC 的 λ 影响。** 在深层选择中，重复使用 R-编码器的特征进行相似度计算，因此受 REMAC 的 λ 影响。

留出参考集 [7]（轻度差异）；(2) 从同一参考集中随机选择（中度差异）；(3) 从由美国国家航空航天局火星2020任务的未见图像组成的分布外参考集中随机选择 *NASA's Mars 2020Mission*[1] （显著差异）。这代表了一种极具挑战性且符合现实的失配情形。这个新构建的火星参考集与用于训练REMAC的实际训练集（即排除留出参考图像后的MIC训练集）之间的FID [46] 为62.54，证实存在显著的分布偏移。

相比之下，实际训练集与留出参考集之间的 FID 为 33.98，显著低于 62.54，表明存在较小但不可忽视的分布差异，这种差异构成了轻度和中度失配情形的基础。这些设置使我们能够系统地分析 REMAC 在轻度、中度和显著输入-参考失配情景下的泛化能力。

具体而言，为了构建一个分布外参考集，我们从美国国家航空航天局的火星*2020*任务[1],中收集了386张未见过的火星图像,其尺寸与我们原始参考集

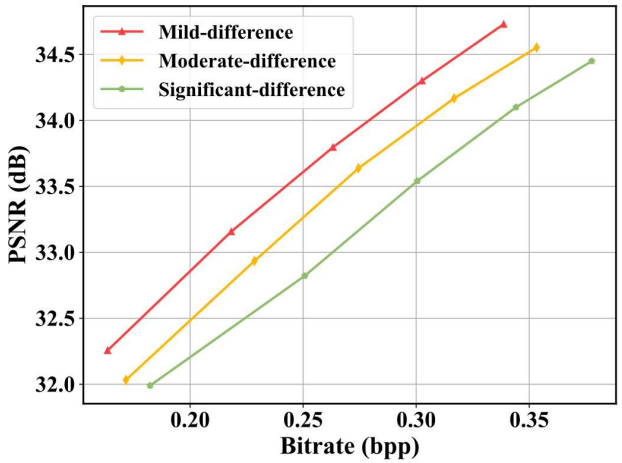


图20：不同程度的输入-参考失配对应的率失真曲线。

相同。为确保统计可靠性，我们独立重复随机参考选择过程10 次，每次都构建新的输入-参考图像，并在图20中报告平均压缩结果。从该图可以看出，在峰值信噪比方面，轻微差异情形（采用我们基于特征级相似度的选择方法）在所有每像素比特数点上始终取得最佳率失真性能。同时，随着输入-参考失配从轻微到显著，率失真性能逐渐下降。峰值信噪比的下降幅度很小，这表明 REMAC 能够始终重建出与原图高度相似的图像，并在输入-参考失配情景下具备泛化能力。

VI. 局限性与未来工作

尽管与现有压缩方法相比，REMAC 实现了先进的率失真性能，但其当前的直接实现需要大量内存和计算资源，这主要是由于使用了全精度（32位浮点）权重，以及深度学习方法固有的高强度推理过程。这一局限凸显了提升模型效率的必要性。未来，我们计划探索轻量化策略，例如权重量化和模型剪枝，以降低内存和计算成本，同时保持压缩性能。

VII. 结论

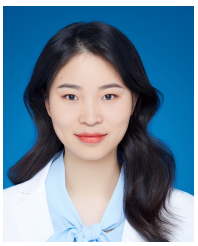
本文提出了 REMAC 方法，用于实现高效的火星图像压缩。通过对 784幅火星图像进行经验分析，我们发现它们在颜色、纹理和语义方面具有很强的图像内和图像间相似性。因此，我们设计了 REMAC，以利用这些相似性提升压缩性能并降低编码器复杂度。借助编码器和解码器两端均可获得的参考图像，REMAC 将计算复杂度从资源受限的编码器转移至资源充足的解码器。为利用图像间相似性，我们提出了参考引导熵模块和参考解码器，通过利用有价值的参考信息，同时增强熵估计和重建质量。为利用图像内相似性，参考解码器采用具有大感受野大小的深度、多尺度架构，以建模单幅图像内的长程空间依赖。此外，我们开发了潜在特征循环利用机制，以重复使用预提取特征，从而减少参考图像选择过程中的特定计算开销。最后，我们开展了大量实验，证明了 REMAC 方法在码率-失真-复杂度性能方面的优越性。

参考文献

[1] S. Liu, Y. Lin, K. Du, J. Zhang, X. Tong, H. Xie, B. Wan, and Z. Zhang, “一种用于分析祝融号火星车获取的多光谱相机图像的新型原位尘埃覆盖指数”，*IEEE*地球科学与遥感汇刊，卷 63，第 1–13 页，2025年。[2] T. Panambur and M. Parente, “利用深度约束聚类增强火星地形识别”，*IEEE*地球科学与遥感汇刊，第 1–1 页，2025年。[3] X. Tan, B. Xi, H. Xu, J. Li, Y. Li, C. Xue, and J. Chanussot, “一种用于零样本火星场景分类的知识蒸馏轻量级框架”，*IEEE*地球科学与遥感汇刊，卷 62，第 1–16 页，2024 年。[4] J.-H. Kim, J.-H. Choi, J. Chang, and J.-S. Lee, “通过非对称自编码器和剪枝实现高效深度学习有损图像压缩”，载于*2020*年 *IEEE* 国际声学、语音与信号处理会议（*ICASSP*）。IEEE，2020年，第 2063–2067 页。[5] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, “高效视频编码（HEVC）标准概述”，*IEEE* 视频技术电路与系统汇刊，卷 22，第 12 期，第 1649–1668 页，2012年。[6] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, and J.-R. Ohm, “多功能视频编码（VVC）标准概述及其应用”，*IEEE* 视频技术电路与系统汇刊，卷 31，第 10 期，第 3736–3764 页，2021年。[7] Q. Ding, M. Xu, S. Li, X. Deng, Q. Shen, and X. Zou, “一种基于学习的火星图像压缩方法”，载于*2022*年 *IEEE* 国际视觉通信与图像处理会议（*VCIP*），2022年，第 1–5 页。[8] Z. Cheng, H. Sun, M. Takeuchi, and J. Katto, “结合离散化高斯混合似然和注意力模块的学习式图像压缩”，载于*2020*年 *IEEE/CVF*计算机视觉与模式识别会议（*CVPR*），2020年，第 7936–7945 页。[9] R. Zou, C. Song, and Z. Zhang, “细节决定成败：用于图像压缩的基于窗口的注意力”，载于*2022*年 *IEEE/CVF*计算机视觉与模式识别会议（*CVPR*），2022年，第 17 471–17 480 页。[10] J. Ball´e, D. Minnen, S. Singh, S. J. Hwang, and N. Johnston, “具有尺度超先验的变分图像压缩”，*arXiv* 预印本 *arXiv:1802.01436*，2018年。[11] J. Ball´e, V. Laparra, and E. P. Simoncelli, “端到端优化图像压缩”，*arXiv* 预印本 *arXiv:1611.01704*，2016年。[12] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly 等，“一张图像胜过16×16 个词：用于大规模图像识别的变换器”，*arXiv* 预印本 *arXiv:2010.11929*，2020年。[13] J. Liu, H. Sun, and J. Katto, “采用混合变换器-CNN架构的学习式图像压缩”，载于*2023*年 *IEEE/CVF*计算机视觉与模式识别会议（*CVPR*），2023年，第 14 388–14 397 页。[14] S. Xiang and Q. Liang, “基于高频和低频分量的遥感图像压缩”，*IEEE*地球科学与遥感汇刊，卷 62，第 1–15 页，2024年。[15] Y. Chong, L. Zhai, and S. Pan, “将高阶马尔可夫随机场用作注意力网络以实现高分辨率遥感图像压缩”，*IEEE*地球科学与遥感汇刊，卷 60，第 1–14 页，2022年。[16] C. Shi, K. Shi, F. Zhu, Z. Zeng, and L. Wang, “用于遥感图像压缩的多层域相似度增强引导网络”，*IEEE*地球科学与遥感汇刊，卷 62，第 1–19 页，2024年。[17] X. Wang, R. Girshick, A. Gupta, and K. He, “非局部神经网络”，载于*2018*年 *IEEE/CVF*计算机视觉与模式识别会议，2018年，第 7794–7803 页。

[18] D. Minnen, J. Ball´e, and G. Toderici, “用于学习式图像压缩的联合自回归先验与层次先验”，CoRR，卷 abs/1809 02736，2018年。。[19] D. Minnen and S. Singh, “用于学习式图像压缩的逐通道自回归熵模型”，载于 2020年 IEEE 国际图像处理会议（ICIP），2020年，第 3339–3343 页。[20] D. He, Y. Zheng, B. Sun, Y. Wang, and H. Qin, “用于高效学习式图像压缩的棋盘格上下文模型”，载于 2021年 IEEE/CVF计算机视觉与模式识别会议（CVPR），2021年，第 14 766–14 775 页。[21] D. He, Z. Yang, W. Peng, R. Ma, H. Qin, and Y. Wang, “Elic：采用非均匀分组空间-通道上下文自适应编码的高效学习式图像压缩”，载于 2022年 IEEE/CVF计算机视觉与模式识别会议（CVPR），2022年，第 5708–5717 页。[22] H. Zheng, M. Ji, H. Wang, Y. Liu, and L. Fang, “CrossNet：一种采用跨尺度变形的端到端基于参考图像的超分辨率网络”，载于《欧洲计算机视觉会议论文集》（ECCV），2018年，第 88–104 页。[23] Z. Zhang, Z. Wang, Z. Lin, and H. Qi, “通过神经纹理迁移实现图像超分辨率”，载于 2019年 IEEE/CVF计算机视觉与模式识别会议（CVPR），2019年，第 7974–7983 页。[24] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “注意力就是一切”，载于《神经信息处理系统进展》，2017年，第 5998–6008 页。[25] F. Yang, H. Yang, J. Fu, H. Lu, and B. Guo, “用于图像超分辨率的纹理变换器网络学习”，载于 2020年 IEEE/CVF计算机视觉与模式识别会议（CVPR），2020年，第 5790–5799 页。[26] L. Lu, W. Li, X. Tao, J. Lu, and J. Jia, “Masa-sr：面向基于参考的图像超分辨率的匹配加速与空间自适应”，载于 2021年 IEEE/CVF计算机视觉与模式识别会议（CVPR），2021年，第 6364–6373 页。[27] M. Li, L. Shen, P. Ye, G. Feng, and Z. Wang, “Rfd-ecnet：参考特征字典的极端水下图像压缩”，载于 IEEE/CVF 国际计算机视觉会议论文集，2023年，第 12 980–12 989 页。[28] M. Li, L. Shen, X. Hua, and Z. Tian, “Euicn：一种高效水下图像压缩网络”，IEEE 视频技术电路与系统汇刊，卷 34，第 7 期，第 6474–6488 页，2024年。[29] A. Luo, H. Sun, J. Liu, and J. Katto, “采用剪枝超先验模块的内存高效学习式图像压缩”，载于 2022年 IEEE 国际图像处理会议（ICIP）。IEEE，2022年，第 3061–3065 页。[30] A. Luo, H. Sun, J. Liu, F. Lin, and J. Katto, “Pts-lic：用于轻量级学习式图像压缩的剪枝阈值搜索”，载于 2023年 IEEE 国际视觉通信与图像处理会议（VCIP）。IEEE，2023年，第 1–5 页。[31] Y. Fang, W. Fei, S. Li, W. Dai, C. Li, J. Zou, and H. Xiong, “全整数端到端学习式图像压缩”，载于 2023年数据压缩会议（DCC）。IEEE，2023年，第 337–337 页。[32] Z. He, L. Luo, L. Zhang, H. Guo, and C. Zhu, “基于轻量级注意力的高效学习式图像压缩”，载于 2023年 IEEE 国际视觉通信与图像处理会议（VCIP）。IEEE，2023年，第 1–5 页。[33] H. Fu, F. Liang, J. Liang, B. Li, G. Zhang, and J. Han, “结合多尺度残差块、重要性缩放和后量化滤波的非对称学习式图像压缩”，IEEE 视频技术电路与系统汇刊，卷 33，第 8 期，第 4309–4321 页，2023年。[34] S. Wang, Z. Cheng, D. Feng, G. Lu, L. Song, and W. Zhang, “Asymllic：非对称轻量级学习式图像压缩”，载于 2024年 IEEE 国际视觉通信与图像处理会议（VCIP）。IEEE，2024年，第 1–5 页。[35] Y. Bao, W. Tan, C. Jia, M. Li, Y. Liang, and Y. Tian, “Shiftlic：采用空间-通道移位操作的轻量级学习式图像压缩”，IEEE 视频技术电路与系统汇刊，2025年。[36] E. Agustsson and R. Timofte, “NTIRE 2017 单幅图像超分辨率挑战赛：数据集与研究”，载于 2017 年 IEEE 计算机视觉与模式识别会议研讨会（ CVPRW ），2017 年，第 1122–1131 页。[37] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “通过稀疏三维变换域协同滤波实现图像去噪”，IEEE 图像处理汇刊，卷 16，第 8 期，第 2080–2095 页，2007年。[38] R. M. Haralick, K. Shanmugam, and I. Dinstein, “用于图像分类的纹理特征”，IEEE 系统、人类与控制论汇刊，卷 SMC-3，第 6 期，第 610–621 页，1973年。

[39] K. Simonyan 和 A. Zisserman，《用于大规模图像识别的超深卷积网络》，计算机科学，2014年。[40] D. Minnen、J. Ball´e 和 G. D. Toderici，《用于学习式图像压缩的联合自回归与分层先验》，神经信息处理系统进展，第31卷，2018年。[41] J. B´egaint、F. Racap´e、S. Feltman 和 A. Pushparaja，《CompressAI：用于端到端压缩研究的 PyTorch 库与评估平台》，2020年11月。[42] D. P. Kingma 和 J. Ba，《Adam：一种随机优化方法》，*arXiv* 预印本 *arXiv:1412.6980*，2014年。[43] Z. Wang、E. Simoncelli 和 A. Bovik，《用于图像质量评估的多尺度结构相似性》，载于第三十七届阿西洛马会议：信号、系统与计算机，*2003*年，第2卷，2003年，第1398–1402页，第2卷。[44] R. Zhang、P. Isola、A. A. Efros、E. Shechtman 和 O. Wang，《深层特征作为感知度量的惊人有效性》，载于*2018*年 *IEEE/CVF* 计算机视觉与模式识别会议，2018年，第586–595页。[45] G. Bjontegaard，《计算率失真曲线之间的平均 PSNR 差异》，国际电信联盟 *SG16* 文件 *VCEG-M33*，2001年。[46] M. Heusel、H. Ramsauer、T. Unterthiner、B. Nessler 和 S. Hochreiter，《通过双时间尺度更新规则训练的生成对抗网络收敛至局部纳什均衡》，神经信息处理系统进展，第30卷，2017年。[47] G. K. Wallace，《JPEG 静态图像压缩标准》，*ACM* 通讯，第34卷，第4期，第30–44页，1991年。



Qing Ding 于2018年和2021年分别获得中国上海上海大学通信与信息工程学院的学士学位和硕士学位。她目前正在中国北京北京航空航天大学电子与信息工程学院攻读博士学位。她的研究兴趣包括视频编码、深度学习、视频增强和学习型图像压缩。



Mai Xu （IEEE高级会员）于2003年获得中国北京北京航空航天大学的学士学位，2006年获得北京清华大学的硕士学位，并于2010年获得英国伦敦帝国学院的博士学位。2010年至2012年，他在清华大学电气工程系担任研究员。自2013年以来，他一直任职于北京航空航天大学，并于2019年由副教授晋升为正教授。2014年至2015年，他曾担任微软亚洲研究院访问研究员。他已发表

或在国际期刊和会议论文集中合著了200多篇技术论文，发表刊物和会议包括《国际计算机视觉期刊》、IEEE模式分析与机器智能汇刊、IEEE 图像处理汇刊、IEEE信号处理选题期刊、计算机视觉与模式识别会议、国际计算机视觉会议、欧洲计算机视觉会议以及AAAI。他的主要研究兴趣包括图像处理和计算机视觉。他曾获得IEEE/ACM会议的最佳论文奖或顶级论文奖，包括ACM多媒体会议。他曾担任IEEE 图像处理汇刊和IEEE多媒体汇刊的副编辑、IEEE信号处理选题期刊的首席客座编辑，以及多个会议的领域主席或技术程序委员会委员，包括国际多媒体与博览会和AAAI。他是多媒体信号处理技术委员会和IEEE信号处理学会经选举产生的委员。



Shengxi Li （IEEE会员）于2021年在英国伦敦的伦敦帝国学院获得电气与电子工程博士学位。他目前是中国北京北航大学电子与信息工程学院的教授。他的研究兴趣包括生成模型、统计信号处理和机器学习。他曾获得国际神经网络学会的青年研究者奖。



Xin Deng （IEEE会员）于2016年获得北京航空航天大学（中国北京）电气信息工程硕士学位，并于2020年3月获得伦敦帝国学院（英国伦敦）电气与电子工程博士学位。她目前是北京航空航天大学网络空间科学与技术系副教授，工作地点为中国北京。她的研究兴趣包括稀疏编码及其在图像与视频处理、机器学习和多模态信号处理中的应用。



Xin Zou 于2009年获得长春理工大学硕士学位。此后，她一直在北京航天器系统工程研究所工作。她的研究兴趣包括航天器系统设计和图像处理。