

摘要

随着信息技术的不断发展，图像和视频数据呈现爆发式增长，如何高效地处理和传输海量的图像数据成为当前信息技术领域的重要挑战。图像压缩技术应运而生，它通过去除冗余信息有效减小数据量，在存储和传输过程中起到了关键作用。传统图像编码基于香农信息论，通过手工设计变换、量化和熵编码等模块压缩图像，通常能取得不错的压缩效果。然而，在极低码率条件下传统编码方法容易出现严重的块效应、高频细节丢失等问题。近年来深度学习方法在图像压缩中取得了一些进展，提升感知质量的同时降低了像素失真，但这些方法在极低码率场景下仍面临着诸多问题。端到端压缩方法通常会导致图像的过度平滑，缺失高频细节，而基于生成对抗网络的方法尽管能够提高感知质量，但存在训练不稳定和语义不一致的问题。扩散模型因其强大的图像生成能力在图像重建中展现出巨大潜力，但在极低码率压缩任务中，现有工作未能充分利用多模态数据增强，限制了其在细节恢复和语义一致性方面的表现。

针对现有方法在极低码率下存在的局限性，本文提出了两种极低码率图像压缩方法，分别是基于双模引导的扩散式图像压缩方法和多模态编码增强的极低码率图像压缩方法。其中，双重引导扩散式解码框架通过图像的隐空间特征和文本提示引导扩散模型的去噪过程，使得图像重建兼顾细节保真和语义一致性。该方法在编码阶段将图像的潜在特征进行高效压缩，降低码率负担，同时保留关键内容变量以指导解码端的图像重建。在解码阶段结合文本提示和内容变量，引导去噪过程，逐步恢复图像的语义一致性和细节纹理。多模态编码增强框架通过引入图像的文本提示和隐空间特征增强编码器的能力，从而提升图像的重建效果。在编码阶段，采用了交叉注意力机制，将图像特征、文本特征和潜在特征进行多尺度对齐融合。其中，图像特征和文本特征通过交叉注意力机制进行双向融合，而隐空间特征则直接引导图像特征的生成，这种融合策略使得编码器生成的潜在表示在语义和视觉特征上更加丰富，从而在解码阶段实现高质量的图像重建。

实验结果表明，本文提出的方法在极低码率条件下，能够显著提高图像的重建质量，保留更多的图像细节和语义信息，优于传统和现有的深度学习图像压缩方法。此外，本文方法通过充分利用图像和多模态数据的协同作用，进一步推动了图像压缩技术向更高质量、更低码率的方向发展。

关键词：图像压缩，极低码率，扩散模型，多模态融合，端到端压缩